

FOCO

ARTE Berndnaut Smilde, *Nimbus Green Room*, 2013  
Impressão digital cromogênica, 75 x 102 cm/125 x 170 cm.  
Cortesia do artista e da Ronchini Gallery.

# A disciplina da experimentação em negócios



**Stefan Thomke** é professor de administração de negócios da Harvard Business School.

**Jim Manzi** é fundador e presidente da Applied Predictive Technologies, que fornece *softwares* para projetar e analisar experimentos de negócios.

## ***Aumente suas chances de sucesso com test-drives de inovação.*** Stefan Thomke e Jim Manzi

Logo depois que Ron Johnson saiu da Apple para se tornar CEO da J. C. Penney, em 2011, sua equipe implementou um plano ousado que substituiu cupons e prateleiras cheias de produtos em liquidação por lojas com butikues de marca e tecnologias que eliminavam caixas, máquinas registradoras e balcões de atendimento. No entanto, apenas 17 meses depois de Johnson ter ingressado na Penney, as vendas caíam vertiginosamente, as perdas haviam disparado e Johnson foi despedido. A loja de departamentos então promoveu uma mudança radical.

Como a Penney tinha enveredado por caminhos tão errados? Ela não dispunha de toneladas de dados de transações com clientes que indicavam seus gostos e preferências?

Provavelmente sim, mas o problema é que grandes dados fornecem pistas apenas sobre o comportamento passado dos clientes — e não sobre como reagirão a mudanças ousadas. Geralmente, quando se chega à inovação, a maioria dos executivos precisa

trabalhar num ambiente em que os dados são insuficientes para basear suas decisões. Consequentemente, muitas vezes eles apelam para sua experiência ou intuição. Mas as ideias verdadeiramente inovadoras — isto é, que podem reestruturar o setor — geralmente seguem na contramão da experiência de executivos e do senso comum.

Executivos podem, no entanto, descobrir se um novo programa de produto ou de negócio será bem-sucedido submetendo-o a um teste rigoroso. Pense no seguinte exemplo: uma empresa farmacêutica jamais introduziria um medicamento no mercado sem antes realizar uma bateria de experimentos baseados em protocolos científicos estabelecidos. (Na verdade, a Administração de Alimentos e Medicamentos — FDA, na sigla em inglês — dos Estados Unidos exige rigorosos testes clínicos). E é essencialmente isso que muitas empresas fazem ao lançar novos modelos de negócios e novos conceitos. Se a J.C. Penney tivesse testado as mudanças propostas pelo CEO, a empresa provavelmente teria percebido que elas seriam rejeitadas pelos clientes.

Por que mais empresas não realizam testes rigorosos de suas propostas de reestruturação geral, quase sempre arriscadas e dispendiosas? Porque a maioria das organizações reluta em financiar experiências adequadas de negócios e tem grande dificuldade de executá-las. Embora o processo de experimentação pareça simples, é surpreendentemente difícil pô-lo em prática, devido a uma infinidade de desafios técnicos e organizacionais. Esta é a principal conclusão de nossa experiência conjunta de mais de 40 anos conduzindo e estudando experimentos de negócios em dezenas de

empresas, incluindo Banco da América, BMW, Hilton, Kraft, Petco, Staples, Subway e Walmart.

Realizar um teste padrão A/B num canal direto como a internet — comparando, por exemplo, a taxa de resposta para a versão A de uma página da *web* com a taxa de resposta para a versão B — é um exercício relativamente simples que utiliza matemática desenvolvida há um século. Mas a grande maioria (mais de 90%) de interações com consumidores é realizada por meio de sistemas de distribuição mais complexos, como redes de lojas, áreas de vendas, agências bancárias, franquias de *fast-food*, e assim por diante. Nesses ambientes, a experimentação de negócios está sujeita a uma série de complexidades analíticas, sendo que a mais importante é o tamanho da amostra: geralmente muito pequeno para produzir resultados válidos. Enquanto um grande varejista *online* pode simplesmente selecionar 50 mil consumidores de forma aleatória e detectar suas reações a uma oferta experimental, mesmo os maiores varejistas de lojas físicas não conseguem selecionar aleatoriamente 50 mil lojas para testar uma nova promoção. Para eles, um grupo de teste real é geralmente contado às dezenas, e não aos milhares. Na verdade, descobrimos que a maioria dos testes de novos programas de consumo são muito informais. Eles não se baseiam em métodos científicos e estatísticos comprovados e, por isso, os executivos acabam interpretando o ruído estatístico como a causa — e tomam decisões erradas.

Num experimento ideal, o experimentador separa uma variável independente (suposta causa) a partir de uma variável dependente (efeito observado), mantendo constantes as demais possíveis causas, e em seguida manipula a primeira para estudar mudanças na última. Manipulação seguida de análise e observação criteriosa permite obter *insights* sobre as relações entre causa e efeito que, idealmente, podem ser aplicados e testados em outras condições.

Para chegar a esse conhecimento — e garantir que a experimentação de negócios compense os gastos e os esforços — as empresas precisam responder a várias questões cruciais: a experiência tem um propósito claro? Os acionistas se comprometem a aceitar as consequências dos resultados? O experimento é viável? Como garantir resultados confiáveis? Extraímos o máximo de valor possível do experimento? (ver quadro “Lista de checagem para realizar um experimento de negócio”). Embora estas questões pareçam óbvias, muitas empresas começam a realizar testes antes de ter todas as respostas.

## EXPERIMENTO DA KOHL'S

# Varejista deseja testar a hipótese de que abrir as lojas uma hora mais tarde não provoca queda significativa nas vendas.

## Ideia em resumo

### O PROBLEMA

Na falta de dados suficientes para corroborar decisões sobre inovações propostas, executivos muitas vezes confiam em sua experiência, intuição ou senso comum convencional — nenhum deles é necessariamente relevante.

### A SOLUÇÃO

Num teste científico rigoroso, as empresas separam a variável independente (causa presumida) da variável dependente (efeito observado), enquanto as demais possíveis causas permanecem constantes, e então manipula a primeira para estudar variações na última.

### A ORIENTAÇÃO

Para aproveitar seus experimentos ao máximo, as empresas devem se perguntar: o experimento tem um propósito claro? Os acionistas se comprometeram a aceitar os resultados? O experimento é factível? Como garantir resultados confiáveis? Conseguimos extrair do experimento o máximo de valor possível?

## O experimento tem um propósito claro?

As empresas devem realizar experimentos, se essa for a única forma prática de responder às questões específicas sobre ações gerenciais sugeridas.

Considere a grande varejista Kohl, que em 2013 estava preocupada em reduzir seus custos operacionais. Uma sugestão foi abrir as lojas uma hora mais tarde de segunda a sábado. As opiniões dos executivos da empresa sobre o assunto divergiam. Uns argumentavam que reduzir horas de atendimento das lojas resultaria em queda significativa de vendas. Outros acreditavam que o impacto sobre as vendas seria mínimo. A única forma de discutir e chegar a alguma conclusão foi realizar um experimento rigoroso. Um teste envolvendo cem lojas mostrou que abrir as lojas uma hora mais tarde não resultaria em qualquer redução significativa nas vendas.

Para determinar quando um experimento é necessário, antes de tudo, os executivos precisam definir exatamente o que desejam saber. Só então poderão decidir se o teste será a melhor abordagem e, se for, qual o escopo da experiência. No caso da Kohl, a hipótese a ser testada era objetiva: abrir as lojas uma hora mais tarde para reduzir custos operacionais não provoca queda significativa nas vendas. Mas, com frequência, falta às empresas a disciplina para aperfeiçoar suas hipóteses, levando a testes ineficientes, desnecessariamente dispendiosos, ou, pior, ineficazes em responder a uma questão tão simples. Uma hipótese fraca (como, por exemplo, “é possível expandir nossa marca sofisticada”) não representa uma variável independente específica para testar uma determinada variável dependente, por isso é difícil apoiá-la ou rejeitá-la. Uma boa hipótese ajuda a descrever as variáveis.

Em muitas situações, executivos precisam ir além dos efeitos diretos de uma iniciativa e investigar seus efeitos colaterais. Quando, por exemplo, a Family

Dollar queria saber se deveria investir em unidades de refrigeração para poder vender ovos, leite e outros produtos perecíveis, ela descobriu que um efeito colateral — o aumento nas vendas de mercadorias secas tradicionais aos clientes que seriam atraídos pelos itens refrigerados — na verdade, teria um impacto maior nos lucros. Efeitos colaterais também podem ser negativos. Há alguns anos, a Wawa, cadeia de lojas de conveniência da Costa Leste dos Estados Unidos pretendia introduzir no café da manhã o item pão árabe, que tinha sido bem recebido em testes pontuais. Mas o projeto foi abandonado antes do lançamento, quando um experimento rigoroso — apoiado em grupos de teste e de controle seguido de análise de regressão — mostrou que o novo produto provavelmente canibalizaria outros itens mais lucrativos.

## Os acionistas se comprometeram a aceitar os resultados?

Antes de realizar qualquer teste, os acionistas devem concordar sobre como procederão quando os resultados surgirem. Eles devem se comprometer a pesar todos os resultados, em vez de aproveitar somente os melhores dados que apoiam um determinado ponto de vista. Talvez seja mais importante eles se disporem a se afastar de um projeto que não seja sustentado pelos dados.

Quando a Kohl pensou em acrescentar uma nova categoria de produto — móveis —, vários executivos foram extremamente favoráveis, prevendo uma receita adicional significativa. No entanto, um teste em 70 lojas durante seis meses mostrou uma redução efetiva nos lucros. Produtos que agora dispunham de menor área de exibição (para ceder espaço aos móveis) sofreram queda nas vendas, e a Kohl estava, de modo geral, perdendo clientes. Esses resultados negativos foram muito decepcionantes para os defensores da ideia e o programa foi descartado. O exemplo da Kohl

## Lista de checagem para realizar um experimento de negócio



### Propósito

- O experimento foca numa ação de gestão específica?
- O que as pessoas esperam aprender com o experimento?



### Adesão

- Que mudanças específicas deveriam ser feitas com base nos resultados?
- Como a organização garante que os resultados não são ignorados?
- Como o experimento se encaixa na agenda geral de aprendizagem e nas prioridades estratégicas da organização?



### Exequibilidade

- O experimento tem uma hipótese testável?
- Qual o tamanho da amostra?
- Nota: o tamanho da amostra depende do efeito esperado (5% de aumento nas vendas, por exemplo).
- A organização consegue realizar o experimento de forma viável nos locais de teste durante o período necessário?

evidencia o fato de que experimentos, muitas vezes, são necessários para avaliar objetivamente iniciativas apoiadas por pessoas com respaldo organizacional.

Obviamente deve haver boas razões para se lançar um projeto novo, mesmo que os benefícios previstos não sejam suportados pelos dados — por exemplo, mesmo que os experimentos mostrem que um programa não impulsionará substancialmente as vendas, ele ainda poderá ser útil para construir fidelidade do cliente. Mas, se a iniciativa proposta for um negócio garantido, por que gastar tempo e dinheiro realizando um teste?

É preciso instituir um processo para garantir que os resultados do teste não sejam ignorados, mesmo quando eles contradizem hipóteses ou a intuição de altos executivos. Na Publix Super Markets, uma cadeia do sudeste americano, praticamente todos os grandes projetos de varejo, especialmente os que exigem consideráveis investimentos de capital, devem ser submetidos a experimentos formais para receber sinal verde. As propostas passam por um processo de filtragem, sendo que o primeiro passo é o financeiro, para determinar se vale a pena realizar o experimento.

Para os projetos qualificados, analistas desenvolvem testes adequados e os submetem a um comitê que inclui o vice-presidente de finanças. Os experimentos aprovados pela comissão são então realizados e supervisionados por uma equipe de teste interna. O financeiro aprova verba significativa somente para iniciativas que tenham seguido o processo e cujos experimentos produziram resultados positivos. “Projetos são avaliados e aprovados muito mais rapidamente — e com menos detalhamento — quando os resultados de nossos testes são promissores”, observa Frank Maggio, executivo sênior de análise de negócios da Publix.

Ao criar e implementar esse processo de filtragem, é importante lembrar que experimentos podem formar uma agenda de aprendizagem para apoiar

prioridades organizacionais da empresa. Na Petco, cada solicitação de análise é acompanhada por uma descrição sobre como aquele experimento deverá contribuir para tornar a estratégia global da empresa mais inovadora. No passado, a empresa realizava cerca de cem testes por ano, mas esse número foi cortado para 75. Muitas solicitações de testes são negadas porque a empresa já realizou um teste semelhante no passado. Outras são rejeitadas porque as mudanças pretendidas não são suficientemente radicais para justificar as despesas com o teste (por exemplo, um aumento no preço de um único item de US\$ 2,79 para US\$ 2,89). “Queremos testar coisas que de fato farão a diferença para o negócio”, ressalta John Rhoades, ex-diretor de análise de varejo da empresa. “Queremos tentar novos conceitos ou novas ideias.”

### O experimento é factível?

Experimentos precisam ter predições testáveis. Mas a “densidade causal” do ambiente do negócio — ou seja, a complexidade das variáveis e suas interações — pode dificultar muito a determinação das relações de causa e efeito. A aprendizagem extraída de um experimento de negócio pode não ser exatamente tão fácil como isolar uma variável independente, manipulá-la e observar as mudanças na variável dependente. Ambientes estão em constante mudança, as possíveis causas em experimentos de negócios muitas vezes são incertas ou desconhecidas e, por isso, relações entre elas geralmente são complexas e mal compreendidas.

Imagine uma cadeia de varejo hipotética com dez mil lojas de conveniência, sendo oito mil chamadas QwikMart e duas mil denominadas FastMart. O faturamento anual das lojas Qwik-Mart beirava a média de US\$ 1 milhão e das lojas FastMart US\$ 1,1 milhão. Um executivo sênior propõe uma questão aparentemente simples: mudar o nome das lojas QwikMart para FastMart levaria a um aumento de US\$ 800 mil



## Confiabilidade

- Que medidas serão utilizadas para contabilizar o viés sistêmico, seja ele consciente ou inconsciente?
- As características do grupo de controle são compatíveis com as do grupo de teste?
- O experimento pode ser realizado na forma “cego” ou “duplo cego”?
- Todos os vieses restantes foram eliminados por meio de análises estatísticas ou outras técnicas?
- Se o mesmo teste for realizado por outros, eles obterão resultados similares?



## Valor

- A organização pensou num lançamento com objetivo direcionado — isto é, um lançamento que leve em consideração o efeito da iniciativa proposta sobre diferentes clientes, mercados e segmentos — para concentrar investimentos em áreas onde potencial retorno financeiro seja o mais alto?
- A organização implementou somente os componentes da iniciativa com maior retorno sobre o investimento?
- A organização tem uma boa percepção sobre quais variáveis causam quais efeitos?

no seu faturamento? Obviamente, inúmeros fatores afetam as vendas da loja, incluindo suas dimensões físicas, o número de habitantes num certo raio e seu poder aquisitivo, o número de horas que a loja permanece aberta por semana, a experiência do gerente, o número de concorrentes nos arredores, e assim por diante. Mas o executivo só está interessado numa variável: o nome das lojas (*QwikMart versus FastMart*).

A solução óbvia é realizar um experimento mudando o nome de algumas lojas *QwikMart* (digamos, dez) para avaliar o resultado. Mas determinar o efeito da mudança de nome naquelas lojas pode ser uma tarefa complexa, porque vários outros parâmetros podem ter variado simultaneamente. O tempo estava muito ruim em quatro das localidades, por exemplo, ou, numa delas o gerente foi substituído, ou um grande conjunto residencial foi inaugurado perto de outra, e um concorrente iniciou uma campanha promocional agressiva perto de outra. A menos que a empresa possa separar o efeito da mudança de nome das outras variáveis, o executivo não saberá com certeza se a mudança de nome ajudou (ou prejudicou) os negócios.

Para operar em ambientes de alta densidade causal, as empresas precisam analisar se é viável selecionar uma amostra suficientemente grande para promediar os efeitos de todas as variáveis, exceto as que estão sendo estudadas. Infelizmente, esse tipo de experimento nem sempre é factível. O custo de um teste envolvendo uma amostra adequada pode ser proibitivo, ou as mudanças nas operações podem ser perturbadoras demais. Nesses casos, como discutiremos mais adiante, muitas vezes, executivos utilizam técnicas analíticas sofisticadas, envolvendo grandes dados, para aumentar a validade estatística dos resultados.

Com isso em mente, deve-se notar que executivos muitas vezes pressupõem, equivocadamente, que uma amostra maior levará automaticamente a melhores dados. Na verdade, um experimento pode

envolver muitas observações. Mas, se estiverem fortemente aglomeradas ou correlacionadas, o verdadeiro tamanho da amostra pode, na verdade, ser bem pequeno. Quando uma empresa utiliza um distribuidor, em vez de vender diretamente aos clientes, por exemplo, esse ponto de distribuição poderia facilmente fornecer correlações entre dados do cliente.

O tamanho da amostra depende, em grande parte, da intensidade do efeito esperado. Se uma empresa espera que a causa (por exemplo, uma mudança no nome da loja) tenha um efeito grande (um aumento substancial nas vendas), a amostra pode ser menor. Se o efeito esperado for pequeno, a amostra deve ser maior. Isso pode parecer contraintuitivo, mas pense da seguinte forma: quanto menor o efeito esperado, maior o número de observações necessário para detectá-lo sobre o ruído circundante, com a confiabilidade estatística desejada.

Selecionar o tamanho correto da amostra é mais importante que garantir a validade estatística dos resultados. Uma amostra adequada também pode contribuir para reduzir custos dos testes e aumentar a inovação da empresa. *Softwares* facilmente disponíveis podem ajudar as empresas a escolher o tamanho ideal da amostra. (Informação necessária: a empresa de Jim Manzi, a *Applied Predictive Technologies*, produz um, o *Test & Learn*.)

### Como garantir resultados confiáveis?

Na seção anterior, descrevemos os conceitos básicos para realizar um experimento. A verdade é que as empresas normalmente precisam combinar confiabilidade, custo, tempo, e outros itens de ordem prática. Três métodos podem ajudar a reduzir essas compensações e aumentar a confiabilidade dos resultados.

**Testes de campo randômicos.** O conceito de randomização na pesquisa médica é bastante simples: imagine um grupo grande de pessoas com as mesmas

características e doenças e divide-o aleatoriamente em dois subgrupos. Ofereça tratamento somente para um subgrupo e monitore cuidadosamente a saúde de todos. Se o grupo tratado (de teste) apresentar, estatisticamente, melhor resultado que o grupo não tratado (de controle), então a terapia será considerada eficaz. Da mesma forma, testes de campo randômicos podem ajudar as empresas a descobrir se determinadas mudanças resultarão em melhor desempenho.

A empresa de serviços financeiros Capital One utiliza há muito tempo experimentos rigorosos para testar até mudanças aparentemente triviais. Por meio de testes de campo randômicos, por exemplo, a empresa testa a cor dos envelopes utilizados nas ofertas de produtos, enviando dois lotes (um na cor de teste e outro branco) para detectar quaisquer diferenças na resposta.

A randomização desempenha papel importante: ela ajuda a evitar que um viés sistêmico, introduzido consciente ou inconscientemente, afete um experimento, e distribui uniformemente quaisquer potenciais causas remanescentes (e provavelmente desconhecidas) do resultado entre grupos de teste e de controle. Mas testes de campo randômicos não estão livres de desafios. Para que os resultados sejam válidos, os testes devem ser realizados com rigor estatístico.

Em vez de identificar uma população de sujeitos de teste com as mesmas características e, em seguida, dividi-los aleatoriamente em dois grupos, executivos,

às vezes, cometem o erro de selecionar um grupo de teste (por exemplo, um grupo de lojas de uma cadeia) e, depois, supor que o restante das lojas forma o grupo de controle. Ou, ao selecionar os grupos de teste e de controle, eles, inadvertidamente, introduzem vieses no experimento. A Petco costumava selecionar suas 30 melhores lojas (como grupo de teste) para testar uma nova iniciativa e compará-las com suas 30 piores lojas (como grupo de controle). Iniciativas testadas desta forma geralmente parecem muito promissoras, mas fracassam quando são lançadas.

Atualmente a Petco leva em conta uma ampla gama de parâmetros — espaço físico da loja, dados demográficos do cliente, presença de concorrentes nos arredores, etc. — para comparar características de grupos de controle e de teste. (A Publix também faz isso.) Resultados desses experimentos têm se mostrado muito mais confiáveis.

**Teste cego.** Para reduzir vieses e aumentar ainda mais a confiabilidade, a Petco e a Publix realizaram testes “cegos”, que ajudam a prevenir o efeito Hawthorne: a tendência que os participantes do estudo apresentam de modificar seu comportamento, consciente ou inconscientemente, quando sabem que são sujeitos de um experimento. Na Petco nenhum funcionário das lojas de teste é informado sobre o experimento, e a Publix realiza testes cegos sempre que possível. Em testes simples que envolvem apenas alteração de preços, a Publix pode aplicar testes cegos porque as lojas estão continuamente praticando novos preços, por isso os testes não são distinguíveis das práticas normais de funcionamento.

Mas os procedimentos cegos nem sempre são práticos. Nos testes de novos equipamentos ou métodos de trabalho, a Publix normalmente informa as lojas de que elas foram selecionadas como grupo de teste. (Obs.: um padrão experimental mais exigente utiliza teste “duplo cego”, nos quais nem o experimentador nem os sujeitos sabem quais participantes integram o grupo de teste e quais participam do grupo de controle. Testes duplo-cegos são amplamente utilizados em pesquisas médicas, mas não são comuns em negócios.)

**Grandes dados.** Em ambientes *online* e outros canais diretos, a matemática necessária para realizar um experimento aleatório rigoroso é bem conhecida. Mas, como já discutimos, a grande maioria das transações comerciais ocorre em outros canais, como lojas de varejo. Em testes nesses ambientes, o tamanho da amostra geralmente é inferior a 100, violando

## EXPERIMENTO DA WAWA

Um novo pão árabe teve sucesso em pontos de teste, mas a cadeia o interrompeu depois que experimentos rigorosos mostraram que ele canibalizaria outros produtos.

## Como Big Data pode ajudar

Para filtrar o ruído estatístico e identificar relações de causa e efeito, experimentos de negócios deveriam empregar idealmente amostras da ordem de milhares. Mas essa condição pode ser proibitivamente onerosa ou impossível. Uma nova abordagem para comercializar produtos em estoque poderá ter de ser testada em apenas 25 lojas, um programa de treinamento em vendas com 32 vendedores e uma reestruturação proposta para dez unidades hoteleiras. Nesses casos, *Big Data* e outras técnicas computacionais sofisticadas, como “aprendizado da máquina”, podem ajudar. Vejamos como:

### O começo

Se um varejista deseja testar o novo *layout* de sua loja, ele deverá coletar dados detalhados (como, por exemplo, proximidade de concorrentes, carreira dos funcionários e demografia do cliente) em cada unidade de análise (cada loja e sua área de atuação, cada vendedor e sua contabilidade, e assim por diante). Essa informação criará num grande conjunto de dados. Determinar quantas e quais lojas, clientes ou funcionários devem participar do teste e qual sua duração depende da volatilidade dos dados e da precisão necessária para impactar as estimativas.

### Formação do grupo de controle

Em experimentos envolvendo pequenas amostras, é fundamental compatibilizar corretamente os sujeitos de teste (como lojas individuais ou clientes) com sujeitos de controle, e isso depende da capacidade do experimentador de identificar completamente dezenas ou até centenas de variáveis que caracterizam os sujeitos de teste. Fontes de *Big Data* (registro completo de transações com cliente, dados climáticos detalhados, tendências das mídias sociais, etc.) podem ajudar nessa tarefa. Uma vez determinadas as características, pode-se criar um grupo de controle que contenha todos os elementos do grupo de teste, exceto o que está sendo testado. Isso permite que o varejista determine se os resultados do teste foram influenciados somente por um elemento — o novo *layout* — ou por outros fatores (variâncias demográficas, melhores condições econômicas, clima mais quente).

### De olho nas melhores oportunidades

Os mesmos dados podem ser usados para identificar situações nas quais o programa testado é eficaz. O novo *layout* da loja, por exemplo, pode funcionar melhor em áreas urbanas altamente competitivas, mas pode ter um sucesso apenas moderado em outros mercados. Ao identificar esses padrões, o experimentador pode implementar o programa para situações onde ele funciona e evitar investimentos onde o programa pode não gerar o melhor ROI.

### Modelagem do programa

Fontes adicionais de *Big Data* podem ser usadas para caracterizar componentes relativamente eficazes do programa. Um varejista que está testando os efeitos de um novo *layout* de loja, por exemplo, pode usar dados capturados pelo fluxo de vídeos interno da loja para determinar se o novo *layout* está encorajando clientes a se deslocar mais pelo recinto ou se está gerando mais tráfego perto de produtos com margens de lucros mais altas. Ou o experimentador pode descobrir que deslocar itens logo na entrada da loja e instalar novas prateleiras pode causar um impacto positivo, mas deslocar as caixas registradoras tumultua a saída e prejudica os lucros.

as regras de diversos métodos estatísticos-padrão. Para minimizar os efeitos dessa limitação, as empresas podem utilizar algoritmos especializados e combiná-los com vários conjuntos de grandes dados (ver quadro “Como grandes dados podem ajudar”).

Imagine um grande varejista analisando a reestruturação de uma loja que custaria US\$ 0,5 bilhão para ser aplicada em 1.300 localidades. Para testar a ideia, o varejista reestruturou 20 lojas e acompanhou os resultados. A equipe financeira analisou os dados e concluiu que a reformulação aumentaria as vendas em parcos 0,5%, o que resultaria num retorno negativo sobre o investimento. A equipe de *marketing* realizou uma análise independente e previu que a reestruturação resultaria num saudável aumento de 5% nas vendas.

Como se constatou depois, a equipe financeira havia comparado os locais de teste com outras lojas da cadeia com aproximadamente o mesmo tamanho, renda demográfica, e outras variáveis, mas não estavam necessariamente no mesmo mercado geográfico. Ela também usou dados correspondentes a seis meses anteriores e seis meses posteriores à reestruturação. E a equipe de *marketing* havia comparado lojas dentro da mesma região geográfica, mas considerando um período de 12 meses antes e depois da reestruturação. Para saber em quais resultados confiar, a empresa utilizou grandes dados, incluindo dados de nível transacional (itens da loja, horários em que a venda ocorre, preços), atributos da loja e dados sobre o ambiente em torno das lojas (concorrência, dados demográficos, condições climáticas). Desta forma, as lojas incluídas no grupo de controle eram mais compatíveis com aquelas em que a reestruturação tinha sido testada, o que validou estatisticamente o pequeno tamanho da amostra. A empresa utilizou então métodos estatísticos objetivos para rever as



duas análises. Conclusão: os resultados da equipe de *marketing* foram os mais precisos.

Mesmo quando a empresa não pode seguir um protocolo de teste rigoroso, analistas podem ajudar a identificar e corrigir certos vieses, falhas de randomização, e outros erros experimentais. Uma situação comum ocorre quando a função de teste da organização é introduzida com experimentos naturais não randômicos — o vice-presidente de operações, por exemplo, pode querer saber se o novo programa de treinamento de funcionários, que foi introduzido em cerca de 10% dos mercados da empresa, é mais eficiente que o antigo. Como se verifica nessas situações, os mesmos algoritmos e grandes conjuntos de dados usados para resolver o problema de amostras pequenas ou correlacionadas também podem ser aproveitados para revelar *insights* valiosos e aumentar a precisão dos resultados. Essa análise pode então ajudar os experimentadores a projetar um teste de campo randômico real para confirmar e refinar os resultados, principalmente quando são relativamente contraintuitivos, ou são necessários para tomar uma decisão com grandes desdobramentos econômicos.

Em qualquer experiência, o protótipo da perfeição é a reprodutibilidade, isto é, se outros realizarem o mesmo teste, deverão obter resultados semelhantes. Geralmente se torna impraticável repetir um teste caro, mas as empresas podem checar os resultados de outras formas. A Petco, às vezes, promove um lançamento ensaiado de grandes iniciativas para confirmar os resultados antes de implementar o programa na empresa toda. A Publix utiliza um mecanismo próprio para acompanhar resultados de um lançamento e compará-los com os benefícios esperados.

### Conseguimos extrair do experimento o máximo de valor

Muitas empresas investem em experimentos, mas depois não conseguem retirar dele o máximo de valor possível. Para evitar esse inconveniente, executivos devem levar em conta o efeito do programa sobre vários clientes, mercados e segmentos e concentrar os investimentos em áreas onde a potencial recuperação de investimentos seja a mais alta. Normalmente a pergunta correta não é “o que funciona?”, mas “o que funciona onde?”

A Petco normalmente lança um programa somente em lojas mais semelhantes às lojas de teste que apresentaram os melhores resultados. Ao fazer isso, ela não só economiza nos custos de implementação,

## MELHOR CONDUTA DA PETCO

# O varejista especializado garante resultados confiáveis em experimentos combinando características dos grupos de teste e de controle.

mas também evita envolver lojas onde o novo programa pode não produzir os benefícios esperados, ou até ter consequências negativas. Graças a esses lançamentos com alvo certo, a Petco tem conseguido, consistentemente, dobrar as vantagens previstas em novas iniciativas.

Outra tática útil é a “engenharia de valor”. A maioria dos programas inclui vários componentes. Alguns criam benefícios que compensam os custos, outros não. O truque, então, é implementar apenas componentes com um retorno sobre o investimento (ROI, em inglês) atraente. Para dar um exemplo simples, vamos supor que os testes de um varejista de uma promoção com descontos de 20% mostram um aumento de 5% nas vendas. Que parte desse aumento pode ser atribuída aos produtos em si e que parte é devida à propaganda e capacitação das equipes da loja, sendo que ambos contribuíram para direcionar os clientes para aqueles produtos específicos? Nesse caso, as empresas podem realizar experimentos para investigar várias combinações de componentes (por exemplo, promoções com propaganda, mas sem treinamento adicional de pessoal). Uma análise de resultados pode desmembrar os efeitos, e permitir que componentes (por exemplo, treinamento adicional de pessoal) que resultem em ROI baixo ou negativo sejam descartados.

Além disso, uma cuidadosa análise de dados gerados por experimentos permite que as empresas entendam melhor suas operações e testem hipóteses sobre variáveis que causam efeitos. Com grandes dados a ênfase está na correlação — descobrindo,

por exemplo, que a venda de certos produtos tende a coincidir com a venda de outros. Mas a experimentação em negócios pode permitir que as empresas olhem além da correlação e investiguem a causalidade — revelando, por exemplo, fatores que causam aumento (ou redução) de compras. Esse conhecimento fundamental da causalidade pode ser crucial. Sem ele, a compreensão que os executivos têm de seus negócios é fragmentada e as decisões que tomam podem facilmente sair pela culatra.

Quando a Cracker Barrel Old Country Store, cadeia de restaurantes especializados em culinária sulista, realizou um experimento para decidir se deveria substituir lâmpadas incandescentes por lâmpadas de LED em seus restaurantes, os executivos se surpreenderam ao descobrir que o tráfego de clientes, na verdade, tinha *diminuído* nos locais onde lâmpadas de LED foram instaladas. A ideia da iluminação poderia ter parado aí, mas a empresa aprofundou um pouco mais sua análise para tentar entender as causas subliminares. Como se verificou, a nova iluminação atenuava a luz do pórtico frontal dos restaurantes e muitos clientes erroneamente pensavam que os restaurantes estavam fechados. Essa foi uma constatação intrigante — os LEDs deveriam ter tornado os pórticos mais brilhantes. Um aprofundamento maior da investigação revelou que antes os gerentes dos restaurantes não seguiam os padrões de iluminação da empresa. Eles utilizavam suas próprias soluções, muitas vezes iluminando demais os pórticos de entrada. Por isso a luminosidade diminuiu quando os restaurantes aderiram à nova política de uso de LEDs. A questão aqui é que somente a correlação teria deixado a empresa com a impressão errada — de que LEDs são ruins para os negócios. Foi preciso que a experimentação revelasse a real relação causal.

De fato, sem compreender completamente a causalidade, as empresas estão mais sujeitas a cometer grandes erros. Você se lembra do experimento da Kohl para investigar os efeitos de abrir as lojas uma hora mais tarde? No início do teste a empresa acusou uma queda nas vendas. Nesse momento, os executivos poderiam ter interrompido a iniciativa. Mas a análise mostrou que o número de transações com clientes tinha se mantido constante. O que houve foi uma queda em unidades por transação. Finalmente, as unidades por transação se recuperaram e as vendas totais retornaram aos níveis anteriores. O experimento da Kohl não explicou completamente a queda inicial, mas seus executivos resistiram à tentação de culpar a redução

nas horas de funcionamento da loja. Eles não se precipitaram em comparar correlação com causalidade.

Mas o mais importante é que muitas empresas estão descobrindo que realizar um experimento é só o começo. Valor decorre de análise e posterior exploração de dados. Nos últimos anos, a Publix passou 80% do tempo de teste reunindo dados, e 20% analisando-os.

### Desafiando o senso comum

Se as empresas levarem em conta o tamanho da amostra, grupos de controle, randomização e outros fatores, elas poderão garantir a validade dos resultados de seus testes. Quanto mais validados e reproduzíveis os resultados, melhor eles se sustentarão diante de resistência interna, que pode ser particularmente forte quando os resultados põem em xeque condutas do setor e senso comum que prevalece há muito tempo.

Quando os executivos da Petco investigaram novos patamares de preço para um produto vendido por peso, os resultados foram inequívocos. Disparado, o melhor preço foi para 100 gramas do produto e com um valor de centavos que terminasse em 25, como US\$ 3,25. Este resultado contrapunha-se nitidamente ao senso comum, que normalmente dá preferência para valores que terminam em 9, como US\$ 4,99 ou US\$ 2,49. “Isso quebrou a regra do varejo de que você não pode ter um preço ‘feio’”, observa Rhoades. No início, os executivos da Petco estavam céticos sobre os resultados, mas como o experimento fora extremamente rigoroso, eles finalmente se dispuseram praticar os novos preços. Um lançamento direcionado confirmou os resultados, levando a um salto nas vendas de mais de 24% em seis meses.

A lição não é simplesmente de que experimentação em negócios pode levar a formas melhores de fazer as coisas. Ela também confere às empresas a confiança necessária para neutralizar o senso comum convencional perverso e falhas de intuição para negócios que executivos maduros costumam apresentar. E tomada de decisão mais inteligente fatalmente melhora o desempenho.

A J.C. Penney poderia ter evitado o desastre testando rigorosamente os componentes de sua reestruturação geral? Neste momento, é impossível saber. Mas uma coisa é certa: antes de tentar implementar um programa tão ousado, a empresa deveria ter se certificado de que conhecimento — e não intuição — estava orientando a decisão. ▣