

Квазиэкспериментальные планы и прикладные исследования

Обзор задач главы

В этой главе мы будем рассматривать прикладные исследования, и в частности исследования по оценке программ. С прикладными исследованиями вы встречались в главе 3 при сравнении их с фундаментальными исследованиями. Из этой главы вы узнаете, что прикладные исследования представляют собой мощное направление в американской экспериментальной психологии и отображают один из центральных моментов американской идеологии — идею полезности. Оценка программы, призванной определенным образом помочь людям, — это вид прикладного исследования, направленный на изучение ее эффективности. В данной главе также описываются так называемые квазиэкспериментальные планы, часто используемые в прикладных исследованиях, в том числе в исследованиях по оценке программ¹. Изучив данную главу:

- вы сможете определить двойное назначение прикладных исследований;
- поймете, почему прикладная психология всегда являлась важной частью американской психологии;
- сумеете охарактеризовать планы и этические проблемы, связанные с прикладными исследованиями, и поймете, какие из планов, описанных в предыдущих главах, были квазиэкспериментальными;
- сможете описать особенности плана с неэквивалентными контрольными группами и поймете, почему такой план обязательно содержит осложнители;
- поймете, почему уравнивание неэквивалентных групп по оценкам предварительного тестирования может вызвать эффект регрессии;
- научитесь описывать особенности планов с прерванной временной последовательностью и поймете, как с их помощью можно оценивать существующие тенденции;

¹ Очень важно обратить внимание на то, что совместное рассмотрение квазиэкспериментальных и прикладных исследований, осуществляемое в этой главе, не означает, что квазиэкспериментальные планы встречаются только в прикладных исследованиях. Чаще всего их можно встретить именно в таких исследованиях, но фундаментальные исследования также могут проводиться с использованием названных планов. Более того, их совместное рассмотрение не означает, что в прикладных исследованиях нельзя использовать экспериментальные планы.

- сумеете описать стратегии проведения анализа потребностей;
- поймете цели и процедуру проведения предварительной оценки, суммарной оценки и оценки затрат-эффективности;
- сможете охарактеризовать особые этические проблемы, часто сопутствующие исследованиям по оценке программ.

Как было отмечено в самом начале, главная моя цель —развить в вас интерес к изучению законов поведения и желание стать психологом-экспериментатором. Мой преподавательский опыт подсказывает мне, что некоторые из вас действительно будут участвовать в проведении исследований, но большинство не станет заниматься этим. Тем не менее многие из вас станут профессиональными психологами и посвятят себя работе консультанта, специалиста по работе с персоналом или другой психологической или связанной с охраной здоровья деятельности (например, физиотерапии). Поэтому вы встретитесь в своей работе со сферой прикладных исследований и оценки программ. Очень возможно, что вам придется:

- читать, понимать и критически оценивать литературу, посвященную исследованиям по оценке программ, которая будет использоваться службой, в которой вы работаете;
- участвовать в планировании новых программ, информируя (тактично) тех, кто незнаком с техникой исследований, об адекватности их оценки программ;
- участвовать в усовершенствовании методов и подходов, используемых в вашем учреждении для получения аккредитаций. Возможно, это будет связано с вашими знаниями по курсу методов исследований;
- разрабатывать исследование по оценке программы какого-либо учреждения и нести за него ответственность.

Другими словами, эта глава важна вне зависимости от того собираетесь ли вы когда-либо проводить психологические исследования.

За пределами лаборатории

Впервые вы узнали о различии между фундаментальными и прикладными исследованиями из главы 3. Важнейшей целью фундаментальных исследований по психологии является расширение наших знаний о законах поведения. Эти знания могут затем найти применение, но не это является основным мотивом работы — знание ценится само по себе. Прикладные исследования призваны увеличить наши знания о причинах существующих проблем и нацелены на их решение. Второе различие заключается в том, что фундаментальные исследования обычно проводятся в стенах лаборатории, находящейся в университетском городке или в кампусе колледжа, а прикладные исследования чаще всего проходят вне академической обстановки, в клиниках, учреждениях социального обеспечения, тюрьмах и на предприятиях. Конечно, существуют и исключения: некоторые фундаментальные исследования проводятся в «полевых» условиях, а некоторые прикладные — в лабораториях.

Чтобы дать вам представление о разнообразии прикладных исследований, рассмотрим заголовки нескольких статей, опубликованных в двух видных психоло-

гических журналах, «Журнале прикладной психологии» и «Журнале экспериментальной психологии»:

«Пол и взаимосвязи между оценкой справедливости оплаты труда или между продвижением по службе и удовлетворением от работы» (Witt & Nye, 1992);

«Зрительные пространственные способности пилотов» (Dror, Kosslyn, & Waag, 1993);

«Осознавание обстановки, сопутствующей вождению автомобиля: эксплицитное и имплицитное знание в динамической пространственной памяти» (Gugerty, 1997);

«Явление вероятности и практические аспекты программ по сохранению видов, находящихся под угрозой вымирания» (DeKay & McClelland, 1996).

Эти заголовки отображают две особенности прикладных исследований. Во-первых, в соответствии с приведенным выше определением данные исследования фокусируются на легко локализуемых проблемах: от удовлетворения работой до охраны вымирающих видов. Во-вторых, такие названия показывают, что хотя главной целью прикладных исследований является решение конкретных проблем (например, улучшение навыков вождения), они также способствуют углублению наших знаний о фундаментальных психологических процессах (например, о пространственной памяти). И действительно, как показывает исследование, разработанное для оценки так называемой «теории справедливости» (Adams, 1965), между фундаментальными и прикладными исследованиями существует тесная связь. Основное допущение названной теории состоит в том, что для оценки взаимоотношений с другими людьми мы сравниваем относительный вклад, сделанный нами в эти взаимоотношения, с относительной полученной пользой. Если затраты не соответствуют полученной пользе, мы ощущаем ситуацию как несправедливую и стремимся восстановить баланс. Например, если в супружеской паре оба супруга работают, а домашним хозяйством занимается в основном жена, то она может почувствовать несправедливость и попытаться прибегнуть к различным процедурам (например, договориться с мужем), чтобы нормализовать положение.

Большинство прикладных исследований выполняют двойную функцию: они одновременно решают прикладные задачи и подтверждают фундаментальные психологические явления, тем самым способствуя разработке теорий. Справедливость, к примеру, часто изучается именно прикладными исследованиями. Указанное выше исследование корреляции между полом и оценкой оплаты труда или продвижения по службе затрагивает важную производственную проблему (как увеличить удовлетворение от работы), но также дает информацию, касающуюся более общего вопроса адекватности теории справедливости. Пример, иллюстрирующий оценку метода развития способности к опознанию, также показывает эту особенность прикладных исследований.

Пример 23. Прикладное исследование

Вспомните исследование с участием шотландского клуба ныряльщиков (глава 8), показавшее, что запомненная информация воспроизводится лучше всего, если и запоминание, и воспроизведение происходят в одном и том же месте (например, под водой). Важность контекста для запоминания и общеизвестная неточность свидетельств очевидцев навели Эварда Гейзельмана и Рональда Фишера на мысль

о разработке особого подхода, который они назвали «когнитивным интервью». Одна из особенностей этого подхода — принцип, обозначенный исследователями как «соответствие воспроизводимого контекста обстоятельствам события (*event-interview similarity*)», применяя который исследователь старается «восстановить в памяти свидетелей внешние (например, погода), эмоциональные (чувство страха) и когнитивные (мысли) аспекты, переживаемые в момент преступления» (Fisher, Geiselman, & Amador, 1989, p. 723). Говоря кратко, исследователь пытается помочь свидетелю мысленно воспроизвести контекст произошедших событий.

Гейзельман и его коллеги ранее продемонстрировали эффективность своего подхода в условиях лаборатории, но они хотели оценить метод «за пределами знакомых стен» (Fisher et al., 1989, p. 742). Они обучили методу когнитивного интервью семь полицейских из отдела ограблений полиции Майами штата Флорида. В сравнении с девятью полицейскими из контрольной группы обученные полицейские при опросе свидетелей получали более надежную информацию (т. е. подтверждаемую другими источниками). Из рис. 10.1 видно, что обе группы показали приблизительно одинаковые результаты на этапе предварительного тестирования, включавшем 4 месяца, на протяжении которых полицейские записывали свои опросы свидетелей, а независимые судьи оценивали полученные сведения. После того как экспериментальная группа завершила обучение, обе группы провели и записали несколько опросов, оцененных таким же образом. Выяснилось, что обученные методу когнитивного интервью полицейские способны получить от свидетелей больше информации.

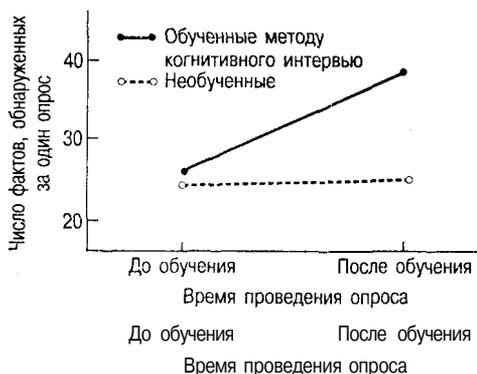


Рис. 10.1. Проверка эффективности метода когнитивного интервью в ходе прикладного исследования (из работы Фишера и др., 1989)

Исследование Фишера и др. — это превосходный пример того, как прикладное исследование помогает решить конкретные проблемы, а также способствует расширению наших знаний о фундаментальных психологических явлениях. Оно показывает, как особый метод опроса может увеличить эффективность работы полиции, и в то же время дает новые подтверждения важности контекста для запоминания.

Прикладная психология в историческом контексте

Поскольку психология в Америке формировалась в академических условиях, вы можете решить, что она традиционно развивалась в направлении фундаментальных исследований. Это не так. С тех пор, когда в конце XIX в. сформировалась

новая наука, психологи в Соединенных Штатах интересуются прикладными исследованиями. Во-первых, под давлением общества в начале XX в. психологам пришлось показать, что они способны сделать что-то для изменения общества к лучшему. Чтобы получить приличный кусок от пирога академического финансирования в те времена, когда психологические лаборатории только открывались, психологи должны были продемонстрировать, что их идеи могут найти себе применение.

Психологи-исследователи занимались не только расширением знаний, но и нередко старались использовать методы фундаментальных исследований для решения вопросов образования, психического здоровья, воспитания детей и, как в случае Уолтера Майлза, спорта. Майлз в 1920 г. был директором лаборатории при Стэнфордском университете. Он посвятил свою работу в основном фундаментальным исследованиям и тем не менее в 1927 г. оказался на футбольном поле, где практиковалась университетская футбольная команда (рис. 10.2, Майлз в пиджаке). Легендарный стэнфордский тренер «Поп» Уорнер («Pop» Warner) был известен своим новаторским подходом к тренировкам и интересовался всеми методами, способными улучшить игру команды. Появляется Майлз, разработавший «сложный хронограф» для одновременного определения скорости реакции у семи футболистов (Miles, 1928). По сигналу семь мячей для гольфа падали во вращающийся с постоянной скоростью барабан (по одному мячу на каждого игрока) и семь игроков выполняли рывок вперед, толкая раму, которая тянула за нить, выбрасывающую в барабан еще семь мячей. Мячи оставляли отметки на барабане и, зная скорость его вращения, а также расстояние между отметками, Майлз определял скорость реакции игроков. Ученый провел серию исследований со сложным хронографом (см., например, Miles, 1931) и показал, что его можно использовать для отбора игроков с наивысшей скоростью реакции, но прибор не получил широкого применения.

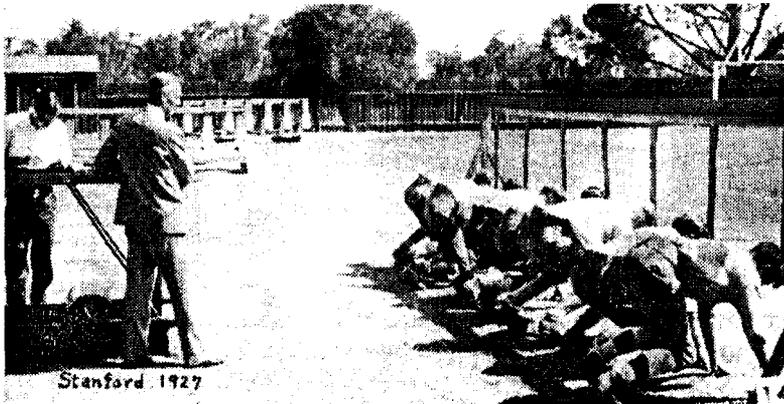


Рис. 10.2. Одновременная проверка скорости реакции у членов стэнфордской футбольной команды, около 1927 г. (из архивов по истории американской психологии, Университет Экрон, город Экрон, штат Огайо)

Более основательная попытка применения психологических принципов была произведена в первой клинике для коррекции психических проблем, основанной Лайтнером Уитмером при Университете штата Пенсильвания. Уитмер учился на психолога-экспериментатора в лаборатории Вундта в Лейпциге и с 1890 г. стал активно заниматься экспериментами. Однажды (в 1895 г.) он выступил за отделение психологов-экспериментаторов от недавно образовавшейся Американской психологической ассоциации и за образование отдельной организации, так как видел недостаточную научную подготовку *АРА* (Goodwin, 1985). В марте 1896 г. школьный учитель привел в лабораторию Уитмера маленького мальчика с необучаемостью (так мы называем подобное явление сегодня), и это событие отметило основание первой американской психологической клиники (McReynolds, 1987). Уитмер посвятил свою работу применению всех своих знаний о психических процессах для лечения различных психологических дисфункций, а в особенности нарушений способности к обучению, влияющих на школьную успеваемость.

Интерес психологов к прикладной области хорошо виден на примере движения по умственному тестированию. В предыдущей главе мы говорили о том, что тестирование интеллектуальных и личностных особенностей играет в Америке очень большую роль. Корни этого направления можно найти в работах первых психологов. Например, Джеймс Кеттелл, так же как и Уитмер, выпускник лаборатории Вундта, в 1890 г. ввел в использование термин «умственный тест» и в 1890-х гг. с учетом открытий Гальтона, сделанных в Англии, разработал в Колумбии расширенную (но обладавшую непоправимыми изъянами — см. Sokal, 1987) программу тестирования. В начале 20-х гг. XX в. он основал существующую и сегодня Психологическую корпорацию — гиганта индустрии психологического тестирования (например, она издает все тесты интеллекта Векслера).

И в завершение рассмотрим научную жизнь Гуго Мюнстерберга — идеальный для Америки пример прикладного психолога. Как и многие другие первые психологи, он учился на психолога-исследователя, но стал известен как эксперт по применению психологических подходов в бизнесе, клинической психологии и даже в суде. Одно время он был одиозной фигурой в Америке, и поэтому до недавних пор при историческом рассмотрении прикладной психологии его имя обычно опускалось. Чтобы более подробно узнать о человеке, заслужившем титул первого настоящего прикладного психолога Америки (несмотря на то, что он был немец), прочитайте вставку 10.1.

В начале XXI в. психологи заинтересованы в практическом применении своих открытий так же сильно, как и их предшественники в XX в. Профессиональные психологи не просто применяют научные знания о поведении для решения проблем, с которыми они сталкиваются. Они также выполняют исследования, соответствующие приведенному выше определению прикладного исследования. Это означает, что они разрабатывают и проводят исследования, чтобы найти решение конкретных проблем и одновременно чтобы обогатить фундаментальную базу психологических знаний. Однако прикладным исследованиям свойственны различные трудности, которые не встречаются в лабораториях.

ВСТАВКА 10.1

История — Гуго Мюнстерберг, экстраординарный психолог

Неудивительно, что прикладная психология представляет собой центральную силу американской психологии. В конце концов американцы славятся своим неизменным прагматизмом. «Какая в этом польза?» - такой вопрос является для нас вполне естественным. Примечательно поэтому, что одним из самых выдающихся прикладных психологов Америки считается Гуго Мюнстерберг (1863—1916), немец по национальности. Когда в 1892 г. он приехал в страну с целью основать психологическую лабораторию в Гарварде, он едва говорил по-английски. К моменту своей смерти он написал около 20 английских книг и был признан пионером производственной и судебной психологии и ученым, внесшим заметный вклад в клиническую психологию.

«Психологию и эффективность производства» Мюнстерберга, опубликованную в 1913 г., иногда считают первой вехой в развитии производственной психологии (Goodwin, 1999). Она включает в себя советы и описания исследований на такие темы, как подбор людей для определенных видов работ, факторы, наиболее сильно влияющие на производительность труда, а также предложения по маркетингу и рекламе. В работу также вошли его собственные исследования по разработке тестов для подбора персонала, в которых использовались процедуры проверки валидности, ставшие сегодня стандартными. К примеру, когда Мюнстерберга попросили разработать систему отбора для Новой английской телефонной компании, он проанализировал работу операторов с точки зрения решения познавательных задач (он выделил 14 видов задач), разработал тесты для проверки способности к решению каждого из видов и подтвердил валидность тестов, показав, что люди, получающие наивысшие оценки, становятся лучшими операторами.

Мюнстерберг был также первопроходцем в области судебной психологии - применении психологических принципов в юриспруденции. Его книга «Сточки зрения свидетеля», вышедшая в 1908 г., содержала поразительное описание того, как опасно верить свидетельским показаниям, в ней он высказывал предостережения, которые позднее были сформулированы в XX в. в ходе экспериментальной работы над этим вопросом. Не имея непосредственного экспериментального подтверждения несовершенства памяти свидетелей, для того чтобы показать ненадежность наблюдений людей, Мюнстерберг использовал свои знания об иллюзиях восприятия. Он также утверждал, что различные психологические методы могут быть использованы, чтобы уличить во лжи. Однако, как это часто бывает, если ученый попадает в поле зрения широкой публики, деятельность Мюнстерберга была поставлена под сомнение. Ссылаясь на исследование, показавшее, что суждение группы более адекватно только в случае испытуемых-мужчин, Мюнстерберг заключил, что суд присяжных не должен включать женщин. Его вывод получил широкую огласку в прессе и навлек на него гнев публики, выраженный заголовком газетной статьи, процитированной в книге Хотерсэла (1990 г.): «Разгневанные Мюнстербергом суффражистки считают, что женщины вполне подходят для суда присяжных».

Еще большую полемику вызвала попытка Мюнстерберга нарисовать в благоприятном свете портрет своей родной Германии. К сожалению, он сделал это прямо перед началом Первой мировой войны, что привело к лавине пышущих ненавистью писем с обвинениями в том, что Мюнстерберг - немецкий шпион. Согласно газете Хотерсэлу (1990), некто даже предложил Гарварду 10 млн долларов за увольнение Мюнстерберга, но отказался от своего предложения, когда ученый согласился подать в отставку, если этот человек заплатит 5 млн университету и 5 млн самому Мюнстербергу! Когда ученый в 1916 г. перед самым окончанием войны умер, его считали одним из злодеев Америки. «Американский психологический журнал», обычно печатающий длинные некрологи о психологах гораздо меньшего значения, чем Мюнстерберг, даже не упомянул о его смерти. Все это привело к тому, что значение этого ученого как первопроходца и человека, всячески способствовавшего развитию прикладной психологии, было забыто на долгие годы (Landy, 1992).

Проблемы прикладных исследований

Поскольку вы изучили главы 2, 5 и 6, для вас не будет неожиданностью материал данного раздела. Прикладные исследования связаны со следующими проблемами.

- *Этические дилеммы* (глава 2). Ученые, проводящие полевые исследования, могут столкнуться с проблемами обоснованного согласия и ограждения личной жизни. Также не всегда возможно проведение дебрифинга. Исследования, проходящие в производственных условиях, могут содержать элемент принуждения, если работники решат, что их служебное положение зависит от участия в исследовании (более подробно об этике прикладных исследований см. вставку 10.3).
- *Компромисс между внешней и внутренней валидностью* (глава 5). Поскольку исследования по прикладной психологии часто проходят в полевых условиях, исследователь может отчасти потерять контроль за переменными, влияющими на ход исследования. Поэтому опасность возможного осложнения снижает внутреннюю валидность исследования. С другой стороны, прикладные исследования обычно имеют высокую внешнюю валидность, так как условия их проведения более близки к ситуациям реальной жизни и в их ходе решаются повседневные проблемы,
- *Проблемы, свойственные только межсубъектным планам* (глава 6). В прикладных исследованиях для формирования групп не всегда возможно использовать случайное распределение, поэтому нередко приходится сравнивать неэквивалентные группы. Это, конечно, может снизить внутреннюю валидность исследования вследствие проблем отбора или наличия взаимодействия между отбором участников и другими факторами, такими как взросление. При попытке использовать уравнивание, чтобы добиться определенного уровня эквивалентности, может возникнуть проблема регрессии (глава 5).
- *Проблемы, свойственные только внутрисубъектным планам* (глава 6). В прикладных исследованиях с внутрисубъектными факторами не всегда возможно провести необходимое позиционное уравнивание. Поэтому в исследовании может возникнуть неконтролируемый эффект последовательности. Исследования, продолжающиеся длительный период, нередко также сталкиваются с проблемой истощения.

Прежде чем продолжить дальнейшее изучение главы, вам необходимо вновь обратиться к соответствующим разделам глав 2, 5 и 6 и повторить материал, посвященный указанным идеям. Кроме того, вам стоит повторить раздел главы 5 о том, какие выводы могут быть сделаны в случае управляемых и субъективных переменных.

Квазиэкспериментальные планы

Строго говоря, «настоящие» экспериментальные исследования проводятся с использованием управляемых независимых переменных и либо эквивалентных групп в случае межсубъектных планов, либо позиционного уравнивания в случае внутрисубъектных планов. Все остальные исследования являются квазиэкспериментальными («почти» экспериментальными). В главе 5 обсуждение управляемых переменных основывалось на том, что субъектные переменные используются в каче-

стве независимых. В данной главе обсуждение этой темы будет несколько расширено. В целом, о **квазиэксперименте** говорят тогда, когда вследствие неполного контроля за переменными нельзя сделать выводы о существовании причинно-следственной связи. Мы уже встречались с некоторыми квазиэкспериментальными планами:

- однофакторный план с неэквивалентными группами, включающий один или несколько уровней;
- факторный план с неэквивалентными группами;
- факторный план $P \times E$;
- все корреляционные планы.

В этой главе мы рассмотрим два вида планов, которые обычно описываются в работах, посвященных квазиэкспериментальным планам (например, Cook & Campbell, 1979): планы с неэквивалентными группами и планы с прерванной временной последовательностью. Существуют и другие квазиэкспериментальные планы, но эти два встречаются наиболее часто.

Планы с неэквивалентными контрольными группами

Такие исследования проводятся для оценки эффективности воздействия программ. Люди, участвующие в программе, сравниваются с членами контрольной группы, **не** подвергающимися воздействию. Этот вид планов используется, когда нельзя применить случайное распределение. Кроме разницы в значениях независимой переменной контрольная группа имеет и другие отличия от группы воздействия. Это значит, что группы неэквивалентны с самого начала исследования. В плане такого исследования вы должны узнать особый случай, обозначенный в главе 7 как план с «неэквивалентными» группами, в котором сравниваются неидентичные группы, нередко отбираемые на основании таких субъектных переменных, как пол, возраст или свойства личности. В случае **плана с неэквивалентной контрольной группой** группы не равны друг другу с начала исследования, а *кроме того*, в ходе исследования они участвуют в разных событиях. Поэтому использование такого плана влечет за собой осложнение, способное затемнить интерпретацию исследования. Тем не менее подобный план может быть весьма полезен для оценки программы воздействия, когда невозможно использовать случайное распределение.

В соответствии со схемой, предложенной Кэмпбеллом и Стэнли (Campbell & Stanley, 1963), план с неэквивалентной контрольной группой можно представить следующим образом.

Экспериментальная группа:	N_1	V	N_2
Неэквивалентная контрольная группа:	N_1		N_2

где N_j и N_2 обозначают соответственно доэкспериментальные и послеэкспериментальные наблюдения или измерения, а V обозначает оцениваемую программу воздействия. Поскольку группы могут различаться на стадии предварительного тестирования, сравнение двух групп заключается не просто в изучении разницы результатов, полученных на этапе заключительного тестирования, а в сравнении степени изменений, произошедших между предварительным и заключительным тестированием. Таким образом, статистически сравниваются оценки изменений (разница между N_1 и N_2) для каждой группы. Рассмотрим это на конкретном примере.

Предположим, что руководство компании по производству электрических сковородок хочет ввести новый свободный график рабочего дня. Рабочие будут продолжать работать по 40 часов в неделю, но в соответствии с новым расписанием они смогут начинать и заканчивать рабочий день в разное время или отработать все 40 часов за 4 дня и отдыхать 3 дня в неделю. Руководство надеется, что это нововведение поднимет рабочий настрой, а следовательно, увеличит производительность труда, и поэтому разрабатывает квазиэксперимент для проверки предположения. Компания владеет двумя фабриками: одной вблизи от Питтсбурга и второй неподалеку от Кливленда. Бросив монетку, руководители компании решили сделать из питтсбургской фабрики экспериментальную группу, а из кливлендской — неэквивалентную контрольную группу. Очевидно, что квазиэкспериментальное исследование проводится по той причине, что рабочих невозможно случайным образом распределить по двум фабрикам (только представьте себе расходы по перемещению рабочих, затраты на удовлетворение жалоб профсоюзов и т. д.). Независимая переменная принимает два значения: наличие или отсутствие свободного графика работы, а зависимой переменной является особым образом измеряемый показатель производительности труда. Предположим, что окончательный вариант плана следующий:

Фабрика в Питтсбурге:	Предварительное тестирование	Средняя производительность труда за 1 месяц до введения свободного графика работы
	Экспериментальное воздействие	Свободный график работы введен на 6 месяцев
	Заключительное тестирование	Средняя производительность труда в течение 6 месяцев работы по свободному графику
Фабрика в Кливленде:	Предварительное тестирование	Средняя производительность труда за 1 месяц до введения свободного графика работы в Питтсбурге
	Экспериментальное воздействие	Отсутствует
	Заключительное тестирование	Средняя производительность труда в течение 6 месяцев работы по свободному графику на фабрике в Питтсбурге

Результаты

На рис. 10.3 приведены четыре результата данного квазиэксперимента. Все графики показывают увеличение производительности труда на фабрике в Питтсбурге. Вопрос в том, вызвано ли это увеличение экспериментальной программой или другим (другими) фактором (факторами). Прежде чем продолжить чтение, попробуйте определить, какой из графиков лучше всего демонстрирует, что введение свободного графика работы увеличивает производительность труда, а также попытайтесь выявить факторы, угрожающие внутренней валидности исследования и затрудняющие интерпретацию результатов.

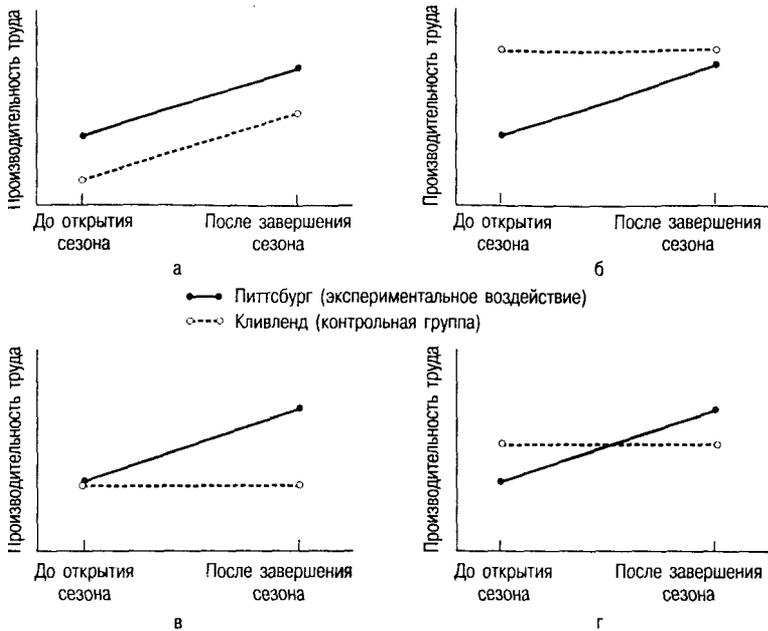


Рис. 10.3. Гипотетические результаты исследования по плану с неэквивалентными группами

Я думаю, что вам не составило труда увидеть, что на увеличение производительности труда кроме свободного графика работы повлияли какие-то другие факторы. Этот график говорит о важности использования контрольной группы, даже если она является неэквивалентной. Да, действительно, питтсбургская производительность труда увеличилась, но то же самое произошло и в Кливленде, а следовательно, данные о росте производительности в Питтсбурге могут быть связаны не с успехами экспериментальной программы, а с различными факторами, угрожающими внутренней валидности, которые вы изучали ранее. История и взросление — одни из таких возможных факторов. Может быть, в период между предварительным и заключительным тестированием проводились национальные выборы и рабочие разных предприятий испытали связанный с этим эмоциональный подъем, что привело к росту производительности труда. А возможно, производительность увеличилась с приобретением опыта.

Рисунок 10.3, б говорит о том, что производительность в Кливленде поддерживалась высокой на протяжении всего исследования, а в Питтсбурге сначала была очень низкой, но затем увеличилась вследствие введения свободного графика работы. Однако здесь есть две проблемы. Во-первых, результаты, полученные в Кливленде, могут отображать эффект потолка (см. главу 8). Это значит, что уровень производительности изначально был настолько высок, что эксперимент не смог показать его дальнейшее увеличение. Если бы можно было увидеть рост производительности (т. е. если бы значения на оси Y могли увеличиться), то вы бы увидели две параллельные линии, как на рис. 10.3, а. Вторая проблема заключается в том, что поскольку начальная производительность в Питтсбурге была настолько низкой, ее увеличение вызвано скорее эффектом регрессии, чем экспериментальной программой.

Рисунок 10.3, в на первый взгляд кажется отображением идеальных результатов. Обе группы начинают с одного уровня производительности труда, но ее рост наблюдается только в экспериментальной группе (Питтсбург). Такой результат может действительно иметь место, и исследователи будут просто счастливы его получить, но тем не менее возможна проблема. Поскольку две группы неэквивалентны, эффект отбора субъектов может сочетаться с влиянием других факторов. Это значит, что определенный фактор, например история или взросление может воздействовать на рабочих одной фабрики и не влиять на рабочих другой. Например, несложно представить здесь проблему отбор χ история: некоторое событие повлияло на рабочих Питтсбурга, а на рабочих Кливленда не оказало никакого влияния. Возможно, на питтсбургских рабочих повлияло то, что они знали об участии в исследовании (помните Хоторн?). А может быть, в период между предварительным и заключительным тестированием Стилеры выиграли кубок *Super Bowl* по американскому футболу, а так как рабочие в Питтсбурге очень активно болеют за свою команду, то общее ощущение успеха могло повысить у них рабочий настрой, а следовательно, и производительность труда. А Брауны, которые никогда не получали кубок, не смогут вызвать рост продуктивности у рабочих Кливленда.

Вы, вероятно, заметили сходство рис. 10.3, в и 10.1, иллюстрирующих исследование метода когнитивных интервью. Фишер и др. (Fisher et al., 1989) не проводили случайного распределения полицейских по группам, одной, в которой проводилось обучение по данному методу, и второй, контрольной, но исследователи считали обе группы приблизительно эквивалентными, так как применяли процедуру уравнивания с использованием «информации, собранной в... ходе предварительных опросов и сведений, полученных от вышестоящих офицеров полиции» (Fisher et al. 1989, p. 724). Таким образом, они считали, что результаты, приведенные на рис. 10.1, отображают реальное влияние обучения и что эффект отбора был минимальным.

Результаты, представленные на рис. 10.3, г, лучше всего подтверждают эффективность программы. В этом случае производительность труда у группы воздействия в начале исследования *ниже*, чем у контрольной группы, а в конце становится *выше*, чем у нее. Регрессию как причину улучшения можно отбросить, так как она возможна только до уровня контрольной группы, но не выше. Конечно, нельзя совсем исключить проблему отбора и возможность взаимодействия между отбором и другими факторами, но такой вариант эффекта взаимодействия в целом расценивается как подтверждение эффективности программы (Cook & Campbell, 1979).

Регрессия и уравнивание

При попытке снизить неэквивалентность групп с помощью процедуры уравнивания появляется особый вид угрозы внутренней валидности исследования. В только что рассмотренном примере эта проблема не возникла, но в прикладных исследованиях она иногда появляется. Уравнивание описывалось в главе 6 как альтернатива случайному распределению, подходящая для создания эквивалентных групп в случае управления независимой переменной. При использовании уравнивания участников разделяют на пары в зависимости от значения переменной уравнивания, а затем случайным образом распределяют по группам (см. о проце-

дуре уравнивания в главе 6). Однако, если при использовании плана с неэквивалентной контрольной группой две группы были набраны из популяций, имеющих значимые отличия друг от друга в отношении переменной уравнивания, уравнивание может вызвать проблемы. Использование процедуры уравнивания в этом случае может усилить **эффект регрессии** и даже привести к отказу от эффективной программы. Рассмотрим гипотетический пример.

Предположим, вы разрабатываете программу развития навыков чтения у плохо читающих детей. Вы объявляете о наборе добровольцев для участия в новой обучающей программе и отбираете тех, кто наиболее нуждается в помощи (т. е. детей с очень низкими средними результатами). Чтобы держать под контролем социоэкономический класс, вы набираете добровольцев, живущих в таких же районах, но в других городах, и создаете из них контрольную группу. Главная задача заключается в том, чтобы создать группы с одинаковыми начальными навыками чтения, поэтому вы решаете уравнивать группы в отношении этой переменной. Вы проводите предварительный тест на навыки чтения у добровольцев из изучаемого района и у потенциальных членов контрольной группы, а затем по полученным результатам формируете две группы с одинаковыми средними оценками. Допустим, что результаты теста варьируются от 0 до 50 баллов. Вы решаете для обеих групп отобрать детей со средней оценкой 25 баллов. Далее группа воздействия занимается по разработанной программе, а контрольная группа не подвергается воздействию. Используется обычный план с неэквивалентной группой:

Экспериментальная группа	Предварительное тестирование	Программа по чтению	Заключительное тестирование
Контрольная группа	Предварительное тестирование	—	Заключительное тестирование

Вы уверены, что ваша программа совершенна и действительно способна помочь множеству детей, и с нетерпением ждете результатов исследования. Поэтому вы шокированы, когда получаете следующие данные:

Экспериментальная группа	Предварительное тестирование = 25	Программа по чтению	Заключительное тестирование - 25
Контрольная группа	Предварительное тестирование = 25	-	Заключительное тестирование = 29

По-видимому, разработанная программа не просто не работает, но даже препятствует развитию навыков чтения — у контрольной группы обнаружено заметное улучшение! Что же произошло?

Очень вероятно, что эффект регрессии, вызванный процедурой уравнивания, перекрыл эффект экспериментального воздействия. Не забывайте, что экспериментальная группа была сформирована из детей со слабыми навыками чтения как наиболее нуждающихся в этой программе. Если провести тестирование всех детей, попадающих в эту категорию (т. е. в эту популяцию), то средняя оценка может оказаться довольно низкой, скажем 17. Однако при использовании процедуры уравнивания вы были вынуждены отобрать детей с гораздо более высокими оценками, чем в среднем у детей из популяции «плохо читающих». Возможно, что некоторые

дети из этой группы в ходе предварительного тестирования получили оценки выше, чем они должны были бы получить, ведь никакой тест не обладает абсолютной надежностью и могла возникнуть некоторая ошибка измерений. Поэтому при заключительном тестировании многие из этих детей получили более низкие оценки вследствие регрессии к среднему. Предположим, что программа была действительно эффективной и увеличивала оценку навыка чтения в среднем на 4 балла. Однако если эффект регрессии состоял в снижении оценки в среднем на 4 балла, то в результате будет обнаружено отсутствие изменений результатов от предварительного к заключительному тестированию:

$$25 + (+4) + (-4) = 25.$$

С участниками контрольной группы могло произойти обратное. Возможно, средняя оценка для их популяции была намного выше, чем 25 (например, 35) и они довольно хорошо читали (т. е. образовывали популяцию, отличающуюся от экспериментальной группы). Отбор участников, чьи оценки гораздо ниже, чем в среднем у популяции, проведенный с той целью, чтобы результаты предварительного тестирования соответствовали оценкам экспериментальной группы, мог привести к тому, что из-за эффекта регрессии оценки заключительного тестирования оказались выше. На результаты этих детей, полученные при заключительном тестировании, повлиял такой же эффект регрессии, заключающийся в изменении оценок на 4 балла. Таким образом:

$$25 + 0 + (+4) = 29.$$

На рис. 10.4 эта проблема представлена наглядно. Регрессия и вызванные программой улучшения взаимно исключили друг друга в экспериментальной группе, а на контрольную группу повлияла только регрессия, сдвинувшая оценки в сторону увеличения. В целом, программа по чтению могла быть весьма хорошей, но процедура уравнивания вызвала эффект регрессии, сделавший незаметной ее эффективность¹.

Этот вид регрессии обнаружил свой эффект при первой крупномасштабной попытке оценить эффективность программы «Рывок» (60-х гг. XX в.) — одной из центральных программ законодательной инициативы президента Линдона Джонсона «Великое общество» (Campbell & Erlebacher, 1970). Программа была начата в 1965 г. и представляла собой масштабную попытку дать возможность дошкольникам из бедных семей сделать «рывок» к школе: развить у детей различные, связанные со школой навыки, а также привлечь к этому процессу их родителей. К 1990 г. около 11 млн детей участвовали в программе, и сегодня «Рывок» считается самой успешной социальной программой, проводимой правительством (Horn, 1990). В начале 70-х гг. ее критиковали за неспособность вызывать продолжительный эффект. Критики преимущественно основывались на «Вестингхаузовском исследовании» (названном так, поскольку его позволил начать грант, выданный Вестингхаузовской корпорации обучения и Университету штата Огайо), проведенном Виктором Цицирелли и его коллегами (Cicirelli, V.G., Cooper, W.H. & Granger, R.L., 1969).

¹ Хотя проведение прикладного исследования в полевых условиях может предотвратить подобную проблему, лучше всего проверить навыки чтения у большой группы детей, уравнивая их по результатам данного теста, а затем случайным образом распределить их на экспериментальную и контрольную группы.

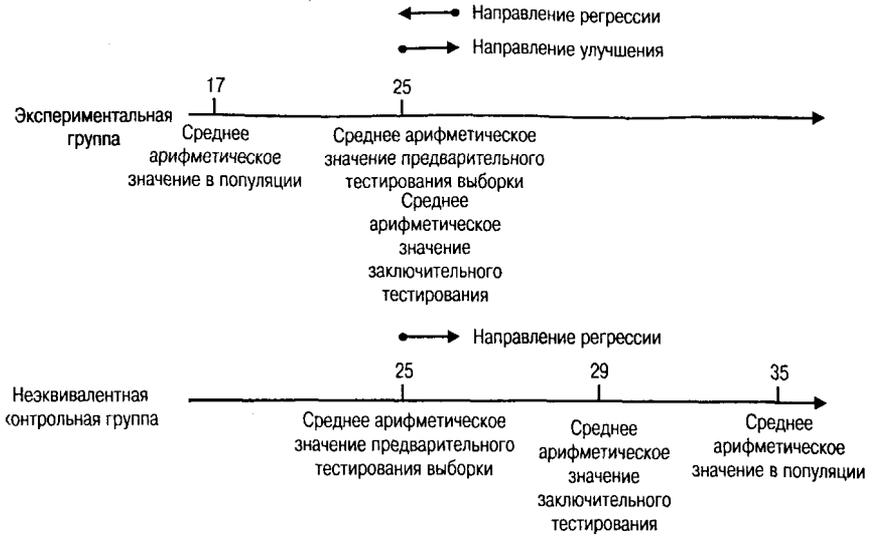


Рис. 10.4. Гипотетическое влияние эффекта регрессии при уравнивании неэквивалентных групп

В ходе Вестингхаузовского исследования был обнаружен «эффект затухания» — раннее развитие способностей у детей, произошедшее под влиянием программы «Рывок», к третьему классу стиралось. Вывод, конечно, был сделан такой, что федеральные доллары тратятся на неэффективную социальную программу. Именно на это в своем обращении к конгрессу указал президент Никсон, ссылаясь на вестингхаузовское исследование. Впоследствии в годы правления Никсона финансирование «Рывка» неоднократно ставилось под сомнение. В то же самое время главный фундамент критики, вестингхаузовское исследование, получил отрицательную оценку социологов.

Из-за того, что программа «Рывок» уже давно действовала к моменту, когда стартовал вестингхаузовский проект по ее оценке, в ходе его было невозможно случайным образом распределить детей на экспериментальную и контрольную группы. Вместо этого исследовательская группа набрала группу детей, участвующих в программе «Рывок», и уравнивала их по когнитивным показателям с группой детей, не принимающих участия в этой программе. Но для того чтобы уравнивать детей по когнитивным показателям, из участников программы «Рывок» были отобраны дети, чьи оценки заметно превышали среднее значение, и контрольные дети, чьи оценки были заметно ниже средних для их группы. А это в точности повторяет ситуацию, только что описанную в приведенном гипотетическом примере программы развития навыков чтения. Таким образом, отсутствие улучшений в третьем классе у группы из программы «Рывок» было, но крайней мере отчасти, связано с регрессией, вызванной процедурой уравнивания (Campbell & Erlebacher, 1970).

В защиту вестингхаузовской исследовательской группы следует отметить, что они отчаянно возражали политикам, которые собирались закрыть программу. Цицирелли (Cicirelli V.G., 1984) настаивал на том, что исследование «не показало, что "Рывок" был неудачным» (р. 915; курсив автора), что необходимы дополнительные

исследования, а также что должны быть предприняты «серьезные усилия по расширению и усовершенствованию этой программы» (р. 916). Не так давно (Cicirelli, V.G., 1993) Цицирелли отметил, что исходя из вестингхаузовского исследования можно порекомендовать «не останавливать программу "Рывок", а на основании открытий, сделанных в ходе применения этой программы, постараться сделать ее более эффективной» (р. 32).

Планы с неэквивалентной контрольной группой не всегда приводят к разногласиям, подобным тем, которые возникли в ходе вестингхаузовского исследования. Нижеследующий пример демонстрирует попытку поднять самооценку бейсболистов малой лиги с помощью особого курса занятий для тренеров.

Пример 24. План с неэквивалентной контрольной группой

Тренер может играть очень важную роль в жизни детей, занимающихся спортом. На детей, у которых только начали формироваться навыки, тренер может оказать весьма деструктивное влияние, особенно если единственная его цель победа в чемпионате, а любые недостатки в игре детей вызывают его порицание и насмешки. Но тренеры могут быть и прекрасными учителями, а также могут служить примером для детей в годы их становления. Смолл, Смит, Барнетт и Эверетт (Smoll, Smith, Barnett & Everett, 1993) провели исследование по оценке программы подготовки тренеров, призванной помочь им в формировании самоуважения у своих юных спортсменов. За две недели до открытия бейсбольного сезона в малой лиге несколько тренеров прошли программу, названную «Тренинг эффективности для тренеров». В программе были выделены несколько целевых способов поведения, которых необходимо придерживаться тренерам: поощрение усилий наравне с хорошей игрой, ободрение и инструкции о правильных действиях, если ребенок совершил ошибку, поощрение действий, направленных на коллективную игру команды. Программа оценивалась в ходе исследования, проведенного по плану с неэквивалентной контрольной группой. В экспериментальную группу вошли восемь тренеров из одной лиги, а в контрольную — десять тренеров из разных лиг. Почему нельзя было использовать экспериментальный план, набрать тренеров из одной лиги и случайно распределить их по группам? На это есть три причины. Во-первых, при использовании плана с неэквивалентной контрольной группой легче координировать тренинговую программу. Во-вторых, такое решение позволило, чтобы количество побед и поражений в каждой группе составило 50%. А в-третьих, это обеспечило «отсутствие утечки информации и передачи сведений о программе тренерам в контрольной группе» (р. 603). Таким образом, две группы были неэквивалентны по определению, но исследователи постарались сделать их как можно более похожими, за исключением использования тренинга. Например, они контролировали социоэкономическое положение тренеров.

Дети, занимающиеся у тренеров из экспериментальной и контрольной групп, были исследованы до и после сезона малой лиги. Мальчики прошли стандартизованный тест на самооценку, а в ходе заключительного тестирования было опрошено их мнение о действиях тренеров (исследователи, проводящие интервью, не знали, к какой из групп принадлежали мальчики, к экспериментальной или контрольной). Оказалось, что игрокам из экспериментальной группы тренеры и занятия

бейсболом нравятся больше, чем детям из контрольной группы. В отношении самооценки различий между двумя группами обнаружено не было, но когда Смолл и др. (Small et al., 1993) исследовали только мальчиков, чьи оценки, полученные при заключительном тестировании, лежали ниже медианы, они получили зависимость, представленную на рис. 10.5. Хорошо видно, что если с мальчиками с относительно низкой самооценкой занимался тренер из экспериментальной группы, то в начале сезона у них был обнаружен заметный рост этого показателя, а если тренер был из контрольной группы, то самооценка не менялась (видимое снижение не было значимым). Если допустить, что мальчики с низкой самооценкой легче других падают духом при спортивных неудачах, то такие результаты покажутся весьма впечатляющими. Если тренер правильно обращается с детьми, он может оказать положительное влияние на жизнь маленьких игроков. Это действительно хорошее исследование.

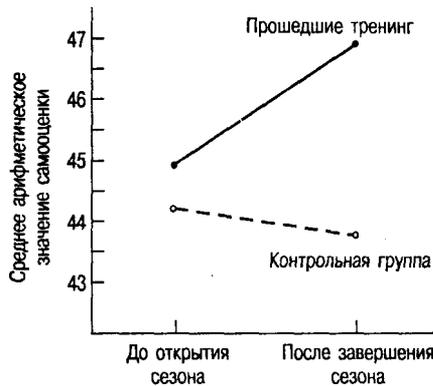


Рис. 10.5. Изменение самооценки у игроков малой лиги до и после того, как тренеры были обучены способам повышения этой характеристики (из работы Small et al., 1993)

В планах с неэквивалентной контрольной группой обычно используется предварительное тестирование, но иногда обстоятельства складываются так, что его провести невозможно. Одним из таких непредвиденных обстоятельств явилось землетрясение, произошедшее в Сан-Франциско в 1989 г. Джеймса Вуда и Ричарда Буцина из Университета штата Аризона это событие навело на мысль об исследовании ночных кошмаров — тема, давно интересовавшая обоих исследователей (Wood & Bootzin, 1990). Вместе со своими коллегами из Стэнфордского университета (расположенного неподалеку от эпицентра землетрясения) они быстро разработали исследование, чтобы выяснить, влияет ли переживание такого травмирующего события на содержание снов в целом и ночные кошмары в частности (Wood, Bootzin, Rosenhan, Nolen-Hoeksema & Jourden, 1992). Было необходимо использовать план с неэквивалентной контрольной группой. Как это обычно происходит в исследованиях с таким планом, группы были неэквивалентны с самого начала (студенты из двух различных штатов), кроме того, на одну группу было оказано воздействие одного типа (они пережили землетрясение), а на другую — другого (отсутствие непосредственного переживания землетрясения).

Экспериментальная группа состояла из студентов Стэйфорда и Государственного университета города Сан-Хосе, переживших землетрясение. Неэквивалентная контрольная группа была сформирована из студентов Университета штата Аризона. Они не столкнулись с землетрясением, но находились под влиянием многочисленных сообщений средств массовой информации. Все участники записывали содержание своих снов, которые затем анализировались по тематическому содержанию и частоте появления кошмаров. Были получены очень интересные результаты. Через три недели после начала исследования по меньшей мере один кошмар видели около 40% участников, перенесших землетрясение, и только 5% контрольных испытуемых. Из всех ночных кошмаров, пережитых первой группой, приблизительно одна четверть была о землетрясениях. Практически никто из контрольной группы не видел таких кошмаров. Кроме того, частота кошмаров значимо коррелировала с уровнем тревоги, испытанной участниками в момент землетрясения.

Прекрасно зная о проблемах интерпретации квазиэкспериментальных исследований, Вуд и др. (Wood et al., 1992) понимали трудность сравнения неэквивалентных групп. Например, не имея полученной при предварительном тестировании (до землетрясения) информации о частоте ночных кошмаров, они не смогли «исключить возможность того, что жители Калифорнии видят по ночам больше кошмаров о землетрясениях, чем жители Аризоны, даже если в ближайшее время не было зафиксировано ни одного подобного явления» (р. 222). Возможно, что кошмары о землетрясениях — это обычное дело для жителей Калифорнии. Однако, полагаясь отчасти на свой исследовательский опыт в области ночных кошмаров, авторы пришли к выводу, что частота кошмаров была особенно высока в группе из Калифорнии и, по всей видимости, была связана с их недавними травматическими переживаниями.

Планы с прерванной временной последовательностью

Если бы Вуд и его коллеги могли предвидеть землетрясение в Сан-Франциско, они бы начали собирать данные о ночных кошмарах у своих участников за несколько месяцев до этого события и продолжали бы делать это еще несколько месяцев после него. Это помогло бы им определить: а) действительно ли землетрясение увеличивает количество ночных кошмаров у людей, живущих в зоне землетрясения, и б) действительно ли частота кошмаров достигает пика сразу после землетрясения, а затем возвращается к обычному уровню. Конечно, даже талантливые психологи-исследователи не могут предсказать землетрясение, а поэтому Вуд и его сотрудники сделали все, что могли, и разработали исследование с неэквивалентной контрольной группой. Если бы они могли проводить измерения какое-то время до и после воздействующего на поведение события, то план их исследования носил бы название **плана с прерванной временной последовательностью**.

В соответствии с обозначениями, использованными Кэмпбеллом и Стэнли (Campbell & Stanely, 1963), типичное исследование с прерванной временной последовательностью можно представить следующим образом:

$$I_1 \quad H_2 \quad H_3 \quad H_4 \quad H_5 \quad V \quad H_6 \quad H_7 \quad H_8 \quad H_9 \quad H_{10},$$

где все Н означают измерения, проведенные до и после V — момента, в который была начата программа воздействия или произошло определенное событие (напри-

мер, землетрясение). В — это «прерывание» в прерванной временной последовательности. Конечно, количество измерений, проводимых до и после В, варьируются от исследования к исследованию и их число не ограничено пятью для каждого из этапов. Также не обязательно, чтобы количество измерений до и после прерывания было одинаковым. Как правило, чем больше данных собрано в различные моменты, тем лучше, и некоторые специалисты (например, Orwin, 1997) советуют проводить по крайней мере 50 измерения до момента прерывания.

Результаты

Основное преимущество плана с прерванной временной последовательностью заключается в том, что он позволяет оценивать тенденции — предсказуемые события, которые возникают с течением времени. Предположим, к примеру, что вы хотите обнаружить влияние двухмесячной антитабачной кампании на количество курящих подростков. Программа может включать определенные методы убеждения, личные беседы, показ подросткам изображений легких курильщиков и т. д. Допустим, что у вас есть хороший способ выявления курильщиков. Вы проводите измерения за месяц до и через месяц после антитабачной программы и получаете результаты, представленные на рис. 10.6.



Рис. 10.6. Частота курения непосредственно до и после гипотетической антитабачной кампании

Успешна ли программа? Снижение частоты курения от предварительного тестирования к заключительному очевидно, но оценить его при отсутствии контрольной группы довольно сложно. Но даже без контрольной группы можно понять, эффективна ли программа, если провести не одно, а несколько измерений до и после ее проведения. На рис. 10.7 представлены возможные результаты исследования по оценке антитабачной кампании, в ходе которого частота курения измерялась каждый месяц в течение года до и года после проведения программы.

На рис. 10.7, а ясно показано, как с помощью плана с прерванной временной последовательностью можно выявить тенденцию. В данном случае видно, что уменьшение числа курильщиков, которое выглядело так замечательно на рис. 10.6, — это не более чем общая тенденция к снижению числа курильщиков среди взрослых. Это говорит о важной особенности планов с прерванной временной последовательностью — они позволяют исключить (т. е. фальсифицировать, помните?) альтернативные объяснения изменений, произошедших от предварительного к заключительному тестированию.

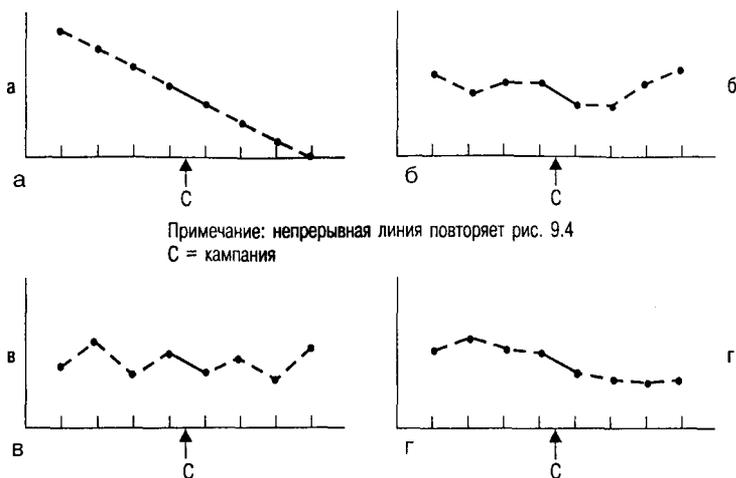


Рис. 10.7. Гипотетическая антитабачная кампания с планом с прерванной временной последовательностью - несколько возможных результатов

На рис. 10.7, б представлены два других результата, вызывающие вопросы об эффективности программы. На рис. 10.7, б количество курильщиков держалось довольно постоянным до кампании, а затем снизилось, но лишь немного. Другими словами, если антитабачная программа вообще имела какой-нибудь эффект, он был краткосрочным. На рис. 10.7, в снижение числа курильщиков после проведения программы является частью другой общей тенденции, па этот раз — периодических колебаний между более высоким и более низким уровнем курения. Идеальный результат показан на рис. 10.7, г. Здесь число курильщиков постоянно и довольно велико до начала программы, падает после того, как антитабачная программа приводится в исполнение, и сохраняется низким некоторое время после ее окончания. Обратите внимание, что показанный на рис. 10.7, г относительно неизменный базовый уровень, выявленный до начала кампании, позволил исследователям исключить эффект регрессии.

Пример 25. План с прерванной временной последовательностью

Результат, аналогичный последнему, был получен в исследовании производительности труда, проведенном в железолитейном цехе Вагнером с соавторами (Wagner, Rubin & Callahan, 1988). Исследователи изучали влияние введения стимулирующего плана, в соответствии с которым рабочие рассматривались не индивидуально, а как члены небольших групп, каждая из которых несла ответственность за определенную линию производства. Данные о производительности труда собирались в течение 10 лет, 4 года до введения стимулирующего плана и 6 лет после этого. Как видно из графика на рис. 10.8, где по оси X отложены отрезки времени, производительность до введения плана была довольно стабильной и не производила большого впечатления, но после введения плана она стала равномерно увеличиваться.

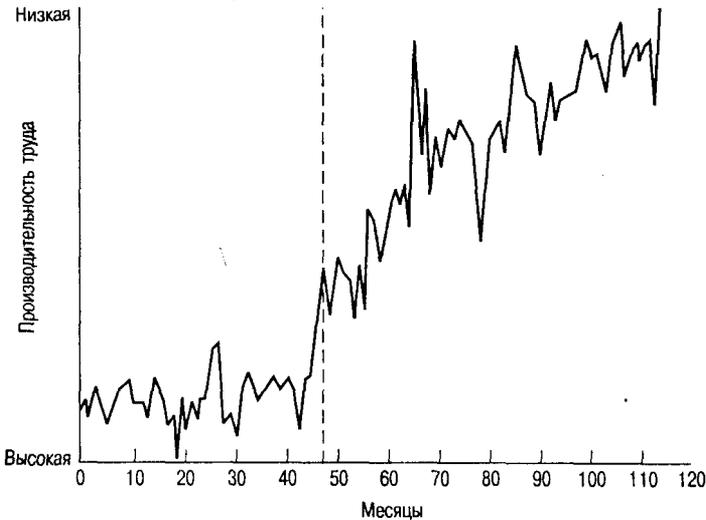


Рис. 10.8. План с прерванной временной последовательностью; влияние стимулирующего плана на производительность труда рабочих железолитейного цеха (из работы Wagner et al., 1988)

Это исследование также показывает, как ученые, использующие план с прерванной временной последовательностью, справляются с факторами, угрожающими внутренней валидностью™ их работы. Рисунок 10.8 определенно показывает, что стимулирующий план сотворил чудо, но на обнаруженные изменения могли повлиять другие факторы, в том числе история, инструментарий и отбор. Авторы данного исследования указали, что история не повлияла на изменения, так как, тщательно изучив практически все события, имевшие место в период до и после произошедших изменений, они не обнаружили повода подозревать, что некоторое необычное событие привело к скачку производительности труда. В действительности определенные события, от которых можно было ожидать негативного влияния на производительность труда (например, спад в автомобильной индустрии, повлиявший на продажи отливаемых из железа изделий), не оказали подобного влияния. Также исследователи исключили влияние инструментария: он мог представлять проблему, если бы значения среднего арифметического производительности труда рабочих с годами изменялось. Но этого не произошло. В-третьих, хотя мы обычно думаем, что отбор испытуемых является потенциальным осложнителем только в исследованиях с двумя и более неэквивалентными группами, эта проблема может проявиться и в плане с временной последовательностью, если во время действия нового плана будет обнаружена значимая текучесть кадров. В таком случае коллектив рабочих до внедрения нового плана будет отличаться от группы, работающей после этого события. В исследовании Вагнера и др. этого не произошло. В целом, планы, подобные использованному в данном исследовании, чувствительны к различным факторам, угрожающим внутренней валидности, но их можно исключить, если систематически изучать доступную информацию, как это делали Вагнер и его коллеги.

Разновидности планов с временной последовательностью

Иногда выводы, сделанные на основании исследования, проведенных по плану, с прерванной временной последовательностью, можно укрепить, проведя особое контрольное сравнение. Один подход заключается в объединении самых полезных особенностей плана с неэквивалентной контрольной группой (контрольной группы) и плана с прерванной временной последовательностью (долговременного анализа тенденций). Такой план можно представить следующим образом:

$$\begin{array}{cccccc} H_1 & H_2 & H_3 & H_4 & H_5 & \mathbf{B} & H_6 & H_7 & H_8 & H_9 & H_{10} \\ H_1 & H_2 & H_3 & H_4 & H_5 & & H_6 & H_7 & H_8 & H_9 & H_{10} \end{array}$$

Если забежать вперед и посмотреть на рис. 10.10 из вставки 10.2, можно увидеть классический пример применения данной стратегии: исследование по оценке мер, применяемых в штате Коннектикут для снижения скорости движения автомобилей, путем сравнения результатов, полученных в этом штате, с данными из других штатов.

Вторая стратегия по укреплению выводов исследования с временной последовательностью применяется, если программа вводится в разных местах в разное время. Такой план Кук и Кэмпбелл (Cook & Campbell, 1979) назвали **планом с прерванной временной последовательностью с переключающими повторениями** и обозначили следующим образом:

$$\begin{array}{cccccc} H_1 & H_2 & H_3 & \mathbf{B} & H_4 & H_5 & H_6 & H_7 & H_8 & H_9 & H_{10} \\ H_1 & H_2 & H_3 & H_4 & H_5 & H_6 & H_7 & \mathbf{B} & H_8 & H_9 & H_{10} \end{array}$$

В соответствии с этой процедурой одно и то же воздействие или программа проводится в двух различных местах в разное время. Контрольная группа не используется, но этот план обладает всеми достоинствами планов с повторениями. Если результаты, полученные в месте 2, соответствуют результатам места 1, то исследователи имеют больше оснований для уверенности в универсальном характере изучаемого явления. Такой результат был получен в неопубликованном исследовании, отчет о котором представлен у Кука и Кэмпбелла (Cook & Campbell, 1979). Это исследование было проведено в конце 40-х — начале 50-х гг. XX в., когда телевидение только начинало оказывать влияние на нашу жизнь. Нескольким сообществам штата Иллинойс было выдано разрешение на постройку телевизионных станций, но с 1951 по 1953 г. разрешения не выдавались. Это позволило исследователям изучить влияние появления телевидения на сообщества в два различных момента времени: в конце 40-х гг., до прекращения выдачи разрешений, и сразу после 1953 г., когда их снова ввели в употребление. Предполагая, что новое изобретение снизит процент читающих, исследователи изучили данные о потоке людей в библиотеках и получили подтверждение своей теории. С момента проникновения телевидения в жизнь поток людей в библиотеках снизился, причем оба изучаемые момента времени практически не отличались в этом отношении.

Третий подход к проведению исследований по плану с прерванной временной последовательностью без использования контрольной группы заключается в измерении нескольких зависимых переменных, причем предполагается, что одни из них находятся под влиянием «прерывания», а другие остаются без изменений. Имен-

но эта стратегия была использована в исследовании Штольценберга и Далессиио (Stolzenberg & D'Alessio, 1997). Ученые решили изучить влияние калифорнийского закона об окончательном приговоре к тюремному заключению — так называемой политики «три удара, и ты вылетел», — на рост преступности. Суть этой политики состоит в том, что человек, признанный виновным в трех серьезных уголовных преступлениях, автоматически наказывается тюремным заключением. Собранные по 10 крупнейшим городам штата Калифорния, Штольценберг и Далессиио исследовали два вида преступности (т. е. две зависимые переменные). Они изучили серьезные преступления, число которых предположительно должно было снизиться благодаря введению закона об окончательном приговоре, и относительно несерьезные преступления (проступки), на которые новый закон не должен был повлиять. На рис. 10.9 показаны полученные результаты, которые отражают достоинства использования временных последовательностей. Если посмотреть на кривую серьезных преступлений после введения закона, то обнаружится снижение уровня преступности, особенно в сравнении с горизонтальной линией несерьезных преступлений. Но если посмотреть на кривую серьезных преступлений в целом, то станет ясно, что снижение числа серьезных преступлений является частью тенденции, обозначившейся около 1992 г. В целом исследователи обнаружили, что закон «о трех ударах» не оказал заметного влияния ни на один из видов преступности.

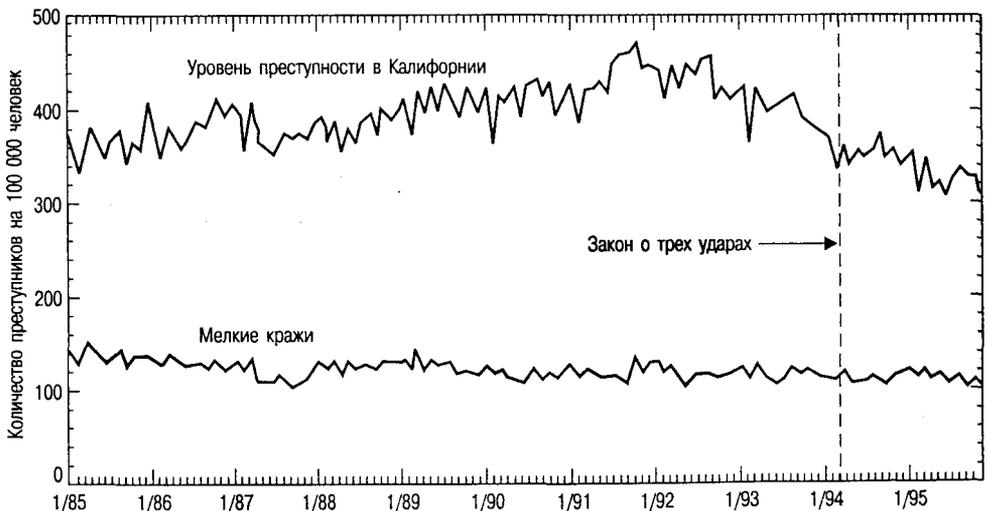


Рис. 10.9. План с прерванной временной последовательностью с использованием двух зависимых переменных; влияние закона об окончательном приговоре на уровень преступности (из работы Stolzenberg & D'Alessio, 1997)

Оценка программ

Попытка оценить изменения, вызванные законом о тюремном заключении, — это лишь один пример прикладного исследования, носящего название **оценки программы**. Эта область исследований стала разрабатываться в 60-х гг. в связи с необ-

ходимостью оценки таких социальных программ, как «Рывок». Данный вид исследований призван не просто ответить на вопрос «Эффективна ли программа X?», но имеет гораздо более широкие цели. В целом, оценка программы включает: а) процедуры для выявления необходимости в конкретной программе и определения того, кто получит пользу от ее введения, б) оценку того, проводится ли программа в соответствии с планом, и если нет, то какие изменения можно сделать, чтобы облегчить ее реализацию, в) методы для оценки результатов проведения программы и г) анализ затрат с целью определить, оправдывает ли полученная польза произведенные финансовые вложения. Рассмотрим более подробно каждый из этих пунктов. Но сначала вам необходимо прочитать вставку 10.2, посвященную документу, составленному Дональдом Кэмпбеллом в 1969 г. Этот документ считается одним из «самых важных документов в недолгой истории оценки программ».

ВСТАВКА 10.2

Классические исследования — реформы как эксперименты

Статья Дональда Кэмпбелла «Реформы как эксперименты», написанная им в 1969 г., примечательна по трем причинам. Во-первых, в ней убедительно доказывается, что нам необходимо выработать экспериментальный подход по отношению к социальным реформам. В самом начале статьи Кэмпбелл выступил в защиту применения научного образа мышления, написав, что «мы должны быть готовы к экспериментальной проверке социальных реформ, проверке, в ходе которой мы испытываем новые программы, направленные на решение особых социальных проблем, узнаем, эффективны ли эти программы, и на основании этого оставляем без изменений, копируем, модифицируем или отвергаем их» (р. 409).

Во-вторых, в статье Кэмпбелла было описано несколько исследований, ставших классическими в области оценки программ, и явившихся прототипами других исследований с использованием таких планов, как план с прерванной временной последовательностью. Вероятно, самым известным примером является его описание исследования по оценке мер, направленных на снижение скорости движения автомобилей в штате Коннектикут (Campbell & Ross, 1968). 1955 г. был годом рекордного количества смертей в автокатастрофах (324) и губернатор штата Коннектикут Абрахам Рибикофф ввел ограничения скорости, разумно предположив, что скорость движения и количество смертей в автокатастрофах коррелируют. В следующем году количество смертей упало до 284. Для Рибикоффа этого было достаточно, чтобы заявить, что на основании «сохранения 40 жизней в 1956 г. и снижения количества похоронных маршей... на 12,3% по сравнению с 1955 г. мы можем сказать, что эта программа действительно стоящая» (цитировано по Campbell, 1969, р. 412). Но действительно ли это так?

Я надеюсь, вы уже подумали о том, что здесь возможны и другие объяснения снижения уровня смертности. Например, мог сказаться фактор истории. Возможно, в 1956 г. погода была лучше, а дороги суше. Еще более вероятно влияние регрессии: 324 - это хороший пример максимальной оценки, вслед за которой обычно проходит регрессия к среднему. И действительно, Кэмпбелл указал, что на полученных в Коннектинуте результатах сказались регрессия и что «в литературе, посвященной экспериментальным социальным реформам, регрессия - это, вероятно, наиболее часто встречающийся вид самомистификации» (р. 414). Это характерно для данного вида исследований, так как меры, подобные снижению допустимого предела скоро-

сти, часто принимаются сразу после того, как происходят особо трагические события. Теория вероятности говорит нам о том, что в следующем году дела не будут обстоять так же плохо. Только ли регрессия повлияла на полученные результаты? Вероятно, нет. Проведя исследование по плану с прерванной временной последовательностью и неэквивалентной контрольной группой (близлежащие штаты, в которых не было принято мер по снижению скорости автомобилей), Кэмпбелл сделал вывод, что эти меры имели некоторый эффект, хотя и не такой значительный, как думал губернатор. Результаты этого исследования представлены на рис. 10.10.

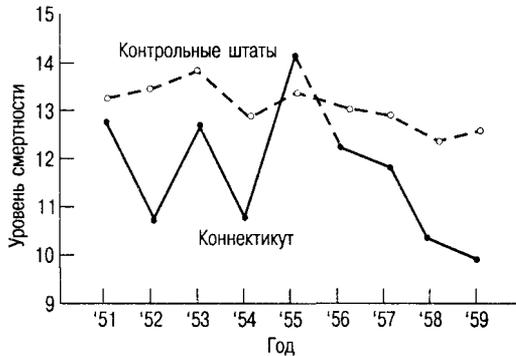


Рис. 10.10. Введение ограничения скорости в Коннектикуте - классический пример плана с прерванной временной последовательностью и неэквивалентной контрольной группой (из Campbell, 1969)

Третья причина особой важности статьи Кэмпбелла заключается в том, что она дает представление о политической реальности исследований, посвященных социальным вопросам. Политики часто предлагают программы, в эффективность которых они просто верят. Они говорят о необходимости тщательной оценки таких программ, но не слишком обрадуются, узнав о получении отрицательных результатов. Ведь поддержав программу, они оказываются заинтересованы в ее успехе и продолжении, особенно если она приносит пользу родному штату или округу этого политика. По этой причине политики и администрация, принявшие на работу для проведения конкретной программы, редко настаивают на скрупулезной оценке и склонны довольствоваться положительными результатами, даже если они получены в ходе плохо разработанных исследований. Например, губернатору Рибикоффу было достаточно информации о смертности в результате автокатастроф, полученной непосредственно до и после принятия мер по ограничению скорости.

Кэмпбелл рекомендует изменить отношение к этому вопросу и сместить акцент с важности конкретной программы на важность проблемы. Это заставит политиков и администрацию думать о программах как о попытках экспериментального решения проблем. В таком случае, прежде чем выбрать одну эффективную, будут испытаны несколько различных программ. Как выразился Кэмпбелл в заключительной части своей статьи: «Директивная администрация настолько заранее уверена в эффективности реформ, что просто не может позволить себе честную оценку... Экспериментирующая администрация подтверждает эффективность реформ исходя из важности проблемы, а не на основании уверенности в своей правоте и готова рассмотреть другие возможные решения этой проблемы, если первое оказывается неудовлетворительным. Поэтому она не боится строгого анализа проводимых реформ») 1969, р. 428; курсив автора).

Планирование программ — анализ потребности

Учреждение начинает использовать определенную программу, так как его руководители находят, что в этом есть потребность. Как определяется эта потребность? Очевидно, что мнения руководства об осмысленности применения программы не достаточно. Программа физических упражнений для пенсионеров кажется весьма подходящей, но если никто не захочет в ней участвовать, то время и деньги будут потрачены зря. Прежде чем планировать любой проект, необходимо провести анализ потребности.

Анализ потребности — это набор процедур, с помощью которых определяет-ся, существует ли достаточно большая популяция, которой будет полезна предложенная программа, может ли данная программа решить строго определенную проблему и будут ли члены этой популяции пользоваться программой. Существуют различные методы оценки необходимости, и очень важно прибегнуть по крайней мере к нескольким из них, поскольку необходимость очень легко переоценить. Один из поводов для осторожности — **переоценка по** запоминаемости, с которой вы познакомились в главе 1 при обсуждении типов познания. События, попадающие на первые страницы, привлекают наше внимание и лучше запоминаются. Поскольку они так легко закрепляются в памяти, мы склонны переоценивать частоту их появления. Достаточно было широкой огласки пары историй о детях, оставленных взявшими отпуск родителями, чтобы раздался звонок с предложением о проведении новой программы для решения этой весьма распространенной проблемы. Также потребность в новой программе может переоцениваться людьми, которым выгодно ее существование (она сохраняет им рабочие места).

Как указали Позавак и Кэри (Posavac & Carey, 1997), есть несколько способов определить потенциальную необходимость в программе.

- *Данные переписи населения.* Если предлагаемая вами программа предназначена для пожилых людей, а в данном сообществе живут лишь несколько пенсионеров, то очевидно, что программа будет мало полезной. Данные переписи населения (www.census.gov) дадут вам базовую демографическую информацию о количестве людей, соответствующих различным категориям. Кроме того, эта информация достаточно подробна, чтобы вы могли определить такие показатели, как количество матерей-одиночек в возрасте до 21 года, количество людей с различными физическими или психическими нарушениями или число пожилых людей, находящихся за чертой бедности. Кроме данных переписи местные организации могут иметь определенные архивные данные. Например, руководство корпорации, которая планирует организовать детский сад при производстве, может оценить количество сотрудников, имеющих маленьких детей. Также важную информацию о необходимости проведения программы могут дать опубликованные отчеты об исследованиях. Если исследование показало устойчивое снижение уровня интеллекта детей, живущих в нищете, то это говорит о необходимости программы по улучшению экономических условий жизни.

- *Изучение доступных услуг.* Нет необходимости начинать программу доставки пищи на дом инвалидам, если такая программа уже успешно действует. Таким образом, один из очевидных шагов в процессе анализа необходимости — это составление перечня существующих услуг, включая информацию о том, кто предоставляет услуги, какие именно услуги предоставляются, и оценку количества людей, которым предоставляются данные услуги. К примеру, вы можете обнаружить, что действующая программа доставки пищи проводится организацией, имеющей слишком большое финансирование и небольшой штат сотрудников, а это приводит к тому, что охвачено не более 10% нуждающихся. В таком случае разработка второй программы может оказаться весьма полезной.
- *Изучение потенциальных пользователей.* Третья стратегия анализа необходимости заключается в проведении опроса общественности: либо высоко репрезентативной выборки, либо целевой группы, определяемой по данным переписи населения. В ходе опроса можно выяснить мнение участников о необходимости проведения дайной программы. Но более важно, что опрос может помочь оцепить количество людей, которые действительно будут участвовать в предлагаемой программе.
- *Ключевые информаторы, фокус-группы и общие собрания.* **Ключевой информатор** — это человек, имеющий опыт и специальные знания о существующей проблеме, которая иначе не станет известна исследователю (Gilchrist & Williams, 1999). Это могут быть общественные активисты, священники, люди, входящие в правление различных социальных служб, и т. д. Иногда они могут стоять слишком близко к проблеме, чтобы оценивать ее реалистично, но в целом они представляют собой очень важный источник информации. **Фокус-группа** — небольшая (обычно 7-9 человек) группа людей, отвечающих на ряд вопросов, посвященных определенной теме, например необходимости в конкретной программе. Изучение мнения фокус-группы часто проводится после опроса общественности; предварительно это делают, чтобы уточнить вопросы, используемые при опросе общественного мнения. Члены фокус-групп обычно сходны по возрасту и социоэкономическому статусу, что делает группу в целом подходящей для выполнения определенного задания. Кроме того, полезную информацию иногда можно получить на **общих собраниях** — открытых встречах, куда приглашаются все члены сообщества, в котором хотят ввести определенную программу ключевые информаторы. Фокус-группы и общие собрания могут очень помочь в сборе информации, но исследователям необходимо с осторожностью подойти к этому занятию и не переоценить слова речистых информаторов, членов фокус-групп и лидеров общих собраний.

В последние два десятилетия во многих корпорациях распространилось мнение о том, что прибыль имеет связь со здоровьем рабочих. Поэтому компании принялись разрабатывать, реализовывать и проводить оценку программ по охране здоровья рабочих. В нижеследующем примере описывается крупномасштабная программа, которая началась с тщательного анализа потребности.

Пример 26. Оценка потребности

В компании «Дюпон» перед началом проведения программы, призванной улучшить здоровье рабочих, был проведен анализ потребности (Bertera, 1990). Программа предусматривала введение ряда изменений, касающихся 110 000 сотрудников, работающих на 100 заводах фирмы. Было ясно, что затраты на проведение такой масштабной программы будут огромны, а поэтому было важно четко продемонстрировать потребность в ней.

Проведенная компанией «Дюпон» оценка потребности включала анализ существующей информации о частоте различных видов заболеваний сотрудников фирмы, о причинах смерти рабочих, а также об отсутствии на рабочих местах и нетрудоспособности сотрудников за последние 15 лет. Был получен один непредвиденный результат: обнаружилось, что сотрудники, выполняющие самую низкооплачиваемую работу, имели самые высокие показатели по всем основным категориям заболеваний. Это открытие показало исследователям, что данная группа рабочих требует особого внимания.

Еще одно свидетельство необходимости программы по охране здоровья было получено при изучении действующих в фирме программ по улучшению здоровья сотрудников. Исследование выявило широкий набор программ, проводимых медицинским персоналом заводов фирмы «Дюпон», в том числе программы по снижению веса, антитабачные и антистрессовые программы и др. Однако в основном они заключались в проведении лекций и показе фильмов без последующей работы с сотрудниками и без систематической оценки эффективности этих программ. Также был проведен опрос сотрудников, чтобы выявить их знания об охране здоровья, оценку своей деятельности по оздоровлению, готовность изменить привычки питания и предпочтения по отношению к различным видам программ по охране здоровья.

На основании полученной информации в фирме разработали обширную серию программ, направленных на улучшение здоровья рабочих. В нее вошли тренинговые программы, далеко превосходящие рамки проведения однократных лекций, включающие создание местных комитетов по охране здоровья, изменение обстановки на рабочих местах (например, удаление автоматов по продаже сигарет). Также все рабочие прошли опрос о риске для здоровья. По результатам опроса была составлена оценка риска, включившая в себя индивидуальные планы по улучшению здоровья. Эта оценка вошла в личные дела служащих.

После того как был окончен анализ потребности и принято решение о продолжении начатой работы, были разработаны детали программы и начата ее реализация. В ходе проведения программы было предпринято еще один вид оценки.

Наблюдение за программами — предварительная оценка

Программы нередко продолжают довольно долгое время. С методологической точки зрения может казаться правильным ждать в течение года до проведения заключительной оценки эффективности программы, но что если уже в первый месяц ясно, что существуют проблемы, которые можно легко решить? Почему бы вместо ожидания окончания программы тщательно не понаблюдать за ней, пока

она находится в процессе реализации? Такое наблюдение называется **предварительной оценкой**. Согласно одному исследованию (Securest & Figueredo, 1993), она представляет собой наиболее часто встречаемый вид оценки программ.

В предварительную оценку входит несколько компонентов. Во-первых, она показывает, соответствует ли плану реализация программы. Предположим, к примеру, что местная организация, создавшая «горячую линию», решила разработать программу помощи детям, которые остаются одни после школы, так как их родители работают. Одна из очень важных задач — это сделать номер телефона горячей линии доступным и хорошо известным всем. В ходе предварительной оценки будет выявлено, помещена ли в надлежащее время реклама в газетах и были ли разосланы письма с наклейками, на которых указан телефонный номер горячей линии, как это было запланировано. Кроме того, можно выборочно обзвонить жителей и спросить, слышали ли они о новой горячей линии для детей. Нет никакого смысла проводить оценку эффективности программы, если люди даже не слышали о ней.

Вторая функция предварительной оценки — обеспечить получение данных о том, как используется данная программа. Исследователи позаимствовали название из области бухгалтерии и иногда называют такую процедуру **программным аудитом**. Так же как финансовый аудитор ищет случаи несоответствия того, как должен проводиться учет, с тем, как он в действительности проводится, аудитор программы изучает, соответствует ли реализуемая программа описанию, представленному в документации данной организации.

Заключительная часть предварительной оценки — это пилотажное исследование. Реализацию программы и начальные результаты можно оценить на небольшом примере, прежде чем проводить ее полномасштабное внедрение. Так и было сделано в ходе исследования в фирме «Дюпон». Пилотажное исследование на одном из заводов показало значимое снижение количество дней на больничном после введения программы по улучшению здоровья. Это воодушевило авторов программы и привело к ее реализации на других заводах (Bertera, 1990).

Оценка результатов — суммарная оценка

С политической точки зрения предварительная оценка менее опасна, чем **суммарная оценка**, представляющая собой общую оценку эффективности программы. Предварительная оценка направлена на усовершенствование программы и вряд ли поставит под вопрос само ее существование. Суммарная оценка вполне может это сделать. Если программа неэффективна, то зачем продолжать ее реализацию и платить ее руководителю и персоналу (понимаете, что я имел в виду, называя такую оценку «опасной»)? По мнению Сикреста и Фигереди (Sechrest & Figueredo, 1993):

Суммарная оценка и даже причина ее проведения ставят под сомнение само существование организаций. Предварительная оценка, наоборот, просто отвечает на вопрос «Как можно сделать лучше?», практически не касаясь вопроса «Приносим ли мы вообще какую-нибудь пользу?» (р. 661).

Несмотря на эти политические трудности, суммарная оценка является ядром процесса оценки и важнейшей особенностью всех программ, имеющих федеральное финансирование. Любая организация, собирающаяся потратить средства, полученные от сбора налогов, на разработку программы, обязана показать, что деньги будут потрачены с пользой.

В действительности суммарная оценка подразумевает применение методов, с которыми вы уже успели познакомиться, а в особенности подходов, связанных с квазиэкспериментальными планами. Но иногда можно провести более точный эксперимент с использованием случайного распределения, особенно при оценке программы, на которую рассчитывают больше людей, чем существует возможностей. В таком случае случайное распределение, проводящееся в форме лотереи (случайные победители становятся участниками программы, а остальные записываются в лист ожидания), — это не просто методологический идеал, а единственная справедливая процедура.

Исследователи, проводящие оценку программы, иногда сталкиваются с вопросом о том, как объяснить то, что между экспериментальной и контрольной группами не обнаружено значимых различий. На языке статистики это означает «невозможность исключить H_0 ». Такой результат может оказаться весьма трудным для интерпретации. Вероятно, различий просто не существует, но всегда есть возможность совершения ошибки 2-го рода (см. главу 4), особенно если измерительные инструменты не слишком точны или надежны. Программа может иметь небольшой, но важный эффект, но в ходе анализа обнаружить его не удастся.

Хотя невозможность обнаружить различия может вызвать проблемы, большинство исследователей считают, что такой результат (особенно если его удалось воспроизвести) дает важную информацию, необходимую для принятия решения, особенно в случае прикладных исследований. Например, человек, выступающий за продолжение проведения новой программы, обязан показать, чем эта программа лучше других, уже существующих. Если не будут показаны различия между новой программой и уже существующей, то новую программу стоит прекратить, особенно если она требует большего финансирования. Как указали Итон и Сикрест (Yeaton & Sechrest, 1986), это может стать вопросом жизни и смерти. Они привели в пример исследование, сравнивающее виды лечения рака груди, различающиеся степенью хирургического вмешательства. Процент выживших женщин, перенесших минимальные операции, значительно не отличался от процента выживших из числа подвергшихся полной мастэктомии. Такой результат является примером «невозможности исключить H ». Стоит ли проигнорировать его патом основании, что не было обнаружено значимых различий? Конечно, нет. Если нужный результат достигается при менее травмирующей процедуре лечения, зачем делать более радикальные шаги? Конечно, уверенность в правильности решения, принятого на основании отсутствия различий, непосредственно зависит от регулярности возможного воспроизведения данного открытия. Рекомендации об использовании минимального хирургического вмешательства могут быть сделаны только после многократного подтверждения отсутствия различий между умеренными и радикальными процедурами.

«Невозможность исключить H_0 » может помочь правильно оценить утверждения защитников новой программы. Человек, утверждающий, что обязательное тюремное заключение после третьего уголовного преступления заметно снизит число серьезных преступлений, обязан продемонстрировать наличие подобного эффекта. Результаты рассмотренного ранее исследования этого вопроса, проведенного по плану с прерванной временной последовательностью (Stolzenberg & D'Alessio, 1997), в ходе которого не было обнаружено никакого влияния программы «трех ударов» на уровень преступности, важны, несмотря на то что «различия не были обнаружены». Воспроизведение этих результатов привело бы к серьезным вопросам о значении закона об обязательном тюремном заключении.

Пример «невозможности исключить H_0 » говорит о том, что отсутствие различий очень важно, когда при принятии решения учитываются возможные затраты на проведение программы. Таким образом, мы подошли к последнему виду оценки программ.

Оценка стоимости: анализ затрат и эффективности

Предположим, что исследователь занимается изучением вопроса здоровья и физической подготовки рабочих и проводит сравнение двух оздоровительных программ. Одна из них включает создание условий для физкультурных занятий в рабочее время, семинары, посвященные умению справляться со стрессом, и запрет курения. Вторая программа более обширная (и более дорогостоящая). Она включает оценку состояния каждого рабочего, разработку индивидуальных программ физических занятий и оплату таких процедур, как снижение кровяного давления и уровня холестерина. Программы проверяются на двух заводах; третий завод используется в качестве контрольной группы. Таким образом, используется план с неэквивалентной контрольной группой и не с одной, а двумя экспериментальными группами. Суммарная оценка не показала различий в улучшении здоровья рабочих между двумя экспериментальными группами, но по сравнению с контрольной обе экспериментальные группы продемонстрировали улучшение. Это значит, что обе программы оздоровления эффективны, но версия Шевроле так же хороша, как и Мерседес-версия. Если две программы приводят к одинаковым результатам, но различаются по стоимости, то к чему использовать более дорогую?

Данный пример с физическими занятиями на предприятии иллюстрирует один из видов **анализа затрат и эффективности**: изучение действительной стоимости программы и сравнение ее с эффективностью. Если две программы, имеющие одну цель, одинаково эффективны, но первая стоит наполовину меньше, чем вторая, то очевидно, что лучше использовать первую. Второй вид анализа стоимости проводится на этапе планирования программы. Оценка стоимости в самом начале проекта помогает определить пригодность данной программы и является основой для позднейшего сравнения планируемых и реальных затрат.

Исследование Эрфурта с соавторами (Erfurt, Foote & Heirich, 1992) иллюстрирует анализ затрат и эффективности. Исследователи сравнили несколько различа-

ющихся по стоимости «оздоровительных программ для рабочих» на четырех автомобильных заводах. Они обнаружили, что «в сравнении с общеобразовательными лекциями (рабочая площадка В) предоставление рабочим возможности для занятий спортом (площадка В) не привело к изменению степени риска для здоровья...» (р. 5). Так как стоимость программы на площадке А равнялась 17,68 долларов на рабочего, а на площадке В — 39,28 на рабочего, создание условий для занятий спортом было либо излишним (т. е. не оправдывало затрат), либо проведено неверно.

Сравнение затрат с результатами — процесс довольно сложный и нередко требует участия экономиста. Подробное обсуждение процедур для такого сравнения не входит в цели данной главы. Это мнение разделяет моя жена. Она по профессии экономист и находит мои зачаточные знания по этому вопросу весьма забавными. Основные понятия анализа затрат рассматриваются в главе 11 (Posavac & Carey, 1997).

Примечание по качественному анализу

В главе 3 были рассмотрены различия между количественным (численным) и качественным (числа не так важны) анализом. Хотя при оценке программ применяется в основном количественный анализ, качественный анализ также активно используется, особенно в первых трех описанных видах оценки. Так, при анализе необходимости количественные данные, собранные при опросе общественности, и данные переписи могут сочетаться с информацией, полученной в ходе подробных опросов ключевых информаторов и фокус-групп. При предварительной и суммарной оценке количественные данные могут быть дополнены качественным анализом опросов сотрудников и клиентов организации и непосредственным наблюдением работы программы. В целом, в исследованиях по оценке программ редко возникает вопрос о том, какой вид анализа лучше, качественный или количественный. Чаще всего исследователи используют (и высоко ценят) оба этих вида.

Как уже отмечалось в главе 5 при обсуждении внешней валидности, психологические исследования нередко критикуются за недостаточное внимание к решению практических проблем. Эта глава, посвященная прикладным исследованиям, должна показать, что подобная критика необоснованна. В действительности проблему применения и обобщения результатов исследований понимают все психологи, даже те, которые занимаются в основном фундаментальными исследованиями. История психологии показывает, что применение сделанных открытий — это основа американской психологии, хотя бы потому, что американцы не могут не думать о практической стороне дела. Поиск возможностей для практического применения так же неотъемлем от Америки, как яблочный пирог. В следующей главе вы познакомитесь с другим направлением психологических исследований — подробным изучением отдельных индивидуумов. Как вы уже знаете, корни прикладных исследований следует искать в работах первых психологов. Эксперименты с малым количеством участников (малым *hi*) также зародились на заре развития этой науки. Но прежде, чем переходить к изучению главы II, прочитайте вставку 10.3, где обобщаются некоторые этические проблемы, с которыми сталкиваются исследователи в ходе оценки программы.

ВСТАВКА 10.3

Этика – оценочные исследования и этика

Вне зависимости от того оценивается ли программа, реализуемая на предприятии или определенная правительственная программа, исследователи, занимающиеся ее оценкой, часто сталкиваются с этическими проблемами, которые никогда не встанут перед психологами, работающими в лабораториях. Вот некоторые из них,

- *Осведомленное согласие.* Социальную помощь часто получают люди беспомощные. Когда их просят «добровольно поучаствовать» в исследовании и подписать документ об осведомленном согласии, они могут решить, что отказ «подписаться» будет означать потерю социальной поддержки. Исследователи должны совершать хорошо обдуманное действие, чтобы в такой ситуации убедить участников в ошибочности подобного мнения.
- *Сохранение конфиденциальности.* В некоторых исследованиях для сохранения конфиденциальности полученные от участников данные могут оставаться неподписанными. В других исследованиях необходимо знать, от кого именно получены данные. Например, исследователю может понадобиться вновь связаться с участниками, особенно если проводится лонгитюдное исследование. Или может потребоваться узнать, кто прошел опрос, чтобы связаться с теми, кто не прошел его. В таком случае следует разработать систему кодировки, чтобы скрыть имена участников. Иногда участники лонгитюдных исследований пользуются псевдонимами. Заполнив опросник, участник может анонимно послать его исследователю и сообщить о своей участии в отдельном письме (Sieber, 1998).
- *Воспринимаемая несправедливость.* Как было отмечено во вставке 7.2, некоторые люди могут возражать против участия в исследовании в качестве члена контрольной группы на том основании, что они не подвергаются предположительно благотворному воздействию. Хотя чаще всего члены контрольных групп подвергаются обычному воздействию, а не остаются вовсе без никакого, может возникнуть проблема контрольной группы. Например, если члены контрольной группы получают важную информацию о программе, в которой участвуют другие люди, может произойти «утечка» (см. вставку 8.2). Их возмущение по поводу того, что с другими «обращаются особенным образом», может серьезно повлиять на результаты исследования. Например, в исследовании, посвященном оценке изменений в работе угольной шахты, шахтеры из контрольной группы быстро начинают завидовать членам группы воздействия, которым, по их мнению, уделяют особое внимание и которые не должны так тяжело работать при той же зарплате (Blumberg & Pringle, 1983). Неприязнь распространилась и на самих исследователей. Рабочие из контрольной группы посчитали, что те находятся в сговоре с владельцем шахты с целью нарушить единство внутри профсоюза. Исследование в том виде, как оно было запланировано, пришлось прекратить.
- *Избегание конфликтов с заинтересованными лицами.* Заинтересованные лица - это люди, связанные с программой и материально заинтересованные в ее проведении, - заказчики, персонал и руководители программы. Исследователям, занимающимся оценкой программы, необходимо избегать возможных конфликтов с заинтересованными лицами. Это означает знать требования заинтересованных лиц и явным образом учитывать их на этапе планирования оценки. Например, исследователь может внести в контракт особые положения о том, в каком виде будет представлен отчет и каковы будут последствия получения различных видов результатов (Posavac & Carey, 1985). Говоря словами Кэмпбелла, четкие представления, составленные заранее, помогут превратить «директивную» администрацию в «экспериментирующую» (вставка 10.2).

Резюме

За пределами лаборатории

Задача прикладных исследований — пролить свет на причины возникновения и способы решения практических проблем. Но аналогично фундаментальным исследованиям их результаты важны для построения общих теорий о поведении (например, исследование, проведенное методом когнитивного интервью, расширяет наши знания о влиянии контекста на запоминание). Американские психологи всегда интересовались прикладными исследованиями, отчасти вследствие общественного давления, заставляющего их продемонстрировать, что психология — «новая», возникшая в конце XIX в. наука — способна быть полезной обществу. Прикладные исследования сталкиваются с этическими проблемами (например, обоснованное согласие) и проблемами внутренней валидности (например, неэквивалентность групп), но внешняя валидность у таких исследований всегда высока.

Квазиэкспериментальные планы

Исследования, в которых участники не могут быть случайно распределены по условиям, называются квазиэкспериментальными. Один из примеров — исследование с неэквивалентной контрольной группой. Обычно проводят сравнение изменений, произошедших от предварительного к заключительному тестированию в группе, подвергающейся некоторому воздействию, с подобными изменениями в контрольной группе, сформированной без использования случайного распределения. Эффект регрессии может затруднить интерпретацию результатов, если неэквивалентные группы уравниваются по оценкам предварительного тестирования. При использовании планов с временной последовательностью исследователи проводят несколько измерений до и после применения воздействия, которое необходимо оцепить. Исследования с прерванной временной последовательностью позволяют оценить влияние различных тенденций. Иногда к стандартному плану с временной последовательностью добавляют неэквивалентную контрольную группу, «переключающее» повторение или дополнительные зависимые переменные.

Оценка программ

Оценка программ — это один из разделов прикладной психологии. В ходе исследований по оценке программ собираются эмпирические данные об эффективности различных услуг, предоставляемых населению, или правительственных программ. Анализ необходимости выявляет, должна ли разрабатываться новая программа. Данные переписи, различных опросов и другие данные, полученные от населения, помогают оценить необходимость. Предварительная оценка показывает, проводится ли программа в соответствии с планом, а в ходе суммарной оценки оцениваются результаты исследования. Анализ затрат помогает определить, стоит ли польза, полученная от программы, затрат на ее реализацию. Исследование по оценке программ объединяет качественный и количественный подходы.

Задания для повторения

Выбор ответа

1. Прикладные исследования:
 - 1) как правило проводятся за пределами лабораторий;
 - 2) молодой подход в истории американской психологии;
 - 3) дают результаты, не подходящие для проверки общих теорий поведения;
 - 4) это синоним исследований по оценке программ.
2. Какие из перечисленных ниже проблем встречаются в прикладных исследованиях?
 - 1) невозможно получить осведомленное согласие участников;
 - 2) результаты редко оказываются полезными для построения, подтверждения или опровержения психологических теорий;
 - 3) низкая внутренняя валидность;
 - 4) результаты редко обладают внешней валидностью.
3. В плане с неэквивалентной контрольной группой:
 - 1) важно, чтобы обе группы получили одинаковые оценки предварительного тестирования;
 - 2) результаты позволяют исследователю оценить линейные тенденции;
 - 3) ключевым измерением зависимой переменной является оценка различий между результатами предварительного и заключительного тестирования;
 - 4) осложнение отбор x история можно исключить вследствие наличия контрольной группы.
4. Предположим, по причине большого количества травм у детей в штате Огайо решили ввести закон о ремнях безопасности для детей. Какое из следующих утверждений верно?
 - 1) скорее всего, количество травм возрастет в ближайшие год-два, так как тенденции не могут быстро измениться;
 - 2) для оценки эффективности данного закона лучше всего использовать план с неэквивалентной контрольной группой;
 - 3) сравнение со сходными штатами, в которых не действует подобный закон, поможет оценить возможный эффект регрессии;
 - 4) в такой ситуации качественный анализ гораздо эффективнее количественного.
5. Служба помощи семьям ввела новую программу по консультированию по вопросам задолженностей. Исследователь, занимающийся оценкой этой программы, периодически присутствует при таких обсуждениях, чтобы определить, соответствует ли происходящее плану. К какому виду относится такая процедура по оценке программы?
 - 1) анализ потребности;

- 2) суммарная оценка;
- 3) анализ затрат;
- 4) предварительная оценка.

Короткие эссе

1. Используя пример с когнитивными интервью, покажите: а) различия фундаментальных и прикладных исследований и б) что прикладные исследования связаны с разработкой теорий.
2. Опишите основные вклады Гуго Мюнстерберга в развитие прикладной психологии.
3. Опишите важнейшие особенности плана с неэквивалентной контрольной группой и объясните, почему рис. 10.3, в не позволяет заключить, что программа была успешной.
4. Первоначальная оценка программы «Рывок» показала, что успехи, которых добились участвующие в программе дети, были кратковременны: к третьему классу между детьми, принимающими участие в программе, и остальными не существовало никаких различий. Но такой результат мог быть вызван эффектом регрессии, который появился из-за того, что для формирования групп использовалась процедура уравнивания. Объясните данное явление.
5. Опишите пример, в котором оценивалась программа применения «тренинга эффективности для тренеров», и объясните, почему исследователи решили использовать не экспериментальный, а квазиэкспериментальный план.
6. Опишите важнейшие особенности плана с прерванной временной последовательностью и три разновидности основной процедуры, которые делают выводы исследования более основательными.
7. Опишите две количественные и две качественные процедуры, которые можно использовать при проведении анализа потребности.
8. Опишите различия между предварительной и суммарной оценками и объясните, почему организации предпочитают использовать первую.
9. Иногда в ходе исследований по оценке программы не удается обнаружить никаких различий. Объясните, почему такой результат не всегда плох.
10. Кратко опишите основные черты четырех видов исследований по оценке программ.
11. Кратко опишите этические проблемы, с которыми можно встретиться при проведении оценочных исследований.

Упражнения

Упражнение 10.1. Обнаружение факторов, угрожающих внутренней валидности

В квазиэкспериментальных исследованиях угроза внутренней валидности — явление весьма распространенное. Ниже приведен список факторов, угрожающих внутренней валидности, с которыми вы встретились в этой главе и главе 5. Для каждо-

го из описанных ниже гипотетических экспериментов определите, какой из этих факторов дает наиболее разумное альтернативное объяснение результатов.

Некоторые факторы, угрожающие внутренней валидности:

история	взросление
регрессия	отбор
истощение	отбор χ история

1. Директор колледжа озабочен низким процентом студентов, продолжающих обучение в колледже на втором курсе. Обычно их число было около 75%, но в начале учебного года в колледж вернулось лишь 60% бывших первокурсников. Директор ввел особую кураторскую программу, а когда на следующий год количество вернувшихся студентов составило 65%, похвалил ее эффективность.
2. Два соседних колледжа решили совместно провести оценку новой компьютеризированной системы обучения. Колледж *A* проводит программу, а колледж *B* использует уже существующие подходы к обучению. Когда исследование было проведено наполовину, колледж *B* заявил о своем банкротстве. Через год в колледже *A* повысился уровень компьютерной грамотности.
3. Двенадцать женщин, участвующих в программе домашних родов, сравнили со случайной выборкой других женщин, у которых роды проходили в больнице. Женщины из первой группы рожали средним по 6 часов, а женщины из контрольной группы — в среднем по 9 часов.
4. Разработана шестинедельная программа по управлению экзаменационной тревогой. В ней приняла участие выборка из студентов колледжа первого семестра. По окончании программы их уровень тревожности стал значительно ниже, чем он был вначале.
5. Преподаватель решил использовать новаторскую методику обучения, согласно которой в течение семестра студенты будут обучаться с наиболее подходящей для них скоростью. Курс содержит 10 разделов и студенты переходят к изучению раздела *N*, только изучив раздел *N-1*. По окончании изучения всех 10 разделов курс закончен и студент получает оценку *A*. Для 30 студентов, записавшихся в группу, получено следующее распределение оценок:

A: 16 F: 2 W: 12

Преподаватель решил, что новая методика обучения не имела успеха.

6. Компания решила ввести свободный график работ. В январе была измерена производительность труда, затем в течение 6 месяцев использовалась новая программа, а в июне снова была измерена производительность. Был обнаружен рост производительности труда.

Упражнение 10.2. Интерпретирование исследований с неэквивалентными контрольными группами

Производитель роликовых подшипников владеет двумя заводами в штате Иллинойс и хочет проверить, можно ли снизить затраты на здравоохранение, если ввести программу охраны здоровья. Один завод (*E*) выбран для проведения годичной экспериментальной программы, включающей оценку здоровья рабочих и разработку индивидуальных физкультурных программ. Второй завод (*C*) — это контрольная

группа. Уровень «отсутствий по болезни» операционально определен как количество дней в году, проведенных 100 рабочими на больничном. Он измеряется в начале и конце экспериментального года. Ниже представлены четыре набора результатов. Постройте график для каждого из наборов и определите, какой из них (возможно, ни один из них) наиболее явно говорит об эффективности программы. Для результатов, не подтверждающих эффективность программы, приведите альтернативные объяснения видимого улучшения показателей экспериментальной группы:

Результат 1.	Е: предварительное тестирование = 125	Заключительное тестирование = 100
	С: заключительное тестирование = 125	Заключительное тестирование = 140
Результат 2.	Е: предварительное тестирование = 100	Заключительное тестирование = 100
	С: заключительное тестирование = 125	Заключительное тестирование = 100
Результат 3.	Е: предварительное тестирование = 125	Заключительное тестирование = 100
	С: заключительное тестирование = 110	Заключительное тестирование = 110
Результат 4.	Е: предварительное тестирование = 125	Заключительное тестирование = 100
	С: заключительное тестирование = 130	Заключительное тестирование = 105

Упражнение 10.3. Интерпретирование исследований с временными последовательностями

Представьте себе исследование с временной последовательностью, в ходе которого оценивается влияние закона об обязательном ношении шлема на количество травм головы у хоккеистов. На следующий год после введения закона количество травм значимо снизилось по сравнению с предыдущим годом. Постройте три графика с временными последовательностями, один для каждого из следующих результатов:

1. Закон о ношении шлемов был эффективен.
2. Закон о ношении шлемов был эффективен, но его влияние было кратковременным.
3. Закон о ношении шлемов был неэффективен. Видимое снижение количества травм, скорее всего, связано с эффектом регрессии.
4. Закон о ношении шлемов был неэффективен. Видимое снижение количества травм, по-видимому, отражает общую тенденцию к снижению уровня насилия в спорте.

Упражнение 10.4. Планирование оценки необходимости

Вы — глава психологического отделения из пяти человек в гуманитарном университете искусств. Однажды декан говорит вам: «Почему бы вам не разработать программу для получения степени магистра по консультативной психологии?» В колледже уже действуют магистерские программы по администрированию в бизнесе и по лечебной физкультуре. Поскольку вы прочитали эту главу, вы отвечаете, что необходимо провести тщательный анализ потребности. Декан соглашается и даже выделяет небольшую сумму для реализации проекта. Опишите факторы, которые необходимо рассмотреть, прежде чем вводить программу получения новой степени, и методы, которые вы будете использовать при оценке потребности.