

Изменит ли вашу жизнь поступление в Гарвардский университет

Блестящие исследователи в области социальных наук блестящие вовсе не потому, что умеют выполнять в уме сложные вычисления и выигрывают в телевикторине Jeopardy* больше денег, чем обычные исследователи (хотя, возможно, они преуспели и в том и в другом). Блестящие исследователи — это те, кто существенно меняет наши знания и представления о мире и находит творческие способы проведения управляемых экспериментов. Чтобы измерить чье-либо влияние, нам требуется нечто такое, относительно чего мы будем выполнять измерение. Как сказало бы на вашей жизни поступление в Гарвардский университет? Чтобы ответить на этот вопрос, вам нужно знать, что произойдет после того, как вы поступите в Гарвардский университет, и что произойдет *после того, как вы в него не поступите*. Очевидно, вы не можете располагать данными для *обоих* случаев. Тем не менее умные исследователи находят возможность сравнить то или иное «воздействие» (например поступление в Гарвардский университет) с его противоположным сценарием.

Чтобы проиллюстрировать это положение, давайте поразмышляем над казалась бы простым вопросом: приведет ли к снижению преступности увеличение количества полицейских на улицах? Это социально значимый вопрос, поскольку преступность обходится обществу слишком дорого. Если рост числа полицейских на улицах позволит ее снизить (либо потому, что окажется сдерживающим фактором для преступников, либо за счет поимки большего количества «плохих парней»), то инвестиции в наращивание численности полицейских могут обернуться большой выгодой для общества. С другой стороны, рост числа полицейских — весьма дорогостоящее удовольствие; и если эта мера не даст нужного результата или он будет совсем незначительным, то общество может пожалеть о том, что не нашло

* Русский аналог этой телевикторины называется «Своя игра». *Прим. перев.*

более эффективного применения своим ресурсам (например, внедряя современные технологии борьбы с преступностью, такие как видеореамеры наблюдения).

Проблема в том, что ответить на этот вроде бы простой вопрос о последствиях наращивания численности полицейских на улицах не так уж просто. На основании прочитанного к данному моменту материала вы должны это признать, по крайней мере мы не можем ответить на него исходя из информации о тех населенных пунктах, где число полицейских на душу населения существенно превышает средний показатель. Цюрих — не Лос-Анджелес. Даже сравнение крупных американских городов окажется совершенно некорректным: Лос-Анджелес, Нью-Йорк, Хьюстон, Майами, Детройт и Чикаго — слишком разные города с разным демографическим составом населения и разными проблемами, порождающими преступность.

Нашим обычным подходом было бы попытаться определить уравнение регрессии, в котором учитывались бы все эти различия. Но увы, даже множественный регрессионный анализ здесь не поможет. Если мы попробуем объяснить уровень преступности (нашу зависимую переменную) путем использования числа полицейских на душу населения в качестве объясняющей переменной (наряду с другими объясняющими переменными), то возникнет серьезная проблема с обратной причинно-следственной зависимостью. У нас есть надежное теоретическое основание полагать, что увеличение количества полицейских на улицах приведет к снижению преступности, но возможно и обратное: уровень преступности «обуславливает» рост численности полицейских, то есть в городах с повышенной криминогенной обстановкой будет больше служителей порядка. Мы легко можем обнаружить положительную — но вводящую в заблуждение — взаимосвязь между уровнем преступности и количеством полицейских: в районах, где больше всего полицейских, как правило, самый высокий уровень преступности. Аналогично, там, где больше всего медицинских работников, обычно высокий уровень заболеваемости. Разумеется, люди там болеют вовсе не потому, что там слишком много врачей; просто врачи сконцентрированы в местах, где в них существует особая потребность (с другой стороны, больные люди направляются в места, где они могут получить надлежащую медицинскую помощь). Подозреваю, что во Флориде непропорционально большое число онкологов и кардиологов; но даже если выслать половину из них за пределы Флориды, проживающие в этом штате пенсионеры здоровее не станут.

Итак, добро пожаловать в программы статистического оценивания, представляющие собой процесс, посредством которого мы пытаемся измерить результат того или иного воздействия, коим может быть все что угодно, от нового лекарства от рака до программы обеспечения занятости тех, кто бросил школу. Воздействие, о котором я веду речь, обычно называется «активированием»*, хотя в статистическом контексте это слово используется в более широком значении, чем в повседневной жизни. Активирование может быть воздействием в буквальном смысле (то есть медицинским вмешательством того или иного рода), или чем-то наподобие поступления в колледж, или обучением какой-либо профессии после выхода из тюрьмы. Дело в том, что мы пытаемся изолировать влияние этого единственного фактора; в идеале нам хотелось бы знать, как чувствует себя группа, получающая такое «активирование», по сравнению с абсолютно идентичной группой, которая его не получает.

Программы статистического оценивания предлагают совокупность инструментов, позволяющих обособить влияние активирования, когда невозможно установить причину и следствие. Ниже описано, как Джонатан Клик и Александер Табаррок, исследователи из Пенсильванского университета и Университета Джорджа Мейсона, изучали влияние наращивания численности полицейских на улицах на уровень преступности. Стратегия их исследования предусматривала использование системы оповещения о терроризме (terrorism alert system). Если конкретнее, то полицейское управление Вашингтона в дни «повышенной опасности терроризма» выводит на улицы определенных районов города дополнительные наряды полиции, поскольку столица США является естественной мишенью терроризма. Мы можем предположить, что между уличной преступностью и угрозой терроризма нет никакой зависимости, поэтому такое увеличение количества полицейских на улицах Вашингтона не связано с уровнем обычной преступности, то есть обусловлено «внешними» причинами. Самым ценным стало то, что исследователи на основе естественного эксперимента смогли ответить на вопрос: что происходит с обычной преступностью в дни «повышенной опасности терроризма»?

* В оригинале приведено слово *treatment*, которое имеет множество значений. Эти значения: обработка, решение, лечение, трактовка, активизация и др. Мы выбрали слово «активирование» как наиболее подходящее по смыслу для использования в данном тексте. *Прим. ред.*

Ответ оказался таким. Количество преступлений, совершенных в дни «оранжевой» угрозы (высокая опасность и использование дополнительных нарядов полиции), было примерно на 7% ниже, чем в дни, когда уровень террористической угрозы был «желтым» (повышенная опасность, но никаких дополнительных мер по обеспечению правопорядка не принимается). Авторы также пришли к выводу, что самое резкое снижение уровня преступности наблюдалось в полицейском округе, который пользуется самым пристальным вниманием со стороны полиции в дни высокой опасности терроризма (этот округ включает Белый дом, Капитолий и Эспланаду*). Важный вывод заключается в том, что мы можем ответить на сложные социально значимые вопросы, если подходить к этому делу с умом. Ниже перечислены некоторые из наиболее типичных подходов к обособлению влияния активирования.

Статистический управляемый эксперимент. Самый простой способ создать подопытную (по какому-либо методу активирования) и контрольную группы — это... создать подопытную и контрольную группы. Но у этого подхода есть две крупные проблемы. Во-первых, существует много видов экспериментов, которые мы не можем проводить на людях. И это ограничение (я надеюсь) в обозримом будущем никуда не денется. Таким образом, мы можем проводить управляемые эксперименты на людях лишь тогда, когда у нас есть основания полагать, что соответствующее активирование принесет потенциально положительный результат. Зачастую у нас такой уверенности *нет*, именно поэтому нам необходимы стратегии, о которых будет рассказано в этой главе.

Во-вторых, люди отличаются между собой гораздо больше, чем лабораторные крысы. На эффект воздействия, который мы проверяем, вполне могут накладываться другие различия в испытуемой и контрольной группе: там обязательно окажутся высокие люди и «коротышки», больные и здоровые, мужчины и женщины, преступники и законопослушные граждане, алкоголики и трезвенники, банкиры и малообеспеченные люди и т. п. Как мы можем гарантировать, что различия по этим и другим характеристикам не скажутся на результатах тестирования? У меня есть для вас хорошая новость: это один из тех редких случаев жизни, когда наилучший подход предполагает минимум усилий! Оптимальный способ создания любой подопытной группы, которая подвергается активированию, и контрольной группы заключается

* Эспланада — отрезок музейно-парковой зоны в центре Вашингтона между Капитолием и памятником Джорджу Вашингтону. *Прим. перев.*

в случайном (рандомизированном) распределении по ним участников исследования. Прелесть рандомизации в том, что она приведет к более или менее равномерному распределению между этими двумя группами переменных, не связанных с активированием, — как очевидных характеристик, таких как пол, расовая принадлежность, возраст и образование, так и ненаблюдаемых характеристик, которые могли бы исказить интересующие нас результаты.

Это можно представлять себе так. Если в нашей большой выборке присутствуют 1000 женщин, то после того как мы произвольно разделим ее на две группы, в каждой из групп, скорее всего, окажется по 500 женщин. Разумеется, утверждать наверняка это нельзя, но и здесь теория вероятностей придет нам на помощь. Вероятность того, что в какой-то из двух групп будет непропорционально большое число женщин (или непропорционально большое число людей с какой-либо другой характеристикой), очень мала. Если, например, в выборке из 1000 человек половину составляют женщины, то вероятность того, что в какой-то из двух групп окажется менее 5 женщин, будет меньше 1%. Понятно, что чем больше выборка, тем эффективнее (с точки зрения создания похожих, в широком смысле, групп) рандомизация.

Медицинские испытания, как правило, стремятся проводить в духе рандомизированных управляемых экспериментов. В идеале они представляют собой двойное контрольное (слепое) испытание; это означает, что ни пациент, ни врач не знают, кто получает лечение, а кто — плацебо. Разумеется, двойное контрольное испытание невозможно, когда речь идет о хирургических процедурах (надеюсь, кардиохирургу будет заранее известно, к какому из пациентов применяется операция шунтирования). Хотя даже в этом случае иногда удается держать пациентов в неведении относительно того, к какой именно группе (подопытной или контрольной) они относятся. Одно из моих любимых исследований предполагало оценивание определенного вида коленной хирургии, призванной облегчить боль. Участникам «лечебной» группы делали соответствующую хирургическую операцию на колене. А членам контрольной группы хирург, имитируя выполнение операции, делал три небольших надреза в области колена пациента*. Оказалось, что реальная хирургическая операция была ненамного эффективнее, чем мнимая¹.

Рандомизированные статистические исследования могут применяться для тестирования некоторых интересных явлений. Например, улучшают ли

* Участники этого эксперимента знали, что участвуют в клиническом испытании и что им могут сделать фиктивную хирургическую операцию.

постхирургические исходы молитвы людей, незнакомых с прооперированными? Разумные люди по-разному относятся к религии, но авторы исследования, результаты которого были опубликованы в *American Heart Journal*, провели управляемый эксперимент, который должен был ответить на вопрос, будет ли меньше послеоперационных осложнений у больных, перенесших операцию шунтирования на сердце, если большая группа совершенно незнакомых им людей будет молиться за их скорейшее выздоровление². В исследовании участвовали 1800 пациентов и членов трех религиозных общин со всей страны. Пациентов, перенесших операцию коронарного шунтирования, разделили на три группы: за одну группу никто не молился; за вторую молились, и им сообщили об этом; за третью молились, но ее участникам сказали, что за них могут молиться или не молиться (таким образом обеспечивалась управляемость плацебо-эффекта молитвы). Тем временем членов религиозных конгрегаций попросили молиться за определенных пациентов по их имени и первой букве фамилии (например, Чарли У.). Какие именно молитвы они будут произносить, прихожане решали самостоятельно; единственное условие, чтобы молитва содержала фразу «за успешную хирургическую операцию с быстрым выздоровлением и без осложнений».

Ну и?... Станут ли молитвы экономически эффективным решением проблем здравоохранения в Америке? Наверное, нет. Исследователи не обнаружили какой-либо разницы в частоте осложнений в течение тридцатидневного послеоперационного периода между теми, за кого молились, и теми, за кого не молились. Критики этого эксперимента указывали на переменную, которую не учли ученые: молитвы, исходившие от других источников. Как резюмировала газета *The New York Times*: «Эксперты сказали, что это исследование не смогло преодолеть самое, пожалуй, крупное препятствие к изучению эффективности молитв: неизвестный объем молитв, получаемых каждым участником эксперимента от своих друзей, родственников, членов семьи и конгрегаций по всему миру, которые ежедневно молятся за страждущих».

Проведение экспериментов на людях может закончиться арестом или даже международным трибуналом. О такой «перспективе» никогда не следует забывать. Однако в области социальных наук всегда найдется место для статистических управляемых экспериментов с участием людей. Одним из самых знаменитых стало исследование *Tennessee's Project STAR*, в ходе которого изучалось влияние уменьшения количества учащихся в группах на степень усвоения ими знаний. В наши дни буквально все страны борются

за повышение качества своих систем образования. Если уменьшение количества учащихся в группах способствует более эффективному обучению, то при прочих равных условиях (*ceteris paribus*) общество должно направлять дополнительные средства на подготовку большего числа преподавателей, которые при этом понадобятся. В то же время дополнительные преподаватели — это дополнительные расходы (и весьма немалые); если учащиеся в небольших группах демонстрируют лучшие результаты по причинам, не *зависящим от размера этих групп*, то государственные деньги окажутся попросту выброшенными на ветер.

Связь между величиной учебной группы и эффективностью усвоения учащимися материала, как ни странно, изучить не так просто. Учебные заведения, формирующие компактные группы студентов, как правило, имеют больше ресурсов; это означает, что учащиеся и преподаватели в них отличаются от учащихся и преподавателей в учебных заведениях с большими размерами учебных групп. К тому же меньшие учебные группы обычно формируются в силу определенных причин. Например, директор может создать такую группу для отстающих учеников (в этом случае может наблюдаться ложная отрицательная зависимость между небольшим количеством учеников в классе и их успеваемостью). Еще один вариант: опытные преподаватели могут отдать предпочтение небольшим группам; в этом случае преимущество последних будет следствием выбора преподавателей.

Начиная с 1985 года в ходе исследования Tennessee's Project STAR был проведен управляемый эксперимент по выявлению последствий использования меньших учебных групп³. (В то время губернатором штата Теннесси был Ламар Александер, будущий министр образования в правительстве Джорджа Буша.) Ученики из семидесяти девяти разных подготовительных школ* были случайным образом распределены либо в небольшой класс (13–17 учеников), либо в обычный класс (22–25 учеников), либо в обычный класс с обычным преподавателем и помощником преподавателя. Учителей также распределили случайным образом по разным классам. До третьего класса включительно ученики оставались в тех классах, в которые они изначально попали. В процессе рандомизации влияние тех или иных жизненных реалий в какой-то мере нивелировалось. Одни ученики входили в эту систему в середине эксперимента, другие «выходили из игры». Кого-

* В Соединенных Штатах в подготовительных школах учатся дети пяти-шести лет.
Прим. перев.

то из детей переводили из одного класса в другой по дисциплинарным соображениям; кому-то из родителей удалось перевести своих чад в меньшие по численности классы. И так далее.

Тем не менее исследование Tennessee's Project STAR остается единственным рандомизированным тестом статистически и социально значимых результатов сокращения численности учебных групп. В целом при сдаче стандартизированных экзаменов успеваемость учеников в меньших по размеру классах оказалась на 0,15 среднеквадратических отклонений лучше, чем в обычных классах; а успеваемость чернокожих учащихся меньших по численности классов — в два раза выше, чем в обычных классах. А теперь плохая новость. Эксперимент Tennessee's Project STAR обошелся примерно в 12 миллионов долларов. Стоимость исследования влияния молитв на послеоперационные осложнения вылилась в 2,4 миллиона долларов. Самые точные и полезные исследования характеризуются тем же, что и все «точное и полезное», — очень высокими затратами.

Натурный эксперимент. Далеко не каждый располагает несколькими лишними миллионами долларов, которые он готов потратить на проведение крупномасштабного статистического исследования. А поскольку жизнь иногда совершенно случайно создает подопытную и контрольную группы, ученые стараются воспользоваться столь привлекательной ситуацией. Эта более экономичная альтернатива называется натурный эксперимент, яркий пример которого — наш рассказ о полицейских Вашингтона, приведенный в начале главы. Рассмотрим интересную, но сложную взаимосвязь между образованием и долголетием. Более образованные люди обычно живут дольше, даже если зафиксировать такие факторы, как уровень дохода и доступ к медицинским услугам. Как отмечала The New York Times: «Единственным социальным фактором, который, по признанию ученых, безусловно связан с продолжительностью жизни человека во всех странах, где изучалось его влияние, является образование. Уровень образования более важен, чем принадлежность к той или иной расе; он нивелирует любое влияние дохода»⁴. Правда, до сих пор речь шла лишь о корреляции. Но можно ли утверждать, что более высокий уровень образования, *ceteris paribus*, способствует улучшению здоровья? Если образование как таковое представить как «активирование», то можно ли утверждать, что повышение его уровня приведет к увеличению продолжительности вашей жизни?

Поначалу кажется, что исследовать этот вопрос практически невозможно, поскольку люди, стремящиеся повысить свой уровень образования,

отличаются от тех, кто этого не хочет. Разница между выпускниками средней школы и выпускниками колледжей не только в том, что вторые учились на четыре года дольше. У людей, испытывающих тягу к знаниям, вполне могут оказаться какие-то ненаблюдаемые (то есть не поддающиеся наблюдению) общие черты, которые могут объяснять их более высокую продолжительность жизни. Если это действительно так, то предлагать повышать образовательный уровень тем, кто к этому не склонен, бессмысленно: это не улучшит состояния их здоровья. То есть крепкое здоровье не зависит от уровня образования и может быть присуще той категории людей, для которых стремление учиться *естественно*.

Мы не можем провести рандомизированный эксперимент, чтобы решить эту головоломку, так как это предусматривало бы уход некоторых его участников из школы раньше, чем им хотелось бы. (Попробуйте кому-то объяснить, что он не сможет — никогда! — поступить в колледж, поскольку находится в контрольной группе!) Единственно возможным тестом причинно-следственной связи между уровнем образования и продолжительностью жизни мог бы стать эксперимент, в рамках которого достаточно большой сегмент совокупности оставался бы в школе дольше, чем это было бы в противном случае. Такой вариант по крайней мере приемлем с этической точки зрения, потому что мы рассчитываем на положительный эффект исследования. Тем не менее мы не можем заставлять детей оставаться в школе — это не американский стиль жизни.

Правда, элемент принуждения все же присутствует в реальной жизни. В каждом штате действуют законы, предусматривающие обязательный минимальный уровень образования, причем на разных исторических этапах *эти законы менялись*. Такого рода экзогенные (внешние) изменения обязательного минимального уровня образования — настоящий подарок для ученых. Адриана Ллерас-Муни, выпускница Колумбийского университета, усмотрела значительный исследовательский потенциал в том, что разные штаты вносили изменения в законы в разное время. Она выполнила исследование в историческом аспекте и изучила связь между периодами, когда штаты вносили изменения в законы об обязательном минимальном уровне образования, и последующими изменениями продолжительности жизни в штатах (для этого ей пришлось перелопатить горы данных по переписям населения). Тем не менее перед Ллерас-Муни оставалась серьезная методологическая проблема: увеличение продолжительности жизни жителей

какого-то штата после повышения обязательного минимального уровня образования нельзя объяснять именно повышением последнего. Продолжительность жизни людей со временем увеличивается (это связано с достижениями науки и медицины, улучшением качества питания и т. п.). В 1900 году люди жили дольше, чем в 1850-м, независимо от того, какие законы принимались теми или иными штатами.

Однако в распоряжении Ллерас-Муни был естественный управляющий фактор: штаты, в которых законы об обязательном минимальном уровне образования не менялись. Ее работа соответствует гигантскому лабораторному эксперименту, в котором жители Иллинойса были обязаны учиться в школе не менее семи лет, тогда как их соседи из штата Индиана имели право бросить учебу через шесть лет. Разница в том, что этот управляемый эксперимент стал возможен в силу исторической случайности — отсюда термин «натурный эксперимент».

Что же оказалось в результате? Продолжительность жизни взрослых, достигших тридцатипятилетнего возраста, увеличилась примерно на полтора года только за счет того, что они проучились в школе один дополнительный год⁵. Результаты, полученные Ллерас-Муни, удалось воспроизвести в других странах, где изменения законов об обязательном школьном образовании создавали аналогичные натурные эксперименты. Однако у нас остаются основания для некоторого скептицизма в отношении итогов этого эксперимента. Нам по-прежнему непонятен механизм, посредством которого продолжительность учебы повышает продолжительность жизни.

Неэквивалентный контроль. Иногда оптимальным вариантом для изучения эффекта активирования является создание нерандомизированных подопытной и контрольной групп. Наша надежда/ожидание заключается в том, что обе группы будут похожи в широком смысле слова даже несмотря на то, что обстоятельства не позволяют нам такой «статистической роскоши», как рандомизация. Радует уже то, что у нас есть подопытная и контрольная группы. Хотя плохо, что неслучайный характер назначения в эти группы создает как минимум возможность ошибки. Между подопытной и контрольной группами могут существовать ненаблюдаемые различия, связанные с распределением участников. Отсюда и название — «неэквивалентный контроль».

Тем не менее неэквивалентная контрольная группа представляет собой чрезвычайно полезный инструмент. Давайте поразмышляем над вопросом,

поставленным в подзаголовке к этой главе: обеспечивает ли значительное жизненное преимущество учеба в каком-либо из элитных колледжей или университетов? Очевидно, что выпускники Гарвардского, Принстонского и Дартмутского университетов имеют все основания рассчитывать на успех. В среднем они зарабатывают больше и имеют более широкие перспективы, чем студенты, посещающие менее престижные учебные заведения. (Исследование, выполненное PayScale.com в 2008 году, показало, что средняя заработная плата выпускников Дартмутского университета с рабочим стажем от десяти до двадцати лет составляет 134 000 долларов; это самая высокая заработная плата среди выпускников высших учебных заведений, обладающих степенью бакалавра. На втором месте по этому показателю — выпускники Принстонского университета, их средняя заработная плата 131 000 долларов.)⁶ Надеюсь, вы понимаете, что столь впечатляющие показатели средней заработной платы абсолютно ничего не говорят нам о ценности образования, получаемого в Дартмутском или Принстонском университете. Студенты этих учебных заведений, несомненно, талантливы — *именно поэтому им удалось туда поступить*. Но они наверняка преуспели бы в жизни независимо от того, в каком университете или колледже учились.

Нам неизвестно, каков эффект от учебы в таких учебных заведениях, как Гарвардский или Йельский университет. Их выпускники преуспевают в жизни потому, что просто необыкновенно талантливы? Или это университеты и колледжи сделали этих изначально талантливых молодых людей еще более продуктивными? Или, может быть, сыграло свою роль то и другое?

Мы не можем провести рандомизированный эксперимент, чтобы ответить на этот вопрос. Вряд ли среди выпускников средней школы найдется много желающих попасть в какой-либо выбранный наугад колледж; весьма сомнительно и то, что Гарвардский и Дартмутский университеты обрадуются идее произвольного набора студентов. Похоже, у нас нет какого-либо механизма проверки ценности эффекта рассматриваемого нами активирования. Талант под угрозой! Но экономисты Стейси Дейл и Алан Крюгер нашли способ ответить на этот вопрос, воспользовавшись* тем

* Исследователям нравится слово «воспользоваться» (exploit). Оно, в частности, применяется в значении «воспользоваться какой-либо возможностью, связанной с данными». Например, когда исследователи обнаруживают какой-либо натурный эксперимент, который создает подопытную и контрольную группу, они пишут, как собираются «воспользоваться разбросом в соответствующих данных».

фактом, что многие из выпускников средней школы подают документы сразу в несколько колледжей⁷. Некоторых выпускников принимают в то или иное элитное учебное заведение, и они решают поступить именно в него; другие в аналогичной ситуации выбирают менее престижный колледж или университет. Замечательно! Теперь у нас есть подопытная группа (студенты, которые предпочли элитные колледжи и университеты) и неэквивалентная контрольная группа (*достаточно талантливые студенты, тем не менее* решившие поступить в менее престижный колледж или университет)*.

Дейл и Крюгер изучили повторные данные о доходах в обеих группах. Это нельзя назвать идеальным сравнением яблок с яблоками, а уровень доходов, конечно же, не самое главное жизненное достижение человека, однако результаты, полученные Дейлом и Крюгером, должны развеять тревоги переутомленных напряженной учебной работой старшекласников и их родителей. Выпускники элитных колледжей зарабатывают приблизительно столько же, сколько выпускники менее престижных учебных заведений — если, конечно, речь идет о людях примерно одинаковых способностей. Единственное исключение — студенты из малоимущих семей, которые зарабатывали больше, если учились в элитных колледжах или университетах. Подход Дейла и Крюгера кажется мне весьма элегантным способом

* Здесь существует вероятность ошибки. Обе группы студентов достаточно талантливы для того, чтобы быть принятыми в один из элитных колледжей или университетов. Однако одна группа студентов решила поступить в элитное учебное заведение, а другая предпочла менее престижный колледж или университет. Вторая группа студентов может быть менее мотивирована, менее трудолюбива или может отличаться в каких-то других, ненаблюдаемых отношениях. Если бы Дейл и Крюгер обнаружили, что студенты, поступившие в элитные учебные заведения, впоследствии зарабатывали больше, чем студенты, принятые в одно из элитных учебных заведений, но выбравшие менее престижный вуз, мы все же не могли бы быть уверены, что разница в их будущих доходах объясняется учебной работой в элитном учебном заведении, а не особенностями человека, получившего шанс поступить в элитное учебное заведение и воспользовавшегося им. Но в исследовании Дейла и Крюгера эта потенциальная ошибка не играет существенной роли. Дейл и Крюгер обнаружили, что студенты, которые поступили в элитные учебные заведения, впоследствии зарабатывали ненамного больше тех, кто выбрал какой-либо другой вариант продолжения учебы, *несмотря на то обстоятельство, что студенты, отказавшиеся поступить в элитные учебные заведения, могли обладать другими (помимо образования) особенностями, которые мешали им зарабатывать больше*. Как бы то ни было, упомянутая мною ошибка заставляет авторов данного исследования скорее *преувеличивать* денежные выгоды учебы в элитных колледжах и университетах, которые в любом случае оказываются несущественными.

отделить эффект активирования (четырёхлетняя учеба в элитном учебном заведении) от эффекта престижности учебного заведения (в эти учебные заведения принимают самых способных студентов). Подводя итоги исследования в статье для *The New York Times*, Алан Крюгер косвенно ответил на вопрос, поставленный в подзаголовке этой главы: «Вам придется признать тот факт, что ваша собственная мотивация, амбиции и способности будут определять ваш успех в большей степени, чем название колледжа в вашем дипломе»⁸.

Разница в различиях. Один из наилучших способов наблюдать причину и следствие — это сделать что-то, а затем посмотреть, к чему это действие приведет. Именно так маленькие дети (а подчас и взрослые) познают окружающий мир. Например, мои дети довольно быстро поняли, что если бросать куски пищи из одного конца кухни в другой (причина), то собака будет носиться как угорелая по кухне за этими кусками (следствие). Вполне возможно, что та же сила наблюдения помогает человеку в течение жизни. Если сокращение налогов приводит к оздоровлению экономики, значит, оно могло исправить в ней ситуацию.

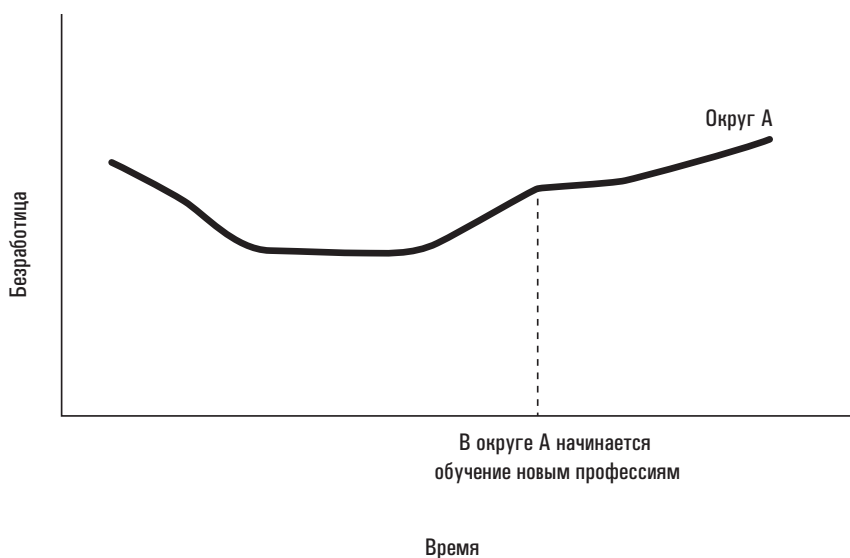
Может быть, может быть... Весьма внушительный подводный камень такого подхода заключается в том, что жизнь, как правило, гораздо сложнее швыряния обглоданных куриных ножек из одного конца кухни в другой. Да, на каком-то этапе мы могли сократить налоги, однако примерно на том же этапе разворачивались другие «активизации». В колледжи могло поступить большее число женщин, интернет и другие технологические новшества привели к повышению производительности американских рабочих, стоимость китайской национальной валюты оказалась заниженной, профессиональный бейсбольный клуб *Chicago Cubs* уволил своего генерального менеджера и т. д. Все, что произойдет после сокращения налогов, нельзя приписывать исключительно сокращению налогов. Проблема любого анализа, который можно отнести к категории «до и после», заключается в том, что если одно событие (событие А) произошло после какого-то другого (события В), то это отнюдь не означает, что событие В послужило причиной события А.

Подход, который мы обозначили как «разница в различиях», может помочь идентифицировать последствия некоторой активизации при выполнении двух условий. Во-первых, мы анализируем данные «до» и «после» для той группы или юрисдикции (подведомственной области), которые

получили соответствующее активирование, например данные по безработице для округа, реализовавшего программу обучения новым профессиям. Во-вторых, сравниваем эти данные с показателями безработицы за тот же период времени для какого-либо похожего округа, который не реализовал данную программу.

Предполагается (что довольно важно), что две группы, используемые для такого анализа, в целом сопоставимы — за исключением активирования; таким образом, любое значимое различие в исходах между этими группами может быть обусловлено оцениваемой нами программой или политикой. Допустим, что один округ в штате Иллинойс реализует программу обучения новым профессиям в надежде снизить уровень безработицы. Однако в течение двух последующих лет безработица продолжает расти. Означает ли это, что программа провалилась? Как это выяснить?

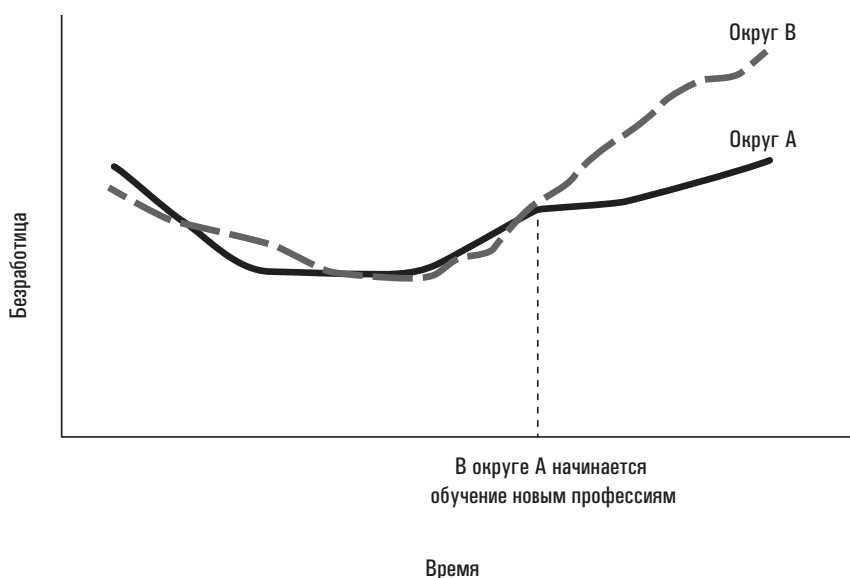
Влияние обучения новым профессиям на уровень безработицы в округе А



Разумеется, на этой арене могут действовать множество экономических сил, в том числе и вероятность продолжительного экономического спада. Подход «разница в различиях» сравнил бы изменение уровня безработицы в течение какого-то времени в оцениваемом нами округе с уровнем безработицы в каком-либо из соседних округов, где не реализуется программа обучения новым профессиям. Эти два округа должны быть похожи во всех

остальных важных аспектах, таких как структура промышленности, демографическая картина и т. п. Как уровень безработицы в округе, где реализуется программа обучения новым профессиям, изменяется со временем *в сравнении с округом, в котором такая программа не внедрена*? Мы можем оценить эффект данной программы, сравнив изменения в этих округах за время, в течение которого проводилось данное исследование («разница в различиях»). Другой округ в этом исследовании действует, по сути, как контрольная группа, что позволяет нам воспользоваться данными, собранными до и после воздействия. Если контрольная группа выбрана правильно, то она будет подвергаться воздействию тех же более широких сил, что и подопытная группа. Подход «разница в различиях» может быть особенно поучителен, когда активирование сперва оказывается неэффективным (после начала реализации программы обучения новым профессиям уровень безработицы повышается), хотя контрольная группа показывает, что в отсутствие активации эта тенденция могла быть еще хуже.

Влияние обучения новым профессиям на уровень безработицы в округе А, для сравнения используется округ В



Анализ на основе разрывности. Один из способов создать подопытную и контрольную группы — сравнить исходы для какой-то группы, которая едва подходит для воздействия или активирования, с исходами для группы,

которая едва превысила порог непригодности и не получает активирования. Те, кто окажется чуть выше или чуть ниже некоторого случайно выбранного предела, например результата сдачи экзаменов или минимального дохода семьи, будут почти идентичны во многих важных отношениях; то обстоятельство, что одна группа получила активирование, а другая — нет, является, по сути, делом случая. В итоге мы можем сравнить их исходы таким способом, который позволял бы сделать значимые выводы относительно эффективности соответствующего воздействия.

Допустим, какому-то учебному округу требуется летняя школа, в которой могли бы заниматься отстающие ученики. Руководство округа хотело бы знать, представляет ли собой какую-либо долгосрочную учебную ценность такая программа летнего обучения. Как обычно, простое сравнение учеников, посещающих и не посещающих эту школу, было бы не только бесполезно, но даже вредно. Ученики, которые посещают такую школу, ходят туда *именно потому, что являются отстающими*. Даже если программа обучения в этой школе весьма эффективна, успеваемость ее учеников в долгосрочной перспективе все же, по-видимому, будет хуже, чем успеваемость учеников, не нуждающихся в дополнительных занятиях. Мы хотим знать, какой будет успеваемость отстающих учеников после посещения летней школы *по сравнению с их успеваемостью в случае, если бы они ее не посещали*. Да, мы могли бы провести нечто вроде управляемого эксперимента, в ходе которого случайным образом отобрали бы отстающих учеников для посещения летней школы, и, соответственно, отстающих учеников, которые не будут в нее ходить (контрольная группа). Однако это бы предполагало отказ контрольной группе в доступе к программе обучения в летней школе, между тем как мы предполагаем, что обучение в летней школе будет полезно *всем* отстающим ученикам.

Вместо этого подопытная и контрольная группы создаются путем сравнения учеников, которые оказались чуть ниже порога для летней школы, с учениками, которые находятся чуть выше этого порога. Это можно представлять себе так: ученики, провалившие экзамены по итогам полугодия, существенно отличаются от учеников, которые успешно их сдали. Но ученики, набравшие 59% от максимальной суммы баллов (что не дает им права перейти в следующий класс), не *столь уж сильно* отличаются от учеников, у которых 60% от максимальной суммы баллов (проходной балл). Если для тех, кто провалил экзамены по итогам полугодия, организовать некую программу

обучения (например, обязательная подготовка к выпускным экзаменам под руководством преподавателя), то у нас появились бы приемлемые подопытная и контрольная группы, где бы мы сравнили результаты выпускных экзаменов тех, кто недобрал самую малость до проходного балла (и получил возможность готовиться под руководством преподавателя), с результатами тех, кто едва превысил проходной балл в ходе сдачи экзаменов по итогам полугодия (и не пользовался помощью преподавателя).

Этот подход был применен, чтобы определить эффективность тюремного заключения малолетних преступников в качестве сдерживающего фактора от повторного совершения преступлений. Очевидно, при выполнении такого анализа нельзя просто сравнивать уровень повторной преступности среди несовершеннолетних, получавших тюремные сроки, и тех, кто отделался более легким наказанием. *Малолетние преступники, которых сажают в тюрьму, как правило, совершают более серьезные преступления, чем малолетние преступники, получающие более легкие приговоры; именно поэтому их и сажают за решетку.* Естественно, мы не можем создать подопытную и контрольную группы путем случайного назначения тюремных сроков (если не хотите на двадцать пять лет угодить за решетку за то, что повернули на красный свет). Исследователь Рэнди Хьялмарссон, которая в настоящее время работает в Лондонском университете, воспользовалась статистикой вынесенных приговоров для малолетних преступников в штате Вашингтон, чтобы понять, как влияет тюремное заключение на будущее криминальное поведение этих людей. В частности, она сравнила уровень повторной преступности среди несовершеннолетних, получивших тюремный срок буквально «на грани» действующего уголовного законодательства, с уровнем повторной преступности среди тех малолетних преступников, кто избежал тюремного срока также буквально «на грани» (то есть «отделался выплатой штрафа или условным сроком»)⁹.

Система уголовных наказаний в штате Вашингтон создает для каждого осужденного некую систему координат, используемую для администрирования меры наказания преступника. Ось X измеряет признанные судом прежние преступления данного осужденного. Например, каждое прежнее тяжкое уголовное преступление оценивается в один балл; каждое прежнее уголовное преступление, не относящееся к категории тяжких, оценивается в четверть балла. Сумма баллов округляется до ближайшего меньшего целого значения (почему это важно, станет понятно из дальнейшего объяснения).

По оси Y измеряется тяжесть текущего преступления по шкале от E (наименее тяжкое) до A+ (самое тяжкое). Приговор осужденному вычисляется (в буквальном смысле) путем нахождения подходящей клетки в такой системе координат. Осужденный, предыдущие преступления которого оцениваются в два балла, совершивший тяжкое уголовное преступление класса B, получает от пятнадцати до тридцати шести месяцев тюремного заключения. Осужденный, предыдущие преступления которого оцениваются лишь в один балл, совершивший такое же уголовное преступление, не отправляется за решетку. Именно такая разрывность шкалы наказания определила выбор стратегии исследования. Рэнди Хьялмарссон сравнила исходы для осужденных преступников, которые оказались чуть выше и чуть ниже порога получения тюремного срока. Как объясняет Хьялмарссон в своей статье: «Если есть два преступника с текущим преступлением класса C+ и суммами баллов за предыдущие приговоры $2\frac{3}{4}$ и 3, то лишь последний из этих двоих будет отправлен в тюрьму».

С точки зрения данного исследования эти два человека, по сути, одинаковы — пока один из них не попадет за решетку. С этого момента их поведение резко разнится. Вероятность осуждения малолетних преступников, попавших за решетку, за какое-либо другое преступление после их освобождения из тюрьмы оказывается существенно ниже.

Нас интересуют прежде всего действенные способы анализа. Это касается медицины, экономики, бизнеса, уголовного судопроизводства и всего остального. Все же причинно-следственные зависимости — крепкий орешек, который не так-то просто расколоть даже в случаях, когда причина и следствие кажутся совершенно очевидными. Чтобы уяснить подлинную эффективность того или иного активирования, нужно знать, «что было бы в противном случае», то есть в отсутствие соответствующего активирования, или воздействия. Зачастую узнать это не так-то легко, а порой и невозможно. Рассмотрим пример не из области статистики. Оказалась ли Америка в большей безопасности после вторжения в Ирак?

Существует лишь один честный ответ на этот вопрос: мы никогда не узнаем это наверняка. А не узнаем потому, что нам не дано знать, что случилось бы, если бы Америка не вторглась в Ирак. Да, Соединенные Штаты не нашли в Ираке оружия массового поражения. Но не исключено, что на следующий день после того, как США *не вторглись в Ирак*, у Саддама Хусейна во время

принятия душа могли бы возникнуть следующие мысли: «А не обзавестись ли мне водородной бомбой? Может быть, Северная Корея продаст мне парочку?» Кто знает, что случилось бы потом...

Правда, на следующий день после того как Соединенные Штаты *не вторглись в Ирак*, Саддам Хусейн во время принятия душа мог бы подумать: «А не обзавестись ли мне...», после чего ему на голову могла бы упасть отвалившаяся от стены кафельная плитка и он бы умер, так и не додумав до конца интересную мысль о возможности покупки водородной бомбы.

В этом случае мир избавился бы от Саддама Хусейна, не понеся колоссальных издержек, связанных с вторжением Америки в Ирак. Короче говоря, вариантов — множество, но никто никогда не сможет сказать наверняка, что случилось бы, если бы Америка не вторглась в Ирак.

Задача любой программы статистического оценивания — узнать, «что было бы в противном случае». Только так мы можем измерить эффективность того или иного активирования, или воздействия. В случае статистического управляемого эксперимента роль «что было бы в противном случае» исполняет контрольная группа. В случаях, когда проведение управляемого эксперимента нецелесообразно или невозможно по этическим соображениям, нужно найти какой-то другой способ приближения того, «что было бы в противном случае». Наше понимание окружающего мира зависит от нахождения разумных способов решения этой задачи.



[Почитать описание, рецензии
и купить на сайте](#)

Лучшие цитаты из книг, бесплатные главы и новинки:

