

ТОМАС ДЭВЕНПОРТ

КИМ ДЖИН ХО



# О ЧЕМ ГОВОРЯТ ЦИФРЫ

КАК ПОНИМАТЬ И ИСПОЛЬЗОВАТЬ ДАННЫЕ



Thomas H. Davenport and Jinho Kim

# **Keeping Up With The Quants**

YOUR GUIDE TO UNDERSTANDING  
AND USING ANALYTICS

HARVARD BUSINESS REVIEW PRESS

Томас Дэвенпорт и Ким Джин Хо

# **О чем говорят цифры**

**КАК ПОНИМАТЬ  
И ИСПОЛЬЗОВАТЬ ДАННЫЕ**

Издательство  
«Манн, Иванов и Фербер»

## Информация от издательства

*Издано с разрешения издательства Harvard Business Review Press (США) при участии Агентства Александра Корженевского (Россия)*

*На русском языке публикуется впервые*

### **Дэвенпорт, Т.**

О чем говорят цифры. Как понимать и использовать данные / Томас Дэвенпорт, Ким Джин Хо ; пер. с англ. Э. Кондуковой — М. : Манн, Иванов и Фербер, 2014.

ISBN 978-5-00057-118-7

Неважно, чем вы интересуетесь, в какой сфере работаете и каких размеров ваша компания — цифры и аналитика сегодня повсюду, и всем приходится иметь с ними дело. Эта книга в доступной форме познакомит вас с количественным анализом, его терминами и методами, поможет развить аналитические навыки и разговаривать на одном языке с количественными аналитиками.

*Все права защищены. Никакая часть данной книги не может быть воспроизведена в какой бы то ни было форме без письменного разрешения владельцев авторских прав. Правовую поддержку издательства обеспечивает юридическая фирма «Вегас-Лекс»*

© 2013 Harvard Business School Publishing Corporation

© Перевод на русский язык, издание на русском языке, оформление. ООО «Манн, Иванов и Фербер», 2014

## ГЛАВА 1

# Почему аналитические способности нужны каждому

Мы живем в мире информации, объем которой нарастает с поразительной скоростью — все больше и больше данных сваливается на наши головы, причем большую их часть собирают, чтобы улучшить качество принимаемых решений в бизнесе, государственном управлении или общественной деятельности. Если нам не удастся это сделать методами количественного анализа, тогда данные пропадают впустую, а уровень эффективности нашей деятельности далек от желаемого. Цель этой книги в том, чтобы показать читателю, как работает количественный анализ (даже если у читателя нет математической подготовки) и как использовать его для улучшения принимаемых решений.

### Актуальность данных и аналитических исследований

Рост объема информации характерен почти для каждой сферы общественной деятельности. Если вы занимаетесь спортом, то наверняка знаете о *бейсбольной статистике Moneyball*<sup>[1]</sup> и революции в профессиональном бейсболе, которую позволил совершить анализ данных об эффективности действий отдельных игроков. Сейчас такая статистика внедрена практически во всех популярных видах спорта. Если вы увлекаетесь сетевыми компьютерными играми, то наверняка знаете, что разнообразные сведения о вашем игровом поведении накапливаются и анализируются компании Zynga и Electronic Arts. Любите кино? Возможно, слышали о методике, применяемой компанией Netflix для прогнозирования предпочтений в области кино. Может быть, вы не знаете, что некоторые голливудские киностудии (например, Relativity Media) используют похожие методики, принимая решение о том, какие кинопроекты финансировать.

Важно, что существуют различные типы данных. Некоторые

используют для бухгалтерского учета. Например, ваша компания наверняка сохраняет информацию о том, с какого числа вы зачислены в штат или сколько дней ежегодного отпуска использовали. Но по мере накопления все большего объема данных того или иного типа у компаний возникает понятное желание найти им полезное применение, в частности использовать для принятия решений. Обычно активно пользуются базой данных о персонале. На основе этих данных руководство может задать вопрос: какое количество сотрудников, возможно, уволится в следующем году? Есть ли взаимосвязь между полным использованием отпуска за текущий год и результатами работы сотрудника за тот же период?

Но сбор данных и аналитика не просто способствуют принятию оптимальных управленческих решений. Многие интернет-компании — Google, Facebook, Amazon, eBay и прочие — используют так называемые большие данные о текущих онлайн-операциях не только когда нужно обосновать принимаемое решение, но и при разработке новых продуктов и их свойств. Стремитесь ли вы создать дополнительную ценность для потребителей или оптимизировать принимаемые решения — все это задачи для аналитиков. Они обобщают данные, ищут в них внутреннюю логику и на ее основе разрабатывают модели. Найти в данных логику и эффективно их использовать можно лишь с помощью методов математического или статистического анализа, обобщенно называемых *аналитикой*.

## Что такое аналитика

**Аналитикой** мы называем всестороннее использование баз данных, статистический и количественный анализ, объяснительные и прогнозные модели, а также доказательный менеджмент, применяемые для поддержки решений и увеличения ценности для потребителей.

В зависимости от цели и методов аналитику можно разделить на описательную (дескриптивную), предсказательную (предикативную) и нормативную (прескриптивную). **Описательная аналитика** включает сбор, систематизацию, представление данных в табличной форме, а затем выделение их основных характеристик. Этот вид аналитики всегда ориентировался на информирование о характеристиках данных. Он может оказаться весьма полезным, но ничего не говорит о причинах сложившейся ситуации или о том, что произойдет в будущем.

**Предсказательная аналитика** выходит за рамки простого описания данных и

зависимостей между переменными (в виде показателей, которые могут иметь целый ряд значений) и прогнозирует динамику показателей в будущем на основе данных за прошлые периоды. Сначала определяются связи между переменными, а затем на основе их анализа оценивается вероятность того или иного события: например, насколько вероятно, что потребитель отреагирует на рекламу и купит данный продукт. Хотя связи между переменными используются для прогнозирования будущего, явная причинно-следственная связь обнаруживается далеко не всегда. По сути, она совсем не обязательна для получения точного прогноза.

**Нормативная аналитика** ориентируется на более широкий круг задач и включает такие методы, как проведение экспериментов и оптимизация. Подобно тому как доктор выписывает рецепт, нормативная аналитика предлагает направление действий. **Эксперимент** призван ответить на вопросы о **причине** тех или иных явлений. Чтобы с уверенностью делать выводы о причинных связях, исследователи изменяют одну или несколько независимых переменных и наблюдают реакцию зависимой переменной, одновременно контролируя внешние по отношению к исследуемой системе факторы. Если тестовая группа, подчиняющаяся условию эксперимента, показывает существенно лучшие результаты по сравнению с контрольной группой, то ответственный менеджер может принять решение о широком внедрении этого условия.

Еще один вид нормативной аналитики — **оптимизация**. Она направлена на выявление оптимального значения конкретной переменной во взаимосвязи с другой переменной. Например, нам необходимо рассчитать цену продукта, обеспечивающую максимальную рентабельность его продаж. В розничной торговле оптимизационный подход таким же образом позволяет выявить уровень запасов, гарантирующий отсутствие сбоев из-за временного отсутствия какого-либо товара.

В зависимости от применяемых методов и видов данных, подлежащих сбору и анализу, аналитику можно разделить на количественную и качественную. Цель **качественной аналитики** состоит в углубленном понимании причин и мотивов тех или иных явлений. Обычно для этого собирают ограниченное количество неструктурированных данных на основе нерепрезентативной выборки<sup>[2]</sup>, а анализ проводят нестатистическими методами. Качественная аналитика полезна при проведении поисковых исследований, то есть на первом этапе исследований аналитических. **Количественная аналитика** представляет собой систематическое изучение событий при помощи статистических, математических и вычислительных процедур. На основе сбора информации о большом количестве репрезентативных событий с последующей статистической обработкой обычно получают структурированные данные.

В зависимости от целей исследователи применяют различные аналитические методы:

**статистика** — сбор, систематизация, анализ, интерпретация и оглашение данных;

**прогнозирование** — оценка динамики той или иной переменной в определенный момент в будущем на основе данных о ее динамике в прошлом;

**интеллектуальный анализ данных (Data mining)** — автоматизированное или полуавтоматизированное выявление ранее неизвестных зависимостей в больших массивах данных с помощью специальных вычислительных алгоритмов или статистических методов;

**интеллектуальный анализ текстов** — выявление неизвестных зависимостей или тенденций в тексте методами, подобными интеллектуальному анализу данных;

**оптимизация** — использование математических методов для того, чтобы найти оптимальные решения на основе заданных критериев и установленных ограничений.

**эксперимент** — формирование тестовой и контрольной групп методом случайного отбора и выявление причин и степени влияния независимых переменных на зависимую переменную.

В этом списке приведены широко известные аналитические методы, причем многие из них используют одни и те же аналитические приемы и процедуры. Например, регрессионный анализ — наиболее распространенный аналитический прием в предсказательной аналитике — не менее популярен и в статистике, прогнозировании и интеллектуальном анализе данных. Точно так же анализ временных рядов, специальная аналитическая процедура из арсенала статистики, предназначенная для анализа меняющихся во времени значений переменных, используется не только в статистике, но и в прогнозировании.

Учетные данные, помогающие принимать решения по персоналу (мы уже говорили о них), являются структурированными (легко представляются в виде таблицы), количественными и относительно небольшими по объему (не более терабайта или двух даже в очень крупных компаниях). Такие данные традиционно использовались в аналитике, поэтому назовем их *малыми данными*. Долгое время аналитики ни с чем другим дела не имели.

Но сегодня крупные компании, некоммерческие организации и даже стартапы сталкиваются с так называемыми *большими данными* — неструктурированными массивами информации колоссальных объемов. Их источниками могут быть онлайн-обсуждения в интернете, видеоматериалы или данные анализа ДНК пациентов больницы. У данных такого рода объем намного больше — иногда тысячи петабайт<sup>[3]</sup>. Например, Google обрабатывает порядка 24 петабайт интернет-данных ежедневно, а AT&T<sup>[4]</sup> передает по телекоммуникационным сетям около 30 петабайт музыки и прочих данных в день. Благодаря новым прикладным компьютерным программам и техническим новшествам мы можем анализировать огромные массивы данных и извлекать из них полезную информацию.

## Что такое большие данные

Термин **большие данные** применяется для обозначения данных уникально большого объема или неструктурированных данных. Приведем несколько примеров:

- За месяц 600 миллионов пользователей Facebook добавили в сеть 30 миллиардов единиц контента.
- Компания Zynga, занимающаяся сетевыми виртуальными играми, ежедневно обрабатывает более петабайта игровой информации.
- Пользователи YouTube просматривают более двух миллиардов видеоклипов в день.
- Пользователи Twitter выполняют 32 миллиарда поисковых запросов в месяц.
- Пользователи Google в 2011 году выполняли почти 5 миллиардов поисковых запросов в день.
- В 2009 году ежедневно отсылалось более 2,5 миллиарда текстовых сообщений.
- В 2010 году население планеты использовало 5 миллиардов мобильных телефонов.
- Объем файла с полной расшифровкой человеческого генома составляет около одного терабайта.
- Беспроводной датчик для контроля физических параметров одной коровы передает около 200 мегабайт данных в год.
- В 2008 году количество подключенных к интернету устройств превысило численность населения земного шара.
- По оценке компании Cisco Systems, к концу 2011 года двадцать типичных домохозяйств генерировали более интенсивный интернет-трафик, чем все пользователи интернета в 2008 году.
- McKinsey & Company считает, что почти в каждой отрасли американской экономики компании с численностью персонала более 1000 человек накапливают в среднем больший объем информации, чем Библиотека Конгресса США.

Большие данные и основанная на них аналитика способны существенно изменить практически каждую отрасль экономики и бизнес-процессы в течение следующих десяти лет. Любая организация (и любой ее сотрудник), если вовремя ознакомится с сутью и методами обработки больших данных, получит огромное конкурентное преимущество. Точно так же как компании, в свое время первыми освоившие методы обработки малых данных, опередили своих конкурентов, сейчас на позиции лидеров выйдут те, кто раньше других сумеет использовать возможности больших данных.

Потенциал больших данных можно реализовать благодаря глобальным устройствам их сбора и обработки. Сенсоры и микропроцессоры в скором времени будут везде. Практически каждое механическое или электронное устройство регистрирует свои действия, местонахождение или состояние. Эти устройства и люди, их

эксплуатирующие, поддерживают связь через интернет, а это еще один колоссальный источник данных. Если добавить к этому объемы информации, проходящие через прочие средства связи (беспроводные и проводные телефонные линии, кабели, спутники и т. п.), трудно даже оценить все перспективы.

Доступность всех этих данных означает, что практически любая предпринимательская или управленческая деятельность может рассматриваться либо как проблема больших данных, либо как возможность их обработки. Производство, в котором значительная часть оборудования оснащена одним или несколькими микропроцессорами, все чаще становится средой, где функционируют большие данные. Потребительский маркетинг с мириадами покупок и историй посещений покупателей также сталкивается с проблемой их обработки. Google даже описывала свою самоходную повозку из будущего как проект, связанный с обработкой больших данных.

Гэри Лавмен, CEO<sup>[5]</sup> компании Caesars Entertainment (известный своим выражением «Мы так полагаем или мы знаем?»), глава Amazon Джефф Безос («Мы никогда не пренебрегаем данными») и Рид Хоффман, руководящий LinkedIn («Web 3.0 — это сеть с новым типом данных»), публично заявляли, что аналитическое мышление и принятие решений — это надежный способ обеспечить успех компании и свой личный успех. Любая компания в любой отрасли заинтересована в том, чтобы извлечь пользу из вала данных. Для этого требуются люди, умеющие провести их детальный анализ. У них разные имена, но всех их называют *квантами*<sup>[6]</sup>, и эта книга предназначена не им. А еще компаниям нужны люди, способные принимать оптимальные решения на основе анализа и воплощать их в жизнь. Именно для них написана эта книга. Это вовсе не те самые *кванты*, не аналитики, у них нет математической подготовки, но им приходится работать с количественными данными и принимать решения на основе их анализа.

## Какую пользу вы извлечете из этой книги

У нас, авторов, по этому вопросу разные мнения, но общая цель: расширить применение аналитического мышления в бизнесе и

обществе, в первую очередь помогая *не-квантам* извлекать больше пользы из массивов данных. Том вообще не очень-то серьезно относится к квантам, будучи социологом по образованию и не слишком хорошо разбираясь в статистике, но ему приходится заниматься аналитикой и определять направления ее применения в бизнесе. В течение двадцати лет Том вел исследования, писал книги, обучал студентов и консультировал компании по вопросам формирования аналитического потенциала. Его опыт работы с менеджерами и обучения их аналитическим исследованиям лег в основу этой книги. Кроме того, Том — автор и соавтор бестселлеров *Competing on Analytics* («Аналитика как конкурентное преимущество»)<sup>[7]</sup> и *Analytics at Work* («Аналитика в работе»), в которых рассказывается о том, какую роль аналитика играет в разработке стратегии больших компаний. В предлагаемой вниманию читателей книге основное внимание уделено тому, как помочь сотрудникам развить аналитические навыки и способности.

Джин Хо в университете учился именно деловому администрированию и статистике и, естественно, стал убежденным квантом. Он исследовал применение аналитических методов для решения разнообразных проблем в бизнесе и общественной деятельности. Кроме того, он разработал и вел образовательный курс по развитию аналитических способностей сотрудников. Джин Хо получил в Корее звание профессора делового администрирования и статистики, написал шесть книг, в том числе *100 Common Senses in Statistics* («100 здравых смыслов в статистике») и *Freak Statistics* («Причудливая статистика»). Главная их цель — помочь людям без статистической или математической подготовки лучше понимать и интерпретировать статистические данные и результаты.

Мы надеемся, что эта книга поможет вам разобраться в аналитике и уверенно использовать статистическую информацию. Она облегчит сотрудничество с квантами и научит на равных дискутировать с ними об аналитических процедурах и методиках. Вы заговорите языком количественного анализа и сможете задавать правильные вопросы. Может быть, она даже вдохновит вас на то, чтобы стать квантом!

## Информированный потребитель аналитических данных

У менеджеров, тесно сотрудничающих с аналитиками, множество вариантов того, как можно использовать количественно-аналитические данные для обоснования своих действий. Вот, например, какие решения недавно приняла Дженнифер Джой, вице-президент отделения клинических исследований и управляющая колл-центром компании Signa, ведущей в отрасли здравоохранения. Колл-центр активно работает с клиентами, стремясь улучшить их состояние, особенно если у них хроническое заболевание (например, диабет или болезни сердца), требующее постоянного лечения и контроля. Это способствует повышению качества медицинских услуг. Джен в свое время работала медицинской сестрой и не занималась аналитикой. Но она получила степень магистра делового администрирования и считает, что аналитика очень важна для компании. На ее примере легко убедиться в том, что два ключевых аспекта аналитического мышления — умение определять проблему и задавать правильные вопросы — помогают сберечь деньги для компании и ее клиентов.

Ключевой вопрос в работе Джой — сколько времени следует уделять консультированию клиентов Signa через колл-центр по поводу того, как предотвратить хроническое заболевание или замедлить его развитие. Она стремится доказать, что можно одновременно создавать дополнительную ценность для потребителей и контролировать издержки компании. Ключевая итоговая переменная для Джой — это частота первоначальной и повторной госпитализации пациентов. С одной стороны, она просматривает десятки страниц ежемесячных отчетов, где отражена динамика этого показателя. Процент повторной госпитализации то повышается, то понижается, и Джой не уверена, что понимает, почему так происходит: «Мне приходится изучать множество разных показателей, но это всего лишь цифры, они не отвечают на вопрос, что из этого следует». Но ей действительно хотелось бы выяснить, помогают ли пациентам звонки в колл-центр компании и влияют ли они на частоту повторных госпитализаций.

Чтобы лучше понять причинно-следственную связь между этими показателями, Джой обратилась к экспертам по аналитике в Signa. Группу аналитиков в компании возглавляет Майкл Казинс. Он и его

коллеги решили помочь Джой разобраться в возникшей проблеме. По словам Майкла, «хотя Джен и не специалист по методике анализа причинно-следственных связей, у нее блестящая логика, пытливым ум и она умеет задавать правильные вопросы. У нее нет математической подготовки, но она придает огромное значение аналитике». Джен относится именно к тому типу лиц, принимающих решения, с которым аналитики особенно любят работать.

Группа Казинса специализируется на применении аналитики для нужд бизнеса Cigna. Майкл и его коллеги целиком разделяют мнение Дженнифер о том, что показатель частоты госпитализаций не слишком полезен сам по себе, вне связи с вопросом о влиянии консультаций ее врачей на здоровье пациентов. Иными словами, предшествующие отчеты, показывавшие то повышение, то понижение частоты госпитализаций, были основаны на методике, которая не искала решения проблемы. В частности, не проводилось обоснованное сравнение с контрольной группой, результаты которого можно было бы использовать при принятии решения. В прошлом, например, отчеты не учитывали степень тяжести заболевания тех или иных пациентов. Группа Казинса разработала методику *парной группировки пациентов* в зависимости от тяжести заболевания, демографических параметров, образа жизни, региона проживания. Один из пары пациентов, у которых эти показатели были похожими, пользовался консультационными услугами колл-центра, а второй — нет. Казинс подчеркнул, что «Джой потребовалась определенная решимость, чтобы проверить, действительно ли эффективна консультационная служба, ее любимое детище. И все же она без колебаний приступила к выяснению истины».

Полученные результаты свидетельствовали: консультации врачей колл-центра по поводу некоторых болезней оказались вовсе не столь эффективны, как ожидалось; зато по другим заболеваниям картина была обратной. Джой решила сократить время телефонных консультаций для пациентов из целевой группы с определенными заболеваниями на тот период, пока не удастся выяснить, как сделать их более эффективными. Для контрольной группы предполагалось внедрить консультации, способные принести реальную дополнительную ценность.

В то же время Джой продолжала совместную работу с группой Казинса над другими аналитическими проектами. Один из них — контролируемый эксперимент с разными подходами к консультированию, например с включением в процесс консультаций персонального врача пациента. Вместо того чтобы полагаться на интуицию, решая, какой подход сработает лучше, Джой с энтузиазмом реализовывала идею структурированной аналитики, например пилотные программы типа «тестируй и учись», в соответствии с которыми в год выполнялось двадцать или тридцать тестов.

Дженнифер Джой всегда отличалась аналитическим складом ума, но благодаря сотрудничеству с аналитиками она получила инструменты, позволяющие проверять различные гипотезы и принимать обоснованные решения. И хотя так и не удалось отыскать идеальный способ лечить болезнь по телефону, но Сигна по крайней мере тратит меньше денег на неэффективные формы работы с пациентами и больше на те, которые доказали свою эффективность. С другой стороны, Майкл Казинс и его коллеги извлекли большую пользу из работы с Джой и другими аналитически мыслящими топ-менеджерами компании. Они научились излагать результаты исследований языком бизнеса и объяснять их смысл пациентам.

Вот почему эту книгу мы писали для множества Дженнифер Джой в разных уголках мира, а не для Майклов Казинсов. Мы не предлагаем вам стать квалифицированным количественным аналитиком или специалистом по базам данных. Для этого потребуется намного больше, чем просто прочесть книгу, да и не у каждого есть к этому интерес и способности. Но мы предлагаем стать компетентным пользователем данных и аналитики. Тогда вы сможете правильно ставить задачу аналитикам, использовать данные в своей работе, принимать на их основе решения и пропагандировать их применение в вашей организации. Мы хотим, чтобы о вас говорили как о человеке, глубоко увлеченном аналитикой, или как о Джен Джой из Сигна: «Сама она не аналитик, но понимает их и ценит их работу». Мы не ожидаем, что вы сами начнете проводить сложный анализ баз данных, но очертить поле поиска решения, поставить нужные вопросы о данных и методологии их анализа, интерпретировать полученные ответы и на их основе повысить эффективность действий компании вы сможете. Если

перефразировать заведующего кафедрой статистики Гарвардского университета Сяо Лименга, цель этой книги не в том, чтобы сделать из вас винодела (так он называет обладателей степени PhD<sup>[8]</sup> по статистике), а в том, чтобы привить вкус к хорошему вину<sup>[9]</sup>.

В прошлом в большинстве случаев было куда трудней стать компетентным пользователем информации, не ориентируясь в методах и приемах ее получения и обработки. Но сегодня в этой области произошли большие перемены. Теперь не надо до тонкостей разбираться в устройстве двигателя внутреннего сгорания, чтобы стать хорошим водителем; точно так же не обязательно вникать в детали статистического анализа, чтобы использовать статистические данные для принятия решений. Аналитическое программное обеспечение взяло на себя черновую работу, иногда даже может выбирать методику анализа, соответствующую характеру данных и переменных. Некоторые новые программы (например, от компании SAS) имеют справочную функцию, простым и понятным языком объясняющую смысл тех или иных зависимостей или характер методов, применяемых для прогнозных расчетов.

Хотя потребность в квалифицированных потребителях аналитики высока, на текущий момент нет книг, просто и без математического сленга написанных для новичков в области количественного анализа. В этой книге говорится о том, что такое аналитика, как можно ее использовать во многих жизненных ситуациях и как развить свои аналитические способности. Это поможет вам не только лучше разбираться в аналитике, но и значительно эффективнее обсуждать со специалистами различные аналитические методы и их применение для решения проблем компании. В соответствии с отчетом международной консалтинговой компании McKinsey Global Institute о больших данных за 2011 год, экономике требуются более полутора миллионов компетентных в аналитике менеджеров, чтобы эффективно использовать данные, накапливаемые обществом<sup>[10]</sup>. Надеемся, что вы станете одним из них.

## Роль аналитики в принятии решений

Решения в коммерческих и некоммерческих организациях принимают

исходя из целого ряда факторов: опыта, интуиции, результатов экспериментов, аналитических исследований и накопленных данных. В книге Moneyball, посвященной применению аналитических процедур в профессиональном бейсболе, говорится, что одно это отнюдь не гарантирует неизменно положительного результата. Команда Oakland Athletics выигрывала далеко не каждую игру описанного в книге сезона, да и всех последующих тоже. Тем не менее аналитика способна обеспечить некоторое конкурентное преимущество тем, кто в ней разбирается. Oakland Athletics добивается гораздо лучших результатов, чем можно было бы ожидать с учетом ее более чем скромного бюджета.

Конечно, ответственный менеджер вполне способен принять удачное решение, руководствуясь лишь интуицией и опытом, особенно когда решение лежит в сфере его непосредственной компетенции. Но почти в каждой области деятельности можно найти доказательства того, что решения, принятые на основе анализа данных, более точны и эффективны, обеспечивают больший выигрыш для организации<sup>[11]</sup>. В настоящее время в профессиональном бейсболе практически каждая команда применяет аналитические подходы, разработанные в Oakland Athletics. Даже команда New York Yankees, некогда чуть ли не гордившаяся отказом от аналитики в вопросах подбора игроков и определения стратегии игры, сейчас пригласила на работу 21 специалиста по спортивной статистике.

В коммерческих организациях традиционная аналитика чаще всего применяется для поддержки внутренних решений компании: «Сколько должен стоить этот продукт?» или «Как стимулировать покупателей совершать у нас покупки?» Аналитика в среде больших данных часто используется для разработки новых видов продуктов или дополнительных потребительских свойств. Например, Google создала PageRank — алгоритм ранжирования для поиска, социальная сеть LinkedIn — функцию «Люди, которых вы можете знать» или «С кем я могу связаться в сети», а компания Zynga — новые игры. Все эти продукты и свойства стали результатом управленческих решений или компаний, или их потребителей. (На рис. 1.1 «Типы управленческих решений, которые требуют аналитической поддержки» приведены еще

некоторые примеры.)

**Рис. 1.1. Типы управленческих решений, которые требуют аналитической поддержки**

### **Маркетинг**



- Ценообразование
- Размещение торговых точек и региональных подразделений
- Целевое продвижение продуктов
- Дизайн сайта компании
- Размещение рекламы в электронных СМИ

### **Поставщики**



- Объем складских запасов
- Размещение дистрибьюторских центров и складов
- Маршруты доставки продуктов или движения транспорта
- Загрузка транспорта

### **Финансы**



- Факторы финансовой деятельности
- Сбалансированная система показателей
- Различные виды прогнозов

### **Персонал**



- Каких сотрудников нанимать
- Кто из сотрудников собирается уволиться
- Какой должна быть сумма выплат и компенсаций
- Какое образование предпочтительно для сотрудников

### **Исследования и разработки**



- Какие потребительские свойства продукта больше всего привлекают потребителей
- Насколько эффективно производство и продажа продукта
- Какой дизайн продукта максимально привлекает потребителей

И это только то, что лежит на поверхности. В других отраслях и секторах экономики (государственном управлении, здравоохранении, спорте и других) можно найти массу подобных образцов.

Тот, кто принимает решения, оценивает возможные варианты решений с учетом информации как количественного, так и качественного характера. Источники качественной информации включают интуицию, опыт, здравый смысл, житейскую мудрость, слухи и предположения. Иногда они оказываются полезными, однако нужно помнить об их недостатках. Даже если у вас богатый опыт принятия решений в той или иной области, очередная ситуация может оказаться непохожей на предыдущие. Принимать решения на основе предположений всегда рискованно, а интуиция может вас обмануть. Большинство людей слишком полагаются на нее при принятии решений. Экономическая школа «Поведенческая экономика» исходит из постулата о том, что интуиция — далеко не лучший советчик при принятии экономических решений.

Несмотря на все достоинства аналитических исследований, бывают ситуации, когда нецелесообразно опираться на них при принятии решения. Если проблема слишком незначительна, носит единовременный характер или решение зависит от личных предпочтений, нет особого смысла тратить время на сбор и анализ данных, построение аналитической модели. Если действовать нужно быстро, то заниматься анализом просто некогда. Напротив, если решения приходится принимать периодически, есть время на проведение анализа, а проблема достаточно серьезна, чтобы оправдать и затраты, и труд, то аналитический подход будет весьма полезен.

## Три этапа аналитического подхода и порядок их выполнения

В центре внимания в этой книге — три основных этапа аналитического подхода, каждому из которых посвящено по главе.

Приводятся примеры аналитических задач, для которых особенно важен тот или иной этап. Схематически эти этапы вместе с входящими в них шагами показаны на рис. 1.2.

**Рис. 1.2. Три этапа и шесть шагов количественного анализа**

### **Формулирование проблемы**

1. Определение проблемы



2. Изучение предыдущих поисков решения



### **Решение проблемы**

3. Моделирование ситуации



4. Сбор данных

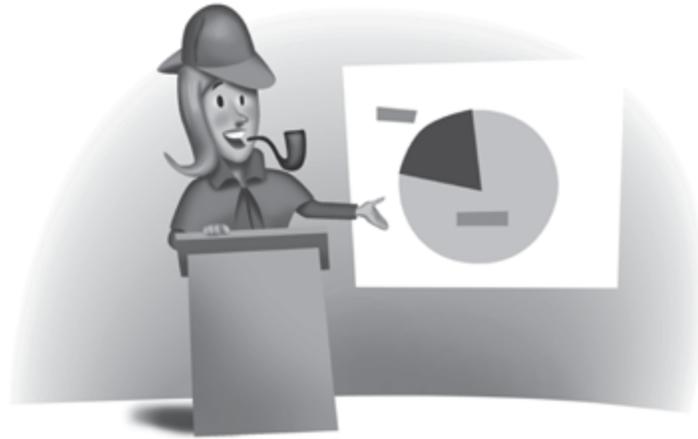


5. Анализ данных



## Результаты и необходимые меры

### 6. Демонстрация результатов и дальнейшие действия



В [главе 2](#) рассказывается о первом этапе — *формулировании проблемы*. Он включает в себя постановку вопросов, на которые предстоит ответить аналитику, и поиск формулировки для проблемы, которую нужно решить. Понятно, что это очень важный этап, ведь если сформулировать проблему некорректно, никакие данные или процедуры анализа не помогут ее преодолеть. Этот этап включает два шага: *определение проблемы* и *изучение предыдущих поисков решения*. В [главе 2](#) мы увидим, что после того, как проблема определена и сформулирована для решения аналитическими методами, как правило, оказывается, что некоторые ее аспекты уже рассматривались другими исследователями ранее, и это помогает уточнить формулировку.

Второй этап рассматривается в [главе 3](#) и охватывает наиболее интересующие читателей методы *решения проблемы*. На этом этапе вам предстоит отобрать переменные для будущей модели, собрать данные об их значениях за определенные периоды, а затем провести анализ. Допустим, вы вовсе не квант и не имеете особого желания им становиться. Тогда выход один: поручить большую часть этих работ количественному аналитику. Тем не менее очень полезно знать хотя бы в общих чертах, что и как тот собирается делать. Вам не обязательно решать всю проблему самостоятельно, но если вы будете уметь задавать правильные вопросы и понимать общие принципы анализа, то это, безусловно, поможет вам найти лучшее решение.

В [главе 4](#) рассматривается третий и последний этап количественного анализа, ничуть не менее важный, чем остальные, но часто игнорируемый, — *оформление результатов анализа* и принятие необходимых мер. От того, как вы расскажете о результатах анализа, в определяющей степени зависит, станут ли они толчком к дальнейшим действиям. Если тот, кто принимает решения (возможно, это вы и есть), не понимает, в чем именно заключается проведенный анализ и что означают полученные результаты, то вряд ли они станут основой для управленческих решений. Тогда не стоит и тратить время на прохождение первых двух этапов. Мы живем в информационно перегруженном мире, в котором трудно привлечь к чему-то внимание. Вот почему особенно важно подать результаты анализа в интересной, захватывающей форме. Сейчас уже нельзя представлять информацию в виде толстого отчета, переполненного сухими цифрами, и при этом надеяться, что для кого-то они станут стимулом к действию.

В остальных главах рассматриваются некоторые частные вопросы аналитического подхода. В частности, в [главе 5](#) мы говорим о роли креативности в аналитической работе: оказывается, эти понятия вполне совместимы! В [главе 6](#) описаны несколько способов развития аналитических способностей для тех, кто почувствовал вкус к аналитике. [Глава 7](#) посвящена налаживанию конструктивных взаимодействий менеджеров и квантов для принятия более эффективных решений. Нечего и говорить, что успех этих взаимодействий зависит от обеих сторон. Мы приводим много разнообразных примеров, иллюстрирующих возможности использования аналитических методов для решения проблем (или, напротив, то, как из-за пренебрежения ими возможности были упущены; см. вставку «Почему опасно пренебрегать аналитическими методами»). Кроме того, в этой главе показаны примеры расчетов и описаны самые простые возможности применения аналитики, для внедрения которых не требуется много времени и сил.

**Почему опасно пренебрегать аналитическими методами**

Мы учимся как на положительных, так и на отрицательных примерах, но почему-то отрицательные обычно производят более сильное впечатление. Трудно найти более яркий образец, чем история Джо Кассано, практически в одиночку доведшего до кризиса огромную компанию, а заодно и экономику США, да черт возьми, всего мира!

Кто же такой Кассано и чего он не знал об аналитике? Возможно, вы вспомните этого человека, если назвать его должность — глава AIG Financial Products (AIGFP), подразделения с четырьмя сотнями сотрудников, входящего в состав гиганта страховой отрасли AIG. Именно он несет ответственность за потерю компанией колоссальной суммы денег, о точной ее величине все еще идут споры, но что-то около 85 миллиардов долларов. Столько американским налогоплательщикам пришлось заплатить, чтобы удержать AIG на плаву и погасить ее долги.

Не один Кассано терял деньги, но, по словам расследовавшего эту историю репортера Мэтта Тайбби из журнала Rolling Stone, он был «виновником № 1 мирового экономического кризиса»<sup>[12]</sup>. Тайбби описывал его как «толстого лысеющего коротышку с глазами-бусинками и высоким лбом, окончившего Бруклинский колледж», хотя это не имеет особого отношения к нашей истории. Уверены, если бы он зарабатывал деньги вместо того, чтобы их терять, его внешность показалась бы репортеру более привлекательной.

Итак, к чему же привело пренебрежение аналитикой и количественным анализом? AIGFP потеряла эти деньги в результате активной продажи финансового продукта, называемого **кредитным дефолтным свопом (CDS)** и представляющего собой своего рода страховку ипотечных деривативов. Репортер New York Times Гретчен Моргенсон вскоре после этих событий писала: «Хотя крах рынка жилой недвижимости в США часто называют непосредственной причиной кризиса, надо отметить, что экономическая система в целом отличалась неустойчивостью из-за широкого распространения загадочных ценных бумаг — кредитных деривативов, предоставлявших кредиторам страховку на случай неплатежеспособности должников. Их выпускали частным образом, не привлекая внимания регуляторов фондового рынка, а иногда и не слишком включая мозги менеджеров, ответственных за их выпуск»<sup>[13]</sup>.

Кассано уж точно оказался одним из топ-менеджеров, не перегружавших мозги размышлениями о последствиях. И деривативы, и CDS принадлежат к числу сложных финансовых инструментов, разрабатываемых методами статистического и математического анализа, но, как позже выяснилось, на этот раз математика и статистика подвели. Если ипотечные заемщики оказывались неплатежеспособными, деривативы обесценивались и AIG сталкивалась с необходимостью выплачивать их держателям страховую стоимость ценных бумаг. Стоит ли говорить, что именно это и произошло?

Майкл Левис описал суть проблем подразделения AIGFP в журнале Vanity Fair<sup>[14]</sup>:

«Трейдерам AIGFP считают загадкой причины, приведшие их компанию к катастрофе. А начиналось все просто — с изменения процедуры принятия решений новым руководством. В конце 2001 года второй по счету генеральный директор компании Том Саваж ушел в отставку и на эту должность назначили его бывшего заместителя Джо Кассано. Саваж имел солидную математическую подготовку и хорошо разбирался в моделях оценки риска, применявшихся трейдерами AIG для гарантии получения достаточной компенсации. Кроме того, Саваж любил и часто

проводил внутрикорпоративные обсуждения достоинств тех или иных моделей, а также сделок. Новый генеральный директор намного хуже разбирался в математике и не имел особого желания вести дискуссии по этим вопросам.

Кассано не слишком беспокоился о том, насколько точна оценка риска и адекватны ли модели, на основе которых она получена. Он не задавал лишних вопросов аналитикам AIG и Уолл-стрит, разрабатывавшим эти модели. Он просто продавал CDS всем, кто желал их купить. Мысль о том, что их цена рассчитана на основе неверных допущений, в частности о том, что люди, бравшие эти недорогие ипотечные кредиты, окажутся неспособны их выплатить, кажется, никогда не приходила ему в голову. И даже если приходила, то не слишком его беспокоила».

Развить в себе аналитические способности путем самостоятельных занятий очень сложно. Мы создали специальный сайт (<http://keepingup-withthequants.weebly.com>), где читатели могут задавать любые вопросы об аналитике, аналитических процедурах или трудностях, с которыми приходится сталкиваться при решении конкретных проблем. Не стесняйтесь обращаться туда со всем, что показалось непонятным. На сайте подробно проанализированы некоторые ситуации, описанные в этой книге. Периодически мы будем ссылаться здесь на этот сайт.

## ГЛАВА 2

# Формулирование проблемы

Хотя в мире существует множество разных видов количественного анализа, все они имеют некоторые общие черты и порядок проведения. Как мы уже говорили в [главе 1](#), количественный анализ включает три основных этапа и шесть шагов.

### Формулирование проблемы

- Определение проблемы
- Изучение предыдущих поисков решения

### Решение проблемы

- Отбор переменных и разработка модели
- Сбор данных
- Анализ данных

### Результаты и необходимые меры

- Демонстрация итогов и дальнейшие действия

В этой главе и в главах [3](#) и [4](#) мы рассмотрим по отдельности каждый этап и шаг, а также приведем несколько примеров количественного анализа, которые включают в себя все шесть шагов и показывают их отличительные особенности в той или иной ситуации. В конце каждой главы для иллюстрации шести шагов даны два примера, обычно один — из сферы бизнеса, а второй — из сферы общественных или личных отношений. Наша трехэтапная и шестишаговая методика выполнения количественного анализа не единственная в своем роде (например, существует методика анализа отклонений качества продукции «Шесть сигм», направленная на обеспечение уровня брака не выше 3,4 единицы на миллион единиц произведенной продукции), но мы думаем, что большая часть экспертов-аналитиков одобряет ее. К тому же она достаточно гибкая, чтобы помочь выявить и

проанализировать множество разнообразных проблем в бизнесе.

## Шаг 1. Определение проблемы

Количественный анализ начинается с идентификации проблемы и подходов к ее решению. В анализе принятия решений этот этап называется *формулированием проблемы* и считается одним из наиболее важных для получения оптимального решения. Информацию для формулирования проблемы можно получить разными способами:

- обыкновенное любопытство (здравый смысл, наблюдение за событиями);
- опыт работы;
- потребность в решении либо действии;
- актуальные события, требующие внимания (сотрудника, организации в целом, нации);
- ранее проводившиеся исследования и уже существующие концепции;
- разработка проектов решений и анализ имеющегося и необходимого финансирования.

### 1. Формулирование проблемы



Заметим, что на этом этапе анализ как таковой пока отсутствует. Решение о проведении анализа часто принимают на основе интуитивных догадок. Особых доказательств их правильности не

требуется. Главная цель количественного анализа в том и состоит, чтобы путем исследования массива данных проверить правильность интуитивной догадки. Аналитики отличаются от других людей именно тем, что тестируют свои предположения, используя массивы данных и аналитические процедуры.

На этапе формулирования проблемы самое главное — глубоко осознать, в чем ее суть и что делает ее актуальной. Именно правильный ответ на эти два вопроса позволяет не только оценить, что нам может дать ее решение, но и спланировать действия на последующих этапах.

## Кто заинтересован в результатах?

Очевидно, что в наибольшей степени в результатах анализа заинтересованы менеджеры и те, кто принимает решение (владельцы бизнеса и менеджеры, ответственные за конкретные корпоративные вопросы). Тем не менее даже на этом этапе опытные количественные аналитики, имеющие представление о сути проблемы, процедурах принятия решений и возможных методах количественного анализа, могут оказать существенную помощь. Если не удастся найти одного человека, разбирающегося во всех этих вопросах, надо создать группу, включив в нее тех, кто обладает необходимыми компетенциями.

Стоит серьезно задуматься: кто составит основной круг заинтересованных в результатах предполагаемого анализа и что они думают по поводу проблемы, для решения которой анализ проводится?

### Оцените тех, кто заинтересован в результатах анализа

Если вы не можете ответить «да» на большинство приведенных вопросов, предполагаемый аналитический проект с самого начала под угрозой провала.

1. Знаете ли вы, кто именно из топ-менеджеров кровно заинтересован в результатах количественно-аналитического проекта?
2. Имеют ли они представление о проблеме и возможных вариантах ее решения?
3. Могут ли они предоставить необходимые ресурсы и внедрить преобразования, необходимые для успеха проекта?
4. Готовы ли они использовать аналитику и анализ данных в процессе принятия решений?

5. Соответствуют ли предлагаемая методика анализа и способы ее представления их традиционному образу мышления и методам принятия решений?
6. Предусмотрели ли вы меры для регулярного информирования и обратной связи с теми, кто заинтересован в результатах анализа?

Обладают ли они необходимыми полномочиями, чтобы предпринимать какие-либо действия на основе полученных результатов? Может быть, они не уверены в существовании проблемы, для решения которой разрабатывается аналитический проект? Можно ли их убедить в необходимости принятия мер, диктуемых результатами анализа?

Аналитики вообще склонны сразу переходить к выбору методики и способов анализа, не слишком задумываясь о тех, кто будет пользоваться результатами их труда. Чем больше они полагаются на свои аналитические навыки, тем меньше беспокоятся о распространении результатов и переходе к действиям.

Если вы убеждены, что вашему аналитическому проекту нужно внешнее руководство, то следует принять некоторые меры. Они, в частности, включают:

- 1) выявление всех заинтересованных в результатах анализа;
- 2) документирование их требований;
- 3) оценку и анализ интересов и сфер влияния тех, кто заинтересован в результатах анализа;
- 4) управление их ожиданиями;
- 5) принятие мер;
- 6) контроль результатов предпринятых действий и повторный анализ [\[15\]](#).

Анализ пользователей позволяет установить, кто является ключевыми лицами, принимающими решения, и достаточно ли убедительными окажутся для них результаты аналитического проекта. Даже наиболее корректный и точный аналитический подход окажется бесполезным, если его результаты не побудят к действиям тех, кто принимает решения. Иногда имеет смысл применить даже сомнительный с методологической точки зрения аналитический

подход, если он окажется единственным средством убедить пользователей в достоверности полученных результатов.

Вот пример. Рон Дубофф возглавляет компанию по маркетинговым исследованиям и стратегиям Hawk Partners. Вообще говоря, он убежден в том, что количественный анализ полезен и продуктивен практически в любой ситуации. Но он давно уже понял, что некоторые топ-менеджеры не понимают, какие он дает возможности для изучения покупательских потребностей и пожеланий, и гораздо больше доверяют качественным методам изучения спроса, например методике целевых групп. Суть ее такова: отбирают небольшую группу фактических или потенциальных покупателей, которым предлагают ответить на несколько вопросов о продуктах компании, при этом наблюдая за их реакцией и фиксируя ответы. Дубофф считает, что целевые группы — это методологически сомнительный подход. Специалистам по маркетинговым исследованиям хорошо известно, что покупатели склонны говорить то, что интервьюеры хотят от них услышать. Так что если они утверждают, что им нравится продукт, это еще не означает, что они будут его покупать. Опытный руководитель исследования целевой группы в некоторой степени может сгладить эту проблему, но результаты опроса нельзя распространять на более широкую аудиторию. Тем не менее Дубофф считает, что хоть какое-то изучение лучше, чем никакого, и если топ-менеджер доверяет результатам опроса целевых групп и не доверяет результатам количественного анализа, то пусть будет целевая группа.

Подобным же образом те, для кого предназначены результаты анализа, могут помочь определить, как и в какой форме их представить в отчете. У всех разные предпочтения о том, как лучше преподнести количественные данные: кто-то предпочитает табличную форму, кто-то графики и диаграммы, а еще кто-то — текст, описывающий количественные закономерности. Очень важно выявить эти предпочтения на относительно ранней стадии. Если данные предназначены для компьютерной обработки (а такое бывает все чаще и чаще по мере полной или частичной автоматизации процедур принятия решений), то не имеет смысла разрабатывать идеальный формат визуального представления. Просто скормите компьютеру цифры, и это окупится!

Может оказаться, что те или иные аналитические подходы лучше других помогают вовлечь пользователей в процесс анализа. Например, в компании Cisco Systems при выборе метода прогнозирования было установлено, что статистические методы позволяют получить намного более точные и надежные прогнозы (мы приведем описание всех шести шагов этого проекта в [главе 7](#)). Некоторые топ-менеджеры сразу поддержали эту идею, но кое-кто сомневался, что прогноз окажется более качественным. Руководитель проекта Энн Робинсон решила применить поэтапный подход: каждые несколько недель появлялись новые результаты, которые немедленно сообщали всем заинтересованным. Такое постоянное предъявление новой порции результатов помогло пользователям осознать преимущества проекта и втянуться в его выполнение. В итоге даже наиболее скептически настроенные менеджеры убедились в том, что прогнозы, полученные с помощью статистических методов, надежнее, делаются быстрее и охватывают более широкий круг продуктов, чем при применении прежних качественных методов.

## Сконцентрируйтесь на решении

Мы обнаружили, что фокусировать внимание на конкретных решениях, которые будут приниматься по итогам анализа, весьма полезно уже на этапе формулирования проблемы. Тому есть много причин. Во-первых, внимание к будущим решениям заставляет всех участников проекта помнить о том, что количественный анализ проводится не просто из любопытства, а с конкретной практической целью. Во-вторых, внимание к ключевым решениям помогает определиться с ключевым «потребителем» результатов анализа — человеком или группой людей, которые будут принимать решение на основе полученных результатов. В-третьих, если не удастся определить, какие решения будут приняты по итогам, то возникает вопрос: целесообразно ли проводить исследование?

Вот как описывает переговоры с клиентом на этапе формулирования проблемы Майк Томпсон, SEO фирмы First Analytics. Клиент, представитель сети ресторанов, считал, что первоочередным вопросом для анализа должна стать рентабельность продуктов. Топ-

менеджеры сети ресторанов хотели, чтобы First Analytics оценила, насколько рентабельно каждое блюдо в их меню. Майк разделяет мысль о том, что необходимо сосредоточиться на будущих решениях уже на этапе формулирования проблемы, поэтому он спросил, какие решения его собеседники собираются принять по итогам анализа рентабельности. Последовало долгое молчание. Один менеджер предположил, что ключевым должно стать решение о том, оставить ли блюдо в меню. Но другой отметил, что за последние двадцать лет в их ресторанах не было случаев исключения блюд из меню. После короткой дискуссии представители клиента пришли к выводу, что в фокусе анализа должна быть не рентабельность, а цена блюд. «Мы периодически меняем цены на протяжении всего времени существования», — заявил один из менеджеров.

## Какой проект количественного анализа вам нужен

Если вы определили, какие решения собираетесь принимать, можно переходить к следующему этапу анализа — изучению предыдущих попыток решить проблему. В [главе 4](#) мы поговорим о том, *какую историю могут рассказать данные*, именно в этом состоит лучший способ ознакомить с результатами анализа неспециалистов. Уже на первом этапе нужно начинать думать о том, какого рода эта история и как вы будете ее рассказывать, хоть многие ее детали и станут известны позже, уже в процессе анализа. Конечно, она связана с числами. Существует по меньшей мере шесть видов проектов количественного анализа. Рассмотрим их и приведем примеры.

**Проект CSI: полиция Майами.** Некоторые проекты количественного анализа напоминают детективные телесериалы, только в них вопросы бизнеса «расследуются» методами количественного анализа. Обнаруживается определенная проблема, и анализируются данные, для того чтобы удостовериться в правильности ее идентификации и найти пути решения. Часто в такой ситуации не нужен углубленный статистический анализ — достаточно корректно подготовленных и представленных данных. В интернет-магазинах, например, каждый щелчок покупателя мышью несет в себе огромное, иногда даже слишком, количество информации для анализа.

Одним из главных действующих лиц в нашей «следственной истории» станет Джо Меджибов — вице-президент и SEO онлайн-туристического агентства Expedia из США. Когда-то Джо был специалистом по веб-аналитике. Он и сейчас им остается, и его подходы к решению бизнес-проблем на основе количественного анализа данных привели к нескольким блестящим решениям.

Во многих онлайн-исследованиях Expedia возникал вопрос о том, почему транзакции по некоторым заказам не завершаются. В частности, руководство отеля хотело выяснить, почему уже в процессе оформления заказа не удается довести дело до перечисления средств на счет компании. Анализ данных показал, что после выбора отеля и заполнения формы с данными о сроках путешествия и стоимости номера, а затем щелчка по кнопке «Оплатить сейчас» часть транзакций оказываются незавершенными. Группа аналитиков под руководством Меджибова решила разобраться в причинах неудач, используя систему показателей интернета и лог-файлы операций, хранящиеся на сервере.

Очевидно, проблемы возникали из-за поля «Компания», размещенного под полем фамилии клиента. Некоторые клиенты считали, что в нем надо указать название банка, эмитировавшего их кредитную карту, а в поле «Адрес, на который будет направлен счет» указывали адрес этого банка. В результате в процессе списания средств с карты клиента операция отменялась как некорректная. После того как поле «Компания» просто убрали из формы, прибыль Expedia возросла на 12 миллионов долларов. Меджибов говорит, что специалисты Expedia расследовали несколько таких историй, и это всегда либо приносило фирме дополнительную прибыль, либо эффективность операционной деятельности повышалась.

Иногда расследование требует более глубокого статистического и количественного анализа. Один из сотрудников Меджибова изучал вопрос о том, какие точки взаимодействия с покупателями важнее всего для стимулирования продаж. Аналитик использовал *регрессионную модель Кокса* — метод, обычно используемый для прогнозирования вероятности дожития пациентов до определенного момента в будущем; «анализ выживаемости». Оказалось, что более простые модели, применявшиеся ранее, давали искаженную информацию о том, какие маркетинговые подходы наиболее

эффективны. Меджибов прокомментировал это так: «Мы и не знали, сколько денег утекает сквозь наши пальцы»<sup>[16]</sup>.

**Проект «Эврика!».** Проекты этого типа напоминают проекты CSI, но только целенаправленно подходят к выявлению и решению проблемы в отличие от ситуаций, когда трудности возникают неожиданно. Обычно рассматриваются проблемы, порождаемые глобальными переменами в стратегии компании или ее бизнес-модели. Как правило, такие аналитические проекты требуют больше времени для реализации и предполагают большой объем статистического и математического анализа. Иногда проекты типа «Эврика!» включают в себя элементы проектов других видов, поскольку ожидаемые результаты очень важны для заказавших анализ организаций.

Вернемся к анализу в компании Expedia. Встретился там и проект типа «Эврика!», когда требовалось отменить штрафы за перенос сроков и отмену бронирования номеров в отелях, туров и автомашин. До 2009 года Expedia и ее конкуренты взимали до 30 долларов за перенос сроков или отмену брони — в дополнение к соответствующим санкциям со стороны отелей. При заказе отеля через Expedia или другие онлайн-агентства номер обходился клиентам значительно дешевле, чем при заказе непосредственно в отеле, поэтому они мирились со штрафами за перенос или отмену заказа. Но к 2009 году стало ясно, что это превратилось в проблему. Стоимость номера при заказе через Expedia существенно приблизилась к расценкам самих отелей, поэтому Expedia сделала упор на удобство обслуживания, а штрафы за перенос и отмену бронирования стали неудобны. Аналитики изучили коэффициенты удовлетворенности клиентов, и оказалось, что у тех, кому пришлось платить этот штраф, уровень удовлетворенности существенно ниже. Сотрудники колл-центра Expedia имели право отменить штраф только по одной причине — в случае смерти кого-либо из членов семьи клиента. Темпы роста численности освобожденных от штрафа по этой причине за последние три года составляли двузначную цифру. То ли в это время свирепствовала эпидемия с высоким процентом смертельных исходов, то ли клиенты поняли, что это единственный способ вернуть свои деньги.

Топ-менеджеры агентства поняли, что на рынке туристических услуг произошли существенные изменения, но штрафы за отмену и перенос сроков бронирования составляли значительную часть общей выручки. Возник вопрос, как поведет себя коэффициент перехода контактов в продажи (то есть процент оплативших номер по отношению к числу приславших заявку), если отменить эти штрафы. В апреле 2009 года Expedia объявила о временном (сроком на месяц) моратории на штрафы за перенос сроков и отмену бронирования (в чем-то похоже на историю об эксперименте сумасшедшего ученого, описанную ниже). Коэффициент перехода контактов в продажи существенно вырос. Топ-менеджеры поняли, что получено достаточно доказательств того, что отмена штрафов целесообразна, и очень скоро примеру Expedia последовали другие компании отрасли.

В центре Сиэтла находится штаб-квартира компании Zillow, предоставляющей информацию о жилой недвижимости в регионе. Вероятно, эта компания известна в среде квантов прежде всего благодаря разработанному ее сотрудниками алгоритму Zestimates, позволяющему рассчитать стоимость объектов недвижимости. Но, как и в Expedia, корпоративная культура Zillow построена на культе данных и аналитики, что и неудивительно, поскольку основателем обеих этих компаний является Рик Бартон.

Один из проектов типа «Эврика!» посвящен глобальной проблеме: реорганизации отношений с агентами по недвижимости. Zillow начала работать с агентами по недвижимости в 2008 году, а до этого взаимодействовала непосредственно с покупателями. Особенность бизнес-модели, построенной на работе с агентами, в том, что компания рекламирует своих агентов и направляет к ним потенциальных покупателей. За каждого покупателя с агента взимаются комиссионные, но, с точки зрения топ-менеджеров, их размер недостаточен. Директор Zillow по продуктам и стратегии Хлоя Харфорд особенно заинтересована в разработке адекватной модели оптимизации комиссионных за направленных к агентам покупателей.

Харфорд, получившая ученую степень по вулканологии, уже проводила довольно сложные математические анализы раньше. Тем не менее она и ее коллеги первоначально полагались на методы, которые называли «расчетами на салфетке», чтобы оценить другие пути,

позволяющие привлечь больше потенциальных покупателей и установить справедливые комиссионные с агентов. В апреле 2010 года Zillow внедрила новую модель взаимоотношений с агентами, немедленно скопированную конкурентами и включавшую помимо прочего продажу рекламных услуг агентам. В результате поток контактов с потребителями резко возрос, причем они были переключены непосредственно на агентов. Zillow также внедрила интеллектуальный алгоритм расчета комиссионных за потенциальных покупателей с учетом их экономической стоимости и коэффициента перехода контактов в продажи. Конкуренты в той или иной степени старались повторить эти новшества, но не в таком объеме, как Zillow. Контакты потенциальных покупателей и определение комиссионных за их направление к агентам настолько важны для Харфорд и ее коллег, что постоянно тестируются различные подходы к их оценке, в том числе и с использованием методов, описанных в истории о безумном ученом. Коротко говоря, проекты «Эврика!» тесно связаны с моделью бизнеса компании и ее коммерческим успехом.

**Проект «Сумасшедший ученый».** Мы знаем, как широко распространены научные эксперименты в высокотехнологичных отраслях, например фармацевтической. Производящие лекарства компании тестируют продукты на целевых и контрольных группах, давая членам последних плацебо (лекарства-«пустышки», вещества без лечебных свойств). Они уделяют огромное внимание соблюдению случайного метода распределения участников между целевой и контрольной группами, чтобы их состав был однородным и не влиял на оценку эффективности лекарства. Этот действенный аналитический прием делает возможным причинно-следственный анализ и распространение выводов, сделанных на основе данных, полученных в целевой группе, на генеральную совокупность.

Строгий эксперимент больше не является прерогативой одних только ученых; сейчас он стал аналитическим приемом, необходимым каждой крупной компании. Сейчас широко распространено программное обеспечение, помогающее менеджерам и аналитикам проводить анализ. Компании получили возможность принимать решения на основе строго научных экспериментов. В прошлом любое

вторжение в область рандомизированного тестирования (случайного распределения участников целевой и контрольной групп, о котором мы только что говорили) требовало приглашения дипломированного специалиста по статистике или по разработке научных экспериментов. Теперь магистр делового администрирования, прошедший курс статистического анализа, вполне может организовать процесс с помощью нужного программного обеспечения, помогающего определить численность целевой и контрольной групп, сайты для тестирования и контроля, а также сделать оценку статистической значимости любых отклонений, выявленных в ходе эксперимента.

«Проекты сумасшедших ученых» особенно удачно подходят для розничных сетей с многочисленными супермаркетами, банков с множеством отделений и других подобных компаний. Это упрощает использование части торговых точек или отделений в качестве целевых, а остальных — в качестве контрольных. Стало легко проводить эксперименты на сайтах, где часть посетителей можно направить на одну версию интернет-страницы, а вторую часть — на другую ее версию, а потом проверить, окажутся ли результаты существенно отличными (это называется *A/B тестированием* в сфере исследования сайтов).

Некоторые примеры «проектов сумасшедших ученых» приведены ниже<sup>[17]</sup>.

- Способствует ли установка аквариумов с живыми омарами их продажам в супермаркетах Food Lion? Видимо, ответ будет утвердительным, если покупатели этого супермаркета уже привыкли покупать здесь омаров (то есть принадлежат к группе лиц со сравнительно высокими доходами), и отрицательным, если обеспеченные покупатели не заходят сюда.
- Увеличится ли общая выручка супермаркета Kmart, если часть его торговых площадей отвести под магазины супермаркета Sears? Председатель совета директоров компании Sears Holdings Эдди Ламперт является большим поклонником рандомизированного тестирования. Он протестировал различные комбинации использования торговых площадей. На этот конкретный вопрос у нас ответа нет, но можно предположить,

что если бы он был положительным, то таких комбинированных супермаркетов было бы гораздо больше.

- Какие из сети ресторанов морепродуктов Red Lobster (с высоким, средним или низким уровнем цен) обеспечивают максимальный объем продаж и что важнее для привлечения клиентов: внешний вид ресторана или его внутренняя отделка? Топ-менеджеры Red Lobster утверждают, что наибольшие продажи обеспечивают рестораны средней ценовой категории. Внешний вид ресторана играет очень большую роль в привлечении новых клиентов, но если они увидят, что его внутренняя отделка не соответствует внешнему виду, то второй раз в этот ресторан не придут.

**Проект «Опрос».** Опросы — это классический метод количественного анализа. Аналитики, проводящие их, имеют дело с уже произошедшими или происходящими в данный момент событиями. Аналитик не пытается повлиять на результаты, он только наблюдает, классифицирует и анализирует их. В типичном случае интервьюер стремится выявить статистически значимую зависимость между рядом исходных и рядом изучаемых факторов или переменных. Самый простой пример — опрос в выборке покупателей конкретного продукта об их личных характеристиках, в том числе демографических (возраст и пол). Задавая вопросы о том, какие продукты они предпочитают, можно выяснить, пользуется ли конкретный продукт спросом в большей степени у мужчин, чем у женщин, будут ли определенные продукты пользоваться спросом преимущественно у молодых покупателей.

Опросы весьма популярны и несложны с точки зрения организации и проведения. Однако следует помнить, что полученные результаты могут существенно отличаться в зависимости от постановки вопросов и изменения их формулировок с течением времени. Например, Бюро переписи населения США десятилетиями работает над формулировкой вопроса о национальной принадлежности граждан. Количество вариантов ответа на этот вопрос постоянно увеличивается; в 2010 году предлагалось выбрать из пятнадцати вариантов, в том числе «другая национальность». Этот ответ пользовался большой популярностью

среди граждан США — латиноамериканцев: 18 миллионов из более чем 50 миллионов отметили именно его<sup>[18]</sup>. Если уж вопрос о национальности вызывает столько сомнений, то что говорить о таких скользких темах, как политика, религия, социальный статус или сексуальные привычки!

Кроме того, нам следует помнить, что сама по себе связь между двумя переменными еще не говорит о ее причинно-следственном характере. Мы поговорим об этом подробнее в [главе 6](#), а пока просто отметим, что вполне могут существовать и другие переменные, оставшиеся за рамками анализа; именно они, возможно, обуславливают выявленную зависимость. Опросы предполагают выяснение убеждений или отношений людей к тем или иным событиям, но их вопросы не должны задевать эмоции интервьюируемых. Рассмотрим в качестве классического примера (приводимого во многих учебниках по статистике) устроенный во время Второй мировой войны опрос об организации ремонта самолетов.

Во время Второй мировой войны нужно было обеспечить максимально длительное функционирование самолетов, поэтому было решено выяснить, можно ли сократить количество трудоемких ремонтов моторов без роста аварийности. Был проведен ретроспективный опрос о самолетах, потерпевших катастрофу, и вопреки всем ожиданиям оказалось, что количество аварий из-за проблем с моторами достигало максимума непосредственно после ремонтов, а затем постепенно снижалось. В результате было принято решение существенно удлинить интервалы между ремонтами и, естественно, пересмотреть их характер, чтобы убедиться в том, что все гайки и болты затянуты как надо<sup>[19]</sup>.

Если вы планируете провести опрос или проанализировать его результаты, сначала убедитесь, что смысл включенных в анкету вопросов, равно как и отобранных для анализа переменных, тщательно продуман. Переменной называется любая количественно измеримая характеристика параметров людей, ситуаций или поведения с двумя или больше уровнями или вариантами значений. Пол, балл на экзамене, температура в помещении, любовь, счастье, сплоченность команды — все это примеры переменных.

Кроме того, важно убедиться, что выборка для опроса репрезентативна для той группы населения, которую вы собираетесь

тестировать. Способы проведения опроса могут повлиять на результаты. Например, если хотите изучить поведение молодежи или ее отношение к чему-то, не стоит нанимать маркетинговую компанию, проводящую опросы исключительно по стационарным телефонам. Да, это стандартный способ их проведения, но мы ведь знаем, что у многих молодых людей просто нет стационарных телефонов, да они и не собираются ими обзаводиться. В результате выборка для молодежи будет нерепрезентативной<sup>[20]</sup>.

**Проект «Предсказание».** Все проекты этого вида имеют целью прогнозирование того, что должно произойти в будущем. Получить надежную информацию о грядущих событиях довольно трудно, но если речь идет о прошлых событиях и их причинах, то для количественного аналитика это несложно. Обычно проекты такого рода относятся к предсказательной аналитике или предсказательному моделированию.

Проекты вида «Предсказание» весьма разнообразны. Приведем некоторые ситуации, в которых они целесообразны.

*Реакция на коммерческое предложение.* Кто из покупателей мог бы отреагировать на разосланное по электронной почте коммерческое предложение с бесплатной доставкой продукта в течение двух рабочих дней при сумме заказа 50 долларов или больше?

*Кросс-продажи и продажи более дорогих версий продукта.* Кто из клиентов, имеющих чековый счет с остатком более 2000 долларов, мог бы купить одногодичный депозитный сертификат под 1,5 процента в год в течение одного месяца с момента рассылки коммерческого предложения?

*Убыль персонала.* Кто из сотрудников, проработавших более шести месяцев и еще не подписавшихся на программу страхования 401(k), уволится в течение следующих трех месяцев?

Существует много других вариантов того, как можно применить результаты предсказательного анализа. В бизнесе чаще всего нужно определить, какое именно коммерческое предложение, скорее всего, примет потребитель. Более сложные варианты анализа «следующего по привлекательности предложения» все чаще проводятся с помощью прикладного программного обеспечения. Во-первых, содержание

коммерческих предложений нужно хранить в тайне, пока не наступит время довести его до потребителя; во-вторых, таких предложений могут быть сотни и даже тысячи.

Например, Microsoft, как мало кто другой, постоянно разрабатывает все новые усовершенствования для своей поисковой системы Bing (поисковик бесплатный, поэтому компания старается таким образом просто привлечь как можно большее количество пользователей). Эти новшества побуждают вас испробовать предлагаемый поисковик, ввести панель инструментов Bing в браузер, испытать дополнительные свойства Bing и т. п. Модификация поисковика в соответствии с предпочтениями пользователя производится по разным параметрам: возраст, место жительства, пол, часто посещаемые сайты и т. п., определяемым на основе *cookies*<sup>[21]</sup> пользователя и других источников. Если вы когда-нибудь получали идентификатор Microsoft Passport, то компания располагает еще большим объемом информации о вас, и это позволяет формировать «коммерческое предложение» целенаправленно. С помощью программы Infor Epiphany Interaction Advisor компания Microsoft имеет возможность мгновенно формировать и отсылать адресное электронное сообщение в тот момент, когда вы щелкаете мышкой по заголовку сообщения в почте: это занимает около 200 мс. По словам представителей компании, это средство отлично работает на повышение коэффициента конверсии веб-узла.

Зачастую проекты вида «Предсказание» весьма напоминают ловлю рыбы сетью. Мы точно не знаем, какие именно факторы позволят сделать обоснованный прогноз, поэтому тестируем все и отбираем те, которые срабатывают. Иногда это дает неожиданные результаты. Например, в ситуации с усовершенствованием поисковика Bing, о котором мы только что рассказали, оказалось, что в зависимости от количества контактов пользователя в Microsoft Messenger можно с уверенностью предсказать, захочет ли он использовать Bing.

Google поставила цель выяснить, какими особенностями отличаются наиболее эффективные сотрудники компании. Анализ показал, что те критерии, на которые компания обращала внимание изначально, — диплом университета и рейтинг по итогам

собеседования — практически бесполезны при прогнозировании будущей эффективности работы. Поскольку неизвестно было, какие критерии окажутся более надежными, сотрудникам предложили заполнить анкету из трехсот вопросов. Как отметил начальник отдела персонала Google Ласло Бок, «мы решили забросить очень широкую сеть. В нашей компании нет ничего необычного в том, чтобы выйти в коридор и наткнуться на человека с собакой. Может быть, у владельцев собак есть общие черты характера, говорящие об их творческих качествах?»<sup>[22]</sup>

Вряд ли привлечение собак к прогнозированию даст какой-либо результат, но Google все же удалось найти некоторые неожиданные критерии. Например, если претендент на рабочее место ставил мировые или национальные рекорды в любой области, учреждал некоммерческую организацию или клуб, то чаще всего он оказывался высокоэффективным работником. Сейчас Google включил вопросы об этом в свои онлайн-анкеты для претендентов на вакансии.

Конечно, если обнаруживаются факторы, которые демонстрируют связь с анализируемыми параметрами, но при этом сами по себе бессмысленны, стоит вернуться к самому началу и проверить доброкачественность исходных данных и корректность методики анализа. Однако в большинстве случаев анализ массивов данных более эффективен, чем прогнозы на основе качественного анализа. Только учтите, что предсказательные проекты основываются на информации за прошлые периоды для прогнозирования событий в будущем. Если со времени проведения последнего анализа в мире что-то изменилось, то полученный прогноз может оказаться недостоверным.

**Проект «Что случилось, когда...?»** Проекты, описывающие ряд событий и показателей на основе собранных данных, распространены наиболее широко. Они представляют структурированные данные: сколько единиц продукта было продано, за какой период и где, какие финансовые результаты продемонстрировала компания в прошлом квартале, сколько человек взяли на работу в прошлом году. Поскольку эти проекты ориентированы в основном на отчетные данные, сложные математические методы в них, как правило, не используются, и может показаться, что они очень просты в исполнении. Но колоссальный рост

данных, генерируемых современными организациями, привел к тому, что аналогично возрос и объем отчетности на их основе. Вот почему иногда бывает сложно привлечь внимание целевой аудитории к создаваемым и распространяемым отчетам.

Этот вид проектов особенно полезен, когда требуется наглядно представить данные. Достаточно сказать, что если в ваших отчетах в основном таблицы, заполненные цифрами, вряд ли вы привлечете внимание аудитории. Кое-кто скажет, что устал от обилия графиков и цветных диаграмм, но большинство наверняка считают, что они более понятны, чем цифры на бумаге. Поскольку в [главе 4](#) мы подробно рассматриваем, какими способами можно проинформировать о результатах анализа, там же поговорим и о том, как сделать цифровые отчеты более интересными и привлекающими внимание.

## Масштаб проблемы

По определению, у аналитического проекта, основанного на количественном анализе, узок круг рассматриваемых вопросов, поскольку требуется сбор данных и проверка на их основе некой гипотезы (см. вставку «Примеры проверяемой гипотезы»). Трудно собирать данные по широкому кругу. Но на этом этапе важно не сузить преждевременно масштабы рассматриваемой проблемы или будущего решения. Взгляд на проблему должен быть достаточно широким для того, чтобы выделить несколько возможных вариантов решения. Например, если организация считает, что в отдельном подразделении или регионе возникла проблема с эффективностью деятельности, то причин этого может быть много — начиная от недовольства потребителей до производственных факторов или особенностей продуктов и услуг.

### Примеры проверяемой гипотезы

- Анализ видов продуктов, пользовавшихся спросом в прошлом году, лучше всего поможет понять, на какие коммерческие предложения, рассылаемые по электронной почте, мы получим максимальный отклик в будущем.
- Продолжительность учебы является надежным критерием будущей эффективности деятельности сотрудников на должностях, требующих высокого уровня

компетентности.

- Десятипроцентные скидки с цены продуктов за неделю до праздников менее эффективны, чем скидки, введенные в другое время.
- Размещение товара на витрине в конце прохода наиболее эффективно с точки зрения стимулирования продаж.
- наших покупателей можно разделить на четыре группы в зависимости от того, какие продукты они предпочитают.
- Возможность повышать цены на продукты стандартного качества и при этом не вызвать сокращения спроса существенно уменьшается в годы экономической рецессии.
- Наши подразделения, централизовавшие склады, предпочитают поддерживать меньший операционный запас материалов для производственного процесса.

В примере с компанией Transitions Optical, приведенном в конце этой главы, к идентификации и формулированию проблемы менеджеров подтолкнуло смутное ощущение, что маркетинговые расходы избыточны. Однако из области решений было выбрано только одно, включавшее комплексную оптимизацию маркетинговых расходов и привлечения средств массовой информации.

Мы называем первый этап количественного анализа формулированием проблемы, но его же можно назвать и выявлением возможностей. Британский инженер Джозеф Джаггер (1830–1892) понял, что есть возможность сорвать банк в казино Монте-Карло<sup>[23]</sup>. Он приобрел опыт работы с машинами и механизмами на хлопкопрядильных фабриках Йоркшира. Это помогло ему понять принципы движения колеса рулетки и предположить, что под действием механической неисправности одни цифры выпадают чаще, чем другие. А что если использовать эту неисправность в своих интересах? Он поехал в Монако, чтобы проверить свое предположение.

На колесе рулетки французского/европейского образца есть цифры от 1 до 36, а также 0. Если колесо делает один оборот, то теоретическая вероятность выпадения каждого номера равна  $1/37$ . Таким образом, удельный вес числа выпадений каждого номера при большом количестве оборотов колеса также равен  $1/37$ . Джаггер предположил, что разбалансированность колеса приведет к тому, что вероятность

выпадения определенных чисел превысит  $1/37$ .

Эти рассуждения побудили Джаггера нанять шестерых помощников для наблюдения за шестью столами с рулеткой в легендарном казино Beaux-Arts в Монте-Карло. Каждый помощник получил инструкции относительно записи результатов всех розыгрышей на своем столе. Проанализировав результаты, Джаггер понял, что на пяти столах числа, как и ожидалось, выпадали случайным образом. Но на шестом столе девять чисел (7, 8, 9, 17, 18, 19, 22, 28 и 29) выпадали чаще, чем остальные. Джаггер сделал вывод, что у колеса на этом столе имеется какой-то дефект, нарушающий его балансировку. Свои первые ставки на этом столе он сделал 7 июля 1875 года и быстро выиграл приличную сумму (14 тысяч фунтов стерлингов — в 2012 году она равнялась бы примерно 840 тысячам фунтов стерлингов, или более 1,3 миллиона долларов с учетом инфляции). Казино разобралось, в чем состоит выигрышная стратегия Джаггера, и в конце концов нейтрализовало ее. Но к этому моменту он уже успел выиграть сумму, равную более чем шести миллионам современных долларов. Это и есть возможности аналитики!

## Решите, что конкретно вы хотите выяснить

Хотя на этапе формулирования проблемы полезно мыслить широко, но к его окончанию надо иметь четкое понимание ее сути, конкретную формулировку и перечень показателей, которые подлежат анализу. Это необходимо, поскольку в количественном анализе от формулировки проблемы зависит очень многое. Например, представьте себе, что вы топ-менеджер телекомпании, желающий выяснить, какие каналы предпочитают ваши зрители. Два аналитика-консультанта предложили вам разные подходы к проведению исследования. Из чистого любопытства вы решили нанять обоих и посмотреть, насколько будут отличаться их результаты.

Один из консультантов предложил зрителям каждый день в течение недели записывать (в онлайн-овом режиме либо на бумаге) те каналы и программы, которые они смотрят. Второй провел анкетирование зрителей, попросив их расположить телеканалы, которые они смотрели в течение нескольких последних месяцев, в порядке

предпочтения. Оба отобрали репрезентативную выборку зрителей, отражавшую особенность всей аудитории.

Хотя оба консультанта пытаются получить ответ на один и тот же вопрос, они, скорее всего, получат весьма различающиеся результаты. Тот, кто предложил участникам целевой группы записывать просмотренные телепередачи, вероятно, получит результаты более точные, но зато процент откликнувшихся на это предложение будет ниже, поскольку процедура опроса связана с дополнительными затратами времени. (Компания Nielsen Media Research ведет постоянный мониторинг телеканалов и программ, и среди участников целевой группы каждый раз около 50 процентов отказываются вести записи, а ведь эта процедура выполняется автоматически.) Кроме того, этот консультант столкнется с проблемой влияния на зрительские предпочтения времени года или программ телеканалов в ту неделю, когда предполагается вести записи.

Точность исследования, проведенного вторым консультантом, возможно, будет ниже, но зато и сезонные факторы окажут меньше влияния за счет охвата более продолжительного периода. В результате расхождения в результатах будут настолько значительными, что сгладить их не удастся. Поэтому очень важно на этапе идентификации проблемы четко определить, что именно вы собираетесь изучать.

## Шаг 2. Изучение предыдущих поисков решения

После того как проблема определена, следует выяснить, проводились ли ее исследования ранее и каковы были их результаты. Это второй шаг первого этапа количественного анализа (формулирование проблемы), поскольку информация о предшествующих исследованиях помогает аналитику и менеджеру оценить разные варианты формулировки проблемы и ее концептуализации.

Довольно часто аналитики находят в отчетах о ранее проводившихся исследованиях нечто такое, что заставляет их пересмотреть собственный вариант определения проблемы. Не исключено, что после этого придется пересмотреть отчеты о более ранних исследованиях.

Обычно на этом этапе аналитик задает себе вопрос: «Похож ли мой проект на те, которые выполнялись раньше?» Если да, то из отчетов

можно почерпнуть идеи для собственного анализа. Обзор предшествующих исследований помогает уточнить ряд вопросов:

- В чем особенности нашего исследовательского проекта? Включает ли он опрос, предсказание, эксперимент, отчет?
- Какие данные нам необходимо собрать?
- Какие параметры изучали в предшествующих исследованиях?
- Какие виды анализа нам придется провести?
- Будут ли результаты нашего анализа отличаться от полученных ранее и как представить их в интересной форме?

Одна из ключевых особенностей количественного анализа (и вообще научного метода исследований) — это учет результатов более ранних исследований. Например, поиск относящейся к теме информации в книгах, отчетах и статьях очень важен для всестороннего понимания проблемы. Это помогает установить ключевые параметры и связи между ними.

Комплексный обзор результатов любых предыдущих исследований той же тематики обязателен для любого вида количественного анализа. В аналитике невозможно получить нечто из ничего. Приступить к решению проблемы можно только ознакомившись с опытом тех, кто делал это до вас. Помните простую вещь: любая проблема не настолько уникальна, как вам кажется, и не исключено, что многие уже сделали то, что вы только собираетесь делать. Не стоит заново изобретать колесо, достаточно выяснить, кто сделал это до вас. В наше время при помощи поисковика вроде Google нетрудно получить большую часть материалов по теме. Одна только систематизация и оценка имеющейся информации играет важную роль в уточнении модели анализа или подходов к решению проблемы.

## **2. Изучение предыдущих поисков решения**



Пример успешного обзора результатов более ранних исследований можно найти в истории Второй мировой войны. Адольф Гитлер приказал начать производство новой мощной управляемой ракеты ФАУ-2, и с ее помощью в 1944 году люфтваффе терроризировало жителей Лондона. В течение нескольких следующих месяцев 1358 из 3172 запущенных по целям в Англии ФАУ-2 упали на Лондон, в результате чего погибли 7250 человек — военных и гражданского населения.

Во время обстрела Лондона многие наблюдатели отмечали, что большая часть ракет падала в определенных районах города. Британцы хотели выяснить, является ли падение ракет случайным или управляемым. Если поражение целей возможно только случайно, то размещения бомбоубежищ и укрытий в окрестностях города будет вполне достаточно для обеспечения безопасности населения Лондона. Если же ракеты управляемые, то опасность намного больше и равномерно распределенные убежища особой пользы не принесут. Британское правительство пригласило известного статистика Р. Д. Кларка для решения этой задачи. Кларк провел простой статистический тест, основанный на информации о предыдущих попаданиях ракет. В частности, он понимал, что для этого анализа

можно применить так называемое *распределение Пуассона*, которое показывает вероятность совершения ряда событий в фиксированный период времени, в определенном регионе или объеме, если эти события происходят с известной частотой. Если ракеты падают случайным образом, то количество ракет, попадающих в каждый квадрат местности, будет соответствовать распределению Пуассона. Например, если на каждый квадрат приходится в среднем по одной ракете, то с помощью формулы Пуассона можно подсчитать вероятность попадания в один квадрат одной, двух, трех, четырех и более ракет или непопадания ракет в один квадрат вообще.

Чтобы рассчитать количество ракет, которые могут попасть в определенный квадрат местности, Кларк разделил Южный Лондон на 576 квадратов площадью в четверть квадратного километра каждый и подсчитал количество квадратов, в которые попадали 0, 1, 2, 3, 4 и более ракет.

Если обстрел производился наугад, тогда количество квадратов с определенным количеством попаданий ракет подчинялось бы распределению Пуассона. Совпадение реального количества попаданий с рассчитанными по распределению Пуассона оказалось очень тесным, что опровергло предположение об управляемом обстреле определенных целей (см. [сайт](#) авторов этой книги). Выводы Кларка принесли британцам большое облегчение. К счастью, в 1945 году Германия капитулировала, так и не успев доработать управляемые ракеты ФАУ-2. (В скобках заметим: несмотря на то что их нельзя было точно наводить на цель, именно ФАУ-2 стали технической основой развития космической программы США.)

Вы можете последовать примеру Кларка, то есть вернуться назад и пересмотреть содержание этапа формулирования проблемы после изучения предыдущих попыток ее решить (см. вставку «Некоторые методы изучения предыдущих исследований»).

## Некоторые методы изучения предыдущих исследований

- Интернет-поиск по ключевым терминам, используемым в анализе.

- Поиск в учебниках по статистике примеров анализа, сходного с предполагаемым.
- Собеседование с вашими аналитиками на предмет того, не приходилось ли им делать что-либо подобное.
- Анализ системы управления знаниями в вашей компании, если таковая имеется.
- Обсуждение проблемы с аналитиками из других (не конкурирующих с вами) компаний.
- Посещение конференций по аналитике (или хотя бы просмотр программ подходящих по тематике конференций в интернете) и выступлений по смежной тематике.

Возможно, вам придется изменить аналитический проект, масштаб анализа, варианты решений, а может быть, даже пересмотреть состав заинтересованных. Если вы это сделали или, наоборот, остались при первоначальном мнении по всем этим вопросам, то можно считать, что формулирование проблемы завершено, и переходить к ее решению методами количественного анализа.

## Уточнение формулировки проблемы

Хотя мы обрисовали аналитический процесс решения проблемы как линейную последовательность из шести шагов, объединенных в три этапа, но так или иначе этот процесс должен быть итеративным, предполагающим в случае необходимости возврат к предыдущим шагам и повторение их. Каждый следующий шаг способен выявить новые грани исходной проблемы, и всегда полезно задуматься над тем, как новая информация может изменить принятые на ранних этапах решения. Конечно, нельзя без конца пересматривать уже принятые решения, но полезно иногда задумываться над тем, как изучение предыдущих попыток решения способно повлиять на формулировку проблемы.

## Этапы формулирования проблемы

Адекватна ли разработанная формулировка проблемы? Если да, то ответ на большинство этих вопросов будет положительным.

- Удалось ли вам выявить проблему или возможности, касающиеся действительно важных для вашей организации вещей?

- Рассматривали ли вы альтернативные варианты решения проблемы?
- Определены ли те, кто сильнее всего заинтересован в решении, проводилось ли с ними обсуждение?
- Уверены ли вы в том, что способ, который выбран для решения проблемы, придется по вкусу всем заинтересованным и они смогут использовать полученные результаты для обоснования управленческих решений?
- Отдаете ли вы себе отчет в том, какие именно решения и в какие сроки будут приниматься на основе полученных в результате анализа данных и кто будет это делать?
- Вы начали с широкого определения проблемы, затем сузили его до конкретных, четко сформулированных вопросов, определили круг информации, которую надо собрать и представили возможные результаты?
- Способны ли вы определить вид аналитического проекта, который нужен для решения проблемы?
- Есть ли у вас сотрудники, способные помочь в реализации этого конкретного вида аналитического проекта?
- Провели ли вы комплексный анализ предыдущих исследований, проводившихся внутри или вне вашей организации по этой тематике, ознакомились ли с их результатами?
- Пытались ли вы пересматривать формулировку проблемы на основе изучения предыдущих поисков ее решения?

В своем блоге аналитик в сфере розничной торговли Рама Рамакришнан, в настоящее время CEO стартапа CQuotient, приводит прекрасный пример, в котором потребовалось переформулировать проблему для анализа<sup>[24]</sup>.

Рассмотрим проблему таргетирования (определения целевой аудитории), типичную для прямого маркетинга. Необходимо принять решение, кому из потребителей разослать коммерческие предложения, потому что рассылка по всем адресам обойдется слишком дорого. Эта проблема возникла давно; ее изучением занимались многие исследователи и практики. Наиболее распространенный подход к ее решению такой.

1. Рассылка тестовых предложений только выборке покупателей.
2. Использование результатов тестовой рассылки для построения «модели реагирования» потребителей на рассылку в зависимости от их характеристик, истории покупок и т. п.

3. Рейтинговая оценка каждого потребителя, включенного в базу, и рассылка предложений тем, кто получил максимальный рейтинг.

Эта последовательность действий выглядит вполне логичной и способной обеспечить необходимый результат. А может быть, и нет.

Термин «модель откликов» предполагает, что рассылка станет непосредственной причиной покупки. Но в действительности потребитель мог бы просто зайти в супермаркет и купить этот продукт (речь идет о торговцах, имеющих несколько каналов сбыта, а не о фирмах, торгующих по каталогам).

Модели откликов нацелены в первую очередь на то, чтобы отделить покупателей, склонных совершать покупки по собственной инициативе, от тех, кто совершает их под влиянием рассылок коммерческих предложений. Менеджерам требуется выявить как раз последнюю категорию. Рассылка предложений тем покупателям, которые купят продукт так или иначе, а также тем, кто не станет покупать его ни при каких обстоятельствах, это пустая трата времени, которая лишь напрасно раздражает адресатов. Поэтому особенно важно выделить ту категорию покупателей, которых именно рассылка коммерческих предложений заставит совершить покупку.

Такая модернизированная формулировка проблемы выделения целевой категории потребителей, а также методов ее решения появилась относительно недавно. У нее несколько названий: инкрементное моделирование, моделирование поведения потребителей. Разработана она гораздо менее детально по сравнению с традиционной моделью управления потенциальными продажами. Тем не менее для многих розничных сетей инкрементное моделирование — намного более подходящий и полезный метод, позволяющий сформулировать и решить проблему выделения целевого сегмента потребителей.

В этом примере изучение предыдущих поисков решения выявило новые методы, что и стало толчком к пересмотру формулировки. Рамакришнан в таких ситуациях предлагает следующее: «Хотя новые методы исследований по определению еще не опробованы в достаточной мере, но благодаря своей простоте могут дать результат

достаточно быстро».

Мы завершим главу, посвященную формулированию проблемы, двумя примерами: один из области бизнеса, а другой — из области права. В обоих формулировка проблемы сыграла решающую роль с точки зрения конечного результата, но в одном случае она оказалась удачной, а во втором некорректной. Правда, мы пока не обсуждали остальные этапы количественного анализа, но думаем, что это не помешает читателям разобраться в сути приводимых примеров.

## Пример аналитического мышления: Transitions Optical

Одна из наиболее распространенных проблем в бизнесе — определить сумму затрат на тот или иной вид деятельности. Особенно сложно рассчитать расходы на маркетинг. Автор концепции универсамов Джон Ванамейкер и кое-кто из розничных торговцев в Европе до него любили говорить: «Я знаю, что половину денег на рекламу трачу зря; проблема в том, что я не знаю, какую именно». Но сегодня розничные компании применяют количественный анализ, чтобы установить, какие сбытовые затраты эффективны, а какие нет и каким должно быть их распределение по отдельным статьям, чтобы общая сумма была потрачена с максимальной пользой. Обычно такой анализ называют *маркетинг-микс анализом*, и он приобретает все большую популярность среди компаний, продающих товары непосредственно населению.

**Определение и формулирование проблемы.** Компания Transitions Optical продает фотохромные линзы для очков и является дочерней компанией PPG и Essilor, которые постоянно давят на менеджеров, желая сократить маркетинговые расходы. В частности, PPG не занимается маркетингом розничных продаж, поэтому скептически воспринимает информацию о стоимости и ценности рекламных мероприятий и акций по продвижению продуктов. В целом они считают, что расходы на маркетинг неоправданно высоки, но не могут привести данные, на основе которых можно было бы определить их оптимальный уровень. Менеджеры Transitions Optical решили сформулировать проблему так: оптимизация отдельных видов затрат на маркетинг с целью стимулирования роста объема продаж на

каждый инвестированный доллар. По словам директора по маркетингу Грейди Ленски, в компании «считали маркетинг своего рода искусством, но теперь требуется научный подход».

**Изучение предыдущих поисков решения.** Никакого поиска просто не было. В компании имелись данные о потребителях, пригодные для такого анализа, но в отрывочном виде, рассредоточенные по разным подразделениям. Ленски и некоторые его коллеги понимали, что в принципе анализ эффективности различных маркетинговых подходов провести можно, но не знали как.

**Моделирование (выбор переменных).** Модели оптимизации маркетинг-микса, приобретавшие все большую популярность в крупных компаниях и применявшиеся для оптимизации маркетинговых расходов, обычно включают показатели маркетингового отклика, расходов на сбыт и рентабельности отдельных видов продуктов. Ежемесячную или еженедельную сумму расходов на рекламу и сбыт, а также оптимальный уровень цен на продукты можно рассчитать с помощью оптимизационных моделей методами линейного и нелинейного программирования. В качестве критериев принимаются максимальное увеличение объема продаж, рентабельности или обоих показателей. Модели позволяют определить, какие средства массовой информации выбрать для размещения рекламы, чтобы ее эффект был максимальным. Как правило, в них включается ряд «контрольных» переменных, которые могут повлиять на расходы и поведение потребителей: это, например, погода или макроэкономические факторы.

**Сбор данных.** Для Transitions Optical этот шаг оказался наиболее трудным, поскольку компания работает с посредниками (например, с оптическими лабораториями), а с конечными потребителями практически не контактирует и не имеет о них информации. Соответственно, невозможно адекватно определить, заметил ли потребитель рекламу и повлияло ли это на объем продаж. Transitions Optical пришлось в течение нескольких лет вести сбор данных о потребителях через своих торговых партнеров (а среди них были и

конкуренты их материнских компаний). Ленски когда-то возглавлял дистрибьюторскую фирму, поэтому занимался организацией сбора данных. Данные о потребителях попадали в Transitions Optical в тридцати разных форматах, но в итоге их удалось конвертировать и объединить в общую базу. Ленски замечал, что его департаменту маркетинга пришлось упорно убеждать различные подразделения в необходимости предоставить необходимые данные. Предыдущий аналитический проект пришлось проводить, не опираясь на общую базу данных.

**Анализ данных.** Transitions Optical наняла внешнего консультанта для анализа данных, поскольку никто из постоянных сотрудников не имел опыта работы с оптимизационными моделями маркетинг-микса. Сначала такой анализ занимал несколько месяцев, поскольку требовалось сначала собрать данные, затем разработать модель с учетом многочисленных внешних факторов, влияющих на поведение потребителей (погода, маркетинговые акции конкурентов и т. п.). Сейчас такие модели в достаточной степени апробированы и отработаны, поэтому результат можно получить в течение нескольких дней.

**Результаты и необходимые меры.** Менеджеры компании чувствовали, что интерпретация результатов анализа и оформление их для клиента — очень серьезная работа, так что для этого в штат наняли специальных сотрудников. Они ознакомились с разработанной внешним консультантом оптимизационной моделью, совместно с топ-менеджерами обсудили ее особенности и степень соответствия их взглядам на ситуацию на рынке. В итоге Transitions Optical приняла решение увеличить расходы на сбыт, в частности на телевизионную рекламу.

## Пример аналитического мышления: People vs. Collins

Под названием People vs. Collins известен суд присяжных, слушавший на шумевшее дело супругов Коллинз, в ходе которого математические методы и теория вероятности были использованы для доказательства виновности подсудимых. Однако проблему идентифицировали

неверно, и результат оказался неутешительным<sup>[25]</sup>. Суд признал Малкольма Коллинза и его жену Джанет Коллинз виновными в ограблении второй степени. Малкольм обжаловал это решение, и Верховный суд штата Калифорния отменил его, подвергнув сомнению результаты статистических выкладок и способ их представления в суде. Рассмотрим этот случай в соответствии с требованиями к стандартной процедуре количественного анализа.

**Определение проблемы.** Сделав кое-какие покупки, миссис Хуанита Брукс шла домой по переулку в районе Сан-Педро. Внезапно ее швырнул на землю человек, лица которого ей не удалось разглядеть. Придя в себя от шока и болезненного ушиба, миссис Брукс обнаружила пропажу кошелька, в котором оставалось от 35 до 40 долларов. Очевидец происшествия сообщил, что грабителей было двое: чернокожий мужчина с бородой и усами и белая женщина с белокурыми волосами, завязанными в конский хвост. Они скрылись с места преступления на желтой машине. Подозреваемых задержали быстро, но идентифицировать их как лиц, совершивших ограбление, не удалось и через неделю. Потерпевшая не смогла опознать Джанет Коллинз и не видела лица мужчины. Показаний свидетеля также оказалось недостаточно. Прокурор, видимо, стремясь спасти разваливающееся дело, решил построить обвинение на расчете вероятности случайного совпадения внешности обвиняемых с описанной свидетелем.

**Изучение предыдущих поисков решения.** Известно, что обычно суд не считает несовместимыми математические расчеты и право и не отрицает, что математику можно использовать в качестве инструмента для поиска доказательств. Существует несколько судебных прецедентов, когда обвинение использовало математическую вероятность для доказательства вины подсудимого.

**Моделирование (выбор переменных).** Прокурор решил рассчитать математическую вероятность того, что внешность и машина задержанных супругов Коллинз случайно совпали с описанием преступников.

**Сбор данных.** Прокурор вызвал в качестве свидетеля преподавателя математики из местного колледжа. В своих показаниях он постарался убедить присяжных в том, что вероятность совпадения характерных особенностей случайной пары жителей города с описанием преступников ничтожна. В данном случае частные вероятности совпадения каждой из особенностей, упомянутых в описании преступников, можно оценить следующим образом:

Чернокожий мужчина с бородой	1 из 10
Мужчина с усами	1 из 4
Белая женщина с прической «конский хвост»	1 из 10
Белая женщина с белокурыми волосами	1 из 3
Автомашина желтого цвета	1 из 10
Мужчина и женщина разной расовой принадлежности в автомобиле	1 из 1000

**Анализ данных.** Преподаватель математики предположил, что все эти частные вероятности независимы, и тогда вероятность их совпадения можно рассчитать, перемножив показатели из таблицы.

$P(A)$  = вероятность того, что произвольно выбранная пара совпадет по описанию с описанием подозреваемых равна

$$\frac{1}{10} \times \frac{1}{4} \times \frac{1}{10} \times \frac{1}{3} \times \frac{1}{10} \times \frac{1}{1000} = \frac{1}{12\,000\,000},$$

или один шанс из 12 миллионов.

Презентация результатов и проведение мероприятий. Прокурор заключил, что вероятность наличия у случайно выбранной пары всех указанных характерных особенностей составляет один шанс на двенадцать миллионов. Соответственно, отсюда можно сделать вывод о том, что у пары, у которой присутствуют все эти особенности, лишь один шанс из двенадцати миллионов оказаться невиновными. Жюри присяжных вынесло вердикт: «Виновны».

Коллинзы подали апелляцию. Верховный суд штата Калифорния решил, что жюри присяжных оказалось под чрезмерным впечатлением от мистики математических терминов и расчетов и не сумело оценить их значимость и связь с делом. Верховный суд отменил приговор,

подвергнув сомнению статистические выкладки и способ их представления присяжным. В частности, расчеты имели два существенных недостатка. Во-первых, следствие не привело никаких доказательств надежности исходных данных. Во-вторых, существенным недостатком этих расчетов было отсутствие уверенности во взаимной независимости частных вероятностей (так, например, мужчины с бородой, как правило, носят и усы).

Еще более важно то, что прокурор неправильно сформулировал проблему. Даже если допустить, что сами расчеты выполнены корректно, из них не следует, что Коллинзы виновны, поскольку они не дают ответа на вопрос о том, что если в мире найдется несколько пар с такими характерными особенностями, то какая из них действительно виновна в преступлении?

Но самое главное — это тот факт, что расчеты отвечают вовсе не на тот вопрос, который ставило обвинение, а именно какова вероятность того, что подозреваемые случайно подошли под описание преступников. На самом деле расчеты отвечают на вопрос, какова вероятность существования других пар с аналогичными характерными особенностями при том, что обвиняемые тоже подходят под описание. В зависимости от того, сколько всего пар в Лос-Анджелесе, вероятность существования хотя бы одной пары с такими же характерными особенностями может достигать 40 процентов (см. [сайт книги](#)). В этом случае об отсутствии обоснованных сомнений в виновности Коллинзов речь уже не идет; напротив, вероятность существования в Лос-Анджелесе более чем одной такой пары и того, что именно другую пару видели на месте преступления, становится весьма существенной.

После рассмотрения всех обстоятельств дела, в том числе свидетельских показаний, Верховный суд отменил обвинительный приговор. Некорректная формулировка проблемы привела к принятию неверного решения.

## ГЛАВА 3

# Решение проблемы

Многие считают это словосочетание, по крайней мере в некоторых отношениях, главным в количественном анализе — ведь именно здесь проводятся аналитические процедуры и проблема получает решение. Конечно, все это очень важно. Но операции на этом этапе более структурированы и точнее определены, чем на предшествующей и последующей фазах. Если у вас нет математической и статистической подготовки, то, скорее всего, вы передадите выполнение этих операций людям с необходимыми навыками и знаниями (см. вставку «Как найти кванта»). Но вне зависимости от ваших личных познаний в математике полезно получить общее представление об основных этапах решения проблемы.

### Как найти кванта

Если для решения вашей проблемы требуется количественный аналитик, то существует несколько способов отыскать его.

- Если вы работаете в большой компании, наверняка несколько квантов найдутся в штате. Загляните в отдел маркетинговых исследований, производственную лабораторию, отдел бизнес-аналитики.
- Если ни одного кванта среди сотрудников отыскать не удалось, то можно обратиться к целой армии независимых консультантов. Проведите интернет-исследование по запросу «консультанты по бизнес-аналитике».
- Если вы хотите привлечь кванта из-за рубежа, то лучше всего обратить внимание на Индию, в частности на компании Mu Sigma, Fractal Analytics и Genpact.
- Возможно, в местном университете удастся найти профессоров или студентов-старшекурсников, специализирующихся на количественном анализе; позвоните заведующему кафедрой статистики, к примеру.
- Если вы считаете необходимым взять кванта на постоянную работу, можно просмотреть объявления на сайтах вакансий, где, как правило, есть соответствующие предложения: например, на сайте Simply Hired есть страница с резюме количественных аналитиков, а на сайте [analyticrecruiting.com](http://analyticrecruiting.com) — с резюме статистиков. Можно обратиться в специализированную рекрутинговую фирму.

Сначала ознакомимся с последовательностью выполняемых этапов. Мы ведь помним, что речь идет об аналитическом проекте, направленном на проверку гипотезы. Сначала мы формулируем проблему ([глава 2](#)), затем переходим к моделированию и выбору переменных (первый шаг на этом этапе решения проблемы), а в результате можно будет выдвинуть гипотезу, требующую подтверждения или опровержения. Затем аналитик собирает данные и решает проблему. На каждом из этих шагов необходимо понимать или хотя бы предполагать, как функционирует исследуемый мир, и тогда на основе анализа данных можно будет сделать вывод о том, была ли правильной исходная гипотеза. Однако есть несколько видов анализа, не требующих предварительного выдвижения гипотезы. В интеллектуальном поиске данных и *машинном обучении* (когда модели разрабатываются на основе закономерностей, выявленных в имеющихся данных, с помощью программного обеспечения давая быстрый и оптимальный результат) аналитик просто вводит в компьютер массив данных и запускает поиск закономерностей. Все гипотезы выдвигаются уже потом, на этапе интерпретации и распространения результатов.

Нам не слишком нравится этот подход: в основном потому, что зачастую он дает необъяснимые результаты. А поскольку ни один аналитик не пытался использовать анализ данных для подтверждения своих взглядов на происходящее вокруг, то и комментировать результаты анализа или убеждать в необходимости изменить решение на их основе никто не будет. Однако иногда случаются обстоятельства, в которых подход к анализу как к «черному ящику» может сэкономить немало времени и труда аналитикам. В среде больших данных, где постоянно генерируются колоссальные массивы информации, у аналитика не всегда есть возможность формулировать гипотезы до проведения анализа данных. Например, при размещении рекламы на сайтах издательств решения принимает автоматизированная система в тысячные доли секунды, а компании, занимающиеся этой работой, генерируют несколько тысяч статистических моделей каждую неделю. Очевидно, такой вид анализа не рассчитан на выдвижение гипотез и рассмотрение результатов людьми, поэтому машинная работа здесь абсолютно необходима. Но по большей части в дальнейшем изложении

мы будем иметь дело с этапами и методикой анализа на основе проверки гипотез.

### Шаг 3. Моделирование (выбор факторов)

Модель — это *преднамеренно упрощенное* представление определенного события или ситуации. Термин «преднамеренно» означает, что модель разрабатывается специально для решения конкретной проблемы. Термин «упрощенно» говорит о том, что следует исключить из рассмотрения все банальные и несущественные детали, выделив важные, полезные и ключевые особенности, определяющие специфику проблемы. Проиллюстрируем процедуру выбора факторов на примере.

#### 3. Моделирование



Модель можно сравнить с карикатурой. Она заостряет внимание на некоторых чертах — носе, улыбке, кудрях, — и на их фоне другие черты теряют выразительность. Хорошая карикатура отличается тем, что отдельные черты выбираются обдуманно и эффективно. Точно так же модель акцентирует внимание на отдельных особенностях реального мира. При построении любой модели вам придется действовать избирательно. Нужно выбрать именно те особенности, которые имеют отношение к решению вашей проблемы, и пренебречь остальными. Модель носит схематичный характер, чтобы помочь пользователю сфокусироваться на исследуемой проблеме <sup>[26]</sup>.

Отсюда следует, что модели не могут быть абсолютно корректными.

Знаменитый статистик Джордж Бокс как-то заметил, что «...все модели некорректны, но некоторые при этом полезны»<sup>[27]</sup>. Ключевая проблема в том, чтобы определить, когда модель приносит пользу, а когда она некорректна настолько, что искажает реальность. В [главе 5](#) мы подробнее поговорим об этом. А пока заметим, что одним из ключевых является вопрос о выборе факторов для включения в модель.

Каким образом отбираются факторы для модели и прогнозируются их взаимосвязи? По большей части мы в этом вопросе руководствуемся субъективными соображениями. Гипотеза, то есть априори разработанная концепция анализа, представляет собой не более чем научнообразные предположения о том, какие факторы имеют наибольшее значение в каждом конкретном случае. На этом этапе разработка модели требует логического мышления, опыта и знакомства с предшествующими исследованиями. Только в этом случае можно с большой долей уверенности предположить, какие зависимые (те, которые нужно прогнозировать или объяснить) или независимые факторы сыграют основную роль. Можно попытаться протестировать модель — именно это отличает аналитическое мышление от менее точных методов принятия решений вроде интуиции.

Например, если вы социолог и пытаетесь прогнозировать динамику дохода семьи (зависимая переменная), то можно предположить, что независимыми переменными в вашей модели будут возраст, образование, семейный статус и количество работающих постоянно членов семьи. Именно эти переменные имеют смысл при прогнозировании семейного дохода. Впоследствии, в процессе количественного анализа (а точнее, на этапе анализа данных) вы можете обнаружить, что модель недостаточно точно отражает реальную ситуацию, и захотите пересмотреть состав переменных при условии, что по новым переменным можно получить данные.

Даже очень субъективные модели и переменные могут быть полезны для уточнения проблемы. Например, Гарт Сандем, известный популяризатор науки, математики, юморист и писатель на темы гик-культуры, многие жизненные проблемы решал путем анализа субъективно отобранных, но все равно полезных переменных<sup>[28]</sup>. В частности, так он подходил к решению вопроса о том, какое именно

домашнее животное лучше выбрать и стоит ли его заводить вообще.

Какие переменные человек принимает во внимание, решая, заводить ли домашнее животное? Сандем отобрал следующие:

- Постоянная жизненная потребность в любви ( $D$ , 1–10, где 10 баллам соответствует жизнь как у начальника тюрьмы днем и честного налогоплательщика ночью).
- Общий уровень ответственности ( $R$ , 1–10, где 1 балл соответствует убежденности в том, что «дети, налоговый инспектор и дела как-нибудь сами устроятся, если оставить их в покое»).
- Наиболее продолжительная поездка в последние шесть месяцев ( $T$ , дней).
- Продолжительность сверхурочных ( $H$ , часов в день).
- Ваша терпимость к проделкам других существ ( $M$ , 1–10, где 1 балл означает, что вы ведете себя как Стервелла де Виль, а 10 баллов — как доктор Дулиттл).
- Насколько вы заботливы ( $N$ , 1–10, где 1 балл означает «мой кактус засох»).

Все эти переменные весьма субъективны, но они, по всей видимости, полезны и, уж конечно, забавны. Сандем вывел следующее уравнение (выглядит довольно устрашающе!), где обобщающим показателем является  $F_{ido}$  — индекс готовности к заведению домашнего питомца:

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

Наиболее важной переменной в этом уравнении является  $D$  — потребность в любви, которая прямо пропорционально связана с результирующим показателем. Неплохо также, если у вас есть немного свободного времени ( $H$ ), чтобы проводить его с питомцем, и вы ответственный человек ( $R$ ). Эти две переменные также прямо пропорционально влияют на  $F_{ido}$ . Но если вам приходится много ездить,

значение вашего индекса существенно снизится. В зависимости от итогового результата Сандем предлагает выбрать одно из следующих домашних животных:

- если  $F_{ido}$  менее 1, то даже морские рачки будут слишком обременительны;
- если  $F_{ido}$  составляет от 1 до 2, попробуйте завести золотых рыбок;
- если  $F_{ido}$  составляет от 2 до 3, можно завести кошку;
- если  $F_{ido}$  превышает 3, то можно взять собаку.

Джин Хо подставил собственные значения в это уравнение и получил значение индекса готовности к заведению домашнего питомца 0,7, а значит, ему не стоит рисковать даже с кактусом.

Конечно, кто-то может сказать, что слишком большая точность расчетов при решении данного вопроса не требуется, но так или иначе этот пример показывает, что даже очень субъективные и банальные решения можно оценить количественно и смоделировать.

Какие переменные отобрать, а какие отбросить — зависит от цели разработки модели и того, связана ли переменная непосредственно с решением проблемы. Например, если вы рисуете карту Нью-Йорка, то расстояния между точками имеют большое значение и должны быть пропорциональны реальным расстояниям. Однако если вы рисуете схему нью-йоркского метро, то расстояния между станциями на карте совсем не обязательно должны быть пропорциональны расстояниям на местности. Ведь главная цель схемы метро — это показать, как можно добраться от одной станции до другой.

Еще один прекрасный пример важности тщательного выбора переменных модели — это спор по поводу того, кто является автором серии опубликованных в 1861 году писем. Десять писем, подписанных Квинтусом Куртиусом Снодграссом, появились в *New Orleans Daily Crescent*. В них мистер Снодграсс (ККС) описывал свои военные приключения во времена службы в Национальной гвардии Луизианы. Сразу после публикации письма не привлекли особого внимания. Они

впервые попали в поле зрения широкой публики лишь в 1934 году, то есть спустя семьдесят три года после выхода из печати. О них в своей книге Mark Twain, Son of Missouri упомянула Минни Брашер. В частности, она привела текст одного из писем, пересказала содержание трех других и сделала смелый вывод о том, что «письма ККС имеют огромное значение в качестве свидетельства становления Марка Твена как юмориста; именно Марка Твена следует признать их автором, а некоторые различия в стиле можно объяснить его стремлением выработать свой собственный литературный стиль»<sup>[29]</sup>. Оставшиеся шесть писем ККС опубликовал и проанализировал Эрнст Лейзи в 1946 году<sup>[30]</sup>. Проведенный им тщательный анализ аналогий позволил утверждать, что письма действительно написаны Твеном, но кое-кто из литературных исследователей до сих пор считает, что у них был другой автор.

В русле исследований вопроса о том, действительно ли Шекспир был автором всех приписываемых ему произведений, Томас Менденхолл в конце двадцатого века опубликовал две статьи, в которых изложил статистический подход к проблеме определения авторства. Топ-менеджер нефтяной компании Клод Бринегар, имевший хорошее университетское образование и увлекавшийся коллекционированием первых изданий книг Марка Твена, изучил историю вопроса и применил метод Менденхолла, впоследствии получивший название *стилометрии*, или количественного анализа литературного стиля, к письмам ККС.

Этот метод основан на предположении о том, что, хочет он того или нет, каждый автор чаще использует одни слова, чем другие, и сохраняет одинаковый литературный стиль, по крайней мере в долгосрочной перспективе. С позиций количественного анализа это означает, что доля слов определенной длины будет постоянной во всех текстах, написанных данным автором. Если доля слов определенной длины в двух разных текстах существенно отличается, это можно считать подтверждением того, что тексты написаны разными авторами. В качестве переменных для анализа писем ККС выбирались слова различной длины, и их удельный вес сравнивался с аналогичными показателями из работ, определенно принадлежавших

перу Твена. Для проверки авторства проводился тест по критерию согласия. Результаты тестирования показали, что расхождения по набору переменных слишком велики, чтобы считать их случайными, — поэтому вряд ли Марк Твен является автором этого произведения (подробности см. на [сайте книги](#))<sup>[31]</sup>.

Далее в этой главе мы еще поговорим об анализе текстов (в противоположность анализу чисел), а пока отметим, что Бринегар в процессе анализа перевел слова в числа.

## Шаг 4. Сбор данных (измерения)

На следующем шаге анализа проводится сбор данных и измерения выбранных переменных. Измерение — это определение значения переменной; массив данных — это набор таких значений. Существуют разные способы измерения переменных (см. вставку «Способы измерения переменных»). Сформулированная проблема сначала представляется в виде набора переменных в процессе моделирования, а затем приобретает вид массива данных в результате измерения.

### Способы измерения данных

Известны три основных метода измерения данных.

**Двоичные переменные.** Такие переменные имеют только два значения, и для целей статистического анализа лучше определять их как наличие или отсутствие определенного фактора со значениями 0 и 1. В качестве примера можно привести данные о поле респондентов, когда возможен выбор двух значений: женщина или мужчина (в первом случае переменная приобретает значение 1, во втором — 0), или о наличии гражданства США (либо гражданин, либо нет).

**Категориальные (также называемые номинальными) переменные.** В этом случае переменная может приобретать одно из нескольких заранее определенных значений. Так измеряются цвет глаз, вкус мороженого, штат или район проживания. Поскольку перевод таких значений в количественную форму представляет определенные сложности, существует отдельное направление статистики, занимающееся анализом категориальных данных.

**Ординальные переменные.** Эти переменные имеют упорядоченные количественные значения, причем чем оно больше, тем сильнее выражен соответствующий признак. Таким образом, у этих переменных разница между 1 и 2 — это не то же самое, что разница между 5 и 6. Типичный пример ординальных переменных — шкала Ликерта, получившая название в честь автора, социолога Ренсиса Ликерта. Обычно применяется в опросах и включает такие значения, как «полностью согласен», «отчасти согласен», «не могу выразить отношение», «отчасти

не согласен», «не согласен». Несколько ординальных переменных, сведенных вместе, носят название шкалы Ликерта.

**Количественные (интервальные и рациональные) переменные.** Значения этих переменных выражены числами, обычно в стандартных единицах: вес в фунтах или килограммах, рост в дюймах или сантиметрах. Чем больше значение, тем сильнее выражен соответствующий параметр. Количественные переменные хорошо подходят для традиционных видов статистического анализа, например корреляционного или регрессионного.

Таким образом, массив данных организован с учетом переменных, выбранных на предыдущем шаге.

Если значения нужных вам переменных часто собирает и анализирует кто-то еще (иногда такие факты всплывают во время изучения предыдущих поисков решения), то этот этап будет несложным. Можно просто позаимствовать результаты измерений, полученные вашими предшественниками. Однако в некоторых случаях приходится вести работу самостоятельно. Нужно помнить, что даже субъективные события можно систематически измерять.

#### 4. Сбор данных



Предположим, что вам нужно собрать данные по волнующей в наше время многих (если судить по телевизионной рекламе) проблеме мужской потенции. Оказывается, что вам повезло: на эту тему уже проводился сбор данных, которые вполне подходят для ваших целей. Однако если бы вы были первопроходцем в этой области, то пришлось бы проводить сбор данных самостоятельно.

В 1990-е годы Р. С. Розен и его коллеги разработали компактный,

надежный и простой для изучения критерий потенции, чувствительный к изменениям в состоянии здоровья пациентов в результате лечения<sup>[32]</sup>. О проблемах с потенцией можно узнать только от самого пациента. Объективных диагностических тестов не существует, и это весьма усложняет жизнь практикующим врачам. Розен и его коллеги определили, что ключевыми переменными для анализа проблемы мужской потенции являются:

- регулярность эрекции
- сила эрекции
- частота возбуждения
- способность к половому акту
- удовлетворение

В их разрезе был организован сбор информации с использованием вопросов, приведенных в табл. 3.1.

*Таблица 3.1*

### Ключевые переменные для диагностирования эректильной дисфункции

За последние шесть месяцев					
Оцените вашу уверенность в том, что вы можете ощутить и поддерживать эрекцию	Очень низкая — 1 балл	Низкая — 2 балла	Средняя — 3 балла	Высокая — 4 балла	Очень высокая — 5 баллов
В тех случаях, когда с помощью сексуальной стимуляции удавалось добиться эрекции, была ли она достаточной?	Никогда или почти никогда — 1 балл	Изредка (гораздо меньше половины случаев) — 2 балла	Иногда (примерно в половине случаев) — 3 балла	В большинстве случаев (гораздо чаще, чем в половине случаев) — 4 балла	Всегда или почти всегда — 5 баллов

За последние шесть месяцев					
Насколько часто в процессе полового акта вам удавалось поддерживать эрекцию после введения?	Никогда или почти никогда — 1 балл	Изредка (меньше половины случаев) — 2 балла	Иногда (примерно в половине случаев) — 3 балла	В большинстве случаев (гораздо чаще, чем в половине случаев) — 4 балла	Всегда или почти всегда — 5 баллов
Насколько сложно вам поддерживать эрекцию до завершения полового акта?	Очень сложно	Довольно сложно	Сложно	Есть некоторые затруднения	Затруднений не испытываю
Как часто половой акт приносит вам удовлетворение (от общего числа попыток)?	Никогда или почти никогда — 1 балл	Изредка (гораздо меньше половины случаев) — 2 балла	Иногда (примерно в половине случаев) — 3 балла	В большинстве случаев (гораздо чаще, чем в половине случаев) — 4 балла	Всегда или почти всегда — 5 баллов

Вопрос о том, возможно ли ответы на них перевести в диагноз, решается довольно просто. Каждому варианту ответа присваивается балл от 5 до 25. Проблему с потенцией классифицировали по пяти степеням: серьезная (5–7), умеренная (8–11), от умеренной до незначительной (12–16), незначительная (17–21) и отсутствие проблемы (22–25). Этот простой в применении диагностический тест называется ПЕФ-5 (вариант Международного индекса эректильной функции из пяти вопросов) и прекрасно иллюстрирует способы сбора субъективной информации.

Неважно, каким объемом данных вы располагаете, — всегда остаются возможности собрать еще больше или расширить круг показателей, по которым собирались данные. После начала работы над проектом обычно выявляется, что тех показателей, которые были отобраны на этапе идентификации проблемы, недостаточно. Талантливый квант Рама Рамакришнан, о котором мы уже говорили в [главе 2](#), в своем блоге описал интересный способ улучшить качество данных: «Одно из моих любимых занятий — улучшать качество данных. Это означает не увеличивать их количество, а, скорее, получать новые по характеру данные по сравнению с теми, которые использовались до этого момента. Если у вас имеются

демографические данные, добавьте данные об объемах закупок. Если у вас и те и другие, попробуйте добавить функцию их свободного просмотра. Если у вас есть количественные данные, добавьте к ним текстовые (кстати говоря, в последней работе мы получили весьма обнадеживающие результаты, добавив к традиционным данным об объемах продаж и сбытовых мероприятиях текстовые данные о покупателях с целью их персонификации и моделирования потребительского поведения)»<sup>[33]</sup>.

Специалист по интеллектуальному поиску данных Ананд Раджараман также писал в своем блоге о возможностях улучшения качества анализа за счет включения новых данных.

Я веду курс по интеллектуальному поиску данных в Стэнфордском университете. Студентам поручают выполнить аналитический проект, включающий нетривиальный вариант интеллектуального поиска данных. Многие из них пытались разработать более совершенную методику подбора рекомендаций по поводу кино, чем в проекте Netflix Challenge.

Это яркий пример того, как действует конкуренция. Netflix предоставляет огромный массив данных о рейтингах 18 тысяч фильмов, выставленных почти полумиллионом посетителей сайта. Основываясь на этой информации, надо спрогнозировать рейтинги, которые выставят пользователи тем фильмам, которые они еще не оценивали. Первая группа аналитиков, которой удастся разработать методику, работающую лучше, чем Netflix Challenge, получит миллион долларов!

Студенты в моей группе пытались применить разные подходы для решения этой проблемы, причем одна команда использовала уже известные алгоритмы, а вторая — новые идеи. Их результаты позволяют взглянуть на проблему шире. Первая команда предложила очень сложный алгоритм, основанный на имеющихся данных. Вторая использовала довольно простой алгоритм, но зато на основе не только имеющихся, но и новых данных, которых в базах Netflix не было. Их позаимствовали из онлайн-базы данных о фильмах (Internet Movie Database). Какая из команд, по вашему мнению, добилась лучших результатов? Представьте себе, вторая! Ее результаты оказались почти так же хороши, как и результаты лучших участников конкурса Netflix!<sup>[34]</sup>

В том же посте Раджараман отмечает, что появившийся недавно источник информации — гипертекстовые ссылки — стал отличительной чертой поискового механизма Google по сравнению с прочими поисковиками, использовавшими только текст на веб-страницах. В своем высокорентабельном алгоритме AdWords, предназначенном для размещения рекламы, Google также использовал дополнительные данные, которыми на тот момент не интересовался ни один из конкурентов — коэффициент эффективности баннеров

(отношение числа щелчков к общему числу показов), рассчитывавшийся для каждого баннера рекламодателей.

Раджараман и Рамакришнан в один голос утверждают, что больший объем и лучшее качество данных почти в любом случае важнее, чем лучший алгоритм расчетов. Оба ссылаются на опыт розничного бизнеса и электронной коммерции, но и в других областях существует множество подобных примеров. Топ-менеджер команды НБА Houston Rockets Дэррил Морей является одним из лучших аналитиков в профессиональном баскетболе (мы вспомним о нем в [главе 6](#)). Он считает, что «реальное преимущество обеспечивают лишь эксклюзивные данные», и держит в штате нескольких квантов, анализирующих действия соперников в защите в каждой игре<sup>[35]</sup>. Кроме того, Морей стал одним из первых менеджеров в НБА, которые начали анализировать видеозаписи отдельных матчей.

В страховом бизнесе одним из факторов, долгое время отличавших компанию Progressive от менее склонных к аналитике компаний, стала ее уникальная база данных. Компания первой стала использовать кредитный рейтинг агентства FICO (этот пример рассматривается в [главе 4](#)) в качестве одной из переменных в модели страховых тарифов, а также в течение долгого времени использовала гораздо больше данных и переменных в анализе клиентского риска и расчете страховых тарифов, чем ее конкуренты. Progressive выступила первопроходцем в сборе данных о манере вождения автомобилей клиентами (конечно, с разрешения последних) и расчете страховых тарифов в зависимости от их водительских привычек (эту программу компания сейчас называет Snapshot). Вы можете не захотеть сообщать страховой компании такие сведения, но если проявите себя осторожным водителем, то получите скидку по страховке.

## Ценность вторичных данных

Многие аналитики самостоятельно собирают, а затем анализируют данные. Но иногда можно воспользоваться данными, собранными кем-то другим (так называемыми *вторичными данными*), и существенно сэкономить время. Обычно вторичные данные получают из результатов переписей, опросов, внутренней документации и других

подобных источников. Таких данных везде очень много, и они просто ждут, когда аналитики обратят на них внимание.

Иногда вторичные данные помогают получить очень важные результаты. Достаточно вспомнить, например, работу астронома Иоганна Кеплера. Он родился в бедной семье, но ему повезло получить очень точные вторичные данные о движении астрономических объектов, тщательно собиравшиеся в течение нескольких десятилетий. Необыкновенный математический талант и удача помогли ему разгадать тайны планет.

Данные достались Кеплеру в основном от датского дворянина и блестящего астронома Тихо Браге (1546–1601), который сумел сделать точные астрономические наблюдения при помощи уникальных инструментов еще до изобретения телескопа. При поддержке датского короля Браге построил исследовательский центр, получивший название Ураниборг (Небесный замок), и разместил в нем лучшую на тот момент в Европе обсерваторию. Он сам разработал и изготовил высокоточные измерительные инструменты, откалибровал их и каждую ночь в течение более чем двадцати лет вел астрономические наблюдения.

В 1600 году Браге пригласил Кеплера, блестящего, но бедного учителя, в помощники. Они не очень-то ладили: сказывалась разница в характерах и жизненном опыте. Браге опасался, что его умный молодой помощник со временем затмит его и станет лучшим астрономом своего времени. В следующем, 1601 году Браге внезапно заболел и умер. Разгорелся спор о его наследстве, и Кеплер понял, что если не будет действовать быстро, то навсегда потеряет возможность воспользоваться данными, собранными учителем. Он немедленно забрал результаты наблюдений (по его выражению, узурпировал их) и уже не выпустил из рук. Через два дня после похорон Браге Кеплер был назначен на его должность придворного математика. Наконец-то уникальная коллекция записей об астрономических наблюдениях была полностью в его распоряжении! Анализируя их, Кеплер сделал вывод, что орбиты планет имеют форму эллипса, а затем сформулировал свои знаменитые законы движения планет<sup>[36]</sup>.

Конечно, можно привести массу более современных примеров

использования вторичных данных. Например, источник вторичных данных компании Recorded Future прекрасно известен: интернет. Основатель компании — консультант по аналитике Кристофер Альберг, а основной вид деятельности — анализ информации в интернете на предмет частоты упоминания и классификации тех или иных событий и субъектов. Особое внимание компания уделяет подсчету предсказаний — упоминаний о будущем. Данные и аналитика пользуются спросом у государственных разведывательных служб, интерес которых к частоте упоминания террористических актов и войн легко объясним. Среди клиентов есть и финансовые компании, которые интересуются данными, отражающими настроения инвесторов и потребителей.

## Первичные данные

Но если вам не так повезло, как Кеплеру или Recorded Future, и не досталось ценных вторичных данных (а может быть, данных, имеющих отношение к вашей проблеме, пока просто не существует), то вам придется собрать их самостоятельно (это *первичные данные*). Существует несколько методов получения первичных данных: опрос, включающий разработку анкет и проведение интервью; наблюдения, в ходе которых наблюдатель открыто или скрытно фиксирует информацию; тщательно спланированные и контролируемые «сумасшедшие» эксперименты, предназначенные для изучения специфических проблем. Выбор метода сбора данных зависит от особенностей сформулированной проблемы и включенных в анализ переменных.

**Структурированные и неструктурированные данные.** В течение долгого времени почти все количественные аналитики работали со *структурированными данными*: данными в числовой форме, которые легко можно представить в табличном виде. Независимо от того, проводится ли анализ с помощью электронных таблиц, мощной статистической программы или старомодного калькулятора, все равно данные структурируются при помощи строк и столбцов (обычно в строках отражаются события или наблюдения, а в столбцах — значения соответствующих переменных). Все, что вам оставалось

выяснить, это сколько наблюдений следует сделать и сколько знаков после запятой показывать в таблице.

Но положение дел стало меняться с распространением в последние годы XX века анализа текстов. На примере истории с письмами Марка Твена мы показали, что в тексте можно искать не только числа, но и логические закономерности. Типичный вопрос: как часто повторяется в тексте то или иное слово? Текст представляет собой пример неструктурированных данных. Поскольку он состоит из определенной последовательности слов, его трудно разложить по строкам и столбцам таблицы. Однако лишь после 2000 года резко возросли объем и разнообразие неструктурированных данных. Именно этот год стал началом массированного использования интернета, когда компании вроде Recorded Future приступили к анализу огромных массивов данных в виде текста, изображений и щелчков мышки. Телекоммуникации и социальные медиа поставляют огромные объемы информации социальной направленности. Объем аудио- и видеоданных, которые хотели проанализировать организации, рос в геометрической прогрессии. Революция в генетике привела к необходимости анализировать большие объемы сведений о генах.

Сейчас мы официально вступили в век больших данных, когда обработка нескольких петабайт информации стала для организаций рутинным делом. (1 петабайт равен 1000 терабайт, или  $10^{15}$  байт, то есть 1 000 000 000 000 000 единиц информации.) Например, хранилище информации eBay имеет объем более чем в 40 петабайт. Каждое ваше нажатие на изображение видеокамеры или украшенной цветочным орнаментом вазы фиксируется в общей базе данных.

Анализ данных такого рода имеет существенные отличия от анализа структурированных количественных данных, особенно на первых шагах. Во многих случаях, прежде чем приступить к подсчету, требуется провести тщательную фильтрацию и классификацию, а также другие подготовительные операции. *Специалист по базам данных* — это человек, глубоко разбирающийся не только в анализе данных, но и в процедурах их подготовки к проведению анализа. Такие программные инструменты, как Hadoop и MapReduce, получают все большее распространение в организациях, сталкивающихся с

необходимостью анализа больших данных. Они предназначены для такой фильтрации и классификации данных, которая позволит применять количественные методы анализа. Видео- и аудиоинформация также требует серьезной обработки, прежде чем можно будет ее анализировать количественными методами. Во многих случаях после подготовки организация будет анализировать эти массивы данных при помощи традиционных статистических приложений.

Билл Франк из компании Teradata в своем посте в блоге Международного института аналитики подчеркивает<sup>[37]</sup>:

Неструктурированные данные в последнее время очень популярный предмет для обсуждения, поскольку слишком многие распространенные источники больших данных предоставляют их в неструктурированном виде. Но зачастую забывают об очень важном обстоятельстве: никакая аналитика не имеет дела напрямую с большими данными. Последние могут стать толчком к проведению анализа, но когда дело доходит до собственно аналитических процедур, то неструктурированные данные не обрабатываются. «Как же так?» — спросите вы. Позвольте объяснить.

Вот пример: отпечатки пальцев. Если вы любите сериалы вроде «CSI: полиция Майами», то постоянно видите, как эксперты идентифицируют их. Отпечатки пальцев представляют собой неструктурированные данные, причем довольно большого объема — если изображение высококачественное. Когда полицейские — в сериале или в жизни — сравнивают их, то есть ли смысл накладывать одно изображение на другое? Нет. Сначала они определяют несколько ключевых точек на каждом отпечатке. Затем по этим точкам формируется карта (многоугольник). Именно по этим картам производится сравнение. Особое значение имеет тот факт, что карта представляет собой структурированные данные, к тому же небольшого объема, даже если исходное изображение «весило» много. Как видите, хоть неструктурированные данные и необходимы для начала анализа, но в самом процессе обрабатываются не они, а полученные из них структурированные данные.

Всем понятный пример такого рода — анализ текстов. В общедоступных средствах массовой информации в последнее время принято вести смысловой анализ множества сообщений. Но можно ли непосредственно анализировать твиты, посты в Facebook и прочие посты и комментарии в соцсетях на предмет их смысловой оценки?

В действительности — нет. Текст необходимо разбить на фразы или слова. Затем определенным фразам и словам присваивается определение «положительный» или «отрицательный». В простом случае фразе или слову, определенному как «положительное», присваивается значение 1, «отрицательному» — -1, а «нейтральному» — 0. Смысл сообщения оценивается по сумме значений входящих в него слов или фраз. Таким образом, оценка ведется на основе структурированных количественных данных, полученных из первоначально неструктурированного источника — текста. Любой дальнейший анализ тенденций или стандартных моделей полностью основывается на структурированном, количественном выражении текста, но не на самом тексте.

Так же как в ситуациях, приведенных Франксом в качестве примера, многие приложения для обработки больших данных первоначально предназначались для обработки неструктурированных данных, но после того как те проходят через такие приложения, как Hadoop и MapReduce, можно их анализировать как структурированные данные с использованием статистических программ или инструментов визуализации.

## Шаг 5. Анализ данных

Поскольку сами по себе данные ни о чем не говорят, нужно проанализировать их и определить значения и взаимосвязи. Анализ данных включает выявление устойчивых моделей, или взаимосвязей между переменными, значения которых введены в массив данных. Если удастся выявить взаимосвязи, тогда можно объяснить динамику переменных. Тогда будет легче решить проблему.

### 5. Анализ данных



Предположим, что мы собрали данные по выборке избирателей относительно их намерения голосовать за того или иного кандидата. Метод сбора данных — опрос по телефону. Но в процессе анализа мы пытаемся выявить, каким образом регион проживания, образование, уровень дохода, пол, возраст и партийная принадлежность способны повлиять на выбор того или иного кандидата. Для обнаружения зависимостей в данных можно использовать целый ряд методов, начиная с достаточно простых — графиков, расчета удельного веса и средних значений переменных — и заканчивая сложными

статистическими исследованиями.

Параметры массива данных и сложность предстоящего анализа подскажут, какими именно методами лучше воспользоваться. В [главе 2](#) мы привели примеры таких методов. Если вы просто описываете сложившуюся ситуацию, то достаточно составить отчет или разработать набор графиков, показать, сколько анализируемых событий случилось в каждом временном интервале, и прокомментировать эту информацию. Обычно приходится приводить сведения о некоторых *показателях, отражающих основную тенденцию*, в частности о средних значениях — медианах.

Исходя из этих условий, потребуется программное обеспечение, ориентированное на составление отчетов. Сбалансированные системы показателей, сводные таблицы, тревожные сигналы — это все формы отчетов. Во вставке «Основные поставщики аналитического программного обеспечения» мы перечислили ключевых поставщиков программного обеспечения, обеспечивающего визуальное представление результатов анализа.

## Основные поставщики аналитического программного обеспечения

### ПРОГРАММЫ — ГЕНЕРАТОРЫ ОТЧЕТОВ

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

### ИНТЕРАКТИВНАЯ ВИЗУАЛЬНАЯ АНАЛИТИКА

- QlikTech QlikView
- Tableau
- TIBCO Spotfire

## КОЛИЧЕСТВЕННЫЕ МЕТОДЫ И СТАТИСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

- IBM SPSS
- R (свободно распространяемое программное обеспечение)
- SAS

У всех этих поставщиков программного обеспечения есть программы для графического представления данных, но некоторые из них специализируются именно на *интерактивной визуальной аналитике*, то есть визуальном представлении данных и отчетов. Иногда такие программы используют для простого построения графиков, иногда для исследования данных: законов распределения данных, позволяющих идентифицировать *выбросы* (точки с нетипичными значениями) и визуальную взаимосвязь между переменными. Таких поставщиков мы выделили в отдельный список.

Кроме того, в перечне выделена группа поставщиков, специализирующихся на еще одной категории аналитических программ — *количественных методах и статистическом моделировании*. В них статистика используется для выявления взаимосвязи между переменными и переноса закономерностей выборки на генеральную совокупность. Его формы — предсказательная аналитика, рандомизированное исследование и различные формы регрессионного анализа. Программное обеспечение для статистического моделирования и для генерирования различных отчетов разрабатывается разными группами поставщиков, хотя со временем они начинают смешиваться между собой.

Например, самая распространенная в мире аналитическая компьютерная программа Microsoft Excel (хотя большинство пользователей считает ее всего лишь электронной таблицей) способна решать некоторые задачи статистического анализа (и визуальной аналитики), равно как и генерировать отчеты. Однако если вам необходимо обработать большой массив данных или построить сложную статистическую модель, то возможностей Excel не хватит. Поэтому к данной категории программного обеспечения она не относится. В корпоративной среде для решения аналитических задач в

дополнение к Microsoft Excel часто используют и другие программы Microsoft, в том числе SQL Server (главным образом предназначенную для работы с базами данных и решения некоторых аналитических задач) и SharePoint (обеспечивает совместную работу над проектом и решение некоторых аналитических задач).

## Типы моделей

Аналитики и компании для решения аналитических задач и принятия решений на основе анализа используют множество типов моделей. Мы не собираемся учить читателей статистике, но считаем, что им было бы полезно знать, какие критерии применяют количественные аналитики, выбирая наиболее адекватную модель. Это поможет читателям сделать первые шаги в бизнес-аналитике и твердо усвоить ее основы. Если мы хотим знать, какие типы моделей лучше всего подойдут в том или ином случае, надо оценить специфику ситуации с точки зрения тех, кто принимает решения (или их аналитиков).

- Чтобы правильно выбрать модель, надо ответить на три основных вопроса.
- Сколько переменных подлежат анализу? Возможны такие варианты ответа: одна переменная (*одномерная модель*), две переменные (*двумерная модель*), три и более переменных (*многомерная модель*). Последний вариант ответа достаточен для решения любой проблемы.
- Требуется ли нам описание решения проблемы или просто ответы на поставленные вопросы? *Описательная статистика* просто описывает имеющиеся данные и не пытается делать выходящих за их рамки обобщений. Средние значения, медианы и стандартные отклонения — вот классический пример описательной статистики. Они весьма полезны, но не слишком интересны с математической или статистической точки зрения. *Индуктивная статистика* исследует выборку из какой-либо совокупности и распространяет выводы о средних характеристиках ее объектов на всю совокупность. Примеры такой статистики — корреляционный и регрессионный анализ

(см. далее): они включают оценку вероятности того, что взаимосвязи, выявленные на основе выборки, характерны и для всей совокупности. Статистики и количественные аналитики обычно отдают предпочтение индуктивной статистике по сравнению с описательной.

- Насколько точно можно оценить значения интересующих переменных? Некоторые методы оценки описаны во вставке [«Методы измерения данных»](#).

Конкретный тип используемой вами (или вашими квантами) модели зависит от того, какого вида ваш аналитический проект и какого типа данные. Некоторые характеристики проектов и массивов данных, а также моделей, выбранных для их обработки, описаны ниже. Мы рассмотрели далеко не все типы моделей, но из тех, которые изо дня в день используются организациями для аналитики, здесь представлены примерно 90 процентов.

*Модели с двумя числовыми переменными.* Если требуется установить взаимосвязь между двумя числовыми переменными, то проще всего это сделать с помощью *корреляционного анализа*. Это один из простейших видов статистического анализа. В типичном случае с его помощью можно установить, меняется ли одна переменная с изменением другой. Для примера возьмем рост и вес человека. Можно ли утверждать, что вес человека увеличивается с увеличением его роста? Как правило, так и бывает, поэтому можно утверждать, что эти две переменные коррелируют между собой. Поскольку корреляционный анализ является одним из методов индуктивной статистики, существуют способы определить: может ли определенный уровень корреляции быть случайным? Если вам, например, говорят, что «статистическая значимость связи равна 0,05», то это означает, что в пяти случаях из ста наблюдается согласованное изменение анализируемых показателей.

*Две категориальные переменные или больше.* Если вы используете данные опросов и они представлены *номинальными категориями* (например, мужской и женский пол; молодой, средний или пожилой возраст), то вам понадобится ряд аналитических процедур для анализа категориальных данных. Результаты этого вида анализа часто

оформляют в виде таблицы, в ячейках которой указано количество наблюдений. Например, если вы устанавливаете связь между полом и продолжительностью жизни, то обнаружите, что численность мужчин и женщин в молодом и среднем возрасте примерно одинакова, но поскольку женщины обычно живут несколько дольше, чем мужчины, то в старшем возрасте их численность будет выше. Если эта или подобная закономерность присутствует в вашем массиве данных, то таблица покажет значимую (то есть вряд ли случайную) взаимосвязь в соответствии со значением такого статистического критерия, как хи-квадрат. Взаимосвязь может быть значимой при уровне значимости 0,05 или 0,01. Такие бинарные категориальные переменные, как пол, можно также обрабатывать с помощью регрессионного анализа, используя при этом фиктивные переменные: то есть такие, которые получают значение 0 при отсутствии признака (например, мужского пола), и 1 при его наличии.

*Более чем две количественные переменные.* Если количественных переменных более двух, то проводится углубленный анализ корреляционной связи, называемый *регрессионным анализом*: иногда множественной регрессией (если для объяснения динамики одной переменной используются несколько других переменных), а иногда линейной регрессией (если взаимосвязь между переменными остается стабильной (линейной) во всех интервалах их значений). Регрессия представляет собой метод подбора уравнения (или линии, если речь идет о графическом выражении), описывающего совокупность собранных в прошлом данных. Если вам это удалось, то с помощью уравнения регрессии можно прогнозировать поведение переменных в будущем. В регрессионной модели каждой независимой переменной приписывается определенный коэффициент, отражающий (или прогнозирующий) ее «вес» в модели.

В качестве примера множественной линейной регрессии можно привести случай из практики экономиста из Принстона Орли Ашенфельтера. Он использовал регрессионный анализ для прогнозирования аукционных цен на марочные французские вина. Его прогноз аукционных цен основывался на погоде в период сбора урожая вин этого года — и вызвал шок в среде экспертов по винам и

даже привел их в ярость. (Газета New York Times опубликовала на первой странице статью об этом прогнозе под названием «Уравнение цены на вино вывело из строя многие носы»<sup>[38]</sup>.) Если у вас есть хорошее уравнение, то зачем вам эксперты?

Большинство экспертов сходятся в том, что хорошее вино получается в том случае, если предшествующая зима была дождливой, в сезон созревания винограда стояла теплая погода, а в сезон его сбора — сухая. Таким образом, Ашенфельтер выбрал три независимые переменные, относящиеся к погоде и влияющие на качество винограда: средняя температура воздуха в период созревания и количество осадков в период сбора винограда, а также количество осадков в предшествующую зиму. Кроме того, поскольку вкус вина, как правило, зависит от его выдержки, еще одной независимой переменной стала продолжительность выдержки в годах.

Качество сбора винограда влияет на цену зрелого вина, которая и становится зависимой переменной, которую Ашенфельтер пытался предсказывать. Он собрал информацию о ценах на лондонском аукционе за шесть бутылок бордо шато в 1960–1969 годы. Этот период был выбран потому, что вина, сделанные из урожая сборов этих лет, уже созрели, а в их качестве не было сомнений. Данные о значениях независимых переменных предоставило бюро прогнозов погоды из района выращивания винограда.

Ашенфельтер составил регрессионное уравнение логарифма цены вина, включающее показатели возраста вина и параметров погоды. Он получил такое выражение:

$$\text{Качество вина} = 12,145 \text{ (константа)} + 0,0238 \times \text{Возраст вина} + 0,616 \times \text{Средняя температура периода созревания} + 0,00386 \times \text{Количество осадков в период сбора урожая} + 0,0017 \times \text{Количество осадков предшествующей зимой.}$$

Как показывают значения коэффициентов при переменных, возраст вина, умеренная температура в период созревания и количество осадков в течение предшествующей зимы оказывают прямое положительное влияние на цену вина. Осадки в период сбора урожая оказывают негативное влияние на качество вина. *Коэффициент детерминации R-квадрат* (подробнее см. во вставке «Основные статистические концепции и аналитические приемы») для этого

уравнения составляет 0,828, что означает, что включенные в уравнение переменные на 83 процента объясняют отклонения в ценах на вино. Коротко говоря, эти переменные в совокупности играют определяющую роль в процессе установления цен. Легко понять, почему эксперты сочли эти результаты до некоторой степени спорными и менее интересными, чем бесконечные разговоры о *терруаре*<sup>[39]</sup>, дубовых бочках и переспевшем винограде.

## Основные статистические концепции и аналитические методы<sup>[40]</sup>

**Дисперсионный анализ (ANOVA).** Статистический тест на равенство средних значений двух и более групп.

**Причинно-следственная связь.** Взаимосвязь между двумя событиями (причиной и следствием), когда второе событие считается последствием первого. В типичном случае причинно-следственная связь — это зависимость между рядом факторов (причинами) и результирующим фактором (следствие). Наличие причинно-следственной связи требует соблюдения трех условий:

- Событие-причина должно предшествовать событию-следствию во времени и пространстве.
- При наличии причины наступает следствие.
- При отсутствии причины следствие не наступает.

**Кластеризация, или кластерный анализ.** Распределение результатов наблюдений (записей в базе данных) по группам (кластерам) таким образом, что результаты в одной группе имеют сходные черты, в то время как результаты разных групп отличны друг от друга. Кластеризация является основной задачей интеллектуального поиска данных и стандартным приемом анализа статистических данных в самых разных областях.

**Корреляция.** Степень зависимости двух или более переменных друг от друга. Степень зависимости выражается коэффициентом корреляции, принимающим значения в интервале от 1,0 до -1,0.

Если коэффициент корреляции равен +1 (полная положительная корреляция), то это означает, что обе переменные пропорционально изменяются в одинаковом направлении.

Коэффициент корреляции равен 0 — между переменными нет связи.

Если коэффициент корреляции равен -1 (полная отрицательная корреляция), то это означает, что при возрастании одной переменной вторая уменьшается.

Наличие корреляции не обязательно означает, что имеется причинно-следственная связь. Иначе говоря, корреляция является необходимым, но не достаточным условием причинности.

**Факторный анализ.** Статистический метод, раскрывающий взаимосвязь между многими переменными или объектами. Это позволяет объединить взаимосвязанные переменные в группы, называемые факторами. Такой прием часто используется для структурирования и/или сокращения количества видов данных. Например, если исследователю предстоит проанализировать более сотни переменных, факторный анализ позволит объединить их в десяток комбинированных показателей, каждый из которых отражает динамику десятка исходных переменных.

**Зависимая переменная.** Переменная, значение которой неизвестно и подлежит прогнозированию или объяснению. Например, если вы хотите предсказать качество вина урожая определенного года на основе среднегодовой температуры периода созревания, количества осадков в период сбора урожая и в предшествующую зиму, то качество вина будет зависимой переменной. Иногда используются еще термины «объясняемая переменная» и «результатирующий фактор».

**Независимая переменная.** Переменная, значение которой известно и применяется для прогнозирования или объяснения динамики зависимой переменной. Например, если вы хотите предсказать качество вина на основе исследования различных переменных (средняя температура в период созревания, количество осадков в период сбора и предыдущей зимой, возраст вина), то эти переменные и будут независимыми. Иногда их называют еще объясняющими переменными, переменными регрессии, фактор-аргументами.

**Регрессия.** Статистический метод, позволяющий построить уравнение для оценки неизвестного значения зависимой переменной через известные значения одной или более независимых переменных. Простая регрессия означает, что для оценки зависимой переменной используется одна независимая переменная. Множественная регрессия означает, что для прогнозирования зависимой переменной используются несколько независимых переменных. Логическая регрессия использует несколько независимых переменных для прогнозирования бинарной категориальной зависимой переменной (то есть переменной вида да/нет, за/против, покупать/не покупать).

**R-квадрат (R<sup>2</sup>).** Наиболее популярный показатель для оценки степени совпадения рассчитанной регрессии с данными выборки, по которой произведен расчет. R-квадрат отражает также степень изменчивости зависимой переменной по сравнению с рассчитанной линией регрессии. Его значение колеблется в интервале от 0 до 1, и если оно равно, например, 0,52, то это означает, что 52 процента вариации зависимой переменной объясняется независимыми переменными, включенными в уравнение регрессии. В общем случае чем выше значение R<sup>2</sup>, тем более адекватной считается модель.

**Проверка гипотез.** Системный подход к проверке исходного предположения об окружающей реальности. Он включает сопоставление исходной гипотезы или утверждения с доказательствами истинности и на этом основании принятие решения о том, следует ли признать ее истинной или ложной. Гипотезы можно разделить на два вида: нулевая гипотеза и альтернативная гипотеза. Суть нулевой гипотезы (H<sub>0</sub>) состоит в том, что между результатами приведенных наблюдений не существует статистически значимой связи <sup>[41]</sup>.

Альтернативная гипотеза (H<sub>a</sub> или H<sub>1</sub>) исходит из предположения о наличии такой связи. Проверка гипотез включает в себя сравнение эмпирически выявленных закономерностей в выборке с теоретически предполагаемыми (то есть предполагаемыми для случая, если нуль-гипотеза верна). Например, если вы хотите

предсказать качество вина на основе его возраста, то нулевая гипотеза будет звучать следующим образом: «Возраст вина не влияет на его качество», в то время как альтернативная гипотеза такова: «Возраст вина существенно влияет на его качество». Данные собираются и анализируются с целью установления соответствия Н0. Редкие или нестандартные результаты наблюдений (часто определяемые по р-значению ниже определенного уровня) являются показателем того, что Н0 ложная; это означает, что существует статистически значимая вероятность того, что альтернативная гипотеза истинна.

**Р-значение.** В процессе проверки гипотез р-значение показывает вероятность подтверждения данными истинности нулевой гипотезы. Невысокое р-значение указывает на небольшое количество или нестандартный характер данных, подпадающих под нулевую гипотезу, что, в свою очередь, говорит о ее ложности (отсюда можно сделать вывод, что истинна альтернативная гипотеза). При тестировании гипотез мы «отбрасываем нулевую гипотезу», если р-значение меньше, чем уровень значимости  $\alpha$  (альфа греческого алфавита), который обычно равен 0,05 или 0,01. Если нулевая гипотеза отбрасывается, то результат считается статистически значимым.

**Уровень значимости альфа ( $\alpha$ ).** Уровнем значимости называется такое максимальное отношение количества нетипичных выборочных значений (выбросов) ко всему объему выборки, что нулевая гипотеза отклоняется [\[42\]](#).

Иными словами, уровень значимости показывает количество нетипичных наблюдений (выборочных значений), необходимых для признания ложности нулевой гипотезы. Обычно уровень значимости задается как 5 процентов (0,05), но в ситуациях, когда предъявляются особенно строгие требования к доказательству истинности альтернативной гипотезы, этот показатель может быть задан и на более низком уровне, например 1 процент (0,01). Значение  $\alpha$ , равное 5 процентам, означает, что для отбрасывания нулевой гипотезы как ложной достаточно наличия менее 5 процентов нетипичных данных от их общего количества (при условии истинности нулевой гипотезы). На практике это требование часто проверяется путем расчета р-значения. Если р-значение меньше, чем  $\alpha$ , то нулевая гипотеза признается ложной, а альтернативная гипотеза — истинной.

**Ошибка первого рода, или ошибка  $\alpha$ .** Эта ошибка возникает, когда нулевая гипотеза истинна, но тем не менее отбрасывается. В традиционной проверке гипотез нулевая гипотеза отбрасывается в том случае, если р-значение меньше, чем  $\alpha$ . Таким образом, вероятность ошибочного отбрасывания нулевой гипотезы как ложной равняется  $\alpha$ , почему эта ошибка и называется ошибкой  $\alpha$ .

**Тест (статистический критерий)  $\chi$ -квадрат.** Статистический тест, отражающий соответствие данных выборки определенному типу распределения. Измерение этого критерия обычно показывает расхождение между фактическим распределением событий и ожидаемым исходя из некоего заданного распределения. Наиболее часто используется для проверки соответствия фактического распределения заданному.

**t-тест, или t-критерий Стьюдента.** Метод статистической проверки гипотез путем проверки равенства средних значений двух выборок или проверки равенства среднего значения одной выборки некоторому заданному значению.

Во вставке «Основные статистические концепции и аналитические

приемы» мы описали наиболее часто встречающиеся индуктивные статистические модели (мы уже говорили, что описательные и ориентированные на отчеты модели полезны, но не слишком интересны с точки зрения количественного анализа). Конечно, написано множество книг на эту тему, поэтому мы сделаем только краткий обзор.

## Изменение модели

Нетрудно понять, что ни одну модель нельзя использовать неограниченно долго. Если мир в своих основных проявлениях изменился, то очень вероятно, что и модель больше не является его адекватным отражением. Мы уже говорили о том, насколько важны исходные допущения в моделях, а также о том, что проверять их нужно так, чтобы все заинтересованные лица знали, можно ли еще их применить (более подробно об этом поговорим в следующих главах). Достаточно сказать, что любая организация или частное лицо, использующие количественные модели, должны их регулярно пересматривать, чтобы убедиться, что они по-прежнему имеют экономический смысл и соответствуют данным. Если же это не так, то их следует модифицировать. Под словом «регулярно» мы имеем в виду ежегодно, если только нет причин делать это чаще.

В некоторых случаях модели следует пересматривать с еще меньшей периодичностью. Например, если на основании модели вы определяете стратегию торговли ценными бумагами, то придется пересматривать их очень часто. Владелец компании Renaissance Technologies Джеймс Симонс управляет одним из крупнейших в мире хеджевых фондов и занимается пересмотром моделей постоянно. Он приглашает на работу профессоров, хакеров, интересующихся статистикой инженеров и ученых. С момента основания в марте 1988 года материнская компания Симонса Medallion Fund, располагающая капиталом в 3,3 миллиарда долларов и продававшая все, начиная с фьючерсов на соевые бобы и до французских государственных облигаций, обеспечила ежегодную доходность в размере 35,6 процента. За полных одиннадцать лет, до декабря 1999 года, кумулятивная доходность Medallion Fund достигла ошеломляющей

величины в 2478,6 процента. В 2008 году Симонс получил рекордную прибыль в сумме 2,5 миллиарда долларов, а чистая стоимость его компании достигла 8,7 миллиарда. Журнал Forbes поставил Симонса на восьмидесятое место в списке богатейших людей планеты и на двадцать девятое место в списке богатейших людей США. В 2006 году Financial Times назвала его самым умным миллиардером планеты»<sup>[43]</sup>.

Симонс понимал, что выгодные возможности по своей природе невелики и непостоянны. На одном из семинаров он так высказался по этому поводу: «Эффективная теория рынка права в том, что в глобальном смысле рынок действительно эффективен. Тем не менее мы видим незначительные и краткосрочные аномалии. Мы делаем прогноз. Вскоре после этого мы еще раз оцениваем ситуацию и пересматриваем прогноз, а также инвестиционный портфель. Мы тратим на это целый день. Мы всегда считаем и пересчитываем, считаем и пересчитываем. Именно благодаря нашей активности мы и зарабатываем деньги». Чтобы сохранять позиции, Симонс еженедельно меняет свои модели.

Мир вокруг меняется, и именно способность приспосабливаться к этим изменениям сделала Симонса столь успешным бизнесменом. Он говорит: «Временной горизонт статистических прогнозов охватывает несколько лет — может быть, пять или десять. Вам приходится постоянно внедрять что-то новое, потому что рынок играет против нас. Если вы не совершенствуетесь, значит, вы становитесь хуже».

## Пример аналитического мышления: модель ценообразования опционов Блэка и Шоулза

Фишер Блэк и Майрон Шоулз решили проблему ценообразования ценных бумаг<sup>[44]</sup>, долгое время доставлявшую неудобства инвесторам. Блэк получил степень PhD по прикладной математике в Гарвардском университете, затем работал в консалтинговой фирме Arthur D. Little, Inc. Получив степень по экономике в Чикагском университете, Шоулз недавно приступил к работе на кафедре финансов в MIT.

Терминология по ценообразованию опционов в значительной степени специализированная. *Опцион* — это ценная бумага, дающая право, но не обязывающая купить или продать определенный вид

активов на установленных условиях в течение указанного времени. Цена, уплачиваемая за актив в момент исполнения опциона, называется *ценой исполнения*, или *страйк-ценой*. Последний день, когда возможно исполнение опциона, называется *сроком погашения*. Простейший вид опциона, часто называемый *колл-опционом*, представляет собой право на покупку обычных акций компании. *Премия за риск* — это сумма, уплачиваемая инвестором за акции или другие виды активов сверх цены аналогичных безрисковых активов.

В целом чем выше цена акций, тем больше будет цена опциона. Если цена акций намного превышает цену исполнения опциона, то опцион наверняка будет исполнен. С другой стороны, если цена акций намного ниже цены исполнения опциона, владелец вряд ли будет его исполнять, и тогда его цена стремится к нулю. Если срок погашения опциона очень отдален во времени, то цена опциона приблизительно равна цене акций на текущий момент. Обычно цена опциона падает по мере приближения срока его погашения даже при том условии, что цена самих акций может и не изменяться. Но размер премии за риск предугадать трудно.

**Определение и формулирование проблемы.** Необходимое условие эффективного управления рисками, связанными с опционами и другими деривативами, это корректное установление цены на них. Предыдущие попытки разработать эффективную модель ценообразования на деривативы по целому ряду причин оказались неудачными. Возник вопрос о поиске нового метода — научно обоснованного и подкрепленного фактическими данными.

**Изучение предыдущих поисков решения.** Ценообразование на деривативы имеет долгую историю, начиная с 1900 года. В большинстве случаев речь шла об установлении цены на так называемые *варранты* (*колл-опционы*, выпускаемые компаниями и предоставляющие владельцу право выкупить у компании акции по определенной цене), причем методики расчета цены базировались на аналогичных формулах. Эти формулы, как правило, включали в себя один или более произвольно выбранный параметр, вследствие чего отличались неполнотой и страдали одним и тем же глубоким

недостатком: отсутствием объективной методики расчета премии за риск. К сожалению, модели ценообразования на ценные бумаги в условиях равновесия рынка, которая была бы основана на адекватной методике расчета премии за риск, просто не существовало. Блэк и Шоулз впервые в истории попытались вывести формулу цены опциона исходя из условия равновесия рынка.

**Моделирование (выбор переменных).** Было установлено, что на цену опциона влияют пять переменных, в том числе:

- срок погашения
- спот-цена соответствующего актива (цена, по которой в данное время и в данном месте продаются реальный товар или ценные бумаги на условиях немедленной поставки)
- цена исполнения опциона
- ставка процента по безрисковым ценным бумагам
- волатильность доходности соответствующего актива (показатель, характеризующий изменчивость цены).

Отметим, что среди переменных отсутствовало отношение инвесторов к риску. Блэк и Шоулз внесли существенный вклад в развитие темы, по сути дела, показав, что нет необходимости учитывать премию за риск при установлении цены на опцион. Это не значит, что премия за риск вообще отсутствует, но ее величина уже учтена в текущей цене акций.

**Сбор данных (измерения).** Модель Блэка и Шоулза основана на некоторых технических допущениях и признании взаимосвязей между переменными. На этапе разработки модели никаких измерений не проводилось. Однако Блэк и Шоулз провели эмпирические тесты своей теоретической модели на большом массиве данных о колл-опционах и опубликовали результаты в статье *The Pricing of Options and Corporate Liabilities*<sup>[45]</sup>.

**Анализ данных.** Блэк и Шоулз вывели дифференциальное

уравнение с частными производными на основе некоторых технических допущений и теоретических предположений (с использованием методов дифференциального исчисления, а не статистики). Решением этого уравнения и стала формула Блэка и Шоулза, показывающая, каким образом можно рассчитать цену колл-опциона как функцию ставки процента по безрисковым ценным бумагам, вариации цен на базовый актив и параметров опциона (страйк-цены, срока погашения и рыночной цены базового актива). Формула основана на том предположении, что чем выше текущая цена акций и ее волатильность, а также ставка процента по безрисковым ценным бумагам и чем дольше период до погашения опциона, тем выше будет его цена. Аналогично этому рассчитывается цена и других деривативов.

**Результаты и необходимые меры.** Блэк и Шоулз пытались опубликовать результаты своих исследований, отправив их сначала в *Journal of Political Economy*, но редакция отклонила статью. Будучи уверенными в ценности своих изысканий, они послали работу в журнал *Review of Economics and Statistics*, где ее постигла та же участь. Большинству экспертов мысль о том, что можно математически рассчитать цену опциона, не учитывая при этом отношение инвесторов к риску, казалась неприемлемой и слишком неординарной. Изучив развернутые высказывания нескольких знаменитых экономистов по этому поводу, Блэк и Шоулз опять отправили статью в *Journal of Political Economy*, и на этот раз там ее приняли. Через некоторое время профессор MIT Роберт Мертон опубликовал статью, развивавшую некоторые математические аспекты модели Блэка и Шоулза.

Несмотря на проблемы с публикацией, основные выводы статьи получили широкое распространение во всем мире среди тысяч трейдеров и инвесторов, применявших их для рутинных расчетов цены опционов. Модель проста в расчетах и подробно раскрывает взаимосвязи между всеми входящими в нее переменными. Она обеспечивает полезную аппроксимацию, особенно при анализе направленности движения цен на опционы в критических точках. Даже если результаты нельзя считать абсолютно точными, их можно использовать в качестве первого приближения, а затем уточнить.

Модель Блэка и Шоулза стала незаменимой не только при прогнозировании цен на опционы, но и при решении многих других проблем экономики. Ее можно назвать самой успешной экономической концепцией во всей экономической теории. Мертон и Шоулз в 1997 году получили Нобелевскую премию по экономике за развитие новых методов определения цены деривативов. Хотя умерший в 1995 году Блэк не смог стать нобелевским лауреатом, но его заслуги были специально отмечены Академией наук Швеции.

## Пример аналитического мышления: подозрительный муж

В 1973 году в разделе «Советы читателям» газеты Dear Abby появилась такая заметка<sup>[46]</sup>:

Dear Abby, в вашей колонке написано, что женщина вынашивает ребенка 266 дней. Кто вам это сказал? Я вынашивала своего ребенка десять месяцев и пять дней; в этом не может быть сомнений, поскольку я точно знаю, когда он был зачат. Мой муж — флотский офицер, и ребенок не мог быть зачат в другой день, поскольку я видела своего мужа всего лишь в течение часа и в следующий раз мы встретились уже после рождения ребенка. Я не пью и не гуляю с мужчинами, поэтому отцом ребенка может быть только мой муж. Пожалуйста, напечатайте опровержение этой заметки насчет 266 дней, иначе у меня будут большие неприятности.

Читательница из Сан-Диего

В ответной заметке газета постаралась ободрить читательницу, но о сроках беременности было написано немного.

Дорогая читательница! Средний период беременности действительно составляет 266 дней. В некоторых случаях дети рождаются недоношенными, а в некоторых — переношенными. В вашем случае ребенок родился переношенным.

Если бы газета уделила больше внимания количественной стороне вопроса, то в ответной заметке содержалось бы больше чисел. Последние всегда более убедительны, а в данном случае речь идет об относительно простой проблеме, связанной с теорией вероятности. Рассмотрим ее в рамках стандартного шестишагового подхода к проблеме количественного анализа.

**Формулирование проблемы.** В данном случае вопрос не в том, что ребенок родился переношенным, это и так понятно. Десять месяцев и

пять дней — это примерно 310 дней, что существенно больше среднего срока беременности в 266 дней, о котором упоминала газета. Вопрос в том, насколько нетипичен этот случай (или какова его вероятность). Достаточно ли он нетипичен, чтобы заподозрить женщину во лжи?

**Изучение предыдущих поисков решения.** Мы можем с уверенностью предположить, что распределение продолжительности беременности является нормальным (то есть график распределения напоминает колокол). Вероятность того, что беременность будет продолжаться 310 дней, легко рассчитать с помощью *Z-критерия* (количество стандартных отклонений от среднего значения) для нормального распределения, что является азбукой статистических расчетов.

**Моделирование (выбор переменных).** Вероятность того, что беременность может длиться по крайней мере 310 дней.

**Сбор данных (измерения).** Имеющиеся данные позволяют сделать вывод о том, что среднее значение продолжительности беременности составляет 266 дней со стандартным отклонением 16 дней.

**Анализ данных.** Если средняя продолжительность беременности составляет 266 дней со стандартным отклонением 16 дней, то вероятность ее продолжительности в 10 месяцев и пять дней (300 и более дней) составляет 0,003 (если принять, что распределение нормальное).

**Результаты и необходимые меры.** Это значит, что три ребенка из тысячи рождаются более чем через 300 дней после зачатия. Казалось бы, вероятность очень невелика, но только не в случае больших чисел. В Америке ежегодно рождается около четырех миллионов детей. Соответственно, около двенадцати тысяч из них рождаются с таким большим опозданием. Видимо, Dear Abby стоило написать нечто вроде следующего: «Каждый год в США со столь большим запозданием рождаются примерно двенадцать тысяч детей, и одним из них стал ваш

ребенок». Это успокоило бы не только читательницу, но и ее мужа.

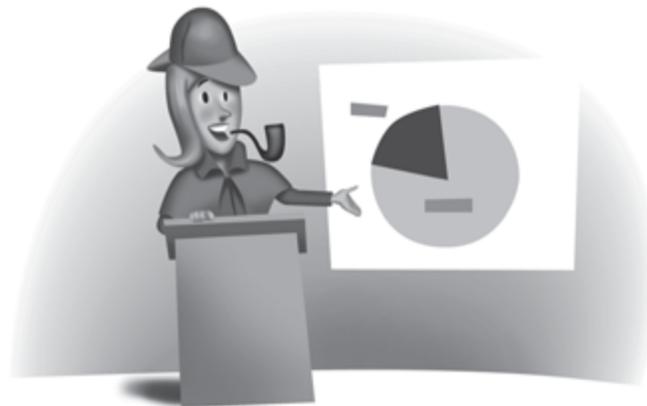
В статистическом тестировании гипотез рассчитанное выше значение вероятности 0,003 называется *p-значением*, что равно вероятности получения данного значения критерия (в данном случае *Z-значения*, равного 2,75) в предположении, что нулевая гипотеза истинна. В данном случае нулевая гипотеза ( $H_0$ ) звучит следующим образом: «Отцом ребенка является муж». В стандартной методике проверки гипотез нулевая гипотеза отбрасывается как ложная, если *p-значение* меньше уровня значимости. В данном случае *p-значение* равно 0,003, а это значит, что нулевая гипотеза будет отброшена, даже если уровень значимости составит 1 процент. Вообще говоря, мы должны были бы отбросить гипотезу об отцовстве мужа читательницы. Как можно объяснить этот ошибочный результат проверки гипотезы? Это типичный пример ошибки первого вида (или ошибки альфа), когда отклоняется нулевая гипотеза ( $H_0$ ) при ее истинности. Этот пример показывает, что жизнь может не укладываться в рамки теории вероятности.

## ГЛАВА 4

# Результаты и необходимые меры

Оформление результатов анализа и представление их всем, кто в них заинтересован, это последний этап в нашей модели количественного анализа из трех этапов и шести шагов. Он крайне важен. Если вы блестяще выполнили предыдущие этапы, но провалили этот, то ничего хорошего в итоге не выйдет. Аналитики, не желающие трудиться впустую, оставаться в стороне от принимаемых решений и предпринимаемых действий, очень серьезно относятся к этому этапу, не жалея сил и времени для его успешного завершения. Аналитики, которых не заботят такие вещи (на наш взгляд, это плохие аналитики!), думают, будто результаты говорят сами за себя, и не уделяют большого внимания этому этапу.

### 6. Результаты и необходимые меры



Сложность еще и в том, что вопросы оформления и обнародования результатов не рассматриваются в курсе обучения будущих аналитиков. В результате даже те из них, кто искренне увлечен исследованиями и учебой в своей области, традиционно сосредоточены на аналитических методах и не слишком заботятся о том, как лучше преподнести результат. К счастью, такое положение дел постепенно меняется. Вот как видит задачу воспитания специалистов, которые могут эффективно представить результаты статистических

исследований, заведующий кафедрой статистики Гарвардского университета Сяо Лименг (недавно назначенный деканом Гарвардской школы науки и искусств):

В последние годы возобладал более широкий взгляд на подготовку статистиков в Гарвардском университете. Упор нужно перенести с подготовки нескольких блестящих студентов к получению степени PhD по количественному анализу на обучение всех первокурсников основам статистических исследований как неотъемлемой составляющей критического образа мышления в гуманитарных науках. Как ни странно, мысль о том, что можно стать экспертом по вину, не имея представления о том, как это вино производится, позволила нам подготовить множество будущих виноделов вместо того чтобы пытаться самостоятельно вырастить урожай винограда<sup>[47]</sup>.

Исходя из этой мысли, Менг и его коллеги разработали для студентов младших курсов учебный цикл под названием «Статистика в реальной жизни: ваш шанс на счастье (или несчастье)». В нем через призму статистики рассматриваются темы «Романы», «Вино и шоколад», «Финансы», «Медицина» (включая статистические аспекты клинических испытаний виагры), «Фондовый рынок» и т. п. Менг старается сделать статистику «не просто вкусным блюдом, а деликатесом»<sup>[48]</sup>.

Материал этой главы пригодится и аналитикам, и адресатам их труда (или, в соответствии с аналогией Менга, и виноделам, и ценителям вина). Аналитики смогут сделать результаты своих исследований более интересными и привлекающими внимание, что позволит работать эффективнее. Те же, кто пользуется результатами их исследований, например менеджеры, давшие добро на аналитический проект, часто настаивают на том, чтобы эти результаты были представлены в интересной и современной форме. Если аналитики заставляют слушателей скучать или путаться в специальной терминологии, то, наверно, это их вина. Заказчики должны вместе с аналитиками работать над тем, чтобы сделать результаты анализа понятными и легкими в использовании. И конечно, именно они обычно принимают решения и принимают меры, основываясь на результатах анализа.

Суть этого этапа состоит в том, чтобы описать проблему и пути ее решения, разработанную модель, необходимые данные и взаимосвязи между переменными. Если эти связи выявлены, то их нужно

интерпретировать применительно к поставленной проблеме. Чем яснее представлены результаты, тем больше вероятность того, что они приведут к обоснованным решениям и действиям — в конечном счете именно это первоочередная цель аналитического проекта.

Сообщая результаты, обязательно опишите процесс исследования, выводы, дайте рекомендации по поводу решения исходной проблемы, хотя, возможно, и не в таком порядке. Как правило, лучше начинать с описания результатов и рекомендаций. Существует два отличных способа демонстрации результатов: можно организовать совещание, пригласив всех заинтересованных и проведя для них сессию вопросов и ответов, а можно подготовить формальный отчет о результатах проекта. Если исходная проблема и результаты анализа имеют некоторую академическую ценность, можно написать статью и опубликовать ее в журнале соответствующего профиля.

Как мы уже говорили в [главе 2](#), оформление данных в виде черно-белых таблиц — это прекрасный способ заставить людей проигнорировать их, даже если это результаты совсем простого описательного анализа. Результаты легко представить в простой графической форме (гистограммы, диаграммы, графики), а можно выбрать более затейливый вариант: например, интерактивный дисплей. Кое-кто предпочитает простую табличную форму более наглядным графическим способам представления информации, но таких людей не так уж много. Если можно использовать цвет и анимацию, чтобы оживить презентацию и сделать ее более доходчивой, — что ж, тем лучше!

## О чем говорят данные

Наиболее успешными аналитиками становятся те, кто способен понять, о чем рассказывают данные. В [главе 2](#) мы рассматривали разные виды аналитических проектов. Но независимо от их вида и процедур, с помощью которых они выполняются, основные составляющие всегда одинаковы. Это логичный сюжет, в основе которого лежит бизнес-проблема или цель компании. Например, рассказ о результатах аналитического проекта, посвященного исследованию лояльности потребителей, должен начинаться так: «Как

вам известно, в течение долгого времени мы собирались выяснить, кто же наши самые лояльные потребители и как сделать их еще более лояльными. Сейчас мы можем ответить на эти вопросы».

Хорошие истории надо рассказывать понятным для слушателей языком. Если ваша аудитория состоит из аналитиков или высококвалифицированных технических специалистов, тогда вполне допустимо употреблять статистические и математические термины и даже некоторые формулы. Но значительно чаще слушатели не имеют математической подготовки, поэтому надо выбирать понятную им терминологию, связанную со знакомыми концепциями и задачами. Применительно к бизнесу это понятия «прибыль», «сбережения» и «доходность инвестиций».

Хорошие истории завершаются выводами о том, какие действия следует предпринять, а также описанием их возможных последствий. Конечно, прежде чем обсуждать различные сценарии действий, нужно проконсультироваться с ключевыми пользователями. Никому не хочется слышать от количественного аналитика что-то вроде «Вы должны сделать это и еще вот это».

Руководитель аналитической группы Дэвид Шмитт из глобальной сети отелей IHG (Intercontinental Hotels Group) считает очень важным умение слушать, что говорят данные. О некоторых особенностях действительно эффективного рассказа о результатах аналитических проектов он написал в своем блоге<sup>[49]</sup>:

Итак, что же отличает по-настоящему увлекательную историю? При каждом удобном случае я обращаюсь к экспертам. В наше время лучшие рассказчики работают на студии мультипликационных фильмов Pixar. Именно они поведали нам такие замечательные истории, как «В поисках Немо», «Суперсемейка» и, конечно, «История игрушек». Художник-мультипликатор Эмма Коутс опубликовала в Твиттере список из двадцати двух правил успешного показа истории. Не все из них можно применить к аналитике, но мы выбрали три наиболее близкие к нашей теме.

«Надо знать, чем закончится история, даже если не дошел еще и до середины. Это правда. Закончить историю труднее всего, и развязка требует больше всего сил и времени». Результаты анализа — это единственная причина, по которой вы беретесь рассказывать историю. Как вы считаете, что должны знать зрители, выслушав ваш рассказ? Еще более важно: что они должны предпринять? Пусть эти вопросы станут своеобразным тестом для каждой части вашей истории, и помните о том, чем вы собираетесь ее завершить.

«Изложение истории на бумаге помогает сохранить ее». Если история хранится лишь в вашей голове, она может быть прекрасной идеей, но никто больше о ней не узнает.

Креативная история зарождается в вашей голове, но будучи изложена на бумаге, способна повести вас в такие места, о которых вы даже не догадывались. Я могу почти двинуться умом, так и сяк крутя историю в голове, но стоит заставить себя записать ее (не имеет значения: на бумаге, в Word или в PowerPoint), и творческие идеи появляются одна за другой.

«В чем суть вашей истории? Сделайте короткую выжимку из нее. Если вы это знаете, остальное приложится». Вы должны уметь уложить вашу историю в три-пять предложений. Если сделать это удачно, то разработка деталей становится гораздо более простой задачей.

Очень полезно определить структуру презентации с теми, для кого предназначены результаты анализа. Это поможет четко представить, чего они ожидают от аналитика и чего он ожидает от людей, принимающих решение. Например, Джордж Румелиотис возглавляет группу обработки данных в компании Intuit. Он и его сотрудники анализируют и проектируют потребительские свойства продуктов, основываясь на большом объеме интернет-информации, собираемой компанией. Аналитические проекты, выполняемые его группой, предназначены для внутренних пользователей, и каждый раз он разрабатывает методологию исследования и демонстрации результатов. Большинство ее этапов ориентировано на бизнес-проблемы компании.

1. Мое понимание бизнес-проблемы.
2. Как я собираюсь оценивать ее влияние на бизнес.
3. Какие данные можно собрать.
4. Первоначальная гипотеза относительно решения.
5. Решение.
6. Влияние решения на бизнес.

Используя эту методологию, специалисты по базам данных могут создавать вики-сайты<sup>[50]</sup>, где отражаются результаты всех шагов и этапов. Те, кому необходимы их результаты, имеют возможность просматривать вики-сайт и комментировать действия специалистов. Румелиотис говорит, что даже притом что это сайт для ознакомления сотрудников с результатами исследований, все равно он побуждает аналитиков и пользователей обмениваться информацией напрямую.

## О чем лучше умолчать

Аналитики хорошо знакомы со статистической терминологией (описание статистических методов, определение коэффициентов регрессии, расчет значения  $R^2$  и т. п.) и зачастую полагают, что их слушатели разбираются во всем этом не хуже. Однако это большая ошибка. Большинство не поймет, о чем идет речь в насыщенном специальной терминологией докладе или презентации. Как сказал один из аналитиков компании ING, «никому не интересны ваши  $R$ -квадраты».

Часто аналитики стараются изложить результаты в виде последовательности процедур и операций, которые они проделали для их получения: «Сначала мы исключили выбросы из массива данных, затем сделали логарифмическое преобразование. Выявился высокий уровень автокорреляции, поэтому мы ввели переменную с лагом в один год», — вы уже представили себе, как это выглядит? Повторим еще раз: аудиторию, как правило, не интересует, как вы получили те или иные результаты, поскольку им важны только сами результаты. Полезно привести подобную информацию в приложении к докладу или презентации, но не стоит включать ее в ваш рассказ. Лучше начните непосредственно с того, что хотят знать ваши слушатели.

## Примеры оформления результатов

То, как именно вы оформите и преподнесете результаты количественного анализа, может и убедить слушателей, и полностью подорвать доверие к вашим словам. Это давно известный факт. Рассмотрим несколько примеров, иллюстрирующих обе ситуации.

### Флоренс Найтингейл: удачная форма результатов

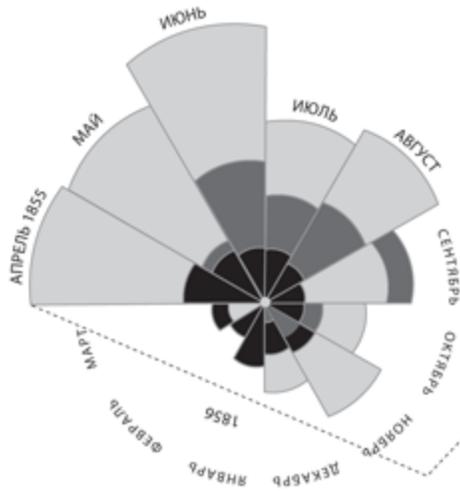
Флоренс Найтингейл широко известна как родоначальница профессии медсестры и реформатор в области санитарии и методов ухода за ранеными в госпиталях. Однако помимо этого она еще известна тем, что одна из первых применила количественные методы. Когда в октябре 1854 года, в период Крымской войны, ее и еще 38 медсестер направили в Британский военный госпиталь в Турции, она ужаснулась сложившейся там ситуацией с лечением раненых. Большинство

случаев летального исхода в госпиталях было вызвано эпидемиями, эндемическими заболеваниями и инфекционными болезнями, а вовсе не ранами, полученными в бою. В феврале 1855 года смертность в госпитале составила 43 процента. Найтингейл считала, что ситуацию нужно срочно исправлять и что для этой цели может пригодиться статистика. Она организовала сбор и обработку данных, ведение подробных ежедневных записей о назначениях врачей, характере ранений, заболеваемости, лечении и причинах смертей.

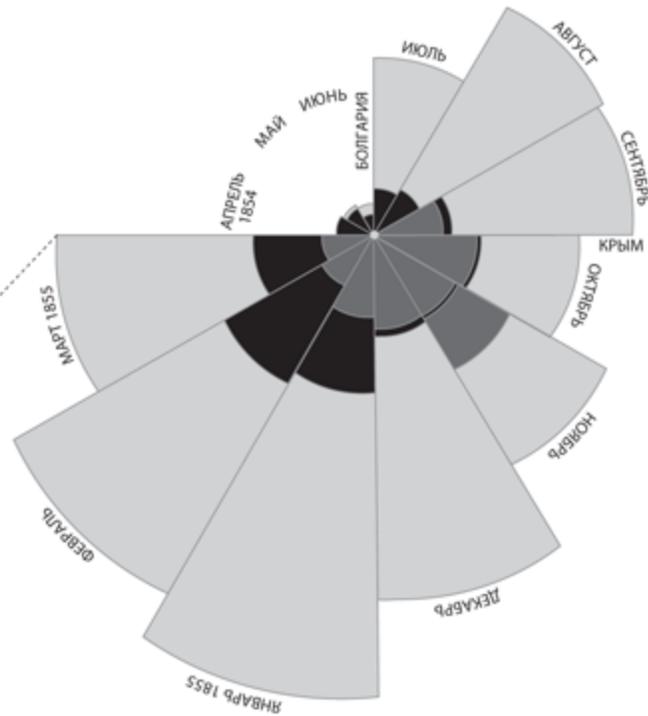
Однако самым крупным нововведением Найтингейл стали методы оформления результатов. С раннего возраста она интересовалась цифрами и любила таблицы. Она прекрасно понимала важность подкрепленных цифрами аргументов, но при этом отдавала себе отчет в том, что далеко не все разделяют ее увлечение таблицами (особенно если учесть, что в то время они не были столь популярны, как теперь!). Обычный человек, скорее всего, просто не будет читать отчет, перегруженный таблицами, а значит, ее аргументы окажутся бесполезными. Поскольку она хотела во что бы то ни стало донести до читателей свои статистические выкладки, то разработала линейку диаграмм, наглядно показывающих, как антисанитарные условия становятся причиной неоправданных смертей в госпитале. Они лучше всяких таблиц свидетельствовали о необходимости реформ (рис. 4.1).

**Рис. 4.1. Диаграммы «Причины смертности раненых в Восточной армии»  
Флоренс Найтингейл**

2. Апрель 1855 г. к марту 1856 г.



1. Апрель 1854 г. к марту 1855 г.



Светло-серые, темно-серые и черные секторы имеют общую вершину, находящуюся в центре диаграммы.

Светло-серые секторы, площадь которых рассчитывается от вершины в центре диаграммы, представляют собой уровень смертности от болезней, поддающихся профилактике, или инфекционных болезней, течение которых можно облегчить; темно-серые секторы характеризуют уровень смертности от ранений, а черные — от прочих причин.

Черная линия, пересекающая темно-серый сектор в ноябре 1854 г., обозначает границу смертности от прочих причин в течение месяца.

В октябре 1854 г. и апреле 1855 г. черный сектор совпадает с темно-серым, в январе и феврале 1856 г. светло-серый сектор совпадает с черным.

Площадь секторов можно сравнить, наложив их друг на друга по ограничивающим серым линиям.

Сейчас это считается само собой разумеющимся, но во времена Флоренс это был относительно новый способ демонстрации данных. Ее диаграммы были разновидностью круговых диаграмм, состоявших из нескольких радиальных секторов. Найтингейл сделала их цветными, чтобы наглядно показать динамику смертности от тех или иных причин по месяцам. Эти показатели, равно как и диаграммы, оказались весьма убедительными.

Найтингейл периодически докладывала в Лондон о результатах своих изысканий и неуклонно настаивала на проведении реформ. Она широко использовала свои новаторские диаграммы, чтобы проиллюстрировать членам парламента реальное состояние службы медицинской помощи в Крыму; последние вряд ли прочли и поняли бы длинные доклады, изобилующие цифровой информацией. Людей шокировали сведения о том, что в госпиталях раненых не столько лечили, сколько доводили до смерти. В итоге уровень смертности начал постепенно снижаться, о чем говорят систематически собиравшиеся Найтингейл данные. После завершения Крымской войны она вернулась в Англию в июне 1856 года и оказалась в центре внимания — публика считала ее героиней.

Найтингейл первой из женщин стала членом Королевского статистического общества в 1859 году и почетным членом Американской ассоциации статистиков в 1874-м. Знаменитый статистик и основатель первой в мире кафедры статистики Карл Пирсон назвал Найтингейл «предвестницей» будущего развития прикладной статистики<sup>[51]</sup>.

## Грегор Мендель: неудачный пример преподнесения результатов

Гораздо менее удачный пример преподнесения итогов анализа (при этом напоминающий о важности этого этапа) можно найти в биографии Грегора Менделя<sup>[52]</sup>. Родоначальник учения о наследственности Мендель за несколько месяцев до смерти в 1884 году сказал: «Научные исследования приносили мне огромную радость и удовлетворение, и я совершенно уверен в том, что в скором времени весь мир признает результаты моего труда». Мир действительно признал работы Менделя, но на это потребовалось несколько десятилетий. Возможно, если бы великий генетик лучше умел рассказывать о результатах своих исследований, он добился бы признания намного раньше: возможно, даже и при жизни.

Монах Мендель стал одним из первопроходцев в генетике. На тот момент в биологии господствовала теория о том, что наследственные признаки становятся все менее выраженными у каждого

последующего поколения. Исследование Менделя, посвященное наследованию тех или иных отличительных признаков, показало, что этот процесс подчиняется определенным законам (позже получившим его имя). Работы Менделя получили заслуженное признание лишь на рубеже XX века, когда те же результаты были получены независимыми исследователями и сформировалась генетика как наука.

Мендель вывел математические принципы наследования отличительных признаков путем прекрасно разработанного тщательного эксперимента, который он проводил на полях монастыря. Для исследований он выбрал обычный горох, поскольку он занимал мало места, был дешев и давал два урожая в год. Мендель выборочно проводил перекрестное опыление растений, обладавших определенными признаками, и наблюдал их проявление у последующих поколений.

В период 1856–1863 годов исследователь терпеливо проводил опыты по перекрестному опылению, закрывая опыленные растения колпачками, чтобы предотвратить случайное опыление насекомыми. После восьми лет упорного труда Мендель вырастил 12 980 кустов гороха от исходных 225 кустов. Он собрал обширную базу данных об их наследственных признаках, чтобы выявить закономерность их наследования. Он обнаружил доминантные и рецессивные наследственные признаки, которые позже были названы генами.

К сожалению, обнародование результатов исследований оказалось далеко не столь эффективным, как сами исследования. Мендель опубликовал результаты своих экспериментов в малоизвестном моравском научном журнале. Экземпляры журналов были разосланы 130 европейским и американским научным обществам и авторитетным исследователям. Тем не менее большого влияния на научную среду того времени это не оказало, и работа Менделя цитировалась всего лишь три раза в последующие 35 лет. Результаты сложного и скрупулезного труда Менделя не были поняты даже современными ему признанными авторитетами в этой области науки. Если бы Мендель был не монахом, а ученым, то, возможно, добился бы больших успехов в продвижении своего проекта и опубликовал бы работу за рубежом. Он пытался установить контакты с зарубежными учеными, отправил свою статью Дарвину и некоторым другим

научным светилам. История сохранила имена примерно дюжины из них. Говорят, что Дарвин даже не разрезал страницы в присланном журнале, чтобы прочесть статью Менделя.

Незадолго до смерти Мендель говорил: «Мое время еще придет», — но трудно сказать с уверенностью, верил ли он сам в эти слова. Он умер, так и не узнав, что его открытия изменят последующую историю человечества.

Хотя работа Менделя была столь же блестящей, сколь и беспрецедентной, ученым-биологам потребовалось более тридцати лет, чтобы оценить ее значимость. Лишь в начале XX века его идеи получили распространение, и биологи признали его приоритет в изучении наследственности. Какой отсюда следует вывод? Если вы не хотите, чтобы ваши выдающиеся исследования постигло столь долгое забвение, вы должны уделить достаточно внимания их презентации.

## Современные способы представить результаты

В наше время существует множество способов оформления и обнародования результатов анализа, неизвестных во времена Найтингейл и Менделя. Если вы ограничиваетесь распечаткой или выводением на монитор таблиц и графиков, то вы явно не используете весь арсенал средств, имеющихся в вашем распоряжении. Конечно, выбор подходящего инструментария зависит от обстоятельств, особенностей аудитории, и вряд ли вы станете использовать изображения сексапильных девушек в презентации только потому, что они сексапильны.

Тем не менее визуальная аналитика (известная также как *визуализация данных*) в последние годы развивается очень быстрыми темпами. Если вы используете только секторальные диаграммы или пиктограммы, вы, видимо, лишь поверхностно знакомы с возможностями визуализации данных. Во вставке «Цели и виды визуализации», возможно, описаны не все возможные варианты ее использования, но по крайней мере это наиболее полный список способов визуализации из тех, что мы знаем<sup>[53]</sup>. Выбор наиболее подходящего вида графиков или диаграмм для решения конкретной задачи может оказаться делом трудным, но прикладное программное

обеспечение для визуальной аналитики может отчасти взять решение этой задачи на себя с учетом выбора переменных для анализа. Например, программа SAS Visual Analytics уже это делает: в нее вмонтирована функция под названием Autochart. Если данные содержат, например, «Одну категорию даты/времени и любое количество других категорий и показателей», программа автоматически генерирует гистограмму<sup>[54]</sup>.

## Цели и виды визуализации

### ЕСЛИ ВЫ ХОТИТЕ

#### **Увидеть взаимосвязи между координатами опорных точек:**

- точечная диаграмма показывает взаимосвязь между двумя переменными в двумерной системе координат;
- матричная диаграмма показывает взаимосвязи иерархических переменных;
- карта показывает значения переменных на двумерной карте разными цветами;
- сетевая диаграмма показывает наличие и прочность взаимосвязей между объектами.

#### **Сравнить набор значений или частот (как правило, для одной переменной):**

- столбиковая диаграмма, где высота столбцов отражает значение показателя;
- гистограмма — разновидность столбиковой диаграммы, в которой высота столбцов отражает частоту попадания переменной в определенный интервал;
- пузырьковая диаграмма показывает набор количественных переменных в виде кругов, площади которых соответствуют значениям переменных.

#### **Показать динамику одного показателя относительно другого (обычно времени):**

- линейчатый график показывает график одной переменной или нескольких переменных со стандартизированными значениями на плоскости;
- график с наложением — линейчатый график с прямоугольниками различной длины, указывающими на изменения в значениях переменных; может также указывать на изменения различных категорий данных с помощью различного цвета.

#### **Видеть структуру целого, оценить соотношения различных его частей:**

- круговая диаграмма показывает распределение значений одной переменной в виде секторов круга;
- карта дерева — технология визуализации, при которой значения данных показаны в иерархической подчиненности, например: Весь мир — Континенты — Страны — Население каждой страны.

**Наложить данные на географическую карту:**

- накладывать обобщенных данных на географические карты при помощи цветов, пузырьков или выносов, отражающих значения соответствующих переменных.

**Анализировать частоту использования слов в тексте:**

- облако тегов — визуальное представление частоты употребления тех или иных слов в тексте, когда более часто встречающиеся слова выделяются более крупным шрифтом;
- сеть фраз показывает частоту комбинаций слов, используемых вместе, когда более употребительные словосочетания показаны более крупным шрифтом.

Способы визуализации, перечисленные в этом списке, статичные. Но в последнее время все чаще применяются их динамичные и интерактивные разновидности. Шведский профессор Ханс Рослинг популяризирует их в своем выпуске видеоконференции TED Talk, где методы визуализации применяются для отражения изменения показателей здоровья развитых и развивающихся стран с течением времени<sup>[55]</sup>. Рослинг создал сайт Gapminder ([www.gapminder.org](http://www.gapminder.org)), где приведено множество аналогичных примеров. Скорее всего, популярность использования интерактивной аналитики для отражения динамики показателей в течение определенного периода будет только расти; тем не менее она подходит далеко не для всех типов данных и далеко не всегда нужна.

Иногда можно найти более эффектные способы визуализации данных, чем графика. Например, маркетолог и специалист по стратегиям Винс Барабба, работающий с несколькими крупными компаниями (в том числе с General Motors, Херох и Kodak — проблемы у них возникли не из-за него!), весьма креативно подошел к вопросу о том, как лучше всего презентовать результаты маркетинговых исследований. В частности, он знал, что топ-менеджеры General Motors используют трехмерные модели автомобилей для презентаций. В какой-то момент, получив особо важные результаты маркетинговых исследований, он решил разработать их трехмерную графическую модель, по которой топ-менеджеры могли совершить виртуальное путешествие. Возможность своими глазами увидеть пики рыночного спроса и виртуально «потрогать» их сделала данные очень

выразительными.

В компании ING работают несколько групп аналитиков. Дэвид Шмитт возглавляет одну из них — в отделе стратегии и планирования. Ее задачи — оперативно информировать топ-менеджеров о том, каковы текущие дела компании. Иначе говоря, группа ориентирована на оформление всевозможных отчетов. Ее сотрудникам важно привлекать как можно больше внимания к полученным результатам и стимулировать к принятию решений на их основе. Для этого в их распоряжении целый ряд инструментов, в зависимости от особенностей аудитории. Один из них — создание «музыкальных клипов»: пятиминутных видеороликов, обосновывающих полученные результаты с помощью картинок, аудиоклипов и видео. За показом такого клипа, как правило, идет устный рассказ с необходимой дополнительной информацией.

Например, не так давно группа Шмитта создала видеоклип с описанием того, какие предпосылки определяют прогнозы покупательского спроса летом. Видео получило название «Путешествие по дорогам лета». В нем автомобиль ехал по дороге мимо дорожных знаков «Внимание! Впереди пик спроса!» и билбордов со статистическими данными<sup>[56]</sup>. Целью создания видеоклипа было привлечь внимание к основным факторам операционной деятельности в приближающемся летнем сезоне и их особенностям в разных регионах страны. Как отмечает Шмитт, «данные не самоцель и цифры не самоцель, главное — это идея, которая за ними стоит». Если основную идею удалось довести до сознания слушателей, то Шмитт может переключиться на более традиционные формы презентации. Но он надеется, что благодаря видеоклипу аудитория уже сформировала мнение о теме доклада.

Еще один способ донести до аудитории результаты анализа и объяснить суть аналитических моделей — это ролевые и имитационные игры. Их часто применяют для иллюстрации взаимодействия переменных в сложных моделях. Например, «Пивная игра» — симуляция, основанная на модели сбыта продукции пивоваренной компании, — была разработана в MIT в 1960-х годах. Тысячи компаний и студентов на ее примере учились организовывать

сбытовую сеть и постигали такие принципы, как «эффект хлыста»: колебание объема заказов по причине неполной информации об участниках сбытовой цепочки. Другие компании приступают к разработке собственных обучающих игр для решения конкретных задач. Занимающаяся грузовыми перевозками компания Schneider National разработала имитационную игру, чтобы убедить персонал в важности аналитического мышления в управлении грузовиками и трейлерами. Цель игры — минимизировать простаивание машин и переменные издержки при заданном объеме выручки. Решения о том, принимать ли груз или допустить холостой пробег грузовика, участники принимают, имея в распоряжении вспомогательную аналитику. Шнейдер использует игру, чтобы сообщить сотрудникам, каковы текущие результаты деятельности, и поменять психологию «исполнителей заказов» на позицию «получателей прибыли». Некоторые клиенты компании Шнейдера тоже участвовали в этой игре.

Компании используют современные технологии, чтобы дать возможность тем, кто принимает решения, непосредственно работать с данными. Например, Deloitte Consulting создала для iPad виртуальную симуляцию для сотрудников аэропорта: выполняемых ими рабочих операций и отчетов. В ней используется программа Google-карты, в которой можно отметить те аэропорты, где может совершить посадку определенный самолет. Разным цветом самолетов показывается неэффективность (красный) или эффективность (зеленый) работы аэропорта. Щелкнув по символу того или иного аэропорта на карте, можно получить финансовые и операционные данные о результатах его работы. Интерфейс содержит пиктограммы, означающие численность персонала, уровень обслуживания пассажиров, финансовые результаты, операционные проблемы и пр. Это приложение — лишь один пример того, чего могут добиться современные интерактивные и удобные технологии.

## Чего не найдешь в отчете

Презентации и доклады, конечно, не единственно возможные способы рассказать о результатах аналитических проектов. Чем глубже

аналитики вовлечены в принятие мер по результатам анализа, тем более успешным будет проект. Например, все чаще компании «встраивают» аналитику в среду автоматизированного принятия решений<sup>[57]</sup>. В страховании, в финансовой сфере, а также в отраслях, где цены на услуги зависят от клиентов (например, в отельном бизнесе и авиаперевозках), автоматизированные процедуры принятия решений на основе аналитики получили широкое распространение. И действительно, каждый знает, как сложно заставить сотрудника страховой компании или банка заняться наконец вашей заявкой на страховку или кредит. В таких условиях аналитика будет распространяться все больше и больше, поскольку у клиента практически нет выбора (или он очень невелик; немногочисленные исключения можно оставить сотрудникам). Если вы количественный аналитик или ответственный за принятие важного решения менеджер и ваша задача — разработать и внедрить такого рода системы, то это будет намного эффективнее, чем просто подготовить отчет.

В отрасли онлайн-поиска информации существуют колоссальные базы данных, измеряемые многими петабайтами. Новая информация поступает в таких объемах и с такой скоростью, что человеческий мозг не способен справиться с ней. В таких условиях специалисты по базам данных (обычно количественные аналитики с навыками в области IT выше средних) часто являются сотрудниками отделов исследований и разработок. Их задача состоит в разработке новых потребительских свойств продуктов и опытных образцов, а отнюдь не в создании бумажных отчетов или презентаций.

Например, группа по работе с базой данных в социальной сети для профессионалов LinkedIn входит в состав отдела по разработке новых продуктов, создавшего целый ряд новых свойств и функций, которые основаны на взаимосвязях между социальными сетями и работой. Это в числе прочего «Люди, которых вы можете знать», «Рабочие места, которые могут вас заинтересовать», «Мы ищем таланты», интерактивная визуальная презентация профессионального сообщества InMaps «Группы, которые могут вас заинтересовать». Некоторые из этих функций (в частности, «Люди, которых вы можете знать») в огромной степени повлияли на рост и стабильность

клиентской базы LinkedIn.

Если вы или количественные аналитики в вашей организации занимаетесь преимущественно внутренними процессами и проблемами, то и в этом случае результатом анализа вовсе не обязательно будет лишь отчет или презентация. Ведь конечная цель в том, чтобы повысить эффективность внутреннего бизнес-процесса или управленческого решения. А это значит, что вы или ваш количественный аналитик должны принимать самое активное участие не только в обосновании, но и практической реализации результатов анализа.

Том провел небольшое исследование того, каким образом несколько лет назад компании усовершенствовали 57 управленческих решений, и выяснил, что именно аналитика чаще всего упоминалась в качестве фактора, позволившего добиться успеха<sup>[58]</sup>. На втором месте по частоте упоминаний оказались «изменения в корпоративной культуре и лидерстве», на третьем — «более качественные данные» и на четвертом «реорганизация бизнес-процессов». В среднем представители опрошенных компаний называли пять и более различных факторов, благодаря которым удалось добиться обоснованных решений. Это означает, что аналитикам приходится быть больше чем аналитиками. Они становятся консультантами по преобразованиям в бизнесе.

## Когда результаты не влекут за собой действий

Мы ожидаем, что по результатам количественного анализа будут приняты какие-то меры, но иногда даже отличные результаты анализа не влекут за собой соответствующих действий (хотя знать их все равно полезно). Например, Джин Хо как-то принимал участие в исследовании преимуществ первого шага в игре го. Это одна из старейших игр-стратегий в мире. В игре простые элементы (черные линии, черные и белые чечевицеобразные фишки, или «камни») и столь же простые правила. Но, несмотря на очевидную простоту, в игре масса тонкостей, которые тысячелетиями захватывают игроков<sup>[59]</sup>.

В Азии, особенно в Корее, Японии и Китае, десятки миллионов

людей увлекаются игрой го. Почти тысяча профессиональных игроков зарабатывают себе на жизнь, участвуя во всевозможных турнирах с призовыми фондами в миллионы долларов.

В начале игры доска пуста. Один игрок выбирает черные камни, а второй — белые. Черные и белые камни ставятся на доску попеременно, но у черных есть право первого хода. Это дает определенное преимущество, но никому прежде не приходило в голову попытаться оценить, насколько оно велико. С 1974 года в профессиональных состязаниях белым в начале игры дается фора в 5,5 очка в качестве компенсации за право первого хода черных. Эта фора называется *коми*.

Чтобы понять, насколько компенсация адекватна, Джин Хо проанализировал данные о 577 профессиональных матчах, проводившихся в Корее. Результаты анализа позволяют предположить, что в действительности этой компенсации мало. Однако в процессе анализа возникла одна проблема. Данные позволяли утверждать, что право первого хода становится преимуществом только для опытного игрока, способного им воспользоваться. Теоретически на этом основании было бы вполне оправданно диверсифицировать размер коми, то есть для тех игроков, у которых фора, как правило, не становилась решающим фактором выигрыша, оставить ее размер 5,5, а для всех остальных увеличить. Но в профессиональных турнирах на это вряд ли пойдут. Таким образом, оригинальный анализ (позволивший Джин Хо и его коллеге опубликовать прекрасную статью в академическом журнале) не привел ни к каким конкретным действиям<sup>[60]</sup>.

## Подведем итоги

Итак, мы описали каждый из шести шагов и трех этапов модели количественного анализа. Если вам удастся на практике осуществить все действия, описанные в главах [2](#), [3](#) и [4](#), то это означает, что вам по плечу практически любая аналитическая проблема. Даже если вы не слишком хорошо разбираетесь в количественных методах, вы вполне сможете провести большинство рекомендуемых процедур. По сути, продвинутый математический или статистический анализ ведется

лишь на одном-двух шагах анализа. К тому же большую часть тяжелой работы за вас сделают компьютеры.

Ключ к успеху аналитического проекта лежит в продуманности его первого и последнего шага. Если проблема определена и сформулирована правильно, то промежуточные шаги, как правило, реализуются без особых проблем. Если вам не удастся эффективно довести до общего сведения результаты, то и действий по итогам анализа никаких не последует. Вот почему последний этап имеет такое большое значение.

### **Пример аналитического мышления: практический способ предсказать развод**

Молодожены пришли к блестящему математику, славившемуся умением предсказывать, узнать, ждет ли молодоженов долгая и счастливая семейная жизнь или они в скором времени разведутся. Их попросили сесть друг напротив друга в пустой комнате и поговорить на какую-нибудь острую тему пятнадцать минут. Проанализировав запись этой беседы, математик предсказал: «Вы, ребята, с тем же успехом могли бы развестись прямо сейчас». Было прекрасно известно, что его прогнозы на удивление точны. Математиком, разработавшим модель прогнозирования будущего новобрачных, был профессор Оксфордского университета Джеймс Мюррей. Он проводил это исследование совместно с профессором и психологом Вашингтонского университета Джоном Готтманом. Тот разрабатывал гипотезу и собирал данные: записанные на видео и закодированные наблюдения за поведением многих семейных пар. Кроме того, он постоянно интересовался вопросом: что же делает брак удачным? Мюррей занимался разработкой нелинейных моделей. После завершения исследования Готтману неплохо удавалось объяснять результаты новобрачным из разных уголков мира.

Рассмотрим подход Готтмана и Мюррея к решению этой проблемы в соответствии с нашей общей трехэтапной схемой количественного анализа.

**Определение и формулирование проблемы.** Рост количества

разводов в развитых странах, включая США, явление широко известное и плохо изученное. Но, не имея теоретического обоснования процессов, приводящих либо к устойчивому браку, либо к его скорому распаду, трудно проанализировать и оценить перспективы новых семейных пар. Если получится разработать математическую модель прочного брака, ее можно будет положить в основу научной теории семейных отношений. Мюррей и Готтман сделали попытку разработать формулу семейного счастья.

**Обзор предшествующих исследований.** Попытки рассчитать вероятность счастливого брака предпринимались неоднократно, но прогнозирование разводов не отличалось ни точностью, ни научностью. Мюррей и Готтман пошли дальше остальных исследователей. Они объединили усилия, чтобы создать математическую модель, способную с поразительной точностью предсказать, разведутся молодожены или нет. Готтман исследовал семейные отношения и случаи разводов в течение многих лет, поэтому предмет был ему хорошо знаком.

**Моделирование (выбор переменных).** Мюррей и Готтман заметили, что разговор на острую тему отражал скрытые проблемы в отношениях семейной пары; приводимые аргументы, смех, подшучивание и проявления любви во время беседы характеризовали эмоциональную связь между собеседниками. В частности, исследователи обращали особое внимание на проявления таких эмоций, как юмор, согласие, радость, любовь, интерес, гнев, доминирование, печаль, жалобы, агрессия, отвращение, пренебрежение и презрение.

**Сбор данных (измерения).** Исследователи пригласили семейные пары поучаствовать в лабораторном эксперименте. Семьсот недавно поженившихся пар согласились. Их просили сесть друг напротив друга и обсудить заведомо острую тему, например секс, деньги или отношения с родственниками, а затем оставляли наедине. Заранее зная, что по этому вопросу у супругов нет согласия, им предлагали беседовать не менее пятнадцати минут, причем их беседа записывалась

на видео. Наблюдатели просматривали запись и выставляли супругам положительные или отрицательные баллы за каждое высказывание. Пары, проявлявшие любовь, чувство юмора, счастье, получали максимум баллов, за презрение и агрессию начислялись самые низкие баллы. Презрение расценивалось как более разрушительное чувство по сравнению даже с отвращением, гневом или печалью.

Проявляемые эмоции расценивались следующим образом:

Юмор: +4	Печаль: -1
Согласие: +4	Жалобы: -1
Радость: +4	Агрессия: -2
Любовь: +4	Оборонительная реакция: -2
Интерес: +2	Соппротивление: -2
Гнев: -1	Отвращение: -3
Доминирование: -1	Презрение: -4

**Анализ данных.** Задача Мюррея и Готтмана состояла в том, чтобы выявить модели поведения, приводящие к разводу или, напротив, к счастливой семейной жизни. Сначала полученные каждым супругом баллы наносились на график так, чтобы через них можно было провести линию. Точка, в которой обе линии пересекались, указывала на успех или провал брака. Если муж или жена постоянно получали отрицательные баллы, то становились кандидатами на развод. Как выяснилось, ключевое значение имело соотношение количества положительных и отрицательных эмоций, проявленных во время беседы. Нижнее значение этого отношения равно 5:1. Если оно падает ниже, то пара, скорее всего, разведется. Результаты ввели в математическую модель, используя разностные уравнения, отражавшие личностные особенности супругов и на этой основе предсказывавшие успешность брака.

Используя итоговые показатели, исследователи выделили пять типов супружеских пар.

*Поддерживающая семья.* Спокойные, глубокие отношения; супруги всячески поддерживают друг друга, сохраняют дружеские чувства. Такие пары предпочитают обмениваться опытом, а не полагаться

каждый на себя.

*Дипломатичная семья.* Делают все возможное, чтобы избежать конфронтации и конфликта. Положительно реагируют друг на друга.

*Нестабильная семья.* Супруги обычно страстно и романтично относятся друг к другу, но очень быстро втягиваются в ссору. Такая семья представляет собой смесь стабильности и нестабильности, но в целом чаще оказываются несчастливы в браке, чем наоборот.

*Враждебная семья.* Один из партнеров не желает обсуждать противоречия, второй партнер с этим согласен. Общение в семье отсутствует.

*Враждебно-отстраненная семья.* Один из супругов — страстный спорщик, а второй просто не заинтересован в обсуждении вопроса.

Математическая модель объяснила различия, с одной стороны, между двумя положительными типами семейных пар (поддерживающей и дипломатичной), чей брак, скорее всего, будет долгим и счастливым, а с другой — между двумя негативными типами (враждебной и враждебно-отстраненной). Нестабильные семьи, невзирая на все свои противоречия, разводятся реже, чем сохраняют брак.

После окончания эксперимента и на протяжении двенадцати лет исследователи выходили на связь с каждой из семисот пар раз в год или два. Предсказание Мюррея и Готтмана оказалось верным в 94 процентах случаев. Шестипроцентная погрешность вызвана тем, что некоторые нестабильные пары, которым прогнозировали долгую, хотя и несчастливую жизнь в браке, все же развелись.

**Результаты и необходимые меры.** Модель семейного счастья была опубликована в книге Готтмана, Мюррея и их коллег под названием «Математика супружества: динамические нелинейные модели» (The Mathematics of Marriage: Dynamic Nonlinear Models). Книга предназначалась в основном для научных работников. Однако в отличие от многих ученых (и в отличие от Мюррея) Готтман был очень заинтересован в практическом применении своей теории. Он опубликовал несколько книг<sup>[61]</sup> и статей, а потом совместно с женой Джулией основал Институт семейных отношений Готтмана

([www.gottman.com](http://www.gottman.com)), в котором устраивали тренинги, создавали учебные фильмы о проблемах совершенствования семейных отношений, организовывали другие подобные мероприятия.

Эта работа весьма полезна для практикующих психотерапевтов, поскольку предоставляет в их распоряжение новые методы преодоления разрушительной модели поведения, доводящей семью до развода. Институт Готтмана разработал комплекс методических рекомендаций и проводил семинары специально для врачей. Помимо всего прочего, модель позволяет исследователям прогнозировать реакцию семейной пары в той или иной ситуации. Таким образом, модель позволяет проводить эксперименты типа «что... если?», а они, в свою очередь, помогают разрабатывать новые научно обоснованные стратегии восстановления отношений в неблагополучных семьях.

Готтман помог провести крупнейшее клиническое исследование методом случайной выборки («проект сумасшедшего ученого») с участием более десяти тысяч пар. Он дал ответ на вопрос, каким образом это исследование способно помочь реальным людям: «В последние восемь лет мы вместе с моей талантливейшей женой полностью погрузились в работу, стараясь свести воедино все наши идеи ради помощи семьям и детям. Теперь мы знаем, что наше вмешательство в жизнь несчастливых семей действительно способно ее изменить. Мы можем помочь наладить отношения примерно 75 процентам пар, проведя для них двухдневный семинар и девять сеансов семейной терапии»<sup>[62]</sup>.

Это мы называем эффективным сообщением о результатах и успешными действиями!

## Пример аналитического мышления: рейтинг компании FICO

Рейтинг FICO — это трехзначный рейтинг, колеблющийся в интервале от 300 до 850 и характеризующий индивидуальное финансовое положение заемщика на данный момент<sup>[63]</sup>. Когда вы подаете заявку на кредит (неважно, на кредитную карту, автомобильный или ипотечный), кредиторы, конечно, хотят знать, насколько рискованно давать вам деньги. Кредитный рейтинг FICO используется большинством банков

для оценки кредитного риска заемщика. Ваш индивидуальный рейтинг влияет на множество решений банка: максимальную сумму, на которую вы можете рассчитывать, условия ее предоставления (процентная ставка и т. п.). Этот рейтинг представляет собой поразительный пример того, как аналитика немедленно конвертируется в действие. Неудивительно, что почти все кредиторы в США и растущее их количество за пределами страны его используют. Посмотрим, как же был разработан этот рейтинг с точки зрения нашей стандартной процедуры из трех этапов и шести шагов.

**Определение и формулирование проблемы.** Кредитные рейтинги позволяют кредиторам быстро и объективно оценить кредитный риск конкретного заемщика. До появления рейтинга FICO процесс одобрения кредитной заявки был долгим, медленным, непоследовательным и зачастую необъективным. Инженер Билл Фэйр и математик Эрл Исаак выдвинули идею о том, что управленческие решения могли бы быть куда более эффективными, если бы можно было статистически оценить риск провала с учетом разнообразных обстоятельств личной жизни и финансового положения заемщика. В 1956 году они основали компанию, занялись разработкой модели, а через два года уже продавали свои системы оценки кредитного риска всем желающим. Первый рейтинг FICO общего назначения появился на рынке в 1989 году.

**Изучение предыдущих попыток решения.** Кредитная история появилась более ста лет назад, когда мелкие розничные торговцы объединялись для обмена информацией о своих покупателях. Ассоциации торговцев постепенно преобразовались в небольшие кредитные агентства, которые впоследствии консолидировались в более крупные компании, использовавшие все преимущества компьютеризации. Но до появления кредитного рейтинга общего назначения было еще очень далеко.

**Моделирование (выбор переменных).** Из многочисленных показателей кредитоспособности клиента Фэйр и Исаак отобрали показанные в табл. 4.1 переменные, объединив их в пять групп (пол,

возраст, национальность, раса и семейный статус были исключены).

Таблица 4.1

## Переменные для анализа кредитоспособности в модели Фэйра и Исаака

Группы	Переменные
Кредитная история	История оплаты счетов, факты невозврата долга, сроки уплаты и/или уклонения от уплаты долга; степень тяжести противозаконных поступков; сумма просроченного долга или сроки просрочки; период после просрочки долга, количество просроченных кредитов; количество счетов, оплаченных вовремя
Сумма задолженности	Общая сумма задолженности по счетам; задолженность по отдельным видам счетов; в некоторых случаях — неплатежеспособность; количество вовремя погашаемых счетов; доля использованных кредитных линий; доля еще не оплаченных взносов по кредиту
Динамика кредитной задолженности	Время, прошедшее с момента возникновения задолженности; время, прошедшее с момента возникновения определенного вида задолженности; время, в течение которого отсутствуют платежи по счету
Новые кредиты	Количество недавно взятых кредитов; удельный вес недавно взятых кредитов в их общей сумме по разным видам кредитов; количество заявок на кредиты в последнее время; период времени с момента предоставления кредитов различных видов; восстановление кредитной истории путем урегулирования прошлых проблем с кредитами
Виды кредитов	Количество используемых видов кредитов

**Сбор данных (измерения).** Финансовые привычки американцев тщательно изучают несколько национальных кредитных агентств (CRA): Equifax, Experian и Trans Union. Ежемесячно финансовые институты и кредиторы отсылают в кредитные агентства файлы заемщиков, содержащие данные о количестве кредитов у потребителей, виды кредитов (ипотечный кредит, займы на кредитных картах, кредиты на покупку автомобиля), балансы, мероприятия по сбору задолженности и истории погашения задолженности. Данные об одном и том же заемщике могут довольно сильно различаться от агентства к агентству. С тех пор как при расчете FICO стали

использоваться данные кредитных агентств, для каждого из нас существует три варианта этого рейтинга — по одному для каждого крупнейшего агентства.

**Анализ данных.** В целом кредитный рейтинг учитывает множество разнообразной информации из кредитного досье клиента, но не вся она одинаково необходима. Некоторые аспекты вашей кредитной истории более важны, чем другие, и, соответственно, более сильно влияют на итоговый рейтинг. Хотя составляющие формулы и методика расчетов хранятся в тайне, а каждое кредитное агентство применяет свою собственную методику, известно, что значимость переменных в формуле характеризуется такими показателями:

35%	Кредитная история
30%	Сумма задолженности
15%	Динамика кредитной задолженности
10%	Новые кредиты
10%	Виды кредитов

**Результаты и необходимые меры.** FICO не хранит результатов оценки кредитного риска заемщика в прошлом. Вместо этого при запросе рейтинга заемщика кредитное агентство генерирует его заново. FICO предоставляет кредитным агентствам программное обеспечение, содержащее алгоритм расчетов, а также математическую формулу, выведенную на основе анализа случайным образом выбранных кредитных историй. На основе этого алгоритма производится расчет рейтинга конкретного заемщика. Однако с тех пор как кредитные агентства начали до некоторой степени модифицировать этот алгоритм, результаты их расчетов по конкретному заемщику могут различаться.

Рейтинг FICO колеблется в интервале от 300 до 850. Лучшие кредитные условия получают заемщики с рейтингом свыше 770, но и рейтинг около 700 тоже считается хорошим. В среднем он составляет 725. Если его значение падает ниже 550, то заемщик может рассчитывать только на различные варианты субстандартного кредитования, где ставки процента значительно превышают обычные.

Наиболее распространенный в мире рейтинг FICO резко повысил эффективность кредитного рынка в США, дав банкам надежный инструмент для оценки риска. Теперь кредитор может уверенно ожидать возврата кредита на основе анализа кредитной истории заемщика. Все большее число компаний, не имеющих никакого отношения к кредитованию (например, страховые компании, мобильные операторы, компании, сдающие в аренду земельные участки или оказывающие разнообразные финансовые услуги), заинтересовано в анализе кредитной истории заемщика, чтобы решить, стоит ли иметь с ним дело и какие условия можно ему предложить. Некоторые работодатели даже интересуются кредитными рейтингами нанимаемых на работу сотрудников, чтобы выявить среди них наиболее надежных. Многие страховые компании используют этот рейтинг для общей оценки финансовой ответственности потенциального получателя автокредита, поскольку считают, что она прямо связана с ответственным поведением за рулем. Короче говоря, рейтинг FICO можно считать одним из крупнейших успехов аналитического мышления за всю его историю.

### Пример аналитического мышления: концепция «магазин ценности»

В мае 1999 года компания Homeplus вошла в сегмент супермаркетов на рынке Кореи, создав совместное предприятие с компанией Tesco. К этому времени сегмент супермаркетов на местном рынке представлял собой поле битвы не на жизнь, а на смерть, и это поле обильно поливали кровью одиннадцать основных конкурентов. Не только местные розничные компании (E-mart, Lotte Magnet, Kim's Club, Hanaro Mart, LG Mart, Mega Mart, Aram Mart и Top Mart), но и западноевропейские гиганты, располагавшие колоссальным капиталом, участвовали в этой борьбе. Homeplus стала двенадцатым конкурентом, но именно ей удалось добиться потрясающего успеха «10 × 10»: добиться объема продаж в десять триллионов долларов в течение десяти лет. За эти десять лет Homeplus обеспечила среднегодовой рост объемов продаж на 47 процентов, а прибыли — на 175 процентов, далеко опередив прежних игроков рынка. Секрет

успеха таился в лидерских качествах генерального директора компании Ли и его стратегии «нет другого пути, кроме победы». Она стала результатом тщательного маркетингового анализа, а в ее основе лежала концепция супермаркетов, полностью перевернувшая представления об этом типе магазинов. Хотя исследование не было чисто статистическим, но все равно можно о нем говорить в рамках шестишагового подхода к количественному анализу.

**Определение и формулирование проблемы.** Homeplus прекрасно понимала, что невозможно добиться успеха, просто создавая такие же магазины, как у конкурентов. Требовалось разработать совершенно новую концепцию супермаркетов, которая могла бы отличить магазины Homeplus от всех прочих. Но в чем ее суть? В море всевозможных идей и маркетинговых теорий Homeplus решила выбрать девиз «Покупатель прежде всего». Супермаркеты конкурентов принадлежали к типу магазинов-складов, копируя американские и европейские образцы. Этажи магазинов были забиты скучными вешалками и полками с горами упакованных коробок. Тщательное исследование и анализ потребностей покупателей — вот единственный способ выяснить, чего же покупатели действительно хотят от супермаркетов.

**Обзор предшествующих исследований** Результаты проведенных на тот момент маркетинговых исследований покупательского спроса были недоступны: их организовывали конкуренты, и результаты были конфиденциальной информацией. Ли любил повторять: стать лучшей розничной компанией мира можно только «взяв лучшее и сделав его еще лучше». Для этого требовалось понять, кто же на текущий момент самый лучший в розничном бизнесе, а затем превзойти лидера.

Чтобы изучить последние течения на глобальном розничном рынке, Homeplus провела сравнительный анализ различных каналов розничной торговли: супермаркетов, гипермаркетов, недорогих специализированных магазинов, универсальных магазинов и многоцелевых торговых центров в двадцати пяти странах мира: в Америке, Японии, странах Европы и Юго-Восточной Азии. Цель анализа состояла в изучении прогрессивного опыта и

совершенствовании своей компании.

**Моделирование (выбор переменных).** В маркетинговых исследованиях используются следующие переменные:

- места для шопинга и причины их выбора;
- цель шопинга, уровень удовлетворенности, типичные жалобы.

**Сбор данных (измерение).** Данные для исследований получены преимущественно двумя основными способами:

- посещение и анализ деятельности различных розничных компаний в разных уголках мира;
- структурированное и углубленное маркетинговое исследование, а также опросы покупателей в Южной Корее независимыми маркетинговыми агентствами.

**Анализ данных.** Результат анализа собранных данных оказался шокирующим. Покупателям требовались не только низкие цены. Конечно, этот фактор тоже важен, но покупатели стремились в первую очередь получить высокую ценность, то есть удовлетворительную цену при минимальных затратах времени с учетом назначения продукта. Они хотят чувствовать заинтересованность в них, поэтому дизайн и атмосфера супермаркета, а также качество обслуживания должны быть на уровне. Еще более важен тот факт, что покупатели хотят не только купить разнообразные продукты по низким ценам, но и получить удовольствие от приобретения всего необходимого в одном месте и в приятной обстановке. Таким образом, новая концепция супермаркета от Homeplus базировалась на понятии «магазин ценности»: здесь не просто продавался широкий ассортимент товаров по низким ценам, но предоставлялись «жизненные ценности», к которым покупатели действительно стремились: любезное обслуживание, приятная обстановка для шопинга, разнообразные удобства, образовательные мероприятия (безусловная культурная ценность в Южной Корее), а также интересные развлечения и культурные программы.

**Результаты и необходимые меры.** Дизайн первых супермаркетов Homeplus в Ансане обеспечивал такую же приятную обстановку для покупок, как в универсальных магазинах, зону отдыха для покупателей на первом этаже с ресторанным двориком на 400 посадочных мест, аптекой, медпунктом, магазином оптики, автоматической прачечной, отделением банка, детской игровой площадкой, комнатой ухода за младенцами и даже государственным центром по обслуживанию населения, выдававшим удостоверения личности. Культурный центр сделал супермаркет не просто местом для покупок, но и крупнейшим центром общения в округе. Все эксперты по розничной торговле, увидев планировку этажей, дружно качали головами и заявляли о нарушении базовых принципов дизайна супермаркетов. С их точки зрения, размещать относительно малоприбыльные общественные удобства на весьма дорогой площади первого этажа не имело никакого смысла. Для супермаркетов считалось особенно важным эффективное управление торговыми площадями, чтобы получать максимум прибыли с каждого метра торговой площади. Поэтому размещение на первом этаже Культурного центра было расценено как нонсенс. Посетители из компании-партнера Tesco тоже выражали сомнения: «Не лучше ли было бы разместить на первом этаже скоропортящиеся продукты, фрукты и овощи, чтобы они сразу попадались на глаза покупателям? Не кажется ли вам, что такая планировка в большей мере соответствовала бы сути розничного бизнеса?» Но Ли ответил: «Генеральный директор уже принял такое решение, а у нас эту должность занимают покупатели», — и продолжил в том же духе.

Homeplus в Ансане, ставший первым «магазином ценности» в Корее, уже в день открытия побил рекорд по объему дневных продаж: этот показатель в полтора раза превысил показатели двух расположенных поблизости конкурирующих супермаркетов, вместе взятых, а число покупателей достигло ста тысяч и примерно равнялось численности населения в радиусе пяти километров. Покупатели немало удивлялись дизайну торговых залов, царившей в них атмосфере, ну а таких общественных удобств они нигде ранее не встречали.

Несмотря на это, эксперты продолжали утверждать, что супермаркет не просуществует и года. Но Homeplus продолжала расти

и ставить рекорды с открытием каждого нового супермаркета. Магазин, открывшийся в районе Сеула Yongdeongpo в 2001 году, обеспечил более высокий объем продаж, чем в универсальном магазине. Взрывной рост продаж и бизнеса существенно укрепил репутацию компании не только на местном, но и на международном рынке. Об этом много писали инвестиционные аналитики.

«Homeplus сумел уникальным образом объединить удивительное маркетинговое чутье с передовыми технологиями розничных продаж».

«Концепция этого супермаркета резко отличалась от концепций конкурентов».

«Объем продаж Homeplus просто невероятен».

Конкуренты начали преобразовывать свои «магазины-склады» в «магазины ценности», но и на этом фоне супермаркеты Homeplus выделялись уникальным сервисом (например, созданием школ для обучения взрослых). По состоянию на 2011 год, такие школы открыты в 110 из 124 супермаркетов компании, расположенных во всех уголках страны. В них обучаются более миллиона человек в год и работают около шести тысяч инструкторов. Это крупнейшая программа по обучению взрослых в мире. Homeplus видит огромное преимущество школ для взрослых в том, что они укрепляют лояльность покупателей. Очень многие утверждают, что именно эти школы — главная причина их приверженности компании. В среднем слушатель школы тратит за один визит в супермаркет Homeplus в два раза больше, чем обычный покупатель, и заходит за покупками в два раза чаще. Ну а положительные отзывы миллиона слушателей школ Homeplus, данные их друзьям и знакомым, — это мощное средство укрепления бренда компании.

## ГЛАВА 5

# Креативность в количественном анализе

В один из дней 275 года до н. э. в общественную баню вошел молодой человек. Погрузившись в ванну с горячей водой, он заметил, как вода начала переливаться через бортики. Внезапно он подскочил и, как был голым, побежал домой, крича «Эврика! Эврика!» Видевшие эту сцену прохожие наверняка решили, что парень сошел с ума. Это был Архимед, а *эврика* на греческом языке означает «Я нашел!» Расскажем вкратце, что же нашел Архимед и почему его это так взволновало. Предварительно отметим, что для решения своей проблемы он использовал одновременно количественный анализ и творческий подход.

Часто считается, что количественный анализ и креативность несовместимы. Креативный подход отличается стремлением к поиску, свободой мышления, вдохновенностью и способностью к провидению. Количественный же анализ воспринимается как скучные, рутинные упражнения с цифрами. Поэтому мы интуитивно ощущаем, что креативный подход и количественный анализ — это противоположные по сути явления, хотя и тесно связанные. Наиболее успешные примеры применения аналитики очень креативны (надеемся, что это уже доказано приведенными в книге историями), да и вообще креативность — важная составляющая аналитического подхода к проблеме. Мы попробуем доказать, что одна только креативность, без сбора информации и аналитики, не может обеспечить оптимального решения. Нам постоянно встречались образцы того, что наиболее успешные люди и организации сочетали креативность с количественным подходом.

В наши дни Apple часто называют одной из самых творческих компаний на планете. Действительно, продукты компании выглядят очень нестандартно. Но тем не менее это не мешает компании строго контролировать производственный процесс и применять аналитику в организации поставок, чтобы удостовериться в том, что необходимые продукты будут готовы к моменту отгрузки. В своих розничных магазинах Apple, к примеру, собирает и анализирует огромное

количество данных. В одной статье говорилось: «Когда продукт поступает в продажу, компания может отслеживать спрос в каждом магазине сети за определенный период и на этой основе ежедневно корректировать планы производства»<sup>[64]</sup>. В описании требований к претенденту на вакансию менеджера в розничной цепи магазинов Apple среди прочего сказано, что успешный кандидат должен «сочетать актуальные знания о цепи поставок продуктов компании мирового уровня, выдающиеся аналитические способности и предпринимательскую жилку»<sup>[65]</sup>. Если даже высококreativeвные компании вроде Apple требуют аналитических навыков от большинства своих сотрудников, то в будущем мы, скорее всего, увидим еще больше разнообразных сочетаний креативности и аналитики.

Конечно, креативность в сочетании с аналитикой может представлять угрозу. Возможно, вы слышали о книге Даррела Хаффа «Как обмануть с помощью статистики» (How to Lie with Statistics), вышедшей в 1954 году. Даже из названия понятно, что знающий основы аналитики мошенник способен использовать количественный анализ для искажения правды<sup>[66]</sup>. Мы часто слышали, как в шутку говорят: «Мы просто пытали статистиков, пока они не сознались». Между допустимой и недопустимой креативностью применительно к аналитике грань довольно тонкая. Критерием тут можно считать искреннее стремление выяснить правду. Если же вы творчески используете аналитику, чтобы доказать правильность вашей (или вашего босса) идеи и при этом обращаете мало внимания на то, что числа упорно не хотят ложиться в заданную схему, то лучше не усердствовать с креативностью и переключиться на другую гипотезу.

## Краткий обзор шести шагов количественного анализа

Сначала сделаем краткий обзор того, каким образом творческий подход мог бы вписаться в наши шесть шагов реализации аналитического проекта (описанных в главах [2](#), [3](#) и [4](#)). Затем рассмотрим те применения креативного подхода, которые выходят за рамки стандартного процесса количественного анализа.

На шаге *определения и формулирования проблемы* творческий подход чрезвычайно важен и полезен. Половина дела в решении проблем и принятии решений состоит в креативном подходе к формулированию

проблемы; именно это позволяет решить ее максимально эффективно. Именно на этом шаге аналитического процесса аналитик разрабатывает гипотезу о закономерностях, имеющихся в данных. Это творческий и интуитивный акт. В определенном организационном и деловом контексте, с учетом ограничений креативная формулировка проблемы может изменить сам контекст, заставить рассматривать его с других точек зрения, сократить или исключить ограничения. Например, в уравнении цены вина, описанном в [главе 3](#), Орли Ашенфельтер сформулировал проблему цены как показателя, который можно спрогнозировать на основании одних только погодных факторов и возраста вина. Аналогично в исследовании продолжительности браков, описанном в [главе 4](#), Джеймс Мюррей и Джон Готтман весьма творчески предположили, что продолжительность брака можно прогнозировать на основе анализа отношений между супругами. Очень часто применительно к аналитическим исследованиям креативность означает, что целый ряд сложных факторов можно прогнозировать и объяснить на основе намного более простых и легко измеримых факторов.

Очевидно, что *обзор результатов* предшествующих исследований не назовешь самым творческим шагом, но и здесь существует возможность проявить креативность, решая вопрос о том, какие именно предшествующие исследования теснее всего связаны с текущим. Например, процедура анализа «дожития» традиционно применяется в тех ситуациях, когда требуется выяснить причины и уровень смертности людей или других живых существ. Однако один из исследователей, Хун Сянлу, успешно применил ее для прогнозирования ценности, получаемой зрителями на протяжении всей жизни в области телекоммуникаций<sup>[67]</sup>. Другие исследователи применили ту же процедуру для решения иных проблем маркетинга, в частности оценки готовности покупателя купить тот или иной товар<sup>[68]</sup>.

*Моделирование (выбор переменных)*, несмотря на аналитическое название, также может быть творческим делом, особенно если вы впервые используете данную модель. Выбор переменных для модели иногда очевиден и определен предшествующими исследованиями или интуицией, а иногда может стать по-настоящему креативным. Вспомним хотя бы анализ длины употребляемых в тексте слов для определения авторства Марка Твена, описанный в [главе 3](#). Для Клода Бринегара это

был умеренно креативный шаг, поскольку о таком методе анализа он и раньше читал в книгах. Зато для Томаса Менденхолла это определенно был весьма творческий подход, поскольку он впервые применил его для установления авторства Шекспира. Конечно, если вы используете такой же подход к разработке модели и выбору переменных, как и все остальные исследователи, то, скорее всего, получите такие же результаты, как у остальных. Тогда зачем вообще тратить время на этот анализ?

*Сбор данных* сам по себе довольно скучен, но решение вопроса о том, какие именно данные собирать, несет в себе массу креативности. Хотите ли вы изучать поведение людей, крыс или атомов, нужно найти какой-то способ наблюдать и оценивать интересующие вас явления, которыми, может быть, никто до вас и не интересовался.

Социальные психологи Михай Чиксентмихайи и Рид Ларсон хотели исследовать эмоции и отношения в среде тинейджеров. Чтобы собрать данные о том, как меняются эмоции подростков на протяжении дня, пришлось разработать уникальный метод. Исследователи раздали пейджеры 75 старшеклассникам и набрали группу студентов университета, поручив им в течение дня в произвольные моменты посылать школьникам на пейджер запрос о том, какие чувства они испытывают сейчас. Этот метод сбора данных получил название *метода выборки переживаний* и ныне широко используется в психологических исследованиях<sup>[69]</sup>.

Исследователи не слишком удивились тому обстоятельству, что в течение большей части дня тинейджеры чувствовали себя несчастливymi. Но то, что их настроение переключается на позитив, когда им нужно выполнить трудное и ответственное задание, оказалось по-настоящему неожиданным. В 1984 году по итогам эксперимента вышла книга *Being Adolescent: Conflict and Growth in the Teenage Years*, в которой впервые Чиксентмихайи описал особое состояние увлеченности и вдохновения, получившее название *потокового*. Впоследствии именно оно оказалось предметом большей части исследований психолога<sup>[70]</sup>. По сути, Михай весьма творчески собирал данные о креативности!

*Анализ данных* считается неподходящим шагом для проявления креативности, если только вы не разбираетесь профессионально в математике и статистике, но и в этом случае лучше действовать

осторожнее. Это именно тот этап аналитического процесса, креативность на котором может закончиться большими неприятностями. Любой статистический эксперимент или математический анализ имеет исходные предпосылки и ограничения; пренебрегать ими не стоит, разве что вы действительно знаете, что вы делаете.

В противоположность этому на шаге *обнародования результатов и принятия мер* творческий подход жизненно важен, но не так уж часто к нему прибегают. Поскольку слушатели без специальной подготовки обычно не воспринимают результаты анализа, изложенные математическим или техническим языком, хороший аналитик должен проявить креативность, придумав, как оформить их в понятном и даже забавном виде. Например, не стоит рассуждать о значениях коэффициентов или доле объясняемой вариации. Вместо этого формулируйте выводы примерно так: «Если мы увеличим расходы на рекламу на доллар, то в среднем получим 1,29 доллара дополнительной выручки». Это звучит гораздо более доступно для понимания и принятия необходимых мер, чем технические рассуждения. Для того чтобы перевести аналитику на доступный язык, как раз и требуется проявить творческий подход.

## Четыре этапа креативного аналитического мышления

Мы не сторонники выделения очень уж большого количества этапов и шагов, но, возможно, будет полезно поговорить о том, каким образом процесс креативного аналитического мышления вписывается в те шесть шагов, о которых мы говорим в этой книге. В целом креативность включает четыре последовательных этапа.

*Подготовка.* Подготовительная работа для решения проблемы.

*Погружение.* Интенсивное погружение в решение проблемы и анализ имеющихся данных. Обычно от аналитика требуются настойчивые и длительные усилия для оптимального решения.

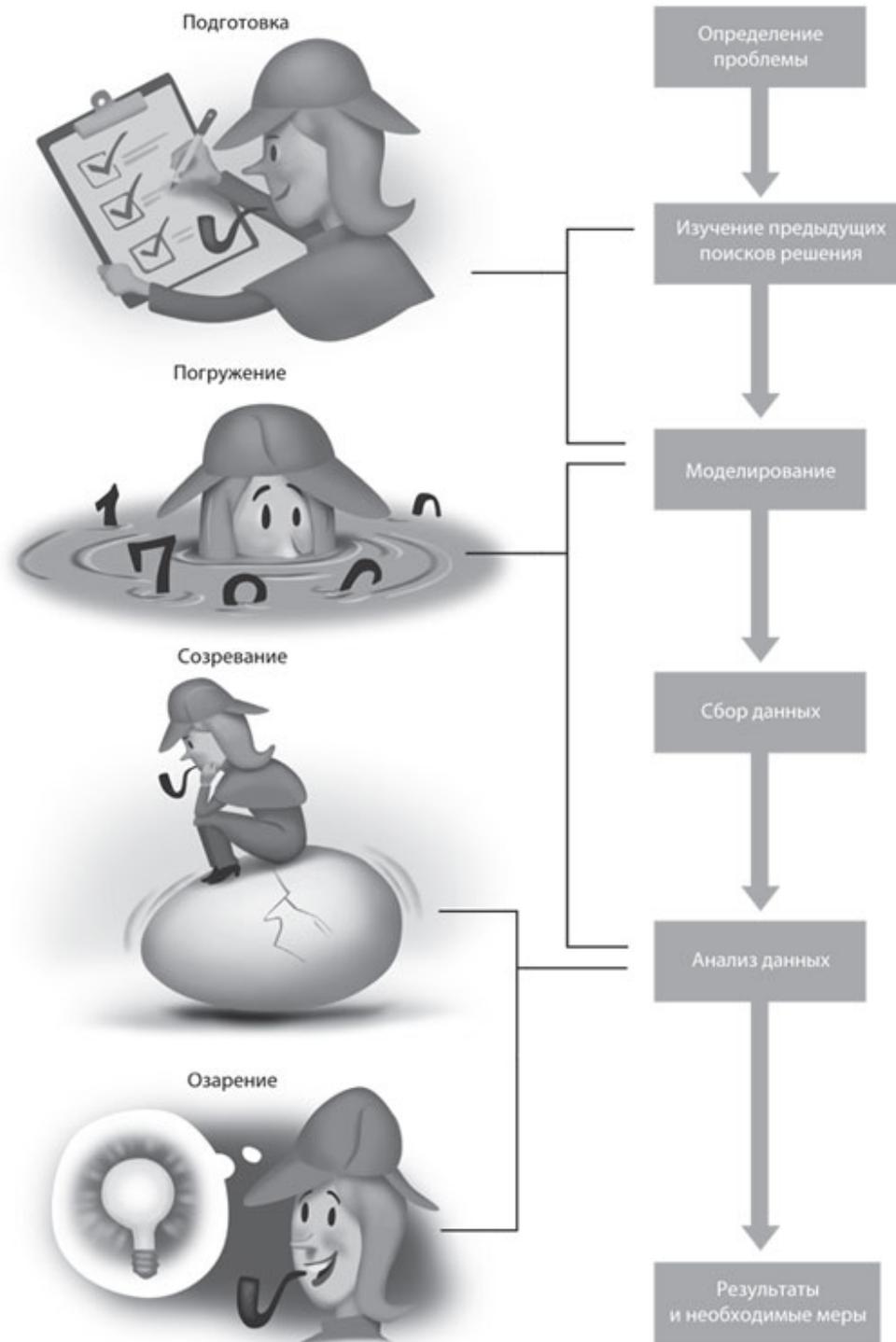
*Созревание.* Проблема «отлеживается» в подсознании, что включает нестандартное мышление (нередко проявляющееся в тот момент, когда аналитик растерян и готов сдаться).

*Озарение.* Большой прорыв в понимании путей решения проблемы с помощью методов количественного анализа.

Обычно большая часть *обзора предшествующих исследований* и

*выбора переменных для модели по данной классификации приходится на этап подготовки. Этап погружения включает часть шага моделирования, весь шаг сбора данных и часть шага анализа данных. Этап созревания начинается тогда, когда аналитик заходит в тупик на шаге анализа данных. Затем, когда внезапно приходит озарение, все части пазла сходятся. Графически все вышеизложенное представлено на рис. 5.1.*

**Рис. 5.1. Качественный анализ и креативность**



## Пример аналитического мышления: Архимед и корона

Проиллюстрируем наложение четырех фаз процесса аналитического мышления на шесть этапов количественного анализа на примере истории Архимеда — одного из первых известных в истории примеров

креативного аналитического мышления<sup>[71]</sup>.

**Определение и формулирование проблемы.** Тиран Сиракуз (Сицилия) Гиерон пожелал украсить храм, посвященный бессмертным богам, золотой короной в форме лаврового венка. Он отмерил точное количество необходимого золота и отдал золотых дел мастеру. В назначенный день мастер, к восторгу Гиерона, принес изящную корону тонкой работы, по весу равную полученному количеству золота. Но когда король уже готовился к церемонии подношения храму, до него дошли слухи, что корона сделана не из чистого золота. Якобы мастер заменил часть золота серебром (в те времена, как и сейчас, ювелиры частенько прибегали к такому мошенническому трюку). Гиерон заподозрил, что стал жертвой обмана. Но как это выяснить? Гиерон задал этот вопрос своим советникам, но те не смогли порекомендовать надежный способ. Наконец Гиерон поручил Архимеду до назначенного дня церемонии выяснить, не было ли разбавлено золото, и при этом *ни в коем случае не повредить корону*. Архимеду на тот момент было всего двадцать два года, но он уже был широко известен своими работами по математике и физике. Он принял вызов.

**Обзор предшествующих исследований** Тогда не существовало способа измерить объем предмета произвольной формы, поэтому Архимеду предстояло самостоятельно его найти. С этого начался этап *подготовки*. Архимед, возможно, один из величайших математиков и изобретателей всех времен, умел измерять объем большинства предметов правильной формы. Он рассудил, что, поскольку золото весит больше, чем серебро, то корона, изготовленная из сплава золота и серебра, при одинаковом весе должна иметь больший объем, чем корона из чистого золота. Проблема в том, как точно измерить объем предмета неправильной формы (короны), не повредив его.

**Моделирование (выбор переменных).** На том же этапе подготовки Архимед решил, что объем короны и будет ключевой переменной, вычислив которую можно будет ответить на поставленный вопрос. Но чтобы сделать это, требовалось *полностью* погрузиться в имеющиеся данные и возможные варианты измерения объема. Самый простой способ — переплавить корону в куб и измерить его грани. Но этого

делать нельзя, ведь Гиерон приказал не повреждать ее. Ученый постоянно думал об этой проблеме и в конце концов решил отложить поиск решения на некоторое время. Однако, как он ни старался, найти ответ не удавалось.

**Сбор данных.** Погрузившись в проблему, Архимед должен был собрать некоторые данные. Он выяснил плотность золота и серебра и предположил, что сплав, из которого изготовлена корона, содержит 30 процентов серебра. Но как бы то ни было, требовалось измерить объем короны в форме лаврового венка, а он не знал, как это сделать.

**Анализ данных.** День церемонии подношения короны храму приближался. Расстроенный Архимед уже готов был признать поражение, пусть даже его репутация гениального математика и физика серьезно пострадала бы. Но в результате глубокого погружения в проблему мозг продолжал работать над ее решением на подсознательном уровне. Начался этап *созревания*. Как-то раз Архимед решил дать отдых телу и духу, изнуренному бесплодной борьбой с проблемой, отправившись в общественную баню. Когда он погрузился в горячую ванну (это другая форма погружения!), вода начала переливаться через борта. Внезапно математик понял, что объем вытесненной воды равен объему той части его тела, которая погружена в воду. Это означало, что найден способ точно измерить объем предметов неправильной формы: если погрузить корону в заранее известный объем воды и она вытеснит больше воды, чем такое же по весу количество золота, то отсюда можно будет сделать вывод о том, что объемы короны и бруска золота не равны. Он выскочил из ванны и голым помчался по улицам Сиракуз, радостно крича: «Эврика! Эврика!» Очевидно, что в этот момент на Архимеда снизошло *озарение*. Он провел эксперимент, погрузив в воду сначала корону, а потом брусок золота равного веса, и обнаружил, что их объем различен. Следовательно, можно было утверждать, что корона изготовлена не из чистого золота и мастер действительно добавил более дешевый металл, стремясь присвоить часть драгоценного материала.

**Результаты и необходимые меры.** Архимед сообщил о своем эксперименте тирану Гиерону, и тот восхитился изяществом найденного решения. Репутация Архимеда как гениального математика и физика еще

более упрочилась. Однако далеко не для всех история кончилась благополучно, ведь нечистого на руку мастера казнили.

Креативность можно определить как *способность генерировать оригинальные и полезные идеи*. Пример Архимеда говорит о том, что с точки зрения количественного анализа креативность — это всего лишь *способность видеть неочевидные связи между переменными*, предварительно отобрав их для анализа и собрав необходимые данные. В соответствии с этим определением креативности можно научить и научиться, внедрить в практику повседневной деятельности. Человек может развить в себе творческий подход, приобретя аналитические навыки. Компании могут добиться того же, внедрив ряд образовательных программ по аналитике для сотрудников и создав корпоративную культуру, поощряющую аналитическое мышление.

## Погружение и упорный труд как источники креативности и озарения

Томас Эдисон произнес знаменитую фразу: «Гений — это на 99 процентов труд и лишь на 1 процент — вдохновение». Творение — это озарение, нисходящее к человеку неожиданно и позволяющее решить проблему. Но откуда берутся творческие находки? Мы считаем, что это продукт упорного труда. Озарение и ведущее к нему креативное аналитическое мышление отнюдь не даются человеку от природы случайно — это результат упорного труда на ниве количественного анализа: анализа предшествующих исследований, отбора и измерения переменных, настойчивого поиска закономерностей, управляющих динамикой данных. Если вы каждый день без усталости бьетесь над решением задачи и не сдаетесь, то в один прекрасный день сможете воскликнуть: «Эврика!» Именно это случилось с Архимедом, Ньютоном и многими другими гениями.

Американский ученый Барбара Макклинток в 1983 году получила Нобелевскую премию за открытие геной транспозиции. Она единственная женщина, единолично получившая Нобелевскую премию в физиологии и медицине. Во время своих полевых экспериментов она часто испытывала моменты «Эврика!» и говорила, что источник ее вдохновения — это забвение себя, полная погруженность в работу и подчинение всей жизни ей.

Макклинток настолько хорошо изучила кукурузные стебли, что, исследуя хромосомы, идентифицировала себя с ними:

Я обнаружила, что чем больше я работаю, тем большая часть моей личности ассоциируется с ними. Я уже не воспринимала себя отдельно от них; казалось, что я где-то там, внутри. Я стала частью системы. Казалось, я могу рассмотреть внутреннее устройство хромосом — все, что там было. Это было удивительно, ведь была полная иллюзия вхождения в клетку, а все, что там было, стало моими друзьями. Когда вы смотрите на них, то чувствуете, что они становятся частью вас. Вы забываете себя. Это самое главное: вы забываете себя<sup>[72]</sup>.

Конечно, креативность и тяжелый труд ассоциируются не только с миром количественного анализа. Например, они точно так же присутствуют в мире искусства и литературы. Известный корейский автор Те Дзюн Ра тоже считает, что вдохновение не вспыхивает внезапно на пустом месте, а требует долгой и трудной подготовительной работы, погружения в тему.

Вы концентрируетесь и погружаетесь в одну тему, продолжая накапливать разнообразные идеи, а затем в один прекрасный момент вас озаряет вспышка, неожиданный полет вдохновения и вдруг становится понятно, что же вы так отчаянно искали все это время. Говорят, что вдохновение приходит внезапно. Это правда, если говорить о самом моменте озарения. Но до этого может потребоваться долгий и кропотливый труд. В конечном счете можно сказать, что глубина погружения в предмет и объем проделанной кропотливой работы предопределяют приход вдохновения<sup>[73]</sup>.

Чаще всего интуицию определяют как непосредственное восприятие истины без всякого объяснения или логического обоснования. Однако мы верим в то, что интуиция — это качество, которое можно развить в себе путем постоянного анализа, как количественного, так и качественного. Нейробиологи считают, что нейроны можно «обучать» путем постоянного повторения анализа. Мы считаем, что точно так же способность к интуитивному восприятию развивается, если человек на основе накапливаемого со временем опыта пытается выявить взаимосвязи между переменными, не собирая и не анализируя для этого данные. Немецкий философ Георг Гегель считал, что только тот, кто способен мыслить глубоко аналитически, обладает чистой истинной интуицией<sup>[74]</sup>.

## Поиск моделей методами аналитической креативности

Суть креативного анализа данных состоит в выявлении модели отношений между переменными. Модель — это устойчивая неочевидная

взаимосвязь переменных. На этом этапе математическое мышление весьма полезно, поскольку математика сама по себе является наукой о моделях: эвклидова геометрия, первая великая наука Древней Греции, развилась из наблюдения за геометрическими фигурами в окружающей природе. Теорема Пифагора — это не что иное, как модель зависимости между тремя сторонами прямоугольного треугольника.

Поиск моделей в современных организациях осуществляется главным образом путем статистического, а не чисто математического анализа. Некоторые статистические инструменты помогают выявить наличие в данных модели — связи, встречающейся чаще, чем можно было бы ожидать, если бы распределение было случайным. Анализ дает основания полагать, например, что потребители с определенной моделью покупательского поведения (или ее отсутствием) в тот или иной момент могут вообще прекратить покупки. Или что покупатели, приобретающие определенную книгу, часто покупают и другие: например, на сайте Amazon.com имеется встроенная функция рекомендаций. (Друг Тома Дэвенпорта получил от Amazon.com рекомендацию купить вместе с книгой Тома шуточный сувенир в виде собачьих экскрементов — по мнению сайта, именно такое сочетание предпочитают другие покупатели!) В табл. 5.1 представлены различные виды моделей, а также бесплатное и коммерческое программное обеспечение для их построения.

Таблица 5.1

### Программное обеспечение для интеллектуального поиска данных для построения моделей на основе баз данных

Задачи	Описание задач	Бесплатное		Коммерческое универсальное (решение множества задач)
		специализированное (решение одной задачи)	универсальное (решение множества задач)	
Регрессионный анализ	Поиск функции, описывающей данные		RapidMiner R KNIME Weka	Microsoft Analysis Services SAS: Enterprise Miner Oracle Data Mining

Задачи	Описание задач	Orange		IBM-SPSS Modeler Statistica Коммерческое универсальное решение для анализа данных
		Бесплатное	универсальное (решение множества задач)	
Классификация	Распределение данных по одной или нескольким заданным категориям	CART ML-Flex		
Кластеризация	Поиск групп объектов (кластеров)	Carrot2 ELKI		
Последовательный ассоциативный анализ	Поиск взаимосвязей между переменными	JHepWork ARtool		
Выявление аномалий	Поиск нетипичных значений в базе данных	ELKI JHepWork		
Интеллектуальный анализ текстов	Поиск моделей и тенденций в тексте	GATE NLTK UIMA RapidMiner		

### Пример аналитического мышления: пиво и пеленки

Более подробно механизм выявления моделей на основе анализа массива данных можно проиллюстрировать на хорошо известном примере из области интеллектуального поиска данных: мужчины, заходящие в продовольственные магазины только по выходным, склонны вместе с пивом покупать пеленки. Это открытие было сделано отнюдь не в Древней Греции, а в Чикаго в 1992 году. Его определенно можно считать примером выявления модели поведения; другое дело, что вряд ли это удачный пример аналитического мышления (правда, мы считаем, что изучение и положительных, и отрицательных примеров в равной степени поучительно)<sup>[75]</sup>.

**Определение и формулирование проблемы.** Том Блисчок, на тот момент менеджер группы консультантов по розничной торговле в компании Teradata, создающей аппаратно-программные комплексы для обработки и анализа данных, вместе с командой проводил анализ

розничной торговой точки (point-of-sale — POS) для своего клиента, компании Osco Drug. По словам одного из членов этой группы, Джона Эрла, целью их работы был поиск закономерности в том, какие продукты чаще всего приобретаются вместе: «Мы предложили несколько экспериментов по размещению товара в торговом зале, чтобы посмотреть, как это будет влиять на склонность потребителей приобретать определенные товары вместе»<sup>[76]</sup>. Эрл предполагает, что истинной целью анализа было не столько решить какую-то конкретную проблему или обосновать управленческое решение для клиента, сколько продемонстрировать ему преимущества технологий, применяемых Teradata.

**Обзор предшествующих исследований** Подобные исследования ранее практически не проводились. Единственное, что было известно, это то, что товары для малышей высокорентабельны, поэтому было бы хорошо найти товары, вместе с которыми они бы чаще продавались и, соответственно, рос бы их объем продаж.

**Моделирование (отбор переменных).** Частота покупок различных видов товара по данным кассовых аппаратов аптек.

**Сбор данных.** Osco Drug располагала данными с кассовых аппаратов в своих аптеках (розничных торговых точках) и предоставила их аналитикам. База данных содержала информацию о перечне и ценах купленных по одному чеку товаров по 1,2 миллиона таких чеков, пробитых в двадцати пяти аптеках сети.

**Анализ данных.** Сегодня доступны значительно более сложные технологии интеллектуального поиска данных, но в 1992 году они еще не получили распространения. Команда аналитиков Teradata сформулировала запросы по базе данных, чтобы выявить товары, покупаемые вместе чаще, чем другие. Такой анализ, проведенный К. Хис, одной из членов группы, показал, что покупатели (вообще говоря, не обязательно мужчины), заходившие в магазин в интервале между 17:00 и 19:00 по четвергам и субботам, очень часто покупали вместе пиво и пеленки. Однако никаких статистических тестов для подтверждения того, что это не случайное явление, проведено не было.

**Результаты и необходимые меры.** Это именно тот случай, когда анализ показал свою неэффективность. В досужих рассуждениях об этой истории частенько проскальзывали разные предположения — например, что магазинах пиво и пеленки располагались на соседних стеллажах или что, наоборот, они находились в разных концах магазина и покупателям приходилось пересекать весь торговый зал. По сути, ни одно из этих предположений не подтвердилось. Результат анализа сочли забавным курьезом, и ни аналитики Teradata, ни менеджеры Osco Drug даже не пытались произвести какие-то действия на их основе или хотя бы оценить потенциальные их последствия.

У нас недостаточно информации, чтобы судить о том, почему этот интересный пример поиска моделей в данных так и не получил достойного продолжения. Но он наглядно свидетельствует о том, что любой этап аналитического процесса будет эффективным только в том случае, если приведет к какому-то результату. Компьютеры способны найти модели в базе данных, но только человек может сказать, есть ли в этих моделях какой-то смысл, и принять соответствующие меры (см. вставки «Компьютеры и модели: число  $\pi$ » и «Закон первой цифры — способ обнаружения мошенничества»).

## Компьютеры и модели: число $\pi$

Число  $\pi$  — это отношение длины окружности к ее диаметру.  $\pi$  приблизительно равно 3,141592 в обычном десятичном исчислении. Многие формулы математики, инженерного дела и науки используют это значение, что и делает его самой важной математической концепцией после теоремы Пифагора<sup>[77]</sup>.

$\pi$  — иррациональное число, что означает, что его значение не может быть выражено обыкновенной дробью, а последовательность знаков после запятой никогда не заканчивается и не является периодической. Однако это не значит, что человечество, начиная с древних вавилонян и до современных математиков, прекратило попытки обнаружить повторяемость в десятичных знаках  $\pi$ .

Конечно, появление компьютеров в XX веке привело к новым попыткам поставить рекорд, рассчитав число  $\pi$  до еще большего количества знаков; компьютеры полностью заменили в этом деле людей. Нынешний рекорд количества разрядов, до которых рассчитано число  $\pi$ , составляет пять триллионов. Дальнейшее увеличение числа разрядов зависит не столько от математических способностей человека, сколько от технических возможностей компьютеров. Тем не менее до сих пор повторяющихся групп цифр в знаках числа  $\pi$  не обнаружено.

Зато математики нашли множество новых применений этому замечательному числу, что говорит о важной роли креативности в аналитическом мышлении. Например,

математик Дэвид Ачесон рассказывает такую историю.

Представьте себе удивление математиков, когда в середине XVII века они обнаружили появление числа  $\pi$  в разных областях, подчас весьма далеких от геометрии окружности. Один из самых замечательных фактов этого рода состоит в необычной связи между  $\pi$  и рядом нечетных чисел.

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

...Удивительные взаимосвязи такого рода принадлежат к явлениям, которые всегда волновали математиков»<sup>[78]</sup>.

Была обнаружена также взаимосвязь между рядом четных чисел и числом  $\pi$ . Более того,  $\pi$  появляется в уравнении «знаменитая пятерка»<sup>[79]</sup>, выведенном великим швейцарским математиком XVIII века Леонардом Эйлером. В 1988 году читатели авторитетного математического журнала признали это уравнение «самой красивой математической формулой в истории математики».

Все эти изыскания в области истории расчета и применения числа  $\pi$  говорят о том, что даже сейчас, когда компьютеры выполняют многие математические расчеты, креативность человеческого ума всегда найдет себе выход. Применительно к количественному анализу в сфере бизнеса и корпораций компьютеры взяли на себя подавляющее большинство статистических расчетов. Но творчеству по-прежнему есть место там, где заходит речь об использовании этих расчетов для принятия решений.

## И в заключение об аналитике и креативности

Хотелось бы надеяться, что нам удалось показать читателям: аналитическое мышление и креативность не только вполне совместимы, но и тесно взаимосвязаны. Вы не сможете ни стать хорошим количественным аналитиком, ни компетентно пользоваться аналитическими данными, если не умеете подключать к делу все ваши творческие способности. Однако помните, что в манипулировании цифрами и интерпретации результатов анализа баз данных креативность должна иметь предел. Творческий подход очень важен, но правда еще важнее.

## Закон первой цифры — способ обнаружения мошенничества

Профессор математики в Политехническом институте штата Джорджия Тед Хилл в начале курса дает студентам задание на дом: или подбросить монетку двести раз и записать результаты, или представить, как подбрасываешь монетку двести раз, и сфальсифицировать результаты.

На следующем семинаре он просматривает отчеты о домашней работе и, к восторгу аудитории, легко обнаруживает почти всех, кто занимался фальсификациями. Как это ему удается? На основании небольшого эксперимента он знает, что в некоторый момент при длительном подбрасывании монетки начинают выходить серии из шести-семи орлов или решек подряд. «Фальсификаторы» об этом не знают и интуитивно пытаются не писать подряд слишком много одинаковых результатов, поскольку считают, что это маловероятно. Хилл с первого взгляда выявляет записи о шести-семи орлах или решках, выпавших подряд (или их отсутствие), на основе чего и делает вывод о действительно проведенном студентом эксперименте или о фальсификации его результатов. На первый взгляд, это просто небольшой фокус для привлечения внимания студентов, но на самом деле в этом есть глубокий смысл. Если в данных отсутствуют те модели и зависимости, которые вы ожидали там увидеть, логично предположить фальсификацию или мошенничество.

Мы хорошо знаем, что наша система исчисления использует цифры от 1 до 9. Поэтому можно предположить, что вероятность выбора любой из этих цифр в качестве первой значащей в числе равна  $1/9$ . Но, как ни странно, это не так. «Закон первой цифры», называемый также законом Бенфорда, гласит, что в списке чисел, взятых из реальных баз данных, частота распределения той или иной цифры на первое место в числе подчиняется специфической закономерности: примерно в 30 процентах случаев такой цифрой будет 1, а вероятность появления на первом месте остальных цифр тем меньше, чем цифра больше<sup>[80]</sup>. В соответствии с законом Бенфорда вероятность распределения цифр на первом месте в числе такова:

Первая цифра	1	2	3	4	5	6	7	8	9
Вероятность распределения на первое место в числе, %	30,1	17,6	12,5	9,7	7,9	6,7	5,8	5,1	4,6

Этот довольно удивительный факт был обнаружен в 1881 году американским астрономом Симоном Ньюкомбом, заметившим, что в

справочнике логарифмов первые страницы всегда гораздо более потрепанные, чем последующие. В 1938 году физик Франк Бенфорд сделал то же открытие на основе анализа гораздо большего массива данных, чем Ньюкомб. Он рассмотрел 20 229 наборов данных, включая географические координаты рек, бейсбольную статистику, количество статей в журналах, и номера домов первых 342 человек, перечисленных в рейтинге «Деятели науки Америки». Анализ всех этих вроде бы не связанных друг с другом баз данных показал, что вероятность распределения цифр на первое место в числе та же, что и для потрепанных таблиц логарифмов. Эта модель определения первой цифры в числе получила впоследствии название закона Бенфорда в честь ее первого исследователя. Стало общепризнанным действие закона Бенфорда во многих ситуациях реальной жизни.

Многие статистики и бухгалтеры твердо убеждены в том, что закон Бенфорда является очень простым, но надежным способом выявления возможных случаев мошенничества, хищений, неуплаты налогов и бухгалтерской небрежности. Идея проста: если кто-то фальсифицирует базу данных, то вряд ли он сможет имитировать распределение первой значимой цифры в числах в соответствии с законом Бенфорда. Поэтому простое сравнение распределения по закону Бенфорда и фактического распределения первых значимых цифр в базе данных поможет выявить сфальсифицированные блоки чисел. Как правило, в них частота распределения 1 как первой значимой цифры намного меньше 30 процентов, зато частота распределения 6 — намного больше, чем в доброкачественных массивах данных.

В 1972 году докторант университета Беркли Хэл Вэриан показал, что этот закон применим для выявления возможных искажений в массивах социологических данных, предоставляемых для подтверждения общественного мнения по готовящемуся политическому или экономическому решению. Судебный эксперт по бухгалтерским вопросам Марк Нигрини получил известность благодаря применению разработанной им на основе закона Бенфорда системы выявления мошенничества в некоторых громких делах в Бруклине. В наши дни многие налоговые управления используют специальное программное обеспечение для выявления фальсифицированных данных, разработанное на основе закона Бенфорда. Точно так же поступают

крупные компании и аудиторские фирмы. В США доказательства фальсификаций на основе закона Бенфорда официально признаны в судебных разбирательствах на государственном уровне, уровне штата и местном уровне.

## Пример аналитического мышления: языковые способности и болезнь Альцгеймера

Болезнь Альцгеймера — это дисфункция мозга, вызывающая проблемы с памятью, мышлением и поведением. Симптомы обычно развиваются медленно, с течением времени усиливаются и в конце концов начинают мешать человеку выполнять обычные домашние дела; затем пациент умирает по неизвестной причине. Болезнь Альцгеймера лежит в основе 60–80 процентов всех случаев приобретенного слабоумия. Около 5,3 миллиона американцев, в том числе каждый восьмой в возрасте старше 65 лет (а их 13 процентов от общей численности населения), страдают от болезни Альцгеймера. В США эта болезнь является шестой по распространенности причиной смерти. Помимо страданий самого пациента необходимо принять во внимание эмоциональные и физические страдания его семьи, необходимость повседневного ухода, изменение социальных ролей в семье, трудное решение о помещении в специальное лечебное учреждение.

Причины возникновения и развития болезни Альцгеймера до конца не выяснены. Многие исследователи пытались найти взаимосвязь между развитием болезни и характеристиками (или маркерами) группы повышенного риска. Например, люди с низким уровнем образования в большей степени подвержены заболеванию, чему способствует их образ жизни, в том числе питание, ежедневные занятия, употребление алкогольных напитков и риски профессиональной деятельности. Профессор центра геронтологии Сандерс-Браун при Кентуккийском университете Дэвид Сноудон и его коллеги считают, что языковые способности — более надежный показатель для прогнозирования вероятности болезни Альцгеймера, чем образ жизни<sup>[81]</sup>. Они предположили, что высокий уровень лингвистических способностей становится неким буфером для развития когнитивного снижения мнемонических процессов в мозгу, обеспечивающих декодирование, организацию и поиск информации. Они провели образцовое и весьма

креативное исследование, связав когнитивные способности в молодом возрасте с риском возникновения болезни Альцгеймера в пожилом. В качестве целевой выборки они, что весьма необычно, взяли членов монашеского ордена, причем в качестве источников информации использовались их автобиографии. Рассмотрим это исследование с точки зрения наших шести шагов количественного анализа.

**Определение и формулирование проблемы.** Определить, ассоциируются ли языковые способности в раннем возрасте с когнитивными функциями и низкой вероятностью болезни Альцгеймера в пожилом.

**Обзор предшествующих исследований** Многие аналитические процедуры, применявшиеся командой Сноудона, ранее были описаны в работах доктора Дэвида Векштейна и доктора Уильяма Марксбери. В 1989 году они проводили исследование изменений когнитивной функции и поведения, связанных с возрастом, в группе пожилых пациентов, согласившихся после смерти пожертвовать свой мозг для научных исследований. Цель исследования состояла в том, чтобы установить, как изменения ткани мозга связаны с развитием болезни Альцгеймера и других неврологических заболеваний.

**Моделирование (отбор переменных).** Участниками исследования Сноудона стали сестры ордена Нотр-Дам из Милуоки (штат Висконсин). С 1991 по 1993 год сестер монастыря, родившихся до 1917 года, попросили принять участие в долгосрочном исследовании причин болезни Альцгеймера в пожилом возрасте. Из 1027 подходивших по возрасту сестер предложение приняли 678 (66 процентов), дав письменное согласие. Процент согласившихся довольно высок, особенно с учетом того факта, что все участницы обязались пожертвовать свой мозг для научных исследований после смерти, а также ежегодно проходить обследование когнитивных функций и общего физического состояния. Далее Сноудон и его коллеги обследовали выборку из 93 участниц, при вступлении в монастырь написавших автобиографию. Эти бумаги сохранились в монастырском архиве. Были выбраны следующие переменные для анализа:

— Языковые способности в молодости (способность формулировать

мысли и грамматическая сложность речи).

- Когнитивные функции (семь различных показателей) и наличие болезни Альцгеймера в пожилом возрасте.

**Сбор (измерение) данных.** Автобиографии сестер использовались для оценки их языковых способностей в молодости. После примерно четырех лет жизни в монастыре каждая из них написала автобиографию незадолго до принятия монашеского сана. Из архивных документов стало известно, что всех их попросили написать краткий очерк о своей жизни. «По объему он не должен был превышать двухсот-трехсот слов и одного листа... в нем должны быть указаны место рождения, имена родителей, памятные и поучительные события детства, посещение школы, факторы, повлиявшие на решение уйти в монастырь, религиозная жизнь с ее замечательными событиями».

Каждую автобиографию оценивали по двум параметрам: *способность формулировать мысли* и *грамматическая сложность речи*. *Способность формулировать мысли* оценивалась путем подсчета «идей», приходящихся на каждые десять слов текста. Под идеями понимались элементарные предложения, в состав которых входили глагол, наречие или прилагательное, существительное с предлогом. Сложные фразы — те, что содержат констатацию или предположение о существовании причинно-следственных, мирских и церковных или других взаимосвязей между элементарными идеями.

*Грамматическая сложность* рассчитывалась на основе методики оценки развития, классифицирующей предложения по восьми уровням грамматической сложности, начиная от 0 (простые односоставные предложения) и до 10 (сложные предложения со вставными оборотами и подчиненностью).

Когнитивные функции оценивали на основе комплекса из семи нейропсихологических тестов. Они включали оценку памяти, способности к концентрации, речи, способности ориентироваться в пространстве и времени. Девяносто три участницы написали первые биографии в возрасте около 22 лет и прошли обследование когнитивных функций в среднем 53 спустя, в возрасте от 75 до 87 лет.

**Анализ данных.** Неспособность формулировать мысли и низкая грамматическая сложность биографий, написанных в юные годы,

ассоциировались с невысокими результатами когнитивных тестов в пожилом возрасте. При этом неспособность формулировать мысли имела большее влияние на результаты когнитивных тестов по сравнению с низкой грамматической сложностью. Из четырнадцати умерших сестер подтвержденная неврологом болезнь Альцгеймера имела у всех тех, кто когда-то продемонстрировал низкие способности формулировать мысли; из тех, кто показал хорошие результаты по этому критерию, болезнью Альцгеймера не страдал никто.

**Результаты и необходимые меры.** Сноудон и его коллеги пришли к выводу, что способность письменно излагать свои мысли, принятая в качестве показателя когнитивных способностей в молодости, «является эффективным маркером когнитивных проблем, болезни Альцгеймера и заболеваний мозга в пожилом возрасте». Это означает, что слабые лингвистические способности в молодости — это первый симптом изменений в мозге, впоследствии ведущих к болезни Альцгеймера. Такие исследования помогают точнее определить группу риска для болезней, связанных с возрастом.

Сноудон и его коллеги опубликовали статью *Linguistic Ability in Early Life and Cognitive Function and Alzheimer's Disease in Late Life* в ведущем медицинском журнале *Journal of the American Medical Association*. Сноудон, кроме того, написал популярную книгу о сестрах ордена Нотр-Дам под названием *Aging with Grace: What the Nun Study Teaches Us About Leading Longer, Healthier, and More Meaningful Lives*. Ее высоко оценили, в том числе журнал *Library Journal*.

Книга Сноудона написана с симпатией к этим монахиням и восхищением перед сестрами, благородно согласившимися пожертвовать свой мозг для научных исследований после смерти. Результаты исследования Сноудона показали, что патологические изменения поведения не всегда вызываются видимыми изменениями в мозге, а лингвистические способности в молодости, по всей видимости, предотвращают в старости развитие болезни Альцгеймера. Предупреждение инсультов и сердечных заболеваний помогает избежать деменции (приобретенного слабоумия), а наследственность, диета и физические упражнения также играют в этом значительную роль. Изложение научных фактов на фоне реальных человеческих судеб, вдохновляющий и восхищенный взгляд на процесс старения будут весьма интересны читателям<sup>[82]</sup>.

Работа Сноудона с сестрами-монахинями стала темой статьи, проиллюстрированной на обложке журнала *Time* в 2009 году<sup>[83]</sup>. Это еще раз говорит о том, что творческое аналитическое мышление способно

заинтересовать самую широкую аудиторию.

## Пример аналитического мышления: продажа инсайдерской информации

Симон Хайнес в прошлом работал инвестиционным банкиром в банке Masquarie. Под вымышленным именем Марк Бус он купил опцион на пакет акций транспортной компании TNT на общую сумму около 90 тысяч долларов. Банк Masquarie предоставлял услуги по финансовому консультированию компании TNT, и перед самым увольнением из банка Хайнес получил косвенные свидетельства того, что TNT в скором времени будет поглощена. Так и случилось буквально через пару дней после покупки опциона. Хайнес исполнил опцион и получил прибыль около двух миллионов долларов за одну эту операцию. Кроме того, он умело замел следы, и инспекторам из Австралийской комиссии по ценным бумагам и инвестициям пришлось проделать большую аналитическую работу, чтобы разобраться в сути махинации.

**Определение и формулирование проблемы.** Нетипичная торговая активность с опционами TNT была отмечена за три дня до объявления о поглощении компании. Это объявление привело к росту рентабельности на инвестиции в акции компании до 200 процентов. Хотя инспекторы биржи называли нескольких физических и юридических лиц, купивших и продававших акции компании на протяжении этих трех дней, но они не могли определить, кто из них был первым. Дело было передано в Комиссию по ценным бумагам и инвестициям — орган, ответственный за состояние фондового рынка. Несмотря на то, что к расследованию привлекли все возможные ресурсы и обычные судебные процедуры, Комиссия три месяца не могла добиться результата. В конце концов она пришла к выводу, что операции проводились по фальшивым документам.

**Изучение предыдущих поисков решения.** Конкретно в рамках этого случая предыдущие исследования отсутствовали, поскольку он оказался исключительным. Но у сотрудников Комиссии был большой опыт проведения аналогичных расследований. Они предположили, что информация об операции может «протечь» в социальную сеть. Члены Комиссии умели искать информацию о людях, компаниях и адресах во

внутренних корпоративных и социальных сетях.

**Моделирование (отбор переменных).** В качестве двух основных переменных в модель включены доступ неустановленного лица к информации о грядущем поглощении компании TNT и наличие у него необходимых предпосылок (счет в банке, деньги и т. п.) для торговли ценными бумагами на момент поглощения TNT.

**Сбор (измерение) данных.** Традиционные методы расследования позволяли установить того, кто имело или мог иметь доступ к инсайдерской информации («информированные лица»). Кроме того, с их помощью можно было отследить операции по снятию соответствующих денежных сумм со счетов («люди с наличностью»). Имея эту информацию, они могли выявлять и накапливать сведения о связях между людьми, компаниями, адресами и активами, принадлежащими «информированным людям» и «людям с наличностью». Процесс анализа этих связей привел к созданию аналитической базы данных, содержащей сведения о более чем 160 тысячах человек, компаний, адресов, активов и операций по снятию наличных, между которыми установлено более миллиона разнообразных контактов.

**Анализ данных.** Одни и те же элементы многократно встречаются в базе данных в связи с различными операциями, поэтому, прежде чем переходить к следующему этапу, аналитики решили выделить операции, относящиеся к одному и тому же элементу (то человеку, компании, активу или адресу нахождения). Для выполнения этой задачи было задействовано более сотни разработанных специалистами Комиссии алгоритмов. Энтони Вьель, на тот момент главный следователь Комиссии по данному делу, а ныне партнер по аналитике и расследованиям в австралийском отделении аудиторской фирмы Deloitte, так прокомментировал этот процесс: «После того как все повторяющиеся элементы были объединены, мы запустили специальный алгоритм для выявления связей между “информированными людьми” и “людьми с наличностью”. При этом характер связей оценивался как либо “слабый”, либо “прочный”, чтобы можно было отсортировать полученные результаты. На первом этапе анализа мы выявили 65 элементов со слабыми и прочными связями, на втором из них было

отобрано только два элемента с прочными связями. Один из них оказался ложной переменной, появившейся в результате некорректно выполненного объединения операций, зато вторым и был тот человек, который нас интересовал»<sup>[84]</sup>.

**Результаты и необходимые меры.** У Симона Хайнеса, найденного сложными методами сетевого анализа, был проведен обыск. В его доме обнаружили улики, достаточные для предъявления обвинения по нескольким статьям уголовного кодекса. Состоялся суд, и жюри присяжных признало его виновным. Поданная апелляция была отклонена. Хайнес провел два с половиной года в тюрьме и выплатил 100 тысяч долларов штрафа. Кроме того, у него конфисковали всю прибыль, полученную в результате незаконной сделки с использованием инсайдерской информации.

Вьель и сейчас использует те же методы сетевого анализа при расследовании случаев мошенничества.

## ГЛАВА 6

# Развитие навыков количественного анализа

От чего зависит судьба человека или, если хотите, линия его жизни? Давным-давно Аристотель сказал, что привычка — вторая натура. «Нравственные, или этические, добродетели (добродетели характера) рождаются из привычек-нравов: человек действует, приобретает опыт, и на основе этого формируются черты его характера»<sup>[85]</sup>, — писал он. Значит, привычки и определяют нашу судьбу. Можно изобразить это следующим образом:

Мысли → Действия → Привычки → Характер → Судьба

Ваш привычный образ мыслей предопределяет ваши действия. Ваши действия вырабатывают привычки. Другие оценивают вас исходя из того, как вы имеете обыкновение поступать. Наконец, мнение окружающих о вас определенным образом влияет на вашу судьбу. Процесс, в результате которого вы можете стать квалифицированным количественным аналитиком, включает примерно такие же этапы (рис. 6.1).

Рис. 6.1. Как стать квалифицированным количественным аналитиком



Привычка подходить к решению проблем с количественно-аналитической точки зрения формирует ваши навыки. Если вы постоянно работаете с числами, то со временем станете квалифицированным количественным аналитиком. Знания в области количественного анализа постоянно расширяются, поскольку используются в процессе реализации навыков и подходов к нему. Далее мы поговорим о том, какие навыки и подходы к количественному анализу надо освоить, чтобы стать квалифицированным специалистом в этой области, а также о том, какие знания и умения необходимы на каждом из этапов.

## Количественный подход

В то время как знание методов количественного анализа, безусловно, полезно при принятии аналитических решений, не менее важно выработать в себе количественный подход к решению проблем, то есть стремление оперировать количественными показателями и находить надежные аргументы в поддержку той или иной гипотезы. Если ваш мозг не выключается при виде цифр или уравнения, то вы сами удивитесь, сколько процедур количественного анализа можно освоить.

## Не бойтесь цифр

Мы давно подозревали, что аллергия на цифры поражает многих компетентных и разумных людей. Бывает, что они испытывают непреодолимое отвращение к математике. Это не только печально, но и контрпродуктивно. Тем не менее математика — это не главное для количественного образа мыслей. Хорошего количественного аналитика отличают не столько познания в области чистой математики, сколько подход к классификации количественной информации. Количественно мыслящие люди обладают определенными навыками, способностями и привычками, которые помогают им принимать решения, основываясь на ней. Некоторые специалисты даже утверждают, что полученных в средней школе математических знаний вполне достаточно, чтобы научиться мыслить количественно: «Математические знания и количественное мышление — это разные вещи... Очень важно это понимать, поскольку для того, чтобы стать квалифицированным

количественным аналитиком, требуется не так уж много знаний по математике, выходящих за пределы программы средней школы»<sup>[86]</sup>.

Если не верите, вот пример Эмили Роуз. Она в девять лет уже провела эксперимент в духе «сумасшедшего ученого», анализируя концепцию бесконтактного массажа. Ее работа оказалась достаточно сложной и актуальной, чтобы получить право на публикацию в профессиональном журнале *Journal of the American Medical Association*<sup>[87]</sup>. Выходящий постоянно с 1883 года, JAMA имеет один из самых крупных тиражей и принадлежит к числу наиболее престижных медицинских журналов в мире. Как научная работа четвероклассницы могла попасть в такой журнал? Эмили видела, как ее мать Линда смотрит видеофильм о набирающем популярность виде медицинских услуг — *бесконтактном массаже* (БМ). Это неоднозначный способ лечения недомоганий путем управления «энергетическими полями» пациента. Поборники БМ утверждают, что уже более 100 тысяч человек во всем мире овладели его техникой. БМ преподается более чем в ста колледжах и университетах, расположенных в 75 странах. Говорят, что эта процедура получила наибольшее признание среди процедур, практикуемых сторонниками холистической медицины. Во время сеанса врач делает пассы руками в нескольких сантиметрах от тела пациента, чтобы выявить и устранить в его организме «застой энергии», способный послужить причиной различных заболеваний. Эмили предложила своей маме провести эксперимент. Линда, медсестра по профессии, дала дочери полезные советы по его методике. Проведя некоторые изыскания, Линда и Эмили поняли, что еще никто не пытался установить, действительно ли специалист по бесконтактному массажу способен обнаруживать «энергетические поля» человека.

Эмили сосредоточила свои усилия на поиске ответа на этот вопрос. Если практикующие бесконтактный массаж действительно способны делать то, что обещают, то они должны хотя бы чувствовать «энергетическое поле» пациента. Если они на это не способны, тогда медицинская ценность этого способа лечения по меньшей мере сомнительна. Таким образом, если процент правильно выявленных бесконтактным массажистом «энергетических полей» окажется

существенно выше уровня, объясняемого случайностью, то заявления о преимуществах этого вида лечения имеют под собой основание.

Изучив газетные объявления и другие источники информации, Эмили обнаружила 25 физиотерапевтов, практикующих бесконтактный массаж в северо-западном Колорадо; двадцать один из них согласился участвовать в эксперименте. Потенциальных участников предупредили, что исследование будет представлено на ярмарке научных проектов школьников. Во время тестирования физиотерапевт располагал руки ладонями вверх на плоской поверхности, примерно в 25–30 сантиметрах друг от друга. Чтобы тестируемый не видел своих рук, перед его лицом размещался овальный непрозрачный экран с вырезами у основания; к экрану крепилось полотенце, покрывавшее его руки. Эмили подбрасывала монетку, чтобы определить, какая рука будет целевой (этот метод известен как рандомизированное распределение по группам в количественном эксперименте). Затем она протягивала правую руку ладонью вниз в восьми-десяти сантиметрах над целью и говорила «Готово!». После этого физиотерапевт говорил, какая его рука находится ближе к руке Эмили. Этот тест повторялся 280 раз при участии 21 физиотерапевта.

Хотя все участники утверждали, что они способны почувствовать руку пациента, лишь в 122 случаях (44 процента) из 280 они правильно ответили на поставленный вопрос. Это даже хуже, чем если бы они отвечали на вопрос просто наугад (см. [сайт этой книги](#)), ведь в таком случае процент угаданных правильно ответов составил бы около 50. Эмили сделала вывод, что физиотерапевты не смогли подтвердить наиболее важный навык БМ, что говорит о беспочвенности их заявлений. Дальнейшее использование БМ в медицинских целях, таким образом, неоправданно. В апреле 1998 года одиннадцатилетняя Эмили (учась в шестом классе) опубликовала результаты эксперимента в JAMA. Редактор журнала Джордж Лундберг сказал, что эксперты по статистике были восхищены простотой эксперимента и очевидностью его результатов<sup>[88]</sup>. Эмили была занесена в *Книгу рекордов Гиннеса* как самый молодой автор исследования, опубликованного в ведущем научном журнале.

Эмили отнюдь не гений и не вундеркинд; просто она обладает здравым смыслом и ясным количественным мышлением. Наиболее ценный навык квалифицированного количественного аналитика — это умение извлекать информацию из баз данных, что зависит совсем не от математической подготовки, а от количественного мышления. Первый шаг в этом направлении состоит в том, чтобы преодолеть страх перед числами и чувствовать себя комфортно в цифровой среде. Вы можете быть уверенными, что необходимые познания в математике у вас наверняка есть или, в крайнем случае, их можно легко получить. Главная черта квалифицированного количественного аналитика — это умение обращаться с данными и не испытывать перед ними страха.

Даже если вы не демонстрировали особых математических успехов в школе, это никогда не поздно исправить. Кто сказал, что вы до конца жизни обречены плохо разбираться в числах? Если в школе вам не давалась математика, это не значит, что вы не можете ликвидировать этот пробел в дальнейшей карьере. Например, Ди-Джи Патил, придумавший термин «ученый по данным», чтобы описать свою роль в управлении аналитическими исследованиями в интернет-компании LinkedIn, в школьные годы немало пострадал от математики: «Впервые я завалил экзамен по математике в восьмом классе. То же самое периодически случалось в старших классах, поэтому пришлось подать заявку на повторный курс. Но и его я завалил, в результате чего с трудом получил аттестат о среднем образовании. Сразу в университетский колледж с такими баллами поступать было нельзя, поэтому пришлось пойти в местный колледж низшей ступени... Там я записался на курс по математике, благодаря чему смог перейти в университетский колледж. Там стало ясно, что я не понимаю ничего из того, что говорят преподаватели. Я учился в слишком слабом с точки зрения математической подготовки классе. Было очень стыдно, поэтому я пошел в библиотеку»<sup>[89]</sup>.

Патил взял в библиотеке несколько книг и за одни выходные самостоятельно изучил школьный курс математики. В итоге он добился неплохих успехов и был принят в колледж университета в Сан-Диего, где окончил университетский курс математики за три года. Потом он получал степень PhD по прикладной математике в

Мэрилендском университете.

Вначале он столкнулся с некоторыми трудностями. «Я получил образование в США, а конкурировать приходилось с русскими, израильтянами и корейцами. На первом экзамене я катастрофически провалился, получив, кажется, второй балл с конца списка. Самый низкий балл был у студента, не явившегося на экзамен». Зато на второй попытке наш герой набрал самые высокие баллы. Окончив докторантуру, Патил начал преподавать в Мэрилендском университете, одновременно занимаясь исследованиями по моделированию погоды. Ему пришлось поработать и на правительственные разведывательные организации. Финансирование в то время было ограниченным, и он уволился, перейдя работать в Skype, а затем в eBay. Потом он стал ведущим аналитиком в LinkedIn, где люди, занимающие эту должность, имеют огромное влияние на процесс разработки продуктов.

Сейчас Патил — «исследователь данных, проживающий по месту службы» (наверно, первый человек, чья должность носит такое название) в венчурной фирме Greylock Partners; он помогает клиентам компании решать проблемы с данными и аналитикой. Возможно, он представляет собой лучший пример человека со скрытыми математическими способностями.

## **Используйте поиск в интернете для обнаружения неизвестных концепций и понятий, связанных с вашими данными**

В эру глобализации экономические и деловые термины мелькают в новостях каждый день. То же можно сказать о повседневных разговорах. Поскольку многие незнакомы с общеупотребительными понятиями и характеристиками, они просто пропускают их мимо ушей. Но если вы собираетесь стать количественным аналитиком, то вам придется запоминать их, записывать, а впоследствии пытаться при помощи Google найти их значение. Большой объем информации такого рода есть в Википедии; существуют онлайн-курсы и электронные учебники по многим предметам. Распечатайте результаты поиска и систематизируйте их в отдельный файл для последующего изучения.

Это хороший способ учиться и с течением времени преодолеть страх перед числами. Вероятно, вы не поймете всего, что прочитаете, но тем не менее чему-то научитесь. Если проявлять упорство в течение хотя бы шести месяцев, то вы немало удивитесь, когда узнаете, что ваши коллеги считают вас весьма информированным человеком, особенно в том, что касается чисел.

## Проявляйте любопытство

Если вам действительно интересны числа, то со временем вы будете понимать их все лучше и лучше. Например, когда вы слышите, что Усейн Болт самый быстрый человек в мире, у вас возникает вопрос, каково его лучшее время. Несложно выяснить, что на стометровке оно составляет 9,58 секунды. Теперь задайте себе вопрос, чему равна его скорость в пересчете на километры (или мили) в час. Проведя несложные расчеты, вы обнаружите:

9,58 секунды  $\approx$  10 секунд = 1/6 минуты = 1/360 часа.

100 м = 0,1 км.

Таким образом,

$$\frac{0,1 \text{ км}}{\frac{1}{360} \text{ часа}} = 36 \text{ км/час.}$$

Теперь у вас возникает вопрос, чему равен мировой рекорд в марафоне: 2 часа 3 минуты 59 секунд. Простой расчет показывает, что для этого марафонец должен бежать со скоростью около 20,4 километра в час (12,7 мили в час). Сравнив эти две средние скорости, вы поймете, как быстро может бежать человек на самой короткой и самой длинной олимпийской дистанции. Любопытство во всем, что касается чисел, — вот отличительная черта хорошего количественного аналитика.

## Количественный подход ↔ Количественные знания

Чтобы расширить свои познания в количественном анализе, на следующем этапе полезно ознакомиться с начальным курсом

статистики для широкого круга читателей. Одна из наиболее популярных книг такого рода — уже упоминавшаяся нами «Как обмануть с помощью статистики» (How to Lie with Statistics) Даррела Хаффа. Это наиболее продаваемая книга по статистике во второй половине XX века<sup>[90]</sup>. Она представляет собой краткое, увлекательное, иллюстрированное изложение наиболее часто встречающихся ошибок, как намеренных, так и непреднамеренных, имеющих отношение к статистике и способных привести к неверным выводам. Это действительно прекрасная книга, рассказывающая о том, как можно лгать языком статистики и как проверить достоверность статистических данных. Можно порекомендовать также опубликованную в 1974 году книгу Стивена Кемпбелла «Недостатки и заблуждения в статистическом мышлении» (Flaws and Fallacies in Statistical Thinking). С учетом даты выхода это удивительно полезная книга для всех, кто читает газеты и обращает внимание на содержащиеся в них статистические данные<sup>[91]</sup>. Гораздо более новая книга на ту же тему — «Распространенные ошибки в статистике и как их избежать» (Common Errors in Statistics and How to Avoid Them) Филипа Гуда и Джеймса Хардина<sup>[92]</sup>. Таких книг множество, почитайте рецензии и купите лучшие.

## Мыслить вероятностно

Ключевой аспект количественного мышления — это понимание законов вероятности и случайности. Отсутствие такого понимания — наиболее серьезный недостаток мышления взрослого человека, который, к сожалению, отмечается у большинства. Нассим Николас Талеб считает, что большинство из нас «одурачены случайностью» (такое же название носит его книга<sup>[93]</sup>), ведь мы придаем огромное значение вещам, которые вполне могут оказаться случайностью. Например, это может быть совпадение дней рождения у кого-то из гостей на вечеринке. Допустим, что на вечеринке присутствуют 23 гостя, и вдруг двое из них обнаруживают, что родились в один день. Они могут задаться вопросом о том, насколько вероятно такое событие. Оказывается, вероятность достаточно велика: более 50 процентов (см. <http://keepingupwiththequants.weebly.com>).

Понимание законов вероятности весьма полезно не только на вечеринке, но и во множестве других случаев. Если вы их не знаете, то не сможете понять, что фондовый рынок не что иное, как *прогулка среди случайностей* (колебания цен на акции не подчиняются каким-либо закономерностям), и некоторые лидеры по темпам роста в течение нескольких лет подряд могут показывать гораздо лучшие результаты, чем средние компании, но рано или поздно произойдет обвал их показателей. Вам не обязательно понимать феномен возврата к норме: например, если ваш доход намного превышает средний уровень, то доход вашего ребенка, скорее всего, будет ниже, чем ваш. Возможно, то, что вы не знакомы с теорией вероятности, обойдется вам в кругленькую сумму, если вы частый гость в казино Лас-Вегаса. Вам также будет трудно получить работу в страховом бизнесе или решить, стоит ли покупать аннуитет. И конечно, вся статистика основана на вероятности. Короче говоря, теорию вероятности следует понимать не только для того, чтобы стать количественным аналитиком, но и для того, чтобы добиться успеха в жизни.

Конечно, существует много источников информации по теории вероятности. В учебниках обычно она рассматривается вместе с основами статистики, поскольку там требуется умение мыслить вероятностно. Если вы хотите сконцентрироваться именно на теории вероятности, а не на статистике в целом, то лучше выбрать книгу Ричарда Исаака «Удовольствия от вероятностей» (*The Pleasures of Probability*)<sup>[94]</sup>. Если вы терпеть не можете учебники, то забавное и хорошо написанное введение в статистику для широкой публики содержится в книге *The Drunkard's Walk: How Randomness Rules Our Lives*<sup>[95]</sup> Леонарда Млодинова.

## Снова в школу

Если вам трудно разбираться с аналитикой, теорией вероятности и количественным мышлением самостоятельно, то совсем не обязательно это делать. Существует много возможностей дистанционно и очно изучить их. Университеты все чаще размещают учебные материалы в открытом доступе, в том числе и разнообразные курсы по статистике. Например, если вы хотите поучиться в

университете, который известен высоким уровнем преподавания и исследовательской работы в области количественного анализа, то зайдите на страничку курса по статистике и теории вероятности на сайте Массачусетского технологического института<sup>[96]</sup>. Если речь идет о платном дистанционном образовании, то можно приобрести у издательства Harvard Business Publishing курс-самоучитель по количественным методам, разработанный Яном Хаммондом, коллегой Тома Дэвенпорта по Гарвардской школе бизнеса (это стоит чуть больше сотни долларов). Если вы планируете потратить более крупную сумму, можно записаться на дистанционное обучение на степень магистра по предсказательной аналитике в Северо-Западном университете.

Если вы предпочитаете общаться с профессорами лицом к лицу, иметь возможность обсудить с ними проблемы и извлекать пользу из повседневного общения с однокурсниками, то для вас в американских университетах есть все растущее число программ на степень по аналитике, обычно магистерскую. Их прародительница — основанная пять лет назад магистерская программа по аналитике в Университете Северной Каролины<sup>[97]</sup>, одна из лучших в этой области. Почему именно Северная Каролина? Основатель и генеральный директор компании по разработке аналитического программного обеспечения SAS Джим Гуднайт учился именно в этом университете и любезно пожертвовал деньги, чтобы основать эту программу.

Магистерская программа по аналитике — это программа подготовки профессионалов с глубоким пониманием инструментов, методов, приложений и практического применения. Принципы ее организации, по нашему мнению, в полной мере соответствуют задачам и потребности компаний и бизнеса в целом в количественно ориентированных специалистах.

- *Короткая по времени.* Десять месяцев интенсивного обучения, разбитых на три семестра (лето, осень и весна). Зачисление производится в июле, выпуск в мае следующего года. Заочной и вечерней формы обучения нет.
- *Напряженный график обучения.* Обучение в течение полного дня (понедельник — пятница с 9:00 до 17:00) в кампусе.

Комплексный учебный план для группы студентов; предусмотрена работа в группах; во внеаудиторное время планируется работа над проектами.

- *Широкое и рассчитанное на практическое применение содержание курсов.* Комплексный, мультидисциплинарный учебный план (разработанный на основе опыта кафедр и колледжей Северной Каролины) направлен на обучение практическим навыкам, пригодным для решения актуальных проблем в статистике, прикладной математике, компьютерных науках, исследовании операций, финансах и экономике, маркетинге.
- *Обучение на собственном опыте.* Использование практикумов вместо стандартных форм обучения (студенты работают в группах по пять человек, решая реальные проблемы на основе заданий и данных, предоставленных промышленными спонсорами; напряженная работа в течение семи месяцев завершается отчетом перед спонсором).

Магистерская программа по аналитике Университета Северной Каролины основана на инновационном учебном плане, состоящем из специально разработанных для нее курсов. Они посвящены интеллектуальному поиску данных и анализу текстов, прогнозированию, оптимизационным задачам, базам данных, визуализации данных, безопасности баз данных, финансовой и потребительской аналитике. Студенты приходят в программу с разным базовым образованием и опытом, хотя предпочтительна определенная ориентация на количественные исследования. Средний возраст студентов 27 лет, примерно 26 процентов уже имеют университетское образование. Примерно половина студентов до поступления на программу работала на полную ставку. Несмотря на сложную ситуацию и медленный рост в национальной экономике, выпускники программы 2011 года поставили рекорд по количеству интервью с работодателями — 469 (то есть на каждого выпускника пришлось по 12 интервью; в 2012–2013 годы численность студентов возросла до 80 человек). Все выпускники обеспечены заявками работодателей на пять лет вперед. Учитывая, что спрос работодателей на этих выпускников

постоянно растет, вполне понятно, что аналогичные программы открываются и в других университетах. Одно из недавних исследований выяснило, что 59 университетов предлагают программы по деловой аналитике или бизнес-разведке, в том числе 37 магистерских и 22 бакалаврские программы<sup>[98]</sup>. Школы также начинают предлагать курсы по изучению данных, а в скором времени надо ожидать и программ с выдачей диплома.

## Количественные навыки

Количественный подход важен, но не менее важны и количественные навыки. Говорят, что легче привести свои действия в соответствие с новым способом мышления, чем придумать новый образ действий. Если постоянно тренировать количественные навыки, то через некоторое время окажется, что у вас выработался количественный подход.

## Требуйте чисел

Хорошие количественные аналитики (а также организации, стремящиеся воспитать их в своем коллективе) всегда должны требовать количественную информацию, когда им представляют идеи, теории и случайные наблюдения. Приучитесь задавать вопрос: «Есть ли у вас данные в поддержку этой гипотезы?» Ну а если вы действительно круты, то периодически вставляйте фразу «Множество слухов — это еще не данные». Требование данных играет большую роль, поскольку позволяет ответить на вопросы: «Как человек мыслит?», «Какие инструменты и процессы помогли обеспечить это направление действий?» В дополнение к глобальным задачам спасения мира попытайтесь бороться с желанием перескакивать сразу к выводам, не требуя представления данных.

То же справедливо при презентации ваших идей. Если никаких данных нет, проведите хотя бы небольшой эксперимент в стиле «сумасшедшего ученого» и соберите их. Постоянно стремитесь собрать как можно больше данных до того, как начнете развивать свои теории. Это поможет поставить их на надежную основу количественного анализа и одновременно увеличит шансы убедить слушателей в правильности

ваших идей. Навык опираться на цифры и подтверждать ими свои идеи абсолютно необходим любому, кто хочет стать квалифицированным количественным аналитиком.

## Никогда не доверяйте числам

Только что мы писали о том, что аналитик всегда стремится собрать числа и прочие данные в поддержку своих теорий и взглядов на мир. Но сейчас мы хотели бы поговорить о необходимости разумно-скептического подхода к ним. Как и в случае с новым знакомым, не доверяйте данным до тех пор, пока не узнаете о них больше. Никогда не принимайте за чистую монету никакие числа, которые вам представили. Существует несколько причин, по которым данные не заслуживают доверия. Люди (а особенно политики) иногда лгут и мошенничают, прикрываясь цифрами, или в крайнем случае неправильно интерпретируют их, стремясь добиться своих скрытых целей. Шотландский поэт и критик Эндрю Ланг писал: «Он использует статистику, как пьяница использует столб — больше в качестве опоры, а не источника света»<sup>[99]</sup>. Числа могут быть неправильно отобранными, и в таком случае они не отражают состава генеральной совокупности. Критичное отношение к данным оправданно, особенно если они нетипичны и вызывают удивление. Лучший способ получить надежные данные — относиться к ним критично и стараться узнать как можно больше об их происхождении. В частности, скептицизм относительно данных должен включать следующие аспекты.

**Релевантность.** Представленные данные должны иметь непосредственное отношение к проблеме, для решения которой они собирались. Они должны быть репрезентативными по отношению к той группе или организации, которую они предположительно представляют. Если данные не дают хотя бы некоторых ответов на вопросы, они бесполезны.

**Точность.** Если данные релевантные, но неточные, их необходимо отбросить. Точность данных можно оценить, изучив вопрос о том, кто и как их готовил. Если данные не проходят эту проверку на точность, они опять-таки бесполезны.

**Правильная интерпретация данных.** Даже точные данные могут

ввести в заблуждение, если их неправильно истолковать. Особенно склонны неверно интерпретировать данные те люди, у которых есть скрытые мотивы и цели. Рассмотрим пример, в котором интерпретация данных способствовала формированию субъективного мнения.

Критик из журнала Newsweek, писавший рецензию на книгу «Лучшая половина» (The Better Half) о первых суфражистках, завершил свой опус несколько провокационно. Он задал риторический вопрос, что сказали бы Сьюзан Энтони и ее подруги, если бы узнали о том, что пятьдесят лет спустя после предоставления американским женщинам политических прав социологи Колумбийского университета обнаружили: лишь одна из двадцати двух женщин голосовала не за того кандидата, за которого голосовал ее муж. Один из читателей в ответ на это написал: «Я думаю, что суфражистки были бы весьма довольны. Их движение проделало большую работу, если менее чем через пятьдесят лет после предоставления женщинам политических прав только один муж из двадцати двух имел мужество голосовать не так, как его жена» <sup>[100]</sup>.

Таким образом, всегда стоит задавать себе вопрос, насколько правильна интерпретация данных с учетом проблем и мотивов того лица, которое ее представило.

## Особенная осторожность с причинно-следственным аргументами

Стоит с особой осторожностью относиться к аргументации, основанной на причинно-следственных связях: их выявить очень сложно. Как мы уже упоминали во вставке об экспериментах «сумасшедшего ученого», если вы формируете контрольную и целевую группу, случайным образом распределяя в них людей, и эти группы демонстрируют различные результаты, то обычно аналитик приписывает это воздействию тестируемого фактора. Но если вы просто выявили статистическую связь между двумя факторами, вряд ли она окажется причинно-следственной. Возможно, вы слышали фразу «корреляция — еще не причина». Это важно помнить.

Когнитивные психологи Кристофер Чабрис и Даниэль Симонс предложили эффективную процедуру для выявления причинно-следственной связи в книге «Невидимая горилла и другие способы нашей интуиции обмануть нас» (The Invisible Gorilla and Other Ways Our Intuitions Deceive Us): «Когда вы слышите или читаете о наличии связи между двумя факторами, задумайтесь, можно ли говорить о том,

что в тестовую группу для их проверки люди отбирались действительно случайно. Если это невозможно, слишком дорого или этически неприемлемо, то проводить эксперимент нельзя и причинно-следственная связь считается неподтвержденной»<sup>[101]</sup>.

Например, вы прочитали в газете: «В ходе десятилетнего эксперимента доказано, что запойное пьянство приводит к раку». Задумайтесь над тем, возможно ли в данном случае случайное распределение участников на тестовую и контрольную группы с последующей просьбой к одним запойно пить, а к другим — соблюдать трезвость в течение десяти лет. Наверно, нет. Куда более вероятно, что исследователь обнаружил корреляционную зависимость между запойным пьянством (по всей видимости, по собственным словам опрашиваемого) и случаями рака в группе населения, которую мониторили в течение десяти лет. Возможно, исследователь учитывал, что выявленная корреляция может объясняться и другими факторами (например, сильно пьющие люди часто курят), но уж репортер точно об этом не задумывался.

Если вы заподозрили, что кто-то в вашей организации является сторонником софизма *cum hoc ergo propter hoc* (после этого — значит по причине этого), то для предотвращения неправильных выводов следует тщательно следить за разработкой условий экспериментов, хорошо знать статистику и эконометрику. Именно в этом случае лучше пригласить эксперта.

## Задавайте вопросы

Задавать вопросы стоит для того, чтобы лучше понять суть проблемы и оценить методику ее решения. Аналогично, если у вас возникли сомнения по поводу представленных данных, следует без колебаний задавать вопросы. Многие стесняются задавать вопросы по поводу чисел, поскольку боятся показаться глупыми. Эти страхи преувеличены. Некоторые идеи по поводу того, какие вопросы лучше задавать, приведены во вставке «Хорошие вопросы о количественном анализе».

Хорошие вопросы о количественном анализе

Перечень приведенных далее вопросов, конечно, не исчерпывающий, но может помочь вам с чего-то начать. Они пригодны практически для любых видов количественного анализа или такого неколичественного, который должен стать количественным.

- Есть ли у вас данные в поддержку вашей гипотезы?
- Что вы можете сказать об источнике данных, использованных в анализе?
- Вы уверены, что выборка репрезентативна для генеральной совокупности?
- Присутствуют ли какие-либо выбросы в распределении данных? Как они влияют на конечный результат?
- Какие предположения вы положили в основу своего анализа?
- Есть ли какие-либо обстоятельства, способные повлиять на достоверность предположений и модели анализа?
- Почему вы выбрали этот подход к анализу?
- Какие преобразования данных вы провели, чтобы добиться соответствия между ними и моделью?
- Рассматривали ли вы какие-либо иные подходы к анализу данных, и если да, то почему от них отказались?
- Насколько вероятно, по вашему мнению, что независимые переменные действительно обуславливают динамику зависимых переменных? Можно ли провести еще какие-либо аналитические процедуры для подтверждения причинно-следственной связи?

Представьте, что кто-то задает вопросы по поводу представленных данных. Такой человек скорее будет выглядеть смелым и заслужит уважение, чем покажется глупым. Поэтому, если вам встретились непонятные числа, отважно задавайте вопросы. Более того, в некоторых случаях вопросы ожидаемы. Например, если вам показывают средние значения, спросите о дисперсии, или распределении, или стандартном отклонении. Вы можете также спросить, присутствуют ли в данных какие-либо выбросы, являющиеся результатом ошибок или экстраординарных событий, а также не пропущены ли данные, которые могут оказаться значимыми. Вы можете спросить, чему равна медиана, если представлены средние значения. Интерпретировать среднее значение, не имея данных о его дисперсии, опасно, поскольку результаты отдельных наблюдений могут существенно отличаться друг от друга. Кроме того, если кто-то говорит о данных из конкретного обследования, следует спросить, кто

проводил это обследование, какова его методика и как были сформулированы вопросы. Говорить о результатах обследования, не зная ответы на эти вопросы, нельзя. Коротко говоря, имеет смысл приобрести привычку задавать вопросы и исследовать данные: это очень важно для выработки навыков количественного анализа.

## Практика количественного анализа

Подобно молодому человеку, спрашивавшему, как найти Карнеги-Холл в Нью-Йорке, если вы хотите выработать навыки, искусство и дисциплину хорошего аналитика, вам следует практиковаться, практиковаться и еще раз практиковаться. Возможно, достигнуть прогресса удастся не сразу. Но трудности неизбежны в любом обучении, и только преодолевая их, вы сможете усовершенствовать свои навыки количественного аналитика — медленно, но неуклонно. Работу эту можно представить в виде тех же шести шагов решения любой проблемы, о которых мы говорили на протяжении всей книги.

**Определение и формулирование проблемы.** Вероятно, у вас на работе немало проблем. Определите их приоритет исходя из срочности и применимости методов количественного анализа для их решения. На этапе определения проблемы самое важное понять, в чем суть проблемы и почему необходимо ее решить. Ответы на эти два вопроса не только дают представление о преимуществах решения проблемы, но и помогают спланировать следующие этапы.

**Обзор предшествующих исследований.** Если проблема определена, то следует изучить все имевшие место попытки ее решить. Хотя, как правило, этот этап требует довольно много времени, но поисковики вроде Google обычно весьма полезны. Поиск относящейся к проблеме информации очень важен для правильной оценки ее масштаба и для подбора определяющих переменных. Если вы глубоко разобрались в наработках предшественников, то получите более ясную картину того, как можно проблему решить. Это можно сформулировать так: «Проблема сформулирована правильно, предыдущие попытки изучены, значит, полдела сделано».

**Моделирование (выбор переменных).** Если очертить круг определяющих переменных удалось еще на этапе изучения предыдущих поисков, то на этом этапе остается отбросить те переменные, которые не связаны непосредственно с поиском ответа на поставленный вопрос. То, какие переменные отбросить, а какие оставить, зависит главным образом от целей построения модели. Если вы хотите собрать игрушечный поезд, то размеры и пропорции поезда реального имеют большое значение. Если же вы хотите оценить экономическую эффективность поезда, то на первый план выдвигаются такие параметры, как скорость, грузоподъемность, потребление топлива. Практикуясь в сосредоточении внимания на отдельных особенностях реального объекта (как это делает карикатурист), вы постепенно нащупываете решение, а ваши количественные навыки совершенствуются.

**Сбор данных (измерение).** Прежде чем собирать данные по отобранному переменным, надо проверить: а не делал ли этого кто-либо ранее? Очень часто оказывается, что в соседнем департаменте, а иногда и в вашем собственном такие данные уже собирались. Не исключено, что их можно найти в открытом доступе. Даже если придется их купить, это иногда обходится дешевле, чем собирать самому. Если данных, непосредственно относящихся к вашей проблеме, нет, то следует определиться с методологией их сбора. Если вы решили провести опрос, то содержание анкеты и даже формулировку отдельных вопросов необходимо тщательно изучить. Если решено провести эксперимент, то лучше пригласить специалиста для консультаций по методике его проведения. Получение точных и актуальных данных для анализа имеет огромное значение для конечного результата, поэтому на данном этапе не стоит жалеть времени и усилий.

**Анализ данных.** Анализ данных заключается в поиске устойчивой модели взаимосвязей между ними или между переменными. Статистические методы для анализа подбирают на этапе определения проблемы: после того как будет определена суть проблемы, выбор наиболее подходящего метода анализа становится очевидным. Если

для решения проблемы необходимо провести сравнение между группами данных или показателей, то потребуется соответствующая процедура. Если проблема во взаимной связи переменных, то логично выбрать процедуру регрессионного анализа или ей подобную. Поскольку все эти аналитические процедуры широко используются в самых разных ситуациях, имеет смысл потратить время на овладение их теоретическими и практическими аспектами. Наверняка в вашей организации найдется специалист, способный помочь вам советом в этом деле.

**Результаты и необходимые меры.** Не пренебрегайте этим важным этапом количественного анализа. Успешные аналитические команды уделяют ему столько же времени и внимания, сколько предыдущим пяти. Попросите коллег в компании поделиться опытом демонстрации результатов анализа, обсудите с ними ваши идеи и проблемы. Освойте язык визуализации статистических исследований, например найдите постоянное место на рабочем столе книгам Эдварда Тафти. А если сам Тафти приедет в ваш город с однодневным семинаром «Представление данных и информации», не упустите случай пойти.

## Количественные навыки ↔ Количественные знания и методы

На этом этапе следует ознакомиться с теоретическими основами количественного анализа. В первую очередь стоит обратить внимание на такие курсы, как «Элементарная статистика» и «Методы исследования». Если вы сможете записаться на них дистанционно, это будет прекрасно. Можно поэтапно осваивать их в интернете или с использованием обычных учебников каждый раз, когда появится требующая решения практическая задача. Многие испытывают трудности, осваивая понятия элементарной статистики, поскольку в некоторых учебниках ничего не пишут об их применении в реальном мире. Поэтому следует очень тщательно подходить к выбору учебника, содержащего информацию о прикладных аспектах применения статистических методов в тех или иных реальных ситуациях. Одной теории тут мало. В качестве основного мы рекомендуем учебник

Хайнца Кохлера Statistics for Business and Economics<sup>[102]</sup>. Один из рецензентов так отозвался о нем на Amazon.com:

Лучший учебник по статистике всех времен и народов!

Грядущий экзамен по статистике пугал меня до смерти. Было такое чувство, что освоить все эти понятия и выводы будет невероятно трудно. Но компоновка материала и структура этой книги настолько удобны для изучения... Статистика никогда больше не будет проблемой для меня. Изумительно, как Кохлеру удалось превратить статистику в легкий и очень интересный предмет<sup>[103]</sup>.

То же самое можно сказать и о методах исследования: по ним написано множество учебников и рецензий. Самое важное в изучении этих базовых курсов — постараться решить приведенные в учебнике примеры. Поскольку большинство задач предназначено для иллюстрации связи теории с практикой, теоретические концепции, закрепленные решением примеров, усваиваются намного прочнее. Если вы читаете каждый из выбранных учебников не менее трех раз, то можете считать, что необходимый минимум знаний количественного аналитика вы почти освоили. Теперь остается только практиковаться (практиковаться и еще раз практиковаться).

## Количественный аналитик

В бизнесе процедуры количественного анализа осуществляются непрерывно и в социальном контексте. Обычно они охватывают не только саму процедуру анализа, но и подготовку отчета (или проведение презентации), совместную работу в сообществе аналитиков, обмен опытом на семинарах и других подобных мероприятиях, работу с экспертами. Все эти виды деятельности крайне важны для развития аналитических навыков.

## Подготовьте отчет

Первый шаг в деле реализации результатов анализа — это подготовка отчета, который расскажет «историю данных». Вы многому научились, читая учебники, но гораздо большему можно научиться, упорно стараясь решить реальную проблему, а затем готовя отчет о результатах. Если вы занялись проблемой, которую многие уже пытались безуспешно решить, то это произведет впечатление; но если

это не так, нет причин сожалеть. Уделите пристальное внимание выбору подходящей проблемы, поскольку, как говорил Вольтер, люди оценивают вас по вашим вопросам, а не по ответам. Затем педантично соблюдайте последовательность шести шагов количественного анализа и представьте результаты и предложения в форме отчета. Его подготовка преследует три основные цели:

- вы научитесь чему-то новому и усовершенствуете аналитические навыки, самостоятельно решая реальную проблему;
- вы можете внести серьезный вклад в решение проблемы;
- вы сможете создать атмосферу уважения к аналитике, побудить других мыслить и действовать аналитически.

Не ждите слишком многого от первой же попытки, но убедитесь в том, что в отчете отражены результаты предшествующих исследований, ваши собственные усилия по решению проблемы и доказательства аналитической строгости результатов. Помните о том, что слушатели будут в первую очередь оценивать способ, при помощи которого вы решали проблему, а не только ваши выводы и заключения.

## Создайте благоприятную аналитическую среду

Найдите товарища или нескольких коллег и создайте группу количественного анализа. Существует множество причин для создания такой группы: вместе изучать теорию анализа эффективнее, равно как и выполнять аналитические проекты. Группа может более успешно продвигать идеи количественного анализа. В кругу единомышленников можно обсудить приоритет тех или иных аналитических задач, обсуждать проблемы, распределять задания в соответствии с шестишаговой схемой анализа, объяснять суть операций, по очереди проводить презентации. Такие группы — самый надежный и быстрый способ освоить количественный анализ. Как сказал Генри Форд, «собраться вместе — это начало, остаться вместе — это прогресс, работать вместе — это успех». Действительно, небольшая группа людей, обладающих аналитическим мышлением, способна изменить компанию, создав аналитическую атмосферу и мотивируя других думать и действовать аналитически.

## Регулярно проводите занятия

Цель таких занятий очевидна: дать возможность представить результаты анализа членам группы, обсудить ход выполнения аналитических проектов и особенности проектов, планируемых к выполнению. Кроме того, занятия помогают поддерживать и развивать атмосферу, способствующую аналитическим изысканиям и принятию решений (на их основе). Многие компании приглашают лекторов со стороны для освещения специальных аналитических проблем. Регулярно проводимый тематический семинар полезен для всех участников с точки зрения получения новой информации, поддержания интереса к аналитическим пробелам, получения дополнительной мотивации к совершенствованию аналитических навыков. Постарайтесь сделать эти семинары доступными для всех желающих, не имеющих технической подготовки.

## Количественный аналитик ↔ Количественные знания и методы

На этом этапе главная задача состоит в обучении групп методике анализа на конкретных примерах из практической жизни. Начать можно с деловых ситуаций, приведенных в этой книге. Подготовьте материалы, разработайте методику анализа, установите очередность презентаций и обсудите детали деловых ситуаций в рамках шестишаговой последовательности анализа. Если появляется информация о новом методе количественного анализа, то это удобный случай изучить его всем вместе на конкретном примере. Можно пригласить специалиста с подробным объяснением сути метода, преимуществ и недостатков конкретных процедур — это очень эффективно. Вполне возможно, что такой специалист найдется в вашей компании или хотя бы в местном университете.

Если вы проделали все или большинство процедур, описанных в этой главе, то наверняка продвинулись по пути приобретения навыков хотя бы полупрофессионального количественного аналитика. Это, несомненно, поможет в продвижении по карьерной лестнице и расширит ваш кругозор. Кроме того, вы сможете органично влиться в глобальную трансформацию деловой и корпоративной среды, как раз

сейчас охватывающую страны мира.

## Пример аналитического мышления: как получить должность преподавателя

В университетском мире конкуренция за получение постоянной должности преподавателя на кафедре все сильнее давит на научных работников, вынуждая их как можно чаще публиковаться. Фраза «публикуйся или погибни» хорошо характеризует степень этого давления, особенно в престижных и ориентированных на научные исследования университетах. Публикации необходимы как для продвижения по карьерной лестнице, так и для сохранения занимаемой должности. Однако подготовка статьи для научного журнала, особенно ведущего в отрасли, дело трудное и отнимает много времени. Сотрудничество с коллегами в процессе подготовки повышает продуктивность работы и постепенно все больше распространяется. Типично и сотрудничество профессоров со студентами старших курсов при подготовке публикаций. Таким образом, получение стипендии на старших курсах зависит от уровня подготовки или наличия достаточных исследовательских способностей, чтобы помогать профессору в его работе.

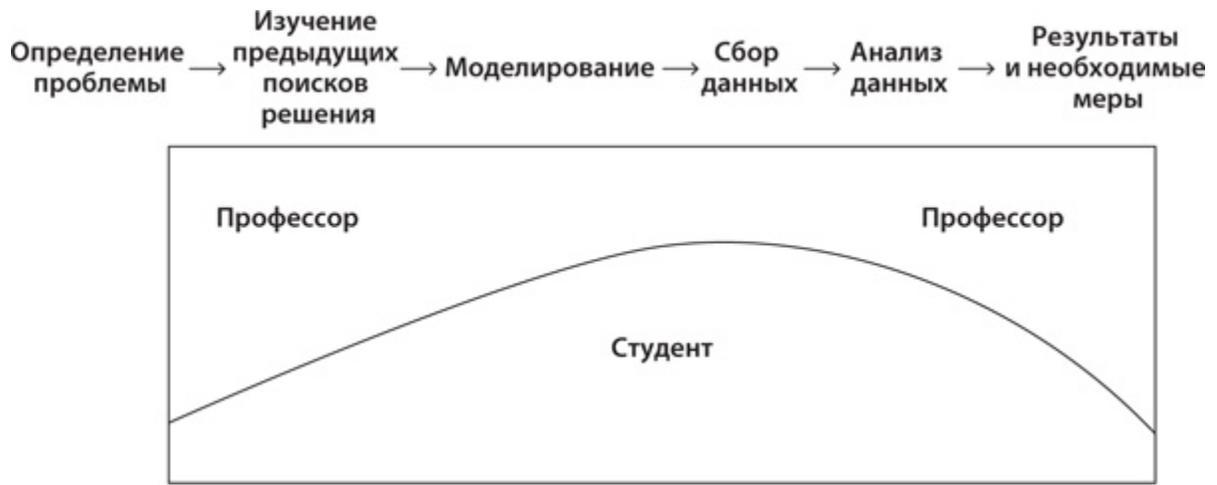
Джин Хо живет в Южной Корее, а две его дочери учатся в США. По его словам, он не мог себе позволить оплатить обучение обеих дочерей в этой стране, но зато систематически изучал все возможности получения стипендии для обучения в магистратуре. Сейчас девушки учатся в аспирантуре: в Стэнфордском и Мичиганском университетах соответственно. Сначала Джин Хо колебался, стоит ли упоминать о его семейных делах, но решил, что это важно — показать, как хорошее знание основ количественного анализа помогает получить стипендию и как окупается систематическая и осознанная работа над совершенствованием своих количественных навыков.

Когда Нури, старшая дочь Джин Хо, оканчивала колледж, она хотела продолжить образование в США. Девушка прекрасно понимала, что для того, чтобы поехать в эту страну учиться, надо получить стипендию. Но иностранным студентам, специализирующимся в области коммуникаций, очень трудно

получить ее. Джин Хо всячески поддерживал Нури, уверяя, что при достаточно высоком уровне подготовки она вполне сможет этого добиться. Нури послушалась его совета и упорно готовилась к подаче заявления на стипендию.

Но каким должен быть уровень подготовки студента, чтобы убедить стипендиальную комиссию в том, что он сможет принимать участие в научных исследованиях и потому заслуживает положительного решения? Вспомним о наших шести шагах количественного анализа и попробуем в этом контексте обсудить ведущую роль профессоров и их сотрудничество со студентами старших курсов (рис. 6.2).

**Рис. 6.2. Сможет ли Нури претендовать на стипендию? Шесть шагов количественного анализа**



На этапе определения проблемы обычно принимается решение о выборе темы исследования, которая имеет шансы на публикацию в научном журнале. Как правило, ведущую роль в этом играют профессора. На этапе изучения предыдущих поисков следует обосновать актуальность и новизну исследования по сравнению с ранее проводившимися. На этом этапе ведущую роль также играют профессора.

Они же обычно определяют, какие данные следует собрать. Студенты старших курсов помогают в сборе данных и их анализе. Как правило, эти функции поручаются студентам потому, что они требуют много времени и труда. Таким образом, шансов на получение

стипендии гораздо больше, если, во-первых, студент способен понять, почему именно эта тема выбрана для исследования и какие методы будут использованы; во-вторых, он может продемонстрировать умение собирать данные и анализировать их под наблюдением профессора. Наконец, профессора обычно пишут статью в журнал на основе собранных студентами данных.

Чтобы подготовиться к подаче заявления на стипендию, Нури записалась на вводный курс по статистике и исследовательским методам. Для углубленного изучения ключевых концепций в этой области Джин Хо и Нури провели сессию вопросов и ответов, а также обсуждение материала каждой главы базовых учебников. Она также прошла интенсивную программу изучения статистической программы SPSS, чтобы получить навыки анализа данных. Наконец, Джин Хо и Нури вместе прочли последние двадцать статей, опубликованных в ведущих научных журналах коммуникационной сферы, затем критично оценили их основные положения, методы исследования, использованные данные, методiku проведенного анализа и особенности интерпретации их результатов. Это помогло Нури получить ясное представление, как изученные в ходе трех курсов понятия и концепции применяются в реальном научном исследовании. К концу года Нури настолько овладела исследовательскими навыками, что могла и была готова ассистировать профессору практически в любом исследовании.

Свои цели и уровень подготовки Нури описала в разделе «Цели и задачи обучения» заявления о зачислении на магистерскую программу и выиграла стипендию от Висконсинского университета. Там, благодаря полученной подготовке, Нури работала с несколькими профессорами, готовя статьи к публикации. Позднее она получила приглашение в аспирантуру Стэнфордского университета и сейчас занимается подготовкой диссертации. Младшая дочь Джин Хо — Юнгри пошла по стопам сестры и получила приглашение в аспирантуру Мичиганского университета. Она работает в химико-технологической лаборатории.

**Пример аналитического мышления: Дэрил Мори и Шейн**

## Баттье из Houston Rockets

Этот пример посвящен двум менеджерам профессионального баскетбола, овладевшим количественным подходом. Дэрил Мори — генеральный директор баскетбольной команды НБА Houston Rockets. Тем, кто знаком с книгой и одноименным фильмом Moneyball, Мори известен под кличкой Билли Бин профессионального баскетбола<sup>[104]</sup>. Он окончил колледж по специальности «Статистика и компьютерные науки» на Северо-Западе США, а затем получил степень MBA от Слоановской школы бизнеса при Массачусетском технологическом университете. Он всегда стремился применить идеи Билли Джеймса о бейсбольной статистике к другим видам спорта. Затем он стал старшим вице-президентом по информационным системам и операциям клуба Boston Celtics. В 35 лет он возглавил клуб Houston Rockets и внедрил целый ряд статистических и количественных методов для улучшения результатов команды. Дэрил — председатель ежегодной Конференции по спортивной аналитике в Массачусетском технологическом университете, на которую собираются около двух тысяч участников.

Шейн Баттье — форвард НБА, в настоящее время играет за Miami Heat. В 2006–2011 годы он играл за Houston Rockets. Он отличался выдающимися аналитическими способностями как игрок и занимал седьмую позицию в рейтинге самых интеллектуальных игроков профессионального спорта по версии журнала Sporting News<sup>[105]</sup>. В статье Moneyball Майкла Левиса Дэрил Мори говорит, что Баттье

...получал специальный подбор данных. «Он единственный игрок, которому мы их предоставляли, — говорит Мори. — Они обрушивались на него, как вода из пожарного шланга, но он ухитрялся фильтровать их и извлекать полезные выводы. Большинство игроков в этом плане напоминают игроков в гольф — вы не хотите, чтобы они двигались, пока не решат, куда ударить». Данные классифицировались по ряду категорий, и рассчитывалась вероятность забрасывания мячей из разных точек поля, при разной степени активности обороны противника, при разных видах взаимодействия с другими игроками: насколько удачно он забрасывает из-за заслона, после обводки игрока соперника, броском сразу после принятия мяча и т. п. Баттье извлек много полезного из данных об игроках, которых должен был опекать на поле<sup>[106]</sup>.

Тем не менее, полезность Баттье для его команды не была бесспорной, и об этом нам говорит количественный анализ.

**Определение и формулирование проблемы.** Должна ли Houston Rockets купить такого игрока, как Баттье, даже если его индивидуальная статистика не столь впечатляющая?

**Изучение предыдущих поисков решения.** Спортивная аналитика получает все большее распространение, и существует множество книг (и еще больше сайтов) на эту тему. Но такой анализ провести намного легче в тех видах спорта, где индивидуальная статистика игрока непосредственно влияет на результат команды. В баскетболе оценить эффективность действий команды в целом и взаимодействие игроков намного сложнее. Баттье имел относительно невысокие индивидуальные показатели (в течение пяти лет выступлений за команду Memphis Grizzlies он в среднем приносил команде 10 очков за игру всего при пяти подборах мяча). Однако его команда и партнеры, казалось, действовали лучше, иногда намного лучше, когда он был на площадке.

**Моделирование (выбор переменных).** Переменными при решении вопроса о приобретении такого игрока, как Баттье, будут его стоимость (непосредственно или в обмен на других игроков), аванс, который потребуется выплатить вперед, разнообразные показатели индивидуальных действий и, в идеале, некоторые показатели эффективности деятельности команды в то время, когда Баттье находился на площадке, и в то время, когда его не было.

**Сбор данных (измерение).** Индивидуальную статистику и финансовые аспекты сделки выяснить несложно. Через них можно оценить влияние действий игрока на результаты команды. Статистика rebounds, адаптированная Роландом Бичем из [82games.com](http://82games.com) на основе аналогичной статистики в хоккее, позволяет оценить успешность действий команды в то время, когда определенный игрок находится на поле, и в то время, когда он сидит на скамейке.

**Анализ данных.** Мори и его статистики решили использовать так называемый анализ плюс-минус, чтобы оценить Баттье. Но во время сезона 2006/07, его первого сезона в Houston Rockets, Мори сказал, что

Баттѐе имел показатель  $-8$  по методике плюс-минус. Это означало, что в случае замены Баттѐе игроком со средними показателями результаты команды снижались на восемь очков<sup>[107]</sup>. По этому показателю игрок занимал семнадцатую позицию в рейтинге NBA, но при этом получал среднюю по лиге заработную плату, а тот игрок, в обмен на которого его продали, занимал в том же рейтинге лишь 45-ю позицию.

**Результаты и необходимые меры.** Конечно, Мори принял решение купить Баттѐе, и это решение оправдалось. В сезоне, предшествовавшем его появлению, Houston Rockets финишировали со счетом 34–48, а уже в следующем сезоне этот показатель составил 52–30. В 2008/09 году был поставлен рекорд — 56–27, и при этом команда продемонстрировала уникальную серию из двадцати двух побед подряд, несмотря на то что несколько ключевых игроков получили травмы. Баттѐе играл за Houston Rockets пять лет, пока не получил серьезную травму и не был продан обратно в Grizzlies, а затем в Miami Heat. За команду из Майами Баттѐе также выступал весьма удачно в своей обычной не бросающейся в глаза манере игры, в результате чего команда стала чемпионом NBA в 2012 году.

## ГЛАВА 7

# Работа с квантами

Поскольку эта книга адресована не-квантам, то мы считаем полезным описать принципы оптимальной организации совместной работы с профессионалами в области аналитики и специалистами по данным. Даже притом что вы уже многому научились, прочитав эту книгу и изучив примеры аналитического мышления из предыдущей главы, все же этого недостаточно, чтобы самостоятельно реализовать продвинутые аналитические проекты. Так или иначе, придется сотрудничать со специалистами в области количественного анализа. Количественные аналитики и специалисты по данным часто имеют ученую степень по статистике, математике или даже физике. Это косвенным образом характеризует уровень знаний и навыков квантов, необходимых для серьезной аналитической работы.

Основное внимание в этой главе мы уделим взаимоотношениям между тремя группами профессионалов, так или иначе имеющих отношение к аналитике:

- лица, принимающие коммерческие и организационные решения;
- бизнесмены или сотрудники компании;
- количественные аналитики или специалисты по данным.

Предполагаем, что большинство читателей попадает в одну из двух первых групп и должны налаживать тесное сотрудничество с представителями третьей группы. Если вы представитель третьей группы, то, возможно, тоже найдете эту главу полезной для себя, поскольку в ней содержатся рекомендации по эффективному взаимодействию с не-квантами.

В пользу того, чтобы все три группы устанавливали конструктивные взаимоотношения вместо того, чтобы перекладывать ответственность на кого-то одного, есть весомый довод. Авторы (если читатели этого еще не заметили) горячие сторонники использования аналитики в целом (и собранных данных, в частности) как основы

принятия управленческих решений. Но для многих топ-менеджеров важным фактором принятия решений по-прежнему остаются интуиция и опыт. Иногда они приводят к принятию ошибочных решений, но их полезность при выборе адекватных финансовых показателей, разработке важных деловых сценариев «что... если» и выборе условий, при которых аналитические модели релевантны, не подвергается сомнению.

Таким образом, задача заключается в том, чтобы принимать решения на основе аналитики, но с учетом интуиции менеджеров. Немногие топ-менеджеры в равной мере могут полагаться и на свою аналитическую подготовку, и на интуицию. Это означает, что придется работать в тесном контакте с количественными аналитиками, если они хотят принимать эффективные решения. По сути, мы могли бы доказать, что качество сотрудничества между топ-менеджерами и их консультантами по количественному анализу — это ключевой фактор принятия эффективных управленческих решений.

Научный сотрудник Intel Карл Кемпф (эта должность в компании предусмотрена для сотрудников, имеющих заслуги в науке и получивших право на относительную независимость в работе) возглавляет группу разработки технических решений. Он считает, что эффективные количественные решения «принимаются не на основе математики, а на основе взаимоотношений»<sup>[108]</sup>. Это весьма примечательное заявление из уст Кемпфа, известного в компании под дружескими прозвищами Суперквант и Главный математик. Если кто-то, кого, пусть и в шутку, зовут Главным математиком, заявляет, что математика тут не главное, то к этому надо отнестись серьезно.

Кемпф заметил, что математические и статистические алгоритмы, применяемые количественными аналитиками, могут быть как предельно простыми, так и очень сложными. Но в любом случае это результат тщательной работы очень умных людей в течение десятилетий (или столетий, как это было в некоторых из приведенных в этой книге примеров), который проверялся, перепроверялся и анализировался снова и снова другими очень умными людьми.

Математика работает всегда, а вот человеческие аспекты принятия решений гораздо менее отработаны. Организации, создавшие группы

количественных аналитиков (а в наш аналитический век это должно быть повсеместным явлением), должны представлять, какой именно тип аналитиков им нужен и какие задачи они будут решать для топ-менеджеров. Судя по всему, аналитики с необходимым набором навыков всегда в дефиците. Кроме того, топ-менеджерам придется пересмотреть свои ожидания относительно аналитиков и отношений с ними. Наконец, методы и инструменты, используемые аналитическими группами, следует пропагандировать и совершенствовать таким образом, чтобы они могли постепенно стать частью системы принятия решений.

## Взаимоотношения квантов и предпринимателей в принятии решений в компании Intel

Карл Кемпф и его аналитическая команда из Intel на личном опыте убедились в правильности многих изложенных в этом разделе уроков. Группа интенсивно работает над формированием конструктивных взаимоотношений между менеджерами и аналитиками. Первоочередная задача — добиться взаимного уважения: чтобы, с одной стороны, те, кто принимает решения, проявляли хотя бы минимальный интерес и уважительное отношение к навыкам и опыту количественных аналитиков, а с другой — аналитики интересовались проблемами менеджеров и серьезно относились к ним. Такая асимметрия в отношениях допускается преднамеренно. В то время как для менеджеров достаточно иметь некоторый вкус к математике, для количественных аналитиков Кемпф и его группа считают абсолютно необходимым как можно глубже понимать механизм интуиции менеджеров. Это означает, что менеджеру не обязательно становиться специалистом в математическом анализе, но аналитику можно и нужно стать экспертом по проблемам бизнеса.

Хотя количественный аналитик вряд ли сможет понять происхождение и природу деловой интуиции, но он должен понимать, что это такое, и уметь говорить на языке предпринимателя. Подход компании Intel состоит в том, чтобы периодически направлять кого-то из аналитиков в коммерческие подразделения, по крайней мере для того, чтобы слушать, наблюдать, учиться, а через некоторое время

задавать вопросы. В большинстве случаев этот человек проходит такое же обучение, как и только что принятый в это подразделение сотрудник. В обоих случаях главная цель обучения — добиться, чтобы человек понимал формальный и неформальный механизм функционирования организации, модели мотивации персонала, материального и морального поощрения и т. п.

Кемпф считает, что если количественный аналитик уверен в том, что разбирается в деловых проблемах, то это еще ничего не гарантирует. Шансы на успех повышаются, если менеджер уверен в том, что количественный аналитик разбирается в деловых проблемах. В этом случае тот завоевывает авторитет в глазах менеджеров и сотрудников компании («впервые кто-то пришел и потратил время, чтобы разобраться в наших проблемах, — этот парень, кажется, действительно заинтересован в том, чтобы помочь»). Менеджеры, в свою очередь, завоевывают авторитет в среде квантов своими математическими познаниями («это не так легко, как я думал, — должно быть, этот парень действительно умен»).

Можно предположить, что наглядно преимущества использования аналитики на практике видны количественному аналитику при непосредственном наблюдении или во время работы в составе коммерческого подразделения; это же помогает попутно достичь еще одной цели — выявить и привлечь к работе самых больших скептиков в группе. В худшем случае они окажутся правы и задачу решить невозможно; но в лучшем случае вы заранее будете знать, кто из наиболее влиятельных членов группы будет в первую очередь критиковать форму подачи результатов или принятые на их основе решения.

Энн Робинсон, ранее возглавлявшая группу аналитиков в Cisco Systems, а сейчас занимающая аналогичную должность в Verizon Wireless, также подчеркивает важность «перевербовки» скептиков в сторонников: «Они не дают вам соврать и стимулируют высокоэффективную работу всей команды. И если вы сумеете убедить их, значит, сумеете убедить кого угодно». (Об опыте работы Энн в Cisco Systems мы поговорим в конце этой главы.)

В Intel следующий этап на пути формирования отношений между квантами и менеджерами — это сотрудничество на этапе разработки

базовой модели анализа. Ведущий аналитик проводит мозговые штурмы для принятия решения о необходимых вводных условиях (исходных данных, источниках их получения, методах выявления и исключения недоброкачественных данных), результатах (предпочтительные критерии классификации данных, какие методы демонстрации результатов будут наиболее убедительны для предполагаемой деловой аудитории), а также выборе ключевых переменных и определении базовых взаимосвязей между ними.

Повторим еще раз, что в таких ситуациях менеджеру не обязательно знать методы решения, скажем, гиперболических дифференциальных уравнений в частных производных, но по крайней мере на доске в зале совещаний должна быть представлена диаграмма с результатами и решениями, отвечающими на вопросы вроде следующих:

- Поскольку  $A$  и  $X$  взаимосвязаны, то какова будет динамика  $X$  в случае роста  $A$ ?
- Каковы максимальные и минимальные значения, которые может принимать переменная  $B$ ?
- Если между причиной (фактор  $Y$ ) и наступлением следствия (фактор  $Q$ ) существует временной лаг, то какова его возможная продолжительность?

Как и в случае с любым другим типом моделей, несколько конкретных примеров (из жизни или теоретических) будут очень полезны для объяснения принципов построения базовой модели. В этом случае квант должен внимательно слушать, задавать уточняющие вопросы и стараться усвоить как можно больше деловой информации, имеющей отношение к принимаемому решению. Налаживание взаимоотношений между людьми тут так же важно, как и построение самой модели.

К этому моменту группа квантов должна быть готова выйти на сцену. Требуется выбрать правильный математический метод, формализовать модель таким образом, чтобы можно было передать ее решение компьютеру: собрать данные и ввести их в него. Затем аналитик должен протестировать модель, выполнив анализ чувствительности переменных и взаимосвязей, а также рассмотрев

возможные альтернативы. Если присутствующий менеджер указывает на какие-либо погрешности в функционировании модели, то аналитик должен развеять его сомнения или внести в модель необходимые коррективы. Наиболее важно на этом этапе как можно быстрее получить работающую модель и продемонстрировать ее потенциальным пользователям для получения замечаний. Обычно хороший результат дает проведение нескольких демонстраций модели для групп пользователей, представляющих разные подразделения и службы. Это помогает оценить степень завершенности модели и ее восприятие менеджерами и сотрудниками компании.

Затем модель усовершенствуется на основе полученных откликов и представляется еще раз. Иными словами, важно получить как можно больше критических замечаний на возможно более ранней стадии разработки модели. Ведь каждый раз всплывают нюансы, о которых квант забыл, неправильно интерпретировал или просто оценил ошибочно, а также обстоятельства, о которых менеджеры забыли предупредить, о чем предупреждали, но им не нравится, как это было реализовано, и тому подобное. Независимо от того, посвящен ли аналитический проект решению какой-либо уникальной задачи или периодически повторяется, успех его реализации, как правило, способствует формированию конструктивных отношений в коллективе. Менеджер, принимающий решения, вначале нуждается в некоем авансе доверия со стороны коллег, но если дела идут хорошо, то вместо аванса возникает настоящее доверие, основанное на опыте совместной успешной работы. Возникновение взаимного доверия, уважения и взаимопонимания требует усилий и времени, особенно со стороны количественных аналитиков, поскольку у менеджеров на это отпущено времени, как правило, гораздо меньше. Часто случается так, что хорошо сложившиеся взаимоотношения позволяют реализовать несколько последовательных успешных аналитических проектов, что, в свою очередь, ведет к углублению доверия и взаимопонимания.

## Окупаемость сочетания искусства и науки в Intel

В Intel описанный выше подход разрабатывался и совершенствовался в течение более чем двадцати лет, в процессе реализации самых разных

аналитических проектов. Он доказал свою продуктивность. Сначала проекты касались, прежде всего, производства, в том числе проектирования производственных мощностей, строительства, модернизации и текущих операций компании. Анализируемые проблемы и внедряемые управленческие решения охватывали широкий спектр проблем, начиная с расчета количества необходимого оборудования и его планировки до управления объемом незавершенного производства и ремонтом оборудования.

Второе ведущее направление аналитической работы касается комплексного планирования запасов, производства и логистики во всей сети заводов компании. Ее производственные мощности расположены в США, Коста-Рике, Ирландии, Израиле, Китае, Малайзии, а в последнее время и во Вьетнаме, причем все они работают 24 часа в сутки и 365 дней в году. При этом аналитика должна учитывать не только часовые пояса, но и культурные и языковые отличия.

Следующие аналитические проекты касались движения запасов в цепи поставок. Проекты анализа структуры контрактов на оборудование нацелены на обеспечение взаимовыгодного сотрудничества с поставщиками и гибкости производства. Прогнозирование спроса на продукты, позиционирование новых продуктов ориентированы на обслуживание потребителей и оптимизацию затрат, требуемых цепочкой поставок. Эти системы по-прежнему активно используются или совершенствуются на основе аналитических систем и процедур принятия решений второго и третьего поколения. В последнее время такие проекты распространились и на многочисленные подразделения Intel, занимающиеся разработкой новых продуктов. Специалисты по техническим решениям устанавливают отношения со старшими менеджерами, наделенными интуицией и имеющими опыт разработки инновационных продуктов. Последние аналитические проекты включают разработку отдельных потребительских свойств будущих продуктов, проигрывание сценариев внедрения продуктов с учетом распределения технических ресурсов между отдельными проектами.

Эти проекты и их результаты в 2009 году принесли Intel премию Института исследования операций и менеджмента за «последовательное применение принципов исследования операций и

менеджмента лидерским, инновационным, нестандартным образом и на постоянной основе». Принимая награду, тогдашний председатель совета директоров Крейг Барретт назвал применение прогрессивной аналитики причиной роста конкурентоспособности компании в течение предшествовавших двух десятилетий, что обеспечило миллиарды долларов прибыли.

## Ваши аналитические обязанности

Успешное решение аналитической проблемы требует соблюдения определенных обязанностей как количественными аналитиками, так и лицами, принимающими решения (то есть вами). Большая часть глав этой книги посвящена описанию того, чем занимаются кванты и как вам лучше разобраться в их данных и отчетах. Сейчас пришло время обрисовать роль тех, кто принимает решения, в аналитических разработках — независимо от уровня их математической подготовки.

### Чего количественные аналитики могут ожидать от тех, кто принимает решения

Принимая решения, вы должны:

- уделить аналитикам достаточно времени и внимания, чтобы удостовериться в том, что они способны увидеть проблему с вашей точки зрения;
- выделить в их распоряжение время и внимание людей из вашей организации, способных помочь в понимании деталей деловой ситуации;
- иметь четкое представление о времени и средствах, необходимых для выработки решения, и достичь согласия с коллегами по этому вопросу;
- в достаточной степени изучить математику и статистику, чтобы в общих чертах понимать принцип работы модели и возможные причины того, что она неадекватна реальной ситуации;
- вежливо, но твердо требовать объяснить вещи, которые вам непонятны;
- посещать все брифинги, совещания и демонстрации, имеющие отношение к аналитике;
- проинформировать сотрудников о том, что эффективное использование модели необходимо как для успеха компании, так и для их собственного успеха.

## Изучить основы математики и статистики

В [главе 6](#) мы предложили несколько способов изучения основ статистики людьми, принимающими решения. Мы считаем, что это неременная обязанность менеджеров любого уровня, включая топ-менеджеров. Почему? В нашем насыщенном информацией обществе и деловой культуре просто невозможно представить применение данных и аналитики для решения управленческих проблем без достаточно сложного математического инструментария.

Те менеджеры, которым недостает математической подготовки, легко могут попасть в неприятные ситуации, что наглядно подтверждает пример Джо Кассано из AIG Financial Products, приведенный в [главе 1](#). Многие компании все в большей степени используют статистические и математические модели для ведения бизнеса. Отсюда логически следует вывод о том, что менеджер, не понимающий принципов построения математических моделей, не может успешно внедрить их в практику. Как говорит выпускник Йельского университета Роберт Шиллер (обосновывая причины финансового кризиса 2008–2009 годов, который ему удалось предсказать), «если вы руководите компанией, то должны уметь обращаться с цифрами. Количественная информация действительно имеет значение»<sup>[109]</sup>.

Некоторые компании настаивают на том, чтобы их менеджеры имели базовые знания в математике и построении моделей. Например, генеральный директор TD Bank Group Эд Кларк, получивший степень по экономике в Гарвардской школе бизнеса, сумел избежать многих проблем, с которыми столкнулись другие американские банки в период финансового кризиса. Он так описывал эти проблемы в интервью газете Toronto Star: «Общаясь с коллегами по сфере операций со структурированными финансовыми продуктами, я столкнулся с весьма опасным фактом: они просто не понимали сути этих инструментов. Им никогда не приходило в голову самостоятельно провести математические расчеты, положенные в основу каждого такого продукта. Отчасти это следствие того, что они делегировали их разработку и понимание на слишком низкие уровни организационной иерархии»<sup>[110]</sup>.

По мере того как во всех отраслях находят все более широкое применение аналитика и базы данных, топ-менеджерам вменяется в обязанность овладеть более или менее сложными аналитическими приемами. Иначе они просто не смогут вмешаться, когда какой-нибудь трейдер ввяжется в операции, связанные с чрезмерным и неосознаваемым риском, или когда маркетолог предложит предсказательную модель, требующую сбора слишком большого объема аналитических данных. В результате их компании и потребители оказываются в весьма опасном положении.

В частности, топ-менеджеры должны разбираться в таких концепциях:

- показатели общей тенденции (среднее значение, мода, медиана);
- вероятности и распределение;
- выборка;
- основы корреляционного и регрессионного анализа;
- основы постановки экспериментов;
- интерпретация визуальной аналитики.

Топ-менеджеры могут освоить эти концепции теми же способами, что и их подчиненные; кроме того, топ-менеджеры располагают достаточными средствами для того, чтобы пригласить профессоров или консультантов провести занятия для группы старших менеджеров или даже индивидуальные.

## Понимание и тестирование исходных предположений

Мы уже приводили знаменитое высказывание статистика Джорджа Бокса: «Все модели некорректны, но некоторые при этом полезны». Тогда же мы заметили, что очень важно уловить тот момент, когда некорректная модель перестает быть полезной. Чаще всего это происходит тогда, когда заложенные в модель исходные предположения оказываются неверными или недействующими. Мир постоянно меняется, и обязанность скептически настроенного топ-менеджера в том, чтобы определить, не привели ли эти изменения к недостоверности модели. Далее приведены некоторые примеры

исходных предположений для количественных моделей, на практике используемых организациями.

- Готовность покупателя купить продукт по определенной цене (известной под названием модели эластичности цены) не изменилась, хотя общеэкономические условия ухудшаются.
- Предпочтения покупателей сегодня не отличаются от таковых по выборке покупателей, на которой мы тестировали различные версии дизайна веб-страниц несколько лет назад.
- Разработанная нами предсказательная модель вероятности банкротства ипотечных банков в период роста цен на недвижимость продолжает действовать и в период снижения цен (видимо, это несколько проблематично).
- Вероятность мощного урагана в Южной Флориде не исчезла, несмотря на то что, похоже, постепенно происходят глобальные климатические изменения на планете.
- Формирование выборки приверженцев политического деятеля из владельцев стационарных телефонных номеров по-прежнему удовлетворяет требованиям репрезентативности, несмотря на то что многие избиратели не имеют стационарных телефонов (как мы и предполагали, это тоже проблематично).

Не все из этих исходных предположений оказались необоснованными. По сути, поскольку практически все модели разрабатываются на основе данных за прошлые периоды (как мы помним, получить надежные данные о будущем трудно), они по умолчанию базируются на том предположении, что будущее в основных чертах будет повторять прошлое. Такие модели часто оказываются адекватными в долгосрочной перспективе. Как отметил Чарльз Дахигг в книге *The Power of Habit: Why We Do What We Do in Life and Business*<sup>[111]</sup>, установившись, поведение человека остается на удивление постоянным в течение долгого времени. Это позволяет нам предсказывать будущее на основе информации о прошлом.

Некоторые организации платят немалые деньги талантливым прогнозистам только за тестирование исходных предположений. Взять,

к примеру, Ларри Саммерса. Бывший консультант по экономике в администрации Клинтона и Обамы, экс-президент Гарвардского университета работал в качестве консультанта хеджевого фонда D.E. Shaw. Том встретился с Саммерсом на каком-то общественном мероприятии и поинтересовался его функциями в этом фонде. Тот ответил: «Я прихожу на работу раз в неделю и брожу по кабинетам квантов, разрабатывающих математические модели торговли ценными бумагами. Я задаю им вопросы относительно исходных предположений, лежащих в основе этих моделей, а также ситуаций, когда они могут оказаться нерелевантными. Вы удивитесь, если узнаете, как часто мне не могут дать внятного ответа». Говорят, за эту работу Саммерс получал пять миллионов долларов в год, так что, надо думать, руководство фонда считало ее важной.

Вы тоже можете последовать примеру Саммерса. Если кто-то представляет математическую модель, весьма умно с вашей стороны было бы поинтересоваться, на каких исходных предположениях она основана и при каких условиях выдаст недостоверные результаты. Если ответ перегружен математической терминологией, попытайтесь еще раз поставить вопрос о том, что должно измениться в мире, чтобы модель утратила адекватность.

## Не стесняйтесь переспрашивать

Последнюю мысль предыдущего пункта можно сформулировать несколько иначе: очень важно попросить о дополнительных объяснениях, если что-то непонятно. Особенно важно просить дополнительных объяснений, если вместо данных и анализа вам предлагается чье-то мнение или истории из жизни. Как говорит CEO Caesars Entertainment Гэри Лавмен, «я не обязан получить сто процентов ответов, но моя работа состоит в том, чтобы задавать массу неприятных, глубоких, а иногда почти оскорбительных вопросов, поскольку они являются частью аналитического процесса, приводящего к точному и глубокому знанию»<sup>[112]</sup>.

Чтобы стимулировать собеседника к использованию аналитики, можно поставить такие вопросы:

— Вы помните ваши данные?

- Как вы думаете, можно ли протестировать эту гипотезу с помощью конкретных данных?
- Задумывались ли вы над возможностью эмпирического анализа этой идеи?
- У нас примерно ... покупателей. Проверяли ли вы вашу идею хотя бы на ком-то из них?
- Может быть, вам стоит рассмотреть возможность проведения небольшого, но научно строгого эксперимента для проверки этой концепции?

Думаем, вам понятна основная идея. Если в организации найдется достаточно людей, постоянно задающих такие вопросы, корпоративная культура быстро и существенно изменится к лучшему.

Количественные аналитики зачастую пытаются описывать модели и проблемы на математическом жаргоне. Это не значит, что вам остается только молча слушать или самим осваивать их терминологию. В качестве удачной иллюстрации этого тезиса можно привести фильм «Предел риска», где рассказывается о драматических событиях, приведших к финансовому кризису 2008–2009 годов. Место действия — инвестиционный банк, весьма напоминающий Lehman Brothers. Главный герой — специалист по двигателестроению и техническим наукам. Он разработал новую методику оценки кредитного риска и демонстрирует ее директору отдела торговых операций банка (его играет Кевин Спейси). Этот некоронованный король кредитования заявляет: «Вы же знаете, что я ничего в этих формулах не понимаю. Объясните простым английским языком, что это означает»<sup>[113]</sup>.

На его месте любой менеджер заявил бы то же самое.

Профессор по маркетинговым стратегиям Лиам Фахи опубликовал в журнале *Strategy and Leadership* статью, посвященную выявлению роли топ-менеджеров в организации аналитических исследований путем специального опроса<sup>[114]</sup>. Они станут хорошим завершением раздела о том, чего можно ожидать квантам от топ-менеджеров. Приведем полный список этих вопросов:

- Какой деловой проблеме или необходимости посвящен

аналитический проект?

- Какая новая информация, аспекты деловой проблемы или условия ее возникновения проанализированы?
- Каким образом я могу использовать новую информацию в своей работе?
- Каким образом новая информация повлияет на решения, которые нам предстоит принять?
- Каким образом новая информация повлияет на решения, находящиеся на этапе разработки и предполагаемые в будущем?

Когда появляются предварительные результаты, топ-менеджерам стоит спросить:

- Чем нас удивили предварительные результаты?
- Можете ли вы провести дальнейший анализ, чтобы подтвердить или опровергнуть предварительные выводы?
- Следует ли нам привлечь других сотрудников для дальнейших исследований?
- Можно ли считать, что мы далеко продвинулись в решении проблемы?
- Если предварительные результаты подтвердятся, как это повлияет на мое мнение об этой проблеме или каких-либо других?

При получении каждого нового результата топ-менеджер должен задать вопросы:

- Что нового мы узнали?
- Что мы думали об этом до получения результата?
- Насколько существенна разница?
- Чем обоснована связь между исходным массивом данных и полученным результатом?

После получения окончательных результатов стоит уточнить следующие моменты:

- Кто принимал/принимает участие в формировании нового взгляда на проблему?
- Как исполнители повлияли на результат?
- В чем может заключаться принципиальная разница между людьми или подразделениями?

Если вы как топ-менеджер задаете эти вопросы, то будете гораздо глубже вовлечены в аналитическую работу и аналитики будут считать вас заинтересованным и компетентным пользователем. Ну а если аналитики сумеют ответить на ваши вопросы, то можно считать, что и они неплохо поработали!

## Чего можно ожидать от количественных аналитиков

Поскольку мы достаточно подробно рассмотрели обязанности людей, принимающих решения, по организации аналитической работы, логично поставить вопрос о том, чего менеджеры могут ожидать от количественных аналитиков, чтобы их сотрудничество было продуктивным.

### Чего можно ожидать от количественных аналитиков

Если вы топ-менеджер, работающий с количественными аналитиками, то вы с полным основанием можете ожидать от них следующего:

- Хорошего понимания закономерностей бизнеса и отдельных бизнес-процессов, в частности тех, для которых формулируются аналитические проблемы.
- Понимания логики вашего мышления, знания видов и методов анализа вкупе с их результатами, которые способны повлиять на ваше мнение.
- Способности устанавливать конструктивные взаимоотношения с ключевыми сотрудниками в вашей организации.
- Объяснения преимуществ и усовершенствований, которые способна обеспечить аналитика, исходя из интересов бизнеса.
- Предоставления точных оценок времени и затрат, необходимых для разработки модели и связанных с ней инструментов.
- Терпения и умения объяснить иначе или с других позиций в том случае, если вы не понимаете сути их предложений или не верите в реальность прогнозируемых преимуществ.
- Умения построить структурированный процесс сбора информации и внедрения

правил, необходимых для разработки и эксплуатации модели.

- Помощи в осмыслении аналитической проблемы в широком контексте, в том числе формулировки предполагаемого решения, определения лиц, заинтересованных в ее решении, организационных ресурсов, необходимых для ее решения.
- Умения быстро разработать демоверсию новой модели и набора аналитических процедур (если только нет причин поступить иначе), чтобы представить пользователям наглядное подтверждение работоспособности модели и получить на нее отзывы.
- Настойчивого поэтапного совершенствования модели до тех пор, пока она не будет удовлетворять конкретным требованиям пользователей.
- Понимания того, что для освоения модели менеджерам требуется определенное время, в течение которого следует терпеливо отвечать на возможные вопросы и развеивать недоверие к рабочим свойствам модели.

## Стремление изучать особенности бизнеса и заинтересованность в разрешении его проблем

Некоторые количественные аналитики интересуются не столько деловой проблемой, которую следует решить, сколько методами количественного анализа и самой процедурой анализа. Отчасти это результат особенностей нашей системы образования, в которой принято изучение математики и статистики вне привязки к практике их применения. Но если кванты не сосредотачивают внимание на проблемах бизнеса, то вряд ли смогут эффективно решить их и дать полезную информацию для принятия решения.

Лучше всего убедиться в заинтересованности аналитика в решении проблем бизнеса еще на этапе подбора кандидатур и найма на работу. Если человек уже принят в штат, то изменить что-либо бывает очень сложно. Например, руководитель группы аналитиков из компании Verizon Wireless Энн Робинсон просит любого претендента на вакансию описать конкретную деловую проблему, которую тому приходилось решать в прошлом, а также выделить ее наиболее интересные особенности. Примерно такие же вопросы задает соискателям Карл Кемпф из Intel. Если соискатели затрудняются ответить на эти вопросы, то их кандидатуры отклоняются — Робинсон и Кемпф утверждают, что, к несчастью, это происходит довольно часто.

## Умение говорить на языке бизнеса

Мы уже несколько раз говорили об этом на протяжении предшествующих глав, и это оправданно, поскольку проблема действительно сложная. Но количественные аналитики должны научиться излагать аналитические подходы и результаты языком бизнеса. Во многих случаях это означает использование терминологии, знакомой бизнесменам и менеджерам: доходность инвестиций, поведение потребителей, экономия средств и полученная прибыль. Постоянные разговоры о деньгах могут показаться несколько торгашескими, но это язык бизнеса. В государственном учреждении или некоммерческой организации своя терминология, включающая, например, слова «граждане», «избиратели» или «бюджеты».

Патрик Мур, возглавляющий группу коммерческой аналитики в компании Merck (в конце этой главы приводится пример работы его группы), говорит, что пытается следовать трем простым правилам, продиктованным здравым смыслом, когда объясняет клиенту результаты анализа, стремясь помочь ему принимать более эффективные управленческие решения.

- Не представлять процедуру анализа как своего рода «черный ящик», поскольку в этом случае клиент пытается уклониться от его проведения. Поэтому Мур старается говорить предельно прозрачно и доступно.
- Доводить до деловых клиентов ту мысль, что для получения соответствующих целям данных следует пользоваться соответствующими методами; иными словами, он и его аналитики стараются демонстрировать уверенность в правильности применяемых аналитических методов.
- Предоставлять клиенту краткие и емкие выводы, основанные на результатах анализа, которые тот мог бы использовать для убеждения своего руководства.

Группа Мура также широко использует графические методы представления данных, например для того, чтобы наглядно показать относительную значимость отдельных переменных в модели. Даже если клиент не вполне понимает суть используемых в анализе

показателей или статистических данных, он, глядя на столбиковую диаграмму, способен понять, какие из них максимально повлияли на результат.

## Способность объяснять специальные термины

Иногда квантам все же приходится прибегать к специальной терминологии, чтобы объяснить суть проведенных процедур. Но даже если без этого не обойтись, кванты должны быть готовы к «переводу» специальных терминов на понятный язык, и это не должно заставить их врасплох. Если речь идет об аналитической процедуре или методе, который используется часто, то кванты в вашей организации могут организовать совещание с коллегами и совместно решить, как понятно и просто объяснить его суть. Конечно, для относительно простых видов анализа, включающих одну-две переменные, визуальная аналитика представляет собой наиболее эффективный способ объяснить взаимосвязи между показателями.

## Стремление налаживать отношения

Как уже упоминалось в этой главе, оптимальные решения принимаются не столько благодаря математике, сколько благодаря налаженным взаимоотношениям между квантами и менеджерами. Если ваши кванты не стремятся к установлению таких отношений, то им, возможно, стоит переквалифицироваться в астрофизики, лесники или поискать профессию, предполагающую работу в одиночестве, вместо того чтобы пытаться делать карьеру в бизнесе.

Конечно, легче сказать, чем сделать; однако действительно большинство квантов исторически предпочитали цифры людям. Если вы ищете и проводите интервью с ориентированными на командную работу квантами или отбираете кандидатов в аналитическую программу для нужд бизнеса (вроде той, которая проводится в Университете Северной Каролины и описана в [главе 6](#)), то так или иначе вам придется решать эту проблему.

## Не заставлять менеджеров чувствовать себя дураками

Нам приходилось видеть не одну организацию, в которой кванты

прямо-таки получали удовольствие от того, что «нормальные» предприниматели чувствуют себя дураками. Они могли заявить что-нибудь вроде: «Конечно, вам приходилось слышать о регрессионном анализе!», «Извините, но показатель R-квадрат — это настолько элементарная вещь, что у меня нет времени ее объяснять». Некоторые «суперкванты» (как одна организация называла своих наиболее продвинутых в количественном анализе сотрудников) даже позволяли себе смотреть сверху вниз на рядовых квантов, занимавших аналогичные должности в соседней компании.

Конечно, подобное поведение неприемлемо и крайне вредно для эффективного решения проблем. Как и во многих других случаях, оно говорит о том, что человек не чувствует достаточного уважения к себе. В организациях, где кванты глубоко вовлечены в предпринимательские аспекты деятельности и пользуются искренним уважением людей, принимающих решения, каждый из них оказывается прекрасным человеком и приятным в общении коллегой. В организациях, которые зачем-то пригласили их на работу, но при этом игнорируют каждый раз, когда дело доходит до принятия важных решений, отношения с ними складываются весьма непросто. Как и большинство других людей, кванты уважают тех, кто с уважением относится к ним.

## Пример аналитического мышления: прогноз спроса в Cisco

Прогнозирование потребительского спроса представляет серьезную проблему для многих компаний, особенно в машиностроительной отрасли<sup>[115]</sup>. Для Cisco Systems, лидера на рынке телекоммуникационного оборудования, это особенно важно. Компания имеет очень сложную глобальную цепь поставок и не производит большинство продуктов, которые продает. Вице-президент по глобальным операциям в Отделении управления цепочками ценности Cisco Кевин Харрингтон говорит: «Прогнозирование потребительского спроса, конечно же, центральное звено управления цепью поставок и ключевой фактор гибкого производства. Это приобретает особое значение в такие времена, как сейчас, когда в макроэкономике происходят быстрые изменения, сопровождаемые хаотичными

колебаниями спроса и предложения. По сути, возникшая у Cisco необходимость списать некоторое количество неиспользуемых запасов (стоимостью около 2,25 миллиарда долларов) после краха доткомов в 2001 году стала своего рода толчком к глобальной трансформации нашей цепи поставок»<sup>[116]</sup>.

Этот проект стал удачной иллюстрацией не только аналитического мышления, но и конструктивных взаимоотношений между менеджерами, принимающими решения, и количественными аналитиками.

**Определение и формулирование проблемы.** Проблема Cisco состояла в том, чтобы разработать надежную методику прогнозирования спроса по каждому из более чем десяти тысяч продуктов. Менеджеры из разных подразделений компании, включая отделы продаж, маркетинга и финансов, уже разработали «согласованный прогноз» на основе некоего сочетания интуиции и экстраполирования тенденций спроса за предыдущие периоды. Но вице-президент Cisco по управлению спросом и планированию Карл Брейтберг понимал, что статистический прогноз, основанный на данных, вполне может стать полезной альтернативой интуитивной разработке сотрудников, ведь в последнем случае на показатели спроса мог повлиять избыточный оптимизм разработчиков. Он поручил старшему менеджеру отдела планирования и прогнозирования спроса Энн Робинсон и ее команде из шести человек попытаться разработать методику статистического прогноза. Робинсон поняла, что для успеха этого проекта ей надо не только разработать надежную модель, но и заставить менеджеров компании поверить в нее и использовать при принятии решений. Она определила ключевых сотрудников, заинтересованных в получении надежной методики прогнозирования, и составила план проекта, согласно которому надежные результаты ожидалось через восемнадцать месяцев после начала эксплуатации предложенной модели. На каждом этапе разработки она демонстрировала результаты заинтересованным лицам, учила их использовать модель и, как она надеялась, получила поддержку на будущее.

**Изучение предыдущих поисков решения.** Существует несколько подходов к статистическому прогнозированию. На основе изучения имеющегося опыта таких исследований можно утверждать, что наилучшие результаты достигались в результате одновременного применения разных подходов — так называемого сборного прогноза. Робинсон знала из опыта предыдущих исследований, что сборный прогноз представляет собой достаточно мощный и универсальный инструмент исследования, поэтому она удостоверилась, что все рассмотренные на этом этапе аналитические инструменты обладают аналогичными качествами.

**Моделирование (выбор переменных).** Ключевыми переменными модели, по всей видимости, должны стать объем заказов и объем продаж в прошлые периоды. Эти переменные достаточно широко используются для прогнозирования спроса в целом ряде отраслей.

**Сбор данных (измерение).** Притом что выбор переменных с самого начала был очевиден, источники информации об объеме текущих заказов достаточно разнообразны и каждый из них следовало оценить с точки зрения соответствия потребностям модели. Например, Cisco группирует заказы потребителей по отраслям, емкости потребительских сегментов, регионам и фактическому объему поставок отдельным потребителям. Эти показатели не всегда соответствуют друг другу. К счастью, группировка продаж по всем возможным критериям уже содержалась в имеющейся базе данных компании. Тем не менее группе аналитиков под руководством Робинсон предстояло разработать несколько новых показателей удовлетворения спроса по категориям потребителей, что могло бы стимулировать ориентацию на потребителя во всей цепи поставок Cisco.

**Анализ данных.** Статистическое прогнозирование на выходе дает прогнозируемый уровень спроса с доверительным интервалом для каждого уровня. Например, спрос на конкретную модель роутера в месяц по прогнозу составит от 3 до 3,5 тысячи единиц с 95-процентной вероятностью попадания в этот интервал. Гибкий подход к

разработке модели включает ряд шагов, на выполнение каждого из которых требуется два-три месяца, чтобы удостовериться в работоспособности модели и в том, что ее можно масштабировать в соответствии с количеством и разнообразием продуктов Cisco. Некоторые из этих шагов включают выполнение следующих операций:

- выбор программного продукта, удовлетворяющего требованиям (Cisco выбрала SAS Forecast Server с поддержкой функции сборного прогноза);
- оценка вероятности того, что статистическая модель представит более надежный результат, чем согласованный прогноз аналитиков, и проверка этого утверждения;
- настройка моделей с целью повышения точности прогнозов;
- оценка возможности охвата номенклатуры с тысячами позиций, относящихся к трем сотням линий, в рамках данного метода прогнозирования (это возможно);
- автоматизация модели (процедура ручного расчета будет слишком трудоемкой, но в случае необходимости менеджеры и эксперты смогут это сделать).

На каждом шаге предусмотрен контроль со стороны сотрудников, заинтересованных в результатах анализа, что увеличивает их вовлеченность в реализацию нового подхода.

**Результаты и необходимые меры.** В настоящее время методом статистического прогнозирования еженедельно готовится прогноз спроса по более чем 18 тысячам позиций номенклатуры на 24 месяца вперед. В результате сочетания статистического и согласованного прогноза его точность повысилась в среднем на 12 процентов. По словам Кевина Харрингтона, проект оказался очень успешным:

Результатами проекта стали повышение точности прогноза, рост оборачиваемости запасов и общее повышение согласованности спроса на продукты и их поставки, что привело к сокращению избыточных складских запасов и более быстрому, надежному обслуживанию как Cisco, так и ее потребителей. В худшие времена недавнего финансового кризиса Cisco получила возможность сократить объем складских запасов в сбытовой цепи, не прибегая к их уценке или ухудшению качества обслуживания. В настоящее время наши эксперты по статистическому прогнозированию работают над

дальнейшим совершенствованием модели в условиях роста спроса в результате глобального экономического оживления<sup>[117]</sup>.

В дополнение к результатам, описанным Харрингтоном, Энн Робинсон отмечает, что теперь менеджеры компании уверенно используют показатели уровня спроса и вероятности для характеристики спроса. Они привыкли использовать интервалы возможных значений спроса в противоположность одному прогнозному показателю и упоминают о них на любом важном совещании. Коротко говоря, культура прогнозирования в Cisco резко изменилась в направлении перехода на аналитическую основу.

На всем протяжении проекта Робинсон пыталась вовлечь всех занимающихся прогнозированием сотрудников в новый аналитический процесс. Она провела интенсивный мозговой штурм с участием заинтересованных в получении прогноза, чтобы идентифицировать новые показатели, ориентированные на потребителя. Она устраивала панельные дискуссии с более широкой аудиторией, разрабатывала «дорожные карты» реализации проекта для отдельных групп, а также много раз проводила презентацию «Прогнозирование 101» (Forecasting 101). Для нее разрабатывалось визуальное представление результатов, полученных благодаря функционирующей модели, а сотрудникам предлагалось на их основе «рассказать их собственную историю» о данных. Робинсон наладила тесное сотрудничество с IT-отделением Cisco и отметила, что иногда сложно было найти отличия в заданиях ее группы и IT-отделения.

## Пример аналитического мышления: оптимизация сбытового персонала в компании Merck

Определение оптимальной численности сбытового персонала в ведущей фармацевтической компании наподобие Merck — сложная аналитическая задача. Регулярно появляются новые продукты, стимулируя спрос и одновременно потребность в сбытовом персонале; столь же регулярно заканчивается срок действия патентов на выпускаемые продукты, что приводит к сокращению потребности в сбытовом персонале. Для новых продуктов данных об объемах продаж за прошлые периоды не существует, поэтому методов надежного

прогнозирования будущей потребности в сбытовом персонале нет.

Многие фармацевтические компании нанимают внешних консультантов для решения этой задачи. Однако когда количественный аналитик Пол Каллукаран, имеющий опыт анализа данных по объемам продаж в фармацевтических компаниях, пришел на работу в отдел коммерческой аналитики, топ-менеджеры компании решили, что пора провести анализ численности сбытового персонала собственными силами.

**Определение и формулирование проблемы.** Какой должна быть оптимальная численность сбытового персонала при наличии лекарств с истекшим сроком действия патента и новых лекарств, только что выведенных на рынок? Сбытовой персонал сгруппирован по регионам и брендам продуктов. Поэтому задача усложняется тем, что нужно определить его численность не в целом, а в разрезе регионов и брендов.

**Изучение предыдущих поисков решения.** Если вспомнить о предлагаемых консалтинговыми фирмами услугах, то понятно, что кое-какая литература по теме исследования имеется. На предыдущей работе Каллукаран занимался немного иными проблемами, но интересовался тем, как разные компании решают этот вопрос. Однако по большей части методика анализа оставалась за кулисами презентаций, поскольку последние были рассчитаны на восприятие маркетологов и специалистов по продажам. Каллукарану и начальнику отдела коммерческой аналитики Патрику Муру такой подход категорически не нравился. В прошлом разные отделения Merck приглашали консультантов, а те использовали свои методы прогнозирования численности сбытового персонала; на этот раз впервые предпринималась попытка выработать централизованный подход к решению этой проблемы.

**Моделирование (выбор переменных).** Каллукаран решил использовать несколько методов для расчета оптимальной численности сбытового персонала. В дополнение к традиционному использованию статистических моделей он и его команда решили

внимательнее присмотреться к процессу обслуживания отдельного покупателя. Они расспрашивали сбытовых агентов об их работе с покупателями-врачами, пытаясь оценить объем нагрузки для одного такого потребителя. Они также рассчитывали прогнозы по различным продуктам и разрабатывали нелинейные модели откликов продвижения сбытового персонала и возможных изменений в составе прописываемых врачами лекарств. Они анализировали воздействие связанных со сбытовым персоналом факторов, способных повлиять на модель поведения врачей в отношении выписываемых лекарств в противоположность всем остальным факторам (привычки, бренд лекарственных средств, давление со стороны пациентов). Аналитики рассматривали данные на уровне пациентов, чтобы понять причины их приверженности тем или иным лекарствам. Эти данные говорили и о том, что многие пациенты с течением времени прекращают или уменьшают прием прописанных лекарств, что существенно влияет на объем их продаж в долгосрочной перспективе. Наконец, они разработали комплексную оптимизационную модель, способную оптимизировать ресурсы на обслуживание отдельного врача, продукта и каждого из сотен регионов, где продавались лекарства.

**Сбор данных.** Обычно фармацевтические компании получают данные о выписываемых врачами лекарствах от третьих лиц: от компаний, распространяющих базы данных. У Merck эти данные были. Но для разработки модели функционирования сбытовых агентов требовалось провести анкетирование о моделях их рабочего поведения и затратах времени на потребителя. Для этого следовало завоевать доверие сбытовых агентов, то есть так провести анкетирование, чтобы это не выглядело прелюдией к сокращению персонала. Агенты сообщат достоверные данные только в том случае, если не будут чувствовать угрозу для себя.

**Анализ данных.** Как мы отмечали, решение этой сложной задачи включало несколько аналитических подходов, в том числе целочисленную оптимизацию и непараметрическую (не предполагающую определенный тип распределения данных) модель, которые рассчитывали характеристики для каждого продуктового

сегмента на основе данных за прошлые периоды. Поскольку проект предполагал разработку моделей для каждого бренда и региона, первоначально на это потребовалось 16 часов. Но группа Каллукарана стремилась получить результат быстрее, поэтому ее сотрудники распределили расчетные задания между сотнями дополнительных компьютеров. На каждом из них рассчитывались показатели для определенного региона. Мобилизовав столь мощный компьютерный парк, Каллукаран добился проведения расчетов за 20 минут.

**Результаты и необходимые меры.** Хотя это был совершенно новый подход к проблеме расчета численности сбытового персонала в Merck, но сама проблема отнюдь не отличалась новизной. Это помогло привлечь сотрудников к внедрению модели и обучить их принимать решения на основе расчетов. Со стороны отдела продаж вице-президент по стратегическому планированию владел аналитическим мышлением и также интересовался наработками прежних консультантов в этой области. В отделе маркетинга уже предпринимались попытки разработать модель для решения этой задачи, но до внедрения дело не дошло. Всегда находился кто-то недовольный тем, что модель не дает ответа то на один, то на другой вопрос; это классический пример того, что лучшее — враг хорошего. В процессе внедрения своей модели расчета численности сбытового персонала Каллукарану удалось убедить маркетологов «использовать то, что уже есть». Сначала он поработал с небольшой командой маркетологов одного из брендов и показал им преимущества новой модели перед интуитивным подходом. Он сопоставил интуитивный подход с основанным на данных аналитическим; но не пытался навязать им использование модели: «Рассматривайте ее как еще один аргумент при принятии решения». Разъяснение преимуществ новой модели группам и отделам опиралось на системное внедрение аналитического подхода на уровне компании.

К этому времени президенту Merck понравилась возможность сравнивать группы агентов, отвечавшие за различные бренды, поскольку это позволяло объективно оценить подаваемые ими заявки на ресурсы. Постепенно почти все отделы в Merck внедрились модель Каллукарана. Его группа аналитиков получала заявки на повторный

расчет численности агентов, когда в каком-либо регионе открывались вакансии. Региональные менеджеры получили больше свободы действия одновременно с повышением ответственности за финансовые результаты. Исходя из этих требований, аналитическая группа, включавшая системных разработчиков, создала специальное «аналитическое приложение», подсказывавшее менеджеру, стоит ли брать кого-то на освободившееся место. Приложение получило широкое распространение, поскольку освобождало группу Каллукарана от многочисленных заявок на расчеты.

# Итоговые выводы об аналитическом мышлении

Мы надеемся, что к этому моменту — эффектного завершения книги, если, конечно, вы не увлекаетесь чтением сносок — нам удалось убедить вас в истинности некоторых положений. Во-первых, аналитическое мышление и принимаемые на основе данных решения будут играть всевозрастающую роль и в бизнесе, и в обществе. Нам нужно множество менеджеров и профессионалов, способных мыслить аналитически, и очень хорошо, если вы станете одним из них. Во-вторых, надеемся, что вы убедились в своей способности принять участие в этой игре, даже если математика или статистика не является вашей сильной стороной. Если вы понимаете, какова последовательность этапов аналитического мышления и что делает качественным аналитический процесс принятия решения, вы можете сотрудничать с квантами и помогать в совершенствовании их моделей. При этом вы и сами научитесь лучше мыслить и принимать решения.

В-третьих, хотя большинство людей обычно думают о «решении проблемы» как основе аналитического мышления, на самом деле это лишь один из шагов на пути к эффективному управленческому решению. Если проблема сформулирована неверно или неоптимально, то и предложенное решение не принесет особой пользы. Если вы не смогли эффективно сообщить людям о полученных результатах, то вряд ли на их основе будет принято какое-либо решение или предприняты какие-либо действия. Если вы работаете над аналитическим проектом и задаетесь вопросом, как наилучшим образом распределить время, то для начала неплохо будет, если вы распределите его поровну между тремя этапами количественного анализа.

Наконец, в-четвертых, многие считают, что в мире аналитики и аналитических решений речь идет лишь о цифрах и статистических закономерностях, доступных людям с хорошо развитым левым полушарием мозга. Но и работой правого полушария пренебрегать нельзя. Мы старались показать, особенно в [главе 5](#), что креативность

очень важна для аналитического мышления, и в этой главе на основе богатого опыта и наблюдений мы постарались показать, что и взаимоотношения между людьми в процессе анализа имеют не меньшее (если не большее) значение для его результатов, чем способность щелкать числа как орешки. Если вы прочитали эту книгу и обдумали приведенные в ней примеры и идеи, то можно считать вас готовым к вступлению в ряды аналитически мыслящих людей. Поздравляем! Это очень увлекательно! Объем и значение данных будут только расти с течением времени, и вы будете расти вместе с ними.

Мы надеемся, что обретенные вами аналитические навыки принесут пользу и вашей карьере, и той компании, где вы будете работать.

# Благодарности

Том выражает благодарность (и посвящает написанные им страницы этой книги) своей любящей жене и спутнице жизни в течение более чем тридцати лет Джоан Пауэлл Дэвенпорт. Сыновья Хейс и Чейз уже вылетели из гнезда, но не перестают поддерживать отца в его работе, пусть и на расстоянии. Отделение научных исследований Гарвардской школы бизнеса хоть и оказалось менее любящим, но зато предоставило возможность проведения научных исследований на своей базе.

Джин Хо благодарит Джона и Бонни Ризов за их энтузиазм, и вдохновляющую поддержку, в том числе при чтении корректуры рукописи. Особую признательность Джин Хо выражает Джанг Ха Шин за неизменную поддержку и увлеченность идеей книги. Большое спасибо старшей дочери Нури и зятю Донг Вуку, а также младшей дочери Йонгри за их любовь и вдохновение. Джин Хо посвящает написанные им страницы книги любящей и мудрой матери Тай Ха Чо.

Том и Джин Хо благодарны своему редактору Мелинде Мерино другим сотрудникам Harvard Business Review Press за блестяще выполненную работу. Именно эти люди сделали Harvard Business Review Press лучшим издателем деловой литературы. Кроме того, огромное спасибо Марку Аллену за прекрасные иллюстрации. Авторы с огромным удовольствием выражают благодарность всем потрясающим аналитикам и увлеченным анализом топ-менеджерам, живущим и покинувшим нас, которые предоставили увлекательные примеры для этой книги. Авторы горды возможностью стать летописцами их героических свершений.

# Об авторах

Томас Дэвенпорт — приглашенный профессор в Гарвардской школе бизнеса, почетный профессор в области компьютерных технологий и управления в колледже Бэбсона, а также научный сотрудник Центра виртуального бизнеса при Массачусетском технологическом институте. Он старший консультант в компании Deloitte Analytics и соучредитель и директор-исследователь Международного института аналитики. Дэвенпорт написал книги «Аналитика как конкурентное преимущество. Новая наука побеждать» (в соавторстве с Джоан Харрис) и *Analytics at Work* («Аналитика в работе»). Книга, которая у вас в руках, семнадцатая по счету; над ней Том работал в качестве автора, соавтора и редактора.

Джин Хо Ким — профессор бизнеса и статистики в Корейском национальном университете обороны и директор-исследователь лаборатории аналитических исследований при этом же университете. Он получил степень PhD в Школе бизнеса Уортон. Ким — автор шести книг, опубликованных в Корее, в числе которых бестселлеры *100 Common Senses in Statistics* («100 здравых смыслов в статистике») и *Freak Statistics* («Причудливая статистика»). Он разработал и открыл учебную программу по формированию аналитических навыков. В настоящее время занимается исследованием применения аналитических методов для решения разнообразных проблем в бизнесе и обществе.

- [1] Этой системе посвящена книга: Льюис М. [Moneyball. Как математика изменила самую популярную спортивную лигу в мире](#). М.: Манн, Иванов и Фербер, 2013. 464 с. *Прим. ред.*
- [2] Нерепрезентативная — выборка, не отражающая характеристик генеральной совокупности. *Прим. перев.*
- [3] Один петабайт равен 1 073 741 824 мегабайтам. *Прим. перев.*
- [4] AT&T Inc — одна из крупнейших телекоммуникационных компаний в США, провайдер телефонной и беспроводной связи. *Прим. ред.*
- [5] CEO — высшее должностное лицо компании, ближайший российский аналог — генеральный директор.
- [6] От англ. quant — специалист по применению математических методов, методов количественного анализа в решении финансовых вопросов, оценке рисков и пр. Сленговое сокращение от quantitative analyst, quantitative analysis — количественные методы анализа. *Прим. ред.*
- [7] Издана на русском языке: Харрис Д., Дэвенпорт Т. Аналитика как конкурентное преимущество. М. : BestBusinessBooks, 2010. 264 с.
- [8] Ученая степень PhD (Doctor of Philosophy) примерно соответствует степени кандидата наук в России. *Прим. ред.*
- [9] Цитата из выступления Сяо Лименга на семинаре «Статистическое образование и обучение статистиков: подготовка знатоков вина и виноделов», проводившемся в Университете Миннесоты 28.10.2011 г. URL: <http://catalystsumn.blogspot.com/2011/11/statistics-education-seminar-presented.htm>.
- [10] Big Data: The Next Frontier for Innovation, Competition, and Productivity // McKinsey Global Institute, May 2011 URL: [http://www.mckinsey.com/insights/mgi/research/technology\\_and\\_innovation/big\\_data\\_the\\_next\\_frontier\\_for\\_innovation](http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation).
- [11] Dawes R., Faust D., and Meehl P. Clinical Versus Actuarial Judgment // Science. March 1989. Vol. 243, no. 4899. P. 1668–1674.

- [12] Taibbi M. AIG and the Long Con // Rolling Stone. 2009, March 23.
- [13] Morgenson G . Behind Insurer’s Crisis, Blind Eye to a Web of Risk // New York Times. September 27, 2008.
- [14] Lewis M . The Man Who Crashed the World // Vanity Fair, August 2009. URL:  
<http://www.vanityfair.com/politics/features/2009/08/aig200908>.
- [15] Babou S. What Is Stakeholder Analysis? // The Project Management Hut. URL: <http://www.pmhut.com/what-is-stakeholder-analysis>.
- [16] Expedia Travels to New Heights // SASCom Magazine. Third Quarter 2011, no. 14.
- [17] Все эти примеры взяты из практики деятельности клиентов компании по разработке прикладного программного обеспечения Applied Predictive Technologies, хотя мы получили их самостоятельно. Более подробно методы анализа описаны в книге: Davenport T. How to Design Smart Business Experiments. Harvard Business Review, November 2009.
- [18] Navarro M. For Many Latinos, Race Is More Culture Than Color // New York Times. January 13, 2012.
- [19] Wallis W. and Roberts H. Statistics: A New Approach. New York : Free Press, 1960.
- [20] Carley-Baxter L. et al. Comparison of Cell Phone and Landline Surveys: A Design Perspective // Field Methods. February 2010. Vol. 22, no. 1. P. 3–15.
- [21] Cookie (англ.) — небольшой фрагмент данных, отправленный веб-сервером и хранимый на компьютере пользователя. Используется для аутентификации пользователя и хранения персональных предпочтений и настроек пользователя. *Прим. ред.*
- [22] Hansell S. Google Answer to Filling Jobs Is an Algorithm // New York Times. January 3, 2007. URL: [www.nytimes.com/2007/01/03/technology/03google.html](http://www.nytimes.com/2007/01/03/technology/03google.html).

- [23] Joseph Jagger // URL: ; Joseph Jagger. URL: [www.realmoneycasinos.net/joseph-jagger.html](http://www.realmoneycasinos.net/joseph-jagger.html); Roulette—The Men Who Broke the Bank at Monte Carlo—Joseph Jagger. URL: [www.wiseguyroulette.com/roulette-history/joseph-jagger/](http://www.wiseguyroulette.com/roulette-history/joseph-jagger/).
- [24] Ramakrishnan R. Three Ways to Analytic Impact // The Analytic Age blog. July 26, 2011. URL: <http://blog.ramakrishnan.com/>.
- [25] People vs. Collins, 68 Cal. 2d 319 (1968). URL: [http://scholar.google.com/scholar\\_case?case=2393563144534950884](http://scholar.google.com/scholar_case?case=2393563144534950884); People vs. Collins. URL: [http://en.wikipedia.org/wiki/People\\_v.\\_Collins](http://en.wikipedia.org/wiki/People_v._Collins).
- [26] Starfield A., Smith K., and Bleloch A. How to Model It: Problem Solving for the Computer Age. — New York: McGraw-Hill, 1994. P. 19.
- [27] Box G. and Drape N. Empirical Model-Building and Response Surfaces. New York : Wiley, 1987. P. 424.
- [28] Sundem G. Geek Logik: 50 Foolproof Equations for Everyday Life. New York : Workman, 2006.
- [29] Brashears M. Mark Twain: Son of Missouri. Whitefish, MT : Kessinger Publishing, 2007.
- [30] Leisy E. (ed.). The Letters of Quintus Curtius Snodgrass. Irving, TX : University Press of Dallas, 1946.
- [31] Brinegar C . Mark Twain and the Quintus Curtius Snodgrass Letters: A Statistical Test of Authorship // Journal of the American Statistical Association. 1963, no. 58.
- [32] Rosen R. et al. The International Index of Erectile Function (IIEF): A Multidimensional Scale for Assessment of Erectile Function // Urology. 1997. Vol. 49, no. 6. P. 822–830; Rosen R. et al. Development and Evaluation of an Abridged, 5-item Version of the International Index of Erectile Function (IIEF-5) as a Diagnostic Tool for Erectile Dysfunction // International Journal of Impotence Research. 1999. Vol. 11. P. 319–326.
- [33] Ramakrishnan R. Three Ways to Analytic Impact // The Analytic Age (blog), July 26, 2011. URL: <http://blog.ramakrishnan.com/>.

[34] Rajaraman A. More Data Usually Beats Better Algorithms // Datawocky (blog), March 24, 2008. URL: <http://anand.-typepad.com/datawocky/2008/03/more-data-usual.html>.

[35] Morey D. Success Comes from Better Data, Not Better Analysis // Harvard Business Review (blog). August 8, 2011. URL: [http://blogs.hbr.org/cs/2011/08/success\\_comes\\_from\\_better\\_data.html](http://blogs.hbr.org/cs/2011/08/success_comes_from_better_data.html).

[36] Tycho Brahe // Fowler M. Tycho Brahe. URL: <http://galileoandeinstein.physics.virginia.edu/1995/lectures/tychob.html>; Koestler A. The Watershed: A Biography of Johannes Kepler. Doubleday, 1960; Johannes Kepler // Johannes Kepler // Encyclopædia Britannica Online Academic Edition. URL: <http://www.britannica.com/EBchecked/topic/315225/Johannes-Kepler>.

[37] Franks B. Why Nobody Is Actually Analyzing Unstructured Data // International Institute for Analytics (blog post). March 9, 2012. URL: <http://iianalytics.com/2012/03/why-nobody-is-actually-analyzing-unstructured-data/>.

[38] Passell P. Wine Equation Puts Some Noses Out of Joint // New York Times. March 4, 1990.

[39] Микроклимат в контексте виноделия, то есть совокупность природных факторов (тип почв, количество солнца, средняя температура воздуха и другие особенности местности), которые могут повлиять на качество вина (букет и даже вкус). *Прим. ред.*

[40] Определения взяты из Википедии, учебника Хайнца Кохлера (Heinz Kohler) Statistics for Business and Economics (2002), «шпаргалки» от компании Dell по аналитике (2012, табл. 6 и 8). Рекомендуем для быстрого ознакомления со статистической терминологией и методами книгу: Минько А. А. Статистика в бизнесе. Руководство менеджера и финансиста. М. : Эксмо, 2008. *Прим. ред.*

[41] Строго говоря, нулевой гипотезой может быть любое предположение о генеральной совокупности. Предположение, что между наблюдениями не существует значимой связи, только одна из возможных гипотез, которая не обязана быть нулевой. *Прим. ред.*

[42] Здесь авторы пытаются на пальцах дать определение критической области, то есть той части выборочного пространства, которая приводит к отклонению нулевой гипотезы. *Прим. ред.*

[43] Alternative Rich List // [FT.com](http://ft.com). September 22, 2006.

[44] Black F. and Scholes M. The Pricing of Options and Corporate Liabilities // Journal of Political Economy. 1973. Vol. 81, no. 3. P. 637–654; Black–Scholes // The Prize in Economics 1997 // Пресс-релиз, Nobelprize.org. URL: [http://nobelprize.org/nobel\\_prizes/economics/laureates/1997/press.html](http://nobelprize.org/nobel_prizes/economics/laureates/1997/press.html).

[45] Black F. and Scholes M. The Pricing of Options and Corporate Liabilities // Journal of Political Economy. May 1973. Vol. 81, no. 3.

[46] Larsen R. and Marx M. An Introduction to Mathematics Statistics and Its Applications. Englewood Cliffs, NJ : Prentice-Hall, 1981. P. 159. Эту заметку впоследствии процитировали во многих учебниках по статистике и курсах лекций.

[47] Выдержка из выступления Сяо Лименга на семинаре Statistical Education and Educating Statisticians: Producing Wine Connoisseurs and Master Winemakers, проходившем в Миннесотском университете 28.10.2011 г. URL: <http://catalystsumn.blogspot.com/2011/11/statistics-education-seminar-presented.html>.

[48] Лименг С. Statistics: Your Chance for Happiness (or Misery). Описание курса приведено на сайте [http://www.stat.harvard.edu/Academics/invitation\\_chair\\_txt.html](http://www.stat.harvard.edu/Academics/invitation_chair_txt.html).

[49] Schmitt D. Tell a Story. URL: [http://www.allanalytics.com/author.asp?id=2092&doc\\_id=246428](http://www.allanalytics.com/author.asp?id=2092&doc_id=246428) (дата обращения: 27.06.2012).

[50] Вики-сайт — веб-сайт, содержание которого посетитель может редактировать. При этом изменения, внесенные в содержание, отображаются немедленно. Часто является общедоступной доской объявлений. *Прим. ред.*

[51] Cohen I. The Triumph of Numbers: How Counting Shaped Modern Life. New York : W.W. Norton, 2006. Chapter 9; Florence Nightingale // Nuttall P. The Passionate Statistician // Nursing Times. 1983. Vol. 28. P. 25–27.

[52] Mendel G. Experiments in Plant Hybridization. URL: <http://www.mendelweb.org/>; Gregor Mendel // Seung Yon Rhee. Gregor Mendel, Access Excellence. URL: [http://www.accessexcellence.org/RC/AB/BC/Gregor\\_Mendel.php](http://www.accessexcellence.org/RC/AB/BC/Gregor_Mendel.php).

[53] Этот перечень модифицирован авторами на основе материалов одного из сайтов IBM ManyEyes site; см. [http://www-958.ibm.com/software/data/cognos/manyeyes/page/Visualization\\_Options.html](http://www-958.ibm.com/software/data/cognos/manyeyes/page/Visualization_Options.html).

[54] Пример взят из руководства пользователя по визуальной аналитике версии 5.1 компании SAS Working with Automatic Charts. URL: <http://support.sas.com/documentation/cdl/en/vaug/-65384/HTML/default/viewer.htm#n1xa25dv4fiyz6n1etsfkbz75ai0.htm>.

[55] Rosling H. Stats That Reshape Your Worldview // TED talk. February 2006. URL: [http://www.ted.com/talks/hans\\_rosling-shows\\_the\\_best\\_stats\\_you\\_ve\\_ever\\_seen.html](http://www.ted.com/talks/hans_rosling-shows_the_best_stats_you_ve_ever_seen.html).

[56] Хотя группа Шмитта периодически производит такие видеоклипы самостоятельно, этот клип снят на заказ независимой компанией.

[57] Taylor J. Decision Management Systems: A Practical Guide to Using Business Rules and Predictive Analytics. — IBM Press, 2011.

[58] Davenport T. How Companies Make Better Decisions // International Institute of Analytics, 2010. URL: [www.sas.com/decisions](http://www.sas.com/decisions).

[59] Ch'ananda S. The Origin of Go // British Go Journal. 1990. Vol. 79, Spring. P. 18–19.

[60] Kim J. and Kim H. The Advantage of Playing First in Go // Chance. 1995. Vol. 8, no. 3. P. 22–26.

[61] На русском языке изданы такие книги Джона Готтмана: Карта любви. М. : Эксмо, 2011 и Мужчины и женщины с одной планеты. М. : Эксмо, 2012. *Прим. ред.*

[62] Интервью с Готтманом приведено на сайте [http://www.edge.org/3rd\\_culture/gottman05/gottman05\\_index.html](http://www.edge.org/3rd_culture/gottman05/gottman05_index.html).

[63] Информация о рейтингах FICO взята из FICO Credit Score Chart. URL: [www.myfico.com/crediteducation/whatsinyourscore.aspx](http://www.myfico.com/crediteducation/whatsinyourscore.aspx); Credit Score in the United States // Wozniacka M. and Sen S. Credit Scores: What You Should Know About Your Own // PBS Frontline, November 2004. URL: <http://www.pbs.org/wgbh/pages/frontline/vshows/credit/more/-scores.html>.

[64] Satariano A. and Burrows P. Apple's Supply Chain Secret? Hoard Lasers // Bloomberg BusinessWeek. November 03, 2011. URL: <http://www.businessweek.com/magazine/apples-supplychain-secret-hoard-lasers-11032011.html>.

[65] Описание вакансий на сайте <http://jobs.apple.com/index.ajs?BID=1&method=mExternal.showJob&RID=81197> (дата обращения: 5.01.2012).

[66] Huff D. How to Lie with Statistics. New York : Norton, 1954.

[67] Lu J. Predicting Customer Churn in the Telecommunications Industry: An Application of Survival Analysis Modeling Using SAS. SAS User Group International Proceedings, 2002.

[68] Drye T., Wetherill G., and Pinnock A. When Are Customers in the Market? Applying Survival Analysis to Marketing Challenges // Journal of Targeting, Measurement and Analysis for Marketing. 2001. Vol. 10. P. 179–188.

[69] Hektner J., Schmidt J., Csikszentmihalyi M. Experience Sampling Method: Measuring the Quality of Everyday Life. Thousand Oaks, CA : Sage, 2007.

[70] Csikszentmihalyi M. and Larson R. Being Adolescent: Conflict and Growth in the Teenage Years. New York : Basic Books, 1984.

[71] Rohini Chowdhury. ‘Eureka!’: The Story of Archimedes and the Golden Crown. Long, Long Time Ago. URL: [http://www.longlongtimeago.com/lta\\_greatdiscoveries\\_archimedes-eureka.html](http://www.longlongtimeago.com/lta_greatdiscoveries_archimedes-eureka.html); John Monahan. Archimedes Coins ‘Eureka!’ in the Nude—and Other Crazy Ah-Ha Moments of Science // Scientific American, December 7, 2010. URL: <http://www.scientificamerican.com/article.cfm?id=archimedes-coins-eurekaeureka>.

[72] Root-Bernstein R. and M. Sparks of Genius: The Thirteen Thinking Tools of the World’s Most Creative People. Boston : Houghton Mifflin, 1999. P. 4.

[73] Rae C. The Enchanted Prison of Writing. Seoul : Sisain Books, 2009. P. 394.

[74] Идеи Гегеля, конечно, глубоки и сложны для понимания. Детальное обсуждение этих концепций приведено в книге: Rockmore T. Hegel, Idealism, and Analytic Philosophy. New Haven : Yale University Press, 2005.

[75] Существует много версий этой истории, но наиболее убедительную изложил Дэн Пауэр в выпуске DSS News 10.11.2002 г. URL: <http://dssresources.com/newsletters/66.php>.

[76] John Earle, процитировано в блоге: Riggs Eckleberry. Of Interest. 21.12.1998. URL: [http://www.riggs.com/archives/1998\\_12\\_01\\_OIarchive.html](http://www.riggs.com/archives/1998_12_01_OIarchive.html).

[77] Ask Dr. Math FAQ: About Pi, математический форум: <http://mathforum.org/dr.math/faq/faq.pi.html>; Facts About Pi, <http://mathforum.org/library/drmath/view/57543.html>. Расчет числа пи: <http://mathforum.org/library/drmath/view/55815.html>. Докомпьютерная история числа пи: <http://personal.bgsu.edu/~carother/pi/Pi2.html>. Прим. ред.

[78] Acheson D . 1089 and All That // A Journey into Mathematics. Oxford : Oxford University Press, 2002. P. 14.

[79] Пять самых знаменитых чисел в математике, объединенных в уравнении  $e^{i\pi} = -1$ . *Прим. перев.*

[80] Benford's Law // Browne M. Following Benford's Law, or Looking Out for No. 1 // New York Times, August 4, 1998. URL: <http://www.nytimes.com/1998/08/04/science/following-benford-s-law-or-looking-out-for-no-1.html?pagewanted=all&src=pm>; Benford's Law // Wolfram MathWorld. URL: <http://mathworld.wolfram.com/BenfordsLaw.html>; Hill T. P. The First-Digit Phenomenon // American Scientist, July–August 1998; Nigrini M. J. Benford's Law. URL: <http://www.nigrini.com/benfordslaw.htm>.

[81] Snowdon D. et al. Linguistic Ability in Early Life and Cognitive Function and Alzheimer's Disease in Late Life // JAMA. 1996. Vol. 275. P. 528–532; Landmark Study Links Cognitive Ability of Youth with Alzheimer's. URL: <http://archive.hhs.gov/news/press/1996pres/960220b.html>.

[82] Janes J. Review of David Snowdon's Aging with Grace // Library Journal. June 2001. Vol. 126, no. 11. P. 96.

[83] Sharples T. Can Language Skills Ward Off Alzheimer's? A Nuns' Study // Time. July 09, 2009. URL: <http://www.time.com/time/health/article0,8599,1909420,00.html#ixzz1p7bYUKSu>.

[84] Переписка по электронной почте с Энтони Вьелем 9.11.2011 г.

[85] Аристотель. Никомахова этика (серия «Философы Греции»). М. : Эксмо-Пресс, 1997.

[86] Niederman D. and Boyum D. What the Numbers Say: A Field Guide to Mastering Our Numerical World. New York : Broadway Books, 2003. P. 233.

[87] Rosa L., Rosa E., and Sarner L. A Close Look at Therapeutic Touch // Journal of American Medical Association. 1998. Vol. 279, no. 13. P. 1005–1010; Hubbard D. How To Measure Anything. Hoboken, NJ : Wiley, 2010. P. 13–17.

[88] Courcey K. Therapeutic Touch: Further Notes. URL: [www.-quackwatch.org/01QuackeryRelatedTopics/tt2.html](http://www.-quackwatch.org/01QuackeryRelatedTopics/tt2.html).

[89] Приведено по: The Phenom List — The Top Big Data Scientists: DJ Patil. URL: [http://thephenomlist.com/Big\\_Data\\_Scientists/DJ\\_Patil](http://thephenomlist.com/Big_Data_Scientists/DJ_Patil) (дата обращения: 23.02.2012). Остальная информация взята из интервью с Патилом.

[90] Huff D. How to Lie with Statistics. New York : Norton, 1954.

[91] Campbell S. Flaws and Fallacies in Statistical Thinking. Englewood Cliffs, NJ : Prentice-Hall, 1974.

[92] Good P., Hardin J. Common Errors in Statistics. New York : Wiley, 2003.

[93] Издана на русском языке: Талеб Н. [Одуроченные случайностью. О скрытой роли шанса в бизнесе и в жизни](#). М. : Манн, Иванов и Фербер, 2012. 320 с. *Прим. ред.*

[94] Isaac R. The Pleasures of Probability. New York : Springer, 1995.

[95] Издана на русском языке: Млодинов Л. (Не)совершенная случайность. Как случай управляет нашей жизнью. М.: Livebook/Гаятри, 2010.

[96] См. <http://ocw.mit.edu/courses/mathematics/18-05-introduction-to-probability-and-statistics-spring-2005/>.

[97] Информация о магистерской программе по аналитике в Университете Северной Каролины получена на встречах с администраторами и преподавателями факультета, а также из журнала Institute for Advanced Analytics, North Carolina State University. URL: [analytics.ncsu.edu](http://analytics.ncsu.edu).

[98] Эти предварительные цифры взяты из доклада State of Business Intelligence in Academia 2011, представленного Барбарой Виксом на III Конгрессе по бизнес-разведке, состоявшемся в Орландо (штат Флорида) 16 декабря 2012 года.

[99] Цитируется по книге: Dietz T. and Kalof L. Introduction to Social Statistics: The Logic of Statistical Reasoning. New York : Wiley, 2009. P. 28.

[100] Campbell. Flaws and Fallacies in Statistical Thinking. 108 p.

[101] Chabris C. and Simons D. The Invisible Gorilla and Other Ways Our Intuitions Deceives Us. New York : Crown, 2010. P. 163.

[102] Kohler H. Statistics for Business and Economics. London : Longman, 1988. Из книг на русском языке можем предложить относительно простую книгу по статистике: Минько А. А. Статистика в бизнесе. Руководство менеджера и финансиста. М. : Эксмо, 2008. *Прим. ред.*

[103] Обзор сайта Amazon.com А 'Umea University' student (Sweden) give ratings на 24.08.1999 г. (дата обращения: 30.12.2012).

[104] Lewis M. Moneyball: The Art of Winning an Unfair Game. New York : Norton, 2003.

[105] SN Names the 20 Smartest Athletes in Sports // The Sporting News. Sept. 23, 2010. URL: <http://aol.sportingnews.com/mlb/feed/2010-09/smart-athletes/story/sport-ing-news-names-the-20-smartest-athletes-in-sports>.

[106] Lewis M. The No-Stats All Star // New York Times. February 13, 2009. URL: [www.nytimes.com/2009/02/15/magazine/15Battier-t.html](http://www.nytimes.com/2009/02/15/magazine/15Battier-t.html).

[107] Frei F. and Perlberg M. Discovering Hidden Gems: The Story of Daryl Morey, Shane Battier, and the Houston Rockets (B) // Harvard Business School case study. Boston: Harvard Business Publishing. September 2010, no. 1.

[108] Из частных бесед.

[109] Surveying the Economic Horizon: A Conversation with Robert Shiller // McKinsey Quarterly. April 2009. URL: [http://www.mckinseyquarterly.com/Surveying\\_the\\_economic\\_horizon-A\\_conversation\\_with\\_Robert\\_Shiller\\_2345](http://www.mckinseyquarterly.com/Surveying_the_economic_horizon-A_conversation_with_Robert_Shiller_2345).

[110] Olive D. Getting Wise Before That ‘One Big Mistake’ // Toronto Star. December 17, 2007.

[111] Издана на русском языке: Дахигг Ч. Сила привычки. Почему мы живем и работаем именно так, а не иначе. М. : Карьера Пресс, 2013. 416 с.

[112] Предисловие Гэри Лавмена к книге: Дэвенпорт Т., Харрис Д. Аналитика как конкурентное преимущество. Новая наука побеждать. М. : BestBusinessBooks, 2010. 264 с.

[113] Более подробно с сюжетом фильма и действующими лицами можно познакомиться на сайте <http://business-ethics.com/2011/11/23/0953-margin-call-a-small-movie-unveils-big-truths-about-all-street/>.

[114] Fahey L. Exploring ‘Analytics’ to Make Better Decisions: The Questions Executives Need to Ask // Strategy and Leadership. 2009. Vol. 37, no. 5. P. 12–18.

[115] Информация для этого примера получена из нескольких интервью с Энн Робинсон, а также сборника деловых ситуаций Блейка Джонсона Leveraging Enterprise Data and Advanced Analytics in Core Operational Processes: Demand Forecasting at Cisco, разработанного на кафедре менеджмента и технологий Стэнфордского университета.

[116] Harrington K. Seeing the Future in Value Chain Management // Analytics Magazine. March/April 2010.

[117] Harrington K. Seeing the Future in Value Chain Management // Analytics Magazine. March/April 2010.

## **Над книгой работали**

Руководитель направления *Артем Степанов*  
Художественный редактор *Алексей Богомолов*  
Ответственный редактор *Ольга Копыт*  
Редактор *Ирина Зубкова*  
Дизайн обложки PUNK YOU BRANDS  
([www.punk-you.ru](http://www.punk-you.ru))  
Верстка *Юлия Рахманина*  
Корректоры *Лев Зелексон, Наталья Лазариди*

**ООО «Манн, Иванов и Фербер»**  
[mann-ivanov-ferber.ru](http://mann-ivanov-ferber.ru)

Электронная версия книги  
подготовлена компанией Webkniga, 2014  
[webkniga.ru](http://webkniga.ru)

## **Максимально полезные книги от издательства «Манн, Иванов и Фербер»**

Если у вас есть замечания и комментарии к содержанию, переводу, редактуре и корректуре, то просим написать на [be\\_better@m-i-f.ru](mailto:be_better@m-i-f.ru), так мы быстрее сможем исправить недочеты.

### **Наши электронные книги:**

<http://www.mann-ivanov-ferber.ru/ebooks/>

### **Заходите в гости:**

<http://www.mann-ivanov-ferber.ru/>

<http://blog.mann-ivanov-ferber.ru/>

<http://www.facebook.com/mifbooks>

<http://vk.com/mifbooks>

<https://twitter.com/mifbooks>

[Дерево знаний](#)

[Предложите нам книгу.](#)

[Ищем правильных коллег](#)

### **Для корпоративных клиентов:**

[Полезные книги в подарок](#)

[Корпоративная библиотека](#)

[Книги ищут поддержку.](#)

# Оглавление

[ГЛАВА 1. Почему аналитические способности нужны каждому](#)

[ГЛАВА 2. Формулирование проблемы](#)

[ГЛАВА 3. Решение проблемы](#)

[ГЛАВА 4. Результаты и необходимые меры](#)

[ГЛАВА 5. Креативность в количественном анализе](#)

[ГЛАВА 6. Развитие навыков количественного анализа](#)

[ГЛАВА 7. Работа с квантами](#)

[Итоговые выводы об аналитическом мышлении](#)

[Благодарности](#)

[Об авторах](#)