

УДК 338.22.021.1:519.2

ББК 65.050.2:22.172

X121

Хаббард Дуглас У.

X121 Как измерить все, что угодно. Оценка стоимости нематериального в бизнесе / Дуглас У. Хаббард / [Пер. с англ. Е. Пестеревой]. — М.: ЗАО «Олимп—Бизнес», 2009. — 320 с.: ил.

ISBN 978-5-9693-0163-4 (рус.)

Среди людей бытует мнение, что многие вещи в мире не поддаются количественной оценке. Это широко распространенное заблуждение наносит существенный ущерб экономике, общественному благосостоянию, окружающей среде и даже национальной безопасности. В результате убежденности в невозможности измерения таких «нематериальных активов», как качество, отношение к делу, удовлетворенность клиентов, бренд и т. п., принимаемые решения часто оказываются необоснованными, а порой и ошибочными.

Дуглас Хаббард пытается развеять это вредное заблуждение, предлагая свой подход к оценке «неизмеряемого», названный им «прикладная информационная экономика». Он знакомит читателей с понятием «калиброванная оценка», оценкой риска (метод Монте-Карло), способами выборочного исследования, другими необычными инструментами измерений (Интернет, экспертные оценки, рынки предсказаний и др.), а также с оценкой стоимости информации. Свой подход автор применяет в разных областях и приводит ряд примеров успешного решения задач по количественному измерению. В книге содержатся ценные инструкции и рекомендации, которые без труда может использовать любой человек, принимающий решения, а также приложения, позволяющие проверить способность читателя давать калиброванные оценки.

Книга предназначена широкому кругу читателей, интересующихся процессами обоснования и принятия решений. Она будет полезна руководителям, менеджерам, преподавателям и студентам.

УДК 338.22.021.1:519.2

ББК 65.050.2:22.172

Охраняется Законом РФ об авторском праве. Воспроизведение всей книги или ее части в любом виде воспрещается без письменного разрешения издателя.

©2007 by John Wiley & Sons, Inc.
All rights reserved.

© ЗАО «Олимп—Бизнес», перевод
на рус. яз., оформление, 2009
Все права защищены.

ISBN 978-5-9693-0163-4 (рус.)

ISBN 978-0-470-11012-6 (англ.)



Содержание

Предисловие к русскому изданию	X
Об авторе	XI
О книге	XII
Отзывы о книге «Как измерить все, что угодно. Оценка стоимости нематериального в бизнесе»	XIV
Предисловие	XVI
От автора	XX
 ЧАСТЬ I. Измерение: решение существует	1
Глава 1. Нематериальное и проблема его измерения	3
Глава 2. Интуитивное умение измерять все: Эратосфен, Энрико и Эмили	7
Как один древний грек определил размеры земного шара	8
Как определить неизвестную: берите пример с Ферми	9
Эксперименты — не только для взрослых	12
Чему нас учат Эратосфен, Энрико и Эмили	17
Примечания	18
Глава 3. Почему неизмеримость нематериального — всего лишь иллюзия	19
Сущность измерения	20

Объект измерения.....	24
Методы измерения	27
Экономические возражения против проведения измерений	32
Возражение против полезности «статистики» вообще	34
Этические возражения против проведения измерений	36
Навстречу универсальному подходу к измерению	38
Примечания	40
Часть II. Прежде, чем приступить к измерениям.....	41
Глава 4. Формулирование задачи по измерению	43
Уточним терминологию: что на самом деле означают понятия «неопределенность» и «риск»	45
Примеры уточнения объекта измерения: чему бизнес может поучиться и у государства.....	47
Примечания	53
Глава 5. Калиброванные оценки: что вам известно уже сейчас?	54
Хорошо ли вы оцениваете шансы?.....	54
Упражнение «Калибровка»	57
Дальнейшее повышение точности калибровки	61
Понятийные заблуждения, мешающие калибровке.....	63
Эффект калибровки.....	67
Примечания	72
Глава 6. Оценка риска: введение в моделирование методом Монте-Карло.....	73
Пример применения метода Монте-Карло и расчета риска	76
Инструменты и другие ресурсы для моделирования методом Монте-Карло	82
Парадокс риска.....	86
Примечание.....	87
Глава 7. Оценка стоимости информации.....	88
Вероятность ошибиться и цена ошибки: ожидаемые потери от упущенных возможностей.....	89

Стоимость информации для переменных величин	92
Мир несовершенен: стоимость частичного снижения неопределенности	96
«Уравнение прозрения»: стоимость информации меняет все.....	99
Связь между неопределенностью, риском и стоимостью информации: первые измерения	103
Примечание.....	104
Часть III. Методы измерения	105
Глава 8. Переход от объекта к способу измерения.....	107
Инструменты наблюдения: введение в инструментарий измерений.....	109
Разложение на составляющие	113
Вторичные исследования: предположим, что до вас этот объект уже измеряли.....	118
Основные методы наблюдения: не работает один — попробуйте другой	119
Не измеряйте точнее, чем нужно.....	122
Учитывать погрешность	124
Выберите или разработайте нужный инструмент.....	129
Глава 9. Выборочная реальность: как наблюдение за частью рассказывает нам о целом	132
Развиваем интуитивное понимание случайных выборок: пример с леденцами.....	137
Кое-что о малых выборках: подход пивовара	138
Простейшие статистические методы получения выборок.....	143
Пристрастный отбор методов выборочного обследования.....	147
Определите порог	154
Эксперимент	157
Выявление взаимозависимости параметров: введение в регрессионное моделирование	161
Примечания	167
Глава 10. Кое-что о Байесе	168
Простая байесовская статистика.....	168
Используйте свой природный байесовский инстинкт.....	172

Неоднородный бенчмаркинг и его использование для оценки «ущерба бренду».....	178
Кое-какие тонкости: байесовская инверсия для диапазонов	182
Примечания	188
Часть IV. Не только основы.....	189
Глава 11. Предпочтения и подходы:	
«мягкие» аспекты измерения	191
Наблюдение за мнениями, ценностями и поиски счастья.....	192
Готовность платить: определение стоимости через компромиссы	195
Ставим все на карту: количественное определение склонности к риску	200
Количественное определение субъективных компромиссов: решение проблемы нескольких взаимоисключающих предпочтений.....	203
Не забывать о главной цели: максимизация прибыли и субъективные компромиссы.....	208
Примечания	210
Глава 12. Решающий инструмент измерения:	
людские суждения.....	211
Хомо абсурдус: странные мотивы наших решений.....	212
Значение систематизации: пример оценки эффективности деятельности	215
На удивление простые линейные модели	217
Как стандартизировать любую оценку: модели Раша.....	219
Устранение непоследовательности людских суждений: модель линзы	222
Панацея или плацебо? Сомнительные методы измерения	228
Сравнение методов.....	235
Примечания	236
Глава 13. Новые инструменты измерения для менеджмента.....	239
Маркеры XXI века: в ногу с техническим прогрессом	239

Интернет как инструмент измерения мира	242
Рынки предсказаний: проведение измерений с эффективностью Уолл-стрит	246
Примечание.....	253
Глава 14. Универсальный метод измерения:	
прикладная информационная экономика	254
Сводим все воедино	255
Подход прикладной информационной экономики.....	256
Пример из практики: стоимость системы, следающей за качеством питьевой воды.....	260
Пример из практики: прогнозирование потребности морской пехоты в топливе.....	266
Советы начинающим: несколько последних примеров	274
Подведем итоги.....	281
Примечание.....	282
Приложение. Тесты на калибровку (и ответы на них)	283



Предисловие к русскому изданию

Уважаемый читатель!

Как известно, в реальном мире все можно измерить — вопрос только в том, с какой точностью. Очевидно, что степень точности тех или иных измерений/оценок будет зависеть от глубины наших знаний об объекте измерений и о последующем применении результатов этих измерений. В свою очередь, требуемая степень точности результатов измерений определяет выбор методов измерений. Но, как уже было отмечено, данный выбор ограничен глубиной наших знаний в каждом конкретном случае. Такой естественный замкнутый круг ни в коем случае не означает тупиковую ситуацию, а лишь только указывает на логическую связь между знанием предмета измерений, применением их результатов и требуемой точностью данных измерений.

Эта идея красной нитью проходит через всю книгу — все можно с какой-то точностью измерить/оценить.

Идеи и практические советы, изложенные в данной книге, будут весьма полезными для управленцев различного уровня при принятии решений по бизнесу в целом и по нематериальным ценностям в частности.

Приятного чтения.

Акоп Саркисян
Ernst & Young CIS
Партнер, Руководитель Департамента
консультационных услуг
по сделкам в странах СНГ



Об авторе

Дуглас У. Хаббард — изобретатель прикладной информационной экономики, метода измерения, высоко оцененного компаниями The Gartner Group, Giga Information Group и Forrester Research. Международно признанный эксперт в области определения стоимости информационных решений, он нередко выступает с докладами на научных конференциях. Его статьи публиковались в журналах «Information Week», «CIO Enterprise» и «DBMS Magazine». Бывший сотрудник Coopers & Lybrand, он обладает более чем 20-летним опытом консалтинга в области управления информационными технологиями. Из них 12 лет Хаббард обучает своему методу сотрудников разных организаций. Десятки компаний из списка «Fortune 500» и государственных учреждений применяли прикладную информационную экономику для решения проблем в области инвестиций в ИТ, военной логистики, венчурных инвестиций, авиации и космонавтики, а также защиты окружающей среды. Более полную информацию можно получить на сайте: www.howtomeasureanything.com



О книге

Книга «Как измерить все, что угодно» показывает значение измерений в самых разных областях — от информационных технологий до финансовой отчетности — для нашего понимания бизнеса и мира в целом. Эта изобилующая идеями, яркая книга учит, как измерить в своем бизнесе то, что вы до сих пор считали не поддающимся количественной оценке, например удовлетворенность потребителей, организационную гибкость, связанный с новой технологией риск и доходность инвестиций в технологию.

Примеры, рассказывающие и о том, как ихтиобиологи определяют численность популяции рыбы в большом озере, и о том, как Корпус морской пехоты США выяснял, что на самом деле имеет значение при прогнозировании потребности в топливе в боевых условиях, помогут вам поближе познакомиться с «универсальным подходом» к измерению «неизмеряемого», а также с некоторыми интересными методами решения конкретных проблем.

Из этой книги вы узнаете об:

- иллюзии по поводу нематериальных активов: как их можно измерять;
- калиброванных оценках: что вы знаете уже сейчас;
- оценке риска: введение в метод Монте-Карло;
- реалиях выборочного исследования: как наблюдение за несколькими объектами рассказывает нам обо всех;
- необычных инструментах измерения, таких как Интернет, экспертные оценки, рынки предсказаний и др.;

- оценке стоимости информации: какой показатель стоит определять.

Написанная Дугласом Хаббардом, признанным экспертом и создателем метода прикладной информационной экономики, книга «Как измерить все, что угодно» рассказывает о том, как ее автор использовал свой подход в различных областях, и о том, как любая задача по измерению, какой бы сложной, запутанной или плохо сформулированной она ни была, поддается решению проверенными методами. Содержащая четкие и легкие для выполнения инструкции, эта книга — ценный информационный ресурс, без которого не может обойтись ни один руководитель.



Отзывы о книге «Как измерить все, что угодно. Оценка стоимости нематериального в бизнесе»

Мне очень нравится эта книга. Дуглас Хаббард помогает находить путь к ответам практически на любые вопросы о бизнесе, науке или жизни... Он доказывает: проводя поиск показателей в процессе решения проблемы, мы на самом деле пытаемся узнать об интересующем нас предмете больше, чем знаем сейчас. «Как измерить все, что угодно» вооружает читателей именно теми инструментами, в которых они нуждаются для более качественных оценок, более глубокого понимания, прогресса и успеха.

*Доктор Питер Туннетт,
директор по технологиям компании CyberTrust
и разработчик первой антивирусной программы*

Дуг Хаббард предложил легковоспринимаемое объяснение, проливающее свет на процесс, которым пользуются менеджеры, чтобы получать информацию, необходимую для принятия менее рискованных и более выгодных деловых решений. Мы советуем своим клиентам испытать этот эффективный и удобный подход.

*Питер Скей, исполнительный вице-президент
и директор по операциям The Advisory Council*

Читая эту книгу, вы быстро понимаете: измерить можно практически все, но в то же время учиться оценивать только то, что имеет значение. Эта книга продирается сквозь устоявшиеся клише и бизнес-риторику и предлагает практические шаги по использованию измерений как инструмента принятия более удачных решений. Хаббард заполняет существовавшие ранее пробелы и доказывает возможность использования университетской статистики при принятии деловых решений.

*Рей Гилберт,
исполнительный вице-президент Lucent*

Книга Хаббарда замечательна тем, что показывает возможности методов измерения в самых разных областях и отличается замечательной доходчивостью. Это обязательное чтение для всех специалистов, кто когда-либо восклицал: «Конечно, этот показатель очень важен, но можно ли его рассчитать?»

*Доктор Джек Стеннер, один из основателей
и главный исполнительный директор MetraMetrics, Inc.*



Предисловие

Я написал книгу, чтобы развеять дорого обходящееся и разделяемое сегодня многими заблуждение, что некоторые вещи измерить невозможно. Это широко распространенное мнение наносит значительный ущерб экономике, общественному благосостоянию, окружающей среде и даже национальной безопасности. Такие «нематериальные» ценности, как качество продукции, отношение сотрудников к делу или экономический эффект от использования более чистой воды, нередко должны учитываться компаниями и государственными органами при принятии важнейших решений. Зачастую принятие решения требует количественной оценки предполагаемых нематериальных активов, но когда руководитель считает что-то не поддающимся такой оценке, он и не пытается это делать.

В результате принимаются далеко не столь обоснованные решения, как ожидалось. Возрастает вероятность ошибки. Ресурсы распределяются не по назначению, хорошие идеи отвергаются, а плохие принимаются. Деньги бросаются на ветер. В некоторых случаях возникает угроза жизни и здоровью людей. Убеждение, что некоторые, очень важные, вещи измерить невозможно, — песок в механизме функционирования всей экономики.

Любому человеку, принимающему важные решения, полезно знать, что вся необходимая ему информация поддается измерению. В то же время в условиях демократии и свободного предпринимательства к числу этих «важных лиц» относятся и избиратели и потребители. Вполне вероятно, что ваши профессиональные и личные решения оказались бы более удачными, проведи вы более точные измерения.

И почти наверняка лично вам уже не раз приходилось ощущать негативные последствия того, что *кто-то другой* принимал недостаточно взвешенное решение.

Всю свою жизнь я занимался измерением вещей, которые многие считали не поддающимися количественной оценке. Впервые необходимость в совершенствовании приемов измерения я заметил в 1998 г., вскоре после того, как получил диплом МВА и пришел работать в отделение консалтинга по менеджменту компании Coopers & Lybrand. Меня поразило, как часто клиенты даже не пытались определить важный для них показатель, который мог заставить их иначе оценить инвестиционный проект или изменить свою политику, считая его абсолютно неизмеримым количественно. Курсы лекций по статистике и методам количественного анализа были еще свежи в моей памяти, и в ряде случаев, когда какой-то объект называли не поддающимся измерению, я вспоминал конкретный пример его успешной оценки. Я стал подозревать все подобные утверждения в поспешности и проводить исследования, чтобы подтвердить или опровергнуть свое предположение. И каждый раз обнаруживалось: якобы не поддающийся измерению параметр уже оценен учеными или практиками, пускай и в других областях.

Одновременно я заметил, что ни в одной из книг по количественным методам анализа авторы не пытаются доказать, что измерить можно абсолютно все. Не стараются они и сделать этот материал доступным для тех, кто в нем действительно нуждается. Обычно авторы подобных книг исходят из того, что читатель уже убежден в измеримости какого-то явления, нужно только вооружить его подходящим алгоритмом. А еще, по их мнению, читатели стремятся к такому уровню строгости, который будет достаточен для публикаций в научном журнале, им не нужно простое снижение неопределенности при принятии какого-нибудь важного решения способом, доступным для понимания нестатистиков.

В 1995 г. после нескольких лет наблюдений я пришел к выводу, что существует спрос на способы измерения, адаптированные к потребностям менеджеров. Чтобы решить эту задачу, мне пришлось объединить методы, используемые в разных областях. Многочисленные проекты, связанные с измерениями, над которыми я работал с 1995 г., позволили усовершенствовать данный подход. При этом выяснилось, что не только считающееся неизмеримым таковым не является, но и что наиболее «неподатливые» нематериальные объекты нередко можно оценить на удивление простыми приемами. Пришло время опровергнуть стойкое убеждение, что некоторые важные показатели рассчитать невозможно.

В процессе работы над книгой мне казалось, что я раскрываю большой секрет, и как только он станет всеобщим достоянием, многое

изменится. Я даже воображал, что это станет своего рода небольшой «научной революцией» для менеджеров, дальней родственницей революции «научного управления», произведенной век назад Фредериком Тейлором. В самом деле, для менеджеров XXI века этот материал должен оказаться даже более актуальным, чем в свое время методы Тейлора. Если научное управление имело своей целью прежде всего оптимизацию трудовых процессов, то сегодня нам нужно оптимизировать измерения, чтобы принимать лучшие управленческие решения. Формальные методы количественной оценки подобных вещей, которые менеджеры обычно игнорируют, едва достигли в своем развитии уровня алхимии, а нам необходимо поднять их до уровня химии и физики.

У нас с издателем было несколько вариантов названия этой книги. Все они начинались с «Как измерить все, что угодно», но далее не обязательно следовало «Оценка стоимости нематериального в бизнесе». Я веду семинар, который называется «Как измерить все, что угодно, но только то, что вам нужно». Поскольку в данной книге рассказывается и о методе расчета экономической стоимости измерений (с тем, чтобы понять, что заслуживает измерения, а что нет), нынешнее название представляется особенно удачным. Рассматривался вариант «Как измерить все, что угодно: оценка нематериального в бизнесе, государственном управлении и технике», так как в книге приводится много примеров из подобной практики. Но мы решили, что название «Как измерить все, что угодно. Оценка стоимости нематериального в бизнесе» лучше привлекает нужную аудиторию и передает основную идею и содержание книги.

Книга состоит из четырех частей. Главы и части следует читать по порядку, поскольку главы трех первых частей основаны на предыдущем материале. В части I доказывается, что измерению поддается абсолютно все, и приводятся примеры, призванные вдохновить читателя на количественную оценку параметров даже тогда, когда это кажется невероятным. В ней излагается основная идея книги, и если вы решите прочитать лишь одну часть, то советую выбрать эту. Для правильного понимания остального материала крайне важно ознакомиться, в частности, с предлагаемым здесь определением измерения.

В начале части II подробнее говорится о том, как измерить такие, например, явления, как неопределенность, риск и ценность информации. Эти измерения интересны не только сами по себе; в рамках предлагаемого мной подхода они должны предварять все прочие оценки. Читатель научится оценивать субъективную неопределенность путем «калиброванной оценки вероятностей» и использовать эту информацию для расчета риска и стоимости дополнительных измерений. Крайне важно, чтобы вы усвоили этот материал, прежде чем перейти к следующей части.

Часть III посвящена различным методам наблюдения, позволяющим снизить неопределенность, в том числе случайной выборке и проведению управляемых экспериментов. В ней рассказывается о нескольких методах упрощенной аппроксимации, применяемых, когда возможно. Также в ней обсуждаются способы повышения точности измерений, основанные на рассмотрении результата каждого наблюдения в качестве корректировки с незначительным снижением предыдущего уровня неопределенности. Материал, возможно, знакомый читателю по университетскому курсу статистики, изложен так, чтобы показать его связь с предметом обсуждения из предыдущей части. Наиболее сложный материал по регрессионному моделированию и управляемым экспериментам читатель может в зависимости от своих потребностей либо пропустить, либо изучить более подробно.

В части IV рассказывается об интересных решениях различных задач по измерению и приводятся конкретные примеры из практики. В ней обсуждаются методы оценки таких вещей, как предпочтения, моральные ценности, гибкость или качество, и описываются некоторые новые или малоизвестные инструменты измерения, в том числе суждения калиброванных специалистов и даже Интернет. Здесь резюмируются и сводятся воедино подходы, рассмотренные в других частях книги, и подробно обсуждаются примеры из практики.

В главе I я предлагаю читателям проделать одно полезное упражнение и заранее хочу упомянуть о нем сейчас. Составьте перечень задач по измерению, с которыми вы столкнулись в своей профессиональной или частной жизни, и попытайтесь найти в книге способы их решения. Если эти оценки помогут принять хоть сколько-нибудь важное решение, то время и деньги, потраченные вами на эту книгу, окупятся сторицей.



От автора

Появлению этой книги способствовали многие, поделившись со мной своими соображениями или рассказав об интересных решениях задач по измерению. Не придерживаясь особого порядка, хочу выразить свою благодарность следующим людям:

Скипу Бейли, Тому Бейкуэллу, Биллу Биверу, Джеффу Брайану, Робину Дейвесу, Фримону Дайсону, Марку Дею, Робу Донату, Рею Гилберту, Дэвиду Греттеру, Терри Каннеману, Арту Койнзу, Бетти Коулсон, Мери Лундз, Рику Мелберту, Чаку Маккею, Майку Макши, Барри Назбауму, Риду Оглиеру, Эндрю Освальду, Пату Планкетту, Аркалгуду Рамапрасаду, Джеймсу Рэнди, Линде Роза, Джею Эдварду Руссо, Сэму Сэвиджу, Джеку Стеннеру, Питеру Типпетту, Луису Торресу, Тодду Уилсону, Робину Хансену, Джулиане Хейл, Джеймсу Хэммитту, Лео Чампиону, Генри Шафферу, Питеру Шею, Доминику Шилту, Эмили Серван-Шрайбер, Джорджу Эберштадту, Рею Эпичу и Гарри Эпштейну.

Я особенно благодарен Доминику Шилту из Riverpoint Group LLC, поверившему в перспективность данного подхода еще в 1995 г. и с тех пор неизменно оказывавшему мне поддержку.

Посвящаю эту книгу тем, кто неизменно вдохновляет меня на все, что я делаю: моей жене Джанет и нашим детям Ивену, Маделин и Стивену, сулящим со временем превратиться в людей с разносторонними интересами.

Хочу также посвятить свою книгу американским военно-служащим — мужчинам и женщинам, со многими из которых я знаком. Я уже давно не состою в национальной гвардии сухопутных войск, но надеюсь, что мои усилия по совершенствованию боевого материально-технического снабжения морской пехоты США повысили ее эффективность и безопасность.

Измерение:
решение существует



Нематериальное и проблема его измерения

Если вы можете измерить то, о чем говорите, и выразить это в цифрах, значит, вы что-то об этом предмете знаете. Но если вы не можете выразить это количественно, ваши знания крайне ограничены и неудовлетворительны. Возможно, это начальный этап, но еще не подлинно научное знание.

Лорд Кельвин

Измерить можно все, что угодно. Когда какой-либо объект или явление удастся наблюдать тем или иным образом, значит, существует метод для его измерения. Каким бы приблизительным ни было это измерение, оно все равно будет им, если расскажет вам больше, чем вы знали до сих пор. А то, что чаще всего считается не поддающимся измерению, практически всегда можно оценить сравнительно простым способом.

Как показывает название этой книги, мы будем обсуждать способы определения стоимости вещей, которые в бизнесе нередко называют нематериальными. Существуют два основных толкования слова «нематериальное». Во-первых, нематериальными, по традиции, считают вещи, буквально не являющиеся материальными (телесными, осязаемыми), но, тем не менее, поддающиеся измерению. Хорошими примерами объектов, которые нельзя потрогать, но можно измерить, служат время, бюджет, право собственности на патент и т. д. Сегодня уже сложилась целая индустрия оценки нематериальных активов, таких как авторское право и торговая марка. Но со временем слово «нематери-

альный» стали употреблять и в значении «не поддающийся никакому измерению — прямому или косвенному». И вот я утверждаю, что объектов, нематериальных в этом смысле слова, вообще не существует.

Вам, конечно, доводилось слышать в своей компании о нематериальных активах, то есть объектах, предположительно абсолютно неизмеряемых. Убежденность в правильности данной идеи порой настолько сильна, что никто и не пытается наблюдать за этими объектами — а ведь они могут рассказать вам кое-что удивительное. Возможно, вы в своей практике уже сталкивались со следующими примерами из реальной жизни:

- «гибкостью», необходимой в создании новых продуктов;
- риском неудачи при реализации проекта по информационным технологиям (ИТ);
- эффектом, который новая политика государства в области защиты окружающей среды оказывает на здоровье населения;
- эффективностью научных исследований;
- стоимостью информации;
- вероятностью того, что та или иная политическая партия победит в борьбе за Белый дом;
- качеством;
- мнением общественности.

Любая из этих проблем может оказаться весьма актуальной для организации, собирающейся принять важное решение. Иногда речь идет даже о единичном, но очень серьезном эффекте от дорогостоящей новой программы, принятой компанией или государственными органами. И все же многие организации, считая «нематериальное» не поддающимся измерению, принимают далеко не такие информационно обоснованные решения, как могли бы.

Эту ситуацию мне не раз доводилось наблюдать в так называемых управляющих комитетах, которые рассматривают предлагаемые проекты и решают, какие из них следует принять, а какие — отклонить. Нередко обсуждаемые проекты так или иначе связаны с ИТ. В ряде случаев управляющие комитеты категорически отвергали любые инвестиционные предложения, когда ожидаемые выгоды были, скорее, моральными. Такие важные факторы, как расширение рекламы, создаваемой потребителями, снижение стратегического риска или позиционирование премиум-бренда, при их оценке не учитывались как «не поддающиеся измерению». При этом проект отвергался вовсе не потому, что его автор не рассчитал ожидаемую выгоду (веская причина для отказа), а вследствие убежденности в невозможности этого расчета. Соответственно многие важнейшие стратегические предложения были

пропущены в пользу слабых, но низкозатратных идей только потому, что в одних случаях методы оценки ожидаемых эффектов были всем известны, а в других — нет.

Между тем некоторые организации сумели успешно проанализировать и количественно оценить все перечисленные выше явления, используя для этого совсем не такие сложные методы, как можно было ожидать. Цель данной книги состоит в том, чтобы продемонстрировать два положения:

- 1) нематериальные явления и факторы, представляющиеся абсолютно неизмеряемыми, измерить можно;
- 2) подобные измерения можно произвести экономически обоснованным способом.

Чтобы полностью охватить тему, книга с названием «Как измерить все, что угодно», наверное, должна быть многотомной. Я отнюдь не ставил перед собой задачу описать все методы измерения, используемые в областях естествознания или экономической науке, особенно если они уже хорошо разработаны. Специалисты в этих областях располагают приемами, позволяющими решать самые разнообразные интересующие их задачи, и уже гораздо менее склонны навешивать ярлык «нематериальное» на занимающие их проблемы. В центре внимания книги — методы измерения, знать которые важно (и даже крайне важно!), чтобы принимать серьезные деловые решения. Речь идет, прежде всего, о таких способах, которые позволяют оценить то, что не поддается стандартным и принятым подходам.

Предлагаемая вашему вниманию книга развеивает ряд распространенных заблуждений в отношении измерения нематериального и описывает «универсальный подход» к его оценке, подтверждаемый интересными методами решения конкретных задач. Я постарался привести в ней «вдохновляющие» примеры того, как на практике были оценены самые неизмеряемые объекты, какие мне удалось найти.

Не в ущерб содержанию в книге дается простое объяснение некоторым понятиям статистики, имеющим отношение к измерениям и кажущимся доступными только посвященным. Там, где возможно, математические выкладки приводились в виде более понятных формул, графиков и таблиц. Некоторые из этих методов настолько проще преподаваемых во вводном университетском курсе статистики, что читатель сумеет преодолеть многочисленные фобии, которые многим внушают количественные методы измерения. Никакой особой математической подготовки вам не потребуется. Все, что понадобится, — это способность четко формулировать проблемы.

Рекомендую читателям посетить веб-сайт книги:
www.howtomeasureanything.com. Там вы найдете целую библиотеку

электронных таблиц для проведения некоторых более точных расчетов из описанных в книге примеров, загружаемых из Сети. Там же находятся дополнительные учебные пособия, примеры и форум для обсуждения вопросов по книге или вообще задач, связанных с измерением. Сайт, кроме того, позволяет мне дискутировать на тему новых способов и методик, не существовавших во время работы над книгой.

Советую также читателям проделать одно полезное упражнение. Составьте и по мере прочтения этой книги пополняйте список того, что кажется не поддающимся измерению или, по крайней мере, способы измерения чего вам неизвестны. Надеюсь, что, прочитав эту книгу, вы сумеете количественно оценить любой объект из вашего списка.



Интуитивное умение измерять все: Эратосфен, Энрико и Эмили

Развить в себе умение измерять — совсем не простое занятие, поэтому поставить такую цель легче, если видишь свет в конце туннеля. Нам нужны лучшие примеры людей, интуитивно решавших задачи по измерению и при этом нередко находивших на удивление простые способы для этого. К счастью, история знала немало личностей, продемонстрировавших такое поразительное умение, — людей одновременно и вдохновленных, и вдохновляющих. Показательно, однако, то, что многие из них не имели отношения к бизнесу. На самом деле, в поисках методов, применимых в бизнесе, книга широко заимствует материал и примеры из других областей человеческой деятельности.

Вот лишь несколько человек, которые, не занимаясь измерениями в бизнесе, могут, тем не менее, показать бизнесменам, что такое интуитивная способность к количественным исследованиям.

- Один древний грек сумел определить длину земной окружности, измерив полуденные тени в разных городах и проведя простейшие геометрические расчеты.
- Один лауреат Нобелевской премии по физике учил своих студентов измерять, показывая, как можно оценить число настройщиков пианино в Чикаго.

- Одна девятилетняя девочка поставила эксперимент, положивший конец популярности метода «бесконтактного массажа» из нетрадиционной медицины, и два года спустя стала самым молодым автором, когда-либо публиковавшимся в «Journal of the American Medical Association».

Возможно, вам приходилось слышать об этих людях или, по крайней мере, о ком-то из них. Но даже если вы что-то припоминаете, такие сведения стоит освежить и посмотреть, что же их объединяет. Эти люди не были знакомы друг с другом и даже жили в разное время, но все они проявили способность сформулировать задачу по измерению и быстро сделать несложные исследования, давшие поразительные результаты. Они сумели в короткое время определить неизвестные величины с помощью простых наблюдений. Очень важно сравнить их подход с тем, что вы обычно наблюдаете в деловой обстановке. Те, о ком пойдет речь, — это реальные люди, которых звали Эратосфен, Энрико и Эмили.

КАК ОДИН ДРЕВНИЙ ГРЕК ОПРЕДЕЛИЛ РАЗМЕРЫ ЗЕМНОГО ШАРА

Наш первый наставник в измерениях сделал то, что в его время многие наверняка считали невозможным. Древний грек по имени Эратосфен (примерно 276—194 гг. до н. э.) был, насколько это известно, первым, кому удалось измерить длину окружности Земли. Если его имя кажется вам знакомым, то, скорее всего, потому, что оно упоминается во многих университетских учебниках по геометрии и тригонометрии.

Эратосфен не пользовался точным геодезическим оборудованием, и, конечно, у него не было лазеров и спутников. Он и не помышлял о кругосветном путешествии, которое могло не только оказаться рискованным, но и затянуться на долгие годы. Взамен он вычитал в одной книге из Александрийской библиотеки, что дно глубокого колодца в Сиене, городе в Южном Египте, целиком освещается полуденным солнцем только раз в году. Это означало, что в полдень солнце должно находиться прямо над колодцем. Но Эратосфен также заметил, что в полдень вертикальные предметы в Александрии, которая находится почти прямо на север от Сиены, отбрасывают тень, то есть в другом городе в это же время солнечные лучи падают на Землю под несколько иным углом. Эратосфен догадался, что данную информацию можно использовать для оценки кривизны Земли.

Он установил, что в то время года полуденные тени в Александрии образуют угол, соответствующий дуге в $\frac{1}{50}$ окружности. Поэтому

если расстояние между Сиеной и Александрией было $1/50$ окружности, то вся длина окружности Земли должна превышать это расстояние в 50 раз. Более поздние попытки повторить расчеты Эратосфена отличаются друг от друга только тем, какие были взяты углы и расстояния между древними городами, а также результатами перевода древних единиц измерения в современные, но обычно выходит, что погрешность оценки Эратосфена составляет $\pm 3\%$ ¹. Расчеты этого ученого стали огромным шагом вперед по сравнению с имевшимися ранее знаниями, а ошибка его измерения меньше допущенной нашими учеными всего несколько десятилетий назад при оценке размера и возраста Вселенной. Погрешность, допущенная при аналогичных расчетах Колумбом 1700 лет спустя (который, очевидно, не знал или не учел результата Эратосфена), составила 25% (по этой причине мореплаватель думал, что находится вблизи Индии, а не у другого колоссального континента). Фактически, ошибка Эратосфена была исправлена только через 300 лет после плавания Колумба. Тогда результат Эратосфена сумели, наконец, уточнить два француза, располагавшие самыми точными геодезическими приборами, имевшимися во Франции в конце XVIII века, значительными финансовыми ресурсами и множеством помощников².

Вот вам и урок для бизнеса: Эратосфен произвел измерения, казавшиеся невозможными, остроумно использовав данные простейших наблюдений. Когда я спрашиваю слушателей своих семинаров по оценке и анализу риска, как они определили бы длину земной окружности, не пользуясь современными инструментами, они обычно предлагают какой-нибудь сложный путь, например кругосветное плавание. Но ведь Эратосфен выполнил свои расчеты, *не покидая окрестностей библиотеки*. На поставленный им вопрос могли ответить гораздо более сложные исследования, но его оценка основывалась на других, простых наблюдениях. Ученый извлек всю возможную информацию из тех немногих фактов, которые мог проверить сам, не «зацикливаясь» на том, что эта задача решается только сложным путем.

КАК ОПРЕДЕЛИТЬ НЕИЗВЕСТНУЮ: БЕРИТЕ ПРИМЕР С ФЕРМИ

Другой не имеющий отношения к бизнесу человек, способный вдохновить предпринимателей на измерения, — это Энрико Ферми (1901–1954), физик, получивший Нобелевскую премию в 1938 г. У него был настоящий талант к интуитивным измерениям, иногда казавшимся даже случайными. Как-то он продемонстрировал его при испытании атомной бомбы на полигоне Тринити 16 июля 1945 г., где вместе с другими учеными-атомщиками наблюдал за взрывной волной из ба-

зового лагеря. Пока другие окончательно настраивали приборы для измерения мощности взрыва, Ферми разорвал на мелкие кусочки страничку из своего блокнота. Когда после взрыва подул сильный ветер, он подбросил эти кусочки в воздух и заметил, куда они упали (обрывки, улетевшие дальше всех, должны были показать пик давления волны). Ферми пришел к выводу, что мощность взрывной волны превысила 10 килотонн. И эта информация оказалась очень важной, так как другим наблюдателям нижний предел данного параметра был неизвестен. После длительного анализа показаний приборов мощность взрывной волны была в конце концов оценена в 18,6 килотонн. Как и Эратосфен, Ферми сумел определить требуемый показатель, проведя одно простое наблюдение — за рассеиванием обрывков бумаги по ветру.

О важности быстрых оценок Ферми знал на протяжении всей своей карьеры. Он славился тем, что учил студентов навыкам приблизительных расчетов самых фантастических величин, о которых те, казалось бы, не могли иметь никакого представления. Самым известным примером такого «вопроса Ферми» является определение числа настройщиков пианино в Чикаго. Студенты (будущие ученые и инженеры) начали с того, что у них нет для этого расчета никаких данных. Конечно, можно было просто пересчитать всех настройщиков, прочитав объявления, справившись в каком-нибудь агентстве, выдающем лицензии на такие услуги, и т.д. Но Ферми пытался научить своих студентов решать задачи и тогда, когда проверить результат будет не так просто. Ему хотелось, чтобы они поняли, что все-таки *знают что-то* об искомой величине.

Для начала Ферми попросил определить другие имеющие отношение к пианино и их настройщикам показатели — тоже неизвестные, но более легкие для оценки. Это были численность населения Чикаго (составлявшая в 1930–1950-х годах чуть более 3 млн человек), среднее число человек в одной семье (два или три), процент семей, регулярно пользующихся услугами настройщиков пианино (максимально — каждая десятая, минимально — каждая тридцатая семья), требуемая частота настройки (в среднем, вероятно, не менее раза в год), число пианино, настраиваемых настройщиком за день (четыре или пять инструментов с учетом затрат времени на дорогу), а также число рабочих дней настройщика в году (скажем, 250). Эти данные позволили рассчитать число настройщиков по следующей формуле:

$$\begin{aligned} & \text{Число настройщиков пианино в Чикаго} = \\ &= (\text{Численность населения} / \text{Число членов одной семьи}) \times \\ & \times \text{Процент семей, пользующихся услугами настройщиков} \times \\ & \times \text{Число настроек в году} / (\text{Число пианино, настраиваемых одним} \\ & \quad \text{настройщиком за день} \times \text{Число рабочих дней в году}). \end{aligned}$$

В зависимости от цифр, подставляемых в это уравнение, вы получите ответ в интервале 20—200, скорее всего, правильный ответ составлял примерно 50 человек. Когда эту цифру сравнивали с реальной (которую Ферми мог узнать из телефонного справочника), она всегда была ближе к реальной, чем думали студенты. Полученный интервал значений выглядит слишком широким, но разве это не огромный шаг вперед по сравнению с позицией «неужели это вообще можно определить?», которую студенты занимали поначалу?

Данный подход позволял людям, производившим расчеты, понять, откуда берется неопределенность. Какие переменные характеризовались наибольшей неопределенностью — процент семей, регулярно пользующихся услугами настройщиков пианино, частота настроек, число инструментов, которые можно настроить за день, или что-то еще? Самый крупный источник неопределенности указывал на то, какие измерения позволят максимально снизить ее.

Поиск ответа на «вопрос Ферми» не предполагает проведения новых наблюдений и поэтому не может безоговорочно считаться измерением. Скорее, это оценка того, что вам уже известно о проблеме, способом, позволяющим несколько приблизиться к цели. Вот еще один урок для бизнесмена — не считайте неопределенность неустранимой и не поддающейся анализу. Вместо того чтобы впадать в уныние по поводу своего незнания, спросите себя: а что же вы все-таки знаете о проблеме? Как мы увидим позже, оценка имеющейся количественной информации о предмете — очень важный этап измерения явлений, которые выглядят неизмеряемыми.

«ВОПРОСЫ ФЕРМИ» ДЛЯ НОВОГО ПРЕДПРИЯТИЯ

Чак Макей из Wizard of Ads всячески поощряет компании использовать «вопросы Ферми» для оценки размера своего рынка в том или ином районе. Недавно один страховой агент попросил Чака дать совет, стоит ли его компании открывать офис в Уичита-Фоллз (штат Техас), где до сих пор у нее не было представительства. Будет ли на данном рынке спрос на услуги еще одного страховщика? Чтобы проверить реализуемость плана, Макей покопался в Интернете в поисках ответов на ряд «вопросов Ферми». Как и тот, Макей начал с проблемы численности населения.

По данным City-Data.com, жители Уичита-Фоллз владели 62 172 автомобилями. А согласно Институту страховой информации (Insurance Information Institute), средняя годовая автомобильная страховая премия в штате Техас составляла 837,40 дол. Макей предположил, что почти все машины застрахованы, поскольку это обязательное требование, поэтому общая выручка от страхования составляла ежегодно 52 062 833 дол. Агент узнал, что средняя коммиссионная ставка составляет 12%, так что все годовое коммиссионное вознаграждение составляло 6 247 540 дол. По сведениям Switchboard.com, в городе

действовали 38 страховых агентств, и примерно столько же их было по информации Yellowbook.com. Если разделить все комиссионное вознаграждение на 38 агентств, то окажется, что годовые комиссионные одного из них составляют в среднем 164 409 дол.

Рынок, по всей видимости, был уже достаточно насыщен, поскольку, по сведениям City-Data.com, численность населения Уичита-Фоллз сократилась со 104 197 человек в 2000 г. до 99 846 человек в 2005 г. Кроме того, на данном рынке уже работало несколько крупных фирм, поэтому доходы нового агентства были бы еще меньше — и все это без учета накладных расходов.

Вывод Макея: скорее всего, новое агентство в этом городе вряд ли будет прибыльным, поэтому от плана следует отказаться.

ЭКСПЕРИМЕНТЫ — НЕ ТОЛЬКО ДЛЯ ВЗРОСЛЫХ

Еще одним человеком, обладавшим, по всей видимости, настоящим талантом к измерению, была Эмили Роза. Свой отчет о проведенных измерениях она опубликовала в «Journal of American Medical Association» (JAMA), но у нее не было ни ученой степени, ни даже диплома о высшем образовании. Когда девочка провела свою оценку, ей было всего девять лет, она училась в четвертом классе и работала над проектом для школьной ярмарки научных идей. Ей исполнилось 11, когда исследование было опубликовано, и это сделало Эмили самым молодым автором, когда-либо печатавшимся в престижном медицинском, а возможно, и вообще в научном журнале.

В 1996 г. Эмили увидела, как ее мать Линда смотрела видеофильм о «бесконтактном массаже» — довольно спорном методе лечения путем воздействия на «энергетические поля» пациента. На экране сцена: пациент лежит неподвижно, а врач водит руками в нескольких сантиметрах от его тела, пытаясь обнаружить и устранить «нежелательные энергии», якобы вызывающие различные заболевания. Эмили предложила своей матери экспериментально проверить это утверждение. Линда, которая была медсестрой и членом Национального совета по борьбе с мошенничеством в сфере здравоохранения (National Council Against Health Fraud, NCAHF), дала ей кое-какие советы.

Сначала Эмили привлекла к своему эксперименту для школьной ярмарки научных идей 15 врачей, практикующих бесконтактный массаж. Опыт состоял в следующем: Эмили и врач сидели за столом друг против друга. Их разделял картонный экран. В нижней части экрана имелись отверстия, через которые целитель просовывал руки ладонями вверх. Эмили подбрасывала монетку и в зависимости от того, как она

падала, протягивала свою руку так, чтобы та оказывалась сантиметров на 10 выше правой или левой руки медика. Врачи, не видя девочку за экраном, должны были определить, над какой их рукой — правой или левой — она держит свою ладонь, только на основании собственных ощущений от ее энергетического поля. О результатах Эмили доложила на ярмарке научных идей и получила в награду голубую ленточку — впрочем, как и все остальные участники.

Линда рассказала об эксперименте Эмили доктору Стивену Барретту, которого она знала по работе в NCAHF. Барретт, заинтересовавшийся и простотой метода проведения опыта, и его первыми результатами, упомянул о нем продюсерам телешоу «Scientific American Frontiers» («Передовые рубежи американской науки»), которое шло по каналу государственного телевидения (Public Broadcasting System). В 1997 г. продюсеры сняли сюжет о методе Эмили, а девочка привлекла к участию в шоу еще 13 целителей, так что всего их стало 28.

Все эти врачи сделали в общей сложности 280 попыток ощутить энергетическое поле Эмили. Правильно определить положение руки девочки им удалось только в 44% случаев. А по теории вероятности они должны были угадать в 50% случаев с 95-процентным доверительным интервалом (confidence interval, CI) и погрешностью $\pm 16\%$ (если вы подбросите 280 монет, есть 95-процентная вероятность того, что орел выпадет в 44–66% всех случаев). Так что врачам, по всей видимости, немного не повезло (поскольку полученные ими данные оказались ближе к нижней границе интервала), но эти результаты не вышли из пределов, объясняемых одной только теорией вероятности. Иными словами, люди, не владеющие методом бесконтактного массажа (вы или я), могли бы просто попробовать угадать и добились бы таких же или даже лучших результатов, чем врачи-специалисты, участвовавшие в эксперименте.

Получив информацию, Линда и Эмили решили, что ее, наверное, стоит опубликовать. В апреле 1998 г. Эмили, которой тогда было всего 11 лет, напечатала экспериментальные данные в JAMA. После этого ее имя занесли в книгу рекордов Гиннеса как самого молодого автора, когда-либо печатавшегося в крупном научном журнале, и девочка получила в награду 1000 дол. от Образовательного фонда Джеймса Рэнди (James Randy Educational Foundation, JREF).

Джеймс Рэнди, бывший фокусник и известный скептик, основал фонд для научной проверки утверждений о наличии паранормальных способностей. Он предложил в качестве приза миллион долларов любому, кто научно докажет наличие у себя экстрасенсорного восприятия (extrasensory perception, ESP), ясновидения, способности находить воду или минералы с помощью «волшебной лозы» и т. п. Рэнди не нравится, когда говорят, что он занимается «развенчанием» притязаний на паранормальные способности, поскольку он просто тестирует

эти заявления объективными научными методами. Но поскольку сотни претендентов на миллион так и не сумели получить приз, не пройдя простейшие научные испытания, до сих пор подобные утверждения только опровергались. Рэнди заинтересовался методом бесконтактного массажа еще до публикации данных Эмили, и сам пытался его проверить. Но, в отличие от Эмили, ему удалось найти только одного целителя, согласившегося на этот эксперимент, кстати, прошедший для врача неудачно.

После публикации опытных результатов сторонники метода бесконтактного массажа стали говорить, что методика проведения эксперимента не свободна от недостатков, а сам он ничего не доказывает. Одни указывали на то, что энергетическое поле ощущается на расстоянии 2–7, а вовсе не 10–12 см, как в эксперименте Эмили³, другие — на то, что энергетическое поле подвижно, а не статично, поэтому обнаружить неподвижную руку Эмили было объективно невозможно (хотя во время «лечения» пациенты обычно лежали спокойно)⁴. Рэнди все это совсем не удивило. «Люди всегда находят оправдания потом, — сказал он, — но ведь каждого целителя до начала испытания спрашивали, согласен ли он с условиями. Все не только согласилось, но и выразили уверенность, что пройдут тест успешно». Конечно, лучшим опровержением результата Эмили стал бы другой управляемый эксперимент, который окончательно доказал бы действенность бесконтактного массажа. Но провести такой опыт никто не предложил.

Рэнди так часто сталкивался с запоздалыми оправданиями участников подобных экспериментов, что стал просить их предварительно подписывать специальный официально заверяемый документ. В нем говорилось, что испытуемый согласен с условиями опыта, что впоследствии не будет подвергать его критике и что рассчитывает на успех в предлагаемых условиях. Одновременно он вручал каждому запечатанный конверт. После эксперимента, когда человек пытался оспорить его результаты, ссылаясь на плохую методику, Рэнди просил открыть конверт. В нем было письмо, в котором говорилось буквально следующее: «Вы согласились с тем, что данные условия оптимальны, и обещали не искать отговорок после проведения опыта. А теперь вы все же это делаете». «Испытуемым было исключительно неприятно», — замечал Рэнди.

Бизнес может извлечь из этого несколько уроков. Во-первых, даже такие кажущиеся эфемерными вещи, как расширение полномочий сотрудников, креативность или согласованность со стратегией, должны иметь поддающиеся обнаружению последствия, если только они вообще имеют какое-то значение. Я вовсе не утверждаю, что эти вещи «паранормальны», но здесь действуют те же правила.

А во-вторых, эксперимент Эмили продемонстрировал эффективность простых методов, традиционно используемых в научных иссле-

дованиях, таких как эксперимент в контролируемых условиях, выборка (даже если она невелика), рандомизация* и испытания «вслепую», позволяющих избежать необъективности со стороны испытуемого или исследователя. Эти простые приемы могут использоваться как по отдельности, так и в любых сочетаниях, давая возможность наблюдать и измерять самые разные явления.

Кроме того, Эмили доказала, что простейшие эксперименты, не требующие больших расходов, может проводить даже ребенок (Линда Роза сказала, что затраты на эксперимент составили 10 дол.). Интересно отметить, что Эмили могла провести гораздо более сложные клинические испытания действенности бесконтактного массажа, используя тестовые и контрольные группы для проверки того, насколько метод улучшает состояние здоровья пациентов. Но ей не пришлось этого делать, потому что она задала более фундаментальный вопрос. Если целители могут делать то, о чем говорят, то, рассудила Эмили, они должны, *по крайней мере, чувствовать энергетическое поле*. Если же они его не чувствуют (а это основная исходная посылка выгоды, приписываемой методу), значит, все, что связано с бесконтактным массажем, вызывает большие сомнения. Девочка могла найти способ потратить намного больше, если бы, например, располагала обычным бюджетом для проведения непродолжительного медицинского исследования. Но она и так выяснила все, что хотела, и с более чем достаточной точностью. Для сравнения: о скольких методах оценки показателей вашей деятельности можно написать в научном журнале?

Пример Эмили доказывает, что полезный результат можно получить и простыми методами. Временами мне приходится слышать советы не использовать такие «продвинутые» методы измерения, как контролируемые эксперименты, поскольку высшее руководство этого не поймет. По-видимому, авторы советов считают, что правило Дилберта⁵ (согласно которому повышение всегда получает самый некомпетентный сотрудник) действительно выполняется в случае топ-менеджеров. По своему опыту могу сказать, что обычно руководители высшего звена всегда все понимают, если вы хорошо объясняете.

— Эмили, объясни им это, пожалуйста.

ПРИМЕР. ИНФОРМАЦИОННАЯ ИНФРАСТРУКТУРА КОМПАНИИ MITRE

Интересным примером того, как организация может оценить свои нематериальные активы, убедившись предварительно, существуют ли они вообще, может служить информационная инфраструктура Mitre (Mitre Information

* Внесение случайности. — *Примеч. редактора.*

Infrastructure, MII). Эту систему разработала в конце 1990-х годов Mitre Corporation — некоммерческая компания, предоставляющая федеральным органам власти консалтинговые услуги в области инжиниринга систем и информационных технологий. MII была корпоративной базой знаний, позволявшей наладить сотрудничество отделов, ранее полностью изолированных друг от друга.

В 2000 г. в журнале «CIO Magazine» была опубликована статья о MII. В этом журнале было принято, чтобы всю трудную работу по подготовке такой статьи готовил штатный сотрудник, а затем независимый эксперт высказывал свое мнение в колонке под названием «Критический анализ». Когда публиковались статьи об измерениях, оценке риска и подобных вопросах, в роли такого эксперта нередко выступал я. Так и произошло в случае с MII.

Колонка «Критический анализ» должна была служить неким противовесом, поскольку, рассказывая о своих новых программах, компании обычно не жалеют розовых красок. В статье было процитировано следующее высказывание Эль Грассо, тогдашнего руководителя информационной службы Mitre Corporation: «Наши важнейшие достижения (качество и инновационность решений, которые становятся реализуемыми, когда вы овладеваете всей информацией) легко не измеришь». Однако я в своей колонке предложил простой способ измерения и качества, и инновационности:

Если MII действительно улучшает качество предоставляемых услуг, то оно должно влиять на восприятие этих услуг клиентами и, в конечном счете, на доходы⁶. Просто попросите случайную выборку клиентов проранжировать качество каких-либо услуг до и после создания MII (так, чтобы они не знали, за какой период оценивают данный параметр) и узнайте, заставило ли их повысившееся качество приобретать у Mitre больше услуг⁷.

Таким образом, я предложил, чтобы компания Mitre, как в свое время Эмили, задала себе не тот вопрос, с которого, наверное, начал бы руководитель информационной службы, а более простой, но связанный с ним. Если качество и инновационность действительно возросли, то разве эта разница не должна, по крайней мере, чувствоваться? Если те, кому положено об этом судить, то есть потребители, в ходе испытания вслепую не смогут заявить, что исследования после создания MII стали качественнее или инновационнее, чем до нее, то это будет означать, что данная информационная структура никак не влияет на удовлетворенность клиентов, а следовательно, и на доходы. А вот если они заметят отличие, то можно будет поставить следующий вопрос: увеличились ли доходы настолько, чтобы оправдать инвестиции, составившие к 2000 г. 7 млн дол.? Если такие преимущества MII, как качество, инновационность и любые другие, невозможно обнаружить, то они не имеют значения. Нынешние и бывшие сотрудники Mitre рассказывали, что моя колонка наделала много шума. Однако никто из них не слышал, чтобы кто-либо в компании действительно пытался оценить качество и инновационность. Как вы помните, руководитель информационной службы сказал, что эти достижения MII были *важнейшими*. И вот эти преимущества так никто и не определил количественно.

ЧЕМУ НАС УЧАТ ЭРАТОСФЕН, ЭНРИКО И ЭМИЛИ

Все эти люди — Эратосфен, Энрико и Эмили — демонстрируют нечто кардинально отличающееся от того, с чем мы обычно сталкиваемся в бизнесе. Руководители часто говорят: «Ни о чем подобном мы не могли бы даже догадываться». Они заранее пасуют перед неопределенностью. Вместо того чтобы, по меньшей мере, попытаться провести измерения, они бездействуют, обескураженные кажущейся невозможностью устранить ее. Ферми в подобном случае мог бы сказать: «Да, вы многого не знаете, но *что-то* же вы все-таки знаете?»

Иные менеджеры возражают: «Чтобы определить этот показатель, нужно потратить миллионы». В итоге они предпочитают не проводить и менее масштабные (с малыми затратами) исследования, потому что их погрешность обычно выше, чем у дорогих комплексных научных работ. Между тем, даже небольшое снижение неопределенности может принести миллионы в зависимости от важности решения, принятию которого оно способствует, и от частоты принятия подобных решений. Эратосфен и Эмили могли бы сказать, что полезные наблюдения (даже при небольшом бюджете) могут сообщить то, чего вы не знали ранее, если только избавиться от пораженческих настроений и подойти к делу чуть более креативно.

Примеры Эратосфена, Энрико и Эмили учат нас разным вещам. Эратосфен не мог рассчитать ошибку своей оценки, поскольку 2000 лет назад статистических методов расчета погрешности еще не существовало. Однако будь он знаком с одним из таких методов, то знал бы, что эта ошибка объясняется неточностью оценки расстояний между городами и углов наклона теней. Представление об измерении как об уменьшении погрешности — главная идея данной книги.


Пример Энрико Ферми имеет для нас другое значение. Поскольку он был лауреатом Нобелевской премии, у нас есть все основания считать его выдающимся физиком, причем не только теоретиком, но и экспериментатором. Но «вопросы Ферми» показали даже тем, кто далек от науки, как можно проводить измерения, кажущиеся на первый взгляд настолько сложными, что не стоит и пытаться. Хотя предложенные им прогрессивные экспериментальные методы имеют большое значение, на мой взгляд, нематериальное представляется не поддающимся измерению не только из-за отсутствия каких-то особо сложных способов оценки. Обычно вещи, считающиеся в бизнесе неизмеряемыми, можно количественно определить с помощью простейших приемов наблюдения, как только люди поймут, что неизмеримость — всего лишь иллюзия. С этой точки зрения ценность подхода Ферми состоит, прежде всего, в том, что оценка современного уровня наших знаний о предмете — необходимое условие последующих измерений.

В отличие от примера Ферми, пример Эмили касается вовсе не значения оценки исходных знаний, поскольку до своего эксперимента она не делала никаких предположений об эффективности метода бесконтактного массажа. Не учит он нас, в отличие от примера Эратосфена, и тому, что неосуществимые наблюдения можно заменить остроумными расчетами. Ее собственные выкладки основывались на стандартных методах выборки и сами по себе не требовали такой гениальной изобретательности, как простые геометрические вычисления Эратосфена. Однако Эмили доказала, что полезные наблюдения необязательно должны быть сложными, дорогостоящими или, по мнению некоторых, недоступными пониманию высшего руководства, даже если речь идет об оценке таких эфемерных явлений, как бесконтактный массаж или стратегическая согласованность.

Как бы ни были полезны эти уроки, мы попробуем, опираясь на них, сделать шаг вперед. Мы научимся оценивать нынешнюю неопределенность относительно какого-либо показателя, что повысит эффективность подхода Ферми, познакомимся с методами выборки, в некоторых отношениях даже более простыми, чем у Эмили, а также с несложными способами, которые позволили бы даже Эратосфену уменьшить ошибку расчета.

■ ПРИМЕЧАНИЯ

- ¹ *M. Lial, C. Miller. Trigonometry. 3rd ed. Chicago: Scott, Foresman, 1988.*
- ² Два француза, Пьер-Франсуа-Андре Мешен и Жан-Батист Жозеф, рассчитали длину окружности Земли за семь лет, пришедшихся на период Французской революции, преследуя цель определить стандартную длину метра (метр сначала был принят равным одной десятимилионной расстояния от экватора до полюса).
- ³ Letter to the Editor // New York Times, 1998, April 7.
- ⁴ Therapeutic Touch: Fact or Fiction? // Nurse Week, 1998, June 7.
- ⁵ *Скотт Адамс. Принцип Дилберта. Взгляд из офисной кабинки на начальство, совещания, причуды дирекции и прочие бедствия. Минск: Попурри, 1999; 2003.*
- ⁶ Являясь некоммерческой организацией, Mitre все равно должна получать доходы от консалтинговых услуг, оказываемых федеральным органам власти, чтобы продолжать свою деятельность.
- ⁷ *Doug Hubbard. Critical Analysis column in: An Audit Trail // CIO Magazine, 2000, May 1.*



Почему неизмеримость нематериального — всего лишь иллюзия

Обычно люди считают вещи неизмеримыми по трем причинам, проистекающим из ошибочных представлений о разных аспектах количественной оценки. Речь идет о сущности измерения, его объекте и методах.

Сущность измерения. Многие неверно понимают саму идею измерения. Если бы люди осознали, что она означает на самом деле, то гораздо больше вещей стали бы для них измеримыми.

Объект измерения. Зачастую объект измерения бывает определен недостаточно четко. Измерению мешают неточность и неоднозначность используемых терминов.

Методы измерения. Многие техники эмпирического наблюдения известны недостаточно хорошо. Будь люди лучше с ними знакомы, стало бы очевидно, что многие вещи, считающиеся неизмеримыми, не только могут быть, но и уже кем-то количественно оценены.

Для легкого запоминания этих трех аспектов весьма распространенного заблуждения по поводу неизмеримости можно использовать название сайта howtomeasureanything.com (буквы *c*, *o* и *m* в «.com» могут обозначать сущность (concept), объект (object) и метод (method). Как только мы поймем, что люди часто сомневаются в возможности

проведения измерения по этим трем причинам, станет ясно, что на самом деле измерению поддается абсолютно все.

Наряду с сомнениями в возможности проведения измерения существует убеждение, что порой проводить количественную оценку и не следует. Выдвигаются следующие причины:

- экономические возражения (ссылки на то, что это потребует слишком больших затрат);
- возражения против полезности и значимости статистики вообще (когда, например, утверждают, что «с помощью статистики можно доказать все, что угодно»);
- морально-этические возражения (утверждение, что оценивать некоторые вещи просто аморально).

В данном случае речь идет не о том, что измерение невозможно в принципе, а о том, что оно экономически нецелесообразно, бесполезно или нежелательно по моральным соображениям. Я постараюсь доказать, что внимания заслуживают лишь экономические возражения, но и ими нередко злоупотребляют.

СУЩНОСТЬ ИЗМЕРЕНИЯ

Пока законы математики остаются определенными, они не имеют ничего общего с реальностью; как только у них появляется нечто общее с реальностью, они перестают быть определенными.

Альберт Эйнштейн

Как это ни парадоксально, но всякая точная наука основывается на приближительности. Если кто-то говорит вам, что точно знает что-то, можете смело делать вывод: вы разговариваете с человеком, понятия не имеющим о точности.

Бертран Рассел, британский математик и философ

Тем, кто считает что-либо не поддающимся количественной оценке, труднее всего преодолеть свое непонимание сущности измерения. Если мы будем придерживаться ошибочного мнения, что измерение означает соответствие каким-то практически недостижимым критериям, то измеримым окажется совсем немного. На семинарах и лекциях я часто спрашиваю своих слушателей, что, по их мнению, означает измерение. (Интересно наблюдать, как глубоко этот вопрос заставляет задуматься людей, отвечающих в своих компаниях за различные программы по количественным оценкам.) Обычно я получаю ответы: «количественно выразить что-либо», «рассчитать точную стоимость», «свести к одному числу» или «выбрать репрезентативное число» и т. д. Явно или неявно все эти ответы подразумевают, что измерение — это определенность, точная величина без права на ошибку. Если бы об-

суждаемый термин действительно имел этот смысл, то измеримые объекты можно было бы перечесать по пальцам.

Однако когда измерения проводят ученые, актуарии* или статистики, они, похоже, пользуются де-факто совершенно иным определением. Каждый из этих специалистов знает, что в его области термины часто имеют иные значения, чем вкладывают в них широкие массы населения. И у экспертов понятие «измерение» не вызывает особых затруднений. А объясняется это тем, что в своих сферах деятельности они пользуются специальной терминологией, выходящей за рамки односложных толкований, а все ее определения являются элементами общей теоретической схемы. В физике сила тяжести — это не просто статья из специального словаря, но параметр конкретных уравнений, связывающих ее с такими понятиями, как масса, расстояние, а также определяющих ее влияние на пространство и время. Поэтому, желая истолковать понятие «измерение» с такой же точностью, мы должны знать кое-что о той теоретической схеме, элементом которой оно является. Без этого мы в нем не разберемся.

ОПРЕДЕЛЕНИЕ ИЗМЕРЕНИЯ

Измерение — это совокупность снижающих неопределенность наблюдений, результат которых выражается некой величиной.

Для всех практических целей научное сообщество рассматривает измерение как совокупность *снижающих неопределенность наблюдений, результат которых выражается некой величиной*. Таким образом, измерение — это не только полное устранение, но и частичное сокращение неопределенности. Даже явно не декларируя данную идею, ученые используют методы, оставляющие сомнения в том, что под измерением они подразумевают именно это. Факт присутствия ошибки, избежать которой полностью не удастся, при том что полученный результат все равно станет шагом вперед по сравнению с прежними представлениями, — ключевая идея проведения экспериментов, опросов и прочих научных измерений.

Фактические различия между таким толкованием измерения и его наиболее распространенными определениями колоссальны. Истинное измерение не только не обязано быть абсолютно точным, чтобы считаться таковым. Однако отсутствие информации о погрешности (подразумевающей строгость оценки величины) может служить признаком того, что эмпирические методы, такие как выборочное и экспериментальное исследования, не использовались (а значит, на самом деле это нельзя

* Эксперты-статистики страховых учреждений. — *Примеч. редактора.*

считать измерением вообще). Настоящие научные методы описывают результаты в определенном интервале значений, например говорят, что «использование нового сорта кукурузы позволило повысить среднюю урожайность кукурузных ферм на 10–18% (95-процентный CI)». Точные значения без указания погрешности могли бы быть рассчитаны «принятым способом» (как, например, оценивались активы Enron). Однако если они не представляют собой 100-процентный итог полного подсчета (как мы пересчитываем мелочь в кармане), их не всегда можно рассматривать как основанные на эмпирическом наблюдении.

Для многих читателей такое понимание сущности измерения может оказаться в новинку, но есть веские математические основания, как и практические причины трактовать данный термин именно таким образом. По крайней мере, мы можем сказать, что измерение — это способ получения информации, к тому же наука уже дала строгую теоретическую концепцию информации. Математическую теорию информации создал в 1940-х годах Клод Шеннон — американский инженер-электротехник, математик и разносторонний ученый, интересовавшийся робототехникой и компьютерными шахматными программами.

В 1948 г. Шеннон опубликовал работу под названием «Математическая теория связи» («A Mathematical Theory of Communication»), заложившую основы теории информации и измерений в целом. Нынешнее поколение знает о Шенноне немного, но на самом деле его заслуги переоценить невозможно. На теории информации основаны все современные теории обработки сигналов и технологии систем электронной связи, в том числе и производство всех когда-либо собранных микропроцессоров.

Шеннон предложил математическое определение информации как снижения неопределенности в сигнале, которое он обсуждает в терминах энтропии, сокращаемой данным сигналом. Согласно Шеннону, получателя информации можно описать как объект, находившийся ранее в состоянии неопределенности. Это означает, что получатель уже обладал ранее какими-то сведениями, а затем получил новую информацию, которая устранила существовавшую неопределенность, но не обязательно полностью. Прежний уровень знаний, или состояние неопределенности, может быть использован, например, для расчета объема информации, который можно передать сигналом, минимальной величины сигнала для поправки на шум и оценки максимально возможного сжатия данных.

Представление об информации как о снижении неопределенности имеет огромное значение для бизнеса. Ведь многие решения (например, стоит ли внедрять новую информационную технологию или разрабатывать новый продукт) принимаются компаниями в условиях неопределенности, и даже незначительное ее уменьшение способствует более удачному выбору. Такое снижение неопределенности может дать многомиллионный экономический эффект.

Итак, измерение не должно устранять неопределенность полностью. Одно только *снижение* ее уже считается измерением, эффект которого порой многократно превышает затраты на проведение подобной оценки. Но есть и другая сторона дела, наверное, удивительная для большинства читателей: измерение — это не обязательно количественная оценка в традиционном смысле слова. Предложенное мной определение гласит, что результат измерения должен выражаться некой величиной. При этом неопределенность следует выразить числом, в то время как сам объект наблюдения может оставаться в виде «качества», скажем принадлежности к какой-либо категории. Например, можно «измерить», получит ли компания патент, произойдет ли слияние и т.д., и то, что мы будем для этого делать, точно удовлетворит нашему определению «измерение». Но неуверенность в результатах наблюдений должна быть выражена количественно (так, необходимо указать, что вероятность получения патента составляет 80% или что вероятность улучшения имиджа компании в глазах общественности после слияния составит 93% и т.п.).

Точка зрения, что измерение позволяет ответить на вопросы, которые требуют лишь утвердительного или отрицательного ответа, или выявить другие качественные различия, вполне согласуется с воззрениями еще одной школы научной мысли, получившей признание. В 1946 г. психолог Стенли Смит Стивенс написал статью под названием «On the Theory of Scales and Measurement» («О теории шкал и измерений»). В ней он описал различные шкалы измерения, включая номинальную (шкалу наименований) и порядковую. Номинальные измерения просто показывают принадлежность объекта к той или иной категории, например пол эмбриона — мужской или женский, или наличие у объекта того или иного заболевания. Шкалы наименований не предполагают ранжирования или сравнения, в частности по размерам. Объект просто относят к той или иной категории.

Порядковые шкалы позволяют утверждать, что одна величина «больше» другой, не уточняя, однако, на сколько. Примером может служить четырехзвездочная система оценки кинофильмов или шкала твердости минералов по Моосу. В обеих шкалах 4 больше, чем 2, но не обязательно вдвое. Вместе с тем однородные единицы, такие как доллары, километры, литры, вольты и т.п., можно суммировать и получить содержательный результат. И если просмотр четырех однозвездочных фильмов не обязательно доставит вам такое же удовольствие, как от одного четырехзвездочного, то камень массой четыре тонны весит ровно столько же, сколько четыре камня массой по одной тонне. Однородные единицы позволяют нам рассчитывать имеющие смысл пропорции (так, четыре километра действительно вдвое превышают два километра).

Концепция номинальной и порядковой шкал может противоречить сложившемуся представлению о том, что такое шкала, но позволяет сделать полезные наблюдения об изучаемых предметах. Иногда геоло-

гу важно знать, что один камень тверже другого (при этом неважно насколько), и такую информацию дает шкала твердости по Моосу.

Стивенс и Шеннон, каждый по-своему, опровергают распространенное представление об измерении. Стивенса больше занимала таксономия различных видов измерения, и он ничего не писал об измерении как о снижении неопределенности. Работавший в другой области Шеннон, возможно, не знал или не интересовался тем, что двумя годами ранее сделал для теории измерений психолог Стивенс. Однако я не думаю, что практическое определение измерения, пригодное для оценки всего, с чем приходится сталкиваться компаниям, было бы разработано без этих двух концепций.

Наряду с некоторыми другими концепциями теория измерения использует обе эти идеи. Измерение она трактует как нечто среднее между измеряемостью и численным выражением величины. Со временем она все более становится понятной только посвященным, но если говорить о вкладе Шеннона и Стивенса, то менеджеры найдут здесь для себя много полезного. Общее представление об измерении как об определении точного количества не учитывает важности простого снижения неопределенности в тех случаях, когда ее полное устранение невозможно или экономически нецелесообразно. В бизнесе решения принимаются в условиях неопределенности. Когда такая неопределенность велика и касается важных, рискованных решений, ее снижение играет большую роль. Вот почему в дальнейшем мы будем пользоваться именно этим определением измерения.

ОБЪЕКТ ИЗМЕРЕНИЯ

Правильно поставленная проблема уже наполовину решена.

*Чарльз Кеттеринг (1876–1958),
американский изобретатель, получивший 300 патентов,
в том числе патент на автомобильную систему зажигания*

Ничто так не мешает прогрессу знания, как расплывчатость терминологии.

Томас Рейд (1710–1769), шотландский философ

Даже если принять более полезную концепцию измерения, кое-что по-прежнему будет казаться не поддающимся оценке, поскольку мы просто не знаем, что имеем в виду, когда впервые ставим перед собой проблему. На самом деле часто нам неясен сам объект измерения. На вопрос, как измерить стратегическую согласованность, гибкость или удовлетворенность потребителей, я отвечаю: «А что конкретно вы имеете в виду?» Интересно наблюдать, как зачастую, уточняя используемый ими термин, люди фактически сами отвечают на свой вопрос. На семинарах я часто прошу присутствующих поставить передо

мною максимально сложную задачу, связанную с измерением. Как-то мне предложили измерить наставничество, и я сказал: «Что ж, такая проблема может возникнуть в реальной жизни. Первое, что я заметил бы, — это: больше наставничества всегда лучше, чем меньше наставничества. Я знаю людей, которые тратят на это деньги, поэтому не удивлюсь, если кому-то потребуется оценить данное явление. Однако что *вы* имеете в виду под наставничеством?» Слушатель ответил, не задумываясь: «Не уверен, что смогу это сформулировать», на что я заметил: «Вот поэтому, наверное, вам и трудно его измерить. Сначала необходимо сформулировать, что такое наставничество».

Когда я просто спрашиваю людей, что они подразумевают и насколько важно измерить интересующий их объект или явление, они сами нередко отвечают на свой вопрос. Обычно это мой первый этап анализа — то, что я называю семинарами по выявлению объекта измерения. Дело в том, что клиенты нередко просят меня измерить вещи, которые можно толковать по-разному. В таких случаях я, как правило, задаю вопрос: «А что конкретно вы имеете в виду под... (вставьте нужное)?»

В 2000 г., когда сотрудники Управления по делам ветеранов (Department of Veterans Affairs) попросили меня помочь в разработке показателей информационной безопасности, я спросил: «А что вы имеете в виду под информационной безопасностью?», и на первых двух или трех встречах они объясняли мне, что это такое. В конце концов выяснилось, что под информационной безопасностью они понимают такие вещи, как сокращение числа несанкционированных вторжений и вирусных атак. Кроме всего прочего, подобные вещи влияют на финансовое положение Управления за счет убытков от мошенничества, снижения производительности труда и даже возможных случаев правовой ответственности (чего они едва избежали в 2006 г., вернув украденный ноутбук, в памяти которого содержались номера карточек социального страхования 26,5 млн ветеранов).

Очевидно, что почти все упомянутые факторы поддавались измерению. Понятие «безопасность» допускало неоднозначное толкование до тех пор, пока сотрудники Управления не разложили его составляющие, которые можно реально наблюдать. Однако довольно часто клиенты нуждаются в дополнительном руководстве, когда пытаются охарактеризовать объекты таким образом, чтобы их можно было измерить. В сложных случаях я использую «цепочку уточнений» или, если это не срабатывает, провожу своего рода мысленный эксперимент.

Цепочка уточнений — это просто короткий ряд ассоциаций, помогающий представить себе нематериальный объект как материальный. Сначала мы осознаем, что если объект X имеет для нас значение, то он, по определению, должен в чем-то проявляться. Разве могли бы такие реальности, как качество, риск, безопасность или репутация,

иметь для нас какую-то ценность, не проявляй они себя прямо или косвенно? Если возникает причина интересоваться неизвестной величиной, значит, мы думаем, что она каким-то образом приводит к желательным или нежелательным последствиям. Кроме того, если эта величина как-то себя проявляет, значит, она обнаруживает себя в каком-то количестве. Если нечто вообще наблюдается, то его бывает больше или меньше. Как только мы все это примем, последний шаг наверняка окажется самым легким. Раз что-то наблюдается в каком-то количестве, значит, это можно измерить.

Например, решив, что нужно оценить нашу деловую репутацию, потому что она влияет на рекламу, которую нам делают сами клиенты, а это, в свою очередь, влияет на продажи, мы уже приступили к поискам способа оценки своей репутации. Ведь рекомендации клиентов можно не только наблюдать, но и фиксировать каким-то количеством. А это означает, что они поддаются измерению. Иногда я провожу людей не через все звенья цепочки уточнения объекта измерения, но этот метод все равно дает хорошие результаты, если помнить три важных правила.

ЦЕПОЧКА УТОЧНЕНИЯ

1. Если это вообще имеет какое-то значение, значит, оно обнаруживается или наблюдается.
2. Если это обнаруживается, значит, оно обнаруживается в каком-то количестве (или количественном интервале).
3. Если это обнаруживается в количественном интервале, его можно измерить.

Когда цепочка уточнения не срабатывает, я провожу мысленный эксперимент. Представьте себе, что вы иностранный ученый, способный клонировать не только овец и людей, но даже целые организации. Предположим, предметом вашего исследования является конкретная сеть предприятий экспресс-питания и вас интересует, как на ее работу повлияет такая нематериальная вещь, как расширение полномочий сотрудников. Вы создаете два клона изучаемой организации, один из которых называете тестируемой, а второй — контрольной группой. При этом сотрудники тестируемой группы получают дополнительные полномочия, а в контрольной группе все остается по-прежнему. Какие изменения — прямые или косвенные — вы наблюдаете в первой организации? Ожидаете ли вы, что решения будут приниматься на более низком организационном уровне? Станут ли решения качественнее и будут ли они приниматься быстрее? Приводит ли передача полномочий к ослаблению контроля за работниками? Означает ли это, что

теперь структура организации станет более «плоской» и накладные расходы на управление сократятся? Если удастся обнаружить хотя бы одно различие между двумя организациями-клонами, то вопрос о способе измерения вашего объекта будет наполовину решен.

На самом деле точное определение объекта измерения — исходный пункт любого научного исследования, даже самого революционного. Менеджеры компаний должны понять, что некоторые вещи кажутся нематериальными только потому, что люди сами толком не решили, о чем они говорят. Определитесь с объектом — и половина работы по измерению будет проделана.

МЕТОДЫ ИЗМЕРЕНИЯ

Некоторые вещи кажутся не поддающимися измерению только потому, что тот, кто хочет это сделать, незнаком с основными методами измерения, такими как выборочные исследования или управляемые эксперименты разного типа. Традиционное возражение звучит примерно так: данная задача уникальна, ранее такие измерения не проводились и подходящего метода просто не существует. Надежду вселяет тот факт, что существует целый ряд проверенных способов, позволяющих оценить даже то, что сейчас кажется неизмеримым. Вот несколько примеров.

- *Измерения при очень малой случайной выборке* (например, можете ли вы узнать нечто важное от маленькой выборки из потенциальных потребителей, своих работников и т. д. в условиях большой неопределенности?).
- *Оценка некой совокупности, или множества, объектов, которые невозможно увидеть все сразу* (например, численность рыб одной породы в океане, растений одного вида в ливневых лесах, сбоев при производстве нового продукта или же таких попыток несанкционированного доступа в вашу систему, которые остаются необнаруженными, и т. д.).
- *Измерения при наличии множества переменных, в том числе неизвестных* (например, определение причины роста продаж — реализация новой программы повышения качества, экономия, ошибки конкурентов, новая политика ценообразования и т. д.).
- *Оценка риска наступления маловероятных событий* (например, вероятность неудачного запуска новой ракеты, повторения атаки террористов 11 сентября 2001 г. или еще одного наводнения в Новом Орлеане).
- *Оценка стоимости предметов искусства, свободного времени или повышения вашей личной безопасности на основе того, сколько на самом деле люди готовы за это заплатить.*

Многие из перечисленных подходов к измерению — просто разновидности базовых методов, включающих выборку и управляемые эксперименты, иногда концентрирующих внимание на других типах вопросов. Такие методы наблюдения практически не используются в принятии бизнес-решений, вероятно, потому, что представляются довольно сложными и формализованными. Считается, что это не те способы, которые можно при необходимости применить сразу же без подготовки и с минимальными затратами. Однако на самом деле это совсем не так.

Вот очень несложный пример быстрого измерения, которое каждый может провести и при этом легко рассчитать статистическую погрешность. Предположим, что вы подумываете над тем, чтобы перевести часть своих сотрудников на дистанционную работу. Один из факторов, который вам необходимо учесть, — сколько времени средний служащий ежедневно тратит на дорогу до работы и домой. Чтобы выяснить это, вы можете официально опросить всех работников, потратив много времени и денег. При этом, скорее всего, ответ будет более точным, чем вам необходимо. Допустим теперь, что взамен вы выберете наугад пять человек. Позже мы еще поговорим о том, что означает «выбрать наугад», а сейчас представим, что вы просто закроете глаза и ткнете пальцем в список сотрудников. Вызовите этих людей и спросите, сколько времени они обычно тратят на дорогу. Предположим, будут получены следующие ответы: 30, 60, 45, 80 и 60 минут. Возьмем самое высокое и самое низкое значения в выборке — 35 и 80. Вероятность того, что медиана значений продолжительности поездок на работу и домой для совокупности работников находится в этом интервале значений, составляет 93%. Я называю это **Правилом пяти**. Правило пяти несложно, оно работает, и можно доказать его статистическую обоснованность для решения целого ряда задач. Когда выборка настолько мала, интервал значений может быть очень широким, но если он окажется значительно уже предыдущего интервала, значит, вы провели измерение.

ПРАВИЛО ПЯТИ

Существует 93-процентная вероятность того, что в любой случайной выборке медиана для всей совокупности находится в интервале между наименьшим и наибольшим значениями.

Кажется, что нельзя быть уверенным в чем-то на 93%, располагая всего пятью случайными значениями, но это возможно. Чтобы понять, почему это так, важно учесть, что **Правило пяти** позволяет определить медиану совокупности. Медиана — это такая точка, которая делит всю

совокупность на половины со значениями выше и ниже нее. Если выбрать наугад такие пять значений, которые все располагались бы выше или ниже медианы, то медиана оказалась бы вне интервала. Но какова вероятность подобного выбора?

Вероятность случайного выбора значения, превышающего медиану, составляет, по определению, 50% (вероятность выпадения орла при подкидывании монеты). Вероятность случайного выбора пяти значений, которые все окажутся выше медианы, равна вероятности выпадения орла при подкидывания монеты пять раз подряд, то есть равна $1/32$, или 3,125%. Такова же вероятность того, что пять раз подряд выпадет решка. Тогда шанс *не* получить все решки или все орлы составляет $100\% - 3,125\% \times 2 = 93,75\%$. Поэтому вероятность случая, когда по крайней мере одно значение из пяти окажется больше или меньше медианы, составляет 93,75%, или, если округлить с понижением, 93% (или даже 90%, если вы хотите быть очень осторожным). Некоторые читатели, возможно, еще помнят тот раздел статистики, где обсуждаются методы выборочного обследования для малых выборок. Эти методы несколько сложнее Правила пяти, но по причинам, о которых я подробнее расскажу позже, не позволяют получить намного более точные результаты.

Дополнительно к такому эмпирическому правилу можно использовать простые методы, позволяющие учесть определенные виды смещения (ошибок). Предположим, что начавшееся строительство удлинит время, затрачиваемое каждым сотрудником на дорогу, или же люди, которые дольше добираются до работы, чаще болеют или не вошли в вашу выборку по другим причинам. Однако даже при всех его недостатках любой человек, желающий развить у себя умение измерять, обязательно должен знать Правило пяти.

Позднее мы поговорим о разных подходах, позволяющих, как доказано, еще более снизить неопределенность. Это и более сложные методы выборочного обследования или проведения экспериментов, и приемы, которые просто снижают погрешность субъективных оценок экспертов. В своем стремлении повысить точность измерений мы должны учесть множество факторов, но важно помнить одно: когда наблюдение сообщает нам нечто, чего мы раньше не знали, это означает, что произведено измерение.

Между тем ответим на вопрос, почему неправ тот, кто считает, что «метода измерения подобного объекта еще не существует». В бизнесе принято приклеивать ярлык нематериального всему, чего нет в бухгалтерской отчетности или базах данных. И даже если измерения считают возможными, то полагают, что они по силам только специалистам, владеющим сложными методами, практически не доступными для бизнесменов. К счастью, обычно это не так. Свой способ измерения может придумать каждый.

Важную подсказку дает этимология слова «эксперимент», которое происходит от латинского *ex*, что означает «из» или «от», и *periri*, что означает «попытка». Иными словами, оно означает получение чего-то путем осуществления попытки. Статистик Дэвид Мур, возглавлявший в 1998 г. Американскую статистическую ассоциацию (American Statistical Association), как-то сказал следующее: «Измеряйте, даже если не знаете, что измерять. Тогда вы и узнаете, что вам нужно измерить»¹. Принцип Мура мало чем отличается от подхода Nike «просто сделай это». Звучит как «сначала измерь, а вопросы будешь задавать потом», и если довести данный принцип до крайности, то окажется, что он далеко не идеален. Однако у него есть ряд значительных преимуществ перед нынешним неконструктивным отношением некоторых менеджеров к измерению.

Многие из тех, кто принимает решения, думают только о том, что мешает провести измерения, и поэтому даже не пытаются осуществлять наблюдения. Если вы предлагаете провести опрос и выяснить, сколько времени люди тратят на какой-то вид деятельности, они говорят: «Хорошо, но кто же точно запоминает такие вещи?» Если вы хотите опросить потребителей, чтобы узнать их предпочтения, они заявляют: «У нас такие разные потребители, что вам потребуется огромная выборка». Пытаетесь показать, какая программа совершенствования привела к росту продаж, и слышите: «Но продажи зависят от множества факторов, разве можно узнать, как именно повлияла именно эта программа?» Такие люди заранее уверены в бесполезности измерений, чего заранее знать нельзя. Они просто предполагают.

Подобные критики делают определенные допущения о трудности проведения измерений. Они могут даже ссылаться на свою подготовку в этой области (например, на то, что 20 лет назад они в течение двух семестров изучали статистику). Я не стану утверждать, что их предположения всегда оказываются неверными. Но скажу, что если это только допущения, то они контрпродуктивны. Давайте воспользуемся другими предпосылками, которые, как и любые другие, не всегда оказываются верными в конкретных случаях, но на практике приносят намного больше пользы.

ЧЕТЫРЕ ПОЛЕЗНЫЕ ПРЕДПОСЫЛКИ ИЗМЕРЕНИЯ

1. Ваша проблема совсем не так уникальна, как вы думаете.
2. У вас гораздо больше информации, чем вам кажется.
3. Вам нужно меньше данных, чем вы предполагаете.
4. Существует удобный способ измерения, который намного проще, чем вы представляете.

Предпосылка № 1

Такие измерения уже проводились. Какой бы сложной или «уникальной» ни казалась вам проблема, предположите, что кто-то уже занимался ею до вас, возможно в другой области. Если эта предпосылка окажется неверной, пусть утешением послужит то, что, возможно, вы осуществляете попытку получить за свое открытие Нобелевскую премию. А если серьезно, то я заметил, что специалисты любого профиля склонны считать свою сферу деятельности уникальной с точки зрения уровня неопределенности. Обычно они говорят: «В отличие от других отраслей в нашей каждая задача уникальна и непредсказуема» или «В моей отрасли факторов слишком много, чтобы их можно было выразить количественно» и т. п. Мне доводилось много работать в разных отраслях и слышать одно и то же. Но до сих пор проблемы измерения везде оказывались стандартными и ничем не отличались друг от друга.

Предпосылка № 2

На самом деле у вас намного больше информации, чем вы думаете. Предположите, что необходимые для ответа на поставленный вопрос данные вполне доступны, и если вы потратите время и подумаете, то обязательно их найдете. Мало кто из руководителей даже отдаленно представляет себе, какую информацию можно найти в собственной организации. То, что так важно измерить, обычно оставляет следы, которые вы обязательно обнаружите, если окажетесь достаточно изобретательным.

Предпосылка № 3

На самом деле вам нужно намного меньше информации, чем кажется. Правило пяти помогает снизить неопределенность при решении множества задач. Я знал статистиков, которые не верили в это правило до тех пор, пока сами не выполняли расчеты. Но, как нам показал Эратосфен, при умелом подходе интересные выводы можно извлечь и из минимальной информации.

Предпосылка № 4

Существуют полезные способы измерения, гораздо менее сложные, чем вы предполагаете. Представьте, что первый пришедший вам на ум способ — самый трудный. Представьте, что, проявив чуть больше изобретательности, вы сможете найти более простой метод. Например, Кливлендский оркестр захотел оценить, улучшается ли со временем качество его исполнения произведений. Многие бизнес-аналитики пред-

ложили бы руководству оркестра проводить периодические опросы случайно выбранных постоянных слушателей, пришедших на концерт. Возможно, они предложили бы этим людям оценить то или иное выступление оркестра (если те его помнят) по шкале от «плохое» до «блестящее» или по нескольким параметрам, на основе которых потом можно было бы рассчитать общий показатель удовлетворенности. Но Кливлендский оркестр подошел к этой проблеме творчески и начал подсчитывать, сколько раз публика устраивала овации стоя. Если слушатели вставали не два раза, а три, то, очевидно, выступления отличались не слишком сильно. Но если с появлением нового дирижера публика стала устраивать бурные овации, поднимаясь из кресел намного чаще, то мы можем сделать весьма ценные выводы по поводу этого дирижера. Да, это было измерением в полном смысле слова, и оно потребовало гораздо меньше усилий, чем проведение опроса, дав, как сказали бы многие (и я бы с ними согласился), более содержательные результаты.

Итак, не стоит исходить из того, что единственный способ уменьшить неопределенность — это использовать какой-то сложный, практически нецелесообразный метод. Что вы собираетесь делать — публиковаться в научном журнале или сокращать неопределенность при принятии реального бизнес-решения? Отнеситесь к измерению как к итеративному процессу. Начните измерять то, что вам нужно. Получив первые результаты, вы всегда сможете скорректировать свой метод.

Самое главное, как это следует из этимологии слова «эксперимент», интуитивный экспериментатор *делает попытку*. Сделайте ее и вы. За исключением тех случаев, когда заранее можно предсказать результат какого бы то ни было наблюдения, оно обязательно расскажет то, чего вы не знали ранее. Сделайте еще несколько наблюдений — и узнаете больше.

Конечно, бывают такие редкие случаи, когда объекты или явления кажутся не поддающимися количественной оценке только из-за отсутствия современных изощренных методов. Но обычно для измерения большинства вещей, называемых нематериальными, не хватает вовсе не передовых замысловатых способов. Как правило, неопределенность в отношении подобных объектов так велика, что уменьшить ее позволяют и базовые методы измерения.

ЭКОНОМИЧЕСКИЕ ВОЗРАЖЕНИЯ ПРОТИВ ПРОВЕДЕНИЯ ИЗМЕРЕНИЙ

Как мы видели, сомнения в концепции, объекте и методе измерения зачастую связаны с ошибочным восприятием проблемы. Однако иногда возражения против измерений основываются не на уверенности в их невозможности, а на убеждении, что проводить их *не следует*.

На мой взгляд, измерения не следует осуществлять в одном случае — если издержки, связанные с их проведением, превышают получаемые выгоды. Однако на практике, конечно, встречается и такое. В 1995 г. я разработал метод оценки неопределенности, рисков и других нематериальных факторов, влияющих на принятие любого крупного рискованного решения, который назвал прикладной информационной экономикой (applied information economics, AIE). Ключевой этап данного процесса (и этим объясняется название метода) — расчет экономической стоимости информации. Позже поговорим об этом подробнее, но проверенная формула из теории решений позволяет рассчитать денежную стоимость любого снижения неопределенности. Я вставил эту формулу в программу Excel и годами рассчитываю экономическую стоимость измерения любых величин, учитываемых при принятии десятков важных бизнес-решений. Благодаря этим расчетам я обнаружил ряд любопытных закономерностей, но пока упомяну только об одной: информационная ценность большинства переменных, учитываемых при принятии бизнес-решений, равна нулю. В каждом случае оказывалось, что число факторов, которые имеют для принятия решения значение, оправдывающее усилия по их оценке, не более четырех.

ЗНАЧЕНИЕ ИМЕЮТ ЛИШЬ НЕСКОЛЬКО ВЕЩЕЙ

В каждом случае лишь несколько ключевых переменных имеют значение, оправдывающее усилия по их определению. Информационная ценность остальных равна или практически равна нулю.

Конечно, существуют величины, не стоящие усилий, затраченных на их измерение, и все же бытует устойчивое заблуждение, что если полученная количественная оценка не соответствует неким произвольным критериям (например, результаты нельзя опубликовать в научном журнале или они не согласуются с общепризнанными стандартами бухгалтерского учета), то она не имеет никакой ценности. Это, разумеется, излишнее упрощение, поскольку на самом деле ценность измерению придает большая неопределенность неизвестной в сочетании с высокой ценой принятия неверного решения. Отвечает ли измерение другим стандартам, значения не имеет. Когда вы рискуете большими деньгами в расчете на результат действия фактора, поведение которого отличается серьезной неопределенностью, даже небольшое ее снижение имеет вполне поддающуюся расчету денежную стоимость. Предположим, например, что вы собираетесь потратить крупные средства на придание своему продукту свойства, благодаря которому его продажи в данном демографическом секторе могут воз-

расти на 12%, хотя, возможно, рост будет не так заметен. Кроме того, вы считаете, что затраты окупятся только в том случае, если объемы продаж повысятся не менее чем на 9%. Если в результате инвестиций объемы продаж не возрастут на 9%, значит, вы не добились доходности. А если рост продаж будет еще меньше или реализация упадет, то вся затея обернется катастрофой. В этой ситуации предварительная оценка имела бы очень высокую ценность.

Когда вам говорят, что определить величину «слишком дорого» или «слишком трудно», сразу спрашивайте: «По сравнению с чем?». Разумеется, если информационная ценность измерения буквально или практически равна нулю, проводить его не имеет смысла. Но когда эта ценность велика, задайте себе вопрос: «Существует ли вообще метод, позволяющий снизить неопределенность настолько, чтобы оправдать затраты на проведение измерения?» Как только ценность даже частичного снижения неопределенности становится понятной, ответ на этот вопрос обычно бывает утвердительным.

ВОЗРАЖЕНИЕ ПРОТИВ ПОЛЕЗНОСТИ «СТАТИСТИКИ» ВООБЩЕ

В конце концов, факты — это факты. И хоть мы напоминаем друг другу, посмеиваясь, слова одного мудрого политика про «ложь, проклятую ложь... и статистику», есть все же такие цифры, в которых разберется даже последний простак и которые не сможет извратить и первый ловкач.

*Леонард Кортни, первый барон Кортни,
президент Королевского статистического общества (1897–1899)*

Иногда высказывается и такое возражение, что даже если измерение возможно, оно ничего не даст, поскольку статистика и вероятность сами по себе бессмысленны (так сказать, «ложь, проклятая ложь... и статистика»²⁾). Глубокое непонимание основ статистики проявляют порой даже квалифицированные специалисты. Иногда приходится слышать такие поразительные вещи, что не знаешь, как на них реагировать. Вот несколько высказываний, которые мне довелось выслушать:

«Все события равновероятны, так как мы не знаем, что произойдет»
(сказано слушателем моего семинара).

«Я совершенно нетерпим к риску, потому что никогда не рискую»
(фраза менеджера среднего звена страховой компании, которую я консультировал).

«Как узнать интервал, если я даже не знаю среднего значения?» (вопрос клиента Сэма Сэвиджа, ученого, моего коллеги и пропагандиста методов статистического анализа).

«Откуда известно, что вероятность выпадения орла — 50%, если никто не знает, что произойдет?» (*вопрос, заданный дипломником на лекции, которую я читал в Лондонской школе экономики, — это не шутка.*)

«С помощью статистики можно доказать все, что угодно» (*известная поговорка о статистике*).

Давайте сначала обсудим последнюю фразу. Я прямо сейчас предлагаю приз в 10 тыс. дол. любому, кто, используя статистику, докажет следующее утверждение: «С помощью статистики можно доказать все, что угодно». Под словом «доказать» я подразумеваю, что это доказательство можно будет опубликовать в любом крупном математическом или вообще научном журнале. Критерием будет именно такая публикация (уверен, что подобное важное открытие обязательно будет напечатано). Под словами «что угодно» я подразумеваю буквально все, что угодно, включая любое давно опровергнутое утверждение из области математики или другой науки. А вот под статистикой я понимаю статистику в самом широком смысле слова. Тот, кто получит этот приз, может использовать любое направление математики или другой науки, имеющее малейшее отношение к теории вероятности, методам выборочного обследования, теории решений и т. д.

Дело в том, что когда люди говорят: «С помощью статистики можно доказать все, что угодно», они, скорее всего, имеют в виду не статистику как таковую, а использование цифр вообще (особенно, по какой-то причине, процентов). На самом деле они подразумевают не совсем «что угодно» или не совсем «доказать». Реальный смысл поговорки заключается в том, что «цифрами можно сбить с толку людей, особенно легковых, кто не в ладах с математикой». Вот с этим я полностью согласен.

Другие высказывания из моего списка свидетельствуют о незнании основ теории вероятности, риска и измерений в целом. Ясно, что вероятности мы используем исключительно потому, что не можем быть уверены в полученном результате. Очевидно и то, что нам приходится принимать какие-то риски, даже при обычной поездке на работу, так что все люди в известной мере терпимы к риску.

Иногда я обнаруживаю, что люди, делающие подобные заявления, противоречащие здравому смыслу, не вполне понимают, о чем говорят, поскольку их собственные шаги идут вразрез с теми убеждениями, которых они якобы придерживаются. Заключите с кем-нибудь пари, что он не угадает, сколько раз выпадет решка, если подбросить монету 12 раз, — и вы увидите, что даже тот, кто считает, что вероятность определить невозможно, назовет цифру 6 или близкую к ней. Тот, кто говорит, что никогда не рискует, тем не менее, летит в Москву самолетом Аэрофлота (компания, статистика авиапроисшествий которой

хуже любого американского перевозчика), чтобы получить в качестве приза миллион долларов. С какими только заблуждениями по поводу статистики и теории вероятности не приходится сталкиваться! Этим ошибочным мнениям, характерным даже для руководителей крупнейших компаний, а также путям их преодоления посвящены почти все статьи в таких изданиях, как, например, «Journal of Statistics Education». Но тот, кто дочитает книгу до конца, наверняка избавится хотя бы от некоторых своих заблуждений относительно статистики.

ЭТИЧЕСКИЕ ВОЗРАЖЕНИЯ ПРОТИВ ПРОВЕДЕНИЯ ИЗМЕРЕНИЙ

Обсудим теперь последний аргумент, приводимый теми, кто считает, что осуществлять измерения не следует. Это аргумент *этического* характера. Сопротивление проведению измерений порождается страхом перед возможной ответственностью и уверенностью в окончательном характере чисел в сочетании с ранее приобретенным недоверием к «статистике». Иногда измерения воспринимают даже как нечто антигуманное. Попытки оценить такие деликатные вещи, как стоимость исчезающих видов животных или даже человеческой жизни, вызывают праведное негодование. Тем не менее подобные оценки проводятся, причем регулярно, поскольку они необходимы.

Агентство по защите окружающей среды (Environmental Protection Agency, EPA) и другие государственные органы должны распределять весьма ограниченные средства на экологию, охрану нашего здоровья и даже жизни. Я помогал EPA оценивать некоторые инвестиционные проекты, в том числе проект создания Географической информационной системы (Geographic Information System, GIS) для улучшения мониторинга концентрации диметилртути — по мнению многих, вещества, высокая концентрация которого приводит к снижению коэффициента интеллекта (IQ) у детей.

Чтобы оценить, следует ли создавать эту систему, мы должны были задать важный, хотя и не очень приятный вопрос: целесообразно ли тратить в течение пяти лет 3 млн дол. на то, чтобы предотвратить возможное снижение IQ? Кого-то может возмутить одно то, что такой вопрос вообще был поставлен. Многие скажут, что миллионы стоит потратить даже в том случае, если IQ детей удастся повысить хотя бы на один балл.

Однако позвольте! Ведь EPA приходится финансировать и другие системы, призванные наблюдать за эффектом воздействия новых загрязнителей, которые иногда приводят к преждевременной смерти. Ресурсы EPA ограничены, а программ по улучшению общественного здравоохранения и состояния окружающей среды, а также спасению

исчезающих видов животных очень много. Агентству приходится сравнивать программы, задавая вопросы: «Скольким детям мы сумеем помочь, и насколько повысится их IQ?» и «Сколько преждевременных смертей можно будет предотвратить?»

Иногда даже приходится спрашивать: «Насколько преждевременной является такая смерть?» Когда ограниченность ресурсов вынуждает делать выбор, должны ли мы относиться к смерти пожилого человека так же, как к гибели молодого? В определенный момент ЕРА даже подумывала о том, чтобы использовать поправку на смерть пожилых людей (*senior death discount*). Это означает: смерть человека в возрасте за 70 лет «стоит» на 38% меньше смерти человека в возрасте до 70. У некоторых это вызвало возмущение, и в 2003 г. администратору ЕРА Кристин Тодд Уитмен пришлось объявить, что данная поправка использовалась для общей ориентации, а не для разработки политики, и теперь эта практика прекращена³. Конечно, даже говоря об одинаковой стоимости жизни людей разных возрастов, мы проводим измерения, поскольку выражаем эти величины количественно. Но если они одинаковы, то всегда ли действует этот принцип? Стоит ли потратить на спасение жизни 99-летнего больного человека столько же усилий, сколько на спасение здорового пятилетнего ребенка? Каким бы ни был ваш ответ, это будет измерение относительной стоимости жизни каждого.

Если мы будем настаивать на том, чтобы не рассчитывать относительную стоимости многочисленных государственных социальных программ (что является естественным результатом отказа от определения ценности каждой из них), то почти наверняка будем распределять ограниченные ресурсы так, что, решив менее значимые задачи, потратим гораздо больше денег. Это связано с огромным числом возможностей для подобных инвестиций, и без цифр определить лучший вариант очень трудно.

В других случаях кажется, что само существование погрешности (которая, как мы знаем, почти всегда существует в эмпирических измерениях) делает попытку измерения неприемлемой с точки зрения морали. Стивен Гулд, автор книги «*The Mismeasure of Man*» («Ложное измерение человека»), страстно возражает против полезности и даже этичности оценки интеллекта путем определения IQ или единого показателя интеллекта *g* (проявляющегося при различных тестах, в том числе при расчете IQ). По его словам, *g* — не что иное, как артефакт математических манипуляций⁴. При определении баллов IQ и *g*, разумеется, возникает ошибка, но все-таки они не только отражают процедуру их расчета, но и основаны на наблюдениях (тестах). И поскольку теперь мы понимаем, что измерение вовсе не означает «полного отсутствия ошибки», возражение о невозможности количественной оценки интеллекта из-за погрешностей тестов, по меньшей мере, беспомощно.

Более того, другие исследователи считают: точка зрения, что показатели интеллекта не являются характеристиками реального явления, не согласуется с хорошей корреляцией таких показателей, рассчитанных разными способами⁵, а также с социальным феноменом криминального поведения людей или индивидуальными доходами⁶. Как IQ может быть совершенно произвольным числом, если он соответствует наблюдаемой реальности? Я не пытаюсь здесь решить этот спор, но мне интересно, как Гулд распорядился бы средствами, выделяемыми на защиту окружающей среды от токсических веществ, влияющих на умственное развитие. Рассмотрим пример с диметилртутью, которая приводит к снижению IQ у детей. Считает ли Гулд, что на самом деле она такого эффекта не оказывает, или же он думает, что даже если эффект наблюдается, мы не вправе его оценивать, поскольку значения IQ были определены с погрешностью? В любом случае в итоге нам придется проигнорировать наносимый токсическими веществами вред здоровью, а отсутствие полной информации заставит направить ресурсы на другую программу. Бедные дети!

Когда между полным невежеством и возможностью хоть немного восполнить пробелы в своих знаниях выбирают полное невежество, вряд ли это можно назвать высокими моральными устоями. Если решения принимаются в условиях высокой неопределенности, хотя ее и можно снизить, то политики (или даже руководители таких компаний, как авиастроительные) сильно рискуют неправильно распределить ограниченные ресурсы и в результате создать угрозу для нашей жизни. В измерениях, как и во многих других областях человеческой деятельности, невежество не только расточительно, но и опасно.

Невежество никогда не может быть лучше знания.

*Энрико Ферми,
лауреат Нобелевской премии по физике (1938)*

НАВСТРЕЧУ УНИВЕРСАЛЬНОМУ ПОДХОДУ К ИЗМЕРЕНИЮ

Итак, мы узнали о людях, обладавших интересными интуитивными подходами к измерению. Мы познакомились с тем, как следует рассматривать главные возражения против проведения измерений, в том числе некоторые «измеренческие» максимы, и познакомились с несколькими интересными примерами количественной оценки. Мы выяснили, что люди считают объект не поддающимся измерению только из-за своих ошибочных представлений. Все это вместе очерчивает общие рамки схемы измерения. Чтобы завершить дело, нам нужно добавить к ней еще несколько понятий. Данная схема, как оказалось,

лежит в основе разработанного мною метода прикладной информационной экономики.

При всем разнообразии возникающих на практике задач по измерению мы все же предложим такую последовательность шагов, которой следует придерживаться при проведении количественной оценки практически любого вида. Таким образом, мы можем разработать универсальный подход. Каждый элемент этого подхода хорошо известен специалистам разных областей науки или отраслей экономики, но никто еще не объединял их в единый логически последовательный метод. Предлагаемый универсальный подход требует ответов на шесть перечисленных далее вопросов.

1. Что вы пытаетесь измерить? Что на самом деле представляет собой этот якобы неизмеримый объект?
2. Почему вы хотите его измерить? Какое решение будет принято по результатам измерения, и каким должно быть «пороговое значение» определяемого показателя?
3. Что вам известно сейчас — какие интервалы или вероятности представляют нынешнюю неопределенность?
4. Какую ценность имеет данная информация? К каким последствиям приведет ошибка, какова ее вероятность и какие усилия, связанные с измерением, будут оправданы с экономической точки зрения?
5. Какие наблюдения, затраты на которые будут оправданы ценностью требуемой информации, позволят подтвердить или исключить различные возможности? Что именно мы должны увидеть сразу, если сбудется тот или иной сценарий?
6. Как учесть такие ошибки при измерении, которых можно избежать (опять при условии, что затраты оправдаются ценностью информации)?

Более подробно каждый шаг описываемого подхода будет рассмотрен в последующих главах, но частично мы о них уже говорили.

Выгоды взгляда на мир «калиброванными» глазами, которые видят все через призму количественной оценки, были той исторической силой, которая двигала вперед науку и способствовала повышению экономической эффективности. Измерение — один из основных инстинктов человека, однако этот инстинкт подавляется в условиях, когда люди предпочитают создавать комитеты и добиваться консенсуса вместо того, чтобы делать простые наблюдения. Многим менеджерам даже не приходит в голову, что «нематериальное» можно измерить с помощью простых и остроумных наблюдений.

Из предыдущего знакомства с концепцией измерения люди часто выносят свои предубеждения. Все мы сталкивались с основными по-

нениями теории измерений, например в университетской химической лаборатории, но вряд ли научились чему-то, кроме того, что измерения всегда точны и что количественно оценить можно только то, что поддается непосредственному наблюдению и количественному выражению. Однако, видимо, университетская статистика приводит в замешательство столько же людей, сколько информирует. Когда мы идем на работу, то видим, что специалисты любого профиля и квалификации просто завалены задачами по измерению совсем не столь легко поддающихся оценке объектов, с какими мы имели дело на семинарах. Взамен мы узнаем, что некоторые вещи просто не поддаются измерению. Однако, как мы видели, «неизмеримость» — это миф. Задачу по измерению можно решить. Вопрос «Сколько?» позволяет взглянуть на любую задачу под нужным углом зрения, и даже самые сложные проблемы измерения решаемы, если только осознать, к каким последствиям приведет отказ от их решения.

■ ПРИМЕЧАНИЯ

- 1 *George W. Cobb*. Reconsidering Statistics Education: A National Science Foundation Conference // *Journal of Statistics Education*, 1993, 1, p. 63–83.
- 2 Это утверждение часто неправильно приписывают Марку Твену, который, безусловно, помог его популяризовать. Твен позаимствовал его у двух британских политиков XIX века — Бенджамина Дизраэли и Генри Лабушера.
- 3 *Katharine Q. Seelye, John Tierney*. Senior Death Discount' Assailed: Critics Decry Making Regulations Based on Devaluing Elderly Lives // *New York Times*, 2003, May 8.
- 4 *Stephen Jay Gould*. *The Mismeasure of Man*. New York: W. W. Norton & Company, 1981.
- 5 Размышления о книге Стивена Джея Гулда «The Mismeasure of Man»: *John B. Carroll*. A Retrospective Review // *Intelligence*, 1995, 21, p. 121–134.
- 6 *K. Tambs, J. M. Sundet, P. Magnus, K. Berg*. Genetic and Environmental Contributions to the Covariance between Occupational Status, Educational Attainment, and IQ: A Study of Twins // *Behavior Genetics*, 1989, March 19, № 2, p. 209–222.

Прежде, чем приступить к измерениям



Формулирование задачи по измерению

Когда вы сталкиваетесь со сложной задачей по измерению, полезно поместить ее в определенный контекст. Прежде чем приступить к измерению, задайте себе следующие пять вопросов.

1. Какое решение будет принято с учетом результатов данного измерения?
2. Что на самом деле представляет собой объект измерения?
3. Почему данное измерение необходимо для принятия решения?
4. Что мы знаем об объекте измерения в настоящий момент?
5. Какова ценность проведения дальнейших измерений?

При работе с созданным мною и используемым с 1995 г. методом прикладной информационной экономики я методично отвечаю на эти вопросы, что бы ни приходилось измерять. Я применял метод прикладной информационной экономики при решении более чем 50 сложных проблем в целом ряде организаций¹. Если прекратить процедуру «вопросы — ответы», то можно полностью изменить не только *способ*, но и *объект* измерения.

Первые три вопроса помогают понять, что представляет собой объект оценки в рамках решений, которые планируется принимать по результатам измерения. Если эти результаты вообще имеют значение, то только потому, что они понятным образом влияют на решения и линию поведения. Когда не удастся установить, какие решения будут

затронуты итогами задуманного измерения и способ их воздействия, значит, это измерение просто не имеет никакой ценности.

Например, вы захотели оценить качество продукта. Тогда придется выяснить, на какие факторы повлияет результат оценки, и ответить на более общий вопрос о том, что вообще подразумевается под качеством продукта. Вы хотите использовать полученную информацию для решения об изменении действующего производственного процесса? Если да, то насколько низким должно оказаться качество продукта, чтобы это решение было принято? Вам нужны данные о качестве, чтобы рассчитать премии менеджеров по программе качества? Если да, то по какой формуле будут рассчитываться эти премии? И конечно, прежде всего ответы на вопросы зависят от того, какой смысл вы вкладываете в понятие «качество продукта».

Когда в конце 1980-х годов я работал в компании Coopers & Lybrand, мы консультировали небольшой региональный банк, решивший упростить свои процессы отчетности. Банк использовал систему микрофильмирования для хранения отчетов из филиалов (их поступает более 60 еженедельно). Большинство отчетов были необязательными и составлялись не по требованию регулирующих органов, а потому, что кто-то из руководства считал эту информацию необходимой. В наши дни любой хороший программист на Oracle может сказать, что готовить и хранить такие справки очень легко, но в то время на составление отчетов уходила масса времени. Когда я спросил менеджеров банка, какие решения принимаются на основе этих документов, они смогли вспомнить лишь несколько случаев, когда необязательная отчетность изменила или *только могла изменить* сложившееся мнение. Неудивительно, что отчеты, не использовавшиеся непосредственно для принятия реальных управленческих решений, редко даже просматривали. Кто-то когда-то попросил составлять их, но первоначальная цель была, по всей видимости, забыта. Осознав, что многие отчеты совершенно не нужны для принятия решений, менеджеры поняли и то, что эти отчеты не имеют никакой ценности.

Годы спустя аналогичный вопрос возник у сотрудников аппарата министра обороны (Office of the Secretary of Defense, OSD). Они решили выяснить, какую ценность имеют многочисленные отчеты, составляемые еженедельно и ежемесячно. Когда я попросил их назвать хотя бы одно решение, принимаемое на основе каждого отчета, они обнаружили, что многие из них вообще не используются. Информационная ценность этих отчетов тоже оказалась нулевой.

Прежде чем выбрать конкретный способ измерения, вы должны задать себе еще два вопроса: что известно об объекте измерения сейчас, и какова стоимость информации, которую предполагается получить? Знать объем затрат, безусловно, следует заранее, ведь выбор метода оценки качества зависит от того, стоят эти сведения 10 тыс.

или 10 млн дол. Но определить стоимость информации можно только тогда, когда станет ясно, что мы знаем в настоящий момент и как результаты оценки повлияют на конкретные решения.

В следующих главах мы обсудим несколько примеров, связанных с поиском ответов на вопросы, перечисленные выше. Рассказывая о проблемах этапа, предшествующего измерению, мы покажем, что ответы на эти вопросы, в частности о неопределенности, риске и стоимости информации, сами по себе являются полезными оценками.

Уточним терминологию: что на самом деле означают понятия «НЕОПРЕДЕЛЕННОСТЬ» И «РИСК»

Как уже говорилось, прежде чем приступить к измерению, целесообразно определить, о чем, собственно, идет речь и почему это нас интересует. Хороший пример задачи по измерению, с которой может столкнуться любая компания и которая требует существенных предварительных уточнений, — это оценка надежности информационных технологий. Приступая к решению этой задачи, мы должны задать следующие вопросы: «Что мы понимаем под безопасностью?» и «Какие решения зависят от этой оценки?»

Для большинства людей повышение безопасности означает не только рост числа работников, прошедших соответствующее обучение, или число компьютеров, на которых установлено новое защитное программное обеспечение. Когда безопасность повышается, какие-то риски должны снижаться. Раз это так, нам нужно также знать, что подразумевается под словом «риск». Вообще-то, именно поэтому я и начинаю с примера, связанного с надежностью информационных технологий. Уточнение этой задачи требует того, чтобы мы вместе выяснили смысл понятий «неопределенность» и «риск». И то и другое не только поддается измерению, но и служит ключом к пониманию измерения в целом.

Риск и неопределенность нередко считаются неизмеримыми факторами, однако специалисты любой процветающей отрасли постоянно занимаются их оценкой. Чаще всего мне доводилось консультировать страховые компании. Помню, как-то я анализировал один проект по поручению директора по ИТ некой чикагской страховой компании. Он сказал: «Дуг, проблема с информационными технологиями состоит в том, что они рискованны, а способа оценить существующие риски нет». Я ответил: «Но вы же работаете в страховой компании. В вашем здании целый этаж занимают актуарии. Чем же, по-вашему, они занимаются целый день?» Выражение лица моего собеседника было сродни прозрению. До него дошла вся абсурдность ситуации: человек,

работающий в компании, ежедневно оценивающей риски наступления страховых событий, считает риск не поддающимся измерению.

Смысл слов «неопределенность» и «риск», а также их отличие друг от друга, похоже, не до конца ясны даже некоторым специалистам. Вот, например, что писал в начале 20-х годов прошлого века экономист Чикагского университета Фрэнк Найт:

Неопределенность следует употреблять как нечто кардинально отличающееся от Риска в его привычном понимании, хотя до сих пор четкую грань между ними никто не провел... Существенным фактом является то, что в одних случаях «риск» означает величину, поддающуюся измерению, а в других — нечто совсем иного характера и что возникают важные и принципиальные отличия в развитии явления в зависимости от того, присутствует в нем неопределенность или риск²...

Именно поэтому, давая определение своим терминам, так важно понять, для принятия каких решений мы проводим свои измерения. Найт говорит о непоследовательном и неоднозначном использовании понятий «риск» и «неопределенность» некими не названными им группами людей. Однако это не означает, что и мы обязаны использовать их таким же образом. На самом деле эти понятия довольно регулярно используют в науке принятия решений, причем трактуют их непоследовательно и неоднозначно. Но как бы ни толковали их другие, мы дадим им определения с учетом решений, которые придется принимать нам.

ОПРЕДЕЛЕНИЯ НЕОПРЕДЕЛЕННОСТИ, РИСКА И ИХ ПОКАЗАТЕЛЕЙ

Неопределенность. Отсутствие полной определенности, то есть существование более чем одной возможности. «Истинный» результат (состояние, последствие, стоимость) неизвестен.

Показатель неопределенности. Ряд вероятностей, приписанных ряду возможностей. Например: «Есть 60-процентная вероятность того, что данный рынок за пять лет более чем удвоится, 30-процентная вероятность того, что он вырастет не так заметно, и 10-процентная вероятность того, что за тот же период рынок сократится».

Риск. Такое состояние неопределенности, когда в число возможностей входят убытки, катастрофы или другие нежелательные исходы.

Показатель риска. Набор возможностей с приписанными ими количественными вероятностями и количественно определенным ущербом. Например, «мы полагаем: существует 40-процентная вероятность того, что нефтяная скважина окажется сухой, и в результате мы понесем убытки в размере 12 млн дол. в виде затрат на разведочное бурение».

О том, как определяются эти вероятности, расскажем немного позже, а пока мы, по крайней мере, уточнили, что имеем в виду, то есть выполнили предварительный этап любого измерения. Мы выбрали именно такие определения как наиболее соответствующие той задаче, которую решаем в нашем примере, — задаче оценки информационной защищенности и стоимости безопасности. Но, как мы увидим, эти определения наиболее удобны и при решении *любых* встающих перед нами проблем, связанных с измерениями.

Будут ли другие и впредь использовать неоднозначные термины и вести бесконечные теоретические споры, мало интересует того, кто должен срочно решить поставленную задачу. Слово «сила», например, использовалось в английском языке задолго до того, как сэр Исаак Ньютон дал этому понятию математическое определение. Сегодня оно иногда используется как синоним терминов «энергия» или «мощность» — но только не физиками и не инженерами. Когда термин «сила» используют авиаконструкторы, они точно знают, что имеют в виду в количественном смысле (и те из нас, кто часто путешествует по воздуху, ценят их стремление быть точными).

Теперь, сформулировав, что такое неопределенность и риск, мы получили лучший инструментарий для определения такого понятия, как «безопасность» (или «защищенность», «надежность» и «качество», но об этом чуть позже). Говоря «безопасность повысилась», мы обычно имеем в виду, что отдельные риски снизились. Если исходить из данного нами определения риска, то его снижение должно означать уменьшение вероятности наступления нежелательных событий и (или) масштабов связанного с ними ущерба. Как я уже сказал ранее, именно этот подход позволил мне оценить целесообразность вложения Управлением по делам ветеранов крупной суммы (100 млн дол.) в ИТ с целью повышения надежности информационных технологий.

ПРИМЕРЫ УТОЧНЕНИЯ

ОБЪЕКТА ИЗМЕРЕНИЯ:

ЧЕМУ БИЗНЕС МОЖЕТ ПОУЧИТЬСЯ

И У ГОСУДАРСТВА

Многие государственные служащие представляют себе бизнес как некий сказочный мир высокой эффективности и мотивации, где страх проиграть в конкурентной борьбе заставляет людей трудиться изо всех сил. Как часто можно услышать от них сожаления, что они не работают в бизнесе! А для представителей деловых кругов органы власти (федеральные, штата или иные) — синоним бюрократической неэффективности и немотивированности сотрудников, считающих дни, остав-

шиеся до пенсии. Мне доводилось консультировать и государственные учреждения, и частные компании, и я бы не назвал ни ту, ни другую точку зрения полностью правильной или абсолютно неверной. Многие представители этих двух сторон удивились бы, узнав мое мнение: бизнес мог бы поучиться у государства (или, по крайней мере, у некоторых государственных учреждений) очень многому. На самом деле, в крупных компаниях с их сложной внутренней структурой немало сотрудников, очень далеких от экономических реалий бизнеса, а их работа не менее бюрократизирована, чем у служащих любого государственного органа. И я готов прямо сейчас засвидетельствовать где угодно и перед кем угодно, что в федеральном правительстве США, хотя это, конечно, и самая крупная в истории бюрократическая машина, работает немало мотивированных и любящих свое дело людей. Поэтому я приведу здесь несколько весьма поучительных для бизнеса примеров из практики моих клиентов — государственных учреждений.

Расскажу подробнее о работе по оценке надежности информационных технологий, которую я выполнял для Управления по делам ветеранов и о которой говорилось в предыдущей главе. В 2000 г. Совет директоров по информационным технологиям при Федеральном правительстве США (Federal CIO Council) решил провести своего рода испытания, чтобы сравнить различные методы оценки эффективности. Как следует из его названия, Совет директоров по информационным технологиям — это организация, объединяющая руководителей информационных служб федеральных учреждений и их непосредственных подчиненных. У Совета есть свой бюджет, и иногда он финансирует исследования, представляющие интерес для всех директоров по информационным технологиям федеральных органов. Проанализировав несколько подходов, Совет решил, что должен испытать метод прикладной информационной экономики.

Было решено проверить этот метод на большом пакете мер по повышению информационной безопасности, который был предложен Управлению по делам ветеранов. Моя задача состояла в подборе показателей эффективности для каждой системы, связанной с безопасностью, и оценке самого пакета под пристальным наблюдением Совета. Всякий раз, когда я проводил семинар или презентацию своих результатов, на них присутствовало несколько наблюдателей от Совета — сотрудников разных федеральных органов. В конце каждого проекта они готовили свои отчеты, в которых сравнивали мой метод с другим популярным подходом, использовавшимся в то время в других организациях.

Прежде всего, я озадачил специалистов Управления по делам ветеранов вопросом, который задаю, приступая к решению большинства проблем по измерению: «Что именно вы подразумеваете под информационной безопасностью?» Иными словами, в чем должно будет

проявиться усиление этой безопасности? Что нового мы увидим или обнаружим, если безопасность улучшится или ухудшится? Более того, что такое, по нашему мнению, «величина» безопасности?

Информационная безопасность, возможно, и не такое уж эфемерное, расплывчатое понятие, но участники проекта вскоре обнаружили, что *не вполне уверены* в том, какой смысл в него вкладывают.

Было очевидно, например, что уменьшение частоты и масштабов воздействия «пандемических» вирусных атак можно считать усилением безопасности, но что такое в данном случае «пандемические» и что такое «воздействие»? Также было очевидно, что несанкционированный вход хакера в систему — это нарушение информационной безопасности, но является ли таким нарушением кража ноутбука? А пожар в информационном центре, наводнение или торнадо? На первой же нашей встрече участники проекта установили один факт: хотя все они и считали, что безопасность могла бы быть выше, единого понимания того, что это такое, у них не было.

И дело было вовсе не в том, что разные стороны уже выработали свои, отличные от других представления о безопасности. Проблема заключалась в том, что до этого момента *никто и не задумывался* над смыслом слова «безопасность». Как только члены группы столкнулись с поиском специфических конкретных примеров информационной безопасности, они достигли согласия по поводу однозначной и полной ее модели.

Специалисты из Управления по делам ветеранов решили, что повышение безопасности означает снижение частоты определенных нежелательных событий и уменьшение ущерба от них. Они договорились, что в Управлении к таким событиям относятся вирусные атаки, несанкционированный доступ (логический и физический), а также некоторые другие происшествия (например, утрата центра обработки и передачи данных в результате пожара или урагана). Каждый из этих типов событий влечет за собой определенный тип издержек. В таблице 4.1 перечислены предложенные системы повышения безопасности, события, которые они были призваны предотвратить, и возможные последствия этих событий.

Каждая из предложенных систем уменьшала частоту или тяжесть воздействия конкретных событий. Каждое из этих нежелательных событий привело бы к ряду негативных последствий. Так, вирусная атака обычно снижает эффективность труда, в то время как несанкционированный доступ приводит одновременно к снижению эффективности, убыткам от мошенничества и даже возникновению юридических обязательств в результате неправомерного раскрытия частной информации, например медицинского характера, и т. п.

Выработав эти определения, мы добились более конкретного представления о том, что такое усиленная информационная безопасность,

ТАБЛИЦА 4.1.

Меры по повышению информационной безопасности в Управлении по делам ветеранов

Система безопасности	Предотвращенное событие или с «усеченными» последствиями	Предотвращаемые издержки
Инфраструктура открытого ключа (шифровка и дешифровка и т. д.) Биометрическая регистрация (устройства для распознавания отпечатков пальцев, считывания карточек безопасности и т. д.) Системы обнаружения сетевых атак Программа сертификации соответствия новых систем действующим требованиям безопасности Новое антивирусное программное обеспечение Система отчетности о происшествиях Дополнительная подготовка по безопасности	Пандемические вирусные атаки Несанкционированное проникновение в систему извне (хакеры) или изнутри (служащие) Несанкционированный физический доступ на объекты Прочие инциденты: пожар, наводнение, торнадо и т. д.	Затраты, связанные со снижением эффективности Убытки от мошенничества Возникновение юридических обязательств и неправомерное раскрытие информации Невозможность дальнейшего выполнения миссии (которая в случае Управления по делам ветеранов заключается в заботе о ветеранах)

а значит, и о том, *как ее можно измерить*. На мой вопрос «Что вы замечаете, когда информационная защищенность повышается?» руководство Управления по делам ветеранов могло теперь ответить вполне конкретно. Специалисты поняли: наблюдая за усилением безопасности, они обнаруживают снижение частоты и тяжести последствий перечисленных в таблице 4.1 событий. Они реализовали первый этап измерения.

Конечно, к этому определению можно предъявить какие-то претензии. Вы можете (вполне обоснованно) возразить, что риск пожара не является, строго говоря, риском информационной безопасности. И все же специалисты Управления решили, что в своем учреждении им нужно учитывать и этот фактор. Я думаю, что, если оставить в стороне мелкие разногласия по поводу того, что еще можно включить в данное определение, нам действительно удалось выработать такое базовое определение, которое может использоваться при любой оценке информационной безопасности.

Ранее в Управлении применялся совершенно иной подход к измерению безопасности. Использовались такие показатели, как число сотрудников, закончивших курсы подготовки, и число компьютеров, на которые были установлены определенные программы. Иными словами, *результаты* вообще не измерялись. Все предыдущие усилия были направлены на количественную оценку того, что было легче измерить. До моей работы с Советом директоров по информационным технологиям при федеральном правительстве некоторые считали конечный результат реализации мер по повышению безопасности не поддающимся измерению, и никто не пытался уменьшить неопределенность хоть ненамного.

Разработав необходимые критерии, мы приступили к измерению вполне конкретных вещей. Мы построили на компьютере модель, учитывающую все эти последствия. В сущности, мы просто задали ряд «вопросов Ферми». В контексте вирусных атак эти вопросы звучали следующим образом:

- Как часто происходит средняя пандемическая (охватывающая все Управление) вирусная атака?
- Сколько человек страдает от такой атаки?
- Как снижается эффективность работы этих людей по сравнению с нормальными условиями?
- Как долго эффективность остается сниженной?
- Какие издержки от неэффективности труда несет при этом компания?

Знай мы ответы на все эти вопросы, удалось бы рассчитать, во что обходится компании вирусная атака, по следующей формуле:

$$\begin{aligned} & \text{Среднегодовые убытки от вирусных атак} = \text{Число атак} \times \\ & \times \text{Среднее число пострадавших сотрудников} \times \text{Среднее снижение} \\ & \text{эффективности} \times \text{Средняя продолжительность простоев} \times \\ & \times \text{Годовые затраты на оплату труда} / 2080 \text{ часов в год}^3. \end{aligned}$$

Конечно, эта формула учитывает только потери, связанные со снижением производительности труда. Она ничего не говорит о том, как вирусная атака сказывается на заботе о ветеранах, или о другом ущербе. Тем не менее, хотя некоторые виды потерь и не учитываются, формула, по крайней мере, дает нам осторожную оценку нижней границы убытков. В таблице 4.2 приведены ответы на все эти вопросы.

Указанные интервалы значений отражают неопределенность в оценках экспертов по безопасности, сталкивавшихся ранее с вирусными атаками в Управлении по делам ветеранов. Этими интервалами эксперты пытаются сказать: есть 90-процентная вероятность того, что

ТАБЛИЦА 4.2.

**Последствия вирусных атак в Управлении
по делам ветеранов**

Неизвестная переменная	Интервал значений переменной (вероятность 90%)
Число вирусных атак в Управлении за год (для следующих пяти лет)	2–4
Среднее число пострадавших сотрудников	25 000–65 000
Снижение эффективности (%)	15–60
Средняя продолжительность снижения продуктивности (ч)	4–12
Годовая зарплата и премии одного человека (дол.; сильнее пострадают люди с низкой зарплатой)	50 000–100 000

истинные значения окажутся где-то между верхней и нижней границами. Я научил экспертов оценивать неопределенность. Приобретая данный навык, они, по сути, стали «калиброванными», как любой научный инструмент.

Приведенные интервалы значений могут показаться чисто субъективными, но субъективные оценки некоторых людей бывают очень точными. Мы сочли возможным рассматривать эти интервалы как верные, поскольку серия испытаний доказала: когда наши эксперты говорили, что уверены на 90%, они оказывались правыми в 90% случаев.

Как вы могли убедиться, можно взять такое неоднозначное понятие, как «безопасность», и разложить его на несколько конкретных, вполне поддающихся наблюдению составляющих. Определив, что такое безопасность, Управление по делам ветеранов сделало большой шаг к ее оценке. На тот момент его специалисты еще не проводили наблюдения, которые позволили бы снизить неопределенность. Все, что они совершили, — это количественно оценили имеющуюся неопределенность с помощью вероятностей и интервалов значений.

Оказывается, способность человека оценить шансы можно калибровать — точно так же, как любой научный инструмент калибруется для получения правильных показаний. Калиброванная оценка вероятности — ключ к измерению степени приближенности (достоверности) ваших нынешних знаний о чем-либо. Обучение количественному выражению приближенности своих знаний о неизвестном показав-

теле — важный этап в выборе такого способа его оценки, который отвечал бы вашим потребностям. Воспитанию этого навыка посвящена следующая глава.

■ ПРИМЕЧАНИЯ

- ¹ С августа 1995 г. по август 2006 г. я выполнил 30 договоров с 15 компаниями или государственными агентствами, в том числе ряд договоров, предусматривавших анализ важных решений.
- ² *Фрэнк Найт*. Риск, неопределенность и прибыль / Пер. с англ. М.: Дело, 2003.
- ³ 2080 часов в год — стандарт, используемый Службой управления и бюджета (Office of Management and Budget) и Главным бюджетно-контрольным управлением (Government Accountability Office) для пересчета годовой заработной платы и премий в почасовые ставки.



Калиброванные оценки: что вам известно уже сейчас?

ХОРОШО ЛИ ВЫ ОЦЕНИВАЕТЕ ШАНСЫ?

Сколько часов в неделю работники тратят на разбор жалоб потребителей? Насколько повысятся объемы продаж после новой рекламной кампании? Даже не имея точных ответов на подобные вопросы, вы все-таки что-то знаете. Например, одни значения интересующего вас показателя кажутся более вероятными, чем другие. Понять, что именно вам известно о чем-то сейчас, очень важно: нередко от этого зависит выбор подхода к измерению или даже сама необходимость этого измерения. Нам очень нужен способ выразить то, сколько мы знаем сейчас, каким бы малым ни было наше знание. Кроме того, обязательно следует понять, насколько хорошо нам удастся выразить неопределенность.

Один из способов показать неточность определения величины — выразить ее в виде интервала возможных значений. В статистике интервал, в котором с некоторой вероятностью может содержаться правильный ответ, называется доверительным интервалом (confidence interval, CI); 90-процентный доверительный интервал — это диапазон значений, содержащий правильное с вероятностью 90%. Например, не зная точно, сколько потенциальных потребителей действительно ста-

нут вашими клиентами в следующем квартале, вы думаете, что, вероятно, договоры подпишут не меньше трех и не больше семи человек. Если существует уверенность на 90%, что реальное число окажется больше трех, но меньше семи, то можно утверждать, что ваш 90-процентный доверительный интервал составляет три—семь. Границы такого интервала рассчитывают различными, довольно сложными статистическими интерференционными методами, но вы можете выбрать их, руководствуясь собственным опытом. В любом случае они отражают достоверность ваших знаний об искомом показателе.

Выразить неопределенность наступления конкретных событий в будущем вы можете и с помощью вероятностей. Например, существует вероятность 70%, что данный потенциальный клиент подпишет договор в следующем месяце, однако правильно ли выбрано это значение? Чтобы понять, насколько хорошо человек умеет количественно оценивать неопределенность, нужно проверить судьбу всех потенциальных клиентов, по поводу которых делался квартальный прогноз, и ответить на вопрос: «Менеджер был на 70% (80%, 90%) уверен, что все эти люди вот-вот подпишут договоры, но сколько их реально сделали это — 70% (80%, 90%)?» Таким образом, чтобы узнать, насколько умело мы субъективно оцениваем вероятности, ожидавшиеся результаты нужно сравнить с фактическими.

ДВЕ КРАЙНОСТИ СУБЪЕКТИВНОЙ ОЦЕНКИ

Чрезмерная уверенность наблюдается, когда человек постоянно переоценивает точность своих знаний и оказывается правым реже, чем ожидает. Например, когда такого специалиста просят оценить что-то в виде 90-процентного доверительного интервала, в его пределах оказываются гораздо меньше, чем 90%, правильных ответов.

Недостаточная уверенность проявляется, когда человек постоянно недооценивает точность своих знаний и оказывается правым намного чаще, чем ожидает. Например, когда такого специалиста просят оценить что-то в виде 90-процентного доверительного интервала, в его пределы попадают гораздо больше, чем 90%, правильных ответов.

К сожалению, очень немногие из нас являются от природы калиброванными оценщиками. Большинство либо чересчур уверены, либо, наоборот, слишком не уверены в правильности своих результатов. Определение вероятности наступления недостоверных событий или интервала возможных значений некоего показателя — не такой навык, который приобретается по мере накопления опыта и развития интуиции.

К счастью, теоретические исследования доказали, что если людей специально учат преодолевать систематические ошибки и необъективность, их оценки становятся существенно точнее¹. Калиброванные оценки вероятности активно изучались в психологии принятия решений в 1970-х и 1980-х годах. Чуть менее интенсивно изучаются они и сегодня. В отличие от многих преподаваемых в школах бизнеса методов «науки управления» или «количественного анализа», предлагающих оптимальные пути решения конкретных, четко определенных задач, психология принятия решений занимается тем, как люди реально принимают свои решения, какими бы иррациональными они ни были.

Исследователи обнаружили, что букмекеры обычно оценивают вероятность наступления события точнее, чем, например, менеджеры. Кроме того, они сделали ряд неутешительных открытий о том, как плохо врачи диагностируют непонятные явления: «есть шанс, что эта опухоль окажется злокачественной» или «эта боль в груди, вероятно, вызвана сердечным приступом». Наблюдаемые различия между людьми разных профессий позволили исследователям сделать вывод, что оценка вероятности — это навык, которому можно научиться.

Ученые установили, какими способами эксперты могут определить, страдают они чрезмерной уверенностью, недостаточной уверенностью в себе или имеют еще какие-то отклонения. После самоанализа с помощью ряда методов они способны устранить выявляемые недостатки и оценить достигнутый прогресс. Короче говоря, оказалось, что оценка неопределенности — *это навык, который можно приобрести и который можно совершенствовать*. Таким образом, когда калиброванные менеджеры по продажам говорят, что они на 75% уверены, что новый конкурент не переманит к себе вашего крупного клиента, значит, действительно есть 75-процентная вероятность того, что вам удастся удержать его.

Давайте проверим с помощью небольшого теста, насколько хорошо вам удастся количественно выразить неопределенность. В таблице 5.1 приведены 10 вопросов, на которые нужно ответить с 90-процентной уверенностью, и 10 бинарных вопросов, предполагающих ответы типа «правильно» или «неправильно». Это вопросы из разных областей знания, и если только вы не чемпион по игре «Jeopardy», ответить на них со всей определенностью вряд ли удастся. Но некоторое представление об этих вопросах у вас все же должно быть. Аналогичные упражнения я предлагаю участникам своих семинаров. Разница только в том, что я даю тесты, в которых больше вопросов каждого вида, причем слушатели выполняют их подряд и после каждого я сообщаю о результатах. Обычно продолжительность семинара по калиброванию составляет половину рабочего дня.

Но даже при такой малой выборке удастся получить информацию о ваших навыках. И, что важнее, это упражнение подтолкнет вас к осознанию того, что саму приблизительность ваших оценок можно определить количественно.

УПРАЖНЕНИЕ «КАЛИБРОВКА»

Указание: в таблице 5.1 приведено по 10 вопросов двух видов.

Вопросы с 90-процентным доверительным интервалом. Отвечая на каждый вопрос, укажите нижнюю и верхнюю границы интервала. Помните: интервал должен быть таким широким, чтобы вы были на 90% уверены, что правильный ответ попадает в него.

Бинарные вопросы. Ответьте «правильно» или «неправильно» на каждое из приведенных выше утверждений, а затем обведите кружком число, отражающее степень вашей уверенности в своем ответе. Например, если вы абсолютно уверены в своем выборе, то должны указать, что его вероятность составляет 100%. Когда вы совсем не знаете темы, вероятность правильного ответа должна быть равна вероятности выпадения орла или решки при подбрасывании монеты (50%). В противном случае (возможно, в большинстве случаев) эта вероятность должна составлять 50–100%.

Конечно, легко подглядеть ответы на все эти вопросы, но ведь мы воспользовались данным упражнением, чтобы понять, насколько хорошо вы находите ответы, которые нельзя найти нигде (например, как точно вы оцениваете продажи следующего месяца или возможное повышение производительности труда в результате внедрения новой информационной системы).

Важная подсказка: вопросы различаются по сложности. Одни кажутся легкими, другие — слишком трудными. Но каким бы сложным ни выглядело задание, вы все равно сможете как-то его выполнить. Сосредоточьтесь *на том, что знаете*. А вы, в частности, знаете, какие границы интервала наверняка неправильны (например, всем точно известно, что Ньютон не был современником древних греков или нашим современником). Аналогично, отвечая на бинарный вопрос, вы представляете себе более правдоподобный вариант.

Покончив с этой работой, не торопитесь искать правильные ответы. Сначала проведите небольшой эксперимент: убедитесь, что указанные вами доверительные интервалы — действительно 90-процентные. Возьмите один из вопросов с таким доверительным интервалом, скажем, когда Ньютон опубликовал свою работу о всемирном законе

ТАБЛИЦА 5.1. Тест «Калибровка»

№	Вопрос	90-процентный CI:	
		нижняя граница	верхняя граница
1	Какую скорость (мили/ч) развил в 1938 г. британский паровоз, установивший новый рекорд скорости?		
2	В каком году сэр Исаак Ньютон опубликовал свою работу о всемирном законе тяготения?		
3	Какова длина в дюймах обычной визитной карточки?		
4	В каком году военные создали Интернет, называвшийся тогда Арпанетом?		
5	В каком году родился Уильям Шекспир?		
6	Каково расстояние в милях по воздуху между Нью-Йорком и Лос-Анджелесом?		
7	Какой процент площади квадрата можно закрыть кругом с диаметром, равным стороне квадрата?		
8	В каком возрасте умер Чарли Чаплин?		
9	За сколько дней Луна совершает полный оборот вокруг Земли?		
10	Какого числа по телевидению впервые была показана программа «Gilligan's Island»?		
	Утверждение	Ответ (правильно или неправильно)	Вы уверены на (обведите нужное кружком)
1	Древние римляне были завоеваны древними греками		50% 60% 70% 80% 90% 100%
2	Трехгорбых верблюдов не бывает		50% 60% 70% 80% 90% 100%
3	Галлон нефти весит меньше галлона воды		50% 60% 70% 80% 90% 100%
4	Марс всегда более удален от Земли, чем Венера		50% 60% 70% 80% 90% 100%

ТАБЛИЦА 5.1.
 (окончание)

	Утверждение	Ответ (правильно или неправильно)	Вы уверены на (обведите нужное кружком)
5	Команда «Boston Red Sox» стала первым победителем чемпионата США по бейсболу		50% 60% 70% 80% 90% 100%
6	Наполеон родился на острове Корсика		50% 60% 70% 80% 90% 100%
7	«М» — одна из трех наиболее часто используемых букв		50% 60% 70% 80% 90% 100%
8	В 2002 г. персональный компьютер стоил в среднем не более 1500 дол.		50% 60% 70% 80% 90% 100%
9	Прежде чем стать президентом, Линдон Джонсон был губернатором		50% 60% 70% 80% 90% 100%
10	Килограмм больше фунта		50% 60% 70% 80% 90% 100%

тяготения? Допустим, я дал вам шанс выиграть 1000 дол. одним из двух следующих способов (см. рис. 5.1):

- 1) вы получите 1000 дол., если год публикации книги Ньютона окажется в пределах между верхней и нижней границами указанного вами интервала. Если границы выбраны неверно, то не выиграете ничего;
- 2) вы должны раскрутить круглый диск, разделенный на два неравных сектора, площади которых составляют 90 и 10% соот-

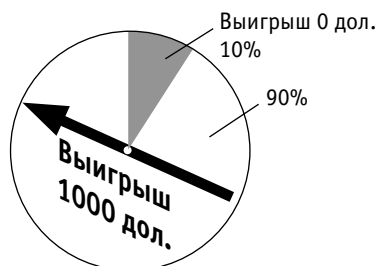


РИСУНОК 5.1.
 Вращайте и выигрывайте!

ветственно. Если указатель остановится на большем секторе диска, то вы выиграете 1000 дол., если же на меньшем, то не выиграете ничего (то есть вероятность выигрыша — 90%).

Какой из вариантов вы выберете? Во втором случае шанс выиграть составляет 90%. Если вы такой же, как большинство (примерно 80%) людей, то предпочтете вращать диск. Но почему? Единственным объяснением может служить ваша убежденность в том, что шансов выиграть, раскручивая диск, гораздо больше. Отсюда вывод: доверительный интервал с вероятностью 90% на самом деле — неправильная оценка, сделанная вами. Эта вероятность, скорее, составляет 80, 65, а то и 50%. С точки зрения статистики это называется чрезмерной уверенностью. Вы выразили неопределенность таким образом, который указывает на гораздо меньшее ее значение по сравнению с фактическим.

Так же нежелателен выбор варианта 1, в котором вы получите 1000 дол., если указанный год попадет в границы вашего интервала. Данный выбор демонстрирует, что на самом деле вы уверены в своем ответе более чем на 90% (хотя и указали, что убеждены в нем именно на 90%).

Правильный ответ в данном случае — сказать, что ни один из вариантов не является предпочтительным. Так вы покажете, что уверены в своем ответе именно на 90% (а не на больше или меньше). Но для этого тем, кто слишком уверен в себе (то есть большинству из нас), пришлось бы предварительно расширить свой доверительный интервал.

То же проверочное упражнение можно проделать и с бинарными вопросами. Допустим, что вы на 80% уверены в правильности своего ответа на вопрос о месте рождения Наполеона. Сделайте снова выбор между способами получения приза (1 или 2). Но пусть в этом случае вероятность выигрыша при вращении диска будет составлять только 80%. Если вы предпочтете раскручивать диск, значит, уверены в своем ответе менее чем на 80%. Пусть вероятность выигрыша при вращении диска сократится до 70%. Если вы и в таком случае сделаете этот выбор, то действительно убеждены в своей правоте только на 70%.

На тренингах по калибровке я называю такую процедуру проверкой с помощью эквивалентных ставок. Как видно из самого названия, она помогает установить, действительно ли вы на 90% уверены в названном интервале значений, предлагая сделать другую ставку, представляющуюся эквивалентной. Исследования показывают: даже когда люди только воображают, что рискуют деньгами, это значительно улучшает их способность оценивать шансы². На деле же выяснилось, что, реально рискуя деньгами, они оценивают шансы лишь ненамного точнее, чем когда рискуют ими понарошку (подробнее об этом в той части главы 13, где обсуждаются рынки предсказаний).

Методы, подобные проверке с помощью эквивалентных ставок, помогают людям точнее оценивать неопределенности. Тех, кто точно определяет степень своей уверенности (то есть тех, кто оказывается прав в 80% случаев, когда говорит, что уверен на 80%), называют калиброванными специалистами. Существуют и другие простые приемы подготовки калиброванных экспертов, но давайте сначала посмотрим, как вы справились с этим тестом. Ответы даны в приложении А.

Чтобы понять, насколько вы калиброваны, необходимо сравнить ожидаемые вами результаты с фактическими. Поскольку вас просили указать 90-процентный CI, вы, в сущности, полагаете, что 9 из 10 правильных ответов окажутся в названных вами интервалах значений. Однако если вы похожи на большинство людей, то процент «прямых попаданий», скорее всего, окажется меньше. Конечно, это очень малые выборки, не позволяющие точно оценить вашу калибровку, но примерное представление они дают. Если даже при такой малой выборке в ваши доверительные интервалы попали менее семи верных ответов, вы, скорее всего, страдаете преувеличенной уверенностью. А когда их оказывается менее пяти (как у большинства людей), вы чрезмерно самонадеянны.

Итак, отвечая на вопросы с 90-процентным доверительным интервалом, вы ожидали, что в предложенные вами границы попадут девять правильных ответов, но фактически их оказалось меньше. Теперь необходимо определить «ожидаемые» значения для бинарных вопросов. Вы указали, что уверены в своих ответах на каждый из них на 50, 60, 70, 80, 90 или 100%. Переведите все обведенные вами кружочком проценты в десятичные дроби (0,5; 0,6 ... 1,0) и суммируйте их. Предположим, что вы были уверены в своих ответах на 1,0; 0,5; 0,9; 0,6; 0,7; 0,8; 0,8; 1,0; 0,9 и 0,7, тогда сумма этих показателей составляет 7,9. Таким образом, «ожидаемый» показатель равен 7,9. Конечно, 10 — тоже небольшая выборка, но если ваш фактический показатель составил, например, 2,5, то, скорее всего, вы слишком самонадеянны.

ДАЛЬНЕЙШЕЕ ПОВЫШЕНИЕ ТОЧНОСТИ КАЛИБРОВКИ

Как показывают теоретические исследования, точность калибровки можно повысить путем обучения. Мы уже упоминали о проверке с помощью эквивалентных ставок, позволяющей человеку представить себя лично заинтересованным в точности оценок. Исследования (и мой опыт) также доказывают возможность повысить способность оценивать неопределенности методом повторения и обратной связи. Он состоит в том, что я задаю обучаемым ряд несложных вопросов, подобных тем, на которые вы только что отвечали. Слушатели дают

свои варианты, затем я знакоблю их с истинными значениями, и они выполняют тест снова.

Однако не похоже, чтобы присущую большинству людей врожденную излишнюю уверенность можно было исправить каким-то одним способом. Поэтому я объединил несколько методов и обнаружил, что в большинстве случаев все вместе они позволяют добиться практически идеальной калибровки человека.

Кроме того, я обычно прошу своих слушателей привести аргументы «за» и «против» для обоснования сделанных ими оценок. Довод «за» — это причина, почему оценка разумна; довод «против» — основание, почему она может быть чрезмерно самоуверенной. Например, ваша оценка продаж нового продукта коррелирует с данными о продажах аналогичных новинок с такими же расходами на рекламу. Однако когда вы задумываетесь о неопределенности прогнозов неожиданных провалов или ошеломительных успехов других компаний наряду с неопределенностями, связанными с общими темпами роста рынка, часто приходится корректировать исходные доверительные интервалы. Исследования показали, что один только этот метод способен значительно повысить точность калибровки³.

Наконец, я прошу экспертов, определяющих диапазон значений, отнестись к своей задаче так, как будто им задали отдельные бинарные вопросы о каждой границе диапазона. Доверительный интервал 90% означает, что есть вероятность 5% того, что истинное значение окажется выше верхней границы предлагаемого диапазона, и такая же вероятность того, что истинное значение окажется ниже нижней границы данного диапазона. Это означает, что эксперт должен быть на 95% уверен, что истинное значение оцениваемой величины будет ниже верхней границы интервала. Если этой уверенности нет, ему придется повысить верхнюю границу интервала так, чтобы такая уверенность появилась. Похожую процедуру выполняют и с нижней границей. Похоже, что выполнение этого теста решает проблему заикливания эксперта на каком-то значении. Заикленность — сужение интервала значений оцениваемой величины до одного пришедшего вам в голову числа. Специалисты часто признаются: называя интервал значений величины, они обычно думают о каком-то одном числе, к которому просто добавляют или из которого вычитают погрешность оценки, чтобы получить границы этого интервала. Такой подход выглядит разумным, но на самом деле заставляет человека сужать диапазоны значений. Избежать подобного заикливания можно, задавая себе бинарные вопросы типа: «Уверен ли я на 95%, что правильное значение окажется выше или ниже данного значения?», касающиеся как верхней, так и нижней границ интервала.

Выполнив ряд калибрующих тестов и попрактиковавшись в использовании таких методов, как составление списка всех «за» и «про-

ТАБЛИЦА 5.2.
 Методы повышения точности калибровки вероятности

Метод	Содержание
1. <i>Повторение и обратная связь</i>	Выполните подряд несколько тестов, оценивая результаты каждого, и повторите их, чтобы улучшить в следующий раз
2. <i>Эквивалентные ставки</i>	Придумайте эквивалентную ставку для каждой оценки, чтобы проверить, действительно ли предложенный диапазон значений или вероятность отражает неопределенность
3. <i>Перечисление двух доводов «за» и двух «против»</i>	Приведите, по крайней мере, две причины, по которым вы уверены в своей оценке, и две причины, по которым вы могли ошибиться
4. <i>Преодоление заикленности</i>	Представьте проблему диапазона оценки в виде двух бинарных вопросов типа: «Уверен ли я на 95%, что истинное значение лежит выше нижней (ниже высшей) границы предложенного мною диапазона?»

тив», проверка с помощью эквивалентных ставок и антизацикливание на одном числе, слушатели приобретают навык в точной настройке «чувства вероятности». Большинство людей становятся практически идеально калиброванными после тренинга продолжительностью всего половину рабочего дня. И, что еще важнее, несмотря на то, что обучение осуществляется на самых общеизвестных проблемах, полученные навыки позволяют проводить оценки в любой сфере деятельности.

В приложении приведены еще два калибрующих теста с вопросами двух видов — бинарными и на интервал значений оцениваемой величины. Попробуйте использовать перечисленные в таблице 5.2 методы для повышения точности своей калибровки.

ПОНЯТИЙНЫЕ ЗАБЛУЖДЕНИЯ, МЕШАЮЩИЕ КАЛИБРОВКЕ

Описанные только что методы оказываются бесполезными, когда у слушателя безотчетно сложились неправильные представления о калибровке или о вероятностях в целом. Хотя, по моим наблюдениям, большинство людей, по долгу службы принимающих важные решения, уже знакомы с теорией вероятности или вполне могут освоить ее, некоторые по-прежнему находятся в плену поразительных заблуждений. Вот, например, что мне доводилось слышать в ходе проведения

занятий по калибровке или когда я по их окончании просил произвести калиброванные оценки:

- моя уверенность на 90% вовсе не означает, что я прав в 90% случаев, поскольку субъективная уверенность на 90% — совсем не то же самое, что объективная;
- это мой 90-процентный доверительный интервал, хотя я понятия не имею, содержит ли он правильный ответ;
- оценить то, что вы просите, вообще невозможно — мы ведь ничего об этом не знаем;
- не зная точного ответа, нельзя подсчитать шансы.

Первое утверждение, сделанное инженером-химиком, свидетельствует о том, что поначалу у него были проблемы с калибровкой. До тех пор, пока человек считает, что субъективная вероятность в чем-то уступает объективной, он не может откалибровать свои оценки. Однако, выполнив ряд калибрующих упражнений, инженер убедился, что его субъективные оценки вероятности оказывались правильными. Другими словами, его 90-процентные доверительные интервалы содержали правильные значения в 90% случаев.

Остальные возражения нам уже знакомы. Все они так или иначе основываются на идее, что тот, кто не знает точного значения какого-либо показателя, вообще ничего о нем не знает. Дама, заявившая, что «понятия не имеет» о том, содержал ли ее 90-процентный доверительный интервал правильное значение, имела в виду один из вопросов на экзамене по калибровке. Речь шла о размахе крыльев «Боинга-747» в футах. Ее ответ был 100–120 футов. Тогда между нами состоялся примерно следующий разговор.

Я: Уверены ли вы на 90%, что правильное значение больше 100 и меньше 120 футов?

Участница семинара: Понятия не имею. Это не более чем догадка.

Я: Но интервал 100–120 футов означает, что вы достаточно уверены в своих знаниях. Данный интервал слишком узок для человека, утверждающего, что он не имеет о предмете ни малейшего представления.

Участница семинара: О'кей. Я не очень уверена в предложенном интервале.

Я: Это означает только, что ваш реальный 90-процентный доверительный интервал значений размаха крыльев самолета должен быть шире. Не кажется ли вам, что размах крыльев может составлять, например, 20 футов?

Участница семинара: Нет, думаю, он не может быть таким маленьким.

Я: Хорошо, тогда, наверное, менее 50 футов?

Участница семинара: Маловероятно. Я бы приняла это за нижнюю границу.

Я: Что ж, мы делаем прогресс. А может ли размах крыльев превышать 500 футов?

Участница семинара (после паузы): Нет, не думаю, чтобы он был так велик.

Я: О'кей. А может ли он превышать размер футбольного поля — 300 футов?

Участница семинара (которая, кажется, поняла, куда я клоню): Ну, хорошо, думаю, что моей верхней границей будет 250 футов.

Я: Значит, вы на 90% уверены, что размах крыльев «Боинга-747» составляет от 50 до 250 футов?

Участница семинара: Да.

Я: Значит, на самом деле ваш 90-процентный доверительный интервал значений размаха крыльев составляет 50–250, а не 100–120 футов.

В ходе нашей беседы эта слушательница скорректировала границы первоначально указанного неправдоподобно узкого интервала значений и определила реальный 90-процентный доверительный интервал, безусловно, улучшив результат. Теперь она уже не могла сказать, что «понятия не имеет», попадает ли в него правильный ответ, так как новый диапазон действительно отражал ее знания.

Этот пример объясняет также, почему я стараюсь не использовать в своем анализе слово «допущение». Допущение предполагает, что в целях доказательства мы считаем некий факт верным независимо от того, так ли это на самом деле. Допущения необходимы, если вы используете детерминистские методы расчета с точными значениями в качестве величин. Точно знать данные значения вы не можете, поэтому каждое такое значение обязано быть допущением. Но если возможно моделировать неопределенность через интервалы значений и вероятности, то пропадает необходимость утверждать что-то, чего вы не знаете наверняка. Когда вы «понятия не имеете», правилен ли этот узкий диапазон, просто расширяйте его до тех пор, пока он не будет отражать ваши знания.

Очень легко потеряться в неизвестном о предмете изучения и забыть, что какие-то факты вы *все-таки знаете*. А там, где границы нашего объекта — плюс-минус бесконечность, вряд ли вообще потребуются что-либо измерять.

Следующий пример несколько отличается от предыдущего диалога с дамой, предложившей нереально узкий диапазон оценки. Разговор состоялся со специалистом по безопасности Управления по делам ветеранов. Сначала он вообще не дал никакого интервала значений, настаивая на том, что объект оценить невозможно. Все началось с утвержде-

ния, что он ничего не знает о величине, но впоследствии оказалось, что некоторые границы ее значений не вызывают у него сомнений.

Я: Если в ваши программы проникает компьютерный вирус, то сколько времени в среднем продолжается сбой в работе? Как всегда, все, что мне нужно, — это 90-процентный доверительный интервал.

Специалист по безопасности: Это определить невозможно. Иногда сбой бывает короткий, а иногда продолжается очень долго. На самом деле мы за этим особенно и не следим, поскольку главная задача — восстановить нормальную работу, а не задокументировать происходящее.

Я: Конечно, точно вы знать не можете. Вот почему мы стремимся определить диапазон, а не точное число. А как долго продолжался самый длительный из случавшихся у вас простоев?

Специалист по безопасности: Не знаю, в разных случаях по-разному...

Я: Было ли когда-нибудь так, чтобы вы не работали целых два рабочих дня подряд?

Специалист по безопасности: Нет, никогда.

Я: А более одного дня?

Специалист по безопасности: Не уверен... Возможно.

Я: Мы хотим определить 90-процентный доверительный интервал для средней продолжительности сбоя. Если вспомнить обо всех сбоях, связанных с вирусами, то могла бы средняя продолжительность перерыва превысить один день?

Специалист по безопасности: Я понял, что вы имеете в виду. Я бы сказал, что среднее значение составляет менее одного рабочего дня.

Я: Так что вашей верхней границей для среднего простоя будет...?

Специалист по безопасности: О'кей. Думаю так: очень маловероятно, чтобы средняя продолжительность сбоя превышала 10 часов.

Я: Давайте теперь подумаем о нижней границе. Какой она может быть?

Специалист по безопасности: Некоторые повреждения устраняются за пару часов, на исправление других уходит намного больше времени.

Я: Отлично, но как вы думаете, могла бы средняя продолжительность сбоев составлять два часа?

Специалист по безопасности: Нет, не думаю, чтобы она была так мала. Думаю, что не менее шести часов.

Я: Хорошо. Итак, ваш 90-процентный доверительный интервал для средней продолжительности сбоев составляет 6–10 часов?

Специалист по безопасности: Я ведь выполнял ваши калибрующие тесты. Дайте мне подумать. Думаю, этот диапазон составляет 4–12 часов.

Такие разговоры совсем не редкость, когда нужно определить величины, характеризуемые высокими неопределенностями. Сначала эксперты вообще отказываются давать интервалы значений — ведь, наверное, все они усвоили в своих компаниях, что отсутствие точного числа равносильно полному незнанию, или это связано с тем, что обычно они несут ответственность за точные данные. Но на самом деле *отсутствие точного числа вовсе не означает полной неосведомленности*. Специалист по безопасности знал, что средняя продолжительность вирусной атаки и устранения последствий никак не может составлять 24 часа (три полных рабочих дня). Кроме того, было известно, что сбой не может составлять один час. Таким образом, кое-какой информацией эксперт обладал, а приблизительность его знаний можно было выразить количественно. Диапазон 6–10 часов говорит о меньшей неопределенности, чем диапазон 2–20 часов. В любом случае эта степень неопределенности интересует нас сама по себе.

Я называю метод, использованный мной в двух приведенных выше диалогах, проверкой на абсурдность и применяю его всякий раз, как слышу, что «этого мы знать не можем» или «я предлагаю диапазон значений, но это только догадка». Какими бы скудными ни считал эксперт свои знания о каком-либо объекте, всегда оказывается, что он четко представляет значения этой величины, невозможные в исследуемых условиях. То значение, которое эксперт считает уже не абсурдным, а лишь маловероятным, и определяет границу его доверительного интервала. А напоследок я предлагаю слушателям сделать эквивалентную ставку и убедиться, что полученный ими в результате доверительный интервал оценки — действительно 90-процентный.

ЭФФЕКТ КАЛИБРОВКИ

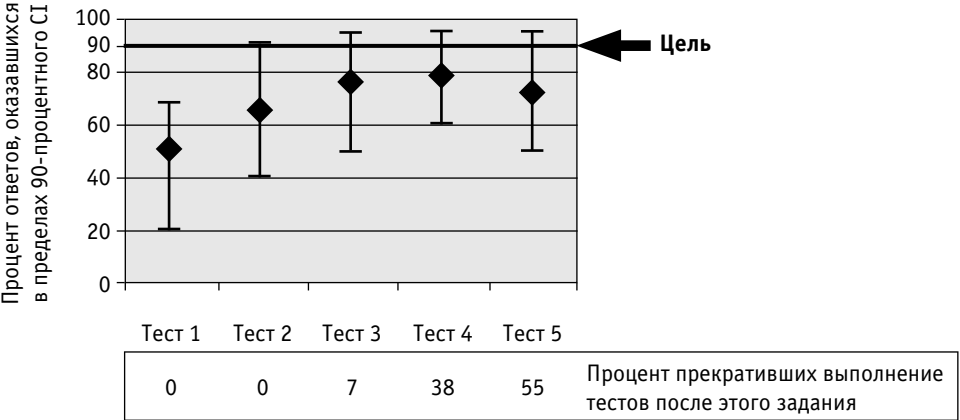
Я занимаюсь такого рода консалтингом с 1995 г., поэтому имею возможность сделать определенные выводы о том, как мои простые тесты и калибровка помогают людям оценивать неопределенность, возникающую в реальных жизненных обстоятельствах.

До 2001 г. мои методы калибровки и тесты постоянно совершенствовались, но с тех пор остаются практически неизменными. За этот период обучение по моей методике прошли в общей сложности 142 человека. Все они выполняли на семинарах продолжительностью половину рабочего дня по несколько тестов подряд, а я сравнивал их ожидаемые и фактические результаты. Будучи хорошо знаком с данными исследований в этой области, я представлял, что слушатели добьются значительных, пусть и не идеальных, успехов. В чем я был меньше уверен, так это в том, какого прогресса добьются разные люди. В теоретических работах обычно приводятся общие результаты всех участ-

ников, так что мы видим лишь средний показатель по группе. Когда я рассчитываю такой параметр для слушателей своих семинаров, то получаю итог, весьма похожий на информацию других исследователей. Но поскольку у меня были данные и по отдельным участникам, я обнаружил очень интересный феномен.

Рисунок 5.2 демонстрирует обобщенные ответы всех 142 слушателей на вопросы, связанные с доверительными интервалами, во всех предложенных на семинаре тестах. Те, кто быстро учился давать калиброванные оценки, освобождались от выполнения последующих заданий (и это оказалось очень хорошим стимулом). В рамке под графиком приведен процент участников, освобожденных от дальнейшего тестирования после выполнения теста под данным номером. Вертикальные линии показывают результаты выполнения каждого теста 90% средних участников, а черные ромбы — среднее значение для группы. Цель слушателей, конечно, заключалась в том, чтобы оказаться на жирной горизонтальной линии, показывающей, что 90% их ответов попадают в указанные 90-процентные доверительные интервалы.

На первый взгляд, график иллюстрирует значительное улучшение результатов выполнения первых трех заданий и последующую стабилизацию на уровне, далеком от идеальной калибровки. Даже принимая во внимание то, что тесты 4 и 5 выполняли только отстающие,



142 участника — результаты 90% участников и средний результат по каждому тесту

РИСУНОК 5.2.

Обобщенные результаты выполнения участниками семинаров по калибровке заданий на 90-процентные доверительные интервалы

складывается впечатление: три-четыре часа интенсивного тренинга еще не позволяют достичь желаемого уровня.

Но взяв данные о каждом слушателе, я обнаружил, что большинство из них добились к концу тренинга большого прогресса, а средние показатели снижаются за счет нескольких отстающих. Статистически мы допускаем, что даже идеально калиброванный эксперт в своих оценках может отклоняться от цели. С учетом одной этой погрешности при тестировании семинар помогает стать идеально калиброванным оценщиком не менее 70% его участников. Больше они не страдают ни излишней самонадеянностью, ни недостатком уверенности в себе. Если эти люди указывают 90-процентный CI, то вероятность попадания правильного ответа именно в этот интервал значений действительно составляет 90%. Еще 20% участников достигают заметного прогресса, но не состояния идеальной калибровки. И только 10% слушателей, похоже, вообще не повышают свой уровень. Так что же, для одного из десятка этот тренинг проходит бесполезно? Нет, это не так. Все, на кого мы когда-либо полагались в фактических оценках, попали в две первые группы и почти все — в группу идеально калиброванных экспертов. А те, кто на первый взгляд вообще не поддался калибровке, еще до тестирования не считались хорошими специалистами в своих областях или людьми, принимающими решения. Возможно, они не были мотивированы в достаточной мере, зная, что их мнения ни на что не влияют. А, возможно, те, кто не склонен к подобным задачам, не стремятся дорасти до уровня экспертов, услугами которых пользуются для проведения оценок. В любом случае это уже вопрос дальнейших исследований.

Мы видим, что большинству занятия идут на пользу. Но свидетельствуют ли успехи на семинарах о способности оценивать шансы в реальной жизни? На этот вопрос можно ответить только утвердительно. У меня было немало возможностей наблюдать, как ведут себя хорошо калиброванные специалисты в жизненных ситуациях, но особенно мне запомнился один управляемый эксперимент. В 1997 г. меня попросили провести курс занятий по оценке вероятности будущего наступления недостоверных событий для аналитиков Giga International Group, компании, занимавшейся консалтингом в области ИТ (впоследствии поглощенной Forrester Research, Inc.). Giga первой из подобных фирм стала продавать результаты своих исследований другим компаниям — своим постоянным подписчикам. Она приняла некий метод определения вероятности событий, наступление которых прогнозировала для своих клиентов, и хотела удостовериться в его корректности.

Я обучил 16 аналитиков Giga использованию описанных выше приемов. В конце занятий я попросил их высказать свое согласие или несогласие с 20 конкретными прогнозами, касавшимися индустрии информационных технологий, и указать степень уверенности в сво-

их оценках. Это задание они выполняли в январе 1997 г., а во всех прогнозах речь шла о событиях, которые могли произойти до 1 июня того же года (например, участники должны были указать, верно или неверно предположение, что Intel выпустит до 1 июня свой процессор Pentium 300 МГц и т.д.). В качестве контрольного эксперимента я познакомил с этими предсказаниями 16 руководителей информационных служб различных организаций. После 1 июня мы могли уже определить, какие прогнозы сбылись, а какие нет. О полученных результатах я рассказал на крупном симпозиуме Giga World (см. рис. 5.3).

По горизонтали отложены указанные участниками занятий вероятности правильной оценки прогноза, по вертикали — процент сбывшихся прогнозов. Результаты идеально калиброванного эксперта представлены пунктиром. Такой специалист оказывается прав в 70% случаев, если уверен в своих предсказаниях на 70%; в 80% случаев, если уверен в своих предсказаниях на 80%, и т.д. Вы видите, что результаты аналитиков Giga (обозначенные квадратами) в пределах допустимой погрешности были очень близки к идеальной уверенности. Сильнее всего эти результаты отклоняются от «идеальной калибровки» в нижней части шкалы, но и здесь это отклонение не превышает допустимой ошибки (в левой части графика интервал допустимой ошибки шире, а в правой он сужается до нуля). Участники оказались правыми в 65% всех случаев, когда они указали, что уверены на 50%. Это означает, что они знали больше, чем полагали, и что были недостаточно увере-

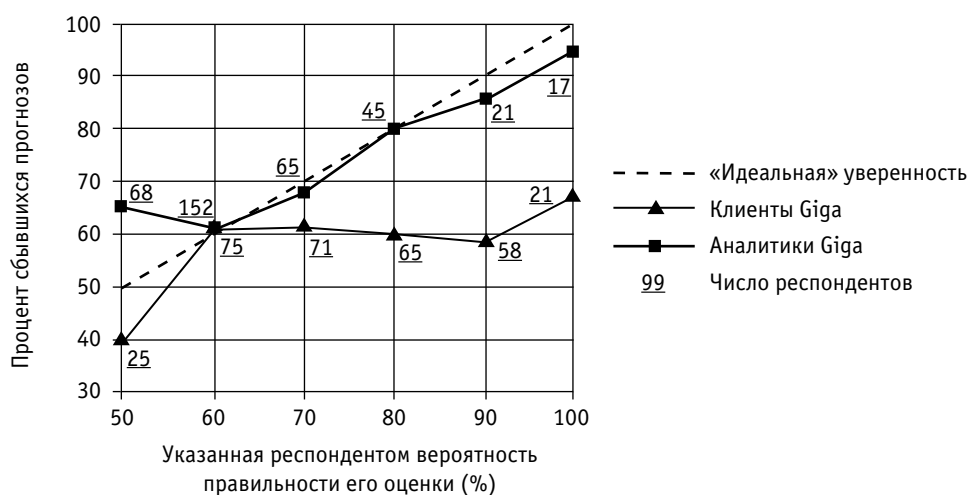


РИСУНОК 5.3.

Результаты калибровочного эксперимента по оценке точности 20 прогнозов развития ИТ-индустрии на 1997 г.

ны в себе (только в этой части шкалы). Но это отклонение невелико и вполне могло объясняться случайностью. Имеется вероятность 1% того, что не менее 44 человек из 68 угадают ответ, просто подбросив монету. В другой части шкалы это отклонение более значительно, по крайней мере статистически, если не визуально. Случайностью могло бы объясняться несколько меньшее отклонение от ожидаемого, поэтому в этой части шкалы слушатели несколько самоувереннее. Но в целом они очень хорошо калиброваны.

Для сравнения, результаты клиентов Giga (обозначенные треугольниками), не прошедших тренинга по калибровке, показали их чрезмерную самонадеянность в прогнозировании. Цифры рядом с результатами калибровки означают, например: тот или иной клиент 58 раз указывал, что уверен в правильности данного прогноза на 90%. Однако сбылось только 60% таких предположений. Клиенты (21%), заявившие, что убеждены в правильности прогноза на 100%, угадали только в 67% случаев.

Не менее интересен тот факт, что у аналитиков Giga правильных ответов оказалось не больше (вопросы были составлены по отрасли в целом, а не по узким направлениям специализации аналитиков). Просто они проявили большую (но не чрезмерную) осторожность, указывая процент уверенности в своей оценке прогноза. Однако до участия в занятиях аналитики, отвечая на вопросы общего характера, так же плохо оценивали неопределенность, как и клиенты, оценивавшие достоверность прогнозов реальных событий. Вывод очевиден: разница в точности определяется только занятиями по калибровке способности слушателей оценивать шансы, весьма действенной в реальных жизненных ситуациях.

Хотя у некоторых участников занятий и возникали трудности с калибровкой, большинство воспринимают ее с готовностью и считают способность оценивать шансы важнейшим навыком, необходимым для проведения измерений. Пат Планкетт, менеджер по оценке эффективности информационных технологий Министерства жилищного строительства и городского развития (Department of Housing and Urban Development), пожалуй, лучше всех в американском правительстве разбирается в использовании показателей эффективности. Он знаком со многими специалистами различных учреждений, прошедшими калибровку с 2000 г. Планкетт в 2000 г. еще работал в Управлении служб общего назначения (General Service Administration, GSA), и именно он стоял за экспериментом Совета директоров по информационным технологиям при федеральном правительстве США и рекомендовал Управлению по делам ветеранов внедрить эти методы. Планкетт считает калибровку серьезным шагом вперед в решении проблемы неопределенности. Он сказал: «Калибровка открыла нам глаза. Многие, включая меня самого, обнаружили, что излишне оптимистичны, делая

оценки. Калибровка делает вас другим человеком. Вы приобретаете обостренную способность оценивать степень неопределенности».

Возможно, единственный сотрудник правительственного учреждения США, повидавший больше прошедших калибровку людей, чем Планкетт, — это Арт Койнз, старший консультант по политике Агентства по защите окружающей среды, десятки специалистов которого прослушали мои семинары. Как и Планкетт, он был удивлен отношением обучаемых к калибровке: «Люди высидели до конца все занятия и поняли их ценность. Их явное желание научиться делать калиброванные оценки стало для меня большим сюрпризом — я ожидал, что они вообще откажутся отвечать на вопросы о таких неопределенных вещах».

Навык в калибровке очень пригодился и команде Управления по делам ветеранов, оценивавшей проект повышения надежности информационной технологии. Команде нужно было понять, что она уже знает, а что остается неизвестным, и выяснить неопределенность, связанную с надежностью. Первоначальные оценки (все интервалы значений и приписанные им вероятности) отражают приблизительность имеющихся данных о рассматриваемых величинах. Эта неопределенность служит основой для следующих этапов: использования вероятностей в модели принятия решений и расчета стоимости информации.

Теперь, научившись вычислять калиброванные вероятности, вы знаете, как количественно рассчитать текущую неопределенность. Умение правильно оценивать калиброванные вероятности крайне важно на следующих этапах измерения. Из глав 6 и 7 вы узнаете, как использовать калиброванные оценки вероятности для определения риска и стоимости информации.

■ ПРИМЕЧАНИЯ

- ¹ *B. Fischhoff, L. Phillips, S. Lichtenstein*. Calibration of Probabilities: The State of the Art to 1980 // *Judgement under Uncertainty: Heuristics and Biases* / Eds. D. Kahneman, A. Tversky. New York: Cambridge University Press, 1982.
- ² Ibid.
- ³ Ibid.



Оценка риска: введение в моделирование методом Монте-Карло

Лучше быть приблизительно правым, чем точно неправым.

Уоррен Баффетт

Мы установили различие между неопределенностью и риском. Первоначально оценка неопределенности величины сводится к присвоению нами неизвестным переменным калиброванных интервалов значений или вероятностей. При последовательных измерениях чего бы то ни было одновременно оценивается и неопределенность, при этом каждое следующее измерение все больше снижает ее.

Риск — это просто состояние неопределенности, которое влечет за собой возможный ущерб любого рода. Как правило, подразумевается, что этот ущерб довольно значителен. Владение методами калибровки позволяет точно рассчитывать первоначальную неопределенность с помощью диапазонов и вероятностей. То же можно сказать и о применении этих методов для оценки риска.

Приемы, которыми многие организации «измеряют» риск, нельзя назвать достаточно информативными. Методы оценки риска, о которых я собираюсь рассказать, хорошо знакомы актуариям, статистикам и финансовым аналитикам. Но некоторые наиболее популярные способы даже отдаленно не напоминают страховую статистику. Многие организации просто характеризуют риск как «высокий», «средний» и

«низкий» или же оценивают его по пятибалльной шкале. Обнаруживая, что дело обстоит именно так, я иногда спрашиваю, насколько «средним» является данный риск. Вероятность 5% понести убытки, превышающие 5 млн дол., — это какой риск: низкий, средний или высокий? Никто не знает. Что лучше — среднерисковые инвестиции с доходностью 15% или высокорисковый вклад с доходностью 50%? Опять никто не знает.

Чтобы доказать, почему подобные классификации не так полезны, как могли быть, я предлагаю участникам своих семинаров в следующий раз при оплате чеком (или через Интернет) страхового взноса за новую автомашину или дом написать в графе «сумма» вместо числа в долларах слово «средняя» и посмотреть, что из этого выйдет. Тем самым они сообщат своим страховщикам, что желают снизить риск до среднего значения. Скажет ли это страховщику что-нибудь конкретное и разумное? Думаю, что ничего, как и любому другому человеку.

Количественное выражение неопределенности с помощью интервалов, а не точных значений, определить которые невозможно, очевидно, имеет свои преимущества. При использовании интервалов значений и вероятностей не нужно делать допущения, в которых вы не уверены. Но точные значения имеют то преимущество, что их легко суммировать, вычитать, умножать и делить в электронной таблице. А можно ли выполнить подобные действия в электронной таблице, если вместо точных чисел есть только диапазоны? К счастью, это можно достаточно просто осуществить методом Монте-Карло, используя разнообразные компьютерные модели, разработанные специально для таких целей.

Наш учитель измерения Энрико Ферми одним из первых применил на практике то, что впоследствии было названо моделированием по методу Монте-Карло. Метод позволяет генерировать на компьютере большое число сценариев на основе вероятностных исходных данных. Для каждого сценария наугад выбираются конкретные значения каждой неизвестной переменной. Затем их подставляют в формулу для расчета исхода данного сценария. Обычно такие расчеты выполняют для тысяч сценариев.

Ферми использовал метод Монте-Карло для того, чтобы рассчитать поведение большого числа нейтронов. В 1930 г., начиная работу над данной проблемой, он осознавал, что ее нельзя решить обычными методами интегрального исчисления. Однако можно было присвоить некие вероятности определенным результатам при заданных условиях. Ученый понял, что можно взять наугад несколько таких ситуаций и выяснить, как будет себя вести большое число нейтронов в системе. В 1940-х и 1950-х годах несколько математиков продолжили работу над аналогичными проблемами ядерной физики и начали использовать компьютеры для составления случайных сценариев. Наибольшую известность получили труды Станислава Улама, Джона фон Нейманна

и Николаса Метрополиса. Сначала они работали над созданием атомной бомбы (Манхэттенский проект), а позднее — водородной бомбы в Лос-Аламосе. По предложению Метрополиса Улам назвал свой метод компьютерного генерирования случайных сценариев методом Монте-Карло в честь своего дяди — азартного игрока¹.

То, что начал Ферми и в последствии продолжили Улам, фон Нейманн и Метрополис, сегодня широко используется в бизнесе, государственном управлении и научных исследованиях. Этот метод применяется, в частности, для определения доходности инвестиций в случае, когда точные затраты и выгоды не известны. Как-то руководитель информационной службы одной чикагской инвестиционной компании советовалась со мной по поводу того, как определять ценность информационной технологии. Она сказала: «Мы очень неплохо научились оценивать риски, но даже не представляем себе, как можно рассчитать выгоды».

При ближайшем рассмотрении ситуация оказалась очень любопытной. Она объяснила, что, вкладывая средства в информационные технологии, компания прежде всего старается добиться повышения на несколько базисных пунктов (1 б.п. = 0,01%) доходности инвестиций, которыми она управляет по поручению своих клиентов. Руководство надеется, что грамотные инвестиции в ИТ позволят добиться конкурентного преимущества в сборе и анализе данных, на основе которых принимаются инвестиционные решения. Но когда я спросил свою собеседницу, как компания оценивает выгоды сейчас, она ответила, что специалисты «просто выбирают число наугад». Иными словами, пока достаточное число людей соглашались или, по крайней мере, не возражали против предложенного значения, отражающего рост базисных пунктов, именно на нем и строилось экономическое обоснование решения. Конечно, выбранный параметр мог отражать предыдущий опыт, но очевидно, что в нем руководство могло быть уверено не больше, чем в любых других значениях. Но если это так, то как же компания измеряет свои риски? Ясно, что самой серьезной опасностью оказалась бы связанная с неопределенностью получения данной выгоды по сравнению с другими. Компания не пользовалась интервалами значений для выражения неопределенности в росте базисных пунктов, поэтому включить ее в расчет риска было невозможно. Таким образом, хотя руководитель ИТ-службы была убеждена, что в фирме хорошо анализируют риски, похоже, что на самом деле их там даже не рассчитывали.

В действительности *весь* инвестиционный риск можно выразить количественно одним-единственным способом: задав интервалы значений неопределенностям, связанным с затратами и выгодами от решения. Когда вы точно знаете сумму и сроки осуществления затрат и получения выгод (что предполагается при традиционном экономиче-

ском обосновании проектов, когда используются точные числа), риск буквально не существует. Ведь выгоды или затраты не могут оказаться ниже или выше ожидавшихся. Но на самом деле все, что мы о них знаем, — это интервалы, а не точные параметры. А раз у нас есть только широкие интервалы значений, значит, существует и вероятность отрицательной доходности. Вот что лежит в основе оценки риска, и вот для чего служит моделирование методом Монте-Карло.

ПРИМЕР ПРИМЕНЕНИЯ МЕТОДА МОНТЕ-КАРЛО И РАСЧЕТА РИСКА

Приведем очень простой пример моделирования методом Монте-Карло для тех, кто никогда не работал с ним ранее, но имеет определенное представление об использовании электронных таблиц Excel. А люди, уже пользовавшиеся данным методом, могут просто пропустить несколько следующих страниц.

Предположим, что вы хотите арендовать новый станок для некоего этапа производственного процесса. Стоимость годовой аренды станка 400 000 дол., и договор нужно подписать сразу на несколько лет. Поэтому, не достигнув точки безубыточности, вы не сможете сразу вернуть станок. Вы собираетесь подписать договор, думая, что современное оборудование позволит сэкономить на трудозатратах и стоимости сырья и материалов, а также поскольку считаете, что материально-техническое обслуживание нового станка обойдется дешевле.

Ваши калиброванные специалисты по оценке дали примерные интервалы значений ожидаемой экономии. Кроме того, они рассчитали ожидаемые годовые объемы производства для данного процесса:

- | | |
|------------------------------------------------------------------------------|---------------------------------------------------|
| • экономия на материально-техническом обслуживании (maintenance savings, MS) | от 10 до 20 дол.
на единицу продукции; |
| • экономия на трудозатратах (labour savings, LS) | от –2 до 8 дол.
на единицу продукции; |
| • экономия на сырье и материалах (raw materials savings, RMS) | от 3 до 9 дол.
на единицу продукции; |
| • объем производства (production level, PL) | от 15 000 до 35 000
единиц продукции
в год; |
| • стоимость годовой аренды (точка безубыточности — breakeven) | 400 000 дол. |

Рассчитаем теперь годовую экономию по следующей очень простой формуле:

$$\text{Годовая экономия} = (MS + LS + RMS) \times PL.$$

Конечно, этот пример слишком прост, чтобы быть реалистичным. Объем производства каждый год меняется, какие-то затраты снизятся, когда рабочие окончательно освоят новый станок, и т. д. Но мы в этом примере намеренно пожертвовали реализмом ради простоты.

Если мы возьмем медиану каждого из интервалов значений, то получим:

$$\text{Годовая экономия} = (15 + 3 + 6) \times 25\,000 = 600\,000 \text{ (дол.)}.$$

Похоже, что мы не только добились безубыточности, но и получили кое-какую прибыль, но не забывайте — существуют неопределенности. Как же оценить рискованность этих инвестиций? Давайте прежде всего определим, что такое риск в данном контексте. Помните? Чтобы получить риск, мы должны наметить будущие результаты с присущими им неопределенностями, причем какие-то из них — с вероятностью понести ущерб, поддающийся количественному определению. Один из способов взглянуть на риск — представить вероятность того, что мы не добьемся безубыточности, то есть что наша экономия окажется меньше годовой стоимости аренды станка. Чем больше нам не хватит на покрытие расходов на аренду, тем больше мы потеряем. Сумма 600 000 дол. — это медиана интервала. Как определить реальный интервал значений и рассчитать по нему вероятность того, что мы не достигнем точки безубыточности?

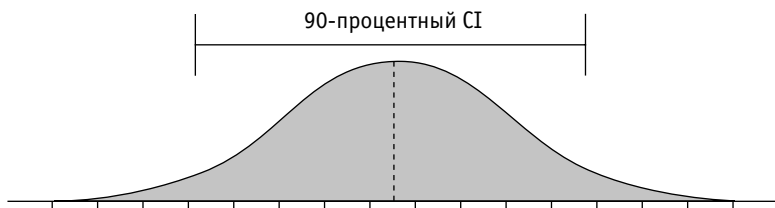
Поскольку точные данные отсутствуют, нельзя выполнить простые расчеты для ответа на вопрос, сможем ли мы добиться требуемой экономии. Есть методы, позволяющие при определенных условиях найти интервал значений результирующего параметра по диапазонам значений исходных данных, но для большинства проблем из реальной жизни такие условия, как правило, не существуют. Как только мы начинаем суммировать и умножать разные типы распределений, задача обычно превращается в то, что математики называют неразрешимой или не имеющей решения обычными математическими методами проблемой. Именно с такой задачей столкнулись физики, работавшие над расщеплением ядра. Поэтому взамен мы пользуемся методом прямого подбора возможных вариантов, ставшим возможным благодаря появлению компьютеров. Из имеющихся интервалов мы выбираем наугад множество (тысячи) точных значений исходных параметров и рассчитываем множество точных значений искомого показателя.

Моделирование методом Монте-Карло — превосходный способ решения подобных проблем. Мы должны лишь случайным образом

выбрать в указанных интервалах значения, подставить их в формулу для расчета годовой экономии и рассчитать итог. Одни результаты превысят рассчитанную нами медиану 600 000 дол., а другие окажутся ниже. Некоторые будут даже ниже требуемых для безубыточности 400 000 дол.

Вы легко сможете осуществить моделирование методом Монте-Карло на персональном компьютере с помощью программы Excel, но для этого понадобится чуть больше информации, чем 90-процентный доверительный интервал. Необходимо знать *форму* кривой распределения. Для разных величин больше подходят кривые одной формы, чем другой. В случае 90-процентного доверительного интервала обычно используется кривая нормального (гауссова) распределения. Это хорошо знакомая всем колоколообразная кривая, на которой большинство возможных значений результатов группируются в центральной части графика и лишь немногие, менее вероятные, распределяются, сходя на нет к его краям (см. рис. 6.1).

Вот как выглядит нормальное распределение:



Особенности:

- значения, располагающиеся в центральной части графика, более вероятны, чем значения по его краям;
- распределение симметрично, без переходов. Медиана находится точно посередине между верхней и нижней границами 90-процентного доверительного интервала;
- «хвосты» графика бесконечны. Значения за пределами 90-процентного доверительного интервала маловероятны, но все же возможны.

Как построить случайное распределение данного вида в среде Excel:

= norminv(rand(), A, B).

Здесь A (медиана) = (Верхняя граница 90-процентного CI + Нижняя граница 90-процентного CI)/2;

B (стандартное отклонение) = (Верхняя граница 90-процентного CI – Нижняя граница 90-процентного CI)/3,29.

РИСУНОК 6.1.

Нормальное распределение

Говоря о нормальном распределении, необходимо упомянуть о таком связанном с ним понятии, как стандартное отклонение. Очевидно, не все обладают интуитивным пониманием, что это такое, но поскольку стандартное отклонение можно заменить числом, рассчитанным по 90-процентному доверительному интервалу (смысл которого интуитивно понимают многие), я не буду здесь подробно на нем останавливаться. Рисунок 6.1 показывает, что в одном 90-процентном доверительном интервале насчитывается 3,29 стандартного отклонения, поэтому нам просто нужно будет сделать преобразование.

В нашем случае следует создать в электронной таблице генератор случайных чисел для каждого интервала значений. Следуя инструкциям рисунка 6.1, мы можем выбрать случайные значения параметра MS (экономия на материально-техническом обслуживании) по формуле Excel:

$$=norminv(rand(),15,(20 - 10)/3,29).$$

Продедаем то же самое для остальных интервалов значений. Некоторые предпочитают пользоваться генератором случайных цифр из Excel Analysis Toolpack. Я привожу эту формулу в таблице 6.1 на случай, если вы захотите попрактиковаться.

ТАБЛИЦА 6.1.

Представление простого расчета методом Монте-Карло в Excel

Сценарий №	Экономия на материально-техническом обслуживании (дол.)	Экономия на трудозатратах (дол.)	Экономия на сырье и материалах (дол.)	Объем произведенной продукции (шт.)	Общая экономия (дол.)	Достигается ли безубыточность?
1	9,27	4,30	7,79	23 955	511 716	Да
2	15,92	2,64	9,02	26 263	724 127	Да
3	17,70	4,63	8,10	20 142	612 739	Да
4	15,08	6,75	5,19	20 644	557 860	Да
5	19,42	9,28	9,68	25 795	990 167	Да
6	11,86	3,17	5,89	17 121	358 166	Нет
7	15,21	0,46	4,14	29 283	580 167	Да
						
9 999	14,68	(0,22)	5,32	33 175	655 879	Да
10 000	7,49	(0,01)	8,97	24 237	398 658	Нет

Построим на основе наших данных таблицу, аналогичную таблице 6.1. Два последних столбца показывают результаты расчетов на основе данных других столбцов. В столбце «Общая экономия» показана годовая экономия, рассчитанная для каждой строки. Например, в случае реализации сценария 1 общая экономия составит $(9,27 + 4,30 + 7,79) \times 23\,955 = 511\,716$ (дол.). Столбец «Достигается ли безубыточность?» вам на самом деле не нужен. Я включил его просто для информативности. Теперь скопируем таблицу и сделаем 10 000 строк.

Чтобы оценить полученные результаты, можно использовать пару других простых инструментов, имеющихся в Excel. Функция «=countif()» позволяет рассчитывать значения величин, отвечающие определенному критерию, — в данном случае те значения годовой экономии, которые меньше 400 000 дол. Или же для полноты картины вы можете воспользоваться инструментом построения гистограмм из Analysis Toolpack. Он подсчитает число сценариев в каждой области, или инкрементной группе значений. Затем вы строите график, отображающий результаты расчета (см. рис. 6.2). Этот график показывает, сколько из 10 000 сценариев будут иметь годовую экономию в том или ином интервале значений. Например, в более чем 1000 сценариев годовая экономия составит от 300 000 до 400 000 дол.

Из всех полученных значений годовой экономии примерно 14% будут меньше 400 000 дол. Это означает, что вероятность ущерба составляет 14%. Данное число и представляет содержательную оцен-

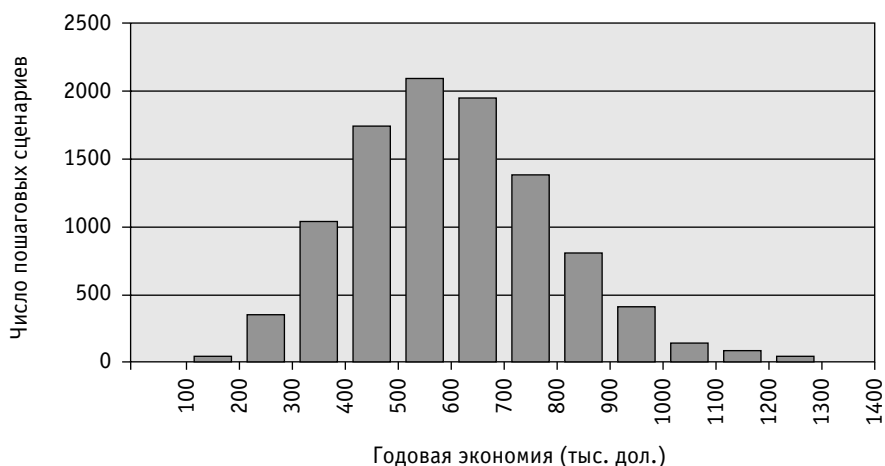


РИСУНОК 6.2.

Гистограмма результатов расчета

ку риска. Но риск не всегда сводится к возможности отрицательной доходности инвестиций. Оценивая размеры вещи, мы определяем ее высоту, массу, обхват и т.д. Точно так же существуют и несколько полезных показателей риска. Дальнейший анализ показывает: есть вероятность 3,5% того, что завод вместо экономии будет терять ежегодно по 100 млн дол. Однако полное отсутствие доходов практически исключено. Вот что подразумевается под анализом риска — мы должны уметь рассчитывать вероятности ущерба разного масштаба. Если вы действительно измеряете риск, то должны делать именно это. С дополнительным материалом к данному примеру можно ознакомиться на сайте www.howtomeasureanything.com

В некоторых ситуациях можно пойти более коротким путем. Если все распределения значений, с которыми мы работаем, будут нормальными и нам надо просто сложить интервалы этих значений (например, интервалы затрат и выгод) или вычесть их друг из друга, то можно обойтись и без моделирования методом Монте-Карло. Когда необходимо суммировать три вида экономии из нашего примера, следует провести простой расчет. Чтобы получить искомый интервал, используйте шесть шагов, перечисленных ниже:

- 1) произвести вычитание среднего значения каждого интервала значений из его верхней границы. Для экономии на материально-техническом обслуживании — $20 - 15 = 5$ (дол.), для экономии на трудозатратах — 5 дол. и для экономии на сырье и материалах — 3 дол.;
- 2) возвести в квадрат результаты первого шага — $5^2 = 25$ (дол.) и т.д.;
- 3) суммировать результаты второго шага — $25 + 25 + 9 = 59$ (дол.);
- 4) извлечь квадратный корень из полученной суммы (получится 7,68 дол.);
- 5) сложить все средние значения: $15 + 3 + 6 = 24$ (дол.);
- 6) прибавить к сумме средних значений или вычесть из нее результат шага 4 и получить в итоге верхнюю и нижнюю границы диапазона: $24 + 7,68 = 31,68$ (дол.) — верхняя граница; $24 - 7,68 = 16,32$ (дол.) — нижняя граница.

Таким образом, 90-процентный доверительный интервал для суммы трех 90-процентных доверительных интервалов по каждому виду экономии составляет 16,32–31,68 дол. В итоге область значений (размах) суммарного интервала равна квадратному корню из суммы квадратов областей значений отдельных интервалов.

Иногда нечто похожее делают, суммируя все «оптимистические» значения верхней границы и «пессимистические» значения нижней

границы интервала. В данном случае мы получили бы на основе наших трех 90-процентных доверительных интервалов суммарный интервал 11–37 дол. Этот интервал несколько шире, чем 16,32–31,68 дол. Когда такие расчеты выполняются при обосновании проекта с десятками переменных, расширение интервала становится чрезмерным, чтобы его игнорировать. Брать самые «оптимистические» значения для верхней границы и «пессимистические» для нижней — все равно что думать: бросив несколько игральные кости, мы во всех случаях получим только «1» или только «6». На самом же деле выпадет некое сочетание низких и высоких значений. Чрезмерное расширение интервала — распространенная ошибка, которая, несомненно, часто приводит к принятию необоснованных решений. В то же время описанный мной простой метод прекрасно работает, когда у нас есть несколько 90-процентных доверительных интервалов, которые необходимо суммировать.

Однако наша цель не только суммировать интервалы, но и умножить их на объем производства, значения которого также даны в виде диапазона. Простой метод суммирования годится только для вычитания или сложения интервалов значений.

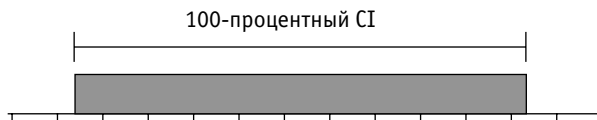
Моделирование методом Монте-Карло требуется и тогда, когда не все распределения являются нормальными. Хотя другие типы распределений не входят в предмет данной книги, упомянем о двух из них — равномерном и бинарном (см. рис. 6.3 и 6.4). И то и другое нам еще встретится, когда мы будем обсуждать стоимость информации.

ИНСТРУМЕНТЫ И ДРУГИЕ РЕСУРСЫ ДЛЯ МОДЕЛИРОВАНИЯ МЕТОДОМ МОНТЕ-КАРЛО

К счастью, сегодня вам не придется строить модель Монте-Карло с нуля. Существуют многочисленные инструменты, облегчающие труд знакомого с теорией аналитика и повышающие его эффективность. К ним относятся и простые наборы макросов для Excel (которыми пользуюсь я), и очень сложные пакеты программ.

Горячим сторонником использования метода Монте-Карло в бизнесе является профессор Стэнфордского университета Сэм Сэвидж, разработавший инструмент, названный им Insight.xls. Сэвидж предлагает интуитивный подход к использованию вероятностного анализа. Кроме того, у него есть кое-какие соображения по поводу формализации процедуры разработки моделей методом Монте-Карло. Если такими моделями пользуются разные подразделения одной и той же организации, то, по мнению Сэвиджа, ей следует иметь объединенную

Вот как выглядит равномерное распределение



Особенности:

- вероятность всех значений одинакова;
- распределение симметрично, без перекосов. Медиана находится точно посередине между верхней и нижней границами интервала;
- значения за пределами интервала невозможны.

Как построить случайное распределение данного вида в среде Excel:

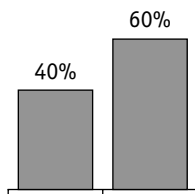
`=rand()*(UB — LB) + LB.`

Здесь UB — верхняя граница;

LB — нижняя граница.

РИСУНОК 6.3. Равномерное распределение

Вот как выглядит бинарное распределение



Особенности:

- возможны только два значения;
- существует единственная вероятность одного значения (в данном случае 60%); вероятность другого значения равна единице минус первая вероятность

Как построить случайное распределение данного вида в среде Excel:

`=if(rand() <P,1,0).`

Здесь P — вероятность выпадения «1» (вероятность выпадения «0» равна 1–P).

РИСУНОК 6.4. Бинарное распределение (распределение Бернулли)

базу общих распределений, а не придумывать каждый раз новые распределения для одних и тех же величин. Более того, он считает, что само определение распределения величины часто представляет собой особую задачу, требующую некоторой математической подготовки.

У Сэвиджа есть интересный метод, который он называет управлением вероятностями: «Предположим, описание распределений вероятностей мы возьмем на себя. Чем тогда вы сможете оправдать свое нежелание пользоваться распределениями вероятностей? Некоторые говорят, что не умеют описывать распределение вероятностей. Однако как выработать электроэнергию, они тоже не знают, но все же ею пользуются».

Его идея заключается в том, чтобы ввести в компаниях такую должность, как ведущий специалист по вероятностям. Он будет отвечать за управление объединенной библиотекой распределений вероятностей, которой может пользоваться каждый, кто занимается моделированием по методу Монте-Карло. Сэвидж ввел такое понятие, как стохастический информационный пакет (stochastic information packet, SIP) — заранее разработанный набор из 100 000 случайных значений того или иного показателя. Иногда разные пакеты связаны друг с другом. Например, доход компании может согласовываться с темпами роста национальной экономики. Набор таких коррелирующих между собой стохастических информационных пакетов называется стохастическими библиотечными модулями с сохраненными зависимостями (stochastic library units with relationships preserved, SLURPs). Ведущий специалист по вероятностям будет управлять SIP и SLURPs таким образом, чтобы пользователям распределений вероятностей не пришлось изобретать велосипед всякий раз, когда нужно моделировать инфляцию или затраты на здравоохранение.

Я бы добавил еще несколько аспектов, тогда модели Монте-Карло будут создаваться и использоваться в организациях так же официально, как методы бухучета. Этими аспектами являются следующие:

- *официальная процедура сертификации калиброванных специалистов.* Проводя занятия, я заметил, что даже высококвалифицированные специалисты, постоянно пользующиеся моделями Монте-Карло, мало что слышали о калиброванной оценке вероятностей. Как мы уже говорили, некалиброванный эксперт обычно слишком уверен в своих оценках. Любой расчет риска на основе его предположений приведет к недооценке;
- *хорошо задокументированная процедура построения моделей* начинающая с исходных оценок калиброванных специалистов. На отладку этой процедуры всегда уходит какое-то время. Организации в большинстве своем не нуждаются в разработке модели с нуля всякий раз, когда приходится анализировать новый инвести-

ционный проект. Можно воспользоваться результатами других предприятий или, по крайней мере, своими предыдущими моделями;

- *единый набор компьютерных инструментов.* Некоторые из них перечислены в таблице 6.2.

ТАБЛИЦА 6.2. Инструментарий метода Монте-Карло

Инструмент	Кем разработан	Описание
@Risk	Palisade Corporation, Итака, штат Нью-Йорк	Достаточно совершенный инструмент для работы на основе Excel; описывает большое число распределений; широкая база пользователей, предоставляется техническая поддержка
AIE	Hubbard Decision Research, Глен-Эллин, штат Иллинойс	Набор макросов на основе Excel; также позволяет рассчитывать стоимость информации и оптимальный портфель; подчеркивает приоритетность методологии над инструментарием; предоставляются консалтинговые услуги по практическим вопросам внедрения
Crystal Ball	Decisioneering, Inc, Денвер, штат Колорадо	Еще один инструмент на базе Excel, продукт, успешно конкурирующий с @Risk. Много пользователей, предоставляется техническая поддержка
Risk Solver Engine	Frontline Systems, Инклин-Вилладж, штат Невада	Уникальная платформа разработки на базе Excel, позволяющая выполнять моделирование методом Монте-Карло с беспрецедентной скоростью. Поддерживает форматы SIP и SLURPs, необходимые для управления вероятностями
SAS	SAS Corporation, Роли, штат Северная Каролина	Пакет программ высшей степени сложности, используемый многими профессиональными статистиками и далеко выходящий за рамки метода Монте-Карло
SPSS	SPSS Inc., Чикаго, штат Иллинойс	Также выходит за пределы метода Монте-Карло; весьма популярен среди ученых
XLSim	Профессор Стэнфордского университета Сэм Сэвидж, AnalyCorp	Недорогой пакет программ, предназначенный для легкого изучения, удобен в применении. Сэвидж проводит в организациях семинары по методу Монте-Карло

ПАРАДОКС РИСКА

Построить модель методом Монте-Карло вряд ли сложнее, чем создать любую другую компьютерную экономическую модель. На самом деле модели, которые я разрабатывал для оценки рисков крупных проектов в области информатики, всегда оказывались более простыми, чем анализируемые информационные системы.

Сложны ли вообще модели Монте-Карло? Конечно, по некоторым меркам. Слишком ли они сложны для того, чтобы их можно было использовать в бизнесе? Нет, по современным деловым стандартам. К тому же, как и в любом другом трудном случае, руководство компании может пригласить специалистов, обладающих необходимыми навыками.

Несмотря на это, количественный анализ рисков с помощью метода Монте-Карло принят далеко не повсеместно. Многие организации пользуются при решении отдельных задач довольно сложными методами анализа риска. Например, актуарии в страховых компаниях определяют требования к страховому продукту, статистики анализируют рейтинги нового телешоу, а менеджеры используют подход Монте-Карло для моделирования изменения методов производства. Однако те же самые организации *не* пользуются регулярно теми же сложными методами анализа риска, принимая гораздо более важные решения в условиях серьезных неопределенностей и высоких вероятностей ущерба.

В 1999 г. мне довелось вести семинар для группы менеджеров, желавших научиться анализировать риски в сфере ИТ. Я начал рассказывать о методе Монте-Карло и спросил, пользуется ли уже кто-нибудь им для оценки рисков. Обычно те, кто утверждает, что рассчитывает риски, просто характеризуют их значения как «высокое», «среднее» или «низкое», даже не пытаясь дать им количественную оценку. Моя цель — помочь слушателям понять разницу между подобным псевдоанализом и расчетом, приемлемым с точки зрения актуария. Один из студентов сказал, что регулярно использует метод Монте-Карло для оценки риска. Это произвело на меня впечатление: «До сих пор мне не встречались специалисты по ИТ, которые пользовались бы данным методом». На это он сказал: «Нет, я не эксперт в информационных технологиях. Я занимаюсь анализом производственных методов в Boise Gascade, целлюлозно-бумажной и деревообрабатывающей компании». На мой вопрос «А что, по вашему мнению, более рискованно — инвестиции в ИТ или в производство бумаги?» он ответил: инвестиции в ИТ рискованнее, но добавил, что его компания никогда не применяет метод Монте-Карло для оценки риска в этой сфере.

ПАРАДОКС РИСКА

Если организация и применяет количественный анализ рисков, то обычно это делается для принятия повседневных оперативных решений. Самые серьезные и опасные решения (чаще всего) формулируются при минимальном использовании полноценного анализа риска.

С годами я обнаружил, что если организация и применяет количественный анализ рисков, то обычно это делается для принятия повседневных оперативных решений. Самые серьезные и опасные решения чаще всего принимаются без предварительного анализа рисков, связанных с ними, — по крайней мере, таких исследований, с которыми согласился бы актуарий или статистик, не проводится. Я назвал этот феномен «парадоксом риска».

Почти все самые сложные методы анализа риска применяются при принятии простых решений, почти не влекущих серьезных негативных последствий, однако решения о слияниях, крупных инвестициях в ИТ, финансировании научных исследований и т. д. обычно формируются без этой процедуры. Почему так происходит? Может быть, из-за существующего мнения о том, что оперативные решения (одобрение кредита или расчет страховой премии) количественно оценить намного проще в отличие от действительно сложных проблем, связанных с рисками, которые с трудом поддаются точному расчету. Это серьезное заблуждение. Как я уже показал, в важных решениях нет ничего «неизмеримого».

Теперь, когда вы уже познакомились с конкретными количественными понятиями неопределенности и риска, мы можем перейти к изучению редко используемого, но очень полезного инструмента измерения: расчета стоимости информации.

■ ПРИМЕЧАНИЕ

- ¹ *Станислав Улам. Приключения математика / Пер. с англ. М.: Регулярная и хаотическая динамика, 2001, 288 с.*

Оценка стоимости информации

С умев рассчитать стоимость самой информации, мы смогли бы определить на ее основе затраты на проведение измерений. Зная же данную стоимость, мы, возможно, предпочли бы измерять совсем другие вещи. Вероятно, мы потратили бы больше усилий и денег, оценивая то, что никогда не оценивали прежде, и потеряли бы интерес к тому, что постоянно количественно определяли до сих пор.

ЗАБЛУЖДЕНИЕ МАКНАМАРЫ

Первый уровень — мерить все, что легко поддается измерению. Этот подход не вызывает возражений. Второй — отбросить то, что трудно измеряется, или приписать ему произвольное количественное значение — искусственный, уводящий в сторону путь. Третий уровень — предположить, что все трудноизмеримое не имеет значения. Это страусиная политика. Четвертый этап — сказать, что измеряемое с трудом вообще не существует. Это самоубийство.

Чарлз Хэнди *,
«The Empty Raincoat» («Пустой плащ»), 1995, с. 219.

* Английский теоретик менеджмента (род. в 1932 г.). — Примеч. переводчика.

На самом деле существуют всего три главные причины тому, что информация имеет свою стоимость для бизнеса.

1. Информация снижает неопределенность в связи с решениями, имеющими экономические последствия.
2. Она влияет на поведение людей, и это также имеет экономические последствия.
3. Иногда информация сама обладает собственной рыночной стоимостью.

Объяснение первой причины из трех перечисленных выше было известно с 1950-х годов. Оно было обосновано в математической теории принятия решений — разделе теории игр. Именно сокращению неопределенности мы и уделим основное внимание, главным образом потому, что оно наиболее актуально для обычных условий, и потому, что две другие причины несколько проще. Ведь стоимость информации, влияющей на поведение людей, просто равна стоимости разницы в их поведении. Если измерение результатов деятельности дает рост производительности труда 20%, то денежное выражение роста производительности и есть стоимость измерения. А если стоимость информации является ее рыночной стоимостью, то перед нами проблема рыночного прогноза, ничем не отличающаяся от оценки продаж любого другого продукта. Если мы собираем информацию об интенсивности движения на городских перекрестках в разное время дня, чтобы продать ее компаниям, ищущим подходящие места для размещения розничных магазинов, то стоимость таких измерений равняется ожидаемой прибыли от продажи этих данных. Однако чаще всего в бизнесе производят измерения потому, что они хотя бы частично связаны с принятием решений. Об этом и пойдет речь в данной главе.

ВЕРОЯТНОСТЬ ОШИБИТЬСЯ И ЦЕНА ОШИБКИ: ОЖИДАЕМЫЕ ПОТЕРИ ОТ УПУЩЕННЫХ ВОЗМОЖНОСТЕЙ

Более 50 лет назад в теории игр — области, понятной лишь посвященным, — была разработана такая формула стоимости информации, которую можно не только вывести математически, но и уяснить интуитивно. Снижение неопределенности (то есть проведение измерений) позволяет делать более удачные ставки (то есть принимать более обоснованные решения). Знать стоимость измерений необходимо, чтобы определить, как можно измерить что-либо и следует ли этим заниматься вообще.

Неуверенность в деловом решении (а калиброванный эксперт должен реалистично оценивать уровень неопределенности) означает, что у вас есть шанс ошибиться. Под ошибкой я понимаю следующее: последствия альтернативного решения могут оказаться предпочтительнее, и, зная об этом, вы, несомненно, выбрали бы его. Цена ошибки — это разница между сделанным вами неправильным выбором и лучшей из имевшихся альтернатив, то есть той, на которой вы остановились бы, обладая полной информацией по вопросу. Например, собираясь вложить деньги в новую смелую рекламную кампанию, вы надеетесь, что эти инвестиции окупятся. Но полностью уверенным в успехе мероприятия вы быть не можете. Известно, что в прошлом многие на первый взгляд прекрасно задуманные и обоснованные рекламные акции не оправдали возлагавшихся на них больших надежд. Некоторые из них даже сыграли на руку конкурентам. В то же время правильно спланированные кампании приводят к значительному росту доходов. Нельзя же сидеть сложа руки и не вкладывать деньги в собственную фирму только потому, что есть вероятность ошибиться. Итак, учитывая всю имеющуюся на данный момент информацию, вы решаете провести свою кампанию, но, возможно, имеет смысл прежде кое-что подсчитать.

Чтобы определить стоимость измерения вероятности успеха намеченных действий, вы должны знать, какие убытки понесете, если инвестиции в кампанию окажутся неудачными, а также какова вероятность провала. Будь эта вероятность полностью исключена, снижать неопределенность вообще не потребовалось бы — ваше решение очевидное и безрисковое.

Чтобы не усложнять наш пример, рассмотрим бинарную ситуацию: вы либо преуспеете, либо провалитесь — вариантов больше нет. Предположим, что вы заработаете 40 млн дол., если реклама сработает, и потеряете 5 млн дол. (затраты на проведение кампании) в другом случае. Допустим также, что ваши калиброванные эксперты говорят, что существует вероятность провала рекламы 40%. Обладая этой информацией, вы можете составить таблицу 7.1.

ТАБЛИЦА 7.1.

Простейший пример расчета потерь от упущенных благоприятных возможностей

	Успех	Провал
Вероятность	60%	40%
План проведения кампании одобрен	40 млн дол.	–5 млн дол.
План проведения кампании отвергнут	0 дол.	0 дол.

Потери от упущенных благоприятных возможностей (opportunity loss, OL) — это просто те затраты, которые мы понесем, если выберем путь, который окажется ошибочным. Ожидаемые потери от упущенных возможностей (expected opportunity loss, EOL) для той или иной стратегии можно рассчитать путем умножения вероятности допустить ошибку на цену ошибки. В нашем примере мы получим такие ответы:

OL (план проведения кампании одобрен) 5 млн дол.

OL (план проведения кампании отвергнут) 40 млн дол.

EOL (план проведения кампании одобрен) $5 \text{ млн дол.} \times 40\% =$
 $= 2 \text{ млн дол.}$

EOL (план проведения кампании отвергнут) $40 \text{ млн дол.} \times 60\% =$
 $= 24 \text{ млн дол.}$

Ожидаемые потери от упущенных благоприятных возможностей возникают из-за того, что вы не знаете, какова вероятность негативных последствий принимаемого решения. Сумей вы снизить данную неопределенность, уменьшится и *EOL*. Именно это и позволяет сделать измерение.

Все измерения, результаты которых имеют некую стоимость, приводят к снижению неопределенности в отношении показателя, влияющего на решение, чреватое экономическими последствиями. Чем сильнее уменьшаются ожидаемые потери от упущенных благоприятных возможностей, тем больше стоимость информации, полученной путем измерения. Разница между значениями *EOL* до и после измерения называется ожидаемой стоимостью информации (expected value of information, *EVI*).

Расчет ожидаемой стоимости информации, получаемой в ходе измерений, до их проведения требует от нас предварительной оценки ожидаемого снижения неопределенности. Иногда это бывает довольно трудно сделать из-за сложности определения некоторых переменных, но возможен и упрощенный подход. Легче всего рассчитать ожидаемую стоимость полной информации (expected value of perfect information, *EVPI*). Если бы существовала возможность полного устранения неопределенности, то значение *EOL* уменьшилось бы до нуля. Таким образом, *EVPI* — это просто *EOL* выбранного вами варианта. В нашем примере решение, принимаемое без осуществления измерений, заключается в одобрении плана проведения рекламной кампании, тогда ожидаемые потери от упущенных благоприятных возможностей составляют 2 млн дол. Таким образом, стоимость устранения любой неопределенности относительно успешности планируемой

акции просто равна 2 млн дол. Если удастся не устранить, а только уменьшить неопределенность, то ожидаемая стоимость информации несколько сокращается.

СТОИМОСТЬ ИНФОРМАЦИИ

Ожидаемая стоимость информации (EVI) = Сокращение ожидаемых потерь от упущенных благоприятных возможностей (EOL):

$EVI = EOL \text{ (до измерений)} - EOL \text{ (после измерений)}$,

где EOL — вероятность ошибиться, умноженная на цену ошибки.

Ожидаемая стоимость полной информации (EVPI) = EOL до измерений (если информация точна и полна, то EOL после измерений равна 0).

Чуть более сложный, но более распространенный и реалистичный метод — это расчет EOL в случае, когда рассматриваются не два экстремума (успех или провал рекламной кампании), а множество значений какой-нибудь величины. Гораздо чаще возникает необходимость рассчитать стоимость измерения, когда неопределенная переменная может принимать значения в некоем интервале. Метод расчета стоимости такой информации принципиально не отличается от того, как мы оценивали стоимость информации в простой бинарной ситуации. Нам, кроме того, потребуется рассчитать EOL.

СТОИМОСТЬ ИНФОРМАЦИИ ДЛЯ ПЕРЕМЕННЫХ ВЕЛИЧИН

Предположим, что в нашем примере с рекламой возможны не два исхода, а результат в виде интервала значений. Калиброванный маркетолог на 90% уверен, что эта рекламная кампания поможет увеличить продажи на 100 тыс. — 1 млн единиц продукции. Однако чтобы достичь точки безубыточности нашей кампании, нужно продать некий объем продукции. Допустим, что с учетом затрат на проведение рекламной акции и валовой прибыли от продукта мы определили наступление точки безубыточности при реализации как 200 тыс. единиц товара. Продав меньше, мы понесем чистые убытки, причем чем меньше объем реализации, тем крупнее эти убытки. Продав ровно 200 тыс. единиц продукции, мы не получим ни прибыли, ни убытков. А если реализовать товар не удастся вообще, то мы потеряем деньги, израсходованные на рекламную кампанию, а именно 5 млн дол. (вы можете сказать, что этим убытки фирмы не ограничатся, но для простоты будем учитывать только их).

Другая точка зрения состоит в том, что на каждой не проданной нами единице продукции, которую надо реализовать для достижения безубыточности, мы потеряем 25 дол. Какова в этой ситуации стоимость снижения неопределенности результата кампании?

Чтобы рассчитать EVPI для подобных интервалов значений, необходимо:

- 1) разбить распределение значений на сотни или тысячи мелких сегментов;
- 2) рассчитать потери от упущенных благоприятных возможностей для медианы каждого сегмента;
- 3) рассчитать вероятность для каждого сегмента;
- 4) умножить потери от упущенных возможностей в каждом сегменте на их вероятности;
- 5) суммировать произведения, полученные на этапе 4 для всех сегментов.

Лучше всего создать для этой цели макрос на базе Excel или написать программу, которая разбила бы распределение значений примерно на 1000 фрагментов, а затем выполнила требуемые расчеты. Так мы гарантированно рассмотрим все важные ситуации и исключения. Чтобы упростить задачу, я уже проделал за вас основную работу. Теперь все, что вам нужно, — это использовать пару следующих графиков и выполнить несколько несложных арифметических расчетов.

Прежде чем приступить к делу, нужно решить, какую из границ 90-процентного доверительного интервала (верхнюю или нижнюю) считать лучшей (best bound, BB), а какую — худшей (worst bound, WB). Ясно, что иногда лучше самое большое число (если, например, речь идет о доходах), а порой — самое маленькое (если мы говорим о затратах). В примере с рекламной кампанией маленькое число — это плохо, то есть WB — 100 тыс., а BB — это 1 млн единиц продукции. По этим данным мы рассчитаем показатель, который я называю «условным порогом» (relative threshold, RT); он указывает, где находится порог относительно остальных значений интервала. Графически RT представлен на рисунке 7.1.

Мы используем условный порог для четырехэтапного расчета ожидаемой стоимости полной информации:

- 1) рассчитаем условный порог: $RT = (\text{Порог} - WB) / (BB - WB)$.
В нашем примере лучшая граница доверительного интервала равна 1 000 000, худшая — 100 000, а порог — 200 000 единиц продукции, поэтому $RT = (200\,000 - 100\,000) / (1\,000\,000 - 100\,000) = 0,11$;

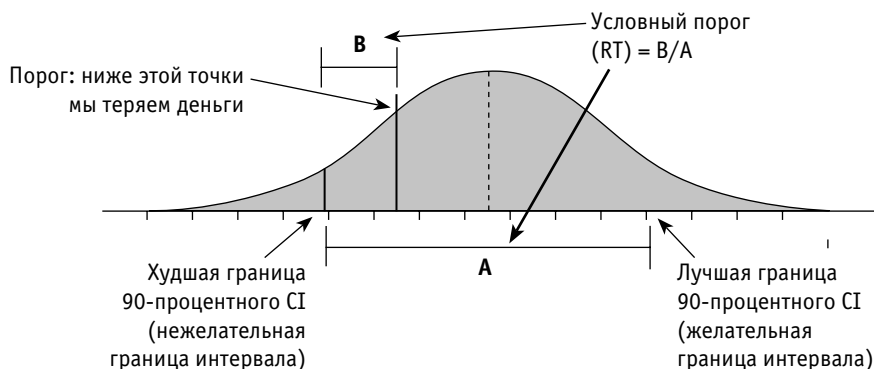
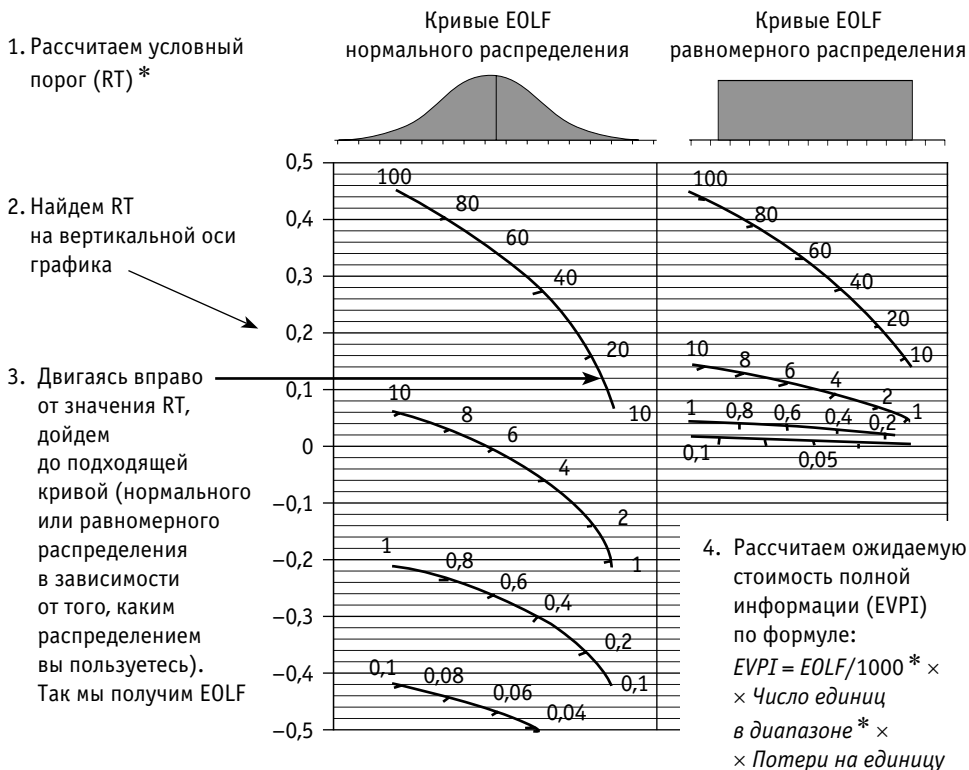


РИСУНОК 7.1. Пример «условного порога»

- 2) найдем местоположение RT на вертикальной оси рисунка 7.2;
- 3) двигаясь вправо от значения RT, мы видим две серии кривых: одну (слева) для нормальных и другую (справа) для равномерных распределений. Поскольку в нашем примере распределение является нормальным, найдем точку пересечения кривой для нормальных распределений с прямой, проведенной через значение RT параллельно горизонтальной оси. Я назову эту величину фактором ожидаемых потерь от упущенной благоприятной возможности (expected opportunity loss factor, EOLF). В данном случае EOLF равняется 15;
- 4) рассчитаем EVPI следующим образом: $EVPI = EOLF / 1000 \times OL$ на единицу продукции $\times (BB - WB)$. В нашем примере OL на единицу продукции равняется 25 дол., поэтому $EVPI = 15/1000 \times 25 \times (1\,000\,000 - 100\,000) = 337\,500$ дол. (см. рис. 7.2).

Расчет показывает, что затраты на проведение измерения (в данном случае на прогноз) объема продаж теоретически составят 337 500 дол. Это абсолютный максимум, определенный исходя из предпосылки, что измерение полностью устранил неопределенность. Хотя сделать это практически невозможно, данный простой метод предоставляет важный ориентир для максимально возможных расходов.

Порядок расчета для равномерного распределения тот же, за исключением, конечно, одного: нам потребуются кривые равномерного распределения. И при нормальном, и при равномерном распределении необходимо учитывать следующие важные оговорки. Во-первых, данный простой метод применим только к убыткам, имеющим линей-



* $RT = (\text{Условный порог} - \text{Худшая граница}) / (\text{Лучшая граница} - \text{Худшая граница})$.
Дополнительную информацию см. на рисунке 7.1.

РИСУНОК 7.2.

График фактора ожидаемых потерь от упущенной благоприятной возможности (EOLF)

ный характер. Это означает, что на каждой непроданной единице продукции мы теряем фиксированную сумму — в нашем примере 25 дол. Если изобразить график зависимости потерь от числа проданных изделий, то он будет иметь вид прямой (окажется линейным). Но когда скорость изменения потерь оказывается непостоянной, график EOLF может оказаться недостаточно точным. Например, если в роли неизвестной выступает сложная процентная ставка, то график потерь при любом пороге не будет представлять собой прямую. Следует отметить и то, что когда речь идет об усеченном нормальном распределении или распределениях, отличающихся от нормальных и равномерных, график может оказаться недостаточно хорошим приближением.

АНАЛИЗ СТОИМОСТИ ИНФОРМАЦИИ НА ВСПОМОГАТЕЛЬНОМ ВЕБ-САЙТЕ

www.howtomeasureanything.com

Перейдите на ссылку «Value of Information Analysis» («Анализ стоимости информации»). Вы можете скачать подробный калькулятор на базе Excel для определения стоимости информации с примерами из этой книги.

Если вы проводите важные измерения с высокой стоимостью информации, возможно, стоит проделать описанные мною дополнительные расчеты и разбить распределение на большое число мелких фрагментов. Но чтобы не составлять такую таблицу с нуля, скачайте таблицы «Анализа стоимости информации» и примеры со вспомогательного веб-сайта www.howtomeasureanything.com

МИР НЕСОВЕРШЕНЕН: СТОИМОСТЬ ЧАСТИЧНОГО СНИЖЕНИЯ НЕОПРЕДЕЛЕННОСТИ

В последнем примере с ожидаемой стоимостью полной информации мы оценили затраты на полное устранение неопределенности, а не ее снижение. Расчет EVPI полезен сам по себе, поскольку, по крайней мере, позволяет узнать потолок стоимости информации, который не должен быть превышен при осуществлении измерений. Однако нередко приходится довольствоваться простым снижением неопределенности, особенно когда речь идет о прогнозе, например, роста продаж в результате проведения рекламных кампаний. В таких случаях полезно знать не только максимальную сумму, которую можно израсходовать в идеальных условиях, но и во что обойдется измерение в реальной жизни (обязательно сопровождаемое реальной погрешностью). Иными словами, нам надо знать ожидаемую стоимость информации, а не ожидаемую стоимость *полной* информации.

Ожидаемую стоимость информации также лучше всего рассчитывать с помощью более сложного моделирования, но мы можем сделать несколько простых оценок. Для этого полезно мысленно представить себе, как выглядит график зависимости EVI от объема информации (см. рис. 7.3).

Кривая EVI обычно имеет выпуклый вид и асимптотический характер. Значит, стоимость информации сначала стремится к быстрому росту с каждым небольшим снижением неопределенности, а затем, когда неопределенность приближается к нулю, стабилизируется. Как

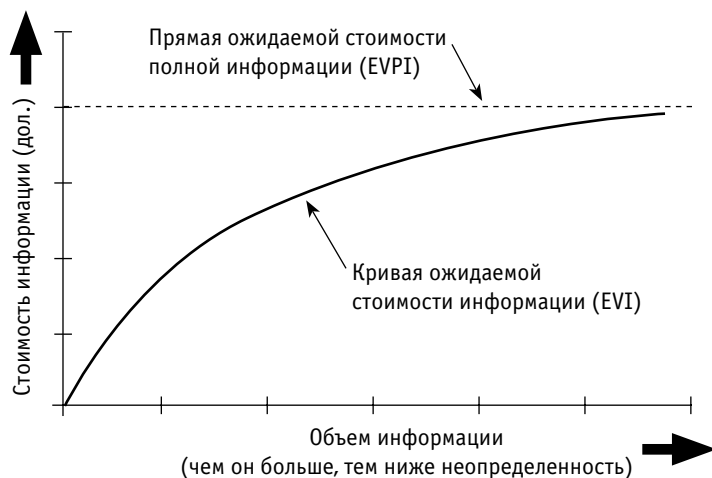


РИСУНОК 7.3. Кривая ожидаемой стоимости информации

видно из рисунка, стоимость информации растет очень быстро, выходя на асимптоту на уровне EVPI (которую, конечно, никогда не превышает).

Необходимо также иметь в виду, что график EVI для нормального распределения обычно ближе к прямой, чем тот же график для бинарного или равномерного распределений. При нормальном распределении ожидаемая стоимость информации, снижающей неопределенность первоначального интервала значений вдвое, равна половине EVPI; информации, снижающей неопределенность в четыре раза, — одной четвертой EVPI и т.д. Кривая EVI для нормального распределения, конечно, не имеет вида прямой, поскольку должна приближаться к значению EVPI. При бинарном или равномерном распределении эта кривая обычно выгнута сильнее, чем при нормальном, поэтому с каждым новым измерением EVI растет быстрее.

Таким образом, стоимость информации, снижающей неопределенность вдвое, обычно превышает половину EVPI. В нашем примере с рекламной кампанией EVPI составляла 337 500 дол. Если вы считаете, что, потратив на исследования в 150 000 дол., удастся сократить неопределенность вдвое, то ваше исследование экономически оправданно (хотя, возможно, и не полностью). А если вы сумеете провести измерение за 30 000 дол., то это, очевидно, будет большой удачей.

Еще одна особенность кривой EVI, о которой необходимо помнить, особенно в случае равномерного распределения, заключается в следующем: график равномерного распределения плоский, резко об-

рывающийся на границах; значения за этими границами невозможны, а все значения внутри них равновероятны. Когда калиброванный эксперт желает отнести наш интервал (100 тыс. — 1 млн проданных единиц продукции) к равномерному распределению, этим он, в сущности, говорит, что вероятность продать больше 1 млн или меньше 100 тыс. единиц продукции равна нулю. Если мы сумеем осуществить измерение, которое, по крайней мере, позволит поднять нижнюю границу до уровня, превышающего порог в 200 тыс. проданных единиц продукции, то возможность убытков будет устранена. В такого рода примерах EVI быстро растет до точки, в которой неопределенность снижается ровно настолько, что появляется возможность устранить вероятность потерь. Разность между стоимостью информации, снижающей неопределенность вдвое и снижающей ее на три четверти, может оказаться весьма небольшой. Как только мы устраним возможность убытков (или определим наверняка, что их избежать не удастся), стоимость результатов любых дополнительных измерений окажется намного ниже.

Хотя описанный метод расчета EVPI с помощью рисунка 7.2 для нормальных распределений является аппроксимацией, погрешность ответа при нормальном распределении не должна превышать 10%. Можно определить ожидаемую стоимость информации, помня о том, что она не должна превысить EVPI, и зная общую форму кривых EVI. На первый взгляд, мы нагромождаем одно приближение на другое, но в итоге получается достаточно точная оценка. Сам по себе расчет EVPI для предложенного измерения включает некую неопределенность, поэтому точность в расчете EVI не имеет большого смысла. Кроме того, стоимость информации о тех величинах, которые необходимо оценить, обычно очень высока. Нередко она в 10 или даже в 100 раз превышает стоимость результатов оценки менее важных переменных. Погрешность оценки EVI обычно ни на что не влияет.

Зная денежную стоимость результатов измерения, мы можем по-новому подойти к вопросу о том, что «измеряемо», а что нет. Когда кто-нибудь заявляет, что измерение стоит слишком дорого, мы должны спросить его: «По сравнению с чем?» Если измерение, которое снизит неопределенность вдвое, обойдется в 50 000 дол., но EVPI — 500 000 дол., то его, безусловно, нельзя назвать слишком дорогим. Но если стоимость информации равна нулю, затраты на любое измерение будут чрезмерными. Стоимость реализации результатов некоторых измерений может быть недостаточной (скажем, несколько тысяч долларов), чтобы экономически оправдать их осуществление, но все же не такой, чтобы эти результаты можно было проигнорировать. В таких случаях я стараюсь придумать приемы, которые позволят быстро снизить неопределенность. Один из них — просто позвонить другим экспертам.

Кривая EVI также показывает значение итеративных измерений. С приближением к обычно недостижимому состоянию полной определенности стоимость информации, как показывает эта кривая, стабилизируется, но кривая затрат на ее получение резко идет вверх. Это говорит о том, что к измерению нужно относиться как к итеративному процессу. Не пытайтесь попасть в десятку с первой же попытки. Каждый следующий этап измерения может сказать вам что-то новое о том, как следует провести очередной этап и стоит ли вообще его проводить.

«УРАВНЕНИЕ ПРОЗРЕНИЯ»:

СТОИМОСТЬ ИНФОРМАЦИИ МЕНЯЕТ ВСЕ

В своей практике консультанта я применяю чуть более сложный вариант описанной выше процедуры.

К 1999 г. я проверил свой метод количественного анализа на базе прикладной информационной экономики примерно на 20 крупных проектах инвестиций в ИТ. В каждом случае нужно было оценить от 40 до 80 величин, таких как первоначальные затраты на разработку, темп восприятия нововведений, рост производительности труда, рост доходов и т.д. При анализе каждого проекта я запускал макрос на основе Excel, который рассчитывал стоимость информации о каждой переменной. Это позволяло мне решить, какие величины необходимо определить в первую очередь.

Работая с этой программой, я стал замечать следующие закономерности.

- Стоимость информации о подавляющем большинстве переменных равна нулю, то есть существующий уровень неопределенности для них вполне приемлем и дальнейшие измерения были бы (это уже упоминалось в главе 3) экономически нецелесообразными.
- Особенно высока стоимость информации о тех переменных, которые клиенты обычно не оценивают. При обосновании предыдущих проектов эти важные величины ни разу не определялись.
- Стоимость информации о переменных, на определение которых обычно тратится больше всего времени и средств, очень невелика или просто равна нулю (то есть крайне маловероятно, чтобы их уточнение влияло на принимаемые решения).

Анализ всей проведенной мной работы по исследованию указанных проектов и расчету стоимости полученной информации позво-

лил подтвердить выявленную закономерность. Я написал на эту тему статью под названием «The IT Measurement Inversion» («Инверсия ИТ-измерений»), которая была опубликована в «CIO Magazine» в 1999 г.¹

Полученные впоследствии данные продолжали подтверждать мои первоначальные наблюдения. Однако я заметил, что данная тенденция характерна для проектов, касающихся не только ИТ, но и военной логистики, защиты окружающей среды, венчурного капитала и расширения производственных мощностей. Клиенты почти всегда удивляются тому, какая информация оказывается для них самой ценной. Снова и снова я убеждался: люди тратят массу времени, сил и денег на измерение того, что не имеет большой информационной стоимости, и игнорируют величины, действительно важные для принятия решений. В конце концов, я отказался от прежнего названия «Инверсия ИТ-измерений» и переименовал этот феномен в инверсию измерений. Ведь тенденция к оценке незначачих вещей и игнорированию важных факторов наблюдается в самых разных областях.

Более того, я часто вижу: начав измерять что-то совершенно иное (осознав его информационную ценность), клиенты рассматривают результат как настоящее открытие. Иными словами, если вы жаждете прозрения, обратите внимание на переменную, которую прежде игнорировали. Все эти мои наблюдения суммированы в рисунке 7.4.

ИНВЕРСИЯ ИЗМЕРЕНИЙ

При обосновании проекта экономическая стоимость результатов измерения переменной обычно обратно пропорциональна тому, какое значение придается ее оценке.

Поскольку организации в большинстве своем незнакомы с методами оценки стоимости проведения измерений, они измеряют совершенно не то, что им нужно. И дело вовсе не в том, что затраты на реализацию проекта не следует измерять, а в том, что им уделяется основное внимание, хотя неопределенность в других вопросах намного выше.

Яркой иллюстрацией инверсии измерений может служить пример моего клиента — крупной британской страховой компании, активно применявшей метод определения сложности и трудоемкости программного обеспечения, называемый балльной функциональной оценкой. Он был популярен в 1980–1990-е годы и использовался для расчета затрат труда на крупные программные разработки. Компания проделала большую работу и собрала первоначальные оценки, балльные

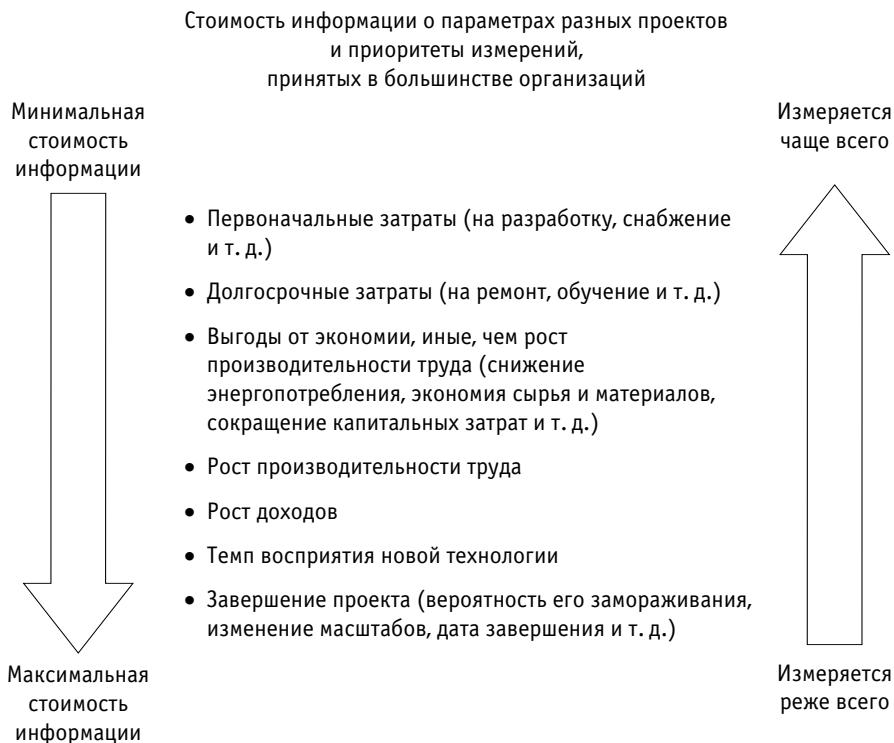


РИСУНОК 7.4. Инверсия измерений

функциональные оценки и данные о фактических затратах труда на реализацию более чем 300 проектов в области информационных технологий. Три-четыре штатных сотрудника занимались исключительно подсчетом баллов. Ранее компании еще не приходилось тратить столько сил на анализ отдельных аспектов планируемых проектов по созданию нового программного обеспечения.

Когда я сравнил балльные функциональные оценки с первоначальными, сделанными менеджерами проектов, и с окончательными затратами, рассчитанными автоматической системой учета рабочего времени, выявилась очень интересная закономерность. Дорогостоящий, занимающий много времени подсчет баллов дал результаты, чуть более точные, чем первоначальные расчеты, но в среднем довольно сильно отличающиеся от фактических затрат. Иными словами, балльная функциональная оценка была иногда ближе, а иногда дальше первоначальной от фактических затрат, определенных по завершении проекта.

Таким образом, компания не только тратила на измерения необычно много времени, но и делала это напрасно, поскольку никакого снижения неопределенности практически не происходило.

Инверсия измерений возникает по нескольким причинам. Во-первых, люди измеряют то, что умеют, или то, что, по их мнению, измерить легче. Наверное, вы знаете старый анекдот о пьяном, который ищет часы на хорошо освещенной улице, хотя знает, что потерял их в темном переулке. Он объясняет это тем, что на улице светлее. Если в компании принято пользоваться для анализа результатами опросов, то все, что определяется другими методами, чаще всего вообще не оценивается. Когда в организации умело пользуются методом анализа баз данных, значит, измеряется только то, что поддается оценке именно этим способом.

На последнем курсе университета мой преподаватель по количественным методам анализа часто говорил студентам: «Если ваш единственный инструмент — молоток, то каждая проблема — это гвоздь». Похоже, сказанное относится и ко многим компаниям и государственным учреждениям. У них есть излюбленные методы измерения. Хотя в некоторых фирмах существуют прекрасно отлаженные способы оценки, например влияния удовлетворенности потребителя на доходы, другие предприятия ими не пользуются и занимаются взамен теми малозначащими измерениями, которые привыкли проводить.

Кроме того, менеджеры любят осуществлять такие измерения, результаты которых могут их порадовать. Ну зачем, в самом деле, оценивать прибыль, если вы подозреваете, что она окажется нулевой? Конечно, в этом случае менеджеры рассуждают, как люди, которые просят деньги или создают видимость работы, а не как руководители, которые подписывают чеки.

Наконец, не зная деловой ценности информации, которая будет получена путем измерения, нельзя оценить и его сложность. Нередко оценки считаются сложными до тех пор, пока не выяснится, что стоимость этой информации намного превысит затраты на ее получение. Одна крупная компания, занимающаяся потребительским кредитованием, как-то попросила меня оценить эффект от инвестиций в ИТ-инфраструктуру на сумму свыше 100 млн дол. Ознакомившись с существом проблемы, я подсчитал, что проведение этой оценки должно обойтись примерно в 100 тыс. дол. Когда компания согласилась потратить не более 25 тыс. дол., я отказался от этой работы. Мой расчет (100 тыс. дол.) составлял менее 0,1% от суммы намеченных этой компанией крайне рискованных инвестиций. В некоторых отраслях принято проводить гораздо более детальный, чем предложенный мной, анализ существенно менее рискованных инвестиционных проектов. По самой осторожной оценке, стоимость информации, полученной в результате таких исследований, составила бы миллионы долларов.

Я называю формулу расчета стоимости информации уравнением прозрения. Дело в том, что, желая совершить действительно важное открытие, вы почти всегда должны обратиться к тому, что раньше не попадало в сферу вашего внимания. Компании, научившиеся рассчитывать стоимость информации, переключают свое внимание совершенно на другие вещи и нередко узнают такое, что заставляет их менять принятые ранее решения.

СВЯЗЬ МЕЖДУ НЕОПРЕДЕЛЕННОСТЬЮ, РИСКОМ И СТОИМОСТЬЮ ИНФОРМАЦИИ: ПЕРВЫЕ ИЗМЕРЕНИЯ

Знание методов измерения неопределенности — ключевой фактор оценки риска. Осознание сути риска на количественном уровне — основа понимания того, как можно рассчитать стоимость информации. Стоимость информации нам нужна для того, чтобы выбрать объект и определить, какие усилия требуются для его измерения. Рассматривая все это в контексте количественного сокращения неопределенности, мы поймем, что же такое измерение.

ЧЕМУ НАС УЧИТ РАСЧЕТ СТОИМОСТИ ИНФОРМАЦИИ

Измерения — процесс итеративный. Самую ценную информацию мы получаем на начальном этапе измерений, поэтому разбейте весь процесс на несколько этапов и подведите итоги каждого из них.

Стоимость информации имеет значение. Не определив заранее эту стоимость, вы, скорее всего, измерите не то и не так.

Резюмируя все сказанное в этой главе, мы можем сформулировать несколько новых идей. Прежде всего, самую ценную информацию получают на начальном этапе измерений. Не стоит планировать масштабные исследования, если требуется измерить то, о чем сейчас вы практически ничего не знаете. Оцените хотя бы что-то, устраните хоть какую-то неопределенность и проанализируйте то, что вы узнали. Результат вас удивил? Нужны ли дальнейшие измерения? Не подсказали ли вам полученные на этом этапе данные, что следует изменить метод измерения? Итеративный подход дает возможность для маневра и максимизирует отдачу от усилий.

Наконец, если вы не рассчитали стоимость полученной информации, значит, измерили то, что, вероятно, не имеет особого значения, и проигнорировали действительно важные факторы. Более того, если

вы не рассчитали стоимость информации, значит, возможно, не знаете, как эффективно измерять что-либо. Вы потратите на это слишком много или слишком мало времени, откажетесь от проведения важных измерений как слишком затратных, так как не можете сопоставить затраты со стоимостью полученной информации.

До сих пор в этой книге мы говорили только о первом этапе измерения того, что нередко часто считается не поддающимся количественной оценке. Взяв как будто слишком расплывчатое понятие, мы определили его смысл с учетом значения, которое оно для нас имеет, и выбрали способы наблюдения. Мы оценили неопределенность, риск и стоимость информации и можем теперь перейти к следующему этапу.

Любопытно, что именно на этом остановилось в свое время Управление по делам ветеранов, реализуя проект оценки надежности информационной технологии. Цель этого проекта состояла только в том, чтобы выявить объекты измерений, которые должны были проводиться в течение нескольких следующих лет. По мнению специалистов Управления, рассчитать стоимость информации было важно само по себе, так как на ее основе в дальнейшем можно было оценить все необходимые показатели надежности.

Далее нам предстоит сделать шаг вперед от простой констатации значения текущей неопределенности и расчета стоимости ее измерения. Теперь, когда мы знаем, что оценивать и сколько можно потратить на измерения, приступим к выбору подходящих способов осуществления этого процесса.

■ ПРИМЕЧАНИЕ

- ¹ *Douglas W. Hubbard. The IT Measurement Inversion // CIO Enterprise Magazine, 1999, April 15.*

Методы измерения



Переход от объекта к способу измерения

Применив уроки предыдущих разделов к своей задаче по измерению, вы определили объект наблюдения, количественно оценили имеющуюся неопределенность, связанную с ним, и рассчитали стоимость информации, которую хотите получить. Все это делается обычно до начала измерений. Теперь же мы должны подумать о том, как сократить неопределенность, иными словами, ее измерить.

Пришло время рассказать о нескольких эффективных и удобных в применении эмпирических методах. Мы уже дали определение термину «измерение», отметим, что часто используемое выражение «эмпирическое измерение», в сущности, избыточно. Ведь эмпирическим является факт, подтверждаемый наблюдениями (возможно, вы даже слышали такую избыточную фразу, как «эмпирическое наблюдение»). Эмпирические методы представляют собой формализованный систематизированный подход к осуществлению наблюдений, позволяющий предотвратить или, по крайней мере, снизить число допускаемых при этом ошибок определенного типа. Вопреки распространенному заблуждению наблюдение вовсе не сводится к визуальному. Наблюдать за явлениями даже не всегда нужно напрямую; эффективность можно повысить, используя инструменты измерения, что почти всегда и делается в естественных науках.

В научной литературе нередко описываются остроумные эмпирические методы, удивляющие и экспертов в данной области, и широкую публику. Многие, наверное, были бы поражены, узнав, что статистики

нашли способ (о котором мы подробнее расскажем позже) определить число танков, произведенных Германией во время Второй мировой войны, на основе всего-навсего серийных номеров захваченных в боях машин. Был найден даже способ установить, когда у младенцев появляется пространственное зрение, хотя они в этом возрасте еще не разговаривают.

Однако мы обсуждаем то, что считают неизмеряемым в бизнесе. К счастью, подобные оценочные задачи, как правило, решаются не самыми сложными методами. Стоит еще раз напомнить, что цель книги — показать: многое из того, что менеджеры компаний считают неподдающимся измерению, на самом деле вполне можно оценить количественно. Вопрос только в том, заслуживает ли данный объект измерения, то есть не превысят ли связанные с этим затраты стоимость полученной в результате информации.

В большинстве своем подобные задачи решаются несколькими сравнительно простыми способами. На самом деле, как мы уже выяснили, реальным препятствием в этом является не отсутствие знания многочисленных сложных методов, а непонимание сущности измерения. В конце концов, в тех отраслях, где используются достаточно сложные методы оценки, дискуссии о возможности измерить тот или иной объект — большая редкость. Ведь все эти сложные методы разрабатывались, поскольку кто-то понял, что какой-то объект измерим. Ну зачем писать двухтомный трактат, например, по количественным методам клинической химии, если автор и целевая читательская аудитория с самого начала были уверены в принципиальной невозможности такого анализа?

Описывать специальные количественные методы, используемые в различных научных дисциплинах, я предоставлю другим. Ведь вы выбрали эту книгу потому, что хотите узнать, как математически строго обращаться с иными, «мягкими» объектами измерения.

В данной главе зададим себе следующие вопросы, которые помогут нам выбрать подходящую категорию методов измерения.

- *Каковы составляющие интересующего вас объекта?* Разложите неизвестный объект так, чтобы по составляющим, характеризующимся собственными неопределенностями, оценить целое.
- *Как интересующий вас объект или его составляющие измерялись ранее?* Вполне возможно, что вы не первый, кто столкнулся с подобной задачей, и что на данную тему даже существует обширная литература. Анализ работ других авторов называется вторичным исследованием.
- *В чем проявляют себя выявленные вами составляющие объекта измерения?* Вы уже ответили на вопрос, как можно наблюдать интересующий вас объект. Прodelайте то же относительно его

составляющих. Возможно, и в этом вам помогут вторичные исследования.

- *Что на самом деле нам необходимо знать, чтобы проводить измерения?* Примите в расчет оцененную ранее нынешнюю степень неопределенности, порог и стоимость информации. Все эти ориентиры помогут вам выбрать правильный подход к измерению.
- *Что может вызвать ошибку?* Подумайте над тем, каким образом наблюдения могут ввести вас в заблуждение.
- *Какой инструмент следует выбрать?* Ответы на предыдущие вопросы помогут вам выбрать или разработать подходящий инструмент измерения. Возможно, как и раньше, полезными окажутся вторичные исследования.

Теперь, не забывая об этих вопросах, мы можем обсудить, какие инструменты используются для измерения.

ИНСТРУМЕНТЫ НАБЛЮДЕНИЯ: ВВЕДЕНИЕ В ИНСТРУМЕНТАРИЙ ИЗМЕРЕНИЙ

Названия предметов и явлений, варианты изменения этих названий со временем немало рассказывают о развитии наших представлений. Хороший тому пример — научные инструменты. До промышленной революции, особенно в эпоху европейского Возрождения, их нередко называли инструментами философского поиска. Научные инструменты использовали для поиска ответов на «глубинные» вопросы того времени. Галилей с помощью маятника и наклонной плоскости, по которой он скатывал шарики, рассчитал ускорение свободного падения (история о том, как он бросал с Пизанской башни разные предметы, скорее всего, вымысел). Даниэль Фаренгейт с помощью ртутного термометра дал количественную оценку тому, что раньше называлось «качеством» температуры. Но эти устройства позволили не только назвать точные числа, но и узнать нечто новое о мире, в котором жили эти ученые. Каждое из устройств предоставило своим создателям возможность наблюдать за той стороной мира, которая ранее была скрыта от всеобщего обозрения.

А тех, кто пользовался этими приборами, называли в те времена даже не учеными, а натурфилософами, которые не пропадали день и ночь, как нынешние исследователи, в научных лабораториях, а были, скорее, хорошо образованными людьми. За немногими исключениями, они располагали и временем, и средствами, необходимыми для конструирования дорогостоящих устройств, или имели состоятельных

покровителей — так же, как многие художники. Используемые ими приборы были нередко не только функциональными, но богато украшенными произведениями искусства и приобретались с целью продемонстрировать хороший вкус, а также научную любознательность (хотя иногда и то и другое было только видимостью).

В конце XIX века, во времена таких изобретателей, как Томас Эдисон и Александр Грейам Белл, исследования и разработки превратились в крупномасштабный бизнес. До этого инструменты нередко изготавливались по индивидуальному заказу; ко времени Эдисона и Белла научные приборы стали стандартными и началось их массовое производство. Кроме того, они приобрели более утилитарный вид. Если раньше просвещенные господа демонстрировали свои новые микроскопы гостям наряду с дорогими предметами искусства, то микроскопы изобретателей времен промышленной революции годились для показа только в лабораториях, условия работы в которых по нынешним меркам были просто каторжными. Неудивительно, что именно тогда наука и научные наблюдения стали ассоциироваться у широкой публики не с увлекательной погоней за глубокими знаниями, а с тяжелым и изнурительным трудом.

Даже сегодня для многих инструмент измерения обычно означает прибор (быть может, элемент сложного электронного оборудования), предназначенный для количественного определения какого-то малоизученного физического явления. Примеры — счетчик Гейгера для регистрации уровня радиации и служащие для определения массы весы. Однако многие специалисты в различных областях вкладывают в термин «инструмент» гораздо более широкий смысл. Так, в сфере образования исследователи называют инструментом опрос, тест или даже отдельный вопрос теста, и такое использование данного понятия вполне правомерно.

Как и любой другой инструмент, инструмент измерения дает своему пользователю определенные преимущества. Простой механический инструмент подобен рычагу, приумножающему мускульную силу человека. Аналогично инструмент измерения улучшает работу его органов чувств, обнаруживая вещи, незаметные при непосредственном наблюдении. Кроме того, он облегчает осмысление и запоминание, ускоряя расчеты и гарантируя сохранение их результатов. Даже тот или иной экспериментальный прием, безусловно, облегчает человеческое восприятие и поэтому тоже может считаться инструментом измерения. И стремясь узнать, как можно измерить все, что угодно, мы должны использовать данный термин именно в самом широком смысле слова.

Нередко менеджеры считают что-либо не поддающимся измерению просто потому, что данные о нем еще не собраны полностью и к ним нет легкого доступа. Они могут придумать, как оценить будущие продажи, — ведь необходимые для этого данные уже собрали и им

предоставили. Но вряд ли таким управленцам сразу придет в голову мысль, что определение некоторых показателей потребует целенаправленного наблюдения с применением таких, например, инструментов измерения, как опрос или эксперимент.

Возможно, менеджеры проявляли бы больше изобретательности в использовании инструментов измерения, обладай они той увлеченностью, с какой охотились за тайнами окружающего мира Галилей и Фаренгейт. Эти естествоиспытатели относились к измерительным приборам не как к сложным приспособлениям, применяемым только посвященными для своих загадочных исследований. Устройства Галилея и Фаренгейта были просты и очевидны. От современных менеджеров они отличались и тем, что не забраковывали инструменты только из-за присущих им недостатков или ошибки измерения. Приборы *всегда* дают какую-то погрешность. Вопрос только: «По сравнению с чем?» По сравнению с наблюдением невооруженным глазом? По сравнению с нежеланием даже попытаться что-то измерить? Не забывайте: цель измерения — снижение неопределенности, а вовсе не обязательно ее полное устранение.

Обычно применение инструментов дает шесть преимуществ. Но если таких преимуществ и меньше, инструмент не перестает быть инструментом. Нередко даже одно преимущество — уже шаг вперед по сравнению с наблюдением невооруженным глазом.

1. *Инструменты обнаруживают то, что вы заметить не можете.* Вольтметр регистрирует напряжение в цепи, микроскоп увеличивает изображение, камера Вильсона фиксирует следы субатомных частиц. Эти свойства инструментов обычно вспоминают первыми, однако значение данного преимущества преувеличивают.
2. *Инструменты более последовательны.* Когда люди полагаются только на свои чувства, их выводы отличаются большой противоречивостью. Любой инструмент, будь то весы или опрос потребителей, как правило, дает более последовательные результаты.
3. *Инструменты можно калибровать для учета ошибки измерения.* Калибровка — это измерение объекта, чьи параметры вам известны, с целью проверки точности самого инструмента. Весы калибруют, поместив на них груз, масса которого, как мы знаем, составляет ровно 1 грамм. Мы калибровали вашу способность оценивать вероятность, задавая вопросы, ответы на которые известны. Таким же образом мы определяем, какую погрешность дает тот или иной инструмент.

Нередко погрешность инструмента можно компенсировать регулировкой. Так, в ходе управляемого эксперимента текущее

состояние измеряемого объекта сравнивают с исходным. Если вы хотите узнать, привело ли внедрение автоматизированной системы продаж к росту числа повторных заказов, то должны сравнить результаты тех торговых представителей, которые ею пользуются, и тех, кто нет. Возможно, одни продавцы применяют данную систему чаще, чем другие, или она внедрена еще не во всех регионах и не для всех продуктовых линеек (подробнее об этом см. следующую главу).

4. *Инструменты намеренно чего-то не замечают.* Инструменты полезны тем, что игнорируют все, что искажает результаты наблюдения невооруженным глазом. Например, когда учителя проверяют экзаменационные работы, на которых не указаны фамилии учащихся, это исключает возможность того, что кто-то поставит завышенную оценку своему любимчику. В клинических научных исследованиях ни доктор, ни пациент не знают, кто принимает лекарство, а кто плацебо. Благодаря этому пациенты не могут исказить информацию о своих ощущениях, а доктор — свой диагноз.
5. *Инструменты регистрируют.* Хороший тому пример — старый добрый электрокардиограф, регистрирующий на длинной бумажной ленте сердечную деятельность. Конечно, сегодня вся документация ведется в электронном виде. Инструменты не полагаются больше на избирательную и ненадежную человеческую память. Игроки, например, обычно переоценивают свои способности потому, что на самом деле не следят за тем, сколько выигрывают и проигрывают. Ведь для них лучший показатель успеха — деньги на банковском счете.
6. *Инструменты измеряют быстрее и дешевле, чем человек.* Теоретически можно нанять достаточное число людей, чтобы они ежедневно и ежечасно пересчитывали все товары на полках большого продовольственного магазина. Но стандартные сканеры делают это дешевле. Полицейский может определить скорость машины с помощью часов с секундомером и указателей расстояния, но радар сделает это точнее, а главное, еще до того, как нарушитель скроется из виду. Даже если инструмент не дает больше никаких преимуществ, уже одно только снижение затрат оправдывает его использование.

С этой точки зрения, пастух, считающий овец с помощью веревки с узелками, тоже пользуется инструментом. Веревка калибрована, она регистрирует, и без нее пастух наверняка сделал бы больше ошибок. Методы выборки и экспериментальные приемы, даже не предполагающие использование каких-либо механических или электронных

приборов, — сами по себе инструменты, причем их так нередко и называют. Кто-то может возразить против подобного широкого толкования понятия «инструмент измерения», указав, например, на то, что опрос клиентов не всегда обнаруживает информацию, которую человек обнаружить не в состоянии. Но опрос, по крайней мере, должен быть последовательным и калиброванным. А если он проводится по Интернету, то обойдется дешевле и проанализировать его результаты будет легче (подробнее об этом см. главу 13). Те, кто не считает опрос покупателей инструментом измерения, забывают о самой сущности измерения, о том, насколько приблизительными были бы их знания, не располагай они этим инструментом.

Существует столько методов измерения, предложенных для решения самых разнообразных задач, что ни одна книга не вместила бы в себя их подробного описания. Но это изобилие только лишний раз убеждает нас, что какая бы задача измерения ни возникла, ее решение уже существует. И хотя данная книга не претендует на звание энциклопедии измерения, несколько приведенных в ней широких категорий методов позволяют решать огромное число проблем. Более того, эти методы можно использовать в различных сочетаниях, получая таким образом разнообразные подходы к решению конкретных задач.

Памятуя о своей решимости измерить все, что угодно, повторим здесь еще раз четыре полезные предпосылки измерения, о которых мы говорили в главе 3.

1. Не изобретайте велосипед — это уже сделали до вас.
2. У вас есть доступ к гораздо большему объему информации, чем вам кажется, — требуются только изобретательность и исходные наблюдения.
3. Вам нужно меньше данных, чем вы предполагаете, если придумать, как их проанализировать.
4. Нужный вам метод измерения наверняка проще, чем казалось вначале.

РАЗЛОЖЕНИЕ НА СОСТАВЛЯЮЩИЕ

Некоторые очень полезные методы, снижающие неопределенность, технически измерениями не являются, так как не предполагают проведения новых наблюдений. Однако нередко они очень помогают продвинуться в выборе подходящего способа измерения. Порой они позволяют оценщику понять, что на самом деле он знает больше, чем думал, когда давал первоначальную калиброванную оценку. Один из таких методов я называю «разложением на составляющие». Разложение на составляющие позволяет понять, как измерить что-то

совершенно неизвестное, предварительно оценив его составляющие с меньшими неопределенностями или, по крайней мере, легче поддающиеся измерению.

РАЗЛОЖИТЕ ОБЪЕКТ ИЗМЕРЕНИЯ НА СОСТАВЛЯЮЩИЕ

Многие измерения начинаются с разложения неизвестной величины на составляющие с целью выявления того, что можно наблюдать непосредственно и что легче поддается количественной оценке.

Если вы вспомните начало книги, то поймете, что Эратосфен в свое время осуществил именно разложение. Ведь он не стал измерять длину окружности Земли напрямую. Вместо этого он определил математическое соотношение между длиной окружности и проявляющимся в длине тени углом падения солнечных лучей на разных широтах. Длина тени и стала тем простым наблюдением, на основе которого Эратосфен сумел рассчитать то, что хотел. «Вопросы Ферми» — это тоже всего лишь прием, позволяющий разложить проблему на составляющие. Столкнувшись с трудной задачей по измерению, Ферми не опустил руки, а просто разделил ее на несколько и последовательно решил каждую из них. В то время как его исходную переменную (число настройщиков пианино Чикаго) определить было как будто невозможно, оказалось, что она является функцией других, более легких для оценки величин.

Фактически, большинство измерений в эмпирических науках осуществляются именно таким образом — опосредованно. Например, масса электрона и масса Земли были определены вовсе не по прямым наблюдениям. Были проведены другие исследования, которые и позволили рассчитать эти величины.

Разложение особенно полезно, например, при оценке затрат на реализацию крупного строительного проекта. Основываясь на прошлых затратах на аналогичные проекты, вы можете оценить их в 10–20 млн дол. Но когда разобьете свой конкретный проект на несколько составляющих и определите примерные диапазоны затрат по каждой из них, то получите, в конце концов, такой интервал совокупных затрат, который будет существенно уже первоначального. При этом вы не сделаете никаких новых наблюдений, а просто построите более развернутую модель на основе уже известных данных. Более того, вы можете обнаружить такую статью будущих затрат, о которой не знаете абсолютно ничего (например, затраты на оплату труда рабочих определенной специальности). И одно только это открытие уже сильно приблизит вас к желаемому результату.

Разложение может использоваться и при прогнозировании роста эффективности. Предположим, появился новый процесс или технология, способная повысить производительность труда определенной группы работников. Но, оценивая ее возможный рост, эксперты указали очень широкий интервал (5–40%), объяснив это незнанием многих факторов, способных повлиять на данный показатель. Например, они не знают, сколько точно человек будет занято на участке, где внедряется данная технология.

Определение численности людей, занятых каким-либо видом деятельности, представляется простым и очевидным этапом этого измерения. Но те, кто уверен в неизмеримости какого-то фактора, обычно возражают и против этого. В подобных случаях большую помощь в преодолении подобных преград оказывают фасилитаторы.

Фасилитатор: Ранее вы сказали, что в случае внедрения новой автоматизированной программы управления документооборотом эффективность труда ваших инженеров увеличится на 5–40%. Поскольку именно этот показатель несет особую информационную ценность для обоснования инвестиций в новое программное обеспечение, мы должны уточнить вашу оценку.

Инженер: Ну, это будет непросто. Разве можно измерить такую нематериальную вещь, как эффективность труда? Мы даже не следим за управлением документооборотом как за отдельным видом деятельности и поэтому не знаем, сколько времени тратится на него сейчас.

Фасилитатор: Однако вы, очевидно, предполагаете, что эффективность повысится — ведь на выполнение каких-то операций у инженеров будет уходить меньше времени, не так ли?

Инженер: Да, это так.

Фасилитатор: На какие виды деятельности инженеры тратят больше всего времени сейчас и на какие будут тратить его намного меньше после установки новой программы? Будьте как можно конкретнее.

Инженер: О'кей. Думаю, что они будут тратить меньше времени на поиск нужных документов. Но это только одна статья.

Фасилитатор: Хорошо. Начало положено. Сколько времени в неделю уходит на такие поиски сейчас и как, по вашему мнению, сократятся временные затраты? Пока что нам подойдут и калиброванные оценки.

Инженер: Ну, думаю, я на 90% уверен, что средний инженер тратит на поиски нужных бумаг от одного до шести часов в неделю. Все технические условия на оборудование, чертежи, руководства и прочее хранятся в разных местах, при этом большинство не в электронном виде.

Фасилитатор: Хорошо. Сколько времени удастся сэкономить, если они будут сидеть за своими столами и просто делать запросы?

Инженер: Даже я, когда пользуюсь такими поисковыми машинами, как Google, все равно трачу много времени, отбрасывая ненужную информацию, поэтому компьютеризация уменьшит временные затраты вовсе не на 100%. Но я уверен, что они сократятся, по крайней мере, вдвое.

Фасилитатор: Зависит ли это от вида деятельности инженеров?

Инженер: Конечно. Инженеры, выполняющие руководящие функции, тратят на это меньше времени. За них поиск документов выполняют подчиненные. А инженерам, занимающимся вопросами соблюдения законодательных и иных требований, постоянно приходится искать разные документы. Программой будут пользоваться и техники.

Фасилитатор: Отлично. Сколько инженеров и техников относятся к каждой из этих категорий и сколько времени каждый из них тратит на подобную деятельность?..

Так мы продолжали до тех пор, пока не выявили разные категории персонала, каждой из которых соответствовали свои временные затраты на поиск документов и свой потенциал снижения этих затрат. Сотрудники компании могут отличаться друг от друга и по темпам освоения новой технологии, и по другим признакам.

Приведенный выше диалог фактически является реконструкцией реального разговора, состоявшегося у меня с инженерами одной крупной американской АЭС. На нашей встрече мы определили и другие функции, такие как дистрибуция, контроль качества и т.д., затраты времени на выполнение которых могут сократиться с внедрением автоматизированной системы управления документооборотом. Время выполнения этих функций также зависело от служебных обязанностей инженера или техника.

Короче говоря, причиной такого широкого интервала значений возможного роста эффективности труда, который указали инженеры АЭС, частично оказалось то, что они знали обо всех этих различиях между инженерами, выполняющими разные обязанности, но не смогли разбить их по категориям. Когда это было сделано, обнаружилось, что некоторые величины (например, число инженеров в каждой группе или разное время поиска документов инженерами разных категорий) были известны и что неопределенность исходного значения объяснялась в основном незнанием всего одного или двух показателей. Если бы мы обнаружили, что в затратах времени на копирование или поиск потерянных документов они сомневаются больше, чем в том, в какие группы нужно объединить инженеров, то начали бы свои измерения с выяснения информации о затратах, а не с распределения работников по категориям.

ЭФФЕКТ РАЗЛОЖЕНИЯ НА СОСТАВЛЯЮЩИЕ

Эффект разложения на составляющие состоит в том, что сам процесс нередко обеспечивает такое значительное снижение неопределенности, что дальнейшие наблюдения становятся ненужными.

За последние 10 лет мне довелось 55 раз анализировать соотношение риска и доходности, при котором приходилось учитывать в общей сложности более 3300 переменных, то есть в среднем примерно по 60 переменных на один проект. Из них лишь около 120 (всего по две на один проект) потребовали дополнительного уточнения после определения стоимости информации. Большинство из этих 120 величин (около сотни) пришлось разложить на более легкие для оценки составляющие. Прочие переменные поддавались оценке более прямыми и очевидными методами. Например, чтобы определить расход топлива у грузовика при движении по дороге, покрытой гравием, было достаточно просто проехать на нем с включенным расходомером, а оценить число вирусов в программном обеспечении можно было, выборочно проверив коды.

Почти треть из тех переменных, что были разложены на составляющие (около 30), не потребовали после разложения дальнейшего уточнения. Иными словами, около 25% всех измерений с высокой стоимостью были произведены с помощью одного только разложения. Калиброванные эксперты уже знали о данной переменной вполне достаточно, они просто нуждались в более развернутой модели, точнее выражавшей имевшиеся у них знания.

Нам пришлось оценить одну или более составляющих примерно двух третей всех разложенных на элементы переменных. Например, разрабатывая прогноз роста эффективности труда, мы попросили группу работников оценить время, затрачиваемое ими на определенный вид деятельности. Для таких величин разложение на составляющие стало важным шагом вперед к пониманию способа, каким их можно оценить. Весь процесс разложения сам по себе — процесс постепенного прозрения скептиков, считающих объекты и явления не поддающимися измерению. Используя разложение, мы, подобно инженеру, перед которым поставили непростую задачу построить подвесной мост совершенно новым способом, методично решаем все проблемы измерения, разлагая их на элементы. Как и этому инженеру-мостостроителю, анализ составляющих на каждом этапе позволяет нам заново сформулировать и уточнить стоящую задачу. Разложение на составляющие переменной, «не поддающейся оценке», — важный этап, иногда он и сам по себе достаточно снижает неопределенность.

ВТОРИЧНЫЕ ИССЛЕДОВАНИЯ: ПРЕДПОЛОЖИМ, ЧТО ДО ВАС ЭТОТ ОБЪЕКТ УЖЕ ИЗМЕРЯЛИ

Похоже, что, приступая к измерениям в бизнесе, некоторые толковые предприниматели обычно исходят из того, что им, раз уж они так умны, придется самим придумывать метод количественной оценки интересующего объекта. Однако на самом деле в большинстве случаев подобные инновации вовсе не требуются.

Ни одно научное исследование не начинается с первой случайной выборки или экспериментального наблюдения. Не начинается оно и с разработки метода выборки или способа проведения эксперимента. Исследование начинается с вторичных исследований, то есть с анализа результатов, полученных до вас. Вторичные исследования отличаются от первичных, во время которых исследователи сами делают свои наблюдения. Несмотря на название, вторичные исследования всегда предшествуют первичным. Все исследователи считают само собой разумеющимся, что проблемой уже кто-то занимался. Об этом говорит и первое из четырех сформулированных нами «рабочих предпосылок» измерения (см. главу 3): «Ваша проблема совсем не так уникальна, как вы думаете. Предположите, что кто-то уже решил ее до вас или, по крайней мере, уточнил». Эта предпосылка почти всегда оказывается правильной.

Анализ имеющейся литературы, похоже, еще не вошел у менеджеров в привычку. Но проводить его стало намного легче. Сегодня почти все мои исследования начинаются с поиска информации в Интернете. Какую бы задачу по измерению я ни пытался решить, обязательным этапом является «домашнее задание» в Google или Yahoo. Потом я, конечно, все равно отправляюсь в библиотеку, но уже лучше представляю, что мне нужно.

Даю несколько советов по использованию Интернета для вторичных исследований. Быстро найти нужную вам информацию по методам измерения вы сможете только при условии, что укажете в строке «Поиск» нужные слова. Умение эффективно пользоваться Интернетом приходит с практикой, но, возможно, вам пригодятся следующие советы:

- *Если предмет мне совершенно незнаком, то я начинаю вовсе не с Google, а с онлайн-энциклопедии Wikipedia.org. Она содержит свыше миллиона статей, многие из которых посвящены проблемам бизнеса и техники, кажущимся слишком спорными для освещения в традиционных энциклопедиях. В хорошей статье обычно имеются ссылки на другие сайты, а по спорным*

вопросам приводятся несколько точек зрения, чтобы вы могли сами выбрать наиболее подходящую.

- *Ищите по словам, которые ассоциируются с исследованиями и количественным анализом.* Если вам необходимо оценить качество программного обеспечения или мнение потребителя, ищите не только по этим словам — найдете только общие рассуждения. Ищите по таким словам, как «таблица», «опрос», «контрольная группа», «корреляция» и «стандартное отклонение», которые обычно фигурируют в более содержательных исследованиях. Кроме того, в серьезных поисках обычно используют такие слова и словосочетания, как «университет», «доктор наук» и «общенациональные исследования».
- *Помните, что в Интернете есть не только поисковые машины, но и хранилища специальной информации.* Недостатком таких мощных поисковиков, как Google, является то, что вы, получив тысячи совпадений, можете все же не найти необходимой информации. Но попытайтесь заходить на сайты отраслевых или научных журналов. Если меня интересует национальная или мировая экономическая статистика, то я отправляюсь напрямую на такие сайты, как сайт Бюро переписи населения США, Министерства торговли и даже ЦРУ («World Fact Book» — «Всемирная книга фактов ЦРУ» — место, где я всегда нахожу самые разнообразные данные международной статистики). Совпадений окажется меньше, а нужных данных будет больше.
- *Используйте несколько поисковых машин.* Даже всесильный на первый взгляд Google пропускает статьи, которые быстро находят другие машины. Лично я помимо Google пользуюсь clusty.com и yahoo.com.
- *Найдя исследования не совсем по своей теме, в которых, тем не менее, упоминается интересующий вас вопрос, обязательно посмотрите библиографию.* Иногда ее изучение — лучший способ отыскать нужные данные.

ОСНОВНЫЕ МЕТОДЫ НАБЛЮДЕНИЯ:

НЕ РАБОТАЕТ ОДИН — ПОПРОБУЙТЕ ДРУГОЙ

Подробное описание того, как вы представляете себе или распознаете предполагаемый объект измерения, — полезный прием при выборе метода его количественной оценки. Наличие любых оснований считать, что объект вообще *существует*, означает: вы его так или иначе наблюдаете. Когда кто-то утверждает, что удовлетворенность потребителей значительно повысится в случае сокращения времени ожидания

ответа у телефона, он делает это, основываясь на каких-то аргументах: возможно, поступали жалобы от клиентов; возможно, с ростом компании наметилась тенденция к сокращению числа довольных покупателей. Измерения почти всегда проводятся для проверки достоверности какой-то идеи, а эти идеи возникают не на пустом месте.

Выявив неопределенность, установив соответствующие пороги и рассчитав стоимость информации, вы уже обнаружили нечто, в принципе поддающееся наблюдению. Обдумайте несколько этапов, связанных с характером наблюдений, — своеобразный каскад эмпирических методов. Если вам не подойдет первый подход, попробуйте следующий и т.д. Вопросы приведены в произвольном порядке, но, возможно, в некоторых ситуациях вам будет удобнее начать с первого, а потом перейти к другим.

1. *Оставляет ли интересующий вас объект какие-либо следы после себя?* Практически все явления, какие только можно вообразить, оставляют после себя некое доказательство своего осуществления. Рассуждайте, как следователь. Приводит ли объект, событие или вид деятельности, который вы хотите измерить (оценить), к последствиям, оставляющим после себя какие-нибудь следы? Пример: длительное ожидание ответа у телефона заставляет некоторых клиентов, позвонивших в службу поддержки, дать отбой. Это должно приводить, по меньшей мере, к убыткам в бизнесе, но каким именно? Почему люди не дожидаются ответа — по каким-то своим причинам или из-за раздражения от проволоочки? В первом случае они перезвонят, а во втором — нет. Выявив хотя бы несколько человек из тех, кто повесил трубку, и связав это с сокращением объемов их покупок, вы получите подсказку. Сумеете ли вы установить корреляцию между числом клиентов, не дождавшихся ответа, и снижением продаж своей продукции? (См. пример «Оставшийся след».)
2. *Если явление не оставило следов, попробуйте понаблюдать за ним или какой-то его составляющей непосредственно.* Вы наверняка не подсчитывали, сколько машин с номерами других штатов скапливается на парковке у магазина, — теперь можете сделать это. Конечно, дежурить на парковке целый день нецелесообразно, но вполне приемлемо подсчитать такие автомобили в какие-то произвольно выбранные часы работы магазина.
3. *Если объект измерения, по всей видимости, не оставил после себя никаких определяемых следов, а непосредственное его наблюдение без дополнительных инструментов кажется невозможным, можете ли вы придумать способ обнаружить последствия сейчас?* Если до сих пор объект никаких следов не оставил, его можно «пометить», чтобы они проявились, по крайней мере, теперь.

Например, Amazon.com предоставляет бесплатную подарочную упаковку, чтобы следить за тем, сколько книг приобретается в подарок. Раньше создателей сайта эта информация не интересовала. Введение новой услуги «мешки» позволило им получить такие сведения. Другой пример — раздача купонов покупателям, давшая владельцам магазинов возможность следить в том числе и за тем, какие газеты читают их клиенты.

4. *Если наблюдение в существующих условиях не дает нужных результатов (когда имевшейся или вновь собранной информации недостаточно), то нельзя ли «заставить» интересующий вас феномен проявиться в других условиях, в которых наблюдать за ним будет проще?* Пример: владелец сети магазинов хочет узнать, как новые правила возврата бракованных товаров повлияют на удовлетворенность покупателей и объемы продажи. Он должен опробовать их сначала в одном магазине, оставив в других все по-прежнему, и сравнить результаты.

НЕКОТОРЫЕ ОСНОВНЫЕ ПРАВИЛА НАБЛЮДЕНИЯ

1. Идите по следу, как заправский детектив. Проводите «криминалистический анализ» уже имеющейся у вас информации.
2. Используйте непосредственное наблюдение. Начните смотреть, подсчитывать и по возможности делайте выборочные оценки.
3. Если до сих пор объект не оставлял после себя следов, «пометьте» его, чтобы следы наконец появились.
4. Если выйти на след никак не удастся, создайте условия, в которых появится возможность наблюдать за объектом (проведите эксперимент).

Перечисленные методы применимы и тогда, когда требуется оценить происходящее в настоящее время (например, текущие продажи, обусловленные положительными отзывами покупателей), и тогда, когда нужно сделать прогноз (ожидаемого совершенствования рекламы, базирующейся на отзывах потребителей о новой характеристике товара, повышения качества обслуживания и т.д.). Когда возникает необходимость оценить текущее состояние, вся необходимая для этого информация уже имеется. Но когда составляется прогноз, приходится решать, что из уже наблюдаемого дает основания ожидать улучшения. Если же в прошлом нельзя найти ничего, что позволяет надеяться на прогресс, то на чем вообще базировались подобные ожидания?

И помните: чтобы обнаружить след, «пометить» объект или провести эксперимент, нужно понаблюдать всего за несколькими явле-

ниями из случайной выборки. Не забывайте и о том, что выявленные составляющие объекта порой измеряются разными способами. И не стоит сразу беспокоиться из-за проблем, которые часто возникают в ходе применения приведенных выше методов. Просто решите, какой из них проще и практичнее применить именно сейчас.

ПРИМЕР «ОСТАВШИЙСЯ СЛЕД»: ЦЕННОСТЬ СКОРОСТИ ОТВЕТА НА ЗВОНКИ КЛИЕНТОВ

Один крупный европейский дистрибьютор лакокрасочной продукции попросил меня оценить влияние скорости работы сети на объемы продаж, так как от этого показателя зависела быстрота ответов на входящие звонки клиентов. Поскольку офисная АТС вела учет всех звонков, в том числе и прерванных абонентами в период ожидания, а торговый персонал хранил данные о степени занятости (а следовательно, о времени отклика) за прошлые периоды, я порекомендовал провести перекрестный анализ обоих массивов данных. В результате оказалось, что когда нагрузка на сбытовую сеть возросла, клиенты стали обрывать звонки чаще. Изучались также ситуации прошлых периодов, когда сеть работала медленнее по другим причинам, а не из-за того, что ею активно пользовалась служба поддержки, а также динамика продаж. В результате компания сумела выявить ту разницу в продажах, которая объяснялась исключительно более медленной работой сети.

НЕ ИЗМЕРЯЙТЕ ТОЧНЕЕ, ЧЕМ НУЖНО

В главе 7 мы рассказали о том, как рассчитать стоимость информации, необходимой для принятия решения. Определенная вами исходная неопределенность, стоимость информации и пороги дают достаточно данных о методе измерения, который действительно подходит для поставленной задачи. Если информация о том, заметили ли покупатели улучшение качества вашей продукции после перехода на новый метод производства (например, после изменения рецептуры изготавливаемого вами напитка), стоит пару тысяч долларов, то вам не удастся экономически обосновать ни небольшое маркетинговое исследование, ни даже дегустацию вслепую. Но когда эта информация стоит миллионы долларов (что более вероятно, если ваша компания крупная или хотя бы средняя), вас не запугает исследование, которое обойдется в 100 тыс. дол. и продлится несколько недель. Знание стоимости информации, порога, текущей неопределенности и решения, которое

придется принять, — необходимое условие правильного выбора цели и условий измерения.

Стоимость информации определяет верхний предел затрат на проведение измерений. Но, как правило, такие затраты оказываются намного ниже. Обычно я рассчитываю, что они приблизительно составят 10% EVPI, хотя порой они не превышают и 2% EVPI (это минимум, из которого вы должны исходить). Я ориентируюсь на эти показатели по трем причинам. Во-первых, EVPI — это стоимость полной информации. Но все эмпирические методы дают какую-то погрешность, а наша цель — снизить неопределенность, а вовсе не получить точные сведения. Поэтому стоимость результатов наших измерений будет, по всей видимости, намного меньше EVPI. Во-вторых, начальные измерения нередко меняют стоимость результатов дальнейших измерений. Если первые итоги оказываются неожиданными, то затраты на продолжение оценки могут упасть до нуля. Это означает одно: измерения следует проводить итеративно. А поскольку у нас всегда есть возможность продолжить измерения, если нужен более точный результат, то обычно в недооценке результатов начальных оценок присутствует управляемый риск. Наконец, не забывайте о том, что кривая стоимости информации обычно круче всего идет вверх в самом начале: первая сотня изученных образцов снижает неопределенность намного больше, чем вторая.

Способ измерения объекта подсказывает порог. Если возникает необходимость оценить возможный спрос, чтобы определить объем производства, или производительность труда (для расчета премиальных выплат), то фактически порога у вас не будет. Каждое повышение точности будет иметь свою стоимость, хотя с приближением к EVPI стоимость информации будет инкрементально расти все медленнее. Но предположим, что требуется рассчитать размер рынка, так как сделанные вашей компанией инвестиции окажутся безубыточными только в случае роста рынка не менее чем на 12% по сравнению с прошлым годом. Выбирая метод измерения, вам следует учесть, что ошибка не выше 1% не так важна, главное — узнать, на какой стороне порога вы окажетесь. Если вы можете с достоверностью установить, что рынок вырастет менее чем на 5%, то какова погрешность этого показателя — 1% или 5% — вам неважно. В этом случае главное, чтобы прогноз не показал, что рост рынка составит больше (или меньше) 12%, когда на самом деле правильно обратное.

Наконец, о том, какой метод измерения вам потребуется, говорит и исходная неопределенность. Помните: чем выше неопределенность, тем больше информации дадут первые наблюдения. Если сначала неопределенность была очень высока, то даже методы, имеющие большую неустранимую погрешность, дадут больше сведений, чем имелось раньше.

УЧИТЫВАТЬ ПОГРЕШНОСТЬ

Все измерения дают погрешность. Чтобы решить эту проблему, ее, как и любую другую, необходимо, прежде всего, признать. После этого мы сможем выработать стратегию, позволяющую, по крайней мере, частично компенсировать ошибки. Однако те, кто быстро пасует перед трудностями измерения, нередко исходят из того, что наличие *любой* погрешности означает невозможность оценки. Будь это так, измерения были бы невозможны ни в одной области науки. Но, к счастью для научного сообщества и для всех остальных, это заблуждение. Энрико Ферми может покоиться с миром.

Ученые, статистики, экономисты и большинство других специалистов, занимающихся эмпирическими измерениями, делят ошибки на две большие категории — систематические и случайные. Систематические ошибки — это постоянные, а не случайные отклонения результатов одного наблюдения от итогов другого. Например, если прогнозы доходов следующего квартала, составляемые менеджерами по продажам, постоянно оказываются завышенными в среднем на 50%, то это ошибка систематическая. То, что завышение не всегда составляет ровно 50%, — ошибка случайная. Случайная ошибка одного наблюдения, по определению, не поддается предсказанию, но случайные ошибки ряда наблюдений подчиняются определенной закономерности и могут быть рассчитаны с помощью теории вероятности.

Термины «систематическая» и «случайная ошибки» связаны с такими понятиями, как достоверность и точность. Достоверность отражает воспроизводимость и повторяемость результатов измерений, в то время как точность показывает, насколько результат измерения близок к «истинному» значению. Большинство считает понятия «точность» и «достоверность» синонимичными, но специалисты по измерению их определенно различают.

Показания домашних весов, специально настроенных, чтобы занижать или завышать вес, могут быть достоверными, но неточными. Они достоверны, так как будут показывать один и тот же вес каждый раз, когда один и тот же человек будет вставать на них несколько раз в течение часа (за который его вес не успеет измениться). И все же они неточны, потому что всякий раз будут показывать вес, специально заниженный, например на 3 кг. Представим себе теперь, что идеально настроенные весы находятся в движущемся автофургоне. Ухабы, ускорение или спуски и подъемы заставят их давать каждый раз иные показания, даже если взвешиваться будет один и тот же человек (становясь на них по два раза в минуту). И все же окажется, что средний ответ очень близок к его реальному весу. Вот пример довольно хорошей точности, но низкой достоверности. Это характерно и для калиброванных экспертов: они могут быть непоследовательными в своих суждениях, постоянно что-то пере- или недооценивая.

КРАТКИЙ ГЛОССАРИЙ ПОГРЕШНОСТИ

Систематическая ошибка, или систематическое отклонение (смещение) — неотъемлемое свойство процесса измерения давать определенный результат; постоянное отклонение.

Случайная ошибка — ошибка, непредсказуемая для отдельного наблюдения, непостоянная и не зависящая от известных величин (хотя в своей массе такие ошибки подчиняются законам вероятности).

Точность — характеристика измерений, дающих низкую систематическую ошибку, то есть таких, когда искомое значение не занижается и не завышается на постоянной основе.

Достоверность — характеристика измерений, дающих низкую случайную ошибку, то есть таких, которые дают аналогичные результаты, пусть и далекие от истинного значения.

Иными словами, достоверность — это низкая случайная ошибка при любой систематической ошибке, а точность — это низкая систематическая ошибка при любой случайной ошибке. Каждый вид ошибки можно учесть и компенсировать. Зная, что весы всегда показывают на 3 кг больше, мы можем скорректировать их показания. Если весы настроены точно, но дают разноречивые показания, то мы можем устранить случайную ошибку, проведя несколько измерений и рассчитать средний результат. Любое снижение того или иного вида ошибки называется контролем точности.

Случайная выборка представляет собой тип контроля точности в случае правильного ее использования. Непредсказуемые по отдельности, взятые в целом, случайные эффекты подчиняются определенным, вполне прогнозируемым закономерностям. Например, я не знаю, как упадет одна подброшенная монета, но могу сказать, что при подбрасывании 1000 монет решка выпадет 500 ± 26 раз (способ определения этой ошибки мы еще обсудим позже). Намного сложнее оценить, даже приблизительно, систематическую ошибку. Систематические ошибки (допускаемые необъективно настроенными экспертами, определяющими качество продукции, или приборами, постоянно дающими завышенные показания) не обязательно порождают случайные ошибки, не поддающиеся количественной оценке с помощью теории вероятности.

Будь у вас выбор, что бы вы предпочли: взвеситься на разрегулированных, но достоверных весах с неизвестной погрешностью или на настроенных, но находящихся на движущейся платформе и дающих всякий раз разные показания? Я обнаружил, что в бизнесе люди нередко предпочитают достоверность с неизвестной систематической ошибкой

недостоверному значению со случайной ошибкой. Например, чтобы определить, сколько времени торговые представители тратят на встречи с клиентами по сравнению с выполнением других административных задач, менеджеры, скорее всего, проанализируют все ведомости учета рабочего времени. Идея провести случайную выборку торговых представителей и изучить структуру их временных затрат вряд ли придет им в голову. Но ведомости учета рабочего времени не дают точной картины, особенно если заполняются в пятницу в 5 часов вечера перед самым уходом с работы за всю неделю сразу. Люди недооценивают время, затраченное на выполнение одних задач, переоценивают продолжительность выполнения других и непоследовательны в классификации этих задач.

Поэтому даже если анализ всех пяти тысяч ведомостей учета рабочего времени (по 50 недельных ведомостей на каждого из 100 торговых представителей) и скажет нам, что они тратят на непосредственное общение с клиентами 34% своего времени, мы не будем знать, правда ли это. И все же эта «точная» цифра, похоже, кажется многим менеджерам вполне убедительной. Предположим теперь, что прямое наблюдение за случайно выбранными торговыми представителями в случайно выбранные моменты времени показало, что они находились на встречах с клиентами или разговаривали с ними по телефону только в 13 из 100 случаев (в этом можно убедиться, и не отвлекая торговых представителей от их занятия, а опросив их, когда они освободятся). Как показано в главе 9, для последнего измерения мы можем статистическими методами рассчитать, что 90-процентный доверительный интервал этого показателя составляет 7,5–18,5%. Хотя метод случайной выборки и даст нам только интервал, его результат будет представлять больший интерес, чем результат анализа ведомостей учета рабочего времени. Последний способ предоставляет точное число, но мы не имеем возможности узнать, как велика погрешность и в какую сторону полученное значение отклоняется от истинного.

**МАЛЫЕ СЛУЧАЙНЫЕ ВЫБОРКИ
ИЛИ КРУПНЫЕ НЕСЛУЧАЙНЫЕ:
ИССЛЕДОВАНИЯ ЧЕЛОВЕЧЕСКОЙ СЕКСУАЛЬНОСТИ,
ПРОВЕДЕННЫЕ КИНСИ**

В 1940-х и 1950-х годах исследования Альфреда Кинси о сексуальном поведении человека дали толчок оживленным дискуссиям о сравнительных достоинствах малых случайных выборок и крупных неслучайных. Книги Кинси вызвали и большой интерес, и много споров. Получив грант от фонда Рокфеллера, Кинси сумел опросить 18 тыс. мужчин и женщин. Но его выборки были не совсем случайными. Он стремился встретиться с рекомендованными ему людьми и побеседовать с каждым представителем ка-

кой-либо группы (команды по боулингу, студенческих сообществ, книжного клуба и т. д.). По-видимому, он исходил из того, что при достаточно большой выборке погрешность компенсируется. Но в случае большинства систематических ошибок это не срабатывает — они не исключаются методом усреднения. Знаменитый статистик Джон Тьюки, которому тот же фонд Рокфеллера не позволил проверить работу Кинси, якобы сказал, что случайный выбор трех человек дал бы лучшие результаты, чем выбранная мистером Кинси группа из 300 человек. По другой версии, он сказал, что предпочитает случайную выборку из 400 респондентов отобранным Кинси 18 тыс. человек. Возможно, первое высказывание Тьюки и преувеличение, но небольшое. Он имел в виду, что выбиравшиеся Кинси группы были нередко очень близки к однородным. Поэтому, с точки зрения статистики, их можно приравнять к одному случайно выбранному человеку. Во второй версии своего высказывания Тьюки был абсолютно прав: ошибка при случайном выборе 400 человек вполне поддается количественной оценке и может быть намного меньше систематической ошибки при неправильном выборе 18 тыс. человек.

Почему люди предпочитают ложное впечатление точности ошибке случайной выборки, поддающейся количественной оценке? Как я обнаружил, нередко это происходит потому, что они путают ошибку одной выборки с ошибкой всего исследования. Да, в нашем примере с торговыми представителями в какие-то моменты вы могли обнаружить, что кто-то занят нетипичной, нерепрезентативной деятельностью, вовсе не характерной для всей группы, например готовится к командировке, хотя обычно почти никуда не ездит. Если бы мы выбрали именно этого человека и опросили его всего один раз, то вряд ли узнали бы что-то полезное о том, на что *он* тратит свое рабочее время. Но если 25 из 100 торговых представителей, опрошенных неоднократно, действительно готовятся к командировке, то можно не сомневаться, что весь торговый персонал тратит на это, в среднем, 25% своего времени, и простейшие расчеты из главы 9 показывают, что 90-процентный доверительный интервал для этого показателя составляет 18–32%. Из ненадежности результата одной выборки люди делают вывод, что в случае нескольких случайных выборок их ошибки не устраняются путем компенсации, а суммируются.

Ошибку, не исключаемую путем усреднения (систематическую ошибку), называют также отклонением, или смещением. Исследования в области психологии принятия решений и эмпирических наук в целом расширяют перечень возможных типов отклонения чуть ли не каждый год. Но есть три основных типа, которые можно ожидать при проведении измерений: отклонение ожидания, отклонение выбора и отклонение наблюдателя.

ВИДЫ ОТКЛОНЕНИЙ ПРИ НАБЛЮДЕНИЯХ

Смещение ожидания — принятие желаемого за действительное. Наблюдатели и испытуемые порой намеренно или ненамеренно видят именно то, что хотят. Люди доверчивы и склонны к самообману. Когда проводятся клинические испытания лекарств, пациенты не знают, кто принимает лекарство, а кто — плацебо. Это упомянутое выше испытание вслепую. Если принимающие лекарство не известны ни пациентам, ни врачам, то это двойное испытание вслепую. Еще один пример испытания вслепую — подход, предложенный мною Mitre Corporation (см. главу 3).

Смещение выбора. Выборка, планировавшаяся как случайная, может оказаться неслучайной. Если мы отберем 500 участников голосования, 55% из которых скажут, что проголосуют за кандидата А, то возникает большая, а точнее, 98,8-процентная вероятность того, что этот кандидат действительно пользуется наибольшей поддержкой населения. Есть только 1,2-процентная вероятность того, что мы случайно отобрали слишком много людей, поддерживающих А, который на самом деле далеко не любимец народа. Но подразумевается, что выборка была случайной и мы не стремились к отбору сторонников одного кандидата. А если отбор осуществляется, например, путем опроса прохожих в деловом районе города, на определенной улице, где расположено много финансовых учреждений, то, скорее всего, все опрошенные избиратели окажутся определенного типа, даже если вы выбирали их наугад.

Ошибка наблюдателя (или искажение Гейзенберга и Хоторна). Субатомные частицы и люди имеют нечто общее: наблюдение заставляет их менять свое поведение. В 1927 г. физик Вернер Гейзенберг вывел формулу, которая показала, что нашим знаниям о положении и скорости движения частицы есть предел. Наблюдая за частицами, мы на них воздействуем (например, направляем пучок света), что заставляет их менять свою траекторию. В том же году началась реализация научно-исследовательского проекта на заводе Western Electric Company в Хоторне (штат Иллинойс). Первоначально им руководил профессор Элтон Мэйо из Гарвардской школы бизнеса, преследуя цель определить влияние условий на производительность труда рабочих. Исследователи меняли уровень освещенности, влажность, график работы и т. д., чтобы определить, при каких условиях рабочие будут работать лучше всего. К своему удивлению, они обнаружили, что производительность труда растет независимо от изменения условий труда. Рабочие работали лучше просто потому, что знали, что за ними наблюдают; или, возможно, предположили исследователи, положительную реакцию вызвало то, что руководство обратило на них внимание. В любом случае, мы не можем больше считать, что наши исследования показывают нам «реальный» мир, если не постараемся компенсировать эффект, оказываемый наблюдениями на изучаемый объект. Простейший выход из положения — проводить наблюдения в тайне от испытуемых.

ВЫБЕРИТЕ ИЛИ РАЗРАБОТАЙТЕ НУЖНЫЙ ИНСТРУМЕНТ

После разложения объекта измерения на составляющие, установления порядка наблюдения за этими составляющими, выбора уровня снижения неопределенности (лишь настолько, насколько это необходимо) и учета основных типов ошибок можно считать, что нужный инструмент оценки практически у вас в руках. Одни только ответы на приведенные ранее вопросы должны были сделать для вас метод измерения более очевидным.

Итак, чтобы выбрать нужный инструмент измерения, необходимы следующие этапы.

1. *Разложите интересующий вас объект на составляющие так, чтобы их можно было оценить по другим измерениям.* Некоторые составляющие оценивать легко, а иногда само разложение уже достаточно снижает неопределенность.
2. *Проведите вторичные исследования своих результатов.* Узнайте, как подобные объекты измерялись до вас. Даже если итоги других исследователей далеки от ваших проблем, нельзя ли воспользоваться их методами?
3. *Примените к одному или нескольким составляющим объекта измерения один или более методов наблюдения (выявление оставленного следа, прямое наблюдение, наблюдение с помощью «меток» или проведение эксперимента).* Придумайте, по крайней мере, три способа их обнаружения и затем действуйте, как настоящий следователь. Если это не удастся, попробуйте прямое наблюдение. Если и это оказалось безрезультатным, пометьте свой объект или внесите в него другие изменения, чтобы хотя бы потом он начал оставлять поддающиеся обнаружению следы. Если и этот случай оказался безуспешным, воспроизведите событие специально для того, чтобы за ним можно было понаблюдать (проведите эксперимент).
4. *Постоянно помните о правиле «Не измеряйте точнее, чем нужно».* Ну зачем вам высокая точность, если вы хотите, например, только удостовериться, что рост производства оправдает сделанные инвестиции? Не забывайте о стоимости информации: низкая стоимость означает, что и затраты на проведение измерений должны быть низкими; а высокая стоимость информации означает, что вы можете не экономить на расходах. Помните и о том, с какого уровня неопределенности вы начинали. Если сначала этот уровень был очень высок, то сколько наблюдений вам потребуется, чтобы снизить его?

5. *Учитывайте ошибки, специфичные для данной задачи.* Если несколько экспертов оценивают качество работы, помните о смещении ожидания и проведите испытания вслепую. Если нужно сделать выборку, проследите за тем, чтобы она действительно была случайной. Если сам факт проведения наблюдений может повлиять на результат, найдите способ скрыть эту информацию от наблюдаемых.

Но если вы и теперь не сумели выбрать себе инструмент измерения, воспользуйтесь следующими советами, приведенными в произвольном порядке. Некоторые из них вам уже известны, но все они вполне заслуживают повторения.

- *Работайте с последствиями.* Что вы увидите, если интересующий вас показатель сильно повысится? Что произойдет, если он заметно упадет? В примере из главы 2 юная Эмили рассудила: раз целители с помощью бесконтактного массажа действительно могут делать то, что утверждают, значит, они должны, по меньшей мере, обнаруживать ауру человека. Оценивая качество, мы можем исходить из того, что с его повышением сокращается число жалоб от потребителей. Решая, стоит ли устанавливать новое программное обеспечение для отдела сбыта, мы должны спросить себя: «Если эта информационная система так хороша, как говорят, и действительно позволяет продавать больше, то почему же объемы реализации в тех фирмах, которые ею пользуются, пошли вниз?»
- *Представьте себе, как это сделали бы другие.* Посмотрите на свою задачу глазами криминалиста, детектива, психолога-экспериментатора, палеонтолога, библиотекаря, военного разведчика, журналиста. Преодолейте свою профессиональную «зашоренность» и обратите внимание на другие, нестандартные для вашей отрасли методы измерения.
- *Пусть ваши измерения будут итеративными.* Не пытайтесь сразу и полностью устранить неопределенность в ходе одного гигантского исследования. Начните с нескольких наблюдений и пересчитайте стоимость информации. Возможно, вы измените прежний план проведения измерений.
- *Используйте многосторонние подходы.* Если какая-то составляющая объекта измерения не поддается наблюдению одним способом, воспользуйтесь другим. Возможностей у вас много. Хорошо, если сработал первый же метод измерения. Но в некоторых случаях мне приходилось измерять свой объект тремя разными способами, и лишь на третий раз я получал желаемое.

Уверены ли вы, что проверили пригодность всех имеющихся методов? Если не удастся оценить одну из составляющих переменной, нельзя ли измерить другую?

- *Каков на самом деле тот простой вопрос, ответ на который может сделать дальнейшие измерения неактуальными?* Эмили не пыталась оценить, какой эффект дает метод бесконтактного массажа и есть ли этот эффект вообще. В обсуждавшемся ранее примере с компанией Mitre я предложил ее специалистам сначала выяснить, смогут ли ее клиенты заметить хоть какое-то изменение качества исследований, и только потом рассчитывать стоимость ожидаемого повышения качества. Существуют такие базовые вопросы, ответы на которые могут избавить вас от необходимости проведения более сложных измерений. Какой вопрос вы должны себе задать, чтобы понять, необходимы ли дальнейшие усилия?
- *Просто сделайте это.* Не позволяйте своей озабоченности по поводу возможных неудач помешать вам *приступить* к каким-либо упорядоченным наблюдениям. Ведь уже первые результаты могут вас удивить и значительно снизить неопределенность, характерную для измеряемого объекта.

К настоящему моменту вы уже должны понять, что следует наблюдать и (в общих чертах) как это сделать. Теперь можно приступить к обсуждению двух категорий методов наблюдения: способов, результаты которых анализируются методами «традиционной» статистики, и метода байесовского анализа. Эти две категории охватывают практически все эмпирические методы измерения, применяемые в физике, медицине, экологии и экономике. Хотя традиционный подход распространен гораздо шире, более новый байесовский анализ обладает рядом очевидных преимуществ.



Выборочная реальность: как наблюдение за частью рассказывает нам о целом

Если вы хотите знать наверняка, каков процент брака в партии кирпичей, вам придется проверить каждый. А поскольку при испытаниях кирпичи подвергнутся нагрузке вплоть до разрушения, после нее не останется ни одного целого. Стремясь сохранить партию кирпичей для продажи или иного использования, мы должны проверить лишь несколько изделий — это даст информацию о всех.

Группа, сведения о которой необходимо получить (в данном случае это партия кирпичей), называется генеральной совокупностью. Изучение всех элементов группы (то есть проверка каждого произведенного кирпича) называется полным обследованием. Очевидно, что проведение полного обследования партии нецелесообразно, так как по его завершении не останется ни одного неповрежденного кирпича. Вместе с тем в других условиях полное обследование весьма обоснованно. Ежемесячная инвентаризация — это обычно именно тот случай, балансовый отчет — тоже результат полного обследования всех активов и обязательств. Бюро переписей населения США (U.S. Census Bureau) старается пересчитать всех проживающих в стране людей, хотя на практике это не вполне удастся.

Но многие вещи скорее напоминают кирпичи, чем бухгалтерские операции. Существуют причины, по которым проверять, измерять, взвешивать или даже подсчитывать все элементы генеральной со-

вокупности бывает нецелесообразно. Но мы все же можем снизить неопределенность, изучив несколько элементов этой совокупности. Любое обследование, не являющееся полным, называется выборочным. В сущности, выборка — это несколько объектов из генеральной совокупности, которые подвергаются наблюдению с целью получить информацию обо всех ее элементах.

Кого-то может удивить, что, исследуя одни объекты, мы получаем информацию о других, даже не рассматриваемых в данный момент, но ведь этим, в сущности, и занимается наука. Проводя эксперименты, мы изучаем отдельные эффекты Вселенной, изобилующей самыми разнообразными явлениями. Однако, открывая «закон», наука говорит, что он относится ко всем элементам исследуемой совокупности, а не только к тем, за которыми велось наблюдение.

Так, скорость света была определена путем выборочного обследования. И какой бы метод при этом ни применялся, он давал погрешность. Поэтому ученые измеряли скорость света неоднократно, и каждое измерение — новая выборка. Тем не менее скорость света — абсолютная константа, характеризующая и свет, отражающийся от этой страницы и попадающий вам в глаза, и свет, изучавшийся в лабораторных условиях. Полное обследование тоже может считаться выборочным, если со временем генеральная совокупность расширилась. Ведь данные полной инвентаризации, как и балансовый отчет, — всего лишь моментальная фотография.

Следующая фраза может огорчить тех, кто хочет, чтобы в мире было больше определенности: все, что мы знаем «по опыту», — не более чем выборка. Ведь на своем опыте мы испытываем не все, а только кое-что, а затем делаем обобщение. Все, что нам доступно, — лишь мимолетные образы по большей части невидимого нами мира, по которым мы и судим о том, что не можем наблюдать. Но люди уверены в выводах, которые делают на основе изучения ограниченных выборок, потому что опыт подсказывает: выборки работают! (Конечно, этот вывод сделан на основе таких же ограниченных наблюдений.)

Иногда полезные выводы обо всей обследуемой совокупности можно сделать лишь по нескольким образцам. Если мы отбираем их, чтобы установить состав однородного объекта, например делаем анализ крови на ДНК или определяем октановое число бензина, то достаточно одного образца крови или одной партии бензина. Но если элементы генеральной совокупности разнородны, как, например, рыбы в озере по размеру или затраты времени разных специалистов на ремонт компьютеров, то выборка должна быть больше — иногда существенно, хотя все же не настолько, как думают многие.

Как изучение нескольких элементов может рассказать о всей генеральной совокупности? Можно ли, отобрав 12 жителей города и спросив, как часто они ходят в кино или доверяют ли мэру, узнать что-ли-

бо о тех, кто *не* был опрошен? Да, можно. И такая малая выборка порой дает на удивление много информации, хотя отчасти это зависит и от того, как проведено исследование. Если мы просто опросим своих друзей или всех мужчин — клиентов парикмахерской, то данная группа, вероятно, не будет репрезентативной для всего населения и наши общие выводы окажутся ошибочными. Необходим метод, гарантирующий нас от систематического выделения однотипных выборок.

Решением этой проблемы является проведение действительно случайной выборки из интересующей нас генеральной совокупности. Осуществляя выбор случайным образом, мы обязательно получим погрешность, но погрешность, подчиняющуюся законам вероятности. Тогда можно рассчитать, например, вероятность того, что случайно выбрали демократов, проводя опрос в районе, где на самом деле преобладают республиканцы. И чем больше людей мы отберем таким случайным образом, тем меньше вероятность нерепрезентативности группы.

Если вам доводилось читать отчеты об опросах общественного мнения или научные статьи, авторы которых проводили выборку, значит, вы уже встречались с понятием «статистическая значимость». Статистическая значимость просто говорит: вы наблюдаете нечто реальное, а не произошедшее по чистой случайности. Насколько крупной должна быть выборка, чтобы получить статистически значимый результат? Достаточно ли нам опросить 1000 потребителей? Будет ли результат статистически значимым, если выборочно проверить сварку на шасси 50 автомашин или испытать действие нового лекарства на 100 пациентах?

Мне не раз доводилось слышать на первый взгляд авторитетные суждения по этому поводу. Кто-то утверждает, что получить статистически значимые результаты позволяет только выборка определенного размера. Однако спроси такого человека, как он определил это число, — он в лучшем случае сошлется на какое-нибудь правило из учебника по статистике, но объяснить, как оно рассчитано, скорее всего, не сможет.

Короче говоря, термином «статистическая значимость» часто злоупотребляют те, кто не вполне понимает, что он означает. Неужели эти люди действительно думают, что снижение неопределенности происходит только при выборке, достигшей этого порогового значения? Или же они считают, что экономическая стоимость информации, полученной при изучении малой выборки, всегда меньше затрат на проведение измерений? По своему опыту могу сказать: когда в любой компании проводится случайная выборка, всегда найдется «эксперт», точно знающий, что можно, а чего нельзя делать в статистике. Но я обнаружил, что процент ошибок в туманных воспоминаниях таких экспертов о статистике, которую они учили когда-то давно, на

первом курсе университета, бывает намного выше ошибки при малой выборке.

Кто действительно разбирается в статистической значимости, так это Барри Нассбаум, главный статистик Службы статистической поддержки Агентства по защите окружающей среды (Statistical Support Services at the Environmental Protection Agency). Мы вместе занимались внедрением некоторых моих методов в EPA. Он отвечает на самые неожиданные вопросы, поступающие от сотрудников Агентства, связанные с проведением статистических анализов при решении проблем разных типов. Как-то он сказал мне: «Обращаясь за помощью к статистикам, люди всегда спрашивают, каким должен быть размер выборки. Это неправильный вопрос, но именно его и задают в первую очередь». Естественно, прежде всего надо выяснить, что они измеряют и зачем, и в этом с Нассбаумом нельзя не согласиться.

На самом деле небольшая выборка может рассказать намного больше, чем вы думаете. Когда исходная неопределенность высока, даже малая выборка позволяет значительно снизить ее. А если вы уже знаете тот узкий интервал значений, в котором находится интересующий вас показатель (например, новой услугой довольны 80–85% потребителей), то уточнить эту цифру вам поможет только крупная выборка (придется опросить более 1000 человек). Но в этой книге мы говорим о том, что считается не поддающимся измерению. В таких случаях неопределенность обычно особенно высока и полезную информацию могут дать уже первые наблюдения.

Для тех, кто хотел бы освежить свои знания статистики, полученные еще на первом курсе, существуют многочисленные учебники. Повторюсь: на освещение этой тематики полностью книга не претендует. Наша задача — описать самые простые и полезные методы и рассказать то, о чем умалчивают обычные учебники по статистике или, по крайней мере, чему они не уделяют особого внимания. Ограниченность подобной литературы — вот часть проблемы измерения, с которой сталкиваются менеджеры компаний, пытаясь решить возникшие перед ними задачи. Дело в том, что специалистов по статистическому анализу вопросы практического измерения «неизмеряемого», похоже, не очень интересуют. Нассбаум подметил эту общую тенденцию у авторов публикаций в «Journal of American Statistical Association» (JASA): «Даже то, что авторы JASA называют „практическими приложениями“, — сплошная теория».

В этой главе мы обсудим некоторые простые способы извлечения важной информации из малой выборки. Но, в отличие от большинства работ по статистике, вначале мы постараемся обойтись без математики и лишь потом приведем кое-какие формулы, да и то немного. А когда перейдем к конкретным расчетам, покажем, что способы быст-

рой оценки и простые таблицы и графики нередко бывают полезнее сложных уравнений.

Все приводимые примеры можно скачать со вспомогательного сайта www.howtomeasureanything.com Воспользуйтесь возможностями этого ресурса.

СКОЛЬКО МАШИН ЕЗДИТ НЕ НА ТОМ ТОПЛИВЕ?

Государственное агентство принимает подход «просто сделай это»

В 1970-е годы некоторые действия Агентства по защите окружающей среды вызвали негативную реакцию общественности. В 1975 г. начался выпуск автомобилей с каталитическими конвертерами, позволяющими использовать неэтилированный, то есть не содержащий тетраэтилсвинец, бензин. Но этилированный бензин был гораздо дешевле, и водители предпочитали использовать именно его в машинах, оборудованных новыми конвертерами. Известный всем сегодня устанавливаемый на горловину бензобака рестриктор был рекомендован Агентством специально для того, чтобы люди не заправляли новые автомашины этилированным бензином (на заправках такой бензин поступал из пистолетов большего диаметра). Однако водитель мог просто снять этот рестриктор и использовать бензин, содержащий свинец. Главный статистик ЕРА Барри Нассбаум сказал: «Мы знали, что люди заправляли новые машины этилированным бензином, потому что когда представители Управления автомобильного транспорта (Department of Motor Vehicle, DMV) проводили проверки, они смотрели, не снят ли рестриктор». Использование свинецсодержащего топлива в новых автомобилях могло привести к еще большему загрязнению воздуха, а не к улучшению экологической обстановки, что было целью программы. Какое-то время специалисты ЕРА просто не знали, что им предпринять. Разве можно определить, сколько людей заправляют свои новые машины старым бензином? Но, в конце концов, они решили «просто сделать это» — понаблюдать за автозаправками. Сначала в округе случайным образом были выбраны несколько заправок. Затем, вооружившись биноклями, представители Агентства принялись следить за машинами, отмечать, каким топливом их заправляли, и сравнивать номера этих автомобилей со списком типов машин, представленным Управлением автотранспорта. Этот метод вызвал всеобщее недовольство — карикатуристы из «Atlanta Constitution» даже изобразили сотрудников ЕРА в виде фашистов, арестовывающих автовладельцев, пользующихся не тем бензином, хотя за людьми только следили и никого не задерживали. И все же, как сказал Нассбаум, у Агентства возникли проблемы с несколькими отделениями полиции. Правда, полицейским пришлось в конечном счете согласиться, что любой имеет право наблюдать за кем-то из-за угла. Но самое главное — специалисты ЕРА установили желаемое: 8% машин, которые должны были работать на неэтилированном бензине, заправляли

горючим с тетраэтилсвинцом. Несмотря на всю сложность стоящей задачи, сотрудники Агентства поняли, что простые наблюдения и выборка помогут им снизить исходную неопределенность.

РАЗВИВАЕМ ИНТУИТИВНОЕ ПОНИМАНИЕ СЛУЧАЙНЫХ ВЫБОРОК: ПРИМЕР С ЛЕДЕНЦАМИ

Попробуйте выполнить небольшое задание. Укажите 90-процентный доверительный интервал для среднего веса одного леденца в граммах, то есть вы должны назвать всего два числа (нижнюю и верхнюю границы интервала), отстоящих друг от друга ровно настолько, чтобы вы были на 90% уверены в том, что средний вес леденца попадет в этот диапазон. Как всегда, делая калиброванную оценку вероятности, вы, конечно, имеете какое-то представление о неизвестном показателе, каким бы сомнительным оно ни было. Между прочим, один грамм — вес одного кубического сантиметра воды. Запишите предполагаемый CI, прежде чем двигаться дальше. Проверьте его, следуя указаниям из главы 5, сделав эквивалентную ставку, рассмотрев все «за» и «против» того, что диапазон значений обоснован, и испытав границы интервала на соответствие практике.

Предположим теперь, что у меня есть обычный пакетик с леденцами, который можно купить в любой кондитерской. Открыв этот пакетик, я вынимаю наугад несколько конфет и взвешиваю их по очереди на весах. Обратите внимание на вопросы, которые изложены далее. Ответьте на каждый, прежде чем перейти к следующему этапу.

1. Допустим, я скажу, что вес первого отобранного мной леденца — 1,4 грамма. Отвечает ли это вашему 90-процентному доверительному интервалу? Если нет, то каким будет ваш новый 90-процентный CI? Запишите измененный интервал, прежде чем двигаться дальше.
2. Теперь я сообщаю результаты взвешивания остальных четырех из пяти наугад отобранных леденцов: 1,4; 1,5; 1,6 и 1,1 грамма. Как теперь изменится 90-процентный CI? Запишите и его.
3. Наконец, я говорю о результатах взвешивания еще трех леденцов (так что всего их стало восемь): 1,5; 0,9 и 1,7 грамма. Каков теперь ваш 90-процентный CI? Запишите окончательный вариант.

Всякий раз при поступлении новой информации ваш доверительный интервал должен сужаться. Если сначала (до взвешивания) он был очень широк, то уже после первого взвешивания должно произойти его существенное сокращение.

Я дал этот тест девяти калиброванным экспертам и получил практически одинаковые результаты. Главное различие между ними состояло в степени неуверенности специалистов в своей первоначальной оценке. Самый узкий интервал составлял 1–3 грамма, а самый широкий — 0,5–50 граммов, однако чаще всего диапазоны выглядели, скорее, узкими. Получив дополнительную информацию, эксперты в большинстве своем (особенно начинавшие с очень широкого интервала) быстро сузили свои диапазоны. Эксперт, указавший интервал 1–3 грамма, не сократил его после первого взвешивания, а тот, кто предложил 0,5–50 граммов, сильно опустил верхнюю границу и в конце концов пришел к интервалу 0,5–6 граммов.

На самом деле один леденец из этого пакетика весил в среднем около 1,45 грамма. Интересно, что эксперты довольно быстро сузили свои интервалы на основании дополнительной информации всего о нескольких конфетах.

Подобные упражнения помогают развить интуитивное понимание сути случайной выборки и интервала. Как мы вскоре увидим, выяснение субъективного мнения калиброванных экспертов без применения того, что кто-то назвал бы «настоящей статистикой», весьма полезно и даже имеет ряд преимуществ перед традиционными статистическими методами. Но давайте сначала узнаем, что говорится о малых выборках в большинстве работ по статистике.

КОЕ-ЧТО О МАЛЫХ ВЫБОРКАХ:

ПОДХОД ПИВОВАРА

Точно рассчитать 90-процентный СИ для среднего веса одного леденца, а не полагаться на мнение калиброванных экспертов позволяет метод, разработанный одним пивоваром. Он преподается в базовом курсе статистики и часто используется для определения погрешности выборки, состоящей всего из двух объектов. В самом начале XX века Уильяму Сили Госсету, химику и статистику ирландской пивоваренной компании Guinness, потребовалось установить, какой из двух сортов ячменя дает лучшее пиво с большим выходом. Ранее был разработан метод z-значения (z-score), или статистики нормальных распределений, позволяющий находить доверительный интервал на основе случайной выборки, состоящей из не менее чем 30 объектов. Метод дает распределения, не отличающиеся по форме от обсуждавшегося выше нормального. К сожалению, у Госсета не было возможности

протестировать большое число партий пива, изготовленных из каждого сорта ячменя. Однако он не отказался от своей затеи измерить то, что как будто не поддавалось оценке, и решил вывести новый вид распределения для крайне малых выборок. К 1908 г. Госсет разработал новый эффективный метод, который назвал *t*-статистикой, и захотел опубликовать результаты своей работы.

Однако у Guinness уже были проблемы с разглашением коммерческой тайны, и служащим компании было запрещено публиковать любую информацию о бизнес-процессах. Госсет понимал значение своей работы, но ему сильнее хотелось рассказать о своей идее, чем добиться немедленного признания. Поэтому он опубликовал статью под псевдонимом «Стьюдент». И хотя истинный автор давно известен, практически во всех работах по статистике метод называется *t*-статистикой Стьюдента.

Вид распределения Стьюдента напоминает обсуждавшееся нами ранее нормальное распределение. Но в случае очень малых выборок его форма становится намного уплощеннее и шире. Рассчитанный с помощью *t*-статистики 90-процентный CI намного шире (то есть неопределеннее), чем в случае нормального распределения. Если размер выборки больше 30, то график *t*-распределения практически совпадает с нормальным распределением.

И для одного, и для другого типа распределения существует сравнительно простой (по сравнению со многими другими статистическими методами) способ расчета 90-процентного доверительного интервала для среднего значения генеральной совокупности. Кому-то наши расчеты могут показаться слишком сложными, а те, кто уже знаком с данным методом, скажут, что мы просто пересказываем содержание учебников по статистике. Пусть первые подождут, пока мы не рассмотрим в следующей главе намного более простое решение, а вторые просто пропустят этот материал. Адресуя свои пояснения читателям, которые отнесут себя к средней категории, я старался сделать их как можно более простыми для восприятия. Вот как мы могли рассчитать 90-процентный CI в предыдущем примере, отобрав всего пять леденцов.

1. Рассчитаем сначала дисперсию выборки (этим понятием мы позднее будем часто пользоваться):

- а) рассчитаем средний вес отобранных леденцов: $(1,4 + 1,4 + 1,5 + 1,6 + 1,1) / 5 = 1,4$;
- б) вычтем это среднее из каждого значения в выборке и возведем полученные результаты в квадрат: $(1,4 - 1,4)^2 = 0$; $(1,5 - 1,4)^2 = 0,01^2$ и т. д.;
- в) суммируем все квадраты и разделим на размер выборки минус единица: $(0 + 0 + 0,01 + 0,04 + 0,09) / (5 - 1) = 0,035$.

2. Разделим дисперсию выборки на ее размер и извлечем из полученного результата квадратный корень. В электронной таблице мы записали бы «=SQRT (0,035/5)» и получили 0,0837.

(В работах по статистике это называется средним квадратичным отклонением.)

3. Найдем в таблице 9.1 (таблице упрощенных значений t-статистики) значение t, соответствующее размеру выборки: для выборки, состоящей из пяти объектов, $t = 2,13$. Обратите внимание, что для очень больших выборок t близко к z-значению (нормальное распределение) 1,645.
4. Умножим найденное t-значение на результат этапа 2: $2,13 \times 0,0837 = 0,178$. Это ошибка выборки в граммах.
5. Суммируем ошибку выборки и средний вес леденца, чтобы получить верхнюю границу 90-процентного CI, а затем вычтем ее из среднего веса, чтобы получить нижнюю границу: верхняя граница $= 1,4 + 0,178 = 1,578$; нижняя граница $= 1,4 - 0,178 = 1,222$.

Итак, выбрав всего пять леденцов, мы получили 90-процентный CI, составляющий 1,222–1,578. Аналогично рассчитывают доверительный интервал и для более крупных выборок. Единственная разница

ТАБЛИЦА 9.1.

ТАБЛИЦА УПРОЩЕННЫХ t-ЗНАЧЕНИЙ
Выберите ближайший размер выборки
(или воспользуйтесь интерполяцией,
если хотите добиться большей точности)

Размер выборки	t-значение
2	6,31
3	2,92
4	2,35
5	2,13
6	2,02
8	1,89
12	1,80
16	1,75
28	1,70
Более крупные выборки	1,645 (z-значение)

заключается в том, что z-значение, необходимое для этого, всегда будет составлять 1,645 (с ростом размера выборки оно не увеличится).

На рисунке 9.1 представлен общий результат решения другой гипотетической задачи с использованием t-статистики. Это могло быть определение средней партии пива, сваренного в компании Guinness, среднего времени, проводимого покупателями в очередях, или среднего размера обуви жителей штата Небраска. В любом случае необходимо определить 90-процентный CI для среднего значения генеральной совокупности, хотя по каким-то причинам (экономические факторы, ограниченность во времени или несогласие жителей штата Небраска на измерение размеров их ног) размер выборки из подобных совокупностей составлял бы не десятки и сотни, а всего несколько образцов.

На этом рисунке нижняя ломаная линия — график значений выборки, пунктир — истинное среднее значение генеральной совокупности. Одни значения выборки выше этого среднего, другие ниже. Фактических значений на рисунке нет, но для наших целей смысл



РИСУНОК 9.1.

Изменение неопределенности в зависимости от размера выборки

понятен. В верхней части рисунка вы видите три кривые, левые края которых образуют своеобразную «воронку торнадо». Средняя линия — график скользящей средней выборки (сначала среднее из первых трех значений, затем среднее из первых четырех значений и т.д.), которая сравнивается с истинным средним генеральной совокупности, представленным прямой пунктирной линией. Две внешние кривые — графики верхней и нижней границ 90-процентного доверительного интервала, пересчитываемого после каждого нового пополнения выборки.

Обратите внимание, что пока выборка невелика, 90-процентный CI достаточно широк, быстро сужаясь всякий раз, как она пополняется. Разброс значений в выборке по-прежнему велик, но скользящее среднее значение все больше приближается к истинному среднему. Заметьте и то, что при 30 членах выборки 90-процентный CI намного уже, но ненамного уже, чем при 20 или 10 членах. Когда выборка насчитывает 30 объектов, снижение ошибки вдвое требует увеличения ее размера в четыре раза (120 объектов). Чтобы сократить ошибку в четыре раза, нужно увеличить размер выборки в 16 раз (480 объектов). Короче говоря, каждое уменьшение ошибки выборки вдвое требует умножения ее размера на четыре.

И t-статистика, и нормальная z-статистика — разновидности так называемой параметрической статистики. Она предполагает конкретную форму распределения. Во многих случаях можно смело допустить, что распределение переменных является нормальным, но это не всегда так. Поэтому хотя параметрическая статистика и не опирается на субъективные оценки калиброванных экспертов, она все равно начинается с достаточно произвольного допущения о типе распределения.

В конечном счете разница между субъективным и объективным может не иметь большого значения, во всяком случае, на эффективность этих методов она не влияет. Единственное, что должно нас беспокоить, — как они работают на практике. Какая оценка интервала на основе малой выборки точнее — сделанная калиброванными экспертами или полученная с помощью параметрической статистики? Не ошибаются ли эксперты чаще, чем параметрическая статистика?

В эксперименте с калиброванными экспертами и леденцами специалисты постоянно давали более широкие (хотя и ненамного) интервалы значений, чем t-статистика. Это означает, что дополнительные несложные расчеты позволяют уточнить оценки калиброванных экспертов. Когда мы выбрали восемь леденцов, интервалы, предложенные самым осторожным и самым уверенным в себе экспертом, составляли 0,5–2,4 и 1–1,7 грамма соответственно. При выборке такого же размера t-статистика дала бы 90-процентный CI для среднего веса конфеты 1,21–1,57 грамма. Мы видим, что последний гораздо уже, чем самый узкий из интервалов, предложенных специалистами. Но хотя сниже-

ние неопределенности в нашем эксперименте и не было таким заметным, как можно ожидать, оно все равно оказалось значительным по сравнению с исходным состоянием. И как мы еще увидим в главе 10, дальнейшие исследования подтверждают эти выводы.

Итак, мы установили, что:

- когда исходная неопределенность высока, для ее существенного снижения достаточно изучить несколько объектов из генеральной совокупности;
- калиброванные эксперты сумели снизить неопределенность, отобрав из генеральной совокупности всего один объект, чего не может традиционная параметрическая статистика;
- оценки калиброванных экспертов обоснованны, но осторожны. Чтобы снизить неопределенность еще больше, нужно провести дополнительные расчеты.

ПРОСТЕЙШИЕ СТАТИСТИЧЕСКИЕ МЕТОДЫ ПОЛУЧЕНИЯ ВЫБОРОК

Ясно, что эксперты могут сильно снизить исходную неопределенность, сделав всего нескольких наблюдений, а параметрический метод позволяет сократить ее еще больше. Но всегда ли калиброванные оценки уступают результатам, полученным с помощью параметрической статистики? Нет, не всегда. И *t*-статистика, и нормальная статистика учитывают только значения в выборочной совокупности, не принимая во внимание условия выборки, предварительную информацию, имевшуюся у эксперта перед проведением оценки, или наличие иных факторов. Другими словами, параметрические методы не позволяют воспользоваться многим из того, что мы называем здравым смыслом, так как не принимают во внимание информацию, интуитивно учитываемую калиброванными экспертами.

Предположим, что вместо взвешивания леденцов мы спросили у руководителей отделов продаж, сколько времени у них отнимает работа с отстающими торговыми представителями. Допустим, мы выбрали всего пять руководителей и они сказали, что тратят на эти цели 1, 6, 12, 12 и 7 часов в неделю. Применив *t*-статистику, получаем 90-процентный CI, составляющий 3,8–13 часов. Однако в статистическом уравнении не учтен тот факт, что 1 час — это временные затраты Боба, у которого, как всем известно, больше проблем с персоналом, чем у кого-либо другого, и который наверняка намеренно занижил этот показатель. А калиброванный эксперт, напротив, обязательно использовал

бы эту дополнительную информацию. Оценка эксперта субъективна и поэтому кажется ненадежной, но в среднем она почти не уступает, а иногда и превосходит по точности оценку «объективной» статистики.

Более того, при очень малых выборках t -статистика часто дает доверительный интервал, который не имеет смысла с учетом известных ограничений. Предположим, что вы проанализировали ведомости учета рабочего времени и увидели, что ни один руководитель отдела продаж не тратит на общение со своими торговыми представителями более 12 часов в неделю, а затраты времени на общение с отстающими работниками еще меньше. И торговые представители, имеющие неважные показатели, утверждают, что общаются с руководителями не более 12 часов в неделю. Тогда верхняя граница нашего 90-процентного CI (13 часов) выглядит нереалистичной, но это лучший расчет t -статистики.

Я придумал другой, более простой подход, позволяющий не привлекать калиброванных экспертов и использующий только значения выборки. Вместе с тем он свободен от недостатков t -статистики. В главе 3 мы вскользь обсуждали Правило пяти. Как вы помните, оно гласит, что если выбрать случайным образом пять объектов из любой генеральной совокупности, то существует вероятность 93,8% того, что медиана этой совокупности окажется в интервале между наибольшим и наименьшим значениями выборки. Но это лишь одно из правил, касающихся малых выборок. Если бы мы нашли метод определения 90-процентного CI непосредственно по значениям выборки, то быстро находили бы его без всяких расчетов.

Если выбрать восемь объектов, то наибольшее и наименьшее их значения зададут CI, гораздо более широкий, чем 90-процентный (в действительности, это был бы примерно 99,2-процентный CI). Но оказывается, что если взять вторые наибольшее и наименьшее значения в этой выборке, то получим интервал, гораздо более близкий к 90-процентному (примерно 93%). При выборке из 11 объектов интервал окажется еще ближе к 90-процентному для третьих наибольшего и наименьшего значений.

Таблица 9.2 показывает, выборки какого размера позволяют получить CI, близкий к 90-процентному, просто взяв одно из самых высоких и одно из самых низких значений из нее. Например, при выборке 18 объектов верхняя и нижняя границы CI, близкого к 90-процентному, будут примерно равны шестым наибольшему и наименьшему значениям. Осторожности ради я специально выбирал размеры выборки, позволяющие получить не 90-процентный, а несколько более широкий CI. Третий столбец «Фактическая степень уверенности» показывает вероятность того, что медиана окажется в интервале, границами которого служат n -ное наибольшее и n -ное наименьшее значения. Не беспокойтесь о том, как использовать этот столбец. Достаточно знать,

ТАБЛИЦА 9.2.

Таблица для определения 90-процентного СИ без математических расчетов

Нижняя граница — n -ное наименьшее значение
Верхняя граница — n -ное наибольшее значение

Размер выборки	n -ное наибольшее и n -ное наименьшее значения выборки	Фактическая степень уверенности (%)
5	1-е	93,8
8	2-е	93,0
11	3-е	93,5
13	4-е	90,8
16	5-е	92,3
18	6-е	90,4
21	7-е	92,2
23	8-е	90,7
26	9-е	92,4
28	10-е	91,3
30	11-е	90,1

что получаемый СИ максимально близок к истинному 90-процентному и не слишком узок (то есть является довольно осторожной оценкой).

Я называю этот 90-процентный СИ определяемым без математических расчетов, так как все, что от нас требуется, — это записать несколько цифр, касающихся наибольшего и наименьшего значений выборки. Вам не придется рассчитывать ее дисперсию, извлекать квадратные корни или использовать таблицы t -значений. Составляя данную таблицу, я пользовался некоторыми непараметрическими методами и проверил полученные результаты методом Монте-Карло для малых выборок. Не стану утруждать вас подробностями, но ее использование существенно упрощает определение 90-процентного СИ на основе малых выборок. Постарайтесь запомнить, что имея выборку из 5, 8, 11 или 13 объектов, для определения на ее основе границ 90-процентного СИ нужно взять, соответственно, ее первое, второе, третье или четвертое наибольшее и наименьшее значения. Теперь можно быстро оценить 90-процентный СИ, отобрав лишь несколько объектов для наблюдения, не доставая из кармана калькулятор.

Если говорить коротко, данный прием работает потому, что при расчете 90-процентного СИ «середина» ряда значений особой роли не

играет. Чтобы объяснить почему, нам придется еще раз вернуться к параметрическим методам. Они, как мы видели на примере параметрической *t*-статистики, требуют расчета дисперсии выборки. Из каждого значения в выборке вычитается среднее по выборке, а полученный результат возводится в квадрат. Сумма всех этих квадратов и есть дисперсия. Выполнив эти простые расчеты, вы обнаружите, что величина дисперсии почти целиком зависит от значений, намного отличающихся от среднего. Даже у больших выборок дисперсия обычно лишь на 2% зависит от средней трети значений и на 98% — от других двух третей наибольших и наименьших значений выборки. Когда размер выборки не превышает 12 объектов, дисперсия определяется всего двумя экстремальными точками — наибольшим и наименьшим значениями выборки.

Данный не требующий расчетов метод позволяет получить 90-процентный CI, лишь чуть-чуть более широкий, чем при использовании *t*-статистики, без связанных с ней проблем. Вспомним пример, в котором пять руководителей отделов продаж указали, что тратят на общение с отстающими торговыми представителями 1, 6, 12, 12 и 7 часов в неделю. С помощью *t*-статистики мы установили, что верхняя граница интервала — 13, но знаем, что по другим соображениям она должна быть ниже, и с помощью нашего не требующего расчетов метода получим интервал 1–12. Мы *знаем*, что 12 — вполне возможное значение верхней границы интервала, так как это одно из значений нашей выборки. Если отобрать еще шесть руководителей с затратами времени 4, 5, 10, 7, 9 и 10 часов в неделю, то выборка составит 11 объектов. Таблица показывает, что при таком размере выборки границами CI, близкого к 90-процентному, служат третьи наибольшее и наименьшее ее значения. С учетом этого получаем 90-процентный CI, составляющий 5–11 часов в неделю. А *t*-статистика в этом (довольно редком) случае даст нам несколько более широкий интервал 4,5–11,3 часа в неделю.

Важно отметить, что использованный мной непараметрический метод в отличие от *t*-статистики определяет 90-процентный CI для медианы, а не для среднего значения. Медианой генеральной совокупности называют такое значение, выше которого располагается ровно одна половина ее значений, а ниже — другая. Среднее генеральной совокупности — это сумма всех ее значений, деленная на размер. При смещенном (асимметричном) распределении генеральной совокупности медиана не совпадает со средним значением. Однако если допустить, что распределение близко к симметричному, то медиана и среднее совпадут. В этом случае наша таблица позволит определить 90-процентный CI и для медианы, и для среднего значения.

В некоторых случаях данное допущение оказывается натяжкой, но вообще-то в параметрической статистике мы делаем гораздо более сомнительные допущения. В параметрической статистике мы обязаны

придать графику нашего распределения вполне определенную форму. А оценивая медиану по таблице 9.2, мы не делаем *никаких* допущений о распределении значений генеральной совокупности. Оно может быть и нерегулярным — горбатым (camel-back) (как график распределения населения США по возрасту, форма которого объясняется произошедшим после войны демографическим взрывом), и равномерным (как график распределения выигрышей при игре в рулетку). Таблица 9.2 позволяет определить диапазон значений медианы и в том, и в другом случаях. Но если распределение к тому же симметрично, неважно, равномерное оно, нормальное, горбатое или типа «бабочка» (bow-tie), то таблица годится и для определения диапазона среднего значения.

ПРИСТРАСТНЫЙ ОТБОР МЕТОДОВ ВЫБОРОЧНОГО ОБСЛЕДОВАНИЯ

Как обычный работник измерил бы популяцию рыб, обитающих в озере? Этот вопрос я всегда задаю участникам своих семинаров. Обычно в ответ слышу: «Осушил бы озеро». По мнению, например, среднего бухгалтера или даже менеджера среднего звена по ИТ, «измерить» означает «пересчитать». Поэтому когда речь заходит о численности (популяции) рыб, такие люди полагают, что их просят назвать точный итог, а не просто уменьшить неопределенность. С этой мыслью они и предлагают осушить озеро и, несомненно, сумели бы организовать дело так, чтобы каждая мертвая рыбешка была подобрана, брошена в кузов грузовика и сосчитана вручную. Возможно, кто-то пересчитал бы рыбу в грузовике еще раз и осмотрел бы дно осушенного озера, чтобы убедиться в точности подсчетов. Затем они сообщили бы, что всего в озере обитали ровно 22 573 рыбы, так что прошлогодние усилия по пополнению рыбных запасов озера не пропали даром. Правда, теперь вся эта рыба погибла.

А вот если поручить биологам измерить численность рыбной популяции в озере, то уж они не спутают слова «измерить» и «пересчитать». Взамен они, скорее всего, воспользуются методом, состоящим в выпуске пойманной рыбы и повторной ловле. Сначала биологи поймают и пометят некое число, скажем 1000, особей и снова выпустят их в озеро. После того как меченая рыба перемешается с немеченой, они отлавливают еще некое число особей. Допустим, поймали опять 1000 рыб, из которых 50 меченых. Это означает, что помечено 5% всех имеющихся в озере рыб. Зная число первоначально меченых рыб — 1000, биологи делают вывод: в озере около 20 тыс. рыб (1000 — это 5% от 20 000).

Такого рода выборка подчиняется так называемому биномиальному распределению, но для больших чисел можно считать такое рас-

пределение нормальным. Ошибку оценки рассчитывают практически так же, как и ранее. Все, что нужно сделать, — это изменить порядок расчета дисперсии выборки, остальное не меняется. В данном случае дисперсию определяют как долю объектов в группе, численность которой мы пытаемся узнать, умноженную на долю объектов вне группы. Иными словами, берем долю меченой рыбы (0,05), умножаем ее на долю немеченой рыбы (0,95) и в результате получаем 0,0475.

Дальнейший порядок расчетов тот же. Делим дисперсию выборки на размер выборки и извлекаем квадратный корень из суммы: $\text{SQRT}(0,0475:1000) = 0,007$. Чтобы получить 90-процентный CI для доли меченых рыб в озере, к этой доле (0,05) прибавляем и вычитаем из нее 0,007, умноженные на 1,645 (z-значение 90-процентного CI). Результат: помечено 3,8–6,8% всех рыб в озере. Мы знаем, что пометили 1000 рыб, тогда всего в озере от $1000/0,062 = 16\,256$ до $1000/0,032 = 25\,984$ особей.

Кому-то такой интервал покажется широким. Но предположим, что первоначально он выглядел как 2000–50 000 особей. Допустим также, что цель состояла только в том, чтобы определить, растет популяция или вымирает, и что ранее в озеро запустили 5000 мальков. Любое число свыше 6000 скажет нам о том, что популяция, как минимум, растет, а цифра свыше 10 000 — что запускать новых мальков больше не нужно. С учетом первоначального интервала значений и соответствующего порога полученная новая степень неопределенности — явный шаг вперед, а погрешность в данном случае вполне приемлема. На самом деле, и в первый, и во второй раз улов мог состоять не из 1000, а только из 250 рыб, но все равно мы были бы уверены, что популяция увеличилась так, что ее численность превысила 6000 особей.

Приведенный пример наглядно показывает, что выборочное обследование позволяет получить информацию даже о том, что никто не видит. Метод дает возможность определять, например, сколько человек остались не учтенными Бюро переписи населения США, сколько видов бабочек до сих пор не открыто в бассейне Амазонки, сколько несанкционированных попыток доступа в информационную систему было предпринято за прошедший период, а также сколько потенциальных клиентов не сумели выявить менеджеры компании. Если нельзя увидеть целиком какую-то группу объектов, это еще не означает, что измерить их численность тоже невозможно.

Метод выпуска пойманной рыбы и повторной ловли — лишь один из многочисленных приемов выборочного обследования. Несомненно, новые, более эффективные способы еще будут разработаны. Однако даже поверхностное знакомство с основными методами выборки позволяет правильно оценивать наблюдения и получать информацию, необходимую для решения самых разнообразных проблем.

Преднамеренная выборка

Преднамеренную (разовую) выборку можно сравнить со случайной моментальной фотографией людей, процессов или предметов в отличие от постоянного наблюдения за ними в течение некоего периода времени. Например, желая узнать, какую часть своего рабочего времени сотрудники тратят на данный вид деятельности, вы проверяете в один из дней, чем наугад выбранные люди заняты *в настоящий момент*. Окажись, что в 12 случаях из 100 работники участвовали в селекторном совещании, вы сделаете вывод: 12% своего времени они тратят на селекторные совещания (90-процентный CI составляет 8–18%). В этом примере, как и в случае с измерением численности рыбной популяции, мы имеем дело с биномиальным распределением. В каждый конкретный момент времени люди либо занимаются данным видом деятельности, либо нет, а вы просто спрашиваете, какую часть своего времени они на это тратят. И здесь размер выборки позволяет, как и ранее, считать распределение нормальным.

Но что, если удалось разово отобрать только 30 человек, из которых лишь пятеро занимаются интересующим вас видом деятельности? Достаточно ли этой информации? И вновь все зависит от степени исходной неопределенности до проведения выборочного обследования. Для ее снижения, возможно, этих данных и хватит. Рисунок 9.2 предоставляет способ быстрого определения 90-процентного CI для доли генеральной совокупности на базе малой выборки. Следуя указаниям на данном рисунке, вы получите такие результаты: нижняя граница 90-процентного CI — примерно 9%, а верхняя — около 31%. Предположим, что предыдущая оценка интервала дала итог 5–50%, а порог, начиная с которого на базе той же информации придется принимать другое решение, — 40%. Выборка такого размера вполне достаточна для принятия подобного решения.

Кластерная выборка

Кластерное выборочное обследование — это случайная выборка из групп, проведение полного наблюдения или осуществление более концентрированной выборки в полученной группе. Например, если вы хотите узнать, какой процент семей имеет спутниковые тарелки или отделяет пластмассу от других бытовых отходов, иногда проще всего выбрать случайным образом в городе несколько кварталов, а затем провести сплошное обследование семей, проживающих в каждом из этих кварталов (на обследование семей, проживающих в разных концах города, ушло бы слишком много времени). В подобных случаях мы не можем исходить из того, что число объектов в случайной

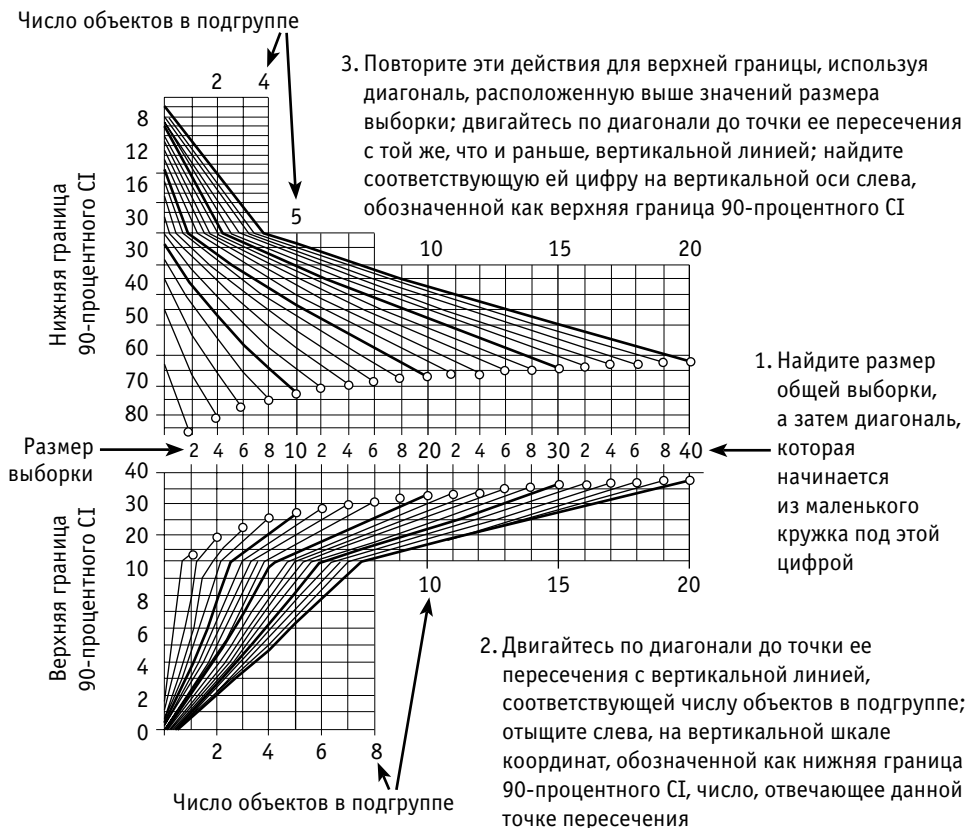


РИСУНОК 9.2.

**Малые выборки: 90-процентный CI
для доли генеральной совокупности**

выборке должно равняться числу объектов в группах (в данном случае числу семей). Семьи, проживающие в одном квартале, могут быть очень похожими, поэтому нельзя считать, что размер случайной выборки определяется их числом. Когда семьи в квартале практически одинаковы, за размер выборки лучше принять число кварталов.

Стратифицированные выборки

При стратифицированном отборе разные методы и (или) размеры выборки применяются к разным группам в составе одной генеральной совокупности. Этот подход имеет смысл, когда генеральная совокупность состоит из нескольких сильно отличающихся друг от друга, но

внутренне однородных групп. Если вы владеете рестораном быстрого питания и хотите определить демографический состав своих клиентов, то, возможно, следует применять разные методы выборочного обследования к клиентам за столиками и к тем, кто берет еду на вынос. Если вы директор завода и хотите проверить, как соблюдается техника безопасности, то, возможно, методы наблюдения за сварщиками, мастерами и уборщицами должны различаться.

Серийная выборка

В учебниках по статистике метод серийной выборки обычно не освещается. Не обсуждали бы его и мы, называясь наша книга «Как измерить большинство вещей». Но в период Второй мировой войны данный подход широко использовался для сбора разведывательной информации¹. Он может пригодиться и сегодня для решения некоторых бизнес-проблем. В годы Второй мировой разведчики союзников собирали сведения о производстве вражеской военной техники, в том числе немецких танков «Mark V». Сведения о производстве этих тяжелых танков были очень противоречивыми, и руководство разведки союзников не знало, чему верить. В 1943 г. работавшие на союзников статистики разработали метод определения объема производства на основе серийных номеров захваченных танков. Серийные номера следовали друг за другом и включали какую-нибудь дату. Однако по одному серийному номеру еще нельзя было точно сказать, когда был произведен первый танк серии (поскольку серия могла начинаться не с номера 001). Здравый смысл подсказывает, что минимальное месячное производство должно равняться, по крайней мере, разности между последним и первым серийными номерами машин, захваченных в данном месяце. Но нельзя ли узнать из этих номеров что-нибудь еще?

Подойдя к захваченным танкам как к случайной выборке из всей генеральной совокупности, статистики поняли, что могут рассчитать вероятность производства в разных объемах. Например, если в каком-то месяце было выпущено 1000 танков, то вероятность случайно захватить 10 произведенных в этот период машин с серийными номерами, отличающимися друг от друга не более чем на 50 единиц, крайне мала. Разброс номеров 10 танков, случайным образом выбранных из 1000, должен быть больше. Если же в этом месяце было произведено только 80 танков, то захват 10 из них с такими близкими номерами, по крайней мере, возможен.

Таблица 9.3 демонстрирует данные о выпуске танков «Mark V» по сведениям разведки и расчетам статистиков в сравнении с фактическим производством (сведения из захваченных после войны докумен-

ТАБЛИЦА 9.3.

**Производство немецких танков «Mark V»
в период Второй мировой войны**

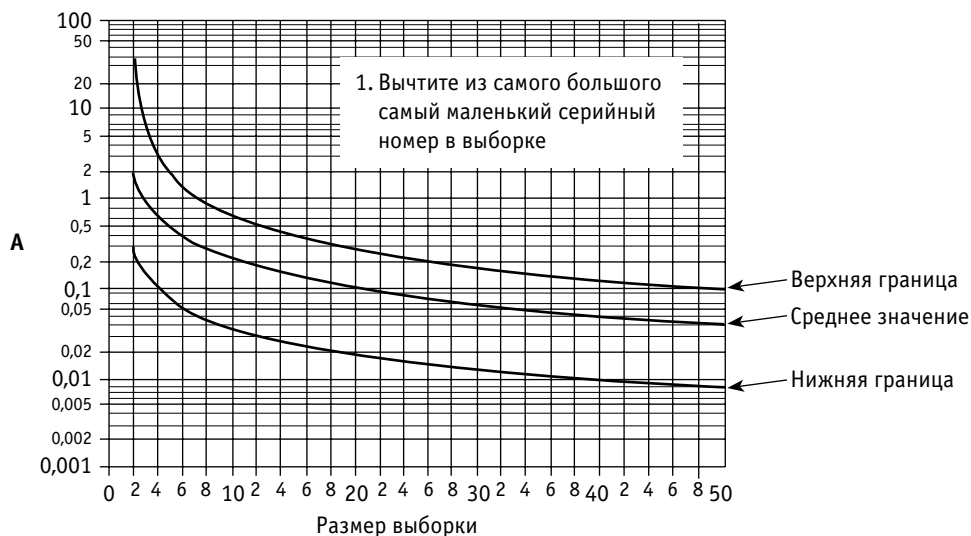
Месяц, год	Оценка разведки	Расчеты статистиков	Фактическое производство ^a
Июнь, 1940	1000	169	122
Июнь, 1941	1550	244	271
Август, 1942	1550	327	342

^a Согласно захваченным после войны документам.

тов). Сравнение подтверждает эффективность статистического метода, основанного на анализе серийных номеров захваченных машин.

Более того, дать оценку, значительно более точную, чем исходные данные разведки, наверное, можно было по номерам нескольких танков. Рисунок 9.3 показывает, как по случайной выборке предметов с серийными номерами определяют размер всей генеральной совокупности. Следуя указаниям рисунка, рассмотрим пример, когда число трофеев составляет восемь объектов (которыми могут быть товары конкурирующей фирмы, найденные в мусоре страницы ее отчета и т.д.). Самый большой серийный номер — 100 220, а самый маленький — 100 070, так что в результате этапа 1 мы получаем 150. Результат этапа 2 — около 1,0 (в этой точке кривая верхней границы пересекает вертикальную линию для нашего размера выборки — 8). На этапе 3 мы производим простые вычисления $(1+1,0) \times 150 = 300$ и получим значение верхней границы CI. Повторив эти шаги для нижней границы и среднего значения, получаем 90-процентный доверительный интервал 156–300 со средним значением 195 (обратите внимание, что среднее — это не середина диапазона, поскольку распределение асимметрично). Так что статистики могли сделать свои выводы, располагая всего восемью номерами захваченных танков.

Два предостережения: если захвачены машины одной бронетанковой части, ни одна из них не может считаться отобранной случайно, поскольку у них могут быть близкие номера. Однако обычно это сразу видно по самим номерам. Кроме того, когда на самом деле нумерация серии не совсем последовательная (то есть каждый следующий номер присваивался не следующему танку) и какие-то номера пропущены, данный метод требует определенной модификации. Отметим: необходимо, чтобы распределение используемых номеров легко обнаруживалось. Например, если используются только четные номера



2. Найдите размер выборки на горизонтальной оси и соответствующую ему точку на кривой верхней границы

3. Найдите на вертикальной оси значение **A**, самое близкое к этой точке на кривой, и добавьте к нему 1; умножьте полученный итог на результат шага 1. Это — верхняя граница 90-процентного CI для всех предметов, имеющих серийные номера

4. Повторите шаги 2 и 3 для нижней границы и среднего значения

РИСУНОК 9.3.

Выборочное обследование серийных номеров

или номера, отличающиеся друг от друга на пять, то это должно быть очевидным из выборки.

Как это можно применить в бизнесе? «Серийные номера» (то есть последовательные серии) можно найти в современном мире где угодно. Так, компании бесплатно предоставляют конкурентам информацию о своем объеме производства, просто указывая на товарах серийные номера, которые может увидеть любой покупатель. (Однако, чтобы быть случайной, такая выборочная совокупность должна состоять из товаров, купленных в разных магазинах.) Аналогичным образом несколько страниц из выброшенного отчета конкурента или цифр из квитанции могут многое рассказать об остальных страницах отчета или обо всех квитанциях за данный день. Я вовсе не призываю вас копаться в отбросах, но исследование содержимого мусорных контейнеров нередко позволяет решить интересные задачи по измерению.

ОПРЕДЕЛИТЕ ПОРОГ

Обычно мы хотим что-то измерить, так как требуется обосновать какое-то решение. И для всех решений обычно имеется некое пороговое значение: если интересующий нас показатель окажется выше его, то мы примем одно решение, а если ниже, то другое. Но статистические методы в большинстве своем не занимаются выяснением, при каком значении X следует принять то или иное решение. Я хочу познакомить вас с таким статистическим методом, который позволяет не только снизить неопределенность в целом, но и сравнить интересующий показатель с важным пороговым значением.

Предположим, требуется определить средние затраты времени сотрудников на совещания, которые в наш век Интернета могли бы проводиться и дистанционно. Работники не теряли бы время на дорогу, а совещания не срывались бы из-за проблем с транспортом. Чтобы решить, следует ли проводить данное совещание дистанционно, нужно выяснить, что на нем происходит. Если сотрудники, которые и так постоянно общаются друг с другом, обсуждают рутинные вопросы, но ради этого кому-то приходится ехать издалека, то подобное мероприятие, наверное, лучше проводить дистанционно. Начнем с калиброванной оценки времени, необходимого среднему сотруднику на то, чтобы попасть на совещание, которое могло бы проводиться дистанционно (3–15%). Далее определяем, что если этот показатель превысит 7%, то инвестировать немалые средства в подобные виртуальные мероприятия стоит. Расчет ожидаемой стоимости полной информации показывает, что необходимо израсходовать не больше 15 000 дол. на исследования по этому вопросу. Согласно нашему правилу определения затрат на проведение измерений, можно потратить на эти цели примерно 1500 дол., так что если в компании тысячи сотрудников, то сплошной опрос всех участников совещаний абсолютно исключен.

Предположим, что мы выбрали 10 человек и после подробного анализа времени, которое они тратят на дорогу и проводившиеся в последнее время совещания, выяснилось, что только один сотрудник расходует менее 7% своего времени на эти виды деятельности. Какова, с учетом этой информации, вероятность того, что интересующий нас средний показатель действительно меньше 7% и подобные инвестиции совершенно неоправданны? Ответ на уровне здравого смысла — один к десяти, или 10%. Но это как раз тот случай, когда здравый смысл совсем не так полезен, как немного математики. На самом деле эта вероятность намного меньше.

Рисунок 9.4 показывает, как можно оценить вероятность того, что медиана генеральной совокупности находится по одну сторону порога при условии, что половина или большинство значений малой выборки — по другую сторону.

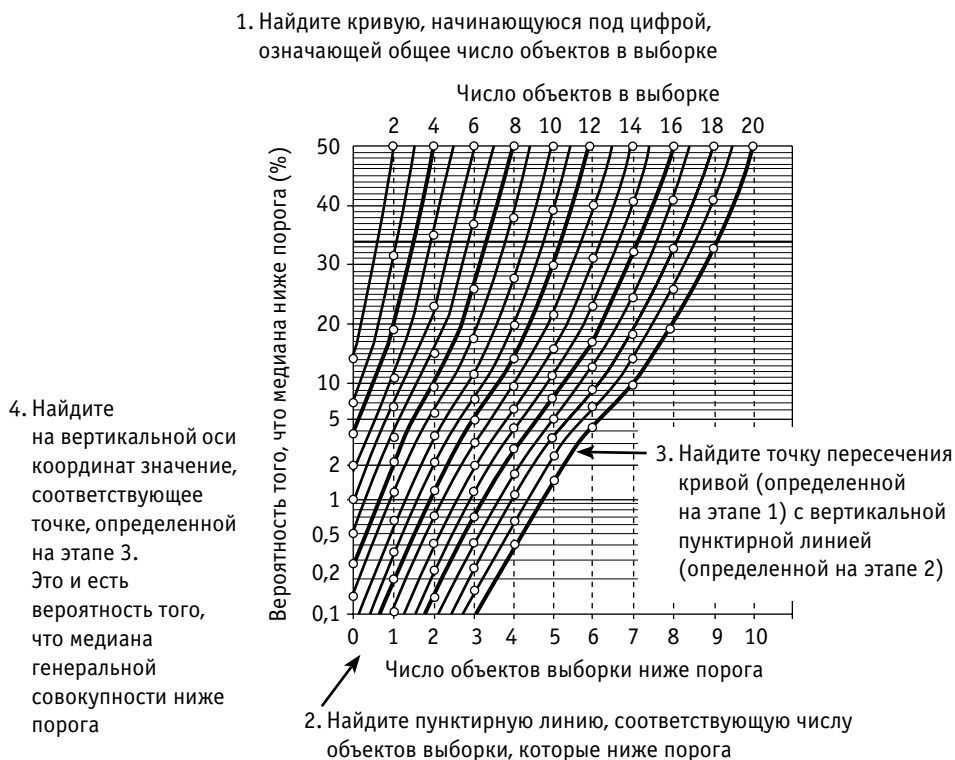


РИСУНОК 9.4.

Расчет вероятности положения медианы относительно порогового значения

Попрактикуйтесь в обращении с рисунком 9.4.

1. Найдите в верхней части диаграммы, где указаны размеры выборок, цифру 10. Проследите, куда ведет сплошная кривая, соединяющая эту цифру с вертикальной осью координат.
2. Найдите в нижней части рисунка, где указано число объектов выборки ниже порогового, цифру 1. Проследите, куда ведет соответствующая этой цифре вертикальная пунктирная линия.
3. Найдите точку пересечения кривой и пунктирной линий.
4. Соответствующий этой точке процентный показатель на вертикальной оси координат (0,6%) показывает вероятность того, что медиана такой выборки меньше порогового значения.

Данная малая выборка сигнализирует: вероятность того, что среднее значение совокупности окажется ниже порога, заметно меньше

1%. Хотя эта статистика и кажется контринтуитивной*, но факт остается фактом: неопределенность положения медианы (или даже среднего значения) генеральной совокупности относительно порога можно снизить очень быстро. Предположим, что мы отобрали из генеральной совокупности всего четыре объекта и ни один из них не оказался ниже порогового. Обратившись снова к рисунку 9.4, мы обнаружим, что вероятность нахождения медианы ниже порога составляет чуть менее 4%, а вероятность ее положения выше него — соответственно 96%. То, что выборочная совокупность всего из четырех объектов настолько снижает неопределенность, может показаться удивительным, но несложные расчеты или моделирование методом Монте-Карло это подтверждают.

Обратите внимание, что неопределенность, связанная с порогом, может снижаться гораздо быстрее, чем неопределенность по поводу самого интересного для нас показателя. Бывает, после нескольких выборок остается довольно широкий диапазон, однако когда порог находится за его пределами, неопределенность, связанная с ним, снижается буквально до нуля.

Серьезным ограничением данного подхода является предположение о максимальной неопределенности порогового значения. Метод исходит из допущения, что изначально у нас нет никакой информации о том, по какую сторону порога может находиться медиана генеральной совокупности. Это означает, что придется начинать с 50-процентной вероятности того, что медиана находится по ту или иную сторону порога. Знай мы заранее, что медиана, скорее всего, ниже порога, наша диаграмма, хотя и неточная, все же дала бы полезный результат. Если же вероятность того, что значение медианы ниже порогового будет меньше вероятности ее положения выше порога, то диаграмма завысит вероятность того, что в действительности медиана ниже порога. В нашем примере диапазон 3–15% указывает на то, что медиана, скорее всего, больше порогового значения 7%. Диаграмма указывает, что вероятность обратного — 0,6%, но, обладая информацией о нашем диапазоне, мы можем сказать, что эта вероятность даже меньше.

Если бы, однако, диапазон составлял 1–8%, то с самого начала было бы понятно, что интересующий нас показатель расположен ниже порогового значения 7%. В данном случае рисунок 9.4 занижает вероятность того, что этот показатель ниже порога. Попытаемся использовать другой ориентир для уточнения искомой величины. Найдем фактическое среднее значение исходного диапазона и рассчитаем вероятность его нахождения по ту или иную сторону порога. При данном диапазоне можно утверждать, что существует 50-процентная вероятность того, что это значение меньше 4,5%. Допустим, что

* Нестереотипной. — *Примеч. редактора.*

из 10 отобранных нами служащих ни у одного затраты времени не оказались меньше 4,5%. Рисунок 9.4 свидетельствует: в этой ситуации вероятность того, что истинное значение на самом деле меньше 4,5%, составляет менее 0,1%. Хотя эта информация и не дает точных данных о том, насколько маловероятно, что искомое значение меньше 7%, все же становится очевидно, что этот вариант практически невозможен.

Итак, как правило, если выборочное обследование убедительно подтверждает имевшуюся ранее информацию (например, лишь одно из 10 значений выборки оказывается ниже порога, а вы уже знаете, что медиана вряд ли может быть ниже порога), то неопределенность снижается даже быстрее. Когда же исследование опровергает имевшиеся ранее сведения, для аналогичного снижения неопределенности придется увеличить размер выборки. Не забывайте и о том, что рисунок 9.4 позволяет оценить вероятность того, что медиана (но не среднее значение диапазона) ниже или выше порога. Конечно, вы можете проделать дополнительные расчеты и еще больше снизить неопределенность. Если четыре значения в выборке окажутся намного больше порогового, то это даст вам большую уверенность, чем если бы они едва превысили его.

ЭКСПЕРИМЕНТ

Мой первый опыт покупки по Интернету относится примерно к середине 1990-х годов. В моей библиотеке было несколько работ по эмпирическим методам, применяемым в различных областях, но мне нужна была книга по общим основам научных измерений — такая, которую можно порекомендовать своим клиентам — менеджерам компаний. Я прочитал всех философов (Куна, Поппера и др.), но не нашел того, что искал. И тогда я увидел на сайте www.amazon.com книжку под названием «How to Think like a Scientist» («Мыслить, как ученый»)². Отзывы были прекрасные, и казалось, что именно ее я смогу порекомендовать типичному руководителю. Оплатив покупку, через пару недель я получил по почте совсем не то, что ожидал. Книжка оказалась детской — для читателей от восьми лет. Я понял, что попал в глупое положение, и зарекся делать покупки по Интернету, пока электронная торговля не вышла из зачаточного состояния. Ведь в книжном магазине я, не имея в то время детей, просто проигнорировал бы секцию детской литературы. А заметь я эту книгу на книжном развале, ее обложка³ подсказала бы, что это вовсе не то серьезное издание из серии «Наука для бизнеса», которое мне нужно. Повозмущавшись, я начал просматривать книгу. И оказалось, что хотя она на две трети состояла из картинок, в ней освещались все основные понятия и каждому давалось как нельзя более простое объяснение. Например, авторы очень доступно говорили о том, что такое проверка гипотезы и проведение

наблюдений. Я изменил свое мнение, теперь покупка книги уже не казалась ошибкой. Я понял, что нашел в Интернете эту жемчужину именно потому, что принял ее за издание для взрослых. Стало ясно: главная идея книги была сформулирована на обложке — научные методы годятся для всех, кому уже исполнилось восемь.

Мысль провести эксперимент, чтобы определить какой-либо важный бизнес-показатель, к сожалению, приходит в голову менеджерам нечасто. А ведь эксперименты (что мы видели на примере Эмили Роза) могут быть очень простыми. При правильном подходе даже горсть конфетти, как показал Энрико Ферми, позволяет выяснить такой параметр, как мощность атомной бомбы. Сама идея очень проста. Как говорится в главе 3 о выборе инструментов измерения, когда вам нужно узнать то, что еще никем не измерено и что нельзя наблюдать без явного вмешательства, постарайтесь создать условия для наблюдения с помощью эксперимента.

Используемое в широком смысле слово «эксперимент» означает любое явление, созданное специально для цели наблюдения. Вы «экспериментируете», когда проверяете систему безопасности, чтобы понять, как быстро она отреагирует на угрозу. Но главной особенностью *управляемого* эксперимента является то, что у вас обычно не один, а два объекта изучения. Вы наблюдаете за тем, что испытываете (группа испытуемых), и за тем, с чем результаты испытания можно будет сравнить потом (контрольная группа). Это идеальный подход к ситуации, когда слишком сложно проследить за существующим явлением или когда объект, требующий измерения, еще не возник (например, необходимость оценки эффекта от изменения композиции продукта или внедрения новой информационной технологии).

Можно запустить новый опытный продукт или новую технологию и без оценки последствий. Но как узнать, понравилась ли потребителям ваша новая продукция, действительно ли повысилась производительность труда? Доходы могут вырасти вовсе не из-за изменения композиции продукта, а производительность труда — измениться совсем по другим причинам. На самом деле, если положение компании в какой-то момент зависит только от одного фактора, то отпадает необходимость в контрольных группах. Тогда можно менять этот фактор, исследовать положение компании и приписать возникшие сдвиги исключительно этой причине. Но, конечно, нужно уметь оценивать одновременное влияние разных сил и тогда, когда на сложные системы влияют многочисленные факторы, в том числе и такие, которые даже нельзя выявить и определить.

Изменяя какое-либо свойство своего продукта и желая понять, как это изменение повлияет на мнение потребителей, мы, возможно, должны будем провести эксперимент. Удовлетворенность потребителей, а следовательно, и число повторных заказов зависят от целого

ряда причин. Но когда необходимо узнать, оправдано ли экономически это новое свойство продукта, приходится оценивать роль только этого фактора, а не совокупности разных причин. Сравнивая мнения потребителей, купивших усовершенствованный продукт, и тех, кто остался верен старому варианту, мы сможем количественно определить эффект именно нового свойства.

Для интерпретации результатов экспериментов используются в основном уже обсуждавшиеся нами методы: проведение разных выборочных обследований, возможно, испытаний вслепую и т. д. Но особенно важно уметь рассчитывать разницу результатов, полученных в группе испытуемых и контрольной группе. Если мы уверены, что группа испытуемых действительно отличается от контрольной, то должны быть убеждены и в том, что разница в результатах возникает не случайно. На самом деле, сравнение откликов двух групп очень похоже на расчет среднего квадратичного отклонения оценки, с которым мы уже знакомы, но с одним небольшим изменением. В данном случае среднее квадратичное отклонение, которое надо определить, — это отклонение различия результатов групп. Рассмотрим следующий пример.

Предположим, что одна компания решила оценить воздействие тренинга «Управление взаимоотношениями с клиентами» на качество работы службы поддержки потребителей. Сотрудники этой службы обычно отвечают на звонки покупателей, у которых появились вопросы или возникли проблемы с новым продуктом. Есть мнение, что плохая работа данной службы приводит не столько к снижению объемов продаж именно этим покупателям, сколько к той устной антирекламе, которую компания получает в результате. Как обычно, эта компания, прежде всего, оценила существующую неопределенность эффекта от проведения тренинга, затем определила соответствующий порог и рассчитала стоимость информации.

Изучив несколько возможных инструментов измерения, менеджеры решили, что лучше всего провести опрос позвонивших в службу поддержки покупателей и поинтересоваться не только их впечатлениями, но и вероятностью того, что они порекомендуют компанию своим друзьям. Используя ранее собранную маркетинговую информацию, калиброванные менеджеры определили, что новый тренинг в области управления взаимоотношениями с клиентами может увеличить объем продаж на 0–12%, но расходы на его проведение окупятся даже в случае, если рост продаж составит лишь 2%.

К проведению опроса приступили еще до начала тренинга, чтобы представить себе исходное положение дел. Из клиентов каждого сотрудника службы отбирали только одного. Главный вопрос был сформулирован так: «Какова вероятность, что вы порекомендуете нас своим друзьям, учитывая свой опыт обращения в службу поддержки?» В случае высокой вероятности такого поступка респонденты долж-

ны были выбрать цифру 1, если мнение о работе службы не изменилось — цифру 2, а в случае малой вероятности — 3. Каждому ответу присваивался соответствующий балл (1, 2, 3). Зная результаты некоторых предыдущих исследований роста продаж в результате повышения удовлетворенности потребителей, отдел маркетинга определил, что улучшение среднего балла ответов на этот вопрос на 0,1 пункта приведет к повышению объема реализации на 2%.

Тренинг, о котором идет речь, был не из дешевых, поэтому сначала руководство решило послать на обучение 30 случайным образом выбранных сотрудников службы поддержки потребителей в качестве группы испытуемых. Тем не менее затраты на обучение этой группы оказались существенно меньше рассчитанной стоимости информации. В контрольную группу вошли все остальные сотрудники службы. После того, как тест-группа прошла программу тренинга, менеджеры продолжили опрос покупателей, осуществив выборочное обследование по принципу, описанному выше. Были рассчитаны среднее значение и дисперсия исходных данных по группе испытуемых и контрольной группе (как это показано в примере с леденцами в начале главы). Результаты расчетов представлены в таблице 9.4.

Ответы респондентов вроде бы продемонстрировали, что обучение дало эффект; но не является ли эта картина просто случайностью? Может быть, 30 случайным образом выбранных сотрудников уже работали лучше, чем персонал в среднем, или этим 30 случайно пришлось иметь дело с менее «трудными» покупателями? Пересчитаем полученные результаты обеих групп следующим образом:

1. Разделим дисперсию выборки каждой группы на размер выборки. Получим: $0,392 / 30 = 0,0131$ для группы испытуемых и $0,682 / 85 = 0,008$ для контрольной группы.
2. Сложим полученные для каждой группы результаты: $0,0131 + 0,008 = 0,021$.

ТАБЛИЦА 9.4.

Эксперимент по оценке эффективности обучения сотрудников группы поддержки клиентов

Сотрудники	Размер выборки	Среднее значение	Дисперсия
Тест-группа (участники тренинга)	30	2,433	0,392
Контрольная группа (не участвовала в тренинге)	85	2,094	0,682
Исходное состояние (до тренинга)	115	2,087	0,659

3. Извлечем из полученного числа квадратный корень. Так мы получим среднее квадратичное отклонение разницы между группами. В данном случае она составит 0,15.
4. Рассчитаем разницу между средними значениями двух сравниваемых групп: $2,433 - 2,094 = 0,339$.
5. Рассчитаем теперь вероятность того, что разница между группой испытуемых и контрольной группой больше 0, то есть группа испытуемых показала действительно, а не случайно, лучшие результаты, чем контрольная группа. Используем для этого расчета формулу normdist в Excel:

$$=normdist(0, 0,339, 0,15, 1)$$

и получим вероятность, равную 0,01.

Таким образом, существует вероятность всего 1% того, что сравниваемые группы одинаково хороши или плохи. Значит, мы можем быть на 99% уверены, что люди, прошедшие тренинг, действительно работают лучше остальных сотрудников службы.

Аналогично можно сравнить контрольную группу с исходным состоянием. Разница между ними составляет всего 0,007. Применяв метод, только что использованный нами для сравнения тест-группы и контрольной группы, найдем, что есть 48-процентная вероятность того, что контрольная группа работает хуже исходного состояния, или 52-процентная вероятность того, что лучше. Таким образом, различие между этими группами пренебрежимо мало, а для всех практических целей его вообще не существует.

Мы определили с высокой степенью уверенности, что тренинг способствует повышению удовлетворенности потребителей. Поскольку разница между группой испытуемых и контрольной группой составляет около 0,4, отдел маркетинга пришел к выводу: обучение приведет к росту продаж примерно на 8%. Это означает, что затраты на обучение всего персонала экономически целесообразны. Вспомним, что мы вполне могли бы взять и меньшую выборку, воспользовавшись t-распределением Стьюдента для выборок размером до 30.

ВЫЯВЛЕНИЕ ВЗАИМОЗАВИСИМОСТИ ПАРАМЕТРОВ: ВВЕДЕНИЕ В РЕГРЕССИОННОЕ МОДЕЛИРОВАНИЕ

На семинарах мне часто задают примерно такой вопрос: «Если благодаря внедрению новой информационной системы продажи увеличатся, то откуда мы будем знать, что это произошло благодаря именно этой системе?» Просто удивительно, как часто этот вопрос возникает при

том, что последние несколько столетий специалисты по научным измерениям только и делают, что пытаются выделить эффект одной переменной. Могут лишь предположить, что эти люди незнакомы с основными понятиями научного измерения. Приведенный ранее в этой главе пример эксперимента ясно показывает: то, что объясняется многими причинами, вполне можно проследить до одного-единственного фактора влияния, сравнивая тест-группу с контрольной группой. На самом деле использование этих групп — лишь один из имеющихся способов выделения эффекта одной переменной из всей массы информации, существующей в любой компании. Другой способ — рассмотреть, насколько одна переменная коррелирует с другой.

Корреляция между двумя наборами данных обычно характеризуется числом, варьирующим от $+1$ до -1 . Корреляция, равная $+1$, означает, что две переменные меняются совершенно одинаково: с ростом одной увеличивается и другая. Корреляция -1 также указывает на тесную связь между переменными, но когда одна из них увеличивается, другая уменьшается. Нулевая корреляция означает отсутствие всякой связи.

Чтобы понять, как выглядят коррелирующие друг с другом показатели, рассмотрим четыре примера, приведенных на рисунке 9.5. По горизонтальной оси отложены баллы теста на профессиональную пригодность или объем телевизионной рекламы за месяц, а по вертикаль-

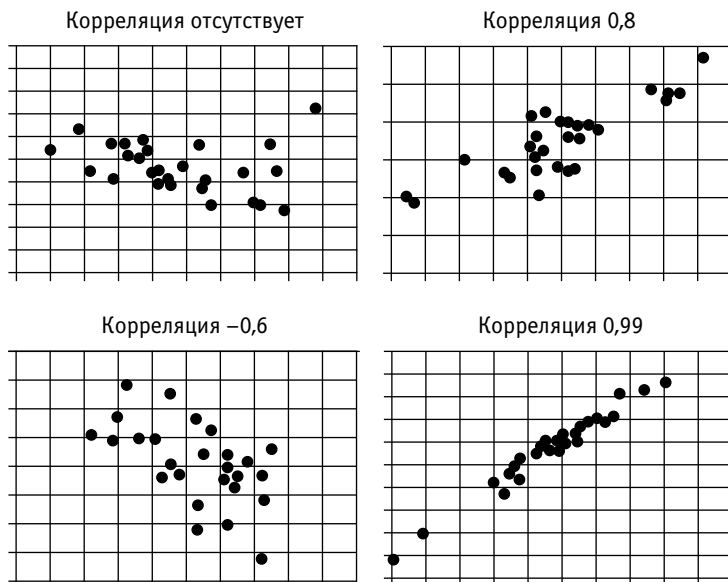


РИСУНОК 9.5.

Примеры корреляции данных

ной — производительность труда или объемы продаж за месяц. Иными словами, на осях могут быть отложены любые критерии. Но из рисунка видно, что на одних диаграммах между параметрами существует более тесная связь, чем на других. В верхнем левом углу — график изменения двух переменных, никак не связанных друг с другом, между которыми нет корреляции, что видно из отсутствия наклона. Плоской и вытянутой в длину совокупность точек кажется потому, что изменчивость переменной по горизонтали больше таковой по вертикали. Будь их изменчивость одинаковой, точки разместились бы по кругу, но наклона все равно мы не наблюдали бы. График в нижнем правом углу указывает на сильную взаимозависимость двух переменных.

Прежде чем приступить к расчетам, необходимо построить график, чтобы убедиться в наличии или отсутствии заметной корреляции. Если при сравнении ожидаемых затрат на реализацию проекта с фактическими издержками ваш график оказался аналогичен изображенному в нижнем правом углу рисунка 9.5, значит, оценка была необыкновенно точной. А если он похож на набор точек в верхнем левом углу, то можно сказать, что игрок в кости оценил бы эти затраты не хуже.

Использование регрессионных моделей и данных за прошедшие периоды освобождает нас от необходимости проведения управляемого эксперимента. Допустим, что нам трудно увязать реализацию проекта в области информационных технологий и рост продаж. Однако имеется много данных о том, как влияет на продажи *что-то еще*, например более быстрый выход нового продукта на рынок. Если знать, что автоматизация некоторых процессов позволит сократить время от разработки до выхода продукта на рынок, то можно найти интересующую нас зависимость.

Однажды по просьбе одной крупной компании кабельного телевидения я анализировал ее проект инвестирования в новое программное обеспечение. Компания собиралась автоматизировать выполнение ряда административных задач, связанных с выпуском новых телешоу. Она надеялась, что это, в частности, приведет к повышению рейтингов подобных передач, а следовательно, и к росту доходов от рекламы. Но как компания могла оценить влияние проекта на рейтинги, зависящие от столь многих факторов?

По идее система автоматизации производства должна была ускорить выполнение определенных административных задач. Если они станут выполняться быстрее, компания сможет раньше начинать раскручивать новые шоу. У телекомпании были данные о рейтингах в прошлые периоды, и изучив старые производственные графики, мы смогли определить, сколько недель каждое новое шоу рекламировалось до выхода в эфир (ранее мы рассчитали стоимость этой информации и определили, что затраты на ее оценку полностью себя оправдают). Рисунок 9.6 представляет собой график роста рейтинга с увеличением

продолжительности раскрутки (в неделях). Он построен не на основе реальных данных моего клиента, но иллюстрирует примерно ту же корреляцию.

Можем ли мы сразу, еще до анализа имеющихся данных, сказать, что между двумя переменными существует корреляция? Если да, то на какой график рисунка 9.5 больше всего будет похожа эта корреляция? Построение такой зависимости — всегда мой первый шаг при регрессионном анализе, поскольку чаще всего корреляция (если она существует) бывает очевидной. В Excel это очень просто сделать: заполните два столбца числами (в данном случае — недели раскрутки шоу и пункты рейтинга), каждая пара данных соответствует одному ТВ-шоу. Просто выберите весь набор данных, щелкните в меню программы Excel по кнопке «график», выберите опцию «XY (Scatter)» («график разброса»), выполните остальные указания, и вы увидите такой же график, как на рисунке 9.6.

Похоже, что корреляция имеется, но насколько она тесная? Ответ на этот вопрос требует знания некоторых тонкостей. Но я не стану здесь объяснять, на чем основано регрессивное моделирование, а взамен просто расскажу, что нужно сделать в Excel.

В программе Excel для расчета корреляции можно просто использовать функцию «=correl()». Предположим, что данные о продолжительности рекламы и рейтингах содержатся соответственно в первых 28 строках столбцов А и В электронной таблицы. Вы написали бы: =correl(A1:A28, B1:B28). С нашими данными мы получим коэффициент корреляции примерно 0,7. Поэтому можно с уверенностью считать, что увеличение продолжительности рекламы нового шоу повы-

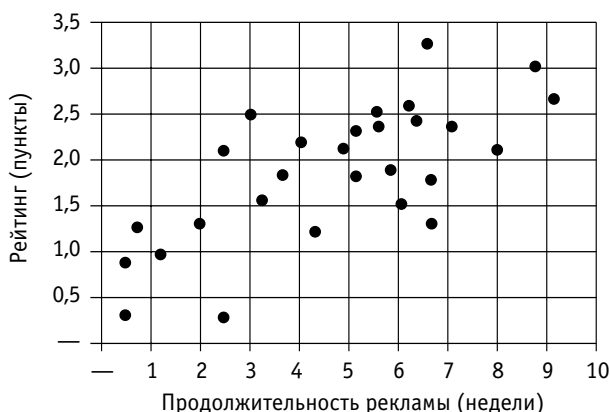


РИСУНОК 9.6.

**График зависимости
«рейтинг—продолжительность рекламы»**

сит его рейтинг. Теперь пора сосредоточиться на упрощении процесса производства шоу и увеличении времени, которое можно потратить на рекламу.

Другой способ проделать это в Excel — воспользоваться опцией «Regression Wizard» («Мастер регрессии») из «Data Analysis Toolpack» («Набор инструментов для анализа данных») (перемещение к этой опции в разных версиях Excel разное, поэтому воспользуйтесь опцией «Справка»). Мастер регрессии подскажет вам выбрать «Y range» («интервал Y») и «X range» («интервал X»). В нашем примере это, соответственно, рейтинг и реклама (в неделях). Мастер создаст таблицу с результатами регрессионного анализа. Некоторые из них поясняются в таблице 9.5.

На основе приведенной информации можно вывести формулу наилучшего приближения для связи между продолжительностью раскрутки и рейтингом шоу. Далее мы рассчитаем прогнозируемый рейтинг по числу недель рекламы. Удобно называть это значение (в данном случае прогнозируемый рейтинг) «зависимой» переменной, а величину, по которой она определяется, — «независимой».

$$\begin{aligned} & \text{Прогнозируемый рейтинг (пункты)} = \\ & = \text{Переменная } X_1 \times \text{Продолжительность раскрутки (недели)} + \\ & + \text{Отрезок, отсекаемый на оси координат.} \end{aligned}$$

Если мы построим кривую, которую дает нам эта простая формула, на уже построенном нами графике, то он приобретет вид, представленный на рисунке 9.7.

ТАБЛИЦА 9.5.

Выделенные объекты из набора инструментов «Summary Output» («Сводка результатов») программы Excel

Переменная	Смысл переменной
Коэффициент множественной корреляции (multiple R)	Корреляция одной или более переменных с «зависимой» переменной (например, с рейтингом): в данном примере R=0,7
R ²	Коэффициент R в квадрате. Его можно интерпретировать как величину дисперсии пунктов рейтинга, объясняемую разной продолжительностью рекламы
Отрезок, отсекаемый на оси координат	Значение рейтинга (в пунктах) при сокращении времени раскрутки шоу до нуля. Точка, в которой линия наилучшего соответствия пересекает вертикальную ось координат
Переменная X 1	Коэффициент (вес) для времени рекламы шоу

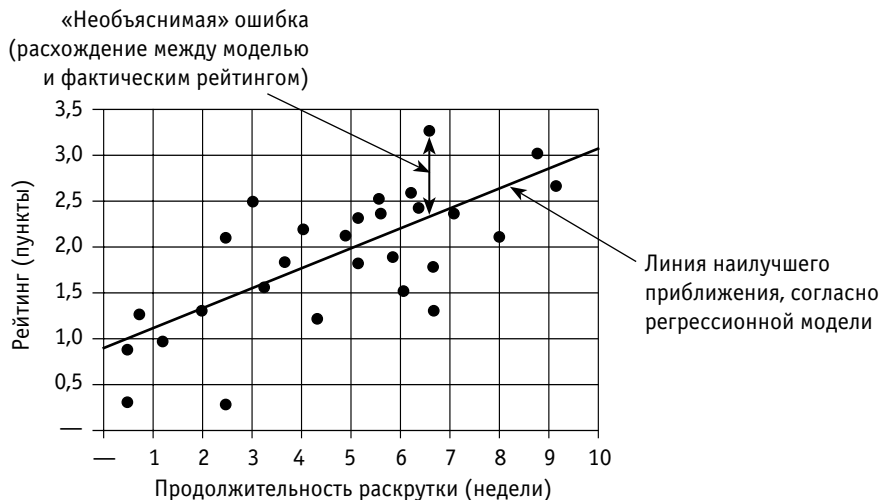


РИСУНОК 9.7. График зависимости рейтинга от продолжительности раскрутки телешоу и линия наилучшего приближения

Согласно рисунку 9.7, хотя корреляция и имеется, рейтинг зависит не только от продолжительности рекламы. Эта информация вместе с итогами управляемого эксперимента позволяет нам ответить на пресловутый вопрос: «Откуда мы это знаем, если есть и другие факторы?» Ясно, что длительность раскрутки имеет значение для рейтинга, и неважно, определены ли количественно эффекты от действия других факторов и даже *выявлены ли они вообще*.

Преимущество инструмента «Regression» («Регрессия») программы Excel над такими более простыми функциями, как `=correl()`, заключается в том, что он позволяет выполнять так называемую множественную регрессию. Таким образом, возникает возможность одновременно рассчитывать коэффициенты для нескольких независимых переменных. При желании мы могли бы создать модель, связывающую рейтинг не только с продолжительностью рекламы, но и со временем года, категорией шоу, откликами фокус-группы и несколькими другими факторами. В таблице 9.5 каждая из этих дополнительных переменных имела бы свой коэффициент «Переменная X 2», «Переменная X 3» и т.д. В итоге мы получили бы следующую формулу:

$$\begin{aligned}
 & \text{Прогнозируемый рейтинг (пункты)} = \\
 & = \text{Переменная X 1} \times \text{Продолжительность раскрутки (недели)} + \\
 & \quad + \text{Переменная X 2} \times \text{Результаты фокус-группы} + \dots + \\
 & \quad + \text{Отрезок, отсекаемый на оси координат.}
 \end{aligned}$$

Сказав все это, необходимо также сделать ряд предостережений. Во-первых, корреляция не означает причину. То, что одна переменная коррелирует с другой, не обязательно означает, что одна из них обуславливает другую. Церковные пожертвования и продажа алкоголя между собой коррелируют, но вовсе не потому, что между производителями алкоголя и духовенством есть какой-то сговор, а потому, что и то и другое зависит от состояния экономики. Как правило, чтобы утверждать, что между какими-то явлениями существует причинно-следственная связь, помимо наличия корреляции нужны дополнительные основания. В случае соотношения рейтинга и длительности рекламы такие основания у нас действительно есть.

Во-вторых, не забывайте о том, что это простые линейные регрессии. Взяв не саму переменную, а какую-нибудь ее функцию (например, ее квадрат, инверсию, произведение двух переменных и т.д.), можно рассчитать корреляцию еще точнее. При желании читатели могут с этим поэкспериментировать. Наконец, в моделях множественной регрессии вы должны следить за тем, чтобы независимые переменные не коррелировали друг с другом. В идеале между независимыми переменными не должно быть никакой связи.

Я лишь коснулся основ множественного регрессионного моделирования. Этот инструмент очень полезен, но пользоваться им необходимо с осторожностью.

■ ПРИМЕЧАНИЯ

- 1 *Goodman*. Serial Number Analysis // Journal of the American Statistical Association, 1952, 47, p.622–634.
- 2 *Kramer Stephen P*. How to Think Like a Scientist. New York: HarperCollins, 1987.
- 3 В то время увидеть обложку книги на этом сайте еще было нельзя.



Кое-что о Байесе

ПРОСТАЯ БАЙЕСОВСКАЯ СТАТИСТИКА

Курс экономической статистики в первом семестре знакомит студентов с рядом методов, основанных на нескольких «делаемых для простоты» допущениях, хотя нередко эти допущения мало что упрощают. А в дальнейшем студенты узнают о таких более «продвинутых» методах, которые лично мне всегда казались более интуитивными, чем составляющие содержание предыдущих разделов.

Главное допущение, которое делается в большинстве вводных курсов статистики, заключается в следующем: единственное, что вы знаете о некоей генеральной совокупности, — это образцы, которые вы собираетесь из нее выбрать. Но на практике это допущение почти всегда неверно.

Предположим, что вы отбираете несколько торговых представителей для опроса на тему, существует ли связь между произошедшим недавно ростом продаж и проведенной ранее рекламной кампанией. Вы хотите оценить «вклад рекламной кампании в объем продаж». Для этого можно просто опросить весь торговый персонал. Но ведь изначально вы знаете больше, чем то, что расскажут вам эти люди. Вам и до опроса было кое-что известно о прошлой динамике продаж и об эффекте, который давали рекламные компании. Вы имеете сведения о сезонных колебаниях объема сбыта, влиянии экономического цикла и роли мер по повышению доверия потребителей. Имеет ли это какое-либо значение? Интуитивно мы понимаем, что предварительные

данные также должны учитываться. Но пока студенты не доберутся до последних разделов своего учебника, им так и не расскажут, что нужно делать с этим знанием.

ПАРАДОКС ПРЕДВАРИТЕЛЬНОГО ЗНАНИЯ

1. Вся традиционная статистика исходит из того, что наблюдатель ранее не располагал никакой информацией об объекте наблюдения.
2. В реальном мире данное допущение почти никогда не выполняется.

Проблему прежних знаний изучает так называемая байесовская статистика. Автор этого метода — Томас Байес, британский математик и пресвитерианский священник XVIII века, самые известные работы по статистике которого были опубликованы только после его смерти. Байесовская статистика занимается вопросом: как мы корректируем свое предварительное знание с учетом новой информации? Байесовский анализ начинается с того, что известно сейчас, и затем рассматривает, как это знание изменится с получением новых сведений. А небайесовская статистика, преподаваемая в большинстве курсов по методам выборочного наблюдения, исходит из следующего: все, что известно о некоей группе объектов, — это выборка, которую вы только что из нее сделали.

Фактически, именно байесовский анализ лежит в основе большинства иллюстраций, приведенных мною в главе 9, в том числе таблицы для определения 90-процентного CI без математических расчетов. Например, составляя рисунок 9.2, изображающий 90-процентный CI для доли в генеральной совокупности при малой выборке, я сначала предположил, что, если нет иных данных, внутри подгруппы значения этого CI распределены от 0 до 100% равномерно. Рассчитывая вероятность нахождения медианы по ту или иную сторону порогового значения (см. рис. 9.4), я начинал с исходного соображения: существует вероятность 50%, что истинная медиана генеральной совокупности лежит по одну из сторон от порога. И в том, и в другом случаях я исходил из максимальной неопределенности.

Теорема Байеса гласит, что вероятность наступления «события» при условии проведения «наблюдения» равна произведению вероятности наступления события и вероятности проведения наблюдения при условии наступления события, деленному на безусловную вероятность проведения наблюдения (см. рис. 10.1).

Предположим, решается вопрос о выпуске нового продукта. Согласно данным за прошедшие периоды, новые продукты приносили прибыль в первый год только в 30% случаев. Математик записал бы

$$P(A|B) = P(A) \times P(B|A)/P(B),$$

где

$P(A|B)$ — вероятность А при условии В;

$P(A)$ — «безусловная» вероятность А;

$P(B)$ — «безусловная» вероятность В;

$P(B|A)$ — вероятность В при условии А.

РИСУНОК 10.1. Теорема Байеса

это утверждение следующим образом: $P(FYP^*) = 30\%$, то есть вероятность получения прибыли в первый год выпуска продукта составляет 30%. Нередко до старта массового производства осуществляется тестирование сбыта. Для всех случаев, когда новый продукт дал прибыль уже в первый год реализации, пробные продажи были удачными (под удачными продажами мы подразумеваем достижение определенного порогового объема реализации) только на 80%. Математик записал бы это следующим образом: $P(S|FYP) = 80\%$, то есть «условная» вероятность успеха тестирования сбыта (S, successful — успешный) при условии, что производство продукта оказалось прибыльным уже в первый год (черта «|» означает «при условии»), равна 80%.

Однако значение вероятности успеха пробных продаж при условии, что производство данного продукта принесло прибыль в первый же год, — вовсе не тот показатель, который нас интересует. Что мы в действительности хотели бы знать, так это вероятность получения прибыли в первый же год при условии, что тестирование сбыта окажется удачным. Таким способом рынок подскажет нам, стоит ли запускать серийное производство продукта. Данный вопрос и позволяет выявить теорема Байеса. Перепишем уравнение теоремы Байеса, подставив в него следующие обозначения интересующих нас функций:

- $P(FYP|S)$ — вероятность получения прибыли в первый же год при условии удачного тестирования сбыта, иными словами, вероятность наступления события FYP при условии S;
- $P(FYP)$ — «безусловная» вероятность получения прибыли в первый же год;
- $P(S)$ — «безусловная» вероятность удачного тестирования сбыта;
- $P(S|FYP)$ — вероятность удачного тестирования сбыта при условии получения прибыли в первый же год.

* FYP — first year profit, прибыль первого года. — *Примеч. переводчика.*

Допустим, что пробные продажи были удачны в 40% случаев. Чтобы рассчитать вероятность получения прибыли от продукта в первый же год при условии удачного тестирования сбыта, перепишем приведенное выше уравнение следующим образом:

$$P(\text{FYP}|\text{S}) = P(\text{FYP}) \times P(\text{S}|\text{FYP}) / P(\text{S}) = 30\% \times 80\% / 40\% = 60\%.$$

Если пробный рынок показал успех, то вероятность получения прибыли в первый же год составляет 60%. Заменяя два числа в уравнении, мы можем рассчитать и вероятность получения прибыли в первый же год в случае неудачных пробных продаж. Как было показано, вероятность успеха тестирования реализации прибыльного продукта 80%. Поэтому вероятность провала пробных продаж равна 20%. Это можно записать следующим образом: $P(\sim\text{S}|\text{FYP}) = 20\%$. Аналогично, если вероятность удачных пробных продаж всех продуктов 40%, то общая вероятность неудачи, или $P(\sim\text{S})$, равна 60%. Заменяя в нашем уравнении $P(\text{S}|\text{FYP})$ и $P(\text{S})$ на $P(\sim\text{S}|\text{FYP})$ и $P(\sim\text{S})$, получим:

$$P(\text{FYP}|\sim\text{S}) = P(\text{FYP}) \times P(\sim\text{S}|\text{FYP}) / P(\sim\text{S}) = 30\% \times 20\% / 60\% = 10\%.$$

Таким образом, провальный результат тестирования рынка дает вероятность получения прибыли в первый же год в размере всего 10%.

Иногда, не зная вероятности получения некоего результата, мы можем оценить вероятности других событий и затем рассчитать на их основе нужный показатель. Предположим, что данных о коэффициенте успеха тестирования рынка в прошедшие периоды у нас нет, так как это первые пробные продажи. Мы можем рассчитать данную величину на основе других. Калиброванный эксперт уже оценил $P(\text{S}|\text{FYP})$ — вероятность успешных пробных продаж продукта, который принесет прибыль в первый же год: $P(\text{S}|\text{FYP}) = 80\%$. Допустим теперь, что эксперт оценил и вероятность удачных пробных продаж продукта, выпуск которого окажется в итоге убыточным (классическим примером может служить «New Coke»): $P(\text{S}|\sim\text{FYP}) = 23\%$. Как и ранее, мы знаем, что вероятность прибыли от продукта в первый же год $P(\text{FYP})$ составляет 30%, значит, вероятность того, что это не случится, $P(\sim\text{FYP})$ будет равна 70% — $[1 - P(\text{FYP})]$. Если мы суммируем произведения каждой условной вероятности на вероятность выполнения данного условия, то получим общую вероятность наступления данного события. Тогда:

$$\begin{aligned} P(\text{S}) &= P(\text{S}|\text{FYP}) \times P(\text{FYP}) + P(\text{S}|\sim\text{FYP}) \times P(\sim\text{FYP}) = \\ &= 80\% \times 30\% + 23\% \times 70\% = 40\%. \end{aligned}$$

Этот этап может оказаться очень полезным, потому что в некоторых случаях расчет вероятности получения определенных результатов при определенных условиях прост и очевиден. Составить большин-

ство иллюстраций из приведенных в главе 9 мне помогли такие вопросы, как: «Если к данной группе действительно относятся только 10% всех объектов генеральной совокупности, то какова вероятность того, что из 12 случайно выбранных человек пятеро будут принадлежать к этой группе?» или: «Если медиана затрат времени на анализ жалоб потребителей составляет более часа, то какова вероятность того, что временные затраты 10 из 20 случайно выбранных человек окажутся менее часа?»

В каждом из этих примеров мы можем рассчитать вероятность наступления события А при условии наступления события В, если знаем эти вероятности и вероятность наступления события В при условии наступления события А. Данный математический прием называется байесовской инверсией, и те, кто начинает использовать его в одной области, быстро обнаруживают применимость инверсии и во многих других сферах. Особенно полезной байесовскую инверсию находят те, кто рассматривает проблемы измерения так же, как в свое время это делали Эмили, Энрико и Эратосфен. Более специальные вопросы, связанные с инверсией, мы рассмотрим позднее, а пока попытаемся объяснить ее на интуитивном уровне. Ведь, возможно, и вы, сами того не осознавая, уже применяли этот прием. Вполне вероятно, что вы обладаете врожденным байесовским инстинктом.

ИСПОЛЬЗУЙТЕ СВОЙ ПРИРОДНЫЙ БАЙЕСОВСКИЙ ИНСТИНКТ

Проблему иного качественного знания о выборочной совокупности, которым вы обладаете, не решают даже некоторые передовые методы статистики. В описанном ранее примере с рекламной кампанией вы могли бы проработать с людьми отдела сбыта весьма продолжительное время и узнать (и это знание было бы качественным), что Боб обычно оценивает ситуацию слишком оптимистично, Мануэль всегда все взвешивает, а Моника любит осторожничать. И, конечно, вы по-разному отнеслись бы к мнениям того сотрудника, которого знаете очень хорошо, и новичка. Как статистика учитывает эти знания? Если отвечать односложно, то она их вообще не учитывает, во всяком случае, тот ее вводный курс, который изучают тысячи людей.

К счастью, существует способ справиться с этой проблемой, причем намного более простой, чем любой раздел статистики за первый семестр. Назовем его инстинктивным байесовским подходом, суть которого заключается в следующем:

- 1) сначала нужно дать объекту (явлению) свою калиброванную оценку;

- 2) затем необходимо собрать дополнительную информацию (провести опрос, изучить работы других исследователей и т. д.);
- 3) далее нужно чисто субъективно скорректировать свою калиброванную оценку без дополнительных расчетов.

Я называю это инстинктивным байесовским подходом, так как есть основания считать, что когда люди получают новую информацию и уточняют свои прежние знания, они делают это способом, который можно охарактеризовать как байесовский. В 1995 г. психологи-бихевиористы Калифорнийского технологического института Махмуд А. Эль-Гамаль и Дэвид М. Гретер изучали, как люди учитывают первоначальные знания и новые сведения, оценивая вероятность каких-либо событий¹. Они попросили группу из 257 студентов угадать, из какого из двух лотерейных барабанов были извлечены шарики. В каждом барабане находились шарики, помеченные буквами «N» и «G». В одном барабане их было поровну, а в другом шариков с буквой «N» было больше. Шарики вынимались шесть раз, и студентам объявляли, сколько всего шариков каждого вида было вынуто.

Итак, задача состояла в том, чтобы определить, из какого барабана были взяты шарики. Студент, который видел, что в выборке из шести шариков, например, пять с буквой «N» и только один с буквой «G», мог решить, что они взяты из барабана с преобладанием шариков, помеченных буквой «N». Однако перед каждым извлечением шести шариков присутствующим говорили, что сами барабаны выбираются случайным образом с вероятностью один к двум, один к трем и два к трем. И вот ответы студентов показали, что они как будто интуитивно использовали байесовскую инверсию и при этом слегка переоценивали значение новой и недооценивали значение старой информации. Иными словами, они не были идеальными байесианцами, но все же, скорее, были ими.

Я также думаю, что будь на их месте калиброванные оценщики, они проявили бы байесианские качества лучше. Ведь студенты, принимавшие участие в исследовании, как и большинство обычных людей, были слишком уверены в своих ответах. А калиброванный специалист, не будучи слишком самоуверенным, все же обладал бы этим базовым байесовским инстинктом.

В нескольких построенных мною моделях использовались определенные калиброванными оценщиками условные вероятности самых разных событий. В 2006 г. я задал калиброванным экспертам из одной государственной структуры следующие пять вопросов.

- А. Какова вероятность того, что через четыре года президентом будет демократ?

- В. Какова вероятность того, что ваш бюджет через четыре года увеличится при условии, что президентом будет демократ?
- С. Какова вероятность того, что ваш бюджет через четыре года увеличится при условии, что президентом будет республиканец?
- Д. Какова вероятность того, что ваш бюджет через четыре года увеличится?
- Е. Если ваш бюджет через четыре года увеличится, то какова вероятность того, что это произойдет в период президентства демократа?

Отвечая на эти вопросы, инстинктивный байесианец руководствовался бы теоремой Байеса. Если бы первые три вероятности (А, В и С) он оценил как 55, 60 и 40%, то, чтобы быть последовательным, четвертую и пятую вероятности (D и E) он должен был бы определить, соответственно, в 51 и 64,7%. Ответ на четвертый вопрос следовало бы записать так: $A \times B + (1 - A) \times C$, строго говоря, не из-за теоремы Байеса, а из-за необходимости правильно сложить условные вероятности. Иными словами, вероятность наступления некоего события равна вероятности выполнения некоего условия, умноженной на вероятность наступления данного события в случае выполнения этого условия, плюс вероятность того, что это условие не будет выполнено, умноженная на вероятность наступления этого события в случае невыполнения этого условия. Поэтому байесианец ответил бы на вопросы А, В, D и E таким образом, чтобы $B = D / A \times E$.

Это не всегда очевидно, но, тем не менее, большинство калиброванных экспертов по принятию решений интуитивно дают ответы, удивительно близкие к удовлетворяющим этому требованию. Допустим, что в нашем последнем примере ответы калиброванного эксперта на вопросы А, В и С были 55, 70 и 40%. Но его ответы на вопросы D и E были 50 и 75%, хотя, по логике, при таких ответах на предыдущие вопросы они должны были быть 56,5 и 68,1%, а не 50 и 75%. На рисунке 10.2 мы показываем, как субъективные ответы на эти вопросы соотносятся с расчетными байесовскими значениями.

Обратите внимание: для того, чтобы согласовываться с другими субъективными ответами, одно из пары байесовских значений должно быть меньше нуля, а другое — больше 100%. Очевидно, что эти значения противоречили бы здравому смыслу, но когда калиброванные эксперты давали свои субъективные оценки, они не знали, что возникнет такая проблема. Однако в большинстве случаев полученные ответы оказались даже более близкими к «собственно байесовским», чем ожидали калиброванные эксперты (см. рис. 10.2).

На практике для того, чтобы сделать субъективные калиброванные оценки условных вероятностей внутренне непротиворечивыми,

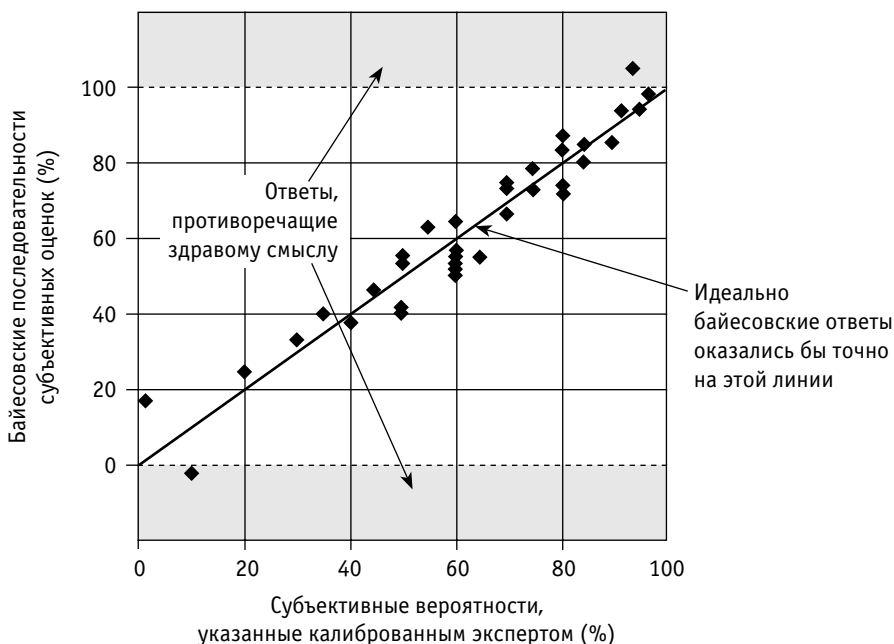


РИСУНОК 10.2.

Субъективные вероятности, указанные калиброванным экспертом, в сравнении с байесовскими вероятностями

я применяю специальный метод, который называю байесовской коррекцией. Я сообщаю калиброванным экспертам, какими могли быть байесовские ответы на некоторые вопросы с учетом их ответов на другие вопросы. Затем они меняют свои оценки до тех пор, пока все их субъективные калиброванные вероятности не окажутся, по крайней мере, совместимыми друг с другом.

Интересно, что, сравнивая новую информацию со старой, люди обычно мыслят максимально логично, что крайне важно, поскольку человек способен учитывать качественную информацию, не вписывающуюся в стандартную статистику. Например, при составлении прогноза возможного влияния новой политики на имидж компании в глазах общественности (что измеряется снижением числа жалоб потребителей, ростом доходов и т. п.) калиброванный эксперт должен был бы дополнить имевшиеся у него сведения «качественной» информацией о том, какую роль эта политика сыграла в других компаниях, результатами обсуждения в фокус-группах и т. д. Даже когда имеется

информация только о выборке, калиброванный специалист, обладающий байесовским инстинктом, принимает во внимание такую качественную информацию о выбранных объектах, которую не учитывают обычные статистические методы.

Убедитесь в этом сами, попытавшись ответить на следующий вопрос: увеличатся ли доходы вашей компании в следующем году? Укажите сначала свою калиброванную вероятность. Затем опросите двух-трех компетентных в этом вопросе сотрудников. Пусть они не только выскажут свое мнение, но и как-то его аргументируют, приведут какие-то детали. Теперь предложите другую субъективную вероятность того, что доходы компании увеличатся. Эта новая оценка наверняка будет отражать полученную вами новую информацию, даже если она и была по большей части качественной.

На рисунке 10.3 калиброванный эксперт (обладающий байесовским инстинктом и не страдающий как избытком, так и недостатком уверенности) сравнивается с тремя другими специалистами — использующим традиционные небайесовские методы выборки, такие как

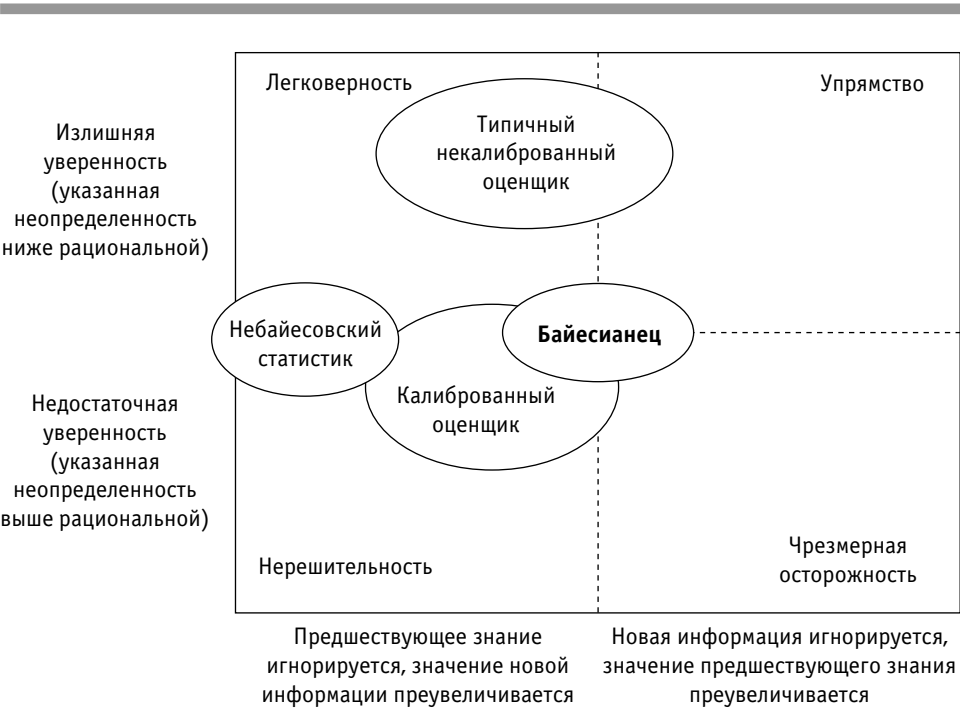


РИСУНОК 10.3. Уверенность или информация: разные акценты при проведении оценок

t-статистика, некалиброванным оценщиком и чистым байесовским оценщиком. Эта концептуальная схема показывает нам, как соотносятся с байесовским используемые ими подходы. Одна ось показывает, насколько специалисты уверены в своих оценках по сравнению с реальной вероятностью правильности их оценок, другая ось — насколько они учитывают предыдущую информацию.

Метод может вызвать опасения тех, кто считает себя сторонником «объективного» измерения, но это беспокойство беспочвенно. Во-первых, я уже показал, что субъективные оценки калиброванных экспертов обычно ближе к разумным величинам, чем к противоречащим здравому смыслу. Во-вторых, этот метод работает там, где «объективная» статистика из первого семестра оказывается бесполезной и единственная альтернатива — вообще ничего не оценивать. В-третьих, те же самые люди постоянно бессознательно пользуются данным методом, принимая личные решения. Например, читая статью о возможном снижении цен на жилье, они принимают в результате решение купить или продать дом вовсе не потому, что используют приведенные в ней данные в своей модели, а потому, что извлекают из нее определенную качественную информацию.

В то же время существуют способы устранения тех недостатков, которые действительно имеются у описываемого метода. Ведь он опирается на субъективные суждения, что создает возможности для различного типа искажений, обсуждавшихся ранее. Вот некоторые приемы, позволяющие избежать искажений при использовании инстинктивного байесовского подхода.

- *Отбирайте самых беспристрастных экспертов.* Если от результатов исследования зависит бюджет отдела, не привлекайте его руководителя к качественной оценке новой информации.
- *Проводите испытания вслепую.* Иногда специалистам предоставляют необходимую информацию, скрывая от них, какую именно проблему они оценивают. Например, из стенограммы обсуждения нового продукта в фокус-группе удаляют все ссылки на этот продукт, с тем чтобы судьи определяли лишь то, позитивна или негативна реакция группы.
- *Разделяйте обязанности.* Этот прием хорошо сочетается с испытаниями вслепую. Пусть одна группа экспертов оценит качественную информацию, а вторая группа, изучив выводы первой и не зная, о каком конкретно продукте, отделе, проекте и т. д. идет речь, даст окончательный байесовский ответ.
- *Помните о байесовских последствиях.* Просите экспертов указать заранее, как определенные результаты могут повлиять на их оценки, и применяйте байесовскую поправку до тех пор, пока они не окажутся внутренне непротиворечивыми.

НЕОДНОРОДНЫЙ БЕНЧМАРКИНГ И ЕГО ИСПОЛЬЗОВАНИЕ ДЛЯ ОЦЕНКИ «УЩЕРБА БРЕНДУ»

Все, что вам нужно представить в количественной форме, можно каким-то образом измерить. В любом случае это даст лучший результат, чем если не проводить измерений вовсе.

Закон Гилба

Одна из трудностей, с которыми столкнулись эксперты в задаче по определению среднего веса леденца (см. главу 9), заключалась в невозможности сравнить его с весом другого объекта для наглядности. Один эксперт заявил: «Не представляю себе, как может выглядеть один грамм леденцов», а другой отметил: «Я вообще плохо определяю на глаз вес маленьких предметов».

А что, если я подсказал бы им: визитная карточка весит примерно 1 г, 10-центовая монета — 2,3 г, а большая скрепка для бумаги — ровно 1 г? Помогло бы это сузить диапазоны предлагаемых ими значений? Кое-кому это было очень полезно, особенно если первоначально указанный ими диапазон был достаточно широким. Получив мою информацию, один человек, который сначала считал, что верхняя граница диапазона может составлять 20 г, сразу же опустил ее до 3 г. Люди корректируют свои оценки потому, что, как мы теперь знаем, все они, особенно калиброванные оценщики, являются интуитивными байесианцами. Они склонны довольно рационально обновлять первоначальную информацию, которой обладали, учитывая новые сведения, даже если те носят качественный характер или имеют к оцениваемому объекту отдаленное отношение.

Я называю этот метод обновления прежнего знания, основанный на сравнении с другими, непохожими, но неким образом связанными с объектом предметами, «неоднородным бенчмаркингом». Когда люди не могут представить себе, как выглядит какая-то величина, подобная информация о масштабе, пусть даже относящаяся к другим предметам, может оказаться очень полезной. При оценке возможного спроса на ваш продукт в новом городе вам пригодятся данные о спросе на него в других городах и даже сравнительные данные об экономическом уровне разных городов.

ПРЕДСТАВЛЕНИЕ О ПОРЯДКЕ ВЕЛИЧИН

Неоднородный бенчмаркинг — метод, при котором калиброванным экспертам, оценивающим неизвестную величину, предоставляют в качестве ориентиров

другие количественные показатели, даже если связь между ними и кажется отдаленной.

Пример: прогнозирование продаж нового продукта на основе информации о сбыте других товаров или аналогичных конкурентных продуктов.

Неоднородный бенчмаркинг проводился, в частности, в нашем примере с информационной безопасностью. В главах 4–6 я показал, как можно моделировать разные риски для безопасности, используя диапазоны значений и вероятности. Но похоже, что область информационной безопасности — неисчерпаемый источник как курьезных представлений о неизмеряемости многих вещей вообще, так и примеров подобных «нематериальных» объектов. Одна из таких неизмеримых величин — «мягкие» затраты, которыми чреваты определенные катастрофические события.

Кому не раз доводилось сталкиваться с сопротивлением проведению измерений в области информационной безопасности, так это Питеру Типпетту из компании Cybertrust. Работая над своим дипломом и кандидатской диссертацией по биохимии, он сделал то, что не пришло в голову никому из его сокурсников: создал первую антивирусную программу, получившую впоследствии известность как Norton Antivirus. Затем Типпетт провел ряд исследований с участием сотен организаций с целью сравнительной оценки рисков для разных угроз безопасности. Казалось бы, мнению такого человека об измеряемости безопасности, безусловно, можно доверять. Тем не менее у многих специалистов в сфере IT сама идея проведения подобных измерений, похоже, вызывает острое неприятие.

Типпетт предложил свой подход к решению проблемы, состоящий в том, чтобы задаться вопросом: «Насколько ужасно будет, если...?» Согласно такому подходу, специалисты по информационной безопасности решают, существует ли малейшая вероятность наступления такого катастрофического события, которое необходимо предотвратить любой ценой. Типпетт замечает: «Поскольку катастрофа может произойти в любой области, превентивные меры должны приниматься везде. Думать о приоритетах здесь не приходится». Он приводит следующий конкретный пример. «Одна компания из списка „Fortune 20“ выделила на реализацию 35 своих проектов в сфере информационных технологий 100 млн дол. Руководитель ее информационной службы захотел узнать, какие из проектов важнее, и получил от своих подчиненных ответ, что этого никто не знает и знать не может».

Одно из тех ужасных событий, наступление которых предвидит Типпетт, — это ущерб для бренда, ухудшение имиджа компании в глазах общественности. По мнению эксперта по безопасности, не ис-

ключено, что хакеры могут взломать сервер, украсть и использовать в своих целях какую-либо важную информацию — истории болезней из медицинского учреждения или данные о владельцах кредитных карт. Далее можно вообразить, что раз это происшествие так повредит имиджу компании, его необходимо предотвратить любой ценой и при любой, самой малой вероятности наступления такого события. Поскольку точно оценить вероятность такого ущерба или его сумму невозможно, это позволяет эксперту утверждать, что защита от хакеров так же необходима, как меры по предотвращению любой другой возможной катастрофы, и поэтому средства на защиту должны быть выделены без вопросов.

Но Типпетт не согласен с мнением, что масштабы проблемы ущерба для бренда и других нежелательных событий нельзя различить. Он предложил оценивать то, что объединяет гипотетические примеры ущерба бренду с реально имевшими место событиями. Например, он спрашивает, во что компании обошлись часовой сбой в работе электронной почты и другие нежелательные события. Затем следует новый вопрос: «Насколько велик этот ущерб по сравнению с...?» («примерно такой же», «вдвое меньше», «в 10 раз больше» и т.д.)

Специалисты Cybertrust уже получили некоторое представление о сравнительной шкале ущерба от различных нежелательных событий после анализа материалов экспертиз 150 случаев взлома баз данных о клиентах. В основном это были сведения о кражах данных карточек MasterCard и VISA.

Специалисты Cybertrust провели опросы руководителей компаний и широкой общественности по поводу восприятия ими ущерба бренду. Кроме того, они сравнили фактические убытки от снижения курса акций компании после подобных нежелательных событий. Благодаря этим опросам и сопоставлениям Типпетт сумел доказать, что ущерб бренду, нанесенный кражей хакерами клиентских данных, превышает потери от неправильного хранения резервной копии всей информации.

Сравнение с несколькими ориентирами позволило выявить разницу в масштабе ущерба от разных типов катастроф. Какой-то ущерб бренду был больше урона от одних событий, но меньше потерь от других. Более того, появилась возможность рассчитать «ожидаемые» убытки на основании относительных уровней потерь и их вероятностей.

Заслугу Типпетта в решении данной проблемы переоценить невозможно. До его исследований компании даже не представляли, насколько большим может быть ущерб бренду, даже порядка этой величины. А теперь они, по крайней мере, могут оценить масштабы вопроса и понимают значение снижения различных угроз безопасности.

Сначала руководство одной компании — клиента Типпетта — отнеслось к его результатам с известным недоверием, но, как пишет он

сам, через год число скептиков сократилось до одного, а все остальные уже стали его сторонниками. Наверное, оппонент Типпетта продолжал утверждать, что устранить неопределенность по этому вопросу не смогут никакие наблюдения. Но когда приходится оценивать такие явления, как возможный ущерб бренду, неопределенность обычно столь высока, что определение одного только масштаба чисел уже позволяет ее снизить, а значит, и провести измерение.

Конечно, ваша компания вряд ли станет проводить опросы в 100 других организациях, чтобы осуществить нужную оценку. Но в этом и нет необходимости, так как они уже были проведены и Cybertrust продает полученные результаты. К тому же использование этого метода даже внутри компании позволяет сократить неопределенность независимо от того, купит ваша компания результаты внешних исследований или нет.

ПРИМЕНЕНИЕ МЕТОДА НЕОДНОРОДНОГО БЕНЧМАРКИНГА

Неоднородный бенчмаркинг — идеальный способ оценки «мягких затрат» на преодоление последствий катастрофических событий, особенно в условиях, когда первоначальная неопределенность чрезвычайно высока. Примерами таких событий могут служить:

- кража хакерами данных кредитных карт и карт социального страхования,
- случайное обнародование персональных медицинских данных,
- массовый отзыв продукта из продажи,
- крупная авария на химическом заводе,
- корпоративный скандал.

Может показаться, что мы уделяем слишком много внимания информационной безопасности, но ведь данный метод может применяться в самых разных областях. Он не только годится для оценки ущерба от нарушения безопасности, но и позволяет определить приоритетность инвестиций, необходимых для предотвращения корпоративного скандала, катастрофы на химическом предприятии и т.п. На самом деле метод может использоваться и для оценки положительных последствий каких-либо событий. Какова стоимость того, что наш товар будет считаться эталоном высокого качества в отрасли? Бенчмаркинг — практичный способ определить масштабы проблемы в случаях, когда неопределенность так высока, что устранить ее кажется совершенно невозможно.

Если подобное использование ориентиров кажется «слишком субъективным», вспомним о цели нашего измерения в данном случае. Что

такое ущерб бренду, как не восприятие? Ведь мы оцениваем не физическое явление, а мнения людей. Такая оценка невозможна без понимания того, что ущерб бренду — это, по определению, изменение представлений потребителей. И вы определяете размеры этого ущерба, опрашивая потребителей. С другой стороны, можно проследить, что покупатели делают со своими деньгами, наблюдая за тем, как неблагоприятное событие повлияло на курс акций или объем продаж. В любом случае ущерб бренду оказывается измеренным.

КОЕ-КАКИЕ ТОНКОСТИ: БАЙЕСОВСКАЯ ИНВЕРСИЯ ДЛЯ ДИАПАЗОНОВ

Как уже упоминалось, в основе многих рисунков и таблиц, составленных мной для этой книги, лежит байесовская инверсия. Решая большинство статистических задач и задач по измерению, мы спрашиваем: «Какова вероятность того, что истинное значение данной величины равно X при условии, что я видел то-то и то-то?» Но вообще-то легче ответить на вопрос: «Если истинное значение равно X , то какова вероятность увидеть то, что я видел?» Байесовская инверсия позволяет нам ответить на первый вопрос, ответив сначала на второй. Нередко ответить на последний бывает намного легче.

Сразу хочу предупредить, что далее нам придется коснуться специальных вопросов. Если вы захотите пропустить это описание, то электронную таблицу для байесовской инверсии, составленную в том числе и на основе приводимого ниже примера, вы найдете на вспомогательном веб-сайте: www.howtomesureanything.com, выбрав связь «Bayesian Inversion» («Байесовская инверсия»). Я постарался сделать это описание как можно проще. Расчеты могут показаться довольно длинными, но я свел их к минимуму, перейдя, где это было возможно, сразу к функциям Excel.

Итак, предположим, что у нас есть магазин автозапчастей и возникла необходимость определить коэффициент удержания покупателей. Мы подозреваем о существовании проблемы с удовлетворенностью потребителей. Калиброванная оценка процента покупателей, которые захотят сделать в нашем магазине еще одну покупку, составляет 75–90% (доверительный интервал, как обычно, 90-процентный). Конечно, желательно, чтобы этот показатель был как можно выше, но если он не достигнет 80%, нам придется принять ряд весьма дорогостоящих корректирующих мер. Расчетная стоимость этой информации намного превышает 500 тыс. дол., но мы, естественно, постараемся минимизировать затраты на проведение опросов потребителей, пе-

реложив часть их на плечи своих покупателей. Помня о поэтапном определении интересующего нас показателя, выберем сначала всего 20 потребителей и посмотрим, какую информацию удастся получить. Если из этой выборки 14 человек скажут, что придут к нам за покупками еще, то как мы изменим первоначальный диапазон? Помните, что типичные параметрические, небайесовские методы не позволяют учитывать его при расчетах.

Начнем с более простого вопроса: если 90% покупателей скажут, что вновь придут за запчастями в наш магазин, то сколько человек из 20 сказали бы то же самое? Ответ очевиден — 90% от 20, или 18 человек. Если бы таких людей было 80%, то в нашей выборке их оказалось бы 16. Конечно, мы знаем, что совершенно случайно в числе 20 выбранных нами покупателей желающих вернуться в магазин может оказаться 15 или даже 20 человек. Поэтому нужно узнать не только ожидаемый результат, но и вероятность его получения.

Чтобы определить вероятность получения конкретного результата, используем специальное, уже упоминавшееся распределение, которое называется биномиальным. Напомним, что биномиальное распределение позволяет рассчитать вероятность определенного числа «попаданий» при условии проведения определенного числа попыток и того, что в каждой попытке может быть только один результат. Например, при подбрасывании монетки «попаданием» можно назвать выпадение орла, попытками — подбрасывания, а шанс попадания составляет 50%. Предположим, например, что мы хотим узнать вероятность того, что при 10 подбрасываниях орел выпадет четыре раза при вероятности его выпадения 50%. Вместо того, чтобы объяснять всю формулу и теорию комбинаторики, я сразу перейду к формуле программы Excel. В Excel мы просто запишем:

=binomdist(число попаданий, число попыток,
вероятность попадания, 0).

Взяв числа из нашего примера с подбрасыванием монеты, запишем: binomdist(4, 10, 0,5, 0), и Excel даст нам значение 20,5% (ноль в конце говорит о том, что нас интересует вероятность только этого конкретного результата. Записав вместо нуля единицу, получим накопленную вероятность, то есть вероятность указанного или меньшего числа попаданий). Данный результат означает, что есть 20,5-процентная вероятность того, что в случае 10-кратного подбрасывания монеты орел выпадет точно четыре раза.

В нашем примере с магазином автозапчастей покупатель, заявивший «да, я еще сюда вернусь», — это попадание, а размер выборки — это число попыток. Используя биномиальное распределение, менеджер может определить вероятность конкретного результата, на-

пример вероятность того, что среди 20 выбранных нами покупателей вернутся в магазин только 14, хотя вообще таких людей должно быть 90%. В Excel мы запишем: =binomdist(14, 20, 0,9, 0), что даст нам 88,7-процентную вероятность 14 попаданий при 20 случайно выбранных покупателях, если бы на самом деле 90% посетителей сказали, что готовы сделать еще одну покупку. Отсюда мы уже видим, что верхняя граница нашего первоначального диапазона не слишком правдоподобна.

Предположим теперь, что мы рассчитали эту вероятность для генеральной совокупности, в которой доля повторных покупателей составит сначала 75%, затем 76, 77 и т. д. вплоть до 90% (таким образом, шаг равен 1%). Используя некоторые таблицы в программе Excel, мы сможем быстро рассчитать вероятность конкретного результата при данном «истинном» проценте повторных покупателей. Для каждого приращения на 1% получим вероятность того, что 14 из 20 покупателей ответят утвердительно на вопрос о возвращении за повторной покупкой при данном «истинном» проценте повторных покупателей. Я бы рассчитывал эти вероятности для каждого приращения на 1%, начиная от 60% (что с учетом нашего 90-процентного СИ маловероятно, но возможно) и заканчивая 100%. Для каждого приращения проведем расчет на основе теоремы Байеса. Запишем все это вместе в следующем виде:

$$P(\text{Prop} = X | \text{Попадания} = 14/20) = P(\text{Prop} = X) \times \\ \times P(\text{Попадания} = 14/20 | \text{Prop} = X) / P(\text{Попадания} = 14/20),$$

где

$P(\text{Prop} = X | \text{Попадания} = 14/20)$ — вероятность данного процента повторных покупателей в генеральной совокупности (процента X) при условии, что 14 из 20 случайно отобранных объектов являются попаданиями;

$P(\text{Prop} = X)$ — вероятность того, что определенный процент покупателей в генеральной совокупности вернется снова (например, $X = 90\%$ генеральной совокупности покупателей, которые действительно сказали, что вернутся снова);

$P(\text{Попадания} = 14/20 | \text{Prop} = X)$ — вероятность 14 попаданий из 20 случайно выбранных объектов при данном проценте (проценте X) повторных покупателей в генеральной совокупности;

$P(\text{Попадания} = 14/20)$ — вероятность получения 14 попаданий из 20 попыток при условии, что все возможные проценты повторных покупателей в генеральной совокупности находятся в первоначальном диапазоне.

Мы знаем, как рассчитать $P(\text{Попадания} = 14/20 | \text{Prop} = 90\%)$ в Excel: [=binomdist(14, 20, 0,9, 1)]. Теперь нам нужно придумать, как рассчитать $P(\text{Prop} = X)$ и $P(\text{Попадания} = 14/20)$. Мы можем рассчитать вероятность каждого приращения на 1% доли повторных покупателей в нашем диапазоне, вернувшись снова к функции =normdist() в Excel и используя калиброванную оценку. Например, чтобы получить вероятность того, что 78–79% наших покупателей окажутся повторными (или, по крайней мере, заявят об этом во время опроса), мы можем записать следующую формулу Excel:

$$=\text{normdist}(0,79, 0,825, 0,0456, 1) - \text{normdist}(0,78, 0,825, 0,0456, 1).$$

Число 0,825 — это среднее значение нашего калиброванного диапазона: $(75\% + 90\%)/2$; 0,0456 — среднее квадратичное отклонение (как вы помните, в 90-процентном CI 3,29 среднего квадратичного отклонения): $(90\% - 75\%)/3,29$. Формула normdist дает нам разность между вероятностью получить менее 79% и вероятностью получить менее 78%, которая составляет 5,95%. Мы можем определить это для каждого приращения на 1% в исходном диапазоне, а затем рассчитать вероятность того, что доля повторных покупателей в генеральной совокупности равна X [то есть $P(\text{Prop} = X)$] для каждого мало-мальски вероятного значения X в нашем диапазоне.

Расчет значения $P(\text{Попадания} = 14/20)$ основан на всем, что мы делали до сих пор. Чтобы рассчитать $P(Y)$, когда мы знаем $P(Y|X)$ и $P(X)$ для каждого значения X , суммируем произведения $P(Y|X) \times P(X)$ для каждого X . Зная, как рассчитать $P(\text{Попадания} = 14/20 | \text{Prop} = X)$ и $P(\text{Prop} = X)$ для любого X , мы просто умножаем эти две величины для каждого X , затем суммируем их и получаем, что $P(\text{Попадания} = 14/20) = 8,56\%$.

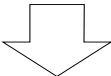
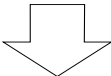
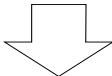
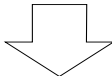
Теперь для каждого значения в исходном диапазоне (и даже немного за его пределами, чтобы получить «хвосты» в уравнении) мы рассчитываем $P(\text{Prop} = X)$, $P(\text{Попадания} = 14/20 | \text{Prop} = X)$ и $P(\text{Prop} = X | \text{Попадания} = 14/20)$, для каждого приращения на 1% повторных покупателей в генеральной совокупности величина $P(\text{Попадания} = 14/20)$ для всех одинакова и равна 8,56% (см. табл. 10.1).

Значения в последнем столбце — вероятности данного процента повторных покупателей в их генеральной совокупности. Если суммировать накопленные значения в последнем столбце (складываем все предшествующие значения в строке), то выяснится, что итог составит около 5%, когда процент повторных покупателей достигнет 79%, и 95%, когда этот процент будет равен 85%. Это означает, что наш новый 90-процентный CI сократится до 79–85%. Это не слишком большое сужение первоначального диапазона (75–90%), но тем не менее достаточно информативное. Теперь, согласно накопленным значениям

ТАБЛИЦА 10.1.

Отдельные строки из таблицы расчетов с использованием байесовской инверсии

$P(\text{Попадания} = 14/20) = \text{sumproduct}(\text{column A, column B}) = 0,0856$

(X) Доля повторных покупателей в генеральной совокупности (%)	(A) =normdist(X, 0,825, 0,0446, 1) – – normdist(X + 0,01, 0,825, 0,0446, 1)	(B) =binomdist(14, 20, X, 0)	=A × B/0,0856
60	8,04E-07	0,122795	1,41 E-05
61	2,25 E-06	0,126974	4,08 E-05
62	6,01 E-06	0,129136	0,000111
			
0,98	0,000189	7,95 E-16	2,15 E-17
0,99	8,59 E-05	3,13 E-20	3,83 E-22

последнего столбца, вероятность того, что мы находимся ниже основного порога в 80%, составляет 61%. Эту электронную таблицу целиком можно найти на веб-сайте: www.howtomesureanything.com

Похоже, что удержание покупателей у нас не на высоте. Но мы пересчитаем стоимость этой информации, и хотя она уменьшится, окажется, что провести дополнительные измерения все равно имеет смысл. Выберем еще 40 покупателей, и тогда в сумме их будет 60 человек. Из этих 60 только 39 скажут, что вернутся в наш магазин. Наш новый 90-процентный CI окажется равным 69–80%. Теперь верхняя граница равняется нашему первоначальному критическому порогу 80%, давая 95-процентную уверенность, что доля повторных покупателей низка настолько, что требует от нас серьезных, дорогостоящих изменений.

Расчеты оказались довольно сложными, но помните, что вы можете воспользоваться таблицами, приведенными на нашем вспомогательном сайте. И вполне возможно, что в данном случае сработал бы обсуждавшийся ранее субъективный байесовский метод, применяемый калиброванными экспертами. Возможно, опрос покупателей выявит такие качественные факторы, которые сумеют учесть наши калиброванные специалисты. Однако стоимость результатов этих важных измерений достаточно высока, чтобы оправдать наши дополнительные усилия.

Можно использовать рисунок, изображающий долю генеральной совокупности из главы 4* (хотя мы и искали бы интервал для покупателей, которые *не захотели бы* вернуться в магазин, так как численность подгрупп меньше половины размера выборки). Но с этим первоначальным диапазоном работать нельзя. Рисунок, приведенный в главе 9, был, между прочим, также составлен на основе байесовской инверсии, если не считать того, что я начал с максимально возможной неопределенности: равномерного распределения этой доли генеральной совокупности в диапазоне от 0 до 100%. Используя такой широкий диапазон в примере из этой главы, мы получили бы более широкий интервал с еще более худшей нижней границей по сравнению с представленным здесь способом. В этом случае мы начали с того, что получение результатов, потенциально таких же плохих, как допущения в главе 9 (даже учитывая наши разочаровывающие результаты), маловероятно. Байесовский интервал, плох он или хорош, учитывает прежние знания. Однако с увеличением размера выборки влияние первоначального интервала уменьшается. Выбрав 60 или более объектов наблюдения, мы получим ответ, весьма близкий к результату параметрического метода определения доли генеральной совокупности.

Овладев такого рода анализом, вы сможете пойти дальше и узнать, как решаются подобные проблемы, когда первоначальное распределение не является нормальным. Например, распределение может быть равномерным или нормально усеченным и не предполагать, что повторными покупателями окажутся более 100% (верхний хвост графика или нормального 90-процентного CI дает малую вероятность того, что это произойдет). С примерами таких распределений можно ознакомиться на нашем вспомогательном веб-сайте.

ИЗБЕГАЙТЕ «ИНВЕРСИИ НАБЛЮДЕНИЯ»

Многие задают вопрос: «Какой вывод я могу сделать из этого наблюдения?» Но Байес показал нам, что нередко полезнее задать вопрос: «Что я должен наблюдать, если будет соблюдаться условие X?» Ответ на последний вопрос позволяет разобраться с первым.

Хотя на первый взгляд байесовская инверсия может показаться весьма трудоемкой, она относится к наиболее эффективным из имеющихся в нашем распоряжении методам измерения. Если удастся сформулировать вопрос «Какова вероятность увидеть X, если справед-

* Неточность: рисунок доли генеральной совокупности приведен в главе 9 (см. рис. 9.2). — *Примеч. редактора.*

ливо Y ?» и превратить его в «Какова вероятность того, что справедливо Y , если мы наблюдаем X ?», то можно решить огромное число задач по измерению. В сущности, именно так мы и находим ответы на большинство научных вопросов. Если предложенная гипотеза верна, то что мы должны наблюдать?

Напротив, многие менеджеры, похоже, считают, что все измерения сводятся к поиску ответов на вопрос: «Какой я должен сделать вывод из того, что вижу?» Когда кажется, что совершена ошибка наблюдения, люди решают: на этом основании делать выводы нельзя, какой бы низкой ни была вероятность такой ошибки. Однако байесовский анализ показывает, что воображаемые менеджерами ошибки крайне маловероятны и что измерение все равно заметно снизило бы существующую неопределенность. Иными словами, отсутствие, по крайней мере, теоретического понимания байесовской инверсии приводит к переворачиванию вопроса и к формированию убеждения, что маловероятные ошибки сводят ценность измерений к нулю — то есть к самой неудачной разновидности «инверсии наблюдения».

■ ПРИМЕЧАНИЯ

- ¹ *David M. Grether, Mahmoud A. El-Gamal. Are People Bayesian? Uncovering Behavioral Strategies // Social Science Working Paper 919, 1995, California Institute of Technology.*
- ² *Tom DeMarco, Timothy Lister. Peopleware: Productive Projects and Teams. 2nd ed. New York: Dorset House Publishing, 1999.*

ЧАСТЬ IV

Не только основы



Предпочтения и подходы: «мягкие» аспекты измерения

Пример расчета ущерба бренду в главе 10 относится к особой группе проблем субъективной оценки. Правда, термин «субъективная оценка», в сущности, избыточен — ведь что на самом деле означает прилагательное «объективная» применительно к цене или ценности? Может ли рыночная стоимость фунта золота считаться «объективной» только потому, что она рыночная? Вовсе нет! Ведь рыночная стоимость — сама по себе результат того, как множество людей делают свои субъективные оценки.

Нередко менеджеры компаний считают, что такие явления, как качество, имидж или ценность, не поддаются измерению. Иногда такое мнение связано с тем, что люди не видят того, что могло бы служить «объективным» показателем качества, имиджа и т.п. Но это говорит только об их недостаточной информированности. Решение любой задачи по оценке качества (мнения общественности, стоимости бренда и т.п.) предполагает определение человеческих предпочтений. В этом смысле предпочтения — единственный источник измерения. Если такое положение означает, что измерение субъективно, значит, субъективность — просто природа измерения, а не физическое свойство какого-то предмета. Субъективность — лишь способ определения предпочтений в отношении этого предмета. А как только мы согласимся с тем, что данная разновидность измерений — оценка человеческих предпочтений, у нас останется только один вопрос: как эти предпочтения себя проявляют?

НАБЛЮДЕНИЕ ЗА МНЕНИЯМИ, ЦЕННОСТЯМИ И ПОИСКИ СЧАСТЬЯ

В широком понимании предпочтения людей обнаруживаются в том, что они говорят, и в том, что делают. *Объявленные* предпочтения — то, что, по их собственным словам, люди предпочитают. *Выявленные* предпочтения — предрасположенность, о которой свидетельствует их реальное поведение. Оба типа предпочтений позволяют значительно снизить неопределенность, но выявленные предпочтения, конечно, делают это более точно.

Спрашивая людей, о чем они думают, во что верят или что предпочитают, мы делаем наблюдение, которое обычно анализируем теми же статистическими методами, что и «объективные» физические свойства Вселенной (и результаты могут оказаться одинаково обманчивыми, ведь мы — люди; применяются только иные средства проверки). Мы просто выбираем группу респондентов и задаем им конкретные вопросы. По своей форме эти вопросы делятся на несколько основных категорий. Профессионалы в области проведения опросов пользуются более детальной классификацией, но начинающим вполне достаточно знать следующие категории.

- *Вопросы, предполагающие использование шкалы Ликерта.* Респондентов просят выбрать один из нескольких ответов, характеризующих их отношение к некоему объекту или явлению, при этом в качестве ответов обычно используются такие выражения, как «очень не нравится», «не нравится», «очень нравится», «абсолютно не согласен» или «совершенно согласен».
- *Вопросы с многовариантным ответом.* Респондентов просят выбрать один из нескольких взаимоисключающих вариантов ответов, например, если речь идет о политиках, то их просят выбрать республиканца, демократа или независимого кандидата.
- *Вопросы ранжирования.* Респондентов просят ранжировать что-либо в порядке убывания или возрастания. Пример: «Расположите в порядке приоритетности следующие восемь видов деятельности — от наименее предпочтительного (1) до наиболее предпочтительного (8)».
- *Открытые вопросы.* Респондентов просят дать ответ в произвольной форме. Пример открытого вопроса: «Есть у вас какие-либо претензии к нашей службе работы с потребителями?»

Специалисты по проведению опросов нередко называют их инструментом. Опросы разрабатываются так, чтобы минимизировать возможность систематической ошибки в ответах — специфического недостатка данного инструмента измерения.

Систематическая ошибка возникает в ответах, когда задания, намеренно или нет, сформулированы так, что респонденты дают ответы, не отражающие их истинное мнение. Иногда разработчик опроса делает это специально, если хочет получить вполне определенный ответ (пример такого вопроса: «Возмущает ли вас преступная небрежность губернатора N?»), но чаще это происходит случайно. Избежать систематической ошибки в ответах позволяет соблюдение следующих пяти простых правил.

1. *Вопрос должен быть максимально точным и коротким.* Многословные вопросы нередко приводят людей в замешательство.
2. *Избегайте многозначных терминов.* Многозначный термин — это слово с позитивной или негативной коннотацией (о которой разработчик опроса может и не догадываться), которая может влиять на ответы респондентов. Вопрос, поддерживают ли люди «либеральную» политику данного деятеля, где слово «либеральная» — пример многозначного термина, задавать не стоит. (Кстати, если суть этой политики не раскрывается, то такая формулировка является еще и примером неточного вопроса.)
3. *Не задавайте наводящих вопросов.* Наводящий вопрос сформулирован так, что заранее подсказывает респонденту, какого ответа от него ожидают. Пример наводящего вопроса: «Следует ли повысить зарплату низкооплачиваемым, перегруженным работой мусорщикам Кливленда?» Иногда эти вопросы задаются намеренно. Как и в случае многозначных терминов, простейшая защитная мера — попросить кого-то еще раз просмотреть ваш список вопросов. Когда я вижу, что наводящие вопросы используются нарочно, каждый раз удивляюсь, зачем вообще проводить такой опрос. Если заранее знать, какой ответ нужен, то какого «снижения неопределенности» можно при этом ожидать?
4. *Избегайте составных вопросов.* Пример: «Что вам больше нравится в машинах А и В: сиденье, рулевое колесо или приборная доска?» Респондент не поймет, ответа на какой вопрос от него ожидают. Разбейте задание на несколько вопросов с многовариантным ответом.
5. *Меняйте вопросы так, чтобы избежать установки на однообразные ответы.* Установка на выбор определенных ответов — это склонность людей отвечать на вопросы (ранжировать ответы) одинаково независимо от их содержания. Если у вас подготовлены наборы вопросов, связанных с баллами от 1 до 5, проследите за тем, чтобы «5» не всегда означало «положительный» ответ (или наоборот). Вы ведь хотите, чтобы респонденты прочитали и действительно ответили на все вопросы, а не автоматически поставили галочки в каждой клетке каждого столбца.

Конечно, узнать, что люди предпочитают, чего хотят и о чем думают, позволяет не только их прямой опрос. Немало информации об их предпочтениях дает наблюдение за тем, что они делают. На самом деле это даже более надежный способ выяснения их истинного мнения и тех моральных ценностей, которых они придерживаются.

Когда человек говорит, что охотнее пожертвует 20 дол. сиротам, чем потратит их на кино, но на самом деле в прошлом году многократно ходил в кинотеатры и ни разу ничего не пожертвовал сиротам, это означает, что его выявленные предпочтения не совпадают с объявленными. Понять, каковы истинные предпочтения человека, позволяют две вещи, ценимые людьми больше всего: время и деньги. Проанализировав, как люди тратят свое время и деньги, вы сразу увидите их истинные предпочтения.

Похоже, что когда респонденты говорят, например, что «совершенно согласны» с такими утверждениями, как: «Елочные украшения стали дорожать в магазинах слишком рано», мы уже не определяем «истинные» значения. Однако подход, предложенный в предыдущих главах, остался тем же. Вам по-прежнему неизвестно точное значение какой-то переменной (например, вы только предполагаете, что часть покупателей, считающих, что елочные украшения начинают дорожать слишком рано, составляет 50–90%), и существует такое ее возможное значение, при котором вы измените свое решение (например, если более 70% покупателей выразят свое абсолютное согласие с тем, что эти товары начинают расти в цене слишком рано, то магазин откажется от планов повышать их цену еще раньше). На основе данных сведений вы рассчитываете стоимость дополнительной информации и выбираете соответствующий ей метод выборки или какой-либо другой способ измерения.

Да, мы действительно отклонились от оценок, выражаемых в единицах измерения, как это было раньше. Когда мы точно определяли, почему нас интересует количество чего-то, обычно всегда можно было четко указать единицы измерения и шкала Ликерта не требовалась. Но есть еще один интересный прием: соотнесение результатов опросов потребителей с другими однозначными и намного более полезными величинами. Ведь раз возникло желание оценить удовлетворенность потребителей, значит, вы хотите сохранить свой бизнес, клиентов и добиться хорошей рекламы, распространяемой довольными покупателями.

На самом деле, субъективные ответы можно соотнести с объективными показателями, и такой анализ проводится постоянно. Кое-кто даже попытался таким образом измерить счастье (см. врезку «Как измерить счастье»). Сумев сопоставить одну вещь с другой, а затем найти корреляцию между одной из них и деньгами, вы сможете выразить и ту и другую в деньгах. А если это окажется слишком сложным, просто спросите человека напрямую: «Сколько вы готовы заплатить за это?»

КАК ИЗМЕРИТЬ СЧАСТЬЕ

Эндрю Освальд, профессор экономики университета в Варвике, придумал, как определить стоимость счастья¹. Он не спрашивал у людей напрямую, сколько те готовы заплатить за свое счастье, а просил их указать свой доход, определить по шкале Ликерта, насколько они счастливы, и рассказать о важных событиях, которые произошли в их жизни недавно (рождении детей, смерти близких, свадьбах и т. д.).

Это позволило Освальду определить, какое влияние оказывают разные события на ощущение испытываемого человеком счастья. Он установил, насколько недавняя смерть родственника снижает это ощущение, а продвижение по службе — повышает. Более того, сопоставив влияние дохода на состояние счастья, он сумел рассчитать эквивалент дохода, соответствующий ощущению счастья, возникающему в результате определенных событий. Так, Освальд обнаружил, что продолжительный брак делает человека таким же счастливым, как если бы его доход ежегодно возрастал на 100 тыс. дол. (Поскольку мы с женой только что отметили 10-летнюю годовщину свадьбы, я почти так же счастлив, как если бы, будучи холостым, заработал за этот же период 1 млн дол. Конечно, это средние цифры, и многое зависит от характера человека. Поэтому я говорю своей жене, что для меня оценка Освальда наверняка занижена и наш брак еще долго будет оставаться счастливым.)

Готовность платить: ОПРЕДЕЛЕНИЕ СТОИМОСТИ ЧЕРЕЗ КОМПРОМИССЫ

Повторим, что оценка в силу самой своей природы всегда субъективна. Даже рыночная стоимость акций или недвижимости — всего лишь результат субъективных суждений нескольких участников рынка. Чтобы получить «объективный» показатель при расчете чистого собственного капитала компании, экспертам приходится суммировать такие параметры, как рыночная стоимость недвижимости (то есть сумму, которую, по их мнению, за нее захотят заплатить на рынке), стоимость бренда (то, насколько больше потребители захотят заплатить за товар данной марки), стоимость используемого оборудования (сколько кто-то захочет за него заплатить) и т. п. Какими бы объективными они ни считали свои расчеты, основная используемая ими единица измерения (доллар) — мерило стоимости.

Вот почему, чтобы оценить большинство вещей, достаточно спросить людей, сколько они готовы за них заплатить, или же, что еще лучше, определить, сколько они *уже* за них заплатили, анализируя действия за прошедшие периоды. Использование метода WTP (willingness to pay — готовность платить) обычно предполагает проведение опро-

сов случайно выбранных людей относительно того, сколько они готовы заплатить за конкретные вещи — обычно такие, которые невозможно оценить другим способом. Этот метод был использован, в частности, для определения стоимости предотвращения исчезновения редких видов животных, а также стоимости улучшения здравоохранения и состояния окружающей среды.

В 1988 г. я впервые участвовал в консалтинговом проекте компании Coopers & Lybrand. Мы оценивали результаты деятельности полиграфического подразделения одной финансовой компании, чтобы определить целесообразность расширения заказа сторонней типографии. Совет директоров финансовой компании считал сотрудничество с местным бизнесом очень ценным. К тому же у президента типографии были друзья в совете директоров. Он спросил: «Вот мы, например, не занимаемся оказанием финансовых услуг, так почему же вы занимаетесь полиграфией?», добиваясь того, чтобы заказ его фирме был увеличен.

Несколько скептически настроенных членов совета директоров пригласили специалистов Coopers & Lybrand, чтобы оценить экономические последствия такого шага. Я был младшим аналитиком и выполнял все расчеты по этому проекту. Выяснилось, что компании следует не расширять, а сокращать масштабы аутсорсинга. Финансовая компания была достаточно крупным предприятием, чтобы привлечь опытных полиграфистов, обеспечить высокую загрузку своего оборудования и заключать с поставщиками выгодные контракты. Она уже располагала обученным персоналом, хорошо разбиравшимся в типографском деле.

Возможно, полиграфия и не должна была стать основным направлением деятельности финансовой компании, но соотношение «затраты/выгода» явно говорило в пользу сохранения полиграфического подразделения в ее составе и даже расширения масштабов его деятельности. Не было сомнений в том, что компании гораздо дороже обойдется аутсорсинг, даже с учетом всех льгот работникам, расходов на ремонт оборудования, аренду помещения и т. п. Предложенный местной типографией вариант означал бы для компании потерю нескольких миллионов долларов в год, что существенно превышало расходы на работу собственного подразделения. Кроме того, появились опасения, что рост аутсорсинга приведет к снижению качества получаемых компанией услуг, поскольку персоналу типографии не придется беспокоиться о каких-либо предпочтениях своего клиента. Чистая приведенная стоимость данного плана аутсорсинга за пять лет превысила —15 млн дол.

В итоге компании пришлось решать, готова ли она заплатить 15 млн дол. за дружбу с типографией и желание поддержать местное сообщество. Будучи младшим аналитиком, я не считал возможным да-

вать членам совета директоров какие-либо рекомендации по поводу суммы, которую стоило за это заплатить, а лишь честно сообщил, во что обойдется такое решение. Если бы руководители сочли, что дружба с общиной стоит 15 млн дол., то такие финансовые потери были бы приемлемыми. Если же они оценили бы ее в меньшую сумму, то данный вариант явно не годился. В конце концов было решено, что выгоды от этой дружбы и поддержки местного сообщества того не стоят. Они не увеличили объем передаваемых на сторону работ и даже решили его сократить.

В то время я называл такие проблемы задачами «покупать искусство». Вы можете думать, что определить стоимость «бесценного» произведения искусства невозможно, но если я назову вам его цену, то вы сами сумеете определить эту стоимость. Если кто-то определяет произведение Пикассо как «бесценное», но никто не заплатит за него больше 10 млн дол., то ясно, что стоимость данного предмета искусства не может быть больше этой суммы. Мы не пытались точно определить стоимость дружбы, но объяснили финансовой компании, сколько за нее придется заплатить, и она смогла сделать свой выбор.

Разновидностью WTP является метод VSL (value of statistical life — стоимость статистической жизни). При этом подходе людей не спрашивают напрямую, во сколько они оценивают свою жизнь, но просят их указать, сколько они готовы заплатить за незначительное снижение риска своей преждевременной смерти. Ведь, принимая свои решения, человек нередко фактически делает выбор между желанием сэкономить и возможностью хоть немного сократить вероятность умереть слишком рано. Вы могли потратить больше и купить чуть более безопасную машину. Допустим, что за 5000 дол. удалось бы на 20% снизить вероятность погибнуть в автомобильной катастрофе, которая и так составляет всего 0,5% (учитывая то, сколько времени вы проводите за рулем, где водите машину, каковы ваши навыки вождения и т.д.). Таким образом, в итоге вы снизили бы общую вероятность умереть преждевременно на одну десятую от 1%. Отказавшись от этой возможности, вы фактически объявляете, что предпочитаете сохранить 5000 дол., а не снижать вероятность своей преждевременной смерти на 0,1%. В этом случае вы оцениваете стоимость своей статистической жизни менее чем в $5000 \text{ дол.} / 0,001 = 5 \text{ млн дол.}$ (ведь вы отказались от необходимых расходов!). Можно потратить 1000 дол. на рентгеновское обследование и с вероятностью 1% обнаружить у себя некое опасное заболевание на ранней стадии, когда своевременно принятые меры приведут к исцелению. В случае осуществления расходов на эту сумму для указанной цели ваша VSL составляет, по крайней мере, 1000 дол./0,001, или 100 000 дол. Можно продолжать анализ других ваших решений по приобретению разных товаров или услуг, повышающих безопасность, и делать дальнейшие предположения о том, во

сколько вы оцениваете возможность снизить угрозы для своей жизни, а следовательно, и саму жизнь.

У данного подхода ряд недостатков. Во-первых, люди очень плохо оценивают риски, которые несут в разных ситуациях, поэтому их выбор не слишком информативен. Доктор Джеймс Хаммитт и специалисты Гарвардского центра анализа риска (Harvard Center for Risk Analysis) сделали следующие наблюдения.

Люди на удивление плохо понимают смысл вероятности, особенно незначительной и связанной с альтернативами в области здравоохранения. В одном общем опросе населения только 60% респондентов правильно ответили на вопрос: «Какая вероятность больше — 5 из 100 000 или 1 из 10 000?» Эта неспособность к количественному мышлению не позволяет людям правильно формулировать свои предпочтения².

Если люди действительно настолько математически безграмотны, оценки, полученные путем опросов населения, не заслуживают доверия. Но Хаммитт не испугался неспособности к математике некоторых людей, а просто сделал на нее поправку. Ответы респондентов, правильно отвечающих на подобные вопросы, он оценивает отдельно от ответов тех, кто незнаком с такими простейшими понятиями, как «вероятность» и «риск».

Помимо математической безграмотности, по крайней мере отдельных респондентов, тем из нас, кому приходится оценивать такие вещи, часто приходится сталкиваться с неуместными проявлениями чувства праведного негодования. Многие исследования показывают, что около 25% участников опросов, проводившихся с целью определения стоимости сохранения окружающей среды, отказались отвечать на том основании, что природа имеет безусловное право на защиту, какой бы ни была ее стоимость³. В результате люди, чьи ответы, наверное, повысили бы среднюю WTP за сохранение окружающей среды, воздержались от участия в опросе и итоговая оценка оказалась меньше, чем могла бы быть. Но мне кажется, что это чувство праведного гнева не более чем личина. Ведь такие люди могут прямо сейчас отказаться от всякой роскоши и сделать пожертвования на защиту окружающей среды. Или же они могут немедленно бросить свою работу и стать волонтерами Green Peace. Однако они этого не делают. Их поведение нередко идет вразрез с теми высокими моральными ценностями, которых они якобы придерживаются. Кое-кто также сопротивляется попыткам оценить в деньгах человеческую жизнь, но, опять-таки, во все не отказывается от удовольствий, чтобы сделать пожертвования на развитие общественного здравоохранения.

Наблюдаемое несоответствие между заявлениями о бесценности некоторых вещей и стремлением к роскоши, которое часто проявляется в выборе людей, может иметь свое объяснение. Как показали ис-

следования Хаммитта (и многих других), на удивление большой процент населения настолько математически безграмотен, что противодействие этих людей определению денежной стоимости человеческой жизни может частично объясняться просто их нелюбовью к числам. Проявление праведного негодования может быть частью их защитного механизма. Вероятно, они думают, что их математическая безграмотность не будет так бросаться в глаза, если задачам количественной оценки вообще и количественной оценки таких нематериальных вещей в частности не придавать особого значения.

Особенно деликатна задача определения стоимости всего, что связано со счастьем, здоровьем и жизнью человека. Поиск в Интернете по фразе «свести все к голой цифре» дает тысячи совпадений, большинство которых — цитаты из возражений против проведения подобных измерений. Создавать математические модели — такая же уникальная способность человека, как писать стихи или рисовать картины, однако вы вряд ли услышите, чтобы кто-нибудь жаловался на «сведение к стихотворению» или «сведение к картине».

Мне не раз доводилось проводить анализ «риск/доходность» для инвестиционных проектов федерального правительства, направленных на улучшение общественного здравоохранения. Во всех случаях мы просто использовали широкие диапазоны значений из различных исследований WTP и VSL. И сколь бы широкими ни были эти диапазоны, обычно никаких дополнительных измерений для их уточнения не требовалось. А те, кому неприятна сама мысль об измерении подобных вещей деньгами, должны подумать о том, что в противном случае при анализе проекта игнорирование какого-либо важного фактора сведет его величину практически к нулю. Это приведет к неправильной расстановке приоритетов и необоснованной недооценке значения определенных мер, намеченных компанией. Из всех многочисленных случаев, когда я работал над анализом проектов, только в одном стоимость информации оказалась такой, что возникла необходимость дальнейшего уточнения интересующих нас переменных. В большинстве случаев основным неизвестным была, как ни странно, вовсе не стоимость общественной безопасности или общественного благосостояния. Первоначальные диапазоны при всей их широте оказывались достаточно точными, а основные измерения проводились для определения совершенно других неизвестных.

Между прочим, использовавшийся многими государственными организациями и заимствованный из различных исследований WTP и VSL интервал значений суммы, позволяющей избежать одной преждевременной смерти, составлял 2 млн—20 млн дол. Если вам кажется, что это слишком мало, то вспомните, сколько вы сами тратите на свою безопасность. Вспомните и о том, как вы расходуете деньги на предметы роскоши, пусть и немногочисленные, вместо того, чтобы боль-

ше жертвовать на борьбу со СПИДом или раком. Если бы вы действительно думали, что жизнь любого человека стоит намного больше этой суммы, то вели бы себя иначе. Анализ собственного поведения позволяет нам понять, что фразы типа «жизнь бесценна» произносят только лицемеры.

СТАВИМ ВСЕ НА КАРТУ: КОЛИЧЕСТВЕННОЕ ОПРЕДЕЛЕНИЕ СКЛОННОСТИ К РИСКУ

Обычно такого рода компромиссы приходится делать, когда оцениваешь такую вещь, как склонность (толерантность, терпимость) к риску*. Никто не может за вас рассчитать, какой риск приемлем для вас или вашей фирмы, но вы имеете возможность его измерить. Как и в методе VSL, главное здесь — определить желательные для вас комбинации (реальные или гипотетические) различных уровней риска и доходности.

Именно этим и занимаются некоторые портфельные менеджеры в процессе управления финансовыми портфелями. В 1990 г. Нобелевская премия по экономике была присуждена Гарри Марковицу (Марковитцу), создавшему современную портфельную теорию (Modern Portfolio Theory, МРТ)**. На этой теории, разработанной Марковицем еще в 1950-е годы, основано большинство современных методов оптимизации портфеля. Возможно, самая простая ее составляющая — кривая риска, на который инвесторы готовы пойти, чтобы получить данную доходность. Когда потенциальная доходность инвестиций высока, инвесторы обычно готовы смириться с более высоким риском. Если же инвестиции малорискованны, то чаще всего инвесторы соглашаются и на более низкую доходность. Это показывает кривая на графике, точки которой соответствуют максимально приемлемому риску и минимально приемлемой доходности. Рисунок 11.1 демонстрирует, как может выглядеть инвестиционная граница.

Наш рисунок несколько отличается от графика Марковица. Его ось риска на самом деле отражала прошлую волатильность доходности определенных акций (складывавшейся из приращения или обесценения капитала и дивидендов). Однако когда мы планируем инвести-

* Уровень риска, приемлемый для компании. В материальном выражении склонность к риску представляет собой сумму, которую предприятие может себе позволить безболезненно потерять. — *Примеч. редактора.*

** В литературе встречаются названия «современная теория портфеля», «теория современного портфеля» и др. — *Примеч. редактора.*

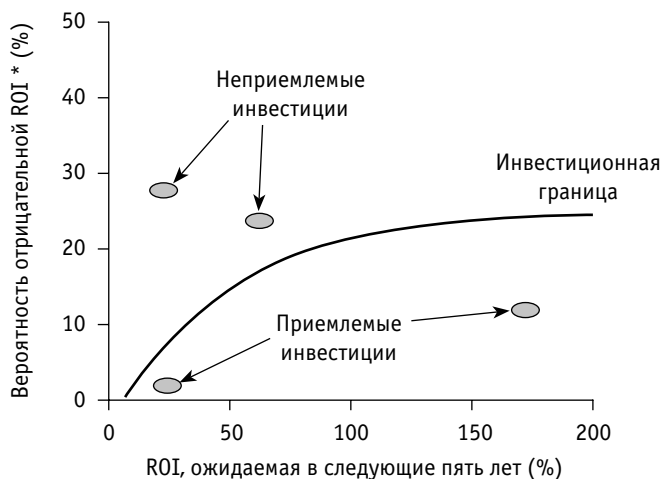


РИСУНОК 11.1. Инвестиционная граница

* Доходность инвестиций.

ции в информационные технологии или разработку нового продукта, «прошлой волатильности» у нас обычно нет, но есть другой вид риска — вероятность понести убытки.

Вы можете быстро построить инвестиционную границу для себя или для своей фирмы. Представьте, что собираетесь сделать крупные инвестиции. Что для вас «крупные» (но все же не необычно крупные) инвестиции: 1 млн или 100 млн дол.? Определите эту цифру, какой бы она ни была, и используйте ее далее в этом примере.

Предположим теперь, что с помощью метода Монте-Карло вы рассчитали доходность для тысяч сценариев. Среднее значение возможной доходности для всех сценариев — годовая доходность инвестиций, которая в течение ближайших пяти лет должна составить 50%. Однако в связи с этой ROI существует некая неопределенность, и даже есть вероятность, скажем 10-процентная, того, что она будет отрицательной. Захотите ли вы сделать инвестиции с таким прогнозом? Если да, то давайте повысим риск до 20%, а если нет, то понизим его до 5%. В каком случае условия инвестирования приемлемы для вас? Будем и далее повышать или понижать риск до тех пор, пока он не окажется максимально приемлемым, а доходность — минимально приемлемой. Точка на графике, соответствующая данному риску и данной доходности, находится на вашей «инвестиционной границе». Повысим теперь ROI до 100%. Каким должен быть риск, чтобы такая доходность стала минимально приемлемой? Так мы найдем еще одну точку на

инвестиционной границе. Предположим, наконец, что вы можете сделать такие инвестиции, доходность которых никогда не будет отрицательной. На какую самую низкую среднюю доходность вы готовы согласиться, чтобы избежать риска, что она окажется отрицательной?

Все три точки расположены на вашей инвестиционной границе. При необходимости можно найти и другие точки, соответствующие более высокой или более низкой ROI. В какой-то момент вы увидите, что точки образовали кривую вполне определенной формы.

Несколько советов инвесторам, пользующимся методом МРТ. Во-первых, инвестициям разных размеров должны соответствовать свои инвестиционные границы. Марковиц первоначально разрабатывал свою инвестиционную кривую для всего портфеля, а не для отдельных инвестиционных проектов. Но я строю всего три кривые (для мелких, средних и крупных оцениваемых мной инвестиций), а остальные можно получить путем интерполяции. (Я составил для себя простую электронную таблицу, позволяющую интерполировать нужную кривую, но вы практически с тем же успехом можете представить ее себе визуально.)

Я часто использую этот простой инструмент для оценки отдельных инвестиций по нескольким причинам. Инвестиционные возможности появляются в течение года в любой момент, в том числе и тогда, когда реализация других проектов еще не завершена. Оптимизировать портфель целиком обычно затруднительно, поскольку запустить или заморозить любой проект по желанию почти невозможно.

В 1997 и 1998 гг. я написал статьи для журналов «Information Week»³ и «CIO Magazine»⁴, где описал свой метод инвестиционной границы, который использовал в прикладной информационной экономике. Я давал задание начертить инвестиционную границу многим менеджерам и получил десятки графиков, составленных для самых разных организаций. В каждом случае на построение инвестиционной границы с нуля уходило от 40 до 60 минут, кто бы ни выполнял эту работу: один человек или 20 членов инвестиционного комитета.

Из всех, кто когда-либо присутствовал на моих занятиях (а это обычно лица, определяющие политику организаций), не было ни одного человека, кому не удалось бы справиться с этим заданием достаточно быстро.

Но я заметил и еще кое-что: даже если задание выполняли общими усилиями 10 членов инвестиционного комитета, они легко достигали консенсуса. Какие бы разногласия ни возникали относительно приоритетности проектов, они быстро приходили к единому мнению по поводу того, насколько склонна к риску их компания.

Применяя инвестиционные границы для оценки инвестиционных проектов, мы обнаруживаем, что требуемая скорректированная на риск ROI должна быть значительно выше типичных «пороговых ста-

вок» (требуемой минимальной доходности), используемых иногда руководителями, санкционирующими вложение средств в информационные технологии (нередко эти пороговые ставки составляют 15–30%). С ростом объемов предполагаемых инвестиций этот эффект быстро усиливается. Доходность самых крупных проектов разработки программного обеспечения должна *намного превышать* 100%. Риск замораживания проекта, неопределенность в связи с выгодами и риск возникновения неожиданных препятствий — все это увеличивает рискованность таких проектов, а значит, и их требуемую доходность. Для руководителей, санкционирующих инвестиции в информационные технологии, этот вывод важен по целому ряду причин.

Не будет преувеличением сказать, что инвестиции в разработку программного обеспечения обычно входят в число самых рискованных проектов вложения средств, которые реализуют компании. Например, вероятность того, что крупный проект такого рода будет заморожен, прямо пропорциональна продолжительности его осуществления. В 1990-х годах закончились ничем около четверти всех существовавших более двух лет проектов разработки программного обеспечения (для сравнения: показатель невыполнения обязательств по мусорным облигациям был ниже 25%).

Тем не менее большинство организаций, применяющих анализ ROI, не принимают во внимание такие риски. Типичные пороговые ставки не корректируются на разные риски, связанные с ИТ-проектами, хотя именно риски должны в основном учитываться при принятии подобных решений. Если бы руководители компаний анализировали инвестиции в разработку программного обеспечения с точки зрения соотношения «риск/доходность», то наверняка принимали бы решения более обоснованные, чем сделанные только на основе фиксированных пороговых ставок.

КОЛИЧЕСТВЕННОЕ ОПРЕДЕЛЕНИЕ СУБЪЕКТИВНЫХ КОМПРОМИССОВ: РЕШЕНИЕ ПРОБЛЕМЫ НЕСКОЛЬКИХ ВЗАИМОИСКЛЮЧАЮЩИХ ПРЕДПОЧТЕНИЙ

Кривая инвестиционной границы — пример тех кривых полезности, с которыми будущие менеджеры компаний знакомятся на первом курсе университета. К сожалению, большинство из них, по-видимому, считают полученные знания чисто теоретическими и не имеющими никакого практического значения. Но кривые полезности — идеальный инструмент, позволяющий определять, какой частью одного стоит по-

жертвовать ради получения другого. Разнообразные виды кривых полезности помогают тем, кто принимает решения, детально выяснять, какой компромисс для них приемлем.

Один из самых распространенных компромиссов, которые приходится делать менеджеру, — это выбор между эффективностью и качеством. Он очень полезен при попытках оценить предпочтения и стойкость. Термины «эффективность» и «качество» толкуются настолько по-разному, что с уверенностью о них можно сказать только одно: высокая эффективность и высокое качество лучше, чем низкая эффективность и низкое качество. Но, как мы уже говорили, причин для такой неоднозначной трактовки не существует, и объяснить содержание этих слов так же легко, как и любых других «нематериальных» понятий.

Когда клиенты просят меня помочь им оценить эффективность, я всегда спрашиваю их: «А что вы подразумеваете под эффективностью?» В ответ они, как правило, предоставляют мне перечень разрозненных наблюдений, которые ассоциируются у них с эффективностью, например: «Этот человек всегда все делает вовремя» или «О ней заказчики всегда отзываются положительно». Могут упоминаться и такие факторы, как небольшое число допускаемых ошибок или высокая производительность труда, например: «За три месяца этот сотрудник сумел собрать целых три модуля без брака». Иными словами, проблема в том, что никто не представляет, как наблюдать эффективность. Один мой клиент высказался: «Я знаю, что должен искать, но как мне суммировать все это? Могу ли я считать, что тот, кто делает всю работу вовремя и почти без ошибок, работает эффективнее того, кто постоянно получает больше положительных отзывов клиентов?»

На самом деле это не проблема измерения, а вопрос документального оформления субъективных компромиссов. Это проблема отражения множества разнородных наблюдений в едином «индексе». И здесь, чтобы такое обобщение оказалось логичным, мы можем воспользоваться кривыми полезности. С их помощью достаточно просто показать, как сформулировать компромиссы по аналогии со следующими примерами:

- Кто работает эффективнее: программист, который 99% заданий выполняет вовремя и в 95% случаев безошибочно, или тот, кто только 92% заданий выполняет вовремя, но в 99% случаев безошибочно?
- Стало ли качество товара выше, если процент брака снизился на 15%, а возврат товара покупателями увеличился на 10%?
- Стала ли «стратегическая согласованность» полнее, если прибыль повысилась на 10%, но «общий показатель качества» упал на 5%?

Для каждого из этих случаев можно составить график, отражающий все возможные компромиссы, по аналогии с определением желаемого соотношения «риск/доходность». Точки на кривой представляют комбинации, одинаково ценные для лица, принимающего решение. В предыдущем примере с инвестиционной границей каждая точка на кривой имела нулевое значение и представляла такое сочетание риска и доходности, что тому, кто принимал решения, было безразлично, принять или отвергнуть данный инвестиционный проект (так как при данной доходности риск был не ниже самого высокого из приемлемых).

Мы могли бы начертить на том же графике и другие кривые полезности для инвестиционных проектов с ненулевой ценностью, каждая из которых соответствовала бы некоей полезности. Иногда экономисты называют эти графики кривыми изопольности, подразумевая, что полезность фиксированна, или постоянна. Поскольку человеку безразлично, какую из двух точек на такой кривой выбрать, экономисты также называют кривую полезности кривой безразличия*. Аналогично тому, как замкнутые линии на картах рельефа показывают точки, расположенные на одинаковой высоте, кривая полезности — геометрическое место точек, отражающих одинаково ценные комбинации.

Рисунок 11.2 представляет собой график с несколькими кривыми полезности. Он иллюстрирует гипотетический пример оценки руководством компромиссов между качеством работы и пунктуальностью выполнения заказов. Рисунок помогает внести ясность в требования, предъявляемые программисту, инженеру, редактору и т. д. Вы видите, что если бы сотрудники А и Б выполняли в срок одинаковое число заданий, но А допускал меньше ошибок, то его считали бы лучшим работником. Кривая обосновывает предпочтения, когда выбор не очевиден, например когда качество работы А выше, но Б более пунктуален.

Эти кривые построены таким образом, чтобы любые две точки на каждой из них отражали одинаково ценные комбинации параметров. Так, верхняя кривая показывает, что, по мнению топ-менеджмента, работник, выполняющий на 96% правильно и на 96% в срок порученные ему задания, имеет для компании ту же ценность, как тот, кто на 93% правильно и на 100% в срок делает свою работу. Не забудьте, что это только гипотетическая оценка какого-то конкретного менедже-

* Кривая безразличия — геометрическое место точек, каждая из которых представляет такую комбинацию двух свойств, что пользователю безразлично, какую ему выбрать. Другими словами, кривая безразличия показывает альтернативные сочетания этих свойств, обеспечивающие одинаковый уровень полезности. Двигаясь от одной комбинации к другой, пользователь увеличивает полезность, обеспечиваемую одним свойством, но уменьшает полезность, обеспечиваемую другим свойством. — *Примеч. переводчика.*

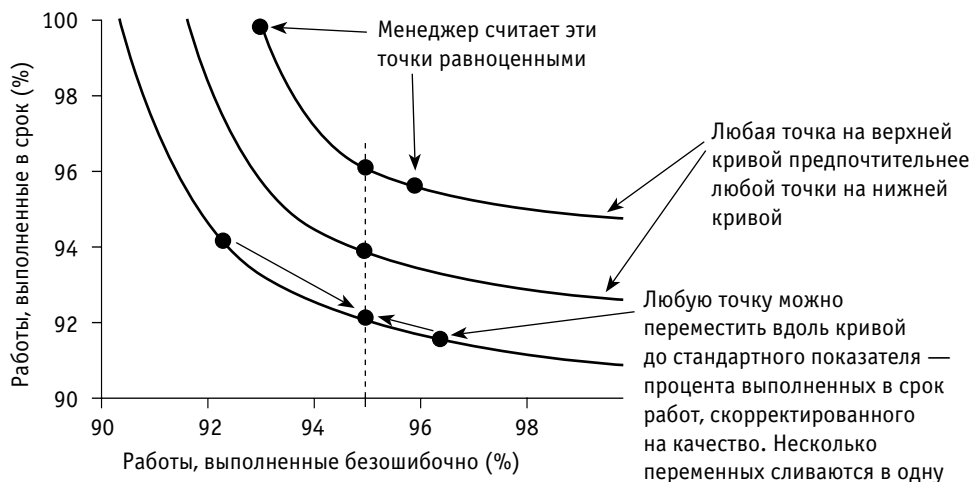


РИСУНОК 11.2. Гипотетические «кривые полезности»

ра, а не общепринятая фиксированная норма. Вполне возможно, что ваши предпочтения оказались бы несколько иными.

Было построено семейство аналогичных кривых таким образом, что любая точка на верхней кривой считается более предпочтительной, чем любая точка на нижней. Для сравнения достаточно начертить лишь несколько кривых, хотя на самом деле между приведенными на рисунке линиями существует бесчисленное множество линий. Просто менеджеры строят их ровно столько, сколько нужно для интерполяции.

Кривая полезности, обеспечиваемой любыми двумя факторами (например, качеством работы и ее своевременным выполнением или низким риском и высокой доходностью), позволяет упростить выражение ценности комбинации факторов, представленной точкой на графике. Поскольку каждую точку можно перемещать вдоль кривой так, что ее ценность не изменится, можно рассматривать все точки эквивалентными одной, расположенной на единой стандартизированной линии. В этом случае мы стандартизируем качество и выражаем сравнительную ценность любой точки на графике через коэффициент своевременности выполнения работ, скорректированный на качество. Заявив, что «сотрудник, выполняющий безошибочно $X\%$ работ и завершающий в срок $Y\%$ заданий, не уступает тому, кто выполняет безошибочно 95% заданий и в срок $___\%$ заданий», мы фактически свернули две переменные в одну.

То же обычно делается с риском и доходностью. С помощью семейства кривых «риск/доходность» мы, взяв эти параметры любых

инвестиционных проектов, можем выразить их как доходность, скорректированную на риск. Этот метод свертывания двух разных показателей в один применяют независимо от числа переменных. Если, например, я построил кривые полезности для факторов X и Y, а затем кривые полезности для факторов Y и Z, то каждый сможет построить на их основе кривую полезности для X и Z. Таким путем можно свернуть в единый стандартизированный показатель несколько разнородных факторов, влияющих на такие проблемы, как эффективность деятельности, оценка местоположения нового офиса, выбор линейки новых продуктов и т. п.

Более того, если в одной из комбинаций участвуют деньги, то все остальные факторы можно выразить в деньгах. Когда оценивают инвестиции разной степени рискованности (например, при возможности отрицательной доходности, низкой доходности в случае неблагоприятного развития событий и т. д.) и при этом используют разные показатели доходности (например, внутренняя норма доходности за семь лет, доходность за первый год и т. д.), то иногда полезно объединить все показатели в особый денежный эквивалент (*certain monetary equivalent*, СМЕ). СМЕ инвестиций — это фиксированная особая сумма в долларах, получить которую инвестор счел бы столь же выгодным, как сделать такое вложение.

Предположим, например, что я хочу приобрести вашу долю в капитале девелоперской компании. За эту долю я предлагаю вам пустующий участок в окрестностях Чикаго стоимостью 200 тыс. дол., с которым вы сможете сделать все, что захотите, *или же* 100 тыс. дол. наличными немедленно. Будь вам безразлично, какой из вариантов выбрать, стало бы ясно: вы считаете, что СМЕ инвестиций в участок составляет 100 тыс. дол. Если же вы решите, что покупка участка по цене 200 тыс. дол. — очень выгодная сделка, значит, особый денежный эквивалент для этих инвестиций вы оцениваете, скажем, в 300 тыс. дол. Иными словами, вы убеждены, что вариант с покупкой участка выгоден не менее, чем немедленное получение 300 тыс. дол. Можно было строить десятки компромиссных сочетаний, чтобы прийти к такому выводу, но результат остался бы тем же. Сколько бы переменных у вас ни было и какие бы их сочетания ни рассматривались, вы всегда предпочтете 300-тысячный СМЕ сотне тысяч наличными.

Именно так я помогаю многим своим клиентам определить приоритетность инвестиций в информационные технологии, когда существуют разные риски и возможны разные подходы к оценке доходности. Мы сводим все переменные в один СМЕ, определяя желательное соотношение между каждой переменной и неким особым денежным эквивалентом. Это очень мощный инструмент, позволяющий, например, решить, как свернуть в единый денежный показатель 12 разных параметров качества. Хотя ваш выбор и субъективен, вы, тем не ме-

нее, получаете количественное представление о компромиссах между разными факторами.

Далее мы рассмотрим ситуации, в которых предпочтительные соотношения между факторами необязательно субъективно определяют лица, принимающими решения.

НЕ ЗАБЫВАТЬ О ГЛАВНОЙ ЦЕЛИ: МАКСИМИЗАЦИЯ ПРИБЫЛИ И СУБЪЕКТИВНЫЕ КОМПРОМИССЫ

Очень часто субъективное определение желаемого соотношения факторов не является обязательным. Иногда имеет смысл свести проблеме к задаче максимизации прибыли или стоимости для акционеров. Грамотный аналитик должен уметь составлять статистически достоверную модель электронной таблицы, показывающей зависимость прибыли от процента допускаемых ошибок, пунктуальности и т.д. Все это сводится к одному аргументу: есть главная преференция — прибыль, и значение прочих факторов, таких как производительность труда и качество, целиком определяется тем, как они влияют на нее. При таком подходе необходимость в субъективном определении желаемого компромисса между такими параметрами, как эффективность и удовлетворенность потребителей, качество и количество или образ бренда и доход, полностью отпадает.

Это и есть суть обоснования любых проектов. В ходе процесса на основе нескольких переменных затрат и выгод рассчитывается некий окончательный показатель, например чистая приведенная стоимость или доходность инвестиций. Место для субъективного выбора по-прежнему остается, но сделать его теперь проще, поскольку он касается конечной цели, к достижению которой должна стремиться компания. Если вы придете к единому мнению по поводу этой конечной цели, то выбор желаемых соотношений между разными показателями деятельности (в данном случае качеством, стоимостью, эффективностью и т.д.) может оказаться и несубъективным. Например, факт, что снижение затрат на 1 млн дол. в одной области деятельности столь же предпочтительно, как сокращение издержек на 1 млн дол. в другой, не означает субъективности решения, поскольку и то и другое одинаково влияет на прибыль. Приведем еще три примера того, как специалисты из разных отраслей количественно оценивали вклад некоей формы «эффективности» в достижение конечной цели.

1. Том Бейкуэлл из Сент-Луиса, штат Миссури, — консультант по управлению, специализирующийся на оценке эффективности работы колледжей и университетов. Бейкуэлл замечает: «Люди

десятилетиями говорили, что эффективность измерить невозможно». Он считает, что конечная цель работы учебного заведения — его финансовое благополучие или, по крайней мере, предотвращение финансового краха. Он рассчитывает финансовые коэффициенты для каждой программы, факультета или профессора, сравнивает результаты с коэффициентами других учебных заведений и таким образом их ранжирует. Кое-кто сказал бы, что этот расчет упускает из виду качественные аспекты эффективности труда преподавателей. Но Бейкуэлл считает, что его подход к измерению продиктован обстоятельствами: «Обычно меня зовут тогда, когда все другие возможности уже исчерпаны и колледж переживает финансовые трудности. Они объясняют мне, почему не могут измениться. Они экономят на всем, на чем можно, но не трогают основную статью расходов — затраты на оплату труда». Его прагматичный подход позволяет выяснить много интересного. Бейкуэлл замечает: «Как правило, люди знают, кто работает плохо, но иногда это становится для них сюрпризом».

2. Гуру менеджмента в области информационных технологий Пол Страссман рассчитывает «доходность управления» путем деления «стоимости, добавленной управлением», на зарплату, бонусы и льготы управляющих⁵. Стоимость, добавленную управлением, он определяет, вычитая из дохода затраты на закупки, налоги, привлечение капитала и другие расходы, не зависящие, по его мнению, от качества управления. Страссман утверждает, что стоимость, добавленная управлением, — это годовая сумма в долларах, величина которой напрямую зависит от политики менеджмента. Даже если не согласиться с тем, какие статьи расходов следует вычитать из дохода компании, чтобы получить данный показатель, этот подход кажется вполне логичным: качество руководства должно проявляться в финансовых показателях компании.
3. Билли Бин, менеджер бейсбольной команды Окленда, решил отказаться от традиционных показателей результативности игроков в бейсбол. Наиболее важным показателем нападающих стала просто вероятность не получить «аут». А показателем результативности игроков обороны стало «генерирование аутов» противника. Улучшение и тех и других показателей способствовало улучшению конечного показателя — вклада игрока в повышение вероятности выигрыша команды, соотнесенного с его зарплатой. На уровне всей команды данный показатель превращается просто в затраты на одну победу. Команда Окленда тратила на одну победу всего 500 тыс. дол., в то время как другие команды — более 3 млн дол.⁶

Во всех описанных случаях тем, кто принимал решение, пришлось изменить свое представление о смысле слова «эффективность». Методы, предложенные Бейкуэллом, Страссманом и Бином, наверное, натолкнулись на сопротивление всех, кто хотел, чтобы эффективность была более качественным показателем. Должно быть, их критики утверждали, что некоторые предложенные методы чересчур упрощены и упускают из виду слишком много важных факторов. Но что означает эффективность, как не поддающийся количественному определению вклад в достижение конечных целей организации? Как эффективность может быть высокой, если этот вклад по сравнению с затратами невелик? Мы уже неоднократно видели: ключ ко всему — уточнение объекта измерения. Поэтому, что бы вы ни имели в виду под «эффективностью», любая детализация ее истинного значения может, как и в этих трех примерах, подсказать нечто очень важное.

■ ПРИМЕЧАНИЯ

- ¹ *Andrew Oswald*. Happiness and Economic Performance // *Economic Journal*, 1997, vol. 107, p. 1815–1831.
- ² *James Hammitt*. Valuing Health: Quality-Adjusted Life Years or Willingness to Pay? // *Risk in Perspective*, Harvard Center for Risk Analysis; *J. K. Hammitt, J. D. Graham*. Willingness to Pay for Health Protection: Inadequate Sensitivity to Probability? // *Journal of Risk and Uncertainty*, 1999, vol. 18, № 1, p. 33–62.
- ³ *Douglas Hubbard*. Risk vs. Return // *Information Week*, 1997, June 30.
- ⁴ *Douglas Hubbard*. Hurdling Risk // *CIO Magazine*, 1998, June 15.
- ⁵ *Paul A. Strassmann*. The Business Value of Computers: An Executive Guide, 1990.
- ⁶ *Michael Lewis*. MoneyBall. New York: W. W. Norton & Company, 2003.



Решающий инструмент измерения: людские суждения

У человеческого разума действительно есть ряд замечательных преимуществ над обычными механическими инструментами измерения. Он обладает уникальной способностью оценивать сложные и неоднозначные ситуации, в которых другие средства бесполезны. Любой пятилетний ребенок справится с задачей распознавания лица или голоса человека в толпе, но разработчикам никак не удастся научить этому программное обеспечение (хотя определенный прогресс в этом направлении уже и достигнут). И мы еще очень далеки от создания искусственного интеллекта, способного написать рецензию на кинофильм или бизнес-план. Человеческий разум — действительно непревзойденный инструмент истинно объективного измерения. Точнее, он был бы таковым, если бы не многочисленные допускаемые человеком систематические ошибки и заблуждения.

Не секрет, что человеческий мозг не просто машина для вычисления. Это сложная система, познающая окружающую среду и приспособляющаяся к ней путем выработки разнообразных упрощающих правил. Практически все эти правила приносят правду в жертву простоте, а многие даже противоречат друг другу. Те, что не вполне обоснованны, но, тем не менее, полезны на практике, называются эвристикой. А те из них, что явно противоречат здравому смыслу, называются заблуждениями.

Если мы все же надеемся использовать человеческий разум в качестве инструмента измерения, то должны придумать, как усилить присущие ему преимущества и в то же время нейтрализовать погрешности. Последствия излишней самоуверенности экспертов устраняет калибровка вероятностей; другие характерные для людских суждений виды систематических ошибок и искажений устраняют специальные методы, особенно эффективные, когда необходимо высказать много мнений по аналогичным вопросам. Примерами могут служить оценка затрат на реализацию новых проектов по информационным технологиям, определение рыночного потенциала новых продуктов либо аттестация работников. Только человеческие суждения позволяют учесть все качественные факторы при проведении этих измерений, однако людям всегда нужна помощь.

ХОМО АБСУРДУС: СТРАННЫЕ МОТИВЫ НАШИХ РЕШЕНИЙ

Упомянутые в главе 8 отклонения — лишь отдельные разновидности ошибок измерения. Отклонения связаны с погрешностями наблюдения, возникающими при проведении случайной выборки или управляемого эксперимента. Но когда измерение пытаются осуществить с привлечением экспертов-оценщиков, то возникает другая проблема — проблема когнитивного искажения. Мы уже видели пример такого искажения, когда говорили о присущей экспертам излишней самоуверенности, но есть и другие — некоторые из них перечислены ниже.

- *Заикленность.* Это когнитивное искажение, уже обсуждавшееся в главе 5 по калибровке, но заслуживающее более детального рассмотрения. Оказывается, что если просто думать о какой-то цифре, *пусть и не имеющей отношения* к вопросу, то это может повлиять на ваш ответ. В одном эксперименте Амос Тверски и лауреат Нобелевской премии по экономике 2002 г. Дэниел Канеман спросили у испытуемых, какой процент стран — членов ООН составляют африканские государства. Одну группу респондентов спросили, превышает ли этот показатель 10%, а вторую — 65%. И тем и другим сообщили, что число, предложенное в вопросе, было выбрано наугад (хотя на самом деле это было не так). Затем каждую группу попросили дать свой ответ. Представители первой (10%) в среднем дали ответ 25%, а члены второй группы (65%) — 45%. Несмотря на уверенность респондентов в случайности предложенных им значений, эти числа все же повлияли на ответы. В своем более позднем эксперименте Канеман показал, что число, на котором заикливаются

участники, может не иметь к делу вообще никакого отношения. Он попросил каждого испытуемого написать на бумаге четыре последние цифры номера своего полиса социального страхования, а затем оценить количество врачей в Нью-Йорке. Как ни странно, Канеман обнаружил корреляцию в 0,4 между оценкой числа врачей и цифрами полиса социального страхования. Эта корреляция была умеренной, но намного превышала объясняемую чистой случайностью.

- *Эффекты ореола.* Если люди сначала замечают некий факт, который настраивает их за или против какого-то варианта решения, то, получив в дальнейшем новую информацию, каким бы ни было ее содержание, они обычно воспринимают ее как аргумент в пользу сделанного ими первого заключения. Например, если при первом контакте у вас сразу же сложилось положительное мнение о человеке, то вы, скорее всего, истолкуете в позитивном свете и любые новые сведения, которые получите о нем позже (эффект ореола святости). А в случае отрицательного впечатления истолкуете эти сведения негативно (эффект дурной славы). Подобные эффекты возникают даже тогда, когда первоначальное ощущение, казалось бы, не должно влиять на оценки, сделанные впоследствии. В эксперименте, проведенном Робертом Капланом из Университета штата Сан-Диего, привлекательная внешность авторов статей заставляла рецензентов давать об их работах лучшие отзывы¹. Испытуемых просили оценить статьи, написанные студентами, причем к каждому экземпляру прилагалась фотография автора. Присвоенные работам оценки сильно коррелировали с баллами внешней привлекательности, определенными особой группой экспертов. Интересно, что всем испытуемым предложили на рецензию статью, написанную одним и тем же человеком, но с приложением разных фотографий.
- *Стадный эффект, или эффект группового давления.* Оказывается, что если вы хотите выяснить мнение экспертов, то лучше опрашивать их по отдельности, а не всех вместе, поскольку в последнем случае возникает дополнительная погрешность. В 1951 г. психолог Соломон Эш предложил группе испытуемых (студентов) проверить их зрение (см. рис. 12.1). Когда он спрашивал каждого в отдельности, длина какой линии ближе всего к контрольному отрезку, то правильный выбор (отрезок С) сделали 99% испытуемых. Одновременно Эш провел опыт: в комнате находились несколько человек, и каждого из них по очереди просили сделать этот выбор. Однако испытуемые не знали, что несколько первых опрашиваемых сами были участниками экс-

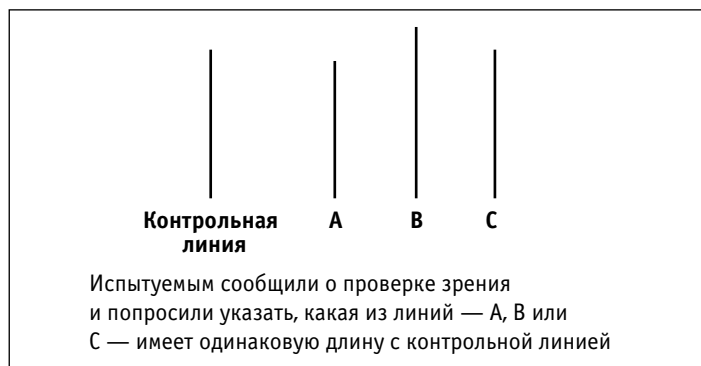


РИСУНОК 12.1. Тест Эша на склонность соглашаться с окружающими

перимента и их проинструктировали выбрать линию А вместо С. После первого, заведомо неверного, ответа наступала очередь реального испытуемого. Если он слышал перед этим один ложный ответ, то только в 97% случаев делал правильный выбор. Когда перед испытуемым ложный ответ давали двое или трое, действительные участники делали правильный выбор в 87 и 67% случаев соответственно. Если же к неверным первым ответам добавлялось некое групповое поощрение (усиливая тем самым давление к проявлению конформизма) на случай, что все участники ответят правильно, только 53% реальных респондентов выбирали нужную линию С.

- *Изменение предпочтений.* Как только люди начинают склоняться к одному варианту, они фактически меняют свои предпочтения по поводу дополнительной информации. Они настраиваются на те сведения, которые поддерживают их первоначальное решение. Это похоже на эффекты ореола святости и дурной славы, но в данном случае связано со сменой предпочтений в процессе принятия решения. Например, если менеджеры решили предпочесть проект А проекту В, а вы скажете им после того, как этот выбор сделан, что первый проект менее рискован, но на его реализацию уйдет больше времени, чем на второй, то вам ответят, что снизить риск для них всегда было важнее, чем сократить сроки реализации. А если вы доложите, что проект А более рискован, но на его реализацию уйдет меньше времени, то получите реплику: всегда считалось, что сократить сроки реализации важнее, чем снизить риск.

К счастью, подобные иррациональные эффекты человеческой способности к оценке всегда можно как-то нейтрализовать. Ряд решений этой проблемы предложил ведущий специалист по когнитивному искажению Джей Эдвард Руссо из Корнелльского университета. Например, для устранения эффекта изменения предпочтений Руссо предложил простую форму слепого перебора: прежде, чем приступить к оценке предлагаемых вариантов, эксперты должны были сформулировать и ранжировать свои предпочтения. Тогда никто уже не смог бы сказать позже, что всегда считал какой-то критерий важнейшим только для того, чтобы не менять свое первое решение.

Как и в случаях уже обсуждавшихся ранее ошибок эксперимента и смещений выборок, первое, что необходимо сделать, — признать само существование проблемы. Только подумайте о том, как перечисленные выше эффекты могут сказаться на экспертных оценках затрат на реализацию проекта, объемов будущих продаж, возможного роста производительности труда и т.п. Специалисты обычно не ощущают, что на их мнение повлияла не имеющая к делу информация, — ведь люди вообще редко осознают, что повинны в искажении. Всем нам хотелось бы думать, что мы не так интеллектуально лабильны, как объекты этих исследований, но лично я обнаружил, что наиболее внушаемы как раз те, кто уверен в непоколебимости своих суждений.

ЗНАЧЕНИЕ СИСТЕМАТИЗАЦИИ:

ПРИМЕР ОЦЕНКИ

ЭФФЕКТИВНОСТИ ДЕЯТЕЛЬНОСТИ

Казалось бы, декан факультета информатики и принятия решений Университета штата Иллинойс (Чикаго) должен предпочитать сложные количественные методы для оценки буквально любого объекта. И все же когда доктору Аркалгуду Рамапрасадугу потребовалось измерить эффективность деятельности преподавателей факультета, он придумал довольно простой подход. «Раньше аттестационные комиссии копались в кучах бумаг, — говорит доктор Рам (он предпочитает, чтобы его называли так). — Члены комиссии усаживались за столом, заваленным личными делами преподавателей, и обсуждали их работу». Публикации, полученные гранты, сделанные каждым сотрудником предложения, присвоение профессиональных званий и степеней обсуждались в произвольном порядке и оценивались по пятибалльной шкале. Подобный бессистемный подход использовался для принятия таких важных решений, как повышение заработной платы профессорско-преподавательскому составу.

Доктор Рам понимал, что главным недостатком этой процедуры является неупорядоченность представляемой информации и что лю-

бое усовершенствование данного процесса, даже простая систематизация данных, может дать большой положительный эффект. Чтобы исправить положение, он проанализировал подаваемые на аттестацию сведения о работе преподавателей и представил их в виде большой матрицы. Каждая ее строка содержала данные об одном преподавателе, а каждый столбец показывал отдельную категорию профессиональных достижений (публикации, награды и т.д.).

Ученый даже не пытался далее формализовать анализ этих данных, он по-прежнему пользовался пятибалльной шкалой. Оценки эффективности в баллах основаны на консенсусе мнений членов аттестационной комиссии, а новый метод просто гарантирует, что они изучают одни и те же сведения. Мне этот метод показался слишком простым, и когда я предложил рассчитывать на основе этих данных какие-нибудь показатели, Рам ответил: «Когда информация представлена в таком явном виде, люди сразу обращают внимание на разницу между собой и своими коллегами, а это совсем не одно и то же, что попытка разобраться в неких условных показателях. Комиссия спорит о присваиваемых баллах, но не о предложенных к рассмотрению данных». Когда ранее ее членам приходилось анализировать разнородную информацию, в их оценки вкрадывалось больше ошибок.

Это еще один полезный пример конструктивного разностороннего подхода к измерению. Наверное, нашлись бы возражения против самой идеи оценивать результаты деятельности преподавателей на основании того, что новый метод принесет с собой новые ошибки и не решит проблему исключений. Не менее вероятно и то, что на самом деле критиков данного подхода беспокоит вероятность оказаться на последнем месте в случае использования подобного подхода. Но доктор Рам знает, что при всех своих недостатках новый метод измерения все же лучше того, что делалось до сих пор. Ведь неопределенность снижается, а значит, то, что он делает, — это измерение. Сегодня доктор Рам в соответствии с таксономией Стивенса (см. главу 3) может, по крайней мере, с определенной уверенностью сказать, что преподаватель А работает лучше преподавателя В. А если учесть, что эти оценки используются для принятия решений о продвижении по службе или повышении зарплаты, то большего и не требуется.

Моя единственная претензия к этому подходу — возможность (и это было бы нетрудно) использовать более аналитический метод расчета и тем самым усовершенствовать процедуру оценки. Доктор Рам не решил ни одну из обсуждавшихся нами проблем когнитивного искажения; он только устранил потенциальный информационный «шум» и погрешность, связанную с анализом разнородной информации о преподавателях. Вот почему я считаю, что систематизация данных — всего лишь необходимое условие применения других способов решения этих проблем.

НА УДИВЛЕНИЕ ПРОСТЫЕ ЛИНЕЙНЫЕ МОДЕЛИ

Существует еще один метод, не самый теоретически обоснованный и даже не самый эффективный, но простой. Когда приходится формулировать суждения по аналогичным вопросам, рассчитывают взвешенные значения. Если сравнивают «деловые возможности», например варианты инвестиций в недвижимость, то можно определить несколько наиболее важных для вас факторов, оценить эти факторы для каждого варианта в баллах и объединить полученные результаты в некую агрегированную величину. Для инвестиций в недвижимость такими критериями могут быть желательное местоположение, затраты, возможный рост спроса на этот вид недвижимости, наличие залогов и т.д. Затем следует «взвесить» каждый фактор путем умножения его баллов на определенный весовой коэффициент и суммировать все результаты, чтобы получить общее значение.

Одно время я категорически отрицал ценность метода взвешенных коэффициентов, уподобляя его астрологии. Однако последующие исследования убедили меня, что он все-таки имеет определенные преимущества. К сожалению, методы, обладающие, на первый взгляд, некими преимуществами, не относятся к тем, которые обычно выбирают компании.

По мнению исследователя в области науки о принятии решений и автора ряда работ Джея Эдварда Руссо, эффективность метода взвешенных коэффициентов «зависит от того, что вы делаете. Людям обычно нужно зайти слишком далеко в своих усилиях, чтобы понять: и простые методы дают хорошие результаты». На самом деле, даже расчет простейших взвешенных коэффициентов, похоже, облегчает процесс принятия решений. В 1979 г. Робин Доуз из Мичиганского университета опубликовал статью под названием «Robust Beauty of Improper Linear Models» («Строгая красота неправильных линейных моделей»)², в которой писал: «Весовые параметры в этих моделях нередко не имеют значения. Главное — знать, что измерить, а затем сложить».

Здесь необходимо сделать два уточнения. Во-первых, опыт доктора Рама в области оценки эффективности преподавателей вполне согласуется с тем, что говорят Руссо и Доуз. Ранее использовавшиеся в университете методы давали такую погрешность, что одна только систематизация исходных данных уже способствовала улучшению измерений. Кроме того, когда Доуз говорит о коэффициенте, он на самом деле говорит о нормированном z-показателе, а не о балле какой-то условной шкалы. Он берет значения одного параметра для всех оцениваемых вариантов и строит их нормированное распределение так, что его среднее значение равно нулю, а каждая величина преобразовывается в ряд средних квадратичных отклонений от среднего в ту или

иную сторону (например, $-1,7$, $+0,5$ и т.д.). Доуз может, например, взять из матрицы доктора Рама число публикаций преподавателя и проделать с этими данными следующие процедуры:

1. Всем значениям столбцов в матрице оцениваемых альтернатив присвоить баллы по какой-либо порядковой или количественной (метрической) шкале. Заметьте, что предпочтительнее использовать количественные шкалы с вещественными единицами измерения (например, доллары, месяцы).
2. Рассчитать среднее значение для всех величин каждого столбца.
3. Использовать формулу Excel `=stdevp(.)` расчета среднего квадратичного отклонения для генеральной совокупности каждого столбца.
4. Рассчитать z-показатель, соответствующий каждому значению в столбце, по формуле:

$$z = \frac{\text{Значение} - \text{Среднее значение}}{\text{Среднее квадратичное отклонение}}.$$

5. В результате получаем средний балл, равный 0, нижнюю границу в пределах -2 или -3 , а верхнюю границу $+2$ или $+3$.

Причиной работоспособности данного подхода является то, что он следит за правильностью выбора весовых коэффициентов. Если не пересчитывать присваиваемый балл в z-показатель, то вы можете использовать для одного фактора более высокое значение, чем для другого, а это окажет такой же эффект, как если бы вы изменили их относительные веса. Предположим, например, что вы оцениваете проекты инвестирования в недвижимость и оцениваете каждый фактор по десятибалльной условной шкале. Однако один из оцениваемых факторов (желаемое местонахождение) довольно сильно варьирует и вы склонны присвоить ему 7 или 8 баллов, в то время как критерию возможного роста спроса на недвижимость дали 4 или 5 баллов. В результате, даже если вы считаете, что рост спроса важнее, желательное местонахождение перевесит. Предложенное Доузом преобразование баллов в z-показателе решает проблему неизбежного искажения при взвешивании.

Хотя этот простой метод и не решает напрямую ни одну из перечисленных нами проблем когнитивного искажения, исследования Доуза и Руссо показали, что принимать решения с его помощью становится легче, пусть и ненамного. Похоже, что данный подход обеспечивает, по крайней мере, некоторое снижение неопределенности и улучшение качества принимаемых решений. Однако для принятия серьезных и рискованных решений, когда стоимость информации очень высока, мы можем и должны использовать гораздо более сложные приемы, чем банальная систематизация данных и расчет взвешенных коэффициентов.

КАК СТАНДАРТИЗИРОВАТЬ ЛЮБУЮ ОЦЕНКУ: МОДЕЛИ РАША

Выбирая из огромного разнообразия имеющихся статистических методов те, о которых следовало бы рассказать в этой книге, я старался не ограничиваться уже ранее знакомыми мне. В частности, для меня новинкой стали приемы дидактического тестирования (*educational testing*), практически неизвестные специалистам по измерению, работающим в других областях. Именно таким приемам посвящена книга с многозначительным названием «Objective Measurement» («Объективное измерение»). Подобная публикация могла быть всесторонним исследованием проблем измерения, одинаково интересным для астронома, инженера-химика и программиста, но в ней говорится лишь об оценке способностей человека и дидактическом тестировании. Это все равно, как если бы вы увидели старую карту с названием «Карта мира», на которой на самом деле изображен богом забытый островок в Тихом океане и которая была составлена людьми, не знавшими, что населяют лишь крошечную часть огромной планеты. Один специалист по дидактическому тестированию как-то рассказал мне об «инвариантном сравнении» — характерной черте измерения, которая, по его словам, «настолько фундаментальна, что ее можно считать на 10% основой статистики». Другой эксперт в этой же области уверял, что это суть физики. Из всех физиков и статистиков, которых я позднее спрашивал об «инвариантном сравнении», что-то слышал о нем только один. По всей видимости, то, что специалисты по дидактическому тестированию считают общей для всех фундаментальной основой, на самом деле имеет значение только для них. Справедливости ради отмечу, что, наверное, кто-то скажет то же самое и о книге, претендующей научить читателя измерять все, что угодно.

Но вообще говоря, у специалистов по дидактическому тестированию есть чему поучиться. Ведь им приходится решать все проблемы по оценке возможностей человека — большую группу задач измерения того, что многие компании считают неизмеряемым. Концепция инвариантного сравнения имеет отношение к ключевой проблеме проведения многих тестов на определение способностей человека, например теста на IQ. Принцип инвариантного сравнения гласит, что если один инструмент измерения говорит, что А больше, чем В, то и другой инструмент должен продемонстрировать то же самое. Иными словами, результаты сравнения А и В должны быть одинаковыми, какие бы инструменты измерения ни использовались. Для физика это очевидно настолько, что даже не заслуживает упоминания. Казалось бы, если на одних весах А весит больше, чем В, то и другие весы должны показывать то же самое, даже если первые весы пружинные, а вторые — цифровые. Аналогичные результаты должны наблюдаться с IQ-тестами

или любыми другими тестами, оценивающими человеческую эффективность. Однако при проведении таких испытаний может произойти нечто совсем иное: результаты теста на IQ с одним набором вопросов могут сильно отличаться в выводах от результатов аналогичного теста с другим набором вопросов. Поэтому один тест может показать, что Боб умнее Шерри, а второй — что Шерри умнее Боба.

Другое проявление той же проблемы наблюдается, когда разным экспертам приходится оценивать способности больших коллективов. Если аттестуемых слишком много, то и их делят на группы, и каждый человек имеет дело с разным составом судей. Возможно, один эксперт оценивает одно качество одного аттестуемого, а другой обращает внимание совсем на другое или же разным людям даются задачи разной степени сложности. Предположим, например, что вы хотите оценить квалификацию менеджеров проекта по тому, как они справятся с порученными им программами. Если таких менеджеров несколько, то вам потребуется не один судья. Экспертами могут стать непосредственные начальники аттестуемых менеджеров как люди, наиболее знакомые с их работой. Отметим, что проекты могут различаться по сложности. Предположим далее, что все менеджеры, независимо от того, какие проекты им поручены и кому они подчиняются, должны конкурировать за один фонд бонусов или за одни и те же карьерные места. Те, кого будет оценивать строгий судья или кому поручены более сложные проекты, окажутся в невыгодном положении по сравнению с другими. Сравнение разных менеджеров не будет инвариантным (то есть независимым от того, кто его проводит и по каким критериям). На самом деле, доминирующими решающими факторами, определяющими рейтинг проект-менеджеров, могут оказаться условия, не контролируемые ими.

Решение этой проблемы в 1961 г. предложил датский статистик Георг Раш³. Он разработал метод предсказания вероятности того, что испытуемый правильно ответит на вопрос бинарного типа «верно/неверно» на основе 1) процента в генеральной совокупности других респондентов, ответивших на этот вопрос правильно, и 2) процента других вопросов, на которые данный испытуемый уже ответил правильно. Выполнение тестов человеком, не знакомым с предлагаемым набором вопросов, можно предсказать с погрешностью, поддающейся вычислению, даже если давать испытуемым разные тесты.

Во-первых, Раш рассчитал вероятность того, что случайно выбранный из группы человек ответит на вопрос правильно. Она просто равна доле тех, кто дал верный ответ, в общем числе ответивших на данный вопрос. Этот показатель называется у Раша уровнем трудности задания (*item difficulty*). Затем Раш рассчитал логарифм отношения вероятностей — натуральный логарифм отношения вероятностей правильного и неправильного ответов. Если трудность задания была 65%, то это означает, что 35% респондентов ответили правильно, а

65% — неправильно. Отношение вероятности ответить правильно к вероятности ответить неправильно — 0,548, а натуральный логарифм 0,548 составляет $-0,619$. При желании можно записать следующую формулу в программе Excel:

$$=\ln(A1/(1 - A1)),$$

где $A1$ — вероятность ответить правильно.

Затем Раш проделал ту же процедуру с вероятностью, что этот человек ответит правильно на любой вопрос. Поскольку данный респондент давал правильные ответы в 82% случаев, соответствующий логарифм составил $\ln(0,82/0,18)$, или 1,52. Наконец, Раш сложил значения двух логарифмов и получил: $(-0,619) + 1,52 = 0,9$. Чтобы снова преобразовать это в вероятность, можно записать следующую формулу в Excel:

$$=1/(1/\exp(0,9) + 1).$$

В результате получится 71%. Это означает, что есть 71-процентная вероятность того, что данное лицо ответит на этот вопрос верно, учитывая трудность задания и правильность его ответов на другие вопросы. При большом числе вопросов и (или) большом числе испытуемых мы обнаружим следующее: когда вероятность получить правильный ответ (уровень трудности задания) 70%, около 70% людей ответят на этот вопрос верно; когда вероятность получить правильный ответ (уровень трудности задания) 80%, около 80% людей ответят на данный вопрос верно, и т.д. Таким образом, модели Раша — просто еще один способ калибровки вероятностей.

Мэри Лунц из чикагской компании Measurement Research Associates Inc. применила модели Раша к решению важной задачи в области общественного здравоохранения, порученной ей Американским обществом клинической патологии (American Society of Clinical Pathology). Использувавшийся этим обществом ранее порядок сертификации патологов давал большую погрешность, которую необходимо было уменьшить. Каждый кандидат должен был разобраться в одном или двух случаях, и каждый его ответ оценивался одним или несколькими экспертами. Практически невозможно одному эксперту оценить все задания, как невозможно гарантировать, что все они будут одинаковой сложности. Раньше получение кандидатом сертификата почти целиком зависело от того, какой экзаменатор ему попадется и какой случай придется разбирать на экзамене. Иными словами, снисходительные экзаменаторы могли пропустить некомпетентных кандидатов. Лунц рассчитала стандартные очки Раша для каждого эксперта, экзаменационного задания, а также кандидата для всех уровней квалификации.

В результате появилась возможность предсказать, пройдет ли кандидат экзамен у среднего экзаменатора при случае средней сложности, или у снисходительного эксперта при легком случае, или, наоборот, у строгого экзаменатора при очень сложном случае. И теперь (наверняка не слишком скоро для самих кандидатов) наконец появилась возможность полностью устранить при сертификации отклонения, связанные с характером экзаменатора или сложностью задания.

ОЦЕНКА УМЕНИЯ ЧИТАТЬ С ПОМОЩЬЮ БАЛЛОВ РАША

Интересное направление применения статистики Раша — оценка сложности прочтения того или иного текста. Доктор Джек Стеннер, президент и основатель компании MetaMetrics, Inc., использовал модели Раша для разработки схемы «Lexile» — способа оценки умения читать и писать, а также анализа сложности текста для восприятия. Система «Lexile» позволяет оценивать навыки чтения, письма, выполненные тесты, тексты и способности студентов, впервые сделав возможным сравнение наиболее распространенных языков. Располагая персоналом всего в 56 человек, компания MetaMetrics добилась в этой сфере гораздо больше, чем любая другая государственная или частная организация. Так:

- баллы «Lexile» используются во всех основных тестах на умение читать. Свои показатели «Lexile» знают около 20 млн американских учащихся;
- система «Lexile» использована для классификации по сложности восприятия около 100 тыс. книг и десятков миллионов журнальных статей;
- программы обучения чтению по учебникам некоторых авторов основаны на системе «Lexile»;
- на систему «Lexile» переходят все новые учебные заведения штатов и местные учебные заведения.

Показатель 100 баллов по системе «Lexile» означает текст первого уровня сложности, а 1700 — это уровень сложности текста решений Верховного суда, научных журналов и т. п. Компания MetaMetrics считает, что читатель, набравший 600 баллов, сумеет на 75% понять содержание текста, уровень сложности которого составляет также 600 баллов.

УСТРАНЕНИЕ НЕПОСЛЕДОВАТЕЛЬНОСТИ ЛЮДСКИХ СУЖДЕНИЙ: МОДЕЛЬ ЛИНЗЫ

В 1950-х годах психолог по имени Эгон Брунстик захотел статистически измерить принимаемые экспертами решения⁴. Большинство его коллег интересовались тем скрытым процессом принятия реше-

ний, через который обычно проходят такие эксперты. А Брунsvику хотелось описать те решения, которые они принимали в реальности. О себе и других специалистах по психологии принятия решений он говорил: «Мы должны быть не столько геологами, сколько картографами». Иными словами, свою задачу он видел в простом описании того, что можно наблюдать, а не в анализе внутренних процессов. В связи с этим Брунsvик начал свои эксперименты, в которых экспертам предлагалось принять какое-то решение (скажем, о приеме выпускника в аспирантуру или о статусе опухоли) на основании определенной представленной информации. Затем Брунsvик подобрал наиболее подходящую регрессионную модель для большого числа собранных экспертных оценок (сейчас это можно легко проделать с помощью инструмента «Regression» в программе Excel, как показано в главе 9). В результате он вывел нечто вроде формулы с набором неявных весов, осознанно или неосознанно использованных экспертами при вынесении оценок.

Удивительно, что он также обнаружил, что эта «формула», хотя в ней использовались вовсе не объективные данные прошлых периодов, а экспертные суждения, позволяет получить более точные оценки, чем сделанные специалистами. Например, эта формула, получившая известность под названием «модель линзы», лучше эксперта определяет, кто будет хорошо учиться в аспирантуре или какая опухоль является злокачественной.

Модель линзы применяется в самых разных целях, например для составления медицинских прогнозов, идентификации самолетов операторами корабельных радаров и расчета вероятности краха компании по ее финансовым показателям. В каждом случае результаты, полученные с помощью модели, ничем не хуже, а в большинстве случаев и значительно лучше, чем оценки экспертов.

Это происходит потому, что модель линзы устраняет непоследовательность в суждениях. Обычно экспертные оценки различаются даже в одинаковых ситуациях. Однако линейная модель экспертной оценки позволяет получать не противоречащие друг другу величины.

Более того, поскольку модель линзы — математическое выражение, элементами которого являются известные исходные данные, можно компьютеризировать и обрабатывать такие объемы информации, которые люди не смогли бы проанализировать по очереди.

Сама семиэтапная процедура расчетов довольно проста. Я слегка изменил ее, чтобы учесть и другие методы (например, калибровку вероятностей), ставшие известными после того, как Брунsvик разработал свой способ (см. рис. 12.3).

1. Выберите экспертов, которые будут участвовать в процессе.
2. Если им придется оценивать вероятность или интервал значений, то калибруйте их.

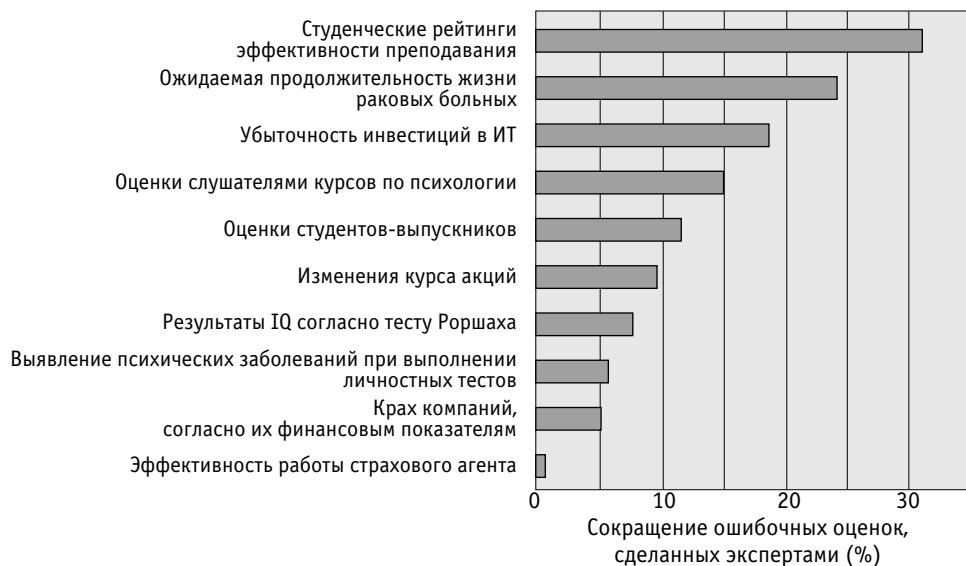


РИСУНОК 12.2. Модель линзы как инструмент принятия более удачных решений



РИСУНОК 12.3. Применение модели линзы

3. Попросите экспертов составить список (не более чем из 10 пунктов) факторов, требующих учета при вынесении оценки (например, «продолжительность реализации проекта разработки программного обеспечения повышает риск неудачи» или «уровень доходов лица, обращающегося за ссудой, влияет на вероятность погашения им взятого кредита»).
4. Разработайте ряд сценариев с разными сочетаниями значений каждого из выявленных факторов воздействия. За основу можно взять и реальные, и чисто гипотетические примеры. Составьте по 30–50 сценариев для каждого эксперта.
5. Попросите экспертов дать оценку каждого сценария.
6. Проведите регрессионный анализ, следуя указаниям, изложенным в главе 9. Независимые переменные «X» — предоставленные экспертам исходные данные. Зависимая переменная «Y» — оценка, которую должен был дать эксперт.
7. Программа Excel создаст таблицу результатов, в которой вы найдете коэффициент для всех столбцов данных из ваших сценариев. Найдите коэффициент, соответствующий каждой переменной, умножьте ее на этот коэффициент, а затем суммируйте все полученные таким образом произведения. Это и есть величина, которую вы пытаетесь измерить.

Результатом процедуры, описанной выше, является таблица с весами для всех переменных, входящих в модель. Поскольку данная модель внутренне непротиворечива, мы знаем, что уменьшили ошибку, по крайней мере частично.

Быстро выяснить, насколько модель линзы снижает неопределенность, можно, оценив непоследовательность экспертных суждений. Для этого нужно предложить экспертам дублирующие друг друга сценарии, что должно остаться для них в тайне. Иными словами, седьмой и двадцать девятый сценарии в списке могут быть одинаковыми. Изучив два десятка сценариев, люди забудут, что уже знакомы с этой ситуацией, и вполне могут дать отличающийся ответ. Вдумчивые эксперты обычно последовательны в своих суждениях о сценариях. Тем не менее именно непоследовательностью объясняется 10–20% ошибок большинства экспертных оценок, их полностью устраняет метод линзы.

Поборник простых, неоптимизированных линейных моделей Робин Доуз согласен с тем, что Брунsvик добился значительного улучшения по сравнению с обычными суждениями экспертов. Вместе с тем он утверждает, что такой результат вовсе не связан с определением «оптимальных» весов при помощи регрессии. В своей статье Доуз на четырех конкретных примерах показал, что модель линзы дает лишь небольшое улучшение по сравнению с моделями, которые он назвал «неправильными», где веса факторов не выводятся из регрессии, а

считаются одинаковыми или, как это ни странно, приписываются *случайным образом*⁵.

Доуз пришел к выводу, что главная ценность экспертов заключается в том, что они определяют подлежащие учету факторы и относят их к «хорошим» или «плохим» (то есть решают, будут ли их веса иметь знак «плюс» или «минус») и что рассчитывать точные значения этих весов с помощью регрессии вовсе не обязательно.

Приведенные Доузом примеры, возможно, и не позволяют судить об эффективности модели линзы как инструмента решения задач, возникающих в бизнесе⁶, но его выводы все равно полезны. Во-первых, собственные данные Доуза доказывают определенное преимущество, пусть и небольшое, оптимальных линейных моделей над «неправильными» моделями. Во-вторых, его выводы подтверждают ту мысль, что некая непротиворечивая модель (с оптимизированными весами или без них) лучше, чем единственно человеческое суждение эксперта. И все же я думаю, что усилия по созданию оптимальных моделей, особенно когда нужно принять действительно важное решение, вполне оправдывают даже то небольшое улучшение, которое они обеспечивают по сравнению с более простыми моделями.

Однако я убежден, что мы часто добиваемся лучших результатов, чем даже «оптимальные» линейные модели. Регрессионные модели, используемые мной для бизнеса, обычно подчиняются нескольким правилам, например такому: «Продолжительность реализации проекта является фактором дифференциации, только если она превышает год. Все проекты, реализуемые в течение года и менее, одинаково рискованны». В этом смысле такие модели не вполне линейны, но позволяют выявлять более тесную корреляцию, чем строго линейные модели линзы. Все модели, о которых Доуз упоминает в своей статье, строго линейны, но, как правило, дают более низкие значения корреляции, чем те, что я получаю с помощью нелинейных моделей.

Одним правилам меня научили эксперты, другие я сформулировал сам, проанализировав их оценки. Например, если специалист, анализирующий вероятность существенного расширения содержания разрабатываемого программного обеспечения, говорит мне, что не проводит грань между проектами, продолжительность которых составит менее года, то я просто не использую в качестве переменной исходную «продолжительность проекта». Взамен я так изменяю эту функцию, чтобы любое значение продолжительности до 12 месяцев было равно 1, 13 месяцев — 2, 14 месяцев — 3 и т.д. Но если эксперт мне этого и не скажет, то я догадаюсь обо всем по его оценкам. Предположим, что мы нанесли экспертные оценки на график зависимости вероятности значительного (требующего, скажем, увеличения объема работ более чем на 25%) изменения спецификаций от продолжительности реализации проекта (в месяцах). Получим следующую картину (см. рис. 12.4).



РИСУНОК 12.4. Пример нелинейного изменения переменной в модели линзы

Если вам кажется, что геометрическим местом этих точек служит, скорее, пунктир, то вы не одиноки в своем мнении. Оценивая проект, на реализацию которого уйдет больше года, придется учесть другой набор факторов. Возможно, с точки зрения эксперта, одни переменные больше или меньше зависят от продолжительности проекта. Модель линзы, учитывающая эти нелинейные зависимости, не только лучше соответствует мнениям специалистов; еще важнее, что она лучше коррелирует с фактическими результатами.

Бывает также, что удачный выбор переменной требует использования еще более сложных правил. Иногда существует тесная корреляция зависимой переменной не с самой независимой функцией, а с ее логарифмом, с обратной к ней величиной или с ее отношением к произведению других независимых переменных. Эксперименты в этой области только приветствуются. Как правило, я пробую несколько вариантов линейных переменных для одних и тех же исходных данных и обычно обнаруживаю, что один из них явно выигрывает на фоне других.

Оказывается, что вы можете пользоваться моделями взвешенных оценок разной степени сложности. Если не боитесь экспериментировать с нелинейными методами, то вам подойдут именно они. Если это для вас слишком сложно, но вы разбираетесь в линейной регрессии, то применяйте ее. Если вы не знакомы с регрессионным анализом, то в вашем распоряжении z-показатели Доуза с одинаковыми весами. Каждый из этих методов эффективнее другого, более простого, и все они эффективнее, чем обычная оценка эксперта.

ПАНАЦЕЯ ИЛИ ПЛАЦЕБО? СОМНИТЕЛЬНЫЕ МЕТОДЫ ИЗМЕРЕНИЯ

ВАЖНЕЙШЕЕ ПРАВИЛО ИЗМЕРЕНИЯ

Самое главное — никогда не используйте метод, *способный увеличить* ошибку первоначальной оценки.

Кое-кто из читателей может подумать: до сих пор автор пытался снизить планку требований к измерениям настолько, что одно только это сделает измеримым любой объект или явление. Ведь я с самого начала заявил: измерением может считаться все, что снижает неопределенность. Разнообразные ошибки наблюдения — вовсе не препятствие, если только неопределенность после измерения ниже, чем до него. Даже методы анализа того, что обычно считается «субъективным» (например, модель Раша и модель линзы), — тоже измерение, если есть убедительные доказательства, что они действительно позволяют получать более точные оценки. Вместе с тем существуют такие способы, которые я не могу считать измерением даже при этих отнюдь не жестких критериях. Поэтому сейчас, прежде чем перейти к описанию новых методов количественной оценки, я хочу благоразумно притормозить и сделать ряд оговорок.

Раз не нужно больше заботиться о точности измерения, наше определение его как «снижение неопределенности», безусловно, делает выполнимой оценку чуть ли не всего, что угодно. Но это определение предполагает и строгое ограничение. Если какой-то метод на самом деле не уменьшает неопределенность или, более того, ее *увеличивает*, то это нельзя считать измерением и его ценность для принимающих решение равна нулю. В подобном духе запоздалого измеренческого скепсиса мы должны обсудить два весьма распространенных метода измерения: анализ «затраты/выгоды» и метод субъективных взвешенных показателей.

Только приступив к написанию этой книги, я попросил своих многочисленных знакомых сообщать мне об интересных решениях в области измерений, которые я смог бы использовать в качестве примеров из практики. Я сказал, что ищу «интересные примеры удачного решения трудных или даже казавшихся неразрешимыми задач измерения, при этом желательно, чтобы результаты оказались поразительными и заставили изменить первоначальное решение». Недостатка в предложениях не было, и я обсудил по телефону намного больше практических примеров, чем в конце концов включил в свою книгу. Однако я заметил, что многие аналитики, консультанты и бизнесмены, похоже, ставят знак равенства между проведением измерений и обоснованием

проекта. Они не приводили примеров интересного использования результатов наблюдений для снижения неопределенности в связи с каким-то неизвестным показателем. Вместо этого они объясняли мне, как готовили обоснование своего любимого проекта.

Справедливости ради скажу, что анализ «затраты/выгоды» вполне может считаться разновидностью разложения на составляющие, о котором мы говорили в главе 8, и сам по себе способен снизить неопределенность без дальнейших измерений. В процессе обоснования проекта проблему разлагают на составляющие точно так же, как в свое время Ферми это делал своими вопросами. При этом не являясь сам по себе измерением, основанным на наблюдениях, процесс все же позволяет узнать что-то новое о том, с чем мы уже были знакомы. Но, как я отмечал ранее, согласно моему многолетнему опыту, при высокой стоимости информации об изучаемых переменных одно только разложение на составляющие достаточно снижает неопределенность всего в 25% случаев. А чаще всего, если стоимость информации оправдывает усилия по снижению неопределенности, все равно требуются определенные эмпирические наблюдения.

Между тем, похоже, что *единственный прием* измерения, которым пользуются многие компании, — это разложение на составляющие (например, при обосновании проекта); применять эмпирические методы они даже не пытаются. Каждая переменная при этом представляет собой просто первоначальную оценку (одного эксперта либо целой рабочей группы) и всегда выражается конкретным значением, а не диапазоном, свидетельствующим о том, что точная ее величина неизвестна. Никакие опросы и эксперименты не проводятся; не применяются и методы, позволяющие уточнить субъективные суждения. Те, кто с таким энтузиазмом рассказывал мне об обосновании проекта как о примере проведения измерений, не смогли, сколько я ни просил об этом, назвать хотя бы один показатель, который использовался ими при анализе «затрат/выгод» и который был бы рассчитан в результате наблюдений реального мира.

Совсем иное поведение наблюдается, когда задача состоит в том, чтобы на этапе обоснования проекта определить точные значения, особенно когда эксперт участвует в нем и заинтересован в результате. Оно отличается от действий калиброванного эксперта, указывающего первоначальный 90-процентный доверительный интервал. Один или несколько собравшихся для анализа проекта специалистов рассматривают со всех точек зрения каждую оценку. Вынуждаемые обстоятельствами выбрать точные значения, несмотря на всю имеющуюся неопределенность и условность ситуации, они задаются вопросом: «Каким должно оказаться это значение, чтобы оно стало приемлемым для других и в то же время подтверждало правоту моей прежней точки зрения?» Это почти то же самое, как если бы мы использовали терми-

ны «консенсус» и «факт». Обсуждавшийся ранее эксперимент Эша со стадным эффектом — лишь один из недостатков подобного подхода.

Еще одна настораживающая тенденция в принятии решений менеджерами компаний — использование взвешенных показателей такого типа, когда и сами показатели, и присвоенные им веса — субъективные, произвольные величины, а не использовавшиеся Доузом z -значения. Как и обсуждавшиеся ранее простые линейные модели, такие методы могут поставить перед менеджером портфеля проектов задачу ранжировать их по категориям типа «организационный риск» или «стратегическое соответствие» и т. д.

Подобные методы в большинстве своем предполагают использование от 4 до 12 категорий оценки, а некоторые — больше 100. Обсуждаемому проекту обычно присваивают балл, например по пятибалльной шкале, для каждой категории. Сумму баллов затем умножают на весовой коэффициент (иногда тоже составляющий от 1 до 5), отражающий относительное значение данной категории. Обычно в компаниях используемые весовые коэффициенты стандартизируют, чтобы можно было оценивать проекты по сопоставимым критериям. Скорректированные на весовые коэффициенты баллы затем суммируют и получают общий показатель обсуждаемой программы.

Присваивание баллов — способ выражения сравнительного значения, предпочтительности и т. д. без помощи реальных единиц измерения. Хотя подсчет баллов справедливо считают разновидностью порядкового измерения, обсуждавшегося в главе 3, я всегда считал, что он в определенном смысле предполагает подмену действительного желаемым, что вносит дополнительные ошибки по следующим четырем причинам.

1. Нередко баллы используют в ситуациях, где вполне оправдано использование обычных количественных показателей, которые были бы намного информативнее (например, иногда в баллы пересчитывают доходность инвестиций или риск вместо того, чтобы использовать эти параметры, как это сделал бы актуарий или финансовый аналитик).
2. Баллы привносят свой собственный тип ошибки в процесс оценки, так как зачастую показатели, определенные по разным шкалам, несопоставимы и неоднозначны. Рассмотрим, например, число звезд, которое кинокритик может присвоить кинофильму или ресторанный обозреватель — ресторану (в первом случае обычно используют шкалу четыре звезды, а во втором — пять звезд). В данной ситуации две звезды не означают, что продукт или услуга в два раза лучше, чем то же, но с одной звездой, а посещение четырех однозвездочных кинофильмов совсем не равнозначно просмотру одного четырехзвездочного.

3. Баллы бывают информативными, если являются элементами опроса большой группы людей (например, при проведении исследования по определению степени удовлетворенности потребителей). Однако они существенно теряют информативность, если используются индивидуумами для оценки возможностей, стратегий, инвестиций и т.п.: людей редко удивляют те баллы, которые они присваивают сами.
4. Баллы лишь отражают порядковый номер, но многие пользователи увеличивают ошибку тем, что расценивают их как реальные величины. Как уже говорилось, более высокое порядковое число означает «больше», но не показывает, насколько больше. Умножение и суммирование порядковых чисел иногда дает результаты, о которых пользователь и не догадывается. Вот почему этот метод может привести к непредвиденным последствиям.

Стоит подробнее остановиться на том, чем такие баллы отличаются от z-показателей, использованных Робинот Доузом, а веса — от весов, получаемых с помощью модели линзы. Во-первых, в «неправильных» линейных моделях Доуза и оптимизированных моделях линзы Брунсвика применяются исходные данные, выраженные в реальных единицах измерения (например, продолжительность реализации ИТ-проекта в месяцах или средний балл поступающего в аспирантуру), а вовсе не баллы, присвоенные экспертами по некоей произвольно выбранной шкале. Во-вторых, такими баллами не были и веса, использовавшиеся Доузом и Брунsvиком. Психология применения произвольных шкал гораздо сложнее, чем кажется. Когда эксперты выбирают веса по пятибалльной шкале, они вовсе не имеют в виду, что балл 4 вдвое важнее балла 2. Из-за этой неоднозначности пятибалльная (семибалльная или какая угодно другая) шкала только добавляет ошибку к процессу оценки.

На мой взгляд, единственное наблюдаемое достоинство систем условных взвешенных коэффициентов то, что менеджерам обычно хватает здравого смысла игнорировать полученные таким образом результаты. Я обнаружил, что лица, принимающие решения, отвергают итоги подобных расчетов очень часто, и мне не удалось найти ни одного доказательства того, что эти баллы когда-либо способствовали принятию *совершенно иного* или даже просто более удачного решения. И это притом что менеджеры нередко тратят массу времени и сил на разработку и применение своих условных шкал.

Один из таких методов иногда используется в информационных технологиях под ошибочным названием прикладной информационной экономики⁷. Его представляют как объективный, систематизированный и формализованный подход, однако, фактически, он не осно-

ван ни на одной общепризнанной экономической модели и на самом деле не может считаться относящимся к экономике. При ближайшем рассмотрении его название оказывается совершенно неправильным. Гораздо точнее выглядит термин «метод расчета субъективных нескорректированных взвешенных коэффициентов для области ИТ».

Полученный этим методом итоговый показатель для оцениваемого ИТ-проекта не имеет смысла и с точки зрения финансов. Описательное определение баллов в каждой категории и весовые коэффициенты, приписываемые этим категориям, никак не привязаны к научным подходам — ни к теоретическим, ни к эмпирическим. На самом деле это всего лишь еще один метод исключительно субъективной оценки, не предполагающий, в отличие от метода Раша или модели линзы, корректировку ошибки. Многие из тех, кто рассчитывает взвешенные коэффициенты для информационных технологий, утверждают, что метод имеет свои преимущества, но доказательств этому пока нет.

Любопытно, что прикладная информационная экономика предполагает преобразование полезных и имеющих экономический смысл величин (например, ROI) в баллы. Этот процесс происходит следующим образом: отрицательной или нулевой ROI присваивается балл 0, ROI, составляющей от 0,1 до 299% — балл 1, ROI от 300 до 499% — 2 и т.д. Иными словами, скромная 5-процентная ROI обозначается таким же баллом, как 200-процентная. В более количественных методах определения приоритетности инвестиционных проектов подобная разница в доходности означала бы, что один проект намного предпочтительнее другого. А здесь два проекта, существенно и очевидно различающихся по доходности инвестиций, попадают в одну категорию. «Чистый эффект» от такой процедуры — «уничтожение» информации.

Того же мнения придерживается специалист по управлению информационными технологиями Барбара Макнарлин. Она проанализировала 25 разных способов оценки преимуществ, в том числе несколько методов взвешенных коэффициентов⁸. Макнарлин охарактеризовала эти методы как бесполезные, причем ни в одном из них не нашла научно-теоретической основы.

Пожалуй, лучше всего это сформулировал Пол Грей, книжный обозреватель «Journal of Information Systems Management». В своей рецензии на книгу «Information Economics: Linking Business Performance to Information Technology» («Информационная экономика: влияние информационной технологии на эффективность компании»), одну из важнейших работ по методу прикладной информационной экономики, Грей написал: «Не обращайте внимания на слово „экономика“ в названии. Все, что в этой книге есть экономического, — это приложение о графиках затрат»⁹. Желая похвалить, Грей на самом деле указал на основной недостаток данного подхода: он не имеет никакого отношения к экономике.

Еще один распространенный вариант метода расчета условных взвешенных коэффициентов называется методом анализа иерархий (analytical hierarchy process, АНР). От других аналогичных методов он отличается в двух отношениях. Во-первых, АНР основан на ряде попарных сравнений, а не на прямом присваивании баллов каким-либо характеристикам. При этом экспертов спрашивают, является ли один критерий «намного более предпочтительным», «чуть более предпочтительным» и т. д. по сравнению с другим, и таким же образом разные варианты выбора сравниваются между собой по одному критерию. Например, возникает вопрос: что предпочтительнее — «стратегические преимущества» нового товара А или «стратегические преимущества» нового товара В? Следом идет другой: что предпочтительнее — «риск, связанный с разработкой» товара А или «риск, связанный с разработкой» товара В? Наконец, требуется найти ответ, что важнее — «стратегические преимущества» или «риск, связанный с разработкой нового товара». Таким образом, сначала эксперты сравнивают проекты по одному критерию, а затем сопоставляются сами критерии. Попарные сравнения не требуют разработки условных шкал, что можно считать преимуществом данного метода. Однако, как ни странно, при использовании АНР результаты сравнений конвертируются в условные баллы.

Второе отличие метода анализа иерархий от других способов расчета условных взвешенных коэффициентов — определение «коэффициента согласованности». Он показывает, насколько ответы не противоречат друг другу. Например, если вы предпочитаете а) возможность получить стратегическое преимущество низкому риску разработки новой продукции и б) низкий риск, связанный с разработкой новой продукции, возможности использовать существующие каналы сбыта, то тогда вы не должны отдавать приоритет возможности использовать существующие каналы сбыта возможности получить стратегическое преимущество. Если таких несоответствий много, то коэффициент согласованности низок. Если ни один ответ не противоречит другому, то коэффициент согласованности равняется 1.

Расчет коэффициента согласованности базируется на одном методе из матричной алгебры — нахождении собственных значений, — применяемом для решения целого ряда математических задач. Поэтому метод АНР нередко называют «теоретически обоснованным», или «математически доказанным». Если бы критерием теоретической обоснованности было просто использование на каком-то этапе математического инструмента (пусть и такого мощного, как нахождение собственных значений матрицы), то тогда доказать правоту новой теории или эффективность нового метода было бы гораздо легче, чем на самом деле. Кто-нибудь нашел бы способ использовать нахождение собственных значений матрицы в астрологии или дифференциаль-

ные уравнения в хиромантии. Но ни в том, ни в другом случае ценность самого метода не повысилась бы только потому, что был применен математический прием, доказавший свою эффективность в других условиях.

На самом деле АНР — просто еще один метод расчета взвешенных коэффициентов, особенностью которого является возможность снижения уровня информационного шума за счет выявления противоречащих друг другу ответов. Однако это вряд ли делает его результаты «доказанными», как часто утверждается. Проблема в том, что сравнение таких критериев, как стратегическая согласованность и риск, связанный с разработкой новой продукции, обычно не имеет смысла. Если бы я спросил, что вы предпочитаете — новую машину или деньги, то вы, прежде всего, спросили бы меня, о какой машине и о каких деньгах я говорю. Если бы речь шла о малогабаритном автомобиле среднего класса с пятнадцатилетней историей и миллионе долларов, то вы, наверное, дали бы один ответ, а если бы о новом «роллс-ройсе» и ста долларах, то другой. Тем не менее, по моим наблюдениям, когда группа людей применяет АНР, никто не спрашивает, о какой степени риска разработки новой продукции и о каких объемах затрат идет речь. Как ни странно, они просто дают ответ, как если бы сравнение было очевидно. Такой подход привносит опасность, что одни люди просто представляют себе совсем иные связи между затратами и риском, чем другие, а значит, уровень шума только повышается.

Последний, особенно странный недостаток анализа иерархий — возможность обратного порядка предпочтений¹⁰. Допустим, вы проанжировали с помощью АНР варианты А, В и С так, что самым предпочтительным оказался вариант А. Предположим, что вы откажетесь от варианта С; изменится ли в результате положение вариантов А и В в списке так, что лучшим станет В, а худшим — А? Нелепо, не правда ли? Как ни странно, применение метода анализа иерархий может привести именно к этому.

Существует только один ограничивающий критерий, позволяющий с уверенностью сказать, являются ли методы анализа «затрат/выгод» или расчета взвешенных коэффициентов способами измерения: результатом должно стать повышение предыдущего уровня знания. Если использованный метод только увеличивает прежнюю ошибку, то это не измерение. Если его считают формализованным и систематизированным, но без научных доказательств уменьшения ошибки и принятия более удачных решений, это не измерение. На проведение псевдоизмерений организации нередко тратят больше времени и сил, чем потребовалось бы на применение способов, гарантированно снижающих неопределенность. Зачем же тогда, спрашивается, даже думать об использовании методов, которые фактически не уменьшают неопределенность?

СРАВНЕНИЕ МЕТОДОВ

В конечном счете, человеческое суждение — совсем не плохой инструмент измерения. Если вы регулярно принимаете большое число аналогичных решений, то модели Раша и линзы, несомненно, помогут вам снизить неопределенность, устранив отдельные типы ошибок, присущие экспертам. Даже простой z -показатель Доуза выглядит, похоже, как определенный шаг вперед по сравнению с мнением эксперта.

Расскажем для сравнения о еще одном подходе, как будто более эффективном, чем все перечисленные методы, — объективной оптимизированной линейной модели. В отличие от других обсуждавшихся в этой главе способов, он никак не зависит от человеческих суждений и поэтому обычно дает намного лучшие результаты. Обычно мы предпочитаем его, но во многих случаях, когда приходится количественно оценивать то, что «не поддается измерению», необходимые для этого подробные, объективные данные за прошлые периоды получить невозможно. Отсюда возникает потребность в таких методах, как модели линзы, Раша и т. д.

В главе 9 мы обсуждали способы проведения регрессионного анализа, нужные, чтобы выделить и оценить эффекты от многочисленных переменных. Имей мы больше данных за прошедшие периоды по некоей периодически возникающей проблеме, полную документацию по каждому фактору, выраженному в реальных единицах измерения (а не в баллах условной шкалы), и возможность зарегистрировать фактические результаты, можно было бы построить «объективную» линейную модель.

Если модель линзы выявляет корреляцию между исходными переменными и экспертными оценками, то объективная модель находит связь между этими переменными и фактическими результатами прошлых периодов. Во всех случаях применения модели линзы, перечисленных в рисунке 12.2, на основе прошлой информации была построена регрессионная модель. Например, врачам были предоставлены медицинские данные о больных раком, а затем на основе оценок их ожидаемой продолжительности жизни была построена модель линзы. Но помимо этого за пациентами продолжали наблюдать и определять их *фактическую* продолжительность жизни. И если погрешность результата, полученного с помощью модели линзы, оказалась всего на 2% меньше человеческого суждения, то ошибка оценки на базе объективной модели была меньше уже на 12%. Средняя погрешность оценок, полученных во всех случаях применения модели линзы (см. рис. 12.2), была на 5% меньше ошибки мнений экспертов, а средняя ошибка объективной модели — на 30%. Конечно, даже объективные линейные модели не являются панацеей от всех бед. Как мы говорили в предыдущих главах, обычно дальнейшее разложение задачи на со-

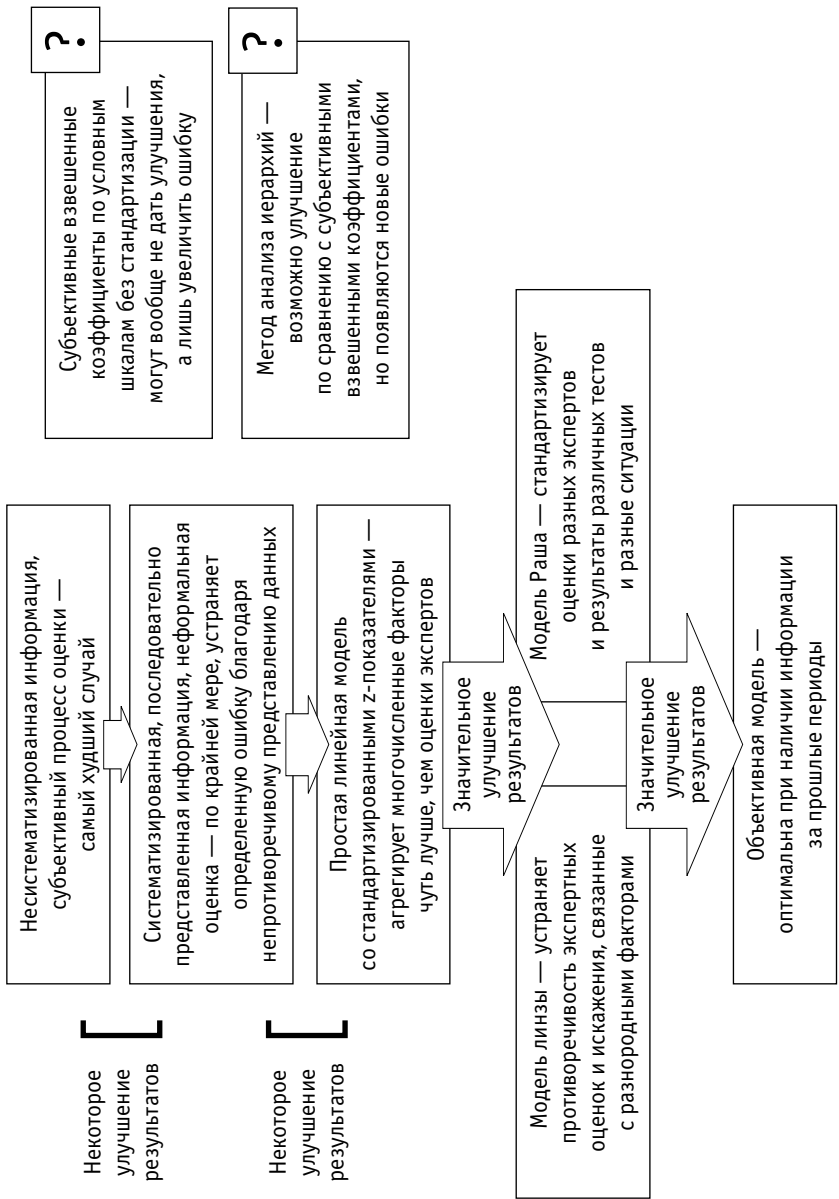


РИСУНОК 12.5. Сравнительная ценность методов оценки групп аналогичных проблем

ставляющие позволяет снизить неопределенность еще больше. Если бы мы расположили все эти методы в определенном порядке, так, чтобы на одном конце спектра оказались простые экспертные оценки, а на другом — объективная линейная модель, то получили бы следующую картину (см. рис. 12.5).

Несмотря на свои недостатки, описанные ранее методы оценки всегда эффективнее простых экспертных мнений. Такие методы, как модели Раша и линзы, устраняют основные погрешности человеческих суждений и превращают эксперта в гибкий, калиброванный и очень мощный инструмент измерения. По мнению многих специалистов по психологии принятия решений, оспаривать эффективность этих методов все равно, что стегать мертвую лошадь. Лучшее всего это сформулировал Пол Мил, профессор психологии Университета штата Миннесота:

Нет ничего странного в том, что в социологии постоянно появляется столько качественно разных исследований, которые ведут к одному заключению. Когда вы проводите 90 исследований [теперь их уже около 150]* с целью предсказания всего, что угодно, начиная от результатов футбольных матчей до диагноза заболевания печени, и когда вы вряд ли можете назвать хотя бы полдюжины работ, доказывающих, что экспертные оценки лучше, то уже пора сделать практические выводы¹¹.

■ ПРИМЕЧАНИЯ

- 1 *Robert Kaplan*. Is Beauty Talent? Sex Interaction in the Attractiveness Halo Effect. Paper presented at the Annual Meeting of the Western Psychological Association. Los Angeles, California, 1976, April 8–11.
- 2 *Robyn M. Dawes*. The Robust Beauty of Improper Linear Models in Decision Making // *American Psychologist*, 1979, 34, p. 571–582.
- 3 *G. Rasch*. On General Laws and the Meaning of Measurement in Psychology. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability. Berkeley: University of California Press, 1980, p. 321–334.
- 4 *Egon Brunswik*. Representative Design and Probabilistic Theory in a Functional Psychology // *Psychological Review*, 1955, 62, p. 193–217.
- 5 *Robyn M. Dawes, Bernard Corrigan*. Linear Models in Decision Making // *Psychological Bulletin*, 1974, Vol. 81, № 2, p. 93–106.
- 6 По крайней мере, в одном из четырех примеров «экспертами» были студенты. В двух остальных примерах эксперты предсказывали мнения других специалистов (патологи предсказывали диагнозы других врачей, а преподаватели предсказывали оценки приемной комиссии). Кроме того,

* Вставка автора. — *Примеч. редактора.*

большинству экспертов, на основе оценок которых была построена первая модель, удалось предсказать результаты несколько лучше, чем это было сделано с помощью баллов Доуза.

- 7 *M. Parker, R. Benson, H. E. Trainor.* Information Economics: Linking Business Performance to Information Technology. Englewood Cliffs, NJ: Prentice-Hall, 1988.
- 8 *Barbara McNurlin.* Uncovering the Information Technology Payoff. United Communications Group, Rockville, MD, 1992.
- 9 *Paul Gray.* Book review of «Information Economics: Linking Business Performance to Information Technology» // Journal of Information Systems Management, 1989, Fall.
- 10 *A. Stam, A. Silva.* Stochastic Judgments in the AHP: The Measurement of Rank Reversal Probabilities // Decision Sciences Journal, 1997, Summer, vol.28, № 3.
- 11 *P. E. Meehl.* Clinical versus Statistical Prediction. Minneapolis: University of Minnesota Press, 1954, p. 372–373.



Новые инструменты измерения для менеджмента

Интересно, что удалось бы измерить таким светлым головам, как Эратосфен, Энрико и Эмили, имей они в своем распоряжении обсуждавшиеся в данной книге методы. Не сомневаюсь, что много всего. Но, к сожалению, эти инструменты используются совсем не так часто, как могли бы, что, конечно, сказывается на качестве многих принимаемых важных и рискованных решений.

Говоря об инструментах измерения, я опять имею в виду не просто приборы, используемые для научных наблюдений. Я говорю о вещах, существование которых вам давно известно, но которые вы наверняка не считаете инструментами измерения. Сюда входят в том числе новые беспроводные устройства и даже Интернет.

МАРКЕРЫ ХХІ ВЕКА: В НОГУ С ТЕХНИЧЕСКИМ ПРОГРЕССОМ

Один из обсуждавшихся нами методов наблюдения — использование специальных средств слежения за тем, что ранее не определяли. Добавив что-либо в наблюдаемое явление, вы могли бы сделать его более легким для наблюдения. Чтобы получить представление о процессах, происходящих в верхних слоях атмосферы, мой отец, со-

трудник Национальной службы погоды США, запускал при сильном ветре шары-зонды с радиопередатчиками и простейшими метеорологическими приборами. В нашем примере с измерением численности популяции рыбы в озеро была выпущена меченая рыба. Когда объект трудно наблюдать в его естественном виде, используют многочисленные приемы: установку на нем датчика, нанесение на него метки или использование маркера.

Массу возможностей создают не столько сами эти инструменты, сколько дешевизна их использования. Например, простая технология радиочастотной идентификации (radio frequency ID, RFID) революционизировала измерения в определенных видах бизнеса, но может применяться еще шире. RFID-метка — крохотное устройство, отражающее радиосигнал и посылающее в отраженном сигнале уникальный идентификационный код. Одна RFID-метка стоит сегодня всего 10–20 центов, а используются они главным образом для облегчения инвентаризации товарно-материальных запасов.

Когда я спросил известного физика и автора книг Фримена Дайсона о самом, по его мнению, важном и интересном достижении в области измерений, он без колебания ответил: «Самое яркое — GPS (Global Positioning System — глобальная система определения местоположения, или глобальная навигационная система). Она изменила абсолютно все». Вообще-то я ожидал другого ответа. Наверно, думал я, он вспомнит о том, как занимался анализом боевых операций ВВС Великобритании в период Второй мировой войны, но GPS — безусловно, хороший пример, поскольку система действительно является революционным инструментом измерения и сама по себе измерение. Экономически GPS доступна практически всем и предлагается в комплекте с различными программными средствами и услугами. Тем не менее, перечисляя новые используемые бизнесом средства измерения, многие и не вспоминают об этой системе, отчасти из-за ее повсеместного использования. Но если такой человек, как Дайсон, считает ее новым словом в измерении, то к его мнению стоит прислушаться.

Предоставляемыми технологией GPS возможностями по измерению пользуются большинство отраслей транспорта. Одна из фирм, помогающих транспортным компаниям полностью использовать GPS, — базирующаяся в Скоттсдейле (штат Аризона) GPS Insight (GPSI). Она поставляет устанавливаемые в транспортных средствах GPS-навигаторы, подключенные к беспроводной сети, в которую можно войти с веб-сайта компании. GPSI показывает местонахождение машин на картах, доступных на Google Earth. Как известно всем знакомым с сайтом Google Earth, в данную систему входят спутниковые фотографии Земли, информация о дорогах, компаниях и прочие самые разнообразные сведения Географической информационной системы (Geographic Information System, GIS), введенные в единую программу. Каждый может бесплатно скачать Google Earth и увидеть

спутниковую фотографию своей улицы или любого другого интересующего его места.

Изображения в Google Earth отражают вид местности не в режиме реального времени: иногда снимкам исполняется уже два года (так, на фото моего микрорайона видна стройка, закончившаяся более двух лет назад); однако дорожная и прочая информация обычно более свежая. Одни районы охвачены системой хуже, чем другие. На снимках многих районов можно с легкостью различить машины, но разрешение фото крохотного городка Йеля (штат Южная Дакота), в котором я жил в детстве, такое низкое, что вы с трудом различаете дороги. Со временем полнота охвата, разрешение и актуальность снимков, конечно, улучшатся.

Высококачественные аэрофотоснимки выкладывают в Интернет разные компании, однако GPS Insight обычно предоставляет их клиентам в виде изображений, наложенных на снимки Google Earth. Цена чаще всего колеблется от 1 до 10 дол. за квадратную милю.

Каждое из этих технических средств могло бы само по себе служить разумному пользователю инструментом измерения. Но, используя одновременно возможности GPS, беспроводных сетей, Интернета и Google Earth, компания GPS Insight в состоянии предоставить такую информацию о местонахождении машин, действиях водителей и их манере вождения, получить которую раньше было практически невозможно. В ее сообщениях кратко указываются время движения, продолжительность стоянки, а также их средние значения и дисперсии, что помогает понять, где «копать дальше». А если «копать дальше», то можно определить точное место, срок и действие, например то, что машина уже два часа стоит на углу 43-й улицы и Центрального парка. Выбрав в Google Earth «бары и рестораны», можно даже узнать, у какого именно ресторана она стоит.

Прочие виды сообщений информируют о том, кто превышает скорость на дороге, как долго разные машины эксплуатируются в течение дня по сравнению с оплаченными часами, используются ли они по окончании рабочего дня, двигаются ли предписанными маршрутами, сколько часов или миль они едут по территории каждого штата (что нужно знать для уплаты налога на горюче-смазочные материалы). Поскольку данная система снижает неопределенность относительно такого числа экономических показателей, она может считаться очень полезным инструментом измерения.

Достижения технического прогресса используются и для оценки интенсивности общения людей в бизнесе. Джордж Эберштадт — один из основателей nTag, компании, разработавшей электронное устройство, способное определять, кто с кем общается. Оно весит не более 150 г и использует одноранговую беспроводную сеть радиосвязи для идентификации каждого, кто носит такое же устройство, когда он оказывается в пределах доступности. Когда люди разговаривают,

система nTag выявляет участников общения, используя инфракрасный строб-импульс для того, чтобы протестировать все имеющиеся в помещении устройства. Эти устройства позволяют определить, кто с кем и как долго разговаривает. Данные передаются без проводов в сеть радиодоступа, а затем в центральную базу данных.

Подход именных электронных значков решает ключевую проблему отношения потребителей к подобным устройствам. Эберштадт говорит: «В то время как большинству людей не нравится носить электронные приборы слежения, устройство nTag — это удостоверение личности, и им никто не отказывается пользоваться». Он называет это устройство обоудовыгодным — вы соглашаетесь им пользоваться потому, что это дает вам преимущества. «Люди соглашаются предоставлять о себе информацию, если получают что-то взамен».

Решив оценить «интенсивность общения» на различных мероприятиях, вы наверняка получите очень полезную информацию. Если бы вы председательствовали на конференции и обнаружили, что представители некоторых групп в аудитории активно общаются между собой, оставляя без внимания других, то смогли бы найти способ преодолеть коммуникационные барьеры. Устройства nTag предназначены прежде всего для проведения конференций, но компания надеется, что они найдут и более широкое применение. Эберштадт говорит: «Обычно основными целями проведения любого совещания люди считают сотрудничество, обучение и мотивацию. Чтобы определить ценность совещания, нужно измерить степень достижения этих целей». Устройства nTag позволяют установить, кто с кем разговаривает и как долго, благодаря чему компания может определить, выполняет ли мероприятие свою функцию развития сотрудничества.

Если Эратосфен сумел определить длину окружности Земли по длинам теней, то, интересно, какие измерения в области экономики, политики или психологии он смог бы провести с помощью GPS? Если Энрико Ферми установил мощность атомной бомбы с помощью десятка обрывков бумаги, то, интересно, что бы он сделал с десяткой RFID-меток? Если Эмили смогла разоблачить метод бесконтактного массажа, проведя простой эксперимент с картонным экраном, интересно, что бы она измерила сегодня, располагай чуть большим бюджетом и парой новых инструментов?

ИНТЕРНЕТ КАК ИНСТРУМЕНТ ИЗМЕРЕНИЯ МИРА

Писатель Уильям Гибсон написал несколько романов в жанре научной фантастики, которыми может заслуженно гордиться. Именно он ввел в обращение термин «киберпространство» как прообраз Интернета,

пользователи которого не только применяли клавиатуру и мышь, но и запросто «входили» в виртуальную реальность. Кое-кто из его персонажей летал над полями данных, чтобы выявить закономерности, например неэффективность рынка, и быстро на них заработать.

Как и многие другие научные фантасты, в некоторых отношениях Гибсон мыслил совершенно нереалистично. Лично я не вижу особого смысла в том, чтобы летать над полями данных в киберпространстве, хотя, наверное, это и было бы забавно. Думаю, что получаю больше полезной информации и делаю это быстрее с помощью старых добрых поисковиков Google и Yahoo. Но мысль Гибсона о том, что киберпространство не просто хранилище данных, но своего рода зеркало того, что в реальном времени происходит на планете, совсем не далека от истины. Мы действительно в любой момент можем получить доступ к целому океану данных. Даже не летая над ними в киберпространстве, мы можем выявлять закономерности и принимать благодаря этому удачные решения.

Восхваление чудесных возможностей Интернета уже давно стало привычным делом. Но одно направление его использования, похоже, пока недооценивается. Интернет — возможно, самый удивительный инструмент измерения, с которым большинству из нас суждено столкнуться за всю свою жизнь. Совсем несложно с помощью поисковой машины найти научную статью о том, что вы хотите оценить. Но есть и другие аспекты применения Интернета как инструмента измерения, объясняющие, почему Всемирная паутина сегодня быстро становится ответом на вопрос о том, как измерить все, что угодно.

Особого упоминания заслуживают несколько новых веб-технологий. Одна из них — метод сбора данных из самого Интернета, а другая — метод использования Сети для получения сведений от других.

В Интернете содержится масса информации, и она очень быстро меняется. Воспользовавшись обычной поисковой машиной, вы получите список веб-сайтов, но и только. Но предположим, что вместо этого необходимо определить, сколько раз название вашей компании упоминается на некоторых новостных сайтах или активно ли ваша новая продукция обсуждается в блогах. Возможно, вам даже нужно использовать эти данные вместе с другими, существующими в систематизированном виде на других сайтах, например с экономическими показателями, публикуемыми правительственными учреждениями.

Программы-анализаторы («screen-scrapers») позволяют регулярно собирать всю эту информацию, не нанимая для этого специалистов и не заставляя их работать круглосуточно. Президент и основатель сайта www.screen-scraper.com Тодд Уилсон говорит: «Есть такие сайты, на которых все меняется каждые 3–4 секунды. Наш инструмент очень подходит для отслеживания изменений, происходящих в Сети со временем». Благодаря таким «экранному шпионам» вы можете срав-

нить цены на ваши новый и проверенный рынком продукты на сайте www.ebay.com, определить зависимость продаж вашей сети магазинов в разных городах от местной погоды или даже час за часом следить за тем, сколько раз название вашей фирмы находят с помощью различных поисковиков (хотя если вы хотите только получать оповещения о новых записях, а не создаете базу данных, то можете просто подписаться на Google Alerts).

Как показывает поиск в Интернете, существует несколько «мэш-апов»*, которые собирают данные из многочисленных источников и представляют их в виде, позволяющем взглянуть на них по-новому. Сегодня многие мэшапы наносят сведения о компаниях, недвижимости, движении транспорта и т. п. на карты таких сайтов, как MapQuest или Google Earth. Я обнаружил на сайте www.housingmaps.com мэшап данных Google Earth и данных о недвижимости, который позволяет вам увидеть на карте цены недавно проданных домов. Еще один мэшап на сайте www.socaltech.com показывает карту расположения компаний, в которые недавно был вложен венчурный капитал. Сначала может показаться, что эти сайты предназначены для тех, кто хочет купить дом или найти работу в новой компании. Но разве их нельзя использовать при проведении исследований для какой-нибудь строительной компании или при прогнозировании темпов роста новой отрасли? Здесь все зависит только от нашей изобретательности.

Можно представить себе практически неограниченное число комбинаций направлений анализа с созданием мэшапов таких сайтов, как MySpace и (или) YouTube, для оценки культурных тенденций или выявления общественного мнения. EBay дает массу бесплатной информации о поведении покупателей и продавцов и о том, что продается и покупается, при этом существует несколько мощных аналитических инструментов, позволяющих обобщить всю имеющуюся на этом сайте информацию. Реплики и отзывы на отдельные товары, выкладываемые на сайтах Sears, Wal-Mart, Target и Overstock.com, служат источниками бесплатной информации, если только у нас хватит сообразительности ее использовать. Информации столько, что голова может пойти кругом.

Или же, вместо того чтобы добывать в Интернете информацию с помощью «экранных шпионов» и мэшапов, вы можете использовать Всемирную сеть для проведения прямых опросов потребителей, работников и т. д. Существует, например, такая интернет-компания по проведению опросов, как Key Survey (www.keysurvey.com). Подобные фирмы предлагают услуги по проведению статистического анализа.

* Мэшап (от *англ.* to mash up — смешивать) — программное приложение, интегрирующее данные из нескольких источников и представляющее его на одной странице. — *Примеч. переводчика.*

Некоторые придерживаются «рационального», или адаптивного, подхода к проведению опросов, когда задаваемые в его ходе вопросы варьируют в зависимости от того, как респонденты ответили на предыдущие задания. Хотя подобные возможности могут быть очень полезными, многие клиенты интернет-служб по проведению опросов находят, что использование таких методов измерения оправдывается уже одним снижением затрат.

Приведем пример. Проведение интернет-опроса фермеров, состоявшего из 40–50 вопросов, обычно обходилось журналу «Farm Journal» в среднем по 4–5 дол. на каждого респондента. Теперь, когда журнал пользуется услугами Key Survey, эти затраты сократились до 25 центов, и у него появилась возможность опрашивать сразу по полмиллиона человек.

NATIONAL LEISURE GROUP

Первый уровень — мерить все, что легко поддается измерению. Этот подход не вызывает возражений. Второй — отбросить то, что трудно измеряется, или приписать ему произвольное количественное значение — искусственный, уводящий в сторону путь. Третий уровень — предположить, что все трудноизмеримое не имеет значения. Это страусиная политика. Четвертый этап — сказать, что измеряемое с трудом вообще не существует. Это самоубийство.

Еще один клиент Key Survey — National Leisure Group (NLG), крупная круизная компания с годовым доходом около 700 млн дол.

Джуллианна Хейл — директор National Leisure Group по человеческим ресурсам и внутренним коммуникациям. Сначала она пользовалась услугами Key Survey только для решения задач, стоявших перед отделом человеческих ресурсов, в частности для оценки удовлетворенности сотрудников, результатов работы коучей и эффективности тренингов, но позже она увидела возможность оценить и удовлетворенность туристов. Она говорит: «Работающим в туризме каждый пенни дается с боем. Норма прибыли в нашей отрасли очень низка». В этих условиях было особенно важно определить, насколько положительным является имидж NLG в глазах клиентов. «У нас была масса турагентов, но клиенты редко снова приходили к нам, — объясняет Хейл. — Поэтому мы создали отдел по изучению впечатлений клиентов и начали оценивать их удовлетворенность. Убедить руководство в необходимости таких измерений удалось не сразу. Пришлось выдержать настоящий бой».

Каждые шесть–восемь месяцев Key Survey проводила опросы клиентов, обслуживаемых разными отделами NLG. Желая сэкономить время потребителем, компания старалась делать это эффективно. Хейл вспоминает: «Было несколько вариантов опроса потребителей, но в конце концов утвердили два». Одна группа вопросов автоматически отсылалась по электронной почте сразу после того, как клиент оформил заказ,

а вторая — по возвращении из круиза. Хейл говорит: «Мы просто хотели посмотреть, какие результаты получим. На первый вариант опроса отвечали 4–5% туристов, но на второй — уже 11,5%». Такой процент отклика считается высоким. Разумно используя простые средства контроля, NLG сравнивает ответы на вопросы типа «Порекомендуете ли вы нас своим друзьям?» до и после того, как клиент побывает в круизе, чтобы определить, повысились ли баллы после путешествия.

Обнаружив, что после круиза удовлетворенность клиентов падает, NLG решила реализовать специальную программу обучения турагентов. Хейл говорит: «Нам нужно было научить их торговать по-новому и предлагать маршруты, более подходящие клиентам». Измерение помогло выявить проблему и уже поэтому оказалось успешным. Теперь компании необходимо оценить эффект реализации новой программы.

РЫНКИ ПРЕДСКАЗАНИЙ: ПРОВЕДЕНИЕ ИЗМЕРЕНИЙ С ЭФФЕКТИВНОСТЬЮ УОЛЛ-СТРИТ

Интернет сделал возможным новый динамичный способ проведения измерений, объединяющий мнения по механизму, аналогичному фондовому рынку. Когда экономист говорит, что фондовый рынок «эффективен», он имеет в виду, что постоянно опережать рынок очень трудно. Курс любых акций в любой момент времени может с одинаковой вероятностью как упасть, так и повыситься. Будь это не так, участники рынка только продавали или только покупали бы эти акции до тех пор, пока равновесие не восстановится.

Результатом процесса сбора суждений является прогноз, более точный, чем индивидуальное мнение любого отдельного участника рынка. Преимущество данного процесса по сравнению с опросами общественного мнения заключается в том, что у участников есть стимул к тому, чтобы не только тщательно продумывать все вопросы, но даже (особенно, если на кону большие деньги) приобретать на собственные средства новую информацию об инвестициях, необходимую для анализа. У людей, которые ведут себя нерационально, быстро кончаются деньги, и они уходят с рынка. Именно они и создают «случайный шум», уравновешивая друг друга на крупном рынке, поскольку из-за своей нерациональности обычно либо переоценивают, либо недооценивают акции. А поскольку рынок поощряет активное участие, новости об изменении стоимости компании быстро находят отражение в курсе ее акций.

Именно такой механизм пытаются использовать новые «рынки предсказаний». По данным исследований, они появились еще в на-

чале 1990-х годов, но широкую известность получили только в 2004 г. благодаря знаменитой книге Джеймса Шуровьески «Мудрость толпы. Почему вместе мы умнее, чем поодиночке, и как коллективный разум влияет на бизнес, экономику, общество и государство»*. Благодаря нескольким программным приложениям и открытым веб-сайтам появились «рынки» предсказаний того, кто, например, получит кинопремию «Оскар» за лучшую женскую роль или кто станет кандидатом в президенты от республиканцев. Примеры существующих рынков предсказаний приводятся в таблице 13.1.

Участники рынка предсказаний покупают или продают акции определенного прогноза, например предположения о том, кто будет кандидатом в президенты в США от республиканцев. Если предсказание сбывается, то держатель акции получает определенную сумму, чаще всего 1 дол. Можно сделать ставку на то, что прогноз осуществится, купив акцию «Да», или на то, что он не реализуется, купив акцию

ТАБЛИЦА 13.1. Обзор существующих рынков предсказаний

Компания	Предназначение
Consensus Point www.consensuspoint.com	Служба для предприятий, решивших создать рынок предсказаний для внутреннего пользования. Основанная теми, кто создавал Foresight Exchange, компания отличается большой гибкостью в организации систем поощрения, в том числе материального, составителей удачных прогнозов
Foresight Exchange www.ideosphere.com	Бесплатный общедоступный веб-сайт и один из результатов первых экспериментов в области рынков предсказаний. В качестве ставок используются только виртуальные деньги. Предсказания делаются всеми желающими и проверяются добровольцами. Это активный рынок с большим числом игроков, дающий хорошее представление о том, как вообще работают рынки предсказаний
NewsFutures www.newsfutures.com	Прямой конкурент Consensus Point, оказывающий компаниям услуги по созданию рынков предсказаний
TradeSports www.tradesports.com	Данный рынок предсказаний начинался как букмекерский сайт со спортивным уклоном, но затем на нем стали принимать ставки на политические, экономические и другие события. Стать членом сайта может любой человек, создав счет, но на кон принимаются реальные деньги. Предсказание может сделать каждый, но и для этого нужны деньги. Все неспортивные сделки теперь перенесены на Intrade

* Книга выпущена в 2007 г. издательством «Вильямс». — *Примеч. редактора.*

«Нет». То есть вы заработаете, если предсказание окажется верным и у вас будет акция «Да» или если предсказание окажется неверным и у вас на руках окажется акция «Нет». Когда решение о том, реализовался ли прогноз, уже принято и выигрыш выплачен, соответствующая акция выбывает из обращения.

Если у вас 100 акций «Да», вы сделали ставку на то, что кандидатом станет определенный человек и он действительно им станет, то вы заработаете 100 дол. Но при покупке этих акций у вас не было никакой уверенности, что ваш прогноз сбудется. За несколько месяцев до объявления кандидата в президенты акция может стоить всего 5 центов; после объявления кандидата ее стоимость возрастает; когда еще один претендент решит баллотироваться в президенты, курс акции несколько уменьшается, а потом подсакивает всякий раз, когда кто-нибудь из кандидатов выбывает из борьбы. Заработать деньги вы можете, просто сохранив акции до самого конца или же продав их в любой момент, когда вам покажется, что рынок их переоценивает.

Но участники рынков предсказаний прогнозируют не только результат выборов, лауреатов премии «Оскар» или нового *Американского идола**. Они могут составить прогноз по любому интересующему вас вопросу, например объединятся ли две конкурирующие компании, какими будут продажи нового продукта, чем закончится какое-нибудь важное судебное разбирательство или даже останется ли данная компания на рынке. На рисунке 13.1 показана цена уже утратившего свою актуальность предсказания «К 2005 г. компания Apple прекратит свое существование», размещенного на веб-сайте Foresight Exchange (www.ideosphere.com). Прекрати компания Apple свое существование как самостоятельное юридическое лицо к 1 января 2005 г., каждая акция «Да» принесла бы своему владельцу один доллар. Точный смысл прогноза (то есть способ определения, перешла ли компания к другому владельцу, объединилась ли с другой фирмой, была ли реструктурирована в ходе банкротства и т. д.) разъясняется в подробном описании, а также в комментариях судьи — того человека, кому предстояло решить, реализовался ли прогноз. Как мы теперь знаем, Apple отнюдь не ушла с рынка и все обладатели акций «Да» обнаружили, что их акции обесценились. Но люди, которые сделали ставку против этого предсказания, купив акции «Нет», заработали по доллару на каждую акцию. Курс такой акции, как и любой другой ценной бумаги, в разные периоды времени отражал появлявшиеся на рынке новости (график показывает некоторые ключевые события истории Apple, произошедшие до того, как прогноз утратил свою актуальность). Однако, в отличие от курсов других акций, цена акций «Да» легко определялась на основе вероятности того, что компания уйдет с рынка. В январе

* Телевизионный певческий конкурс «American Idol». — *Примеч. редактора.*



РИСУНОК 13.1.

Курс акций прогноза «К 2005 г. компания Apple прекратит свое существование» на Foresight Exchange

* Одна из моделей ноутбука от компании Apple. — *Примеч. редактора.*

** Музыкальный плеер. — *Примеч. редактора.*

1999 г. одна акция «Да» стоила около 30 центов, что означало: по мнению рынка, существовала 30-процентная вероятность ухода Apple с рынка до 1 января 2005 г. В 2004 г. цена одной акции «Да» упала ниже 5 центов, поскольку уже было очевидно, что в начале следующего года Apple все еще останется в бизнесе.

Что интересно в рынках предсказаний, так это связь курсов акций с вероятностью правильности прогноза. Если проанализировать большое число уже утративших актуальность предсказаний, легко понять, насколько эффективно работают такие рынки. Чтобы определить, точно ли рассчитывается вероятность, мы, как и калиброванные эксперты, должны сравнить множество сделанных когда-то прогнозов с тем, что произошло на самом деле. Если метод расчета вероятности хорош, то когда он говорит, что вероятность каждого события из группы — 80%, именно 80% из них должны наступить. Аналогично из всех предсказаний, которые продаются по 40 центов, сбыться должны, в конечном счете, около 40%. Рисунок 13.2 показывает, как это правило соблюдается на TradeSports, NewsFutures и Foresight Exchange.

Этот график показывает цены на сайтах TradeSports и NewsFutures для одной и той же серии из 208 игр Национальной футбольной лиги США (National Football League, NFL), приведенные в исследовании, опубликованном в журнале «Electronic Markets»¹. Я наложил на эти данные результаты собственного анализа 353 прогнозов Foresight Ex-

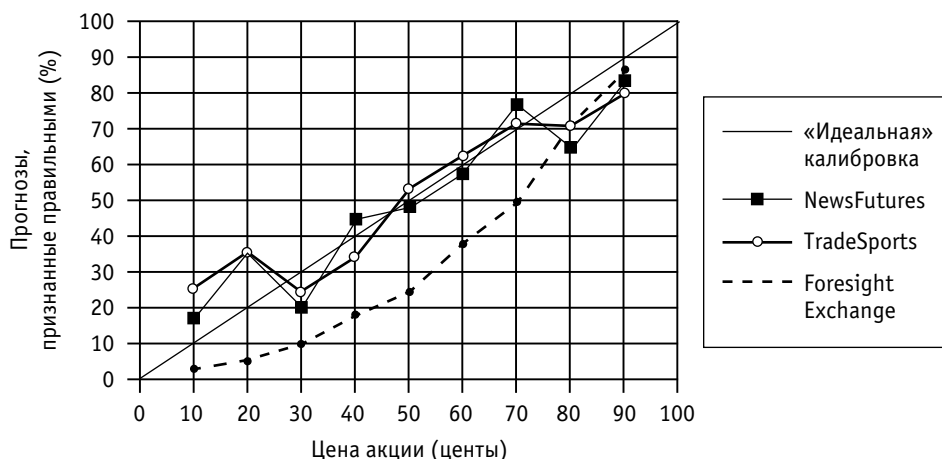


РИСУНОК 13.2. Эффективность рынков предсказаний: курс акций и реальность

change, отобранных из всех (а не только футбольных) источников, причем только тех, по которым было заключено много сделок.

Мы видим, что с ростом цены растет и вероятность наступления данного события. TradeSports, сайт игры на реальные деньги, — пример хорошей калибровки (вероятность события очень близка к цене соответствующей акции). NewsFutures — пример почти такой же хорошей калибровки, хотя используются не настоящие, а виртуальные деньги (лучшим игрокам разрешается покупать на свои «деньги» такие призы, как iPod).

В этом смысле Foresight Exchange сильно отличается от двух других сайтов. На этой электронной бирже в ходу только виртуальные деньги, и возможность купить приз игрокам не предоставляется. Игроки просто получают каждую неделю 50 виртуальных долларов. На эти деньги можно купить только акции прогнозов, а единственное вознаграждение удачливого игрока — репутация хорошего прогнозиста. Возможно, именно поэтому почти все на этом рынке переоценено (цены выше рассчитанных по вероятности наступления данного события). Другая причина — участвовать в качестве прогнозиста на Foresight Exchange может каждый желающий. Большинство предсказаний на этой бирже долгосрочные, многие из них довольно странные и сбываются только 23% из них. Интересно, что переоценка — явление настолько постоянное для этого сайта, что для получения такой же вероятности, как на TradeSports и NewsFutures, нужно просто умножить рыночный курс

на корректирующий коэффициент. Уже по окончании данного исследования TradeSports сделала из своего не имевшего отношения к спорту подразделения самостоятельную компанию, получившую название Intrade (www.intrade.com).

Некоторые компании, такие как General Electric (GE) и Dow Chemical, начинают присматриваться к рынкам предсказаний как к полезному инструменту оценки вероятности наступления в будущем определенных событий. Например, GE уже использовала их для оценки вероятности получения прибыли сотрудниками от внедрения предложенных ими нововведений. Рынки предсказаний можно применять и для определения пороговых значений. Если новый продукт — хорошее вложение средств только при условии, что выручка в первый же год составит 25 млн дол., то компания может сделать прогноз: доходы от продаж продукта X в первые 12 месяцев после выхода на рынок достигнут 25 млн дол.

Ясно, что рынки предсказаний — новый мощный инструмент измерения неизмеримых на первый взгляд объектов и явлений. Энтузиазм поклонников этих рынков столь велик, что они считают их альфой и омегой измерения практически всего, что угодно. По мнению некоторых, чтобы обосновать новый проект, достаточно просто составить прогноз по каждой отдельной переменной проекта и сделать его предметом обсуждения участников рынка. А после того, как вышла в свет книга Шуровьески, их пыл только увеличился.

Памятуя об этом, не стоит забывать и о другом. Рынок предсказаний — отнюдь не волшебная палочка. Это просто способ выяснить мнение группы людей и, особенно если используются реальные деньги, побудить их к анализу интересующего вас вопроса. Другие обсуждавшиеся нами методы также эффективны, и иногда лучше использовать именно их. В таблице 13.2 проводится сравнение всех описанных в нашей книге способов повышения точности оценок.

ПОЛЕЗНЫЙ УРОК: ДЕЛО О «РЫНКЕ ТЕРРОРИЗМА»

В 2001 г. служба информационного обеспечения Агентства передовых оборонных исследовательских проектов (Defense Advanced Research Projects Agency, DARPA) решила изучить возможность использования рынков предсказаний для целей выбора будущей политики. Решение было принято на основании результатов исследований, показавших, что прогнозы таких рынков по целому ряду вопросов оказываются более точными, чем прогнозы отдельных экспертов. И этому эксперименту было суждено вызвать бурную реакцию общественности.

В 2002 г. были созданы пробные рынки для предсказания распространения атипичной пневмонии и уровней угроз нарушения безопасности. Предполагалось, что эти рынки будут использоваться только пра-

ТАБЛИЦА 13.2. Рынки предсказаний и другие субъективные методы оценки

Метод оценки	Преимущества
Тренинг по калибровке	Оптимален, когда необходимо быстро и дешево сделать много оценок. Требуется привлечения только одного эксперта и дает немедленный ответ. Данный метод оценки должен применяться первым в большинстве случаев — более сложные способы могут использоваться только при соответствующей стоимости интересующей вас информации
Модель линзы	Рекомендуется, когда нужно сделать большое число повторных оценок (например, при необходимости оценить несколько инвестиционных возможностей) и когда для каждого суждения можно собрать однотипную информацию. Как только модель линзы построена, она моментально генерирует ответы на такого рода вопросы, не прибегая к услугам экспертов. Модель можно создать даже на основе только гипотетических сценариев
Модель Раша	Используется для стандартизации разных оценок, сделанных разными экспертами, или результатов тестов по различным проблемам. В отличие от модели линзы, требует большого числа реальных (а не гипотетических) оценок, которые затем все учитываются при стандартизации
Рынки предсказаний	Оптимальны для разработки прогнозов, особенно когда полезно следить за изменением вероятностей во времени. Чтобы была заключена хотя бы одна сделка, требуется не менее двух участников рынка. Неидеальны, если нужно быстро получить ответы на множество вопросов, однотипных или нет. Если число предсказаний на рынке превысит число сделок, то справедливость многих из них так и не будет оценена

вительственными учреждениями, но обеспокоенность тем, что трейдеров будет слишком мало, юридические проблемы с условным переводом денег между правительственными учреждениями привели к тому, что торги были открыты для всех желающих.

В одном отчете было приведено изображение дисплея с такими возможными предсказаниями, как убийство Ясира Арафата и ракетная атака со стороны Северной Кореи. Эти примеры не остались незамеченными. 28 июля 2003 г. сенаторы Рон Уайден от штата Орегон и Байрон Дорган от Северной Дакоты написали директору службы информационного обеспечения Джону Пойндекстеру следующее: «Примеры, которые вы приводите в своем отчете, могли бы заставить участников сделать ставку на то, что в следующем году террористы применят в Израиле биологическое оружие. Нет никакого сомнения, что противостоять такой угрозе необходимо путем сбора как можно более надежной разведывательной информации, а вовсе не постановкой вопросов перед людьми, делающими ставки на веб-сайте. Тратить деньги налогоплательщиков на создание букмекер-

ских сайтов, принимающих ставки на терроризм, столь же расточительно, сколь и недостойно». Затем к этой критике присоединилась пресса.

Через два дня программа была заморожена, а Пойндекстер ушел в отставку. Робин Хансен из Университета Джорджа Мейсона, один из членов его команды и авторов идеи рынка предсказаний, по этому поводу сказал: «Никто из конгрессменов не спросил нас, справедливы ли эти обвинения и нельзя ли выбросить из проекта то, что задевает чьи-то чувства. Поддержки от DARPA мы тоже не дождались».

Сенаторы нашли эту программу аморальной и решили, что она не даст положительного эффекта. Также они почему-то предположили, что она заменит собой другие методы сбора разведывательной информации, хотя разведывательные службы, конечно, всегда пользуются сразу несколькими способами. Если их возмущение было вызвано опасением, что, используя этот рынок, террористы смогут обогатиться, то никаких оснований для него не было. Сенаторы не учли, что участники рынка могли заработать весьма небольшие деньги, так как сумма любой сделки не должна была превышать 100 дол. Хансен так резюмировал все это дело: «Они должны были высказать свою точку зрения о малознакомом для них проекте. Как миллионный проект при триллионном бюджете, он был легкой мишенью». Конечным результатом этого морализирования и политиканства стал отказ от использования чрезвычайно эффективного с точки зрения затрат инструмента, способного существенно улучшить результаты анализа разведывательной информации.

■ ПРИМЕЧАНИЕ

- ¹ *Emile Servan-Schreiber et al.* Prediction Markets: Does Money Matter? // *Electronic Markets*, 2004, September, № 14, p. 3.



Универсальный метод измерения: прикладная информационная экономика

В 1984 г. консалтинговая фирма The Diebold Group попросила главных исполнительных (CEO) и финансовых (CFO) директоров 10 крупных корпораций выступить в престижном Чикагском клубе перед своими коллегами из 30 крупнейших чикагских компаний. Представители таких компаний, как IBM, Mobile, AT&T и Citibank, рассказали о своих процессах принятия серьезных инвестиционных решений. Оказалось, что все они делают это одинаково и очень просто: если инвестиции признаются стратегическими, их финансируют. Рассчитывать доходность таких инвестиций никто и не пытается, что стало большим сюрпризом для некоторых присутствовавших на этой встрече представителей чикагского бизнеса.

Присутствовал на ней и Рей Эпич, многоуважаемый гуру в области ИТ, в то время консультант The Diebold Group, а ныне вице-президент RiverPoint Group LLC. Сомневаясь в разумности правила автоматического утверждения «стратегических» проектов, он мог привести немало примеров того, к чему приводит подобный подход к принятию решений. В частности, он рассказал о компании Mead Paper, которая «решила изменить состав бумаги и в результате выбросила на ветер 100 млн дол.».

Эпич также рассказал о своем разговоре с Бобом Прицкером из The Marmon Group — конгломерата, владельцы которого были в то

время третьей богатейшей семьей Америки: «Я спросил его, как он планирует капиталовложения». Ответ Прицкера звучал примерно так: «Мои ребята мне звонят, а я говорю им „да“ или „нет“». Он добавил: «Я не могу себе позволить, чтобы такого рода эксперты рассчитывали ROI». С тех пор осознание значения простых расчетов и определенный здоровый скептицизм по поводу чутья высших руководителей, наверное, уже стали нормой — а впрочем, может быть, и нет.

Во всяком случае, именно таким был тот мир, в который я попал, когда в 1988 г. стал консультантом компании Coopers & Lybrand по управлению. Я работал над несколькими интересными количественными задачами. Даже если поначалу они и не считались количественными, я старался сформулировать их именно так, поскольку таким было и остается мое мировоззрение. Я никогда специально не планировал свою карьеру, но меня, тем не менее, все чаще привлекали к работе над крупными проектами разработки программного обеспечения сначала в качестве аналитика, а затем и руководителя.

Примерно в это время я впервые заметил, что в управлении информационными технологиями крайне редко или вообще не используются количественные методы, широко применяемые в других областях. То, что количественно оценивалось в других сферах деятельности, в ИТ нередко считалось не поддающимся измерению. Именно тогда я решил, что кто-то должен найти способ внедрить в эту область уже апробированные количественные методы.

К тому времени я уже работал в компании DHS & Associates, базировавшейся в Роузмонте, штат Иллинойс. Позднее эта фирма стала называться RiverPoint, и именно в ней сейчас работает Рей Эпич. Руководство DHS & Associates также видело необходимость использования в ИТ более количественно обоснованных решений, а культура этой компании предоставляла консультантам большую свободу в разработке новых идей.

В том же году я приступил к созданию метода, который назвал прикладной информационной экономикой (AIE). Я разрабатывал ее для области информационных технологий, но оказалось, что она позволяет решать задачи по измерению, возникающие в любой сфере.

СВОДИМ ВСЕ ВОЕДИНО

Основные составляющие прикладной информационной экономики — методы оценки неопределенности, риска и стоимости информации, обсуждавшиеся в части II. Словом, метод AIE отвечает на четыре вопроса:

- 1) как смоделировать текущее состояние неопределенности;
- 2) как рассчитать, что еще необходимо измерить;



РИСУНОК 14.1. Цикл проведения измерения

- 3) как измерить это экономически оправданным способом;
- 4) как принять решение?

Чтобы глубже вникнуть в процессы прикладной информационной экономики, обратимся к рисунку 14.1. Вы видите, что АИЕ действительно не более чем обобщение всего, о чем мы говорили до сих пор.

Подход прикладной информационной экономики

С 1995 г. я измеряю с помощью прикладной информационной экономики всевозможные объекты, казавшиеся поначалу трудно или даже вовсе не поддающимися количественной оценке. Довольно длинное название этого подхода было выбрано потому, что я хотел придать

ему описательный характер. Постоянно рассчитывая стоимость информации о каждой неизвестной переменной, подлежащей учету при принятии решения, и пересчитывая ее после каждого нового измерения, мы получаем возможность определить, что именно заслуживает измерения.

На начальном этапе постановки задачи метод АІЕ придает большое значение количественной оценке неопределенности и риска как необходимому условию расчета стоимости информации. А когда выясняется, что проведение измерений экономически оправданно, АІЕ предполагает использование только методов, гарантированно уменьшающих ошибку. Сложность заключалась в том, чтобы собрать все это в один внутренне непротиворечивый метод. После нескольких первых проектов стало очевидно, что процесс должен состоять из следующих этапов:

Этап 0. Подготовка проекта

- *Предварительные исследования.* Чтобы понять характер проблемы, аналитик встречается с заинтересованными лицами и изучает результаты вторичных исследований и отчеты за прошлые периоды.
- *Подбор экспертов.* Обычно необходимо, чтобы свои оценки дали четыре–пять специалистов, но мне доводилось привлекать и по 20 экспертов, хотя я не рекомендую этого делать.
- *Планирование заседаний рабочей группы.* Вместе с отобранными экспертами составляется расписание четырех—шести заседаний рабочей группы продолжительностью в половину рабочего дня.

Этап 1. Построение модели принятия решения

- *Определение проблемы.* На первом заседании рабочей группы эксперты определяют, какую конкретную задачу они на самом деле должны проанализировать. Например, что на самом деле они должны сделать: решить, стоит ли продолжать данный инвестиционный проект, или проблема в том, как его скорректировать? Если задача — одобрить или отвергнуть инвестиционный проект или другую программу, то тогда необходимо встретиться с лицами, принимающими решения, чтобы определить инвестиционную границу для этой организации.
- *Детализация модели принятия решения.* Ко дню проведения второго заседания рабочей группы составляется электронная таб-

лица в программе Excel, учитывающая все факторы, влияющие на анализируемое решение, и их совокупное воздействие. Если принимается решение одобрить или отклонить какой-нибудь крупный проект, следует перечислить все затраты и выгоды, ввести их в общий денежный поток и рассчитать ROI (как это делается при обосновании любого проекта).

- *Первоначальные калиброванные оценки.* На оставшихся заседаниях рабочей группы мы калибруем экспертов и подставляем предложенные ими значения переменных в модель принятия решения. Эти значения не фиксированы (если только нам не известны точные числа), а являются калиброванными экспертными оценками. Все они представляют собой 90-процентные доверительные интервалы или другие распределения вероятностей.

Этап 2. Предварительные измерения

- *Анализ стоимости информации (value of information analysis, VIA).* На этой стадии мы анализируем стоимость информации о каждой переменной, входящей в модель. В результате мы узнаем не только значение каждой неизвестной, но и его порог. Макрос, написанный мной в программе Excel, делает это быстро и точно, но и методы, обсуждавшиеся ранее в этой книге, тоже дают хорошие оценки.
- *Предварительный выбор метода измерения.* В ходе VIA выясняется, что мы обладаем достаточной информацией о большинстве переменных и что их дополнительной оценки, кроме калиброванной, не потребуется. Обычно высокой оказывается стоимость информации лишь о паре переменных (и нередко их выявление приносит сюрпризы). На основании полученных данных осуществляется выбор таких методов измерения, которые обязаны снизить неопределенность, не превышая ожидаемой стоимости полной информации. В ходе VIA также определяется порог измерения, в случае достижения которого приходится принимать иное решение. Наш метод измерения ориентирован на уменьшение неопределенности относительно этого порога.
- *Применяемые методы измерения.* Разложение на составляющие, случайная выборка, субъективно-байесовский способ, проведение контролируемых экспериментов, метод линзы (и т.д.) или любое их сочетание — все это может использоваться для снижения неопределенности переменных, определенных на предыдущем этапе.

- *Усовершенствованные модели принятия решения.* Результаты этих измерений используются для уточнения значений переменных в нашей модели. В модель вводятся величины, появившиеся в результате разложения первоначальной переменной на составляющие (например, неизвестный элемент затрат может быть разложен на более мелкие компоненты с присущими им 90-процентными доверительными интервалами).
- *Конечная стоимость анализа затрат на информацию.* Анализ и измерения (предыдущие четыре шага) часто проходят несколько повторений. До тех пор пока VIA показывает, что стоимость информации превышает затраты на проведение измерений, их можно продолжать. Однако обычно уже после одной-двух итераций, согласно VIA, проведение дальнейших измерений экономически нецелесообразно.

Этап 3. Выбор показателей и конечные результаты

- *Полный анализ соотношения «риск/доходность».* Результатом моделирования методом Монте-Карло являются вероятности возможных исходов. Если необходимо принять решение о судьбе крупных инвестиций, проекта, серьезных обязательств или какой-либо другой программы (как это обычно и бывает), то следует сравнить риск и доходность с инвестиционной границей данной организации.
- *Выбор способов отслеживания показателей.* Нередко бывают такие переменные, рассчитывать которые вначале кажется нецелесообразным, поскольку ценность информации о них становится очевидной лишь впоследствии. Зачастую это величины, характеризующие ход выполнения проекта и внешние условия функционирования компании, например состояние всей экономики. Такие переменные необходимо отслеживать постоянно, так как их изменение может потребовать принятия корректирующих мер. В связи с этим следует ввести процедуры постоянного расчета подобных показателей.
- *Оптимизация решения.* Принимаемое на практике решение редко оказывается итогом простого процесса одобрения по типу «да—нет». А когда это так, существуют многочисленные способы улучшить уже принятое решение. Теперь, с детально разработанной моделью «риск/доходность», можно разработать стратегии уменьшения риска или попытаться повысить доходность инвестиций, проведя анализ по методу «что, если».

- *Заключительный отчет и презентация.* Заключительный отчет должен содержать описание модели принятия решения, результатов анализа стоимости информации, использованных методов измерения, положения на инвестиционной границе, а также всех показателей, требующих постоянного отслеживания, или методов оптимизации принятого решения.

Описанный процесс выглядит довольно сложным, но на самом деле это лишь резюме всего, о чем мы говорили в этой книге до сих пор. Рассмотрим теперь несколько примеров практического применения AIE для измерения в тех областях, которые многим участникам моего исследования казались совершенно неизмеряемыми.

ПРИМЕР ИЗ ПРАКТИКИ: СТОИМОСТЬ СИСТЕМЫ, СЛЕДЯЩЕЙ ЗА КАЧЕСТВОМ ПИТЬЕВОЙ ВОДЫ

В Агентстве по защите окружающей среды (EPA) функционирует информационная служба по безопасности питьевой воды (Safe Drinking Waters Information System, SDWIS) — главная система наблюдения за качеством питьевой воды в Соединенных Штатах, обеспечивающая быстрое реагирование на появление любых угроз здоровью населения. Когда отвечавшему за программу SDWIS руководителю филиала Джеффу Брайану потребовалось больше средств, перед ним встала задача подготовить убедительное обоснование проекта. Однако его беспокоило то, что все преимущества SDWIS были, в конечном счете, связаны с областью здоровья населения и он не знал, как их оценить экономически.

Заместитель руководителя информационной службы и главный специалист по технологии отдела экологической информации Марк Дей предложил Брайану рассчитать этот показатель методом прикладной информационной экономики. Дей, ставший инициатором большинства AIE-проектов в EPA, даже сказал, что его подразделение возьмет на себя часть расходов.

Этап 0

На нулевом этапе, фазе планирования, мы отобрали 12 человек, хорошо знакомых со SDWIS и понимавших ее значение. Мы наметили провести в течение трех недель пять заседаний рабочей группы продолжительностью в половину рабочего дня. Главным человеком в команде, способным привлечь других экспертов и решать возникающие вопросы, был признан Джефф Брайан.

Этап 1

На первом же заседании рабочей группы (посвященном определению стоявшей перед нами задачи) стало очевидно, что на самом деле, вопреки моим ожиданиям, речь идет вовсе не об анализе SDWIS целиком. Ведь эта система существовала уже несколько лет, и никто всерьез не собирался отказываться от нее или менять на другую. Необходимо было просто определить целесообразность трех направлений ее усовершенствования: модернизации системы обнаружения нарушений, обеспечения возможности доступа к информации разных штатов через Интернет и адаптации базы данных к современным требованиям. Три перечисленные программы требовали первоначальных вложений примерно 1, 2 и 0,5 млн дол. соответственно плюс текущие расходы. Мы должны были ответить на вопрос, целесообразны ли экономически данные улучшения, и, если это так, выбрать из них приоритетное.

Таким образом, электронная таблица должна была отразить три разных проекта предполагаемой модернизации SDWIS, каждый из которых имел свои преимущества. Наибольшие трудности вызывало сравнение затрат на реализацию проекта с ожидаемой пользой для здоровья населения. Служба управления и бюджета уже потребовала, чтобы EPA готовило экономические обоснования всех разрабатываемых им экологических правил. EPA приходилось рассчитывать издержки соблюдения предлагаемых норм и правил, а также выгоды для населения от введения каждого нового правила. Авторы нескольких исследований сумели оценить экономические последствия различных наиболее распространенных видов загрязнения питьевой воды. Зачастую в EPA использовали метод готовности платить за сохранение благоприятной экологической обстановки, но иногда рассчитывали убытки от загрязнения воды только по числу дней временной потери трудоспособности. Следующие два заседания рабочей группы были посвящены выяснению того, как SDWIS должна способствовать общественному здравоохранению. В результате нам удалось построить модель в виде электронной таблицы, координировавшую направления совершенствования SDWIS с оценкой стоимости выгод для здоровья населения. В модели, структура которой представлена на рисунке 14.2, было использовано 99 переменных.

На этой схеме каждый блок символизирует несколько переменных табличной модели. Например, для доступа штатов через Интернет мы оценивали временные затраты на разные виды деятельности, насколько сократятся эти затраты, а также насколько быстрее благодаря этому будут устраняться нарушения правил по безопасности питьевой воды.

На последних двух заседаниях рабочей группы этапа 1 мы провели тренинг по калибровке для экспертов и попросили их дать первоначальную

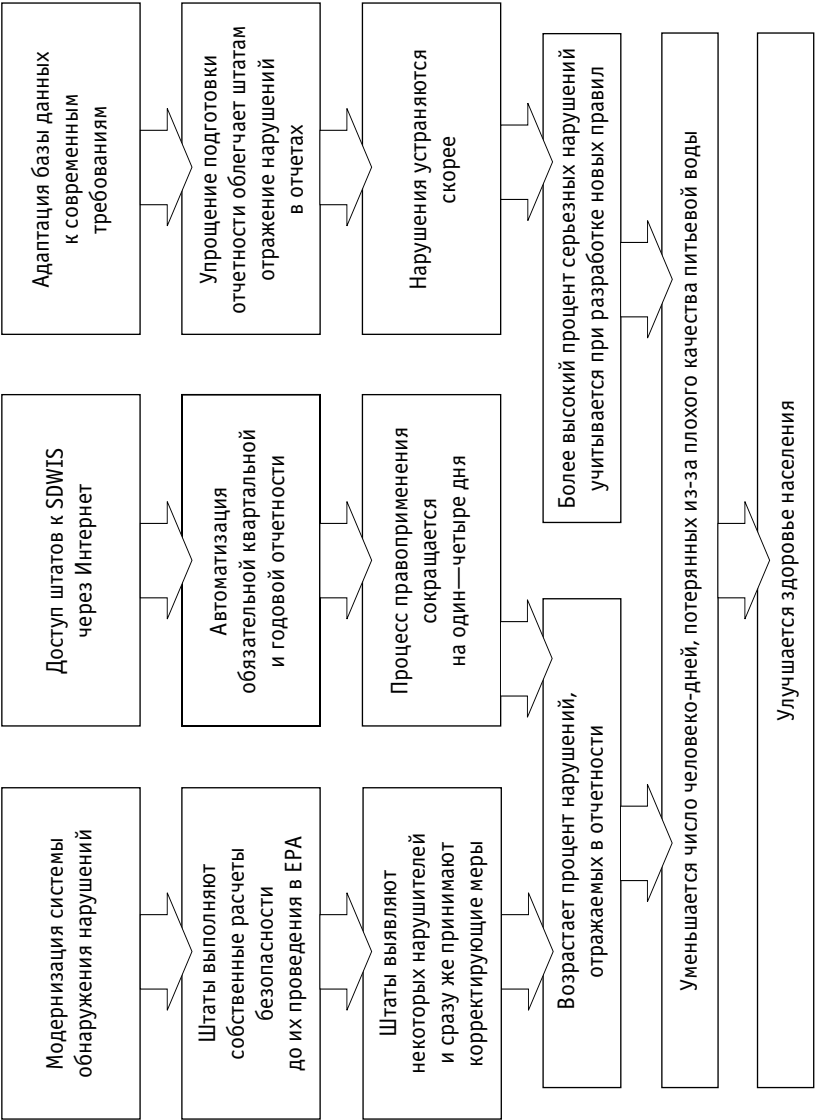


РИСУНОК 14.2. Табличная модель выгод от модернизации SDWIS

чальные оценки каждой переменной в модели. Результаты тренинга показали, что эксперты были хорошо калиброваны (то есть 90% ответов оказались в пределах указанных 90-процентных CI). Каждая величина, введенная в модель, имела свой уровень неопределенности, иногда выражаемый довольно широким интервалом. Например, одной из ожидаемых выгод был рост процента нарушений, отражаемых в отчетности (дело в том, что обычно не все случаи загрязнения воды попадают в официальные отчеты). Неопределенность, связанная с этим ростом, была весьма высока, поэтому эксперты указали 90-процентный доверительный интервал 5–55%.

Программа рассчитала доходность инвестиций для каждого из трех направлений модернизации SDWIS. На этом этапе мы уже имели детализированную модель, отражавшую исходное состояние неопределенности, оцененное экспертами.

Этап 2

На этапе 2 мы провели анализ стоимости информации. Хотя интервалы значений всех переменных оказались достаточно широкими, выяснилось, что измерения заслуживает лишь одна из них — среднее улучшение здоровья населения в результате введения новых норм безопасности питьевой воды. Система SDWIS предназначалась главным образом для того, чтобы следить за загрязнениями тщательнее и принимать корректирующие меры быстрее и эффективнее. Верхняя граница стоимости возможного улучшения здоровья населения в результате введения одного правила составляла 1 млрд дол. в год, но существовала вероятность и того, что она окажется меньше издержек соблюдения новой нормы. Иными словами, экономические преимущества от этих требований были настолько неопределенными, что имелась вероятность отрицательного конечного результата всех предпринятых усилий.

Но если новые правила не дадут положительного эффекта (разность стоимости положительного воздействия на здоровье и издержек соблюдения правил EPA), то бессмысленно внедрять их быстрее или лучше. Стоимость информации о темпах внедрения штатами новой технологии, принятии ими мер по росту эффективности, повышению процента нарушений, отражаемых в отчетности и т.п. оказалась равной нулю. Все, что нам нужно было сделать, — снизить исходную неопределенность относительно чистого экономического эффекта от введения норм безопасности питьевой воды. Но потенциальный эффект для здоровья населения (то есть верхние границы интервалов) был очень высок по сравнению с низкими издержками модернизации SDWIS. В результате порог измерения экономического преимущества оказался чуть выше нуля. Иными словами, фактически мы должны

были снизить неопределенность относительно того, окажется ли положительным чистый экономический эффект от введения новой политики контроля качества питьевой воды. Вот мы и решили заняться только этим и ничем другим.

Поскольку предыдущие исследования экономических последствий введения новых правил безопасности питьевой воды различались применявшимися методами, мы начали с простого инстинктивно-байесовского подхода, предполагающего более глубокое изучение результатов экономического анализа, выполненного до сих пор.

Причина, по которой калиброванные эксперты учли вероятность отрицательного результата введения новых норм, заключалась в том, что одно из проводившихся ранее исследований показало: одно конкретное требование не дало в свое время положительного экономического эффекта. При ближайшем рассмотрении оказалось, что авторы этого исследования рассматривали лишь консервативные экономические последствия загрязнения воды — дни временной потери трудоспособности и ущерб от этих потерь. Однако большинство людей согласились бы с тем, что заболеть намного хуже, чем не получить зарплату за несколько дней. В других исследованиях учитывали не только показатель недополученной заработной платы, но и готовность платить за то, чтобы не заболеть. Во всех работах, где использовались ВТР, экономический эффект новых правил оказался пусть и не очень значительным, но все же положительным.

В результате мы получили более подробную разбивку отдельных эффектов от введения каждой нормы контроля качества воды. Затем был указан калиброванный 90-процентный CI для того, какими оказались бы реальные выгоды от принятия самого бесполезного из введенных ранее правил, будь в нем учтены все благотворные последствия других норм EPA. Стало очевидно: вероятность отрицательного экономического эффекта от введения новой нормы контроля качества воды практически равна нулю. Мы скорректировали свою модель, чтобы отразить эти данные. Проведенный затем анализ стоимости информации показал, что экономическое обоснование любого направления модернизации SDWIS не требует проведения дополнительных измерений.

Этап 3

На этапе 3 мы прогнали окончательную моделирующую программу Монте-Карло для каждого направления инвестирования. Когда неопределенность экономического эффекта от введения новых норм контроля качества питьевой воды снизилась, оказалось, что реализация всех трех инвестиционных проектов крайне желательна. При этом возникла возможность улучшить намеченный ранее график их

осуществления. Потенциальная доходность проекта модернизации отражения нарушений в отчетности была очень высокой (среднее соотношение «выгоды/затраты» составляло в данном случае 3:1), однако имелась высокая (12%) вероятность отрицательной доходности. Вероятность отрицательной доходности двух других проектов составляла менее 1%.

Кроме того, мы выявили необходимость постоянного отслеживания ряда показателей. Самыми неопределенными переменными были темпы перехода пользователей в штатах на новую систему и скорость ее внедрения. В связи с этим данные переменные имели «остаточную стоимость информации» (то есть они представляли определенную, хотя и небольшую, ценность для людей, производивших оценку). Мы рекомендовали ЕРА быстрее реализовать два первых инвестиционных проекта и отложить модернизацию системы отражения нарушений в отчетности. Причем, прежде чем приступить к реализации третьего проекта, следовало учесть темпы перехода пользователей в штатах на новую систему в ходе реализации двух первых проектов. Если эти темпы окажутся достаточно низкими, то реализация проекта модернизации системы отчетности станет нецелесообразной (случай маловероятный, но возможный).

Эпилог

Марк Дей получил от прикладной информационной экономики именно то, что ожидал. Он сказал: «Использование программного обеспечения в оценке последствий для окружающей среды и здоровья населения произвело на меня колоссальное впечатление. То, что отдельные факторы, учтенные программными модулями, через цепь событий могут быть отслежены до определенных полезных последствий для населения, допускалось и ранее, но оценить данную связь количественно еще никому не удавалось. Думаю, все были искренне поражены тем, что кто-то сумел это сделать». Он также коснулся значения количественного анализа для процесса принятия решений. «Результат, который очень удивил меня, — уровень согласованности мнений у людей с диаметрально противоположными представлениями о том, что необходимо делать. На мой взгляд, достижение подобного консенсуса, несмотря на все трудности, — великолепный результат». По словам Марка Дея, этот процесс продемонстрировал и преимущества анализа стоимости информации: «Прежде никто не понимал самой концепции стоимости информации и не знал, что надо искать. Приходилось пытаться измерить все, средств на это не было, поэтому проще было ничего не делать. Число переменных быстро превысило возможности их оценить, поскольку непонятно было, какие из них действительно важны».

В отличие от Дея, Джеффу Брайану раньше не приходилось заниматься прикладной информационной экономикой. Он говорит: «Я активно возражал против этой затеи. Мне не хотелось отвлекать людей от того, чем они занимались, ради этого анализа, но результат оказался стоящим». Кроме того, он скептически относился и к калибровке экспертов, но, по словам Джеффа, «пройдя через этот процесс и увидев, что люди реагируют на оценки, я понял его важность». По мнению Брайана, наиболее полезной оказалась визуализация связи между информационной системой и целями программы. «График (см. рис. 14.2) не только показал связь SDWIS с улучшением здоровья населения, но и дал способ расчета стоимости выгод от ее использования. Я не думал, что одно лишь количественное определение проблемы приведет к чему-либо столь выразительному. Мне не удавалось доходчиво донести свою мысль, а метод AIE позволил сформулировать получаемые преимущества намного лучше. Даже не могу вам сказать, сколько раз я пользовался этим графиком». Наконец, что самое важное, Брайан довел это дело до конца. «Мы следовали всем последним рекомендациям, в том числе их содержанию и срокам».

Я привел здесь данный случай по двум причинам. Во-первых, это пример того, как «нематериальный» параметр — здоровье населения — подвергся количественной оценке для ИТ-проекта. Мне доводилось наблюдать, как при анализе многих подобных проектов из расчета ROI исключались намного более легкие для оценки эффекты по причине их «неизмеряемости». Во-вторых, пример демонстрирует, что многое измерять *и не следует*. Оказалось, что необходимо было снизить неопределенность только в отношении одной величины из 99. Для остальных 98 переменных вполне хватило первоначальных калиброванных оценок. А если бы VIA не проводился, то, как это обычно бывает, наверняка были бы рассчитаны малозначачие показатели (например, затраты и будущий рост производительности труда), а также параметры, связанные с самой большой неопределенностью типа возможного улучшения здоровья населения.

ПРИМЕР ИЗ ПРАКТИКИ:

ПРОГНОЗИРОВАНИЕ ПОТРЕБНОСТИ МОРСКОЙ ПЕХОТЫ В ТОПЛИВЕ

Осенью 2004 г. меня попросили решить с помощью прикладной информационной экономики задачу, сильно отличавшуюся от тех, с которыми я сталкивался раньше, работая с разными компаниями или государственными организациями. Управление научных исследований ВМС США (Office of Naval Research, ONR) и Корпус морской пехоты

США (U.S. Marine Corps, USMC) поручили одной многоуважаемой консалтинговой фирме найти способ повысить точность прогнозов потребности в топливе в условиях ведения боевых действий, которые составлялись военными плановиками и логистиками. В ходе операций в Ираке дневное потребление топлива только наземными подразделениями USMC составляло сотни тысяч галлонов (а авиация потребляла его втрое больше). При этом допустить развитие нежелательного сценария, когда неожиданно иссякнут запасы топлива, было никак нельзя, поскольку это ставит под угрозу и успех боевых операций, и безопасность морских пехотинцев на суше.

Для того чтобы иметь достаточно топлива на месте в любой момент, логистики и плановики должны были начинать подготовку за 60 дней. К сожалению, точно предсказать, какой будет потребность морской пехоты в топливе через два месяца, невозможно. С такой высокой неопределенностью и неприемлемым риском, что имеющихся запасов топлива окажется недостаточно, естественной реакцией было планировать поставки в объемах, в три-четыре раза превышающих ожидаемую потребность.

Старший уорент-офицер 5-го разряда (chief warrant officer 5, CWO5) Терри Каннеман, 27-летний ветеран USMC, в штабе морской пехоты отвечал за расчет потребности в топливе. Он сказал: «Мы знали, что учитываем старые и менее надежные факторы динамики потребления. Во время проведения OIF (Operation Iraqi Freedom — операции «Свобода Ираку») мы обнаружили, что традиционные методики работают плохо. Мусор на входе — мусор на выходе». Луис Торрес, руководивший в Управлении научных исследований ВМС анализом потребления топлива, видел те же проблемы: «Все это было связано с общим указанием сократить потребление топлива. Перед нами поставили задачу устранить внутренние ошибки процесса оценки».

Дополнительное количество топлива, необходимое для стратегического баланса, стало для логистики колоссальным бременем. Строилось множество складов горючего. Ежедневно топливо с одного склада перевозилось на другой, находившийся дальше от береговой линии. Эти склады и транспортные колонны создавали угрозу безопасности; охранявшим топливо морским пехотинцам приходилось рисковать собой.

Если бы USMC мог уменьшить имевшуюся неопределенность относительно потребности в топливе, то не пришлось бы хранить столько горючего и вероятность его нехватки не возрастала бы. В то время в USMC использовали довольно простую модель прогнозирования. Сначала подсчитывались вооружение и техника, имевшиеся в развернутых подразделениях, из полученного результата вычиталась техника, находящаяся в ремонте, переданная другим частям, уничтоженная противником и т.п. Затем определяли, какие боевые единицы в следующие 60 дней будут «наступать», а какие будут «выжидать или

обороняться». Как правило, атакующее подразделение, перемещаясь, сжигает больше топлива. Каждая единица боевой техники имеет свое среднее потребление горючего, измеряющееся в галлонах в час и в часах боевых действий в день. Если подразделение переходит в наступление, то время работы техники обычно возрастает. Для каждой боевой единицы рассчитывали дневное потребление горючего по данным его потребления боевой техникой и плану действий (атака или оборона). Затем определяли общее потребление горючего всеми подразделениями в каждый из 60 дней.

Точность и надежность данного подхода невысоки. Прогнозная потребность в топливе вполне могла занижаться вдвое или даже больше (отсюда и большие резервные запасы). Мне никогда прежде не приходилось прогнозировать запасы, необходимые для ведения военных действий, но я подошел к этой проблеме точно так же, как к любой другой важной задаче измерения, — воспользовался прикладной информационной экономикой.

Этап 0

На данном этапе я изучил результаты проведенных ранее исследований потребности войск в топливе. Ни в одном из них не описывались конкретные статистические методы прогнозирования. В лучшем случае обсуждались возможные способы, и то лишь в общих чертах. И все же я получил хорошее представление о характере проблемы. Мы отобрали несколько специалистов по логистике, способных принять участие в заседаниях рабочей группы, в том числе старшего уорент-офицера Каннемана и Луиса Торреса. Мы решили в течение трех недель провести шесть заседаний рабочей группы, рассчитанных на половину рабочего дня.

Этап 1

Первое заседание рабочей группы на этапе 1 было посвящено определению задачи прогнозирования. Только тогда стало ясно, что USMC интересовало общее потребление топлива в течение 60-дневного периода наземными силами одного только Экспедиционного соединения морской пехоты (Marine Expeditionary Force, MEF), насчитывавшего десятки тысяч солдат. Используя имевшиеся таблицы прогнозирования потребности в топливе, изученные нами на этапе 0, я построил несколько графиков, отвечавших на вопрос: «Куда поступает все топливо?» Эти графики дали членам команды (особенно нам, аналитикам, которые не работали с проблемой каждый день) представление о примерном объеме потребления горючего. Стало ясно, что его основ-

ная часть расходуется не танками и даже не бронетехникой в целом. Танк М-1 «Abrams» действительно сжигает галлон топлива каждую треть мили, но таких танков у MEF было только 58. А вот грузовиков в соединении насчитывалось более 1000 плюс более 1300 ставших теперь известными многоцелевых автомобилей повышенной проходимости (High Mobility Multipurpose Wheeled Vehicles, HMMWV), или хаммеров. Во время боевых действий грузовики сжигали в восемь раз больше топлива, чем танки.

Дальнейшее обсуждение того, что на самом деле делает эта техника, когда сжигает горючее, привело к разработке моделей трех разных видов. Из них самой сложной оказалась транспортная. Подавляющее большинство грузовиков и хаммеров сжигают львиную долю топлива, перемещаясь транспортными колоннами по определенным маршрутам. В составе таких колонн они в среднем дважды в день передвигались по круговому маршруту. Другая модель была «боевая». Бронетехника, например танки М1 и легкие бронемашины, тратя на перемещение по этим маршрутам меньше времени, обычно используют больше топлива в ходе проведения операций. Наконец, все генераторы, насосы и служебные автомашины обычно потребляли топливо более равномерно с меньшим удельным расходом. Для этой группы мы использовали уже существовавшую простую модель почасового потребления.

На одном из заседаний рабочей группы была проведена калибровка экспертов. Все показали хорошую способность оценивать вероятность наступления неизвестных событий. Они указывали интервалы значений для всех интересующих нас показателей, которым ранее присваивали лишь точечные значения. Например, если ранее считалось, что семитонный грузовик сжигает за час ровно 9,9 галлона топлива, то они указали интервал 7,8–12 галлонов. Для техники, обычно двигающейся в составе транспортной колонны, нам пришлось указать интервалы расстояний перемещения и учесть влияние дорожных условий на потребление горючего. Для бронемашин, участвующих в боевых операциях, мы должны были указать диапазон времени (при 60-дневном периоде), в течение которого они действительно участвовали в боях.

В результате все свелось к 52 величинам, которые требовались для расчета потребления горючего за 60-дневный период. Значения переменных были выражены в виде 90-процентных CI. В некотором отношении это мало отличалось от анализа проектов, выполненных мною до сих пор. Но вместо того чтобы рассчитать на основе этих величин денежный поток или доходность инвестиций, нам было нужно просто количественно определить общее потребление топлива за период. Модель Монте-Карло, в которую мы ввели эти интервалы, дала распределение возможных результатов, очень близкое к распределению фактических значений потребления топлива.

Этап 2

На этапе 2 мы провели анализ стоимости информации, воспользовавшись макросом программы Excel (хотя для этого вполне подошел бы и график стоимости информации, приведенный на рисунке 7.2). Поскольку решение не должно было выражаться в денежных прибылях или убытках, VIA дал результаты, означавшие, по сути, уменьшение ошибки прогнозной оценки (в галлонах) дневного потребления топлива. Оказалось, что наиболее высока стоимость информации об особенностях маршрутов транспортных колонн, в том числе сведений о расстояниях и дорожных условиях. Высокой оказалась и стоимость информации о влиянии военных операций на потребление топлива боевыми машинами. Мы придумали способы измерить и то, и другое.

Чтобы снизить неопределенность в потреблении топлива при боевых операциях, была выбрана модель линзы, построенная на основе оценок офицеров-логистиков Первой дивизии морской пехоты. В основном это были батальонные штабные офицеры и некоторые командиры подразделений, все с опытом боевых действий в иракской операции. Они назвали несколько факторов, от которых, по их мнению, зависит потребление топлива боевой техникой, и в том числе вероятность соприкосновения с противником (как это называется в планах проведения операций), знание местности, характер местности (город или пустыня) и т. п. Я провел со всеми офицерами тренинг по калибровке, затем составил список из 40 гипотетических сценариев боевых действий и предоставил им информацию о каждом из названных факторов. Для каждого сценария они указали 90-процентный CI величины расхода горючего теми видами боевой техники, которыми командовали (танками, легкими бронемашинами и др.). Собрав ответы, я пропустил через Excel регрессионную модель и получил формулу расчета потребления топлива каждым видом техники.

Чтобы уточнить переменные транспортной модели, характеризующие дорожные условия, мы решили провести ряд экспериментов на военной базе Twenty-Nine Palms (штат Калифорния). Другие подрядчики, участвующие в проекте, обеспечили нас GPS-навигаторами и счетчиками топлива, которые было решено установить на топливopроводах грузовиков. До этого эксперимента никто из команды не знал о существовании таких счетчиков. Я просто сказал этим консультантам: «Кто-то же постоянно этим занимается. Давайте проявим изобретательность и выясним, кто это делает и как». Незамедлительно по Интернету был найден поставщик цифровых счетчиков топлива, а его представители научили нас ими пользоваться. Они же придумали, как использовать получаемые данные в электронной таблице и синхронизировать работу GPS-навигаторов и счетчиков топлива. На про-

ведение дорожных испытаний и расчеты с помощью модели линзы, включая установку и доработку программы Excel, у трех человек ушло несколько недель с учетом времени на дорогу.

Навигаторы GPS и счетчики топлива были установлены на трех грузовиках двух моделей. Сначала мы опасались, что нужна более крупная выборка, но, вспомнив о принципе инкрементальности измерения, решили: просто посмотрим сначала, какой окажется дисперсия у этих грузовиков — ведь два из них в любом случае были одной модели. Навигаторы GPS и счетчики топлива фиксировали местонахождение грузовика и потребление топлива каждую секунду. Пока грузовик находился в движении, эта информация постоянно передавалась на встроенный портативный компьютер. Испытания проводились в разных условиях: на асфальтированных дорогах, пересеченной местности, на разной высоте над уровнем моря (участки базы располагались на разной высоте), на ровных и бугристых дорогах, скоростных автомагистралях и т.д. К тому времени, как мы закончили, у нас была таблица данных о расходе горючего в разных условиях, состоящая из 500 тыс. строк.

Мы пропустили полученные данные через очень большую регрессионную модель. Строк оказалось намного больше, чем могла обработать программа Excel 2003, но такая подробная информация нам и не требовалась. Мы объединили эти данные в блоки по шесть секунд и провели отдельный регрессионный анализ уже переформатированной таким образом информации для испытаний в разных условиях.

По окончании измерений мы обнаружили удивительные вещи. Прежде всего, основной причиной неточности прогнозов потребления топлива оказался недоучет качества дорог и некоторых других особенностей маршрута. Более того, большинство этих параметров (все, кроме температуры воздуха) обычно известны заранее, поскольку военные располагают подробной картой поля боя, сделанной спутниками и беспилотными самолетами-шпионами. Поэтому неопределенность, связанная с дорожными условиями, — ошибка, которой вполне можно избежать. В таблице 14.1 суммированы ошибки прогнозов из-за недоучета других конкретных факторов.

Не менее удивительные открытия позволила сделать модель для боевых машин. Оказалось, что их будущая потребность в горючем зависит вовсе не от вероятности соприкосновения с противником, а просто от того, дислоцировалось ли данное подразделение в этом районе ранее. Ведь если водители танков плохо ориентируются на местности, то они держат двигатели танков постоянно включенными. Они должны держать гидравлику под давлением, чтобы иметь возможность повернуть орудийную башню и избежать риска, пусть и незначительного, того, что в опасной ситуации двигатель не заведется. Другая боевая техника,

ТАБЛИЦА 14.1.

**Влияние изменения дорожных условий
на расход горючего, производимый
Экспедиционным соединением
морской пехоты**

Изменение дорожных условий	Изменение расхода горючего (галлоны/день)
Гравийное покрытие по сравнению с асфальтом	10 303
Увеличение средней скорости на 3 мили/ч	4685
Подъем дороги на 10 м	6422
Увеличение средней высоты над уровнем моря на 100 м	751
Рост температуры на 10 градусов	1075
Увеличение продолжительности пути на 10 миль	8320
Дополнительная остановка в пути	1980

как и танки, требует небольшого резерва горючего на случай потери ориентации на местности или перемещения по знакомым, но более протяженным маршрутам.

Информация о районе перемещения и дорожные условия — вот что специалисты по планированию должны учитывать всегда. Они знают, было ли подразделение в этом районе прежде. Учет данного фактора уменьшил ошибку прогноза дневного расхода топлива на 3 тыс. галлонов. А учет вероятности соприкосновения с противником сократил эту ошибку лишь на 2,4 тыс. галлонов — меньше, чем все, кроме трех, показатели, характеризующие состояние дороги. Фактически, приняв во внимание вероятность контакта с врагом, мы повысили точность прогноза лишь ненамного больше, чем в случае учета дополнительной остановки на дороге.

Этап 3

На этапе 3 мы разработали для специалистов по планированию и логистике табличный инструмент, куда были включены все новые факторы. В среднем он должен был уменьшить ошибку ранее использовавшегося метода прогнозирования вдвое. В результате, учитывая данные USMC о стоимости топлива (ведь доставить топливо к арене боевых действий гораздо дороже, чем в ближайшую к вам автозаправ-

ку), каждое экспедиционное соединение морской пехоты сэкономило бы ежегодно, по крайней мере, 50 млн дол. Во время написания этой книги в Ираке находились два таких соединения.

Эпилог

Это исследование кардинально изменило подход USMC к прогнозированию потребности в топливе. Даже самые опытные специалисты по планированию в области военной логистики сказали, что были поражены его результатами. Каннеман заявил: «Больше всего меня удивила транспортная модель, показавшая, что больше всего топлива сжигается на маршрутах, связанных с материально-техническим обеспечением. Было обнаружено, что водители танков не глушат двигатели, если не уверены, что смогут достать запасные стартеры. Это то, до чего ни один логистик наверняка не додумался бы и за сотню лет». Более «абстрактные выгоды» использования принципа «измерить можно все, что угодно», похоже, очевидны для старшего уорент-офицера 5-го разряда Каннемана: «Топливо стоит денег. Когда мне говорят, что эти данные трудно найти, я отвечаю, что не верю в это. Сколько вам стоят ошибки ваших прогнозов?» С ним согласен Луис Торрес: «Самым большим сюрпризом для меня стало то, что можно сэкономить такие объемы топлива. У нас появились свободные машины, потому что не нужно больше перевозить столько горючего. Для специалиста по логистике это очень важно. Теперь освободившиеся грузовики смогут перевозить боеприпасы».

Как и в случае SDWIS, это пример того, как можно уменьшить объем работы по измерению. Существовало много других величин, которые при ином подходе пришлось бы анализировать более детально, но нам удалось избежать этого. Пример показывает, как много можно сделать, если подойти к измерению практически, по принципу «берись и просто сделай это». Те самые блестящие программисты из рабочей группы, которые заявляли, что никогда не меняют сами масло в своей машине, закатали рукава и полезли под грязный грузовик, чтобы установить на нем счетчик топлива и навигатор GPS. В итоге оценить расход горючего оказалось нетрудно отчасти и потому, что мы никогда не сомневались: это возможно, если команда достаточно изобретательна. Этим наша работа резко отличалась от предыдущего исследования, проведенного Управлением научных исследований ВМС, которое больше походило на обычные консультации по менеджменту: масса сложных для восприятия понятий и идей, никаких измерений и никакой новой информации.

Главное же, что могло бы разубедить сомневающихся в возможности измерения, — это значение проведенных нами измерений для

здоровья и безопасности людей. Нам не пришлось рассчитывать для последнего проекта стоимость безопасности, и в частности стоимость безопасности морских пехотинцев (хотя мы и могли это сделать, воспользовавшись подходом готовности платить за блага или другими методами). Но если перевозится меньше горючего, то это означает, что меньше морских пехотинцев могут попасть в засаду или пострадать от заложенных на обочинах дорог бомб. Мне нравится думать, что, грамотно осуществив измерения, я, возможно, спас кому-то жизнь. Я рад, что мне не помешали это сделать чей-то страх и незнание существующих методов измерения.

СОВЕТЫ НАЧИНАЮЩИМ: НЕСКОЛЬКО ПОСЛЕДНИХ ПРИМЕРОВ

Мы говорили в этой книге (с рассмотрением конкретных примеров) об оценке результатов деятельности, безопасности, риска, рыночных прогнозов, стоимости информации, а также о подходах к определению стоимости здоровья и счастья. Я познакомил вас с основами эмпирических измерений, в том числе с такими понятиями, как случайная выборка, управляемый эксперимент и регрессионный анализ.

Может показаться, что эта книга перенасыщена информацией. Но, как и практически всегда в бизнесе или в жизни, нередко главное — начать с нескольких примеров, решить задачу от начала и до конца и увидеть результаты. Далее я хочу рассказать вам о нескольких возможных проблемах измерения, вам еще не знакомых. Я остановлюсь на них настолько подробно, насколько это необходимо, чтобы научить вас «идти по задаче» измерения от начала до конца.

Процесс решения подобных задач представляет собой стандартную последовательность шагов, хотя я могу и не упоминать здесь каждый этап отдельно. Я объясню, как решается каждая задача, но вам все равно придется выявить свою первоначальную неопределенность и стоимость информации, разложить переменные на составляющие и выбрать подходящий метод измерения. Однако вы получите достаточно информации, чтобы сделать свой первый шаг на этом пути.

Качество

Один менеджер, по его словам, член профессиональной ассоциации качества, спросил меня, как измерить качество. Он добавил, что споры об оценке качества постоянно ведутся на ежемесячных собраниях их ассоциации. Мне показалось это странным, поскольку тот, кого иногда называют «отцом качества», Эдвард Деминг, расценивал качество

как количество. Этот менеджер, похоже, слышал о Деминге, но не знал, что тот был статистиком. Деминг учил, что если у вас нет плана измерений, значит, нет и программы качества. Для него качество было постоянным соответствием ожиданиям. А несоответствие выявленным ожиданиям — это дефект. Оценка качества в производственном процессе сводилась, по мнению Деминга, к определению частоты возникновения различных дефектов и оценке отклонений от ожидавшейся нормы.

Для понимания концепции измерения качества я считаю точку зрения Деминга абсолютно необходимой, но, возможно, все-таки недостаточной. При всем уважении к Демингу я думаю, что полное определение качества должно включать в себя нечто большее. Ведь товар, произведенный с минимальными затратами, может идеально соответствовать ожиданиям производителя и все же считаться потребителями некачественным. А если потребители не считают продукт качественным, то мнение производителя по этому поводу значения не имеет. Полное представление о качестве любого продукта может дать только опрос потребителей.

Полезно вспомнить и о различии между заявленными и выявленными предпочтениями. Участвуя в опросе, потребители указывают свои предпочтения. Делая (или не делая) покупки, они тоже демонстрируют свои предпочтения. При этом лучшим выражением их мнения о качестве товара является премия, которую они готовы заплатить за это высочайшее качество. Этот «премиум-доход» можно сравнить с сэкономленными расходами на рекламу — ведь обычно люди готовы заплатить больше за товары, считающиеся высококачественными или остроумными даже без той дополнительной рекламы, которая потребовалась бы в ином случае. Качественные товары чаще покупают повторно, и потребители сами делают им хорошую рекламу. Все упомянутое до сих пор поддается, по крайней мере, выявлению путем проведения опроса, а толковый аналитик всегда может рассчитать нечто вроде подразумеваемой ценовой премии на основе покупательского поведения потребителей.

Определение стоимости процесса, отдела или службы

Смысловая нагрузка вопроса «Какова стоимость X?» почти так же сложна, как измерения, позволяющие на него ответить. Обычно определение стоимости кажется трудным делом из-за отсутствия четкого понимания того, для чего это делается. Иногда руководители информационных служб спрашивают меня, как определить стоимость какой-нибудь информационной технологии. Тогда я, в свою очередь,

задаю им ответный вопрос: «А вы что же, собираетесь от нее отказаться?» Ведь все проблемы, связанные с оценкой, будь то в бизнесе или правительственных учреждениях, представляют собой сравнение имеющихся альтернатив. Собираетесь попробовать рассчитать стоимость информационной технологии для компании? По-видимому, вам придется сопоставить затраты и выгоды от ее внедрения и затраты и выгоды работы без нее. Поэтому задаваться подобным вопросом стоит только в том случае, если вы действительно собираетесь обойтись без данной ИТ (или того, чью стоимость вы хотели узнать).

Бывает, однако, что директору информационной службы на самом деле нужно узнать, возросла ли стоимость ИТ с тех пор, как он получил это назначение. В таком случае следует сосредоточиться на расчете чистых выгод от конкретных решений и программ, принятых и реализованных за этот период. К задаче можно также подойти как к оценке эффективности в денежном выражении, о чем мы говорили в предыдущих главах. Если директор информационной службы интересуется стоимостью информационной технологии, потому что хочет найти аргументы против передачи ее функций на сторону, то его на самом деле интересует не стоимость ИТ, а стоимость сохранения отдела по сравнению со стоимостью аутсорсинга.

Обычно вопрос о стоимости возникает только тогда, когда появляются альтернативы. При правильной формулировке альтернатив и решения, которое надо принять, ответ на вопрос о стоимости будет для вас намного более очевиден.

Инновации

Как и все остальное, инновации, если только они реальны, всегда можно наблюдать. По аналогии с многими другими задачами измерения, главная проблема здесь, скорее, в определении решения, которое принимается. Что бы вы сделали иначе, зная заранее результаты оценки инноваций? Если вы можете назвать какое-нибудь реальное решение (например, при оценке работы команды или отдела исследований и разработок для того, чтобы премировать или уволить сотрудников), то читайте дальше. В противном случае никакого экономического смысла в оценке инноваций нет.

Если вам удалось сформулировать, по крайней мере, одно решение, которое будет принято по результатам этой оценки, я советую воспользоваться одним из трех возможных методов. Во-первых, всегда можно получить чисто субъективные, но контролируемые оценки. Привлеките независимых экспертов и используйте для коррекции ошибок модели Раша и другие имеющиеся инструменты. Один из них — испытание вслепую, когда специалист не знает, чье творчество

(рекламу, логотипы, научную работу, архитектурный проект и т. п.) он оценивает. Этот прием полезен, когда необходимо оценить качество исследований и разработок на основе портфеля генерированных идей. Некоторое представление о том, как это делается, мы дали в примере с компанией Mitre (см. главу 2).

Другой способ — использовать известные на тот момент, когда научная работа готовится к публикации, показатели качества инноваций, например число полученных патентов или опубликованных научных статей. В области библиометрии (изучения текстов, в том числе научных работ) применяется такой метод, как подсчет числа прямых и перекрестных упоминаний. Если человек пишет в статье о чем-то действительно революционном, то на его работу обычно ссылаются другие исследователи. В этом случае индекс цитируемости данного автора часто более информативен, чем число опубликованных им работ. Тот же метод используют при выдаче патентов, поскольку в заявке положено ссылаться на существующие аналогичные патенты, чтобы можно было выявить сходства и различия между ними и предлагаемым изобретением. Специалисты в области, называемой наукометрией, пытаются оценить научную эффективность¹. Правда, обычно они сравнивают целые компании или страны, но данный подход может пригодиться и вам.

Последний заслуживающий упоминания метод похож на обсуждавшийся нами в предыдущих главах метод оценки результатов деятельности в денежном выражении. Как сказал гуру с Мэдисон-авеню Дэвид Огилви, «если что-то не продается, значит, оно не креативно». Вещи могут казаться креативными, но не являться таковыми с точки зрения бизнеса. Если цель состояла в том, чтобы модернизировать решение бизнес-проблемы, то каковым оказался экономический (то есть, в конечном счете, финансовый) результат этого решения? Почему бы не оценивать эффективность ученых так, как Том Бейквелл оценивал производительность труда преподавателей или как Билли Бин измерял эффективность бейсболистов (см. главу 11)?

Доступность информации

Мне довелось, по крайней мере, четырежды моделировать доступность информации, и в конце концов в каждой модели оказывались одни и те же переменные. Более высокая доступность информации означает, что вы тратите меньше времени на ее поиск и реже теряете ее. Когда информация потеряна, вы либо обходитесь без нее, либо пытаетесь генерировать ее заново. Поиски документа либо попытки его повторного создания легко измерить через затраты времени на выполнение этой совершенно лишней и нежелательной работы. Если восстано-

вить утерянную информацию невозможно, то вы расплачиваетесь за это менее обоснованными и взвешенными решениями, которые чаще оказываются ошибочными. Для начала калиброванные эксперты могут указать интервалы для средних значений затрат времени на поиск документов, составление их заново или вообще на работу без утерянной информации.

Гибкость

Термин «гибкость» очень широк, неоднозначен и передает суть множества вещей. Здесь я просто расскажу о том, как определяли и оценивали гибкость три моих клиента. Каждый высказал совершенно отличное от других мнение, поэтому стоит привести кое-какие подробности.

Пример 1. Гибкость — это процент сокращения среднего времени реагирования на неожиданные проблемы с доступом к компьютерной сети (например, быстрее устранить вирус или решить задачу неожиданного роста числа запросов на вход в сеть).

Пример 2. Гибкость — это процент сокращения времени разработки нового продукта.

Пример 3. Гибкость — это способность при необходимости добавить новый пакет программ (в предыдущей информационной системе использовались собственные программы, несовместимые с приложениями на базе Oracle).

Все три определения гибкости относились к предполагаемым инвестициям в ИТ — либо в доработку инфраструктуры, либо в модернизацию программного обеспечения. Оценивая три соответствующих инвестиционных проекта, мы должны были, как обычно, рассчитать их годовую денежную стоимость в виде денежного потока, а затем чистую приведенную стоимость и доходность инвестиций.

Пример 1. *Годовая денежная стоимость пятилетнего проекта для расчета ROI =*

= Текущее время простоя (ч/год) × Средняя стоимость одного часа простоя × Сокращение простоев в результате внедрения новой системы.

Пример 2. *Годовая денежная стоимость семилетнего проекта для расчета ROI =*

= [Число новых продуктов, разработанных за год × Процент новых продуктов, выходящих на рынок × Текущее время разработки нового продукта (месяцы) × Дополнительная валовая прибыль от нового продукта, выведенного на рынок в предыдущем месяце + Рост себестоимости] × Снижение затрат времени.

Пример 3. *Годовая денежная стоимость пятилетнего проекта для расчета NPV =*

= Число новых приложений за год \times NPV дополнительного обслуживания собственных приложений в течение среднего срока их службы по сравнению со стандартным пакетом программ + Дополнительные краткосрочные затраты на разработку собственного приложения по сравнению со стандартным пакетом программ.

Поскольку все предлагаемые решения были крупными, а неопределенность — высокой, значения ожидаемой стоимости полной информации (EVPI) в данном случае составляли от сотен тысяч до миллионов долларов. Но, как это часто бывает, важнее всего оказалось измерить совсем не то, что выбрал бы сам клиент. Мы решили эти задачи измерения следующим образом.

Пример 1. Мы разработали для клиента опрос о последствиях простоя сети, который проводили после каждого из пяти сбоев с участием 30 респондентов. Клиент смог определить, повлияли ли сбои на производительность труда его сотрудников вообще, и если да, то сколько времени они не могли работать.

Пример 2. Мы разложили затраты времени на разработку нового продукта на девять конкретных составляющих, соответствующих видам деятельности, попросили калиброванных экспертов оценить затраты времени на каждое из них в процентах от общих временных издержек. Затем мы дали задание калиброванным экспертам, которые получили информацию о результатах дополнительных исследований, оценить сокращение по каждому виду деятельности.

Пример 3. Мы определили конкретные программные приложения, которые будут рассматриваться в ближайшие годы, и рассчитали рост себестоимости и эксплуатационные расходы на каждое из них по сравнению с аналогичным пакетом собственных программ.

В каждом случае затраты на измерение составляли менее 20 тыс. дол., что соответствовало 0,5–0,1% рассчитанной ожидаемой стоимости полной информации. В каждом случае первоначальная неопределенность была снижена на 40% или более. Дополнительный анализ стоимости информации показал, что дальнейшие измерения не будут иметь смысла. Измерения, проведенные для примеров 1 и 3, выявили необходимость дальнейших инвестиций. В случае же примера 2 была установлена высокая рискованность осуществления инвестиций. Доказать обоснованность реализации этого проекта удалось только после сокращения его масштабов и расходов в ходе апробации пилотного варианта.

Гибкость и теория опционов

В 1997 г. Нобелевская премия по экономике была присуждена Роберту Мертону и Майрону Скоулзу за разработку теории опционов и, в частности, формулы Блэка—Скоулза для оценки стоимости финансовых опционов. (Нобелевская премия присуждается только ныне живущим ученым; еще один автор формулы, Фишер Блэк, умер до ее присуждения.) В финансах опцион «колл» дает своему владельцу право, но не обязанность купить другой финансовый инструмент (акции, товар и т. д.) в какой-то момент в будущем по оговоренной цене. Аналогично, опцион «пут» дает своему владельцу право продать его по оговоренной цене. Если, например, у вас есть опцион «колл» на покупку акций по цене 100 дол. через месяц, а к тому времени акции будут продаваться уже по 130 дол., то, исполнив опцион и продав свои акции, вы немедленно заработаете 30 дол. Проблема состоит в том, что вы не знаете стоимости этих акций через месяц и будет ли данный опцион вообще что-то стоить. Лишь после появления формулы Блэка—Скоулза стало ясно, как осуществлять ценообразование подобных опционов.

Пресса подняла из-за этой теории намного больше шума, чем по поводу большинства экономических теорий, и ее стало модно применять не только для оценки опционов «колл» и «пут», но и для принятия решений в компаниях. Появилась так называемая теория реальных опционов, и многие менеджеры попытались представить процесс принятия решений в компаниях как решение задач по оценке опциона. В некоторых случаях (но далеко не всегда) этот подход действительно эффективен. Так, далеко не каждую проблему оценки преимуществ новой технологии можно представить в виде оценки опциона. На практике такой анализ в большинстве случаев сводится к применению не модели Блэка—Скоулза, а более традиционной теории принятия решений. Предположим, вы моделируете с помощью метода Монте-Карло новую программную платформу, позволяющую вносить изменения, если в будущем такие корректировки окажутся выгодными. Тогда модель продемонстрирует, что в среднем иметь эту возможность лучше, чем не иметь. Формула Блэка—Скоулза при этом не используется, но таковы в большинстве своем проблемы оценки реальных опционов. Использование формулы определения цены опциона на акцию оправдано только в случае, если вы можете объяснить, какое значение каждая переменная в формуле Блэка—Скоулза имеет в контексте вашей задачи. В этой формуле фигурируют цена исполнения, цена-страйк* и волатильность курса акций. Если непонятно, ка-

* Фиксированная цена, по которой покупатель опциона может использовать свое право купить или продать. — *Примеч. редактора.*

кие переменные вашей модели принятия решений соответствуют этим показателям, то формула Блэка—Скоулза, скорее всего, не годится (на дополнительном веб-сайте приводятся примеры оценки опционов с помощью и без помощи формулы Блэка—Скоулза).

ПОДВЕДЕМ ИТОГИ

Если вы считаете, что столкнулись с чем-то, что «невозможно измерить», то вспомните примеры SDWIS и USMC. На самом деле решить любую задачу по количественной оценке не так уж сложно, если серьезно обдумать ее.

1. Если нечто действительно важно, значит, вы можете это определить. Если нечто, по вашему мнению, вообще существует, значит, вы это уже каким-то образом наблюдали.
2. Если нечто является и важной, и неизвестной величиной, то существует вероятность ошибки в ее оценке и понесения затрат в случае такой ошибки.
3. Текущую неопределенность вы можете выразить количественно с помощью калиброванных оценок.
4. Рассчитать стоимость дополнительной информации можно, определив пороговое значение интересующего вас показателя, то есть такое его значение, при котором принимаемое решение будет отличаться от решения, возможного в отсутствие этих сведений.
5. Установив, что параметр заслуживает количественной оценки, вы сможете выбрать метод измерения и решить, сколько времени и сил следует потратить на его проведение.
6. Даже поверхностное знакомство с несколькими методами случайной выборки, управляемыми экспериментами или даже просто способами уточнения экспертных оценок позволяет существенно снизить неопределенность.

Оглядываясь назад, я спрашиваю себя: неужели какая-нибудь из обсуждавшихся нами «нерешаемых» задач измерения поставила бы в тупик Эратосфена, Энрико или Эмили? Лично мне сделанное ими демонстрирует, что они, по крайней мере, интуитивно понимали все положения теории измерения, которым в этой книге придается особое значение. Наверное, способы количественного определения текущей неопределенности, методы расчета стоимости информации и ее влияние на выбор приемов измерения оказались бы для них в новинку. Наши учителя в сфере измерения просто не могли быть знакомы с не-

которыми обсуждавшимися нами методами, но я подозреваю, что они все равно нашли бы способ сделать наблюдения, способные снизить неопределенность.

Надеюсь, если не что-то другое, так, по крайней мере, примеры Эратосфена, Энрико и Эмили, а также описанные в книге практические примеры заставят вас усомниться в правоте утверждения: нечто, критически важное для вашей компании, измерить невозможно.

■ ПРИМЕЧАНИЕ

- ¹ *Paul Stoneman et al. Handbook of the Economics of Innovation and Technological Change. Malden, MA: Basil Blackwell Ltd., 1995.*

ПРИЛОЖЕНИЕ

Тесты на калибровку
(и ответы на них)

ОТВЕТЫ НА ВОПРОСЫ ТЕСТА ИЗ ГЛАВЫ 5

№	Вопрос	Ответ
1	Какую скорость развил в 1938 г. британский паровоз, установивший новый рекорд скорости?	126
2	В каком году сэр Исаак Ньютон опубликовал свою работу о всемирном законе тяготения?	1685
3	Какова длина в дюймах обычной визитной карточки?	3,5
4	В каком году был создан Интернет (в те времена Арпанет) как военная система коммуникации?	1969
5	В каком году родился Уильям Шекспир?	1564
6	Каково расстояние в милях по воздуху между Нью-Йорком и Лос-Анджелесом?	2451
7	Какой процент площади квадрата покрывает круг, диаметром равный его стороне?	78,5%
8	В каком возрасте умер Чарли Чаплин?	88
9	За сколько дней Луна совершает полный оборот вокруг Земли?	27,32
10	Какого числа по телевидению впервые была показана программа «Gilligan's Island» («Остров Гиллигана»)?	26 сентября 1964 г.
	Утверждение	Ответ
1	Древних римлян завоевали древние греки	Неверно
2	Трехгорбых верблюдов не бывает	Верно
3	Галлон нефти весит меньше галлона воды	Верно
4	Марс всегда удален от Земли более, чем Венера	Неверно
5	Команда «Boston Red Sox» стала первым победителем чемпионата США по бейсболу	Верно
6	Наполеон родился на острове Корсика	Верно
7	«М» — одна из трех наиболее часто используемых букв	Неверно
8	В 2002 г. персональный компьютер стоил в среднем не более 1500 дол.	Верно
9	Прежде чем стать президентом, Линдон Джонсон был губернатором	Неверно
10	Килограмм больше фунта	Верно

Другие тесты на калибровку вы найдете на следующих страницах.

ДОПОЛНИТЕЛЬНЫЕ ТЕСТЫ НА КАЛИБРОВКУ

Калибровочное исследование на интервалы значений: А

№	Вопрос	Нижняя граница (вероятность того, что реальное значение выше, 95%)	Верхняя граница (вероятность того, что реальное значение ниже, 95%)
1	Какова высота плотины Гувера (в футах)?		
2	Какова длина 20-долларовой банкноты (в дюймах)?		
3	Какой процент алюминия в США перерабатывается для вторичного использования?		
4	Когда родился Элвис Пресли?		
5	Каков массовый процент кислорода в атмосфере?		
6	Какова широта Нового Орлеана? Подсказка: широта экватора 0 градусов, а широта Северного полюса — 90 градусов		
7	Сколькими самолетами располагали ВВС США в 1913 г.?		
8	В каком году был изобретен первый печатный станок в Европе?		
9	Какой процент всего потребленного домашними хозяйствами США электричества в 2001 г. можно отнести на электроприборы?		
10	Какова высота горы Эверест (в милях)?		
11	Какова протяженность ирано-иракской границы (в километрах)?		
12	Какова длина Нила (в милях)?		
13	В каком году был основан Гарвардский университет?		
14	Каков размах крыльев «Боинга-747» (в футах)?		

Окончание таблицы

№	Вопрос	Нижняя граница (вероятность того, что реальное значение выше, 95%)	Верхняя граница (вероятность того, что реальное значение ниже, 95%)
15	Сколько солдат было в римском легионе?		
16	Какова средняя температура (в градусах Фаренгейта) в океанских впадинах, на глубине более 6500 футов?		
17	Какова длина орбитальной ступени «Шаттла» без внешнего топливного бака (в футах)?		
18	В каком году Жюль Верн опубликовал свою книгу «220 тысяч лье под водой»?		
19	Какова ширина ворот в хоккее на траве (в футах)?		
20	Сколько зрителей вмещал Римский Колизей?		

Ответы см. на с. 288.

Ответы на калибровочные тесты на интервалы значений: А

№	Ответ	№	Ответ
1	738	11	1458
2	63/16 (6,1875)	12	4160
3	45%	13	1636
4	1935	14	196
5	21%	15	6000
6	31	16	39°F
7	23	17	122
8	1450	18	1870
9	26,7%	19	12
10	5,5%	20	50 000

Калибровочное исследование на интервалы значений: Б

№	Вопрос	Нижняя граница (вероятность того, что реальное значение выше, 95%)	Верхняя граница (вероятность того, что реальное значение ниже, 95%)
1	В каком году аппарат «Викинг-1» впервые взял пробу грунта с Марса?		
2	Сколько лет было самому молодому космонавту?		
3	Какова высота Сирс-Тауэр (в метрах)?		
4	Какой была максимальная высота воздушного шара «Breitling Orbiter-3», совершившего первый в истории человечества кругосветный полет (в милях)?		
5	Какой в среднем процент общих усилий по разработке нового программного обеспечения затрачивается на создание самого проектного решения?		
6	Сколько людей было переселено из зон отчуждения после Чернобыльской катастрофы?		
7	Какова длина самого большого из существовавших в мире самолета (в футах)?		
8	Каково расстояние по воздуху между Сан-Франциско и Гонолулу (в милях)?		
9	Какую скорость развивает самая быстрая птица в мире — сокол, когда круто пикирует вниз (в милях/час)?		
10	В каком году была открыта двойная спиральная структура молекулы ДНК?		
11	Какова ширина футбольного поля (в ярдах)?		

Окончание таблицы

№	Вопрос	Нижняя граница (вероятность того, что реальное значение выше, 95%)	Верхняя граница (вероятность того, что реальное значение ниже, 95%)
12	На сколько процентов число интернет-хостов увеличилось в 1997 г. по сравнению с 1996 г.?		
13	Сколько калорий в восьми унциях апельсинового сока?		
14	С какой скоростью (в милях/час) надо двигаться на уровне моря, чтобы преодолеть звуковой барьер?		
15	Сколько лет Нельсон Мандела провел в тюрьме?		
16	Каково среднее дневное потребление калорий в развитых странах?		
17	Сколько стран — членов ООН было в 1994 г.?		
18	В каком году в США было образовано Национальное общество имени Одюбона *?		
19	Какова высота (в футах) самого высокого самого крупного в мире водопада (Анхель в Венесуэле)?		
20	На какой глубине (в милях) был обнаружен «Титаник»?		

Ответы см. на с. 291.

Все еще не калиброваны? Дополнительные калибровочные тесты можно найти на сайте: www.howtomeasureanything.com

* Целью общества является охрана местобитания диких животных, воздуха, воды, заболоченных районов. — *Примеч. редактора.*

**Ответы на калибровочное исследование
на интервалы значений: Б**

№	Ответ	№	Ответ
1	1976	11	53,3
2	26	12	70%
3	443	13	120
4	6,9	14	760
5	20%	15	26
6	135 000	16	3300
7	803	17	184
8	2394	18	1905
9	150	19	3212
10	1953	20	2,5 мили

Калибровочное исследование на бинарные вопросы: А

№	Утверждение	Ответ «Правильно/ Неправильно»	Степень уверенности в правильности ответа (обведите нужное кружком)
1	Шоссе Линкольна было первой асфальтированной дорогой в США, которая соединяла Чикаго и Сан-Франциско		50% 60% 70% 80% 90% 100%
2	Шелковый путь соединял древние царства Китая и Афганистана		50% 60% 70% 80% 90% 100%
3	У американских семей больше микроволновых печей, чем телефонов		50% 60% 70% 80% 90% 100%
4	Слово «дорический» — архитектурный термин, обозначающий форму крыши		50% 60% 70% 80% 90% 100%
5	Всемирная туристическая организация предсказывает, что в 2020 г. самым популярным туристическим направлением будет Европа		50% 60% 70% 80% 90% 100%
6	Германия — вторая страна в мире, сумевшая создать атомное оружие		50% 60% 70% 80% 90% 100%
7	Хоккейная шайба поместится в лунку для игры в гольф		50% 60% 70% 80% 90% 100%
8	Сиу — это индейское племя, населявшее равнины		50% 60% 70% 80% 90% 100%
9	Для физика плазма — это разновидность твердого тела		50% 60% 70% 80% 90% 100%
10	Столетняя война на самом деле продолжалась более 100 лет		50% 60% 70% 80% 90% 100%
11	Основная часть пресной воды на Земле заморожена в ледяных шапках Северного и Южного полюсов		50% 60% 70% 80% 90% 100%
12	Премия «Оскар» стали вручать более века назад		50% 60% 70% 80% 90% 100%
13	В мире меньше 200 миллиардеров		50% 60% 70% 80% 90% 100%
14	В программе Excel “^” означает «Вывести показатель степени»		50% 60% 70% 80% 90% 100%

Окончание таблицы

№	Утверждение	Ответ «Правильно/ Неправильно»	Степень уверенности в правильности ответа (обведите нужное кружком)
15	Средняя годовая зарплата пилотов самолетов превышает 150 000 дол.		50% 60% 70% 80% 90% 100%
16	В 1997 г. состояние Билла Гейтса оценивалось более чем в 10 млрд дол.		50% 60% 70% 80% 90% 100%
17	Пушки использовались европейскими армиями до XI века		50% 60% 70% 80% 90% 100%
18	Анкоридж — столица Аляски		50% 60% 70% 80% 90% 100%
19	На горе Рашмор высечены головы четырех президентов — Вашингтона, Джефферсона, Линкольна и Гранта		50% 60% 70% 80% 90% 100%
20	John Wiley & Sons — не самая крупная издательская компания		50% 60% 70% 80% 90% 100%

Ответы см. на с. 294.

**Ответы на калибровочное исследование
на бинарные вопросы: А**

№	Ответ	№	Ответ
1	Правильно	11	Правильно
2	Неправильно	12	Неправильно
3	Неправильно	13	Неправильно
4	Правильно	14	Правильно
5	Правильно	15	Неправильно
6	Неправильно	16	Правильно
7	Правильно	17	Неправильно
8	Правильно	18	Неправильно
9	Неправильно	19	Неправильно
10	Правильно	20	Правильно

Калибровочное исследование на бинарные вопросы: Б

№	Утверждение	Ответ «Правильно/ Неправильно»	Степень уверенности в правильности ответа (обведите нужное кружком)
1	«Большое красное пятно» Юпитера больше Земли		50% 60% 70% 80% 90% 100%
2	Название бейсбольной команды «Brooklyne Dodgers» произошло от trolley car dodgers *		50% 60% 70% 80% 90% 100%
3	Сверхзвуковая скорость выше дозвуковой		50% 60% 70% 80% 90% 100%
4	Полигон трехмерен, а полиэдр — двумерная структура		50% 60% 70% 80% 90% 100%
5	Электрический мотор мощностью 1 ватт производит энергию, равную 1 лошадиной силе		50% 60% 70% 80% 90% 100%
6	В Чикаго жителей больше, чем в Бостоне		50% 60% 70% 80% 90% 100%
7	В 2005 г. продажи Wall-Mart упали ниже 100 млрд дол.		50% 60% 70% 80% 90% 100%
8	Блоки стикеров Post-it были изобретены компанией 3M		50% 60% 70% 80% 90% 100%
9	Альфред Нобель, из средств которого выплачивается Нобелевская премия мира, заработал свое состояние на нефти и взрывчатых веществах		50% 60% 70% 80% 90% 100%
10	BTU — это единица измерения теплоты		50% 60% 70% 80% 90% 100%
11	Средняя скорость победителя первой гонки «Индианаполис-500» была ниже 100 миль/ч		50% 60% 70% 80% 90% 100%
12	В компании Microsoft работает больше человек, чем в IBM		50% 60% 70% 80% 90% 100%
13	Румыния граничит с Венгрией		50% 60% 70% 80% 90% 100%

* Бруклинские пешеходы, катающиеся «зайцами» на подножках трамваев. —
Примеч. редактора.

Окончание таблицы

№	Утверждение	Ответ «Правильно/ Неправильно»	Степень уверенности в правильности ответа (обведите нужное кружком)
14	Штат Айдахо по площади превышает Ирак		50% 60% 70% 80% 90% 100%
15	Касабланка находится в Африке		50% 60% 70% 80% 90% 100%
16	Первая искусственная пластмасса была изобретена в XIX веке		50% 60% 70% 80% 90% 100%
17	Серна — это горное животное		50% 60% 70% 80% 90% 100%
18	Основанием пирамиды является квадрат		50% 60% 70% 80% 90% 100%
19	Стоунхендж находится на самом большом из британских островов		50% 60% 70% 80% 90% 100%
20	Мощность компьютерных процессоров удваивается раз в три месяца		50% 60% 70% 80% 90% 100%

Ответы см. на с. 297.

Все еще не калиброваны? Дополнительные калибровочные тесты можно найти на сайте: www.howtomeasureanything.com

**Ответы на калибровочное исследование
на бинарные вопросы: Б**

№	Ответ	№	Ответ
1	Правильно	11	Правильно
2	Правильно	12	Неправильно
3	Правильно	13	Правильно
4	Неправильно	14	Неправильно
5	Неправильно	15	Правильно
6	Правильно	16	Правильно
7	Неправильно	17	Правильно
8	Правильно	18	Правильно
9	Правильно	19	Правильно
10	Правильно	20	Неправильно

Издательство «Олимп—Бизнес»

105005, Москва, ул. Ф. Энгельса, д. 46

Тел./факс: (495) 411-90-14 (многоканальный)

Интернет-магазин: www.olbuss.ru

e-mail: sales@olbuss.ru

Как купить наши книги:

- В интернет-магазине издательства: www.olbuss.ru
- Сделать заказ по телефону (495) 411-90-14
- Приехать в офис издательства «Олимп—Бизнес»

***Спрашивайте книги нашего издательства
в магазинах вашего города***

Издательство «Олимп—Бизнес»
приглашает к сотрудничеству оптовиков,
книготорговые организации и магазины.
Информацию об условиях работы можно получить
в отделе продаж издательства

Дуглас У. Хаббард

**Как измерить все, что угодно.
Оценка стоимости нематериального в бизнесе**

Издатель В. Стабников

Перевел с английского Е. Пестерева

Редактор Е. Чуракова

Корректор Н. Шерстенникова

Компьютерная верстка С. Родионова

Художник В. Коршунов

Подписано в печать 04.06.2009.

Формат 70×100 1/16. Бумага офсетная № 1.

Гарнитура «Таймс». Печать офсетная.

Печ. л. 20,0. Уч.-изд. л. 16,0. Заказ №

Издательство «Олимп—Бизнес».

105005, Москва, ул. Ф. Энгельса, д. 46.

Отпечатано в ООО «Август Борг».

121009, Москва, Шубинский пер., д. 2/3.