Содержание

Предисловие

Добро пожаловать!

Эта книга – учебник, посвященный *причинно-следственным выводам*, особенно выводам, сделанным на основе *наблюдений*. Представьте, что мы хотим узнать, приводит ли изменение переменной X к изменению переменной Y, и если да, то в какой степени, но не можем или не желаем проводить натурный эксперимент. Как нам спланировать исследование, чтобы ответить на этот вопрос? На самом деле это очень сложная задача. Многие исследователи говорили мне прямо, что она не выполнима и не стоит тратить время напрасно[1].

Именно поэтому я и взялся писать эту книгу. Я намерен рассказать, в чем заключается главный вопрос причинно-следственного исследования и как пройти сложный путь к ответу. Но для этого мне придется начать издалека – задолго до математических формул и доказательств.

В причинно-следственном выводе есть формальная часть, и в этой книге мы обязательно ее обсудим. Но если вам приходилось встречаться с людьми, которые планируют и проводят причинно-следственные изыскания, то вы наверняка заметили, что они рассуждают в первую очередь интуитивно, а не математически. Они начинают с предположений о реальном мире, о том, насколько они разумны, а также о том, какая история стоит за имеющимися данны-

[1] Я думаю, они просто пытаются облегчить себе жизнь.

ми. Лишь разобравшись с предпосылками и допущениями, они переходят к уравнениям и статистическим параметрам. Продумать хорошее исследование и доказать (и особенно понять) теоремы в области статистики – это разные задачи! Я уверен, что их следует рассматривать именно в таком порядке.

В первой части книги мы рассмотрим понятие *идентификации* – процесса выяснения, в какой части данных содержится ответ, чтобы можно было приступить к целенаправленному поиску. На этом этапе исследования вы должны применить все имеющиеся знания о том, как устроен и работает мир. Из первой части вы узнаете о том, что вам нужно сделать, чтобы ответить на вопрос исследования – этот порядок действий называется *планом исследования*. И это еще не все! Вы узнаете, как определить, стоит ли верить причинно-следственным выводам, сделанным другими людьми. Что им нужно было сделать, чтобы доказать свой вывод, и сделали ли они это? Первая часть книги великолепна. Я уверен, вам она понравится. На данный момент я прочитал много книг о причинно-следственных выводах и не думаю, что там есть что-то подобное. Моей маме понравилась первая часть этой книги, а ведь у нее аллергия на статистику.

Вторая часть книги более формальная. В ней я рассматриваю стандартный набор инструментов, которым обычно пользуются исследователи в ходе причинно-следственного вывода. Некоторые из них представляют собой специальные инструменты статистики, такие как регрессия. Но есть и обычные схемы исследований, созданные для вполне конкретных задач и впоследствии оказавшиеся удобным универсальным инструментом.

Хотя я называю это более формальным подходом, но все равно делаю упор на интуитивное понимание. Я не собираюсь убеждать вас в достоинствах какого-либо метода, математически доказывая, что он работает. Моя цель – помочь вам понять, что эти методы делают, почему они полезны и когда их можно использовать. Затем я постараюсь научить вас применять эти методы. В наше время это означает программирование на R, Stata или Python. Во второй части книги вы найдете много разных примеров кода.

Я хочу, чтобы после прочтения второй части этой книги вы почувствовали себя компетентными – го-

товыми реализовать эти методы самостоятельно и понять, что происходит, когда они используются. Если вы приложите немного усилий, у вас это получится.

Разумеется, я не могу судить объективно, но думаю, что эта книга очень увлекательная. Мне было весело ее писать, и я думаю, что читать ее будет настолько интересно, насколько это вообще возможно для учебника по причинно-следственному выводу. Я постарался охватить все методы и способы исследований. Причинно-следственный вывод – это область, в которую внесли важный вклад медицина, эпидемиология, экономика, социология, политология, финансы, наука о данных и т. д., почти до бесконечности. Сам-то я родом из экономики, но могу гарантировать, что уделю внимание всем остальным наукам. Надеюсь, мы пройдем этот путь вместе.

Примечание для студентов

Когда я учился в колледже, меня познакомили с простейшими методами причинного вывода. Но даже они казались мне чрезвычайно мощными. И не случайно: если правильно их использовать, они могут превратить вас из потребителя знаний в производителя. Вы можете найти ответ на вопросы, на которые больше никто не знает ответа. Вы можете понять, как на самом деле устроен мир. Я думаю, это чертовски круто!

Я хочу, чтобы вы помнили об этой силе, читая книгу. На каждой странице я хочу донести до вас простую истину: методы, описанные в этой книге, не похожи на зубило и молоток, которыми можно долбить данные до тех пор, пока наружу не покажется ответ[2]. Они предназначены для расширения представлений исследователя о мире. Они опираются на то, что мы знаем, и говорят нам, как мы можем узнать больше и какие предположения нам нужно для этого сделать.

Поэтому не рассматривайте причинный вывод как формальную задачу со строгими правилами. Да, в ней есть технические элементы, и вам придется проделать некоторую техническую работу, но это не главное. Воспринимайте причинный вывод как задачу рассуждения. Что мы знаем? На какую теорию можно положиться? И как можно использовать все эти знания, чтобы превратить запутанную и не

[2] Если бы эти методы можно было просто применять вслепую, без реального понимания, то вы бы не успели дочитать эту книгу, а кто-то уже написал бы компьютерную программу, которая автоматически делает причинно-следственный вывод, и вы бы зря потеряли время.

всегда надежную массу данных в нечто полезное и познавательное?

Примечание для преподавателей

Я профессор. Мне часто пытаются предложить новые учебники, поэтому я знаю, как это бывает. Если верить авторам, каждый учебник новый, необычный и содержит примеры из реальной жизни, которые увлекут учеников – в отличие от скучных старых книг, которые студенты ненавидят! Затем я начинаю читать и вижу, что это ровно то же самое, что и другие книги, но на этот раз с другими стоковыми фотографиями и карикатурами из жизни Нью-Йорка на полях. Однако моя книга действительно особенная по нескольким основным критериям. Я обещаю! Главное, в ней нет раздражающих стоковых фотографий. Но различия не просто косметические, они структурные. Я подозреваю, что вы либо подумаете, что эта книга – идеальное учебное пособие, которого вы ждали все эти годы, либо подумаете, что она совершенно ошибочна и рассказывает исключительно неправильные вещи[3].

Я старался, чтобы вы получили учебник, который читается довольно легко для такой сложной темы, но без ущерба для строгости изложения и широты охвата. Уровень сложности таков, что он наиболее подходит для изучения причинно-следственной связи, методов наблюдения или прикладной эконометрики. В зависимости от программы его также можно использовать в версиях этих курсов на уровне магистратуры. В сочетании с более углубленными формальными материалами этот учебник может быть полезен аспирантам. Первая часть также подходит для занятий по статистике в старших классах школы, если преподаватель пожелает рассмотреть причинно-следственную связь.

Есть много способов использовать эту книгу, но я сам организую курс так, чтобы первую треть отведенного времени потратить на обсуждение понятия идентификации и приемов работы с причинно-следственными диаграммами. Остальная часть курса посвящена конкретным методам с возможностью рассмотреть и воспроизвести реальные исследования, в которых используются эти методы. Задания и видеоматериалы доступны на сайте учебника www.theeffectbook.net.

[3] Не говоря уже обо всем, что я упустил. Приношу извинения, если для вашего любимого причинно-следственного или статистического метода не нашлось места. Уверяю вас, я удалял из текста каждый отрывок со слезами и болью.

В чем главные отличия этого учебника от остальных?

Первое отличие – уровень математической сложности. По сравнению с существующим набором учебников по причинному выводу (и конечно же, с существующим набором учебников по эконометрике), в этом учебнике очень мало уравнений. По моему опыту, даже среди студентов, которые хорошо разбираются в математике и без труда решают математические задачи, лишь небольшая часть их приходит к пониманию методов статистики через уравнения[4]. Но если вы знаете, как на самом деле работает метод, то уравнения обретают глубокий и ясный смысл, выходящий за рамки домашнего задания, которое нужно решить.

В этой книге я ставлю концептуальное понимание схемы исследования намного выше всего остального. Второй приоритет – возможность немедленного *применения*. Это означает наличие примеров кода для реализации методов, чтобы вы могли в деталях видеть, как они работают. Дополнительная выгода в том, что правильно работающий код вселяет в студентов уверенность. Кроме того, я могу представить более продвинутые и современные методы, чем в обычном учебнике, который сосредоточен только на фундаментальной математике.

С другой стороны, за все приходится платить, и этот учебник не подготовит студентов к выводу доказательств теорем на курсе статистических методов для аспирантов или к разработке собственных оценивателей причинно-следственных связей. Однако я полагаю, что многие студенты, даже те, которые занимаются исследованиями, в любом случае в этом не нуждаются.

Второе отличие – теоретический подход к причинности. В этой книге основное внимание уделяется не только методам, но и концепциям причинно-следственного вывода. Поэтому у студентов будет понимание теоретической базы причинно-следственного вывода!

На самом деле для курса причинно-следственного вывода можно выбрать один из двух фундаментальных базисов. Первый – это концепция потенциальных исходов, связанная в первую очередь с Дональдом Рубином, а второй – концепция структурно-каузальной модели (или причинно-

[4] Но студенты, которые извлекают реальное понимание из уравнений (как и я), с большей вероятностью станут профессорами, и, следовательно, это проблема…

следственной диаграммы), предложенная Джудой Перлом.

Здесь я принимаю два потенциально спорных решения. Во-первых, я почти полностью отказался от концепции потенциальных исходов. Логика потенциальных исходов, конечно, несколько раз упоминается в книге, но я не использую формальную модель. Почему? То, в чем хороши потенциальные исходы, – решение проблемы отсутствующих данных, обработка средних значений эффекта воздействия, выражение условий игнорирования – я либо не использую, либо делаю способами, которые, как мне кажется, более очевидны для студентов. Раньше я преподавал потенциальные исходы студентам и обнаружил, что в данном случае математика мешает пониманию. Я воспользовался правом автора и оставил в книге то, что мне нравится!

Поэтому я в основном использую методику причинно-следственных диаграмм. Во-вторых, у коллег-преподавателей могут возникнуть вопросы к тому, что я называю «упрощенными диаграммами». Действительно, в них нет расчетов, и я делаю некоторые вещи, которые полезны для понимания, но не являются частью построения формальной причинно-следственной диаграммы; например, я время от времени включаю в диаграмму функциональные члены.

В обоих случаях это означает, что студентам, которые хотят продолжить углубленное изучение этих методов, придется проделать некоторую дополнительную работу. Но, надеюсь, они очень хорошо поймут, чего пытаются добиться. Полагаю, вы согласитесь со мной в том, что хотя вещи, которые я упустил, ценны и заслуживают внимания в долгосрочной перспективе, имеет смысл оставить их на потом. Лучше усвоить одно понятие хорошо, чем два плохо.

Примечание для всех остальных

Признак действительно хорошего учебника – когда кто-то решает почитать его просто из любопытства. По-настоящему превосходный учебник – когда хочется сесть и прочитать его до конца. Если вы это сделаете, пожалуйста, дайте мне знать. Я не страдаю от заниженной самооценки, но потешить свое эго всегда приятно.

Хотя я уверен, что упускаю из виду некоторых читателей, я ориентируюсь на три типа людей, которые, вероятно, прочитают эту книгу, не будучи студентами. И для этих трех типов людей у меня есть несколько советов относительно чтения книги.

Специалистам по обработке данных или бизнес-аналитикам с небольшим опытом работы в области причинного вывода, которые хотят ответить на причинно-следственные вопросы: рад, что вы здесь! В этой книге будет использован совершенно иной подход к анализу данных, чем тот, к которому вы, вероятно, привыкли. По большей части наука о данных и бизнес-аналитика – это области, которые в первую очередь определяются данными[5]. Вы ищете закономерности в данных и смотрите, о чем они вам говорят. Ваша цель обычно состоит в том, чтобы сделать определенный прогноз или измерить параметры данных.

[5] Не всегда! Но обычно это так.

Причинно-следственные исследования, с другой стороны, основаны на теории. Вы начинаете с ревизии того, что знаете, и используете это знание для интерпретации данных. Ваша цель – использовать данные, чтобы шире приоткрыть завесу над процессами и законами, которые генерировали данные.

Прочитав эту книгу, вы освоите не только новые методы, но и совершенно новый подход к исследованию! На самом деле это очень большое достижение, потому что всегда трудно начинать использовать принципиально новые ментальные схемы. Для вас ключевыми главами книги будут вторая и пятая. Возможно, вам даже придется перечитать их пару раз, пока не возникнет настоящее понимание. Как только вы этого добьетесь, будет легче, ибо остальное – это методы. С вашим опытом не составит труда выбрать среди них наиболее подходящие. Просто откройте свой разум и позвольте новым знаниям войти в него.

Возможно, вы относитесь ко второму типу читателей – не занимаетесь исследованиями, но хотите понять, как работает причинно-следственный вывод, или научиться лучше интерпретировать и оценивать исследования, в которых он используется: эта книга удобно устроена таким образом, что вы можете узнать все необходимое, не вдаваясь в подробности. Главы с 1 по 9 дадут вам представление о том, как устроены и какие задачи решают исследования, использующие причинно-следственные связи. Вы

научитесь определять, насколько обоснованными с точки зрения причинности являются утверждения отдельных людей или выводы исследований. Даже не будучи исследователем, вы научитесь строить собственные причинно-следственные диаграммы (глава 7), а затем рассуждать о том, что необходимо сделать для подтверждения (идентификации) причинно-следственной связи (главы 8–9). После этого вы можете смело судить о том, насколько обоснованы выводы других людей! Если перспектива столкнуться с математикой вас пугает, возможно, вам удастся пропустить главы 3 и 4, но сперва дайте им шанс и посмотрите, как далеко вы продвинетесь, прежде чем переходить к главе 5.

Вторая часть книги также может быть вам полезна. Если вы не планируете проводить собственное статистическое исследование, вам не нужно будет читать какую-либо главу до конца. Но если есть исследование, в котором вы хотите разобраться, и его схема соответствует одной из схем, описанных в этой книге, вы можете бегло просмотреть раздел «Как это работает» в начале многих глав второй части, чтобы увидеть, что на самом деле пытается сделать этот проект. А если вы хотите проявить фантазию и самостоятельно интерпретировать некоторые из их результатов, вам поможет раздел «Как это реализовано».

Исследователям, имеющим опыт причинного вывода и желающим просмотреть стандартные методы или лучше понять, как они работают, идеально подойдет вторая часть этой книги. Каждая глава, посвященная стандартной схеме причинного вывода, разделена на три части: раздел «Как это работает» познакомит вас с концепциями и теорией, лежащими в основе схемы исследования; раздел «Как это реализовано», вероятно, наиболее похож на описание методов, которые вы уже видели в учебнике по эконометрике; далее следует раздел «Как это делают профессионалы», в котором рассмотрены современные тенденции, проблемы и исправления, о которых вам будет полезно знать. И в любом случае раздел «Как это делают профессионалы» идеально подходит для исследователей, которые давно изучили существующие методы и которым необходимо быть в курсе последних разработок.

Тем не менее я считаю, что даже опытным специалистам не помешает ознакомиться с материалом пер-

вой половины книги о каузальных диаграммах[6]. По моему мнению, предельная ценность изучения каузальных диаграмм для человека, уже обученного анализировать потенциальные исходы, невелика, но это все равно мощный инструмент, который нужно иметь под рукой. И это действительно важный компонент знаний, когда дело доходит до преподавания. Если вы преподаватель, то можете добавить каузальные диаграммы в свой курс, даже если вы не преподаете методы и статистику. Я всегда использовал их на лекциях по экономике системы образования, чтобы студентам было легче понять, в чем суть исследований, о которых они читают в научных статьях и учебниках.

[6] И с главой 5 об идентификации тоже – вы уже знаете эту тему, но я думаю, что глава получилась очень хорошей и вы можете открыть для себя несколько новых взглядов на идентификацию.

Комментарий перед началом чтения

Эта книга была написана в беспокойное время (во всех смыслах этого слова). Я взялся за дело в феврале 2020 года, когда моему ребенку было шесть месяцев, и почти за месяц до того, как пандемия коронавируса охватила Соединенные Штаты. Теперь, когда я дописываю книгу, мой ребенок называет каждое животное «кошкой», а пандемия никак не заканчивается, но я только что получил вторую прививку вакцины в пятницу.

Написание книги было способом отвлечься и продуктивно потратить время в этот странный и очень напряженный год. Если вы читаете это в будущем и понятия не имеете, о чем я говорю, я уверен, что вы можете найти множество книг по истории о событиях в 2020 году. Возможно, по меркам вашего времени, 2020 год был не таким уж беспокойным, и в этом случае я надеюсь, что у вас все в порядке, и потрясен тем, что у вас есть время и силы читать учебники по причинно-следственным выводам.

Вы также можете считать эту книгу замысловатым порождением коллективного разума экономического сообщества в Твиттере (и других академических сообществ, с которыми оно пересекается). Из Твиттера я не только много узнал о причинно-следственных выводах, но и обнаружил, что это прекрасная интеллектуальная среда, и участие в ней стало отличным мотиватором. Очень хочется поделиться своими учебными материалами и услышать, как тысячи людей говорят, что они им нравятся. Поэтому я потратил целый год на написание книги. Ради лести я готов на что угодно.

ЧАСТЬ I

Схема исследования

1
Планирование исследования

1.1. У меня есть вопрос

Как устроен мир? Это невероятно обширный вопрос, не так ли? И наши ответы на этот вопрос всегда будут немного неполными.

Эта незавершенность является проклятием исследователя, но также и его благословением. Конечно, мы никогда не будем знать всего в точности[1]. Но это также означает, что у нас никогда не закончатся вопросы, на которые нужно ответить. Для определенного типа людей поиск ответов на вопросы выглядит как увлекательный способ провести земную часть бытия. Возможно, вы именно такой человек. Я тоже из их числа.

Исследовательский вопрос – это вопрос, который перед вами стоит, на который вы планируете ответить или, по крайней мере, попытаться ответить, проведя исследование. Так просто. Или так сложно – как посмотреть. Хороший исследовательский вопрос должен быть четко сформулирован, на него

[1] Квантовая механика – пожалуй, самая точная из существующих научных областей, с точностью измерений и предсказаний до более чем дюжины десятичных знаков. Но даже в этом случае «с точностью до четырнадцати знаков после запятой» – это не то же самое, что «точно».

должен существовать ответ, и он должен быть понятен – эти требования не всегда легко соблюсти! Подробнее об этом мы поговорим в главе 2.

В качестве примера предположим, что наш исследовательский вопрос звучит так: «Если добавить еще одну полосу движения, уменьшатся ли дорожные заторы?»

Это вопрос о том, как устроен мир. К сожалению, дорожное движение является частью мира. И ответ на этот вопрос можно найти с помощью исследования.

Каким должно быть исследование? Хорошо спланированное исследование способно дать однозначный ответ на вопрос, ради которого мы его затеваем. Это определение кажется очевидным, но на самом деле требует довольно много размышлений и усилий.

И это настоящее искусство.

Как провести исследование так, чтобы по его окончании у вас был ответ на исследовательский вопрос?

Вот о чем эта книга.

1.2. Эмпирические исследования

«Исследование, способное ответить на вопрос, на который пытается ответить» – такое определение, по правде говоря, может означать почти что угодно.

Существует много видов исследований. Вы можете прочитать в книгах, что люди думают по вашему вопросу («что говорят эксперты по дорожному движению о влиянии дополнительной полосы на загруженность шоссе?»), и обобщить полученные знания. Вы можете подойти к этому вопросу с философской точки зрения («если предположить, что люди пытаются сократить время в пути, как они отреагируют на дополнительную полосу движения по шоссе?»). Оба подхода представляют собой исследование. Эта книга будет посвящена эмпирическим исследованиям, в частности *количественным эмпирическим исследованиям*.

Эмпирическое исследование. Исследование, в котором используются структурированные наблюдения, полученные из реального мира.

Эмпирическое исследование – это любое исследование, в котором для ответа на исследовательский вопрос используются структурированные наблюдения из реального мира. Поэтому, вместо того чтобы *рассуждать* о поведении водителей, если бы им предоставили дополнительную полосу движения, мы *наблюдаем* за выбором, который делают водители. Способы наблюдения зависят от контекста и задачи. Возможно, мы решим расспросить водителей о том, как они принимают решения. Или предпочтем со-

брать большой набор данных о нарушениях правил дорожного движения и показателях транспортных потоков на автомагистралях.

Количественные эмпирические исследования – это просто эмпирические исследования, в которых используются количественные измерения (обычно числа). Больше чисел, меньше разговоров.

Количественное эмпирическое исследование, как и любое другое исследование, может оказаться непростым делом! Нужно провести точные измерения и правильно интерпретировать результаты. Статистика – сложная область науки.

Одна особенно неприятная проблема количественных эмпирических исследований заключается в том, что числа, которые мы получаем, часто говорят нам не совсем то, что мы хотим знать.

Допустим, мы решили изучить влияние дополнительной полосы, сравнивая двухполосные и трехполосные шоссе. Но ведь нас не интересует абсолютная величина трафика на трехполосных и двухполосных шоссе. Перед нами стоит вопрос, сможем ли мы снизить интенсивность движения, превратив двухполосное шоссе в трехполосное! Но как бы мы этого ни хотели, доступные количественные наблюдения не дают прямого ответа на вопрос. У нас есть только двухполосные и трехполосные дороги. У нас нет волшебного шоссе, которое говорит нам, какой объем движения был бы, *если бы* мы сделали это двухполосное шоссе на одну полосу шире.

Эта проблема представляет собой большую головную боль для исследователей. Как нам быть, если доступные наблюдения на самом деле не отвечают на исследовательский вопрос?

Оказывается, если все сделать правильно, часто удается собрать нужные числа и проделать с этими числами правильные манипуляции, чтобы получить ответ на нужный вопрос. Но за все приходится платить. Нужно как минимум разработать подходящую *схему исследования*.

1.3. Почему исследованиям нужна схема

Почему так важно правильно выстроить схему исследования? Давайте представим, что произойдет, если мы выберем неправильную схему.

Мы хотели узнать, как изменится дорожное движение, если добавить к шоссе еще одну полосу. Наи-

более очевидный подход состоит в том, чтобы просто сравнить показатели дорожного движения на шоссе с большим количеством полос с показателями движения на шоссе с меньшим количеством полос.

С виду вполне разумно. Но затем вы сравниваете данные, и оказывается, что чем больше полос, тем больше трафик! Удивительно. А теперь задайте себе вопрос: почему на этих автомагистралях вообще больше полос движения? Может, дело в том, что самые востребованные маршруты изначально делают шире, и поэтому неудивительно, что большее количество полос связано с большим трафиком! Конечно, возможно, дополнительные полосы действительно приведут к увеличению трафика[2]. Но чтобы понять, что ваш первоначальный анализ был неправильным, и решить, что нужно делать, требуется тщательно спланированное исследование.

Отсутствие надежной схемы исследования часто проглядывает в результатах. Доводилось ли вам замечать, например, что исследования о здоровом питании периодически приходят к противоположным выводам? Когда я был ребенком в 90-х, нам полагалось есть пищу с высоким содержанием углеводов и низким содержанием жиров, а замороженный йогурт и сухие пресные бублики считались чрезвычайно полезными. Теперь диетологи думают совершенно иначе. А бокал вина каждый вечер полезен или нет? Или кофе? Или масло вместо маргарина? Или сахар вместо кукурузного сиропа[3]? Основные соображения о том, какую пищу полезно есть, вряд ли сильно изменятся, но отдельные научные исследования просто шарахаются из стороны в сторону!

Отчасти в этом мы можем винить новости, которые сверх всякой причины раздувают результаты исследований или совершенно неверно их интерпретируют. Но отчасти это связано с тем, что многие исследования в области питания не придерживаются схемы, позволяющей ответить на вопрос «Какая пища сделает вас здоровее?». Похоже, что разные исследования дают разные ответы на этот вопрос, потому что на самом деле не отвечают на него! У выражения 2 + 2 есть только один ответ[4], но если вы очень сильно хотите получить что-то совершенно отличное от 2 + 2, вы вполне можете вернуться с ответом 6, или 1, или −52. Затем люди просыпаются и видят заголовки новостей о том, что по мнению ученых 2 + 2 = −52.

[2] Исследователи в области транспорта обычно утверждают, что увеличение количества полос по крайней мере приводит к увеличению числа поездок! См., например, Milam et al. (2017).

[3] Вы обращали внимание, что некоторые бренды конфет и газированных напитков рекламируют использование «настоящего сахара», как будто отсутствие кукурузного сиропа делает сахар здоровой пищей? Этот факт мало связан с тем, о чем мы здесь говорим, но, черт возьми, я ненавижу такие манипуляции.

[4] Во всяком случае, в обычной арифметике. Высшая математика способна на неожиданные шутки.

Питание – хорошая тема для примера, потому что на самом деле это не вина самих исследователей в области питания. Когда дело касается питания, просто невозможно выстроить надежную схему исследования[5]. Таким образом, вы получаете целую область науки с шаткой исследовательской структурой. И что это нам дает? Непоследовательные результаты, на которые, к сожалению, люди привыкли не обращать особого внимания сегодня, потому что знают, что завтра все может измениться.

Построить схему исследования нелегко, и даже наличие правильно поставленного вопроса не означает, что существует простой способ найти ответ. Но самое худшее, что может случиться, – обнаружить, что мы не знаем, как получить ответ. Тогда мы будем знать хотя бы это.

Лучшее, что может случиться, – это найти способ ответить на исследовательский вопрос. И сделать это. А потом получить Нобелевскую премию.

[5] Очень сложно точно измерить то, что люди едят, очень трудно отделить влияние одних продуктов от других, очень трудно отделить влияние еды от влияния тех, кто побудил вас принять решение съесть эту еду, и т. д. и т. п.

1.4. Чему научит эта книга

Эта книга преследует несколько целей.

Первая часть научит вас принципам планирования исследования. В частности, речь пойдет о том, как правильно сформулировать исследовательский вопрос, на который существует корректный ответ, а затем подумать о том, какие количественные эмпирические исследования нужно провести, чтобы ответить на этот вопрос. Что вам нужно будет измерить? Как убедиться, что ваш метод действительно дает ответ на нужный вопрос?

Затем, во второй части книги, вы познакомитесь с некоторыми базовыми методами – своего рода «набором инструментов» для разработки исследований с использованием данных наблюдений (т. е. ответа на вопрос о причинно-следственной связи без проведения эксперимента). Эти методы очень часто применяются в современных исследованиях, поскольку они, как правило, подходят для ответа на широкий круг исследовательских вопросов, а предпосылки, на которых они основаны, хорошо понятны.

Надеюсь, что после прочтения этой книги вы сможете уверенно разработать схему исследования, выяснить, какие данные нужны для ответа на исследовательский вопрос, и определить, какие расчеты нужно выполнить с этими данными.

2
Исследовательские вопросы

2.1. Что такое исследовательский вопрос?

Задавать вопросы легко. Просто заговорите с любым пятилетним ребенком, и он задаст вам десятки вопросов. Гораздо сложнее придумать хороший исследовательский вопрос.

Разница, по крайней мере в случае количественного эмпирического исследования, заключается в том, что исследовательский вопрос – это вопрос, *на который можно ответить*, причем *ответ улучшит ваше понимание того, как устроен мир*.

Оба критерия немного абстрактны. Давайте разберем их подробнее.

Что значит иметь вопрос, на который можно ответить? Это означает, что существует некий набор *свидетельств* (или *доказательств*), исходя из которых, можно получить однозначный и проверяемый ответ. Например, на вопрос «Какой фильм о Джейм-

се Бонде лучше всех?» с научной точки зрения ответить невозможно[1]. Независимо от того, какие свидетельства вы найдете, слово «лучше» отражает настолько размытый критерий, что невозможно даже представить данные, которые помогут ответить на этот вопрос. Допустим, вы сможете убедить каждого человека на Земле, что это «Лунный гонщик», но это все равно не будет ответом на вопрос.

С другой стороны, на вопрос «Какой фильм о Джеймсе Бонде был самым кассовым?» существует однозначный ответ. Достаточно посмотреть на статистику продаж билетов. Доступные свидетельства содержат ответ на этот вопрос.

Хорошо, у нас есть вопрос, на который можно ответить. Но улучшает ли ответ наше понимание того, как устроен мир? Это означает, что исследовательский вопрос в сочетании с ответом должен рассказать вам о чем-то новом, раскрыть более широкий взгляд на мир. Вопрос должен тем или иным способом формировать *теорию*. Ваша теория не обязательно должна быть такой же важной, как теория гравитации или теория эволюции. Это может быть простое умозаключение типа «хлеб сегодня стоит дороже, чем в прошлом году, потому что цены на хлеб в целом со временем растут». Наличие теории просто означает, что где-то скрывается «почему» или «потому что». Даже водород – это теория: она утверждает, что такой материал, как вода, имеет определенные свойства, потому что существует тип атома, который ведет себя определенным образом и имеет определенную структуру.

Возьмем, к примеру, теорию микробов. Теория микробов утверждает, что микроорганизмы, такие как бактерии и вирусы, могут вызывать заболевания. Это объясняет, почему у нас есть определенные болезни, а также почему болезнь может передаваться от одного человека к другому. Мы называем это объяснение теорией вовсе не потому, что сомневаемся в его правильности[2]. Мы говорим, что это теория, поскольку она отвечает на вопрос «почему».

Хороший исследовательский вопрос *ведет нас от теории к гипотезе*, где гипотеза представляет собой конкретное утверждение о том, что мы будем наблюдать в мире, например: «Люди, которые моют руки, болеют реже». То есть исследовательский вопрос должен быть таким, чтобы ответ на него помогал улучшить ваше объяснение «почему». Дру-

[1] Я посмотрел два фильма о Джеймсе Бонде и надеюсь, что ни один из них не считается лучшим.

Теория. Объясняет то, почему происходят явления, которые мы наблюдаем, или иным образом обобщает эти наблюдения на другую ситуацию.

[2] Теории, которые в нашем понимании почти наверняка верны, принципиально ничем не отличаются от теорий, которые почти наверняка ошибочны, например что египетские пирамиды помогали строить инопланетяне.

Гипотеза. Конкретное утверждение о том, что вы увидите в данных.

гими словами, вы можете спросить: «Если я найду результат X, что я могу с ним сделать? Изменит ли он мои представления о мире?» Великие исследовательские вопросы часто возникают из самой теории: «Если таково мое объяснение мироустройства, то что я должен наблюдать? Наблюдаю ли я это на самом деле?»

Здесь очень легко свернуть не туда! Давайте продолжим работать с теорией микробов в качестве примера. Мы могли бы поразмышлять над теорией микробов и задать вопрос: «Хм, интересно, насколько мал самый маленький микроорганизм?» Это исследовательский вопрос, на который можно ответить, располагая нужными данными, и он связан с теорией микробов, так что было бы неплохо знать ответ, верно? Однако, найдя ответ на этот вопрос, мы на самом деле не узнаем ничего нового о том, почему у нас возникают болезни или почему болезнь может передаваться от одного человека к другому[3]. Возможно, это поможет нам лучше понять какую-то другую теорию. В таком случае вопрос о самых маленьких микроорганизмах станет лучшим исследовательским вопросом для другой теории[4].

[3] По крайней мере я так не думаю… Я не биолог.

Но позвольте, разве вопрос «Какой фильм о Джеймсе Бонде был самым кассовым?» улучшает наше понимание того, как устроен мир? Да, если он сочетается с правильной теорией. Возможно, у нас есть теория, согласно которой боевики были наиболее популярны в 1980-е годы. Если задаться вопросом о продажах билетов на фильмы про Бонда на определенном историческом интервале, мы сможем получить немного больше информации о том, является ли эта теория точным объяснением продаж билетов.

[4] И даже если исследование не поможет нам понять какую-либо теорию или расширить наше понимание, иногда стоит им заняться только потому, что это довольно изящно. Нет ничего плохого в любопытстве как таковом.

До сих пор мы начинали с вопроса. Давайте попробуем начать с теории. Допустим, у нас есть теория, согласно которой вашему взрослому любопытству вредит воздействие пассивных развлечений, таких как просмотр телепередач и фильмов. Независимо от того, правда это на самом деле или нет, это утверждение можно назвать теорией – оно объясняет, почему любопытство притупляется с возрастом.

Возникает естественный исследовательский вопрос: «Притупляет ли просмотр телевизора в детстве ваше любопытство во взрослом возрасте?»

Давайте проверим, соблюдаются ли два условия для исследовательских вопросов. Можем ли мы от-

ветить на этот вопрос? Да! Данные, необходимые для ответа на этот вопрос, возможно, трудно найти, но мы по крайней мере можем предположить, что они существуют. Если бы мы рандомизировали группу детей, чтобы они смотрели телевизор разное количество часов, а затем проследили бы за ними до взрослой жизни и каким-то образом измерили их любопытство, это было бы довольно убедительным свидетельством по вопросу нашего исследования[5].

Далее, говорит ли нам этот исследовательский вопрос о том, как устроен мир? Да! Утвердительный ответ на этот вопрос послужил бы довольно серьезным подтверждением нашей теории. Если же мы ответим на исследовательский вопрос так: «Нет, регулярный просмотр телевизора в детстве не притупляет ваше любопытство во взрослом возрасте», то нам будет довольно сложно объяснить снижение любопытства взрослых пассивными развлечениями. Исследовательский вопрос действительно помогает нам выяснить, хороша ли теория.

Хороший тест для проверки, влияет ли исследовательский вопрос на теорию, – это представить, что вы обнаружили неожиданный результат, а затем задаться вопросом, заставит ли он вас изменить ваше понимание мира. Допустим, вместо исследовательского вопроса «Притупляет ли просмотр телевизора в детстве ваше любопытство во взрослом возрасте?» мы используем вопрос: «Имеют ли дети, которые много смотрят “Улицу Сезам”, меньший уровень любопытства в дальнейшем?» Мы провели исследование и обнаружили, что на самом деле у детей, которые смотрят «Улицу Сезам», уровень любопытства выше! Ой, кажется, у нас проблема. Должны ли мы изменить нашу теорию, исходя из новых данных? Нет, мы можем придерживаться прежней теории, просто придумав объяснение нестыковки. Мы отвечаем, что «Улица Сезам» может отличаться от большинства телепередач. И есть вероятность, что дети, обожающие смотреть «Улицу Сезам», изначально более любопытны, чем остальные.

Эта возможность придерживаться исходной теории при наличии ответа, противоречащего ожиданиям, говорит нам о том, что исследовательский вопрос был не очень хорош, по крайней мере для данной теории[6]. Ответ на действительно хороший исследовательский вопрос невозможно отвергнуть лишь потому, что он неудобен для теории.

[5] Конечно, в реальном мире очень немногие исследовательские вопросы получают однозначный ответ. Даже после этого эксперимента остаются вопросы, будут ли результаты такими же в другое десятилетие или в другой стране, или если мы выберем разную продолжительность просмотра телевизора. Но даже если бы мы не смогли ответить на этот вопрос окончательно, мы все равно получили бы свидетельства, которые недвусмысленно информируют нас по существу поставленного вопроса.

[6] Это означает, что если результат исследования «Улицы Сезам» совпадет с теорией, наша уверенность в ней не возрастет.

Так что «Притупляет ли просмотр телевизора в детстве ваше любопытство во взрослом возрасте?» – это хороший исследовательский вопрос, на который можно ответить с помощью правильных данных и который послужит источником информации для расширения нашего понимания мира. Конечно, поиск ответа на этот вопрос станет для нас почти непреодолимым препятствием[7]. Но по крайней мере мы знаем, что вопрос сам по себе хорош, даже если ответ недосягаем.

2.2. Зачем начинать с вопроса?

Зачем нам лишняя суета? У нас под рукой масса данных. На самом деле мы утопаем в данных. Данные есть везде. Так почему бы не пропустить сложную часть вывода исследовательского вопроса из теории и вместо этого просто посмотреть, какие закономерности содержатся в данных?

Да, это возможно. На самом деле многие люди так и делают. Это называется *интеллектуальный анализ данных* (data mining), и есть специалисты, которые вполне успешно занимаются именно этим. Они обращаются к данным, ищут закономерности и предоставляют отчеты. Многое из этой деятельности относится к области науки о данных[8], но интеллектуальный анализ данных можно проводить в любое время, когда у вас есть хоть какие-то данные. Просто берете данные, смотрите, что там внутри, и раскапываете связи в обратном направлении.

Звучит неплохо, правда? Что ж, интеллектуальный анализ данных – это прекрасно, только в некоторых вещах он оказывается очень хорош, а в других – очень плох.

Интеллектуальный анализ данных хорош в *поиске закономерностей* и *прогнозировании в условиях стабильности*[9]. Он плохо подходит для *расширения нашего понимания мира*, или, другими словами, *для улучшения теории*. Он также склонен находить *ложные закономерности*, если вы не будете достаточно внимательны и осторожны.

Находить закономерности и делать прогнозы – это очень ценная возможность. И если вам нужно только это, то имеет смысл ограничиться анализом данных. В самом деле, мы ведь не станем обдумывать все возможные зависимости в данных, строить теории, а потом проверять их исследованиями.

[7] Попробуйте провести этот масштабный эксперимент на практике. Желаю удачи.

[8] Это не единственное, чем занимаются специалисты по данным. Еще они зарабатывают много денег. Ладно, ладно, специалисты по анализу данных иногда тоже имеют дело с теориями. Большинство инструментов, предназначенных для интеллектуального анализа данных, при правильном применении можно использовать для работы с теориями. Только это будет уже не совсем анализ данных. Скорее, это ближе к науке о данных.

[9] Что я подразумеваю под «в условиях стабильности»? Я имею в виду, что процесс генерации данных не должен меняться. Если я брошу шестигранную игральную кость тысячу раз, интеллектуальный анализ данных поможет предсказать, что вероятность выпадения 1 равна 1/6. Но если я внезапно переключусь на двадцатигранную кость, прогноз на основе интеллектуального анализа данных резко потеряет достоверность. Он по-прежнему будет прогнозировать вероятность 1/6, пока не получит гораздо больше данных. С другой стороны, теория вероятностей правильно предскажет немедленный переход к шансу 1/20.

Иногда нам достаточно знать, *что* происходит, не задаваясь вопросом, *почему* это происходит. Кроме того, иногда обнаружение закономерностей в данных может дать нам идеи для исследовательских вопросов, которые мы затем дополнительно изучим в других источниках данных.

Если меня не волнует, почему фондовый рынок идет вверх или вниз, и я просто хочу предсказать, в какую сторону изменится цена, чтобы принять решение о покупке или продаже акций, то интеллектуальный анализ данных станет для меня вполне подходящим инструментом.

Но что происходит за пределами подобных сценариев?

Почему интеллектуальный анализ данных плохо работает с теорией? Есть несколько основных причин.

Одна из них заключается в том, что интеллектуальный анализ данных по определению фокусируется на том, *что* находится в данных, а не на том, *почему* это оказалось в данных. Другими словами, он великолепно выявляет корреляции – закономерности, связанные с тем, как наблюдаемые переменные изменялись вместе в прошлом, – но эти корреляции могут иметь мало общего с причинно-следственными связями или пониманием того, почему эти переменные изменялись вместе.

Приведу пример, который несколько раз будет встречаться в этой книге: специалист по интеллектуальному анализу данных, попытавшись понять закономерности продаж мороженого, легко может заметить, что доля людей, носящих шорты, является фантастически точным предиктором динамики продажи мороженого. Но дело ведь не в том, что люди в шортах более охотно покупают мороженое. Они покупают мороженое и носят шорты, потому что жарко. Но для аналитика наблюдаемая корреляция между ношением шорт и поеданием мороженого весьма привлекательна! В конце концов, вычисление доли людей в шортах может стать отличным способом предсказать продажи мороженого, даже если он не отвечает на вопрос «почему».

Однако если нас интересует не предсказание продаж мороженого, а объяснение того, *почему* продажи мороженого коррелируют с ношением шорт, возникает большое искушение попытаться придумать историю, поясняющую, почему люди в шортах чаще

едят мороженое. В случае с мороженым и шортами мы можем сказать, что это смешно, но ситуация становится намного сложнее, когда мы не знаем, что смешно, а что является важным открытием.

Например, нам бы хотелось знать, что заставляет детей действовать агрессивно. Это очень важная проблема! Специалист по интеллектуальному анализу данных может проанализировать все действия детей и проверить, связаны ли некоторые из них с более высоким уровнем агрессии. Возможно, дети, которые много играют в видеоигры, более склонны к агрессии. Итак… несут ли ответственность за это видеоигры? Может, да, а может, нет[10]. Интеллектуальный анализ данных хорошо подходит для обнаружения взаимосвязей и плохо – для объяснения того, почему эти взаимосвязи существуют. Остается лишь надеяться, что политики не запретят все видеоигры, прежде чем исследователь успеет тщательно объяснить разницу между корреляцией и причинно-следственной связью.

[10] Скорее, нет, чем да, если верить последним исследованиям в этой области.

Другая причина заключается в том, что интеллектуальный анализ данных на самом деле не занимается абстракцией, поскольку он максимально сосредоточен на данных. Например, взгляните на стул. Откуда вы знаете, что это стул? У него, вероятно, есть ножки, может быть, спинка, плоская зона для сидения, и он всем своим видом явно предназначен для сидения. Это наша «теория стула» – мы предполагаем, что существуют объекты, называемые стульями, которые обладают определенными свойствами стула, и объединенные возможностью сидеть на них на некотором расстоянии от земли. Стул, на который вы сейчас смотрите, – один из примеров теории стульев[11].

[11] Если бы Платон еще не умер, я уверен, что этот абзац убил бы его.

Но что на самом деле содержится *в данных*? Там нет слова «стул». Есть только плоская площадка и несколько прямых частей сверху и снизу этой площадки. Интеллектуальный анализ данных может пригодиться, если нам важно обнаружить наличие вертикальных деталей выше и ниже горизонтальной площадки, но бесполезен для разработки «теории стула», потому что не может перейти от конкретных данных к абстрактному назначению стула – возможности сидеть на нем. Специалист по интеллектуальному анализу данных никогда бы не догадался, что стул на четырех ножках имеет какое-то отношение, скажем, к креслу-мешку, у которого

вообще нет выраженных ножек и спинки. Ложные взаимосвязи – еще одна причина, почему интеллектуальный анализ данных может быть опасным. Возьмем пример видеоигр и агрессии. Ладно, возможно, дети агрессивны не только из-за видеоигр, но мы все же нашли связь. Наверняка там есть что-то эдакое.

Нет, не обязательно. Интеллектуальный анализ данных фактически означает кропотливый перебор данных в поисках примечательных закономерностей. И здесь есть на что посмотреть! Если вы проверите, скажем, сотню переменных и посмотрите, связаны ли они с агрессией, хоть одна из них обязательно будет выглядеть связанной просто в силу случайного стечения обстоятельств. Эта случайная связь вряд ли возникнет снова, если вы проанализируете другую выборку. Она характерна только для единственной выборки, потому и называется *ложной закономерностью*, или *ложным срабатыванием*.

Это одна из основных опасностей, которая грозит вам, если вы проводите исследование без тщательно продуманного исследовательского вопроса. И даже при наличии такого вопроса нельзя расслабляться. Какая-то случайно возникшая закономерность обязательно всплывет, если вы переберете достаточно много переменных и проверите достаточно много выборок. Нужно быть очень хорошим и честным исследователем, чтобы не притворяться, что именно это вы искали с самого начала, и не придумывать какую-то причину, почему сотое соотношение, которое вы проверили, имеет особый смысл и подтверждает вашу теорию.

Существуют способы избежать ложных закономерностей при интеллектуальном анализе данных. Это одна из важнейших тем в области науки о данных, и ученые придумали много разных методов[12]. Но если вы просто наобум просматриваете набор данных, то, скорее всего, в конечном итоге получите массу ложных закономерностей вперемежку с реальными. У вас не будет возможности отличить одно от другого.

Тем не менее с анализом данных не все так плохо. Интересные теории для проверки не появляются из ниоткуда. Множество теорий возникают в результате изучения данных, обнаружения закономерности и размышлений о том, почему она проявляется и реальна ли она вообще.

[12] Некоторые примеры – «перекрестная проверка» и «наборы для обучения и тестирования». Если это вас интересует, узнайте больше о науке о данных. В этой книге вы познакомитесь с наукой о данных, но не так уж и много.

Например, препарат «Виагра» изначально был разработан как лекарство от повышенного кровяного давления. Исследователи, проверявшие, насколько эффективно он снижает кровяное давление, заметили и другие его эффекты.

Сперва они провели интеллектуальный анализ данных – вместо того чтобы прийти к данным с готовой теорией, они заметили в них интересную закономерность.

Конечно, самое ответственное на этом этапе – не принимать закономерность за безупречный факт. Вот где настоящая проблема интеллектуального анализа данных. Вместо этого исследователи взяли замеченную ими интересную закономерность и посмотрели, сохраняется ли она в других выборках (т. е. обладает ли *повторяемостью*), тщательно убедились, что замеченная ими закономерность реальна, и лишь потом приступили к выяснению того, как действует препарат.

Интеллектуальный анализ данных хорош, когда уместен. Но это плохой последний шаг, если вы пытаетесь объяснить мир. Он по-прежнему может служить источником идей. И кто знает – возможно, вы заработаете на нем миллиарды долларов, как это сделала «Виагра».

2.3. Откуда берутся исследовательские вопросы?

У исследовательских вопросов есть разные источники. В основном это любопытство. Мы хотим знать, как устроен мир, и это, естественно, приводит к вопросам!

Процесс расширения познаний о мире состоит из двух этапов: размышлении о теории и постановки исследовательского вопроса. Любой из них может быть первым.

Возможно, все начинается с теории: «Я думаю, так устроен мир» или «Интересно, так ли устроен мир, как я думаю» – это ваша теория. На самом деле теория может быть какой угодно: от «Я думаю, что люди принимают решения, потому что они стремятся получить вознаграждение» до «Я думаю, что растения не нуждаются в пищеварительной системе, потому что они поглощают энергию солнца» или «Я думаю, что люди перестали покупать компакт-диски, потому что скачивают музыку в интернете».

Когда есть теория, за ней следует гипотеза: «Если мир устроен именно так, что я ожидаю увидеть?» Вышеупомянутые теории могут привести нас к исследовательским вопросам: «Будут ли ученики усерднее учиться в школе, если платить им за хорошие оценки?», или «Погибнут ли растения, если держать их в темной комнате?», или «Популярны ли компакт-диски в регионах с плохим доступом к интернету?» Эти исследовательские вопросы подсказывают гипотезу, которую нужно проверить, а результат проверки так или иначе характеризует теорию.

Можно начать с вопроса. Допустим, вы спросите: «Будут ли ученики усерднее учиться в школе, если платить им за хорошие оценки?» Тогда возникает вопрос, почему вам вообще пришла в голову такая идея. Вероятно, вы предполагаете, что поведение человека зависит от стимулов? Если вы сами плохо понимаете, почему задали этот вопрос, возможно, это не лучший исследовательский вопрос. Или, по крайней мере, вам будет трудно кого-либо заинтересовать ответом, когда вы его получите.

По правде говоря, иногда исследовательские вопросы возникают просто потому, что есть *возможность* их задать.

У вас есть хороший набор данных? Подумайте, какие данные вам доступны и приходят ли на ум какие-либо связанные с ними исследовательские вопросы или теории[13].

Или, возможно, вы узнали о чем-то необычном либо интересном. Скажем, вы узнали, что некоторые школьные округа решили попробовать платить школьникам за хорошие оценки. Услышав о чем-то подобном, вы можете спросить: «На какие исследовательские вопросы я смогу ответить?», следом у вас возникнет исследовательский вопрос, а из него – теория!

[13] Попытайтесь сделать это после того, как поймете, что содержится в данных, но перед их фактическим анализом, если только вашей целью не является интеллектуальный анализ данных.

2.4. Как узнать, хороший ли у вас вопрос?

Допустим, вы играете по правилам. У вас есть исследовательский вопрос. Вы знаете, что на этот вопрос можно ответить с помощью данных, и вы почти уверены, что ответ расширит ваши знания о мире.

Но действительно ли с вашим исследованием все в порядке? Прежде чем перейти к делу, нужно кое-что обдумать и проверить.

1. **Обдумайте потенциальные результаты.** Хороший способ перепроверить связь между вашим исследовательским вопросом и теорией – это *рассмотреть потенциальные ответы, которые вы можете получить*. Затем подумайте о том, какой смысл вы придадите этим ответам или какой вывод вы сделаете. Допустим, вы обнаружили, что учащиеся, как правило, усерднее занимаются в школе, когда им платят за хорошие оценки. Что это скажет нам о том, как учащиеся реагируют на стимулы? Или, скажем, вы обнаружили, что школьники *не* занимаются усерднее, когда им платят. Что это скажет нам о том, как учащиеся реагируют на стимулы? Если вы не можете сказать что-то интересное о своих потенциальных результатах, это, вероятно, означает, что ваш исследовательский вопрос и ваша теория не так тесно связаны, как вы думаете! Допустим, мы обнаружили, что дети, которые часто играют в видеоигры, более агрессивны. Можем ли мы взять этот результат и заявить, что видеоигры являются причиной агрессии? Не совсем, по причинам, которые мы обсуждали ранее. Так что, возможно, этот исследовательский вопрос на самом деле не очень хорошо связан с вашей теорией.
2. **Обдумайте осуществимость.** Исследовательский вопрос должен представлять собой вопрос, на который можно ответить, используя правильные данные. Но доступны ли эти данные? Если ответ на ваш исследовательский вопрос теоретически возможен, но требует неоднократного наблюдения за миллионами людей на протяжении десятилетий или попыток измерить что-то, что действительно трудно измерить точно, например заставить людей вспомнить, что они ели на обед три года назад, или получить доступ к персональным данным тысяч несогласных людей, то этот исследовательский вопрос может оказаться неосуществимым. Хотя иногда эти проблемы удается обойти при помощи изощренных схем исследования, возможно, вам стоит вернуться к чертежной доске.
3. **Учитывайте масштаб.** Какие ресурсы и время вы можете посвятить ответу на исследовательский вопрос? Потратив целую жизнь и значительные ресурсы, вы, возможно, сможете ответить на такие масштабные вопросы, как «Что заставляет

одни страны становиться богатыми, а другие – бедными?». Если же вы намерены ответить на подобный вопрос в курсовой работе, то успеете получить максимум пару весьма поверхностных выводов. Вы достигнете намного более осязаемых результатов за ограниченное время, отвечая на менее сложные вопросы.

4. **Продумайте схему исследования.** Исследовательский вопрос может быть интересным сам по себе, но без ответа он мало чего стоит. Важной частью оценки того, есть ли у вас работоспособный исследовательский вопрос, является выяснение, существует ли разумная схема исследования, которую вы можете использовать для получения ответа. Выяснению того, что представляет собой разумная схема исследования, посвящена остальная часть этой книги.
5. **Будьте проще!** Поиск ответа на исследовательский вопрос бывает сложным. Не усложняйте себе задачу, откусывая больше, чем можете проглотить. *Распространенной ошибкой является объединение нескольких исследовательских вопросов в один*. «Каковы детерминанты социальной мобильности?» То есть вопрос о том, как человек может переходить из одного социального класса в другой на протяжении всей своей жизни. Существует множество факторов, определяющих социальную мобильность. Вряд ли вы сможете дать на этот вопрос исчерпывающий ответ. Вместо этого попробуйте спросить: «Является ли место рождения детерминантом социальной мобильности?» Другой пример: «Как итальянский ренессанс повлиял на живопись?» Да миллионом способов! Вы заблудитесь в этой чаще и плохо справитесь со множеством мелких деталей. Вместо этого лучше сформулировать вопрос так: «В чем заключается сходство стран, которые раньше других начали использовать перспективу в живописи?»

Итак, позаботьтесь об осуществимости, масштабе и схеме. Будьте проще и подумайте, расскажут ли вам ожидаемые результаты что-нибудь интересное о мире. Ведь узнать что-то интересное и новое об окружающем мире – наша цель!