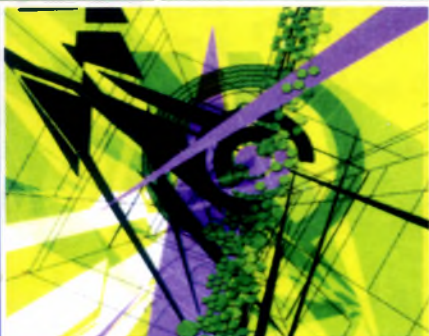


Высшее профессиональное образование

Роберт Готтсданкер

ОСНОВЫ ПСИХОЛОГИЧЕСКОГО ЭКСПЕРИМЕНТА

Учебное пособие



Психология

ВЫСШЕЕ ПРОФЕССИОНАЛЬНОЕ ОБРАЗОВАНИЕ

Р. ГОТТСДАНКЕР

ОСНОВЫ ПСИХОЛОГИЧЕСКОГО ЭКСПЕРИМЕНТА

Перевод с английского

*Учебное пособие
для студентов высших учебных заведений*

Москва
**АКАДЕМА**
2005

УДК 159.9(075.8)
ББК 88я73
Г741

Перевод с английского *Ч. А. Измайлова, В. В. Петухова*

Пособие включено в список минимального числа обязательной литературы по общепрофессиональным дисциплинам для специальностей «Психология», «Клиническая психология»

Gottsdanker, Robert M.

Experimenting in psychology.

Готтсданкер Р.

Г741 Основы психологического эксперимента: Учеб. пособие для студ. высш. учеб. заведений / Пер. с англ. Ч. А. Измайлова и В. В. Петухова; науч. ред. рус. текста Ю. Б. Гиппенрейтер. — М.: Издательский центр «Академия», 2005. — 368 с.
ISBN 5-7695-2005-1

Цель пособия профессора Калифорнийского университета Р. Готтсданкера — ознакомить с общей методологией экспериментального исследования, научить грамотно проводить психологический эксперимент. Изложены наиболее типичные и распространенные методы психологических исследований, дано описание реальных экспериментов, проведен анализ техники экспериментирования задач и проблем, которые встают перед каждым специалистом в области экспериментальной психологии. Первое издание на русском языке вышло в 1982 году.

Для студентов, аспирантов, научных работников в области психологии.

УДК 159.9(075.8)
ББК 88я73

Оригинал-макет данного издания является собственностью Издательского центра «Академия», и его воспроизведение любым способом без согласия правообладателя запрещается

- © Anne Gottsdanker, Gerald Lee Gottsdanker (правопреемники), 2003
- © Издание на русском языке. Издательский центр «Академия», 2005
- © Предисловие. Научное редактирование. Гиппенрейтер Ю. Б., 2005

ISBN 5-7695-2005-1

Предисловие научного редактора

Мы рады представить читателю эту замечательную книгу, за многие годы использования в российских вузах зарекомендовавшую себя как предельно доступное учебное пособие по экспериментальной психологии. Его цель — дать необходимые знания и сформировать необходимые навыки для проведения психологических исследований.

Как отмечает и сам автор, книга построена по методическому принципу: ее главы посвящены описанию и анализу основных типов экспериментов. Последние объединены в три крупных класса: 1) практически направленные эксперименты в реальных условиях; 2) практически направленные эксперименты в специально созданных (воспроизводящих реальные) условиях; 3) лабораторные эксперименты, проводимые для проверки теоретических гипотез.

По мере ознакомления читателя с названными классами экспериментов автор вводит все более сложные экспериментальные схемы, расширяет круг необходимых понятий, описывающих технику экспериментирования, ставит все более сложные проблемы обеспечения полноценности и достоверности экспериментальных результатов.

Книга обладает многими важными дидактическими достоинствами. Автор приводит в каждой главе ряд экспериментов, которые иллюстрируют разбираемую схему или проблему. Список экспериментов, подвергнутых в книге детальному анализу, оказывается в конечном счете довольно обширным и разносторонним. Многие важные исследования, в том числе классические, не только освещены очень подробно, но и подчас повернуты неожиданно новой методической стороной.

С первой же главы автор учит студента, казалось бы, простым, но необходимым и фундаментальным вещам: как вести протокол опыта, оформлять полученные результаты, адекватно их излагать и обсуждать, делать литературные ссылки и т.д. Статистические приложения в конце глав дают возможность учащемуся тут же применять полученные знания в собственных учебных или научных исследованиях. Вопросы, заключающие каждую главу, помогают студенту (равно как и преподавателю) проверить глубину усвоенных знаний.

Существуют, с одной стороны, книги по экспериментальной психологии, построенные по предметному, или содержательному, принципу, — с изложением фундаментальных результатов, которые получены в различных областях психологии. С другой стороны, имеются пособия по планированию эксперимента, не обязательно психологического, где рассматриваются, как правило, схематические, абстрактные примеры. Настоящая книга в свое время стала первым в учебной литературе на русском языке образцом синтетического жанра. В ней строго проанализированные методические схемы наполнены живым, конкретным содержанием, взятым из психологических исследований.

Не вызывает сомнений, что использование этого пособия в университетских курсах экспериментальной психологии и практикумах по общей психологии значительно повысит уровень профессиональной подготовки будущих психологов. Она, конечно, будет небесполезна и для уже работающих исследователей.

Конечно, отработка экспериментаторской техники не может заменить находчивости, изобретательности и глубины ума исследователя. Но тем, кому будут приходиться в голову интересные идеи, эта книга поможет облечь их в форму детально разработанных, грамотных экспериментов.

Не скроем, что освоение этой книги потребует от студентов затраты значительного труда. Мы надеемся несколько облегчить его, дополняя книгу составленным В. В. Петуховым «Словарем экспериментатора», в котором разъясняется смысл основных терминов и понятий.

Проф. Ю. Б. Гинпенрейтер

*Моей жене Джозефине и всем тем,
чье влияние на меня было так велико:
моим родителям и детям, преподава-
телям и студентам.*

Предисловие автора

Эта книга о том, как *проводить* психологические эксперименты. Она не претендует на полное описание частных предметных областей психологии, например исследований сенсорных процессов, научения или социального поведения. В ней рассказывается об экспериментальных методах, которые применяются в любой из этих областей. Иначе говоря, книга построена по методическому принципу.

Те из вас, кто уже знаком с подобными работами, сразу же заметят отличие данной книги от них. В нашей книге гораздо меньше разговоров о Науке с большой буквы. Более того, она строится, как вы убедитесь, скорее по образцу учебников фундаментальных наук. Она начинается с изложения самых простых экспериментальных задач, решить которые нетрудно, и постепенно, шаг за шагом, продвигается к более сложным, спорным и важным проблемам. Каждый следующий шаг обоснован всем предыдущим материалом.

Ориентируясь на эксперименты наиболее высокого уровня, мы никогда не принижаем значения прикладных исследований. Последние занимают почти половину книги. Автор твердо убежден, что хороший прикладной эксперимент стоит дюжины тех, результаты которых не применимы на практике. Безусловно, при современном внимании к качеству планирования и проведения прикладных исследований они приобретают самостоятельный научный статус.

Мы будем излагать методический курс экспериментальной психологии, разделяя его на самостоятельно завершенные, целостные, части. Такое изложение имеет ряд преимуществ. Можно было бы построить книгу «тематически», посвящая отдельные главы *измерению параметров изучаемого поведения, выбору испытуемых, экспериментальным схемам и способам обработки полученных данных*. Но в таком случае просто невозможно получить исчерпывающие знания о каком-либо типе эксперимента, пока не прочтешь книгу до конца. Чтобы полностью понять, как проводятся даже самые простые эксперименты, студенту пришлось бы упорно заучивать каждую отдельную часть материала и бесконечно ждать, когда они сложатся вместе. Здесь же с самого начала на доступном для студентов уровне описываются целые эксперименты. И не

нужно ждать, когда же восполнятся все пробелы. Преимущество такого изложения особенно велико при сочетании чтения курса с лабораторными занятиями. Координации с текстом можно достичь, подобрав постепенно усложняющиеся эксперименты. Полезные лабораторные эксперименты есть фактически уже в первой главе.

Центральная тема нашей книги — тесная взаимосвязь понятий валидности и *безупречного эксперимента*. Мы систематизировали различие внутренней и внешней валидности, введенное Д. Кемпбеллом и Дж. Стенли (1962), предложив более строгие определения, чем употреблявшиеся до сих пор. Хочется верить, что эти определения не просто прибавятся к весьма беспорядочному набору понятий, которых и без того вполне достаточно, но внесут в него какую-то ясность. Мы уточнили довольно противоречивые взгляды Г. Кеппелла (1968) на «идеальный эксперимент» и выделили другие типы безупречных экспериментов. Говоря кратко, эксперимент обладает одним из двух видов валидности постольку, поскольку он приближается к одному из типов безупречного эксперимента.

Большое значение при проведении экспериментов имеет обработка полученных данных и статистические выводы. Поэтому мы снабдили главы нашей книги приложениями по статистике. Статистические приложения написаны так, чтобы отвечать интересам самых разных студентов и преподавателей. Если студенты уже прошли курс статистики, приложения позволят повторить его. Таким студентам, а также тем, кто проходит статистику вместе с курсом экспериментальной психологии, они должны помочь связать эти две области воедино. Для студентов, которые еще будут учить статистику, они должны временно заменить полностью развернутый курс. А если некоторые преподаватели не хотят объединять изучение статистики и экспериментальной психологии, то приложения не обязательно использовать вообще. Книгу можно читать и без них. Кстати, не ищите приложений к главе 5. Их просто нет!

При обсуждении основных идей книги мне очень помогли все мои коллеги по университету в Санта Барбара. Но особо мне хочется поблагодарить проф. Мерилин Бревер. В наших бесконечных дискуссиях по вопросам статистики она проявляла не только блестящую эрудицию, но и завидное терпение. Я искренне признателен проф. Джеймсу Дж. Дженкинсу, старшему консультанту издательства Прентис-Холл за его преданность фундаментальным наукам. Я благодарен также своим рецензентам за их полезные предложения: проф. Чарлзу Е. Клиетту из университета Восточной Каролины, проф. Эдварду Домберу из университета в Дрю, проф. Роберту У. Ферни из Фоникс-Колледжа, проф. Джозефу Б. Хеллиджу из университета Северной Каролины, проф. Марку Собеллу из университета в Вандербилте. Я глубоко признателен м-ру

Теодору К. Джарсеку, заведующему психологической редакцией Прентис-Холла, за то доверие, которое он мне оказал.

Я приношу благодарность распорядителю литературных прав покойного сэра Роналда А. Фишера доктору Франку Уейтсу, а также издательству Лонгмен Групп Лимитед (Лондон) за разрешение перепечатать таблицу IV из их книги «Статистические таблицы для биологических, сельскохозяйственных и медицинских исследований» (6-е изд., 1974).

Роберт Готтсданкер

Глава 1

ЭКСПЕРИМЕНТЫ, КОТОРЫЕ ДУБЛИРУЮТ РЕАЛЬНЫЙ МИР¹

Мощный станок для производства хлопчатобумажной ткани — шумная машина. Представьте себе ткацкий цех в Ланкашире (Англия). Длинные ряды станков, плотно прижатых друг к другу. «Уровень шума на обычной прядильной фабрике превышает тот предел, при котором длительное и непрерывное звуковое воздействие приводит к глухоте» (Текстильная промышленность, 1974). Иначе говоря, грохот оглушающий. Как он влияет на деятельность ткача? Даже если подойти к этому вопросу чисто экономически, не стоит ли уменьшить шум? Такова исходная ситуация первого эксперимента, который мы будем обсуждать.

Наша книга — об экспериментах в психологии как сугубо практических, вроде этого — о ткацком производстве, так и о проводящихся в лабораториях «чистых» исследованиях. Когда вы прочтете ее до конца, вы сможете с пониманием читать типичные экспериментальные работы. Более того, вы будете способны судить о достоверности описанных результатов. И наконец, вы сами сможете проводить эксперименты.

В первой главе мы расскажем вам о том, как нужно проводить самые простые эксперименты. Для проверки ваших знаний будет полезно, если вы самостоятельно проведете эксперимент и дадите его оценить опытному экспериментатору. Впрочем, усвоение основного содержания главы можно проверить гораздо проще. В конце вам будут заданы вопросы по следующим общим темам.

1. Почему эта книга начинается с экспериментов, которые дублируют реальный мир.
2. Организация эксперимента и протоколирование — два основных принципа проведения исследования.
3. Словарь экспериментатора: основные термины.

Эксперимент 1:

ШУМ НА ТКАЦКОМ ПРОИЗВОДСТВЕ

Как повлияет снижение уровня шума на производство ткани? Для определения результатов этого влияния исследователи из

¹ Перевод В. В. Петухова.

Британской службы по охране труда Х. К. Вестон и С. Адамс провели эксперимент в Ланкаширском ткацком цехе. Поскольку процесс производства ткани связан с постоянным передвижением тяжелых челноков, шум ткацких станков просто неизбежен. Исследователи решили понизить уровень шума, предложив каждому рабочему носить специальные заглушки, названные наушниками Мэллока — Армстронга. Эти наушники устроены так, что позволяют воспринимать «негромкие звуки, и можно разговаривать друг с другом, не снимая их» (Weston H. C., Adams S., 1932, с. 40).

Без наушников уровень шума в цехе, измеренный с помощью аудиометра, составлял 96 дБ. Использование наушников понижает уровень шума до 87 дБ. Это гораздо большее понижение, чем может показаться на первый взгляд: по сравнению с первоначальным шумом уменьшился более чем вдвое. И все же если раньше шум был подобен грохоту проносающегося рядом товарного поезда, то теперь он стал примерно таким, будто кто-то кричит вам в ухо, — вот и все, что удалось сделать. При обычном разговоре интенсивность звука составляет около 60 дБ.

В качестве испытуемых в эксперименте приняли участие десять работниц. Эксперимент продолжался 26 недель. Он начался в понедельник, 27 мая 1929 года, и закончился в пятницу, 29 ноября того же года. В течение первой недели ткачихи работали без наушников, во вторую неделю они надевали их, в третью снова снимали и так далее — неделю без наушников и неделю с наушниками.

Каждая ткачиха обслуживает четыре станка. Ее работа заключается главным образом в том, чтобы перезарядить челнок и направлять оборвавшийся конец нити. Челнок укладывает нить внутрь угла, образованного своего рода «полом» и наклонным «потолком» из длинных чередующихся нитей. Каждое полное движение челнока с нитью по станку называется ударом (или кидкой). Передвижение челнока вперед-назад по станку совершается автоматически. Когда очередной заряд нити кончается, челнок нужно остановить и зарядить новой катушкой. Это называется *перезарядкой челнока*. Кроме того, иногда нить обрывается, и ткачихе нужно связать ее концы.

Рассмотрим эти операции более детально. В течение часа челнок мог бы без остановок совершать 11 759 полных движений по станку: это число возможных ударов за час. Эффективность работы ткачихи по перезарядке челнока и связыванию нити определяется по количеству ударов, которые она пропускает в течение часа. В основном испытуемые работали очень эффективно, пропуская лишь 1000 ударов за час, т. е. менее 10 %. Количество фактических ударов челнока измерялось с помощью счетчика. Всякий раз, когда челнок касался одного из краев станка, стрелка счетчика продвигалась на одно деление. Счетчик был установлен на одном из четырех станков, обслуживаемых каждой ткачихой.

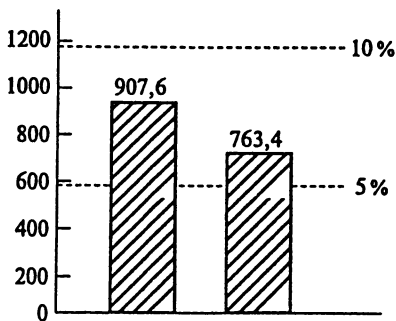


Рис. 1.1. Производительность труда испытуемой Д. при работе без наушников (13 недель) и с наушниками (другие 13 недель). Ось абсцисс — неиспользование (слева) и использование (справа) наушников; ось ординат — количество пропущенных ударов (в среднем за час)

По суммарным результатам производительность труда в течение 13 недель работы с наушниками оказалась несколько выше, чем за 13 недель без них. Однако степень влияния наушников была чрезвычайно индивидуальной, и некоторым ткачихам они почти не помогали. К тому же только половина испытуемых отмечали, что без наушников шум действительно раздражал их, остальные же пять на шум не обращали внимания.

Учитывая эти различия, описанный эксперимент можно было бы рассматривать как десять отдельных экспериментов — по одному с каждым испытуемым. Фактически они и являются ин-

дивидуальными. Каждая работница сама заботилась о производительности своего труда, и решение, носить ли ей наушники дальше, тоже было вполне самостоятельным.

Рассмотрим результаты испытуемой Д. — одной из тех пяти, кому шум мешал работать. После окончания эксперимента она так отозвалась о работе с наушниками: «Мне они очень понравились, и я хочу носить их дальше. С ними я чувствую себя лучше, хотя всегда ощущаю, что они на мне» (там же, с. 47). Другие ткачихи, которым тоже понравились наушники, дали еще более похвальные и «красочные» отзывы. Испытуемая С.: «Очень хорошо. Когда я чувствую себя раздраженной, они помогают мне, успокаивают. Хотелось бы и дальше работать с ними» (там же). Испытуемая К.: «Помощь, которую они оказывают мне, — просто огромна, и я всегда чувствую себя лучше, когда ношу их. С ними я не чувствую себя разбитой. Моя мама говорит, что у меня заметно улучшился характер. Я сама чувствую, как они укрепляют нервы». (Когда эксперимент закончился, ткачиха попросила себе пару наушников и теперь носит их в цехе постоянно.) (там же). Но хотя испытуемые С. и К. чувствовали себя лучше, производительность их труда при использовании наушников повысилась не так, как у испытуемой Д.

Общие результаты испытуемой Д. приведены на рис. 1.1. За 13 недель без использования наушников она пропустила в среднем 907,6 удара в час, а за 13 недель работы с наушниками число ударов сократилось до 763,4. На рис. 1.1 показано также, что оба эти показателя находятся между 5 и 10% от общего числа ударов

за час. Для испытуемой Д. вопрос о дальнейшем использовании наушников практически решен, поскольку оно уменьшает вероятность профессиональной глухоты. Без наушников шум мешает ткачихе работать, а с наушниками она работает эффективнее. Единственный недостаток состоит в том, что она всегда ощущает их на себе.

ОТ ПОВСЕДНЕВНОЙ ЖИЗНИ — К ЭКСПЕРИМЕНТАМ

В только что описанном эксперименте, как и в любом другом, конкретные данные должны приводить к общим выводам. Иначе говоря, эксперимент — это *обобщение*. Для данной конкретной ткачихи вывод о пользе наушников распространился на ее будущую работу, когда сам эксперимент уже завершен.

В нашей повседневной жизни мы фактически всегда экспериментируем и обобщаем, руководствуясь в своих действиях и установках опытом. Если я больше одного раза встречаю человека, который тепло улыбается мне, пожимает руку и беспокоится о моем самочувствии, я делаю вывод, что человек настроен ко мне дружелюбно. Я буду сам стремиться к общению с ним, где бы его ни встретил, — на вечеринке, на конференции или просто на улице. Конечно, мой вывод может оказаться неверным. Возможно, дружелюбный человек хотел уговорить меня застраховать свою жизнь. Но мы просто не знали бы, что делать в каждой новой ситуации, если бы не обобщали предыдущий опыт. Большинство из нас привыкли к тому, что обычно возглас «Привет!» предполагает последующую беседу, а фраза «Простите, я спешу!» — нет.

Наш первый пример эксперимента, так же как и два других, которые мы обсудим в этой главе, относится к повседневной жизни. Эксперименты такого рода фактически дублируют реальный мир. Проведение эксперимента не требовало изменения ни места, ни хода работы ткацкого цеха. И после окончания эксперимента мы видели ту же ткачиху, производящую тот же вид ткани за тем же станком, и т. д. Но для каждой ткачихи эксперимент как бы создал два будущих реальных мира: один — работы с наушниками, а другой — без них.

Полученное обобщение является минимальным. К другим людям или к другим видам работы эти результаты применить нельзя. В большинстве же экспериментов пытаются получить такие результаты, которые можно распространить достаточно широко. Экспериментальный подход полезен в той степени, в какой оправдано обобщение экспериментальных результатов. Собственно говоря, это и есть главная тема нашей книги.

И все же начинать разговор об экспериментах в психологии лучше всего с эксперимента, который дублирует реальный мир, во-первых, потому, что выводы такого эксперимента приложимы

только к каждому конкретному испытуемому. Следовательно, здесь не надо ставить вопрос о сравнении испытуемого с другими людьми, ведь на них экспериментальные выводы все равно распространяться не будут. Оставляя этот вопрос на будущее, мы сможем сконцентрировать внимание на основных принципах, характеризующих любые психологические эксперименты.

Второе основание, побуждающее нас начать именно с данного типа экспериментов, — это их внетеоретическая природа. Цели подобного эксперимента — исключительно практические. Они касаются повышения производительности труда и улучшения условий работы конкретной ткачихи. Два других эксперимента, описанных в этой главе, также имеют практическую направленность. Одному человеку хочется найти лучший метод заучивания фортепьянных пьес, а другому — выбрать более вкусный сорт томатного сока. Такие эксперименты не предназначены для решения отвлеченных теоретических вопросов. Поэтому вы сможете познакомиться с экспериментальными процедурами, не прибегая к изучению психологических теорий.

И наконец, третье основание для анализа экспериментов, дублирующих реальный мир, заключается в том, чтобы научить вас проводить эксперименты уже сейчас. Попробуйте организовать эксперимент, направленный на улучшение каких-нибудь привычных, ежедневно выполняемых вами действий. Предмет своего исследования вы можете найти, посмотрев в зеркало!

ПЛАНИРОВАНИЕ ДЕЙСТВИЙ — ПЕРВОЕ УСЛОВИЕ ЭКСПЕРИМЕНТА

Даже если в эксперименте просто дублируется повседневная жизнь испытуемых — то, что они делают, и то, что их окружает, то в него все равно должна быть внесена система. Исследование в ткацком цехе иллюстрирует первое необходимое условие любого эксперимента — *планирование действий* его проведения. Распределение времени работы с наушниками и без них определялось заранее. Чтобы иметь возможность обсуждать экспериментальные данные о повышении производительности труда, необходим предварительный план, когда именно ткачихи будут носить наушники.

Предположим, что ткачихе предоставили возможность носить наушники, когда захочется, и определять, как они ей помогают. В этом случае она могла бы надевать их в те дни, когда хотела работать особенно хорошо, скажем, если неудачно работала накануне. Если бы в эти дни производительность ее работы повышалась, то скорее всего — благодаря особой мотивации, и наушники здесь были бы ни при чем. В другой раз она может надеть их потому, что чувствует себя «разбитой», и ей хочется уменьшить

шум. Если производительность труда окажется невысокой, то это будет объясняться лишь усталостью испытуемой.

Таким образом, без специальной организации эксперимента мы можем получить только отдельные *корреляции* между производительностью труда и использованием наушников. В первом случае высокая продуктивность работы при использовании наушников была бы результатом влияния побочной переменной — мотивации ткачихи. Во втором — низкая продуктивность объяснялась бы плохим самочувствием испытуемой, т. е. влиянием еще одной переменной. В первом случае продуктивность будет коррелировать с использованием наушников положительно, во втором — отрицательно. Ясно, что подобные корреляции не дают возможности говорить о действительном влиянии основной переменной на деятельность ткачихи. Но поскольку для проведения эксперимента был заранее составлен план действий, согласно которому ткачихи должны носить наушники по чередующимся неделям, полученные результаты можно считать убедительным доказательством влияния использования наушников на производительность труда.

ОСНОВНЫЕ ТЕРМИНЫ ИЗ СЛОВАРЯ ЭКСПЕРИМЕНТАТОРА

Уже несколько раз мы употребляли слово «*переменная*», но еще не определили его содержания. Сейчас мы это сделаем. Для описания, обсуждения и планирования экспериментов любого типа нам нужен словарь общих терминов. Переменная — один из основных терминов этого словаря. Использование или неиспользование наушников — пример переменной. Среднее количество пропущенных ударов за час — еще одна переменная. Слово «*переменная*» означает только то, что в какой-то интересующей нас реальности имеются различия. Использование или неиспользование — очевидное различие: либо человек носит наушники, либо нет. Аналогично 700 и 900 отражают различие в среднем количестве пропущенных ударов за час. Такого рода реальности и называются *переменными*. Однако две из них, приведенные нами сейчас, относятся к разным видам переменных.

Независимая переменная

Использование или неиспользование наушников — независимая переменная. Это значит, что экспериментатор изменяет ее по своему плану. Наша независимая переменная включает в себя два состояния. Использование заглушек — это одно, а неиспользование — другое ее состояние. Когда разные состояния независимой переменной можно оценить количественно, пользуются также словом «*уровень*». Скажем, могло бы быть четыре вида наушников: одни понижали бы шум до 87 дБ, другие — до 78 дБ, а еще два —

до 69 и 60 дБ. Если бы проводился эксперимент по сравнению влияния этих четырех типов наушников на производительность труда, то независимой переменной был бы уровень шума при использовании заглушек. 87, 79, 69 и 60 дБ составляли бы четыре уровня. Иногда вместо слова «состояние» используются еще два термина — «условие» и «фактор».

Зависимая переменная

Среднее количество пропущенных ударов за час — зависимая переменная. В зависимости от использования или неиспользования наушников она может принимать различные значения. Каждому состоянию независимой переменной соответствует одно значение зависимой переменной. Основным результатом эксперимента является определенное отношение между независимой и зависимой переменными. В нашем примере их связь ясно представлена в виде графика, изображенного на рис. 1.1. Отметим, что независимая переменная располагается по горизонтальной оси, а зависимая — по вертикальной. Горизонтальная и вертикальная оси называются *абсциссой* и *ординатой*.

Представляя собственные экспериментальные результаты, вам следует очень точно определять независимую и зависимую переменные. В приведенном эксперименте это было несложно. Два состояния независимой переменной — это просто *использование и неиспользование* наушников. Зависимая переменная тоже была вполне определенной: *производительность, или выработка*, определялась с помощью *измерений* количества ударов. Измерения проводились по нескольку раз каждый день. Зависимая переменная всегда задана конкретным *способом представления результатов* измерения, здесь — вычислением среднего количества пропущенных ударов за час. Иной способ представления этих результатов дал бы другую зависимую переменную. Например, зависимой переменной могло бы стать количество недель, в течение каждой из которых ткачиха пропускала не более 800 ударов в час. В этом случае для испытуемой Д. значение зависимой переменной при использовании наушников было бы 5 недель, а без них — 9.

Экспериментальные гипотезы

Всякий эксперимент начинается с «догадки». В описанном нами эксперименте такая догадка касалась производительности труда ткачих и состояла в следующем. Работа с наушниками позволит использовать время более эффективно, и ткани будет производиться больше, чем без них. Когда догадка конкретизируется в реальном эксперименте, ее называют *экспериментальной гипотезой*. В данном эксперименте гипотеза заключалась в том, что за

время работы с наушниками (13 недель) количество ударов, пропускаемых ткачихами в час, будет меньше, чем за время работы без них (другие 13 недель). Для проверки экспериментальной гипотезы служило полученное отношение между независимой и зависимой переменными.

При выдвижении экспериментальной гипотезы автоматически возникает *контргипотеза* о противоположном отношении названных переменных. Здесь она состояла бы в том, что количество пропущенных ударов при работе без наушников будет меньше, чем при их использовании. В данном случае нам не нужно обсуждать третью конкурирующую гипотезу: использование наушников никак не сказывается на производительности труда. Однако позже, в главе 6, мы убедимся в значимости подобных гипотез при проведении более тщательных, научных экспериментов. Для практического решения — носить ли ткачихе наушники — достаточно рассмотреть только две *конкурирующие гипотезы*. На самом деле наушники не причиняют особых неудобств, не требуют больших расходов, и даже если прирост производительности окажется небольшим, неверное заключение едва ли будет иметь какие-либо серьезные последствия. Во всяком случае решение каждой ткачихи вполне обоснованно: носить наушники, если, судя по общим экспериментальным результатам, они помогают в работе, и не носить, если лучше работает без них.

Цель эксперимента — тщательно проверить две конкурирующие гипотезы и определить, какая из них верна, а какая нет. Основанием для такого заключения служат конкретные экспериментальные данные. Повышение производительности труда в работе конкретной ткачихи — испытуемой Д. — позволяет сделать вывод, что для нее верной была основная гипотеза, а контргипотеза оказалась ложной. Вывод достаточно очевиден. Но вы должны помнить, что время действия экспериментальных гипотез не ограничивается периодом исследования. Выдвинутая экспериментальная гипотеза относится и к будущей работе ткачихи, на годы вперед. Вывод, основанный на результатах эксперимента, всегда имеет более широкое применение.

Эксперимент 2:

СРАВНЕНИЕ ДВУХ МЕТОДОВ ЗАУЧИВАНИЯ ФОРТЕПЬЯННЫХ ПЬЕС — КАК ПРОТОКОЛИРОВАТЬ ЭКСПЕРИМЕНТ

Студент-пианист по имени Джек Моцарт хочет заучивать фортепьянные пьесы наизусть. Для исполнения каждой пьесы Джеку требуется в среднем примерно 15 мин. Обычно он старался запомнить такую пьесу сразу от начала до конца, заглядывая в ноты

только в случае крайней необходимости. Но недавно Джек задумался над тем, нет ли более эффективного метода. Он спросил себя, а что если разделить пьесу на три или четыре равные части и разучивать каждую часть по отдельности? Джек решил сравнить два метода заучивания пьес: целостный и частичный. Давайте посмотрим, как он это делал, посмотрим внимательно, следуя за ним буквально шаг за шагом.

Планирование эксперимента

Итак, Джек хочет сравнить два метода заучивания пьес и определить, какой из них ему больше подходит. Сначала он выбирает для своего эксперимента четыре пьесы, по две на каждый из методов. Вот эти пьесы.

1. Бетховен. Соната № 30.
2. Бетховен. Соната № 31.
3. Дебюсси. Сюита для фортепьяно.
4. Равель. Сонатина.

К счастью для нас, Джек — весьма пунктуальный молодой человек, и мы это скоро увидим. Например, он всегда занимается в одно и то же время. На собственном печальном опыте ему пришлось убедиться, что соседи любят ложиться спать рано вечером, а утром — поспать подольше. Он знает также, что наиболее продуктивно, не слишком уставая и не страдая от скуки, он может заниматься не более трех часов в сутки. Поэтому он занимается от двух до пяти часов дня.

Перед началом эксперимента Джек готовит все необходимое, и прежде всего лабораторный дневник для записи всей экспериментальной информации. Он выбирает сброшюрованную тетрадь в твердом переплете, и это очень разумно, ведь разрозненные листочки часто теряются. На обложке он пишет:

«Джек Амадей Моцарт (это его полное имя).¹ Эксперимент по заучиванию фортепьянных пьес: целостный метод vs частичный метод».

Затем Джек открывает тетрадь и в верхнем углу лицевой страницы пишет: «Стр. 1», а чуть ниже — текущее число. Затем — заглавие: «План эксперимента». В своем плане он отмечает последовательность разучивания пьес и метод, который он намерен применить для каждой из них. Кроме того, он фиксирует время, необходимое для исполнения пьесы. Он разделяет пьесы на пары примерно равной величины, подобного стиля, одинаковой трудности. Полностью его план представлен на рис. 1.2. Можно убедиться, что длина пьес и тип музыки подобраны для обоих методов довольно успешно. Это неслучайно: Джек хочет разучить две пары пьес, причем пьесы каждой пары должны соответствовать друг другу по трудности. И наконец Джек описывает, каким обра-

Стр 1
5 мая 1944

План эксперимента

Расписание

<u>Порядок заучивания</u>	<u>Используемый метод</u>	<u>Пьеса</u>	<u>Время с началом</u>
1	Целостный?	С. Дебюсси	13 мин 30 сек
2	Частичный?	Д. Рахмани	11 мин 30 сек
3	Частичный?	А. Бортняк (3д)	19 мин 30 сек
4	Целостный?	В. Бортняк (3д)	18 мин

Я собираюсь заучить четыре пьесы, одну за другой. Тем, где будет проходить каждый день, с понедельника до субботы, всегда от 8-х до 5-ти часов дня. Каждый в вечер, субботу — и до тех пор, пока не выучу все четыре пьесы. Трижды — два безупречно и окончательно выучу.

Я выбираю такие четыре пьесы, которые можно сыграть с места, но которые раньше не сыграл и их забывал. При заучивании пьесы целостным методом я буду начинать играть сначала всею пьесу, если не смогу выучить, что идет дальше. Но если до конца не смогу сыграть, сначала выучу начало в тысячу раз, то буду продолжать играть. Я для заучивания пьесы частичным методом сначала разобью ее на четыре примерно равные части. Затем буду заучивать первую часть, пока не смогу сыграть ее полностью без ошибок до конца пьесы. Затем точно так же вторую, третью часть и т.д. После того как каждая часть выучена отдельно, я затем попытаюсь сыграть пьесу целиком с начала до конца. Если сделаю ошибку, то мне поработать над частью или парой и начать пьесу сначала.

Рис. 1.2. План эксперимента с методами заучивания фортепьянных пьес

зом он будет применять каждый из методов заучивания. И так, описано все, что необходимо Джеку для проведения эксперимента. И ему уже не нужно держать все это в голове. Он может заглянуть в дневник через годы и получить точные сведения о планировании своего эксперимента. А если его навестит друг, Джек сможет показать ему свой дневник, и тот тоже будет знать об эксперименте буквально все. Правда, выносить тетрадь из комнаты для занятий, скажем, давать ее другу на дом, Джеку не хочется: он опасается ее потерять.

Проведение эксперимента

В первый день эксперимента, в понедельник, перед началом занятий Джек открывает дневник на странице 2 и записывает дату, время, название пьесы и применяемый метод заучивания. Эти за-

Стр 8
17 мая 1977

Начал занятия в 210 дней Раньше на слух —
звонили по телефону
Продолжался разучивать пьесу № 3
Начал с четвертой части
Один раз сыграл от начала до конца наизусть,
во второй — начинаю слушаться
Начал читать все правильно два раза подряд,
засачел в 312
В это время далсон был левый. Выше при-
ступил к делу в 318
Остал в 35 выучил всю пьесу целиком, испол-
нил ее без ошибок два раза подряд. По тому — не-
много подделал первую часть и начался за труд-
ности при переходе от второй к третьей
Затем был старого доброго Чарльз и до 516
орбалева умножением 18 и 25
Суббота: себя, насколько удалось, что по из-
вращениям упустил в начале итальянского раздела.
Но так жалко засачел пьесу сыграл, что в конце
ночи забыл о своем прошлом самоубийстве

Рис. 1.3. Эксперимент с заучиванием фортепьянных пьес: пример стра-
ницы протокола

писи называются протоколом. Затем он разучивает пьесу — как всегда, в течение трех часов или до тех пор, пока не сможет сыграть ее наизусть два раза подряд. Закончив занятие, он фиксирует время и отмечает, насколько хорошо выучил пьесу. Если же он выучил всю пьесу наизусть до окончания занятия, то оставшееся время посвящает упражнениям для тренировки пальцев. Джек не хочет начинать работать с каждой новой пьесой в разное время своих занятий, ведь это может повлиять на общее время разучивания каждой из них. Ежедневно он отмечает все то, от чего могла зависеть эффективность его работы. Пример страницы протокола из середины эксперимента приведен на рис. 1.3. Когда эксперимент полностью завершится, Джек сможет сказать, насколько случайно совпадали «неудачные дни» его занятий с применением того или другого метода.

Анализ результатов

За семь дней занятий Джек выучил наизусть все четыре пьесы. Теперь для сравнения эффективности методов ему нужно обработать полученную информацию (или данные). С этой целью он составляет «Таблицу результатов» (показана на рис. 1.4). Среднее — это просто обычное среднее арифметическое для четырех временных замеров. Лучшим оказался частичный метод — это очевидно. Он позволил Джеку заучить каждую пьесу в среднем за 235 мин, а

Стр. 9
20 мая 1977

Таблица результатов
Время, затраченное на заучивание
пес целостным и частичным методом

Целостный метод

<u>Пьеса</u>	<u>Время ее исполнения (мин)</u>	<u>Всего время заучивания (мин)</u>
C	13,5	220
D	18,0	350
Всего	31,5	570
Среднее	15,75	285
<u>Частичный метод</u>		
D	11,5	160
A	19,5	310
Всего	31,0	470
Среднее	15,50	235

Рис. 1.4. Эксперимент с заучиванием фортепьянных пьес: таблица результатов

целостный — за 285. На этом основании Джек решает применять новый метод для заучивания любых фортепьянных пьес. Эксперимент Джека фактически не потребовал никакого дополнительного времени, разве что пришлось чуть подольше потренировать пальцы.

ПРОТОКОЛИРОВАНИЕ — ВТОРОЕ УСЛОВИЕ ЭКСПЕРИМЕНТА

Давайте еще раз просмотрим всю документацию, которую составил Джек при проведении эксперимента, и отметим, из чего она состоит.

1. Экспериментальная гипотеза: применяя частичный метод, он может заучивать пьесы быстрее, чем применяя целостный.

2. Определение независимой переменной. Джек тщательно описал оба экспериментальных условия — целостный и частичный методы.

3. Изложение способа оценки заучивания пьес и способа представления результатов измерений по каждому методу для определения значений зависимой переменной.

4. Учет других обстоятельств эксперимента: разучиваемые пьесы, время проведения занятий, дни недели и т. п.

5. Расписание занятий с применением каждого из методов.

6. Приемы анализа данных для проверки экспериментальной гипотезы и изложение возможных способов интерпретации результатов.

7. Пожалуй, самое важное — первоначальное описание деятельности испытуемого, протокол на датированных страницах.

8. Анализ результатов, вывод об исследуемой гипотезе.

А теперь давайте посмотрим, как мог бы действовать Джек, не будь он столь пунктуален. Во-первых, он мог бы составить свой план в уме и выполнять его, полагаясь на свою память. Во-вторых, он мог бы не заводить для протоколов специальной тетради, а просто набрасывать данные на любых попавшихся под руку листках (да и не всегда сразу). Неаккуратность так многолика, что мы могли бы говорить о ней очень долго. Но лучше поставить точку. Преимущества тщательно протоколируемого эксперимента перед протоколируемым плохо — неизмеримы. Если бы план не содержал точного описания целостного и частичного методов, то в процессе эксперимента Джек мог просто забыть, в чем конкретно они состоят. Он мог бы, скажем, разделить некоторые пьесы только на две части. Тогда независимая переменная — метод заучивания — для разных пьес была бы неодинаковой. Если бы Джек и дальше полагался на свою память, вспоминая, сколько же времени он заучивал пьесу, то в лучшем случае он допускал бы ошибки, дающие применяемому методу определенное преимущество (или наоборот). А в самом худшем случае он мог стать жертвой предупреждения — за или против этого метода.

Эксперимент не может существовать в уме, факт его проведения оформляется в документах. После эксперимента у исследователя остается одна или несколько переплетенных тетрадей, в которых содержится буквально все, что об этом эксперименте можно сказать.

Эксперимент 3:

ВЫБОР СОРТА ТОМАТНОГО СОКА — КАК ОПИСЫВАТЬ ЭКСПЕРИМЕНТ

Приведем пример еще одного эксперимента, также проведенного «для себя». Подруга Джека Йоко Ойесс не слишком разбирается в музыке, но зато очень разборчива в отношении своего питания. С годами у нее сложился своего рода ритуал: каждое утро она вынимает из холодильника баночку томатного сока в 5,5 унции, тщательно встряхивает ее, наливает сок в стакан (а его она также держит в холодильнике) и затем медленно, со вкусом, выпивает сок, мечтательно глядя в окно, выходящее в сад с розами.

Далеко не всякий томатный сок отвечает ее высоким требованиям. До недавнего времени она считала самым вкусным сортом «Дж. Дж. Риттенхауз». Она закупает его в небольшом специализированном магазине по непомерной цене — 91 цент за упаковку (6 банок). Но однажды за обедом в гостях у Джека Йоко выпила стакан томатного сока, и ей показалось, что сок был даже лучше, чем старый добрый Джи-Джи. К ее удивлению, это был сок из универсама «Бадди'н'Билл». На банке стояла их этикетка, а стоил сок лишь 64 цента за упаковку. Правда, когда речь идет о самом вкусном соке, деньги для Йоко — не самое главное. Но «Бадди'н'Билл» мог понравиться ей случайно, скажем, из-за его температуры или же просто потому, что во время обеда у нее было особенно хорошее настроение.

В библиотеке Йоко узнала, что за последний год различные сорта томатного сока сравнивались дважды. Судя по статье из журнала «Заботы покупателей», «Дж. Дж. Риттенхауз» получил первое место (Лучшая Покупка), а «Бадди'н'Билл» был на четвертом (Товар Текущего Месяца). Однако по анкете другого журнала («Для вас, покупатели») именно «Бадди'н'Билл» получил высшую оценку — четыре очка, а «Дж. Дж. Риттенхауз» — только три.

Подруге Джека необходим эксперимент и как можно более точный. Йоко опасается, что при оценке вкуса ей будет мешать этикетка на банке. Поэтому она попросила Джека помочь ей смыть с банок этикетки и проставить просто номера. Соответствие каждого номера одному из сортов сока Джек зафиксировал в своем лабораторном дневнике. Затем Йоко составила план эксперимента и вела протокол так же, как это делал Джек. Далее мы приводим экспериментальный отчет, который она написала после проведения эксперимента.

ОЦЕНКА ДВУХ СОРТОВ ТОМАТНОГО СОКА

Йоко Ойесс

Сосалито, Калифорния

Тематика: сорта томатного сока

КРАТКОЕ СОДЕРЖАНИЕ

В течение 36 дней был проведен эксперимент, направленный на сравнение двух сортов томатного сока и выбор более вкусного из них. Испытуемым был автор. В случайном порядке было протестировано по восемнадцать банок томатного сока «Дж. Дж. Риттенхауз» и «Бадди'н'Билл». Средняя оценка вкусовых качеств для сока «Бадди'н'Билл» оказалась значительно выше. Предполагается, что причиной разногласий в предыдущих сравнениях этих сортов были разные представления дегустаторов о «естественном» вкусе томатного сока.

ОЦЕНКА ДВУХ СОРТОВ ТОМАТНОГО СОКА

За последние два года у автора сложилось мнение, что самым вкусным сортом томатного сока (из поступающих в продажу) является «Дж. Дж. Риттенхауз». Но недавно она попробовала сорт «Бадди'н'Билл», который показался ей более приятным на вкус. Разумеется, на основании единственного опыта трудно судить о качестве сорта в целом. Во-первых, даже среди банок одного и того же сорта сок может различаться по вкусу. Во-вторых, единственное знакомство с новым сортом произошло в необычное для автора время дня, к тому же сок мог отличаться по температуре. Гипотезу о предпочтении автором сорта «Бадди'н'Билл» нужно проверить экспериментально. Долгая преданность автора фирме «Дж. Дж. Риттенхауз» могла объясняться своеобразным «эффектом ореола», верой автора в «самый вкусный Джи-Джи».

Интересно заметить, что в предшествующих сравнениях указанных сортов были получены противоречивые результаты. Если Дженкинс (1975) отдает предпочтение «Дж. Дж. Риттенхаузу», то, согласно Халлу (1976), выше оценивается «Бадди'н'Билл». Поэтому вторая цель нашего эксперимента (помимо выявления собственных предпочтений автора) — несколько прояснить причины этих разногласий. Разумеется, оценка сока проводилась «вслепую», банки были без этикеток.

МЕТОДИКА

ИСПЫТУЕМЫЙ

В эксперименте участвовал один испытуемый (автор), женщина 22 лет.

ЭКСПЕРИМЕНТАЛЬНЫЙ МАТЕРИАЛ

Три упаковки (по 6 штук) стандартных (по 5,5 унции) банок томатного сока «Дж. Дж. Риттенхауз» и столько же «Бадди'н'Билл». Куплены одновременно. Перед употреблением каждая банка по крайней мере один день находилась в холодильнике.

ОБОРУДОВАНИЕ

Холодильник с температурным контролем был отрегулирован на 3 °С. Использовались одинаковые (по 8 унций) чистые стеклянные прямостенные высокие стаканы, поставленные в холодильник накануне вечером.

ПРОЦЕДУРА

Экспериментальное задание. Каждое утро, между 7.15 и 7.30, испытуемая (почистив зубы и тщательно прополоскав рот) доставала из холодильника банку томатного сока, взбалтывала ее содержимое и наливала сок в стакан, который тоже с вечера находился в холодильнике. Затем она медленно, в течение примерно 45 секунд, выпивала сок. После этого она производила оценку сока по 5-балльной шкале: (1) совсем плохой — жидкий или с «несвежим» привкусом; (2) неплохой — ничего неприятного нет, но не очень вкусный; (3) хороший — приятный вкус, но по аромату ничего нео-

бычного; (4) отличный — необычайный аромат сразу же после принятия; (5) экстраординарный — высшая степень наслаждения от томатного сока (Каллахан Х., 1976).

Кроме этого, испытуемая давала свободное описание своих вкусовых ощущений, отмечая, был ли сок сладким, острым, с привкусом и т. п.

Определение последовательности проб. Перед проведением эксперимента был обеспечен случайный порядок тестирования обоих сортов сока. На 18 бумажных бланках были написаны начальные буквы названия сорта «Дж. Дж. Риттенхауз» — Дж. Дж., на других 18 — Б. Б., «Бадди'н'Билл». Бланки клали в картонную коробку и тщательно перемешивали их в течение минуты. Затем в отсутствие испытуемой другой человек высыпал бланки из коробки. Затем он брал первый попавшийся бланк и согласно указанному на нем названию ставил жирным карандашом цифру 1 на крышке банки соответствующего сорта, затем, взяв следующий бланк, он ставил на банке указанного сорта цифру 2 и т. д., пока все 36 банок не были пронумерованы. Затем он составил список последовательности предъявления сортов, но не показывал его испытуемой до окончания эксперимента. Затем этот человек смыл со всех банок этикетки, отмачивая их около часа в прохладной воде. Затем он аккуратно счистил все остатки этикеток. В остальном банки были одинаковы, они соответствовали стандарту для банок в 5,5 унции, выпускаемых Межконтинентальным картелем по упаковке продуктов. Правда, на основаниях банок с разными сортами сока имелись разные цифры, но они были заклеены небольшими кусочками картона.

Ход опыта. Опыты проводились ежедневно в течение 36 дней, начиная с 1 июня 1977 г. Накануне в 7.15 утра испытуемая поставила во внутренний правый угол нижней полки на двери своего холодильника банку № 1 и рядом с ней — стакан. На следующее утро, снова в 7.15, она достала из холодильника банку № 1 и стакан, а на то же место поставила банку № 2. Затем она хорошо перемешала сок из банки № 1, налила его в стакан и выпила примерно за 45 секунд, наблюдая прекрасный вид за окном своей кухни. После этого она записала в лабораторный дневник свою оценку сока и краткий комментарий о его вкусовых качествах. Вечером она поставила вымытый стакан в холодильник рядом с банкой № 2.

Та же процедура выполнялась на следующий день и т. д. За все время эксперимента испытуемая не пыталась искать различия во внешнем виде банок. Время начала принятия сока испытуемой могло изменяться — от 7.15 до 7.30 утра.

РЕЗУЛЬТАТЫ

Средняя оценка сорта «Бадди'н'Билл» оказалась выше средней оценки сорта «Дж. Дж. Риттенхауз»: 3,6 и 3,2 соответственно. Анализ двух распределений оценок, изображенных на рис. 1.5, показывает, что самым значительным было различие в количестве «пятерок»: шесть у «Бадди'н'Билл» и только одна — у «Дж. Дж. Риттенхауз» (рис. 1.5).

В табл. 1 показаны результаты анализа свободных описаний вкусовых ощущений. Все качественные характеристики, названные по крайней мере 4 раза, попали в один из четырех классов: вкус свежих помидоров, сладкий, горький, острый. Из 36 проб здесь оказались 22, по 11 проб каждого сорта. Интерес-

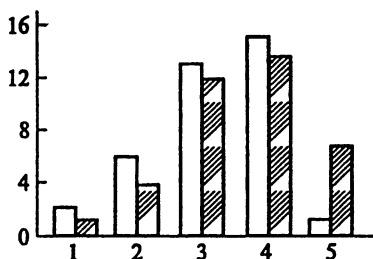


Рис. 1.5. Распределение оценок двух сортов томатного сока. Ось абсцисс — оценки вкусовых качеств сока (по возрастанию); ось ординат — количество оценок для каждого сорта сока; белые столбцы — «Дж. Дж. Риттенхауз»; заштрихованные — «Бадди'н'Билл»

но, что «Бадди'н'Билл» был назван свежим пять раз, а «Дж. Дж. Риттенхауз» — два. С другой стороны, «Дж. Дж. Риттенхауз» пять раз был назван острым, а «Бадди'н'Билл» — всего один раз. Выделяются некоторые качественные вкусовые различия двух тестируемых сортов (табл. 1).

Таблица 1

Классификация свободных описаний (названия, используемые по крайней мере 4 раза)

Сорта сока	Описание сортов				
	Как свежий помидор	Сладкий	Горький	Острый	Итого
«Дж. Дж. Риттенхауз»	2	2	2	5	11
«Бадди'н'Билл»	5	3	2	1	11
Итого	7	5	4	6	22

ОБСУЖДЕНИЕ.

Гипотеза автора о предпочтении томатного сока «Бадди'н'Билл» подтвердилась. Оснований для этого оказалось достаточно. Но нужно заметить, что предыдущие предпочтения автора объяснялись не только высокой престижностью фирмы «Дж. Дж. Риттенхауз».

Обнаруженные различия вкусовых качеств двух сортов сока позволяют высказать предположение о том, почему предшествующие исследователи давали им противоположные оценки. Эксперты, принявшие участие в эксперименте Дженкинса, жили в районе Миннеаполис-Св. Павел штата Миннесота, а те, кто участвовал в эксперименте Халла, — в Даллас-Форт Уорфе штата Техас. Именно с этим фактом и могло быть связано различие в оценках. Так, дегустаторы из Техаса имеют возможность есть свежие помидоры в течение более продолжительного времени и, таким образом, особенно ценят в соке вкус свежего помидора. Как же объяснить то, что группа экспертов из Миннесоты предпочла его «острый» вкус? Можно предположить следующее. Некоторое время в этом штате пытались использовать для продажи сока не только обычные жестяные банки, но и стеклянные. Опыт оказался неудачным. В обычных банках сок раскупали лучше. При этом покупатели могли считать его более «естественным» по вкусу. Иными словами, из-за давней привычки к соку в жестяных банках жители Миннеаполиса воспринимают привкус жести как качество самого сока, причем

оценивают его положительно. Очень может быть, что как раз то, что испытываемая называла «острым», на самом деле было «металлическим». Однако в эксперименте использовались одинаковые банки и любые вкусовые различия между двумя сортами сока можно относить только за счет самих продуктов.

Мы далеки от утверждения, что сорт «Бадди'н'Билл» будет предпочитаться другими людьми. Более того, даже в индивидуальном приложении результатов эксперимента следует соблюдать осторожность. Ведь год от года и даже в течение одного сезона качество сорта может изменяться.

Выводы. Гипотеза о более высокой оценке автором томатного сока «Бадди'н'Билл» подтвердилась. Выдвинуто предположение, что «Дж. Дж. Риттенхауз» может иметь металлический привкус, который некоторым людям нравится, но не понравился испытываемой в проведенном эксперименте.

Использованная литература

Дженкинс Дж. Выиграл Риттенхауз! // Заботы покупателей. — 1975. — № 33. — С. 182—187.

Каллахан Х. Неопубликованная шкала, использованная на ярмарке в округе Контра Коста. — Лето, 1976.

Халл П. Сравнение 16 сортов томатного сока // Для вас, покупатель. — 1976. — № 12. — С. 49—61.

Примечания

Автор хочет поблагодарить м-ра Джека Амадея Моцарта за помощь, оказанную на данной стадии эксперимента. Мы имели возможность провести эксперимент только благодаря его участию.

Заявки о перепечатке направляйте Йоко Ойесс, 139, Виейо ав., Со-салито, Калифорния 94965.

Замечание для студентов: на обороте всех рисунков нужно ясно написать карандашом название статьи (Оценка двух сортов томатного сока), номер рисунка (рис. 1), а также пометить его верх.

ЭКСПЕРИМЕНТАЛЬНЫЙ ОТЧЕТ: ЗАКЛЮЧИТЕЛЬНЫЙ ЭТАП ОФОРМЛЕНИЯ

Даже если эксперимент проводится только «для себя», как в случае с Йоко, экспериментальный отчет имеет важное значение. Он помогает экспериментатору собрать вместе всю информацию и оформить ее так, чтобы представить эксперимент в целом. Для типичного эксперимента, выводы которого должны иметь широкое приложение, отчет не просто полезен, но необходим. Пока не написан отчет, эксперимент не может считаться оформленным до конца. После публикации отчет называется *статьей*.

Как часто нам приходится слышать рассказы об удивительных экспериментальных результатах. Например, вам могут сказать, что у слонов совершенно не развита память. Но часто случается так:

вы бросаетесь искать статью, в которой этот результат опубликован, и нигде ее не находите. Все, что вы слышали об эксперименте, не более чем слух. Именно благодаря экспериментальным отчетам подобные находки получают документальное подтверждение.

Конечно, для заинтересованных читателей экспериментатор может опубликовать весь свой лабораторный дневник. Но даже если смириться с необходимостью большого количества печатного материала, эта идея не очень хороша и сама по себе. Вряд ли кто захочет разбираться во всех многочисленных деталях. Но даже если кто-то и будет все это читать, ему потребуется слишком много времени, чтобы собрать все детали воедино. Короче говоря, эксперимент не имеет никакой ценности, пока он не будет адекватно изложен. Цель экспериментального отчета — как можно эффективнее донести до предполагаемых читателей смысл эксперимента.

Адекватный способ изложения — вот к чему следует стремиться в экспериментальном отчете. Посмотрите еще раз на эксперимент Йоко, и вы убедитесь, что при его описании преследовалась именно такая цель

Во-первых, — и это главное — экспериментальный отчет должен быть написан ясно. В любых случаях автору нужно стремиться не произвести впечатление на читателя, а подробно информировать его. Во-вторых, отчет разделен на стандартные разделы. Такой способ его оформления помогает автору организовать материал. Он помогает и читателю, поскольку тот уже знает, что и где ему искать. Благодаря стандартной структуре отчета читатель может научиться извлекать из статьи необходимые сведения с наибольшей легкостью и наименьшей затратой времени. Как вы увидите из дальнейшего изложения, данная структура, разработанная для описания экспериментов, является очень эффективной. Если тщательно следовать ей, опытный исследователь сможет повторить эксперимент самостоятельно. Это называется воспроизводимостью эксперимента. Далее мы приводим разделы экспериментального отчета и их функции.

1. В «*Названии*» сообщается об изучаемой проблеме.
2. В «*Кратком содержании*» автор передает основной смысл того, что было сделано и что получено в результате.
3. Во «*Введении*» рассказывается о том, ради чего эксперимент был задуман и проведен.
4. В разделе «*Методика*» дается подробное описание того, как проводился эксперимент. Пользуясь материалами этого раздела, эксперимент можно будет повторить. Описание методики должно быть при этом достаточно сжатым, чтобы не отпугнуть и менее посвященного читателя.

Раздел «*Методика*» включает следующие подразделы.

А. Испытуемые. Отмечаются аспекты, важные для данного эксперимента: количество испытуемых, их возраст, пол, уровень слуха и т. п.

Б. *Материал*. Различные пьесы для заучивания, задачи для решения, напитки для дегустации и т. п., используемые в эксперименте, нужно описывать здесь.

В. *Оборудование*. В этом подразделе представлены всевозможные приборы, применяемые для создания экспериментальных ситуаций или для записи ответов испытуемых. Дается описание этих приборов, существенное для эксперимента.

Г. *Процедура*. Здесь приводятся последовательные этапы проведения эксперимента: какая инструкция давалась испытуемым, как именно они ее выполняли и т. д.

5. В разделе *«Результаты»* представлены анализируемые данные, обычно вместе с одной или несколькими таблицами и графиками. Читателю подробно разъясняют, на что именно ему нужно обращать внимание в таблицах и графиках и почему. В первую очередь приводится наиболее важная информация. В разделе можно дать и некоторую интерпретацию результатов, чтобы читатель смог представить себе их общий смысл, однако основной акцент нужно делать на реально полученные данные.

6. В разделе *«Обсуждение»* содержится интерпретация результатов и выводы. Здесь же предлагаются пути дальнейшего усовершенствования эксперимента, если полученные результаты недостаточно ясны. Часто автор комментирует связь своего эксперимента с предыдущим, а также с теми идеями, которые высказывались ранее в соответствующей литературе.

7. В разделе *«Сноски»* указываются неопубликованные материалы, проводящиеся исследования и малотиражные работы. Они перечисляются в порядке их упоминания в статье. В тексте статьи они цитируются по фамилии автора, затем ставится слово «Сноска» и соответствующий номер.

8. *«Цитированная литература»*. Ссылки перечисляются в алфавитном порядке в соответствии с формой, принятой Американской психологической ассоциацией (1976). В статье они приводятся по фамилии автора (иногда в скобках) и году публикации (всегда в скобках).

9. *«Примечания»* в психологических статьях обычно используются автором для выражения благодарности кому-либо и указания своего места жительства. Так, Йоко благодарит за помощь Джека (она должна быть признательна ему и за финансовую поддержку ее исследования) и указывает адрес, по которому ей могут прислать вопросы или заявки на перепечатку статьи.

КРАТКОЕ ИЗЛОЖЕНИЕ

Мы обсудили три эксперимента: использование наушников в ткацком производстве, сравнение двух методов заучивания фортепьянных пьес и оценка вкуса двух сортов томатного сока. Все это были эксперименты, дуб-

лирующие реальный мир. Данный тип эксперимента можно считать моделью общей стратегии, которая помогает нам делать выводы на базе ограниченной практики в нашей повседневной жизни. Как мы обобщаем свой жизненный опыт, так мы обобщаем и результаты эксперимента.

Мы начали с описания этого типа эксперимента, потому что он самый простой. Трудности, которые встречаются в более сложных экспериментах, например вопрос о возможности более широкого распространения экспериментальных выводов и необходимость понимания психологических теорий, нами не обсуждались. Подобные эксперименты читатель сразу же может проводить.

На примере эксперимента в ткацком цехе было показано, что неотъемлемым условием проведения эксперимента является его организация. Без заранее подготовленного плана мы получим лишь отдельные корреляции между переменными. Но если мы организуем определенный порядок применения различных состояний независимой переменной, то сможем обнаружить ее возможные влияния на зависимую переменную. Полученное в эксперименте отношение между независимой и зависимой переменными служит для проверки экспериментальной гипотезы.

Следующее необходимое условие эксперимента — протоколирование. Оно значительно повышает качество полученных данных по сравнению с экспериментом, описанным плохо. Это показано на примере эксперимента Джека Моцарта с заучиванием фортепьянных пьес. Если же экспериментатор не ведет протокола, он вынужден полагаться на свою память, которая никогда не бывает безупречной, а иногда может стать источником нежелательных предубеждений. Шаг за шагом мы описали все этапы работы Джека по составлению документации эксперимента, а затем суммировали ее отдельные части.

Заключительной стадией описания эксперимента является экспериментальный отчет. Его иллюстрация дана на примере эксперимента с томатным соком. Без отчета (который после публикации называется статьей) экспериментальные результаты остаются ничем не подтвержденными слухами. Описана типичная структура экспериментального отчета. Его основная цель — наиболее адекватное изложение эксперимента для читателя.

Контрольные вопросы и задания

1. Зачем проводятся эксперименты, дублирующие реальный мир?
2. Почему эти эксперименты обсуждаются в книге первыми?
3. Как осуществляется организация эксперимента?
4. Как протоколировать эксперимент?
5. Приведите примеры, раскрывающие содержание следующих терминов:
 - а) *независимая переменная*;
 - б) *зависимая переменная*;
 - в) *экспериментальные гипотезы*.
6. Покажите на примере, как эти понятия связаны между собой.
7. В чем состоят этапы протоколирования эксперимента?
8. Что включается в основные разделы экспериментального отчета?

СТАТИСТИЧЕСКОЕ ПРИЛОЖЕНИЕ: ВЫЧИСЛЕНИЕ И ПРЕДСТАВЛЕНИЕ СРЕДНИХ ЗНАЧЕНИЙ ВРЕМЕНИ РЕАКЦИИ¹

В эксперименте по измерению времени реакции испытуемого использовалось два сигнала: вспышка света (условие А) и звучание тона (условие Б). Каждое из этих условий предъявлялось в случайном порядке по 17 раз. Следовательно, всего было 34 пробы. Время реакции (ВР) в каждой пробе измерялось с точностью до 1/1000 секунды (или миллисекунды, сокращенно мс).

Условие А (свет)			
Проба	ВР	Проба	ВР
1	223	10	191
2	184	11	197
3	209	12	188
4	183	13	174
5	180	14	176
6	168	15	155
7	215	16	115
8	172	17	163
9	200		
Условие Б (тон)			
Проба	ВР	Проба	ВР
1	181	10	156
2	194	11	178
3	173	12	160
4	153	13	164
5	168	14	169
6	176	15	155
7	163	16	122
8	152	17	144
9	155		

Вычисление средних значений

Результаты эксперимента приводятся выше в таблице. Представленные числа относятся к первой, второй и т.д. пробам обозначенного условия.

¹ Перевод Ч.А.Измайлова, Г.А.Парамей.

Для определения среднего ВР для условия А вы просто складываете все 17 ВР и делите на 17. Сумма равна 3143. Среднее определяется как 3143, деленное на 17, что равно 185 мс (округленное 184,88235). Эти вычисления можно представить формулой

$$M_X = \frac{\sum X}{N}. \quad (1.1)$$

Значения в отдельных пробах обозначаются как X_1, X_2, X_3 и т. д. до X_{17} . Σ (заглавная греческая буква сигма) означает сложение всех различных X — от X_1 до X_N . N — число проб или X , здесь равное 17. Формула читается следующим образом: «Среднее X равно сумме всех X , деленной на N ». Для условия А

$$M_A = \frac{3143}{17} = 185.$$

Представление средних значений

Независимой переменной в эксперименте был тип сигнала (свет или тон). Зависимой переменной было среднее время реакции. Гистограмма показывает отношение между независимой и зависимой переменными. Высота гистограммы представляет среднее время реакции. (Замечание: Вам необходимо самим вычислить среднее для условия Б.)

Задача: Вычислите среднее время реакции для условия Б.

Ответ: Округленно 162 мс.

Глава 2

ОСНОВЫ ПЛАНИРОВАНИЯ ЭКСПЕРИМЕНТА¹

Если вы хотите экспериментально проверить, помогают ли музыкальные радиопередачи заучивать слова французского языка, вы легко сможете сделать это, повторив один из экспериментов, описанных в предыдущей главе. Скорее всего, вы спланируете свой эксперимент по образцу Джека Моцарта. Вы заранее определите оба условия независимой переменной, будете заниматься в одно и то же время дня и каждый этап эксперимента фиксировать документально. Вместо четырех фортепьянных пьес вы могли бы заучивать четыре списка слов следующим образом: слушая радио, без радио, с радио. Иначе говоря, вы сможете применить ту же *экспериментальную схему*, что и Джек.

Вполне возможно, что вам будут понятны некоторые основания ваших собственных действий. Но что-то наверняка останется неясным, и прежде всего — последовательность условий независимой переменной, т. е. сама экспериментальная схема. Вашей вины в этом нет, ведь экспериментальных схем вы еще не проходили. В настоящей главе этот недостаток будет устранен. Конечно, можно провести эксперимент и путем простого подражания образцу, но гораздо лучше понимать то, что вы делаете. Двух идентичных экспериментов не бывает, и слепое копирование экспериментальной схемы часто приводит к затруднениям. Например, Йоко могла бы применить в своем эксперименте регулярное чередование двух условий (сортов сока), как это делали в эксперименте с ткачихами (использование или неиспользование наушников). Но тогда она наверняка знала бы название тестируемого сока, а как раз этого она и старалась избежать, используя случайную последовательность. К тому же если вы не будете знать оснований различных планов и схем, вам трудно будет оценить качество экспериментов, о которых вы будете читать. А, как вы помните, научить вас этому — одна из главных целей нашей книги.

В данной главе мы сравним те планы, по которым строились эксперименты в главе 1, с менее удачными планами проведения тех же экспериментов. Образцом для их сравнения будет «безупречный» эксперимент (который практически неосуществим). Анализ такого рода позволит рассмотреть те основные идеи, которы-

¹ Перевод В. В. Петухова.

ми мы руководствуемся при создании и оценке экспериментов. В процессе этого анализа мы введем в наш словарь несколько новых терминов. В итоге мы определим, что безусловно и что нет в тех трех экспериментальных схемах, которые использовались в главе 1. А эти схемы представляют три способа упорядочивания, или три вида последовательностей предъявления, различных условий независимой переменной, применяемые в эксперименте с одним испытуемым.

После изучения материала этой главы вы сможете компетентно и не подражая чужому эксперименту спланировать свой собственный. В конце главы вам будут заданы вопросы по следующим темам.

1. Степень приближения реального эксперимента к безупречному.
2. Факторы, нарушающие внутреннюю валидность эксперимента.
3. Систематические и несистематические источники нарушения внутренней валидности.
4. Методы повышения внутренней валидности, способы первичного контроля и экспериментальные схемы.
5. Некоторые новые термины из словаря экспериментатора.

ПРОСТО ПЛАНЫ И ПЛАНЫ БОЛЕЕ УДАЧНЫЕ

Несомненно, первое условие проведения эксперимента — это его организация, наличие плана. Но не всякий план можно считать удачным. Предположим, что эксперименты, описанные в главе 1, проводились иначе, по следующим планам.

1. Пусть в первом эксперименте ткачиха вначале носила наушники 13 недель, а затем 13 недель работала без них.

2. Положим, Йоко решила использовать в своем эксперименте только по две банки каждого сорта сока, и весь эксперимент занял четыре дня вместо 36.

3. Джек решил применить частичный метод заучивания к первым двум пьесам, а целостный — к двум следующим.

4. Или же, сохранив ту же последовательность методов, Джек выбрал для эксперимента короткие вальсы, а не более длинные пьесы, которые он разучивал обычно.

Мы довольно ясно чувствуем, что по сравнению с ранее описанными экспериментами все эти планы — неудачны. А если бы у нас был образец для сравнения, то мы совершенно точно могли бы сказать, почему именно первоначальные планы были лучше. Безупречный эксперимент служит таким образцом. В следующем разделе мы подробно обсудим его, а затем посмотрим, как он применяется для оценки наших экспериментов.

БЕЗУПРЕЧНЫЙ ЭКСПЕРИМЕНТ

Теперь у нас есть примеры удачно и неудачно спланированных экспериментов. А можно ли и дальше усовершенствовать хорошо спланированный эксперимент? И можно ли сделать эксперимент абсолютно безупречным? Ответ: любой эксперимент можно улучшать до бесконечности, или (что то же самое) безупречный эксперимент провести нельзя. Реальные эксперименты *совершенствуются* по мере приближения к безупречному.

Идеальный эксперимент

Лучше всего определить безупречность с помощью понятия идеального эксперимента (Kerppel G., 1973, с. 23). В нем допускается изменение только независимой переменной (и, разумеется, зависимой переменной, принимающей различные значения при разных условиях). Все прочее остается неизменным, и поэтому на зависимую переменную влияет только независимая. В наших трех хорошо спланированных экспериментах это, конечно, не так. Ткачихи носили наушники и работали без них в разное время — по четным или нечетным неделям. Пьесы, которые заучивал Джек с помощью целостного и частичного методов, тоже были разными. И Йоко никогда не выпивала сок обоих сортов в один день. В каждом случае помимо независимой переменной изменялось что-то еще. В последующих главах мы расскажем об экспериментах другого типа, в которых для каждого из условий независимой переменной привлекаются разные испытуемые, что позволяет устранить временные изменения (вроде четных и нечетных недель) и различия в заданиях (заучиваемые пьесы). Но и они не отвечают всем требованиям идеального эксперимента, ведь испытуемые тоже будут разными. Как вы скоро убедитесь, идеальный эксперимент невозможен. Однако сама идея полезна, ею мы руководствуемся при совершенствовании реальных экспериментов.

В идеальном (неосуществимом) эксперименте ткачиха должна была бы работать с наушниками и без них в одно и то же время! Джек Моцарт одновременно заучивал бы одну и ту же пьесу целостным и частичным методами! В обоих этих случаях разница в значениях зависимой переменной была бы обусловлена *только* независимой переменной, различием ее условий. Иначе говоря, все побочные обстоятельства, все другие *потенциальные* переменные оставались бы на одном и том же неизменном уровне.

Бесконечный эксперимент

Бедная Йоко! В ее случае даже идеальный эксперимент не будет безупречным. Недаром она опасается, что в разных банках то-

матный сок одного и того же сорта различается по качеству. Даже если бы она провела идеальный эксперимент, ухитрившись одновременно из одного и того же стакана выпивать сок двух разных сортов, ее оценки все равно относились бы только к частным примерам каждого сорта. И тем не менее Йоко могла бы устранить влияния изменчивости качества сока в разных банках, совершив иной невозможный подвиг. «Все», что ей нужно, — это не прекращать свой эксперимент после 36 дней и продолжать его до бесконечности. Тогда она смогла бы усреднить не только изменчивость каждого из сортов сока, но и возможные колебания в собственных оценках его вкусовых качеств. Это и есть бесконечный эксперимент. Нетрудно видеть, что он не только невозможен, но и бессмыслен. Ведь общий смысл эксперимента заключается в том, чтобы на базе *ограниченного* количества данных делать выводы, имеющие более широкое приложение. Однако бесконечный эксперимент, подобно идеальному, также служит нам руководящей идеей.

Собственно говоря, Джеку Моцарту и авторам исследования в ткацком цехе тоже можно было бы предложить провести бесконечный эксперимент вместо идеального. Ведь даже если в идеальном эксперименте Джек обнаружит, что для данной конкретной пьесы частичный метод более эффективен, останется вопрос, сохраняются ли преимущества этого метода при разучивании других пьес. Те же сомнения вызывает и первый эксперимент: а что если ткачиха лучше работала с наушниками только во время его проведения? Однако их (и вас) нужно предупредить, что бесконечный эксперимент тоже имеет недостатки. Сам факт предъявления испытуемым одного из экспериментальных условий может сказываться (в период исследования) на их работе при другом условии. Возможно, частичный метод был более эффективным во время эксперимента только за счет контраста с целостным методом. А после эксперимента будет применяться один-единственный метод, и фактор контраста исчезнет. Все это доказывает, что полностью безупречными не являются ни идеальный, ни бесконечный эксперименты. К счастью, они имеют не только разные недостатки, но и разные преимущества и могут служить для оценки реальных экспериментов, весьма далеких от безупречного.

Эксперимент полного соответствия

Ни идеальный, ни бесконечный эксперименты не позволяют устранить недостатки неудачного варианта исследования Джека Моцарта — заучивания вальсов вместо сонат. В лучшем случае Джек мог бы провести блестящий эксперимент на вальсах — что, однако, не сделает их сонатами!

Чтобы совершенно исключить недостатки такого рода, нужен эксперимент *полного соответствия*. Этот эксперимент также является бессмысленным, хотя практически он осуществим. В своем исследовании Джек должен был бы заучивать *те же самые* пьесы, которые он будет разучивать и после него. Никакой пользы от такого эксперимента нет, как и от бесконечного. Но зато уж никто не сможет указать Джеку на несоответствие пьес, которые он разучивал в своем эксперименте.

Все три вида безупречного (почти) эксперимента нереальны. Идеальный эксперимент является невозможным, эксперимент полного соответствия — бессмысленным, а бесконечный — тем и другим вместе. Полезны они как эксперименты «мысленные». Они подсказывают нам, что нужно делать для создания эффективного эксперимента. Идеальный и бесконечный эксперименты показывают, как избежать посторонних влияний и тем самым добиться большей уверенности в том, что экспериментальные результаты действительно отражают связь независимой и зависимой переменных. Эксперимент полного соответствия напоминает о необходимости контроля других важных переменных эксперимента, которые мы сохраняем неизменными.

ОБОБЩЕНИЕ, РЕПРЕЗЕНТАТИВНОСТЬ И ВАЛИДНОСТЬ

Как мы установили в главе 1, цель любого экспериментального исследования — сделать так, чтобы выводы, основанные на ограниченном количестве данных, оставались достоверными *за пределами* эксперимента. Это называется *обобщением*. Выполненный нами анализ безупречного эксперимента показывает, что достоверность экспериментальных выводов определяется по крайней мере двумя требованиями. От них же зависит и правомерность возможных обобщений. Первое требование заключается в том, чтобы найденное в эксперименте отношение между независимой и зависимой переменными было свободным от влияния других переменных. Второе требование — чтобы постоянный уровень дополнительной переменной, задействованной в эксперименте, соответствовал ее уровню в более широкой области практики.

Репрезентативность

Мы уже знаем, что безупречный эксперимент невозможен, но он дает нам руководящие принципы для правильного планирования реальных экспериментов. Теперь мы можем задать вопрос о применении этих принципов. Ответ прост — нужно определить, *насколько успешно реально проведенный эксперимент репрезентирует (представляет) эксперимент безупречный*. Прежде всего посмотр-

рим, в какой мере исключена в наших экспериментах возможность посторонних влияний на зависимую переменную.

В оригинальном исследовании, проведенном в ткацком цехе, испытуемая 13 недель работала с наушниками и 13 чередующихся с ними недель — без наушников. При «неудачном» пересмотре эксперимента она носила наушники в первые 13 недель, а следующие 13 работала без них. В идеальном эксперименте испытуемая должна была бы работать и с наушниками, и без них одновременно. Понятно, что схема чередования недель приближается к этому идеалу в большей степени. Чередование двух условий, или АБАБАБАБАБ и т. д., более *репрезентативно* для их одновременного предъявления, чем последовательность, состоящая только из А и Б.

В своем первоначальном эксперименте Джек Моцарт разучивал пьесы в следующем порядке: целостный метод — частичный — частичный — целостный. В «неудачном» эксперименте последовательность была иной: целостный — целостный — частичный — частичный. В первом случае *усредненные* позиции целостного и частичного методов были одинаковы. Целостный метод занимал в последовательности позиции 1 и 4, среднее — 2,5. Позициями частичного метода были 2 и 3, среднее — 2,5. Напротив, в «неудачном» эксперименте целостный метод занимал позиции 1 и 2, среднее — 1,5, а частичный — 3 и 4, среднее — 3,5. Более *репрезентативным* для одновременного предъявления двух условий вновь оказался оригинальный эксперимент.

В первоначальном варианте своего эксперимента Йоко выпивала оба сорта сока — «Дж. Дж. Риттенхауз» и «Бадди'н'Билл» — в случайном порядке в течение 36 дней. В «неудачно» измененном варианте она ограничилась 4 днями. Ясно, что к бесконечности ближе 36, а не 4. Первоначальный план лучше *репрезентирует* бесконечный эксперимент, чем план измененный.

Эксперимент полного соответствия лучше *представлен* в оригинальном исследовании Джека, чем в его измененном варианте с вальсами. Хотя Джек и не разучивал все пьесы, которые он намеревался выучить в дальнейшем, он взял пьесы точно того же типа, т. е. выбрал соответствующий уровень дополнительной переменной. А вариант с вальсами оказывается «неадекватным», поскольку по своему уровню эти пьесы отличаются от тех, которые Джек разучивал бы в эксперименте полного соответствия.

Подводя итоги, можно сказать, что более надежную информацию об отношении между независимой и зависимой переменными дают те эксперименты, которые лучше представляют идеальный и бесконечный эксперименты. А чем ближе уровень значимой дополнительной переменной в проведенном эксперименте к ее уровню в эксперименте полного соответствия, тем лучше представлена в нем изучаемая реальная ситуация.

Валидность

В зависимости от того, насколько реальные эксперименты представляют безупречный, их называют более или менее валидными. Безупречный эксперимент позволил бы безошибочно отделить верную гипотезу от неверной. Если бы Джек Моцарт смог провести безупречный эксперимент, он бы совершенно точно знал, какая из его гипотез верна: *частичный метод лучше* или *целостный метод лучше*. Таким образом, говоря о *валидности* эксперимента, вы оцениваете качество той работы, которую предполагаете провести для определения справедливости одной из конкурирующих гипотез.

Внутренняя валидность. Всем трем описанным нами «неудачным» экспериментам недоставало внутренней валидности. Это значит, что они не позволяют рассматривать полученную картину отношений между независимой и зависимой переменными как достоверную. И виноваты в этом, как мы убедились, всевозможные посторонние влияния. Эксперимент, которому недостает внутренней валидности, не может быть использован для выяснения того, какая гипотеза о связи независимой и зависимой переменных истинна, а какая ложна. Например, если нам не ясно, почему ткачиха работала лучше: потому, что она носила наушники, или потому, что стояла хорошая погода, — мы не вправе считать результаты эксперимента достаточными для определения истинной и ложной гипотез о влиянии наушников на производительность труда.

Термин «внутренний» подчеркивает существо данного вида валидности. Можно сказать, что эксперимент, лишенный внутренней валидности, неудачен, так сказать, изнутри, по самой своей сути. Действительно, если он не позволяет убедиться в достоверности найденного отношения независимой и зависимой переменных, он просто бесполезен.

Внешняя валидность. «Неадекватный» эксперимент, который мог бы провести Джек, разучивая вальсы вместо сонат, не был бы неудачным в принципе. Это был бы вполне нормальный эксперимент по заучиванию вальсов. Бесполезным его считать нельзя. Джек мог бы использовать свои результаты, если бы задним числом решил, что на самом деле он искал наиболее эффективный метод заучивания вальсов. Однако этому эксперименту недостает внешней валидности. Он не обеспечивает достаточных оснований для определения верной и неверной гипотез о лучшем методе заучивания сонат.

Термин «внешний» относится к *определению* тематики проводимого эксперимента — чему именно он посвящен. В данном случае эксперимент не был внешне валидным потому, что «сонаты» — такая же необходимая составная часть проверяемой гипотезы, как независимая и зависимая переменные.

Общие определения. Понятия внешней и внутренней валидности являются центральными для всей нашей книги. Их применение в последующих главах в основных чертах определяется тем, что мы только что сказали. Сейчас мы приведем и более формальные определения этих понятий. Правда, все их значение вы поймете только тогда, когда познакомитесь с экспериментальными проблемами более высокого порядка. Но у вас уже будет основа для общего понимания и дальнейшего уточнения того, что такое валидность и два ее вида.

Начнем со схематического изображения экспериментальной гипотезы.

Независимая переменная... Отношение... Зависимая переменная... Уровни других переменных.

Итак, гипотеза включает в себя само отношение и обозначения обеих его сторон. Определение *валидности* эксперимента, как внутренней, так и внешней, состоит в следующем. Это степень правомерности вывода об экспериментальной гипотезе, которую обеспечивают результаты данного эксперимента по сравнению с результатами эксперимента, *безупречного во всех трех аспектах*.

Понятие внутренней валидности эксперимента касается только самого отношения и не затрагивает того, что именно соотносится. Отсюда *внутренняя* валидность — это степень правомерности вывода об экспериментальной гипотезе, основанного на результатах данного эксперимента, по сравнению с тем выводом, в основе которого лежат результаты *идеального и бесконечного* экспериментов, где изменения независимой и зависимой переменных происходят в одних и тех же условиях, а все другие, побочные, факторы остаются неизменными.

Любой эксперимент сталкивается также с проблемой *соответствия* исследуемой ситуации — реальной. Вопрос о соответствии уровня дополнительной переменной, типа музыки, уже возникал. Несколько позже мы обсудим подобные вопросы для независимых и зависимых переменных. Понятно, что вопросы о соответствии касаются содержания того, что стоит по обе стороны изучаемого отношения. Это и есть вопросы *внешней* валидности. Ее можно определить как степень правомерности данного вывода об экспериментальной гипотезе по сравнению с тем выводом, который основан на результатах эксперимента с *полным соответствием* независимой, зависимой и уровнями всех дополнительных переменных.

В настоящей главе мы обсудим главным образом проблему *внутренней* валидности. В любом эксперименте вы с самого начала столкнетесь с этой проблемой; если внутренняя валидность не достигнута, рассматривать внешнюю не имеет смысла. Вспомните, что в главе I были представлены эксперименты такого типа, для которых вопросы внешней валидности практически не обсуждаются.

А в следующей главе мы рассмотрим эксперименты, в которых именно эти вопросы выступают на первый план.

Никаких гарантий. Мы можем сказать, что эксперимент валиден, не зная фактически, правильны ли выводы. Мы можем сказать, что он невалиден, не зная, что выводы ошибочны. Причина в том, что мы не можем знать наперед, какая из двух конкурирующих гипотез верна. Ведь если бы мы об этом знали, нам не надо было бы проводить эксперимент. Если бы Джек заранее знал, какая из двух его гипотез справедлива: 1) частичный метод лучше или 2) целостный метод лучше, — он мог бы не проводить свое исследование.

При определении валидности реальных экспериментов мы должны сравнивать сами процедуры их проведения с процедурами «проведения» безупречного эксперимента. Валидный эксперимент представляет безупречный эксперимент лучше, чем невалидный. Следовательно, в валидном эксперименте мы с *большой вероятностью* можем получить такие результаты, которых могли бы достичь в безупречном эксперименте. При этом важно помнить, что обобщение ограниченных — и всегда несовершенных — экспериментальных данных связано с риском. Даже самый высоковалидный эксперимент может дать неточную информацию о правильности экспериментальной гипотезы, а информация, полученная в невалидном эксперименте, может оказаться точной. Причины такого риска и его влияние на интерпретацию экспериментальных результатов мы обсудим в следующих главах, прежде всего — в главе 6 («Значимые результаты»).

ФАКТОРЫ, УГРОЖАЮЩИЕ ВНУТРЕННЕЙ ВАЛИДНОСТИ

Теперь мы можем применить понятие безупречного эксперимента (идеального и бесконечного) для описания того, что мешает достижению внутренней валидности в реальных экспериментах. Как мы увидим, некоторые из таких помех устранить нельзя; они необходимо связаны с процедурами проведения наших не вполне безупречных экспериментов. Скажем, если Джеку нужно разучить две пьесы, одну из них он неизбежно будет разучивать первой. Существуют, однако, и такие трудности, которые можно преодолеть, если заранее об этом позаботиться. Так, Джек уже знал, что не стоит применять частичный и целостный методы в разное время дня.

Изменения во времени

Известные побочные факторы. В идеальном эксперименте различные состояния независимой переменной предъявляются испытуемому одновременно. Этого Джек сделать не мог, но он мог

по крайней мере заниматься в одно и то же время. Время дня — это заранее известная *побочная* (т. е. отличная от независимой) *переменная*, которая может повлиять на эффективность занятий, и ее надо *сохранять неизменной*. Если бы Джек был невнимателен, то в разные дни эксперимента он мог бы заниматься то при закрытых, то при открытых окнах. А уличный шум может сильно повлиять на эффективность занятий. Поэтому лучше сохранять его неизменным, держа окна закрытыми. В эксперименте с наушниками, который продолжался более шести месяцев, исследователи знали о возможных изменениях температуры и влажности в ткацком цехе. К сожалению, условия эксперимента не позволяли им исключить эти изменения. Но экспериментаторы фиксировали и старались учесть влияния названных факторов. И что самое главное — чередование двух условий независимой переменной снижало влияние этих факторов. Экспериментатор должен стараться заранее определить все возможные факторы, которые с течением времени могут изменяться. И главное, стараться удерживать их на постоянном уровне при каждой новой пробе.

Нестабильность во времени. Но даже стараясь изо всех сил, экспериментатор не сумеет сделать одну пробу *в точности* (кроме различия уровней независимой переменной) похожей на другие. Некоторая *нестабильность* во времени будет всегда. В эксперименте она проявляется в изменчивости побочных факторов, а также в некоторых вариациях самой независимой переменной. Наконец, всегда остаются совершенно неясные источники сильных колебаний в ответах испытуемых, приводящие к увеличению разброса экспериментальных данных. Давайте рассмотрим конкретные примеры каждой из этих трех форм нестабильности во времени.

Изменчивость побочных факторов. Часто бывает так, что экспериментатор знает о существовании посторонних факторов, влияющих на зависимую переменную, но не может управлять ими непосредственно. Какой-то день в работе ткачихи мог оказаться «не самым удачным» из-за того, что накануне она поздно легла спать. Конечно, экспериментатор мог бы попытаться убедить ее не делать этого, пока эксперимент не завершится. Но ведь эксперимент продолжался шесть месяцев! Поужинав накануне в ресторане, Джек неважно себя чувствовал во время разучивания одной из пьес — в другой раз ему следует быть осторожнее.

От пробы к пробе окружающие условия никогда не остаются неизменными. Описывая эксперимент в ткацком цехе, исследователи утверждают:

«Хорошо известно, что на производительность ткацкого труда могут влиять атмосферные условия. Так, с повышением температуры и относительной влажности уменьшается количество обрывов нити. С другой стороны, дальнейшее повышение того и другого, продолжая благоприятно воздействовать на физические свой-

ства пряжи, неблагоприятно сказывается на физиологическом состоянии людей, работоспособность которых может снизиться так, что это сведет на нет любые положительные влияния» (Weston H. C., Adams S., 1932, с. 56).

Следовательно, даже измеряя температуру и влажность, нельзя установить точно их влияние на производительность труда. Список побочных переменных можно было бы продолжать до бесконечности, включая в него и субъективные факторы, как, например, хорошее или плохое самочувствие испытуемого в течение эксперимента. Добросовестный экспериментатор может фиксировать некоторые из этих изменений, но не может их избежать. Теперь вам понятно, почему экспериментатор стремится уйти из реального мира в прекрасные звуконепроницаемые лаборатории и иметь дело с такими испытуемыми (белыми крысами), поведение которых он может контролировать 24 часа в сутки. Но даже там калориферы иногда остывают, бутылки с водой засоряются, и крысы подхватывают «насморк».

Само пребывание в ситуации эксперимента может вызвать продолжительные изменения в поведении испытуемого. Таков был главный вывод из знаменитых экспериментов «Хауторн», вывод, важный для всех экспериментальных психологов. На Западном электрозаводе в Хауторне (Иллинойс) было проведено исследование влияния освещения в цехах на производительность сборочных работ. Предварительные попытки установить какую-либо закономерность закончились неудачей. Тогда было предпринято систематическое исследование условий труда рабочих (Roethlisberger F. J., Dickson W. J., 1946). Основной частью этого исследования были эксперименты с заданием по сборке переключателей. Оно представляло собой «сборку телефонных реле; это операция, которую обычно выполняют женщины: нужно соединить примерно 35 небольших деталей в «сборную арматуру» и закрепить ее четырьмя винтами» (там же, с. 20).

Для эксперимента была оборудована специальная комната, чтобы исследователи могли контролировать условия работы и адекватно оценивать деятельность операторов. В качестве испытуемых в эксперименте приняли участие пять молодых женщин, вполне освоивших данный вид работы. Исследовались две независимые переменные: распределение периодов отдыха, а также длина рабочего дня и рабочей недели. Оплата труда производилась в соответствии с общим количеством переключателей, собранных бригадой из пяти человек.

Было установлено, что независимо от распределения периодов отдыха и длины рабочего дня и недели производительность труда продолжала расти в течение двух лет! Исследователи сообщают, во-первых, о «постепенном изменении социальных отношений в группе операторов в направлении групповой сплоченности и со-

лидарности и, во-вторых, об изменении отношений между операторами и их контролерами. Организаторы эксперимента стремились создать среди девушек атмосферу взаимной поддержки и сотрудничества, избавить их от излишних волнений и тревог. Эти усилия по созданию необходимых условий эксперимента косвенным путем привели к изменению отношений между людьми» (там же, с. 58 — 59).

Пользуясь нашей терминологией, эту ситуацию можно описать так. До эксперимента социальные условия работы испытуемых находились на одном уровне. В ситуации эксперимента данная «побочная переменная» перешла на другой уровень. Это привело к продолжительному изменению зависимой переменной — производительности труда, несмотря на то что объективно социальные условия в эксперименте оставались неизменными.

Независимая переменная. Рассчитывать на полную идентичность каждого из условий независимой переменной на всем протяжении эксперимента мы не можем. В какие-то дни или даже недели наушники могли быть надеты не так удобно, как в остальные. Несмотря на все старания Джека, он может по-разному относиться, например, к частичному методу, разучивая разные пьесы. И Йоко знала о вариациях каждого из условий своей независимой переменной. Сок одного и того же сорта в любых двух банках не бывает одинаков, и разница подчас очень велика. Какие-то изменения будут встречаться даже в тех экспериментах, в которых, казалось бы, достигнуто полное единообразие условий. Яркость электрического света (как стимула) будет меняться от перепадов напряжения в сети, а они случаются довольно часто. В течение эксперимента могут возникать и закономерные изменения, например с увеличением срока службы лампочки ее свет может становиться все менее ярким.

Зависимая переменная. При действии одной и той же независимой переменной испытуемый не всегда будет давать один и тот же ответ. Так будет даже в том случае, если экспериментатор необычайно искусен и пунктуален в устранении нестабильности побочных факторов и независимой переменной.

Нестабильность зависимой переменной очень эффектно представлена на графиках, отражающих результаты двух экспериментов. На рис. 2.1 показана недельная выработка испытуемой Д. в эксперименте с наушниками. Как видим, меньше всего ударов она пропускала с десятой по двенадцатую неделю и с восемнадцатой по двадцать вторую. А самые неудачные ее показатели — наибольшее количество пропущенных ударов — приходятся на четырнадцатую неделю и окончание эксперимента. И что особенно интересно — для обоих условий работы кривые поднимаются и опускаются совместно. Изменения в производительности труда, происшедшие с течением времени, несомненно, более суще-

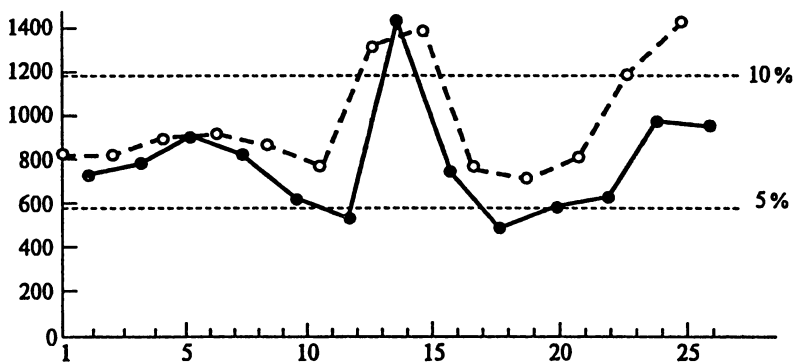


Рис. 2.1. Понедельная производительность труда испытуемой Д. Ось абсцисс — очередность недель проведения эксперимента; ось ординат — количество пропущенных ударов (в среднем за час); пунктирная линия — работа без наушников, сплошная — с наушниками

ственные, чем различия между использованием и неиспользованием наушников.

На рис. 2.2 показаны изменения ответов испытуемого в эксперименте на время реакции выбора. Пробы давались каждые шесть секунд; испытуемый должен был сдвигать рукоятку к себе или от себя и тем самым совмещать две световые точки. Разумеется, точки предъявлялись в случайном порядке. За 70 запланированных последовательных проб во времени реакции испытуемого наблюдались как непродолжительные колебания, так и более регулярные отклонения. Самое короткое время реакции было показано

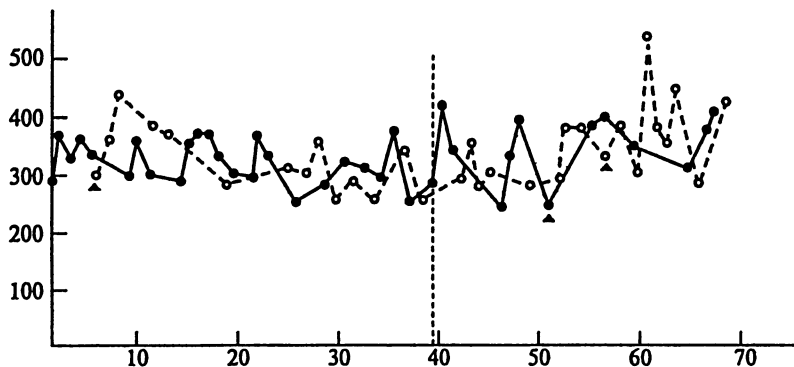


Рис. 2.2. Время реакции выбора для 70 последовательных проб. Ось абсцисс — номера проб (пунктиром отмечен период отдыха); ось ординат — время реакции (в мс); пунктирная линия — сдвиг рукоятки к себе, сплошная — от себя; треугольниками помечены ответы с ошибками

примерно между тридцатой и сороковой пробами, а самое большое — между шестидесятой и семидесятой. И это увеличение нельзя считать результатом усталости, ведь как раз перед сороковой пробой испытуемый отдыхал. В итоге самые большие показатели несколько превышали 400 мс, а самые малые — 200 мс, т. е. время реакции изменялось в отношении два к одному.

Таким образом, в исследовании времени реакции были обнаружены поминутные и даже посекундные его изменения. Они не связаны с усталостью, скорее их можно объяснить *колебаниями внимания*. На графике результатов ткачихи Д. отражены значительные колебания производительности ее труда. При этом взлеты и падения кривых, по-видимому, не зависят от температуры и влажности. Правда, увеличение числа пропущенных ударов к окончанию эксперимента можно объяснить применением искусственного (газового) освещения; оно было необходимым, так как эксперимент заканчивался осенью.

Даже тогда, когда сами ответы испытуемых постоянны, изменения могут вноситься процедурой их *измерения*. Счетчик фиксирует каждое передвижение челнока, совершающего новый удар. Однако приборы не всегда бывают исправны. А если измерения связаны с субъективными суждениями, то они непременно окажутся менее стабильными. Джек считал пьесу выученной окончательно после двух безошибочных исполнений наизусть. Однако в исполнении пьес было довольно много небольших погрешностей, почти ошибок. Иногда Джек мог посчитать их ошибками, а иногда нет. И объяснилось это вполне естественными колебаниями его субъективного состояния. Изменения в оценке исполнения пьес могли быть и регулярными. Например, в процессе эксперимента Джек мог все строже относиться к своим ошибкам.

Различия экспериментальных задач

Одну и ту же пьесу нельзя заучить (как в идеале) двумя различными методами одновременно. Но даже если методы следуют один за другим, их все равно нельзя применить к одной и той же пьесе. Если пьеса заучена, она заучена. Существуют эксперименты, в которых необходимо не только предъявлять разные экспериментальные условия в разное время, но и менять трудность задач. Это весьма значительное отличие от идеального эксперимента. Как Джек может убедиться в одинаковой трудности выбранных им пьес? А ведь в любом эксперименте по исследованию научения с участием одних и тех же испытуемых задачи для разных условий независимой переменной с необходимостью будут разными.

Эффекты последовательности

В неудачном варианте своего эксперимента Джек сначала заучивал две пьесы частичным методом, а затем две другие — целостным. Мы уже знаем, что на качество его игры могут повлиять любые (в том числе только что описанные) факторы, изменяющиеся во времени. Однако есть и другие влияния, связанные с положением каждого из условий независимой переменной в последовательности их предъявления. Влияния одного из условий на следующие за ним называются *эффектами последовательности*, эффектами порядка или эффектами переноса. Они могут быть положительными и отрицательными, иметь общий и специфический характер. Применение частичного метода могло положительно повлиять на дальнейшие занятия Джека по целостному методу за счет увеличения практики или привыкания к режиму эксперимента. Оно могло оказать и отрицательное действие: привычка заучивать пьесы короткими отрывками могла помешать заучиванию больших частей или Джек мог просто устать от занятий.

Предубеждения экспериментатора

Во времена возникновения автомобиля бытовал такой анекдот в форме загадки. *Вопрос:* Какой самый главный винтик в автомобиле? *Ответ:* Тот, что держит руль¹. В том же духе можем спросить и мы. *Вопрос:* Какой из факторов, угрожающих валидности эксперимента, самый опасный? *Ответ:* Экспериментатор.

Если у исследователя есть какие-либо ожидания относительно результатов эксперимента, особенно связанные с предпочтением одного из условий независимой переменной, то эти ожидания так или иначе проявятся при проведении опытов. Йоко хорошо знала, что главное — это составить случайную последовательность обоих сортов сока. Ей хотелось исключить любой намек на то, какой именно сорт она оценивала каждое утро. А вот Джек не проявил должной аккуратности. Сначала он подобрал пары пьес, которые казались ему одинаковыми по трудности (с тем, чтобы разучивать каждую из пьес разными методами), а затем сам расположил их в определенной последовательности. Но если при этом он рассчитывал на большую эффективность частичного метода, то мог невольно отобрать для целостного метода более трудные пьесы из каждой пары.

Кроме того, субъективные оценки качества исполнения пьес могли колебаться не случайным образом (как было показано выше). Джек мог невольно отдавать предпочтение одному из методов.

¹ В оригинале — каламбур, основанный на многозначности слова «nut», в частности «техническая деталь» и «незадачливый человек, чудака». — *Примеч. пер.*

Поэтому при оценке исполнения обеих пьес каждой пары Джеку не следует слишком верить в частичный метод, но и применяя целостный, тоже стараться достичь самых высоких результатов.

В эксперименте с наушниками исследователи, естественно, рассчитывали повысить с их помощью производительность труда и вполне могли передать свою уверенность участникам эксперимента. Поэтому, возможно, с наушниками ткачихи (в среднем) старались работать лучше.

Одно из самых коварных следствий предубежденности экспериментатора — это нежелание учитывать некоторые экспериментальные данные как полученные якобы при нетипичных условиях, например при сильном уличном шуме. К сожалению, мнение экспериментатора о нетипичности условий часто весьма субъективно. Отсюда один и тот же уровень шума будет считаться нетипичным при одном состоянии независимой переменной, но вполне нормальным — при другом.

От предубеждения экспериментатора может зависеть даже точность регистрации данных. Показано, например, что в протоколах экспериментов по изучению экстрасенсорного восприятия встречаются ошибки в пользу наличия соответствующих феноменов, если протоколист верит в их существование. Те же, кто не верит в экстрасенсорное восприятие, не допускают подобных искажений (Kennedy J. L., 1939). Тщательный анализ этой проблемы в целом представлен в книге «Влияния экспериментатора в психологическом исследовании» (Rosenthal R., 1966).

ПОВЫШЕНИЕ ВНУТРЕННЕЙ ВАЛИДНОСТИ С ПОМОЩЬЮ ПЕРВИЧНОГО КОНТРОЛЯ

Термин «*контроль*» используется в нашей книге для обозначения любого способа усовершенствования экспериментов, который приближает их к безупречному. Здесь мы будем иметь дело с угрозами внутренней валидности эксперимента и мерами, которые позволяют их устранять. Если контроль в эксперименте недостаточен, то внутренняя валидность может нарушиться двояким образом. Она может пострадать от *ненадежности* и от *систематического смещения*. Ниже мы рассмотрим способы *первичного контроля* (применимые к любой экспериментальной схеме), которые могут повысить надежность эксперимента и сократить систематическое смещение.

Ненадежность

Предположим, что исследование в ткацком цехе проводилось лишь в течение 11-й и 14-й недель из 26. Для выбора, в какую из этих недель давать наушники, просто бросали монету (чтобы из-

бежать предубеждений). Случилось так, что выбор пал на 14-ю неделю. А теперь посмотрите на рис. 2.1. На этой неделе испытуемая Д. работала плохо: она пропускала более 1400 ударов за час. С другой стороны, 11-я неделя была для нее удачной: менее 800 пропущенных ударов за час. Объективность такого эксперимента сомнений не вызывает, он просто слишком короткий. За две другие недели ткачиха покажет иные результаты и т.д. Двухнедельный эксперимент весьма далек от бесконечного, а 26-недельный гораздо ближе к нему, и у нас есть все основания надеяться, что другой эксперимент в 26 недель даст примерно те же результаты. Конечно, если разброс будет невелик, то большего количества проб не потребуются.

Когда мы говорим о надежности среднего значения зависимой переменной для каждого из условий независимой переменной, то имеем в виду его устойчивость при повторении эксперимента второй, третий, четвертый и т.д. раз. Итак, если надежность высокая, то при многократном повторении эксперимента будет воспроизводиться примерно одно и то же значение зависимой переменной.

Выбор адекватного количества проб. Для достижения высокой надежности эксперимента количество проб должно соответствовать изменчивости изучаемого поведения. В эксперименте с временем реакции, как можно судить по рис. 2.2, требуется много проб — 50, 100 или даже больше. Ведь экспериментальные данные сильно варьируют от пробы к пробе. Джеку Моцарту в эксперименте с заучиванием фортепьянных пьес проб нужно гораздо меньше, поскольку каждая проба включает довольно много отдельных действий, и результат измерения оказывается более стабильным. Таким образом, необходимое количество проб зависит от разброса результатов измерений.

Сокращение изменчивости во времени. Увеличение количества проб — лишь один из способов повышения надежности. Другой — сокращение самой изменчивости изучаемого поведения. Достигается оно прежде всего с помощью *организации* хода эксперимента и его *протоколирования*, уже обсуждавшихся в главе 1. Процедура эксперимента должна строго соблюдаться, а значимая информация о ней (учитывая возможность ошибок при воспроизведении по памяти) — фиксироваться документально. Кроме последовательного выполнения запланированных действий уменьшения разброса экспериментальных данных можно достичь, соблюдая *точность в эксперименте*.

Любые меры, которые позволяют сделать эксперимент более точным, сокращают изменчивость поведения во времени. Так, чтобы не увеличивать эту изменчивость за счет поведения экспериментатора, везде, где это возможно, применяются автоматические методы. Например, в эксперименте с временем реакции

выбора экспериментатор сам мог бы давать испытуемому команду «Приготовьтесь!» примерно за секунду до предъявления сигнала. Разумеется, время подачи такой команды, громкость голоса экспериментатора не могут не изменяться. Поэтому гораздо лучше использовать для предупреждения испытуемого прибор-автомат, зажигающий стандартную световую точку. Далее, экспериментатор мог бы измерять время реакции с помощью обычного секундомера: включать его при предъявлении сигнала и останавливать, увидев, что испытуемый дал ответ. Но это, несомненно, увеличило бы различие между пробами. Использование же электрических или электронных часов, которые автоматически включаются вместе с сигналом и останавливаются с выдачей ответа, позволяет практически предотвратить подобные вариации.

Из сказанного вполне понятно, что любые способы стабилизации побочных переменных повышают надежность эксперимента. Джеку Моцарту, конечно, не следует во время своих занятий то закрывать окна, то широко раскрывать их. Если у него есть возможность поддерживать в комнате нормальную температуру, то окна лучше всегда держать закрытыми, чтобы уменьшить шум транспорта и, что более важно, сократить колебания в уровне этого шума. Мы помним, как внимательно следила Йоко за тем, чтобы томатный сок, который она пила, имел постоянную температуру.

Систематическое смещение

Однако на самом деле ненадежность — еще не самый худший вариант нарушения внутренней валидности. В принципе этот недостаток всегда можно исправить путем увеличения числа проб. В тех же случаях, когда этого сделать нельзя, мы используем термин «*систематическое смещение*». План «неудачного» эксперимента с ткачихами — работа в первые 13 недель с наушниками и следующие 13 без них — не позволяет улучшить эксперимент путем увеличения количества проб. Можно повторить эксперимент полностью, но и тогда останется сомнение, не пришелся ли один из этих 13-недельных периодов на времена, когда зависимая переменная испытывала благоприятные (или неблагоприятные) влияния. Напротив, с помощью плана чередования недель, который применялся в действительности, можно продлить эксперимент настолько, насколько потребуются для достижения надежности. Следовательно, «неудачный» план приводит к систематическому смещению *независимой переменной* (использование или неиспользование наушников) с другими *факторами, изменяющимися во времени* (в первые или вторые 13 недель).

Систематическое смещение означает, что независимая переменная постоянно сопровождается некоторыми побочными пере-

менными. Использование наушников сопровождается изменениями в состоянии ткачихи за первые 13 недель, их неиспользование — изменениями в ее состоянии за вторые 13 недель. По данным такого эксперимента мы не сможем судить о том, была ли высокая производительность труда (за первые 13 недель) результатом воздействия независимой переменной (использования заглушек) или каких-то других факторов. Теперь нам понятно, почему описанный эффект (в отличие от ненадежности) назван *систематическим*. Надежность можно повысить, увеличивая число проб, ведь в течение долгого времени любые благоприятные (или неблагоприятные) влияния окажутся размытыми. Если же эксперимент страдает от систематического смещения, то с увеличением проб этот недостаток лишь усугубится.

Довольно много подобных влияний может возникнуть в случае, если не принять необходимых мер против предубеждений экспериментатора. Если экспериментаторы надеялись на повышение производительности труда с помощью наушников, то они могли передать свою веру рабочим, а быть может, и ошибаться в снятии показаний счетчика ударов. Все это явно нарушило бы равновесие между двумя состояниями независимой переменной. Одно из этих состояний (использование наушников) сопровождалось бы одним уровнем побочной переменной — благоприятным предубеждением экспериментатора, а второе состояние (неиспользование наушников) — другим ее уровнем — неблагоприятным предубеждением. При этом экспериментатору вовсе не обязательно проявлять свои склонности постоянно. Даже эпизодические случаи будут порождать систематические влияния, приводящие к неравенству двух состояний независимой переменной.

СХЕМЫ КОНТРОЛЯ ЗА ФАКТОРАМИ ВРЕМЕНИ В ЭКСПЕРИМЕНТЕ С ОДНИМ ИСПЫТУЕМЫМ

В описанных нами исследованиях разные состояния, или условия, независимой переменной давались одному и тому же испытуемому. Каждый раз реальный эксперимент отличался от идеального тем, что условия давались испытуемому в *разное время*. Следовательно, в эксперименте нужно каким-то образом контролировать довольно большое количество факторов, которые можно объединить под названием «*факторы времени*». Для этой цели есть практически только три схемы последовательного предъявления условий, которые нам нужно подробно обсудить. Такие схемы, как предъявление сначала всех проб одного условия, а затем — всех проб второго условия, вроде «неудачного» эксперимента с наушниками, нас вообще не интересуют. Три экспериментальные схемы — это те, что применялись в трех экспериментах из главы 1: случайная последовательность условий (эксперимент с томатным

соком), их регулярное чередование (наушники) и позиционно уравненная последовательность (заучивание фортепьянных пьес). Сейчас мы разберем основания для применения каждой из этих схем и обсудим, насколько успешно они позволяют контролировать факторы времени.

Схема случайной последовательности

Эта схема особенно хороша для тех экспериментов, где сами пробы достаточно коротки, но в интересах надежности их используется довольно много. В тех случаях, когда испытуемый не должен знать о состоянии независимой переменной в каждой данной пробе, как в эксперименте Йоко, схема случайной последовательности единственно возможна.

Как следует из самого названия схемы, состояния, или условия, независимой переменной предъявляются в случайном порядке. Бросают, скажем, игральную кость и смотрят, на какое число точек (четное или нечетное) она выпала. Если сравниваются два условия, то выпадению нечетного числа может соответствовать условие А, а четного — условие Б. Если количество проб для каждого из условий должно быть одинаковым, то можно распределить их так, как это сделала Йоко, — воспользоваться равным количеством бумажных бланков для того и другого условия. Более строгий способ составления случайной последовательности будет изложен в главе 4.

При использовании этой схемы исключается всякая возможность систематического смещения независимой переменной с *факторами времени*, поскольку в случайной последовательности никакой системы не существует. С увеличением проб повышается надежность эксперимента.

Схема регулярного чередования

Нам понятно, почему в исследовании с наушниками экспериментаторы использовали регулярное чередование условий, а не их случайную последовательность. Они не знали, сколько времени фабрика будет выпускать один и тот же вид ткани, и хотели, чтобы число недель работы с наушниками и без них было одинаковым. Если распределять условия случайно, работа с наушниками могла начаться, скажем, лишь на 10-ю неделю из первых 15. Кроме того, если с течением времени в самочувствии оператора, состоянии станка, влажности воздуха и т. п. происходили систематические изменения, то одно из условий, более часто попадающее в первую (или вторую) половину эксперимента, могло иметь преимущество перед другим. Поэтому для достижения надежности здесь нужно регулярное чередование, а не случайная последовательность.

Данную схему можно использовать при проведении эксперимента с большим числом проб, если только эти пробы не обязательно распределять случайно (чтобы усреднить влияния факторов времени). А вот в эксперименте Йоко никакой регулярной схемы применить было нельзя: нужно было избежать влияния пристрастности испытуемой к одному из экспериментальных условий.

У схемы регулярного чередования есть только один небольшой недостаток. Что если бы в каждую вторую неделю эксперимента происходило некоторое событие, скажем, очередная наладка станка? Это привело бы к систематическому смещению независимой переменной, которое благоприятствовало бы одному из ее условий, следующему сразу же за этим событием. Напротив, при случайном распределении условий любые влияния факторов времени можно компенсировать, увеличивая продолжительность эксперимента. Ведь для этой схемы более актуален вопрос о ненадежности, чем о систематическом смещении с побочными факторами.

Схема позиционно уравненной последовательности

Позиционно уравненная последовательность условий применяется в тех случаях, когда экспериментатор по той или иной причине не использует большого количества проб (или блоков проб). Именно таким был случай Джека Моцарта. Джек хотел найти более эффективный метод заучивания достаточно быстро, не разучивая в эксперименте слишком много пьес. Поэтому здесь позиционно уравненная последовательность состояла только из четырех проб — АББА (целостный — частичный — частичный — целостный). Аналогично, если экспериментатору нужно сравнить время реакции на появление звуковых и световых сигналов, он мог бы составить последовательность из четырех *блоков* по 50 проб в каждом. Во всех пробах каждого блока применялся бы либо только звук, либо только свет. Тогда в позиционно уравненной последовательности АББА: звук — свет — свет — звук — каждую позицию занимала бы не отдельная проба, а целый их блок.

При использовании данной схемы, так же как и двух других, надежность эксперимента зависит от количества проб. Мы имеем в виду либо количество отдельных действий, входящих в одну пробу, либо число самих проб в блоке. А как обстоит дело с систематическим смещением независимой переменной? Если в эксперименте применяется схема позиционного уравнивания, то оба условия следуют в среднем через одинаковые временные промежутки. Это позволяет контролировать любые изменения во времени, которые имеют линейный характер. Как отмечалось ранее, в последовательности из четырех проб в эксперименте Джека каждое условие занимало в *среднем* позицию 2,5. Если предположить,

что все возможные изменения нарастают по прямолинейной функции, то данная схема позволяет их хорошо контролировать. Например, если бы с исполнением каждой новой пьесы Джек становился все более сосредоточенным и уровень его внимания возрастал каждый раз на 4 условные единицы, то влияния данного фактора времени оказались бы контролируемы. Скажем, при разучивании первой пьесы уровень внимания Джека имел значение 90, при разучивании второй — 94, третьей — 98 и четвертой — 102. Среднее значение для условия А (пробы 1 и 4) составило бы 96 и среднее для условия Б (пробы 2 и 3) — тоже 96.

Для проверки предположения о прямолинейном характере происходящих изменений нам стоит еще раз посмотреть на рис. 2.1 и 2.2. Эти изменения совершенно бессистемны и непредсказуемы. Из них можно заключить, что сосредоточенность Джека возрастала от пробы к пробе: на 5 единиц, на 2 и на 1. Тогда ее значения для каждой из четырех проб будут следующими: 90, 95, 97 и 98. Среднее для условия А (90 и 98) составит 94, в то время как для условия Б (95 и 97) — 96. Указанная схема не позволит уравнивать влияния факторов времени для обоих экспериментальных условий. В результате возникнет *систематическое смещение* независимой переменной с этими побочными влияниями.

КАК КОНТРОЛИРОВАТЬ ФАКТОРЫ ЗАДАЧИ

Вы помните, что если бы Джек смог провести идеальный эксперимент, он заучивал бы одну и ту же пьесу двумя разными методами. Поскольку это невозможно, то самое лучшее — найти пару пьес, *одинаковых* по трудности. Такая проблема возникает в любом эксперименте, где из-за влияния научения для разных экспериментальных условий нужно использовать разный материал, т. е. разные задачи. Давайте посмотрим, как можно уравнивать влияния факторов, связанных с различием задач (или, короче, *факторов задачи*), с помощью трех указанных схем, в том числе схемы позиционного уравнивания, которой пользовался Джек.

Случайная последовательность

Начнем с того, что здесь понадобились бы не четыре пьесы, а гораздо больше, пожалуй, даже слишком много, чтобы реализовать эту схему на практике. Предположим, однако, что такой эксперимент можно осуществить. Тогда есть две стратегии подбора различных пьес (задач). Первая — это выбрать для заучивания 30 или 40 пьес и затем расположить их в случайном порядке. Название каждой пьесы можно записать на бланке, положить бланки в коробку, а затем выбирать их случайным образом. Иначе говоря, случайную последовательность пьес можно получить так же, как

и случайную последовательность условий независимой переменной. Другая стратегия заключается в разделении пьес на пары по степени их трудности. Если всего отобрано 30 пьес, то сначала составляют пару из двух самых трудных пьес, затем из двух, следующих за ними по трудности, и т.д., получая таким образом 15 пар. В каждой паре путем случайного выбора, т. е. бросая монету, определяют, какая пьеса будет заучиваться каждым из методов. Затем пары можно расположить в порядке возрастания трудности и приступить к заучиванию, начиная с самой легкой пары. Пьесы, отобранные для каждого метода, можно распределить и случайно. Однако при этом влияние разного уровня трудности задач усилится за счет влияния временных изменений, точнее, факторы задачи просто войдут в состав факторов времени. Если пьесы сильно различаются по трудности, то для достижения надежности эксперимента потребуется большое количество проб, но зато не будет никакого систематического смещения.

Регулярное чередование

Поскольку эта последовательность используется в экспериментах с несколько меньшим количеством проб, то следует применять не простое случайное распределение по каждому из методов, а с предварительным разделением пьес на пары. Разучивание лучше всего начинать с двух самых легких пьес, затем переходить к двум несколько более трудным и т.д. При слишком большом различии заданий, снижающем надежность эксперимента, также следует увеличить количество проб. Здесь систематическое смещение независимой переменной с факторами задачи будет отсутствовать.

Позиционно уравненная последовательность

Если экспериментатор использует только одну серию позиционно уравненной последовательности АББА, как это было у Джека, решающую роль играет подобие пьес в каждой паре. Сначала Джек попытался отобрать для каждого метода две длинные и две короткие пьесы. Затем он составил пары пьес, практически одинаковых по трудности. И все же ему не стоило самому принимать решение о том, каким методом заучивать ту или другую пьесу. Для каждого метода он должен был случайно выбрать по пьесе из обеих пар. Тогда он избежал бы невольного отбора более легких пьес для предпочитаемого им метода. Однако принимая свое решение, Джек не мог иметь точного представления о трудности пьес, поэтому его выбор можно считать удовлетворительным. Но так или иначе, систематического смещения независимой переменной с факторами задачи в данном случае избежать нельзя. Насколько оно серьезно — зависит от того, в какой мере подтвердится пред-

положение о подобии пьес. Экспериментатору всегда стоит стремиться отбирать такие задачи, которые можно как-то измерить. А это очень непросто. Было показано, например, что даже наборы бессмысленных слогов могут различаться по своей семантике и трудности заучивания.

ПРОБЛЕМА ЭФФЕКТОВ ПОСЛЕДОВАТЕЛЬНОСТИ

Важность этой проблемы все более осознается в экспериментальной психологии (Poulton E. C., Freeman P. R., 1966). Сейчас мы обсудим, почему *эффекты последовательности* — это основной фактор, нарушающий внутреннюю валидность в эксперименте с одним испытуемым. Речь идет о тех характеристиках ответа испытуемого в данной пробе, которые вызваны самим фактом предъявления ему предыдущих проб. Эти влияния могут быть положительными и отрицательными. Они могут иметь глобальный характер, как, например, адаптация к режиму эксперимента или усталость испытуемого. Они могут быть и специфическими, скажем, умение предсказывать тот момент, когда в челноке окончится нить. Они бывают недолгими, распространяющимися лишь на одну последующую пробу, и продолжительными, причем их действие от пробы к пробе может накапливаться. Такие влияния обычно называются эффектами переноса.

Однородные и неоднородные эффекты

Рассмотрим эффект переноса, который накапливается по мере предъявления проб. Предположим, что в течение четырех экспериментальных проб (целостный метод — частичный — частичный — целостный) Джек Моцарт постепенно втягивался в режим эксперимента. Если подобное влияние на каждую последующую пробу является *однородным*, то ни один из методов заучивания не получит преимущества перед другим. Например, если величина положительного переноса каждой предыдущей пробы (или переноса от пробы к пробе) равна 2 «единицам», то значения «помощи» испытуемому были бы следующими: первая проба целостного метода — никакой помощи, первая частичного метода — 2 единицы, вторая частичного метода — 4 единицы, вторая целостного метода — 6 единиц. В итоге на каждый метод, целостный и частичный, пришлось бы по 6 единиц. Таким образом, эффект однородного переноса оказывается уравновешенным.

Однако чаще случается так, что процесс научения сначала протекает более интенсивно, а затем замедляется. Поэтому в нашем примере лучше предположить, что перенос первой пробы на вторую равнялся 3 единицам, от второй к третьей пробе он возрастал на 2 единицы, а от третьей к четвертой — только на 1. При таком

неоднородном переносе каждой пробе отвечали бы следующие значения: первая проба целостного метода — никакой помощи, первая частичного метода — 3 единицы, вторая частичного метода — 5 единиц, вторая целостного метода — 6 единиц. Теперь целостный метод по-прежнему получает в итоге 6 единиц, в то время как частичный — 8.

При использовании позиционно уравненной последовательности АББА условия независимой переменной (А или Б) оказываются связанными с *ранним-или-поздним* переносом. Условие А связано с поздним переносом, поскольку оно получает «помощь» только на четвертой пробе, а условие Б — с ранним на второй и третьей пробах. Внутренняя валидность эксперимента пострадает в той мере, в какой предположение об однородности переноса окажется неверным. Приведенное объяснение справедливо не только в случае положительного, но и отрицательного эффекта последовательности, например по причине усталости испытуемого. Только в этом случае преимущество получает условие А.

При использовании схем случайной последовательности и регулярного чередования, когда число проб достаточно велико, проблема неоднородных влияний далеко не так существенна, как при позиционно уравненной последовательности. Ведь каждое из условий независимой переменной проходит в эксперименте по нескольку раз как в ранних, так и в поздних пробах. Гораздо большую опасность для достижения внутренней валидности при использовании любой из трех рассмотренных схем представляют асимметричные влияния. Обратимся к их описанию.

Симметричные и асимметричные эффекты

Если мы уверены в том, что в нашем эксперименте с одним испытуемым эффекты последовательности являются симметричными, то связанные с ними трудности можно преодолеть. Посмотрим, почему это так и что означает в данном контексте слово «*симметричные*». Оно означает, что влияние условия А на последующее условие Б является точно таким же, как и влияние условия Б на последующее условие А. Скажем, использование частичного метода занятий в эксперименте Джека Моцарта точно так же влияет на последующее применение целостного метода, как использование целостного метода на последующее применение частичного.

Предположим, что между двумя методами существует своего рода антагонизм, т.е. негативный перенос, равный 5 единицам. При последовательности методов: целостный — частичный — частичный — целостный — его эффект обнаружится на второй и на четвертой пробах, т.е. на второй (частичный метод) и четвертой (целостный) пробах эффективность заучивания пьес будет снижать-

ся на 5 единиц. Таким образом, при позиционно уравненной последовательности, которой воспользовался Джек, эти симметричные влияния взаимно компенсируются. При большем количестве проб (в случайной или чередующейся схемах) условие А предшествует условию Б, а Б предшествует А примерно одинаковое число раз, и поэтому их взаимовлияния вновь будут уравнены.

Но если перенос условия А на условие Б отличается от влияния Б на А, то экспериментатор оказывается в весьма затруднительном положении. Предположим — и это самый худший случай, — что практика, получаемая при использовании целостного метода, облегчает Джеку разучивание пьес с помощью частичного метода, а практика, получаемая при частичном методе, мешает заучиванию с помощью целостного. Пусть, как и раньше, эти влияния равны 5 единицам. При последовательности АББА качество исполнения пьес повысится на 5 единиц на второй пробе (частичный метод) и понизится на 5 единиц на четвертой пробе (целостный метод). Ясно, что эффекты переноса скомпенсированы не будут, и частичный метод получит преимущество. Для того чтобы это произошло, вовсе не обязательны разнонаправленные влияния, им достаточно быть просто неодинаковыми по величине. В данном случае мы имеем дело с систематическим смещением независимой переменной (метод заучивания) с другой переменной — последовательностью проб: либо условие А — условие Б, либо условие Б — условие А. Одна из проб условия А сопровождается влиянием Б на А, а одна из проб условия Б — влиянием А на Б. И беда в том, что экспериментатор не знает, какой вид влияния имеет место. Все, что у него есть, — это четыре показателя качества исполнения пьес, на каждое из которых воздействуют к тому же факторы времени, а иногда (как в данном эксперименте) — еще и факторы задачи.

Не слишком изменится эта ситуация и при регулярном чередовании проб. Каждая проба условия Б следует за пробой условия А, и наоборот. Если влияния асимметричны, то систематическое смещение независимой переменной будет не в половине проб, как в схеме позиционного уравнивания, а в каждой пробе (кроме первой). И вновь нет практически никаких средств для определения асимметричности этого переноса.

При использовании случайной последовательности примерно половина проб одного условия предшествует пробам другого условия. Возникает хоть какая-то возможность определить само наличие влияний последовательности и их асимметричный характер. Например, для каждого из следующих сочетаний проб: условие А предшествует условию Б, А не предшествует Б, Б предшествует А, Б не предшествует А — можно получить отдельное значение зависимой переменной. Различие между первыми двумя значениями позволит обнаружить величину влияния условия А на усло-

вие Б, а различие между двумя вторыми значениями — величину влияния Б на А. Зная эти величины, можно позаботиться об устранении систематического смещения: определяя значения зависимой переменной при каждом из условий, нужно вычитать соответствующие величины эффектов последовательности.

Следует заметить, что наше обсуждение не дает полного представления о последствиях предъявления обоих условий независимой переменной одному и тому же испытуемому. Здесь возможны влияния более общего характера. Например, целостный метод заучивания может становиться менее эффективным только в сочетании с частичным. По контрасту он может показаться испытуемому слишком утомительным. А если бы применялся один метод, контраста бы не было. Для определения подобных влияний также нет никаких практических средств. Кроме того, в эксперименте с одним испытуемым ни одна из указанных схем не устраняет возможности асимметричного переноса. Именно эти влияния следует признать самым серьезным источником систематического смещения независимой переменной.

Более того, он является и наиболее *общим* видом систематического смещения. Если между двумя условиями независимой переменной существуют асимметричные эффекты последовательности, то они скажутся в любом эксперименте, сравнивающем эти условия. Влияние предубеждений экспериментатора будет сказываться только в данном конкретном эксперименте, в другом эксперименте оно может радикально измениться, если новый экспериментатор имеет противоположные предубеждения. Точно так же смещение с факторами времени и факторами задачи при использовании короткой позиционно уравненной последовательности будет существенно меняться от эксперимента к эксперименту, как и неоднородные эффекты последовательности.

КРАТКОЕ ИЗЛОЖЕНИЕ

Мало иметь заранее подготовленный план эксперимента и хорошо зарегистрировать его ход. Не всякий план является удачным. Реальный эксперимент можно оценить по его близости к эксперименту *безупречному*. Конечно, на практике безупречность недостижима. Значение идеи безупречного эксперимента состоит в том, что она дает образец для оценки реальных экспериментов и точного определения их недостатков. В данной главе мы пользовались этим образцом при сравнении экспериментов, описанных в главе 1, с менее удачными планами проведения тех же экспериментов.

Одним из видов безупречного эксперимента является *идеальный* эксперимент, в котором одному и тому же испытуемому в одно и то же время предъявляются разные условия независимой переменной. Главное в нем — это обеспечение неизменности всех побочных факторов. Другим видом безупречного эксперимента является *бесконечный*, т. е. постоянно продолжающийся, эксперимент. Центральное значение имеет здесь достаточно боль-

шое количество данных. Третий вид безупречного эксперимента — эксперимент полного *соответствия*, все обстоятельства проведения которого суть те же самые обстоятельства, на которые будут распространяться его выводы. В нем важно то, что дополнительные факторы нужно сохранять не просто на неизменном, но и на адекватном уровне. При сравнении первоначально описанных экспериментов с их менее удачными вариантами было обнаружено, что во всех случаях оригинальные планы были ближе к одному из видов безупречного эксперимента.

В любом реальном эксперименте не все полученные данные можно считать достоверными. Однако нужно стремиться получить результаты (основу для будущих выводов), как можно более близкие к результатам безупречного эксперимента. Чем ближе реальный эксперимент к безупречному, безошибочному и по плану, и по процедурам, тем лучше он репрезентирует, или *представляет* его. Во всех случаях сравнения оригинальных экспериментов с их менее удачно спланированными вариантами оказалось, что именно оригинальные лучше представляют безупречный эксперимент. В зависимости от того, насколько реальные эксперименты репрезентируют безупречный, они бывают более или менее валидными. Различают два вида *валидности*. Первый называется *внутренней* валидностью. Речь идет о таком планировании эксперимента, при котором можно получить то же отношение между независимой и зависимой переменными, что и в идеальном или бесконечном экспериментах, т. е. об устранении побочных влияний. Если же эксперимент по своему проекту позволяет получить те же результаты, что и эксперимент полного соответствия, то он обладает *внешней* валидностью. Эксперимент, страдающий недостатком внутренней валидности, можно назвать *неудачным*, несостоятельным, а эксперимент, которому недостает внешней валидности, — *неадекватным*. Примером последнего служит эксперимент, при проведении которого уровень значимой дополнительной переменной не соответствует ее реальному уровню. Было установлено, наконец, что даже когда эксперимент удачно спланирован и успешно проведен, нет полной гарантии того, что полученные в нем результаты подобны результатам безупречного эксперимента. И наоборот, неудачно спланированный эксперимент может дать корректные результаты. Хотя, конечно, у хорошо спланированного эксперимента таких шансов больше.

Особое внимание в этой главе мы уделили внутренней валидности — главному требованию к любому эксперименту. Существует целый ряд факторов, затрудняющих достижение внутренней валидности. В эксперименте с одним испытуемым это прежде всего всевозможные изменения, происходящие с течением времени. Идеальный эксперимент неосуществим, нельзя одновременно предъявить одному и тому же испытуемому различные условия независимой переменной. На практике двух идентичных проб не существует; независимо от того, предъявляются они одному испытуемому или разным. Это связано, в частности, с теми побочными факторами, уровни которых могут изменяться. Некоторые из этих факторов можно зафиксировать и проконтролировать. Однако в большинстве случаев связанная с ними нестабильность результатов от пробы к пробе неизбежна. Как продолжительные, так и кратковременные колебания такого рода являются скорее правилом, чем исключением. Некоторые из них связаны с непостоянством побочных влияний, которые можно определить, но нельзя проконтролировать. Даже сама незави-

симая переменная не всегда остается неизменной в различных пробах. Непостоянство же зависимой переменной может быть связано также с особенностями изучаемого поведения и с его измерениями в эксперименте. Все эти факторы мы объединили под названием «*факторы времени*».

Помимо них мы описали еще три источника нарушения внутренней валидности. Некоторые эксперименты, особенно связанные с научением, требуют применения различных задач для разных условий, и тогда «*факторы задачи*» становятся одним из таких источников. Далее, в любом эксперименте, где разные условия предъявляются одному и тому же испытуемому, существуют *эффекты последовательности*, т. е. влияния ранее предъявленного условия на предъявленное позже. И еще один источник нарушения внутренней валидности — это *предубеждение экспериментатора* о преимущество одного из условий независимой переменной.

Все эти угрозы внутренней валидности в случае их неустранения приведут к одному из двух следствий. Первое — *ненадежность* эксперимента. Она возникает в том случае, если при большом разбросе данных проведено слишком мало проб. Здесь у нас нет полной уверенности в том, что при повторении эксперимента будут получены те же самые результаты. Второе — *систематическое смещение*, когда каждое из условий независимой переменной неразрывно связано со своим уровнем одной из других переменных, и это нарушает внутреннюю валидность.

В главе были обсуждены различные схемы эксперимента с одним испытуемым и то, насколько успешно позволяют они устранить указанные недостатки. Какая бы схема ни применялась, неотъемлемой частью эксперимента является первичный контроль за побочными факторами. Это — организация и протоколирование эксперимента, стабилизация известных побочных переменных, точность экспериментальных процедур и необходимое количество проб.

Схема случайной последовательности особенно удобна в тех экспериментах, где для каждого из условий можно применить большое количество проб. Валидность зависит от числа проб, необходимого для достижения высокой надежности эксперимента. Использование этой схемы исключает возможность всех эффектов последовательности, за исключением асимметричных влияний. Впрочем, последние сохраняются и при использовании других схем.

Если в эксперименте можно использовать не так много проб, то лучше предъявлять их не в случайном порядке, а применить схему регулярного чередования. Надежность эксперимента вновь будет зависеть от соответствия количества проб разбросу показателей. Экспериментатору следует обратить особое внимание на возможные влияния событий, происходящих через равные промежутки времени. Эти влияния могут давать определенное преимущество одному из исследуемых условий и, таким образом, приводить к систематическому смещению.

Схема позиционно уравненной последовательности применяется при относительно небольшом количестве проб (или блоков проб). Надежность зависит от адекватности выбора проб или их блоков изучаемому поведению. Здесь контроль за систематическими влияниями факторов, изменяющимися от пробы к пробе, связан с предположением о линейном характере этих изменений.

Схемы различаются также и по тому, насколько успешно они позволяют преодолеть различия задач. При использовании позиционно уравненной последовательности все задачи разделяют на пары, стараясь подобрать в каждую пару задачи примерно одинаковой трудности. Если это не удастся полностью, то внутренняя валидность эксперимента пострадает из-за неизбежного для данной схемы систематического смещения независимой переменной с факторами задачи. Если используются случайная последовательность или регулярное чередование условий, т. е. число проб достаточно велико, то задачи (или пары задач) можно предъявлять в случайном порядке. Тогда систематического смещения с факторами задачи не происходит. Однако высокая степень изменчивости этих факторов снижает надежность эксперимента.

К систематическому смещению приводят также эффекты последовательности проб. При использовании схемы позиционного уравнивания внутренней валидности угрожает неоднородность влияний ранних и поздних экспериментальных проб. Труднее всего устранить такие случаи систематического смещения, когда влияния последовательности взаимно асимметричны, т. е. влияние условия А на условие Б отличается от влияния условия Б на условие А. И это может произойти в любом эксперименте, где различные условия независимой переменной предъявляются одному и тому же испытуемому.

Контрольные вопросы и задания

1. Как еще можно оценить эксперимент, кроме оценки качества организации его хода и протоколирования?

2. Покажите, как с помощью понятия идеального эксперимента можно определить, что один способ проведения эксперимента лучше другого.

3. Почему для опытов Йоко с томатным соком образцом безупречно-го эксперимента служит бесконечный, а не идеальный эксперимент?

4. Как внутренняя валидность связана с безупречным экспериментом?

5. Можно ли считать хорошим эксперимент, который не вполне представляет эксперимент полного соответствия?

6. Опишите основные факторы, затрудняющие достижение внутренней валидности эксперимента.

7. Почему вопрос о различии задач касался прежде всего эксперимента Джека с заучиванием фортепьянных пьес, а не двух других экспериментов?

8. Покажите различие между ненадежностью и систематическим смещением.

9. Как вы определите, что в эксперименте лучше применить схему регулярного чередования, чем схему случайной последовательности?

10. Сравните возможность систематического смещения при использовании схемы позиционного уравнивания и двух других схем.

11. В чем различие между систематическим смещением, которое может меняться от эксперимента к эксперименту при исследовании какой-то проблемы, и систематическим смещением, которое может произойти во всех экспериментах, направленных на решение данной проблемы?

СТАТИСТИЧЕСКОЕ ПРИЛОЖЕНИЕ: ОПРЕДЕЛЕНИЕ ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ И ПАРАМЕТРОВ НА ОСНОВЕ ВЫБОРОЧНЫХ СТАТИСТИК; СРЕДНЕЕ И СТАНДАРТНОЕ ОТКЛОНЕНИЕ

Определение среднего в популяции (генеральной совокупности)

В эксперименте по определению времени реакции, описанном в приложении к главе 1, были взяты результаты действительного эксперимента. Предполагалось, что они *представляют* такие данные, которые могли бы быть получены в эксперименте с полной внутренней валидностью. Так, среднее время реакции на световой сигнал по 17 пробам представляло среднее, которое можно было бы получить в эксперименте с *неограниченным* числом проб.

Мы используем среднее для ограниченной выборки проб, чтобы сделать вывод о достаточно большой (вплоть до неограниченной) популяции проб. Такая популяция называется генеральной совокупностью. Среднее по генеральной совокупности таких, например, данных, как ВР, обозначается M_X . Такую характеристику генеральной совокупности называют *параметром*. Среднее, действительно вычисленное нами для данной выборки, называется *статистикой*, и обозначается M_X . Является ли статистика M_X наилучшей оценкой параметра M_X , которую мы можем получить на основе нашей выборки проб? Ответ — без доказательства — да. Но прежде чем вы решите, что это всегда так, давайте перейдем к стандартному отклонению, где дело обстоит иначе.

Вычисление стандартного отклонения

Обычно помимо среднего значения оценок мы хотим знать еще кое-что, а именно, какова *несистематическая* вариация оценок от пробы к пробе. Наиболее распространенный способ измерения несистематической вариации состоит в вычислении стандартного отклонения.

Для этого вы определяете, насколько каждая оценка (т.е. X) больше или меньше среднего (M_X). Затем вы возводите в квадрат каждую разность ($X - M_X$) и складываете их. Вслед за этим вы делите эту сумму на N число проб. Наконец, вы извлекаете квадратный корень из этого среднего.

Это вычисление представлено формулой с использованием символа σ_X (малая греческая буква сигма) для обозначения стандартного отклонения:

$$\sigma_X = \sqrt{\frac{\sum (X - M_X)^2}{N}}. \quad (2.1)$$

Эту формулу можно сократить, введя маленькое x для обозначения $(X - M_X)$. Тогда формула выглядит так:

$$\sigma_X = \sqrt{\frac{\sum x^2}{N}}. \quad (2.1 \text{ а})$$

Давайте выпишем данные по условию А из приложения к главе I и одновременно произведем по ним вычисления, указываемые формулой для σ_X .

Проба	X	M_X	$X - M_X$ или x	x^2
1	223	185	+38	+1444
2	184	185	-1	+1
3	209	185	+24	+576
4	183	185	-2	+4
5	180	185	-5	+25
6	168	185	-17	+289
7	215	185	+30	+900
8	172	185	-13	+169
9	200	185	+15	+225
10	191	185	+6	+36
11	197	185	+12	+144
12	188	185	+3	+9
13	174	185	-11	+121
14	176	185	-9	+81
15	155	185	-30	+900
16	165	185	-20	+400
17	163	185	-22	+484
			$\sum x^2$	+5808

Поскольку

$$\sigma_A = \sqrt{\frac{\sum x_A^2}{N}},$$

$$\text{то } \sigma_A = \sqrt{\frac{5808}{17}} = \sqrt{341,7} = 18,5 \text{ мс.}$$

Оценка стандартного отклонения генеральной совокупности

Для определения среднего генеральной совокупности, которое могло бы быть получено в бесконечном эксперименте, наилучшей оценкой фактически было среднее по выборке. Иначе обстоит дело со стандартным отклонением. В любом наборе реальных проб имеет место меньшее число результатов с очень высокими или очень низкими значениями, чем в генеральной совокупности. А поскольку стандартное отклонение является мерой разброса оценок, то его величина, определенная на основе выборки, всегда меньше параметра генеральной совокупности $\bar{\sigma}_X$.

Более точная оценка стандартного отклонения для генеральной совокупности находится по формуле

$$S_X = \sqrt{\frac{\sum x^2}{N-1}} \quad (2.2)$$

или

$$\sigma_X = \sqrt{\frac{N}{N-1}}. \quad (2.2 \text{ а})$$

Для наших числовых данных

$$S_A = \sqrt{\frac{5808}{16}} = \sqrt{363,0} = 19,1 \text{ мс.}$$

В некоторых экспериментах высказывается гипотеза, что поведение в одном условии более *вариативно*, чем в другом. Тогда целесообразнее сравнивать стандартные отклонения, а не средние. Если для обоих условий N одно и то же, можно сравнивать между собой *сигмы*.

Однако когда N различны, сигма для условия с меньшим N дает более заниженную оценку такого параметра генеральной совокупности, как стандартное отклонение. Поэтому следует сравнивать два S .

Таблица, которая приводится ниже, поможет вам запомнить эти положения и формулы.

	Среднее отклонение	Стандартное отклонение
Параметрические характеристики генеральной совокупности (г.с.)	$\bar{M}_X = \frac{\sum X_{гс}}{N_{гс}}$	$\bar{\sigma}_X = \sqrt{\frac{\sum x_{гс}^2}{N_{гс}}}$
Статистические характеристики выборки	$M_X = \frac{\sum X}{N}$	$\sigma_X = \sqrt{\frac{\sum x^2}{N}}$
Оцениваемый параметр генеральной совокупности	$M_X = \frac{\sum X}{N}$	$S_X = \sigma_X \sqrt{\frac{N}{N-1}},$ или $S_X = \sqrt{\frac{\sum x^2}{N-1}}$

Задача: Вычислите σ_X и S_X для условия Б.

Ответ: $\sigma_B = 15,9$; $S_B = 16,4$.

Статистическая таблица 1

Числа от 1 до 100 с их квадратами и корнями квадратными

Число	Квадрат	Корень квадратный	Число	Квадрат	Корень квадратный
1	1	1,000	14	196	3,742
2	4	1,414	15	225	3,873
3	9	1,732	16	256	4,000
4	16	2,000	17	289	4,123
5	25	2,236	18	324	4,243
6	36	2,449	19	361	4,359
7	49	2,646	20	400	4,472
8	64	2,828	21	441	4,583
9	81	3,000	22	484	4,690
10	100	3,162	23	529	4,796
11	121	3,317	24	576	4,899
12	144	3,464	25	625	5,000
13	169	3,606	26	676	5,099

Продолжение статистической табл. 1

Число	Квадрат	Корень квадратный	Число	Квадрат	Корень квадратный
27	729	5,196	57	3249	7,550
28	784	5,292	58	3364	7,616
29	841	5,385	59	3481	7,681
30	900	5,477	60	3600	7,746
31	961	5,568	61	3721	7,810
32	1024	5,657	62	3844	7,874
33	1089	5,745	63	3969	7,937
34	1156	5,831	64	4096	8,000
35	1225	5,916	65	4225	8,062
36	1296	6,000	66	4356	8,124
37	1369	6,083	67	4489	8,185
38	1444	6,164	68	4624	8,246
39	1521	6,245	69	4761	8,307
40	1600	6,325	70	4900	8,367
41	1681	6,403	71	5041	8,426
42	1764	6,481	72	5184	8,485
43	1849	6,557	73	5329	8,544
44	1936	6,633	74	5476	8,602
45	2025	6,708	75	5625	8,660
46	2116	6,782	76	5776	8,718
47	2209	6,856	77	5929	8,775
48	2304	6,928	78	6084	8,832
49	2401	7,000	79	6241	8,888
50	2500	7,071	80	6400	8,944
51	2601	7,141	81	6561	9,000
52	2704	7,211	82	6724	9,055
53	2809	7,280	83	6889	9,110
54	2916	7,348	84	7056	9,165
55	3025	7,416	85	7225	9,220
56	3136	7,483	86	7396	9,274

Окончание статистической табл. 1

Число	Квадрат	Корень квадратный	Число	Квадрат	Корень квадратный
87	7569	9,327	94	8836	9,695
88	7744	9,381	95	9025	9,747
89	7921	9,434	96	9216	9,798
90	8100	9,487	97	9409	9,849
91	8281	9,539	98	9604	9,899
92	8464	9,592	99	9801	9,950
93	8649	9,644	100	10000	10,000

Глава 3

ЭКСПЕРИМЕНТЫ, КОТОРЫЕ «УЛУЧШАЮТ» РЕАЛЬНЫЙ МИР¹

По какой причине высококвалифицированный летчик разбивает торговый реактивный самолет во время обычной посадки в аэропорту спокойной ясной ночью? Такой вопрос задали себе два психолога-экспериментатора Конрад Крафт и Чарльз Элворт вместе с компанией «Боинг». При этом они имели в виду не единственную аварию. Как это ни удивительно, но почти каждый пятый несчастный случай в авиации происходит «во время безопасных на первый взгляд ночных посадок, выполняемых с помощью визуального контроля» (Kraft C. L., Elworth C. L., 1969, с. 2).

Для того чтобы квалифицированно ответить на этот вопрос — и тем самым положить начало практическому решению проблемы, — Крафт и Элворт проанализировали статистику несчастных случаев во время ночных посадок. И им удалось найти ключ к разгадке: по большей части это случается в аэропортах, расположенных несколько ниже по сравнению с близлежащими городами. Примером может служить аэропорт, расположенный на берегу большого озера. Самолет заходит на посадку над водой и приземляется в аэропорту, который лишь немного выше ее уровня. Уровень же, на котором расположен город, постепенно поднимается за аэродромом.

Исследователи выдвинули такую экспериментальную гипотезу. Способ визуальной ориентации, которым пользуется пилот при посадке, позволяет ему вести самолет по нужной траектории только тогда, когда город и аэропорт расположены на одном уровне. Если же город расположен выше, то фактическая траектория посадки оказывается слишком низкой.

Такую гипотезу можно проверить экспериментально. Для этого пилоту нужно посадить самолет в двух разных аэропортах: один из них должен быть расположен на одном уровне с городом, а другой — немного ниже. Однако такой эксперимент, дублирующий реальный мир, не будет удовлетворительной проверкой гипотезы. Основная причина этого — слишком большое количество факторов, не связанных с наклоном земной поверхности, которые трудно устранить. Две реальные ситуации могут различаться по расположению наземных огней, условиям видимости, силе воздушного

¹ Перевод В. В. Петухова.

потока и т. п. Кроме того, подобный эксперимент опасен для пилота. Не исключено поэтому, что в целях собственной безопасности пилот не ограничится только зрительным наблюдением, а будет ориентироваться и по показаниям приборов. Это же, в свою очередь, также понизит надежность результатов эксперимента. Ведь исследователи хотят узнать, как будет совершать посадку пилот, пользуясь только визуальной информацией — наземными огнями, которые расположены либо на горизонтальной поверхности либо на поверхности, идущей в гору.

Таким образом, эксперимент, дублирующий реальный мир, в данном случае неприменим. Необходим эксперимент, который бы «улучшал» этот мир (разумеется, в исследовательских целях). Особая экспериментальная ситуация позволила бы исследователю достичь единообразия в расположении наземных огней, условиях видимости и силе воздушного потока, а также устранить возможность использования высотомера, не подвергая пилота опасности. Именно такой эксперимент провели Крафт и Элворт. Это первый из трех экспериментов, которые мы опишем в настоящей главе. В каждом из них та реальная ситуация, на которую распространяются экспериментальные результаты, определенным образом «улучшается».

Такие эксперименты можно назвать *искусственными*. Они проводятся в том случае, когда простое воспроизведение реальной ситуации не позволяет сделать эксперимент внутренне валидным. Однако возникает вопрос: можно ли применять результаты такого эксперимента к реальности? Какие гарантии имелись у Крафта и Элворта, чтобы считать свои лабораторные наблюдения полностью *соответствующими* реальным событиям во время посадки самолета в аэропорту? Таким образом, на первый план выступает проблема внешней валидности. В первых экспериментах, обсуждаемых в нашей книге, она почти не затрагивалась, поскольку те эксперименты просто дублировали реальный мир. Теперь, когда для достижения высокой внутренней валидности эксперимента реальность нужно улучшить, оказывается под сомнением его внешняя валидность. В одном мы выигрываем, в другом — теряем.

Пока вы вряд ли имеете возможность проводить эксперименты, в которых улучшается реальный мир. Но, изучая их, вы познакомитесь с новыми проблемами, которые не возникают в экспериментах первого типа, и научитесь планировать искусственные эксперименты, обладающие внешней валидностью.

**Вопросы, на которые вы должны суметь ответить,
прочитав эту главу**

1. При каких обстоятельствах возникает необходимость в эксперименте, который улучшает, а не дублирует реальный мир?

2. Как можно повысить внутреннюю валидность, сделав выбор в пользу эксперимента, улучшающего реальный мир?

3. Каковы проблемы репрезентативности по отношению к внешней валидности?

4. В какой мере нужно воспроизводить реальность в искусственном эксперименте?

Эксперимент 1:

НОЧНЫЕ ПОСАДКИ САМОЛЕТОВ ПОД ВИЗУАЛЬНЫМ КОНТРОЛЕМ

Экспериментальная гипотеза

Чтобы понять основания экспериментальной гипотезы, выдвинутой Крафтом и Элвортом, нам нужно разобраться, каким образом пилот заходит на посадку в ночных условиях, руководствуясь зрительным восприятием наземных огней.

Сама посадка состоит в следующем. Сначала пилот должен довольно круто снизиться, а затем продолжать снижаться постепенно, уменьшая угол приземления до тех пор, пока самолет не окажется летящим параллельно земле (над взлетно-посадочной полосой). Всего этого можно достичь в том случае, если самолет будет снижаться по окружности большого радиуса, т.е. находясь как бы на конце маятника, подвешенного высоко в небе. Именно такую посадочную траекторию вы видите на рис. 3.1. Изображенный ее отрезок начинается примерно за 10 миль до аэропорта (немногим более 50 000 футов), когда самолет находится на высоте 5000 футов.

Каким же образом пилот может придерживаться этой траектории, пользуясь наличной визуальной информацией? Среди огней близлежащего города он выбирает и фиксирует взглядом пару световых точек, расположенных вдоль линии полета. Одна точка находится довольно близко, другая — далеко. В поле зрения пилота дальняя точка будет выше ближней. Чтобы убедиться в этом, переведите взгляд с предмета, лежащего у ваших ног, на какой-нибудь дальний предмет. Заметьте, что для этого вам придется немного приподнять голову. Если во время посадки летчик будет видеть, что расстояние между ближней и дальней световыми точками по вертикали примерно одно и то же, то самолет будет следовать по указанной траектории. Этот прием называют «зрительным полетом».

Все это можно изложить иначе. Пилот совершает посадку таким образом, что *угол его зрения* между двумя точками остается постоянным. Для определения этого угла нужно провести прямые линии от ближней и дальней точек до глаза летчика. Угол между

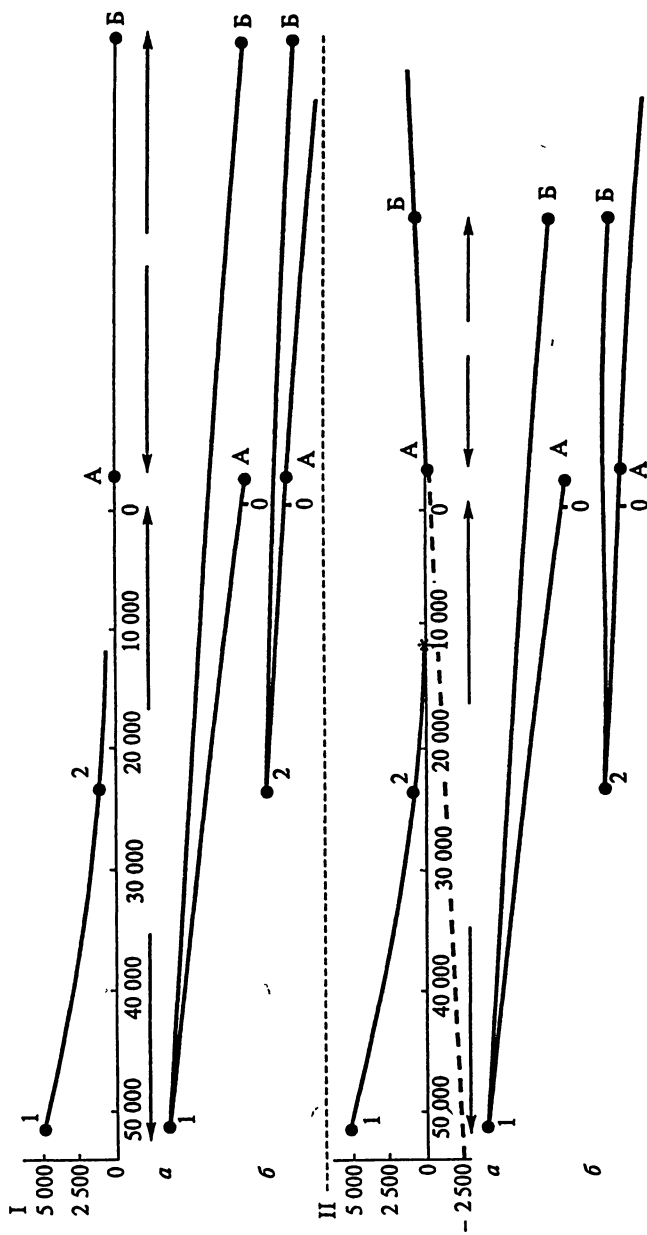


Рис. 3.1. Траектория посадки самолета под визуальным контролем. Ось абсцисс — расстояние до аэропорта (в футах) и поверхность города; ось ординат — высота полета (в футах). А — ближняя световая точка, Б — дальняя световая точка. I. Траектория посадки по ближней и дальней световым точкам, расположенным на горизонтальной поверхности, а и б. В обоих положениях траектории угол зрения остается одним и тем же. II. Траектория посадки по световым точкам, расположенным на наклонной территории, (а) и (б). В обоих положениях траектории угол зрения — один и тот же. Звездочкой отмечено место возможной аварии самолета, если он будет продолжать полет по указанной траектории

двумя полученными отрезками и будет углом зрения. Если в поле зрения пилота расстояние между двумя точками по вертикали остается неизменным, то угол зрения тоже будет одним и тем же по всей траектории посадки.

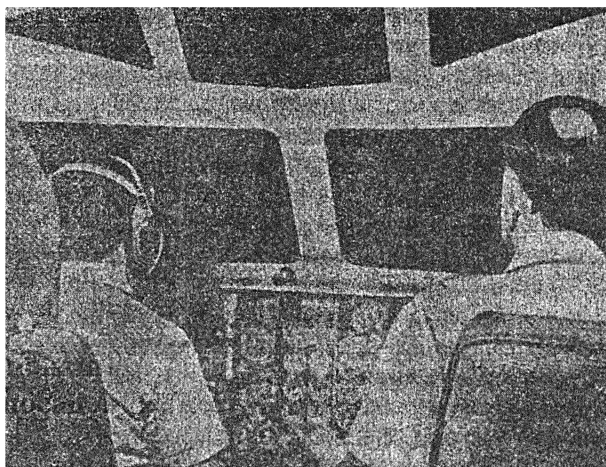
В верхней части рис. 3.1, представляющей посадку над горизонтальной территорией, показано, что когда летчик летит по требуемой траектории, угол зрения действительно не изменяется. Небольшой угол, видимый из начального положения 1 и показанный в (а), — это тот же самый угол, что показан в (б), из положения 2, просто он несколько дальше по траектории. Из этих двух точек, так же как и из всех других точек траектории, расстояние между ближней и дальней световыми точками в поле зрения летчика будет выглядеть одинаковым.

Однако этот прием гарантирует нормальное приземление только в том случае, если выбранные световые точки расположены на горизонтальной поверхности. В нижней части рис. 3.1 изображена посадочная траектория, когда пилот ориентируется по наземным огням города, который расположен немного выше по сравнению с аэропортом. Угол зрения между ближней и дальней точками, показанный в (а) из положения 1 траектории, является здесь тем же самым, что и для посадки на горизонтальную территорию. Не изменяется он и в положении 2, как показано в (б). Однако это положение уже гораздо ближе к земле. Если летчик будет продолжать пользоваться приемом «зрительного полета», то самолет потерпит аварию, не достигнув аэропорта. Летчик окажется жертвой оптической иллюзии. Он всегда воспринимает оба наземных огня как лежащие на горизонтальной плоскости. Пунктирная линия в нижней части рис. 3.1 — вот как представляет себе летчик ту территорию, над которой совершает посадку. Легко убедиться, что она значительно ниже реального уровня земли.

На бумаге данная гипотеза выглядит довольно убедительно. Посмотрим теперь, каким образом Крафт и Элворт проверили ее на практике.

Эксперимент

Тренажер. Для проведения эксперимента Крафт и Элворт использовали специальное устройство — *тренажер*. Это устройство имитирует кабину самолета, оборудованную всем необходимым. В кабине расположены кресла для первого и второго пилотов, все ручки и рычаги для обычного управления самолетом, панель с набором шкал и индикаторов, а также — на ветровом стекле — вид города и аэропорта. Когда пилот «ведет» тренажер, изменения в предъявляемом ему зрительном обзоре точно такие же, какие он мог бы наблюдать в реальном полете. В принципе



а



б

Рис. 3.2. Тренажер (а) и наземные огни, предъявляемые пилотам на ветровом стекле (б) (Kraft C.L., Elworth C.L., 1969)

пилот может совершать посадку, пользуясь либо только зрительными наблюдениями за землей, либо только шкалами, либо тем и другим вместе. В обсуждаемом эксперименте пилоту предъявлялась визуальная информация, а шкала, показывающая высоту полета, высотомер, ему не давалась. На рис. (фото) 3.2, а показан экипаж, который управляет тренажером, наблюдая за наземными огнями. На рис. 3.2, б изображен вид сверху на аэропорт и город. В эксперименте это изображение было еще более

реалистичным — множество цветных сверкающих огней. Как вы, наверное, догадались, устройство такого тренажера довольно сложно и включает в себя компьютер. В дополнение к уже описанным функциям тренажер позволяет осуществлять запись управляющих движений пилота, а также все изменения в показаниях шкальных приборов. В исследовательских целях проводилась также постоянная запись показаний высотомера, который от пилота был скрыт.

Методика

Задача. Пилоту нужно посадить самолет в «аэропорту», двигаясь по правильной траектории, в условиях, показанных на рис. 3.2. «Город» предьявляется ему либо как расположенный на одной плоскости, либо приподнятый по сравнению с аэропортом под углом 3° . (Это такой же подъем, какой был показан в нижней части рис. 3.1.) Конкретный способ выполнения этой задачи пилот выбирал сам, но должен был стараться лететь со скоростью 180 миль в час на высоте 5000 футов за 10 миль до аэропорта и со скоростью 120 миль в час — на высоте 1250 футов за 4,5 мили от него. (На рис. 3.1 эти положения траектории полета обозначены цифрами I и II.) Напоминаем, что пилот пользовался только зрительной информацией, предьявляемой через ветровое стекло; высотомера у него не было, и он не мог точно определить, как высоко находится над землей.

Помимо объективной регистрации «истинной» высоты полета эта высота периодически оценивалась самим пилотом. Оценки записывались на магнитофон. Чтобы приблизить ситуацию эксперимента к условиям реального полета, пилота просили также определять местоположение еще одного самолета, находящегося в поле его зрения, и периодически сообщать о нем «на землю».

Процедура. С каждым участником эксперимента было проведено десять проб. Помимо основной независимой переменной в различных пробах изменялись некоторые другие факторы: расположение наземных огней, степень освещения, начальная высота полета, его общая протяженность.

Результаты

На рис. 3.3 изображены средние траектории посадки по данным 12 пилотов, принявших участие в эксперименте. Для более ясного представления полученных различий шкала высоты несколько расширена. Легко убедиться в том, что траектории посадки на горизонтальную поверхность во многом напоминают предсказанную на рис. 3.1 (верхняя часть), когда пилот пользуется приемом

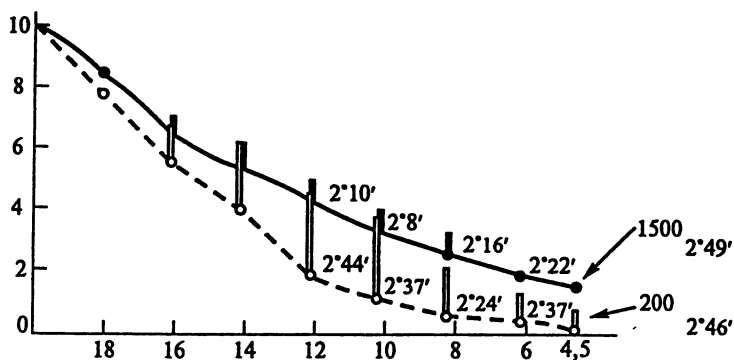


Рис. 3.3. Средние траектории полета при посадке по данным 12 пилотов на тренажере. Ось абсцисс — расстояние до места посадки (в милях); ось ординат — высота «полета» (в тысячах футов); сплошная линия — горизонтальная территория, пунктирная — наклонная территория; верхние границы вертикальных столбиков соответствуют средним субъективным оценкам высоты полета; числами отмечены величины угла зрения пилота между ближней и дальней световыми точками; справа — высота полета и угол зрения в конечных точках траекторий

«зрительного полета» по двум подходящим световым точкам. Если поверхность поднимается (от аэропорта к городу) на 3° , траектория движения самолета также подобна предсказанной на рис. 3.1 (нижняя часть). Правда, она не такая крутая, и самолет не «разбивается». На рис. 3.3 приведены субъективные оценки высоты полета. Можно видеть, что даже при посадке на горизонтальную поверхность наблюдается тенденция переоценивать высоту. Пилоту кажется, что самолет летит выше, чем на самом деле. Если город расположен выше аэродрома, это преувеличение становится очень существенным.

Заметьте, что за 10 миль до аэропорта пилотам кажется, что они находятся на высоте почти 4000 футов, хотя фактически эта высота составляет менее 2000 футов!

Обсуждение и выводы. Общее предположение подтвердилось: расположен ли город на одном уровне с аэропортом или немного выше, летчик совершает посадку примерно по одной и той же траектории. Подобно другим оптическим иллюзиям, описанный эффект нельзя устранить даже с помощью специальных инструкций. «Пилоты практически всегда воспринимают наземные огни города и аэропорта как лежащие на горизонтальной поверхности, хотя реально они могут находиться на разных уровнях» (Kraft C. L., Elworth C. L., 1969, с. 4). Поскольку данную иллюзию нельзя преодолеть, летчику не стоит полагаться на «зрительный полет», а нужно следить за показаниями высотомера.

ПЕРВЫЙ СПОСОБ УЛУЧШЕНИЯ РЕАЛЬНОСТИ: УСТРАНЕНИЕ СИСТЕМАТИЧЕСКОГО СМЕШЕНИЯ

Предположим, что Крафт и Элворт решили провести свой эксперимент не на тренажере, а просто дублируя реальный мир. Для того чтобы сравнить ночные посадки самолета в условиях зрительного наблюдения, они прежде всего выбрали бы два аэропорта, различных по относительному подъему находящихся рядом городов. Конечно, они постарались бы, насколько это возможно, добиться постоянства всех дополнительных факторов. Они попытались бы подобрать похожие поверхности, будь то вода или земля, над которыми будут совершать посадку самолеты. Они постарались бы найти города, примерно равные по площади, с похожим расположением огней, уравнивать условия видимости и т. п. И в конце концов они, наверное, убедились бы, что найти два аэропорта, полностью подобных по всем указанным параметрам, просто невозможно. И это действительно так. Ведь даже если бы они подобрали два таких аэропорта, все равно в ходе испытаний возникли бы новые различия. Нельзя рассчитывать на то, что количество воздушного транспорта, характер инструкций с контрольной вышки и т. д. всегда будут одними и теми же.

Таким образом, эксперимент, дублирующий реальный мир, очень далек от *идеального эксперимента*. Даже там, где в обоих аэропортах все побочные переменные в принципе одинаковы, они никогда не будут такими же идентичными, как на тренажере, «улучшающем» реальный мир. Поэтому при проведении эксперимента в реальных аэропортах исследователи не могут быть уверенными в том, что различие в траекториях посадки объясняется лишь изменением наклона поверхности.

Побочные факторы могут помешать влиянию основной независимой переменной, могут даже полностью определить найденные различия. Другими словами, эксперимент, который дублирует реальный мир, в данном случае не обладает достаточной внутренней валидностью, поскольку не исключает *систематического смещения* независимой переменной (угла наклона) с различными побочными факторами.

Теперь мы можем сформулировать одно из оснований для проведения эксперимента, «улучшающего» реальный мир. Он необходим тогда, когда эксперимент, дублирующий реальность, не может достичь высокой *внутренней* валидности из-за систематического вмешательства побочных переменных. В описанном эксперименте исследователям удалось избежать этого посредством искусственной стабилизации последних.

Сейчас мы обсудим еще два эксперимента, в которых показаны другие преимущества искусственного эксперимента, повышающие его внутреннюю валидность. Эти эксперименты тоже явля-

ются убедительными примерами соответствующих способов улучшения. Процесс планирования каждого из них мы опишем более детально, чтобы понять, почему исследователь решил провести именно искусственный эксперимент. В отличие от эксперимента с ночными посадками самолетов они имеют значительно меньшее техническое оснащение и могли бы быть легко проделаны вами.

Эксперимент 2:

СПАСАТЕЛЬНЫЙ ПОИСК НА МОРЕ

Когда корабль терпит крушение и подает сигнал бедствия, нужно сделать все возможное для спасения людей — найти их и привезти на берег. Этим занимается специальная служба Береговой Охраны. В зависимости от обстоятельств в спасательных операциях используются небольшие катера, самолеты-амфибии или вертолеты, а иногда и все вместе. Правда, в большинстве случаев для поиска оставшихся в живых после крушения не нужны слишком утонченные средства вроде радара или сонара. Чтобы суметь заметить проблеск чего-то плавающего, сами спасатели должны *видеть* поверхность воды. А теперь давайте представим себе следующий диалог между членами экипажа Береговой Охраны в Санто-Томас, происходящий во время уборки палубы патрульного катера БО-99999.

Хокки. Что ты думаешь, дорогая, о приближающихся тренировках?

Дион. Одно слово — рутина. Уже который год мы пользуемся одним и тем же старьем. Я вовсе не уверена, что наши средства вполне подходят для поиска.

Хокки. О чем это ты?

Дион. Да взять хотя бы эти бинокли 7 × 50. По мне, так лучше смотреть просто так, а их все выкинуть за борт.

Хокки. Ты преувеличиваешь.

Дион. Да ты его надень. Во-первых, через это чудовище моря-то увидишь — всего ничего. Пока хорошо просмотришь хотя бы квадратную милю, пройдет целая вечность. Во-вторых, при бортовой да килевой качке нашего катера, да его непрерывных виляниях одно и то же место удержишь разве что на миг. В-третьих, определить, как далеко находится это место, и не надейся, хоть застрелись, — ведь вода то поднимается, то опускается. В-четвертых, достаточно немного в него посмотреть, как глаза у тебя уже лезут на лоб. И наконец, весит он целую тонну.

Хокки. Ха! Вот что бывает, когда слабый пол допускают к мужским занятиям. Ну ладно, а как быть с теми отличными старыми фильмами, где горизонт показывали через два больших смежных круга?

Ди о н. Береговая Охрана подчиняется транспортному министерству, и поэтому я воздержусь от каких-либо комментариев по этому вопросу. Слушай, а что если нам, вместо того чтобы тянуть лямку в бесконечных тренировках, не надоумить капитана Лауфтона провести специальную проверку: как лучше искать — с биноклями или без них?

Хо к и. А помнишь, как жестоко посмеялся капитан над Белделлом, когда тот предложил тренировать для поиска дельфинов?

Ди о н. Не беспокойся, я сумею его так обработать, что он подумает, будто все это пришло в голову ему самому.

Хо к и. Я тебе верю.

Ди о н. Вот что нам нужно сделать. Во-первых, каждый из нас должен провести поиск с биноклем и без него. Это будет несложно: нам понадобится всего один бинокль. Во-вторых, нам нужно набросать в воду много муляжей, причем потом немного подождать, чтобы никакая случайность не помогла нам их заметить. В-третьих, нам нужны муляжи разной величины, напоминающие по размеру плоты и люди. В-четвертых, мы должны проводить проверку в разное время дня, в разных погодных условиях и при различных расстояниях до муляжей. В-пятых...

Хо к и. И не надоест тебе все это перечислять! Ну хорошо, а из чего будут муляжи?

Ди о н. Это я сделаю! Плоты будут из старых надувных матрасов, а люди — из кусочков пенопласта.

Хо к и (воодушевляясь). Точно, вместо людей — кусочки розового пенопласта.

Ди о н. И коричневые тоже.

Хо к и (сглатывая слюну). А если этих муляжей будет много, как мы определим, что не потеряли их из виду?

Ди о н. Гм... Их будут опускать с вертолета через определенные промежутки времени. На это время нам будут завязывать глаза. Как только мы заметим какой-нибудь из них, парни на контроле поймают или загарпунят его. А потом посмотрим, кто из нас работал лучше.

Хо к и. А что будет с катером?

Ди о н. Он будет двигаться по замкнутому кругу или по квадрату — это все равно. Важно только, чтобы рулевой не слишком удалялся от цели. А кто-нибудь на борту будет следить за спуском муляжа и затем отмечать его местонахождение приблизительно каждые полминуты. При обнаружении муляжа каждый из нас также должен отметить место, где он находится, и время.

Хо к и. Ты говорила так долго и так быстро — как же мы все это запоем?

Ди о н. Запомним? Зачем запоминать?! Ведь любой проныраинспектор захочет знать, почему мы отклонились от инструкции и имели ли право проводить свою проверку. Для этого весь экспе-

риментальный материал мы изложим в особом вахтенном журнале. Я напишу оглавление, затем план эксперимента, потом представлю страницы протокола — нам, кстати, нужно два журнала, чтобы ты тоже мог вести наблюдения — и все будет как полагается, вплоть до резюме.

Х о к и. О, боже!

Но здесь мы прекратим этот маловероятный диалог и проанализируем план эксперимента.

План эксперимента

Поскольку сама идея и детали эксперимента уже указаны, в табл. 3.1 приведена лишь его формальная схема.

Таблица 3.1

Схема эксперимента: поиск в спасательных операциях на море (приведены четыре серии из восьми)

Условия:	
Сравнивается успешность поиска с биноклем 7 × 50 и без него. Опыты будут проводиться в течение четырех дней в октябре 1977 года, по возможности — двух солнечных и двух ненастных.	
1. Условия наблюдения (2). Через бинокль 7 × 50 и невооруженным глазом. Испытуемый 1 (с биноклем, первая половина эксперимента). Испытуемый 2 (с биноклем, вторая половина).	
2. Погодные условия (2). Солнечно и ненастно.	
3. Время дня (2). День (11.00—13.00) и вечер (16.00—18.00).	
4. Размер муляжа (2). Большой и малый.	
5. Расстояние до муляжа (3). 0,5; 1 и 4 мили.	
Серия А (день, солнечно)	Серия В (день, ненастно)
1. Малый муляж — 1 миля	1. Малый муляж — 4 мили
2. Большой муляж — 4 мили	2. Большой муляж — 0,5 мили
3. Большой муляж — 4 мили	3. Большой муляж — 1 миля
4. Большой муляж — 1 миля	4. Малый муляж — 1 миля
5. Малый муляж — 0,5 мили	5. Малый муляж — 4 мили
6. Малый муляж — 0,5 мили	6. Большой муляж — 0,5 мили
Серия Б (вечер, солнечно)	Серия Г (вечер, ненастно)
1. Большой муляж — 1 миля	1. Большой муляж — 1 миля
2. Малый муляж — 1 миля	2. Большой муляж — 4 мили
3. Малый муляж — 0,5 мили	3. Большой муляж — 1 миля
4. Большой муляж — 0,5 мили	4. Малый муляж — 0,5 мили
5. Большой муляж — 4 мили	5. Малый муляж — 4 мили
6. Малый муляж — 4 мили	6. Малый муляж — 0,5 мили

Сначала было намерение использовать четыре разных типа муляжей, опускать их по четырем секторам внутри траектории движения катера, проводить поиск в разное время, ну и, конечно, с помощью двух разных методов наблюдения. Но при подсчете оказалось: для того чтобы предъявить каждую из возможных комбинаций всех этих условий хотя бы по разу, потребовалось бы 192 опыта! Первоначальные планы экспериментов почти всегда бывают излишне грандиозны. Капитан Лауфтон признал этот план неосуществимым и сократил его. Точное местоположение спускаемого муляжа было признано несущественным — лишь бы он попадал внутрь траектории движения катера по его правому борту. Начинающие экспериментаторы решили также обойтись только двумя типами муляжей и двумя разными погодными условиями.

Но даже при всех этих ограничениях пришлось опускать муляжи 48 раз. Было проведено восемь экспериментальных серий, в середине дня (высокое солнце) и в поздний полдень (низкое солнце) соответственно. В первые четыре серии (А—Г) биноклем пользовался Хоки, а Дион наблюдала невооруженным глазом, в последующие четыре (Д—З) — наоборот. В каждой серии испытуемым предъявлялось шесть муляжей разной величины.

Анализ результатов

Все экспериментальные данные содержатся в лабораторном дневнике. Результаты Хоки показаны в табл. 3.2. В самой верхней строчке таблицы сравнивается результативность работы Хоки с биноклем и без него при предъявлении больших муляжей на расстоянии 0,5 мили в солнечную погоду в середине дня. При использовании бинокля цель была обнаружена за 3 мин, а без него — за 2,5. Эти, а также общие результаты в нижней части таблицы показывают, что наблюдение невооруженным глазом часто оказывалось более эффективным. На самом деле в 15 случаях муляж был быстрее обнаружен невооруженным глазом и только в 4 — с биноклем. Благоприятствовали поиску с биноклем только такие факторы, как солнечная погода и большие расстояния.

Можно отметить некоторые детали. Как и следовало ожидать, в солнечную погоду поиск в целом проходил успешнее, чем в ненастную. Небольшое различие в результатах в пользу поиска во второй половине дня объясняется, вероятно, наличием ярких бликов при низком солнце.

Анализ результатов Дион также свидетельствует о преимуществе наблюдения невооруженным глазом, хотя различие оказалось не столь высоким: 14 и 8. Дион не смогла обнаружить только два муляжа. Стоит заметить, что хотя для проведения эксперимента хватило бы и одного испытуемого, было очень полезно при-

Таблица 3.2

**Время поиска с помощью бинокля и невооруженным глазом для
сравниваемых условий (один испытуемый)**

Погодные условия	Время дня	Размер муляжа	Расстояние (мили)	Время поиска а)	
				с биноклем	невооруженным глазом
Солнечно	День	Большой	0,5	3,0	2,5
			1,0	6,0	6,0
			4,0	10,5	11,5
		Малый	0,5	4,5	2,0
			1,0	9,0	8,5
			4,0	12,0	14,0
	Вечер	Большой	0,5	2,0	1,0
			1,0	4,0	4,5
			4,0	9,5	9,0
		Малый	0,5	3,0	1,5
			1,0	6,5	6)
			4,0	8,0	9,5
Неясно	День	Большой	0,5	7,0	5,0
			1,0	10,0	9,5
			4,0	10,0	10,0
		Малый	0,5	9,5	6,0
			1,0	в)	13,0
			4,0	—	—
	Вечер	Большой	0,5	6,0	4,0
			1,0	11,5	10,0
			4,0	—	14,5
		Малый	0,5	9,0	7,0
			1,0	—	—
			4,0	г)	12,5
Число сравнений, в которых время поиска было меньше				4	15

Примечание: а) с точностью до 0,5 мин; б) наблюдение за посторонним предметом (вероятно, акулой) 2,5 мин; в) отметка «→» означает, что муляж не был обнаружен в течение 15 мин; г) невозможность поиска вследствие морской болезни.

влекать двух. Факт меньшего различия между двумя условиями наблюдения по результатам Дион (по сравнению с результатами Хоки) говорит о том, что здесь во второй половине эксперимента работа шла лучше: либо легче было обнаруживать муляжи, либо испытуемый был более опытен. А если бы в опытах участвовал только Хоки, мы и не подозревали бы о возможности подобных объяснений различия результатов. Следовательно, хотя результаты исследования в целом и не распространяются на других людей, результаты одного испытуемого являются своего рода контролем для результатов другого.

Краткое изложение эксперимента

Сравнивались две стратегии поиска: с использованием бинокля 7 × 50 и без него (наблюдение невооруженным глазом). Эффективность этих стратегий оценивалась не во время обычных поисковых операций, а в специальной искусственной ситуации, где испытуемым в различных условиях предъявляли большое количество муляжей. Эксперимент проводился в течение четырех дней, по две серии в день при соответствующих погодных условиях. В нем приняли участие двое испытуемых. При наблюдении невооруженным глазом каждый из испытуемых справлялся с заданием лучше, чем при использовании бинокля. Результаты одного испытуемого служат контролем для результатов другого, и это позволяет утверждать, что найденное различие является надежным. Каждый из испытуемых имеет достаточные основания для проведения спасательных поисков без использования бинокля.

ВТОРОЙ СПОСОБ УЛУЧШЕНИЯ РЕАЛЬНОСТИ: БОЛЬШЕ ДАННЫХ — ВЫШЕ НАДЕЖНОСТЬ

Давайте снова предположим, что исследователи решили провести эксперимент, не изменяя условий реальной жизни. Вот как они могли бы это сделать. В каждой спасательной операции один из испытуемых осуществлял бы поиск с помощью бинокля, а другой — наблюдая невооруженным глазом. В следующий раз они менялись бы местами и т. д. К сожалению для экспериментаторов (но к счастью для возможных жертв), подобные спасательные операции случаются не так уж часто. А поскольку для эксперимента необходимо 48 спасательных операций, то он занял бы не одну-две недели, а несколько лет. Весьма маловероятно, что реальные условия, включая сюда и судно, и сам экипаж с его командиром, оставались бы в течение этого времени достаточно стабильными. Вполне возможно, что экспериментаторы вынуждены были бы ограничиться всего лишь несколькими пробами для поисков с биноклем и без него. Нечего и говорить, что здесь едва ли

можно рассчитывать на постоянство таких побочных переменных, как размер и цвет цели, осведомленность испытуемого о месте ее появления, условия видимости, высота волн и т.п. В какой-то степени все эти факторы, конечно, усреднялись бы, поскольку поиск с биноклем и без него проводился бы во время одной и той же спасательной операции, однако такое усреднение было бы все же не очень эффективным. К тому же для каждого участника поиска влияние этих факторов было бы *индивидуальным*, зависящим от его собственного опыта.

Дело не только в том, что в эксперименте, проводимом в реальных условиях, оказалось бы меньше проб, — эти «пробы» не обеспечивали бы необходимой информации об успешности проведения поиска. В реальных спасательных операциях отдельный их участник часто вообще не обнаруживает какой-либо цели. Цель может быть найдена другим кораблем или экипажем вертолета или не найдена вовсе. Поэтому нельзя точно сказать, был поиск неудачным из-за менее эффективного способа наблюдения или из-за того, что в поле зрения испытуемого как с биноклем, так и без него цели просто не было.

В искусственном эксперименте гораздо большая часть факторов, влияющих на успешность поиска, поддается точному измерению. Поэтому искусственный эксперимент ближе к бесконечному эксперименту, чем тот, что проводится в условиях реальной жизни. Нужно еще раз подчеркнуть, что в бесконечном эксперименте (если бы он мог быть проведен) усреднялись бы любые несистематические влияния, изменяющиеся во времени. Таким образом, описанное улучшение — один из способов повышения надежности эксперимента. По сравнению с экспериментом, дублирующим реальный мир, который мог бы включать лишь несколько проб, искусственный эксперимент дает большую уверенность в том, что результаты, полученные на одних 48 пробах, будут теми же, что и на 48 других.

Экспериментаторы старались проконтролировать и другие дополнительные переменные: время года, погодные условия, размер цели и расстояние до нее. Во всяком случае уровень каждого из этих факторов был достаточно постоянным для достижения внутренней валидности. Хотя совсем забывать об их влиянии, конечно, не следует.

Надежность эксперимента повышалась также благодаря возможности опускать муляжи в заданном временном режиме. Размеры муляжа тоже можно было изменять по желанию. Вот, правда, нужную погоду заказать нельзя (как в эксперименте с посадками самолета), но все же можно было проводить опыты с биноклем и без него в одинаковых погодных условиях, равно как и в одинаковое время дня. Мы видели, что экспериментаторы могли сравнивать эффективность наблюдения невооруженным глазом и с по-

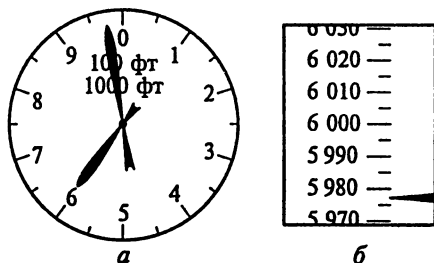
мощью бинокля при каждом конкретном наборе перечисленных условий. А если бы эксперимент проводился во время настоящих спасательных операций, такие тонкости и не могли бы обсуждаться. Необходимое количество проб каждого вида провести было бы просто невозможно.

Эксперимент 3:

ВЫБОР ПОДХОДЯЩЕГО ВЫСОТОМЕРА

Давайте рассмотрим другую возможную, но столь же мало-правдоподобную, ситуацию. Чарлз Аугустус Лендбург (известный своим друзьям под именем Лоун Сперроу) решил, наконец, приобрести небольшой собственный самолет. Он остановил свой выбор на марке «Скайрокет-23», хотя все еще продолжает колебаться. На этом самолете установлен стандартный высотомер с обычным циферблатом и двумя стрелками, показанный на рис. 3.4, а. За небольшую дополнительную плату (42,8 доллара) Чарлз может купить новый унифицированный высотомер моментального действия. На рис. 3.4 он изображен справа (б). Здесь сама шкала передвигается сверху вниз, а стрелка расположена на границе окошка. И хотя при выборе самолета и всего необходимого оборудования Чарлза, конечно, беспокоит его стоимость, в данном случае — жизнь дороже денег. Как раз недавно он перечитал старую измятую газетную вырезку, которую хранил в своем бумажнике.

Рис. 3.4. Слева (а) — шкала стандартного высотомера; справа (б) — шкала унифицированного высотомера моментального действия



Ибиза, Испания (ЮПИ). 1—7—72 В прошедшую пятницу иберийский реактивный самолет «Каравелла» врезался в самую высокую гору на этом средиземноморском курорте. Все 104 пассажира, находившиеся на борту, погибли.

По сообщению авиационной компании, на борту самолета, летевшего по маршруту Мадрид — Валенсия — Ибиза, находились 6 членов экипажа и 98 пассажиров, включая 6 детей. Иберийский представитель сообщил, что среди пассажиров было также двое иностранцев — Джефффри Д. Дессак из Нью-Йорка и Дитер Фрикер из Дюссельдорфа (Германия).

Официальное испанское агентство новостей Кифра сообщило, что во время последнего сеанса радиосвязи пилот указал свои координаты над небольшим островом Балери из архипелага Каниллора в 12 милях от аэропорта Ибиза и попросил разрешения снизиться до 5500 футов.

«Готовьте пиво, мы уже здесь», — цитирует аэропорт Ибиза последние слова капитана Джоуза Луиса Баллестера.

Контрольная служба дала Баллестеру (37 лет, 7000 полетных часов, отец шестерых детей) разрешение на спуск.

Как сообщает агентство новостей Кифра, самолет опустился гораздо ниже, чем предполагал его экипаж. «Самолет не врезался в гору напрямую. Передняя его часть была деформирована не так уж сильно по сравнению со средним и задним отсеками, разрушенными полностью. По всей вероятности, пилот увидел гору в самый последний момент и пытался резко набрать высоту». (Перепечатано с разрешения Юнайтед Стейтс Интернейшенл.)

Решение о типе эксперимента, который нужно провести

Лендбург посоветовался с друзьями, какой из двух высотомеров заслуживает большего доверия. Друзья разбирались в высотомерах по-разному, и единой точки зрения у них не было. Поэтому Чарлз предпочел провести короткий и надежный эксперимент. Сначала его план был довольно обширным. Он решил установить контакты с изготовителями и выяснить, есть ли у них тренажер. Если есть, то он смог бы «полетать» над воображаемой территорией по воображаемому курсу, используя каждую из двух шкал, и посмотреть, какая из них ему больше понравится. Но затем он подумал про себя так: «Скорее всего пилоту этого испанского самолета нравился тот высотомер, что был у него на борту. Нельзя ли поточнее узнать, насколько успешно я буду использовать оба эти высотомера». Тогда ему пришло в голову, что наверное есть какой-то специальный прибор, который позволяет одновременно регистрировать как сами показания индикатора, так и ответы на них. Полной уверенности у Чарлза не было. Так или иначе, для начала он послал в компанию «Скайрокет» запрос о возможности использовать тренажер, который они, по его мнению, сконструировали. Однако фирма ответила ему, что она, конечно, очень польщена, но исследования по данной теме у них пока не финансируются.

Однако вскоре Чарлз сообразил, что в действительности он и не смог бы по-настоящему использовать данные, полученные на тренажере. И тогда его осенило. Он приготовил картонные макеты каждой шкалы и попросил фотографа сделать карточки с самыми разными их показаниями. Таким образом, он получил материал для проведения исследования с двумя высотомерами на самом себе.

Процедура эксперимента

Было использовано по 120 карточек с изображениями каждой шкалы. Набор показаний высоты на одной шкале был тем

же, что и на другой. Лендбург перемешал карточки с обычной шкалой и разделил их на две пачки по 60 штук. Одну пачку он назвал набором А, другую — набором Д. Такие же пачки со шкалой нового высотомера составили наборы В и С. Буквы указывают порядок, в котором должно было проводиться снятие показаний. Для проведения эксперимента Чарлзу понадобились два магнитофона (или один двухдорожечный магнитофон). Он решил снимать показания каждые 5 с. Для этого он приготовил магнитофонную ленту, на которую через каждые 5 с (по секундомеру) он записывал сигналы «Внимание». Во время эксперимента он подавал себе эти сигналы через наушники. Каждый раз, услышав сигнал «Внимание», он смотрел на следующую карточку. Оценки считанных показаний он записывал на второй магнитофон. На экспериментальную серию с каждым набором карточек уходило 5 мин.

Анализ результатов

Первый способ анализа данных, который пришел в голову Чарлзу, — это составить список действительных показаний шкал и рядом с каждым из них написать данную им оценку высоты. Таким образом, можно определить величину ошибки в каждом ответе и затем вычислить среднюю ошибку для той и другой шкалы. Различие оказалось не слишком большим. Средняя ошибка для

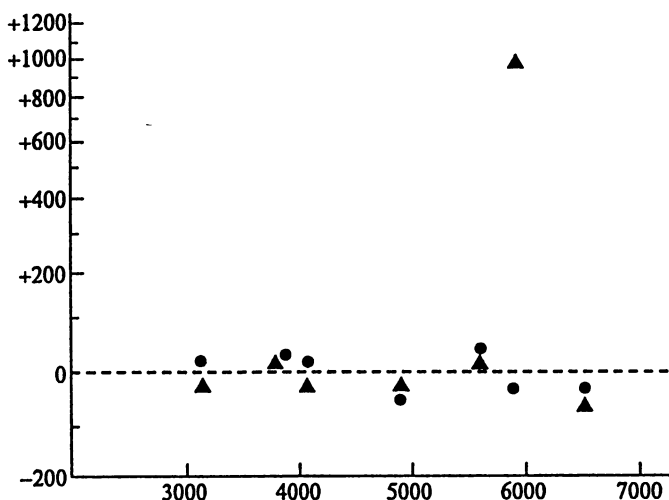


Рис. 3.5. Эксперимент с высотомерами: ошибки при считывании показаний. Ось абсцисс — показания высоты полета на шкалах (в футах); ось ординат — величина ошибки (в футах). Треугольниками отмечены данные по стандартной шкале, кружками — по новой

старой шкалы составила 12 футов, а для новой — 8. Другой способ анализа показан на рис. 3.5. Каждой из оцениваемых высот соответствует отметка на горизонтальной оси. Здесь же показаны ошибки в ответах: положительные — при переоценке высоты и отрицательные — при ее недооценке. В представленных данных также нет значительного различия между шкалами, за одним исключением: по старой шкале высота в 5980 футов была принята за 6975 футов, т.е. ошибка составила почти 1000 футов! Если мы вернемся к рис. 3.4, то сможем понять, отчего происходит такая ошибка. При работе со старой шкалой были и другие ошибки, хотя и не такие грубые. На этом основании Чарлз решил заплатить лишние 42,8 доллара.

Краткое изложение эксперимента

Чарлзу Лендбургу нужно было решить, устанавливать ему на своем самолете новый унифицированный высотомер или нет. Для этого он придумал эксперимент и провел его на себе. Эксперимент состоял в считывании показаний высоты с фотографий шкал в установленном темпе. По величине средней ошибки различия между шкалами оказались небольшими. Однако при использовании традиционного высотомера Чарлз допускал гораздо более грубые ошибки. Иногда эти ошибки можно было объяснить, иногда — нет. Чарлз заключил, что летать с таким прибором небезопасно, и выбрал новый высотомер.

ТРЕТИЙ СПОСОБ УЛУЧШЕНИЯ РЕАЛЬНОСТИ: ВЫСОКАЯ НАДЕЖНОСТЬ ЗА СЧЕТ СОКРАЩЕНИЯ НЕСИСТЕМАТИЧЕСКОЙ ИЗМЕНЧИВОСТИ

Предположим, что Лендбург получил возможность испытать каждый высотомер в реальном полете и решил провести эксперимент первого типа. Что ему пришлось бы для этого сделать? Просто совершить бы несколько полетов с разными высотомерами. Однако при этом оценить качество работы с высотомером было бы довольно трудно. Ведь Чарлзу нужно было бы не только считывать показания шкал, но и контролировать истинность этих показаний, оценивая высоту своего полета визуально, наблюдая территорию, над которой он летал. Положим, ему удалось бы одновременно делать и то и другое. Пусть очень грубо, но он все-таки смог бы определять высоту полета, например по высоте гор, подлетая к ним достаточно близко, причем высоту пришлось бы изменять довольно часто (и терять на этом горючее). Могло случиться и так, что Чарлз, подобно испанскому пилоту, допустил бы очень серьезную ошибку. Даже если бы число проб

в таком полете было бы столь же большим, как в искусственном эксперименте, они дали бы гораздо менее надежные результаты. Оценки высоты зависели бы от таких побочных переменных, как особенность восприятия территории (вспомним эксперимент с ночными посадками), степень сосредоточенности и, наконец, просто умение вести самолет на нужной высоте. Несистематическая изменчивость в поведении испытуемого, а следовательно, и разброс экспериментальных данных были бы очень большими. И если бы эксперименты не продолжались в течение долгого времени, то результаты одного эксперимента явно не сопали бы с результатами другого.

Напротив, искусственный эксперимент, который провел Чарлз, требовал вполне ясных ответов, которые полностью определялись считыванием показаний высотомера. Ошибки в этих ответах можно вычислить совершенно точно. Оценки работы испытуемого в одном эксперименте будут близки к тем, которые могут быть получены в другом эксперименте, проведенном при тех же условиях. Благодаря сокращению несистематической изменчивости этих оценок искусственный эксперимент достигает более высокой надежности по сравнению с экспериментом, дублирующим реальность.

Повторим, что надежность эксперимента можно повысить двумя путями. В эксперименте с поиском эта цель достигалась с помощью увеличения числа проб. В исследовании с высотомерами надежность была повышена благодаря сокращению несистематической изменчивости. Внутреннюю валидность эксперимента можно улучшить, не только повышая надежность, но и устраняя систематическое смещение: это было показано в эксперименте с ночными посадками самолета.

НЕСКОЛЬКО СПОСОБОВ СРАЗУ

В каждом из трех описанных экспериментов внутренняя валидность повышалась главным образом за счет одного из перечисленных способов улучшения реального мира. Но до известной степени в этих экспериментах были реализованы и другие такие способы.

В эксперименте с ночными посадками не только устранялось систематическое смещение. Помимо этого пилоту за короткое время предъявляли довольно много проб, и его работу можно было оценить более точно, чем в реальных полетах.

В эксперименте с поиском было не просто больше проб, чем могло быть в реальных спасательных операциях, но и сокращалась несистематическая изменчивость в поведении испытуемых. Это было достигнуто, во-первых, обеспечением лучшего способа оценки их работы — фиксацией момента спуска муляжа. Во-вто-

рых, пробы с использованием бинокля и без него были уравнены по погодным условиям, времени дня, размеру муляжа и расстоянию до него.

В исследовании с высотомерами сокращение несистематической изменчивости достигалось не только за счет большей точности в оценке работы испытуемого, но и благодаря возможности провести достаточное количество замеров за более короткое время, чем в реальном полете. Более того, поскольку в эксперименте, дублирующем реальность, можно было бы совершить лишь небольшое число полетов, то воздействия независимой переменной (тип шкалы высотомера) неизбежно смешивались бы с побочными факторами (такими, как сила ветра, характер территории, количество воздушного транспорта). В искусственном эксперименте это смешение полностью устранено.

Таким образом, в каждом из трех наших экспериментов, улучшающих реальный мир, применяются все три возможных способа повышения внутренней валидности.

ВНЕШНЯЯ ВАЛИДНОСТЬ: ВОПРОСЫ СООТВЕТСТВИЯ

Хорошим новшествам нередко сопутствуют новые проблемы. Искусственные эксперименты, разумеется, более удачны, чем те, в которых реальный мир просто дублируется. Сама реальность здесь «улучшена», и это очень хорошо. Но *адекватны* ли такие эксперименты? Можем ли мы применять полученные результаты для решения тех реальных проблем, которые, собственно, и давали начало нашим экспериментам? Если ответ отрицательный, то это плохо. Платой за повышение внутренней валидности будет потеря валидности внешней.

Сейчас мы рассмотрим вопросы соответствия искусственных, экспериментальных ситуаций их реальным прототипам для всех трех описанных экспериментов. Вы увидите, что иногда (но не во всех случаях) они имели удовлетворительное решение. Мы последовательно обсудим каждую из составляющих экспериментальной гипотезы. Вы помните, что всякая гипотеза предполагает некоторое отношение между *независимой* и *зависимой* переменными. Поэтому сначала мы проверим на соответствие независимую переменную, а затем зависимую. Однако не менее важная составляющая, которая не всегда отмечается специально, но всегда присутствует, — это уровень *значимой дополнительной переменной*. Вспомните, ведь неадекватный вариант эксперимента Джека Моцарта страдал недостатком внешней валидности именно потому, что по типу используемых в нем пьес он не соответствовал исследуемой гипотезе. Третьим пунктом нашего анализа будет, таким образом, обсуждение соответствия дополнительных переменных.

Проверяя соответствие переменных, необходимо помнить, что мы имеем дело с экспериментами, отвечающими на конкретные практические запросы. В каждом из приведенных случаев было найдено решение, лучшее из возможных. По материалам предыдущей главы вы знаете, что безупречной внутренней валидности достичь нельзя, поскольку реальный эксперимент не может быть ни идеальным, ни бесконечным. Внутренняя валидность реальных экспериментов лишь повышается по мере их приближения к указанным разновидностям безупречного эксперимента. Аналогично невозможна и безупречная внешняя валидность, ведь в реальном эксперименте нельзя достичь полного соответствия всех тех жизненных обстоятельств, к которым прилагаются его результаты. Отсюда можно говорить лишь о большей или меньшей внешней валидности искусственных экспериментов, смотря по тому, в какой степени соблюдаются в них требования эксперимента полного соответствия. Однако искусственные эксперименты ставятся тогда, когда эксперименты, дублирующие реальность, — и, следовательно, более ей соответствующие, — страдают недостатком внутренней валидности. Поэтому мы не вправе отвергать искусственный эксперимент только потому, что он меньше соответствует реальному миру по сравнению с экспериментом, в котором этот мир просто дублируется. Вместо этого следует задаться вопросом, найден ли *самый оптимальный* способ улучшения реального мира. Поэтому для оценки внешней валидности экспериментов, улучшающих реальность, имеет смысл сравнивать их с другими искусственными экспериментами.

Соответствие независимой переменной

Вопрос о соответствии независимой переменной в искусственных экспериментах по большей части довольно прост. Экспериментатору нужно быть твердо уверенным только в том, что введенные им условия в одних реальных случаях являются типичными, а в других — вполне вероятными. Крафт и Элворт выбрали для посадок горизонтальную и наклонную (под углом 3°) поверхности, поскольку эти условия типичны для аэропортов, где приземляются реактивные самолеты. Наблюдение без бинокля в эксперименте с поиском не соответствовало старым инструкциям, но было вполне возможным. Это совершенно реальный способ поиска. А два типа шкал в исследовании с высотомерами — это те самые шкалы, лучшую из которых собирался выбрать Лендбург. Вот если бы он взял такой тип шкалы, который не используется на его самолете, скажем, изображение небольшой модели самолета в трехмерном пространстве, то это условие не было бы соответствующим. Ведь гипотеза Лендбурга касалась только двух высотомеров, которые можно поставить на его будущий самолет.

Соответствие зависимой переменной

Вспомним, что при каждом из условий независимой переменной зависимая переменная принимает определенное значение. Каждое такое значение включает в себя три компонента: *во-первых*, ответы испытуемого, его поведение, *во-вторых*, измеряемые показатели ответов испытуемого и, *в-третьих*, способ представления результатов измерений (дающий нам окончательное значение зависимой переменной). Вот и займемся теперь анализом трех экспериментов, описанных в настоящей главе, по каждому из названных пунктов.

Поведение испытуемых. Соответствует ли поведение испытуемого в эксперименте той его реальной деятельности, на которую будут распространяться полученные результаты? В отношении двух наших экспериментов мы можем ответить на этот вопрос утвердительно. Пилот «ведет» тренажер, ориентируясь по «наземным» огням точно так же, как и в настоящем полете, поэтому соответствие зависимой переменной в эксперименте с ночными посадками является вполне удовлетворительным. И в эксперименте с поиском испытуемые точно так же вели наблюдения за поверхностью моря, как в действительных спасательных операциях. А вот об исследовании с высотомерами разговор особый. Испытуемый работает с показаниями шкал весьма необычным способом. Он не изменяет ни высоту, ни направление полета, он вообще не ведет самолет, т.е., с одной стороны, он выполняет гораздо меньше операций, чем в реальности, а с другой — совершает дополнительные операции. Так, при снятии показаний он называет соответствующее число. В полете же чаще всего показания высотомера нужны пилоту лишь для определения высоты, необходимой для правильного направления полета, т.е. в пределах примерно 200 футов. В полете незачем повторять эти показания, и тем более уделять все внимание высотомеру, как это было в эксперименте. Можно ли оправдать такое значительное отклонение от реальной деятельности? Давайте обсудим это еще раз и кратко напомним о самой проблеме.

Лучше всего сравнить выбранный способ проведения эксперимента с другими возможными альтернативами. Лендбург понимал, что данные, которые можно получить в реальном полете, были бы очень сомнительны. Лендбург предпочел искусственный эксперимент, поскольку условия реального полета не позволили бы ему адекватно оценить собственную работу. Так ли это для эксперимента на тренажере? Тренажер улучшает реальный мир, но только в одном: он позволяет унифицировать погодные условия, а также наземную территорию при использовании обеих шкал. Однако пилоту по-прежнему нужно было бы придерживаться определенной высоты «полета», оценивая ее субъективно. И это вновь

зависело бы от множества факторов: и от восприятия территории в каждом конкретном случае, и от осторожности пилота, и от его умения вести самолет.

По-видимому, простое снятие показаний высотомера действительно отражает тот аспект реальной деятельности, который интересовал исследователя. Весьма разумным было решение проводить испытания в заранее установленном темпе. Как правило, в полете у пилота немного времени для снятия показаний. Темп, конечно, можно было и увеличить, определяя при этом количество показаний, снятых испытуемым за каждую минуту. Однако эта идея не совсем удачна по двум причинам. Во-первых, подобное скоростное считывание меньше соответствует тем реальным операциям, которые выполняет пилот. Во-вторых, возникла бы проблема совместного учета скорости работы и количества совершенных ошибок.

И все же, несмотря на все приведенные аргументы, в данном случае трудно предложить полностью адекватный экспериментальный прием. Любой конкретный прием основан на довольно условных предположениях (скажем, о сравнительной значимости каждого правильного считывания и цене каждой ошибки).

Измеряемые показатели. Первый эксперимент достаточно хорошо соответствовал реальным посадкам самолета и в отношении производимых измерений. При работе испытуемых в каждом из экспериментальных условий — горизонтальной и наклонной территории — фиксировались действительная высота «полета» и ее субъективные оценки. В эксперименте с поиском такой определенности нет. Так ли уж важно на самом деле для спасательной операции, будет найдена цель за 7 или 7,5 мин? Наверное, нет. Правда, когда катер движется по прямой, т.е. ходит туда-сюда, как по длинному коридору, — а обычно это так и происходит — неудачи в нахождении цели за определенный период времени могут означать, что либо цели здесь просто нет, либо нужно плыть по-медленнее, а не разбрасываться на чересчур большое пространство. Поэтому временные характеристики можно связать с успешностью стратегии поиска цели — если она, конечно, существует. В исследовании с высотомерами проблемы выбора показателей не возникало. Фактически ими были сами показания испытуемого по каждому положению шкалы, которые сразу записывались на магнитофон.

Способ представления результатов измерений. На примерах двух описанных экспериментов с оценкой высоты полета вы могли видеть разные способы представления результатов измерений. Впрочем, каждый из этих способов может быть пригоден при распространении экспериментальных выводов на реальный мир. В первом эксперименте гипотеза состояла в том, что пилот совершает систематическую ошибку, недооценивая высоту, и поэтому летит

слишком низко при посадке на наклонную территорию. Графическое изображение результатов эксперимента на рис. 3.3 позволяет проверить эту гипотезу. Здесь представлены усредненные данные 12 пилотов, каждому из которых давали несколько проб. Подобным образом можно было бы отразить выполнение задач любым участником эксперимента в каждом из исследуемых условий. Понятно, что если бы на каждое условие приходилась только одна проба, то выбранный способ представления результатов показал бы лишь изменение высоты полета по мере приближения к аэропорту. Но если дается целая группа проб, то для каждой точки посадочной траектории можно получить среднее значение оценок этой высоты.

Для исследования с высотомерами такой способ представления данных не подходит. Поскольку показатели высоты, которые считывал испытуемый, изменялись не постепенно (как по мере приближения к аэропорту), а случайным образом, графическое изображение последовательности оценок вряд ли имело бы смысл. Простое вычисление среднего для оценок испытуемого по каждой шкале тоже не принесло бы желаемых результатов. Предположим, что при работе со старой шкалой испытуемый допускал грубые ошибки, однако число ошибок с переоценкой и недооценкой высоты было одинаковым. Тогда, не смотря на все ошибки испытуемого, средняя оценка его работы практически равнялась бы средней величине предъявляемых показаний.

Этот факт требует особого внимания, поскольку в значительной части опубликованных экспериментальных работ он не вполне осознается. Покажем, как возникают подобного рода погрешности, на кратком примере. Предположим, что в четырех последовательных пробах испытуемому предъявлялись показания следующих высот: 3200, 6100, 1300 и 4640 футов. Средняя величина для этого набора проб составила бы, таким образом, $3200 + 6100 + 1300 + 4640$, деленное на 4. Она равна 3797,5 фута. А вот результаты испытуемого: 3260, 6040, 1250 и 4590. Средняя оценка тоже равна 3797,5 фута.

Никому и в голову не придет, что испытуемый работал с ошибками. Такая же погрешность сохранится и при вычислении алгебраического среднего, когда переоценки обозначают знаком «плюс», а недооценки — знаком «минус». В приведенных четырех пробах ошибки были следующие: $3260 - 3200$, или +60; $6040 - 6100$, или -60; $1300 - 1250$, или +50; $4590 - 4640$, или -50. Ошибки на +60, -60, +50 и -50 в сумме дадут 0. Вот так и появляется погрешность. Ясно, что необходим какой-то другой способ представления результатов.

Можно было бы не учитывать знаки ошибок — плюс и минус. Тогда, вычислив среднее для указанных проб — 60, 60, 50 и 50, мы получим *абсолютную ошибку*. Она будет равна 55 футам. Нужно

отметить, что и эта средняя оценка может вызвать возражения. В частности, она не позволяет отличить приведенные данные от таких, когда ошибки (все или какая-то часть) имеют одно направление. Например, данные +60, +60, +50 и -50 тоже дадут абсолютную ошибку в 55 футов. В подобных случаях для представления результатов нужно брать сразу два показателя. Первый из них уже описан: это *алгебраическая ошибка*, при подсчете которой пользуются знаками плюс и минус. Она позволяет определить соотношение разнонаправленных ошибок испытуемого. Второй — *стандартное отклонение* — показывает, насколько велик разброс этих ошибок, т. е. характеризует изменчивость в деятельности испытуемого при выполнении задачи.

И все же для эксперимента Лендбурга наиболее адекватен иной способ представления данных. Ведь главное для его автора — не допускать в реальном полете слишком грубых ошибок. Поэтому данные по работе с каждым из высотомеров нужно представить процентным отношением таких ошибок (на 100 футов и больше) к общему числу неверных ответов. Правда, этот способ не подошел бы, если бы нужно было сравнить качество работы с той и другой шкалой при посадке самолета в условиях плохой видимости. Будем надеяться, что в плохую погоду Лендбург не полетит.

Подсчет процентных отношений вполне подходит и для эксперимента с поиском. По данным о времени, затраченном на поиск каждой цели, можно определить процентное соотношение количества целей, быстрее найденных с биноклем или без него. А быстро найти цель — это самое главное в любой спасательной операции.

Соответствие дополнительных переменных

В неадекватном варианте эксперимента Джека Моцарта, когда вместо сонат разучивались вальсы, уровень наиболее важной дополнительной переменной — *типа музыкальных пьес* — был явно несоответствующим. Ведь то, что справедливо для «уровня вальсов», может оказаться неверным для «уровня сонат». Это случай несоответствия ключевой переменной. Давайте рассмотрим три наших эксперимента с точки зрения соответствия ключевых, а также некоторых других дополнительных характеристик.

Ключевые переменные. В одних экспериментах, как, например, у Джека Моцарта, ключевая переменная одна (но очень важная). В других экспериментах их может быть несколько. Скажем, такие характеристики спасательного поиска на море, как размер цели, расстояние до нее, погодные условия и время дня, примерно одинаковы по значимости.

В эксперименте с посадками самолета ключевой переменной была зрительная картина ночного города. Ведь только она и дает

информацию о наклоне территории. Различия понятны: либо это несколько огней, расположенных близко друг к другу, либо целая панорама, где пилот может выбрать любую пару световых точек. Чтобы результаты эксперимента можно было применять для любых аэропортов, Крафт и Элворт предъявляли испытуемым несколько типичных моделей. Они пишут (показывая тем самым, что даже опытные экспериментаторы могут ошибаться): «Мы надеялись повысить эффективность зрительного контроля посадки с помощью расширения и углубления световой картины города. Однако данные показывают, что более обширная и комплексная картина на самом деле может приводить к катастрофе, вводя пилота в заблуждение, что происходит в случае поднимающейся вверх территории» (Kraft C. L., Elworth C. L., 1969, с. 4). Оказалось, что огни, занимая большую площадь, представляются пилоту более надежным показателем уровня земли, и это усиливает иллюзию. Теперь нам ясно, что экспериментаторы не зря копировали огни ночных городов во всех их вариантах.

Столь же аккуратны были исследователи в эксперименте с поиском, добываясь более точного соответствия всех ключевых переменных — размера цели, расстояния до нее, погодных условий и времени дня. А вот в исследовании с высотомерами контролировалась только одна ключевая характеристика — использовался тот же *диапазон высот*, с каким встречается пилот в реальном полете. Передвижение индикаторов — а это тоже ключевая характеристика — не воспроизводилось. Лендбург имел дело со стабильными изображениями шкал, а в реальности индикаторы чаще всего непостоянны. С другой стороны, в реальном полете последовательные показания прибора похожи одно на другое. Самолет не подпрыгивает вверх и не падает вниз случайным образом. А ведь именно так и изменялись показания высот в последовательных пробах эксперимента — случайно. Первое отклонение от реальности облегчало испытуемому работу со шкалой, а второе, наоборот, затрудняло ее. Для более точного воспроизведения реальных изменений шкалы высотомера понадобился бы киноаппарат или видеоманитофон. Это довольно утомительно: снова нужно вырезать картонки, подбирая друг к другу почти одинаковые снимки с чуть измененными положениями индикаторов. А потом можно было бы дополнить подачу словесных команд на магнитофоне соответствующим звуковым сопровождением.

Одновременные действия. В некоторых искусственных экспериментах испытуемому приходится выполнять именно то задание, которым (и только им) он занимается в реальной жизни. Например, во время спасательной операции наблюдатель не имеет никаких других обязанностей, кроме самого поиска. Для пилота это, конечно, не так. Совершая посадку, ему нужно не только держать нужную высоту, но и постоянно корректировать траекторию полета.

та, чтобы самолет находился под правильным углом и не уклонялся в сторону. Пилот должен следить за скоростью, остерегаться столкновения со встречным транспортом. В эксперименте Крафта и Элворта все эти действия воспроизводились. Во-первых, пилот «вел» тренажер как настоящий самолет, а не просто контролировал высоту. Во-вторых, у него была дополнительная задача — «определять местоположение других самолетов и сообщать о них» (Kraft C. L., Elworth C. L., 1969, с. 2).

Другой наш пилот, Чарлз Аугустус Лендбург, ничего этого не делал. Он только считывал показания высоты и не производил никаких дополнительных действий. А было бы неплохо выполнять при этом какую-нибудь другую задачу. Вполне возможно, что новый высотомер становится более надежным лишь в том случае, если все внимание испытуемого уделяется только ему. Материал для дополнительной задачи тоже можно было бы записать на магнитофон (вместе с командами о порядке снятия показаний). Испытуемый мог бы, например, подсчитывать звуковые сигналы.

Напряженность. Все эксперименты, описанные в этой главе, были посвящены практическим проблемам, связанным с жизнью и смертью людей. И это не просто случайное совпадение. Помимо повышения внутренней валидности эксперименты, улучшающие реальный мир, очень часто делают его безопасным для испытуемого. Но тогда возникает вопрос: можно ли переносить результаты, полученные при отсутствии эмоциональной напряженности, на реальную деятельность в стрессовых условиях? Иногда предлагают гипнотизировать испытуемых и внушать им, будто они находятся в реальной ситуации, а не на эксперименте. Однако такое внушение вряд ли будет эффективным для человека, который хорошо знает, что такое гипноз. Давайте посмотрим, насколько серьезен вопрос о недостаточной напряженности для наших экспериментов.

Типичным последствием состояния эмоциональной напряженности является нарушение интеллектуального контроля за поведением. Трудно представить, каким образом недостаток напряженности может усиливать зрительную иллюзию пилота при экспериментальном моделировании посадки над наклонной территорией. Скорее уж можно предположить, что более высокий интеллектуальный контроль уменьшит эту иллюзию. Следовательно, можно сказать, что в эксперименте были получены важные результаты, *несмотря на* отсутствие напряженности.

Известно также, что в состоянии напряженности разрушаются в первую очередь приобретенные и необычные навыки, а не естественные, привычные. Смотреть в бинокль менее естественно, чем без него. Поэтому отсутствие напряженности в эксперименте с поиском было благоприятным для наблюдения с биноклем. И вновь

можно сказать, что именно данные результаты эксперимента были получены, *несмотря на это* преимущество.

Сжатие во времени. Увеличение надежности в искусственных экспериментах по сравнению с теми, которые дублируют реальность, достигается главным образом благодаря возможности предъявить все необходимые пробы за более короткий период времени. Тем самым можно быстрее получить достаточное количество данных. Искусственный мир чаще всего как бы *сжат во времени* по сравнению с реальным. Как это влияет на внешнюю валидность экспериментальных выводов?

Из трех описанных экспериментов меньше всех был сжат во времени эксперимент с посадками самолета. Правда, он и не требовал слишком большого количества проб по сравнению с двумя другими. Известно, что практический опыт до некоторой степени уменьшает зрительные иллюзии. Следовательно, в эксперименте на тренажере влияние иллюзии могло бы в принципе сокращаться быстрее, чем в реальных полетах. Однако, *несмотря на* преимущества, возможные за счет научения, в эксперименте получены результаты, которые свидетельствуют о сохранении этого влияния.

Спасательный поиск на море лучше производить без бинокля — этот результат, полученный в условиях быстрого предъявления всех необходимых проб, также не вызывает никаких сомнений. Конечно, в обычных условиях поиск продолжается дольше, и бдительность его участников будет более изменчивой, чем в своеобразном соревновании между ними, характерном для данного эксперимента. Однако реальная ситуация была бы более жесткой именно для наблюдения с биноклем. Ведь в эксперименте спасатели пользовались им сравнительно недолго, и поэтому влияние веса бинокля, усталость глаз, а также неясность зрительной картины были не столь существенны. Можно сказать, что поиск с биноклем оказался менее эффективным *даже* при коротком испытании, условия которого благоприятствовали его применению.

В эксперименте с высотомерами таких гарантий нет. Очень может быть, что за целую серию проб, между которыми только 5 с, испытуемый просто научится хорошо считывать показания высотомера. А если справляться о высоте полета лишь время от времени, как это происходит в реальности, подобная привычка будет вырабатываться не так скоро. Поэтому по результатам, полученным в условиях сжатого предъявления проб, трудно решить, каким из двух высотомеров удобнее пользоваться в реальном полете. Пожалуй, в этом отношении эксперимент можно было бы улучшить, если все-таки сделать считывание показаний более развернутым, скажем, давать пробы лишь время от времени по мере выполнения другой задачи.

Внешняя валидность в более широком смысле

В этой главе мы часто обращались к проблеме внешней валидности эксперимента, но вместе с тем рассматривали ее лишь с одной точки зрения. В общем виде вопрос о применении экспериментальных результатов к реальной жизни — это систематическое рассмотрение вопроса о степени сходства всех переменных в действительном эксперименте со всеми переменными в *эксперименте полного соответствия*. В следующей главе также будет обсуждаться вопрос относительно соответствия реальности, но уже по отношению к тем людям, на которых распространяются результаты эксперимента. А эксперименты из главы 5 должны будут соответствовать не реальности, а «миру теории». Это соответствие определяется тем, в какой мере конкретные экспериментальные приемы отражают теоретические понятия. Если такой переход затруднителен, то внешняя валидность будет низкой.

КАКОВА ЦЕНА РЕАЛИЗМА?

До сих пор мы почти не затрагивали этот вопрос. Он обсуждался только в эксперименте с высотомерами, который меньше других был сходен с реальной жизнью. Мы показали, что по используемым приемам этот эксперимент был все-таки довольно удачным (по сравнению с возможным экспериментом на тренажере), но что его можно улучшать и дальше.

Однако обсуждение можно продолжить и поставить вопрос так. Не слишком ли страдает внутренняя валидность эксперимента из-за нашего стремления к большему реализму, лучшему воспроизведению действительности? И можно ли провести эксперимент, обладающий высокой внутренней валидностью, не отказываясь от максимального приближения к реальности?

Подобные вопросы относятся, на самом деле, только к эксперименту с поиском. В эксперименте с посадками самолета применение тренажера гарантировало достаточную внутреннюю валидность. А вот второй эксперимент был чрезмерно реалистичен: он проводился на настоящем катере в настоящем море. В таких условиях, конечно, трудно проконтролировать всевозможные побочные факторы. Но будет ли этот контроль более эффективным, если провести эксперимент в лаборатории, используя что-то вроде тренажера?

Можно было бы заснять движущуюся поверхность моря и затем показывать ее испытуемому на широком экране. Ощущение реальности можно усилить, используя вместо катера качающуюся платформу, вроде тех, что применяются при изучении морской болезни. Однако при этом возникает сразу несколько проблем. Во-первых, киносъемка все равно не обеспечит испытуемому воз-

возможность увидеть всю панораму, которую видит перед собой наблюдатель в реальной спасательной операции. Это отклонение особенно значимо для наблюдения невооруженным глазом. А применение панорамной камеры лишь еще больше запутает зрительную картину. Во-вторых, кино не дает полного впечатления пространства. Для наблюдения с биноклем, когда восприятие глубины так или иначе нарушено, это не столь уж большая помеха, а простое наблюдение явно пострадает. Если же вместо кино показывать слайды, то это еще больше удалит испытуемого от реальности. Все изменения морской поверхности и само движение катера будут утеряны. А они могут по-разному влиять на наблюдение с биноклем и без него. Короче говоря, первоначальный вариант эксперимента был не так уж плох. В данном случае эксперимент должен быть высокореалистичным.

Поскольку в этой главе мы вновь рассматривали практические эксперименты, нам следует упомянуть и о финансовой стороне дела. Если для эксперимента нужен корабль (с заранее составленным планом его движения), а также капитан с командой да еще экипаж вертолетов (действия которых должны быть скоординированы), то потребуются, конечно, большие расходы. Проведение экспериментов Крафта и Элворта на тренажере с компьютером стоит, пожалуй, еще дороже. А несколько фотографий и прокат двух магнитофонов для исследования с высотометрами не стоили Лендбургу почти ничего. Понятно, что меньших затрат требуют эксперименты, которым больше недостает реализма. Чаще всего стремление лучше представить в эксперименте реальный мир стоит очень дорого.

КРАТКОЕ ИЗЛОЖЕНИЕ

В этой главе мы обсуждали три эксперимента, которые не дублируют реальный мир, а улучшают его. В первом эксперименте проверялась гипотеза о том, что при посадках самолета на постепенно поднимающуюся территорию у пилотов возникает зрительная иллюзия, и поэтому они снижаются слишком резко. Если проводить этот эксперимент в настоящих аэропортах, неизбежно систематическое смешение независимой переменной с различными побочными факторами. Применение тренажера, имитирующего реальный мир, позволило устранить это смешение.

Гипотеза второго эксперимента состояла в том, что при поисковых операциях на море наблюдение невооруженным глазом даст более эффективные результаты, чем наблюдение с биноклем. Если бы этот эксперимент проводился во время реальных спасательных операций, то за любой практически приемлемый срок было бы собрано лишь небольшое количество данных. Надежность такого эксперимента была бы низкой. Благодаря использованию самодельных муляжей удалось за несколько недель провести эксперимент с достаточным числом проб, повысив тем самым его надежность.

В третьем эксперименте сравнивалось качество работы с двумя высоотомерами. Гипотеза была следующей: использование новой унифицированной шкалы позволит сократить количество грубых ошибок при определении высоты полета по сравнению со старой шкалой (циферблат с двумя стрелками). Если бы этот эксперимент проводился в реальном полете, то на успешность работы испытуемого с высоотомером влияло бы множество побочных факторов. Полученные данные имели бы большой разброс, и это вновь понизило бы надежность. Применение задачи по снятию показаний с фотографий обеих шкал в установленном темпе позволило значительно сократить несистематическую изменчивость экспериментальных данных.

Таким образом, в искусственных экспериментах можно повысить внутреннюю валидность. Описаны три способа улучшения реального мира, позволяющие сделать это возможным. Первый из них — устранение систематического смещения. Второй — возможность получить необходимое количество данных за более короткий срок и тем самым повысить надежность эксперимента. И третий — сократить несистематическую изменчивость данных и, следовательно, их разброс, что также обеспечивает более высокую надежность.

Но именно потому, что искусственные эксперименты не дублируют реальный мир, возникает вопрос об их внешней валидности. Достаточно ли успешно представлена в этих экспериментах реальность, чтобы можно было считать полученные результаты адекватными? Поскольку искусственные эксперименты ставятся в тех случаях, когда эксперименты с простым дублированием реального мира страдают недостатком внутренней валидности, сравнение с последними не может служить критерием их адекватности. Внешняя валидность трех описанных экспериментов оценивалась путем сравнения с другими (но также искусственными) приемами их проведения. Каждая из составных частей экспериментальной гипотезы: независимая, зависимая и дополнительные переменные — была проверена нами на соответствие исследуемой реальности.

Достичь в эксперименте соответствия независимой переменной, как правило, довольно просто. Нужно только, чтобы вводимые условия были либо типичными для реальных ситуаций, либо вполне вероятными.

Соответствие зависимой переменной оценивалось по следующим трем пунктам: 1) соответствует ли работа испытуемого в эксперименте его реальной деятельности? 2) отражают ли измеряемые показатели наиболее важные аспекты этой деятельности? 3) адекватен ли способ представления результатов измерений? Самой важной проблемой оказывается здесь адекватный учет ошибочных ответов испытуемого, имеющих плюсовые и минусовые значения, как, например, переоценка и недооценка высоты полета. Чаще всего применяются два способа представления таких результатов. Один из них позволяет показать преимущественную направленность ошибочных ответов, а другой — определить величину их разброса.

В искусственных экспериментах возникают также вопросы о соответствии дополнительных переменных, стабильных по своему уровню. В целом ряде случаев такие переменные являются ключевыми, и их уровень должен соответствовать реальному миру. Нужно стараться также воспроизводить в эксперименте те дополнительные (по отношению к основной задаче) действия, которые в реальности выполняются одновременно с ней. Кроме того,

следует выяснить, как скажется на внешней валидности искусственного эксперимента отсутствие эмоциональной напряженности (которая, как правило, характерна для соответствующих реальных ситуаций). И наконец, следует специально проанализировать последствия предъявления испытуемому всех экспериментальных проб за короткий (сжатый по сравнению с экспериментом, дублирующим реальность) период времени.

Но есть еще один вопрос — какова цена реализма? Каждый раз необходимо проверять, не слишком ли мы усердствуем, добиваясь в экспериментах как можно более точного воспроизведения реальности, и не страдает ли от этого их внутренняя валидность. Такая проверка также проведена путем сравнения с другими возможными приемами эксперимента. В заключение был затронут вопрос о финансовой стоимости искусственного эксперимента с максимальной имитацией реального мира. Иногда эта стоимость так высока, что реальным может оказаться в действительности менее реалистичный эксперимент.

Контрольные вопросы и задания

1. Почему эксперимент с ночными посадками самолета не мог быть проведен в настоящих аэропортах?
2. В чем состоит основное преимущество эксперимента со спасательным поиском?
3. Приведите пример эксперимента, дублирующего реальный мир, в котором имела бы место чрезмерная несистематическая изменчивость получаемых данных.
4. Перечислите, каким образом в экспериментах, «улучшающих» реальный мир, добиваются большей внутренней валидности, чем в экспериментах, дублирующих реальность.
5. Как вопрос о внешней валидности связан с вопросом о видах безупречного эксперимента, описанных в главе 2?
6. Что имеют в виду, говоря, что решение вопроса о соответствии эксперимента — это проверка соответствия основных составляющих экспериментальной гипотезы?
7. Приведите конкретные примеры соответствующих и несоответствующих способов получения данных для оценки значения зависимой переменной.
8. В связи с тем, что в искусственных экспериментах редко воспроизводятся стрессовые условия реальной жизни, можно ли сделать вывод, что они не могут быть соответствующими?
9. Почему проблема «реалистичности» эксперимента обсуждалась нами на примерах столь различных в этом отношении исследований, как спасательный поиск и сравнение высотометров?

СТАТИСТИЧЕСКОЕ ПРИЛОЖЕНИЕ: ЧАСТОТНЫЕ РАСПРЕДЕЛЕНИЯ

В статистическом приложении к главе 1 значения зависимой переменной (среднее время реакции) для каждого из двух усло-

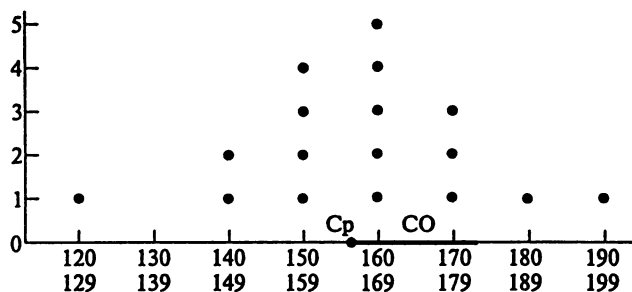


Рис. 3.6. Ось абсцисс — время реакции (по интервалам, в мс); ось ординат — частота. Ср — среднее, СО — стандартное отклонение

вий, вспышек света (А) или звучаний тона (Б), были представлены в виде гистограммы.

Более полная картина оценок ВР, полученных в эксперименте, дается распределением частот. Такое распределение показано для условия Б (звуковой тон) на рис. 3.6.

Мы видим, что в этом распределении каждая оценка представлена не всегда точно, поскольку оценки сгруппированы в классы интервалов: 120—129, 130—139, 140—149 и т.д. Величина всех интервалов в данном случае равна 10 мс.

Это та величина, на которую каждый *нижний предел* увеличивается от интервала к интервалу (например, от 150 до 160 — это 10 мс). Число интервалов здесь равно 8; соответственно имеется 8 колонок. Если бы число оценок показателей времени реакции было больше, чем 17, можно было бы использовать несколько большее число интервалов. Например, если бы было 100 проб, число используемых интервалов могло быть 15 или даже 20. При 15 интервалах нижний интервал был бы 120—124, следующий 125—129 и т.д. до 190—194. В этом случае величина интервала равнялась бы 5 мс.

Как подготовить частотное распределение

Теперь рассмотрим, как было подготовлено данное распределение частот. Во-первых, было принято решение о числе интервалов и величине интервала, а также о нижней и верхней границах. Подобранные интервалы были выписаны в столбик. Затем, начиная с пробы 1, различные показатели времени реакции *распределялись* по соответствующим интервалам. После этого записывалась частота или число показателей, попавших в данный интервал. Наконец, был составлен график распределения частот, который вы уже видели на рис. 3.6. Высота каждой колонки *X* соответствует *частоте* попадания проб в данный интервал. Все эти операции показаны в первых трех колонках табл. 3.3.

Вычисления среднего и стандартного отклонения на основе интервальных данных

Интервал	Отнесение показателя по интервалам	Частоты	Средняя точка X	Произведение средней	M_x	x	x^2	Произведение x^2 на частоту
		3	4	5	6	7	8	9
190—199		1	194,5	194,5	163	+31,5	992,25	992,25
180—189		1	184,5	184,5	163	+21,5	462,25	462,25
170—179		3	174,5	523,5	163	+11,5	132,25	396,75
160—169	— —	5	164,5	822,5	163	+1,5	2,25	11,25
150—159	— —	5	154,5	772,5	163	-8,5	72,25	361,25
140—149		1	144,5	144,5	163	-18,5	342,25	342,25
130—139		0	134,5	0	163	-28,5	812,25	0
120—129		1	124,5	124,5	163	-38,5	1482,25	1482,25

$\sum X_B = 2766,5$
 $\sum x_B^2 = 4048,25$

Используя формулу (1.1): $M_x = \frac{\sum X}{N}$; $M_B = 163$; $\sum \frac{x_B^2}{N} = 238,1$ (округленное из 162,73529) $\sigma_B = 15,4$ мс

Используя формулу (2.1): $\sigma_x = \sqrt{\frac{\sum x_B^2}{N}}$.

Вычисление среднего по данным интервальной классификации

В колонке 4 приводятся значения *средних точек* для каждого интервала. Так, средняя точка интервала 140—149 равна 144,5. Мы можем вычислить среднее методом, который пренебрегает различиями внутри каждого интервала. Во-первых, мы умножаем каждую среднюю точку на частоту внутри интервала. Это показано в колонке 5. Так, для интервала 170—179 средняя точка 174,5 умножается на частоту 3, $\sum X$ показана внизу колонки. Разделенная на N ($N = 17$), она дает среднее, равное 163, что немного отличается от величины 162, полученной сложением показателей ВР в отдельных пробах. Можно не сомневаться, что иногда эти расхождения между средними могут быть еще больше. Но если число интервалов равно 15 или больше, то совпадение бывает достаточно хорошим.

Вычисление стандартного отклонения по данным интервальной классификации

Величина стандартного отклонения вычисляется здесь в основном так же, как и по отдельным показателям ВР. В колонке 6 приводится только что вычисленное среднее. Величина x (т. е. $X - M_x$), полученная для значения средней точки каждого интервала, показана в колонке 7. Например, $194,5 - 163 = +31,5$; $144,5 - 163 = -18,5$. В колонке 8 каждое из значений X возведено в квадрат. Наконец, в колонке 9 каждая из возведенных в квадрат величин умножена на частоту в данном интервале. Например, при средней точке 174,5 и частоте 3 результат в колонке 9 равен 396,75. Это вычисление так же не учитывает различия значений внутри каждого интервала, как и вычисление среднего. Как видно, сумма в данной колонке ($\sum x^2$) равна 4048,25. Вычисление σ_x аналогично тому, как это делалось в статистическом приложении к главе 2, и дает величину 15,4 мс.

Следует заметить, что здесь приведен прямой метод вычисления среднего и стандартного отклонения по данным интервальной классификации. Это было сделано для того, чтобы вы поняли принцип — игнорирование различий внутри каждого интервала. Однако для более строгих вычислений разработаны более простые и быстрые методы.

Мы приводим здесь несколько ссылок: *Guilford J. P., Fruchter B. Fundamental Statistics in Psychology and Education* (5 th. ed.). — N. Y., 1973; *Klugh H. E. Statistics, the Essentials for Research*. — N. Y., 1970; *Wallis W. A., Roberts H. V. Statistics, a New Approach*. — N. Y., 1962.

Графическое представление среднего и стандартного отклонения

Если вы вернетесь к частотному распределению, которое приведено на рис. 3.6, вы заметите на горизонтальной оси большую точку и жирную линию. Точка показывает положение среднего 163 мс. Это немного левее средней точки интервала 160—169, т. е. 164,5 мс.

Жирная линия имеет длину 15,9 мс — величину стандартного отклонения. Мы видим, что в частотном распределении среднее отклонение представлено точкой, а стандартное отклонение — линией. В данном частотном распределении нижняя граница, равная 122, расположена на расстоянии 2,5 стандартных отклонений от среднего, равного 163. Верхняя граница, равная 194, удалена приблизительно на расстояние 2 стандартных отклонений выше среднего. Таким образом, верхняя граница удалена приблизительно на 4,5 стандартных отклонений от нижней. Это в общем-то типично для частотного распределения с малым числом оценок.

Задача: Вычислите σ_x для условия А по данным интервальной классификации.

Ответ: 18,6.

Глава 4

ЭКСПЕРИМЕНТЫ НА ПРЕДСТАВИТЕЛЬНЫХ ВЫБОРКАХ¹

Кто из вас не испытывал волнения в большом универсаме? Придя в магазин со списком нужных вам продуктов, вы, естественно, стараетесь купить их подешевле. Но это не так легко. Положим, один из пунктов вашего списка — маисовые лепешки. Из ценника вы узнаете, что они продаются в упаковках четырех разных размеров. Пакет в 13 унций стоит 76 центов, 10 унций — 59, 9 — 49, а 5 — 48 центов. А теперь ответьте, только быстро: какой пакет лучше купить? Ясно только одно — 5 унций покупать не нужно. Просто не верится — они стоят всего лишь на цент дешевле 9. Если в вашем списке 15 или 20 пунктов, то для выбора хороших покупок потребуется очень много времени, а если вы еще и нетерпеливы, то наверняка наделаете ошибок.

Вы скажете, что здесь должны быть какие-то правила. Правила, конечно, есть. Они сформулированы в «Акте об упаковке и обозначении товаров». Но, как видите, эти правила не очень помогают покупателям. Двое психологов (Gatewood R. D., Perloff R., 1973) решили провести эксперимент и найти способ, который действительно *помогал бы* при выборе покупок. По поводу упомянутого «Акта» они цитируют С. Кохана: «Правила расфасовки и определения стоимости продуктов... это один из ярких примеров полного отсутствия всяких правил» (Cohan S. E., 1969, с. 10). В эксперименте, который мы обсудим в этой главе, Гейтвуд и Перлофф сравнивали существующий способ информирования покупателей о ценах на продукты с двумя другими — усовершенствованными. Один из них был предложен продавцами бакалейных товаров, а другой — самими покупателями.

В отличие от предыдущих экспериментов, рассмотренных в нашей книге, данный эксперимент проводился не на одном или нескольких испытуемых и не с целью найти наиболее удобный способ подачи информации именно для них. В эксперименте приняли участие 75 испытуемых, а его целью было распространить полученные результаты на всех посетителей универсамов в Соединенных Штатах. Любой тип индивидуального эксперимента для этой цели явно непригоден. И конечно, возможность получить результаты, применимые не только к одному человеку, но и к целой популя-

¹ Перевод В. В. Петухова.

ции, — большое преимущество нового эксперимента. Но 75 человек — это далеко не все покупатели. Достаточно ли успешно представляет *популяцию выборка* испытуемых, чтобы считать эксперимент обладающим *внешней* валидностью? Мы постараемся ответить на этот вопрос. Но для начала укажем, что новый тип эксперимента поднимает некоторые новые вопросы в отношении внутренней валидности. В упомянутом эксперименте каждым из трех способов получения информации о ценах на продукты пользовалась отдельная группа испытуемых. С одной стороны, при такой схеме эксперимента снимаются влияния последовательности проб, которые так досаждали нам в главе 2. Но, с другой стороны, на каком основании мы будем сравнивать результаты этих трех групп испытуемых? Ведь люди разные, и группы людей — тоже.

Вы увидите, что существует немало способов проведения групповых экспериментов. Кроме отказа от совершенно неудачных схем, в этой главе не будет каких-либо универсальных рекомендаций. Наше обсуждение будет строиться так, чтобы в результате вы смогли сами принимать решение в каждом конкретном случае. Таким образом, когда вы прочтете эту главу, вы научитесь выбирать наиболее подходящую схему эксперимента, зная о тех трудностях, которые можно преодолеть, и о тех, которые все же остаются. Знакомясь с экспериментальными работами, вы сумеете оценить, насколько успешно справлялись с этими трудностями другие исследователи.

Вот те примерные темы, по которым вы будете опрошены в конце главы.

1. Возможности обобщения экспериментальных результатов при привлечении большого числа испытуемых по сравнению с индивидуальным экспериментом.
2. Преимущества схем межгрупповых сравнений над схемами эксперимента с одним испытуемым.
3. Характер различий между людьми.
4. Логика различных стратегий составления групп.
5. Систематическое смешение независимых переменных с индивидуальными различиями испытуемых.
6. Надежность в схемах межгрупповых сравнений.
7. Внешняя валидность при отборе испытуемых из популяций.
8. Внешняя валидность при проведении эксперимента на испытуемых, имеющих в наличии.

Эксперимент 1:

КАК СООБЩАТЬ ПОКУПАТЕЛЯМ О ЦЕНАХ НА ПРОДУКТЫ?

Было предложено два новых способа, помогающих выбирать покупки из упаковок разного размера. Один из них — воспользо-

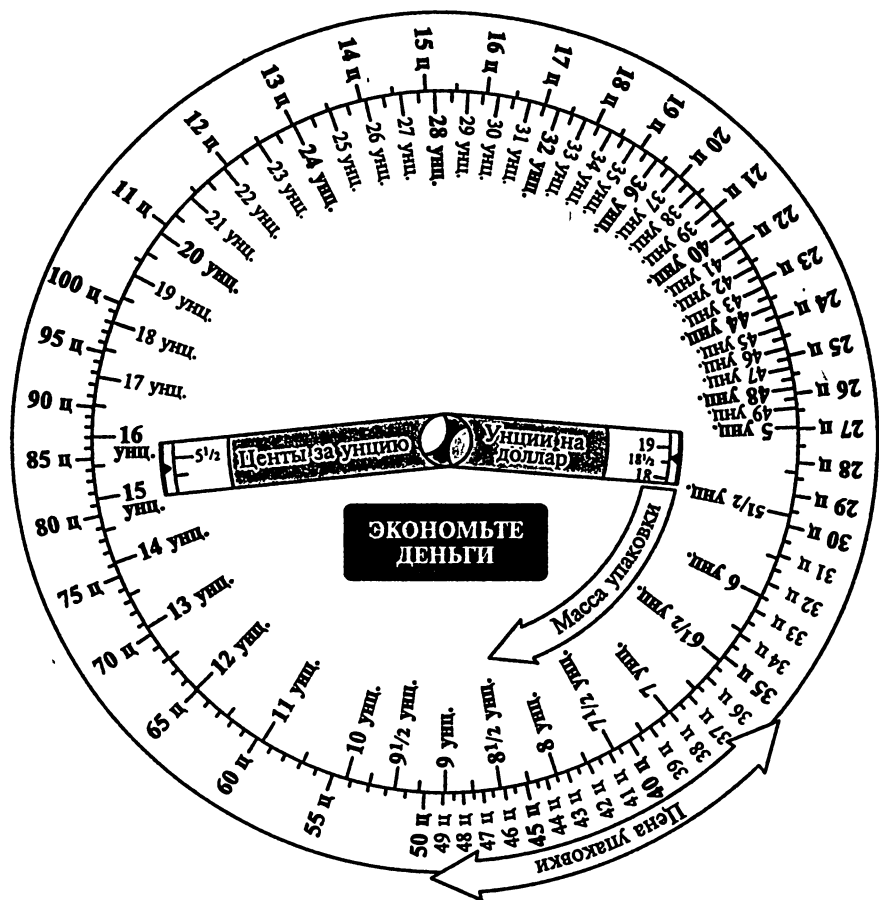


Рис. 4.1. Вычислительное устройство для определения стоимости одной унции маисовых лепешек в разных упаковках. Внутренняя шкала повернута так, что отметка «9 унций» (масса пакета) совмещена с отметкой «49 центов» (цена пакета) на внешней шкале; в левом окошке показана цена за единицу массы — $5\frac{1}{2}$ цента за унцию

ваться круглой счетной линейкой, показанной на рис. 4.1. Покупатель поворачивает внутреннюю шкалу этого устройства до тех пор, пока масса пакета не соединится с его ценой, обозначенной на внешней шкале. Тогда в первом окошке будет показана цена продукта в центах за унцию, а во втором — количество унций, которое можно купить на доллар. Как вы видите (см. рис. 4.1), если пакет с 9 унциями маисовых лепешек стоит 49 центов, то их цена за унцию — меньше $5\frac{1}{2}$ цента. Этот пакет и будет самой дешевой покупкой. Другой предложенный способ — просто поставить на пакете цену продукта за единицу массы. То есть на пакетах в 13, 10 и 9 унций вместе с их общей ценой стоял бы штамп: 5,85, 5,90 и

5,44 цента за унцию. Продавцы согласны, что столь очевидное решение довольно удачно, но отмечают, что для них оно может обойтись слишком дорого. Р. Гейтвуд и Р. Перлофф пишут: «Большинство продавцов были против последнего предложения <...> Управляющие и поставщики универсамов недовольны, — читаем мы в статье “Заботы продавца” (Век рекламы, 1969, с. 3), — ведь, с финансовой стороны подобное нововведение очень невыгодно: оно требует увеличения в два раза рабочей силы, повышения цен, и это может провалить все дело» (Gatewood R. D., Perloff R., 1973, с. 81). Они утверждают, что использование дешевого (а возможно, полезного и для покупателей) вычислительного устройства даст более удачные результаты.

Экспериментальная гипотеза

Исследователи перечисляют пять гипотез, которые проверялись в эксперименте. Для наших целей все эти гипотезы можно сокращенно представить так. Выбор будет производиться более точно и быстро, если сообщать покупателям непосредственно о цене за единицу массы, чем если давать им отдельно информацию об общей цене и размере упаковки (текущая практика), даже если использовать для подсчета вычислительное устройство. Для проверки этой гипотезы проводилось сравнение трех названных способов информирования покупателей. Экспериментальное условие А — сообщение о массе и цене пакета, как в магазине. Условие Б — то же плюс использование вычислительного устройства. И условие В — указание цены продукта за унцию.

Методика

Испытуемые. «В исследовании приняли участие 75 испытуемых-добровольцев, 64 из них — женщины. 60 из 75 закончили одногодичный курс в колледже. Возраст испытуемых: от 20 до 29 лет — 48 человек, от 30 до 39 — 17, от 40 и старше — 10» (Gatewood R. D., Perloff R., 1973, с. 82).

Испытуемых случайным образом распределили по трем группам — по группе на каждое из трех экспериментальных условий. Те, кому нужно было пользоваться вычислительным устройством, «обучались работе с ним и допускались к эксперименту после трех успешных определений цены продукта за единицу массы» (там же).

Задача. В опытах использовались наборы девяти разных пищевых продуктов. Пакеты каждого набора продуктов были пронумерованы. Испытуемого просили выбрать из каждого набора «самый дешевый пакет» (там же). Обстановка эксперимента напоминала универсам. Цены на продукты фактически повторяли цены одного из магазинов данной фирмы. Испытуемому давали листок с

вопросами, на котором он должен был написать номер выбранного им пакета для каждого набора продуктов.

Результаты

Каждый из девяти выборов, произведенных испытуемым, характеризовали по двум параметрам. Это — правильность выбора и затраченное время (в минутах). В левой части табл. 4.1 показаны среднее количество правильных выборов и стандартные отклонения для трех групп испытуемых. Видно, что по средним оценкам лучшим оказалось условие В — около восьми правильных выборов (а условия А и Б — несколько ниже 6). При условии В наименьшим является и значение стандартного отклонения, поскольку большинство испытуемых верно выполнили все девять выборов, а выше этой точки оценки располагаться не могли.

Еще большим оказалось различие во времени, затраченном на каждый из девяти выборов. Это видно по правой части таблицы. В условии В для выбора требовалось в среднем 4 мин, а в условии А — почти 24. Понятно, что работа с вычислительным устройством требовала дополнительного времени, и среднее время для условия Б составило почти 32 мин. Таким образом, для выбора покупки с помощью вычислительного устройства требуется в восемь раз больше времени, чем в том случае, когда цена продукта за унцию проставлена на пакете.

Выводы

Экспериментальная гипотеза подтвердилась. Непосредственное сообщение о цене за единицу массы позволяло выбирать покупки более точно и быстро, чем два другие способа. Использование вы-

Таблица 4.1

Количество правильных выборов и затраченное время для трех видов сообщений о ценах на продукты

Условия	Количество правильных выборов			Затраченное время		
	А	Б	В	А	Б	В
Среднее	5,78	5,96	8,04	23,93	31,72	3,60
Стандартное отклонение	1,31	1,57	0,45	10,00	9,57	1,11

Примечание. А — сообщение об общей цене и массе упаковки, Б — то же с использованием вычислительного устройства, В — о цене за унцию.

числительного устройства для более точного выбора оказывало сомнительную помощь, кроме того, его применение увеличивало время покупки.

ПЕРВОЕ ПРЕИМУЩЕСТВО ЭКСПЕРИМЕНТА С БОЛЬШИМ ЧИСЛОМ ИСПЫТУЕМЫХ: РАСПРОСТРАНЕНИЕ РЕЗУЛЬТАТОВ НА ПОПУЛЯЦИЮ

Предположим, что Р. Гейтвуд и Р. Перлофф очень торопились и не захотели утруждать себя проверкой большого числа испытуемых. Они могли бы поступить «проще» и привлечь только одного испытуемого, как в большинстве ранее описанных экспериментов. Так можно было поступить, если бы они хотели узнать, как будет выбирать покупки именно данный испытуемый. Но при этом нельзя было бы переносить результаты эксперимента на всю популяцию покупателей универсамов. Например, некоторые люди — блестящие «счетчики». Стоит показать такому два пакета маисовых лепешек, где отмечены их цена и масса, как он молниеносно выдает цену за унцию для каждого пакета — до третьего десятичного знака. Другие люди, скажем, никогда не имели дела ни с какими счетными устройствами. Чтобы вывести заключение о популяции, которую исследовали Гейтвуд и Перлофф, недостаточно привлечь одного или нескольких испытуемых. Нужна выборка испытуемых из популяции и притом необходимого размера. Правда, сейчас нам не обязательно знать, что означают слова «необходимый размер» технически, однако ясно, что 25 испытуемых для каждого из экспериментальных условий — это гораздо лучше, чем один, два или десять.

Если в эксперименте привлекается большое число испытуемых, то полученные результаты можно переносить на более широкую популяцию. В предыдущих экспериментах приложение результатов ограничивалось одним человеком — испытуемым и касалось его деятельности в дальнейшем. Но некоторые исследования просто неинтересны, если они не позволяют сделать достаточно широких обобщений. Для того чтобы после проведения эксперимента обозначение цены продукта за единицу массы было введено повсеместно, его результаты должны быть справедливы для всех покупателей. Кроме того, преимущество эксперимента с большим числом испытуемых — в его экономичности. Можно, конечно, тщательно исследовать каждого испытуемого и провести много индивидуальных экспериментов. Но если вы хотите показать справедливость полученных результатов (скажем, того, что спасательный поиск на море лучше осуществлять без бинокля) сразу для целой выборки испытуемых, то каждому из них вовсе не обязательно участвовать во всех экспериментальных условиях

(как было с Дион и Хоки). Хотя, разумеется, кто-то всегда будет продолжать настаивать на проведении индивидуальных экспериментов, поскольку Дион и Хоки в чем-то различны.

Но у эксперимента с большим числом испытуемых есть и еще одно, третье, преимущество. Благодаря количеству собранных данных и возможности применять более эффективные экспериментальные схемы, которые будут описаны ниже, исследования нового типа становятся более качественными, чем эксперименты с отдельными испытуемыми. Короче говоря, возможность распространения результатов эксперимента на популяцию дает ему тройное преимущество. Во-первых, он и ставится для изучения популяции в целом. Во-вторых, групповой эксперимент более экономичен. И, в-третьих, это более удачный эксперимент по сравнению с группой индивидуальных экспериментов.

Эксперимент 2:

МЫСЛЕННАЯ ОТРАБОТКА ДВИГАТЕЛЬНЫХ НАВЫКОВ

Мы вновь поднимаемся в воздух! Вспомните тренажер, о котором мы рассказывали в предыдущей главе, в эксперименте с ночными посадками самолета. Такие тренажеры очень дороги. Экспериментальный психолог Д. Пратер из Академии воздушных сил попытался найти более дешевый способ для обучения начинающих пилотов. Он имел дело с самолетом Т-37. А научить курсантов сажать такой самолет без тренажера, да еще за короткий срок практически невозможно — не хватит ни самолетов, ни инструкторов.

И вот при изучении литературы по формированию навыков у Пратера возникла идея о самом дешевом тренажере — тренировке мысленной. На самом деле, большинство предшествующих исследователей отмечали высокую эффективность мысленной тренировки для совершенствования таких двигательных навыков, как, например, игра в волейбол или точное выполнение штрафных бросков в баскетболе.

Что понимает Пратер под мысленной тренировкой? Прочитав его: «Мысленная тренировка имеет место тогда, когда субъект пытается ясно представить в уме те конкретные двигательные операции, из которых состоит навык» (Prater D., 1973, с. 353). Ключевая идея здесь — *умственное воображение*. Пратер решил провести эксперимент и посмотреть, поможет ли курсантам созданная им программа мысленной тренировки. Подобно Гейтвуду и Перлоффу, он хотел распространить свои результаты на всю популяцию, представителями которой и были испытуемые в его эксперименте. Правда, это не столь обширная популяция: покупателей гораздо больше, чем курсантов летных школ.

Экспериментальная гипотеза

Экспериментальная гипотеза состояла в том, что при дополнении обычных тренировочных программ специальной мысленной тренировкой у курсантов улучшится навык по выполнению посадки самолета Т-37.

Методика

Испытуемые. «Испытуемые были случайно отобраны среди курсантов, проходящих обучение по тренировочной программе «Т-37» на воздушной базе в Уильямсе. В экспериментальную группу вошли 13 человек. Еще 10 человек отобрали для контрольной группы (также случайным образом). Все курсанты были не очень опытные: 20 часов тренировок на самолете Т-41 и 4 часа — на Т-37» (там же).

Экспериментальные условия. Сначала опишем условие Б, предложенное испытуемым «контрольной группы». Это были обычные тренировки, в которых и раньше участвовали обе группы испытуемых, т.е. почти ничего (кроме, пожалуй, времени полетов) не изменялось.

Условием А для испытуемых экспериментальной группы была дополнительная программа специальных мысленных тренировок. Четыре раза за период обучения приземлению испытуемых помещали в макете кабины самолета Т-37. Макет состоял из подвижных рычагов управления и контрольной рукоятки. Через наушники испытуемым подавалась специальная магнитная запись. На ней рассказывалось о выполнении всех необходимых операций, сначала очень подробно, позднее — в самых общих чертах, т.е. «подсказки» постепенно убирались. «Испытуемых экспериментальной группы просили как можно более отчетливо представлять себе ситуацию посадки и выполнять те же действия, что и в реальном полете, вплоть до движения глаз» (там же, с. 354).

Процедура. Испытуемым экспериментальной группы (условие А) предлагали осуществлять мысленную отработку после четвертого, пятого, шестого и седьмого взлетов. Каждая такая серия продолжалась примерно 11—15 мин. Испытуемые контрольной группы (условие Б) выполняли в течение этого периода времени обычные тренировки (там же).

Результаты

Самая высокая оценка работы испытуемых равнялась 7 (лучший курсант), самая низкая — 1 (неудачник). Оценки выставлялись отдельно за знание операций (насколько хорошо понимал курсант, что ему полагается делать) и за технику их выполнения

(насколько успешно операции осуществлялись). При условии А (экспериментальная группа) средняя оценка за знание операций составила 4,53, а при условии Б (контрольная группа) — 4,26. Оценки за технику выполнения были соответственно 4,21 и 3,89.

Выводы

Экспериментальная проверка подтвердила правильность исследованной гипотезы. Основной вывод состоит в том, что мысленная тренировка улучшает реальное выполнение посадки самолета.

ВТОРОЕ ПРЕИМУЩЕСТВО ЭКСПЕРИМЕНТА С БОЛЬШИМ ЧИСЛОМ ИСПЫТУЕМЫХ: ВОЗМОЖНОСТЬ ПРИМЕНЕНИЯ СХЕМ МЕЖГРУППОВЫХ СРАВНЕНИЙ

Возможность проведения новых экспериментов

В исследовании Р. Гейтвуда и Р. Перлоффа (Gatewood R. D., Perloff R., 1973) со способами информирования о ценах на продукты применение схемы межгруппового сравнения не было обязательно. Можно было бы провести 75 индивидуальных экспериментов, предлагая каждому испытуемому все три изучаемых условия, скажем, в позиционно уравненной последовательности АБВВБА. О том, почему экспериментаторы так не поступили, мы скажем далее.

В исследовании же с мысленной тренировкой индивидуальный эксперимент просто невозможен. Вы не можете обучить одного и того же испытуемого сажать самолет двумя методами и затем сравнить эффективность каждого из них. Для каждого метода вы должны привлечь отдельную группу испытуемых!

Иначе говоря, для эксперимента с тренировкой схема межгруппового сравнения необходима. Таким образом, с привлечением большого числа испытуемых становятся возможными некоторые новые эксперименты, которые можно провести только по схемам межгрупповых сравнений.

Устранение некоторых факторов, мешающих достижению внутренней валидности

Предположим, что Гейтвуд и Перлофф применили схему эксперимента с одним испытуемым, используя *позиционно уравненную* последовательность АБВВБА. Тогда, во-первых, возникли бы эффекты последовательности, которые, как мы знаем по главе 2, зачастую просто неустранимы. Вполне возможно, что испытуе-

мый, осознав все выгоды условия В — непосредственного сообщения о цене за единицу массы, просто не захочет пользоваться вычислительным устройством в условии Б. Во-вторых, для каждого из трех исследуемых способов понадобятся разные задачи. Один и тот же список продуктов использовать нельзя. Кроме того, нужно было бы как-то уравнивать отдельные пункты в списках, применяемых для разных экспериментальных условий.

В итоге можно сказать, что эксперименты с привлечением большого числа испытуемых имеют ряд преимуществ по сравнению с индивидуальными экспериментами. Во-первых, они позволяют делать выводы о целой популяции, а не только о самом испытуемом. Во-вторых, в них можно применять схемы межгрупповых сравнений, которые для одних экспериментов очень выгодны, а для других просто необходимы.

Эксперимент 3:

ДВА МЕТОДА ОБУЧЕНИЯ ИСПАНСКОМУ ЯЗЫКУ

Рассмотрим теперь более детально, как планируются эксперименты на представительных выборках. Сделаем это на еще одном вымышленном примере. Подобный эксперимент любой из вас мог бы провести в юности. Особое внимание вы должны обратить здесь на способ распределения испытуемых по группам.

В штате Калифорния большинство студентов старших классов проходят двухгодичный курс иностранного языка (знание второго языка — условие поступления в университет). В местечке Постгейт, где всего одна средняя школа, ежегодно около 100 студентов начинают курс испанского языка. Преподавателей испанского двое — м-с Кроутерс и м-р Рамос. Оба они ведут занятия и на первом и на втором курсе по два потока каждый. И вот однажды в факультетском буфете м-р Рамос обратился к м-с Кроутерс: «Дженни, мне кажется, что за два года мы могли бы добиться гораздо больших успехов, да и студентам было бы интереснее, если бы в первый год мы не допекали их обычным чтением, письмом да грамматикой, а нашли бы что-нибудь посовременнее, вроде того нового метода, что применяют в центральном Канзасе: весь первый год там дают только разговорный язык». «Насколько я поняла вас, Примо, — ответила м-с Кроутерс, — вы предлагаете поработать сверхурочно. Ну что же, развлечений в последнее время здесь немного, и небольшая встряска мне не повредит». Короче, они решили сравнить новый метод обучения с текущим. Им удалось заинтересовать этим заведующего своего отдела, который сумел убедить заведующего учебной частью, тот — директора школы, а уж директор уговорил школьный совет (который можно уговорить на все, что не слишком дорого стоит).

Планирование эксперимента

М-с Кроутерс и м-р Рамос решили обсудить подробности своего эксперимента с м-ром Роджерсом, который прибыл в Постгейтскую среднюю школу в качестве советника и читал спецкурс по «исследованиям личности». В области психологии он был самым крупным авторитетом в школе. М-р Роджерс сказал им, что самое главное при проведении любого исследования — а это он узнал из книги «*Основы психологического эксперимента*» — подробно регистрировать весь его ход, лучше всего в переплетенной тетради с пронумерованными страницами и оглавлением.

Для начала экспериментаторы изложили в своем лабораторном дневнике каждый из методов обучения. Это было несложно. Для описания старого метода они просто воспользовались последним изданием программы курса от 17 апреля 1954 года, перечислив основные приемы обучения. Новый метод они описали по ротапринтной копии учебного пособия, которую м-р Рамос раздобыл в центральном Канзасе. Они обсудили также конкретные детали применения метода для студенческой группы. Для хранения дневника они выбрали специальное место, чтобы иметь возможность после каждого дня занятий отмечать успехи студентов и другие интересные подробности. Естественно, они продолжали вести и обычные классные журналы. Преподаватели понимали, что занятия по каждому из методов нужно стараться проводить как можно лучше и затрачивать на студентов одно и то же количество времени. То, насколько успешно выполняли они свои намерения, также отмечалось в дневнике.

Не забыли экспериментаторы заранее подумать и о том, как они будут оценивать знание языка спустя два года. Думали долго, а ответ оказался неожиданно простым. Если студент может понять вопрос, заданный ему по-испански, и по-испански же ответить на него письменно, то значит, он способен как разговаривать, так и писать на этом языке. Научиться хорошо говорить на иностранном языке, пройдя школьный курс, довольно трудно. А перевод — еще не подлинное знание языка, и это было ясно обоим преподавателям. Выбранные задачи-вопросы, конечно, не позволяют полностью оценить речевые навыки, сформировавшиеся у студентов за почти двухлетний срок, но как *критерий для сравнения результатов* эксперимента они вполне подходят. Перечень контрольных вопросов также вошел в лабораторный дневник.

Разумеется, эксперимент проводился со студентами, изучавшими испанский впервые. Сначала возникла идея распределить их по группам в случайном порядке: по каждому из преподавателей и затем по каждому методу обучения, которые будут называться теперь *письменным* и *разговорным*. Но тут есть одна сложность: некоторые студенты уже немного знали язык — они жили в райо-

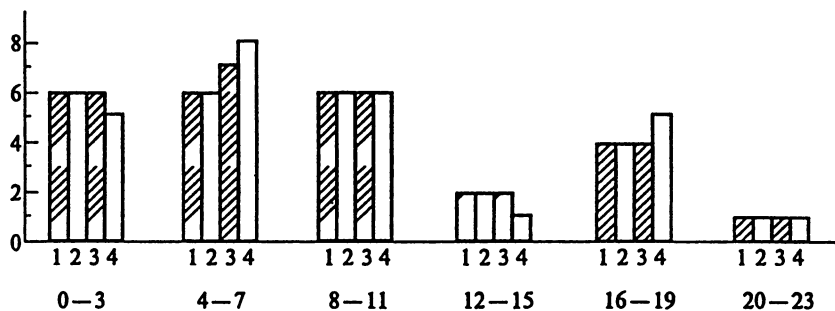


Рис. 4.2. Диаграммы результатов распределения студентов по четырем группам согласно оценкам, полученным в предварительной проверке. Ось абсцисс: первая строка цифр — номера групп; вторая — оценки студентов в предварительной проверке; ось ординат — количество студентов

нах, где говорят по-испански. Вполне возможно, что при случайном распределении в какие-то из четырех групп попадет слишком много таких студентов, а в другие — очень мало. Поэтому экспериментаторы решили провести *предварительную проверку*.

Студентам дали списки из 50 испанских слов и попросили их письменно перевести эти слова на английский. Затем, согласно полученным оценкам, сто студентов были упорядочены по полученным баллам: от самых высоких баллов к самым низким. Если оценки студентов были равными, то их фамилии записывали просто по алфавиту. После этого нужно было избрать стратегию для распределения студентов по четырем группам: 1) Рамос, письмо; 2) Рамос, разговор; 3) Кроутерс, письмо; 4) Кроутерс, разговор. Первые четыре фамилии списка нумеровали в названной последовательности, следующие четыре — в обратном порядке и т.д. Таким образом, перед каждой фамилией был указан номер группы. Распределение студентов по четырем группам показано на рис. 4.2. Видно, что группы очень похожи и что во всех группах у основной массы студентов весьма небольшой словарный запас. Полный список студентов приведен в табл. 4.2.

Таблица 4.2

Распределение студентов по группам в соответствии с их оценками, полученными в предварительной проверке

Оценка	Группа	Фамилия	Оценка	Группа	Фамилия
23	1	Мендоса	19	4	Диас
22	2	Лира	19	3	Руис
22	3	Сарагоса	18	2	Авила
21	4	Оливера	18	1	Бесерра

Оценка	Группа	Фамилия	Оценка	Группа	Фамилия
18	1	Крус	9	2	Келли
18	2	Гуттьерес	9	1	Левин
18	3	Миллер	9	1	Май-Дональд
18	4	Нунес	8	2	Барнес
18	4	Зандовал	8	3	Кровел
17	3	Аргуэллес	8	4	Грэхам
17	2	Крус	8	4	Хавкинс
17	1	Гузман	8	3	Келли
17	1	Жорес	8	2	О'Брайен
17	2	Робинсон	8	1	Миллер
16	3	Якобс	8	1	Шапиро
16	4	Морено	8	2	Уотсон
16	4	Санчер	7	3	Александр
15	3	Инграм	7	4	Д'Амбуаз
15	2	Торрес	7	4	Фукиното
14	1	Рос	7	3	Хаскел
13	1	Хирасима	7	2	Джонс
12	2	Ламб	7	1	Леви
11	3	Сохен	7	1	Мазетти
11	4	Ховертон	7	2	Ойзен
11	4	Смит	7	3	Смит Б.
10	3	Самри	7	4	Смит Г.
10	2	Джонс	7	4	Циммерман
10	1	Ким	6	3	Христенсен
10	1	Валентин	6	2	Даниелс
10	2	Вондерски	6	1	Фишер
9	3	Брэг	6	1	Грехем
9	4	Дуглас	6	2	Джонс
9	4	Гилберт	6	3	Макинтош
9	3	Келлер	6	4	Сандерсон

Оценка'	Группа	Фамилия	Оценка	Группа	Фамилия
6	4	Уи	1	2	Купер
5	3	Кронхайт	1	1	Айленд
5	2	Форрестер	1	1	Мак-Килош
5	1	Фуимото	1	2	Мейер
5	1	Портер	1	3	Шварц
5	2	Янг	1	4	Ван Дейк
4	3	Гарнер	0	4	Бакер
4	4	Мецгер	2	4	Мартин
4	4	Зилиотто	2	3	Вохан
3	3	Андерсон	0	3	Карпентер
3	2	Джонсон	0	2	Дейч
3	1	Петрини	0	1	Кеннеди
3	1	Вудс	0	1	Кирк
2	2	Девис	0	2	Пинеро
2	3	Кац	0	3	Рейес
2	4	Джонсон	0	4	Вильямс

Проведение эксперимента

Занятия с применением каждого из методов обучения и составляли эксперимент. И хотя результаты исследования можно получить лишь после окончания курса, в течение всего учебного периода важно было применять каждый метод одинаково старательно и успешно, а также фиксировать любые наблюдения, которые могли оказаться важными для их последующей оценки. Например, каждый раз нужно было отмечать степень заинтересованности студентов. Могли встречаться также и необычные обстоятельства. Иногда несколько непослушных студентов могут помешать проведению занятий по любому методу обучения.

После окончания занятий был проведен итоговый экзамен. Материал для него заготовили заранее. На магнитофон записали вопросы по-испански. Чтобы исключить влияние дополнительных факторов (например, знакомый или незнакомый голос), вопросы студентам задавал тот преподаватель, у которого они занимались. Экзамен проходил в лингафонном кабинете. Каждый студент отвечал на 100 вопросов в два приема. После каждого вопроса сту-

денту давалось 30 секунд, чтобы записать ответ на листочке бумаги, где стоял номер вопроса и кодовый номер испытуемого. Каждый преподаватель проверял работы студентов другого преподавателя и поэтому узнавание почерка, а возможно, и других индивидуальных черт также было исключено. Работы студентов, занимавшихся по разным методам, складывались в одну пачку и перемешивались. Ответы на каждый вопрос оценивались по шкале от 5 до 0. О содержании каждой оценки преподаватели договорились заранее: «5» — безусловно; «4» — одна незначительная ошибка; «3» — одна грубая ошибка или две незначительных; «2» — несколько грубых ошибок; «1» — в целом перевод неверен, но есть некоторые совпадения; «0» — неверно вообще. Наибольшее количество баллов, которые мог набрать студент, ответив на 100 вопросов, — 500. Ради удобства общая оценка делилась на 5 и округлялась до ближайшего целого числа. Таким образом, наивысший возможный общий балл — 100. Это — критерий оценки работы студента.

Анализ результатов

Было установлено, что у студентов м-с Кроутерс и м-ра Рамоса, обучавшихся по одному и тому же методу, оценки в среднем примерно одинаковы. Это дало возможность представить результаты всего эксперимента в двух частотных распределениях (рис. 4.3):

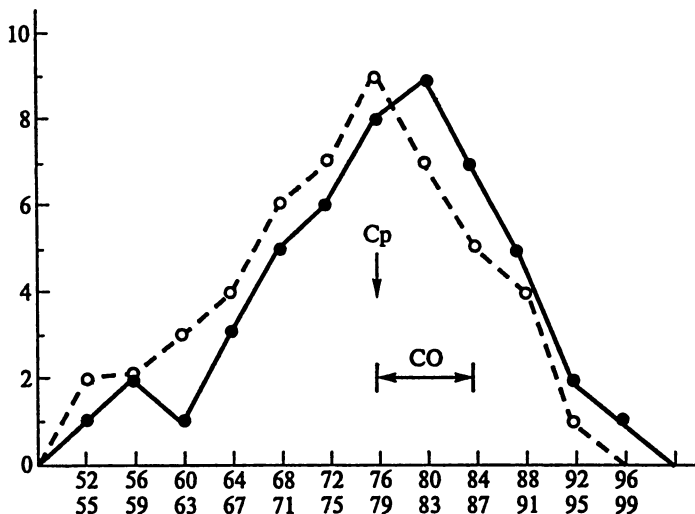


Рис. 4.3. Результаты эксперимента с методами обучения испанскому языку: частотное распределение (показаны средняя оценка (C_p) и величина стандартного отклонения (CO) для разговорного метода). Ось абсцисс — экзаменационные оценки; ось ординат — количество студентов, получивших оценки

одно — для разговорного метода, другое — для письменного. Фактически здесь просто показано, сколько студентов получили оценки от 52 до 99. Для более ясной картины оценки сгруппированы в интервалы по четыре. Так, во второй по возрастанию интервал включены студенты, получившие оценки 56, 57, 58 и 59. За ним идут студенты с оценками 60, 61, 62 и 63; в этот интервал попали три студента, обучавшихся по разговорному методу, и только один, обучавшийся по письменному. Можно убедиться, что у разговорного метода есть небольшое преимущество: его частотное распределение несколько выше — смещено вправо — по сравнению с письменным. Правда, средняя оценка для разговорного метода составила 78 баллов, в то время как для письменного — 75. Это различие не очень показательно, и, уж, конечно, оно вряд ли убедит школьный совет тратить деньги на нововведения. Но Кроутерс и Рамос по-прежнему считают, что в новом методе обучения что-то есть. Быть может, им не хватало опыта для его правильного применения. Но вот однажды в школьном коридоре их приветствовали словами: «Благодарим вас за наши успехи в испанском, новый метод нам очень понравился. Грация!»

ДОСТИЖЕНИЕ ВНУТРЕННЕЙ ВАЛИДНОСТИ В ЭКСПЕРИМЕНТАХ С МЕЖГРУППОВЫМ СРАВНЕНИЕМ

Новая угроза для внутренней валидности: индивидуальные различия испытуемых

Эксперимент с межгрупповым сравнением, вроде только что описанного, не может быть идеальным, но по другой причине, чем эксперимент с одним испытуемым. В идеальном эксперименте *один и тот же* испытуемый или *одна и та же* группа испытуемых *одновременно* работает в различных условиях независимой переменной. В реальности индивидуальный эксперимент ухудшается из-за того, что испытуемый не может работать в различных экспериментальных условиях одновременно. А в реальном эксперименте с межгрупповым сравнением группы могут работать одновременно, но предъявить разные экспериментальные условия одной и той же группе невозможно.

Теперь давайте рассмотрим природу межиндивидуальных вариаций, а затем обсудим, что можно сделать для сокращения влияния этих вариаций на внутреннюю валидность эксперимента.

Известные характеристики

У испытуемых любой потенциальной группы есть довольно много различий, выделить которые очень легко. В только что опи-

санном эксперименте, например, участвовали юноши и девушки. С этим ясно. Другая известная характеристика — возраст. Однако в подобных индивидуальных различиях не всегда можно сразу установить четкие градации. Так, в эксперименте с испанским для определения различий испытуемых по их словарному запасу потребовалась предварительная проверка. Нельзя утверждать, что абсолютно все известные индивидуальные характеристики имеют отношение к изучаемому виду деятельности, но на практике таковой может оказаться почти каждая из них. Длина волос, например, может свидетельствовать об образе жизни испытуемого, а тот, в свою очередь, — влиять на стиль обучения. Место фамилии испытуемого в алфавитном списке может указывать на его этническое происхождение. Трудно найти такую характеристику, которая так или иначе не была бы связана с деятельностью, исследуемой в эксперименте.

Вывод из всего этого в отношении экспериментов с межгрупповым сравнением вполне ясен. Экспериментатор должен быть очень внимательным к индивидуальным различиям при разделении на группы. Можно сказать точно, что если студентов в аудитории разделить на две экспериментальные группы по тому, близко или далеко они предпочитают сидеть от доски, то исследование наверняка будет ненадежным.

Несистематические вариации

Дайте любой группе людей какой-нибудь тест — и вы получите определенное распределение ответов. Чаше всего оно будет иметь куполообразную форму. Так распределились оценки студентов, прошедших предварительную проверку на знание слов испанского языка. То же было справедливо и для покупателей, пользовавшихся вычислительным устройством. Такую же форму имело распределение времени научения, необходимого для правильного выполнения задач на различение в опытах с крысами. В большинстве случаев нельзя заранее выделить те индивидуальные различия, которые объясняли бы подобный разброс. Эти различия непостоянны и изменчивы.

Какую-то часть несистематических вариаций можно отнести за счет изменчивости поведения каждого испытуемого во времени или нестабильности в оценке этого поведения, о чем мы говорили в главе 2. Во-первых, в любой конкретной пробе одни испытуемые будут более работоспособны, другие — менее и т. д. Во-вторых, в каждый конкретный замер обычно попадает лишь какая-то часть изучаемого поведения, и поэтому данная оценка работы одних испытуемых может оказаться удачной, а других — неудачной. В отличие от индивидуальных экспериментов, в экспериментах с межгрупповым сравнением подобная изменчивость сглажи-

вается за счет увеличения количества отдельных действий, оцениваемых в одном замере. В эксперименте с испанским языком это достигалось двумя путями. Во-первых, каждый студент участвовал в опытах на протяжении двух лет (при одном из условий независимой переменной). Таким образом, обучение происходило как в благоприятные, так и в неблагоприятные периоды. Во-вторых, они проводились через длительный тест. Возможно, что для еще большей минимизации факторов во времени было бы лучше повторно проводить тесты по 2—3 в месяц.

Но независимо от того, насколько удастся сократить изменчивость результатов у каждого отдельного испытуемого, все же сохраняются вариации между испытуемыми, причем те, которые мы не сможем объяснить установленными индивидуальными различиями. Причины этих вариаций могут быть разными. Все еще обсуждается извечный вопрос о соотношении врожденных и приобретенных различий. Правда, экспериментатора больше интересуют сами вариации, чем их теоретическое объяснение. Далее, такие вариации могут оказаться связанными с самим экспериментом. Может случиться так, что для большинства испытуемых более эффективным окажется разговорный метод обучения испанскому языку, а кому-то больше поможет письменный. И наконец, есть вариации, возникающие в самом ходе эксперимента. Так, могут меняться личные отношения между студентом и преподавателем, испытуемым и экспериментатором.

Таким образом, разброс данных по группе испытуемых может быть обусловлен как изменчивостью поведения каждого испытуемого, так и различиями между испытуемыми по их «среднему» уровню. Чтобы представить общую картину несистематических вариаций, нужно иметь в виду оба их источника. Если в каждом из экспериментальных условий работает отдельная группа испытуемых, то всегда есть опасность, что полученные различия в результатах объясняются не условиями, а индивидуальными особенностями испытуемых, случайно попавших в каждую из групп.

Еще раз о систематическом смешении и ненадежности

Итак, основным препятствием для достижения внутренней валидности в эксперименте с межгрупповым сравнением является потенциальная возможность влияния на полученные результаты индивидуальных различий испытуемых в каждой из групп. Из-за этих влияний подлинные воздействия экспериментальных условий могут и вовсе не проявиться. Если группы составлены так, что между ними сохраняется устойчивое различие (например, в одной группе юноши, в другой — девушки), то *систематическое смешение* независимой переменной с индивидуальными особенностями

ностями испытуемых неизбежно. А случайные различия между испытуемыми в группе, обусловленные ограниченным размером выборки, будут снижать *надежность* значений тех зависимых переменных, которые исследуются в эксперименте.

Как мы знаем, по своей внутренней валидности реальные эксперименты не могут быть безупречными. Однако реальные группы испытуемых можно сравнивать с такими группами, которые участвовали бы в идеальном и бесконечном эксперименте. Эти мысленные, неосуществимые эксперименты уже знакомы нам по схемам эксперимента с одним испытуемым. В *идеальном* эксперименте с межгрупповым сравнением членами обеих групп должны быть одни и те же люди. Только тогда полученная разница между условиями независимой переменной совершенно не будет связана с индивидуальными различиями испытуемых каждой группы. Однако на практике достичь этого невозможно, так как испытуемые каждой группы — разные люди. И все же нужно стараться сделать группы как можно больше похожими друг на друга, подбирая их буквально испытуемый к испытуемому.

Для проведения *бесконечного* эксперимента с межгрупповым сравнением нужно, чтобы обе группы состояли из бесконечного числа *испытуемых* (причем способ составления групп не должен порождать систематического смещения). Мы уже советовали вам привлекать для эксперимента большое число испытуемых. Как вы помните, в эксперименте с одним испытуемым для достижения высокого уровня надежности с увеличением разброса данных требуется больше проб. А в эксперименте с межгрупповым сравнением соответствующим требованием будет увеличение количества испытуемых. Кроме того, надежность индивидуального эксперимента можно повысить путем усреднения изменчивости во времени. Поскольку какая-то часть различий между испытуемыми целиком зависит от субъективных особенностей, этот метод применяется и при межгрупповом сравнении. Эти различия уменьшаются также в том случае, если для всех испытуемых будут обеспечены примерно равные условия проведения эксперимента. Если в эксперименте с испанским языком применение того или другого метода обучения к разным студентам как-то различалось, то нужно ожидать подобных различий и в успешности обучения. В индивидуальных экспериментах соблюдение строгости в отношении всех необходимых аспектов — независимой, зависимой и дополнительных переменных — будет увеличивать надежность, а в новом случае ее можно повысить путем сокращения межгрупповых различий, уравнивая условия работы испытуемых. И если добиться этого, то высокая надежность эксперимента может быть достигнута и не на столь большом количестве испытуемых.

ТРИ СТРАТЕГИИ СОСТАВЛЕНИЯ ГРУПП, ТРИ СХЕМЫ СРАВНЕНИЯ

Сейчас мы рассмотрим три эффективные стратегии набора различных групп для разных экспериментальных условий. Они называются *схемами сравнения групп*. Мы посмотрим, насколько успешно устраняются в каждой из схем два основных препятствия для достижения внутренней валидности эксперимента — систематическое смещение и ненадежность. Первые три схемы служат для *распределения* испытуемых по группам. И далее мы покажем, как на базе двух из них можно создать две другие схемы, применяемые для *отбора* испытуемых из популяции.

Случайная стратегия

В двух экспериментах, описанных в этой главе, применялась случайная стратегия. В эксперименте с сообщением о ценах на продукты участвовали 75 испытуемых-добровольцев. Случайным образом их распределили в соответствии с тремя экспериментальными условиями, т. е. каждый участник эксперимента мог с одинаковой вероятностью попасть в любую из экспериментальных групп.

Для такого распределения можно использовать таблицу случайных чисел (табл. 4.3). Фамилии 75 испытуемых записывают в алфавитном порядке (чтобы упростить их учет). Затем берут пачку карточек, пишут фамилию первого человека на верхней карточке и т. д. Потом решают, с какой колонки таблицы начать распределять испытуемых по группам. Бросают игральную кость, и если она выпадает тремя очками вверх, значит, нужно начать с третьей колонки таблицы (колонка 9—12). Так, Ааронсон (первая карточка) помечается верхним числом в колонке 5901. Адамс (следующая карточка) помечается числом 4310, стоящим ниже. Когда эта колонка закончится, переходят к следующей, и так до тех пор, пока всем 75 испытуемым не будут присвоены номера из таблицы случайных чисел.

Затем экспериментаторы складывают карточки согласно полученным номерам — по возрастанию. 25 карточек с начальными номерами кладут в стопку «условие А», следующие 25 — в стопку «условие Б» и 25 карточек с последними номерами в «условие В». И наконец, рядом с каждой фамилией в алфавитном списке ставят буквы А, Б и В. Если ваши испытуемые не имеют фамилий (как, например, белые крысы), присвойте им любые имена (скажем, А-7, В-6, М-4), чтобы суметь составить упорядоченный список.

В преимуществах случайной стратегии вы сможете убедиться на следующем примере. Первых 25 человек, пожелавших участвовать в

Таблица 4.3

Таблица случайных чисел

Ряды	Колонки																			
	1—4	5—8	9—12	13—16	17—20	21—24	25—28	29—32	33—36	37—40										
1	23	15	75	48	59	01	83	72	59	93	76	24	97	08	86	95	23	03	67	44
2	05	54	55	50	43	10	53	74	35	08	96	61	18	37	44	10	96	22	13	43
3	14	87	16	03	50	32	40	43	62	23	50	05	10	03	22	11	54	38	08	34
4	38	97	67	49	51	94	05	17	58	53	78	80	59	01	94	32	42	87	16	95
5	97	31	26	17	18	99	75	53	08	70	94	25	12	58	41	54	88	21	05	13
6	11	74	26	93	81	44	33	93	08	72	32	79	73	31	18	22	64	70	68	50
7	43	36	12	88	59	11	01	64	56	23	93	00	90	04	99	43	64	07	40	36
8	93	80	62	04	78	38	26	80	44	91	55	75	11	89	32	58	47	55	25	71
9	49	54	01	31	81	08	42	98	41	87	69	53	82	96	61	77	73	80	95	27
10	36	76	87	26	33	37	94	82	15	69	41	95	96	86	70	45	27	48	38	80
11	07	09	25	23	92	24	62	71	26	07	06	55	84	53	44	67	33	84	53	20
12	43	31	00	10	81	44	86	38	03	07	52	55	51	61	48	89	74	29	46	47
13	61	57	00	63	60	06	17	36	37	75	63	89	51	23	35	01	74	69	93	14
14	31	35	28	37	99	10	77	91	89	41	31	57	97	64	48	62	58	48	69	19
15	57	04	88	65	26	27	79	59	36	82	90	52	95	65	46	35	06	53	22	54

Ряды	Колонки										
	1—4	5—8	9—12	13—16	17—20	21—24	25—28	29—32	33—36	37—40	
16	09 24	34 42	00 68	72 10	71 37	30 72	97 57	56 09	29 82	76 50	
17	97 95	53 50	18 40	89 48	83 29	52 23	08 25	21 22	53 26	15 87	
18	93 73	25 95	70 43	78 19	88 85	56 67	16 68	26 95	99 64	45 69	
19	72 62	11 12	25 00	92 26	82 64	35 66	65 94	34 71	68 75	18 67	
20	61 02	07 44	18 45	37 12	07 94	95 91	73 78	66 99	53 61	93 78	
21	97 83	98 54	74 33	05 59	17 18	45 47	35 41	44 22	03 42	30 00	
22	89 16	09 71	92 22	23 29	06 37	35 05	54 54	89 88	43 81	63 61	
23	25 96	68 82	20 62	87 17	92 65	02 82	35 28	62 84	91 95	48 83	
24	81 44	33 17	19 05	04 95	48 06	74 69	00 75	67 65	01 71	65 45	
25	11 32	25 49	31 42	36 23	43 86	08 62	49 76	67 42	24 52	32 45	

эксперименте, исследователи могли бы включить в группу условия А, вторых 25 — в Б, а третьих — в В. Тогда между тремя группами испытуемых было бы очевидное и устойчивое различие, связанное с *очередностью их записи на участие в эксперименте*. Таким образом, неудачный способ составления групп привел бы к систематическому смещению независимой переменной (способа сообщения о ценах) с другой переменной (индивидуальными различиями испытуемых). Применение же случайной стратегии позволяет сделать влияние этих различий не столь систематичным.

Но если количество испытуемых в каждой группе не будет достаточно большим, то применение случайной стратегии не обеспечит эквивалентности групп. Предположим, что в эксперименте с сообщением о ценах участвовало 15 испытуемых, разделенных на три группы.

В результате случайного выбора в группу условия А могли бы попасть 5 самых лучших испытуемых. И тогда мы не удивились бы, если бы другой эксперимент на 15 испытуемых дал совершенно иные результаты. Из-за недостатка испытуемых такие эксперименты были бы ненадежны. Эксперимент на 75 испытуемых, по 25 в группе, обладает гораздо большей надежностью. Результаты эксперимента на 300 испытуемых были бы еще более достоверными (т. е. надежными). Таким образом, основное условие применения случайной стратегии — это достаточно большое число испытуемых в каждой экспериментальной группе. Первая схема эксперимента с межгрупповым сравнением называется *случайным распределением групп*.

Стратегия подбора пар

Этот способ составления групп связан с выделением очевидных и подлежащих градации индивидуальных характеристик, причем эти характеристики должны быть связаны с исследуемым видом деятельности. Подбираются пары испытуемых, максимально сходных друг с другом по данным характеристикам, а затем каждый из них зачисляется в одну из экспериментальных групп. Показательным примером этого способа был эксперимент с испанским языком. Необходимые индивидуальные различия были установлены в результате предварительной проверки испытуемых на знание испанских слов. Вторая схема эксперимента с межгрупповым сравнением называется *парным распределением групп*.

Ничто не помешало бы исследователям использовать в эксперименте с испанским случайную стратегию. Какая же стратегия лучше? Обе равно хороши для устранения систематического смещения независимой переменной с индивидуальными различиями испытуемых. Ни в том, ни в другом случае индивидуальные характеристики членов каждой группы (занимающихся по одному

из методов) не будут отличаться от характеристик другой группы систематически. Стратегия подбора пар позволит достичь большого подобия групп с одним и тем же числом испытуемых, если основание попарного сравнения (здесь — предварительная проверка) действительно связано с изучаемым видом деятельности (здесь — обучение испанскому языку). Но даже если предположения экспериментатора неверны и выделенные им характеристики не соответствуют виду деятельности, изучаемому в эксперименте, — ничего страшного не произойдет. Ведь в отношении всех остальных индивидуальных характеристик группы набирались случайно, и их подобие будет не хуже, чем при простом использовании случайной стратегии. Опасность возникает тогда, когда экспериментатор слишком полагается на подбор пар, привлекая небольшое число испытуемых, а связь сравнительных характеристик с изучаемым видом деятельности является при этом *недостаточной*.

Однако различие между двумя названными схемами (в отношении внутренней валидности эксперимента) не так уж существенно по сравнению с их огромным превосходством над другой схемой — *использованием реальных групп*. Примером мог бы служить случай, когда преподаватели испанского языка решили бы проводить эксперимент с разговорным методом в одной школе, а с письменным — в другой. Основной характеристикой испытуемых был бы при этом лишь сам факт учебы в школе. Но мы никак не можем рассчитывать на то, что студенты двух школ будут одинаковы по способности к изучению испанского языка. Ведь они живут в разных условиях, воспитываются в разных семьях, имеют разный опыт обучения перед поступлением в школу и т. д. Да и преподаватели в разных школах разные. Но даже если в двух школах занятия вел бы один и тот же преподаватель, систематическое смещение независимой переменной с индивидуальными различиями испытуемых было бы совершенно неизбежным.

Стратегия случайного распределения слоев

Эту стратегию можно назвать смешанной, поскольку в ней объединяются принципы подбора пар и случайного выбора. В эксперименте с испанским языком для выделения двух «слоев» можно было использовать очевидную индивидуальную характеристику — пол испытуемых, юноши и девушки. Существуют данные о том, что девушки в среднем более успевают в обучении языкам, чем юноши, поэтому желательно иметь равное число юношей и девушек в каждой из экспериментальных групп. Если из 100 студентов 56 — девушки, то 28 из них обучались бы по разговорному методу, а другие 28 — по письменному. Подобным образом были бы разделены на две равные группы и 44 юноши.

Для распределения испытуемых *внутри* каждого слоя (здесь — юношей и девушек) используется случайная стратегия. Метод случайного распределения, описанный на примере эксперимента с ценами, применяется к 56 девушкам и 44 юношам. Третья схема эксперимента, с межгрупповым сравнением, называется *случайным распределением групп с выделением слоев*.

Если различие между слоями связано с изучаемым видом деятельности, то данная стратегия будет иметь преимущество по сравнению с простым случайным распределением. Для достижения столь же высокой надежности эксперимента потребуется меньшее количество испытуемых. Если же такая связь отсутствует, то случайное распределение слоев даст те же результаты, что и обычная случайная стратегия. Поскольку выделение слоев — это один из вариантов подбора пар, мы надеемся, что остальные его особенности уже известны вам по предыдущему разделу.

ВНЕШНЯЯ ВАЛИДНОСТЬ: ПРЕДСТАВЛЕННОСТЬ ИЗУЧАЕМОЙ ПОПУЛЯЦИИ

Эффективность любого из описанных нами экспериментов определяется тем, в какой мере его результаты можно распространить на интересующую исследователя популяцию: покупателей универсамов; курсантов, которые будут учиться сажать самолет Т-37; студентов, которым предстоит изучать испанский язык в высшей школе Постгейта. Нам понятно, что выборка испытуемых из популяции не может представлять последнюю *безупречно*. Такой она могла бы стать лишь в безупречном эксперименте, а он, как мы знаем, неосуществим. В *эксперименте полного соответствия* испытуемые фактически и были бы той популяцией, к которой затем будут применяться полученные результаты. С этой точки зрения внешняя валидность реального эксперимента зависит от того, насколько мы приближаемся к столь недостижимой цели благодаря своим способам отбора испытуемых.

Данный аспект вопроса о внешней валидности можно было бы обсуждать на примере эксперимента с привлечением одной и той же группы испытуемых для разных условий (когда оба условия предъявляются каждому испытуемому в последовательности АББА). Но мы продолжим анализ схем экспериментов с межгрупповым сравнением, это позволит нам лучше соотнести вопросы внешней и внутренней валидности.

Отбор из популяции

Популяций существует очень много. Многие будущие покупатели, летчики и студенты, которые будут изучать испанский, пока еще дети или даже не родились. С другой стороны, более и менее

обширные популяции, из которых можно выбирать испытуемых, уже *существуют*, например все студенты Соединенных Штатов, начавшие изучать курс испанского языка. В одном из трех исследований этой главы — в эксперименте с мысленной тренировкой — использовалась относительно небольшая популяция. Но, конечно, курсантов на воздушной базе в Уильямсе гораздо больше, чем можно было привлечь в качестве испытуемых для эксперимента. Поэтому случайная стратегия применялась в этом случае не для распределения всех имеющихся в распоряжении курсантов, а для *отбора* экспериментальных групп из данной популяции.

Случайный отбор групп

Сейчас можно подробно описать процедуру отбора групп (с равным числом испытуемых для каждого из экспериментальных условий), которой воспользовался Д. Пратер. Это четвертая схема межгруппового сравнения — *случайный отбор групп*.

Предположим, популяция включает 810 курсантов, и из нее надо отобрать 30 человек, по 15 в каждую экспериментальную группу. Как нужно действовать в этом случае? Сначала фамилии 810 курсантов записывают по алфавиту, и каждому присваивается порядковый номер от 1 до 810. Затем, как всегда, бросают игральную кость и определяют, с какой колонки табл. 4.3 начать отбирать испытуемых. Выпадает, скажем, 5, и начинают с пятой колонки сверху (17 — 20). При этом в каждом числе учитывают только три последних знака из четырех. Первое число в колонке — 5993, и три последних знака дадут 993. Это число указывает вам порядковый номер курсанта, которого нужно отобрать для эксперимента. Но такого курсанта нет, их всего 810. Тогда нужно просто перейти к следующему числу — 3508. Из алфавитного списка отбирается курсант под номером 508. Следующим будет курсант под номером 223 и т.д. до тех пор, пока из 810 курсантов не будут отобраны 30 человек для эксперимента.

Если номер случайно повторится, скажем, еще раз встретится 508, его нужно пропустить и перейти к следующему. Зачисляя курсантов поочередно в одну из двух групп, вы получите по 15 испытуемых для каждого экспериментального условия. Чтобы знать, сколько людей отобрано, лучше всего ставить цифру 1 около фамилии первого отобранного курсанта, 2 — около второго и так к каждому до 30.

Оценка внешней валидности. Приведенная процедура обеспечивает вам случайный отбор 30 испытуемых из популяции в 810 человек. Достаточно ли велика такая выборка? Это зависит от степени различий среди популяции курсантов по их способности учиться сажать самолет Т-37. Поскольку перед допуском к тренировочной программе курсанты проходят тщательную проверку, у нас есть

все основания считать, что группа довольно однородна, слишком неудачных испытуемых в ней нет. Поэтому даже небольшая выборка из 30 человек может оказаться достаточной. Мы имеем право рассчитывать на результаты, близкие к результатам эксперимента полного соответствия, где в обоих экспериментальных условиях участвовали бы все члены популяции (эксперимент, впрочем, совершенно бесполезный). Популяция хорошо представлена выборкой испытуемых, поэтому внешняя валидность эксперимента (в данном смысле) удовлетворительна.

Однако всегда есть опасность, что при более широком распространении результатов внешняя валидность окажется нарушенной. Если распространить результаты эксперимента на будущих курсантов-пилотов, может возникнуть вопрос: представительна ли популяция из 810 курсантов, обучающихся сейчас, для популяции курсантов, которые будут учиться через несколько лет? Знать этого наверняка мы не можем, необходимо провести сравнение вступительных экзаменов, сопоставить оценки курсантов по тестам на способности и т. д. Если эти данные сильно изменяются, то эксперимент окажется *не соответствующим* по параметру испытуемых. Легко убедиться, что индивидуальные вариации — одна из тех «дополнительных» переменных, уровень которой должен всегда быть соответствующим. Этот случай хорошо сравнить с «неправильным» вариантом эксперимента Джека Моцарта, в котором тот пытался применить результаты, полученные на вальсах, к будущему разучиванию сонат. Там дополнительной переменной, находившейся на несоответствующем уровне, был тип музыки; здесь ею могут стать индивидуальные характеристики испытуемых.

Связь внешней и внутренней валидности при отборе из популяции

Схема межгруппового сравнения. При использовании схемы эксперимента, которая применялась в исследовании с мысленной тренировкой, когда разные экспериментальные условия даются разным группам испытуемых, существует связь между внешней и внутренней валидностью. Если не только вся выборка, но и отдельные группы испытуемых, отобранные для каждого из условий, достаточно хорошо представляют изучаемую популяцию, то эти группы можно считать эквивалентными. Едва ли они будут соответствовать популяции, не будучи подобными одна другой. Поэтому в экспериментах данного типа достижение внешней валидности (в отношении испытуемых) обеспечивает один из важных аспектов внутренней валидности. Ведь основным источником нарушения внутренней валидности в эксперименте с межгрупповым сравнением являются различия между испытуемыми каждой группы.

Напротив, для достижения внутренней валидности эксперимента, т. е. уравнивания групп испытуемых, участвующих в каждом из условий, вовсе не обязательно, чтобы эти испытуемые были представительной выборкой исследуемой популяции. Предположим, что для эксперимента отобрали тех 30 курсантов, которые первыми вошли в класс для занятий. После этого их можно разделить на две группы с помощью любого ранее описанного приема, в том числе путем случайного распределения, но представительной выборкой популяции курсантов они все равно не станут. Таким образом, мы видим, что подобие групп по индивидуальным характеристикам еще ничего не говорит о представительности популяции. Внешняя валидность эксперимента обеспечивает его внутреннюю валидность, но обратное отношение не сохраняется.

Схема интраиндивидуального сравнения. В эксперименте с сообщениями о ценах можно было бы использовать интраиндивидуальную схему. Каждый испытуемый мог бы пройти все три экспериментальных условия. Но между внешней и внутренней валидностью такого эксперимента не будет абсолютно никакой связи. Если выборка испытуемых хорошо представляет популяцию, то достигается высокая внешняя валидность. Однако внутренняя валидность индивидуального эксперимента нарушается главным образом за счет влияний предшествующих проб, различия задач и изменчивости поведения во времени. На представительность выборки эти факторы не влияют. Так что даже если тщательно контролировать их, это все равно не позволит улучшить репрезентативность выборки, т. е. внешнюю валидность эксперимента.

Послойный случайный отбор

Выделение слоев можно применять вместе со случайным отбором испытуемых внутри каждого слоя точно так же, как при случайном распределении слоев. Это дает пятую и последнюю схему эксперимента с межгрупповым сравнением — *случайный отбор групп с выделением слоев*. В эксперименте с мысленной тренировкой основанием для выделения слоев могут быть, например, оценки курсантов в тестах на пространственное воображение. Тогда по высоким, средним и низким оценкам в популяции из 810 курсантов можно было бы выделить три слоя по 270 в каждом. Из каждого такого слоя можно отобрать по 5 испытуемых для каждого из двух условий и, таким образом, получить выборку из 30 человек.

Если основание выделения слоев (здесь — оценка в тесте на пространственное воображение) связано с изучаемым видом деятельности (здесь — обучение сажать самолет Т-37), то данная процедура дает крупное преимущество. При одном и том же количестве испытуемых она обеспечит лучшую представленность попу-

ляции, и это повысит внешнюю валидность эксперимента. А поскольку применяется схема эксперимента с межгрупповым сравнением, то внутренняя валидность тоже будет выше.

Привлечение имеющихся в наличии испытуемых

Если в эксперименте участвуют не случайно отобранные представители популяции, а уже имеющиеся в наличии, будь то специально привлеченные испытуемые («заложники») или добровольцы, то вопрос о репрезентативности выборки становится весьма серьезным. Мы его уже рассматривали, когда обсуждали репрезентативность группы курсантов-летчиков набора одного года по отношению к наборам будущих лет. То же относится и к воображаемому эксперименту с обучением испанскому языку. Могут ли 100 участников эксперимента представить всех будущих студентов, которые будут учить испанский? Часто такие ситуации достаточно постоянны, и поэтому подобные проблемы не слишком серьезны.

Но что вы скажете о 75 добровольцах как представителях популяции покупателей универсамов? Нет никаких оснований полагать, что такая группа представляет всех покупателей. Хотя возможно, как раз тот факт, что участники эксперимента намного превосходят всю популяцию покупателей по своей заинтересованности, свидетельствует о достоверности полученных результатов. Иначе говоря, если даже эти, в каком-то смысле самые лучшие, испытуемые не смогли успешно воспользоваться вычислительным устройством (условие Б), то не стоит надеяться, что в более обширной популяции покупателей это устройство окажется эффективным. Когда в эксперименте участвует явно не репрезентативная выборка, то лучше прямо об этом сказать.

Обзор схем межгрупповых сравнений

В этой главе мы встретили так много различных схем, что запомнить их все сразу довольно трудно. Вам поможет табл. 4.4. В ней представлена классификация этих схем согласно типам стратегий построения экспериментальных групп и типам набора испытуемых. Приводится также общая оценка каждой из шести схем: насколько успешно решаются в них проблемы внутренней и внешней валидности. Две ячейки в таблице не заполнены, так как соответствующие комбинации типа стратегий и типа набора испытуемых просто невозможны. Подбор пар и привлечение реальных групп имеют смысл только в случае распределения испытуемых, но не их отбора.

Вы видите, что стратегия подбора пар, а также случайного распределения и отбора с выделением слоев обеспечивает более вы-

сокую внутреннюю валидность, чем случайная стратегия. Это объясняется тем, что зачастую при одном и том же числе испытуемых можно сделать группы более эквивалентными. Хуже всего — сравнивать реально существующие группы, в их подобии никогда нельзя быть уверенным из-за множества побочных влияний. При отборе испытуемых из популяции стратегия послойного случайного отбора может обеспечить также более высокую внешнюю валидность по сравнению с обычной случайной стратегией. Нередко она позволяет лучше представить популяцию.

При распределении групп мы не можем оценить внешнюю валидность эксперимента непосредственно, за исключением эксперимента на реальных группах: здесь приходится сомневаться уже в том, что группы представляют одну и ту же популяцию. При распределении выборки испытуемых по разным экспериментальным условиям для ответа на вопрос о внешней валидности эксперимента нужны дополнительные данные. Нам необходимо знать, адекватны ли такие значимые индивидуальные характеристики испытуемых, как возраст, образование, социальные и экономические условия их жизни, соответствующим характеристикам изучаемой популяции.

При оценке *каждой* из приведенных схем мы предполагали наличие достаточно большого числа испытуемых. Если число испытуемых слишком мало, то мала и надежда на хорошую пред-

Таблица 4.4

Сводка и оценка межгрупповых схем для представления популяции

Стратегия построения групп	Валидность	Тип привлечения испытуемых	
		Отбор	Распределение
Случайная	Внутренняя	Хорошая	Хорошая
	Внешняя	Хорошая	Зависит от дополнительной информации
Попарная	Внутренняя	—	Очень хорошая
	Внешняя	—	Зависит от дополнительной информации
Случайная с выделением слоев	Внутренняя	Очень хорошая	Очень хорошая
	Внешняя	Очень хорошая	Зависит от дополнительной информации
Реальные группы	Внутренняя	—	Плохая
	Внешняя	—	Плохая

ставленность популяции (внешнюю валидность), да и достичь подобия групп (внутренней валидности) становится труднее.

КРАТКОЕ ИЗЛОЖЕНИЕ

В этой главе мы обсудили три новых эксперимента, для каждого из которых, в отличие от ранее описанных, привлекалось большое число испытуемых. Первый эксперимент посвящен способам информирования о ценах на продукты. Экспериментальная гипотеза состояла в том, что прямое сообщение о цене за единицу массы (путем маркировки пакетов) позволит покупателям выбрать более дешевую покупку точнее и быстрее, чем два другие способа: (1) сообщение о массе пакета и его общей цене (текущая практика) и (2) то же плюс использование специального вычислительного устройства. Гипотеза подтвердилась. Первое преимущество эксперимента данного типа по сравнению с индивидуальным экспериментом заключается в том, что его результаты, полученные на выборке испытуемых, распространяются на более обширную популяцию, а не только на поведение отдельного человека.

Во втором эксперименте проверяли гипотезу о том, поможет ли специальная мысленная тренировка улучшить навыки пилотов по выполнению посадочных операций. Было обнаружено, что курсанты, прошедшие такую тренировку, выполняли эти операции лучше тех, кто тренировался лишь по обычной программе. Второе преимущество экспериментов нового типа также связано с наличием большого числа испытуемых. Становится возможным использование схемы эксперимента с межгрупповым сравнением. Эксперимент с мысленной тренировкой (в том виде, в каком он был представлен) требовал применения такой схемы. Схема группы индивидуальных экспериментов, когда каждый испытуемый участвует во всех экспериментальных условиях, здесь просто неосуществима. Один и тот же человек не может овладеть одним и тем же навыком двумя различными путями. Но даже там, где в принципе можно провести ряд индивидуальных экспериментов, как в эксперименте с сообщением о ценах, преимущества остаются за схемой межгруппового сравнения. Здесь устраняются такие источники нарушения внутренней валидности эксперимента, как эффекты последовательности и факторы задачи.

Детально описан третий эксперимент — с двумя методами обучения испанскому языку. Для его проведения было очень важно подобрать две группы испытуемых, сходных по способности изучать испанский. И тогда стало ясно, что эксперимент с межгрупповым сравнением, устраняя некоторые источники нарушения внутренней валидности, порождает еще один. Это — индивидуальные различия испытуемых. Описаны два вида таких различий. Во-первых, это такие очевидные характеристики, как пол, возраст испытуемых, их образование. Сюда же относятся различия, выявляемые с помощью тестов, скажем, на знание слов испанского языка. Кроме того, существуют несистематические различия испытуемых, которые включают как изменчивость поведения каждого человека, так и различия в выполнении задания разными людьми. Во всех трех экспериментах с межгрупповым сравнением использовались схемы, устраняющие систематическое смешение независимой переменной с индивидуальными различиями испытуемых. Но

если группы, работающие в разных экспериментальных условиях, имеют явные различия, скажем, взяты из разных школ, то такое смешение неизбежно. Надежность эксперимента повышается за счет привлечения большого числа испытуемых и сокращения несистематических вариаций. Последнее обеспечивается тем, что условия работы каждого испытуемого в группе одинаковы, а также путем сокращения разброса данных каждого испытуемого, как описано в главе 2.

Существуют три стратегии составления групп испытуемых, позволяющие уравнивать эти группы в отношении изучаемого вида деятельности: случайная стратегия, подбор пар и случайная с выделением слоев. Их иллюстрациями служат три схемы экспериментов с межгрупповым сравнением, в которых все имеющиеся в наличии испытуемые распределяются по группам для каждого из экспериментальных условий. При использовании схемы случайного распределения групп каждый испытуемый может с равной вероятностью попасть в одну из них. При попарном подборе групп испытуемые сначала располагаются по степени выраженности одной из характеристик, связанной с изучаемым видом деятельности. Испытуемые с одинаковой степенью выраженности этой характеристики попадают затем в разные группы. Применение схемы случайного распределения слоев начинается с классификации испытуемых по некоторому значимому признаку, также связанному с деятельностью, которая исследуется в эксперименте. А затем внутри каждого класса, или слоя, проводится случайное распределение по экспериментальным группам. Все три метода устраняют систематическое смешение независимой переменной с индивидуальными различиями испытуемых. Надежность повышается за счет увеличения числа испытуемых. Если основа подбора пар или выделения слоев действительно связана с исследуемой деятельностью, то при одном и том же количестве испытуемых соответствующие стратегии позволяют обеспечить более высокую надежность по сравнению с обычным случайным распределением.

Чтобы результаты эксперимента можно было распространить на интересующую нас популяцию, выборка испытуемых должна быть репрезентативной. Эксперимент *полного соответствия* (неосуществимый на практике) потребовал бы участия в нем всей популяции. Внешнюю валидность эксперимента можно оценить по тому, насколько близок к этой недостижимой цели метод отбора испытуемых. Хорошо представить популяцию в целом могла бы случайная выборка достаточно большого размера. Однако подходящие условия для такого выбора встречаются редко: часто исследуемая популяция целиком даже не существует в данный момент времени. Самым удачным из описанных исследований был в этом смысле эксперимент с мысленной тренировкой. Здесь можно было применить стратегию случайного выбора к популяции курсантов. Поскольку эксперимент проводился по схеме межгруппового сравнения с исследованием двух условий независимой переменной, использовались две случайные выборки. Здесь случайная стратегия применялась для *отбора* испытуемых, а не для распределения по группам, это имеет место, когда все имеющиеся в наличии люди должны участвовать в эксперименте. Эта схема называется схемой *случайно отобранных групп*.

Хотя для эксперимента с мысленной тренировкой было отобрано не очень много испытуемых, выборка успешно представляла изучаемую суб-

популяцию, так как индивидуальные различия между курсантами были не слишком велики. Если же исследователь хочет распространить результаты эксперимента за пределы данной популяции, скажем, на курсантов следующих лет, то проблемы внешней валидности станут более серьезными. Возникнет вопрос о соответствии «уровней» субъектных характеристик изучаемой популяции и той популяции, к которой прилагаются полученные результаты.

Связь внешней и внутренней валидности была описана для экспериментов, в которых испытуемые отбираются из популяции. Если способ отбора испытуемых обеспечивает высокую внешнюю валидность (в отношении индивидуальных различий), то тем самым достигается достаточное подобие групп, т. е. повышается внутренняя валидность. Напротив, группы, хорошо уравненные между собой по индивидуальным характеристикам, могут не быть достаточно представительными для исследуемой популяции. Иначе говоря, высокая внутренняя валидность эксперимента не гарантирует его высокой внешней валидности. При использовании схемы индивидуального эксперимента связь между внешней и внутренней валидностью отсутствует, поскольку та и другая зависят от разных факторов.

Вместе со случайным отбором, так же как и со случайным распределением, можно использовать выделение слоев. Это дает пятую эффективную схему эксперимента с межгрупповым сравнением — послойный случайный отбор. Сначала популяцию разделяют на классы, или слои, а затем внутри каждого слоя применяют случайный отбор. Выделение слоев дает то же преимущество, что и при случайном распределении, если его основание связано с изучаемым видом деятельности.

Когда для эксперимента привлекаются испытуемые, имеющиеся в наличии, будь то «заложники» или добровольцы, есть опасность, что выборка будет недостаточно репрезентативной. Для оценки репрезентативности выборки экспериментатор должен подробно проанализировать ситуацию, а также предвидеть последствия случаев, когда выборка недостаточно представляет популяцию.

Контрольные вопросы и задания

1. Чем отличаются возможности обобщения результатов экспериментов, описанных в этой главе, от экспериментов, рассмотренных ранее?
2. В каких случаях нужно проводить эксперимент с межгрупповым сравнением, а не интраиндивидуальный эксперимент?
3. Какие источники нарушения внутренней валидности устраняются в эксперименте с межгрупповым сравнением?
4. Каковы основные источники межиндивидуальных различий?
5. Почему неправильно проводить эксперимент на уже существующих группах испытуемых, например на ученических группах в разных школах?
6. Как можно повысить надежность в межгрупповом эксперименте?
7. В чем отличие двух межгрупповых схем: со случайным распределением по группам и со случайным отбором в группы?
8. Какова связь внешней и внутренней валидности в экспериментах с межгрупповыми схемами?

9. Приведите пример эксперимента, в котором хорошо представлена одна популяция и плохо — другая.

СТАТИСТИЧЕСКОЕ ПРИЛОЖЕНИЕ: СИЛА СВЯЗИ МЕЖДУ НЕЗАВИСИМОЙ И ЗАВИСИМОЙ ПЕРЕМЕННЫМИ

Зададимся вопросом: как велико различие в результатах, которое дает нам предъявление условия А по сравнению с условием Б? К настоящему моменту мы можем дать ответ только в терминах единиц зависимой переменной. Так, например, мы можем сказать следующее.

1. С использованием наушников испытуемая Д. пропустила 763 удара в час; без них она пропустила 908 ударов. Разница составляет 145 ударов в час.

2. Время реакции испытуемого на световой сигнал равно 185 мс; время реакции на звуковой сигнал равнялось 162 мс. Разница составляет 23 мс.

3. С умственной тренировкой испытуемые получили среднюю балльную оценку 4,21; «контрольная» группа получила среднюю балльную оценку 3,89. Разница составляет 0,32 деления на шкале балльных оценок.

В каком из экспериментов действие независимой переменной было наибольшим? Как вы сравните 145 ударов в час, 23 мс и 0,32 деления на шкале балльных оценок? Этого сделать нельзя. Одна из возможностей сравнения величин состоит в использовании отношения вместо разницы. Так, для трех данных экспериментов отношения большего результата к меньшему соответственно равны:

$$\frac{908}{763} = 1,19; \quad \frac{185}{162} = 1,14; \quad \frac{4,21}{3,89} = 1,08.$$

Как видим, отношения почти одинаковы. Фактически они кое-что говорят нам о силе связи, но они неадекватны по двум причинам. Во-первых, этот способ неприменим к экспериментам с более чем двумя условиями, такими, например, как эксперимент с информированием покупателей о стоимости товаров. Во-вторых, — и это более важно — отношение вообще не отвечает на наш вопрос: насколько *отчетливо различаются* показатели для одного условия и для других условий.

Использование частотных распределений

Для того чтобы понять, насколько различаются условия, мы можем нанести частотные распределения для этих условий на один и тот же график. Сравнение становится более понятным, если не

только отмечать высоту колонок, но соединить их вершины линией. (Это называется *полигоном частот.*) Данный метод уже был показан на рис. 4.3 для эксперимента по обучению испанскому языку.

По значительному перекрытию распределений мы можем судить, что тестовые оценки для двух условий — письменного и устного — различались незначительно.

Рассмотрим теперь распределения для эксперимента по изменению времени реакции, описанного в статистическом приложении к главе 1. Они показаны на рис. 4.4. Напомним, что это вымышленные данные. Предположим теперь, что они были получены в *межгрупповом* эксперименте. Тогда каждый из показателей времени реакции представляет среднее для одного из испытуемых, где 17 испытуемым предъявлялось данное условие. Этот пример может быть с тем же успехом представлен в терминах *интраиндивидуального* эксперимента, как он первоначально излагался. Рассмотрение эксперимента как межгруппового мы делаем только для того, чтобы связать наш анализ с тематикой данной главы.

Видно, что в данном случае различия между условиями более отчетливы, чем в эксперименте с испанским языком, т.е. *перекрытие* между распределениями *меньше*. Было бы хорошо иметь количественную меру различия вместо таких неопределенных терминов, как «кажется», «очевидно» и т.д. Такая количественная мера давала бы информацию, насколько сильна связь между независимой и зависимой переменными.

Вычисление ω^2

Мы можем получить численную величину силы связи, вычислив ω^2 (ω — малая греческая буква *омега*; мы называем ω^2 омегой в квадрате). По существу, ω^2 — один из параметров генеральной совокупности, о которых рассказывалось в статистическом приложении к главе 1. Его полное описание можно найти в работе Хейса (Hays W. K., 1973).

Наше вычисление с использованием данных по выборке испытуемых дает оценку ω^2 . Мы будем называть ее *est* ω^2 .

Давайте построим новый график результатов эксперимента по измерению времени реакции. Однако теперь мы не будем делать различий между тем, какое из двух условий — А (свет) или Б (тон) — было использовано. Как видно на рис. 4.5, это комбинированное распределение несколько более растянуто, чем каждое из отдельных распределений для света и тона. Чем больше исходные распределения отличаются друг от друга, тем больше будет растянуто комбинированное распределение.

Если бы мы смогли провести бесконечный эксперимент и при этом получили бы распределения, показанные на рис. 4.4 и 4.5,

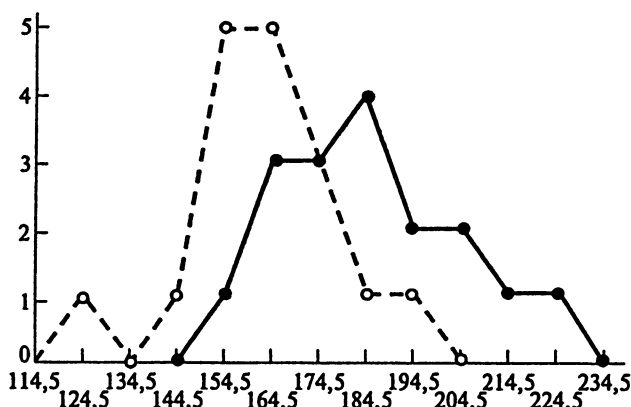


Рис. 4.4. Частотное распределение средних значений времени реакции на световой (условие А) и звуковой (условие Б) сигналы. Ось абсцисс — средние значения времени реакции (в мс); ось ординат — частота; сплошная линия — световой сигнал, пунктирная — звуковой

мы могли бы вычислить ω^2 прямо из параметра σ_x^2 следующим образом:

$$\omega^2 = \frac{\sigma_{\text{комб}}^2 - \sigma_{\text{отд}}^2}{\sigma_{\text{комб}}^2}. \quad (4.1)$$

Однако поскольку наши данные получены только по одной выборке испытуемых, а не в бесконечном эксперименте, мы должны оценивать ω^2 по статистике S_x^2 :

$$\text{est } \omega^2 = \frac{S_{\text{комб}}^2 - S_{\text{отд}}^2}{S_{\text{комб}}^2}. \quad (4.2)$$

Квадрат стандартного отклонения распределения называется *дисперсией*.

Числитель этой формулы дает разность между дисперсиями комбинированного распределения и отдельного распределения, в нашем случае любого из условий А или Б. Делением этой разности на дисперсию комбинированного распределения мы придаем ей форму пропорции. Она отвечает на вопрос, на какую часть уменьшается дисперсия показателей при переходе от комбинированного распределения к отдельному.

Производя вычисления, нет необходимости сначала вычислять S_x и затем возводить его в квадрат, чтобы получить S_x^2 . Вспомните (из формулы 2.2):

$$S_x = \sqrt{\frac{\sum x^2}{N-1}}.$$

Поэтому

$$S_x^2 = \frac{\sum x^2}{N-1}. \quad (4.3)$$

В статистическом приложении к главе 3 мы вычислили для условия А (свет) и условия В (тон):

$$S_A^2 = \frac{5894,25}{16} = 368, \quad S_B^2 = \frac{4048,25}{16} = 253.$$

Используем эти величины для нахождения $S_{отд}^2$. Согласно В. Хейсу (Hays W. K., 1973, с. 418), среднее S_A^2 по S_B^2 и дает величину $S_{отд}^2$ при допущении равенства «истинных» дисперсий двух наборов:

$$S_{отд}^2 = \frac{S_A^2 + S_B^2}{2}. \quad (4.4)$$

Поэтому

$$S_{отд}^2 = \frac{368 + 253}{2} = 310.$$

Такое же вычисление производится для комбинированного распределения:

$$S_{комб}^2 = \frac{14188,5}{33} = 430.$$

Подставляя эти величины в формулу 4.2, получаем

$$est \omega^2 = \frac{S_{комб}^2 - S_{отд}^2}{S_{комб}^2}, \quad est \omega^2 = \frac{430 - 310}{430} = 0,28.$$

Это показывает сильную связь между независимой и зависимой переменными. Даже значение 0,20 уже достаточно существенно. Значение никогда не может превысить 1; однако эта величина достигается редко. В то же время вычисление для эксперимента с испанским языком дает

$$est \omega^2 = \frac{96,6 - 93,9}{96,6} = 0,03.$$

Это очень слабая связь между независимой и зависимой переменными.

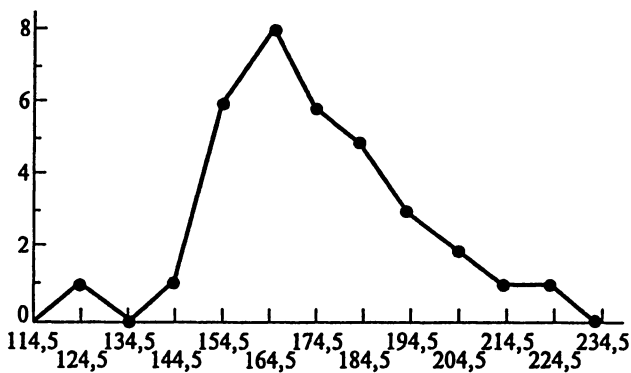


Рис. 4.5. Объединенное частотное распределение средних значений времени реакции. Ось абсцисс — средние значения времени реакции (в мс); ось ординат — частота

Применения ω^2

Обратите внимание, что для $S_{\text{отд}}^2$ необходимо допускать равенство истинных дисперсий для двух условий. В эксперименте по измерению времени реакции это допущение было приемлемым, поскольку дисперсии для двух условий довольно близки по величине. Это справедливо и для эксперимента с испанским языком. Однако для эксперимента с информацией о ценах (см. табл. 4.1) условие В давало значительно меньшее стандартное отклонение (особенно для времени выбора покупки), чем другие условия. Три дисперсии (квадраты стандартных отклонений) были равны 100, 92 и 1,2 для условий А, Б и В. При столь значительном различии прямого способа вычисления ω^2 нет. И здесь дело не в том, что число условий равно трем, а не двум. Если значения дисперсии близки, то величина ω^2 может быть вычислена для любого числа условий с использованием $S_{\text{отд}}^2$ как среднего значения для всех условий. Приводившаяся процедура вычисления ω^2 может быть использована как для интраиндивидуальных, так и для межгрупповых данных. Дисперсия — это то, что относится к пробам, а не к испытуемым.

Интерпретация ω^2

Мы можем рассматривать ω^2 как индикатор величины, на которую уменьшается *неопределенность* за счет того, что нам известно экспериментальное условие. В эксперименте по измерению времени реакции мы кое-что знаем о среднем показателе для каждого отдельного испытуемого благодаря тому, что знаем экспериментальное условие. Но в пределах каждого условия показатели

варьируют, т. е. как-то распределены. Наша неопределенность измеряется дисперсией этого распределения. Если мы не знаем условия, предъявлявшегося испытуемому, наша неопределенность увеличивается: дисперсия комбинированного распределения больше, чем дисперсия для отдельного условия.

Таким образом, зная, какое из условий предъявляется испытуемому, мы уменьшаем *неопределенность*. Как уже говорилось, деление этого уменьшения на уменьшаемую дисперсию ($S_{\text{комб}}^2$) превращает ответ в отношение. Тем самым ω^2 сообщает нам часть, на которую уменьшается неопределенность при знании экспериментального условия. Это и есть мера воздействия на поведение независимой переменной.

Задача: Вычислите *est* ω^2 для эксперимента, сравнивающего условия В и Г, с 18 испытуемыми в каждой группе:

$$\sum x_B^2 = 4700; \quad \sum x_G^2 = 4900; \quad \sum x_{\text{комб}}^2 = 15000.$$

Ответ: *est* $\omega^2 = 0,34$.

Глава 5

ВЫДЕЛЕНИЕ НЕЗАВИСИМОЙ ПЕРЕМЕННОЙ¹

В чем сходны следующие ситуации?

1. Женщина стоит внутри огромного деревянного ящика, в который могло бы поместиться пианино, и смотрит через «глазок» на десятилетнюю девочку-индианку. Если девочка садится на стул у одного угла ящика и многократно нажимает на рычаг, то после каждого десятого нажатия женщина вкладывает шарик в металлическую трубку, через которую тот со стуком скатывается в пластмассовый лоток перед девочкой. Если же девочка садится на стул у другого угла ящика, где никакого рычага нет, женщина все равно периодически посылает ей шарики.

2. Эластичная повязка на голове спокойно сидящего молодого человека прочно удерживает введенные в его ноздри две пластиковые трубки. Несмотря на то что вовсе не холодно, уши человека плотно прикрыты заглушками. Делая вдох, он нажимает кнопку, приводя в действие два насоса, которые поочередно вдувают порции воздуха с запахом гвоздики ему в ноздри.

3. Мать и ребенок находятся в комнате с еще одной молодой женщиной. Мать кладет ребенка в постельку и со словами «бай-бай, Томми» выходит из комнаты. Ребенок начинает кричать. Ровно через 60 с после ухода мать возвращается в комнату.

Вы, конечно, решили, что все это сцены из авангардистского театра абсурда. Ничего подобного! Это описание трех психологических экспериментов из ряда других, которые будут обсуждены в этой главе. На их примере мы разберем проблему *выделения* независимой переменной. Эти эксперименты совершенно отличны от тех, о которых мы говорили ранее.

Во-первых, ни один из них, по всей вероятности, не дает результатов, которые могли бы немедленно быть использованы на практике, как, например, новый метод обучения испанскому языку или лучший способ информировать покупателя о ценах на продукты в универсаме. Во-вторых, здесь независимая переменная более проста, чем описанные выше. Те, как правило, представляли собой сложные по составу «группы факторов». Эту же можно определить как *единичную* переменную. Так, единственное, чем отличаются условия в эксперименте с обонянием, — это вре-

¹ Перевод Ч. А. Измайлова.

мя поступления воздуха в каждую ноздрю. Напротив, когда сравниваются два метода обучения испанскому языку, то они оказываются различными не только по виду речи (разговорная или письменная), но и по используемым упражнениям, по числу проверок, по порядку ознакомления с темами и т.д. В-третьих, в обсуждаемых здесь экспериментах независимая переменная не только более изолирована, но и условия ее предъявления более тщательно «очищены». Так, в эксперименте с обонянием для ослабления посторонних раздражителей на испытуемого надевали наушники.

Напротив, в экспериментах, которые «улучшают» реальный мир (см. гл. 3), экспериментатор иногда специально *вводит* осложняющие обстоятельства, чтобы приблизить обстановку эксперимента к реальности и тем самым повысить внешнюю валидность. Так, в эксперименте с ночными посадками самолета было смоделировано движение второго самолета, о котором нужно было сообщать «на землю».

Новые эксперименты будут способствовать более глубокому пониманию переменных, влияющих на поведение. Такие эксперименты уже можно назвать научными. Их результаты могут не сразу найти практическое применение. Такова первая особенность трех экспериментов, упомянутых в начале данной главы. Они не обязательно связаны с «практикой». Но и слово «научный» следует употреблять с некоторыми оговорками. Весь смысл проведения экспериментов в том, чтобы через четкое и полное описание связей между независимой и зависимой переменными понять, что определяет поведение.

Вторая характерная черта — выделение единичной независимой переменной указывает на то, что если полученные данные будут использованы для лучшего понимания поведения, необходима более *точная* экспериментальная гипотеза. Когда Дион предположила, что поиск лучше вести без бинокля, она привела ряд доводов, как-то: ограниченность обзора в бинокле, отсутствие восприятия глубины. Если бы гипотеза подтвердилась, то это могло бы быть следствием либо всех названных причин, либо некоторых из них, либо каких-то причин, не упомянутых вовсе. Чтобы выделить действительную причину, требуются дальнейшие эксперименты. С этой целью мы должны *изолировать* независимую переменную и в нашей гипотезе, и в нашем эксперименте.

Третья особенность перечисленных экспериментов — использование «чистых» условий. Это один из способов *контролировать «шум»* в эксперименте. Даже идеальный эксперимент в том виде, как он был описан ранее, не всегда имеет «чистые» условия. Другими словами, даже предъявление одному и тому же испытуемому сразу двух условий независимой переменной не обеспечивает строгой проверки ее действия. Оба условия могут быть одинаково

«нечистыми», одинаково «зашумленными». Например, двумя условиями являются два разных количества лекарства — с одними и теми же примесями. В этом случае действие чистого лекарства выявить невозможно.

Очевидно, что чем точнее рабочая гипотеза, тем выше класс эксперимента. Как и ранее, эталоном при этом является безупречный эксперимент. Любой реальный эксперимент должен стремиться к безупречному. Безупречный в отношении *внутренней валидности* эксперимент — это такой вариант идеального эксперимента, который мы можем назвать *чистым* экспериментом. Прежде всего это идеальный эксперимент со всеми вытекающими отсюда последствиями, но этого мало. Чистый эксперимент (также практически недостижимый) требует еще, чтобы экспериментатор оперировал только с выделенной независимой переменной и ни с чем больше. Более того, все другие переменные, которые могли бы воздействовать на поведение, остаются строго постоянными.

Одной из мер достижения этого является «очищение» условий независимой переменной, которое, как вы увидите, может привести даже к надеванию теплых наушников в августе. Другие способы обеспечения достаточно высокой внутренней валидности мы уже обсуждали: это устранение *систематического смешения* с побочными переменными.

Здесь, в лабораторных экспериментах, отличающихся от «естественных» (см. предыдущие главы), мы рассмотрим некоторые новые виды смешения.

Смешения, описанные ранее, в основном проистекали из того факта, что одному и тому же испытуемому не могут быть предъявлены два разных условия одновременно. Конечно, такое «*процедурное*» по своей природе смешение остается проблемой и в этих экспериментах, как и в любых других. Но в лабораторных экспериментах, предполагающих выделение независимой переменной, существует также другая разновидность смешения. Она связана с тем, что экспериментатор не может создать совершенно «чистые» условия независимой переменной, как того требует эксперимент. Г. Хардин, сведущий биолог, хорошо выразил сущность этой трудности, сказав: «Мы никогда не можем иметь дело с чем-то одним» (Hardin G., 1972, с. 68). Поскольку это «одно», что интересует экспериментатора, неизбежно соединено с чем-то «другим», мы говорим о *сопутствующем* смешении.

Первая разновидность такого смешения вызвана *артефактами* эксперимента. Экспериментатор дорого бы дал, чтобы удалить только ту небольшую часть мозга животного, которая фигурирует в его экспериментальной гипотезе. Но он не может этого сделать. Он непременно повредит другую ткань. Если теперь он сравнит свою экспериментальную группу (с частично удаленным мозгом)

с контрольной группой (не подвергавшейся операции), то он смешает экспериментальную переменную с *искусственно* вызванной переменной. Будет неясно, обусловлена ли разница в поведении двух групп животных удалением именно данной части мозга или оперативным вмешательством вообще?

Вторая разновидность сопутствующего смещения порождается не методикой эксперимента как таковой, а тем фактом, что независимая переменная «*естественным*» образом связана с другой переменной. Когда мать выходит из комнаты, оставив ребенка, его плач может быть вызван как тем, что ушла мать, так и тем, что он вообще остался один.

Избавление от подобного «естественного смещения» — одна из главных забот экспериментатора, пытающегося выделить независимую переменную.

В связи с обсуждаемыми новыми экспериментами встает поновому вопрос относительно *внешней* валидности, а именно — вопрос, *тот* ли эксперимент был проведен. Эксперименты, направленные на выделение независимой переменной, часто предпринимаются, чтобы проверить истинность теории. В настоящее время теории, как правило, абстрактны, в то время как эксперименты должны быть конкретными. Необходимо, чтобы *конкретная методика* эксперимента вполне соответствовала экспериментальной гипотезе, изложенной в более абстрактных терминах.

Например, можно выдвинуть гипотезу, что отклонения в поведении подростков преодолеваются более эффективно рациональным путем («убеждением»), чем наказанием. Экспериментатор должен применить какой-то конкретный вид наказания. Предположим, в качестве такового он выбрал неприятный звук зуммера (La Voie J., 1973). Является ли этот звук адекватной экспериментальной операционализацией абстрактного термина *наказание*? Безусловно, нет. Скорее всего он будет действовать на рефлекторном уровне. (При этом может оказаться, что и самому *экспериментатору* придется надеть наушники.) Итак, эксперименты можно оценивать и с точки зрения *операциональной валидности*.

Эта глава научит вас прежде всего более *аналитично* подходить к экспериментам, направленным на выделение независимой переменной. Предположим, вы читаете экспериментальную статью, в которой показывается, что независимая переменная оказалась неэффективной. Вам необходимо внимательно ее изучить, чтобы убедиться, что экспериментальные условия были достаточно «чистыми». А теперь предположим, вы читаете статью, в которой, наоборот, показано, что независимая переменная решающим образом влияла на зависимую переменную. Прежде всего вы должны спросить себя, а не было ли это результатом сопутствующего смещения с другой переменной. Если вы полагаете, что та-

кая возможность не исключена, вы должны суметь показать это на диаграмме (см. ниже). Еще лучше, если вы сможете придумать контрольный эксперимент. Следующее, на что вы должны обратить внимание, — это корректность экспериментальных действий — насколько они представляют переменные, заключенные в экспериментальной гипотезе. Вы еще глубже освоите проблему изоляции независимых переменных, если сами проведете несколько экспериментов (после изучения соответствующей литературы) и разберете их столь же строго, как разбирали эксперименты других.

Ознакомившись с материалом этой главы, вы должны будете ответить на вопросы по следующим темам.

1. Тип экспериментальной гипотезы, проверяемой в экспериментах, которые направлены на более глубокое понимание поведения.

2. Очищение экспериментальных условий.

3. Сопутствующее смешение и что с ним делать.

4. Операциональная валидность.

5. Выбор испытуемых и обращение с ними в эксперименте.

6. Логика тем, представленных в последующих главах книги.

ИЗОЛИРОВАННАЯ НЕЗАВИСИМАЯ ПЕРЕМЕННАЯ

Эксперимент по этике труда

Д. Сингх и В. Квери (Singh D., Query W. T., 1971) предположили, что когда награду можно получить совсем без усилий или при умеренном усилии, то предпочитают последнее. Здесь мы рассмотрим их весьма сложный эксперимент на предпочтение работы детьми — пока с одной этой точки зрения. Другие аспекты будут изложены в последующих главах книги.

Методика. Посмотрим, как было получено предпочтение одного из двух условий. Сначала об испытуемых: 20 девочек-индианок в возрасте от 8 до 12 лет были выбраны наугад из II—IV классов индейской школы в Уоптоне, штат Северная Дакота.

Они вызывались из класса поодиночке «поиграть». В экспериментальном помещении им показывали большой ящик, который может «выдавать» шарики (как описано в начале этой главы). Им посоветовали набрать как можно больше шариков. Такая инструкция была едва ли необходима, поскольку девочки уже знали, что они смогут оставить шарики себе или обменять их на другие заманчивые игрушки, выставленные здесь же.

До того как девочка предпочтет один из вариантов, ее обучали двум разным способам, с помощью которых можно получить шарики. Один способ был более трудный. Девочка садилась у того угла ящика, где находился рычаг, и через каждые десять нажатий

на него получала шарик. Она могла продолжать нажимать, пока не получит 5 шариков. Легкий способ заключался в том, чтобы сидеть на стуле у другого угла ящика и просто ждать, когда выкатится шарик. Через 15 с она получала 5 шариков. Вслед за этим этапом обучения проводился собственно эксперимент на предпочтении.

В этом эксперименте испытуемая сначала указывала на игрушку, которую она хотела бы получить, а затем переходила к добычанию необходимых шариков. Испытуемой разрешалось сесть у любого угла ящика, а также менять место в течение 5 мин, когда ей захочется. Если она садилась на стул у рычага, она, конечно, могла нажимать на рычаг с тем, чтобы получать за каждые 10 нажатий по шарик (10 — здесь показатель применявшейся схемы оперантного обусловливания, называемый «фиксированным отношением» — ФО). Если девочка садилась на другой стул, где не было рычага, шарики выпадали в том же самом темпе и количестве, как и в первом случае. На следующий день с ней проводился такой же 5-минутный опыт. Затем девочка могла либо взять себе шарики, либо получить за них игрушку.

Результаты. Было установлено, что в среднем 60 % своих шариков девочки получили, нажимая на рычаг, а другие 40 % — просто сидя в бездействии. Таким образом, испытуемые не только время от времени нажимали на рычаг, хотя могли получить шарики ничего не делая, но и предпочитали деятельность бездействию. Следовательно, рабочая гипотеза подтвердилась: когда вознаграждение может быть получено без усилий или при умеренном усилии, последнее оказывается предпочтительнее.

Обсуждение. Это был эксперимент, способствующий нашему пониманию природы мотивации. Ни один из двух сравнивавшихся «способов» не может быть использован как таковой в повседневной жизни, вне эксперимента. Ситуации повседневной жизни гораздо сложнее и длятся месяцы и годы, а не 5 мин. И тем не менее полученные результаты позволяют сделать определенные выводы, касающиеся социального обеспечения (пособия по безработице). Многие опасаются, что если люди смогут без труда удовлетворять свои нужды с помощью такого пособия, то они перестанут работать. Однако это не так. Испытуемые в наших опытах предпочитали честно трудиться, чтобы получить вознаграждение. Вот почему этот эксперимент рассматривается как эксперимент по «этике труда». Между прочим, для чиновников из отдела социального обеспечения разговоры о пособиях по безработице, наверное, тоже «честный труд». Для более глубокого понимания поведения важно ответить на вопрос: действительно ли оно мотивируется элементарными потребностями. Ведь думают же некоторые, что люди работают в конечном счете для того, чтобы получить деньги, приобрести пищу и утолить голод. Описанные экспе-

рименты показывают, что деятельность для человека составляет самостоятельную ценность (White R. W., 1959). Не правда ли, интересно? Ну, а теперь вы можете спросить: «Почему привлекались к эксперименту девочки-индианки?» Подождем с ответом, его вы найдете в главе 8.

Необходимость создания специальной ситуации

Та самая особенность эксперимента, которая сделала его малопригодным для непосредственного практического использования, дала возможность проверять гипотезу об изолированной переменной. Этой переменной было — работать или не работать, все равно получая вознаграждение? Допустим, подобный эксперимент был бы поставлен на фабрике с расчетом на немедленное практическое применение его результатов. Рассмотрим, например, работу по сборке мышеловок. Можем ли мы прямо сравнивать продуктивность работы в следующих двух ситуациях: первая — рабочие знают, сколько они получают за каждую собранную мышеловку, и вторая — они получают то же количество денег за день, но безотносительно к числу собранных мышеловок.

Фактически речь идет о форме оплаты труда: почасовой или сдельной. На практике вопрос давно уже решился в пользу сдельной оплаты. Если мы посмотрим на людей, оплачиваемых повременно, мы увидим разное. Одни работают быстро. Другие — медленно, с тем чтобы дирекция не ожидала от них многого. Третьи работают старательно, но неумело, с низкой производительностью. Те, кто работают быстро, недовольны тем, что получают ту же зарплату, что и ленивые или неспособные. Возможно, озлобившись, они в конце концов перестанут так быстро работать. При сдельной же оплате умелые работники гордятся собой — как быстро они умеют собирать мышеловки. Те, кто неповоротливы, возможно, попросят перевести их на другую работу, требующую меньшей ловкости (или менее сильных пальцев). Если при сдельной оплате окажется больше продукции — а так это обычно и случается, — это не опровергнет нашу гипотезу. Различные формы оплаты неизбежно оказываются связанными с разными уровнями других переменных, в особенности социальных. Почасовая оплата не позволяет проявиться трудовой этике, подавляя ее.

ОЧИЩЕНИЕ УСЛОВИЙ

Исследование направленности обоняния

Экспериментатором, вставившим пластмассовые трубки в ноздри студента, был Георг фон Бекеши. Это известный ученый, ла-

уреат Нобелевской премии (в области физиологии и медицины) за 1961 год. Он провел этот эксперимент в 1964 году в Лаборатории психофизики Гарвардского университета. Целью эксперимента было выяснить, действует ли разница во времени, с которой запах достигает рецептора каждой ноздри, на локализацию источника запаха.

Для проведения этого эксперимента нужен был особый исследователь. Во-первых, он должен был считать, что обоняние обязательно направленно. Во-вторых, требовался предельно аккуратный и технически искушенный человек: нужно было обеспечить очень точную разницу во времени поступления воздуха в каждую ноздрю. В-третьих (решающее условие), — этот человек должен был верить, что эксперименты вносят основной вклад в научное понимание поведения.

Вы, разумеется, помните из курса общей психологии, что направление звука определяется благодаря тому, что у нас два уха. Если источник звука находится от нас справа, звук будет немного сильнее в правом ухе и, что более важно, достигнет правого уха немного раньше левого. Не тем же ли способом мы определяем и направление запаха? Бекеши варьировал как силу запаха, поступавшего в разные ноздри, так и время поступления запаха в каждую из них. Здесь мы рассмотрим его исследование, касающееся только временного различия.

Установка. Поскольку эффективность эксперимента зависит от «чистоты» независимой переменной, то решающее значение имеет аппаратура. На рис. 5.1 показана установка для введения запаха в одну ноздрю. В положении, указанном на рисунке, трубка, вставленная в ноздрю, пропущена через *верхнюю* из двух пластинок тефлона и соединена с трубкой, проходящей через нижнюю пластинку и соединенной, в свою очередь, с компрессором, подающим воздух без запаха. Ароматизированный воздух из смесителя также под давлением поступает в вентилятор и уходит наружу. Когда верхняя пластинка сдвигается вправо, в ноздрю сразу попадает ароматизированный воздух. Воздух без запаха выходит через крайнее левое отверстие в верхней пластинке в комнату.

Были использованы две такие системы, по одной для каждой ноздри. Верхняя пластинка передвигалась мощным электромагнитом. Перед началом опыта экспериментатор устанавливал реле времени на необходимую разницу во времени поступления ароматизированного воздуха в каждую ноздрю. Он мог это делать с точностью до одной десятитысячной секунды!

Ход опыта. В каждой пробе испытуемый в момент, когда начал вдох, нажимал кнопку. Это вызывало поступление ароматизированного воздуха поочередно в каждую ноздрю с временным интервалом, заданным экспериментатором. (Испытуемый, конеч-

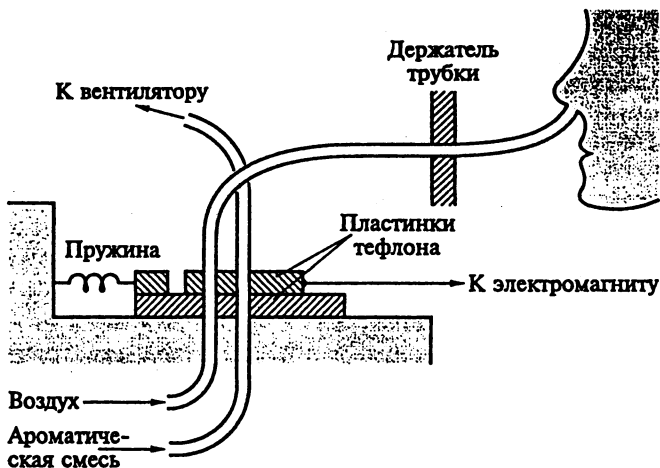


Рис. 5.1. Аппарат для передачи запаха в ноздрю в эксперименте с направленным обонянием (Békésy G. V., 1964)

но, не знал, каков этот интервал.) В ответ на стимул испытуемый проводил линию вперед, влево или вправо, в зависимости от того, откуда, как ему кажется, идет запах.

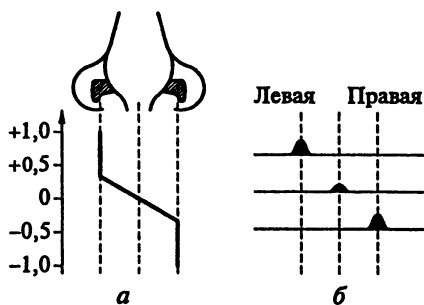
Результаты. На рис. 5.2 показаны временные интервалы (*a*) и ответы испытуемых (*b*). Были использованы следующие три уровня независимой переменной: стимуляция левой ноздри на 0,5 мс раньше стимуляции правой ноздри; одновременная стимуляция обеих ноздрей и стимуляция правой ноздри на 0,5 мс ранее левой ноздри. Как можно видеть, несмотря на очень небольшую разницу во времени ($1/2000$ с), испытуемые четко локализовали кажущийся источник запаха слева или справа. Когда же стимуляция была одновременной, они определяли источник как находящийся прямо впереди, хотя и несколько размытый.

Таким образом, была подтверждена гипотеза о том, что разница во времени поступления ароматического вещества в две ноздри создает впечатление направленности запаха.

Обсуждение. Бекешы отмечает: «В этой связи интересно добавить, что свинья — животное, известное своей способностью отыскивать трюфели под небольшим слоем земли, — имеет широко расставленные ноздри. В поисках этих деликатесов она постоянно двигает головой из стороны в сторону так, как это делают люди, которые пытаются установить источник звука. Похоже, что животное тем самым определяет местоположение источника запаха» (Békésy G. V., 1967, с. 105).

Однако эксперимент был проведен не затем, чтобы объяснить, почему свиньи так хорошо отыскивают трюфели. Скорее, исследователь этим примером стремился показать, что способ-

Рис. 5.2. Временной интервал между стимуляцией ноздрей (а) и определение испытуемыми местонахождения источника запаха (б) в эксперименте с направленным обонянием: а — различия во времени между стимуляцией левой и правой ноздри; б — распределения лево- и правосторонней локализации источника запаха (Békésy G. V., 1964)



ность локализовать источник стимуляции посредством оценки интервала между двумя воздействиями на разные участки тела является общей характеристикой восприятия, связанной с организацией нервной системы. Локализация с помощью слуха не уникальна. Это лишь одна из наиболее развитых способностей. Такое соображение привело Бекешы к изучению «слуха» на коже предплечья, которой он заменил крошечный орган слуха, расположенный в глубине среднего уха! Это дало ему прекрасную большую рецепторную поверхность и интактного испытуемого, от которого он получал субъективные отчеты.

Особая точность эксперимента иногда очень важна

Для корректного проведения эксперимента на направленное обоняние необходимо, чтобы условия независимой переменной (временные интервалы) соблюдались очень точно, а мешающие факторы были бы постоянными и желательно на *низком уровне*. *Достоверные* результаты, как это уже обсуждалось выше, могут получаться и при менее чистых условиях. Однако по поводу своих опытов Бекешы отмечает: «Интересно, что после того, как установка стала работать точнее, работа испытуемых также улучшилась, и временные интервалы, достаточные для латеральной локализации источника запаха, стали короче» (там же, с. 103). Другими словами, проводимый эксперимент должен хорошо представлять *чистый* эксперимент. Наушники в экспериментах Бекешы предназначались частично для предотвращения локализации стимула по щелчкам электродвигателя и шуму насоса, но главным образом для того, чтобы сосредоточить внимание испытуемого на запахе. По крайней мере в некоторых экспериментах, направленных на выделение независимой переменной, последняя должна быть исключительно «чистой»; только тогда будет достигнута *внутренняя валидность* в проверке экспериментальной гипотезы. Это особенно справедливо для психофизиологических экспериментов. Если используются нечистые медикаменты, то результат фактически сводится к нулю. Если операция на мозге проведена не до-

статочны чисто, могут быть повреждены смежные области, которые выполняют прямо противоположные функции. Вот почему мы рассмотрим теперь более тщательно эксперименты, в ходе которых вводится лекарство или разрушается часть мозга.

ИСКУССТВЕННОЕ СМЕШЕНИЕ

В эксперименте, направленном на выделение независимой переменной, причиной систематического смешения может быть неизбежное введение второй переменной, которая также может влиять на поведение. В этом случае употребляется термин *сопутствующее* смешение. Различаются два вида сопутствующего смешения: первое — *искусственное* смешение, возникающее в результате действий, направленных на создание одного из условий независимой переменной. Оно рассматривается в настоящем разделе. Второе — *естественное смешение*, которое возникает как в жизни, так и в эксперименте. Этот тип сопутствующего смешения будет рассматриваться в следующем разделе.

Два психофизиологических эксперимента

Память после рассечения свода. Широкая область исследований связана с мозговыми механизмами поведения. В нашем мире, полном опасностей, человек иногда получает травму мозга, что приводит к временному или постоянному нарушению поведения. Примером может служить амнезия, или потеря памяти. Единственная реальная возможность использовать такие клинические случаи для изучения функций различных частей мозга — это исследовать мозг человека после его смерти. Но при этом нельзя, конечно, рассчитывать на планомерное и тщательное изучение. Вдобавок не всегда можно точно установить, какая именно часть мозга повреждена.

Вследствие этого на животных проводятся контролируемые эксперименты. В зависимости от поставленной задачи подопытными служат различные виды животных. Исследование сложных аспектов памяти, которые связаны с поведением человека, желательно проводить на марышках или даже высших приматах.

Обычно методика заключается в возможно более аккуратном удалении определенной части мозга и последующем сравнении поведения оперированных и неоперированных животных. По окончании опыта оперированных животных необходимо умертвить с тем, чтобы точно установить место экспериментального повреждения.

Нет нужды говорить, что такого рода эксперименты не проводятся кое-как, без подготовки. Они тщательно планируются с тем, чтобы получить ощутимые результаты. Кроме того, подопытные животные должны получать минимальную болевую травму.

Д. Гаффан провел эксперимент «по определению роли гиппокампа в двух различных актах памяти: образовании ассоциаций и узнавании» (Gaffan D., 1974, с. 1100). Далее в этой книге мы увидим, как ему удалось разделить ассоциацию и узнавание. Здесь же мы рассмотрим, как он выяснил, с чем действительно связаны наблюдаемые им явления: с поврежденным или интактным гиппокампом. В качестве подопытных животных он использовал макаков резусов.

В идеальном эксперименте одна и та же обезьяна в одно и то же время должна быть исследована с оперированным гиппокампом и с нетронутым, интактным. Поскольку такое невозможно, то были взяты две группы по три обезьяны в каждой: в одной группе — животные с поврежденным гиппокампом, в другой — с интактным. Предположим, что обе группы абсолютно равноценны. Тогда условия чистого эксперимента требуют, чтобы у животных первой группы был поврежден только гиппокамп и никакая другая часть мозга и чтобы животные второй группы вообще не подвергались операции. Но повредить только гиппокамп невозможно. Гаффан постарался разрушить гиппокамп без повреждения окружающих тканей, насколько это было возможно. Он выбрал специальную часть гиппокампа — узкий пучок нервных волокон, именуемую *сводом*. «Рассечение свода гиппокампа было выбрано потому, что оно в наименьшей степени повреждает экстрагиппокампальную область» (там же). Иными словами, перерезка свода является наилучшим способом нанести локальное повреждение гиппокампа, избегая повреждения других структур.

И все-таки невозможно рассечение свода без оперативного вмешательства, которое требует использования анестезии и влечет за собой значительное повреждение кожи, мышц и костей черепа. «В черепе было проделано пятимиллиметровое отверстие для трепанации. Затем оно было расширено долотом. Полушарие с этой стороны было осторожно отведено, чтобы обнажить мозолистое тело (волокна, связывающие два полушария). В нем после удаления жидкости (высасывания через трубку) была открыта щель, через которую с помощью изогнутой иглы был извлечен и затем разрезан свод» (там же, с. 1102).

Вот что было проделано с «экспериментальными» животными: перерезан свод и нанесено много других повреждений. После того как они поправились, с ними провели эксперименты на проверку памяти, сравнивая с обезьянами второй группы с сохранным сводом. Однако не думайте, что обезьян второй группы вообще не трогали до проверки их памяти. Эту группу животных Гаффан оперировал *точно так же*, как и первую, за исключением того, что свод остался не тронутым. Поэтому обезьяны второй группы называются *оперированными контрольными животными*, тогда как обезьяны с рассеченным сводом — *экспериментальными животными*.

Пищевое поведение, вызванное диазепамом. Еще один способ изучения мозговых механизмов поведения — применение лекарственных веществ. В клинической практике диазепам и другие сходные вещества (из класса бензодиазелинов) часто используются как умеренные транквилизаторы. Наблюдение над пациентами и над подопытными животными позволило сделать вывод, что диазепам помимо успокаивающего действия вызывает повышенный аппетит. Чтобы понять функции мозга, нужно выяснить природу его биохимических процессов и их влияния на поведение.

Исследователи Р. Вайз и В. Досан (Wise R.A., Dawson V., 1974) провели эксперимент со стимулированием аппетита диазепамом. Для изучения этого элементарного вида поведения (по сравнению, например, с памятью, изучаемой Гаффаном) они могли использовать в качестве испытуемых лабораторных крыс. Мы и на этот раз отложим обсуждение полученных ими результатов, снова сосредоточив свое внимание на примененном ими способе контроля.

В отличие от Р. Гаффана, Р. Вайз и В. Досан использовали схему с двумя экспериментальными условиями. Они намеревались сравнить поведение подопытного животного, которому ввели в мозг определенное количество диазепама, с поведением животного, не подвергавшегося инъекции. *Чистый* эксперимент предполагает, что они вводят только диазепам в мозг крысы. Однако этого сделать было нельзя: диазепам должен вводиться путем инъекции в стенку брюшной полости (внутриперитоническая инъекция). Через 15 мин диазепам уже всасывается (но еще не начинает выводиться из организма) и животное готово к проверке.

Другим условием была инъекция нейтрального вещества в брюшину. Обычно это вещество (называемое *плацебо*) состоит из физиологического солевого раствора той же концентрации, что и жидкость, входящая в состав организма. Поэтому раствор не вызывает никаких химических реакций. В эксперименте инъекция плацебо была названа контрольным условием, а инъекция диазепама получила название лекарственного, или экспериментального, условия. Именно в этом смысле термин *плацебо* разъясняется в Американском толковом словаре (American Heritage Dictionary, 1973, с. 1001): это «нейтральное вещество, используемое в экспериментах для контроля».

Контроль искусственного смешения

Предположим, что Гаффан использовал в качестве контрольной группы неоперированных обезьян. В этом случае невозможно было бы установить, насколько зависят и зависят ли вообще обнару-

женные различия от повреждения свода, а не других тканей (т.е. фактически от воздействия самой операции). И такое сравнение мы можем изобразить следующим образом:

Состояние окружающей области	Состояние свода	
	рассеченный	нетронутый
Поврежденная	Оперированная группа	
Неповрежденная		Неоперированная группа

Это сравнение условий совершенно несостоятельно с точки зрения внутренней валидности, так как две группы обезьян различаются не только состоянием свода, но и состоянием окружающей свод области. Корректным будет только такое сравнение, когда условия на диаграмме расположены на одной линии, здесь — горизонтали, поскольку она означает один и тот же уровень побочной переменной. Это достигается путем сравнения группы с рассеченным сводом с контрольной группой, т.е. с группой, прошедшей операцию и отличающейся только состоянием свода. В этом случае получаем следующие результаты:

Состояние окружающей области	Состояние свода	
	рассеченный	нетронутый
Поврежденная	Экспериментальная оперированная группа	Контрольная оперированная группа

Теперь сравнение делается по горизонтали и, таким образом, смешение между состоянием свода и состоянием окружающей области устранено.

Р. Вайз и В. Досан (Wise R. A., Dawson V., 1974) выбрали схему эксперимента, при которой каждое подопытное животное помещалось в оба условия, а не схему, предусматривающую участие в каждом условии отдельной группы животных. Однако и у них проблема смешения и ее решение логически были такими же, что и в эксперименте Гаффана. Если бы они для контроля использовали животных, которым не сделали никакой инъекции, то условия на диаграмме различались бы как по «вертикали», так и по «горизонтали», свидетельствуя о том, что они различаются сразу по двум параметрам.

Применение инъекции	Введение диазепама	
	Да	Нет
Да	Применение диазепама	
Нет		Без какой-либо инъекции

Устранить это смещение можно, изменив контрольное условие таким образом, чтобы оно отличалось от экспериментального только введением диазепама. Итак, получаем следующие результаты:

Применение инъекции	Введение диазепама	
	Да	Нет
Да	Применение диазепама	Применение плацебо

И снова сравнение проводится по горизонтали, так что смещение между применением диазепама и просто инъекцией ликвидируется.

В этих экспериментах независимая переменная имела два уровня: *активный* — рассечение свода, введение диазепама, а также *неактивный* — свод не затронут, диазепам не введен. Уровень переменной называется активным тогда, когда экспериментатор произвел какие-то действия. Искусственно вызванная переменная также имеет как активный уровень (повреждение окружающей области, инъекция), так и неактивный уровень (без инъекции, без повреждения).

В обоих случаях *активное* условие рассечение свода или применение диазепама вызывает *активный* уровень другой переменной. Это называется *артефактом* эксперимента. Связь такого рода в природе не существует. Она вызывается исключительно действиями экспериментатора. Если сравнение производится с группой, на которую не было оказано никакого воздействия (неактивный уровень), различие будет не только в уровне независимой переменной, но также в уровне артефактной переменной. Если бы был возможен чистый эксперимент, то артефакта бы не было, так как экспериментатор манипулировал только с независимой переменной. Просто игнорировать существование артефакта, используя для сравнения только *неактивные* условия воздействия (без операции, без инъекции), нельзя, так как это сразу уведет нас от идеального эксперимента, хотя требования чистого эксперимента будут удовлетворены. Использование активных контрольных условий (частичная операция, плацебо) при-

ближает исследователя к идеальному эксперименту. В отношении артефактной переменной два условия воздействия просто неразличимы.

Такой эксперимент, как бы далек он ни был от чистого эксперимента — ведь в экспериментальных гипотезах ничего не говорится о повреждении других тканей или инъекциях в брюшину, — все же, несомненно, ближе к нему, чем эксперимент, сравнивающий активное условие воздействия с неактивным. Мы можем подытожить суть искусственного смешения и использование контрольных условий для его устранения следующим образом.

1. Эксперимент с искусственным смешением

Артефактная переменная	Независимая переменная	
	Активный уровень	Неактивный уровень
Активный уровень	Активное условие	
Неактивный уровень		Неактивное условие

2. Эксперимент, контролирующий искусственное смешение

Артефактная переменная	Независимая переменная	
	Активный уровень	Неактивный уровень
Активный уровень	Экспериментальное условие	Контрольное условие

Контрольное условие в данном случае содержит *неактивный* уровень независимой переменной и *активный* уровень артефактной переменной. Оно сравнивается с экспериментальным условием, которое состоит из активного уровня обеих переменных.

Когда эксперимент строится по схеме, предусматривающей использование только одной группы испытуемых, каждого испытуемого подвергают воздействию как экспериментального, так и контрольного условия. В случае использования межгрупповой схемы эксперимента две группы испытуемых, которые подвергаются сравнению, как правило, называются соответственно *экспериментальной группой* и *контрольной группой*.

ЕСТЕСТВЕННОЕ СМЕШЕНИЕ

Эксперимент с плачущим ребенком

Спросите мать, плачет ли ее ребенок, когда она уходит, и, скорее всего, вы получите положительный ответ. Но исследовате-

ли в области социологии развития обычно стараются получить более объективное свидетельство. Им бы хотелось установить, насколько важен человек, к которому привязан ребенок (обычно мать), в социальном развитии ребенка. Если в жизни младенца действительно существует период, когда уход матери расстраивает его до слез, то это нужно учитывать при рассмотрении возможных путей социального развития.

Эта идея и послужила основой для эксперимента на привязанность детей, т. е. для изучения поведения, являющегося следствием этой привязанности. Эксперимент был проведен двумя психологами — Д. Флинером и Р. Кернсом (Fleener D. E., Cairns R. V., 1970, с. 215—223). Когда они обратились к литературе, обнаружили значительные противоречия: например, «если один исследователь установил, что ребенок начинает плакать при разлуке с матерью в возрасте от 3 до 6 месяцев, то другие делают вывод, что в среднем это начинает происходить в возрасте 7—9 месяцев» (там же, с. 215).

Методика. Наши исследователи отметили, что в экспериментах их предшественников независимая переменная менялась не в строгом соответствии с тем, что утверждалось в экспериментальной гипотезе. Дети действительно плачут, когда мать уходит. Но, может быть, то же самое произойдет, если уйдет любой другой человек, с кем они провели какое-то время? Исследователи организовали эксперимент таким образом, чтобы с ребенком одновременно находились мать и ассистент экспериментатора (молодая женщина). Затем женщины уходили поодиночке, а поведение ребенка записывалось на магнитофонную ленту и на киноплёнки. Каждая женщина отсутствовала 60 с. Для разных детей, подвергавшихся испытанию, были использованы четыре варианта чередования условий ухода матери (М) и ассистентки (А): АМММ, АММА, МАММ и МАММ.

Балльная оценка плача ребенка производилась «вслепую» экспертами, работавшими независимо с фильмами и магнитофонными пленками. Эксперты не знали, какой из просмотренных эпизодов связан с отсутствием М, а какой — с отсутствием А. Плач оценивался по количеству 5-секундных периодов, в течение которых ребенок, по мнению эксперта, плачет. Применение системы балльных оценок вслепую является важным условием контроля за систематическим *предубеждением экспериментатора*, о котором говорилось в главе 2. Если у эксперта существует априорное мнение, что плач будет сильнее в случае ухода матери, то вероятнее всего он увидит (и услышит) сцену именно таким образом.

Результаты. В эксперименте участвовали разные возрастные группы. Для самых маленьких детей большой разницы в длительности плача не обнаружилось. Однако для детей в возрасте от 12 до 14 месяцев эта разница оказалась значительной. Для 15 детей этого

возраста в условии отсутствия матери оценка равнялась 11,67, а в отсутствие ассистентки — 8,27. Специфическая реакция на отсутствие матери, таким образом, выявилась для более позднего возраста по сравнению с предыдущими данными. Проблема заключается в том, что, раз начав плакать, ребенок будет плакать даже если почти ничего (или ничего) не будет происходить.

Контрольное условие для устранения смешения путем введения расширенной переменной

Д. Флинер и Р. Кернс расценили предыдущие эксперименты как ошибочные, поскольку из результатов не было ясно, плачет ли ребенок потому, что ушла именно мать (М).

Уход человека	Уход матери	
	Да	Нет
Да	Уход М	
Нет		М не ушла

Это может быть установлено путем сравнения условий, которые различаются для обеих переменных. В данном случае *чистый* эксперимент логически невозможен. Здесь нельзя сделать так, чтобы уходила только мать, а *кто-то* другой оставался. Флинер и Кернс устранили это смешение, введя в эксперимент ассистентку А, которая *была некто*, но не М — мать. Таким образом, получилось:

Уход человека	Уход матери	
	Да	Нет
Да	Уход М	Уход А

Эти два результата, по существу, такие же, как описанные нами для искусственного смешения. Первая там была следующей:

Расширенная переменная	Независимая переменная	
	Активный уровень	Неактивный уровень
Активный уровень	Активное условие	
Неактивный уровень		Неактивное условие

Применение активного *контрольного условия*, уход А, позволяет сделать сравнение по горизонтали только по одной-единствен-

ной переменной — независимой переменной из следующей гипотезы:

Расширенная переменная	Независимая переменная	
	Активный уровень	Неактивный уровень
Активный уровень	Экспериментальное условие	Контрольное условие

И снова мы видим, что произошло смешение, когда активное условие воздействия (уход М) сравнивалось с неактивным (М остается). Это значит, что активный уровень независимой переменной включал в себя также и активный уровень расширенной переменной, а неактивный уровень независимой переменной включал в себя также и неактивный уровень расширенной переменной. Это смешение было устранено путем проверки на активном уровне другой переменной — но в данном случае не артефактной, а более широкой переменной, чем та, которая фигурировала в экспериментальной гипотезе.

Контрольные условия для устранения смешения с помощью второй переменной

В обзоре литературы о привязанности человека Л. Коэн (Cohen L. J., 1974) отмечает, что Д. Флинер и Р. Кернс (Fleener D. E., Cairns R. B., 1970) избежали еще одного смешения, что не удавалось сделать экспериментаторам до них. Она тоже указывает, что в предшествующих экспериментах ребенок оставался совсем один, когда уходила мать. Из-за этого трудно было решить, плачет ли ребенок потому, что ушла мать, или потому, что он остался один. С этой точки зрения более ранний эксперимент можно представить следующим образом:

Кто-то остается?	Уход матери	
	Да	Нет
Нет	М уходит	
Да		М не уходит

Именно это смешение было устранено тем, что либо А, либо М остается, когда одна из них уходит. Теперь эксперимент может быть изображен так:

Кто-то остается?	Уход матери	
	Да	Нет
Да	М уходит	А уходит

Кажется, что теперь смешение полностью ликвидировано и можно сделать желаемое сравнение по горизонтали — условий воздействия только независимой переменной. Но Л. Коэн указывает, что контроль все же был недостаточен. Это видно из следующей формы представления эксперимента Д. Флинера и Р. Кернса:

Кто именно остается?	Уход матери	
	Да	Нет
А	Уход М	
М		Уход А

Здесь независимая переменная — «уход матери» (Да или Нет) — смешалась с другой переменной — «кто остается» (А или М). Коэн предлагает план устранения этого смешения. Для этого необходимо ввести третьего взрослого человека, который бы постоянно присутствовал в комнате. Теперь при одном условии мать и этот третий человек будут с ребенком, а затем мать уйдет. При другом условии ассистентка и постоянно присутствующий человек будут с ребенком, а затем ассистентка уйдет.

Диаграмма контроля с использованием постоянно остающегося человека следующая:

Кто остается?	Уход матери	
	Да	Нет
Третий человек	Уход М	Уход А

Наконец-то (мы надеемся!) можно действительно провести сравнение по горизонтали с единственным различием в условиях независимой переменной — «уход матери».

Мы только что рассмотрели два случая смешения со *вторичной* переменной. Эта переменная возникла как результат экспериментальных действий, но она присуща не только эксперименту; она типична и для естественных условий. Вторичная переменная — это чье-либо присутствие вообще. Когда мать уходит, ребенок очень часто остается один. Наступающее одиночество вторично по отношению к уходу матери. Когда мать и знакомая обе находятся с ребенком, они обычно уходят по очереди, а не вместе. В зависи-

мости от того, кто из них уходит, возникает тот или другой уровень вторичной переменной — «присутствия человека».

В более общих понятиях сопутствующего смещения, искусственного или естественного (а для естественных условий — с расширенной переменной или со вторичной), диаграмма будет выглядеть следующим образом:

Эксперимент с сопутствующим смещением

Сопутствующая переменная	Независимая переменная	
	Активный уровень	Неактивный уровень
Активный уровень	Активное условие	
Неактивный уровень		Неактивное условие

Следовательно, два условия различаются по двум переменным. Они же должны различаться только по одной — независимой переменной:

Эксперимент с контролем сопутствующего смещения

Сопутствующая переменная	Независимая переменная	
	Активный уровень	Неактивный уровень
Постоянный активный уровень	Экспериментальное условие	Контрольное условие

Другими словами, экспериментатор должен активно удерживать сопутствующую переменную на одном и том же уровне. После того как мы видели все трудности тонкого выделения независимой переменной, мы должны согласиться с Г. Хардин (Hardin G., 1972), что чистый эксперимент невозможен. Нужно быть сверходаренным, чтобы придумать эксперимент, в котором мы имели бы дело с чем-то одним.

ИСТОЧНИКИ ЭКСПЕРИМЕНТАЛЬНЫХ ГИПОТЕЗ

Для прикладных исследований экспериментальные гипотезы довольно часто возникают из насущных потребностей или, по меньшей мере, из желания что-либо улучшить. Они рождаются непосредственно из повседневной жизни. Именно так возникли эксперименты с шумом в ткацком цехе, с методами заучивания фортепьянных пьес, с ночными посадками самолета. Для научных исследований как направленных на улучшение понимания поведения более типично основываться на предшествующих экспериментах. Эти эксперименты могут быть следствием прикладных ис-

следований, как, например, это уже было с проблемой наблюдения невооруженным глазом и с помощью бинокля. Однако чаще они рождаются непосредственно из предшествовавших научных исследований.

Даже если в голову приходит блестящая оригинальная идея, не стоит тот же час разрабатывать экспериментальную схему: кто-нибудь уже мог быть здесь первым. Какой смысл вторично проверять факторы поведения, которые уже были изучены. Следовательно, необходимо просмотреть *литературу* (так называются опубликованные работы по данной проблеме). Литературу можно грубо разделить на два вида: первичные и вторичные источники. Из-за огромной массы первичных источников по психологии (по меньшей мере около 500 журналов, некоторые из них издаются более 50 лет) лучше начать со вторичных источников. Лучше посмотреть самые поздние научные обзоры по общей проблеме. Они часто используются в качестве учебных пособий для студентов и аспирантов в университете. Так, например, среди этой литературы есть книги, обобщающие исследования по разрешению конфликтов, межличностному восприятию, познавательным процессам, биохимическим факторам поведения и т.д. Вы выясните актуальные проблемы и получите также целый ряд ссылок на оригинальную литературу. Полезным может оказаться обзорный ежегодник «Annual Review of Psychology», в котором за последние один или два года рассматривается 15—20 тем или областей. Подобную информацию по еще большему кругу проблем можно найти в «Psychological Bulletin», который ежемесячно издается Американской ассоциацией психологов. Статьи в журнале «Psychological Review», издаваемом каждые два месяца американской Ассоциацией психологов, носят более теоретический характер, с небольшим числом ссылок. Большинство статей представляет собой скорее изложение оригинальных идей, чем обзор работ других авторов. Все же, чтобы хорошо проводить эксперименты, вы должны знать теоретический контекст. Возможно, что изучение перечисленных источников даст вам ориентировку в интересующей вас проблеме и покажет, что необходимо изучить далее для успешной разработки других проблем.

Приступив к изучению оригинальных работ, указанных в ссылках, вы увидите, что каждая из этих статей, в свою очередь, имеет ссылки на другие работы. Считать, что вы уже достаточно читали литературы, можно только тогда, когда вам перестанут попадаться ссылки на работы, еще не изученные вами. Возможно, для этого придется ограничиться более узкой тематикой, чем вы предполагали сначала. Если вы можете пользоваться хорошей библиотекой, у вас не будет трудностей в отыскании всех необходимых статей. Если же у вас нет доступа в библиотеку, можно получить отдельные оттиски статей, обратившись к авторам.

Главное, помните: никогда не принимайте на веру чей-либо пересказ статьи; вы должны обязательно прочитать ее сами.

ОПЕРАЦИОНАЛЬНАЯ ВАЛИДНОСТЬ

Исследовательская гипотеза, которую мы хотим проверить, формулируется, как правило, в довольно общих и абстрактных понятиях, которые должны репрезентироваться в конкретном эксперименте.

В предисловии к этой главе обсуждался вопрос о том, насколько громкий звук зуммера может репрезентировать абстрактное понятие *наказания*. На самом деле он очень мало похож на наказание. Обычно наказание связывается с представлением о боли, лишении собственности или свободы, общественном осуждении. Именно в вопросе перевода абстрактных или теоретических понятий на язык конкретного эксперимента даже лучшие психологи подчас демонстрируют сомнительную логику. Если такой перевод вызывает сомнения, то говорят, что эксперимент лишен *операциональной валидности*.

Эксперимент в социальной психологии

Хорошим примером такого сомнительного перевода, который сразу же приходит на память, может служить классическое исследование К. Левина, Р. Липпита и Р. Уайта (Lewin K., Lippitt R., White R. K., 1939). В нем сравнивалось влияние авторитарной, демократической и анархической обстановки на групповое поведение 10-летних мальчиков. Посмотрим, как эти понятия переводились на язык конкретных экспериментальных операций. Авторитарная обстановка была представлена взрослым лидером, который сам принимал все решения и делал много персональных замечаний. Демократические условия также обеспечивались взрослым лидером, который, однако, способствовал коллективному решению вопросов и всячески стремился поддерживать ребят. Для создания анархической обстановки предусматривалось полное невмешательство взрослого как в деятельность группы, так и в межличностные отношения. В результате оказалось, что в условиях «демократической» обстановки было меньшее число драк, дети были более счастливы и более деятельны, чем при других режимах. Автор этих строк предпочел бы назвать описанные ситуации соответственно как деспотию, либеральную монархию и демократию. Эти действительные режимы не смогли реализоваться полностью (кроме взрослого лидера) из-за недостатка времени. Другой вопрос, связанный с операциональной валидностью, — насколько эти кратковременные эксперименты с детьми, живущими в очень разной домашней обстановке, могут вообще что-либо

говорить о поведении индивидов, которые постоянно находятся в соответствующих социальных условиях.

Эксперимент по научению

Один из самых выдающихся психологов Е. Торндайк проверил гипотезу о том, что частота упражнения сама по себе (т.е. без знания результата) ведет к научению (Thorndike E. L., 1931, 1932). В опытах на самом себе Торндайк чертил линии заданной длины (например, 4 дюйма) с закрытыми глазами, доведя число проб до нескольких сотен. К сожалению, он использовал два противоречащих друг другу критерия для оценки факта научения.

Первым критерием было увеличение точности действия; вторым — тем большее закрепление движений, чем чаще они осуществлялись в ранних пробах. Когда оказалось, что точность движений не увеличивается (Thorndike E. L., 1931), был сделан вывод, что экспериментальная гипотеза неверна. Когда же точность увеличилась (Thorndike E. L., 1932), но при этом неизбежно уменьшилась частота движений, осуществлявшихся в ранних пробах, то это также было использовано для опровержения экспериментальной гипотезы!

Снова об эксперименте полного соответствия

Были ли эти эксперименты *правильны* в том смысле, что они действительно проверяли экспериментальную гипотезу? Обладают ли они внешней валидностью? Ранее внешняя валидность определялась тем, насколько хорошо эксперимент репрезентирует безупречный эксперимент полного соответствия. Так, в отношении экспериментов, которые «улучшают» реальный мир (глава 3), ставился вопрос, насколько независимая, зависимая и побочные переменные репрезентировали реальную ситуацию, рассматриваемую в экспериментальной гипотезе. Точно так же в случае эксперимента на выборках (глава 4) обсуждался вопрос, насколько была представлена интересующая популяция. Таким образом, обсуждался тот же вопрос, хотя и в более утонченном виде.

Спорный момент исследования социальной обстановки (Lewin K., Lippitt R., White R. K., 1939) в том, насколько адекватно три экспериментальных условия репрезентируют соответствующие социальные системы. Было выдвинуто возражение, что никакая группа с самозванным лидером не может считаться демократической. Однако выполнение настоящего требования, чтобы руководитель выдвигался самой группой, сделало бы эксперимент вообще невозможным. Поэтому предполагается, что при достаточном времени анархическая обстановка может перейти в демократическую. Правда, надо признать, что со временем в той же группе вполне могла сложиться и авторитарная обстановка, как

это описано в «Повелителе Мух» (Golding W. G., 1954). Итак, можно сделать вывод, что хотя в эксперименте не достигнуто полное соответствие, он все-таки был, насколько возможно, близок к этому. Таким образом, внешняя валидность оказалась настолько высокой, насколько это было практически возможно.

Вопрос о репрезентативности в этом эксперименте связан с независимой переменной. В эксперименте по научению, проведенном Е. Торндайком, этот вопрос относится к зависимой переменной. Торндайк был талантливым исследователем, и только веская причина могла затруднить ему определение зависимой переменной. По его мнению, единственно, что можно достичь простым упражнением, без знания результатов, — это закрепить первоначальные реакции. Это должны быть те реакции, которые повторялись наиболее часто. Таким образом, когда точность движений действительно увеличилась, он решил, что это изменение не является свидетельством научения, так как он никак не мог этого объяснить. Тем не менее реально научение может происходить и без знания *внешних* результатов. Например, будучи испытуемым, Торндайк мог все более внимательно относиться к своим движениям и тем самым достигать лучшего соответствия воспроизводимых линий «внутреннему стандарту» в 4 дюйма. Теоретические представления экспериментатора не должны влиять на выбор зависимой переменной. Ведь экспериментальная гипотеза была достаточно ясной: научение означает более точное действие. Поэтому степень точности воспроизводимой линии и была наиболее адекватной зависимой переменной.

Итак, *операциональная валидность* означает, что конкретные экспериментальные операции (создание условий и получение зависимой переменной) репрезентируют независимую и зависимую переменные в недостижимом эксперименте полного соответствия. В данном случае эксперимент полного соответствия невозможен, потому что он существует в мире теорий, а не в сфере конкретных операций.

ИСПЫТУЕМЫЕ

В некоторых случаях из-за стремления к обобщению в экспериментах на выделение независимой переменной в экспериментальной гипотезе отсутствует указание на какую-либо специальную популяцию испытуемых. В других же случаях из экспериментальной гипотезы можно прямо вывести, на каких испытуемых или животных должен проводиться эксперимент.

Выбор животных

Уже по тому, поведение каких животных изучается, гипотезы можно проранжировать от очень общих до узкоспециальных. Иногда эксперименты проводятся только для того, чтобы выяснить, рас-

пространяются ли уже известные феномены поведения на другие виды животных.

Животные вообще. Голубь все чаще будет клевать светлую клавишу, если через каждые 10 ударов он получает зернышко пшеницы (совершенно аналогично тому, как это было в эксперименте по трудовой этике). Если голубь не будет клевать, то он не получит пшеницу. Иначе говоря, получение пшеницы обусловлено клеванием. А клевание обуславливает получение корма. Давайте теперь рассмотрим принцип подкрепления: когда менее вероятное поведение обуславливает более вероятное поведение, то вероятность первого увеличивается (Premack D., Schaeffer R.W., Hundt A., 1964). Так, если животное было ограничено в движениях, но ему давали много пить, то более вероятно, что затем оно будет двигаться, и менее вероятно, что будет пить. Но если возможность бегать будет поставлена в зависимость от питья, то даже не испытывающее жажду животное будет пить чаще, чтобы подвигаться. Здесь мы имеем дело с очень общим законом, применимым и к мышам, и к людям. Трудно сказать, насколько низкой должна быть ступень развития организма (простейшие), чтобы этот закон перестал действовать. Предположим, что эта гипотеза и в самом деле могла бы быть проверена на многих видах. Тогда животные для эксперимента могут выбираться уже просто из соображений удобства. В качестве лабораторных животных, как правило, отбираются особи небольших размеров, понятливые и способные выжить в неволе. Желательно, чтобы они легко размножались, с тем чтобы указанные качества закрепилась. Получаемое при этом единообразие животных удобно еще и тем, что позволяет сравнивать эксперименты, выполненные в разных лабораториях. Однако у такого отбора животных есть и отрицательные стороны. Р. Локард (Lockard R.V., 1968) обратил внимание на то, что лабораторная крыса не является репрезентативным животным. Она развивалась в тесном контакте с людьми и во многих отношениях дегенерировала. Необыкновенная понятливость этих крыс является результатом слабого развития системы эндокринных желез. Эти сомнения поддерживаются наблюдениями К. Бриланда и М. Бриланда (Brealand K., Brealand M., 1961) над нелабораторными видами животных, которые совсем не склонны увеличивать свою активность с увеличением подкрепления. Например, свиньи довольно скоро перестают нажимать на рычаг для получения пищи и вместо этого начинают рыть землю, т. е. переходят к поведению, связанному с добыванием пищи в нормальных условиях.

Специальные животные. Некоторые эксперименты проводятся, чтобы проверить гипотезу, ограниченную только данным видом животных, мужскими или женскими особями, молодыми или старыми и т. д. Если эксперимент проводится для проверки гнездового поведения у голубей, то он должен проводиться именно на голубях.

Другой причиной использования специального вида животного может быть изучение реакции клеток головного мозга на стимуляцию глаза светом разной длины волны. Некоторые методы возможны только при условии, что исследование проводилось на чисто колбочковой сетчатке, которая обнаружена у очень небольшого числа видов, например у сусликов (Jacobs G. H., Anderson D. N., 1972). Точно также принципы обучения чтению могут быть изучены только на маленьких детях.

Люди вообще. Большинство разделов психологической науки, включая экспериментальную психологию, на самом деле относятся только к человеку. Так, существуют эксперименты, рассматривающие поведение людей в социальных группах, их способность определять высоту звука или воспринимать два различных сообщения, поступающих одновременно на оба уха. Мы не будем сейчас обсуждать вопрос о том, что лучше, — психология, которая затрагивает только людей, или психология, охватывающая и много разных видов животных. Следует отметить, что очень многие эксперименты, направленные на изучение человека, проводились на «двойнике» лабораторной крысы — студенте. В некоторых исследованиях это может привести к излишнему обобщению результатов: частное подтверждение гипотезы принимается за общее. Например, вполне возможно, что многие студенты колледжа, в силу того что они впервые оказались далеко от дома и находятся в возрасте, когда межличностные отношения особенно важны, поддадутся давлению группы в эксперименте, тогда как другие лица в этом случае проявили бы большую самостоятельность. С другой стороны, если вопрос заключается в том, может или не может человек в принципе совершать определенные действия, например внимательно слушать одновременно два разговора, то в этом случае студенты колледжей представляют собой богатый материал. Ведь они так расторопны!

Отбор и распределение испытуемых

Когда нам уже известна общая популяция, из которой должны быть выбраны испытуемые, нам все еще нужно произвести специальный их отбор или распределение по подгруппам. В большинстве случаев может быть использована уже существующая группа, такая, например, как класс в средней школе. Используемая схема всегда отражает представление исследователя о том, как достичь наибольшей *внутренней валидности*, т. е. избежать эффектов последовательности или неэквивалентности групп. В экспериментах с выделением независимой переменной экспериментатор редко сталкивается с проблемой выбора таких испытуемых, которые представляют всю популяцию. (Это было основной темой предыдущей главы.) Дело в том, что для выделения независимой переменной можно использо-

вать любых испытуемых. Проблема же обобщения решается за счет повторения экспериментов на других испытуемых.

Типичные межгрупповые стратегии. Предварительные меры. Если испытуемые отбираются из большой популяции, то для формирования экспериментальных групп используется метод случайного отбора. Если в эксперименте должны участвовать все испытуемые, то наиболее подходящим будет метод случайного распределения. Обе эти схемы были описаны в главе 4. До того как эксперимент начнется, должно быть установлено, кто именно будет в группе, относящейся к каждому условию.

Серийное решение. Зачастую для формирования экспериментальной группы применяют метод произвольного выбора испытуемыми удобного для них времени явки на эксперимент из общего экспериментального времени, т. е. метод расписания. Такая процедура может быть применена в том случае, когда в эксперименте участвует как целая группа, так и отдельные индивиды. Предположим, что есть два экспериментальных условия. Испытуемые распределяются поочередно по ним согласно времени их явки. В таком случае надо следить, чтобы условие А в результате не стало предъясняться регулярно в другое время дня или в другие дни недели, чем условие Б. Если испытуемых достаточно много, то наиболее приемлем случайный порядок распределения. Можно провести предварительное испытание, чтобы уменьшить разницу между группами. Например, если в эксперименте на время реакции средней показатель в предварительном испытании для формирующейся группы А оказался равен 140 мс, а для формирующейся группы Б — 150 мс, то новый испытуемый с предварительным результатом в 160 мс должен быть отнесен к условию А. Обычная ошибка состоит в том, что сначала набираются все испытуемые на условие А, а затем — на условие Б. Испытуемые, которые выбраны для эксперимента раньше или позже, могут значительно различаться по параметрам, существенным для результатов эксперимента.

Использование в высшей степени сходных индивидов. Иногда в экспериментах на животных сравниваемые группы состояются из особей одного и того же помета. В этом случае группы оказываются очень похожими. Такой метод особенно эффективен, если выводок состоит из животных с абсолютно одинаковой наследственностью, как у опоссумов. Для некоторых видов амфибий такой результат может быть получен посредством новой техники *клонирования*. В яйцеклетку, удалив предварительно ее ядро, вводят идентичные клеточные ядра. Наибольшее приближение к генетической идентичности достигается у мышей.

В экспериментах на людях следовало бы более широко использовать идентичных (монозиготных) близнецов. Особенно ценно, если они росли вместе и воспитывались одинаково. Правда, если

исследуется относительный вклад наследственности и окружения, то одинаковые условия воспитания окажутся уже нежелательным фактором смешения.

Эксперименты с одним испытуемым

Для выделения независимой переменной можно использовать индивидуальную схему эксперимента. По существу, именно таким был эксперимент по трудовой этике. К изложенным в главе 2 комментариям к схеме этого эксперимента нечего добавить. Остается только указать область применения экспериментов с одним испытуемым.

Мы начали эту книгу с описания экспериментов с одним испытуемым, которые дают прямой практический результат. Такой же метод применялся и в более строгих экспериментах (Dukes W. F., 1965; Gottman J. M., 1973; Holtzman W., 1963). Экспериментальное исследование памяти фактически было начато по такой же схеме Эббингаузом (Ebbinghaus H., 1964), использовавшим бессмысленные слоги. Для большинства изученных им факторов эффект переноса от предыдущего эксперимента (или двух) был, вероятно, не слишком значительным из-за того, что запоминалось несколько дюжинок списков.

Если мы проверяем гипотезу о наличии какой-то способности, достаточно проверить ее на одном человеке. Раз один человек может одновременно слушать и понимать сразу два сообщения, значит, это — в пределах человеческих возможностей. То же самое, если хотя бы один человек проявляет телепатические способности, демонстрируя без использования органов чувств знание того, что думает другой, значит, и эта способность находится в пределах человеческих возможностей. В большинстве случаев необходимо особенно позаботиться о достижении позиционного равновесия условий, предъявляемых одному испытуемому. В исследованиях с оперантным обусловливанием, как, например, у Примака, Шеффера и Хандта (Premack D., Schaeffer R. W., Hundt A., 1964), о котором упоминалось ранее, *фоновый уровень* ответа определялся до ввода подкрепления. Его следовало бы найти и после того, как уровень подкрепляемых ответов снизится с прекращением подкрепления. И только в совокупности обе эти оценки дадут фон, относительно которого можно определять эффективность режима подкрепления.

ИТОГИ И ПЕРСПЕКТИВЫ

В предыдущих главах этой книги мы шли от конкретного и частного к *абстрактному* и общему. Мы начали с вопроса, какое из двух условий лучше для данного человека. Единственное обобщение

ние, которое было здесь возможно, — это использование данным человеком найденного условия в будущем. Когда было необходимо «улучшать» реальный мир, чтобы добиться внутренней валидности, экспериментатору пришлось принимать решение, какие именно свойства реального мира нужно скопировать, т. е. ему пришлось абстрагироваться от реального мира. Большее обобщение в отношении индивидов было достигнуто благодаря замене отдельного конкретного испытуемого выборкой из популяции. Теперь стало труднее решить, что лучше для одного конкретного индивида, но легче — что лучше для любого представителя популяции.

В настоящей и последующих главах продолжается абстрагирование и обобщение. В этой главе абстрагирование и обобщение тесно взаимосвязаны. Здесь мы уже не задавали себе вопрос, какое условие или уровень независимой переменной действует лучше, чем другие. Вместо этого мы отвечали на вопрос, что именно в данном условии обуславливает разницу в поведении.

Давайте рассмотрим две части последней фразы. «Что именно в данном условии» означает выделение из сложной переменной действительной независимой переменной, иначе говоря, абстрагирование существенного свойства. «Обуславливает разницу в поведении» предполагает меньший, чем ранее, интерес к условиям «лучшим, чем другие». Это связано с тем, что в научных исследованиях понятие «лучшее», как правило, неадекватно. Например, в исследовании методов обучения может быть выделена действительная переменная, которая при одном условии приводит к быстрому овладению иностранным языком, а при другом — к быстрому развитию невроза. Подобно этому могут быть найдены одинаковые жизненные процессы и в клетках тела, и в бактерии, вызывающей болезнь. Таким образом, действие изолированной (абстрагированной) независимой переменной может быть широко обобщено.

В данной главе наши основные интересы были сосредоточены на способах контроля, необходимого для выделения независимой переменной. Мы еще вернемся к вопросу о том, отвечает ли выявленное действие независимой переменной определенным критериям надежности. Вслед за этим мы обсудим эксперименты, которые более полно выявляют механизмы влияния независимых переменных на поведение. Прежде всего это будут эксперименты, в которых используется несколько уровней независимых переменных, а не просто два различных условия. Затем мы убедимся в том, что более ясное представление о действии независимой переменной может быть получено в более сложных экспериментах с несколькими независимыми переменными. Таким образом, можно видеть, что к научному пониманию детерминации поведения ведет целая серия шагов.

Последней темой будут корреляционные методы. До сих пор единственная ссылка на них была в главе 1, в которой говорилось, что корреляционный подход не выдерживает сравнения с экспериментальным манипулированием независимой переменной. Тем не менее корреляционный метод тоже служит путем для изучения поведения и внес свой ценный практический и научный вклад. Мы рассмотрим характер этого вклада и возможности контроля в подобных экспериментах.

КРАТКОЕ ИЗЛОЖЕНИЕ

Для научного понимания механизмов поведения необходимы эксперименты, в которых выделяется истинная независимая переменная. Проверяемая в них экспериментальная гипотеза касается уже единичной независимой переменной, в отличие от «группы факторов», которая характерна для экспериментов с немедленным практическим применением результатов. В этих новых исследованиях к экспериментальной гипотезе предъявляются более высокие требования, чем это было раньше, поскольку описываемая независимая переменная должна быть «очищена» от других переменных. Эталоном для оценки внутренней валидности такого эксперимента, в котором делаются попытки выделить единичную независимую переменную, является чистый эксперимент. Он представляет собой прежде всего идеальный эксперимент, в котором испытуемому одновременно предъявляются два разных условия независимой переменной. Кроме того, он требует, чтобы независимая переменная была единичной, а не сложной, и чтобы другие переменные, которые могут влиять на поведение, сохранялись строго постоянными.

Прежде всего мы показали в эксперименте с индианками и в эксперименте по трудовой этике, что для выделения независимой переменной должны быть созданы специальные условия. Например, если эксперимент проводится в реальной рабочей обстановке, то невозможно ограничить условия различием только по одной переменной. В некоторых экспериментах требуется большое мастерство, чтобы «очистить» независимую переменную. Например, при изучении направленного обоняния исследователь очень постарался, чтобы точно задать условия независимой переменной — разницу во времени — и ликвидировать другие переменные, т. е. щелчки реле.

Внутренняя валидность нарушается далее из-за некоторых новых типов систематического смещения. Они проистекают из того печального факта, что мы не можем делать лишь что-то одно. Во всех случаях эти типы смещения возникают от того, что *активное* условие независимой переменной сопровождается активным уровнем побочной переменной. Таким образом, активный уровень независимой переменной связан с активным уровнем побочной переменной, а неактивный уровень независимой переменной связан с неактивным уровнем побочной переменной. В итоге независимая переменная смешивается с побочной переменной, и расчленить их влияние на поведение невозможно. В любом случае средством контроля такого сопутствующего смещения была схема, в которой неактивный уровень независимой переменной включал в себя активный уровень сопутствующей пе-

ременной. Это называется *контрольным условием*; если же используется межгрупповая схема эксперимента, то такая комбинация представлена контрольной группой. Активный уровень независимой переменной в этом случае называется *экспериментальным условием*, а соответствующая группа — *экспериментальной группой*.

Первый тип обсуждавшегося сопутствующего смешения — смешение с артефактной переменной. Он рассматривался на примерах из психофизиологии. Если животные, у которых удалена часть мозга, сравниваются с животными, не оперированными вообще, то независимая переменная (удаление или неудаление части мозга) смешивается с артефактной переменной (повреждение других тканей). Поэтому необходимо ввести в эксперимент прооперированную контрольную группу животных для сравнения с оперированной экспериментальной группой. Эта контрольная группа подвергается той же самой операции, исключая удаление исследуемой части мозга. Подобным же образом в эксперименте, в котором введение лекарства является экспериментальным условием, должно быть контрольное условие с введением *плацебо*.

Для иллюстрации других типов сопутствующего смешения использовался эксперимент, проверяющий гипотезу, что дети плачут из-за ухода близкого им человека (матери). К этим типам относится, во-первых, смешение с *расширенной* переменной, когда вообще кто-нибудь уходит. Во-вторых, смешение с *вторичной* переменной — ребенок находится с кем-то или совсем один. В-третьих, смешение с другой вторичной переменной — с кем ребенок остается. Все это были иллюстрации сопутствующего смешения, называемого *естественным* смешением, в отличие от *искусственного* смешения, описанного выше. Термин «естественное» означает только, что связь уровней независимой переменной и уровней побочной переменной заложена в «природе вещей», а не вызывается исключительно действиями экспериментатора. Способ контроля естественного смешения тот же, что и в случае искусственного смешения: создание контрольного условия, которое включает в себя активный уровень сопутствующей переменной. Если нарисовать диаграмму уровней, по которым различаются условия, то можно увидеть, что контрольное условие делает возможным однолинейное сравнение только по независимой переменной.

Следующей была рассмотрена проблема *операциональной* валидности. Как и в случае с внешней валидностью, вопрос здесь тот же: *правильный ли* эксперимент проведен для проверки экспериментальной гипотезы. Переменные в экспериментах с выделением независимой переменной, как правило, выражены в довольно абстрактных понятиях; экспериментальные же ситуации — гораздо конкретнее. Поскольку в экспериментах приходится делать конкретные вещи, эксперимент полного соответствия невозможен. Можно лишь постараться, чтобы конкретные экспериментальные операции репрезентировали понятия, содержащиеся в экспериментальной гипотезе. Для иллюстрации приводилось исследование влияния социальной атмосферы на поведение людей. В отношении этого исследования рассматривался вопрос, насколько точно условия независимой переменной соответствуют общим понятиям, которые они по предположению репрезентируют. Затем был проанализирован эксперимент по научению с точки зрения репрезентативности независимой переменной.

Мы рассмотрели вопрос об испытуемых, постепенно сужая их контингент. Некоторые эксперименты касаются всех животных, так что выбор конкретного вида зависит главным образом от соображений удобства. Если испытуемыми являются лабораторные крысы, то вопрос об общности результатов становится особенно актуальным. В отдельных случаях выбирается специальный вид животных из-за их характерных особенностей. Для большого количества экспериментов реальный интерес представляет поведение людей. Использование студентов колледжа вполне допустимо, но вопрос об общности результатов необходимо проверять в контексте каждого конкретного эксперимента.

И все же необходимо проводить отбор или распределение испытуемых. Как правило, исследователь в эксперименте с выделением независимой переменной не пытается получить такую выборку испытуемых, которая представляла бы некоторую популяцию. Вместо этого он (или она) сосредотачивается на вопросах внутренней валидности и проблему обобщения результатов решает путем обследования разных групп испытуемых в отдельных экспериментах.

В межгрупповых схемах применяются две различные стратегии. Одна из них включает предварительные решения, которые обсуждались на примере экспериментов с репрезентативными выборками. Другая является стратегией распределения испытуемых по группам. Если число испытуемых в группе невелико, то по мере появления новых испытуемых они распределяются по группам в последовательности регулярного чередования. При большем числе испытуемых лучше случайная последовательность распределения, так как она устраняет возможность нежелательных систематических влияний.

Для улучшения результатов, основанных на сравнении групп, могут быть использованы особенно сходные индивиды. Для этого в сравниваемые условия могут, например, помещаться особи одного помета, а в случае с людьми — монозиготные близнецы.

В современных исследованиях схемы индивидуального эксперимента используются так же часто, как и межгрупповые. Небезынтересно отметить, что эта схема может даже вылиться в эксперимент с одним-единственным испытуемым. В этом случае иногда возможен надежный контроль эффектов последовательности и временных эффектов.

Мы дали краткую характеристику направления нашего движения в предыдущих и последующих главах. Было отмечено, что мы начали с весьма конкретных исследований, позволяющих производить минимальные обобщения. Затем мы перешли к более абстрактным переменным и к более широким обобщениям. В последующих главах будут описаны эксперименты, которые выявляют более полный, менее ограниченный набор факторов, влияющих на поведение.

Контрольные вопросы и задания

1. Сравните эксперимент по спасательным работам на море из главы 3 с экспериментом по трудовой этике с точки зрения характера независимой переменной — изолированная она или составная.

2. Сравните эксперимент на предпочтение сорта томатного сока из главы 1 с экспериментом на направленное обоняние в отношении чистоты экспериментальных условий.

3. Что мы подразумеваем под контрольной группой в эксперименте, требующем удаления части мозга?

4. В чем разница между процедурным смещением, которое возникает из-за того, что эксперимент не может быть идеальным, и сопутствующим смещением как результатом того, что эксперимент не может быть чистым?

5. Приведите пример естественного сопутствующего смещения и покажите, каким образом оно может быть устранено. Составьте таблицу для пояснения ситуации.

6. Как сформулировать гипотезу, которую можно проверить экспериментальным путем?

7. Как используется идея эксперимента полного соответствия в экспериментах на выделение независимой переменной?

8. Как экспериментатор решает вопрос об использовании того или иного вида животных в эксперименте?

9. Что подразумевается под сериальными стратегиями распределения испытуемых по группам?

10. Всегда ли эксперименты на выделение независимой переменной проводятся на репрезентативных выборках? Хорошая ли это идея?

11. Опишите логическую последовательность предшествующих и последующих тем в данной книге.

Глава 6

ЗНАЧИМЫЕ РЕЗУЛЬТАТЫ¹

Давайте рассмотрим две группы результатов, полученных Д. Флинером и Р. Кернсом (Fleener D. E., Cairns R. B., 1970) в эксперименте, описанном в предыдущей главе, где сравнивались сила и продолжительность плача детей при уходе матери и ассистентки экспериментатора. Плач каждого ребенка оценивался в течение 24 периодов по 5 с как после ухода матери, так и после ухода ассистентки. 15 детей в возрасте от 12 до 14 мес плакали в среднем во времени 11,67 из этих 5-секундных периодов, когда уходила мать; когда же уходила ассистентка, эти дети плакали во время 8,27 таких периодов. На основе средней разницы в 3,40 Флинер и Кернс заключили, что дети указанной возрастной группы плачут больше, когда уходит мать.

В младшей группе, состоящей из 13 детей в возрасте от 9 до 11 мес, аналогичные средние данные оказались: 9,08, когда уходила мать, и 8,15, когда уходила ассистентка. Флинер и Кернс сделали вывод, что это небольшое различие — всего лишь 0,93 — незначимо. Их вывод, несомненно, представляется правильным, ибо разница действительно очень мала. Но достаточно ли велика была разница между 11,67 и 8,27 для более старшей группы, чтобы подтвердить вывод о значимом различии? Откуда экспериментаторы знают, какова должна быть разница между двумя условиями, чтобы ее можно было принять как значимую?

Логика их рассуждений была несложной. Они понимали, что разница для старшей группы может оказаться случайной. Как мы уже отмечали, есть много причин, по которым с течением времени может изменяться поведение как одного индивида, так в среднем и целой группы. В любом частном эксперименте более сильный плач при уходе матери мог быть случайным. Да, случайное различие было возможно, но — *маловероятно*. Исследователи имели возможность вывести, что такое или большее различие возможно не более чем в одном эксперименте из двадцати. Их устраивала эта оценка, и они отвергли предположение о том, что их эксперимент был как раз тот самый один из двадцати.

С другой стороны, для более младшей группы разница между 9,08 и 8,15 могла оказаться случайной более чем в одном опыте из

¹ Перевод Ч. А. Измайлова.

двадцати. И экспериментаторы, таким образом, приняли ее за случайную.

Мы увидим в этой главе, что Д. Флинер и Р. Кернс (Fleener D. E., Cairns R. V., 1970) проверяли *нуль-гипотезу*, состоящую в том, что экспериментальные условия не различаются. Термин «*нуль*» в данном случае означает нулевое различие. В случае со старшими детьми они *отвергли* нуль-гипотезу, в случае с более младшими — нет.

Такое действие называется *проверкой на значимость*, или на *статистическую значимость*. Когда нуль-гипотеза отвергается, то говорят, что различие статистически значимо; когда нуль-гипотеза не отвергается, то говорят, что различие (статистически) *незначимо*.

Мы увидим, что статистическое решение, принять или отвергнуть нуль-гипотезу, всегда таит в себе двойкий риск. Мы рассмотрим, как подобные статистические решения приводят к выводам относительно экспериментальной гипотезы. При этом мы снова обратимся к понятию внутренней валидности, и далее к более специальному понятию — надежности.

Наконец, мы попытаемся показать более широкий смысл проверки на значимость. Конечно, она является средством получения валидных выводов об экспериментальной гипотезе, но это еще далеко не все. Главная тема настоящей главы — *значимые результаты* — выходит далеко за пределы технического вопроса о статистической значимости.

Мы собираемся изложить в этой главе вопросы о статистических выводах несколько нетрадиционным способом — без уравнений или вычислений. Последние можно найти, как и в предыдущих главах, в статистическом приложении. Таким образом, вы не сможете сами проводить проверку статистической значимости до тех пор, пока не познакомитесь с этим приложением. Однако те идеи, которые важны для экспериментаторов, рассматриваются достаточно детально. Если вы разберетесь в них, это поможет вам при чтении экспериментальных статей, поскольку вы сможете увидеть, как авторами были сделаны заключения. Вы узнаете, какие статистические решения можно сделать относительно нуль-гипотезы и как они относятся к экспериментальным выводам. Возможно даже, что вы не согласитесь с каким-нибудь исследователем либо в связи с использованным правилом статистического решения, либо в связи с выводом, сделанным на основе применения этого правила.

Основные темы, по которым вам будут заданы вопросы в конце главы, следующие.

1. Как проверяется нуль-гипотеза?
2. Виды риска при принятии статистического решения.
3. Как проверка нуль-гипотезы связана с внутренней валидностью?

4. Как этот вид валидности входит в более общую картину экспериментальной валидности?

НУЛЬ-ГИПОТЕЗА

Кажется весьма странным проверять нуль-гипотезу о том, что интенсивность плача *не* различается в случаях, когда комнату покидает мать и когда уходит ассистентка. Ведь это противоречит тому, что предполагает экспериментатор. Экспериментальная гипотеза состоит как раз в том, что плач *сильнее*, когда уходит мать.

Существуют два основания для такого «хода от противного». Первое состоит в том, что любой реальный эксперимент (который не является ни идеальным, ни бесконечным) не может быть абсолютно доказательным. Мы никогда не сможем сказать, что безусловно и навсегда доказали, что наши условия различные. Мы не в состоянии «доказать» экспериментальную гипотезу. Самое большее, что мы можем сделать, — это показать, что *альтернативные* объяснения неправильны, что приводит нас ко второму основанию обращения к нуль-гипотезе. Это специфическая гипотеза, и ее отвержение имеет большой смысл. Так как она специфическая (разница между условиями равна нулю) в отличие от экспериментальной гипотезы (для одного условия показатель больше), она доступна стандартной статистической проверке. Это и составляет ее смысл. Ведь если неверно, что данные условия не различаются, значит, мы точно знаем, что они в *чем-то* различны.

Третье возможное заключение

Очевидно, вам было не очень приятно узнать из главы 2, что в любом эксперименте приходится делать одно из двух конкурирующих заключений: 1) подтверждена экспериментальная гипотеза о том, что зависимая переменная имеет более высокое значение для условия А, чем для условия Б; 2) подтверждена противоположная гипотеза о большем значении зависимой переменной для условия Б, чем для условия А. (Наверное, все это больше огорчает вашего преподавателя.) Что же тогда можно сказать о выводе, что ни одна из конкурирующих гипотез не подтвердилась? Вообще говоря, мы не должны заботиться об этом третьем возможном заключении при использовании результатов таких простых экспериментов, которые описаны в начале книги. Ткачиха может либо носить наушники, либо нет. Она не выберет компромиссного решения носить только один из них, если разница в пользу наушников окажется очень незначительной. Если нет проблемы стоимости или удобства, почему бы не принять к сведению любую полученную разницу, как бы мала она ни была? Всегда есть неко-

торая вероятность того, что условие, обнаруживающее преимущество в эксперименте, сохранит его и в будущем. Другими словами, в случае таких простых практических решений правило состоит в том, чтобы учитывать любые позитивные данные. Тогда в процессе решения для третьего заключения не остается места.

Однако в экспериментах, подобных описанному в предыдущей главе, где ложное заключение нанесет ущерб научному знанию, необходимо рассматривать третье возможное заключение, состоящее в том, что независимая переменная оказалась просто неэффективной. Итак, на основании результатов эксперимента Д. Флинера и Р. Кернса можно было сделать три заключения, каждое из которых относится к тому, что могло бы быть получено в бесконечном или идеальном эксперименте.

1. Подтвердилась гипотеза, что дети данной возрастной группы плачут больше, если уходит мать.

2. Подтвердилась гипотеза, что дети плачут больше, если уходит ассистентка.

3. Ни одна из приведенных гипотез не подтвердилась.

Указанные исследователи понимали, что в любом реальном ограниченном эксперименте как положительные результаты (плач сильнее, когда уходит мать), так и отрицательные результаты (плач сильнее, когда уходит ассистентка) могут быть чисто случайными. Поэтому только достаточно большое различие в интенсивности плача при уходе матери по сравнению с уходом ассистентки могло бы рассматриваться как подтверждение экспериментальной гипотезы о том, что то же самое обнаружится в идеальном или бесконечном эксперименте. Меньшее различие имело бы весьма высокую вероятность оказаться случайным.

Мы можем представить связь между полученным различием и заключением, сделанным на его основе при помощи следующей диаграммы.

Заключение — результаты подтверждают:

Противоположную гипотезу: плач сильнее, если уходит ассистентка	Ни одну из гипотез	Экспериментальную гипотезу: плач сильнее, если уходит мать
-5 -4 -3 -2	-1 0 +0,93 +1	+2 +3 +3,40 +4 +5

Интенсивность плача при уходе матери минус интенсивность плача при уходе ассистентки

Тонкая вертикальная черточка над значением +3,40 справа показывает, что это различие (для старшей группы) было достаточно для подтверждения экспериментальной гипотезы, что плач

сильнее, если уходит мать. С другой стороны, вертикальная отметка над значением $+0,93$ (различие для более младшей группы) показывает, что это различие недостаточно для подтверждения экспериментальной гипотезы. Для того чтобы подтверждалась либо экспериментальная гипотеза, либо противоположная гипотеза, требуется различие порядка ± 3 единицы.

Отвержение или неотвержение нуль-гипотезы

Правило статистического решения. Три рассмотренных выше возможных заключения из результатов эксперимента делаются на основе *правила статистического решения*. Здесь оно состояло в том, что нуль-гипотеза может быть отвергнута только в случае, если вероятность получения различия, удовлетворяющего нуль-гипотезе, меньше, чем $0,05$ (т. е. меньше 1 из 20).

Основа статистического вывода. Если бы Д. Флинер и Р. Кернс многократно повторяли свой эксперимент на новых группах детей той же возрастной категории, они бы не получали в каждом эксперименте разницу между средними для ухода матери и для ухода ассистентки, в точности равную $3,40$. Из-за случайных вариаций эта разница была бы то больше, то меньше. Если бы для бесконечного числа повторений общая средняя разница равнялась в точности 0 , это означало бы справедливость нуль-гипотезы. Однако для каждого отдельного эксперимента можно было бы ожидать значение, отличное от нуля.

Итак, разность «мать — ассистент» будет варьировать от эксперимента к эксперименту. Величина разброса этих разностей зависит от надежности каждого эксперимента. Как мы видели в главе 2, надежность выше и, следовательно, разброс от эксперимента к эксперименту меньше, чем больше число наблюдений и чем меньше случайных вариаций. Поэтому разброс разностей «мать — ассистент» был бы меньше, если бы каждый эксперимент проводился на большом числе испытуемых и имел небольшое стандартное отклонение.

Из числа испытуемых и стандартного отклонения можно *вывести* величину различия, которая при справедливости нуль-гипотезы может быть превышена лишь с вероятностью $0,05$. Нахождение этой величины называется *статистическим выводом*. Такая величина для эксперимента Д. Флинера и Р. Кернса оказалась порядка ± 3 . (Она была определена с помощью статистической процедуры нахождения так называемого *t-критерия*. Описание ее можно найти в статистическом приложении к данной главе. Это один из многих критериев статистической значимости, используемых экспериментаторами.)

Диаграмма на с. 184 показывает, как в эксперименте Д. Флинера и Р. Кернса применялось *правило статистического решения* для получения одного из трех возможных выводов.

Как видно, различие +3,40 для старших детей попадает в одну из двух областей отвержения нуль-гипотезы. Если бы нуль-гипотеза была верна, только 0,05 части всех экспериментов дала бы различия, попадающие в ту или другую область отвержения. Для данного эксперимента вероятность (p) для каждой области отвержения будет 0,025 и для области «неотвержения» нуль-гипотезы — 0,95. Используя правило решения 0,05, мы говорим, что полученное различие *значимо*, поскольку мы можем отвергнуть нуль-гипотезу. Меньшее различие +0,93, как это видно, не попадает в область отвержения. Следовательно, статистическое решение в этом случае состоит в том, чтобы не отвергать нуль-гипотезу. Этот результат мог бы быть получен с вероятностью более высокой, чем 0,05, если бы нуль-гипотеза была верна.

Итак, мы можем заключить, что старшие дети плачут сильнее, если уходит мать. Что касается младшей группы, то мы не можем сделать ни этого, ни противоположного вывода (что они плачут сильнее с уходом ассистентки).

Заключение

Подтверждается противоположная гипотеза: плач сильнее, если уходит ассистентка				Не подтверждается ни одна из гипотез				Подтверждается экспериментальная гипотеза: плач сильнее, если уходит мать				
-5	-4	-3	-2	-1	0	+0,93	+1	+2	+3	+3,40	+4	+5

Статистическое решение:

Отвергнуть нуль-гипотезу ($p = 0,025$)				Не отвергать нуль-гипотезу ($p = 0,95$)				Отвергнуть нуль-гипотезу ($p = 0,025$)				
-5	-4	-3	-2	-1	0	+0,93	+1	+2	+3	+3,40	+4	+5

Интенсивность плача при уходе матери минус интенсивность плача при уходе ассистентки

Факторы, влияющие на величину требуемого различия

Предыдущая диаграмма продемонстрировала величину различия между средними, необходимую для отвержения нуль-гипотезы в условиях частного конкретного набора данных. При более надежных данных для отвержения нуль-гипотезы оказывается достаточным меньшее различие. Однако если для отвержения нуль-гипотезы используется более строгий критерий, то необходимо большее различие между средними. Эти два фактора иллюстрируются на рис. 6.1.

Статистическое решение

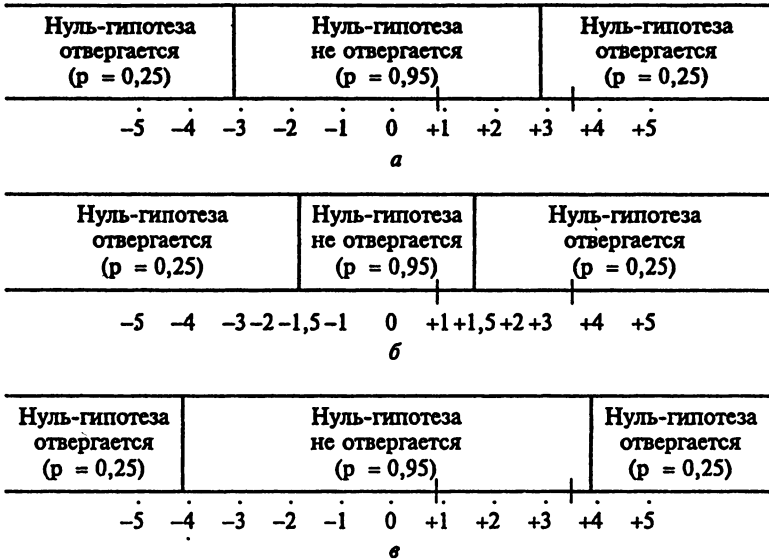


Рис. 6.1. Исходное статистическое решение (а), результат повышения надежности (б) и величина различия, необходимая для отвержения нуль-гипотезы (в). На числовых осях: интенсивность плача после ухода матери минус интенсивность плача после ухода ассистентки

Влияние надежности. Если бы увеличили число обследованных детей или стандартное отклонение внутри каждой группы оказалось меньше, то надежность увеличилась бы. Это означает, что в наборе бесконечно повторяемых экспериментов средние значения варьировали бы меньше. А в таком случае варьировали бы меньше и различия между средними. Вокруг каждой общей средней было бы более плотное распределение. В результате при справедливости нуль-гипотезы две линии, показывающие на диаграмме, куда попадает (с плюсом или с минусом) 0,025 части всех реальных средних, должны оказаться ближе к нулю.

На рис. 6.1 верхняя диаграмма (а) уже была представлена; на ней показаны результаты обследования 15 детей. На средней диаграмме (б) линиями показаны области отвержения нуль-гипотезы при условии, что число детей было увеличено до 60 или что стандартное отклонение для каждой пробы было уменьшено наполовину. И в том и в другом случае требуемое различие уменьшается вдвое, примерно до $\pm 1,5$. Видно, что различие 0,93, обнаруженное в младшей группе, тем не менее остается вне области отвержения нулевой гипотезы, что опять не дает возможность подтвердить гипотезу о том, что ребенок больше плачет при уходе матери. Тем

не менее если бы выборка была увеличена еще больше или стандартное отклонение оказалось меньше, различие 0,93 (если бы оно, конечно, было получено) оказалось бы значимым. Сложность в том, что при более надежных данных было бы менее вероятно получить различие такого уровня, если нулевая гипотеза была верна.

Влияние правила решения. До сих пор фигурировало только одно правило решения: нуль-гипотеза отвергается, если вероятность получения различия, при котором нуль-гипотеза верна, меньше 0,05. Вероятность, используемая для отвержения нуль-гипотезы, называется *альфа-уровнем*. Если применяется более строгий критерий, это означает, что нуль-гипотеза отвергается, если подтверждающая ее разница встречается в меньшей доле случаев. Наиболее часто это 0,01 (1 эксперимент из 100) вместо 0,05 (1 из 20).

Результат использования в правиле решения альфа-уровня 0,01 вместо прежнего 0,05 виден на нижней диаграмме (*а*) рис. 6.1. (Первоначальная выборка из 15 детей и первоначальные стандартные отклонения здесь сохранены.) При альфа-уровне 0,01 для отвержения нуль-гипотезы требуется различие, большее ± 4 . При этом, более строгом, критерии уже нельзя сделать заключение в пользу гипотезы о более сильном плаче с уходом матери.

Способ, каким в большинстве статей сообщается о статистически значимом различии, выглядит так: « $\bar{p} < 0,05$ » или « $\bar{p} < 0,01$ ». Это означает, что вероятность случаев, когда нуль-гипотеза верна, меньше 0,05 или меньше 0,01. Незначимое различие представляется как « $\bar{p} > 0,05$ » или « $\bar{p} > 0,01$ ».

ОТ РЕШЕНИЙ К ВЫВОДАМ

Люди, которые не любят статистику, считают, что все эти модные проверки статистической значимости не имеют особого смысла. Они не правы. С другой стороны, люди, поклоняющиеся статистике, считают, что за каждым статистическим решением автоматически следует экспериментальный вывод. Они тоже не правы. Мы попытаемся показать, что истина находится между этими крайностями. Наш девиз: используйте статистику, но сделайте так, чтобы она стала вашей служанкой, а не вашей госпожой.

Игнорирование проверки на значимость

Предположим, что Д. Флинер и Р. Кернс не стали бы рассматривать нуль-гипотезу. Они решили бы расценивать *любое* различие в пользу ухода матери как подтверждающее их экспериментальную гипотезу. В таком случае они приняли бы различие 0,93 для младшей группы тоже как значимое. Это было бы довольно рискованно. При дальнейшем знакомстве с их статьей можно уви-

деть, что для самой младшей исследовавшейся группы детей (в возрасте от трех до пяти месяцев) различие оказалось равным 1,78 с противоположным знаком, т. е. плач был больше при уходе *ассистентки*. Таким образом, выбирая тактику постоянного игнорирования нуль-гипотезы, экспериментаторы вынуждены были бы прийти к подтверждению противоположной гипотезы, которая в данном случае выглядит довольно бессмысленно.

Мы можем сразу увидеть, к каким последствиям приведет противоположная тактика, при которой нуль-гипотеза никогда не отвергается. Польза от проверки нуль-гипотезы всегда видна сразу. Если нуль-гипотеза окажется верной, экспериментальные выводы, не учитывающие ее, всегда будут ложными: будет считаться, что получила подтверждение либо экспериментальная гипотеза, либо противоположная ей гипотеза. Более того, когда нуль-гипотеза неверна и существует некоторое действительное различие в пользу того или иного условия, выводы тоже зачастую могут оказаться ложными. Предположим, что для младших детей бесконечный эксперимент показал небольшое преобладание плача при уходе матери. В любом конкретном эксперименте это различие вполне могло оказаться с противоположным знаком. Эта последняя ошибка будет делаться тем чаще, чем меньше надежность данных. Вообще говоря, в младшую группу входило только девять детей, так что надежность полученных по ним данных довольно сомнительна.

Итак, в научных экспериментах мы не можем обойтись без проверки на значимость.

Разновидности риска и типы ошибок

Поскольку нам известно, что реальные эксперименты не бывают ни идеальными, ни бесконечными, мы знаем, что некоторые из наших решений окажутся ошибочными независимо от применявшегося правила решения. Может быть, Д. Флинер и Р. Кернс не следовало отвергать нуль-гипотезу для старшей группы. Может быть, нуль-гипотеза была верна. Если бы они использовали 0,01 альфа-уровень, они не смогли бы отвергнуть нуль-гипотезу. И было бы прекрасно, если бы нуль-гипотеза и в самом деле была правильной. Ну а что, если нет? При обоих альфа-уровнях они рисковали бы — но противоположным образом.

Ошибки I типа. Первый риск состоит в возможности *ошибки I типа*: отвержение нуль-гипотезы, когда она верна. Если исследователь использует в правиле решения уровень 0,05, это означает, что он готов сделать такую ошибку не более чем в 5 % своих экспериментов. Когда он затем принимает отвержение нуль-гипотезы в качестве подтверждения экспериментальной гипотезы (например, плач более сильный при уходе матери), это показывает

его чрезмерный оптимизм. Ведь существует один шанс из двадцати, что такое доказательство ошибочно.

В любом эксперименте, направленном на проверку совершенно новой гипотезы, противоречащей общепринятому представлению, можно посоветовать быть более осторожным. Ломать научные традиции — вещь очень серьезная, и для этого нужно быть абсолютно уверенным в своих фактах. В таких случаях рекомендуется использовать более строгое правило решения, с 0,01 альфа-уровнем. Наука еще может выдержать 1 % результатов, которые ошибочно приняты за подтверждающие экспериментальную гипотезу, но 5 % — это уж слишком!

Ошибки II типа. Если мы настаиваем на 0,01 альфа-уровне (или даже более строгом уровне, таком, как 0,001), появляется новый риск: наше желание быть абсолютно уверенными может привести нас к ошибочному неутверждению нуль-гипотезы, когда она на самом деле неверна. Вполне естественно, что это называют *ошибкой II типа*. Если нуль-гипотеза ошибочна, верна должна быть какая-то другая гипотеза. Риск не отвергнуть нуль-гипотезу, когда верна другая гипотеза (например, определенное различие в плаче при уходе матери и ассистентки), может быть также выражен через вероятность, называемую *бета-уровнем*.

Для данной совокупности экспериментальных результатов уменьшение альфа-уровня означает увеличение бета-вероятности для любой ненулевой гипотезы. Использование очень строгого правила решения означает, что экспериментатор готов пойти на значительный риск, заключающийся в неутверждении нулевой гипотезы, когда верна какая-то другая гипотеза. Таким образом, при низком альфа-уровне экспериментатор будет часто ошибочно заключать, что результаты не подтверждают экспериментальную гипотезу. В отличие от альфа-уровня, для бета-уровня невозможно задать некоторое общее значение вероятности; она различается для каждой конкретной ненулевой гипотезы о различии между условиями. Так, если окажется верной гипотеза о большом различии между условиями (скажем, разница в интенсивности плача +5 ед.), вероятность не отвергнуть нуль-гипотезу (бета) будет низкой даже при использовании строгого альфа-уровня 0,01. С другой стороны, если действительная разность окажется небольшой (скажем, +1,0), вероятность ошибочного решения не отвергнуть нуль-гипотезу будет намного больше. Однако логика отношений сохраняется: при одних и тех же данных уменьшение альфа-уровня увеличивает бета-вероятность для всех статистических гипотез, отличных от нуль-гипотезы.

О статистической проверке экспериментальных результатов говорят как об имеющей *силу* в той степени, в какой бета-величина остается низкой для ненулевых гипотез. При хорошей силе выявляются реальные различия. Конечно, сила автоматически повы-

шается с использованием нестрогого правила решения (например, 0,10 альфа-уровня), но это увеличивает риск ошибки I типа. Существует два более удачных способа увеличения силы. Один состоит в увеличении надежности данных. Как мы видели на рис. 6.1, в, даже при небольшом различии между условиями оказывается возможным отвергнуть нуль-гипотезу либо путем увеличения числа испытуемых, либо путем уменьшения случайных вариаций. Другой способ состоит в использовании наиболее эффективных экспериментальных схем и проверок. Те и другие описаны в специальной литературе (см., например, Cohen L., 1977).

В предыдущей рубрике уже говорилось, что ошибки I типа следует избегать в том случае, когда отвержение нуль-гипотезы связано с отрицанием существующих идей или результатов предыдущих экспериментов. С другой стороны, если экспериментатор не обнаруживает значимых различий между условиями, которые обычно признаются эффективными, это его заключение должно основываться на использовании *высокого* (или нестрогого) альфа-уровня, чтобы уменьшить риск ошибки II типа. Почти любой полученный ранее правильный результат может быть «опровергнут» путем ошибочного неотвержения нуль-гипотезы: либо через использование ненадежных данных, либо через применение слишком строгого правила решения, либо (самый худший вариант) через то и другое вместе.

Теперь давайте рассмотрим, какие выводы должен сделать экспериментатор при отвержении нуль-гипотезы.

Если нуль-гипотеза не отвергается. Пожалуйста, заметьте: в отношении нуль-гипотезы принимаются только два статистических решения — отвергнуть ее или не отвергнуть. *Никогда* не бывает решения *принять* нуль-гипотезу. Все же для экспериментатора иногда полезно *заключить*, что независимая переменная не оказывает никакого влияния. Как видно из диаграммы на с. 184, неотвержение нуль-гипотезы привело бы к заключению, что не подтверждается ни экспериментальная, ни противоположная ей гипотеза. Например, для младшей группы детей небольшое различие в интенсивности плача не благоприятствует ни гипотезе о более сильном плаче при уходе матери, ни противоположной гипотезе о более сильном плаче при уходе ассистентки. Однако из подобных неподтверждений можно вывести различные заключения.

Во-первых, экспериментатор может сделать вывод, что он не знает, оказывает ли независимая переменная вообще какое-либо влияние на поведение. Этот вывод особенно подходит к случаю, когда надежность низка из-за небольшого количества испытуемых или из-за большей, чем ожидалось, варибельности поведения. Так, Д. Флинер и Р. Кернс могли бы решить продолжить эксперимент на новых детях, относящихся к той же младшей груп-

пе, и попытаться уменьшить случайные вариации, насколько это возможно.

Во-вторых, экспериментатор может заключить, что надежность была вполне удовлетворительной и что неотвержение нуль-гипотезы означает, что исследовавшиеся условия действительно не различаются. Это заключение может оказаться наиболее справедливым, особенно если более ранние эксперименты показали неэффективность независимой переменной.

Итак, статистическое решение снова состоит в неотвержении нуль-гипотезы. Однако обстоятельства эксперимента заставляют сделать вывод, что независимая переменная оказалась недействительной.

Валидность выводов

Вернемся к определению внутренней валидности, данному в главе 2: это степень уверенности, что заключение об экспериментальной гипотезе совпадает с выводом, который был бы получен в идеальном или бесконечном эксперименте. В предыдущих главах мы видели, как увеличивается внутренняя валидность с помощью процедур, повышающих надежность данных и уменьшающих смещение. Понятно, что заключения из экспериментальных данных не могут быть лучше самих данных. В настоящей главе мы видели, как разумное использование правил статистического решения ведет к обоснованным заключениям об экспериментальной гипотезе. Это — тоже способ увеличения внутренней валидности, поскольку заключение составляет наиболее важную часть эксперимента. Остановимся на этом более подробно.

Бесконечный эксперимент одновременно и определяет полную внутреннюю валидность, и обеспечивает основу для проверки нуль-гипотезы. Конечно, для этой последней цели существует особый вид бесконечного эксперимента. Он разбивается на отдельные конкретные эксперименты. Каждый из них такой же, как и реально проводимый эксперимент, но только в нем берутся другие испытуемые, выбранные из той же популяции (или другие пробы, если мы обращаемся к интраиндивидуальной схеме эксперимента).

При проверке нуль-гипотезы мы должны предположить, что из бесконечного эксперимента, безусловно, следует заключение, что экспериментальные условия не различаются. Естественно, общая средняя разность между условиями по всем этим экспериментам должна равняться нулю. Однако для каждого отдельного эксперимента это будет не так. Разности между средними будут лишь *распределяться* вокруг нуля, но при этом некоторые эксперименты будут благоприятствовать одному условию, другие — другому. А теперь нам нужно соотнести различие, полученное в

нашем собственном эксперименте, со всем *набором* различий, которые могли бы быть получены в этом типе бесконечного эксперимента.

Достоверные заключения, когда нуль-гипотеза верна. Если оказалось, что нуль-гипотеза верна, т. е. если общая средняя разность между условиями в бесконечном эксперименте равна нулю, мы хотели бы иметь возможность прийти к такому же заключению и в нашем эксперименте. Ведь мы не хотим сделать заключение в пользу гипотезы о различии между условиями, если много шансов за то, что на основе бесконечного эксперимента мы пришли бы к выводу об отсутствии различия. Таким образом, из надежных данных мы сделаем вывод, что экспериментальная гипотеза о различии условий неверна, если разность такой величины, как в нашем эксперименте, при верности нуль-гипотезы может появиться с вероятностью 0,05 или 0,01. Мы хотели бы быть уверенными в таком заключении при проверке новой экспериментальной гипотезы, особенно если она идет вразрез с общепринятым убеждением. Вот почему в таком случае альфа-уровень должен быть 0,01 или ниже. При уровне 0,05 будут неверны 5 % заключений. В бесконечном числе экспериментов 5 % дали бы результаты, которые привели бы к отвержению нуль-гипотезы в единичном эксперименте.

Достоверные выводы, когда нуль-гипотеза неверна. Если нуль-гипотеза оказывается ложной, т. е. если общая средняя разность между условиями такова, как предсказывает экспериментальная гипотеза, хотелось бы прийти к такому же выводу и в нашем эксперименте. Это не так важно при проверке какой-то новой идеи. Если она верна, то ее время все равно придет. Однако если различие между условиями ожидается на основе уже имеющихся знаний, мы хотим быть совершенно уверенными в своем выводе в пользу экспериментальной гипотезы.

Как мы уже видели, для этого требуются надежные данные. Кроме того, необходимо использовать менее строгое правило решения, например альфа-уровень 0,05. Если нуль-гипотеза окажется верной, мы, конечно, хотели бы сделать именно такое заключение. Однако мы готовы увеличить риск ошибочного отвержения нуль-гипотезы для того, чтобы уменьшить риск ошибочного вывода об отсутствии различия между условиями, т. е. в ситуации, когда бесконечный эксперимент такие различия бы показал.

НЕПРИЯТНЫЕ ПРОБЛЕМЫ, КОТОРЫЕ ОСТАЮТСЯ

Экспериментатор может благополучно пройти в своих заключениях через минное поле рисков относительно нуль-гипотезы и все же не внести никакого вклада в научное знание. В данном разделе рассматриваются три «трудные проблемы», которые угрожа-

ют внутренней валидности заключений, сделанных на основе правил решения, несмотря на то что данные надежны и проверка значимости осуществлена весьма разумно.

Бросающиеся и не бросающиеся в глаза результаты

Один исследователь предложил в шутку использовать для определения значимости результатов своего рода «интерокулярный травмирующий тест». Согласно этому «тесту» понять, что означают полученные результаты, можно только если вывод «бьет вас прямо в переносицу» (Дж. Берксон, цит. по: Edwards W., Lindman H., Savage L.J., 1963). Конечно, он имел в виду эксперимент, в котором ожидается *сильный эффект*, большое различие между условиями.

В отношении действия независимой переменной обычно имеется одно из двух ожиданий. Первое — что исследуемое поведение сильно зависит от переменной и что присутствие или отсутствие последней повлечет за собой большие различия. Так, мы могли бы ожидать, что человек с закрытыми глазами будет хорошо локализовать звук только в том случае, если звуки, достигающие обоих ушей, будут физически различаться. Если же его оценки будут успешны и без подобного различия, значит, независимая переменная не настолько важна, как предполагалось. Экспериментальная гипотеза, таким образом, не подтвердилась бы, даже если бы нулевая гипотеза была отвергнута.

Второе: иногда все, что имеет смысл ожидать от разных условий, — это небольшое, но стойкое различие. Рассмотрим эксперимент на перцептивную защиту, которая, как предполагают, обнаруживается в том, что нецензурные слова, предъявленные на короткое время, плохо опознаются. Различие в количестве опознанных нецензурных и нейтральных слов (какова бы ни была причина его) не будет предполагаться большим: ожидаемое влияние будет почти *забито* другими факторами. Например, некоторые нейтральные слова могут быть неточно восприняты или воспроизведены из-за смещения с другими словами. Некоторые из «нейтральных» слов также могут относиться к неприятным ситуациям в прошлом данного человека. Наконец, колебания внимания во время эксперимента могут влиять на опознание слов возможно даже больше, чем вариации в «пристойности». При таком ожидании незначительная, но постоянная тенденция к худшему опознанию неприличных слов была бы уже достаточной для демонстрации действия независимой переменной.

Эксперимент Д. Флинера и Р. Кернса был как раз таким, в котором можно было ожидать не бросающегося в глаза результата. Хотя ребенок, естественно, привязан к матери и поэтому очевидно, что с ее уходом он будет плакать сильнее, этот эффект вполне

может затушеваться другими факторами. Возможно, например, что некоторые матери из-за занятости на работе проводят со своими детьми сравнительно мало времени, и поэтому привязанность детей к ним не столь велика. Другие матери ведут себя так, что приходят и уходят на глазах у ребенка по многу раз в день, приучая его к своему отсутствию. Ассистенты по внешнему виду и манере поведения могут оказаться похожими на мать в большей или меньшей степени. Далее Д. Флинер и Р. Кернс (Fleener D. E., Cairns R. B., 1970, с. 218) обнаружили, что некоторые дети вообще плачут почти все время в течение эксперимента: «Пожалуй, наиболее очевидной чертой плача было его постоянство: если ребенок начал сильно плакать, он скорее всего продолжал это делать и дальше». Все эти факторы могли затруднить выявление различия между уходом матери и уходом ассистентки. Таким образом, данный эксперимент относится к тем, в которых нельзя ожидать отчетливого результата. Действие переменной — уход определенного лица — может быть выявлено лишь через статистическое различие в интенсивности плача. И здесь нельзя требовать такого различия, которое «бьет в глаза».

Количество не помогает

По мере значительного увеличения числа испытуемых с целью увеличения силы проверки на значимость происходит любопытная вещь. Оказывается, что любые два условия начинают давать статистически значимые различия (Bakan D., 1967). Если значимость не будет обнаружена для 20 испытуемых, она появится для 200, или 2 000, или 2 000 000. В этом нет никакой мистики. Любые два сравниваемых условия включают много факторов помимо тех, которые они должны представлять. Мы уже говорили, что человек не в силах контролировать все приходящие обстоятельства. Возможно, что буквы распознаются лучше, чем числа, только потому, что в каждой тысяче находится несколько испытуемых, которые реагируют отрицательно на числа в силу неудачного опыта в школьной математике. Возможно, что слова, произносимые каждые 8 с, воспринимаются хуже, чем произносимые в другом темпе, поскольку они немного чаще совпадают с обычными глотательными движениями, уменьшающими способность слышать.

Такие *дополнительные факторы* обнаруживают всегда свое присутствие при анализе больших массивов данных. Наш основной урок состоит в том, чтобы не позволять себе слишком зависеть от тестов на значимость в заключениях об исследуемых факторах. Ведь это только одно из средств. С особенной осторожностью следует подходить к случаям, которые требуют для выявления действия независимой переменной слишком большого количества данных. Гораздо большее впечатление производят статистически значи-

мые различия, полученные на относительно небольшом числе испытуемых или проб.

Справедлив ли вывод для всех испытуемых?

Д. Флинер и Р. Кернс вполне могли бы получить статистически значимые результаты для детей старшей группы, если бы только девять из пятнадцати детей плакали сильнее при уходе матери. Но что в таком случае сказали бы об остальных шести?

Проводя эксперимент, мы рассчитываем, что исследуемый психологический фактор действует эффективно на каждого испытуемого. Читая статьи в журналах, вы обнаружите, что это положение принимается почти всеми. Негласное допущение состоит в том, что если независимая переменная эффективна, она влияет на всех индивидов, подпадающих под данную гипотезу. Если последняя в самом деле верна для некоторых, она верна для всех. Поэтому отсутствие в шести случаях более сильного плача при уходе матери приписывается действию дополнительных факторов, о которых упоминалось выше: прошлому опыту в общении с матерью, предшествующему плачу в эксперименте и т. д.

Однако это не всегда так. Рассмотрим другой эксперимент. Предположим, обнаружено, что испытуемые лучше узнают слова, если они проговаривают их вслух при первом предъявлении. По крайней мере, 13 из 20 испытуемых показали именно этот результат. Тогда причиной, по которой тот же результат не наблюдался у других семи испытуемых, могут быть случайные вариации, связанные, например, с ассоциациями, которые вызывали некоторые из использовавшихся слов. Однако могло оказаться, что одни испытуемые помогали себе немедленным проговариванием, а другие — нет. Проводя индивидуальные эксперименты, можно было бы выявить реальные причины индивидуальных различий каждого испытуемого. Еще лучше, если нам удастся найти некоторые признаки, разделяющие лиц, которым проговаривание помогает и которым — нет. Вполне возможно, например, что проговаривание не помогает лицам с очень хорошей визуализацией. Но мы забегаем вперед, поднимая вопрос, который будет рассматриваться в главе 8, где мы будем свидетелями рождения еще одного или даже двух новых способов контроля.

ДРУГИЕ АСПЕКТЫ ВАЛИДНОСТИ

В этой главе мы говорили о заключениях, основанных на статистических решениях. Однако мы не должны упускать из виду, что существуют и другие важные аспекты валидности. Слишком часто о валидности заключения судят лишь учитывая надежность,

которой мы занимались в данной главе. Но мы, конечно, знаем, что валидность этим не исчерпывается.

Внешняя валидность

Напомним прежде всего, что эксперимент может не обладать *внешней* валидностью по ряду причин. Эксперимент может не быть экспериментом *полного соответствия* из-за несоответствующего уровня другой переменной (например, если бы Джек Моцарт запоминал вальсы вместо сонат при сравнительной оценке методов заучивания). В экспериментах, которые улучшают реальный мир (например, с ночными посадками самолетов), мы хотим также быть уверенными, что искусственные независимая и зависимая переменные представляют те ситуации, к которым затем будут прилагаться результаты. В экспериментах, проводимых на выборке испытуемых (с информацией о стоимости товаров), мы рассматривали вопрос, насколько хорошо данная выборка представляет популяцию покупателей универсама. Анализируя конкретные способы представления экспериментальной ситуации различных схем социальной структуры (авторитарной, демократической или анархистской), мы больше всего сомневались относительно *операциональной* валидности этих ситуаций. Все наши статистические решения имеют отношение к внешней валидности. Тем не менее экспериментальные выводы не могут быть до конца валидными, если они наряду с внутренней валидностью не будут обладать внешней валидностью.

Систематическое смещение

Напомним далее о рассматривавшейся в главе 2 необходимости избегать систематического процедурного смещения (такого как эффекты последовательности) и в главе 5 — сопутствующего смещения. Мы видели, что Флинер и Кернс (Fleener D. E., Cairns R. B., 1970) сделали вывод о более сильном плаче при уходе матери, чем при уходе ассистентки. Однако, как показал Дж. Коэн (Cohen J., 1977), здесь имело место систематическое сопутствующее смещение. Ведь остававшийся человек был разным, когда уходила мать и когда уходила ассистентка. Тем самым ложится тень на внутреннюю валидность независимо от величины различия между двумя условиями. Отвержение нулевой гипотезы ничего не говорит о систематическом смещении. Экспериментальные заключения могут иметь внутреннюю валидность лишь в той степени, в какой удалось избежать систематического смещения.

Когда мы судим о валидности экспериментальных заключений на основе статистических решений (хорошо или плохо они были использованы или они вообще были по ошибке проигнорирова-

ны), мы должны предполагать, что все ранее рассмотренные аспекты валидности удовлетворительно реализованы. Вам следует об этом хорошо помнить или еще лучше — «зарубить себе на носу».

КРАТКОЕ ИЗЛОЖЕНИЕ

Большие различия в действии разных условий независимой переменной приводят экспериментатора к выводу о подтверждении экспериментальной гипотезы. Меньшие различия интерпретируются как случайный результат. Основанием для таких различных выводов является статистическая значимость. Более конкретно это означает, что если бы в идеальном или бесконечном эксперименте различие отсутствовало, то было бы маловероятно получить в конкретном эксперименте большое различие, не так невероятно — меньшее различие.

В научных экспериментах — в отличие от тех, где существуют только два практических исхода, — возможны три заключения из экспериментальных данных. В дополнение к подтверждению экспериментальной или противоположной ей гипотезы возможно заключение о не подтверждении ни одной из них. Какое именно из этих трех заключений будет сделано, зависит от статистического решения относительно нуль-гипотезы.

Если бы был проведен бесконечный эксперимент и нуль-гипотеза оказалась верной, то среднее различие между условиями было бы равно нулю. Однако в отдельных конкретных экспериментах различия могут быть как в пользу одного условия, так и в пользу другого. Если различие настолько велико, что очень редко могло бы быть получено в бесконечном эксперименте, нуль-гипотеза отвергается. Однако если вероятность появления различия, подобного полученному, достаточно высока, нуль-гипотеза не отвергается. Когда нуль-гипотеза отвергается, делается вывод о подтверждении экспериментальной гипотезы (или противоположной ей гипотезы, если различие оказалось с обратным знаком). Когда нуль-гипотеза не отвергается, ни экспериментальная, ни противоположная ей гипотезы не подтверждаются. Это последнее заключение может означать одно из двух. Если данные ненадежны, заключение будет состоять в том, что действие независимой переменной просто не удалось выявить. При надежных данных экспериментатор может быть уверен, что условия не оказывают различного действия.

Величина различия между условиями, необходимая для отвержения нуль-гипотезы, определяется двумя факторами. Первое — это надежность. Чем больше надежность, тем меньше различие, допускающее отвержение. Второй фактор — вероятность того, что экспериментатор рискнет ошибочно отвергнуть нуль-гипотезу, когда она верна. Он называется *альфа-уровнем* правила его решения. Ошибка, которая будет увеличиваться с возрастанием этого риска, называется *ошибкой I типа*. Так, риск ошибки I типа в пять раз выше при альфа-уровне 0,05 по сравнению с альфа-уровнем 0,01.

Однако при уменьшении альфа-уровня увеличивается риск противоположной ошибки. Это риск не отвергнуть нуль-гипотезу, когда верна некоторая другая гипотеза (и, конечно, нуль-гипотеза ошибочна). Это называют *ошибкой II типа*. Для любого конкретного набора данных эта вероятность (называемая бетой) увеличивается с уменьшением альфа-уровня. Но уве-

личивая надежность эксперимента, можно найти приемлемую величину бета даже при строгом альфа-уровне. Говорят, что статистическая проверка имеет силу в той мере, в какой низка вероятность бета и в которой может быть выявлено истинное различие.

Использование строгого альфа-уровня (например, 0,01) рекомендуется в тех случаях, когда различие между условиями должно подтвердить новую гипотезу, противоречащую общепринятому мнению. Эта строгость нужна для того, чтобы не засорять науку слишком большим числом артефактов. Пять ложных утверждений из двадцати — это слишком тяжелое бремя для науки. С другой стороны, если результаты показали влияние независимой переменной, его нельзя сбрасывать со счетов только потому, что различие не достигло уровня значимости 0,01.

Назначение проверок на значимость — повышение внутренней валидности. Ведь внутренняя валидность и проверка нуль-гипотезы могут быть описаны через бесконечный эксперимент. В бесконечном эксперименте, состоящем из множества отдельных экспериментов (таких, какие проводятся реально), общее среднее различие между условиями будет равно нулю, если верна нуль-гипотеза. Однако различия, обнаруживаемые в отдельных экспериментах, не будут равны нулю, а лишь только распределятся вокруг нуля. Экспериментатор может выяснять это распределение. Он соотнесет полученное различие с его вариабельностью, но не будет делать вывода о различии только на том основании, что много отдельных экспериментов дает достаточно большое различие.

Если же верна нуль-гипотеза, экспериментатор также хотел бы обосновать и этот вывод. Но даже чтобы иметь возможность сделать вывод о правильности какой-то другой гипотезы, экспериментатор вынужден идти на некоторый риск. Экспериментатор хочет иметь заключение о верности экспериментальной гипотезы с такой степенью обоснованности, как если бы ожидаемое различие было получено в бесконечном эксперименте. Положение, которое он в конце концов занимает между ошибками I и II типа, отражает его оценку относительной валидности обоих типов обоснованности.

На пути к окончательным выводам остаются три трудные проблемы. Первая состоит в том, что только одного значимого различия недостаточно, если ожидается сильное влияние независимой переменной. Статистическая проверка наиболее пригодна в тех случаях, когда действие исследуемого фактора «зашумлено» другими случайными факторами. Вторая проблема заключается в том, что использование слишком большого числа испытуемых обнаруживает действие определенных дополнительных факторов. Третья проблема касается универсальности результатов. Можно ли отнести выводы ко всей соответствующей популяции, если они справедливы даже не для всех исследовавшихся испытуемых (причем не только по причине случайных изменений)? Наконец, было показано, что мы не можем принять экспериментальные выводы только на основе постоянных и достаточно сильных выявленных различий между условиями. Эксперименту будет недоставать внешней валидности, если он не будет удовлетворять хотя бы одному из целого ряда условий. Более того, он не будет обладать даже внутренней валидностью, если не организовать достаточный контроль за систематическим смещением.

Контрольные вопросы и задания

1. Почему Флинер и Кернс заключили, что старшие дети больше плачут при уходе матери, чем при уходе ассистентки, а у младших детей такого различия нет?
2. Что такое нуль-гипотеза?
3. Почему в эксперименте Флинера и Кернса возможно третье заключение, в то время как в эксперименте Йоки по предпочтению сорта томатного сока только два?
4. Что показывает диаграмма, иллюстрирующая: различие между средними для каждого условия, статистическое решение и заключение об экспериментальной гипотезе?
5. Как влияет уменьшение надежности на величину различия между средними, требующуюся для отвержения нуль-гипотезы?
6. Как влияет альфа-уровень в правиле решения на величину различия между средними, требующуюся для отвержения нуль-гипотезы?
7. Соотнесите альфа-уровень с риском ошибок I и II типов.
8. Когда особенно важно избегать ошибки I типа?
9. Опишите три фактора, влияющих на вероятность бета. Что это означает в отношении риска ошибки II типа?
10. При каких условиях экспериментатор может заключить, что независимая переменная не оказывает действия?
11. Почему говорят, что разумное использование правила статистического решения способствует внутренней валидности?
12. Может ли быть в эксперименте слишком много испытуемых?
13. Если в эксперименте получены надежные данные и высоко значимые различия между условиями, обеспечивает ли это полностью валидность вывода?

СТАТИСТИЧЕСКОЕ ПРИЛОЖЕНИЕ: t-КРИТЕРИЙ

В данном приложении будет описан метод нахождения величины различия между средними, необходимой для отвержения нуль-гипотезы. Фактически мы будем подробно объяснять диаграммы, представленные на рис. 6.1.

Выборочное распределение

Давайте еще раз предположим, что данные по времени реакции, представленные в предыдущих статистических приложениях, получены в межгрупповом эксперименте. Мы, таким образом, имеем среднее время реакции для каждого из 17 испытуемых, которым предъявлялось условие А (свет), и среднее время реакции для каждого из 17 испытуемых, которым предъявлялось условие Б (тон). Более того, известно общее среднее для испытуемых в условии А (185 мс) и общее среднее в условии Б (162 мс). Наконец, мы знаем разницу между этими двумя средними, $M_A - M_B$, равную +23 мс.

Если бы исследовались две другие группы испытуемых, отобранные тем же способом, то, конечно, не следовало бы ожидать, что $M_A - M_B$ будет в точности равной 23 мс. Нельзя было бы ожидать точно такой же разницы +23 мс и в третьем эксперименте. Напротив, мы предполагаем, что это значение $M_A - M_B$ будет несистематически варьировать от эксперимента к эксперименту.

Допустим, что путем повторения этого эксперимента был реализован бесконечный эксперимент, при котором каждое условие предъявлялось 17 испытуемым бесконечное число раз. Предположим далее, что нуль-гипотеза верна. Тогда различие между общими средними — которое есть *параметр* — должно равняться нулю. Другими словами, $\bar{M}_A - \bar{M}_B = 0$. Однако величина статистики $M_A - M_B$ должна варьировать от эксперимента к эксперименту.

Распределение величин $M_A - M_B$ для серии последовательных экспериментов может быть представлено так, как было описано ранее. Обозначим величину +23, которая была получена в реальном эксперименте, номером 1; предположим, что мы провели второй такой же эксперимент и получили величину -4, обозначим ее номером 2; величину, полученную в третьем эксперименте (допустим, 0), — номером 3 и т.д. Таким образом, результаты девяти экспериментов, в случае $M_A - M_B = 0$, могли бы выглядеть следующим образом (рис. 6.2).

К счастью, можно вывести, как это распределение выглядело бы для бесконечного числа экспериментов. Мы можем реально изобразить ожидаемое распределение величин $M_A - M_B$. Более того, мы можем оценить стандартное отклонение, которое имело бы это распределение. Такой тип теоретически выведенного распределения называют выборочным распределением. Описываемое здесь распределение является выборочным распределением разностей между средними (имеются также выборочные распределения для средних, для стандартных отклонений и т.д.).

Приводим выборочное распределение для нашего эксперимента по времени реакции с предположением, что нуль-гипотеза $\bar{M}_A - \bar{M}_B = 0$ верна.

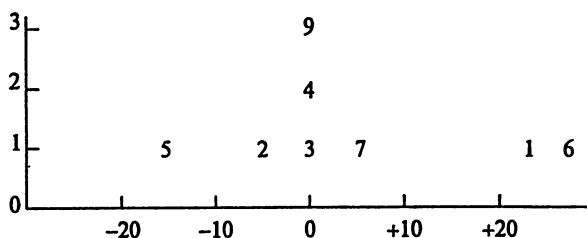


Рис. 6.2. Ось абсцисс — $M_A - M_B$; ось ординат — частота

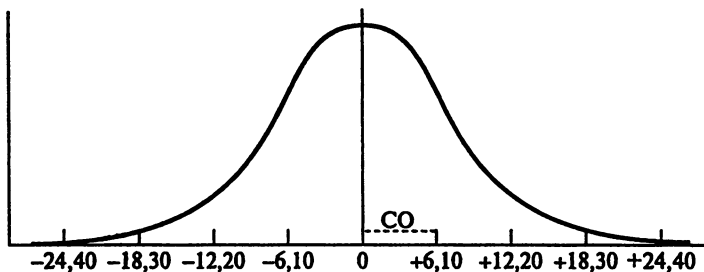


Рис. 6.3. Ось абсцисс — $M_A - M_B$; ось ординат — относительная частота

Заметьте, что стандартное отклонение (CO) равно 6,1. Поэтому разность $M_A - M_B = + 12,20$, полученная в каком-то эксперименте, находится на расстоянии двух стандартных отклонений выше предполагаемой величины $M_A - M_B = 0$, а разность $M_A - M_B$, равная $- 18,30$, — на три стандартных отклонения ниже предполагаемого нуля и т. д. (рис. 6.3).

Стандартная ошибка

До сих пор не объяснялось, как было вычислено стандартное отклонение этого гипотетического выборочного распределения. Вот эта формула:

$$S_{M_A - M_B} = \sqrt{\frac{S_A^2}{N_A} + \frac{S_B^2}{N_B}} \quad (6.1)$$

$S_{M_A - M_B}$ называется *стандартной ошибкой разности между средними*. Использование термина *стандартная ошибка* вместо *стандартного отклонения* показывает, что мы вывели стандартное отклонение, а не пришли к нему через (невозможные) бесконечные вычисления. Заметьте, что здесь используется S , а не σ . Это потому, что популяционный параметр $\bar{\sigma}_{M_A - M_B}$ оценивается на основе выборочных статистик.

Для вычисления в формулу просто подставляют величины S_A^2 и S_B^2 , полученные нами в предыдущих статистических приложениях. Так,

$$S_{M_A - M_B} = \sqrt{\frac{363}{17} + \frac{169}{17}} = \sqrt{21,35 + 15,82} = \sqrt{37,17} = 6,10 \text{ мс.}$$

Вы можете видеть, что формула применима также и в том случае, когда N_A и N_B различны, т.е. когда число испытуемых (или проб в интраиндивидуальном эксперименте) различно для двух условий.

Определение величины t

Следующий шаг состоит в том, чтобы найти, на сколько единиц стандартной ошибки отстоит полученная нами разность $M_A - M_B$ от нуля, представляющего среднюю нуль-гипотезу. Поскольку полученная нами разность равнялась +23, а стандартная ошибка $M_A - M_B = 6,10$, то очевидно, что наша разность находится на расстоянии 3,77 единицы стандартной ошибки выше нуля. Единицы стандартной ошибки называют *t-единицами*. Выражение полученной разности в единицах стандартной ошибки называют *нахождением величины t для данной разности*. Это может быть выражено следующей формулой:

$$t = \frac{(M_A - M_B) - 0}{S_{M_A - M_B}}. \quad (6.2)$$

Подставляя значения из нашего эксперимента по измерению времени реакции, мы имеем

$$t = \frac{(185 - 162) - 0}{6,10} = \frac{23}{6,10} = 3,77.$$

Заметьте, что нуль в числителе при числовых операциях можно опустить. Он служит для того, чтобы напомнить нам, что мы проверяем нуль-гипотезу:

$$\bar{M}_A - \bar{M}_B = 0.$$

Отвержение или неотвержение нуль-гипотезы

Теперь мы готовы (наконец!) описать, как были получены диаграммы на рис. 6.1, показывающие величину разности между средними, необходимую для отвержения нуль-гипотезы. Давайте перерисуем выборочное распределение разностей.

Вы найдете в Статистической таблице 2 в конце данного приложения величину t , достаточную для отвержения нуль-гипотезы. Она дана и для альфа-уровня 0,05, и для альфа-уровня 0,01. Эти *критические величины* зависят от величины N для каждого условия, или, иначе, от числа степеней свободы, $N-1$ для каждого среднего. (Если вы имеете данное среднее, скажем, 179 мс для 17 испытуемых, эта величина *могла бы* быть получена путем *свободного приписывания* любых величин 16 испытуемым. Однако затем вам *придется* приписать 17-му испытуемому совершенно определенную величину, чтобы получить заданное среднее.) Таким образом, поскольку было 17 испытуемых для каждого условия, имели место $16 + 16 = 32$ степени свободы (или *df*).

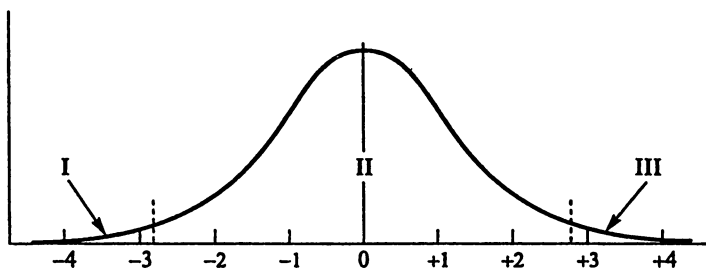


Рис. 6.4. Ось абсцисс: первая — значения t -критерия; вторая $M_A - M_B$. Ось ординат — относительная частота. I, III — $p = 0,005$, нуль-гипотеза отвергается; II — $p = 0,99$, нуль-гипотеза не отвергается

В таблице нет значений именно для $32df$ (но величина для $30df$ вполне годится, так как разница между величинами t для 30 и $35df$ очень мала). Чтобы отвергнуть нуль-гипотезу для $0,05$ альфа-уровня, требуется t , равное $2,04$, для альфа-уровня $0,01$ — t , равное $2,75$. Величина t , равная в нашем эксперименте $3,77$, показывает, что полученная разность $+23$ попадает в область отвержения, даже если использовать альфа-уровень $0,01$.

Вероятности показаны так же, как на рис. 6.1, в. Исходя из этого, наше статистическое решение будет заключаться в отвержении нуль-гипотезы.

Распределение, представленное в величинах t , является выборочным распределением t . Точная форма t -распределения будет разной в зависимости от числа степеней свободы в эксперименте. Вот почему вы должны находить критические величины, чтобы определить, является ли полученное вами различие значимым.

Нуль-гипотеза и ω^2

Из данного статистического приложения видно, что в эксперименте по измерению времени реакции независимая переменная оказывала сильное влияние: $est \omega^2 = 0,28$. Ясно, что получить такую разность между условиями в высшей степени невероятно, если верна нуль-гипотеза. Но не смешивайте эти два понятия — силу действия и статистическую значимость. При очень надежных данных даже небольшая разность между средними позволит отвергнуть нуль-гипотезу. В то же время разность может оказаться статистически значимой даже при слабом действии независимой переменной.

Задача: Вычислите t и проверьте нуль-гипотезу при альфа-уровне $0,01$ для эксперимента по измерению времени реакции выбора между двумя вспышками света (условие В) и выбора между двумя тонами (условие Г).

Условие В (вспышки)				Условие Г (тонны)			
Испыт.	ВР	Испыт.	ВР	Испыт.	ВР	Испыт.	ВР
1	304	10	275	1	272	10	261
2	268	11	268	2	264	11	250
3	272	12	254	3	256	12	228
4	262	13	245	4	269	13	257
5	283	14	253	5	285	14	214
6	265	15	235	6	247	15	242
7	286	16	260	7	250	16	222
8	257	17	246	8	245	17	234
9	279			9	251		

Ответ: $M_B = 265$; $M_G = 250$; $S_B^2 = 292$; $S_G^2 = 337$; $t = 2,47$.

Нуль-гипотеза может быть отвергнута при альфа-уровне 0,05, но не при альфа-уровне 0,01.

Статистическая таблица 2

Величина t -критерия, отвергающая нуль-гипотезу

Степень свободы df	0,05	0,01	Степень свободы df	0,05	0,01
1	12,71	63,66	14	2,14	2,98
2	4,30	9,92	15	2,13	2,95
3	3,18	5,84	16	2,12	2,92
4	2,78	4,60	17	2,11	2,90
5	2,57	4,03	18	2,10	2,88
6	2,45	3,71	19	2,09	2,86
7	2,36	3,50	20	2,09	2,84
8	2,31	3,36	21	2,08	2,83
9	2,26	3,25	22	2,07	2,82
10	2,23	3,17	23	2,07	2,81
11	2,20	3,11	24	2,06	2,80
12	2,18	3,06	25	2,06	2,79
13	2,16	3,01	26	2,06	2,78

Окончание статистической табл. 2

Степень свободы df	0,05	0,01	Степень свободы df	0,05	0,01
27	2,05	2,77	80	1,99	2,64
28	2,05	2,76	90	1,99	2,63
29	2,04	2,76	100	1,98	2,63
30	2,04	2,75	120	1,98	2,62
35	2,03	2,72	150	1,98	2,61
40	2,02	2,71	200	1,97	2,60
45	2,02	2,69	300	1,97	2,59
50	2,01	2,68	400	1,97	2,59
60	2,00	2,66	500	1,96	2,59
70	2,00	2,65	1000	1,96	2,58
			∞	1,96	2,58

Статистическая таблица 2 взята из таблицы IV в работе Фишера и Ятса «Статистические таблицы для биологических, сельскохозяйственных и медицинских исследований».

Глава 7

МНОГОУРОВНЕВЫЕ ЭКСПЕРИМЕНТЫ¹

Предположим, что три исследователя А, Б и В поставили перед собой вопрос: как наилучшим образом использовать фиксированный отрезок времени для заучивания списка объектов (символов)? А решил, что лучше всего подолгу останавливаться на каждом элементе списка, и, таким образом, устраивая длительные интервалы между каждым предыдущим и последующим элементами, пройти весь список всего лишь несколько раз. Б предположил прямо противоположное: лучше всего делать короткие интервалы между элементами и быстро пройти список много раз подряд. В решил, что характер распределения отпущенного времени не играет существенной роли; главное — это общее время заучивания.

Теперь дополним наши предположения некоторыми деталями. Пусть А, Б и В независимо друг от друга решили провести эксперимент по запоминанию списка из 16 элементов, представляющих собой сочетание двузначных чисел с группами из трех букв (называемых *триграммами*). Испытуемый должен усвоить, что при показе, например, триграммы ВАР нужно отвечать «27», а при показе КОМ — «84» и т.д. для каждого из 16 сочетаний. Общая длительность эксперимента — 320 с. Процедура опыта состоит в предъявлении списка — элемент за элементом — в соответствии с выбранным планом распределения времени заучивания и затем, по истечении этого времени (320 с) — в проверке того, сколько пунктов запомнил испытуемый.

Вот что делал А и вот что он обнаружил. Он использовал два условия: а) 1 с между элементами (20 предъявлений списка) и б) 4 с между элементами (5 предъявлений списка). Испытуемые с интервалом 1 с в среднем дали 7 правильных ответов, а испытуемые с интервалом 4 с — 13 правильных ответов. Таким образом, с достаточной уверенностью можно сказать, что его гипотеза о большей эффективности *больших* интервалов времени подтвердилась.

Б также использовал интервал 4 с (условие а), но сравнивал его с интервалом 20 с (условие б), при котором список предъявлялся только один раз. Он также получил около 13 правильных ответов для интервала 4 с, но меньше 10 правильных ответов для

¹ Перевод Ч. А. Измайлова.

интервала 20 с. В результате заключил, что подтверждается его гипотеза о большей эффективности более *коротких* интервалов.

В использовал интервалы 3 с (а) и 10 с (б). При обоих условиях он получил в среднем около 12 правильных ответов. И снова была подтверждена исходная гипотеза, на этот раз о том, что характер распределения экспериментального времени не играет роли.

Наконец, давайте представим, что вы редактор журнала, в который А, Б и В одновременно подали свои статьи. Вы оказались перед необходимостью представить вместе три ряда результатов и показать отношение между независимой переменной (интервал между элементами) и зависимой переменной (среднее число правильных ответов — заученных элементов списка), как это показано на рис. 7.1.

В действительности вы, конечно, не сможете соединить результаты именно таким образом из-за многих различий, имевших место во всех трех исследованиях. На рис. 7.1 представлены данные действительных экспериментов, однако не таких разрозненных, как мы это только что представили. Это — части одного и того же обширного исследования влияния темпа предъявления на заучивание, проведенного Р. Калфи и Р. Андерсон (Calfee R. C., Anderson R., 1971). Значения интервалов в «исследованиях» А, Б и В показаны на графике. Каждое из них — это эксперимент с двумя условиями (иногда называемый *бивалентным*). Эксперимент же Калфи и Андерсон — это многоуровневый эксперимент (иногда называемый *мультивалентным*); в нем было использовано много *уровней* независимой

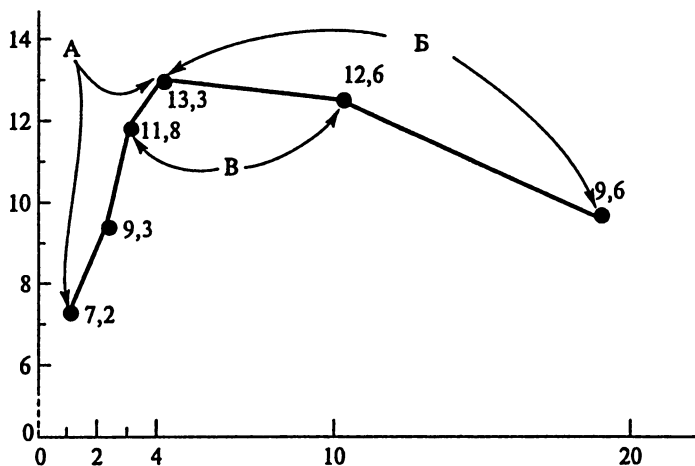


Рис. 7.1. Воображаемые двухуровневые эксперименты А, Б и В на примере результатов исследования Р. Калфи и Р. Андерсон (1971). Ось абсцисс — временные интервалы между предъявляемыми знаками; ось ординат — среднее количество правильных ответов

переменной, а именно шесть временных интервалов между элементами: 1, 2, 3, 4, 10 и 20 с.

Как вы могли убедиться, выход, когда это возможно, за рамки двухуровневого эксперимента весьма плодотворен. Ведь, как оказалось, мы узнали очень немного из каждого отдельного эксперимента А, Б и В. И только применение нескольких уровней независимой переменной позволило установить реальные отношения с зависимой переменной. В этой главе мы рассмотрим эти преимущества более детально. После этого мы разберем экспериментальные схемы, которые могут быть использованы в многоуровневом эксперименте. Мы увидим, что наши прежние схемы, такие как межгрупповые и индивидуальные, применимы и здесь. Однако наиболее употребима здесь новая основная схема *кроссиндивидуального* контроля. Затем мы перейдем к проблемам внутренней валидности в многоуровневых экспериментах.

Из этой главы вы должны вынести новое представление об эксперименте как средстве установления связи между двумя *непрерывными* переменными, а именно о том, что происходит с зависимой переменной по мере того, как *шаг за шагом* меняется независимая переменная. Посмотрите в литературе статьи с двух- или трехуровневыми экспериментами и попытайтесь решить, можно ли превратить их в многоуровневые эксперименты. Попробуйте представить себе, как выглядели бы в этом случае их результаты. Мы надеемся, что вы поймете все «угрозы» внутренней валидности, которые таятся в типичных схемах многоуровневых экспериментов, и, читая статьи, сможете оценить, в какой мере их понимал и сам автор.

Вопросы, на которые вы должны будете ответить в конце главы, относятся к следующим темам.

1. Контрольные функции многоуровневого эксперимента.
2. Более тонкие экспериментальные гипотезы, которые можно проверять только в многоуровневых экспериментах.
3. Различные экспериментальные схемы, использующие межиндивидуальное позиционное уравнивание.
4. Вопросы, касающиеся внутренней валидности при выборе экспериментальной схемы многоуровневого эксперимента.

МНОГОУРОВНЕВЫЙ ЭКСПЕРИМЕНТ КАК КОНТРОЛЬНЫЙ

Сначала мы рассмотрим, что дает многоуровневый эксперимент, когда экспериментальная гипотеза может быть проверена и с использованием только двух условий. Примером может служить эксперимент по рабочей этике, обсуждавшийся в главе 5, где было показано, что девочки-индианки предпочитают активные усилия безделью. Сравнивались два условия: 1) получение шарика без нажатия на рычаг и 2) получение шарика после 10-кратного на-

жатия на рычаг. Можно было бы использовать различное число нажатий, например 2, 5, 10, 50 за каждый шарик. Улучшился ли бы эксперимент при введении большого числа уровней настолько, чтобы его стоило проводить? (Хотя, учитывая гипотезу авторов, ответ заранее ясен: многоуровневый эксперимент должен бы предпочитаться исследователями, поскольку он требует больше активных усилий с их стороны.)

Меньше шансов пропустить эффект

Чтобы увидеть преимущество пятиуровневого эксперимента, следует ответить на вопрос, каким образом экспериментаторы узнали, что для одного из двух экспериментальных условий необходимо использовать именно десять нажатий на рычаг. Возможно, это слишком небольшая работа, чтобы влиять на поведение, а может быть, наоборот, слишком большая. Очевидно, исследователи Д. Сингх и В. Квери (Singh D., Query W. T., 1971) просто решили, что одного нажатия за шарик будет слишком мало. Точно так же нельзя предлагать очень много нажатий за один шарик, ибо в этом случае девочки будут, безусловно, стремиться получать шарики просто так, и лишь где-то между этими крайними значениями будет находиться число нажатий (уровень), при котором активное и пассивное условия окажутся одинаково предпочитаемыми. Если бы Сингх и Квери остановили свой выбор на уровне равного предпочтения, они оказались бы в том же положении, что и воображаемый экспериментатор В, чьи результаты показывают, что независимая переменная не оказывает никакого действия. Прделанный сейчас анализ показывает, что гипотеза, проверявшаяся в исследовании Д. Сингха и В. Квери, в действительности была *количественной* гипотезой, соотносящей количество нажатий на рычаг с величиной предпочтения активных усилий пассивной награде. *Хорошо* проверить такую гипотезу можно только при условии, если независимая переменная будет непрерывна. Это, конечно, невозможно, поскольку тогда потребуется *бесконечное* число уровней с бесконечно малыми различиями. И все же при использовании даже пяти уровней можно приблизиться к выявлению полного отношения между независимой и зависимой переменными. По мере уменьшения числа уровней увеличивается опасность *ошибочного представления* этого отношения. Поэтому можно сказать, что *внутренняя валидность* больше, когда такая гипотеза проверяется при пяти уровнях по сравнению с двумя уровнями независимой переменной. Эта угроза внутренней валидности вытекает из *неполноты* независимой переменной. Угрозы, описанные выше, проистекали либо из ненадежности данных, либо из процедурного или сопутствующего смещения с другими переменными (см. гл. 5,

с. 147). Воображаемые эксперименты А, Б и В служат драматическим примером того, как ложно может быть представлено отношение между независимой и зависимой переменными из-за использования небольшого числа уровней. Кроме того, эксперимент с двумя уровнями сталкивается еще с одной проблемой, касающейся сопутствующего смещения. К показу этого мы сейчас и перейдем.

Лучший контроль над сопутствующим смещением

Доказательство действия независимой переменной неубедительно, если явно возможно сопутствующее смещение, т.е. если активный уровень независимой переменной связан с активным уровнем другой переменной. Контрольные же условия, подобные описанным в главе 5, могут оказаться менее эффективными, чем это кажется.

Предположим, что на студентах колледжа проводится эксперимент для выяснения того, усиливает ли кофеин реактивность нервной системы по отношению к стимуляции. Измеряемым параметром служит время реакции. Если бы использовались только два условия, существовала бы ясная опасность обсуждаемого смещения. Пусть, например, в качестве активного уровня выбрано 3 мг лекарства на каждый килограмм массы испытуемого. (На основании других исследований можно судить, что этот выбор удачный.) Как обсуждалось в главе 5, контрольным условием должно быть нейтральное вещество — *плацебо*. Если кофеин дается в виде таблетки, плацебо тоже должно быть таблеткой, но не содержащей активного вещества. И все же плацебо может не обеспечить контроль за *осведомленностью* испытуемого относительно предъявляемого условия. Здесь не учитывается тот факт, что испытуемый (особенно хороший студент колледжа) может определить, является ли таблетка кофеином или нейтральным веществом. В первом случае появится слабое дрожание пальцев, учащенное дыхание, разольется тепло по лицу и т.д. Осознание этого вполне может оказать косвенное влияние на время реакции. Испытуемый может ожидать, что его ответы станут более живыми, напряжение его повысится, что действительно приведет к сокращению времени реакции. Такое действие кофеина не будет иметь ничего общего с гипотезой экспериментатора о непосредственном действии кофеина на нервную систему. Здесь произойдет сопутствующее смещение. Активный уровень независимой переменной (3 мг кофеина на 1 кг массы) окажется неразрывно связанным с активным уровнем другой переменной — осознанием факта его приема.

Если же вместо двухуровневого эксперимента провести многоуровневый эксперимент с дозами 0, 1, 2, 3, 4 и 5 мг кофеина на

килограмм массы, можно получить кривую, показанную на рис. 7.2. Уровень нулевой дозировки, конечно, обозначает плацебо. На рисунке видно постепенное сокращение времени реакции по мере продвижения от дозы 1 к дозе 5. Более резкое падение кривой от 0 к 1 может быть частично следствием осознания факта приема кофеина. Однако дальнейшее регулярное изменение дает достаточно убедительное доказательство того, что кофеин непосредственно увеличивает реактивность. Маловероятно, что эта кривая может быть результатом нарастающего осознания увеличивающихся доз вещества.

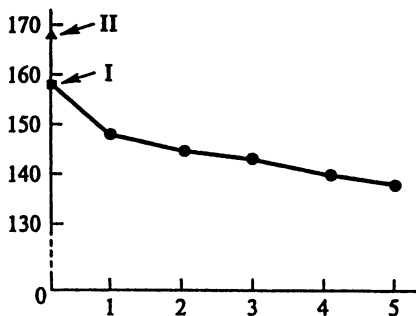


Рис. 7.2. Результаты воображаемого многоуровневого эксперимента с влиянием кофеина на время реакции: ось абсцисс — дозы кофеина (мг/кг); ось ординат — время реакции (мс). I — плацебо, II — без таблетки

Если экспериментатора специально интересует действие плацебо, он может организовать еще одно контрольное условие: вовсе не давать никакой таблетки. Это условие представлено на рис. 7.2 треугольником. Меньшее время реакции при нулевой дозе кофеина (плацебо) показывает, что в самом деле имеет место действие плацебо, равно как и осознания принятия кофеина, равно как и непосредственное действие кофеина. Как много может происходить в простом эксперименте!

В двухуровневых экспериментах вероятность такого сопутствующего смешения очень велика. В эксперименте по трудовой этике, когда девочка нажимала на рычаг, она, наверное, делала нечто большее, чем зарабатывала один шарик за десять нажатий. Таким способом она осуществляла также контакт с невидимым ей экспериментатором (или с тем, что, по ее предположению, находилось в большом ящике). Поэтому в какой-то степени нажатие рычага могло оказаться связанным с активным уровнем второй переменной — установлением контакта с невидимым экспериментатором. В многоуровневом эксперименте с использованием различного числа необходимых нажатий на рычаг постепенное увеличение предпочтения по мере увеличения числа нажатий было бы более четким доказательством экспериментальной гипотезы. Однако если бы небольшое количество работы (например, один шарик за каждое нажатие) выбиралось так же охотно, как и большее количество работы (например, один шарик за каждые пять нажатий), доказательство гипотезы было бы слабым. И снова многоуровневый эксперимент исключил бы опасность сопутствующего смешения.

КОЛИЧЕСТВЕННЫЕ И КАЧЕСТВЕННЫЕ НЕЗАВИСИМЫЕ ПЕРЕМЕННЫЕ

Преимущества многоуровневого эксперимента над двухуровневыми экспериментами выходят далеко за рамки только что описанного контроля. В общем можно сказать, что он способствует дальнейшему развитию понимания, обеспечивая проверку более разработанных экспериментальных гипотез. Эти преимущества будут описаны в данном и в трех последующих разделах.

Покажем прежде всего преимущество, которое определяется не тем, что используется несколько уровней независимой переменной, а тем, что независимая переменная представлена в *количественной* форме. В отличие от нее *качественная* независимая переменная не поддается описанию с помощью чисел. Чтобы иметь качественную независимую переменную, эксперимент не обязательно должен быть двухуровневым. Мы видели в главе 4, что Р. Гейтвуд и Р. Перлофф (Gatewood R. D., Perloff R., 1973) сравнивали три способа сообщения о стоимости товаров покупателям универсама. Различия между тремя условиями были качественными, ибо их нельзя было описать числами. С другой стороны, хотя эксперимент по рабочей этике был двухуровневым (два условия независимой переменной), независимая переменная в нем, как уже говорилось, была по существу количественной.

Уже сама способность вводить количественную или градуированную независимую переменную обнаруживает прогресс в *выделении* существенного фактора. Давайте сравним сходные эксперименты, в одном из которых используется качественная переменная, в другом — количественная. Экспериментатор может выдвинуть гипотезу, что знаковая информация перерабатывается быстрее в зрительном канале, чем в слуховом. Затем он ставит эксперимент, в котором сравнивается скорость чтения и восприятие на слух. Для отобранных им текстов он находит, что его испытуемые могут читать и хорошо понимать текст со скоростью около 1000 слов в минуту; однако при восприятии на слух текста, который произносится со скоростью большей, чем 200 слов в минуту, понимание уже затрудняется. Без сомнения, такой эксперимент показывает, что человек может читать быстрее, чем слушать. Однако это не значит, что подтвердилась гипотеза о более быстрой переработке знаковой информации в зрительном канале.

Чтение и прослушивание различаются по многим параметрам. Один из них состоит в том, что при чтении человек может забежать вперед или возвращаться назад в поисках ключевых слов или идей. При слуховом же восприятии человек вынужден полагаться на свои ожидания или память для расширения его «настоящего времени».

Следуя за этим рассуждением, экспериментатор может попытаться выяснить, ухудшается ли чтение при ограничении объема материала, видимого в данный момент. С этой целью он использует ограничительную рамку, за которой движется строка печатного текста. Независимой переменной является ширина рамки, помещается ли в нее в среднем одно слово, два слова, три слова, пять слов или десять слов. Может оказаться, что скорость чтения будет увеличиваться с расширением окошка (до некоторой величины). В результате такого эксперимента с количественной независимой переменной будут получены некоторые сведения о факторах, определяющих скорость чтения. Он заставит нас также проявлять бóльшую осторожность при отнесении различий в чтении и слуховом восприятии текстов за счет фундаментальных различий в скорости зрительных и слуховых процессов переработки информации.

Конечно, сравнение скорости чтения и скорости слухового восприятия является ценным экспериментом, поскольку оно работает на экспериментальную гипотезу. Но гипотеза сама по себе еще ничего не прибавляет к пониманию, поскольку независимая переменная в ней — это «клубок» факторов. Проверка той же самой гипотезы в эксперименте с ограничительной рамкой была бы более информативной, поскольку различные условия представлены в нем различными количествами лучше выделенной переменной. Благодаря этому можно было бы лучше понять причину различий в результатах.

Как мы уже видели, существует много конкретных способов представления количественной независимой переменной: дозировка вещества, временной интервал между элементами списка, число элементов, которые нужно удержать в памяти, необходимое число нажатий на рычаг, различный вес поднимаемых грузов и т. п. Здесь может быть также использовано психологическое шкалирование. Например, триграммы Р. Калфи и Р. Андерсон имели «высокую степень осмысленности». Большая группа испытуемых предварительно оценила множество триграмм по степени их осмысленности. Все триграммы, использовавшиеся Р. Калфи и Р. Андерсон, имели высокую оценку.

Существуют шкалы оценки вербального материала по таким различным параметрам, как *произносимость* и *эмоциональность*. Ассоциативная сила слова обычно оценивается по продуктивному критерию — количеству ассоциаций, которые может дать испытуемый на заданное слово в течение фиксированного периода времени. Опубликованы таблицы, содержащие разные типы подобных шкал (например, Runquist W. N., 1966). По желанию экспериментатор может использовать их для независимой переменной в исследованиях значимости степени осмысленности, произносимости, эмоциональности или ассоциативности. Например, он

может исследовать зависимость между произносимостью слов и скоростью их запоминания.

В эксперименте можно шкалировать и большие единицы словесного материала. Например, можно оценить предложения по степени их приятности или ясности. Можно даже прошкалировать забавность шуток. Конечно, мнения испытуемых могут разойтись с вашим. Чтобы обойти эту трудность, нужно вначале взять большое число шуток, например 200. Затем предложить пятидесяти судьям оценить их по 10-балльной системе: 10 — самая смешная, 1 — самая плоская. Отобрать затем только те шутки, которые всеми судьями будут оценены одинаково. Так могут остаться только 60 или около того шуток. Шкальная оценка каждой из них будет просто средним баллом. Так может оказаться шутка с высшей оценкой 9,4; другая — со средней оценкой 5,3; еще одна — с низкой оценкой 1,5. На таком материале можно далее проверять гипотезу, например, о том, что очень смешные и очень саркастические шутки запоминаются лучше, чем умеренные. Конечно, первоначальная группа судей не должна участвовать в основном эксперименте.

ГИПОТЕЗА О МАКСИМАЛЬНОЙ (ИЛИ МИНИМАЛЬНОЙ) ВЕЛИЧИНЕ

Эксперимент на скорость предъявления и запоминание

Возвратимся снова к эксперименту Р. Калфи и Р. Андерсон (Calfée R. C., Anderson R., 1971), в котором общее время запоминания списка было распределено различным образом. Как можно видеть на рис. 7.1, шесть различных временных интервалов между элементами списка составили шесть уровней независимой переменной. Каждый интервал исследовался для отдельной группы испытуемых в 20 человек. *Максимальная* величина зависимой переменной — чуть больше 13 правильных ответов — обнаружилась для интервала 4 с, т. е. при пяти повторных предъявлениях списка из 16 элементов в течение 320 с экспериментального времени. Этот результат был как раз таким, который ожидали Р. Калфи и Р. Андерсон. Остановимся вкратце на причинах, по которым экспериментальная гипотеза о максимальном значении при одном из промежуточных уровней независимой переменной имеет смысл.

Эксперимент по исследованию научения

Р. Калфи и Р. Андерсон не были удивлены своими результатами, чего нельзя сказать про двух более ранних исследователей (1908!). Теперь, 70 лет спустя, следуя по стопам этих пионеров, мы могли бы сказать, какой должна была быть их гипотеза. Но обратимся сначала к эксперименту.

Экспериментаторами были Р. Йеркс и Дж. Додсон. Р. Йеркс, который интересовался эволюционным развитием, в то время тщательно изучал любопытное маленькое животное под названием *танцующая мышь* (Yerkes R. M., 1907). Это название связано с тем, что иногда из-за генетического дефекта данная разновидность домашней мыши непрерывно движется по кругу или выписывает восьмерки. Кстати, позже Р. Йеркс провел свои известные исследования на шимпанзе и призывниках на военную службу.

В эксперименте 1908 г. у мыши вырабатывался навык, основанный на различении черного и белого. Перед мышью находились две двери в два туннеля. В каждой пробе мышь осторожно вынуждали войти в тот или другой туннель, постепенно уменьшая с помощью листа картона пространство камеры перед туннелями. Стены и потолок одного туннеля, как и пространство перед входом в него, были выложены белым картоном. Второй туннель аналогичным образом выстилался черным картоном. Пол каждого туннеля представлял собой проволочную сетку. Только в том случае, когда мышь входила в *белый* туннель, через сетку пропускался ток. Пройдя через любой туннель, мышь попадала в свою камеру-гнездо, где ее ожидала мышь противоположного пола. Так что страдание экспериментального животного компенсировалось состраданием.

Мышь не могла научиться избегать удара тока, просто ориентируясь на правую или левую сторону, поскольку местоположение белого и черного туннеля менялось в случайном порядке от пробы к пробе. С каждым животным ежедневно проводилось по 10 проб. Тренировки продолжались до тех пор, пока животное не совершало все пробежки в течение трех дней подряд без ошибок. Например, самец № 128 смог это сделать на 16, 17 и 18-й дни. Очевидно, он научился различать черное и белое к концу 15-го дня. Поэтому можно было считать, что данное животное достигло *критерия* научения после 150 проб — ведь ежедневно было по 10 проб.

Разным мышам подавали ток различной силы. Он измерялся в условных «единицах стимуляции». *Слабый* уровень (125 единиц) почти не замечался мышью. «*Сильный* стимул (500 единиц) был крайне неприятен даже самим экспериментаторам, и мыши также энергично реагировали на него» (Yerkes R. M., Dodson J. D., 1908, с. 467—468). *Средний* уровень составлял 300 единиц. Действие каждого уровня электроудара исследовалось на четырех мышках — двух самках и двух самцах.

В качестве зависимой переменной было выбрано среднее число проб, необходимых для достижения критерия научения всеми четырьмя мышками. Результаты показаны на рис. 7.3. Видно, что *минимальное* значение зависимой переменной, 80 проб, было получено при *промежуточном* уровне независимой переменной — 300 единицах стимуляции. На самом деле научение при самой слабой стимуляции было даже хуже, чем показано, поскольку одна

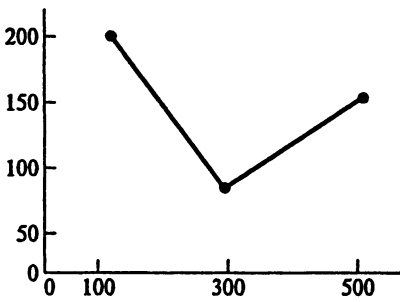


Рис. 7.3. Влияние силы электроудара на успешность научения (опыты на танцующих мышах). Ось абсцисс — сила электроудара (в условных единицах); ось ординат — среднее количество проб, необходимых для достижения критерия успешного решения задач

из четырех мышей так и не научилась различать черный и белый туннели, и ей условно приписали 200 проб, ибо после них эксперимент прекратился.

Конечно, использование только трех уровней независимой переменной с трудом позволяет квалифицировать эксперимент как многоуровневый. Это беспокоило и Р. Йеркса и Дж. Додсона: «Следует иметь в виду, что поскольку использовалось только три интенсивности стимула... возможно, наиболее благоприятная сила стимуляции обнаружена не была» (Yerkes R. M., Dodson J. D., 1908, с. 482).

Прогресс в понимании

Существует множество экспериментов, в которых гипотеза максимума или минимума вполне оправдана. Между прочим, не нужно считать, что термины *максимум* и *минимум* означают в этих экспериментах разное. Если бы в эксперименте Р. Йеркса и Дж. Додсона мерой научения служило число правильных ответов, то при 300 единицах стимуляции достигался бы *максимум*. Если бы Р. Калфи и Р. Андерсон использовали в качестве зависимой переменной среднее число ошибок, а не правильных ответов, то при интервале 4 с достигался бы *минимум*.

Напомним, что количественная гипотеза в эксперименте Сингха и Квери по трудовой этике состояла в том, что предпочтение работы бездеятельности будет расти с увеличением требуемого количества нажатий на рычаг, — но *только до определенного предела*: после того как будет достигнут максимум, дальнейшее увеличение количества нажатий поведет к уменьшению предпочтения этого условия. Основанием для такого предположения была мысль о том, что по мере увеличения необходимого количества нажатий происходят сразу два процесса. Вначале растет «чувство активности», и оно увеличивает привлекательность работы с рычагом. Однако эти усилия не могут быть только приятными; с дальнейшим увеличением числа нажатий возникает и нарастает «неприятное» чувство. Максимальное предпочтение условия с нажатием на рычаг бездеятельному условию будет достигнуто при таком уровне нажатий,

когда разница между чувством активности и неприятным чувством будет наибольшей. Итак, одним из оснований гипотезы максимума (или минимума) является теория двух противоположных основных процессов, определяемых независимой переменной. Причем «негативный» процесс при достижении независимой переменной высокого уровня становится сильнее «позитивного».

Эксперименты Р. Калфи и Р. Андерсон (Calfee R. C., Anderson R., 1971) с запоминанием демонстрируют другой вид противоположных тенденций. Для того чтобы воспроизводить правильные числа при показе триграмм, испытуемый в первую очередь не должен смешивать сами триграммы. Это называется *различением стимулов*. Ему способствует *уменьшение* интервала между элементами списка. Во-вторых, испытуемый должен научиться связывать каждое число с парной ему триграммой. Это называется *ассоциативным процессом*. Ему способствует *увеличение* времени между элементами списка. Таким образом, должен существовать оптимальный интервал. Любой более короткий интервал дает выигрыш в различении в ущерб ассоциативному процессу; любой более длинный интервал дает выигрыш в ассоциативном процессе в ущерб различению. Итак, второе основание для ожидания максимума (или минимума) — это теория, согласно которой увеличение независимой переменной вызывает противоположные изменения в двух основных процессах, каждый из которых «положителен». Максимум или минимум достигается при уровне, который дает оптимальное сочетание этих двух процессов.

Можно показать, что танцующая мышь, учившаяся избегать удара током, находилась точно в таком же положении. Она должна была различать два туннеля и ассоциировать туннель с ударом тока или его отсутствием. По свидетельству Р. Йеркса и Дж. Додсона, различение было плохим при слишком сильном ударе. «Поведение мышей менялось по мере усиления стимуляции. При сильной стимуляции они выбирали не менее быстро, чем при слабой, однако в первом случае они были менее осторожны и действовали с меньшей осмотрительностью и уверенностью» (Yerkes R. M., Dodson J. D., 1908, с. 476). Таким образом, различение стимулов (черного и белого) ухудшалось с увеличением силы удара. Ассоциирование же белого туннеля с ударом (при состоявшемся различении) могло с усилением удара только усиливаться. Следовательно, здесь снова должен был существовать некоторый уровень независимой переменной (силы удара), оптимальный для дискриминационного научения.

Наверное, некоторые из вас уже предвосхитили дальнейшее рассуждение. Оно состоит в том, что более сложное различение требует большей осторожности и осмотрительности, чем простое. Это означает прежде всего, что оно осваивается медленнее. Более того, оптимальным для его освоения будет удар более слабый, чем для простого различения. Как раз такие результаты и получи-

ли исследователи в своей дальнейшей работе. Вот их заключение: «По мере увеличения сложности различения интенсивность стимула, оптимальная для формирования навыка, приближается к порогу» (там же, с. 481). Эта зависимость сегодня известна под названием *закона Йеркса—Додсона*. Но мы забежали вперед, к анализу этих экспериментов мы обратимся в следующей главе.

На примере трех описанных исследований было показано, что многоуровневый эксперимент может обеспечить проверку гипотезы о двух процессах, связанных с уровнем независимой переменной противоположным образом. Действительная экспериментальная гипотеза состоит в том, что максимум (или минимум) зависимой переменной будет достигаться при некотором промежуточном уровне независимой.

ГИПОТЕЗЫ ОБ АБСОЛЮТНЫХ И ОТНОСИТЕЛЬНЫХ ОТНОШЕНИЯХ

В трех экспериментах, которые мы обсудим ниже, экспериментальная гипотеза состояла в том, что зависимая переменная должна изменяться постепенно по мере постепенного изменения независимой переменной. Однако предполагаемые гипотезой отношения во всех трех случаях различны. Мы начнем с наиболее простого отношения и перейдем к наиболее сложному.

Любое изменение можно представить либо в *абсолютных* величинах, либо в *пропорциональных (относительных)*. Так, увеличение от 4 до 6 может быть описано как абсолютное увеличение на 2 или как относительное увеличение в 0,5 (т.е. 6 на 50 % больше, чем 4).

В первых из трех нижеследующих экспериментов предполагалось, что равные *абсолютные* изменения независимой переменной будут приводить к равным абсолютным изменениям зависимой переменной. Во втором эксперименте предполагалось, что равные *относительные* изменения независимой переменной будут приводить к равным *абсолютным* изменениям зависимой переменной. В третьем эксперименте предполагалось, что равные *относительные* изменения независимой переменной будут иметь результатом равные *относительные* изменения зависимой переменной. Все это начинает звучать слишком абстрактно, поэтому перейдем к самим экспериментам.

Эксперимент с гипотезой «абсолютно-абсолютного» отношения: исследование запоминания

Представим эксперимент, в котором испытуемому на короткое время предъявляется какая-то цифра. Если это одна из двух цифр (скажем, 2 или 5), испытуемый нажимает правую кнопку; если цифра не относится к этому набору, который будет называться

позитивным (0, 1, 3, 4, 6, 7, 8 или 9), он нажимает левую кнопку. Время реакции измеряется от момента появления цифры до нажатия на кнопку. В различных сериях используются разные объемы позитивного набора: он может состоять из одной, трех, четырех, пяти или шести цифр. При этом находят среднее время реакции для каждого объема позитивного набора. Затем строится график, выражающий зависимость времени реакции от объема набора. Результаты такого эксперимента (Sternberg S., 1969) показаны на рис. 7.4. Как можно видеть, линия, соединяющая все точки, не совершенно прямая. Однако небольшая волнистость скорее всего связана со случайными вариациями. На основании этих результатов мы можем сказать, что для получения одного и того же абсолютного прироста времени реакции (35 мс) уровень независимой переменной (объем набора) должен быть увеличен на одну единицу (снова в абсолютных единицах).

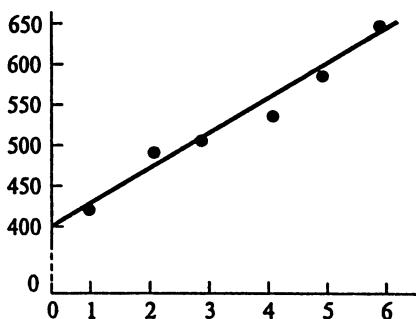


Рис. 7.4. Зависимость времени реакции опознания тестового стимула от количества знаков в ранее предъявленном наборе — эксперимент Стернберга (Sternberg S., 1972). Ось абсцисс — размер предъявленного набора; ось ординат — время реакции (мс)

Эксперимент с гипотезой «относительно-абсолютного» отношения: реакция выбора

Два экспериментатора, У.Хик в Англии и Р.Хаймен в США, провели почти в одно и то же время сходные эксперименты в русле так называемого *информационного подхода* (Hick W. E., 1952; Human R., 1953). Для позитивного набора стимулов они нашли иную закономерность между числом альтернатив и временем реакции, чем С. Стернберг. Методика У.Хика несколько проще методики Хаймена, поэтому мы приводим ее в качестве примера.

Перед испытуемым полукругом располагалось 10 маленьких электрических лампочек. Его пальцы (включая большие) свободно лежали на десяти телеграфных ключах. Когда зажигалась лампочка, испытуемый должен был нажать соответствующий ключ. Десять альтернатив составляли наиболее высокий уровень независимой переменной. В других условиях могли зажигаться 8, 6, 5, 4, 3, 2, либо даже одна из десяти лампочек. Каждый новый сигнал появлялся через 5 с после предыдущего ответа.

Специальные серии проб проводились для 1, 2, 3, 4, 5, 6, 8 и 10 альтернатив. Для каждого из этих уровней определялось сред-

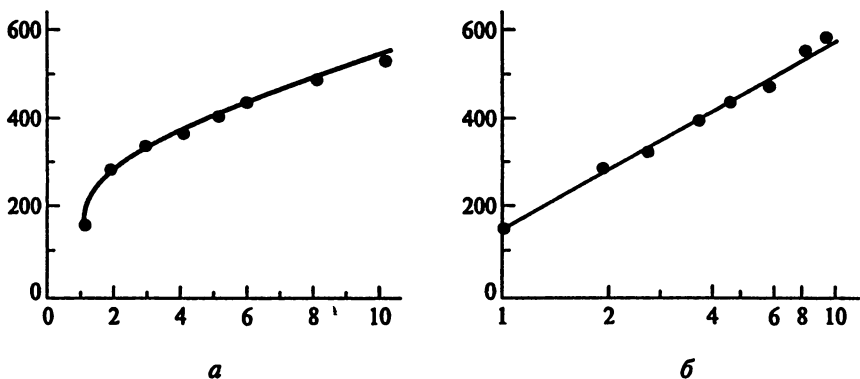


Рис. 7.5. Зависимость времени реакции от количества альтернатив — эксперимент У. Хика (Hick W. E., 1952). Ось абсцисс — количество альтернатив; ось ординат — время реакции (мс)

нее время реакции. На рис. 7.5 представлена двумя различными способами зависимость между числом альтернатив и временем реакции. В случае (а) шкала независимой переменной (ось абсцисс) является абсолютной, как и для данных С. Стернберга. На ней одинаковому увеличению числа альтернатив соответствует одинаковое расстояние на шкале. Однако график получился не прямым, а изогнутым книзу. Каждое новое увеличение числа альтернатив на единицу вызывает все меньший прирост времени реакции. Если же ось абсцисс изменить и представить в том виде, как на правом рисунке (б), график выпрямляется. Деления на рис. 7.5, б представляют собой шаги в относительных единицах (такая шкала называется *логарифмической*). Так, на этой шкале каждому удвоению числа альтернатив — от 1 к 2, от 2 к 4, от 3 к 6 — соответствуют одинаковые расстояния. Чтобы удостовериться, измерьте шкалу.

Таким образом, прямая линия, представляющая отношение между числом альтернатив и временем реакции, подтверждает гипотезу «относительно-абсолютное». При увеличении числа альтернатив в равном отношении мы получаем равное абсолютное увеличение времени реакции. Удвоение числа альтернатив является относительным увеличением на единицу, т. е. на 100%. Каждое такое удвоение увеличивает время реакции на 110 мс — на одну и ту же абсолютную величину.

Эксперимент с гипотезой «относительно-относительное»: субъективная тяжесть

Чувствует ли человек вес в 200 г как половину веса в 400 г? Этот вопрос задал себе С. Стивенс. Проводя серии экспериментов

в течение ряда лет, он использовал не только веса, но также звуковые тоны, свет, запахи и т. д.

Эксперимент с весами был организован Р. Харпером и С. Стивенсом очень просто. Испытуемый стоял возле стола, на котором лежали семь одинаковых закрытых контейнеров. Один из них откладывали в сторону и называли «эталоном». Задача состояла в том, чтобы взвешивая в руке эталон и каждый из шести других весов выбрать тот, «который воспринимался как половина веса эталона» (Harper R. S., Stevens S. S., 1948, с. 344). Заметьте: испытуемые должны были найти не тот вес, который действительно составлял половину веса эталона, а только ощущался таковым.

Было найдено, что вес, который воспринимается как в два раза более легкий, чем эталон 100 г, в среднем составлял 72 г. Экспериментаторы выразили это отношение следующим образом. Пусть вес 100 г имеет субъективную тяжесть 1 вег (по определению), тогда вес 72 г имеет субъективную тяжесть $1/2$ вег. Между прочим, слово «вег» Р. Харпер и С. Стивенс образовали от старого норвежского слова *veg* — поднимать.

В другой серии в качестве эталона использовался вес, отличный от 100 г. Оказалось, что когда в качестве эталона брался вес 140 г, половинное ощущение тяжести соответствовало в среднем 100 г. Поскольку 100 г представлялись как 1 вег, то 140 г было приписано значение тяжести 2 вег.

Всего было 8 серий, в которых эталон изменялся от 20 до 2000 г. В результате всех ответов была получена сглаженная кривая, отражающая отношение между физическими весами и ощущениями их тяжести (рис. 7.6, а).

Можно видеть, что когда горизонтальная и вертикальная оси представлены абсолютными шкалами (например, расстояние между 100 и 200 г такое же, как и между 500 и 600 г, а расстояние между 10 и 20 вег такое же, как между 30 и 40), линия не является прямой, а изгибается кверху. Однако когда обе оси были прокалированы в относительных единицах (например, расстояние между 100 и 200 г такое же, как между 400 и 800 вег, а расстояние между 2 и 4 вег такое же, как между 8 и 16 вег), результаты очень хорошо легли на прямую (рис. 7.6, б).

Таким образом, была подтверждена гипотеза об «относительно-относительной» зависимости. Каждый раз, когда вы увеличиваете физический вес в определенное число раз, ощущение тяжести также увеличивается в определенное число раз. Следует также заметить, что оси разделены на одинаковые относительные единицы. Например, расстояние между 2 и 4 вег такое же, как между 100 и 200 г. Если относительное увеличение веса дает почти такой же относительный прирост в ощущении тяжести, график зависимости будет иметь угол наклона 45° (как показано пунктирной линией). В данном эксперименте относительное увеличение на еди-

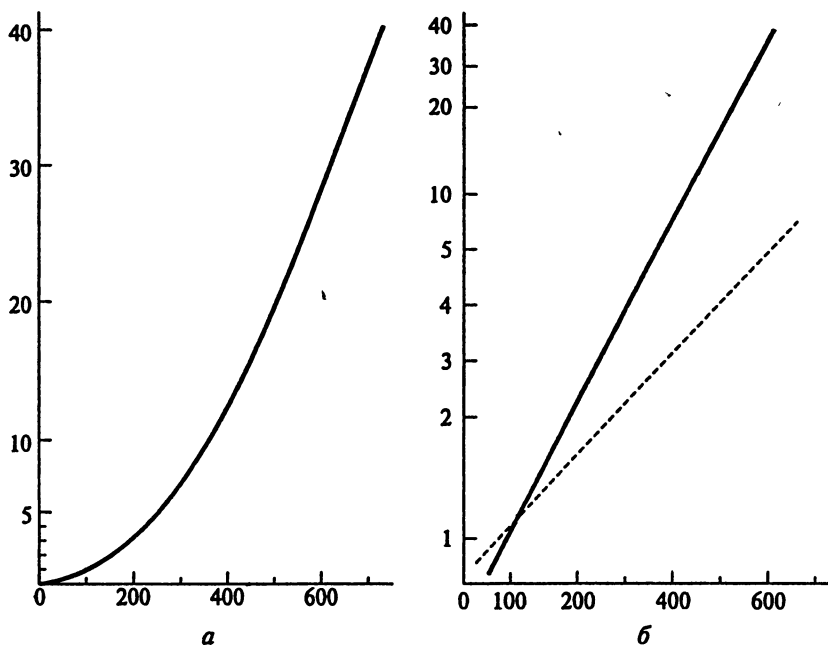


Рис. 7.6. Изменения в ощущении тяжести с увеличением поднимаемого груза — эксперимент Р. Харпера и С. Стивенса (Harper R. S., Stevens S. S., 1948). Ось абсцисс — вес груза (г); ось ординат — ощущение тяжести (в вегах)

ницу (т. е. на 100 %) физического веса дает большое относительное увеличение субъективной тяжести — примерно на 2,5 ед. (т. е. на 250 %). Другими словами, удвоение физического веса почти учетверяет его субъективную тяжесть.

Возвращаясь к поставленному вначале вопросу, мы теперь можем ответить, что вес 200 г не воспримется как половина веса 400 г: он покажется значительно легче.

Прогресс в понимании

Каждое из трех рассмотренных отношений представляет теорию или модель механизмов, лежащих в основе поведения. Механизм, предполагаемый «абсолютно-абсолютной» гипотезой С. Стернберга, следующий. Все цифры позитивного набора фиксируются в памяти испытуемого. Когда появляется тестовая цифра, элементы позитивного набора последовательно «сканируются» для определения того, принадлежит данная цифра к позитивному набору или нет. Другими словами, предъявленная цифра по очереди сравнивается с каждым элементом набора. Если на каждое такое сравнение уходит 35 мс, то при добавлении к позитивному набору еще одной

цифры общее время сканирования возрастет именно на эту величину. Таким образом, тот факт, что каждое абсолютное увеличение позитивного набора на один элемент сопровождается увеличением времени реакции на одну и ту же абсолютную величину, подтверждает модель последовательного сканирования.

Гипотеза У.Хика об «относительно-абсолютной» зависимости времени реакции от числа альтернативных наборов вытекает из модели другого типа. Идея ее состоит в том, что испытуемый совершает выбор, применяя стратегию последовательности простых решений. Так, если существует восемь альтернатив, первое простое решение состоит в выборе между альтернативной группой 1, 2, 3, 4 и альтернативной группой 5, 6, 7, 8. Предположим, правильным выбором является альтернатива 7. Тогда первым решением будет выбор группы 5, 6, 7, 8. Следующее простое решение будет состоять в выборе между группами 5, 6 и 7, 8. Правильным вторым простым решением будет группа 7, 8. Остается только выбор между 7 и 8. Третьим и последним простым решением будет выбор альтернативы 7. В целом для восьми альтернатив мы имели только три простых решения. Для различного числа альтернатив будет сохраняться следующее: две альтернативы — одно решение; четыре альтернативы — два решения; восемь альтернатив — три решения и т.д. Если каждое простое решение требует одного и того же количества времени, которое, как здесь установлено, равно 110 мс, то для каждого увеличения числа альтернатив на одну относительную единицу (т.е. на 100 %) мы будем иметь одно и то же абсолютное увеличение времени реакции (на 110 мс). Конечно, эта теория должна быть уточнена и для другого числа альтернатив, которое не является степенями двойки, например для шести или десяти.

Гипотеза С.Стивенса об относительном приросте ощущения при относительном увеличении стимула основывается на его концепции о механизме преобразования физической энергии в сенсорном органе: пропорциональное увеличение энергии стимула дает почти пропорциональное увеличение нервного возбуждения.

В каждом из этих трех экспериментов проверявшаяся гипотеза и полученные результаты отражают глубину понимания существенно большую, чем простое знание переменной, воздействующей на поведение. Благодаря нахождению точного отношения между независимой и зависимой переменными (обе понимались как непрерывные величины) мы смогли проникнуть в механизмы соответствующих процессов.

ПРЕДЫДУЩИЕ ЭКСПЕРИМЕНТАЛЬНЫЕ СХЕМЫ В ПРИЛОЖЕНИИ К МНОГОУРОВНЕВОМУ ЭКСПЕРИМЕНТУ

В предыдущих главах были описаны две основные экспериментальные схемы. Это — межгрупповая схема, при которой каждое

экспериментальное условие предъявляется отдельной группе испытуемых, и схема интраиндивидуальных проверок. Обе эти схемы будут рассмотрены теперь в связи с многоуровневым экспериментом. Из этого анализа будет понятно, почему исследователей привлекает еще одна, третья, основная схема, которая будет описана в следующем разделе.

Межгрупповые схемы

Р. Калфи и Р. Андерсон (Calfee R. C., Anderson R., 1971) применили в своих экспериментах на запоминание межгрупповую схему. Они использовали шесть различных интервалов между элементами запоминаемого списка: 1, 2, 3, 4, 10 и 20 с. Каждый интервал предъявлялся отдельно группе в 20 человек; таким образом, в опытах участвовало всего 120 человек. (И это была только четвертая часть обширного исследования, где в общем участвовало 480 человек!) Испытуемые распределялись по уровням (независимой переменной) случайно, по мере того как они приходили на опыты. Один недостаток использования межгрупповой схемы в многоуровневом эксперименте очевиден сразу. Поскольку для каждого уровня требуется большое количество испытуемых, то (с целью уравнивания групп) общее необходимое количество испытуемых становится нереально большим.

Схема с интраиндивидуальным контролем

Вы, конечно, помните, что в первых экспериментах, описанных в этой книге, основная «угроза» внутренней валидности заключалась в случайных изменениях во времени и в эффектах последовательности, поскольку различные условия предъявлялись одному и тому же испытуемому. Здесь необходимо было организовать контроль по отношению к тому же испытуемому. Тремя видами интра-, или внутрииндивидуального, контроля были: регулярное чередование (наушники), позиционное уравнивание (фортепьянные пьесы) и случайная последовательность (томатный сок). Схему индивидуального эксперимента использовал и Хик в своем исследовании числа альтернатив и времени реакции: в этом эксперименте он был единственным испытуемым. Однако эта схема может быть использована и при участии нескольких испытуемых, что является более типичным. Тогда описанный выше контроль организуется в отношении каждого испытуемого. Если предъявляется подряд большое количество проб и испытуемый не должен знать уровень независимой переменной в каждой из них, обычно используется случайный порядок предъявлений. Например, в одном эксперименте (Gottsdanker R., Way T. C., 1966) варьировался в случайном порядке интервал между двумя сигналами, на ко-

торые следовало давать ответ. Все интервалы были короткими: 50, 100, 200, 400 и 800 мс. Экспериментаторы проверяли гипотезу о том, что время реакции на второй сигнал будет равномерно сокращаться по мере увеличения интервала (до некоторого значения). Было, конечно, важно, чтобы испытуемый не знал, каким будет следующий интервал. В серии из 100 проб каждый из пяти уровней (длительности интервала) появлялся в случайном порядке 20 раз. Всем восьми испытуемым предъявлялась одна и та же случайная последовательность интервалов.

Однако схема внутрииндивидуального контроля непригодна как в многоуровневом эксперименте, требующем позиционно уравновешенной последовательности, так и в экспериментах, где предъявление каждого условия длится достаточно долгое время (как при заучивании фортепьянных пьес). Позиционно уравненную последовательность для двух условий мы записали ранее как АББА, где А или Б представляют пробу с одним условием. Для многоуровневого эксперимента внутрииндивидуальный позиционно уравненный порядок для шести уровней выглядел бы следующим образом: АБВГДЕДГВБА. Однако при большой длительности каждой пробы предъявление всей последовательности каждому испытуемому было бы практически неосуществимым.

СХЕМЫ С ПОЗИЦИОННЫМ КРОССИНДИВИДУАЛЬНЫМ УРАВНИВАНИЕМ

Мы только что видели, что в многоуровневом эксперименте схема межгруппового сравнения может потребовать слишком много испытуемых для уравнивания групп, а схема внутрииндивидуального контроля — слишком много времени на каждого испытуемого для элиминирования влияний последовательности. Выход состоит в том, чтобы каждому испытуемому предъявлять каждое условие, контроль же влияния последовательности проводить *по всем* испытуемым.

В результате одни и те же испытуемые будут проведены через все уровни и каждому испытуемому каждый уровень будет предъявлен только раз. Однако одному испытуемому (или группе) условия будут предъявлены в последовательности АБВГДЕ, другому же испытуемому (или группе) — в последовательности ЕДГВБА. Такие схемы обычно объединяют со схемами внутрииндивидуального контроля, относя их к одному классу — схемам с повторными замерами, поскольку здесь каждому испытуемому предъявляется больше одного условия. Однако между ними имеется весьма важное различие. При использовании внутрииндивидуального контроля пробы, предъявляемые каждому испытуемому, составляют полный эксперимент. Что же касается внутренней валидности, то группа испытуемых используется для улучшения надеж-

ности, а не для контроля систематического смещения. Если применяется кроссиндивидуальный контроль, то заранее известно, что результаты каждого испытуемого будут искажены систематическим смещением. Для преодоления же этого систематического смещения требуется более одного испытуемого. Теперь мы опишем три наиболее распространенные схемы для многоуровневых экспериментов, использующих кроссиндивидуальное уравнивание.

Реверсивное уравнивание

Реверсивное (обратное) уравнивание — это схема, которую мы только что обсуждали. Она может быть представлена следующим образом.

Группа испытуемых	Последовательность условий (уровней)
1	ВБАГД (вообще любая)
2	ДГАБВ (обратная ей)

Это означает, что используются только две последовательности уровней. Как мы только что показали, они не обязательно должны быть АБВГДЕ и ЕДГВБА, где А означает наименьший уровень независимой переменной и Е — наибольший уровень. Здесь вообще могут быть разные варианты. Например, в другой части экспериментов Р. Готтсданкера и Т. Уэй, о котором говорилось выше, в одном блоке проб временной интервал между двумя стимулами оставался постоянным. Одной группе из четырех испытуемых предъявлялось пять блоков по 100 проб с временными интервалами в следующем порядке: 50, 100, 200, 400 и 800 мс (т.е. АБВГД). Порядок предъявления для другой группы из четырех испытуемых был: 800, 400, 200, 100 и 50 мс (т.е. ДГВБА).

Реверсивное уравнивание обеспечивает для каждого уровня одну и ту же *среднюю* позицию по двум последовательностям. Так, для двух показанных на диаграмме порядков ВБАГД и ДГАБВ уровень Д находится в позиции 5 и 1 при среднем 3; уровень Г — в позиции 4 и 2 при среднем, снова равном 3, и т.д. Это уравнивание обеспечивает хороший контроль влияния последовательности, только если эффект переноса однороден, т.е. если предполагается, что позиция 1 влияет так же на позицию 2, как позиция 2 на 3, или 3 на 4, или 5 на 6.

Однако эффект переноса может быть *неоднороден*, как это было показано в главе 2 применительно к внутрииндивидуальной схеме; тогда возникает серьезная проблема. Предположим, что существуют эффекты научения, которые равномерно улучшают ответ вплоть до третьей пробы, но не дальше. Для испытуемых, которым предъявляется последовательность ВБАГД, последние три

уровня — А, Г и Д — будут в одинаково «выгодном положении». Для испытуемых, которым предъявляется обратная последовательность ДГАБВ, последние уровни — А, Б и В — будут также в одинаково «выгодном положении». Поэтому уровень А, находящийся в середине обеих последовательностей, будет иметь наибольшее преимущество, а В и Д — наименьшее. Если же эффект переноса связан с утомлением, а не научением, то теперь уровень в середине обеих последовательностей окажется в наиболее неблагоприятном положении.

Если эффект переноса различен в различных последовательностях, то величина переноса оказывается переменной, производящей смещение. В только что разбиравшейся последовательности ВБАГД величина переноса для В равна 0 (поскольку это первое условие), для Б — 1 и для А, Г и Д — 2 (поскольку перенос не увеличивается после третьей пробы). Аналогично для обратной последовательности — ДГАБВ — величины переноса будут: 0 для Д, 1 для Г и 2 для А, Б, В. Общий суммарный эффект переноса будет равен: 4 для А, 3 для Б и Г, 2 для В и Д. Из-за неэффективности в подобных случаях схемы реверсивного уравнивания исследователи обратились к схемам, которые обеспечивают лучший контроль. Они и будут сейчас описаны.

Полное уравнивание

Для того чтобы избежать систематического смещения, возникающего при неоднородном переносе в схеме реверсивного уравнивания, можно использовать все возможные последовательности уровней вместо двух. Такая схема с полным уравниванием для трехуровневого эксперимента выглядит следующим образом.

Так, если бы в исследовании Р. Готтсданкера и Т. Уэй было использовано только три уровня независимой переменной (например 50, 100 и 200 мс), различным испытуемым — или группам испытуемых — были бы предъявлены следующие шесть последовательностей: 50, 100 и 200 мс; 50, 200 и 100 мс; 100, 50 и 200 мс; 100, 200 и 50 мс; 200, 50 и 100 мс; 200, 100 и 50 мс. Мы не иллюстрируем полное уравнивание для большего числа уровней независимой переменной (обычно встречающегося в многоуровневых экспериментах) по той причине, что таблица оказалась бы слишком громоздкой. Например, для всех пяти уровней в исследовании Р. Готтсданкера и Т. Уэй

Группы испытуемых	Последовательности
1	АБВ
2	АВБ
3	БАВ
4	БВА
5	ВАБ
6	ВБА

потребовалось 120 последовательностей. Так что если бы даже только один испытуемый проводился через одну последовательность, то число испытуемых оказалось бы равным 120. Число последовательностей, необходимых для полного уравнивания, вычисляется как n -факториал, где n — число уровней. Для шести уровней n -факториал находится следующей серией умножений:

$$6 \times 5 \times 4 \times 3 \times 2 \times 1 = 720.$$

Поскольку кроссиндивидуальное уравнивание было введено для сокращения числа испытуемых по сравнению с их числом в межгрупповой схеме, полное позиционное уравнивание используется крайне редко. Нижеследующая схема позволяет сократить число испытуемых, избегая допущения об однородном переносе, необходимым для схемы реверсивного уравнивания.

Латинский квадрат

Если мы не хотим использовать все возможные последовательности, то естественно прийти к идее о случайном выборе из всего их множества. Иногда это и делается. Однако в случайно выбранном наборе последовательностей маловероятно, что каждый уровень окажется в каждой позиции равное число раз. Поэтому нежелательные последствия неоднородного переноса будут по-прежнему существовать.

Выходом будет случайный выбор среди «квадратов», в которых каждый уровень появляется *один* раз в каждой позиции. Каждый такой квадрат представляет собой полную экспериментальную схему. Он называется латинским квадратом. Приведем пример одного из 8640 таких квадратов для шести уровней независимой переменной.

Поскольку в латинском квадрате каждый уровень оказывается в каждой позиции последовательности, естественно, требуется столько групп испытуемых, сколько уровней независимой переменной.

Если бы Р. Готтсданкер и Т. Уэй использовали (как это им и следовало сделать) латинский квадрат вместо реверсивного уравнивания, их испытуемые должны были разбиться на пять групп соответственно пяти уровням независимой переменной. Значит, в их опыте должны были бы принять участие пять или десять испытуемых вместо восьми, как это было на самом деле (ведь восемь на пять не делится).

Группы испытуемых	Последовательности
1	АБВГДЕ
2	ВДГАЕБ
3	ДВАЕБГ
4	БГЕВАД
5	ГЕБДВА
6	ЕАДБГВ

Исследователи обычно вводят ограничение на латинский квадрат. Оно состоит в требовании, чтобы каждому уровню один раз непосредственно *предшествовал* каждый другой уровень. Такой квадрат называют *сбалансированным квадратом*. В приведенном выше латинском квадрате это условие не соблюдалось. Например, уровню Б только один раз предшествовали уровни А и Д, но три раза Е и ни разу В и Г. Метод получения сбалансированных квадратов приводится в работе У. Уагенаара (Wagenaar W. A., 1969). Вот пример:

Группы испытуемых	Последовательности
1	АБВГДЕ
2	БГАЕВД
3	ВАДБЕГ
4	ГЕБДАВ
5	ДВЕАГБ
6	ЕДГВБА

Если бы все эффекты переноса были связаны с непосредственно предшествующим уровнем, сбалансированный квадрат был бы очень эффективен. К сожалению, нет способа проверить, в действительности ли это так. Рассмотрим теперь систематические смещения (влияния последовательности), которые могут возникать даже при полном уравнивании.

Эффекты ряда

В многоуровневом эксперименте уровни независимой переменной образуют ряд — от наименьшего значения к наибольшему. При любой схеме уравнивания — интра- или кроссиндивидуальной — ответ на данный уровень независимой переменной может различаться в зависимости от того, какими были предшествующие ему уровни: более низкими, более высокими или смешанными.

Асимметричные эффекты. Об этих эффектах уже говорилось в главе 2 в связи с интраиндивидуальными схемами. Такого, например, влияние предшествующего опыта А на Б, но не наоборот. Эта идея может быть распространена на многоуровневые эксперименты с использованием кроссиндивидуального уравнивания. Предположим, имеется пять уровней независимой переменной и использована схема полного уравнивания (т.е. все 120 последовательностей). Поскольку каждому уровню один раз предшествовала каждая из возможных последовательностей остальных уровней, каждому уровню ни разу не предшествовали *идентичные*. В целом более низким уровням предшествовали более высокие уровни и наоборот. Например, самому низкому уровню не может предшествовать серия еще более низких уровней. Если имеется положительный перенос с меньших уровней на большие, но не

наоборот, то больше всего от этого пострадает уровень А. Таким образом, асимметричный перенос в многоуровневом эксперименте будет благоприятно или неблагоприятно влиять на уровни в зависимости от степени их удаления от концов всего ряда уровней.

Эффект центрации. Другой эффект ряда был продемонстрирован в эксперименте Дж. Е. Кеннеди и Дж. Ландесмана (Kennedy J. E., Landesman J., 1963). Они провели два эксперимента, каждый по схеме латинского квадрата с двумя группами испытуемых. Задачей была токарная обработка деталей, независимой переменной являлась высота рабочей поверхности. Диапазон уровней в одном эксперименте пересекался с диапазоном уровней в другом. Независимой переменной служила высота рабочей поверхности. Зависимой переменной было среднее число деталей, обработанных в течение 3-минутной пробы.

На рис. 7.7 отдельно для каждой группы показаны средние количества обработанных деталей. Интересно, что испытуемые в условии А, где наименьший уровень равнялся 45 см, обнаружили наибольшую продуктивность при 15 см, в то время как испытуемые в условии Б работали на этом уровне относительно плохо. Эта вторая группа, для которой наименьшим был уровень 25 см, показала наилучшие результаты при уровне -5 и +5 см.

В этом эксперименте, таким образом, наиболее благоприятными оказались уровни, близкие к *середине* ряда, а не к его краям.

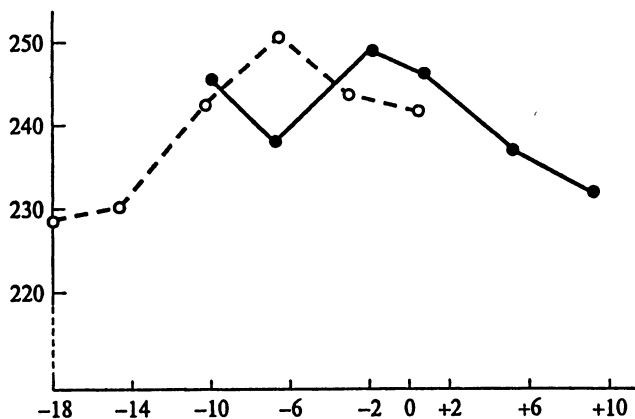


Рис. 7.7. Отношение между высотой рабочей поверхности и количеством обработанных деталей (Kennedy J. E., Landesman J., 1963). Ось абсцисс — высота рабочей поверхности (в дюймах, ниже (-) или выше (+) локтя); ось ординат — среднее количество обработанных деталей. Пунктирная линия — условие А, сплошная — условие Б

Это были как раз те единственные уровни, которым в последовательностях предшествовали как более низкие, так и более высокие уровни. Вы, конечно, можете сказать, что эти средние уровни казались для испытуемых «типичными» и поэтому наиболее удобными. Однако ваше объяснение имеет столько же оснований, сколько и мое. Ясно только одно: в этих опытах обнаружил себя эффект центрации.

Схемы полного позиционного уравнивания и латинского квадрата, в отличие от схемы реверсивного уравнивания, не требуют такого сильного допущения, как однородность переноса от одной позиции к следующей за ней. Однако в них сохраняется допущение, что отношение между настоящим и предшествующими уровнями не играет роли. Целый же ряд данных опровергает это (Poulton E. C., 1973). Оказывается, важно, какие уровни в основном предшествуют: более низкие, более высокие или смешанные.

Как быть?

При использовании кроссиндивидуального уравнивания прежде всего стоит избегать реверсивного уравнивания. Поскольку полное уравнивание, как правило, оказывается непрактичным, стоит обращаться к схеме латинского квадрата, особенно сбалансированного квадрата. Далее, для избежания отрицательного переноса из-за утомления необходимо разнести пробы во времени. Хорошо также разделить эксперимент на две части и использовать два перекрывающихся ряда уровней независимой переменной. Если впоследствии эффектов ряда не обнаружится, это будет хорошим показателем того, что удалось избежать смещения из-за влияния последовательности. Как мы увидим в следующей рубрике, в многоуровневых экспериментах кроссиндивидуальное уравнивание, действительно, имеет одно важное преимущество перед межгрупповыми схемами. Этот подход слишком хорош, чтобы быть оставленным только потому, что он никогда не приводит к безупречному эксперименту. Каковы возможности данного подхода?

МОЖЕМ ЛИ МЫ ДОВЕРЯТЬ КРИВЫМ?

Использование межгрупповых схем полностью исключит влияния или эффекты последовательности, которые мы только что обсуждали. Ведь каждому испытуемому предъявляется один уровень. Однако в многоуровневых экспериментах, которые направлены на проверку гипотез точного отношения между независимой и зависимой переменными, остаются другие угрозы внутренней валидности. Мы имеем в виду эксперименты, подобные исследованию С. Стернберга (Sternberg S., 1969), который проверял гипотезу «абсолютно-абсолютного» отношения между объемом по-

зитивного набора и временем мнемонического поиска, а также экспериментам У.Хика (Hick W.E., 1952), который проверял гипотезу «относительно-абсолютного» отношения между числом альтернатив и временем реакции. Вообще говоря, групповые схемы более уязвимы по отношению к первой из этих угроз, чем схемы, использующие принцип уравнивания.

Представимость индивида

На рис. 7.8, *а* представлены вымышленные данные, демонстрирующие отношение между независимой и зависимой переменными в схеме межгрупповых сравнений.

Каждая маленькая точка соответствует одному испытуемому. Среднее по каждому уровню обозначено большой точкой, а полу-

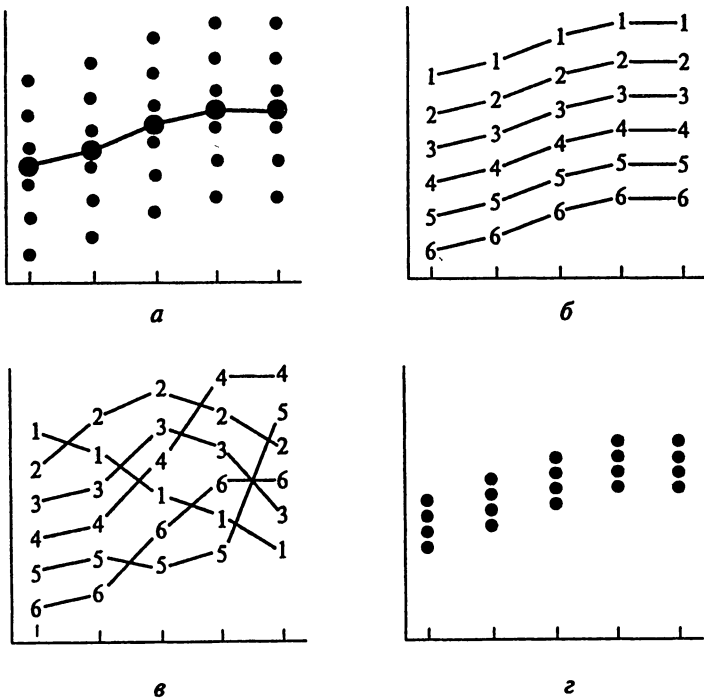


Рис. 7.8. Возможные соотношения усредненной кривой с данными идеального многоуровневого эксперимента, в котором каждому испытуемому одновременно предъявляются все уровни независимой переменной: *а* — индивидуальные данные и усредненная кривая; *б* — кривые по каждому испытуемому аналогичны усредненной кривой; *в* — данные по каждому испытуемому дают различные кривые; *г* — однородные группы испытуемых — высокая вероятность представительности усредненной кривой. Ось абсцисс — независимая переменная; ось ординат — ответы испытуемых

ченная кривая есть линия, соединяющая средние. Теперь посмотрим, как выглядели бы эти данные в идеальном эксперименте, где испытуемый проверялся бы одновременно по всем уровням.

На рис. 7.8, б представлен один возможный вид этих результатов для нескольких испытуемых. Одной цифрой обозначены результаты одного и того же испытуемого при различных уровнях независимой переменной. Линии, соединяющие ответы одного испытуемого, по форме очень похожи на линию, соединяющую средние на рис. 7.8, а. Конечно, возможен и другой вариант, когда линия, проходящая через средние, не обязательно так хорошо представляет все индивидуальные кривые, как это видно, например, на рис. 7.8, в. Когда межгрупповой эксперимент дает результаты, представленные на рис. 7.8, а, невозможно определить, какая из картин — (б) или (в) — имеет место в действительности. Из-за разброса индивидуальных данных в пределах одного уровня форма кривой оказывается неопределенной.

Существуют два способа уменьшения этой трудности при использовании межгрупповой схемы: подбор сходных испытуемых и использование однородных групп. Если испытуемых провести через предварительные испытания, подобрать их по одинаковым уровням показанных результатов и затем предъявить им различные уровни экспериментальной переменной, то вымышленные данные в виде наборов одинаковых цифр на рис. 7.8, б или рис. 7.8, в могут стать действительностью. Цифра 1 будет представлять одну уравненную группу испытуемых, 2 — другую группу и т.д. Тогда мы сможем непосредственно увидеть, какая картина верна — отражающая хорошее соответствие, как на рис. 7.8, б, или довольно хаотическая, как на рис. 7.8, в.

Второй способ основан на использовании одной, но очень однородной группы испытуемых, которая также может быть подобранной в предварительном эксперименте. Пример результатов такой группы приведен на рис. 7.8, г. Теперь уже практически не имеет никакого значения, через какие точки пройдут индивидуальные линии: форма кривых будет примерно одной и той же. Оба описанных метода можно объединить, используя только одну однородную группу и распределяя испытуемых по различным уровням независимой переменной.

В этом пункте может несколько обеспокоить возможное пристрастие экспериментатора при отборе в испытуемые одних индивидов и отвержении других. Однако содержательных выводов о связи исследуемого поведения с уровнем экспериментальной переменной это ни в коей мере не коснется. Конечно, они будут относиться лишь к небольшой части популяции. Однако далее будет уже вопрос обобщения, который можно легко решить, исследуя другие гомогенные группы с более высокими и более низкими уровнями результатов.

Если же вместо всего сказанного будет использовано кросс-индивидуальное уравнивание с предъявлением каждой из пяти последовательностей несколько испытуемым, то можно будет получить более ясную картину. Хотя кривую для каждой определенной последовательности нельзя будет «очистить» от зашумляющих влияний, последние будут одинаковыми для всех испытуемых, которым будет предъявлена эта последовательность. Если все индивидуальные кривые для данной последовательности имеют одинаковую форму, это является хорошим свидетельством того, что вся групповая кривая по всем последовательностям действительно представляет индивидуальные данные. Поскольку одному и тому же испытуемому предъявляется каждый уровень независимой переменной (хотя и не одновременно), кросс-индивидуальная схема больше приближается к идеальному эксперименту — именно в этом отношении, — чем межгрупповая схема. Она имеет лучшую внутреннюю валидность по параметру представленности индивида.

Нет ли искажений?

Если бы вы проводили эксперимент с целью определить, как влияет вес дротика на точность его метания, вы хотели бы быть уверены, что в ваши измерения не вкравлись ошибки. Если вы пользуетесь линейкой для измерения при каждом броске величины отклонения дротика от центра мишени, то, естественно, вам бы не хотелось, чтобы на вашей линейке расстояние между отметками 20 и 25 см было в три раза больше расстояния между 5 и 10 см. (Если бы это было так, вы скорее всего вернули бы линейку в магазин оборудования для фокусов.) Точно так же вы забраковали бы весы, стрелка которых едва отклоняется при помещении на них легкого дротика, но сразу же зашкаливает при чуть более тяжелом дротике. Вы хорошо знаете, что использование подобных искажающих измерительных устройств приведет к тому, что кривая, отражающая отношение между независимой переменной (весом дротика) и зависимой переменной (величиной ошибки попадания в цель), будет весьма неточной. Вообще говоря, может быть вы и обнаружите, что метание становится более точным по мере увеличения веса. Но вы не сможете проверить гипотезу об «абсолютно-абсолютном» отношении (например, что происходит уменьшение ошибки на 5 см с увеличением веса на 1 унцию).

Конечно, вы не собираетесь делать подобных ошибок в своих экспериментах. Однако существует два вида измерений, в которых нужно приложить особые усилия для избежания искажений. Во-первых, это измерения очень маленьких физических величин. Примером может служить регистрация кожно-гальванической реакции — изменений сопротивления кожи к электрическому току,

которые возникают, когда человек пугается или говорит неправду. Чтобы зарегистрировать реакцию, электрическое изменение должно быть усилено. Как мы можем быть уверены в том, что двойное увеличение амплитуды движения пера самописца означает двойное увеличение кожно-гальванической реакции? Обычно усилитель имеет максимальную чувствительность к определенной скорости нарастания или уменьшения тока. Если изменение нарастает либо быстрее, либо медленнее, оно уже не будет усиливаться в такой же пропорции. Итак, существуют такие области психологических исследований, где экспериментатор должен быть совершенно уверен в характеристиках измерительных приборов.

Проблемы искажения возникают и в тех случаях, когда используется психологическое шкалирование. Предположим, мы прошкалировали, как это было описано в одном из предшествующих разделов, шутки от «веселых» до «пустых», используя средние оценки-баллы, данные группой экспертов. Можем ли мы быть уверены в том, что различие в забавности между шутками, получившими оценку «2» и «4», такое же, как между шутками с оценкой «6» и «8»? Вероятно, нет. Следовательно, если бы мы проводили эксперимент для выяснения того, как влияет забавность шутки на ее запоминание, и проверяли бы какую-то точную гипотезу (например, что запоминаемость растет пропорционально росту забавности), мы не могли бы с уверенностью сказать, подтверждает форма кривой гипотезу или нет. Для правильного проведения такого эксперимента вы должны использовать более изощренные методы шкалирования, чем те, которые могут быть описаны в этой книге (см.: Torgerson W.S., 1958). Сейчас же вы должны запомнить, что содержательная интерпретация формы кривых, полученных с помощью субъективного шкалирования переменных, всегда требует доказательства того, что переменные не были искажены.

В идеальном эксперименте, направленном на проверку гипотезы о некотором точном количественном отношении, не должно быть искажений при измерении независимой и зависимой переменных. Однако в реальном эксперименте всегда есть некоторое искажение. Если искажение настолько велико, что отношение, найденное в действительном эксперименте, не представляет отношения, которое могло бы быть найдено в идеальном эксперименте, то внутренняя валидность существенно ослаблена.

Ранее в этой главе было показано, что для проверки любой количественной гипотезы — неважно, сформулирована она в количественных терминах или нет, — необходимо использовать достаточное число уровней независимой переменной. Слишком малое число уровней приводит к плохой представленности *отношения* между независимой и зависимой переменными. Внутренней валидности здесь угрожает не столько ненадежность или смеше-

ние, сколько неполнота независимой переменной. Было показано, что, во-первых, групповая кривая может не представлять индивидуальные кривые и, во-вторых, что искаженные результаты измерения будут давать ложное отношение. В обоих случаях отношение между независимой и зависимой переменными оказывается *невыявленным*. Теперь мы знаем три пути, которые могут угрожать внутренней валидности, три причины того, что результаты реального эксперимента могут плохо представлять отношение между независимой и зависимой переменными, которое могло бы быть обнаружено в идеальном эксперименте: 1) ненадежность, 2) систематическое смещение и 3) неверно найденное отношение.

КРАТКОЕ ИЗЛОЖЕНИЕ

Были рассмотрены три возможных двухуровневых эксперимента, которые оказались совершенно неадекватными по сравнению с аналогичным реально проведенным многоуровневым экспериментом. На этом примере были разобраны преимущества многоуровневого эксперимента.

Во-первых, он обеспечивает большую внутреннюю валидность, чем простые эксперименты, описанные в предыдущих главах. В некоторых экспериментах, где фигурируют только два уровня, независимая переменная по существу является количественной. Использование только двух уровней такой переменной может не выявить истинного отношения, которое могло бы быть найдено в идеальном количественном эксперименте, где используется неограниченное число уровней. Чем больше уровней независимой переменной мы используем, тем больше мы приближаемся к этому невозможному эксперименту и тем больше становится внутренняя валидность. Кроме того, лучше оказывается контроль за сопутствующим смещением. Если активный уровень независимой переменной сравнивается с ее нулевым (или неактивным) уровнем, это может косвенно ввести активный уровень вторичной переменной. В качестве примеров приводилось осознание действия лекарства и установление контакта с экспериментатором. В то же время ступенчатое изменение независимой переменной, приводящее к ступенчатому изменению зависимой переменной, делает маловероятным такое смещение.

Многоуровневые эксперименты превосходят более простые эксперименты и в другом отношении. В них могут проверяться гипотезы, которые ведут к более тонкому пониманию механизмов поведения. Во-первых, в сравнении с экспериментами, использующими качественные независимые переменные, здесь можно лучше выделить единичную переменную. Качественная переменная, такая как чтение (в отличие от прослушивания), может быть только нерасчлененным комплексом факторов. Возможностей для введения количественных переменных — сколько угодно, включая шкалирование поведения.

Дальнейшие преимущества определяются возможностью проверки более тонких гипотез об отношении между независимой и зависимой переменными. Часто наиболее правильной оказывается гипотеза о максималь-

ной (или минимальной) величине зависимой переменной при некотором промежуточном уровне независимой переменной. Она может следовать из теории двух процессов, связанных противоположным образом с уровнем независимой переменной. Например, может случиться, что негативный процесс берет верх над позитивным только на очень высоких уровнях независимой переменной. Именно так был проанализирован воображаемый многоуровневый эксперимент по трудовой этике. В качестве основных здесь были предположены «стремление к активности» и чувство «неприязни» к нажиманию на рычаг. Другим примером служила теория, согласно которой в основе поведения лежат два позитивных процесса, на которые увеличение уровня независимой переменной влияет противоположным образом. Именно так был проанализирован эксперимент по запоминанию списков студентами колледжа с варьированием интервалов между элементами запоминаемого списка и эксперимент, связывающий величину удара током с перцептивным различением у танцующих мышей. Переменные, лежащие в основе поведения в этих двух экспериментах, представляли собой различие стимулов и образование ассоциаций.

Многоуровневые независимые переменные позволяют проверять более *детализованные* экспериментальные гипотезы. Последние создаются на основе моделей и теорий, объясняющих, каким образом ступенчатые изменения независимой переменной приводят к изменению зависимой переменной. Так, на основе сканирующей модели мнемического поиска была выдвинута гипотеза о том, что одинаковые по абсолютной величине приросты объема запоминаемого материала будут сопровождаться примерно равными абсолютными приростами времени поиска: гипотеза «абсолютно-абсолютных» отношений. При исследовании связи между числом альтернатив и временем реакции проверялась гипотеза «относительно-абсолютного» отношения: при каждом увеличении количества альтернатив в одно и то же число раз будет наблюдаться увеличение времени реакции на одну и ту же абсолютную величину. Эта гипотеза была основана на модели наиболее эффективного способа принятия решения. На основе теоретического представления о том, как сенсорные органы превращают физическую энергию стимула в нервное возбуждение, была предсказана гипотеза «относительно-относительного» отношения между величиной поднимаемого веса и субъективным ощущением тяжести. Во всех перечисленных случаях результаты подтверждали гипотезу: при выборе на осях соответствующих шкал получалась линейная зависимость между независимой и зависимой переменными.

В многоуровневых экспериментах могут быть использованы и ранее описанные экспериментальные схемы. Для межгрупповой схемы существует практическая трудность: она состоит в необходимости привлекать слишком большое количество испытуемых. Внутрииндивидуальный контроль наиболее пригоден в случаях, когда предъявляются в случайном порядке короткие пробы на различных уровнях в большом наборе проб. Когда же каждая проба длительна, как это обычно и бывает при использовании внутрииндивидуального уравнивания, в многоуровневом эксперименте возникает практическая трудность: необходимость затраты слишком большого времени на каждого испытуемого.

Эти практические трудности можно преодолеть путем использования кроссиндивидуального реверсивного уравнивания. Однако эта конкретная

схема не обеспечивает контроль эффектов неоднородного переноса от предыдущей пробы к следующей. Такой контроль обеспечивает схема полного позиционного уравнивания, но она требует слишком большого числа различных последовательностей (и групп испытуемых), чтобы быть практически удобной. Типичным методом внутрииндивидуального уравнивания, который также контролирует неоднородный перенос, является латинский квадрат. В этой схеме каждый уровень независимой переменной появляется однажды в каждой позиции последовательности. Более тщательный контроль достигается путем использования только сбалансированных квадратов, в которых каждому уровню независимой переменной только один раз предшествует каждый из остальных уровней.

И все же ни одна схема кросс-индивидуального уравнивания не обеспечивает контроль эффектов ряда. В любой последовательности низким уровням чаще предшествуют более высокие, чем более низкие, уровни, а высоким — низкие, и это порождает угрозу асимметричного переноса. Другим описанным эффектом ряда является эффект центрации. Он возникает в связи с тем, что только уровням, близким к середине ряда, могут в равной мере предшествовать и высокие и низкие уровни. Более благоприятное положение средних уровней было показано в эксперименте с обработкой деталей.

Был рассмотрен ряд полезных советов. Так, при кросс-индивидуальной схеме вместо реверсивного уравнивания лучше использовать латинский квадрат; для того чтобы избежать влияния утомления, необходимо давать достаточный отдых между пробами; и, наконец, для контроля за эффектами ряда нужно использовать перекрывающиеся диапазоны уровней независимой переменной.

При использовании любых описанных ранее экспериментальных схем при проверке точных гипотез все-таки остаются две угрозы внутренней валидности. Одна из них состоит в том, что форма кривой, полученной на группе испытуемых, может не представлять индивидуальные кривые ни одного испытуемого. В идеальном эксперименте один и тот же испытуемый должен был бы проверяться одновременно по всем уровням. Поэтому возможность неверной представленности истинного отношения в полученном отношении является источником внутренней невалидности. Особенно подвержены такой опасности межгрупповые схемы. Опасность может быть уменьшена благодаря уравниванию испытуемых и использованию однородных групп.

Другим источником неверной представленности отношения между независимой и зависимой переменными может быть искажающее действие измерительных приборов и шкал, с помощью которых измеряют зависимую и независимую переменные. С наибольшей вероятностью такое искажение возникает в двух типах измерения. Первый случай — когда необходимо усиление малых физических величин, второй — когда используется субъективное шкалирование.

В предыдущих главах в качестве угроз внутренней валидности описывались ненадежность и систематическое смещение. В этой главе была показана новая угроза — неверно установленное отношение между независимой и зависимой переменными. Оно может быть следствием использования усредненных кривых, которые не представляют индивидуальные, за-

тем — применения слишком малого числа уровней независимой переменной, наконец, проведения неверных измерений.

Контрольные вопросы и задания

1. Чем многоуровневый эксперимент отличается от экспериментов, описанных в предыдущих главах?
2. Что означает утверждение, что многоуровневые эксперименты обеспечивают контроль для проверки экспериментальных гипотез, которые могли бы быть проверены и в двухуровневом эксперименте?
3. Сравните с теоретической точки зрения результаты эксперимента с количественным изменением независимой переменной и эксперимента с условиями, отличающимися только качественно.
4. Что подразумевается под экспериментальной гипотезой максимума или минимума?
5. Почему к эксперименту Стернберга по исследованию памяти приложим термин «абсолютно-абсолютного» отношения? Что лежало в основе этой экспериментальной гипотезы?
6. Определите различие между количественными экспериментальными гипотезами У. Хика (Hick W. E., 1952) о времени реакции и Р. Харпера и С. Стивенса (Harper R. S., Stevens S. S., 1948) о субъективной тяжести.
7. Каковы практические причины использования позиционного уравнивания по всем испытуемым, а не межгрупповой схемы или интраиндивидуального позиционного уравнивания?
8. Что такое латинский квадрат?
9. Может ли предохранить полное позиционное уравнивание от эффектов неоднородного переноса? От эффектов ряда?
10. Какие угрозы внутренней валидности остаются при использовании любых схем проверки гипотезы точного отношения между независимой и зависимой переменными?
11. Понятие идеального эксперимента было вновь введено в связи с угрозой внутренней валидности, отличающейся от ненадежности и систематического смещения. Как это было сделано? Как бы вы в таком случае определили внутреннюю валидность?

СТАТИСТИЧЕСКОЕ ПРИЛОЖЕНИЕ: ОДНОФАКТОРНЫЙ ДИСПЕРСИОННЫЙ АНАЛИЗ И F -КРИТЕРИЙ

Нельзя t -критерий использовать для обнаружения общего действия независимой переменной в многоуровневом эксперименте. Его можно использовать только для проверки различия между средними значениями двух условий. Для того чтобы определить, отличаются ли в целом друг от друга различные уровни, требуется несколько иной подход и другой статистический критерий. Такой подход называют *дисперсионным анализом*; статистическая значимость оценивается F -критерием. Поскольку мы имеем дело с един-

ственной независимой переменной, мы называем анализ *однофакторным*. В статистическом приложении к следующей главе, где будут рассматриваться эксперименты с двумя независимыми переменными, описана техника двухфакторного дисперсионного анализа.

Две оценки $\bar{\sigma}_x^2$

Рассмотрим снова эксперимент по измерению времени реакции, в котором использовались четыре группы испытуемых. Испытуемый дает ответ на звуковой тон; независимой переменной является громкость тона (или, вернее, звуковое давление). Используются четыре уровня звукового давления: 10, 30, 50 и 70 дБ. В каждой группе 17 испытуемых, и для каждого испытуемого определяется среднее время реакции.

Предположим, нуль-гипотеза верна. Тогда в бесконечном эксперименте, т. е. для неограниченного числа тестируемых по каждому уровню испытуемых, мы имели бы всегда одинаковые величины для $\bar{M}_1, \bar{M}_2, \bar{M}_3, \bar{M}_4$. Хотя, конечно же, среднее время реакции для различных испытуемых, которым предъявляется одно и то же условие, было бы различным.

Мы можем сделать две оценки параметра — $\bar{\sigma}_x^2$ по данным нашего эксперимента, снова допуская нуль-гипотезу $\bar{M}_1 = \bar{M}_2 = \bar{M}_3 = \bar{M}_4$. Одна из оценок основана на учете вариаций времени реакции среди испытуемых по всем уровням. *Внутригрупповая* вариация представляет собой просто объединение вариаций по всем уровням. Другая оценка определяет, насколько отдельные групповые средние отличаются от *общего* среднего эксперимента $M_{1+2+3+4}$. Таким образом, существует *внутригрупповая* оценка $\bar{\sigma}_x^2$ и *межгрупповая* оценка $\bar{\sigma}_x^2$.

Выборочное распределение *F*-критерия

Если верна нуль-гипотеза, то при *достаточно длинной выборке* оценки $\bar{\sigma}_x^2$ должны быть идентичны. В бесконечном эксперименте средняя оценка по межгрупповой вариации будет равна средней оценке по внутригрупповой вариации. В каждом отдельном эксперименте, включая рассматриваемый здесь эксперимент, мы не должны ожидать точного совпадения этих оценок. В одном эксперименте две эти оценки могут быть больше похожи, в другом — меньше. Когда две величины идентичны, их отношение равно 1:

$$\frac{\text{Межгрупповая оценка } \bar{\sigma}_x^2}{\text{Внутригрупповая оценка } \bar{\sigma}_x^2} = 1.$$

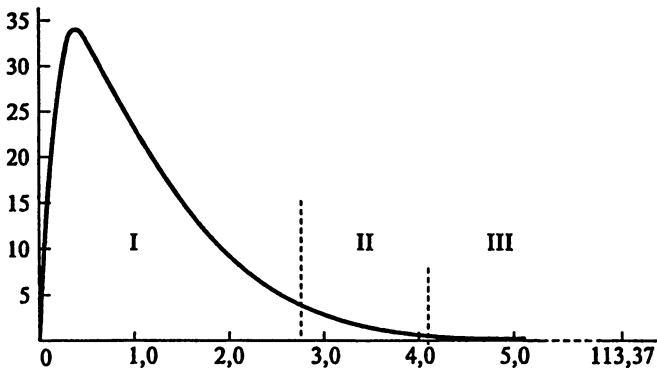


Рис. 7.9. Ось абсцисс — F -отношение; ось ординат — относительная частота. I — область принятия нуль-гипотезы; II — область отвержения с $p = 0,05$; III — область отвержения с $p = 0,01$

Это отношение обозначается как F . В вышеприведенном выражении показан случай, когда $F = 1$. Если нулевая гипотеза неверна, разность между средними для различных уровней будет намного больше, чем та, которую можно было бы объяснить несистематической вариацией данных. Межгрупповая оценка будет больше, чем внутригрупповая оценка; F будет больше 1.

Однако можно ожидать, что отношение F от эксперимента к эксперименту будет отличаться от 1, даже если средняя величина равна 1 (как это предполагается нуль-гипотезой). Распределение величин F в бесконечном ряду экспериментов при допущении верности нуль-гипотезы является еще одним *выборочным распределением*. Это распределение можно представить так же, как распределение для t .

Для примера приводится рис. 7.9.

Вопрос состоит в том, превышает ли полученная в некотором эксперименте величина F критическое значение, соответствующее выбранному альфа-уровню, обычно 0,05 или 0,01. Другими словами, мы отвергнем нулевую гипотезу только если вероятность того, что полученная нами величина F могла бы появиться при правильности нулевой гипотезы, достаточно мала. Для этого наша F должна быть, конечно, больше 1, причем тем больше, чем меньше число испытуемых (или число проб) и чем больше несистематическая вариация.

Нахождение величины F

Давайте сделаем таблицу, показывающую, какие показатели необходимы для вычисления F .

Показатель	Уровень звука			
	1	2	3	4
M_x	M_1	M_2	M_3	M_4
$\sum x^2$	$\sum x_1^2$	$\sum x_2^2$	$\sum x_3^2$	$\sum x_4^2$
n	n_1	n_2	n_3	n_4

Поскольку мы уже делали некоторые вычисления по четырем группам данных, давайте предположим, что они были получены и в эксперименте, где исследовалось влияние уровня громкости на время реакции. Назовем условие В уровнем 1, условие Г — уровнем 2, условие А — уровнем 3, условие Б — уровнем 4. Это избавит нас от большого числа вычислений. Кроме того, это даст нам уменьшение среднего времени реакции с увеличением громкости — как и должно быть. Таким образом, главные показатели нами уже вычислены (см. гл. 6).

Показатель	Уровень звука			
	1	2	3	4
M_x	265	250	185	162
$\sum x^2$	4673	5391	5808	4306
n	17	17	17	17

Сумма квадратов для отдельной группы. Внутригрупповая (ВГ) сумма квадратов (СК) будет использована для определения оценки $\bar{\sigma}_x^2$ внутри группы. Она находится простым сложением членов $\sum x^2$ по строке, поэтому

$$СК_{ВГ} = \sum x_1^2 + \sum x_2^2 + \sum x_3^2 + \sum x_4^2. \quad (7.1)$$

Здесь

$$СК_{ВГ} = 4673 + 5391 + 5808 + 4306 = 20178.$$

Сумма квадратов между группами. Межгрупповая сумма квадратов будет использована при определении оценки между группами. Для того чтобы найти ее, вы сначала вычисляете общее («общ») среднее для четырех условий:

$$M_{\text{общ}} = \frac{M_1 + M_2 + M_3 + M_4}{k}, \quad (7.2)$$

где k — число групп.

Здесь

$$M_{\text{общ}} = \frac{265 + 250 + 185 + 162}{4} = \frac{862}{4} = 215,5.$$

Затем ищется разность между каждым отдельным средним и общим средним. Такие разности обозначаются буквой d . Так,

$$d_1 = M_1 - M_{\text{общ}}, \quad d_2 = M_2 - M_{\text{общ}} \quad (7.3)$$

Для числовых данных:

$$d_1 = 265 - 215,5 = +49,5; \quad d_2 = 250 - 215,5 = +34,5; \\ d_3 = 185 - 215,5 = -30,5; \quad d_4 = 162 - 215,5 = -53,5.$$

Межгрупповая (МГ) сумма квадратов — это просто сумма квадратов величин d , умноженная на число случаев (n) по данному условию:

$$СК_{\text{МГ}} = n(d_1^2 + d_2^2 + d_3^2 + d_4^2). \quad (7.4)$$

Для числовых данных:

$$СК_{\text{МГ}} = 17(2450,25 + 1190,25 + 930,25 + 2862,25) = 17(7433) = 126361.$$

Внутригрупповое среднее квадратичное (СКВ_{ВГ}). Оценка $\bar{\sigma}_x^2$, основанная на внутригрупповой вариации, называется *внутригрупповым средним квадратичным*. Она находится делением суммы квадратов внутри групп на сумму степеней свободы для средних всех групп. Так, она равняется $(n_1 - 1) + (n_2 - 1) + (n_3 - 1), \dots$

Поскольку мы имеем k условий и N испытуемых в целом,

$$df_{\text{ВГ}} = N - k. \quad (7.5)$$

Для нашего эксперимента

$$df_{\text{ВГ}} = 68 - 4 = 64.$$

Как уже говорилось,

$$СКВ_{\text{ВГ}} = \frac{СК_{\text{ВГ}}}{N - k}. \quad (7.6)$$

Для наших данных

$$СКВ_{\text{ВГ}} = \frac{20178}{64} = 315$$

Межгрупповое среднее квадратичное. Оценка $\bar{\sigma}_x^2$, основанная на межгрупповой вариации, называется *межгрупповым средним квадратичным* (СКВ_{МГ}). Она находится делением межгрупповой

суммы квадратов на число степеней свободы для общего среднего, вычисленного из средних для различных условий:

$$df_{\text{МГ}} = k - 1. \quad (7.7)$$

А для числовых данных $df_{\text{МГ}} = 4 - 1 = 3$.

Как уже говорилось,

$$\text{СКВ}_{\text{МГ}} = \frac{\text{СК}_{\text{МГ}}}{df_{\text{МГ}}}, \quad (7.8)$$

или

$$\text{СКВ}_{\text{МГ}} = \frac{126361}{3} = 42120.$$

***F*-отношение.** Последний шаг в вычислении — *F*-деление межгруппового среднего квадратичного на внутригрупповое среднее квадратичное. Вспомните, что чем больше это отношение, тем более вероятно, что нуль-гипотеза может быть отвергнута:

$$F = \frac{\text{СКВ}_{\text{МГ}}}{\text{СКВ}_{\text{вг}}}, \quad (7.9)$$

или

$$F = \frac{42120}{315} = 133,71.$$

Отвержение или принятие нуль-гипотезы

На графике *F*-распределения, приведенном в начале данного статистического приложения, полученная нами величина *F* оказывается расположенной далеко справа. Очевидно, что если бы была верна нулевая гипотеза, то такое большое *F*-отношение должно получаться крайне редко, ведь в бесконечном ряду экспериментов отношение равнялось бы 1. Мы должны обеспечить уверенность, что имеем право отвергнуть нуль-гипотезу, найдя критическую величину в Статистической таблице 3 в конце данного приложения.

Поскольку распределение будет иметь различную форму в зависимости от числа степеней свободы в числителе и знаменателе, таблица разделена на несколько вертикальных столбцов и множество горизонтальных строк. Каждый *столбец* содержит критические величины *F* для альфа-уровня 0,05 и 0,01 при определенном числе степеней свободы в *числителе F*-отношения. Каждая строка показывает то же самое для определенного числа степеней свободы в знаменателе.

Критические значения F для отвержения нуль-гипотезы (верхнее число для $\alpha = 0,05$, а нижнее для $\alpha = 0,01$)

Степени свободы для знаменателя	Степени свободы для числителя									
	1	2	3	4	5	6	7	8	9	10
1	161	200	216	225	230	234	237	239	241	242
2	4052	4999	5403	5625	5764	5859	5928	5981	6022	6056
3	18,51	19,00	19,16	19,25	19,30	19,33	19,36	19,37	19,38	19,39
4	98,49	99,01	99,17	99,25	99,30	99,33	99,34	99,36	99,38	99,40
5	10,13	9,55	9,28	9,12	9,01	8,94	8,88	8,84	8,81	8,78
6	34,12	30,81	29,46	28,71	28,24	27,91	27,67	27,49	27,34	27,23
7	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96
8	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,54
9	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74
10	16,26	13,27	12,06	11,39	10,97	10,67	10,45	10,27	10,15	10,05
11	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06
12	13,74	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87
13	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63
14	12,25	9,55	8,45	7,85	7,46	7,19	7,00	6,84	6,71	6,62
15	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34

Степени свободы для знаменателя	Степени свободы для числителя									
	1	2	3	4	5	6	7	8	9	10
9	11,26	8,65	7,59	7,01	6,63	6,37	6,19	6,03	5,91	5,82
	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13
10	10,56	8,02	6,99	6,42	6,06	5,80	5,62	5,47	5,35	5,26
	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97
11	10,04	7,56	6,55	5,99	5,64	5,39	5,21	5,06	4,95	4,85
	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86
12	9,65	7,20	6,22	5,67	5,32	5,07	4,88	4,74	4,63	4,54
	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76
13	9,33	6,93	5,95	5,41	5,06	4,82	4,65	4,50	4,39	4,30
	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67
14	9,97	6,70	5,74	5,20	4,86	4,62	4,44	4,30	4,19	4,10
	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60
15	8,86	6,51	5,56	5,03	4,69	4,46	4,28	4,14	4,03	3,94
	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55
16	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80
	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,40
	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69

17	4,45	3,69	3,20	2,96	2,81	2,70	2,62	2,55	2,50	2,45
18	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59
19	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41
20	8,28	6,01	5,09	4,58	4,25	4,01	3,85	3,71	3,60	3,51
21	4,38	3,52	3,13	2,90	2,74	2,63	2,55	2,48	2,43	2,38
22	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43
23	4,35	3,49	3,10	2,87	2,71	2,60	2,52	2,45	2,40	2,35
24	8,10	5,85	4,94	4,43	4,10	3,87	3,71	3,56	3,45	3,37
25	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32
26	8,02	5,78	4,87	4,37	4,04	3,81	3,65	3,51	3,40	3,31
27	4,30	3,44	3,05	2,82	2,66	2,55	2,47	2,40	2,35	2,30
28	7,94	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26
29	4,28	3,42	3,03	2,80	2,64	2,53	2,45	2,38	2,32	2,28
30	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21
31	4,26	3,40	3,01	2,78	2,62	2,51	2,43	2,36	2,30	2,26
32	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,25	3,17
33	4,24	3,38	2,99	2,76	2,60	2,49	2,41	2,34	2,28	2,24
34	7,77	5,57	4,68	4,18	3,86	3,63	3,46	3,32	3,21	3,13
35	4,22	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22
36	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,17	3,09

Используя Статистическую таблицу 3 для нашего $F = 133,71$ с $df = 3$ в числителе и $df = 64$ в знаменателе, мы обращаемся к столбцу 3 и строке 65, наиболее близкой к 64. Величина 2,75 показывает значение F , требуемое для отвержения нулевой гипотезы на уровне 0,05; величина 4,10 показывает значение, требуемое для отвержения нуль-гипотезы на уровне 0,01. Этим уровням соответствуют линии, приведенные на графике распределения F . Область отношений отвержения нуль-гипотезы для каждого из этих альфа-уровней лежит справа от каждой линии. Конечно, нет необходимости рисовать распределение, когда мы можем использовать таблицу критических величин. Для наших числовых данных мы можем утверждать, что $p < 0,01$.

Таблица дисперсионного анализа

Только что описанный метод называют *дисперсионным анализом* (или ANOVA при вычислениях на ЭВМ). По существу, все дисперсии данных уже были проанализированы по частям. Вы могли бы вычесть общее среднее из величины реакции, полученной для каждого испытуемого, и возвести в квадрат 68 разностей. Их сложение дает общую сумму квадратов ($СК_{общ}$). Теперь, если вы сложите вместе сумму квадратов внутри групп и сумму квадратов между группами и не сделаете ошибок, эта сумма тоже будет равняться общей сумме квадратов ($СК_{общ}$).

Представлять результаты дисперсионного анализа принято в виде таблицы сумм квадратов и средних квадратичных. Вот как мы могли бы представить наши данные.

Дисперсионный анализ

Эксперимент по исследованию зависимости между громкостью стимула и временем реакции

Источник дисперсии	СК	df	СКВ	F	p
Между уровнями громкости	126361	3	42120	133,71	<0,01
Внутри уровней громкости	20178	64	315		
Общая	146539	67			

Задача. Проведите дисперсионный анализ на основании следующих данных, соотносящих число решенных проблем с величиной денежной награды. Завершите анализ дисперсионной таблицей. Данные получены на различных группах испытуемых.

Награда (от меньшей к большей)

Уровень 1	Уровень 2	Уровень 3	Уровень 4	Уровень 5	Уровень 6
10	8	12	12	24	19
11	10	17	15	16	18
9	16	14	16	22	27
13	13	9	16	18	25
7	12	16	19	20	24
Ответ					

Источник дисперсии	СК	<i>df</i>	СКВ	<i>F</i>	<i>p</i>
Между уровнями	590,8	5	118,16	12,64	<0,01
Внутри уровней	224,4	24	9,35		
Общая	815	29			

Глава 8

ФАКТОРНЫЕ ЭКСПЕРИМЕНТЫ¹

В исследовании Д.Гаффана (Gaffan D., 1974) проводились опыты на шести обезьянах (резусах) с поперечным рассечением свода и — для контроля — на шести других, также оперированных, но без рассечения (см. гл. 5). Проверяемая экспериментальная гипотеза была строго определенной: поражение области гиппокампа приводит к нарушению *узнавания*. Подтверждение этой гипотезы могло бы внести значительный вклад в понимание природы амнезии — довольно частого следствия мозговых расстройств в результате несчастных случаев.

Обратите внимание, насколько специфична приведенная гипотеза. В результате воздействия нарушается не что-нибудь, а *память*, причем не любой из ее видов, а именно *узнавание*. Чтобы выделить в эксперименте только этот результат действия независимой переменной, нужны хорошо продуманные способы контроля. Давайте посмотрим, как проводился эксперимент.

А проводился он не сразу. В течение двух недель обезьяны поправлялись после операции. Затем их обучали одному из вариантов старинной карнавальная игры — «найди орешек» (правда, без фокусов, связанных с ловкостью рук). Экспериментальный вариант этой игры называется *подбором по образцу*. Вот в чем он заключается. Перед обезьяной ставят поднос с тремя расположенными в ряд ячейками. Первая проба служит *образцом*: средняя ячейка прикрыта небольшим предметом, например деревянной лодочкой. Обезьяна поднимает предмет и находит под ним сладкую воздушную кукурузу (сушеные зерна или хлопья в сахаре) — вот и все. Через 10 с дают пробу на *подбор*. Теперь на подносе прикрыты две боковые ячейки, а в средней ничего нет. На одной ячейке лежит лодочка (т.е. предмет, предьявленный в образцовой пробе), а на другой что-нибудь еще, скажем, игрушечный телефон. Если обезьяна поднимет лодочку — и тем самым сделает правильный выбор, — она обнаружит в ячейке еще больше сладкой кукурузы. А если она возьмет телефон, то ничего не найдет — ошибка.

В результате обучения каждая обезьяна выбирала предмет правильно в 81 случае из 90. Это отвечало *критерию* успешности решения задач, установленному экспериментатором. Обучение за-

¹ Перевод Ч.А. Измайлова, В.В. Петухова.

нимало три дня. В распоряжении экспериментатора было 300 наименований всякого «утильсырья» — игрушки, миски, электрические переключатели и т.п. Предметов было так много и они настолько отличались друг от друга, что экспериментатор мог работать без повторений в течение пяти дней. Само собой разумеется, что местоположение предмета (левая или правая ячейка) изменялось случайным образом. Чтобы научиться выбирать нужный предмет, обеим группам обезьян потребовалось примерно одинаковое время. Все обезьяны достигли установленного критерия успешности работы между 330 и 600 парой проб (образец — выбор). Следует отметить, что в течение всего эксперимента обезьяны получали обычный рацион — сладкое лакомство было добавкой. Одна из задач называлась задачей с *отсрочкой*. В предварительном обучении интервал между пробой образцов и пробой-выбором всегда был равен 10 с. Теперь же использовали три разных интервала: 10, 70 и 130 с. Иначе говоря, при самой длинной отсрочке животное должно было помнить предмет, закрывавший приманку, более 2 мин. Полученные результаты представлены на рис. 8.1. Легко убедиться, что при отсрочке в 10 с результаты обезьян с рассеченным сводом и контрольных животных примерно одинаковы. И удивляться тут, конечно, нечему. Ведь это фактически та же практическая задача, которую обе группы обезьян научились решать с равным успехом. Однако с увеличением интервала группы разделились. Когда вторая проба предъявлялась спустя 130 с, контрольные животные по-прежнему давали свыше 90 % правильных ответов, а результаты по группе с рассеченным сводом снижались до 65 %. А это ненамного выше 50%-ного уровня, которого можно достичь случайно. Успешное выполнение задачи с отсрочкой в 10 с свидетельствует о том, что у обезьян с рассеченным сводом процесс *запечатления* сохранен. Иначе говоря, сразу же после предъявления образца обезьяны знали, какой предмет нужно выбрать. Однако контрольная группа продолжала помнить об этом и дальше, а экспериментальная — забывала. Таким образом, было показано, что у животных с рассеченным сводом нарушается не запечатление, а сохранение. Если бы проверочная проба давалась только с отсрочкой в 130 с, исключить воз-

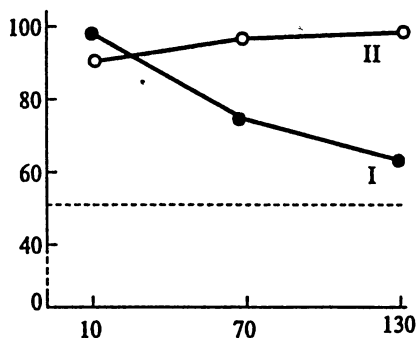


Рис. 8.1. Результаты решения задач с отсрочкой обезьянами с рассеченным (I) и с интактным (II) сводом. Ось абсцисс — интервал отсрочки (в с); ось ординат — проценты правильных ответов (пунктирная линия — результат случайного выбора)

возможность нарушения процесса запечатления было бы нельзя. Для того чтобы отделить его от нарушения собственно памяти, нужно было добавить к основной независимой переменной — *состоянию свода* — еще одну. *Интервал отсрочки* и был второй независимой переменной. Теперь осталось показать, что нарушен именно процесс *узнавания*. О решении этой проблемы мы расскажем в следующем разделе.

В любом эксперименте из предыдущих глав фигурировала только одна независимая (и одна зависимая) переменная. В приведенном эксперименте Д. Гаффана независимая переменная, влияющая на зависимую переменную, была не единственной. Различные состояния свода дополнились разными интервалами отсрочки. Без изменения интервалов, например только при длительной отсрочке, было бы невозможно определить, что же в действительности нарушено — память или восприятие. Мы видим, что привлечение второй независимой переменной позволяет осуществить контроль именно того результата действия первой переменной, которое интересует исследователя. Таково одно из оснований для экспериментов с двумя независимыми переменными.

Мы убедимся, что другим существенным преимуществом этих экспериментов является возможность проверки более сложных, комплексных гипотез, чем те, с которыми мы встречались раньше. Это гипотезы о том, каким именно образом независимые переменные, сочетаясь друг с другом, влияют на изучаемое поведение. Мы будем называть их *комбинированными* гипотезами. Если в эксперименте задействовано несколько (по крайней мере — две) независимых переменных, он называется *факторным*. Термин «факторный» означает только то, что каждая из независимых переменных может быть фактором, определяющим поведение. А поскольку, как мы знаем, поведение определяется многими факторами, то проверка комбинированных гипотез открывает нам новые возможности изучения его природы.

Вы наверняка спросите, существуют ли эксперименты с несколькими *зависимыми* переменными, в которых можно найти связь между независимой переменной и каждой из зависимых? Да, существуют. Есть и такие эксперименты, где и зависимых и независимых переменных несколько. Однако для психологии они являются совершенно новыми, и для их полного понимания наших знаний по статистике недостаточно. Эти перспективные эксперименты называются *многомерными*. Возможно, в будущем кому-нибудь из вас придется их освоить. Однако сейчас подавляющее большинство психологических экспериментов не выходит за рамки факторных.

Для обсуждения организации этих экспериментов привлекается понятие *факторной схемы*. Узнав, что это такое, вы получите новый параметр для классификации экспериментальных схем.

В конце этой главы мы приведем общую классификацию экспериментальных схем, благодаря которой вы сможете хорошо понять и усвоить их разнообразные варианты.

Знакомясь с экспериментальными работами, написанными в последнее время, вы убедитесь, что факторную схему эксперимента применяют почти все. Для полноценного понимания этих работ вам нужно иметь представление о соответствующих понятиях и процедурах. Прочитав эту главу, вы узнаете, что имеют в виду, когда говорят об «*основных результатах действия*» и «*взаимодействиях*». Весьма вероятно, что вы встретитесь с такими исследованиями, которые можно было бы улучшить, привлекая вторую независимую переменную. Ничто не мешает вам применить факторную схему и в собственных экспериментах. Хотя, конечно, чтобы держать в уме сразу две независимые переменные, нужно совершать определенное усилие. Но это — все, что от вас потребуется: в своей книге мы ограничимся экспериментами, где независимых переменных не больше двух. К счастью, некоторые исследователи пытаются проводить эксперименты с тремя и более независимыми переменными, и вы сможете научиться у них, если будете и в дальнейшем заниматься экспериментальной психологией.

Вот те главные темы, по которым будут заданы вопросы в конце главы.

1. Основные понятия: *факторный эксперимент, основные результаты действия, взаимодействие.*

2. Факторные эксперименты как способы контроля при проверке гипотез с одним отношением.

3. Проверка комбинированных гипотез: ожидаемые взаимодействия.

4. Классификация экспериментальных схем.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ И ВЗАИМОДЕЙСТВИЯ

Этот раздел мы начнем с описания еще одного эксперимента Д. Гаффана (точнее, даже двух экспериментов — на разных задачах). На конкретных примерах мы покажем, что такое *основные результаты действия* независимых переменных и *взаимодействия* между ними. А в следующем разделе рассмотрим тот общий способ контроля, благодаря которому в данном эксперименте можно было выделить изучаемые результаты действия.

Экспериментальное разделение способностей к узнаванию и к образованию ассоциаций

Экспериментаторы хотели показать, что с рассечением свода нарушается именно узнавание, а не запоминание по ассоциации.

Для этого они использовали две разные задачи. Одной из них была *задача на ассоциацию*. Каждая такая задача включала в себя серию ознакомительных проб и тест на удержание предъявленного материала. Ознакомительных проб было десять, в каждой предъявлялся какой-либо предмет. В пяти пробах под этим предметом была спрятана сладкая кукуруза, а в пяти других ячейка была пустой. Сразу же после этого давался проверочный тест. Предметы поочередно (в том же порядке, что и при ознакомлении), с интервалом в 20 с помещали над правой ячейкой. Левая же ячейка была прикрыта медным диском, который в ознакомительных пробах ни разу не применялся. Если предмет был одним из тех пяти, которые при первом предъявлении закрывали приманку, то и теперь под ним снова находилась сладкая кукуруза. Если же в ознакомительных пробах предмет закрывал пустую ячейку, то приманку клали под медный диск. Таким образом, чтобы выполнить задачу правильно, нужно было взять предмет, если тот ранее прикрывал приманку, а если нет — поднять диск. Ежедневно с каждым животным проводили по три серии проб. На пятый (и последний) день эксперимента обе группы справились с задачей одинаково успешно, давая свыше 80 % правильных решений.

Во второй задаче проверялась *способность к узнаванию*. Во многом она напоминала первую, но их отличие было принципиальным. Как и раньше, в ознакомительных пробах предъявляли пять предметов с подкреплением, а вот «пустых» проб уже не было. Вместо этого при проверке давали пять совершенно новых предметов. Правильное решение было тем же: если предмет был предъявлен в ознакомительных пробах, нужно снять его с правой ячейки. А если предмет никогда раньше не предъявлялся, приманка находилась под медным диском в левой ячейке. В данном эксперименте между двумя группами обезьян обнаружилось значительное различие. Контрольная группа решала задачи на узнавание так же успешно, как и ассоциативные. Группа же обезьян с рассеченным сводом давала лишь около 60 % правильных решений, т. е. результат оказался значительно ниже прежнего 80%-ного уровня.

Для правильного выполнения задач на ассоциацию нужно было запомнить, как связан каждый из предварительно предъявленных предметов с наличием или отсутствием приманки. А для успешного решения второй задачи этого было мало, поскольку здесь «неподкрепленные» предметы были совершенно новыми. В таком случае нужно суметь определить, предъявлялся ли предмет ранее, т. е. опознать его.

Было обнаружено, что у обезьян с рассеченным сводом образование ассоциаций полностью сохранно, а способность к узнаванию нарушена.

Измерение основных результатов действия

В только что описанном факторном эксперименте мы можем точно определить результаты действия каждой из двух независимых переменных. Это несложно, поскольку и та и другая переменные имеют только по два уровня, или условия. Основным результатом действия переменной — это разность между общими средними оценками для двух ее уровней. По группе контрольных животных с интактным сводом среднее количество правильных решений задач на ассоциацию (с предварительным предъявлением неподкрепленных предметов) составило 83 %, а задач на узнавание (без предварительного предъявления) — 88 %. Общая средняя оценка для этой группы равна 85,5 %. Те же средние оценки по группе обезьян с рассеченным сводом равны соответственно 82 и 62 %, общая средняя — 72 %. Основным результатом действия первой независимой переменной — состояния свода — это разность общих средних оценок, 13,5 %. В дальнейшем обсуждении мы всегда будем брать только по два уровня каждой независимой переменной, так будет легче их проследить. Впрочем, в многоуровневых экспериментах основные результаты действия переменных определяются тем же принципиальным способом, хотя вычисления становятся сложнее.

Таблица 8.1

Вычисление основных результатов действия переменных состояния свода и предварительного предъявления неподкрепленных предметов и взаимодействие между ними

Предварительное предъявление неподкрепленных предметов	Состояние свода		
	рассеченный	интактный	среднее
Есть	82	83	82,5
Нет	62	88	75,0
Среднее	72,0	85,5	76,25

Вычисления.

Основной результат действия: состояние свода:

$$\text{среднее}_{\text{интактный}} - \text{среднее}_{\text{рассеченный}}$$

$$85,5 - 72,0 = 13,5.$$

Основной результат действия: предварительное предъявление неподкрепленных предметов:

$$\text{среднее}_{\text{ссть}} - \text{среднее}_{\text{нет}}$$

$$82,5 - 75,0 = 7,5.$$

Взаимодействие: свод × предварительное предъявление:

(Без предварительного предъявления, интактный — рассеченный) —
(С предварительным предъявлением, интактный — рассеченный)

$$\begin{aligned}(88 - 62) - (83 - 82), \\ 26 - 1 = 25\end{aligned}$$

или:

(Рассеченный: есть — нет) — (Интактный: есть — нет)

$$\begin{aligned}(82 - 62) - (83 - 88), \\ 20 - (-5) = 25.\end{aligned}$$

Все, что было описано, может быть показано непосредственно с помощью таблицы или графика. В табл. 8.1 полученные в эксперименте данные представлены четырьмя числами: 82, 62, 83 и 88. Единицы измерения (здесь — проценты правильных ответов) принято опускать. Общие средние оценки по группам с рассеченным и интактным сводом помещены в *крайней* нижней строчке таблицы. Основной результат действия переменной состояния свода — разность между этими двумя средними, 85,5 и 72, равная 13,5.

Аналогично на другом *крае* таблицы — в правом столбце — помещены общие средние оценки для двух уровней второй независимой переменной — *предварительного предъявления неподкрепленных предметов*: их наличия и отсутствия. Основной результат этой переменной оказался гораздо меньшим: $82,5 - 75 = 7,5$. Формулы для вычисления основных результатов действия приведены под таблицей.

Основные результаты действия независимых переменных представлены также на верхнем графике рис. 8.2. Помимо двух отрезков, отражающих результаты групп с интактным и рассеченным сводом, здесь проведены (и продолжены вправо) еще две пары пунктирных линий. Одна из таких линий идет из середины отрезка результатов группы с интактным сводом и отражает общую среднюю оценку по этой группе — 85,5. Она составляет пару с другой линией, идущей из середины отрезка результатов группы с рассеченным сводом, и это тоже общая средняя оценка — 72,0. В правой части рисунка мы видим, что промежуток, или расстояние, между двумя этими линиями составляет 13,5 (т. е. столько же, сколько мы получили по таблице). Первая линия второй пары берет начало в средней точке между результатами двух групп по выполнению задач с предварительным предъявлением предметов; она соответствует общей средней оценке по этим задачам — 82,5. А парная с ней линия начинается в средней точке между результатами решения задач без предварительного предъявления и тоже отражает общую среднюю оценку — 75,0. Справа указано расстояние между этими линиями, равное 7,5 (как определено по таблице). Таким образом, каждый из основных результатов действия можно представить просто как расстояние между двумя линиями.

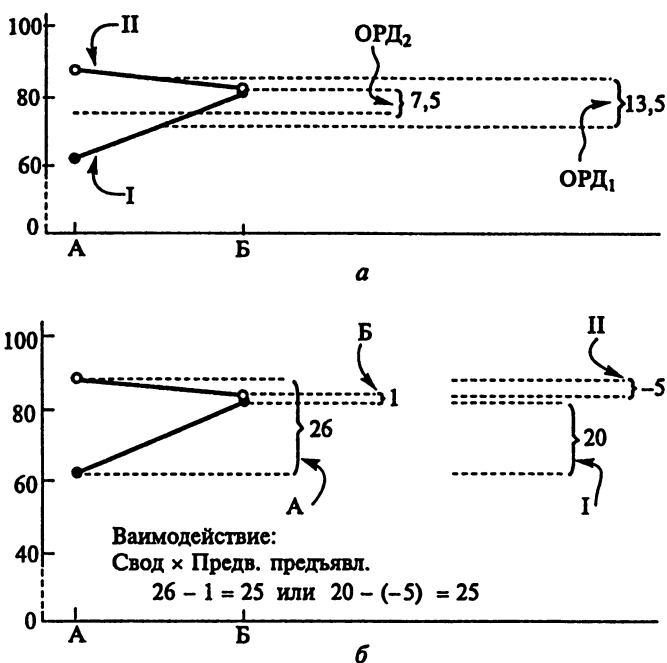


Рис. 8.2. а) Измерение основных результатов действия переменных состояния свода (ОРД₁) и предварительного предъявления неподкрепленных предметов (ОРД₂). I — рассеченный, II — интактный свод; б) измерение взаимодействия переменных. Различия в результатах: I — по группе с рассеченным, II — с интактным сводом; А — по задачам без предварительного предъявления, Б — с предварительным предъявлением. Ось абсцисс — предварительное предъявление неподкрепленных предметов: А — «нет», Б — «есть»; ось ординат — количество правильных ответов (%)

Измерение взаимодействий

Теперь мы рассмотрим способ измерения взаимодействий. Взаимодействие — не просто разность, а разность между двумя разностями. Для его вычисления мы пользуемся числами, расположенными не по краям таблицы, а внутри нее. Давайте еще раз спросим: кто лучше работал — животные с интактным или с рассеченным сводом? Мы видим, что для ответа нужно уточнить, о каких задачах идет речь. Задачи без предварительного предъявления предметов группа с интактным сводом решала лучше, а если все предметы ранее предъявлялись, то обе группы справлялись с задачами одинаково успешно. Чтобы понять, как измеряется такое взаимодействие количественно, обратимся еще раз к данным в табл. 8.1. При выполнении задач без предварительного предъявления предметов группы с интактным и с рассеченным сводом дали 88 и 62 %

правильных решений, разность — 26 %. А в задачах с предварительным предъявлением правильных решений было соответственно 83 и 82 %, разность — 1 %. Взаимодействие — это разность между двумя разностями, т. е. $26 - 1$. Процедуры вычислений приводятся под формулами для определения основных результатов действия. Полное название взаимодействия в данном случае такое: «состояние свода, *помноженное на* предварительное предъявление неподкрепленных предметов». Обычно пользуются сокращенными названиями, здесь — «свод \times предварительное предъявление» (С \times ПП).

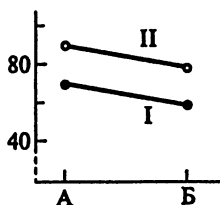
Величина взаимодействия показывает нам, в какой мере основной результат рассеечения свода зависит от предъявленных задач. Но можно задать еще один вопрос: какие задачи решаются лучше — с предварительным предъявлением предметов или без него? Теперь нужно уточнить, какими именно животными: с интактным или с рассеченным сводом. Животным с рассеченным сводом предварительное предъявление предметов помогало выполнять задачу: 82 и 62 % правильных решений, разность — 20 %. А у животных с интактным сводом эта разность не только сократилась, но даже поменяла знак: 83 и 88 %. Если мы будем производить вычисление так же, как и для животных с рассеченным сводом, т. е. по формуле, приведенной в самом низу табл. 8.1, то нам нужно вычесть -5 из 20. Взаимодействие снова будет равно 25. Как бы ни ставились вопросы о различии результатов действия разных уровней каждой из независимых переменных, мы обнаруживаем, что это различие зависит от уровня другой независимой переменной. В этом, по сути дела, и заключается взаимодействие двух независимых переменных.

Оба способа вычисления взаимодействия иллюстрируются на нижнем графике рис. 8.2. Отрезки — те же, что и наверху. Чуть правее сравниваются две пары пунктирных линий. Показано, что различие, или расстояние, между результатами групп животных с интактным и с рассеченным сводом по задачам без предварительного предъявления равно 26, а с предварительным предъявлением — только 1. Разность между двумя этими расстояниями — 25. Теперь посмотрите на правую часть рисунка: нижняя пара пунктирных линий отражает различие результатов решения двух задач группой с рассеченным сводом — 20. Результаты группы с интактным сводом показаны выше: лучше решались задачи без предварительного предъявления предметов, но разница — только 5. Разность между двумя расстояниями вновь равна 25 (ведь вычесть отрицательное число — это прибавить такое же положительное).

Виды взаимодействия

Нужно сказать, что основной результат действия одной независимой переменной (состояние свода), выявленный в обсуждае-

Свод			
Предв. предьявл.	Рассеч.	Интактн.	Среднее
Есть	60	80	70
Нет	70	90	80
Среднее	65	85	



Основные результаты действия:

Свод, $85 - 65 = 20$

Предв. предьявл.

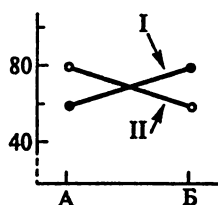
$80 - 70 = 10$

Взаимодействие: $C \times III$

$(90 - 70) - (80 - 60) = 0$

a

Свод			
Предв. предьявл.	Рассеч.	Интактн.	Среднее
Есть	80	60	70
Нет	60	80	70
Среднее	70	70	



Основные результаты действия:

Свод, $70 - 70 = 0$

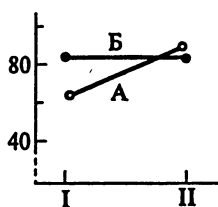
Предв. предьявл.

$70 - 70 = 0$

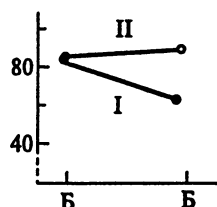
Взаимодействие: $C \times III$

$(80 - 60) - (60 - 80) = 0$

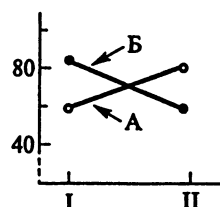
б



в



г



д

Рис. 8.3. Возможные исходы эксперимента с двумя независимыми переменными — состоянием свода (I — рассеченный, II — интактный) и предварительным предьявлением неподкрепленных предметов (A — «нет», B — «есть») — основные результаты и взаимодействия: ось абсцисс: *a*, *б*, *г* — предварительное предьявление неподкрепленных предметов (A — «нет», B — «есть»); *в*, *д* — состояние свода (I — рассеченный, II — интактный); ось ординат — проценты правильных ответов

мых экспериментах Д. Гаффана, оказался относительно небольшим (13,5). Основной результат действия второй независимой переменной (предварительного предьявления неподкрепляемых предметов) — еще меньше (7,5), но между состоянием свода и пред-

варительным предъявлением предметов существует довольно сильное взаимодействие. Давайте посмотрим, что могло бы получиться, если бы данные эксперимента были другими. На рис. 8.3 показано несколько возможных исходов эксперимента как на таблицах, так и графически. На рис. 8.3, *а* основные результаты действия переменных — состояния свода и предварительного предъявления — несколько больше, чем на самом деле, — 20 и 10 соответственно. А вот взаимодействие между ними равно нулю. (Обратите внимание на сокращение — $S \times ПП$.) Мы видим, что в отличие от реальных данных (на рис. 8.2) отрезки результатов обеих групп на рис. 8.3, *а* параллельны. Можно сказать, что на этом рисунке представлено отсутствие взаимодействия между двумя независимыми переменными.

В противоположность нулевому взаимодействию на рис. 8.3, *а* и *б* представлены возможные данные с более сильным взаимодействием, чем в реальном эксперименте. Такое взаимодействие называется *пересекающимся*. А реальные данные дают нам пример *расходящегося* взаимодействия. При *пересекающемся* взаимодействии различия в результатах, показанных животными с рассеченным и с интактным сводом при выполнении задач с предварительным предъявлением предметов и без него, одинаковы, но противоположны по знаку. Группа с рассеченным сводом работала лучше контрольной, если давались задачи с предварительным предъявлением, и хуже, если без него. Те же результаты можно изложить иначе: задачи с предварительным предъявлением решались лучше, чем без него, когда работала группа с рассеченным сводом, но хуже — когда работала контрольная группа. Чтобы провести такое сравнение, данные с рис. 8.3, *б* представлены на рис. 8.3, *д*. Для этого нужно лишь поместить переменную состояния свода на горизонтальную ось, и полученные отрезки будут соответствовать двум уровням переменной предварительного предъявления предметов — его наличия и отсутствия. Какой бы способ графического представления пересекающегося взаимодействия мы ни избрали, изображение этого взаимодействия будет одним и тем же. По вычислениям под рис. 8.3, *б* вы можете видеть, что основные результаты той и другой переменной равны нулю. Иначе говоря, судя по сравнительным результатам решения двух заданий, нет никаких различий между животными с рассеченным и с интактным сводом. Не различается и успешность решения задач с предварительным предъявлением предметов и без него. Однако взаимодействие оказалось очень высоким — 40 (–20 вычиталось из +20, ведь различия в результатах имели противоположные направления).

На рис. 8.3, *в* графически представлены первоначальные данные, причем на горизонтальной оси (так же, как и на рис. 8.3, *д*) помещена переменная состояния свода. По виду они немного напоминают пересекающееся взаимодействие. Группа с рассечен-

ным сводом явно лучше решала задачу с предварительным предъявлением, чем без него, а группа с интактным сводом, в свою очередь, чуть лучше справлялась с задачей без предварительного предъявления. И все же это — расходящееся взаимодействие, поскольку при другом графическом представлении оно уже не даст пересечения. Кроме того, различие в результатах по контрольной группе оказалось весьма небольшим. И наконец, по рис. 8.3, 2 мы видим, что если на горизонтальной оси поместить переменную предварительного предъявления предметов (его наличие и отсутствие), то расхождение отрезков пойдет слева направо. Собственно говоря, расхождение будет столь же высоким, если на вертикальной оси поместить, например, вместо правильных ответов — неверные. В этом случае отрезок, идущий сейчас слева направо, вверх, будет опускаться (и наоборот). Итак, пересекающееся взаимодействие при любом способе графического представления данных будет давать пересечение отрезков. При расходящемся же взаимодействии в зависимости от того, какая из двух независимых переменных помещена на горизонтальной оси, отрезки будут расходиться влево или вправо и по крайней мере в одном случае пересекаться не будут.

Таким образом, в эксперименте с двумя независимыми переменными можно определить основные результаты изолированного действия каждой из них, а также взаимодействие между ними. Взаимодействие бывает трех основных видов: *нулевое (т. е. отсутствие взаимодействия)*, *расходящееся* и *пересекающееся*.

ФАКТОРНАЯ СХЕМА ПРИ ПРОВЕРКЕ ГИПОТЕЗ С ОДНИМ ОТНОШЕНИЕМ

Рассмотрим теперь, как нужно проводить факторный эксперимент для проверки гипотез с одним отношением. Именно о таких экспериментальных гипотезах мы обстоятельно говорили в предыдущих главах. Речь идет о проверке отношения между одной независимой и зависимой переменными, скажем, между состоянием свода (рассечен или нет) и сохранностью способности к узнаванию.

Устранение сопутствующего смешения базисных переменных

Итак, Д. Гаффану удалось показать, что при рассечении свода нарушается память, а не восприятие, причем именно способность к узнаванию. В каждой из двух своих гипотез он *связывал* воздействие независимой переменной с изменением определенной *базисной* переменной, т. е. с нарушением памяти и узнавания. Для проверки этих гипотез ему нужно было устранить смешение базисных переменных с другими. Вы уже знаете (по главе 5), что смешение со второй переменной можно исключить с помощью

введения контрольного условия, сравнивая его действие с действием активного условия основной переменной. Однако такой прием непригоден, если в эксперименте нужно определить не просто какое-то влияние независимой переменной на поведение, но именно влияние на базисную переменную.

Контроль естественного смещения. Чтобы проверить, нарушается ли у животных память, Д. Гаффан должен был давать после пробы-образца достаточно большую отсрочку (130 с). Но если использовать только такую отсрочку, трудно решить, чем именно объясняются различия в результатах групп с рассеченным и с интактным сводом: нарушениями памяти или восприятия? Исследователя интересует одна базисная переменная — память, но ей неизбежно сопутствует вторая базисная переменная — восприятие. С какой из них можно связывать изменение в состоянии свода? Давайте построим диаграмму для случая с использованием только длинной отсрочки и покажем на ней возможность фиксации влияния рассечения свода на память и восприятие.

Интервал отсрочки		Состояние свода	
		рассеченный	интактный
Длинный	память (требуется)	проявление влияния возможно	воздействия нет
	восприятие (требуется)	проявление влияния возможно	воздействия нет

Мы видим, что при использовании длинной отсрочки различие между группами с рассеченным и с интактным сводом можно приписать изменениям любой *потенциальной* базисной переменной, поскольку и та и другая необходимы для выполнения задачи. И если связывать это различие с памятью, то нужно как-то отделить ее от восприятия, т. е. устранить смещение с побочной базисной переменной.

В главе 5 мы показали, как устранить смещение между состоянием свода и состоянием окружающей его области: нужно ввести контрольное условие — повредить у животных эту область, не рассекая свод. Однако смещение, с которым мы встретились сейчас, никакими контрольными условиями для группы с интактным сводом устранить нельзя. Понадобилось бы обеспечить для этой группы ту же самую вероятность нарушения восприятия, что и для группы с рассеченным сводом. А это невозможно сделать, ведь вероятность такого нарушения при рассечении свода нам неизвестна. В данном случае требуется другой прием: нужно подобрать такой *интервал отсрочки*, при котором полученное различие нельзя будет приписать нарушению памяти. Отсрочка между пробой-образцом и пробой-выбором должна быть короткой. В этом

случае для правильного выполнения задачи не нужно запоминать предъявленный предмет надолго, и *возможности* фиксации влияния рассеечения свода будут следующими.

Интервал отсрочки		Состояние свода	
		рассеченный	интактный
Короткий	память (не требуется)	проявление влияния невозможно	воздействия нет
	восприятие (требуется)	проявление влияния возможно	воздействия нет

Теперь различие между группами можно объяснить только нарушением восприятия. Смещение устранено. Однако при короткой отсрочке мы можем исследовать связь состояния свода только с восприятием, а ведь у нашего эксперимента цель другая.

Для того чтобы изучать нарушения памяти при рассечении свода, нужно использовать интервал отсрочки в качестве *второй независимой переменной* (помимо состояния свода). Давайте рассмотрим приведенные диаграммы как две составные части факторного эксперимента. Вернувшись к рис. 8.1, мы увидим, что рассеечение свода не приводит к нарушению восприятия: обе группы одинаково успешно решают задачи с отсрочкой в 10 с. А факт расходящегося взаимодействия между состоянием свода и интервалом отсрочки свидетельствует о том, что в результате рассеечения свода пострадала память. Смещение памяти с восприятием — пример *естественного* сопутствующего смещения, описанного в главе 5. Чтобы запомнить, нужно сначала воспринять, — таков «естественный порядок вещей», как в жизни вообще, так и в эксперименте. Мы взяли в своем примере только два интервала отсрочки ради простоты изложения. Более информативными, как мы показали в главе 7, всегда являются многоуровневые независимые переменные.

Контроль искусственного смещения. Но какой же именно вид памяти оказался нарушен? В реальности (которая, собственно, и дала начало исследованию) — это неспособность людей с дефектами в области гиппокампа опознавать недавно увиденные предметы. Обнаружить такое нарушение у человека очень просто. Достаточно показать ему несколько небольших предметов вроде тех, что предъявлял в своем эксперименте Д. Гаффан, и чуть позже спросить, узнаёт ли он эти предметы среди других, ранее не показанных? Мы уже видели, как организовал задачу на узнавание С. Стернберг (Sternberg S., 1969), работая с людьми: если тестовый стимул входил в предъявленный ранее набор, испытуемые нажимали одну кнопку, если нет — другую.

Но мы не можем поговорить с обезьяной (а ведь *когда-то*, наверное, *могли*). Чтобы исследовать способность обезьяны к опознанию предметов, Д. Гаффан сопровождал их предъявление пищевым подкреплением — сладкой кукурузой. Но тогда различие в результатах решения задач на узнавание между группами с рассеченным и с интактным сводом допускает двоякую интерпретацию: его можно связывать либо со способностью узнавать ранее предъявленные предметы, либо запоминать их по ассоциации с подкреплением. Вспомните, ведь при проверке все «неподкрепленные» предметы были новыми, т.е. ранее не предъявлялись. Поэтому вполне возможно, что у животных с рассеченным сводом нарушена сама способность к образованию ассоциаций. Давайте посмотрим по диаграмме, о каких нарушениях могут свидетельствовать результаты решения задач на *узнавание*.

Предварительное предъявление неподкрепленных предметов		Состояние свода	
		рассеченный	интактный
Нет	узнавание (требуется)	проявление влияния возможно	воздействия нет
	образование ассоциаций (требуется)	проявление влияния возможно	воздействия нет

В этом эксперименте использовалась та же стратегия устранения смещения, что и в приведенном ранее эксперименте по разделению памяти и восприятия.

Мы уже видели, как исследуется способность к образованию ассоциаций — и у людей, и у танцующих мышей. В опытах Р. Калфи и Р. Андерсон (Calfée R. C., Anderson R., 1971) испытуемые связывали числа с триграммами. У испытуемых Р. Йеркса и Дж. Додсона (Yerkes R. M., Dodson J., 1908) вырабатывали связь внешне-го вида туннеля с электроударом. Д. Гаффан (Gaffan D., 1974) подобрал такую задачу на ассоциацию, для успешного выполнения которой *не обязательно* отличать ранее предъявленные предметы от новых, т.е. опознавать их. Животному предъявлялись все предметы, но одни — с подкреплением, а другие — нет. Теперь

Предварительное предъявление неподкрепленных предметов		Состояние свода	
		рассеченный	интактный
Есть	узнавание (не требуется)	проявление влияния невозможно	воздействия нет
	образование ассоциаций (требуется)	проявление влияния возможно	воздействия нет

возможность фиксации влияния рассеечения свода, судя по результатам выполнения ассоциативных задач, можно представить так.

Давая только эти задачи, можно проверить, нарушается ли при рассечении свода образование ассоциаций, — но не больше. А ведь изучаемой базисной переменной было узнавание. Однако в эксперименте использовались *оба типа* задач — и на узнавание, и на ассоциацию, т. е. вводилась вторая независимая переменная — предварительное предъявление неподкрепленных предметов: его наличие или отсутствие.

На рис. 8.2 вы видели, что результаты обеих групп были почти одинаковы, если требовалось просто установить ассоциацию, но когда нужно было определить, предъявлялся ли предмет ранее, т. е. опознать его, группа с рассеченным сводом потерпела крупную неудачу.

Расходящееся взаимодействие между переменными состояния свода и предварительного предъявления свидетельствует о том, что базисной переменной, принимающей разные значения у каждой из групп животных, было именно узнавание, а не образование ассоциаций.

Так же, как и в случае разделения памяти и восприятия, здесь нельзя было ввести контрольное *условие*, т. е. обеспечить одинаковую вероятность нарушения образования ассоциаций у групп с интактным и с рассеченным сводом. Был необходим факторный эксперимент с привлечением второй независимой переменной — предварительным предъявлением негативных (неподкрепленных) предметов. Иначе говоря, нужно было ввести контрольную *переменную*. Смещение между узнаванием и образованием ассоциаций порождалось самой экспериментальной процедурой.

Чтобы иметь возможность проверить способности обезьян к узнаванию, нужно было создать у них ассоциацию между определенными предметами и сладкой кукурузой. В реальности и люди и животные могут опознавать предметы без всякой связи с подкреплением. Следовательно, в данном случае смещение было *искусственным*, вызванным требованиями эксперимента. А смещение памяти с восприятием было *естественным*: ничего нельзя запомнить, не восприняв. Но в обоих случаях стратегия контроля за смещением переменных — одна и та же.

При организации условия, необходимого для проверки предполагаемой базисной переменной, последняя сопровождалась другой, сопутствующей базисной переменной. Тогда вводилось новое условие, при котором разные уровни первичной независимой переменной действовали только на сопутствующую базисную переменную, приводя к различным результатам. Использование и того и другого условия давало вторую, контрольную *независимую переменную* — с двумя уровнями: более и менее активным.

Эксперимент с целью контроля смещения базисных переменных

Контрольная независимая переменная	Первичная независимая переменная	
	активный уровень	пассивный уровень
Более активный уровень		
Предполагаемая базисная переменная	проявление влияния возможно	воздействия нет
Сопутствующая базисная переменная	проявление влияния возможно	воздействия нет
Менее активный уровень		
Предполагаемая базисная переменная	проявление влияния невозможно	воздействия нет
Сопутствующая базисная переменная	проявление влияния возможно	воздействия нет

При «более активном» уровне контрольной переменной активный уровень первичной независимой переменной (рассечение свода) может оказывать действие либо на предполагаемую базисную переменную (память, узнавание), либо на сопровождающую (восприятие, образование ассоциаций). При «менее активном» уровне контрольной переменной предполагаемая базисная переменная вообще никак не проявляется на любом уровне первичной независимой переменной. Расходящееся взаимодействие между первичной и контрольной независимыми переменными позволяет проверить гипотезу об изменении изучаемого базисного процесса.

Контроль сопутствующего смещения повышает *внутреннюю* валидность эксперимента. В чистом эксперименте (как варианте идеального) можно было бы исследовать память отдельно от восприятия и проверять способность обезьян к узнаванию предметов, не прибегая к ассоциированию со сладкой кукурузой. Факторный эксперимент с привлечением второй, контрольной независимой переменной позволяет проверять гипотезы, близкие к тем, что могли бы проверяться в идеальном эксперименте.

Контроль обобщения экспериментальных результатов

Вспомним теперь еще одно исследование, описанное в главе 5, — эксперимент Р. Вайза и В. Досан (Wais R. A., Dawson V., 1974) по изучению влияния диазепама на пищевое поведение. Гипотеза заключалась в том, что данное лекарство оказывает непосредственное действие на процесс питания — оно делает жи-

вотных голодными. При проведении опытов ради удобства каждое животное помещали на 15 мин в специальную клетку, определяя затем количество съеденной там пищи. Эта особенность методики привнесла трудности в проверку исходной гипотезы. Ведь объяснить, почему в специальной клетке крыса с инъекцией диазепама съедает больше, чем крыса с контрольной инъекцией, можно и по-другому. Альтернативной является гипотеза о *снятии тревожности*. Новая ситуация вызывает у крысы состояние тревожности и тем самым — подавляет аппетит. А диазепам — известный транквилизатор — снимает это состояние, и крыса нормально ест. Ясно, что для проверки исходной гипотезы необходим специальный контроль.

Эксперимент проводился на 15 крысах, применялась кросс-индивидуальная схема позиционного уравнивания. Инъекции производились за 15 мин до опыта. «Перед опытом в специальную или домашнюю клетку помещалось 20 (± 5) граммов пищи (точный вес был известен), а после опыта взвешивалась несъеденная пища и оставшиеся крошки» (Wais R. A., Dawson V., 1974, с. 931 — 932). На рис. 8.4 сравнивается количество съеденной пищи при инъекциях диазепама и плацебо. Легко убедиться, что при инъекции диазепама крысы съедали гораздо больше. При тех и других инъекциях в домашней клетке пищи съедалось несколько больше, чем в специальной. Взаимодействие было почти нулевым.

Таким образом, разница в количестве съеденной пищи при инъекции диазепама и без него оказалась весьма значительной как в специальной, так и в домашней клетках (в последней — чуть больше). Следовательно, действие диазепама было непосредственным, не зависимым от снятия тревожности. Это не значит, что связь уменьшения тревожности с повышением аппетита полностью отрицается. Однако если основной результат действия окружающей обстановки (специальная или домашняя клетка) составил лишь 0,7 г, то основной результат действия диазепама был равен 3,5 г.

Мы рассказали лишь об одном из многих экспериментов, вошедших в обширное исследование Р. Вайза и В. Досан. Проверка гипотезы о непосредственном специфическом действии диазепама на

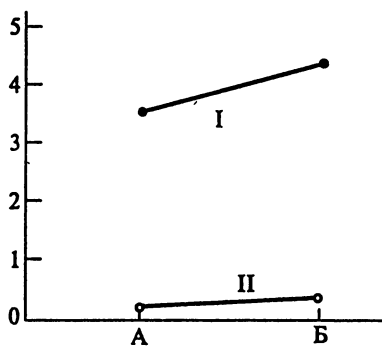


Рис. 8.4. Количество пищи, съеденной при инъекциях диазепама (I) и плацебо (II) в специальных и домашних клетках. Ось абсцисс — окружающая обстановка: А — специальная, Б — домашняя клетка; ось ординат — количество съеденной пищи (в граммах)

пищевое поведение была общей задачей этого исследования. А в данном, только что описанном, эксперименте проверялась альтернативная гипотеза: результат воздействия является не простым, а комбинированным, связанным с двумя переменными — самим введением диазепама и обстановкой, вызывающей тревожность. Но чтобы разделить эти две гипотезы, не обязательно ставить факторный эксперимент. Для проверки гипотезы о снятии тревожности достаточно провести все опыты только в домашних клетках. Мы можем убедиться в этом по следующим диаграммам. Рассмотрим, во-первых, опыты в специальных клетках.

Обстановка, вызывающая беспокойство	Введение диазепама	
	есть активный	нет пассивный
Есть (специальная клетка)	активный	активный

Понятно, что инъекция диазепама — это «активный» уровень соответствующей переменной. Поскольку в специальной клетке у животного возникает состояние тревожности, данный уровень переменной окружающей обстановки тоже будет «активным» (как при введении диазепама, так и без него). Если при введении диазепама пищи будет съедено больше, то это можно объяснить как действием его самого, так и совместным влиянием двух названных переменных. Без диазепама тревожность, возникающая у крысы в специальной клетке, может мешать ей принимать пищу. При инъекции диазепама тревожность снималась бы, и крыса ела бы спокойно. Подобное комбинированное воздействие можно устранить, если проводить опыты в домашних клетках.

Обстановка, вызывающая беспокойство	Введение диазепама	
	есть активный	нет пассивный
Нет (домашняя клетка)	пассивный	пассивный

В этом случае активные уровни двух переменных сочетаться не будут. Поэтому любое различие в количестве съеденной пищи между инъекциями диазепама и плацебо следует интерпретировать как результат непосредственного действия диазепама.

Зачем же понадобилось проводить опыты в специальных клетках, если для разделения гипотез необходимости в этом не было? Ради чего тогда ставился факторный эксперимент? Ответ такой:

если в эксперименте задействовали только один уровень дополнительной переменной, то предлагаемые выводы всегда могут вызвать критику. Безоговорочно *обобщать* результаты, подтверждающие экспериментальную гипотезу, довольно опасно. Скорее нужно предположить, что эти результаты связаны с конкретным *сочетанием* переменных, и поэтому обобщать их не следует. Гипотеза о снятии тревожности — это и есть сомнение в том, что диазепам непосредственно вызывает голод, и его действие можно рассматривать независимо от таких дополнительных условий, как окружающая обстановка. Напротив, гипотеза о непосредственном воздействии диазепама универсальна, приложима к любым обстоятельствам. И когда эксперименты проводятся на нескольких уровнях дополнительной переменной, можно не сомневаться: полученные результаты просты и не зависят от сочетаний с этими уровнями, и их можно обобщать.

Теперь мы можем ответить на вопрос, заданный в главе 5, по поводу эксперимента по трудовой этике, когда испытуемые предпочитали получать награду не просто так, а нажимая на рычаг. Зачем же привлекались девочки-индианки? Вот и ответ: если бы эксперимент проводился только на белых детях, то можно было бы предположить, что нажимать на рычаг предпочитают именно они, т.е. приписать полученный результат сочетанию с дополнительной переменной. Экспериментаторы сделали даже более конкретное предположение: факт предпочтения, обнаруженный в предыдущем исследовании одного из них (Singh D., 1970), может объясняться тем, что белые дети воспитаны в духе протестантской морали, — «чтобы получить награду, надо работать» (Singh D., Query W., 1977, с. 77). Влияние столь специфичного фактора следовало устранить, и для этого Сингх и Кэри провели эксперимент, привлекая не только белых девочек, но и девочек-индианок, а также мальчиков-индейцев и белых мальчиков. Вторую переменную в этом эксперименте можно назвать *культурным фоном*. И поскольку полученные результаты оказались независимыми от культурного фона, можно было с большей уверенностью обобщить исходную гипотезу о том, что люди предпочитают активно действовать, а не сидеть сложа руки. И надо сказать, не только люди. Подобные результаты были получены в опытах с голубями и крысами (Jensen G.D., 1963; Neuringer A.J., 1969; Carder B., Berkovitz K., 1970; Singh D., 1970). Правда, если уж быть совершенно точным, надо признать, что исключения составляют кошки. «Вот она, кошачья лень, — так описывают свои результаты К. Коффер и Г. Коулсон, — кошки предпочитают есть легкодоступную пищу, а не ту, которую нужно специально добывать» (Koffler K., Coulson G., 1971). Шесть подопытных котов (Зеро, Пи-Уи, Галем, Немо, Иван и Зелма) в ситуации выбора сначала съедали свободно лежащую приманку (специальные говяжьи кон-

сервы «Кот в сапогах») и только после этого нажимали на рычаг, чтобы получить точно такую же. Это не первое свидетельство того, что у кошек есть свои причуды (а может быть именно они и ведут себя по-настоящему рационально).

Использование разных уровней второй переменной, включая индивидуальные различия, служит не для устранения смещения, а для обоснованного обобщения результатов. Оно приближает нас к *эксперименту полного соответствия*, т. е. к такому безупречному эксперименту, где мы могли бы проверить свою гипотезу на всех уровнях дополнительных переменных и на всех испытуемых, к которым применяются полученные результаты. Следовательно, привлечение второй переменной повышает *внешнюю* валидность эксперимента.

В главе 5 мы обсуждали связь обобщения результатов с *внутренней* валидностью. Чтобы обобщать полученное отношение между независимой и зависимой переменными, мы должны быть уверены, что наша независимая переменная является чистой, свободной от побочных «примесей». Только тогда можно исследовать результаты действия независимой переменной в любых, отличных друг от друга, ситуациях. Однако теперь мы убеждаемся, что для надежного обобщения результатов этого недостаточно. Если отношение между независимой и зависимой переменными обнаружено нами при вполне определенных уровнях дополнительной переменной, мы должны учитывать, что действие независимой переменной может оказаться непрямым, а его результат — не простым, а комплексным, комбинированным. Распространение этого результата на все остальные уровни дополнительной переменной связано, таким образом, с вопросами *внешней* валидности эксперимента. В сжатой форме все это можно выразить так: для надежного обобщения нужно выделить *простые* результаты действия *изолированных переменных*.

Схемы позиционного уравнивания

Читая экспериментальные работы, вы обнаружите, что схемы позиционного уравнивания часто рассматриваются как факторные, т. е. включающие вторую независимую переменную, даже если в самой гипотезе фигурирует только одна. Например, в каждой паре экспериментов Дж. Кеннеди и Дж. Ландесмана (Kennedy J. L., Landesman J., 1963) было обнаружено определенное отношение между высотой рабочей поверхности и количеством обработанных деталей. Каждая такая высота с одинаковой частотой предъявлялась всем испытуемым в каждой из шести последовательных проб. Средние оценки вычислялись *по каждой пробе*. Тем самым *местоположение пробы в последовательности* их предъявления автоматически оказывалось второй независимой переменной. Таким

образом, мы можем выделить основные результаты действия как высоты рабочей поверхности, так и местоположения каждой пробы. Можно измерить и взаимодействие между этими двумя независимыми переменными. Иногда взаимодействие оказывается настолько большим, что требуется специальное объяснение. Но так или иначе, переменная местоположения проб — это автоматически возникающая особенность применения схем позиционного уравнивания (отличных от реверсивного позиционного уравнивания в многоуровневых экспериментах). Поэтому такие схемы всегда следует анализировать как факторные.

ПРОВЕРКА КОМБИНИРОВАННЫХ ГИПОТЕЗ

С появлением факторных схем в технике экспериментирования совершается еще один шаг вперед. Теперь исследователи попытаются проверять гипотезы о том, каким образом влияют на поведение сразу две независимые переменные, сочетаясь друг с другом. Есть немало примеров, когда мы совершенно уверены, что данная деятельность определяется не единственной независимой переменной. Еще до экспериментов Р. Йеркса и Дж. Додсона (Yerkes R. M., Dodson J. D., 1908) было известно, что на количество проб, необходимых для научения, будет влиять как сила электроудара, так и трудность различения стимулов. С. Стернберг (Sternberg S., 1969), давая задачи на узнавание, тоже заранее знал, что с увеличением количества знаков в предварительно предъявленном наборе для опознания одного знака потребуется больше времени. Знал он и о том, что время опознания увеличится, если тестовый стимул будет трудно различить. Вы помните (по главе 7), что у Р. Йеркса и Дж. Додсона были основания для конкретного предположения об отношении между комбинациями силы электроудара и трудности различия, с одной стороны, и количеством проб, необходимых для достижения критерия научения, — с другой. Мы увидим, что и Стернберг выдвинул гипотезу об отношении между комбинацией двух своих независимых переменных и временем реакции опознания, применив информационную модель этого процесса. И в том и в другом эксперименте экспериментальная гипотеза касалась *взаимодействия* независимых переменных. Мы познакомим вас также с двумя новыми экспериментами, причем в каждой проверяемой гипотезе будет представлен один из основных типов взаимодействия, описанных нами выше.

Закон Йеркса — Додсона: определение оптимального уровня

Предварительное описание этого факторного эксперимента, проведенного много лет назад на танцующих мышах и посвященного изучению влияния силы электроудара на успешность науче-

ния различать черное и белое, было дано в главе 7. Р. Йеркс и Дж. Додсон повторили свой эксперимент, введя новые условия, затруднявшие и облегчавшие различение стимулов. Различение облегчали, сокращая доступ света в черный туннель и делая его тем самым еще чернее, а затрудняли, сокращая доступ света в белый туннель, так что он больше походил на черный.

Затем отдельным группам мышей давали задачи с легким различением (при пяти разных уровнях электроудара) и с трудным различением (четыре уровня электроудара). По каждой группе определяли количество проб, необходимых для достижения критерия успешного решения задач. Все полученные результаты, в том числе и те, что уже были представлены в главе 7, — для задач средней трудности (см. рис. 7.3), показаны на рис. 8.5. Очевидно, что основным результатом воздействия переменной трудности различения очень высокий. Выделяются следующие общие закономерности: во-первых, при всех уровнях интенсивности электроударов быстрее всего научение протекало в тех случаях, когда различать туннели было легко. Во-вторых, можно говорить о некотором преимуществе при сильных ударах: правые ветви каждого из графиков несколько ниже левых. Однако по-настоящему интересно именно взаимодействие между двумя независимыми переменными — силой электроудара и трудностью различения. Гипотеза, о которой мы говорили в главе 7, подтвердилась. Мы уже знаем, что для задач средней трудности минимальное количество проб, необходимых для успешного научения, приходится на силу удара в 300 единиц. Теперь мы видим также, что для задач с трудным

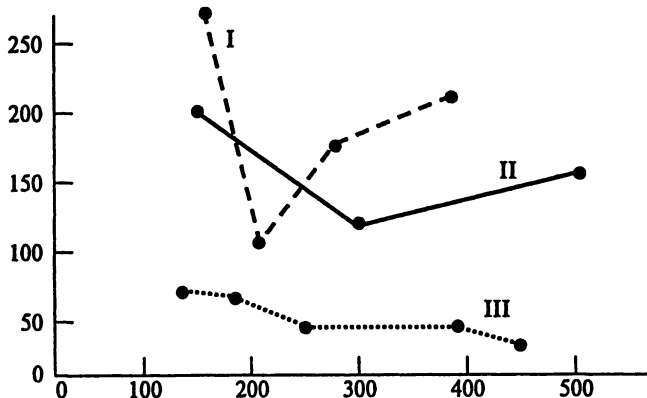


Рис. 8.5. Влияние силы электроударов на успешность научения при различных уровнях трудности задач (опыты на танцующих мышах): I — трудные, II — средние и III — легкие задачи. Ось абсцисс — сила электроудара (в условных единицах); ось ординат — среднее количество проб, необходимых для достижения критерия успешного решения задач

различением этот минимум наступает при более слабом ударе — 195 единиц. И наконец, для самых легких задач результаты продолжают улучшаться даже при 420 единицах; весьма вероятно, что при еще более сильном ударе мышцы будут научиться еще быстрее.

Еще раз об изучении опознания: отсутствие взаимодействия переменных

Вспомните эксперимент С. Стернберга (Sternberg S., 1969), описанный в главе 7. В нем измеряли время, необходимое испытуемым для того, чтобы мысленно просмотреть набор ранее предъявленных цифр и определить, входит ли в него тестовая цифра. Было установлено отношение типа «абсолютно-абсолютное». С каждым увеличением запоминаемого набора на один знак прирост времени, необходимого для опознания, был одним и тем же — примерно 35 мс. Но это было в том случае, когда тестовый стимул был очень четко виден. А что если бы он был нечетким, скажем, плохо отпечатанным? Мы понимаем, что время опознания должно увеличиться. Но будет ли этот прирост одинаковым, независимым от количества цифр в запоминаемом наборе? Возможно, что с увеличением набора воздействие недостаточной четкости тестового стимула будет нарастать. Однако С. Стернберг выдвинул гипотезу о постоянстве прироста времени опознания (нулевое взаимодействие), а не о возрастании этого прироста с увеличением запоминаемого набора (расходящееся взаимодействие).

Чтобы сделать тестовую цифру нечеткой, С. Стернберг проецировал на нее модель шахматной доски. Разглядеть цифру было, конечно, можно, но довольно трудно. На рис. 8.6 показано время реакции опознания при различных сочетаниях двух независимых переменных — размера запоминаемого набора и четкости тестового стимула. Видно, что основные результаты действия и той и другой переменной достаточно велики. При увеличении запоминаемого набора от одного до четырех знаков время реакции возросло в целом на 115 мс. Это лишь немногим больше, чем ожидалось: если с каждой новой цифрой время опознания возрастает на 35 мс, то увеличение на три знака даст общий прирост времени в 105 мс. Видно также, что когда тестовая цифра была нечеткой, время реакции возрастало в среднем примерно на 70 мс.

Напротив, взаимодействие переменных оказалось весьма незначительным. При наборе в один знак различие между опознанием четкой и нечеткой цифр составило 60 мс, а при наборе из четырех цифр — 75 мс. Максимальная величина взаимодействия доходит до 15 мс. В сравнении с основными результатами действия переменных это очень мало и может объясняться простой случайностью. Короче говоря, Стернберг предсказал нулевое взаимодей-

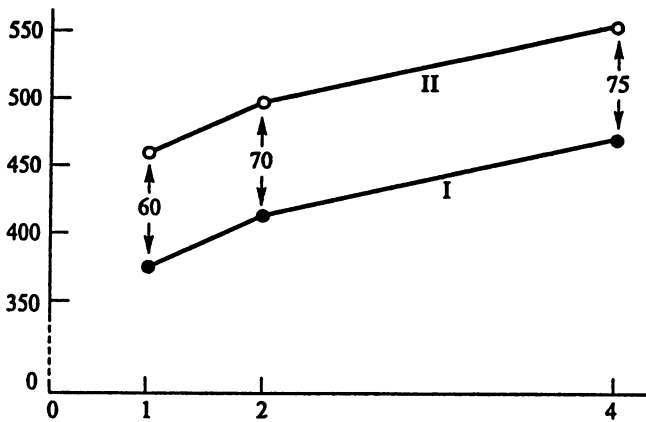


Рис. 8.6. Зависимость времени реакции на четкие (I) и нечеткие (II) тестовые стимулы от количества знаков в ранее предъявленном наборе. Ось абсцисс — размер предъявленного набора; ось ординат — среднее время реакции (в мс)

стве между двумя независимыми переменными, влияющими на время обработки информации, и получил результат, подтверждающий его гипотезу.

Прошлый опыт и решение задач: неравное действие независимой переменной

Гостю, который приходит вторым, всегда легче. Когда эксперимент уже проведен и полученный результат оказался неожиданным, всегда можно сказать исследователям, что они должны были его предвидеть и даже объяснить. Многие из нас, включая и автора этой книги, крепки задним умом. И это стоит учесть при обсуждении еще одного исследования — с прошлым опытом и решением задач в лабиринте, поставленном на крысах В. Дененбергом и Дж. Мортонем (Dennenberg V. H., Morton J. R., 1962).

Эксперимент включал две независимые переменные. Первая переменная — приручение детенышей в период вскармливания. В течение 24 дней после рождения, когда крысята обычно находятся с матерью, половину из них ежедневно забирали из клетки, а другая половина всегда оставалась с матерью и остальным потомством. «Приручение заключалось в том, что крысят вынимали и на 3 минуты помещали каждого в специальную коробку, частично заполненную стружкой, после чего возвращали в домашнюю клетку» (там же, с. 1096).

Было проведено три эксперимента, в каждом из которых несколько изменяли вторую независимую переменную — окружающую обстановку по окончании периода вскармливания. Вот как

описывают В. Дененберг и Дж. Мортон свой третий эксперимент (который мы и обсудим): «После периода вскармливания (в возрасте 25 дней) несколько прирученных и неприрученных детенышей помещали в тесные клетки размером $17 \times 24 \times 17$ см, где они и оставались до полного созревания. Оставшуюся часть детенышей, тоже прирученных и неприрученных, сажали в просторные ящики, где можно было свободно бегать. Таких ящиков было четыре: в двух из них содержались прирученные детеныши, в двух других — нет» (там же, с. 1097). Таким образом, окружающей обстановкой были либо тесные клетки (первый уровень переменной), либо просторные ящики (второй уровень). Клетки были разделены перегородками, и обстановка была очень однообразной. В ящиках же на площади в 4 кв. фута размещали «всевозможные платформы, туннели, скаты и проходы» (там же, с. 1096), и все это обогащало окружающую обстановку. Как в клетках, так и в ящиках крысы жили до полного созревания (до возраста 50 дней).

Затем крыс запускали в лабиринт. Лабиринт Хебба — Вильямса можно составить по-разному и сделать несколько путей к приманке. Каждая трасса считалась отдельной задачей. Крысы по очереди пробегали по лабиринту, сначала для тренировки (в течение 15 дней), а затем — на оценку (в течение 12 дней по одной задаче ежедневно). «Показателем успешности решения задач каждым животным служило общее количество ошибок за весь испытательный период» (там же, с. 1097). Зависимая переменная — среднее число ошибок по каждой из следующих подгрупп: прирученные в клетках, прирученные в свободной обстановке, неприрученные в клетках и неприрученные в свободной обстановке.

Результаты эксперимента, представленные на рис. 8.7, оказались неожиданными. Легко убедиться, что основной результат действия приручения в период вскармливания весьма невелик: прирученные и неприрученные животные решали задачи в среднем примерно одинаково. Основным результатом действия различной окружающей обстановки оказался несколько выше: как прирученные, так и неприрученные животные, содержащиеся в просторных ящиках, допуска-

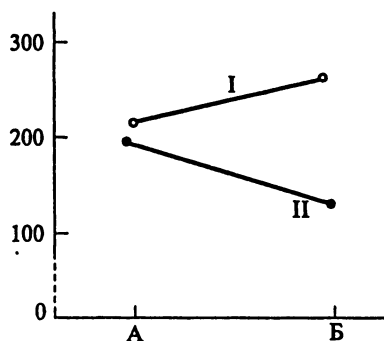


Рис. 8.7. Влияние приручения крыс в период вскармливания и их содержания после вскармливания на успешность решения задач в лабиринте: I — клетки, II — свободное окружение. Ось абсцисс — приручение в период вскармливания: А — прирученные, Б — неприрученные животные; ось ординат — среднее число ошибок

ли меньше ошибок, чем находившиеся в клетках. Однако это различие не было статистически значимым. А вот расходящееся взаимодействие между приручением и окружающей обстановкой выражено настолько ярко, что просто бросается в глаза. Различие по среднему числу ошибок у прирученных животных, содержавшихся в тесных клетках и в просторных ящиках, составило 14 единиц, а у неприрученных — 83; взаимодействие равно 69. Иначе говоря, обнаружено неравное действие окружающей обстановки: небольшое — для прирученных животных и высокое — для неприрученных.

Причины подобного взаимодействия могут быть следующими. Известно, что приручение в период вскармливания делает крыс менее тревожными, менее дикими. Они становятся более зависимыми от человека, менее рискованными. А свободная обстановка только тогда способствует успешному решению задач в лабиринте, когда животное стремится использовать предоставленные ему возможности. Поэтому содержание в просторных ящиках и помогало прежде всего неприрученным животным, неизбалованным человеческой заботой и независимым от нее, но почти не помогало прирученным.

Однако сами исследователи не проявили интереса к указанному взаимодействию, не попытались его объяснить. Их вывод довольно категоричен: опыт, полученный в период вскармливания, не оказывает никакого влияния на дальнейшее решение задач. Они утверждают, что взаимодействие «можно, по-видимому, отнести за счет случайных изменений в поведении» (там же, с. 1097). А как мы знаем из главы 6, для того и нужна проверка статистической значимости (а здесь вероятность нуль-гипотезы — менее 1/100), чтобы получить определенный эталон, а уж затем решать, приемлем полученный результат или его можно приписать случайным изменениям.

Эксперимент с нерелевантной информацией

Итак, мы убедились, что экспериментаторы сталкиваются подчас с совершенно непредвиденными результатами. А иногда (хотя и не часто) полученные данные настолько неожиданны и вместе с тем так ясны, что становятся подлинным открытием. «Способность совершать открытия» — так называют умение разглядеть одно, когда ищешь другое. Нередко оно оказывается просто необходимым. Нечто подобное пережили Р.Саймон и А.Руделл (Simon J. R., Rudell A. P., 1967). Открытый ими результат представится вам сейчас таким очевидным, что, казалось бы, исследователи должны были ожидать его с самого начала. Однако на самом деле эта очевидность свидетельствует о ясности и четкости обнаруженного отношения между независимой и зависимой перемен-

ными. И до проведения эксперимента оно очевидным не было. Помимо всего прочего, данное исследование было посвящено изучению времени реакции, а в этой области эксперименты ведутся уже 100 лет, и никто еще не получал подобных данных.

Р. Саймон и А. Руделл предположили, что время реакции должно быть короче, если сигнал подается в доминантное полушарие мозга. Для правой — это левое полушарие, для левой — правое.

Экспериментальная гипотеза состояла в следующем. Звуковой сигнал, подаваемый на правое ухо (связанное с левым полушарием), будет вызывать у правой более быструю реакцию, чем сигнал на левое ухо (связанное с правым полушарием). У левой это отношение окажется обратным: короче будет реакция на сигнал, подаваемый на левое ухо. Эксперимент по проверке этой гипотезы был прекрасно продуман и организован. В опытах приняли участие студенты колледжа: 16 девушек левшей, 16 девушек правшей, 16 юношей левшей и 16 юношей правшей. Им нужно было выполнять команды, подаваемые в левый или правый наушники: по команде «левая» — нажимать одну кнопку левой рукой, а по команде «правая» — другую кнопку правой рукой. Выполнять команду следовало независимо от того, из какого наушника она поступала. Был составлен блок из 132 проб с равным числом команд «левая» и «правая», набранных в случайном порядке. На каждое ухо команды подавались также в случайной последовательности с одинаковой частотой. Саймон и Руделл ожидали, что у правой время реакции всегда будет короче в тех случаях, когда команда поступает справа, чем когда она поступает слева, а у левой — наоборот.

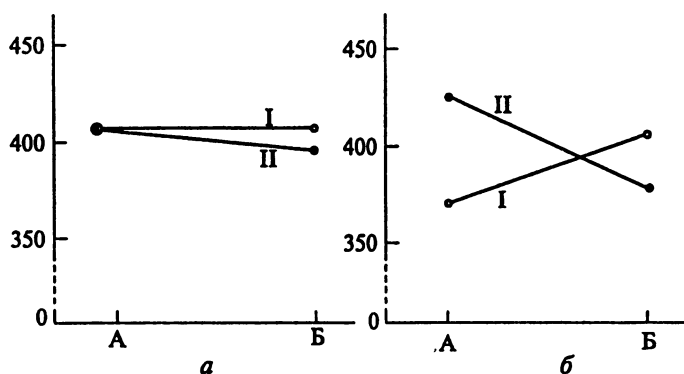


Рис. 8.8. Время ответных реакций на команды, поступавшие на левое и правое ухо: *а* — у левшей (I) и правшей (II); *б* — при командах «левая» (I) и «правая» (II). Ось абсцисс — стимулируемое ухо: А — левое, Б — правое; ось ординат — время реакции (в мс)

Результаты проверки этой гипотезы представлены на рис. 8.8, *а*. Ясно, что ожидания не оправдались. При подаче команд как на правое, так и на левое ухо время реакции у правой и у левой оказалась почти одинаковым — от 404 до 410 мс. Возможное расхождение взаимодействия настолько мало (7 мс), что им с уверенностью можно пренебречь.

Напротив, на рис. 8.8, *б* мы видим, сколь велико оказалось взаимодействие между тем, с какой стороны поступала команда, и самим типом команды — «левая» и «правая». Это, конечно, *пересекающееся* взаимодействие, равное 85 мс. Как мы уже убедились по рис. 8.8, *а*, основной результат подачи команды на разные уши почти неразличим (5 мс). Аналогично, основной результат действия команд разного типа составляет лишь 3 мс.

Следует еще раз подчеркнуть, что испытуемых просили не обращать никакого внимания на то, с какого наушника поступала команда, и реагировать только на само произносимое слово — «левая» и «правая». Однако пренебречь этой *нерелевантной* информацией они все-таки не смогли.

Достигнутые преимущества

Во всех четырех только что описанных исследованиях для проверки экспериментальных гипотез привлекалась вторая независимая переменная. Опираясь только одной независимой переменной, проверить эти гипотезы было нельзя. Мы убедились, что далеко не все гипотезы, получившие подтверждение, были предусмотрены экспериментаторами заранее. Однако без привлечения второй переменной установить что-либо было бы вообще невозможно.

Проверка теоретического предположения о наличии двух базисных процессов, лежащих в основе различительного научения, стала возможной только благодаря использованию различных уровней электроудара. Экспериментальная гипотеза состояла в том, что с увеличением трудности задачи сила электроударов, при которой научение будет проходить наиболее эффективно, т. е. с минимальным числом проб, будет понижаться. Основания такой гипотезы понятны: более трудная задача требует лучшего перцептивного различения (первый предполагаемый базисный процесс). А с понижением силы электроудара танцующие мышцы становились более «внимательными». И даже если связь степени белизны туннеля с электроударом (второй предполагаемый базисный процесс) была небольшой, ее вполне хватало для обеспечения необходимого различения. Напротив, при сильных электроударах животные были слишком встревожены, и их способность к различению не актуализировалась вообще. Закон Йеркса — Додсона дает ясное представление о влиянии силы мотивации на результаты деятельности и не теряет своей значимости по сегодняшний день.

В каждом из трех остальных исследований содержится пример одного из основных видов взаимодействия. Графически эти виды представлены с помощью монотонно возрастающих или убывающих линий (отражающих связь независимой и зависимой переменных), т. е. не имеющих максимума или минимума посередине. Как мы уже говорили, при *нулевом взаимодействии* разные уровни второй независимой переменной представлены параллельными отрезками. Если отрезки расходятся вправо (или влево), то говорят о *расходящемся взаимодействии*, а если отрезки пересекаются, причем независимо от того, какая переменная помещена на горизонтальной оси, то это — *пересекающееся взаимодействие*.

В эксперименте С. Стернберга взаимодействие между количеством знаков в запоминаемом наборе и четкостью тестового стимула приближалось к нулю. Отрезок, представляющий результаты по опознанию нечетких тестовых стимулов, расположен выше второго отрезка, но практически параллелен ему. Экспериментальная гипотеза о нулевом взаимодействии следует из информационной модели процесса опознания, предложенной С. Стернбергом. Мы уже показывали, что прямолинейная зависимость времени реакции от количества знаков в запоминаемом наборе подтверждает идею о сканировании каждого отдельного знака при опознании тестового стимула. Однако, согласно модели С. Стернберга, все это происходит на более поздней *стадии* переработки информации, чем опознание стимула — на стадии его «кодирования». Поэтому если тестовый стимул просто трудно различить, то прирост времени реакции будет постоянным, независимым от количества знаков в ранее предъявленном наборе. Если же это затрудняло само сканирование, то с увеличением числа запоминаемых знаков время опознания одного из них возрастало бы все больше и больше. Иначе говоря, если бы недостаток четкости воздействовал на ту же стадию процесса обработки информации, что и количество знаков, то взаимодействие между этими переменными было бы расходящимся: наблюдалось бы возрастание различий во времени реакции на четкие и нечеткие тестовые стимулы с увеличением ранее предъявленного набора знаков.

В экспериментах В. Дененберга и Дж. Мортонна (Dennenberg V. H., Morton J. R., 1962) было обнаружено расходящееся взаимодействие между влияниями специфического прошлого опыта животных во время и после вскармливания на научение в лабиринте. Если мы сочтем это взаимодействие значимым (чего сами исследователи не сделали), то можем получить ясное представление о комбинированном воздействии изучаемых ранних факторов на последующее поведение животного. Даже если окружающая обстановка, в которой находился детеныш после вскармливания, предоставляет ему возможность свободно передвигаться, это еще не значит, что он впоследствии будет успешно справляться с за-

дачами в лабиринте. Возможности останутся неиспользованными, если еще во время вскармливания детенышей приручали. Благодаря столь раннему опыту их восприятие окружающей среды (и потребность ее обследовать) всецело зависело теперь от участия экспериментатора. Вполне возможно, что, находясь в просторных ящиках, крысята просто ждали, когда их оттуда вынут.

Теперь мы подошли к последнему исследованию. Чтобы реакции были быстрыми, команды «левая» должны подаваться на левое ухо, а команды «правая» — на правое. При перекрестном отношении между ухом и рукой возникают трудности. Чем же это объяснить? Дж. Саймон и А. Руделл говорят о привычке реагировать в том же направлении, откуда получен сигнал, как о «массовом стереотипе» (Simon J. R., Rudell A. P., 1967, с. 300). Он столь же прочен, как хорошо заученный навык. В Соединенных Штатах вы зажигаете свет, поворачивая ручку вверх, в Англии — вниз. Понятно, что подобные традиции могут приводить к образованию прочно усвоенных навыков. Однако трудно себе представить, каким образом можно было научиться столь прочной связи между ухом и рукой. Дж. Саймон и А. Руделл утверждают далее: «Полученные результаты свидетельствуют о наличии сильной природной тенденции связывать стимуляцию правого уха с реакцией правой руки, а стимуляцию левого уха — с реакцией левой руки» (там же, с. 303). Но это значит, что ни о каком научении речи не идет. Несмотря на многочисленные последующие исследования указанной связи, природа процессов, лежащих в ее основе, пока остается неясной. Развитие психологии — это не только поиски разумных ответов, но и постановка новых интересных вопросов.

В своем обсуждении мы ограничились двумя независимыми переменными. Правда, при использовании кроссиндивидуальных схем позиционного уравнивания (например, латинский квадрат в исследовании Дж. Кеннеди и Дж. Ландесмана (Kennedy J. E., Landesman J., 1963) вторая независимая переменная — местоположение проб в последовательности их предъявления — возникает автоматически. И если такая схема применяется в эксперименте с двумя изучаемыми независимыми переменными (скажем, помимо высоты рабочей поверхности изменяется и размер обрабатываемых деталей), то чисто технически независимых переменных окажется три: высота рабочей поверхности, размер деталей и местоположение пробы. Поэтому в литературе по изучению комбинированных действий двух независимых переменных вы будете встречать трехфакторные схемы.

Но есть и такие эксперименты, в которых действительно используются и изучаются больше двух независимых переменных. Когда независимых переменных две, взаимодействие между ними называется *взаимодействием первого порядка*, а если их три, то это

уже *взаимодействие второго порядка*. Например, в одном из своих экспериментов по времени реакции С. Стернберг изменял: 1) количество альтернатив (знаков в наборе); 2) четкость тестового стимула и 3) совместимость между стимулом и ответом. Это дало ему три однопорядковых взаимодействия между независимыми переменными: количество \times четкость, количество \times совместимость и четкость \times совместимость, а также взаимодействие второго порядка: количество \times четкость \times совместимость.

Ему удалось установить, что при четких и нечетких стимулах имеет место одинаковое расходящееся взаимодействие между количеством и совместимостью. Исследование взаимодействия более высокого порядка позволило проверить более детальную модель по сравнению с обсуждавшимися ранее. Но здесь, на трех независимых переменных, давайте и остановимся.

КЛАССИФИКАЦИЯ ЭКСПЕРИМЕНТАЛЬНЫХ СХЕМ

Настало время просмотреть все экспериментальные схемы, приведенные в нашей книге. В последней главе новых схем не прибавится: там пойдет речь о корреляционных исследованиях, а в них экспериментатор уже не изменяет своих независимых переменных самостоятельно. Обсуждавшиеся нами эксперименты отличаются друг от друга по трем основным параметрам. Соответствующие им различия были введены в четвертой, седьмой и настоящей главах.

В самых первых экспериментах, где участвовал только один испытуемый, внутренняя валидность обеспечивалась за счет различного упорядочивания проб: использовались их регулярное чередование, случайная или позиционно уравненная последовательности. Даже если увеличить число испытуемых в подобных экспериментах, контроль внутренней валидности останется прежним. Иначе говоря, сравнение условий, или уровней, независимой переменной проводится здесь интраиндивидуально. Если же внутренняя валидность эксперимента обеспечивалась с помощью привлечения разных групп испытуемых для каждого из условий, или уровней, независимой переменной, то сравнение результатов действия последних проводилось по межгрупповой схеме. А в экспериментах предыдущей главы для обеспечения внутренней валидности каждому испытуемому предъявлялась определенная позиционно уравненная последовательность уровней независимой переменной. В результате сравнение между уровнями проводилось кроссиндивидуально. Таким образом, первый параметр классификации экспериментальных схем можно назвать *способом сравнения условий, или уровней, независимой переменной: интраиндивидуальное, межгрупповое или кроссиндивидуальное*.

В предыдущей главе было введено еще одно различие. Если в одних экспериментах используются два (или несколько) качественно разных условия, или уровня, независимой переменной, то в других экспериментах они могут быть выражены количественно. Мы обсуждали это совсем недавно, и вам должно быть понятно, о чем идет речь. Второй основной параметр экспериментальных схем — это *тип изменения независимой переменной: качественный или количественный*.

Третье различие экспериментальных схем введено в настоящей главе. В экспериментах используются либо одна, либо несколько независимых переменных. *Число независимых переменных* и есть третий основной параметр: *либо эксперимент с одной переменной, либо факторный эксперимент*.

Поскольку у нас есть три основания сравнения условий, два типа независимых переменных и два варианта по их числу, то, по всей видимости, мы должны получить ровно 12 различных ячеек для размещения любых экспериментальных схем. Некоторые эксперименты, описанные в нашей книге, действительно, можно распределить по этим ячейкам. Примеры приведены в табл. 8.2. Скажем, в эксперименте Дененберга и Мортон (Denenberg V. H., Morton J. R., 1962) с прошлым опытом для каждой из независимых переменных — содержания детенышей во время и после вскармливания — использовалось межгрупповое сравнение, причем обе эти переменные изменялись качественно. Этот эксперимент точно попадает в ячейку № 8. Однако не всякий факторный эксперимент можно столь же успешно занести в одну из приведенных ячеек. Ведь используемые независимые переменные (две или несколько) могут различаться по способам сравнения условий или по типу их изменения. Поэтому возможных сочетаний введенных нами параметров гораздо больше 12. В эксперименте Д. Гаффана, где обезьяны решали задачи на выбор по образцу с различной временной отсрочкой, сравнение условий одной независимой переменной — состояния свода — было, естественно, межгрупповым, а сравнение уровней другой независимой переменной — интервала отсрочки — проводилось кроссиндивидуально. Кроме того, состояние свода изменялось качественно (т.е. либо рассечен, либо нет), а интервал отсрочки — количественно (три уровня).

Если обе независимые переменные одинаковы и по способу сравнения условий, и по типу изменения, как в эксперименте В. Дененберга и Дж. Мортон, то схема называется *гомогенной*. А если они различаются по этим параметрам, как в эксперименте Д. Гаффана, то схема становится *гетерогенной*. Факторные эксперименты часто классифицируют по числу уровней каждой независимой переменной. В исследовании В. Дененберга и Дж. Мортон (Denenberg V. H., Morton J. R., 1962) использовалась схема

12 гомогенных экспериментальных схем

Число независимых переменных	Способ сравнения		
	интраиндивидуальное	межгрупповое	кроссиндивидуальное
Одна переменная			
Качественный тип	1. Заучивание фортепьянных пьес (гл. 1)	2. Мысленная отработка навыков (гл. 4)	3. —
Количественный тип	4. Субъективная тяжесть (гл. 7)	5. Темп предъявления (гл. 7)	6. Время реакции (Р. Готтсданкер, Т. Вей), смешанные интервалы (гл. 7)
Факторные схемы			
Качественный тип	7. Рассечение свода (гл. 8)	8. Ранний опыт (гл. 8)	9. Диазепам (гл. 8)
Количественный тип	10. —	11. Закон Йеркса — Додсона (гл. 8)	12. Обработка деталей (гл. 7)

«2 × 2», а Д.Гаффан (Gaffan D., 1974) в своем эксперименте с разными отсрочками использовал схему «2 × 3».

Изложенная трехмерная классификация экспериментальных схем основывается на главных параметрах. Как вы помните, в индивидуальных экспериментах используются различные последовательности предъявления проб, в экспериментах с межгрупповым сравнением — разные способы построения групп: случайное распределение, случайный отбор из предварительно выделенных «слоев» и т. д. Поэтому, прочитав целую подписку психологического журнала, вы можете не обнаружить даже двух одинаковых экспериментов, полностью подобных по своим схемам.

КРАТКОЕ ИЗЛОЖЕНИЕ

Для проверки гипотезы о воздействии рассечения свода у обезьяны на ее способность узнавания нужно было провести два факторных эксперимента: один — чтобы определить, нарушается ли при этом именно память, а другой — чтобы установить специфичность ее нарушения. В каждом из этих экспериментов помимо первичной изучаемой переменной — состоя-

ния свода — привлекалась вторая независимая переменная. *Факторным* называется эксперимент с использованием по крайней мере двух независимых переменных.

В итоге факторного эксперимента определяются *основной результат действия* каждой из независимых переменных и *взаимодействие* между ними. На примере эксперимента с рассеченным сводом мы раскрыли конкретное содержание основных результатов и взаимодействия и показали способы их измерения (когда две независимые переменные имеют по два уровня). Основным результатом действия независимой переменной — это разность между общими средними значениями зависимой переменной по каждому из уровней. Взаимодействие — это разность между двумя разностями по каждому из уровней второй независимой переменной.

Основные виды взаимодействия, которые можно получить в факторных экспериментах, иллюстрировались графически. Здесь же представлены и способы измерения основных результатов и взаимодействий. Одна из независимых переменных помещается на горизонтальной оси, и затем по соответствующим значениям зависимой переменной строятся отдельные графики для каждого из уровней другой независимой переменной. Если отрезки, полученные таким образом для второй независимой переменной, параллельны, *взаимодействие отсутствует (т. е. нулевое)*. Если отрезки расходятся вправо или влево, взаимодействие называют *расходящимся*. А если они пересекают друг друга, то это *пересекающееся взаимодействие*.

Факторные эксперименты часто применяются для проверки гипотез с одним отношением, т. е. для определения результата влияния одной независимой переменной на поведение. Именно такие гипотезы обсуждались нами в предыдущих главах. Центральная задача в таких экспериментах — устранить сопутствующее смешение, как естественное, так и искусственное. Если, согласно гипотезе, независимая переменная действует на определенную *базисную переменную*, то с помощью факторного эксперимента можно обеспечить контроль за выделением этого воздействия. Однако при том условии, которое организуется для определения воздействий различных уровней изучаемой независимой переменной, последняя может по-разному влиять и на вторую базисную переменную. Это показано на примерах экспериментов Д. Гаффана (Gaffan D., 1974) с рассечением свода у обезьян. Чтобы устранить смешение собственно памяти с восприятием, между пробой-образцом и пробой-выбором давались не только длинные, но и короткие интервалы отсрочки. При коротком интервале для правильного выбора не нужно запоминать предмет надолго. Смещение способности к узнаванию и запоминания по ассоциации устранялось путем использования чисто ассоциативных задач, для решения которых не нужно опознавать предьявляемые предметы. В каждом из этих экспериментов второй уровень контрольной переменной не оказывал *никакого* воздействия на предполагаемую базисную переменную. Показателем успешности разделения предполагаемой и сопутствующей переменных служил факт расходящегося взаимодействия между ними. Таким образом, применение факторной схемы позволило получить результаты, близкие к результатам идеального (чистого) эксперимента, в котором можно было бы исследовать только базисную переменную. Тем самым повышается внутренняя валидность эксперимента.

Использование факторных схем позволяет повысить и внешнюю валидность, прежде всего в тех случаях, когда нужно проверить гипотезу о непосредственном простом действии независимой переменной. Согласно альтернативной гипотезе, воздействие является комбинированным, т.е. активный уровень независимой переменной дает результат только благодаря сочетанию с активным уровнем дополнительной переменной. Так, Р. Вайз и В. Досан (Wise R. A., Dawson V., 1974), проводя опыты на крысах, предполагали, что лекарственный препарат диазепам непосредственно приводит к увеличению количества съедаемой пищи, повышая аппетит. Согласно же комбинированной гипотезе, лекарство действует не прямо, а снимая состояние тревожности. В опытах, проводившихся как в новой обстановке, вызывающей у крыс подобное состояние, так и в домашних клетках, было обнаружено значительное различие в количестве съеденной пищи между животными с инъекцией диазепама и с инъекцией плацебо. Тем самым удалось показать, что действие диазепама является непосредственным, оно не зависит от уровня тревожности. А если бы результат действия зависел от уровня дополнительной переменной, то обобщать его было бы нельзя. При проверке обобщения результатов следует учитывать не только факторы, вызванные участием экспериментатора, но и индивидуальные различия испытуемых. Для того и привлекались девочки-индианки в экспериментах по изучению трудовой этики, чтобы распространить полученные результаты на людей, принадлежащих к разным культурным группам.

При использовании схем позиционного уравнивания эксперимент автоматически становится факторным. Помимо изучаемой независимой переменной здесь появляется переменная местоположения каждой пробы в порядке их предъявления. В кроссиндивидуальном эксперименте с позиционным уравниванием определяется среднее значение по данным всех испытуемых для первой пробы, затем для второй и т.д.

Однако главное преимущество факторных экспериментов — это возможность проверять настоящие комбинированные гипотезы. Например, для проверки гипотезы, известной теперь как закон Йеркса—Додсона, оптимальная сила напряжения, способствующего хорошему изучению, должна понижаться с увеличением трудности задач — нужно было использовать задачи различных уровней трудности различения. Далее обсуждаются факторные эксперименты, в каждом из которых был получен один из основных видов взаимодействия, описанных нами выше. Нулевое взаимодействие было установлено в эксперименте С. Стернберга (Sternberg S., 1969) с изменением четкости опознаваемого тестового стимула и размера ранее предъявленного набора знаков. В эксперименте В. Дененберга и Дж. Мортон (Denenberg V. H., Morton J. R., 1962) с использованием двух независимых переменных — условия содержания крыс во время и после вскармливания — было обнаружено расходящееся взаимодействие. А пересекающееся взаимодействие было получено в эксперименте по изучению времени реакции, проведенном Дж. Саймоном и А. Руделлом (Simon J. R., Rudell A. P., 1967). Одной независимой переменной был тип команды («левая» и «правая»), а другой — ухо, на которое она подавалась (левое и правое). Выяснилось, что значительное влияние на время реакции оказывает то, с какой стороны поступает сигнал (а эта переменная была нерелевантной). В тех случаях, когда «ухо» не совпадало с типом команды, время реакции возрастало.

В каждом из упомянутых экспериментов для проверки комбинированных гипотез нужно было использовать факторную схему. Оценить справедливость соответствующих теорий и моделей, привлекая только одну независимую переменную, было бы невозможно. Взаимодействие, полученное Р. Йерксом и Дж. Додсоном, подтвердило теорию о двух базисных переменных — различия стимулов и ассоциирования, на каждую из которых сила электроудара оказывала противоположное воздействие. Результаты В. Дененберга и Дж. Мортонна можно объяснить тем, что приучение в период вскармливания делает крыс неспособными к использованию всех преимуществ свободного передвижения для дальнейшего успешного научения в лабиринте. Результаты, полученные С. Стернбергом, подтвердили предложенную им модель разделения стадий процесса обработки информации — кодирование стимула и сканирование ранее предъявленного набора знаков при опознании одного из них. А обнаруженная Дж. Саймоном и А. Руделлом прочная связь между стимулируемым ухом и типом команды остается пока интригующей загадкой.

Существуют эксперименты, где изучаемых независимых переменных — больше двух, в них проверяются более сложные комбинированные гипотезы. Взаимодействие между двумя переменными называется *взаимодействием первого порядка*, между тремя — *взаимодействием второго порядка* и т. д.

Все обсуждавшиеся нами экспериментальные схемы можно классифицировать по трем основным параметрам. Первый параметр — это способ сравнения условий, или уровней, независимой переменной: сравнение по каждому испытуемому, межгрупповое сравнение или сравнение по всем испытуемым. Второй — это характер изменения независимой переменной, качественный или количественный. И третий — число независимых переменных: либо эксперимент с одной переменной, либо факторный эксперимент. Схемы не обязательно должны быть одинаковы (*гомогенны*) в отношении каждой из независимых переменных. Например, для одной независимой переменной может применяться межгрупповое сравнение, а для другой — интра- или кроссиндивидуальное.

Контрольные вопросы и задания

1. С какой целью в эксперименте с рассечением свода использовалось несколько временных интервалов между пробой-образцом и пробой-выбором?
2. Приведите примеры основных результатов и взаимодействия в факторных экспериментах.
3. Как измеряются основные результаты действия независимых переменных на графике?
4. Опишите, как выглядят на графиках основные виды взаимодействия двух независимых переменных (если у каждой из них — по два уровня).
5. Почему в экспериментах по проверке гипотез о воздействии независимой переменной на определенную базовую переменную часто требуется контрольная переменная?
6. Расскажите, как использовать несколько уровней второй независимой переменной для повышения внешней валидности.

7. Почему для проверки закона Йеркса — Додсона был необходим факторный эксперимент?

8. Каким образом вытекает конкретная комбинированная гипотеза, проверенная в эксперименте С. Стернберга, из предложенной им модели процесса обработки информации?

9. Изобразите на графике такие экспериментальные результаты, которые могли бы быть получены, если бы приручение крыс в период вскармливания не влияло на успешность их научения в лабиринте, а содержание в просторных ящиках после вскармливания помогало бы больше, чем содержание в тесных клетках.

10. Придумайте комбинированную гипотезу, которая подтверждалась бы наличием пересекающегося взаимодействия. (Постарайтесь отвлечься от исследования Дж. Саймона и А. Руделла.)

11. Какое место (и почему) занимает эксперимент Р. Йеркса и Дж. Додсона в системе экспериментальных схем?

СТАТИСТИЧЕСКОЕ ПРИЛОЖЕНИЕ: ДВУХФАКТОРНЫЙ ДИСПЕРСИОННЫЙ АНАЛИЗ

Для того чтобы при наличии двух независимых переменных проверить статистическую значимость двух результатов действия независимой переменной, а также взаимодействие между переменными, применяется F -критерий. Принципы его применения точно такие же, как и описанные в предыдущем приложении. Для того чтобы выявить, достаточно ли величина отношения превышает 1, чтобы отвергнуть нуль-гипотезу, производится сравнение межгрупповой оценки дисперсии генеральной совокупности с внутригрупповой оценкой.

Как получить внутригрупповую оценку, уже было показано. Межгрупповая оценка определяется отдельно для каждого из двух основных результатов действия и для взаимодействия. Таким образом, вычисляются три величины F ; каждая полученная величина сравнивается с табличным значением критерия для альфа-уровня, равного 0,05 или 0,01. Это значение критерия можно найти в Статистической таблице 3.

Эксперимент с двумя независимыми переменными

Давайте по-другому рассмотрим четыре выборки наших данных по времени реакции. Допустим, что на самом деле эксперимент на время реакции проводился с двумя независимыми переменными: одной из них был тип стимула — свет или тон, другой — тип реакции: простая реакция или реакция выбора. Простая реакция означает нажатие левой кнопки, когда сигнал появляется слева, и нажатие правой — когда он появляется справа. Вернемся к исходным обозначениям — условие А представляет простую реакцию на световой стимул; условие Б — простую реакцию

на тон; условие В — реакцию выбора на свет; Г — реакцию выбора на тон. Опыт проводился на четырех группах по 17 испытуемых. Ниже приводятся средние времена реакций, полученные для четырех групп испытуемых.

Тип реакции	Тип сигнала		
	тон	свет	среднее
Простая реакция	162	185	173,5
Реакция выбора	250	265	257,5
Среднее	206,0	225,0	215,5

Различие, связанное с ответом (типом реакции), представлено в этом случае *различием между строками*, а различия, вызванные стимулом, представлены *различиями между столбцами*. Таким образом, произведение реакции на стимул есть произведение строки на столбец (стр \times стл). В матрице r строк и c столбцов, в нашем случае $r = c = 2$.

Внутригрупповое среднее квадратичное

Для тех же четырех групп данных можно использовать предыдущие расчеты для вычисления среднего квадратичного внутри группы (СКВ_{ВГ}).

$$\text{СКВ}_{\text{ВГ}} = \sum x_{r1c1}^2 + \sum x_{r1c2}^2 + \sum x_{r2c1}^2 + \sum x_{r2c2}^2, \quad (8.1)$$

или

$$\text{СКВ}_{\text{ВГ}} = 4306 + 5808 + 5391 + 4673 = 20178.$$

Как вы заметили, индексы у слагаемых уже новые. $\sum x_{r1c1}^2$ означает, что полученная внутри группы величина x^2 соответствует строке 1 (простая) и столбцу 1 (тон). Точно так же $\sum x_{r2c2}^2$ означает величину для строки 2 (выбор) и столбца 2 (свет) и т.д.

Здесь для нахождения среднего квадратичного можно снова применить формулу (7.6) (поскольку $r \times c = k$):

$$\text{СКВ}_{\text{ВГ}} = \frac{\text{СКВ}_{\text{ВГ}}}{N - rc}$$

Из того, что 68 испытуемых делятся на четыре группы, как и ранее, следует:

$$\text{СКВ}_{\text{ВГ}} = \frac{20178}{68 - 4} = 315.$$

Среднее квадратичное по строкам

Вначале найдем сумму квадратов по строкам и из нее найдем среднее квадратичное по строкам. Разности между средним по каждой строке и общим средним вычисляются следующим образом:

$$d_{r1} = M_{r1} - M_{\text{общ}}; d_{r2} = M_{r2} - M_{\text{общ}}, \quad (8.2)$$

или

$$d_{r1} = 173,5 - 215,5 = -42,0; d_{r2} = 257,5 - 215,5 = +42,0.$$

Сумма квадратов по строкам — это сумма квадратов этих d -значений, умноженная на произведение числа случаев в группе n и числа столбцов c :

$$СК_{\text{стр}} = nc(d_{r1}^2 + d_{r2}^2) \text{ и т.д., если есть последующие строки.} \quad (8.3)$$

Здесь

$$СКВ_{\text{стр}} = 17 \cdot 2(1764,0 + 176,40) = 119952.$$

Число степеней свободы для строк равно их числу минус 1:

$$df_{\text{стр}} = r - 1. \quad (8.4)$$

В нашем случае

$$df_{\text{стр}} = 2 - 1 = 1.$$

И здесь также межгрупповое среднее квадратичное находится делением суммы квадратов на число степеней свободы. Поэтому для строк

$$СКВ_{\text{стр}} = \frac{СК_{\text{стр}}}{df_{\text{стр}}}, \quad (8.5)$$

или

$$СКВ_{\text{стр}} = \frac{119952}{1} = 119952.$$

Среднее квадратичное по столбцам

Совершенно аналогичные процедуры могут быть сделаны и относительно столбцов. Вначале

$$d_{c1} = M_{c1} - M_{\text{общ}}; d_{c2} = M_{c2} - M_{\text{общ}}, \quad (8.6)$$

или

$$d_{c1} = 206,0 - 215,5 = -9,5; d_{c2} = 225,0 - 215,5 = +9,5.$$

$$СК_{стл} = nr (d_{c1}^2 + d_{c2}^2) \text{ и т. д., если есть еще столбцы,} \quad (8.7)$$

или

$$\begin{aligned} СК_{стл} &= 17 \cdot 2(90,25 + 90,25) = 6137, \\ df_{стл} &= c - 1 \end{aligned} \quad (8.8)$$

или

$$df_{стл} = 2 - 1 \cong 1.$$

$$СКВ_{стл} = \frac{СК_{стл}}{df_{стл}}. \quad (8.9)$$

В нашем случае

$$СКВ_{стл} = \frac{6137}{1} = 6137.$$

Среднее квадратичное (строки \times столбцы)

Для того чтобы найти сумму квадратов ($СК_{стр \times стл}$), вы должны вначале найти разность между средним каждой подгруппы и общим средним. Затем сложить квадраты этих разностей и умножить полученную сумму на число случаев в группе. Наконец, вычтуть из этого числа сумму квадратов по строкам и сумму квадратов по столбцам. Давайте теперь сделаем эти операции шаг за шагом:

$$\begin{aligned} d_{r1c1} &= M_{r1c1} - M_{общ}; d_{r1c2} = M_{r1c2} - M_{общ}; \\ d_{r2c1} &= M_{r2c1} - M_{общ}; d_{r2c2} = M_{r2c2} - M_{общ}. \end{aligned}$$

В нашем случае

$$\begin{aligned} d_{r1c1} &= 162,0 - 215,5 = -53,5, \\ d_{r1c2} &= 185,0 - 215,5 = -30,5, \\ d_{r2c1} &= 250,0 - 215,5 = +34,5, \\ d_{r2c2} &= 265,0 - 215,5 = +49,5, \end{aligned}$$

$$СК_{стр \times стл} = n(d_{r1c1}^2 + d_{r1c2}^2 + d_{r2c1}^2 + d_{r2c2}^2) - СК_{стр} - СК_{стл}. \quad (8.10)$$

(Замечание: первая часть уравнения уже вычислялась с использованием уравнения 7.4.)

$$\begin{aligned} СК_{стр \times стл} &= 17(2862,25 + 930,25 + 1190,25 + 2450,25) - \\ &- 119952 - 6137 = 126361 - 119952 - 6137 = 272. \end{aligned}$$

Прежде чем мы перейдем к последнему шагу вычисления среднего квадратичного ($СКВ_{\text{стр} \times \text{стл}}$), мы должны найти число степеней свободы для взаимодействия строк и столбцов. Вспомним, что мы сравниваем разности по одной независимой переменной, вызванные действием другой независимой переменной. Существуют $(r - 1)$ разностей по строкам и $(c - 1)$ при сравнении этих строк с разностями по столбцам. Таким образом, общее число df равно произведению $(r - 1)(c - 1)$. В нашем случае, где всего две строки и два столбца, взаимодействие (строки \times столбцы) равно 1:

$$df_{\text{стр} \times \text{стл}} = (r - 1)(c - 1), \quad (8.11)$$

или

$$df_{\text{стр} \times \text{стл}} = (2 - 1)(2 - 1) = 1.$$

Среднее квадратичное по строкам и столбцам равно сумме квадратов по строкам и столбцам, деленное на соответствующее число степеней свободы:

$$СК_{\text{стр} \times \text{стл}} = \frac{СК_{\text{стр} \times \text{стл}}}{df_{\text{стр} \times \text{стл}}}. \quad (8.12)$$

В нашем случае

$$СК_{\text{стр} \times \text{стл}} = \frac{272}{1} = 272.$$

Вычисление F -отношения

Теперь у нас есть четыре оценки популяционной дисперсии σ_{χ}^2 . Это: (1) внутригрупповое среднее квадратичное; (2) среднее квадратичное по строкам; (3) среднее квадратичное по столбцам и (4) среднее квадратичное — строки \times столбцы. Мы можем использовать внутригрупповое среднее квадратичное как знаменатель при вычислении F -отношения относительно каждого из остальных средних квадратичных. Введение знаменателя часто называют показателем ошибки, имея в виду несистематическое изменение, которое невозможно контролировать в экспериментальных условиях:

$$F_{\text{стр}} = \frac{СКВ_{\text{стр}}}{СКВ_{\text{ВГ}}}. \quad (8.13)$$

В нашем случае

$$F_{\text{стр}} = \frac{119952}{315} = 380,80.$$

Таким же образом

$$F_{стл} = \frac{СКВ_{стл}}{СКВ_{вг}}, \quad (8.14)$$

или

$$F_{стл} = \frac{6137}{315} = 19,48.$$

И еще раз соответственно:

$$F_{стр \times стл} = \frac{СКВ_{стр \times стл}}{СКВ_{вг}}, \quad (8.15)$$

или

$$F_{стр \times стл} = \frac{272}{315} = 0,86.$$

Принятие или отвержение нуль-гипотезы

Аналогично тому, как это делалось в статистическом приложении к главе 7, мы воспользуемся Статистической таблицей 3 для нахождения критического значения F . Для $F_{стл}$ имеется $1df$ в числителе и $64df$ в знаменателе. Табличное значение для отвержения нуль-гипотезы для 1 и $65df$ равно $7,04$ на уровне $0,01$. Очевидно, что полученная нами величина $380,80$ позволяет на этом уровне отклонить нуль-гипотезу. Для $F_{стл}$ комбинация в числителе и знаменателе та же самая. И здесь полученная величина $19,48$ позволяет отклонить нуль-гипотезу на альфа-уровне, равном $0,01$.

Для $F_{стр \times стл}$ мы также ищем табличное значение для 1 и $65df$. Полученная нами величина $0,86$ не позволяет отклонить нуль-гипотезу даже для альфа-уровня $= 0,05$. Критическое значение здесь равно $3,99$. F , меньшее единицы, может быть получено лишь для выборочного распределения. В этом случае оно просто не может быть статистически значимым.

Таблица дисперсионного анализа

Дисперсионный анализ можно подытожить в виде следующей таблицы. Обратите внимание, что степени свободы являются аддитивными, так же как и суммы квадратов.

Дисперсионный анализ
Эксперимент на время реакции с разными типами
стимулов и видами реакций

Источник дисперсии	СК	<i>df</i>	СКВ	<i>F</i>	<i>p</i>
Реакции (строки)	119952	1	119952	380,80	<0,01
Стимулы (столбцы)	6137	1	6137	19,48	<0,01
Взаимодействие строки × столбцы	272	1	272	0,86	
Внутригрупповая	20178	64	315		
Общая	146539	67			

Задача. Используйте данные из задачи в статистическом приложении к главе 7 и проведите дисперсионный анализ с составлением таблицы дисперсионного анализа. Снова данные получены для шести отдельных групп испытуемых. Одной переменной является величина награды, второй переменной — трудность задачи. Данные из главы 7 должны быть использованы следующим образом.

Трудность	Величина награды от низкой к высокой				
	А	Б	В		
Легкая	Ур. 4	Ур. 5	Ур. 6		
Трудная	Ур. 3	Ур. 2	Ур. 1		
Ответ:					
Источник дисперсии	СК	<i>df</i>	СКВ	<i>F</i>	<i>p</i>
Трудность (строки)	433,2	1	433,2	46,33	<0,01
Награда (столбцы)	15,8	2	7,9	0,84	
Взаимодействие (трудность × награда)	141,8	2	70,9	7,58	<0,01
Внутригрупповая	224,4	24	9,35		
Общая	815,2	29			

Глава 9

КОРРЕЛЯЦИОННЫЕ ИССЛЕДОВАНИЯ¹

Теперь вы очень хорошо знаете, что такое эксперимент. Вероятно, вы даже убедились, что эксперимент, несмотря на все проблемы контроля, предлагает наилучший путь проверки гипотез относительно переменной, влияющей на поведение. Экспериментатор активно меняет независимую переменную и соотносит значения зависимой переменной с различными уровнями независимой переменной.

С теми знаниями, которые вы приобрели, попробуйте спланировать эксперимент для каждой из следующих трех гипотез, касающихся поведения.

1. Правильные методы воспитания детей приводят к тому, что, повзрослев, люди проявляют лучшую психологическую приспособленность; дети, в отношении которых использовались плохие методы, повзрослев, оказываются и плохо приспособленными. (Гипотеза кажется очевидной.)

2. Самый старший ребенок в семье будет иметь более высокий интеллект, следующий за ним по старшинству будет следующим и по интеллекту и т. д. (Кажется весьма сомнительной.)

3. Лица, получившие высокую оценку по специальному тесту на способности, преуспевают в работе контролера деталей машин более часто, чем лица с низкой оценкой. (Не лишена оснований.)

Если вы хотите сравнить результаты различных способов воспитания детей с помощью активного эксперимента², вам придется убедить одну группу родителей использовать «хорошие» методы воспитания, например спокойные разъяснения (что не очень трудно), а другую группу — «плохие» методы воспитания, например бессмысленную ругань. Конечно, обе группы семей должны быть тщательно уравнены по другим параметрам. Ну что же, желаю удачи!

И все-таки задача такого эксперимента значительно легче, чем задача провести активный эксперимент на тему о порядке рождения. С чего здесь начать? Ведь вы захотите проконтролировать не

¹ Перевод Ч. А. Измайлова.

² Здесь и далее термином *активный эксперимент* обозначается эксперимент, в котором исследователь сам планирует и реализует условия независимой переменной. — *Примеч. ред.*

только количество детей в семье, но и интервалы между их рождениями. Если на то пошло, почему бы не проконтролировать и пол ребенка? Эти намерения уже приближают нас к футуристически планируемому миру, описанному с такой удручающей подробностью Джорджем Орвелом в его романе «1984 год». И при всем этом мы еще оказываемся далеки от эффективного эксперимента.

В *принципе* вышеописанные эксперименты возможны. Однако в зависимости от вашего отношения к ним, они оказываются либо практически неосуществимыми, либо неэтичными. И тем не менее исследования этих гипотез были реально проведены. Только в них использовался не активный эксперимент, а *корреляционный* подход. Это значит, что в данных исследованиях не производилось активных действий с целью вызвать различия в поведении, а лишь отыскивалась корреляция между существующими различиями.

Для проверки первой гипотезы сравнивались психологически хорошо приспособленные взрослые и плохо приспособленные взрослые с учетом информации об их детстве, полученной путем объективных записей, а также с помощью опросов, которые были сделаны много лет назад, и, наконец, воспоминаний. И мы увидим ниже, что хотя гипотеза кажется очевидной и правильной, возникают проблемы в ее доказательстве. Здесь мы столкнемся с теперь уже знакомым нам злом — смешением с сопутствующей переменной.

Во втором случае была взята большая группа молодых людей, родившихся примерно в одно и то же время, которая в возрасте 19 лет прошла испытания по тесту на интеллект. Затем были вычислены отдельно средние для рожденных первыми, вторыми и т. д. И вот *было* обнаружено, что, действительно, существует корреляция между порядком рождения и интеллектом. Хотя это звучит неправдоподобно, но это так.

Обстоятельства этого исследования — главное, очень большое число испытуемых — дали возможность его авторам применить *статистические* методы контроля для выявления смешения с побочными переменными. Однако, как мы увидим, контроль в корреляционных исследованиях никогда не бывает таким же убедительным, как в активных экспериментах.

Гипотеза о том, что тест на способности позволит выбрать хороших контролеров, не может быть исследована с помощью эксперимента по иной причине. В этом случае не существует независимой переменной, о воздействии которой на поведение могла бы идти речь. Здесь перед нами просто две различные оценки поведения одного и того же испытуемого. Индивидуальные различия испытуемых по данным теста на способности соотносятся с их индивидуальными различиями в способности приобретать профессиональные навыки. И тестовые оценки, и оценки качества

работы аналогичны *зависимой переменной* из предыдущих глав. Однако когда нет *независимой переменной*, то нет оснований использовать и термин «зависимая переменная».

Эта глава будет особенно полезна тем из вас, кому в психологии более интересна работа с конкретными личностями, чем лабораторные исследования. Активные эксперименты имеют дело *со сходными* реакциями испытуемых, причем не обязательно людей, тогда как предметом корреляционных исследований, как правило, являются различия между людьми — по интеллекту, способностям или по отдельным чертам личности. Понятие личность полезно только потому, что люди *различаются* по многим параметрам. И тем не менее индивидуальные различия могут быть лучше изучены с помощью активного экспериментирования. Мы уже отмечали, что некоторые корреляционные исследования в принципе могли бы быть проведены как активные эксперименты. Однако, ввиду того что это практически неосуществимо, здесь возникают новые проблемы внутренней валидности. Далее мы увидим, что корреляционное исследование, так же как и активный эксперимент, внутренне валидно в зависимости от того, насколько оно близко к идеальному эксперименту.

Ваше знакомство с этой главой значительно продвинет вас в понимании статей по корреляционным исследованиям. Прежде всего вы сможете увидеть, пытался ли исследователь проконтролировать смещающиеся переменные. Вы сможете оценить, достаточно ли эффективны были использованные методы контроля. Вы даже окажетесь способны сами проводить определенные виды корреляционных исследований. Для этого вам достаточно будет познакомиться с примером исследования, где производится предсказание по тесту, и научиться по статистическому приложению вычислять коэффициент корреляции.

Читая эту главу, приготовьтесь ответить на вопросы по следующим темам.

1. Что понимается под корреляционным исследованием?
2. Почему в корреляционном исследовании всегда присутствует сопутствующее смещение?
3. Методы контроля сопутствующего смещения.
4. Условия, делающие возможным вычисление коэффициента корреляции.
5. Параметры, по которым корреляционные исследования отличаются друг от друга.

ИССЛЕДОВАНИЕ ПРЕДПОСЫЛОК ОПТИМАЛЬНОЙ ПСИХОЛОГИЧЕСКОЙ ПРИСПОСОБЛЕННОСТИ

Вы только что прочитали название статьи, написанной четырьмя исследователями: Элен Сигельман, Джеком Блоком, Джин

Блок и Анной ван дер Липпе (Siegelman E., Block Jack, Block Jeanne, Von de Lippe A., 1970). Под *предпосылками* здесь подразумеваются некоторые факторы на ранней стадии жизни, которые впоследствии сказываются на способности к психологической адаптации. Не относитесь слишком строго к термину «*оптимальная*» приспособленность. Этот термин вовсе не означает, что исследователям удалось найти людей, которые достигли совершенства в этом отношении. Просто они обнаружили несколько человек в возрасте за 30 лет, которые подошли к оптимальному уровню гораздо ближе, чем другие. В этом исследовании проводилось сравнение групп хорошо и плохо приспособленных лиц с точки зрения того, в каких семейных условиях они провели детство.

Удивительно, что исследование такого рода оказалось вообще возможным. Предположим, что вы нашли хорошо и плохо приспособленных взрослых. Как получить полноценные данные о способах их воспитания и тем более выяснить общие установки родителей, которые могли повлиять на эту приспособленность? Исследование, действительно, *было необычным*, поскольку требовало фундаментальной подготовительной работы более чем за 30 лет до того, как были идентифицированы обе группы.

Всего в исследовании был оценен 171 взрослый испытуемый с точки зрения психологической приспособленности. Затем для них искались корреляции с факторами-предпосылками.

Испытуемые

Получить детальную информацию о ранней стадии жизни стольких взрослых людей оказалось возможным в результате того, что все они в свое время были испытуемыми в двух «*продолженных*» исследованиях, проведенных Институтом благосостояния детей при Калифорнийском университете в Беркли. Продолженным называется такое исследование, в котором периодически возвращаются к изучению испытуемого в течение длительного периода времени. Исследовались дети, рожденные в конце 20-х гг. XX в. в двух городах Беркли и Окленд в Калифорнии. Конечно, здесь требовалась помощь родителей. Хотя многие из первоначального набора испытуемых впоследствии оказались недоступны для обследования и некоторые вообще скончались, оставшееся число согласившихся участвовать в исследовании — 171 человек — было достаточно большим.

Большинство теперь уже взрослых испытуемых опрашивались тремя разными психологами, а остальные — только двумя. После опроса психолог составлял личностную характеристику испытуемых. Для этого он раскладывал набор стандартных описательных карточек по девяти классам, руководствуясь тем, в какой мере свойство личности, записанное на карточке, выражено у данного

индивида. Например, на карточке могло значиться: «ищет оправдания своим ошибкам». Если психолог сталкивался с большим числом алиби, которые то и дело приводил для себя испытуемый, он помещал эту карточку в класс 1. Если эта особенность была присуща испытуемому лишь в *небольшой мере*, он помещал карточку в класс 4 или 5. Если же это свойство было совсем не характерно для испытуемого, он откладывал карточку в класс 9.

Всего использовалось 90 карточек из набора «Калифорнийский Q-набор» (Block J., 1961). Экспериментатор предварительно тренировался в том, чтобы в результате его оценок карточки распределялись по классам в соответствии с законом нормального распределения, с наибольшим количеством карточек в средних классах.

До начала основного исследования (Siegelman E. et al., 1970) отдельная группа психологов-клиницистов (всего 9 человек) также классифицировала описательные карточки. Но это делалось не для характеристики какого-то определенного индивида, а для получения общей картины *оптимальной* психологической *приспособленности*. Давайте посмотрим сначала на характеристики, которые в среднем были оценены как наиболее определяющие (положительно) оптимальную приспособленность. В табл. 9.1 слева представлены 13 карточек, которые были помещены в наивысшую категорию.

Каждого, кто произвел бы на интервьюирующего впечатление человека, отвечающего всем этим пунктам, несомненно, можно было бы считать образцом психологической приспособленности. Такой человек доводит дело до конца, остается спокойным, этичным и т. д.

Таблица 9.1

Описательные характеристики теста «Калифорния Q», максимально положительно и максимально отрицательно определяющие оптимальную психологическую приспособленность (Siegelman E. et al., 1970)

№ карточки	Максимально положительные характеристики	№ карточки	отрицательные характеристики
35	Дружественный, способен к близкому общению, к сочувствию	45	Имеет хрупкую систему эгозащиты; слабая интеграция, в условиях стресса или травмы дезорганизуется и снижает адаптивность
2	Исключительно надежный и ответственный человек	78	Чувствует себя жертвой и неудачником в жизни; жалеет себя

№ карточки	Максимально положительные характеристики	№ карточки	Максимально отрицательные характеристики
60	Рефлексирует свои мотивы и поведение	86	Преодолеывает тревожность и конфликты (подавляет их), отказываясь признавать их; склонен к подавленности, избегает общения
26	Продуктивный, доводит дело до конца	22	Не чувствует личной значимости в жизни
64	Социально восприимчив в широком смысле	55	Склонен к самоуничтожению
70	Поведение этически выдержанно. Не противоречит собственным личностным нормам	40	Уязвим как для реальных, так и для воображаемых опасностей; вообще, труслив
96	Ценит свою независимость и самостоятельность	48	Держится с людьми на расстоянии, избегает близкого межличностного общения
77	Откровенен и искренен в общении с другими людьми	68	Высоко тревожен
53	Способен проникнуть в суть важной проблемы	37	Лживый и лукавый, ненадежный
51	Искренне ценит интеллект и знание ¹	36	Склонен к негативизму, противоречиям, созданию препятствий, любит «совать палки в колеса»
33	Спокоен, мягок в обращении	38	Враждебен к другим лицам ³
17	Ведет себя тактично и участливо	76	Склонен проецировать свои чувства и мотивы на других
3	Имеет широкий круг интересов ²	97	Эмоционально слеп. Имеет сглаженные, слабые аффекты

¹ Способности или собственные достижения здесь не рассматриваются.

² Здесь не существенно, глубокие это интересы или поверхностные.

³ Здесь имеется в виду враждебность вообще, формы выражения указаны в других карточках.

Свойства, определяющие *отрицательный* полюс приспособленности, расположены в правой части таблицы. Трудно быть уязвимым одновременно по всем указанным здесь пунктам. Так, например, трудно представить себе человека, очень тревожного и в то же время со сглаженными эмоциями.

Для целей же настоящего исследования реально полезным оказался только положительный набор черт. По нему вычислялась степень приближения данного испытуемого к *идеальной оптимальной* приспособленности. Из общего числа — 171 человек — было выделено две группы: хорошо и плохо приспособленных лиц. В первую группу было отнесено 30 % всех испытуемых с наилучшими оценками, а во вторую группу — тоже 30 % от всей выборки испытуемых, получивших наихудшие оценки по суммарному показателю оптимальной приспособленности. Всего 24 мужчины и 24 женщины были оценены как имеющие высокую оптимальную приспособленность, 24 мужчины и 24 женщины — как имеющие низкую оптимальную приспособленность (Siegelman et al., 1970, с. 285). Обратите внимание: средняя группа испытуемых далее не фигурировала в этом исследовании!

Оценка предпосылок

В активном эксперименте психологическая приспособленность выступала бы как *зависимая переменная*. Давайте рассмотрим теперь факторы-предпосылки, которые соответствовали бы *независимой переменной*.

Оценка атмосферы родительской семьи. Когда самим испытуемым было от 21 до 36 месяцев от роду, некоторых, хотя не всех, родителей посещали психологи и работники социальных учреждений. Они наблюдали «обращение родителей с ребенком». В результате по ряду пунктов была получена средняя балльная оценка. Среди них, например, готовность к браку или раздражительность, проявляемая как матерью, так и отцом.

Оценка матерей. Когда испытуемым было от 1 до 5 лет, психолог опрашивал всех матерей от двух до четырех раз. Были подвергнуты оценке как интеллектуальные, так и эмоциональные их свойства.

Воспоминания об окружающей обстановке детства. Взрослые испытуемые интенсивно опрашивались психологом. При этом «их поощряли вспоминать о характере родителей, их семейной жизни и вообще обо всей атмосфере в семье» (Siegelman et al., 1970, с. 185). После каждого опроса психолог классифицировал 93 карточки с описанием свойств окружающей обстановки: например, теплая и ориентированная на чувства; воспитание с помощью угроз или физических наказаний; придание большого значения положению в обществе, власти, материальным ценностям и т. д. Как

и при измерении психологической приспособленности, карточки с обозначением обстоятельств, наиболее соответствовавших данному испытуемому, помещались в первые несколько классов, а наименее соответствующие — в классы 8 и 9.

Прочие данные. И наконец, была получена информация об IQ (коэффициенте интеллектуальности) испытуемого, социально-экономическом положении семьи, числе браков каждого родителя и т. д.

Результаты

Мы рассмотрим только главные результаты, чтобы не потеряться в деталях. Как утверждают авторы, «в целом как для мужчин, так и для женщин семьи, порождающие высокую степень оптимальной приспособленности испытуемых, были более демократичными, более открытыми, с большей сексуальной совместимостью родителей, с более свободным обсуждением проблем и открытым выражением чувств (например, отцы у мужчин с высокой приспособленностью даже во время гнева были более открытыми), с большей согласованностью мировоззрения и с большей ориентацией на нематериальные и высшие культурные ценности. Эта значительная открытость, так же как и интеллектуальная направленность и большое удовлетворение от своей материнской роли, особенно проявлялась у матерей. Семьи испытуемых с низкой приспособленностью были противоположны по всем этим переменным; в целом их семьи носили признаки конфликта и дисгармонии, и матери были очень неудовлетворены своей ролью» (там же, 1970, с. 287).

Данные табл. 9.2 показывают связь между последующей приспособляемостью испытуемых и жизнью их родителей. В «Оценке атмосферы родительской семьи» мы обнаруживаем, что у высокоприспособленных испытуемых родители также хорошо приспособлены и обладают выраженным чувством «семейности». Для плохо приспособленных испытуемых характерно неприятное сочетание беспокойной матери и сердитого отца. В «Оценке матерей» оказывается, что ярко выраженный «собранный» тип характерен для группы с высокой приспособляемостью, а неорганизованная мать попадает в плохо приспособленную группу.

Для группы с высокой приспособленностью воспоминания детства наполнены теплом, уверенностью и этическими переживаниями в стиле старых фильмов Энди Харди, у группы же плохо приспособленных испытуемых воспоминания больше похожи на переживания антигероя современных фильмов, полного жалости к самому себе, признающего только авторитарность и материальные ценности (обольстительная мать здесь, действительно, часто появляется как неожиданность). И наконец, из «Прочих данных»

Таблица 9.2

Характеристики семей хорошо и плохо приспособленных испытуемых

Группа хорошо приспособленных	Группа плохо приспособленных
Оценка атмосферы родительской семьи	
Хорошая сексуальная совместимость	Усталая мать
У родителей близкие культурные интересы	Беспокойная мать
Мать свободна в приобретении сексуальной информации	Отец отстраняется и замыкается во время конфликта
Отец открыто реагирует на конфликт	
Оценка матерей	
Быстрое и точное мышление	Неорганизованная, болтливая
Интеллигентная, с живым умом	
Дружелюбная, искренняя и честная	
Склонная критиковать ребенка	
Уравновешенная и удовлетворенная	
Воспоминания об окружающей обстановке детства	
Дом и мать: теплота, чувство локтя, любовь и нежность	Дом и мать: безрадостность, разлад
Подчеркивается этическая ответственность	Родители часто ссорятся Большое значение придается социальному положению, власти и материальным ценностям
Матери нравится ее материнская роль	Мать нервная и беспокойная
	Мать авторитарная и категоричная
	Родители не воспитывают в детях независимости
	Мать обольстительна
Статистические данные	
Более высокий уровень образования	Низкий коэффициент интеллектуальности (IQ)

мы видим, что хорошо приспособленные испытуемые, как правило, способны и образованны.

Сопутствующее смешение с побочными переменными

Давайте сначала примем, что корреляции существующей приспособленности с переменными — предпосылками были описаны правильно. Насколько мы можем быть уверены, что если бы оказалось возможным провести активный эксперимент, то была бы обнаружена связь такого же типа? Если бы в эксперименте использовались хорошие методы подбора групп, то группы были бы уравнианы по социально-экономическому положению, интеллекту и даже по психологической приспособленности родителей. Но вспомните, ведь нам надо было бы найти способы убедить одну группу родителей пользоваться так называемыми хорошими методами воспитания, а другую группу — плохими. Только тогда мы имели бы возможность сказать, что именно практикой воспитания детей определяются различия зависимой переменной, т.е. психологической приспособленности. Но вследствие того что две группы не уравнивались по всем остальным переменным, у нас остается возможность для ряда других выводов.

Один вывод будет согласовываться с первым в том пункте, что практика воспитания детей в семьях хорошо приспособленных испытуемых и в самом деле лучше. Однако, согласно ему, как практику воспитания, так и будущую приспособленность ребенка *независимо* определяет совсем другая переменная, скажем, интеллект или социально-экономическое положение родителей. Например, в той мере, в какой интеллект врожден, дети более интеллектуальных родителей окажутся и сами более интеллектуальными. А это, в свою очередь, позитивно повлияет на психологическую приспособленность. У более интеллектуальных родителей ценности будут выше и практика воспитания лучше, но не это определит будущую приспособленность детей.

Другой возможный вывод состоит в том, что никакой действительной разницы в методах воспитания в хорошо и плохо приспособленных группах не было, а различие только показалось психологу. Интеллигентные люди умеют произвести приятное впечатление. Это также справедливо для людей с высоким социально-экономическим статусом. Они кажутся «такими милыми». В результате «гармония» в доме могла быть результатом искусного сокрытия реальной напряженности.

«Воспоминания об окружающей обстановке детства» несут с собой новые возможности смешения. Эта часть данных обнаруживает специфические трудности исследований, базирующихся на материалах памяти. Первая смешивающаяся переменная — это *предубежденность* самого экспериментатора. Вспомните, что две груп-

пы отбирались по результатам опроса испытуемых психологами. Потом другой психолог опрашивал тех же испытуемых. Этот второй психолог также мог оценить приспособленность данного испытуемого. И нет ничего удивительного в том, что эта оценка могла повлиять на оценку домашней обстановки. Психолог мог для хорошо приспособленных испытуемых вообразить более благоприятную обстановку, а для плохо приспособленных испытуемых — более неблагоприятную. Балльная оценка домашней обстановки — это, конечно, не оценка «вслепую», как она описана в главе 5.

Вторая смешивающаяся переменная — это, как ни странно, *предубежденность* самого *испытуемого*. Очень вероятно, что хорошо приспособленный человек будет иметь более благоприятные воспоминания о своем детстве, чем плохо приспособленный. Ведь оценка окружающего мира — в прошлом, настоящем и будущем — это один из основных показателей психологической приспособленности.

Последняя возможность неудачной интерпретации результатов касается уже не внутренней валидности, а внешней. Это может произойти, если приспособленность плохо *операционально* репрезентирована в зависимой переменной. Предположим, что ребенок научился у родителей производить на посторонних хорошее впечатление. Тогда, подобно тому, как родители «провели» одно поколение психологов, их дети могут «провести», т.е. ввести в заблуждение, новое поколение психологов.

Вообще говоря, внешнюю валидность можно усилить, если меньше полагаться на субъективную оценку и больше обращать внимание на объективные показатели приспособленности, которые широко предоставляются повседневной жизнью.

Что же касается внутренней валидности, то в данном исследовании основная угроза проистекала из возможности смешения с другими переменными. И эта возможность здесь никак не контролировалась.

Сейчас мы займемся корреляционным исследованием, где исследователи хорошо осведомлены о других возможных детерминантах поведения и имели возможность организовать их контроль.

ИССЛЕДОВАНИЕ ПОРЯДКА РОЖДЕНИЯ И ИНТЕЛЛЕКТА

Только в редчайших случаях появляется возможность предположить поведенческих различий у целой популяции. Такая возможность представилась Л. Белмонт и Ф. Маролле (Belmont L., Magolla F. A., 1973) в Нью-Йоркском государственном институте душевного здоровья. В их руках оказались материалы обследования интеллектуальных способностей 400 000 молодых людей в Нидерландах, содержащие также дополнительную информацию о порядке рождения каждого испытуемого, о величине семьи, о со-

циальном положении отца. Все это произошло в результате стечения обстоятельств. Упомянутые молодые люди родились в голодные годы (1944—1947) Второй мировой войны. Массовое обследование было проведено с целью выяснить, как тяжелые условия в детстве влияют на последующее умственное и физическое развитие. К счастью, особого вредного влияния не обнаружилось. Все молодые люди в Нидерландах по достижении 19 лет оказались более или менее пригодными к службе в вооруженных силах, и успешно прошли проверку интеллектуальных (так же как и специальных) способностей.

Сэр Фрэнсис Гальтон (Galton F., 1970), двоюродный брат Чарлза Дарвина и тоже блестящий биолог, давно отметил существование непропорционально большого числа перворожденных, достигших выдающихся успехов в науке. Позже аналогичные наблюдения были сделаны в различных областях человеческой деятельности. Относительно недавно У.Альтус (Altus W. D., 1965) нашел, что у молодых ученых наблюдается систематическая связь между порядком рождения и показателем по тесту на вербально-логические способности. Это исследование, основанное на солидных статистических данных по 1878 испытуемым, повлекло за собой ряд более детальных и обширных обследований. Однако, во-первых, другие исследователи не получили такого же определенного результата. Во-вторых, такого рода связь была установлена только для небольшой части популяции. В-третьих, здесь существует возможность смешения с любыми другими переменными, влияющими на величину семьи. Одной из таких переменных является социально-экономическое положение семьи: семьи, находящиеся ниже в этом отношении, как правило, больше по численности. Таким образом, по мере того как вы переходите от второго по порядку рождения ребенка к третьему, четвертому и т.д., растет относительное количество детей из бедных семей.

Уже хорошо установлена зависимость показателя интеллекта от высоты социально-экономического положения семьи. Вследствие этого необходимо провести исследование, чтобы показать, что эффект порядка рождения не связан с величиной семьи. Далее, было бы интересно выяснить, является ли влияние численности семьи чем-то большим, нежели просто отражением социально-экономического положения. Поразительно, что эта тенденция не проявилась в исследовании У.Альтуса: сыновья, рожденные первыми в семьях с четырьмя детьми, при тестировании показывали немного лучшие результаты, чем в семьях с двумя или тремя детьми.

Использовавшиеся показатели

Величину семьи и порядок рождения для каждого индивидуума очень легко было установить по актам регистрации. Следует

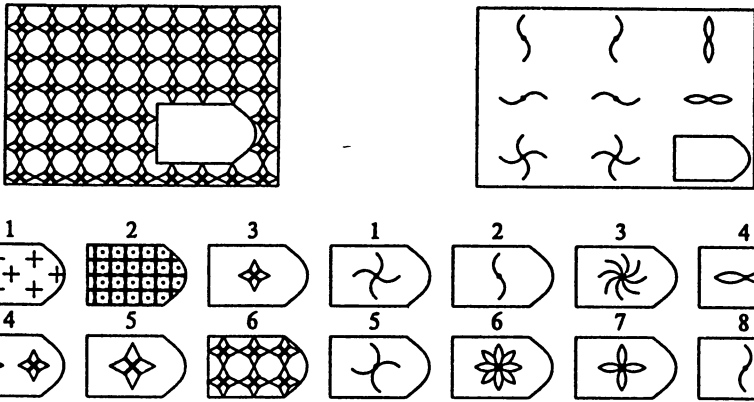


Рис. 9.1. Примеры задач из матриц Равена (Raven J. C., 1947). Нужно правильно заполнить матрицу, выбрав одну из шести фигур (Anastasy A., 1976)

заметить, что использование таких данных допустимо только при абсолютной гарантии истинности имени каждого тестируемого лица. Для описания социального положения отца использовались три широкие социальные категории: «интеллигенция» (специалисты и «белые воротнички»); рабочие (специалисты, полуубу-ченные и чернорабочие) и работники сельского хозяйства (фер-меры и сельскохозяйственные рабочие) (Belmont L., Marolla F. A., 1973, с. 1097).

Оценки интеллекта были получены у испытуемых в возрасте 19 лет. В данном исследовании использовался Датский вариант прогрессивных матриц Равена (Raven J. C., 1947). Образец одной из задач представлен на рис. 9.1. Показатели интеллекта были разби-ты на шесть классов, или уровней; лучшие помещались в класс 1 и худшие — в класс 6. Таким образом, каждый испытуемый полу-чал после тестирования интеллекта оценку 1, 2, 3, 4, 5 или 6.

Обнаруженные различия в группе

Все полученные результаты представлены на трех графиках на рис. 9.2.

На вертикальных осях отложены средние оценки интеллекта по всей группе. График (а) наглядно демонстрирует влияние численности детей и социального положения семьи. Показатель ин-теллекта, по Равену, в общем уменьшается по мере того как се-мья увеличивается (исключение составляют семьи с одним ре-бенком). Одновременно показатели тем выше, чем выше уровень социального положения. На графике (б) мы видим влияние по-рядка рождения — регулярное понижение среднего показателя интеллекта по мере продвижения по ряду вплоть до девятого ре-

банка (сюда относятся и еще большие семьи). Таким образом, были показаны влияния на интеллект численности семьи и порядка рождения при условии, что смешивающаяся переменная — социальное положение — сохранялась постоянной, т. е. контролировалась статистически. И точно так же при сохранении постоянными численности семьи и порядка рождения показано влияние на интеллект социального положения.

До тех пор, пока мы не рассмотрим несколько более сложную группу кривых на третьем графике (в), мы не избавимся полностью от связи между числом детей в семье и порядком рождения. Это потому, что, как уже указывалось, более поздние порядки рождения представлены более многодетными семьями. Однако график (в) показывает, что порядок рождения оказывает влияние даже при сохранении постоянной численности семьи: каждая линия, которая представляет одну численность детей в семье (ЧД),

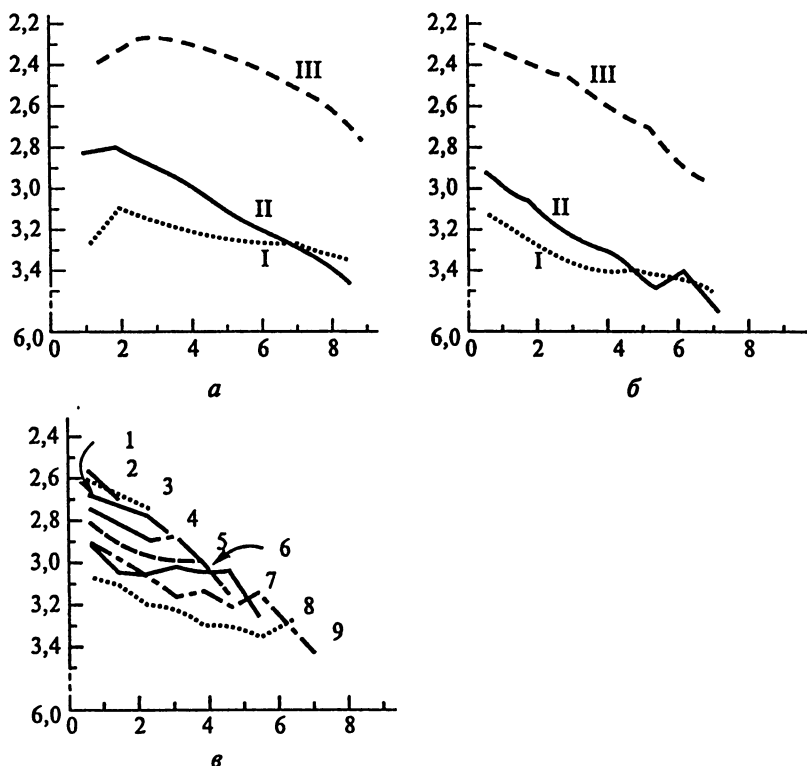


Рис. 9.2. Различия, обнаруженные в исследовании очередности рождения (Belmont L., Marolla F.A., 1973): I — крестьяне, II — рабочие, III — интеллигенция. Ось абсцисс: а — количество детей в семье; б, в — очередность рождения; ось ординат — оценки по тесту Равена (1—6). Цифры на рис. (в) означают численность детей

заметно понижается по мере увеличения порядка рождения. Из этой группы кривых следует также, что влияние размера семьи не есть следствие большей представленности многодетных семей в более поздних порядках рождения. Хотя такой вывод можно сделать при сравнении кривых для двух и трех детей (ЧД 2 и ЧД 3), поскольку положение точки для третьего ребенка на линии ЧД 3 лежит на продолжении линии ЧД 2, тем не менее для семей с большим числом детей это совсем не так: все линии для семей с числом детей 4, 5, 6, 7, 8 и 9 лежат ниже предыдущих.

Объяснение

Давайте ограничимся анализом связи между порядком рождения и интеллектом. Связь же между интеллектом и социальным положением и величиной семьи подведет нас к известной проблеме наследственных и средовых факторов, для обсуждения которой мы располагаем не очень хорошими данными (по сравнению, скажем, с исследованием близнецов). Белмонт и Маролле (Belmont L., Marolla F.A., 1973) не сделали никакой реальной попытки объяснить связь порядка рождения с интеллектом. Их задачей было только статистическое доказательство этой связи.

Нет никаких оснований ожидать у детей из одной семьи врожденной систематической разницы (в среднем) в чертах характера и способностях. Поэтому для влияния порядка рождения на интеллект возможны два объяснения. Одно состоит в том, что порядок рождения создает различия в жизненном опыте подрастающего ребенка. Другое — что эмбриональные условия благоприятнее для детей, родившихся раньше, чем для более поздних. Первое предположение (окружающая обстановка после рождения) основывается на том, что родители уделяют гораздо больше внимания своему первому ребенку и постепенно все меньше и меньше — последующим, что последующие дети учатся говорить у других детей больше, чем у взрослых, и/или что младшие дети обижаются или каким-то иным образом притесняются старшими детьми. Что касается второго объяснения (пренатальные причины), то виной здесь, по мнению авторов, «является ослабление воспроизводящих функций матери с увеличением числа рожденных детей» (Belmont L., Marolla F.A., с. 1101).

Можно выдвинуть еще одно объяснение, которое охватывает период как перед рождением, так и после него. Оно состоит в том, что каждый последующий ребенок имеет возрастающую вероятность быть нежелательным. Это может повлиять на заботу матери о себе в период беременности, например в отношении питания или воздержания от табака или алкоголя. Это может сказываться также на любви и внимании к ребенку после его рождения.

Насколько эффективным был контроль?

Основной целью исследования была проверка гипотезы о влиянии порядка рождения на интеллект. Другие переменные, которые также влияют на интеллект, а именно величина семьи и социально-экономическое положение, сохранялись постоянными путем создания однородных групп. Мы уже видели, что возможно более чем одно объяснение влияния порядка рождения на интеллект при условии, что не было допущено смешение с другими переменными. Давайте же посмотрим, насколько выполнено было это условие.

Что было проконтролировано. Исследователи преуспели в организации контроля двух дополнительных переменных, которые можно было бы просмотреть даже в активном эксперименте, планирующем будущий мир. Допустим, что все исследуемые пары предполагается зарегистрировать в 1983 г. Тогда через несколько лет можно было бы провести сравнение между их точно запланированными первенцами, вторыми детьми, третьими и т.д. Однако и при этом имело бы место смешение с побочными переменными.

Предположим, что дети родились бы в 1985, 1987, 1989 и т.д. Если их всех провести через тесты одновременно, допустим, в 2010 г., то перворожденным будет 25 лет, вторым — 23 года, третьим — 21 год и т.д. Такого смешения с возрастной переменной можно было бы, конечно, избежать, если бы всех испытуемых провести через тест в одном и том же возрасте, скажем, в 19 лет, как в датском исследовании. Но тогда разные группы испытуемых оказались бы принадлежащими к несколько различным поколениям (т.е. людям, родившимся примерно в одно и то же время). Возможно, что пока дети будут взрослеть, изменятся условия в школах. Образование поколения 1985 г. будет несколько отличаться от образования поколения 1989 г.

В исследовании Белмонт и Маролле (Belmont L., Marolla F. A., 1973) все испытуемые тестировались в одном возрасте, так что возрастная переменная контролировалась. Все они принадлежали примерно к одному поколению, поскольку родились с разницей не более трех лет. Поэтому не могло быть никакой систематической связи между порядком рождения и разными поколениями: девятый ребенок в семье имел такую же вероятность оказаться рожденным в 1944 году, как и в 1947.

Что не было проконтролировано. Мы уже познакомились с трудностью контроля всех возможных побочных переменных в исследовании на оптимальную приспособленность. Здесь также, несмотря на чувствительность исследователей к возможным смешениям, все равно остались переменные, которые не были проконтролированы.

Одна возможная переменная, которая оказалась неучтенной, — это возраст родителей. Можно показать (если придерживаться постнатального типа объяснения), что у более младших членов семьи более старые родители, которые менее жизнерадостны и гибки в своем поведении. Если придерживаться пренатального типа объяснения, то можно сказать, что воспроизводящие функции матери ухудшаются не потому, что она до этого рожала детей, а потому, что она просто стареет.

Другой возможной смешивающейся переменной, также не контролирувавшейся в исследовании, является брачный стаж родителей. Общеизвестно, что с течением лет эмоциональная атмосфера дома меняется. Дети по мере отдаления от первого по порядку рождения оказываются и более далекими от медового месяца.

И последнее замечание связано с контролем социального положения, который не был полным. Очевидно, что внутри каждого социального слоя существуют такие же большие различия, как и между слоями. Таким образом, социальное положение не контролировалось (а скорее всего, и не могло контролироваться) в той мере, в какой это можно было сделать для величины семьи, возраста родителей или брачного стажа.

Многие из вас, наверное, удивятся тому, что все испытуемые были мужчинами. Дело, конечно, не в мужском шовинизме исследователей. Просто необходимые данные имелись только для мужчин. Поэтому мы не можем сделать вывод, что порядок рождения имеет такое же влияние на интеллект женщин. Таким образом, внешняя валидность здесь ограничена.

КОНТРОЛЬ В КОРРЕЛЯЦИОННЫХ ИССЛЕДОВАНИЯХ

Давайте теперь, основываясь на описанных двух работах, рассмотрим общую проблему контроля в корреляционных исследованиях. Проблема здесь та же, что и в активных экспериментах, — систематическое смещение с побочной переменной. Однако в корреляционном исследовании мы не можем *активно организовать* контрольные условия, а вместо этого должны воспользоваться *статистическим* контролем смещения. Мы не будем начинать с групп, одинаковых по другим переменным, а постараемся по ходу дела уравнивать группы по этим переменным. Уравнивание групп в корреляционных исследованиях производится двумя способами

Подбор пар испытуемых. В случае если число испытуемых невелико, чаще всего применяется индивидуальный подбор пар. В исследовании психологической приспособленности это вылилось бы в подбор пар хорошо и плохо приспособленных испытуемых, которые были бы сходны по интеллекту родителей и социально-экономическому положению. Допустим, что нам удалось бы най-

ти 50 таких пар. В этом случае между группами уже не было бы различий по названным двум побочным переменным.

Данный метод связан с двумя следующими трудностями.

1. Если побочная переменная и в самом деле существенна, то при подборе групп должна обнаружиться диспропорция в количестве хорошо и плохо приспособленных. Так, большинство хорошо приспособленных испытуемых могут происходить из семей с высоким социально-экономическим положением. И это создаст трудности для нахождения большого числа уравненных пар: среди испытуемых с высоким социально-экономическим положением будет слишком мало людей с плохой психологической приспособленностью. Эта трудность станет особенно серьезной, если мы попытаемся уравнивать индивидов не по одной, а по нескольким побочным переменным. Число соответствующих пар очень резко сократится. Теоретически никогда нельзя быть уверенным, что проконтролированы все значимые побочные переменные. Практически мы крайне ограничены в числе переменных, которые действительно можем проконтролировать.

2. Другая трудность тесно связана с первой — это непредставительность выделенных для исследования испытуемых. Предположим, что для высокоприспособленного индивида характерно иметь родителей интеллигентных, с высоким социально-экономическим положением, а также с хорошей собственной психологической приспособленностью. Однако большая часть таких испытуемых будет отвергнута в процессе составления пар. В то же время это будут те самые испытуемые, чьи родители применяли хорошие способы воспитания. Таким образом, процедура уравнивания может нивелировать влияние процессов воспитания.

Итак, метод подбора пар несет в себе две опасности: мы можем осуществить недостаточный контроль и мы можем переусердствовать с контролем. Недостаточный контроль приведет к тому, что будет упущена значимая побочная переменная, а избыточный контроль — к тому, что диапазон изменений действительной независимой переменной будет ограничен.

Составление однородных подгрупп

В связи с тем, что исследование влияния порядка рождения на интеллект проводилось на очень большом числе испытуемых, в нем не было необходимости составлять индивидуально уравненные пары. Вместо этого были составлены *однородные подгруппы*, которые были уравнены по всем переменным, кроме одной, интересующей исследователей. Так, различие между вторым и пятым ребенком могло сравниваться внутри однородных подгрупп семей, содержащих пять детей, шесть детей, семь детей и т. д. Если бы этого не делалось (т. е. сравнивались бы все дети, родившиеся

вторыми, со всеми пятыми детьми), то произошло бы смешение с величиной семьи. Ведь пятые дети есть только в больших семьях, тогда как второй ребенок есть и в большой, и в маленькой семье.

Основанием для образования подгрупп могли бы быть также и другие переменные, имеющие шанс оказаться значимыми, но опущенные в данном исследовании, например возраст матери и др. Это лучше всего было бы сделать после распределения испытуемых по группам на основании численности семьи. Так, например, в семьях с пятью детьми можно было сравнить между собой детей — от первого до пятого включительно — только таких, матери которых достигли к моменту их рождения 23 лет. Точно так же сравнение могло быть сделано между детьми для 24-летних матерей и т.д. И если влияние порядка рождения на интеллект исчезло бы при сохранении постоянным возраста матери, то мы должны были бы признать, что имели дело с эффектом не порядка рождения, а возраста матери. Более вероятно, конечно, что порядок рождения скажется на различиях между испытуемыми, даже если возраст матери и обнаружит свой самостоятельный вклад.

Переменную брачного стажа тоже можно проконтролировать с помощью однородных подгрупп. Такие подгруппы можно было бы образовать для одно-, двух-, трехлетнего стажа и т.д. к моменту рождения исследуемого ребенка.

Если для различных однородных подгрупп обнаруживается различное влияние интересующей нас переменной, то это может привести нас к более глубокому пониманию механизмов действия данной переменной. Давайте возьмем влияние только очередности рождения как таковое вне связи с возрастом матери или брачным стажем. На рис. 9.2, б мы видим реальные примеры различного влияния этой переменной. Для подгруппы «интеллигенция» наклон кривой оказывается достаточно крутым. То же самое — для подгрупп «рабочие». Однако для подгруппы «работники сельского хозяйства» график оказывается более пологим.

Это, конечно, говорит о том, что существует взаимодействие между влиянием социального положения и влиянием порядка рождения на показатель интеллекта. В предыдущей главе мы уже встречались с примерами взаимодействия. И нет ничего удивительного, что оно обнаруживается в корреляционных исследованиях так же, как и в активных экспериментах. Для вычисления основного результата действия и взаимодействий мы можем использовать тот же самый тип таблиц.

В табл. 9.3 приводятся соответствующие данные для двух уровней социального положения (однородные подгруппы) и двух уровней порядка рождения (1 и 9). Элементами таблицы являются средние значения показателей интеллекта для данной подгруппы. Вы должны помнить, что наилучшая оценка интеллекта была равна

1, а наихудшая — 6. Основной результат действия «порядка рождения», как видите, равен величине 0,59, а результат действия социального положения — 0,67. Взаимодействие «порядок рождения и социальное положение» равно достаточно ощутимой величине — 0,26.

Возможно, что взаимодействие стало бы еще больше, если бы из «работников сельского хозяйства» мы исключили владельцев больших механизированных ферм и земельных управляющих, оставив только сельскохозяйственных рабочих. Возможно также, что в этом случае вообще не наблюдалось бы падения показателя интеллекта в связи с порядком рождения. Иначе говоря, возможно, что порядок рождения — фактор не универсальный, а действующий скорее в городских семьях или, несколько шире, в семьях, не занятых сельским хозяйством. Но уже приведенные данные заставляют учитывать указанное взаимодействие при любом объяснении влияния порядка рождения. Эту информацию о взаимодействии оказалось возможным получить потому, что контроль основывался на методе однородных подгрупп. В случае же использования метода подбора пар такая информация оказывается недоступной.

Возможное объяснение этого взаимодействия заключается в том, что новый член семьи более «желанен» в семье с небольшим, маломеханизированным хозяйством, чем в других семьях. На маленькой ферме каждый новый ребенок представляет собой

Таблица 9.3

Основные результаты действия порядка рождения и социального положения на показатель интеллекта и взаимодействие этих эффектов

Социальное положение	Порядок рождения			Основной результат действия социальное положение
	1	9	среднее	
Интеллигенция	2,28	3,00	2,64	0,67
Работники сельского хозяйства	3,08	3,54	3,31	
Среднее	2,68	3,27	2,975	
Основной результат действия				
Порядок рождения	0,59			
Взаимодействие				
Порядок рождения × × социальное положение:	(3,00 – 2,28) – (3,54 – 3,08) = 0,72 – 0,46 = 0,26			

экономическую ценность. В других же местах новый ребенок означает новые обязанности и еще большую тесноту. Поэтому для сельской семьи восьмой ребенок вполне может быть желанным, но для городской семьи это маловероятно. Возможно, что накапливающееся влияние желательности или нежелательности ребенка к 19-летнему возрасту проявляется в показателе интеллекта.

Теоретически у родителей можно выяснить, хотели они данного ребенка или нет. Допустим, что мы это сделали. Тогда мы сможем выделить в группах семей с одним и тем же количеством детей однородные подгруппы желанных и нежеланных детей для каждого номера в очередности рождения. Допустим далее, что между вторым и пятым ребенком в семьях с пятью детьми нет различия, если они попадают в одну и ту же подгруппу желанных или нежеланных. Тогда вполне можно заключить, что влияние порядка рождения оказалось ослаблено в связи с тем, что пятые дети были такими же желанными, как и вторые.

Однако совсем *не обязательно* делать такой вывод. Вспомним, что в корреляционном исследовании мы имеем дело только с корреляциями — и это тоже одна из корреляций. Возможно, что сельская жизнь менее изматывает, чем городская. Горожанин в общем более утомлен, и поскольку с появлением ребенка связаны новые заботы и новые усилия, понятно его нежелание иметь еще одного ребенка. Здоровый фермер не чувствует этих опасений. Поздний ребенок в городской семье будет страдать от недостатка энергии у родителей, а в сельской семье этого не произойдет. Таким образом, основу эффекта порядка рождения составят не различия в желательности — нежелательности, а скорее величина энергии родителей, которая лишь *коррелирует* с желанием завести нового ребенка.

Итак, благодаря методу составления однородных подгрупп в корреляционных исследованиях оказывается возможным многосторонний контроль. Тем не менее мы никогда не можем знать, действительно ли данный фактор-предпосылка влияет на выбранный поведенческий показатель. Как и в случае подбора пар испытуемых, у нас нет способа узнать, учтены ли все значимые побочные переменные. Более того, когда мы наконец свели наблюдаемый эффект к той переменной, которая кажется решающей, все еще остается возможность, что реальной детерминантой поведения была какая-то другая коррелирующая с ней переменная. Эта трудность и породила известное высказывание, что корреляцию не следует путать с причинностью.

ИССЛЕДОВАНИЕ С ЦЕЛЬЮ ОТБОРА КОНТРОЛЕРОВ

Давайте рассмотрим искусственный пример, как можно применить корреляционное исследование в практических целях. Су-

ществует предприятие, которое сталкивается с серьезной проблемой контроля за качеством сложных механических агрегатов. Большинство контролеров, принимающих агрегаты, пропускают дефекты. Когда их просят работать тщательнее, они теряют уверенность в себе и начинают отбраковывать агрегаты, которые при последующей проверке оказываются вполне удовлетворительными. Эти контролеры неглупы и хорошо мотивированы, но кажется, что им не хватает какой-то специфической способности.

Решить проблему, просто нанимая много контролеров и оставляя только тех, которые работают удовлетворительно, нельзя. Во-первых, это слишком невыгодно с экономической точки зрения — ибо многие окажутся непригодными для работы. Во-вторых, отвергнутые контролеры будут лишены ценного опыта, который за это время они могли бы приобрести на какой-нибудь другой работе. Проблема исчезла бы, если бы 80 % принятых на работу контролеров справлялись с ней успешно.

Такая задача поставлена перед начальником отдела кадров, который имеет опыт обращения с тестами на способности. Он узнаёт, что может за плату приобрести подходящий тест. Последний состоит из чертежей, на которых детали соединены различным образом. В каждом наборе есть один чертеж, на котором какой-нибудь угол между деталями или узел соединения отличается от стандартного чертежа. Задача состоит в том, чтобы найти ошибку в неверном чертеже. Оценки могут распределяться от 0 до 85. Фактически лишь немногие получают оценку ниже 40 или выше 80.

Методика

Людам, согласным на любую работу, сообщают, что они могут получить желаемую работу контролера; если же они не справятся, то перейдут на другую работу. В целом через тест на способности проводится 60 кандидатов. Затем все они половину времени работают контролерами, а оставшуюся половину — на другой должности. После того как они проработали в таком режиме три месяца, в течение четвертого месяца производится оценка их работы в качестве контролеров. Для этого регистрируется число блоков, которые они проверили, и процент сделанных ими ошибок. Возможны два типа ошибок. Во-первых, пропуск дефекта на любой из 40 дефектных деталей агрегата. Во-вторых, обнаружение дефекта на других 40 исправных деталях. Все испытуемые знают, что их работа будет оцениваться. *Итоговая оценка* каждого испытуемого означала число деталей, проверенных в течение последних 20 дней, минус учетверенное количество ошибок. Так, контролер, который проверил 800 деталей с 5 % ошибок, получает оценку 640, т. е. $800 - 4 \times 800 \times 0,05$. Удовлетворительными считаются оценки выше 675.

Результаты

Оценки каждого испытуемого по *тесту на способности* и по *качеству работы* могут быть представлены на диаграмме разброса (рис. 9.3, а). Каждое число на диаграмме обозначает количество испытуемых, которые имеют данную комбинацию тестовой оценки (ось абсцисс) и оценки работы (ось ординат). Например, число «1», обведенное кружком, означает, что существует только один

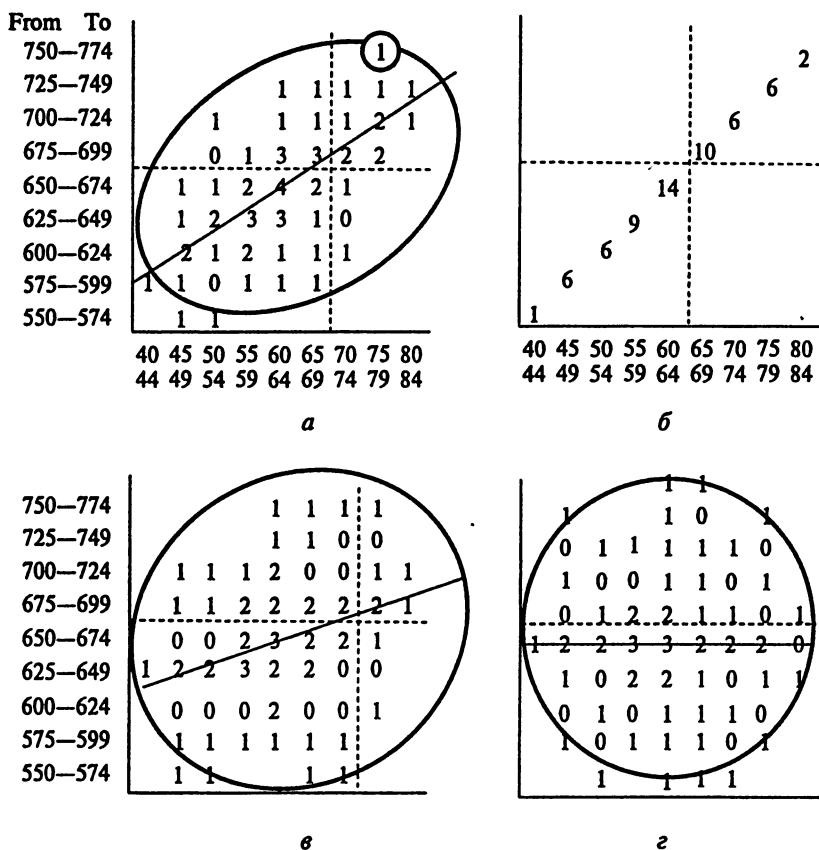


Рис. 9.3. Корреляционное поле тестовых и рабочих оценок выполнения работы. Рабочие оценки, расположенные выше горизонтальной пунктирной линии, характеризуют успешное выполнение. Корреляция между тестовыми и рабочими оценками: а) 0,60; б) 1,00; в) 0,30; г) 0. В случаях (а, б, в) тестовые оценки, расположенные справа от вертикальной линии, позволяют предсказать успешное выполнение работы; в случае (г) такое предсказание невозможно. Ось абсцисс — тестовые оценки (в интервалах); ось ординат — рабочие оценки (в интервалах)

испытуемый, у которого тестовая оценка находится между 75 и 79, а оценка качества работы — между 750 и 774.

В общем все оценки охватываются овалом, который вытянут слева направо и вверх. Это значит, что оценки положительно коррелируют. Величину корреляции можно вычислить. Метод вычисления описывается в статистическом приложении в конце этой главы. Для диаграммы разброса, приведенной на рис. 9.3, а, величина коэффициента корреляции, характеризующего степень связи тестовых и рабочих оценок, равна 0,60, или, если быть пунктуальным, +0,60. В практических задачах мы не должны серьезно рассматривать возможность отрицательной корреляции. Если диаграмму на рис. 9.3, а повернуть слева направо, то мы получим значение корреляции, равное -0,60. Отрицательные корреляции, когда они обнаруживаются, чаще всего настолько невелики, что могут считаться следствием случайной флюктуации «истинного» нулевого значения. Когда же они достаточно велики, то чаще всего связаны с характером шкалы на одной из осей. Так, например, число правильных ответов по одному тесту может иметь отрицательную корреляцию с числом ошибок по другому тесту.

Как это видно из диаграммы, корреляция 0,60 отражает довольно хорошее соответствие между тестовыми и рабочими оценками, хотя, конечно, это соответствие далеко не идеальное. Некоторые испытуемые, имеющие высокую тестовую оценку, оказались худшими контролерами, чем те, у которых тестовая оценка была ниже. На рис. 9.3, б показано абсолютное соответствие между тестовыми и рабочими оценками (реально невозможное). Если один испытуемый имеет оценку по тесту на какую-то величину больше, чем другой испытуемый, то на столько же больше у него будет и рабочая оценка — если обе оценки, конечно, прошкалированы одинаково. Вычисление коэффициента корреляции для данных на рис. 9.3, б даст величину, равную +1.

Тем не менее корреляция 0,60 показывает лучшее соответствие, чем то, которое представлено на рис. 9.3, в. В этом последнем случае есть только некоторая тенденция к совпадению высоких тестовых оценок с высокими рабочими оценками, но не больше. Вычисление коэффициента корреляции здесь дает величину +0,30. Но даже эта низкая корреляция оказывается выше, чем представленная на рис. 9.3, г, где коэффициент корреляции равен 0. Здесь нет совершенно никакой тенденции к соответствию.

Назначение претендентов на должность. Возвращаясь к рис. 9.3, а, мы поймем теперь, как начальник отдела кадров может воспользоваться результатами исследования для отбора претендентов на работу в качестве контролеров. Горизонтальная линия отделяет удовлетворительную работу от неудовлетворительной. Напомним, что удовлетворительной считается оценка от 675 и выше. Если посчитать число лиц над этой линией, то их окажется 24 из 60, т. е. 40 %.

Вертикальная линия проведена чуть левее тестовой оценки 70, через точку, которая называется секущей оценкой. Справа от нее оказывается 14 человек. Из них только 12 показали удовлетворительную работу. Это дает 86 % (от 14), что несколько лучше требуемых 80 %. Если же секущую оценку сдвинуть влево до 65, то справа окажется 24 человека, из которых 17 будут удовлетворительно работающими. Таким образом, число удовлетворительных работников снизится до 71 %, что значительно меньше требуемых 80 %. Итак, начальник отдела кадров может сделать *вывод*, что на работу контролера должны назначаться только претенденты, получившие тестовую оценку 70 и выше. При этом предполагается, что связь между тестовой оценкой и качеством работы, выявленная на данных испытуемых, сохранится и для будущих претендентов

В случае абсолютной корреляции, показанной на рис. 9.3, б, вопрос о проценте удовлетворительно работающих лиц просто не возникает. Здесь нельзя сделать ошибки. Все 24 лица с тестовой оценкой 65 и выше будут работать качественно. И в будущем начальник отдела кадров с нулевым риском может принимать на работу всякого, кто получит оценку 65 и выше.

Однако в случае (в), где корреляция равна только 0,30, видно, что даже если вертикальная разделяющая линия будет сдвинута к оценке 75, начальник отдела кадров не достигнет цели: из 8 лиц, у которых тестовая оценка равна 75 и более, подходят только 6, что составляет 75 %, а не 80 %. Значит, в будущем на работу могут быть приняты только те претенденты, у которых тестовая оценка (секущая) составит 80 и выше.

Теперь вы понимаете, что происходит при изменении величины корреляций. Чем выше корреляция, тем больше людей может быть отобрано при данном правиле решения. В нашем случае правило решения означало, что 80 % из выбранных должны оказаться удовлетворительно работающими. На практике для целей отбора не пользуются перемещением линий на полученной диаграмме разброса, поскольку в ней безусловно содержится множество случайных отклонений. Вместо этого используют таблицы предсказаний. Они основаны на идеализированных диаграммах разброса, представляющих различные корреляции между тестовой и рабочей оценками (Taylor H. C., Russel J. T., 1939).

Предсказание качества индивидуальной работы. Давайте теперь посмотрим на всю ситуацию рабочей оценки, которую он вероятнее всего получит. (Между прочим, он может также узнать, насколько такое предсказание может оказаться ошибочным.)

Наклонная линия, проведенная на рис. 9.3, а, соединяет средние рабочие оценки для различных тестовых оценок. Например, для лиц, находящихся по тесту между значениями 55 и 59, средняя рабочая оценка падает на интервал от 625 до 649. Для тестовых оценок 70—74 средняя рабочая оценка приходится на интервал

675—679. Линия, соединяющая средние рабочие оценки для этих тестовых оценок, точно так же как и для всех остальных тестовых оценок, может быть использована в качестве *линии предсказания*. Любой очередной претендент, будучи проведен через тест на способности, может посмотреть на диаграмму разброса с такой наклонной линией и по ней предсказать, как он будет справляться с работой контролера. Кроме того, он заметит, что такое предсказание чревато некоторой ошибкой. Например, лица, получившие тестовую оценку в интервале 70—74, могут иметь и высокую рабочую оценку — от 725 до 749, и низкую — от 600 до 624, тогда как предсказываемая оценка находится в пределах 675—699.

В случае абсолютной корреляции (рис. 9.3, б) предсказание будет безошибочным. Каждая тестовая оценка дает однозначное предсказание рабочей оценки. В случае же низкой корреляции — 0,30 (рис. 9.3, в) — лица с тестовой оценкой от 70 до 74 могут иметь как высокую рабочую оценку — 750—774, так и низкую — 550—574 (при предсказываемой рабочей оценке порядка 675). По сравнению с более высокой корреляцией в случае (а) здесь ошибка предсказания больше. Для случая нулевой корреляции (г) предсказываемая рабочая оценка всегда равна 660, независимо от тестовой оценки. Овал, имевший место на рис. 9.3, а, здесь превращается в круг, что означает очень большую ошибку предсказания.

Вы уже, наверное, заметили, что чем выше корреляция, тем больше наклон линии предсказания. Положение линии меняется от горизонтального при нулевой корреляции до наклона с угловым коэффициентом, равным 1 (45°) при абсолютной корреляции. Нулевая корреляция означает нулевой наклон, корреляция, равная единице, означает наклон с угловым коэффициентом 1, если тестовые и рабочие оценки прошкалированы в равных единицах. Таким образом, чем выше корреляция, тем больше достоверность рабочих оценок, представленных линией предсказания. И снова эмпирически полученная диаграмма разброса с ее случайными колебаниями не может использоваться для предсказания рабочих оценок или для определения величины ожидаемой ошибки предсказания. Для этого на основании идеализированной диаграммы разброса следует составить таблицу для определенного значения коэффициента корреляции. На самом деле такое предсказание можно сделать и по простой формуле 9.3, как это показано в статистическом приложении к данной главе.

Что обеспечивает предсказывающее исследование?

Используя найденный коэффициент корреляции, можно делать довольно точные предсказания. Из корреляции между тестовой оценкой и рабочей оценкой можно вывести секущую оценку, которая должна обеспечить заданный процент удовлетворительно

работающих лиц. Мы можем также установить для каждой тестовой оценки будущих претендентов наиболее вероятную рабочую оценку, а также величину ошибки предсказания.

В данном примере один вид поведения использовался для предсказания другого вида поведения. Возможно, что в основе корреляции лежало то, что измерявшиеся способности, как и предполагалось, были существенны для данного вида работы. Однако возможно и многое другое. Например, может быть все дело было во внимательности или в количестве усилий, которое данный человек готов был затратить. В практических ситуациях нас не интересуют объяснения: нужны только результаты. При высокой корреляции предсказание будет хорошим, при низкой корреляции — плохим.

Надежность и валидность тестов

Термины *надежность* и *валидность* применительно к тестам несколько отличаются от аналогичных терминов, которые уже использовались в этой книге. Тест называется надежным, если мы можем рассчитывать, что одно и то же лицо каждый раз, при повторных испытаниях, получит примерно ту же самую оценку (относительно других). Как известно, существует много причин непоследовательности поведения, включая факторы времени, которые мы не можем контролировать.

Однако влияние этого непостоянства можно уменьшить, используя достаточно продолжительный тест, конечно, соответствующего уровня трудности.

Один из способов выявления надежности теста заключается в том, чтобы дать один и тот же тест (или очень близкие варианты, если это необходимо) дважды той же самой группе испытуемых. Если коэффициент корреляции между результатами двух применений теста высокий (например 0,90), то тест считается надежным. Однако все еще остается вопрос о надежности самого *исследования*. Это означает, что через тест нужно провести большое число испытуемых. Иными словами, надежность *теста* должна основываться на достаточной надежности самого *исследования*.

Тест называется *валидным* всегда по отношению к некоторой другой оценке, например оценке качества работы, — если он высоко коррелирует с этой оценкой (например 0,60). И снова, чтобы узнать, является ли данное заключение валидным, т. е. следует ли оно из надежного исследования, нужно использовать достаточное число испытуемых.

Типы корреляционных исследований

Мы рассмотрели три различных типа корреляционных исследований. Они, конечно, не исчерпывают всех возможных типов. Од-

нако рассматривая параметры, по которым они различаются, можно будет охватить достаточно широкий круг корреляционных исследований. Но давайте сначала вспомним, что отличает все вообще корреляционные исследования, или, точнее, чего во всех них нет. В них отсутствуют *планируемые изменения* независимой переменной.

Исследователи, изучавшие опыт воспитания детей, не убеждали одних родителей использовать хорошие методы воспитания, а других — плохие. Эти различия в методах уже *существовали*. Порядок рождения ребенка тоже не зависел от экспериментатора. Он тоже уже существовал. Точно так же индивидуальные различия по тесту на способности не задавались экспериментатором. Таким образом, корреляционное исследование — это такое исследование, в котором одни поведенческие различия соотносятся с другими, уже существующими. Давайте теперь обратимся к параметрам, по которым различаются корреляционные исследования.

Степень приближения к независимой переменной

Как уже говорилось, в принципе можно было бы спланировать такой эксперимент, в котором экспериментатор решал бы сам, какие родители будут применять хорошие методы воспитания детей, а какие — плохие. Но вследствие практической неосуществимости такого эксперимента было предпринято корреляционное исследование. Еще большую проблему представляет порядок рождения. Каким активным вмешательством можно добиться, чтобы данный ребенок родился, например, четвертым? И все же порядок рождения — это нечто, похожее на независимую переменную. Ведь это — переменная, предшествующая во времени, следовательно, она может стать причиной различий в поведении. Такого совсем нельзя сказать о тесте на способности, который использовался для предсказания качества работы контролера. Одна и та же причина (какой бы она ни была) обеспечивала данному лицу и высокую оценку по тесту, и возможность стать хорошим контролером. В различиях тестовых оценок нет ничего от независимой переменной. Фактически мы могли бы точно так же из работы испытуемого в качестве контролера вывести, насколько успешно он будет справляться с тестом на способности. И если этого не делаем, то только потому, что это не имеет практического смысла.

Описание корреляции

Все три исследования были названы *корреляционными*, но только в случае отбора контролеров вычислялся коэффициент корреляции. Этот показатель наиболее значим в том случае, если каждая из двух сопоставляемых переменных имеет непрерывное колоколообразное распределение. Это справедливо для оценок почти любого теста. Оценки распределяются от низшей к высшей непре-

рывно и имеют максимум в области среднего значения. Поэтому в исследовании контролеров коэффициент корреляции хорошо подходит для описания корреляции между двумя переменными.

Его можно было бы использовать также и в исследовании приспособленности. Каждый испытуемый имеет оценку приспособленности на почти непрерывной шкале. Почти наверняка эти оценки можно было бы аппроксимировать колоколообразным распределением. Факторы-предпосылки распределялись примерно таким же образом, хотя, в общем, были ступенчатыми, а не постепенно меняющимися. Доход семьи прямо мог бы использоваться как континуальный показатель, хотя распределение оказалось бы отрезанным со стороны высоких доходов. Однако вместо коэффициента корреляции здесь было использовано сравнение групп с высокими и низкими показателями, поскольку не совсем ясно, что значит средняя величина приспособленности.

Кривые на рис. 9.2 — это почти все, что можно сделать для описания корреляций между величиной семьи и порядком рождения — показателями интеллекта. Последний, конечно, имеет континуальное колоколообразное распределение. Однако этого никак нельзя сказать в отношении других названных переменных. Коэффициент корреляции здесь не имел бы большого смысла.

Цель

Исследования приспособленности и порядка рождения проводились для того, чтобы понять, чем определяются различия в поведении. Это не значит, что результаты исследования приспособленности не могут быть использованы в практических целях. Труднее представить сиюминутное практическое использование результатов исследования порядка рождения. (Автор этой книги уже имеет несчастье оказаться самым младшим ребенком в семье!) И конечно же, исследование по отбору контролеров преследует явную практическую цель. Так что давайте не будем «смешивать» тип исследования и его цель. Если в исследовании вычисляется коэффициент корреляции или соотносятся две оценки поведения, это еще не значит, что оно имеет практический характер. Такое делается во многих чисто теоретических исследованиях. Коэффициенты корреляции находятся между тестовыми оценками детей и родителей, между оценками идентичных близнецов и т.д. Все это — теоретические исследования, в которых пытаются разделить влияние наследственности и среды. Предпринимаются также теоретические исследования, в которых члены одной и той же группы испытуемых проводятся через разные тесты — точно так же, как это делалось при исследовании контролеров. Иногда даже используется 40 или 50 различных тестов, и между каждой парой тестов вычисляются коэффициенты корреляции. Для

выявления же значительно меньшего числа базисных переменных, адекватно описывающих различия между индивидами, применяется техника, называемая *факторным анализом*.

КРАТКОЕ ИЗЛОЖЕНИЕ

В тех случаях, когда для проверки некоторых гипотез относительно поведения нельзя использовать активный эксперимент, проводят корреляционные исследования. В двух рассмотренных нами примерах — исследованиях предпосылок оптимальной приспособленности и влияния порядка рождения на интеллект — активный эксперимент был невозможен. В исследовании же по отбору контролеров просто трудно себе представить, какие активные изменения можно было бы внести.

В результате сравнения хорошо и плохо приспособленных испытуемых по ряду параметров было обнаружено, что главный фактор для хорошо приспособленных — это хорошие методы воспитания в детстве. Однако здесь присутствовали также другие переменные, смешивающиеся с методами воспитания. Двумя из них были социоэкономический статус и интеллект родителей. Нельзя сказать, влияли ли они в действительности, так как соответствующий контроль не был организован. Другим источником возможного смешения были субъективные оценки. Поскольку оценки производились не «вслепую», оставалась возможность для предубежденности экспериментатора. Поскольку материал для оценок был получен из бесед с испытуемыми, существовала также опасность предубеждения последних — хорошо приспособленные могли быть склонны к более благоприятным воспоминаниям.

В исследовании, проведенном почти на 400 000 молодых людях, было обнаружено, что показатель интеллекта снижается вместе с порядком рождения. Этот эффект сохранился даже при нивелировании возможного влияния социального положения и размера семьи. В общем, контроль в этом исследовании был достаточно хорошим. Однако было отмечено, что социальные слои состоят из довольно разнообразных групп, которые могут различаться по значимым побочным переменным. Корреляция между показателем интеллекта и порядком рождения может быть объяснена разными причинами; их можно разделить на пренатальные и постнатальные. В этом исследовании не были проконтролированы две переменные: первая — возраст родителей, особенно возраст матери, который может оказаться очень важным; вторая — брачный стаж родителей к моменту рождения испытуемого.

Методы контроля в корреляционных исследованиях могут быть двух видов. Первый — индивидуальный подбор пар испытуемых. Если сравниваются две группы, то каждому индивиду первой группы подбирается индивид второй группы, который имеет тот же уровень определенных побочных переменных. Данный метод связан с двумя трудностями. Первая: чем больше переменных, по которым уравниваются испытуемые, тем меньше испытуемых оказывается в распоряжении исследователя. Отсюда понятно, что много побочных переменных проконтролировать невозможно. Вторая: оставшиеся испытуемые не совсем типичны для сравниваемых групп, и полученные отношения оказываются непредставительными. Таким образом, имеется опасность как слишком слабого, так и слишком строгого контроля.

Другим методом является выделение однородных подгрупп, как это делалось в исследовании порядка рождения. Например, была выделена подгруппа испытуемых с семьей из пяти человек, и уже внутри этой подгруппы выяснялось влияние порядка рождения. Число различных подгрупп совпадает с числом уровней побочной переменной. Таким образом, оказывается возможным обнаружить, как и в случае с двумя независимыми переменными в активном эксперименте, существует ли значимое взаимодействие между рассматриваемыми переменными. В корреляционных исследованиях это может также помочь пониманию отношений между переменными. Например, небольшое понижение линии тестовых оценок интеллекта для работников сельского хозяйства может быть следствием большего желания завести еще одного ребенка. Если бы имелись данные о таком желании, может удалиться показать, что оно лежит в основе влияния порядка рождения. Это следовало бы, во-первых, из понижения линии тестовых оценок интеллекта в зависимости от порядка рождения для однородных подгрупп как городских, так и сельских жителей, не желающих следующего ребенка, и, во-вторых, из отсутствия такого понижения для подгрупп, желающих ребенка.

Однако поскольку доказательство оставалось бы корреляционным, нельзя было настаивать на том, что желание иметь еще одного ребенка — основа рассматриваемого эффекта. Например, желание могло просто коррелировать с физической энергией родителей, и эта последняя могла быть истинной причиной основного эффекта. Таковы корреляционные исследования. Здесь могут быть проверены интересные гипотезы, предложено много разных идей. И все-таки контроль здесь никогда не будет таким же хорошим, как в активном эксперименте.

Последнее исследование приводилось как искусственный пример возможности использования теста на способности для отбора контролеров. Группа испытуемых проводилась через тест на специальные способности. Затем оценивалось качество их работы. Отношение между этими двумя переменными — тестовыми оценками и оценками работы — было представлено на диаграмме разброса. Форма этой диаграммы может быть отражена в коэффициенте корреляции. Положительные значения коэффициента находятся в пределах от 0 до 1. С помощью диаграмм разброса, представляющих разные значения коэффициента корреляции, было показано, как можно отобрать претендентов согласно заданному правилу решения, например, чтобы, по крайней мере, 80 % из всех выбранных оказались удовлетворительно работающими. Чем выше коэффициент корреляции, тем ниже может быть расположена так называемая секущая оценка для теста на способности, что позволяет увеличить число отобранных индивидов при данном правиле решения. Корреляция позволяет также предсказать с помощью линии предсказания, связывающей средние рабочие оценки с тестовыми оценками, какой рабочей оценки может достичь индивид. Чем выше коэффициент корреляции, тем круче эта линия и тем меньше ошибка предсказания.

Были рассмотрены общие черты и различия разных корреляционных исследований. Все эти исследования сходны в том, что переменные в них уже существуют, в отличие от активных экспериментов, где условия независимой переменной активно организуются для выявления влияния последней на зависимую переменную. Различаются же они между собой по многим параметрам. Во-первых, одна из переменных в них в разной степени приближается к

независимой переменной. Например, в исследовании психологической приспособленности факторы-предпосылки вполне могут быть расценены как независимые переменные. Другая крайность представлена исследованием по отбору контролеров, где ни одна переменная не похожа на независимую. Поэтому однонаправленность предсказания определяется только практическими целями. Во-вторых, корреляционные исследования различаются тем, вычисляется ли коэффициент корреляции или нет. Представление степени связи с помощью коэффициента корреляции имеет наибольший смысл в том случае, если значения каждой переменной образуют непрерывное колоколообразное распределение. Этому условию почти всегда удовлетворяют два множества тестовых оценок. Кроме того, такие переменные, как доход или балльные оценки свойств личности, также зачастую достаточно хорошо соответствуют этим условиям. В-третьих, корреляционные исследования различаются по цели: имеет ли оно чисто познавательную цель или предполагает немедленное практическое приложение. Тот факт, что в практическом исследовании, приведенном в качестве примера в этой главе, использовался коэффициент корреляции и соотносились две оценки поведения для каждого испытуемого, вовсе не означает, что исследование такого типа не может служить целям расширения нашего познания.

Контрольные вопросы и задания

1. Почему исследование, в котором сравниваются хорошо и плохо приспособленные группы, называется корреляционным?
2. Почему сопутствующее смещение всегда присутствует в корреляционном исследовании и только иногда — в активном эксперименте?
3. Как можно использовать идеальный эксперимент в качестве эталона внутренней валидности в корреляционном исследовании?
4. С какими трудностями связан контроль смещения путем индивидуального подбора пар? Приведите пример.
5. Дайте пример того, как использование однородных подгрупп может обеспечить сведения о взаимодействии.
6. Какие смешивающиеся переменные были упущены исследователями при определении влияния порядка рождения?
7. Почему даже самое лучшее корреляционное исследование ограничено в выделении переменных, влияющих на поведение?
8. Почему высокий коэффициент корреляции позволяет отобрать пропорционально большее число индивидов, работающих качественно?
9. По каким параметрам различаются корреляционные исследования?

СТАТИСТИЧЕСКОЕ ПРИЛОЖЕНИЕ: КОЭФФИЦИЕНТ КОРРЕЛЯЦИИ

Стандартные оценки

Самая простая формула для вычисления коэффициента корреляции между двумя выборками оценок задается с помощью *стандартных оценок*. Эта формула дает также наиболее ясное представ-

ление о значении коэффициента корреляции. Вот почему в этом приложении вводится понятие стандартной оценки. Кроме того, стандартные оценки, полученные в различных тестах, можно сравнить между собой. Так, если вы скажете кому-либо, что по истории вы получили тестовую оценку 38, а по английскому языку — 221, он мало что поймет. Однако этот «кто-то», если он читал данное приложение, получит точную информацию из сообщения, что ваша стандартная оценка по истории равна +2,1, а по английскому языку —1,3.

Вы уже знаете, что (первичная) тестовая оценка какого-либо испытуемого в группе обозначается через X . Тестовая же оценка данного конкретного испытуемого обозначается с помощью индекса. Так, например, тестовая оценка испытуемого 3 записывается как X_3 . Вы также знакомы с отклонением оценки от среднего $x = X - M_X$. Отклонение оценки испытуемого 3 записывается как $x_3 = X_3 - M_X$. Если отклонение оценки испытуемого разделить на стандартное отклонение σ_X распределения оценок, то оно преобразуется в стандартную оценку (или z -оценку).

Допустим, что испытуемый 3 имеет (первичную) тестовую оценку 60. Средняя оценка для группы равна 49 и стандартное отклонение оценок равно 12, т. е. $X_3 = 60$, $M_X = 49$, $\sigma_X = 12$. Прежде всего $x_3 = 60 - 49 = +11$. Давайте теперь вычислим z_{X_3} , т. е. найдем стандартную оценку для испытуемого 3:

$$z_X = \frac{x}{\sigma_x}. \quad (9.1)$$

Следовательно,

$$z_{X_3} = +\frac{11}{12} = +0,92.$$

Поскольку стандартные оценки редко имеют величину больше +2 и меньше -2, то вы узнаете, что оценка именно этого испытуемого лежит примерно посередине между средней и наивысшей оценкой в группе.

Рабочие оценки, такие, например, как оценки качества работы контролеров, которые необходимо скоррелировать с тестовыми оценками, обычно обозначаются символом Y вместо X . Тогда отклонение оценки обозначается через y , а стандартная рабочая оценка — z_Y . Итак, мы говорим о нахождении корреляции между X и Y тогда, когда каждый испытуемый в группе имеет оценку X и оценку Y . Коэффициент корреляции обозначается символом r_{XY} .

Вычисление r_{XY}

Для вычисления коэффициента мы снова воспользуемся ранее приводившимися данными. Возьмем данные для условия А как

Тестовые оценки помещены в приводимой ниже таблице во втором столбце слева, а рабочие оценки — во втором столбце справа. Они обозначены как X и Y соответственно.

S	X	x	z_x	$z_x z_y$	z_y	y	Y	S
1	223	+38	+2,054	+2,455	+1,195	+190	1810	1
2	184	-1	-,054	-,109	+2,013	+320	1940	2
3	209	+24	+1,297	+,898	+,692	+110	1730	3
4	183	-2	+,108	-,061	-,566	-90	1530	4
5	180	-5	-,270	-,102	+,377	+60	1680	5
6	168	-17	-,919	-,810	+,881	+140	1760	6
7	215	+30	+1,622	+,102	+,063	+10	1630	7
8	172	-13	-,703	+,442	-,629	-100	1520	8
9	200	+15	+,811	-,357	-,440	-70	1550	9
10	191	+6	+,324	-,143	-,440	-70	1550	10
11	197	+12	+,649	+,653	+1,006	+160	1780	11
12	188	+3	+,162	-,020	-,126	-20	1600	12
13	174	-11	-,595	-,075	+,126	+20	1640	13
14	176	-9	-,486	-,214	+,440	+70	1690	14
15	155	-30	-1,622	+,714	-,440	-70	1550	15
16	165	-20	-1,081	+2,720	-2,516	-400	1220	16
17	163	-22	-1,189	+1,346	-1,132	-180	1440	17
M	185						1620	
σ	18,5						159	

$$\sum z_x z_y = +7,336;$$

$$r_{xy} = +0,432.$$

тестовые оценки 17 испытуемых, а данные для условия Б как рабочие оценки для тех же испытуемых. Однако чтобы подчеркнуть относительный характер стандартных оценок, умножим каждое значение для условия Б на 10. К счастью, мы уже сделали много вычислений, необходимых для получения r_{xy} . Для тестовых оценок мы просто используем полученные ранее среднее и стандартные отклонения. Для условия Б полученные среднее и стандартные отклонения нужно просто умножить на 10.

Вы видите, что тестовая оценка (X) первого испытуемого S_1 была 223, а его рабочая оценка -1810 . Сдвинувшись по этой строке от обоих концов к середине, мы обнаружим, что x равно $+38$ (т. е. $223 - 185$) и y равно $+190$ (т. е. $1810 - 1620$). Далее видим, что z_x равно $+2,054$ (т. е. $+38$, деленное на $18,5$), а z_y равно $+1,195$ (т. е. 190 , деленное на 159). И наконец, в среднем столбце мы находим произведение z_x на z_y , которое равно $+2,455$.

Такие же вычисления, сделанные для остальных 16 испытуемых, заполняют всю остальную таблицу. Ниже этих данных приведены величины средних и стандартных отклонений. Еще ниже в центре дается сумма по столбцу $z_x z_y$, равная $+7,336$. Это число, деленное на число испытуемых -17 , и дает величину коэффициента корреляции, равную $+0,432$.

В случае, если вам не хочется запоминать все эти термины, вы можете обратиться к следующей формуле для расчета коэффициента корреляции:

$$r_{xy} = \frac{\sum z_x z_y}{N} \quad (9.2)$$

или для наших данных

$$r_{xy} = \frac{+7,336}{17} = +0,432.$$

Диаграмма разброса (корреляционное поле)

На рис. 9.4 показана диаграмма разброса, каждая точка которой представляет одного испытуемого. Значения шкал даны в единицах стандартных оценок z .

При таких осях наклон линии предсказания прямо показывает величину r_{xy} . В нашем случае r_{xy} равно $+0,432$. Это значение наклона линии: на каждое смещение на единицу вправо точки линии поднимаются вверх на $0,432$ единицы. Так, если данный испытуемый имеет значение z_x , равное $+1$, то предсказываемое значение z_y для него равно $+0,432$. Таким образом, предсказываемая величина значительно ближе к среднему распределения, чем та величина, на основе которой делалось предсказание. Поэтому говорят, что предсказания стремятся (*регрессируют*) к среднему, и линия предсказания называется *линией регрессии* X на Y . Более точно, это предсказание z_y по z_x .

Вы можете заметить, что линия предсказания проходит через пересечение точек $z_x = 0$ и $z_y = 0$. Обе эти точки представляют средние значения соответствующих распределений. Это справедливо, независимо от значения величины r_{xy} . Если испытуемый

оказывается в точке среднего по X , то наилучшим предсказанием всегда будет среднее по Y . Далее видно, что если оценка будет выше среднего по X (положительное значение z_X), то предсказываемая оценка будет также выше среднего по Y (положительное значение z_Y). Точно так же для X ниже среднего значения предсказываемая оценка Y будет ниже среднего значения по Y .

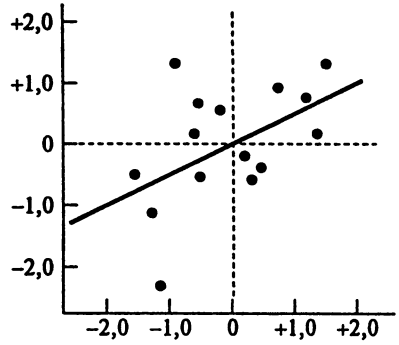


Рис. 9.4. Корреляционное поле. Масштабы осей разные и представлены в единицах стандартных оценок

И наконец, чем выше величина r_{XY} , тем меньше регрессия предсказания. В случае полной корреляции линия предсказания будет иметь наклон $+1$. Так, если, например, z_X равно $+1,5$, то предсказываемое z_Y тоже будет равно $+1,5$, а если z_X равно $-0,8$, то z_Y тоже будет равно $-0,8$. При полной корреляции регрессия к среднему отсутствует. С другой стороны, если корреляция равна 0 , то линия будет иметь нулевой наклон, т.е. она будет представлять собой горизонтальную линию. Она будет проходить на уровне $z_Y = 0$, т.е. среднего значения по Y . Поэтому, какая бы ни была величина z_X , наилучшее предсказание всегда будет $z_Y = 0$. Следовательно, при нулевой корреляции все предсказываемые значения регрессируют к среднему.

Все это может быть представлено посредством следующей формулы:

$$\hat{z}_Y = r_{XY} z_X. \tag{9.3}$$

Эта формула показывает, что стандартную оценку для выборки Y можно получить, умножив стандартную оценку для выборки X на коэффициент корреляции между X и Y . Например, для испытуемого, имеющего стандартную оценку z_X , равную $+0,50$ с коэффициентом корреляции $0,70$, получим

$$\hat{z}_Y = 0,70 (+ 0,50) = +0,35.$$

Задача: Вычислите r_{XY} для данных в задаче, приведенной в статистическом приложении к главе 6. Используйте условие В для X и условие Г для Y .

Ответ: $r_{XY} = 0,576$.

Библиография

Allus W.D. Birth order and scholastic aptitude // *Journal of Consulting Psychology*. — 1965. — 29 (3). — P. 202—205.

American Heritage Dictionary of the English Language / W. Morris (ed.). Boston, 1973.

American Psychological Association. Publication Manual. (2nd. ed.) Washington, 1976.

Bakan D. On Method: Toward a Reconstruction of Psychological Investigation. San Francisco, 1967.

Békésy G. V. Olfactory analogue to directional hearing // *Journal of Applied Physiology*. — 1964. — 19. — P. 369—373.

Békésy G. V. Sensory Inhibition. — Princeton, N. Y., 1967.

Belmont L., Marolla F.A. Birth order, family size and intelligence // *Science*. — 1973. — 182. — P. 1096—1101.

Block J. The Q-sort method in personality assessment and psychiatric research. — Springfield, I 11, 1961.

Breland K., Breland M. The misbehavior of organisms // *American Psychologist*. — 1961. — 16. — P. 681—684.

Calfee R. C., Anderson R. Presentation rate effects in paired-associate learning // *Journal of Experimental Psychology* — 1971. — 88. — P. 239—245.

Campbell D. T., Stanley J. C. Experimental and quasi-experimental designs for research in teaching Gage N. L. (ed.). *Handbook of Research on Teaching*. — Chicago, 1962.

Carder B., Berkowitz K. Rats' preference for learned in comparison with free food // *Science*. — 1970. — 167. — P. 1273—1275.

Cohan S. E. Packaging law is on the book, but ills it aimed to cure are still troublesome // *Advertising Age*. — 1969. — 40. — P. 10.

Cohen J. *Statistical Power Analysis for the Behavioral Sciences* (rev. ed.). — N. Y., 1977.

Cohen L. J. The operational Definition of human attachment // *Psychological Bulletin*. — 1974. — 81. — P. 207—217.

Dennenberg V. H., Morton J. R. C. Effects of preweaning and postweaning manipulation upon problem-solving behavior // *Journal of Comparative and Physiological Psychology*. — 1962. — 55. — P. 1096—1098.

Dukes W. F. «N=1». *Psychological Bulletin*. — 1965. — 64. — P. 74—79.

Ebbinghaus H. *Memory*. (H. A. Ruger and C. E. Bussenius, Trans). — N. Y., 1964. (Originally published 1885).

Edwards W., Lindman H., Savage L. J. Bayesian statistical inference for psychological research // *Psychological Review*. — 1963. — 70. — P. 193—242.

- Fleener D. E., Cairns R. B.* Attachment behaviors in human infants: Discriminative vocalization on maternal separation // *Development Psychology*. — 1970. — 2. — P. 215—223.
- Gaffan D.* Recognition impaired and association intact in the memory of monkeys after transection of the fornix // *Journal of Comparative and Physiological Psychology*. — 1974. — 86. — P. 1100—1109.
- Galton F.* *English Men of Science: Their Nature and Nurture.* (2nd ed.) — Forest Grove, Ore., 1970. (Originally published 1874).
- Gateway R. D., Perloff R.* An experimental investigation of three methods of providing weight and price information to customers // *Journal of Applied Psychology*. — 1973. — 57. — P. 81—85.
- Golding W. G.* *Lord of the Flies.* — N. Y., 1954.
- Gottman J. M.* N-of-1 and N-of-2 research in psychotherapy // *Psychological Bulletin*. — 1973. — 80. — P. 93—105.
- Gottsdanker R., Way T. C.* Varied and constant intersignal intervals in Psychological refractoriness // *Journal of Experimental Psychology*. — 1966. — 72. — P. 792—804.
- Grocers moan, but New York moves on unit prices // *Advertising Age*. — 1969. — 40. — P. 3.
- Guilford J. P., Fruchter B.* *Fundamental Statistics in Psychology and Education.* (5th ed.) — N. Y., 1973.
- Hardin G.* *Exploring New Ethics for Survival: The Voyage of the Spaceship Beagle.* — N. Y., 1972.
- Harper R. S., Stevens S. S.* A psychological scale of weight and a formula for its derivation // *American Journal of Psychology*. — 1948. — 61. — P. 342—351.
- Hays W. K.* *Statistics for the Social Sciences.* (2nd ed.) — N. Y., Rinehart and Winston, 1973.
- Hick W. E.* On the rate of gain of information // *Quarterly Journal of Experimental Psychology*. — 1952. — 4. — P. 11—26.
- Holtzman W.* Statistical models for the study of change in a single case // C. Harris (ed.). *Problems in Measuring Change.* — Madison, 1963.
- Hyman R.* Stimulus information as a determinant of reaction time // *Journal of Experimental Psychology*. — 1953. — 45. — P. 188—196.
- Jacobs G. H., Anderson D. H.* Color vision and visual sensitivity in the California ground squirrel (*Citellus beecheyi*) // *Vision Research*. — 1972. — 12. — P. 1995—2004.
- Jensen G. D.* Preference for bar pressing over «freeloading» as a function of number of rewarded presses // *Journal of Experimental Psychology*. — 1963. — 65. — P. 451—454.
- Kennedy J. E., Landesman J.* Series effects in motor performance studies // *Journal of Applied Psychology*. — 1963. — 47. — P. 202—205.
- Kennedy J. L.* A methodological review of extra-sensory perception // *Psychological Bulletin*. — 1939. — 36. — P. 59—103.
- Keppel G.* *Design and Analysis, A Researcher's Handbook* // Englewood Cliffs, N. Y., 1973.
- Klugh H. E.* *Statistics, the Essentials for Research.* — N. Y., 1970.
- Koffler K., Coulson G.* Feline indolence: Cats prefer free to response-produced food // *Psychonomic Science*. — 1971. — 24. — P. 41—42.

Kraft C. L., Elworth C. L. Night visual approaches // Boeing Airliner. — 1969 (March-April). 2—4.

La Voie J. Punishment and adolescent self-control // *Developmental Psychology*. — 1973. — 8. — 16—24.

Lewin K., Lippitt R., White R. K. Patterns of aggressive behavior in experimentally created «social climates» // *Journal of Social Psychology*. — 1939. — 10. — P. 271—299.

Lockard R. B. The albino rat: A defensible choice or a bad habit? // *American Psychologist*. — 1968. — 23. — P. 734—742.

Neuringer A. J. Animals respond for food in the presence of free food // *Science*. — 1969. — 166. — P. 399—401.

Orwell G. *Nineteen Eighty-Four: A Novel*. — N. Y., 1949.

Poulton E. C. Unwanted range effects from using within-subject experimental designs // *Psychological Bulletin*. — 1973. — 80. — P. 113—121.

Poulton E. C., Freeman P. R. Unwanted asymmetrical transfer effects with balanced experimental designs // *Psychological Bulletin*. — 1966. — 66. — P. 1—8.

Prather D. C. Prompted mental practice as a flight simulator // *Journal of Applied Psychology*. — 1973. — 57. — P. 353—355.

Premack D., Schaeffer R. W., Hundt A. Reinforcement of drinking by running: effects of fixed ratio and reinforcement time // *Journal for the Experimental Analysis of Behavior*. — 1964. — 7. — P. 91—96.

Raven J. C. *Progressive Matrices*. — London, 1947.

Roethlisberger F. J., Dickson W. J. *Management and the Worker*. — Cambridge, Mass., 1946.

Rosenthal R. *Experimenter Effects in Behavioral Research*. — N. Y., 1966.

Runquist W. N. Verbal behavior // J. B. Sidowski (ed.). *Experimental Methods and Instrumentation in Psychology*. — N. Y., 1966.

Siegelman E., Block Jack, Block Jeanne, Von der Lippe A. Antecedents of optimal psychological adjustment // *Journal of Consulting and Clinical Psychology*. — 1970. — 35. — P. 283—289.

Simon J. R., Rudell A. P. Auditory S-R compatibility: the effect of an irrelevant cue on information processing // *Journal of Applied Psychology*. — 1967. — 51. — P. 300—304.

Singh D. Preference for bar pressing to obtain reward over freeloading in rats and children // *Journal of Comparative and Physiological Psychology*. — 1970. — 73. — P. 320—327.

Singh D., Query W. T. Preference for work over «freeloading» in children // *Psychonomic Science*. — 1971. — 24. — P. 77—79.

Sternberg S. The discovery of processing stages: Extension of Donders' method // *Acta Psychologica*. — 1969. — 30. — P. 276—315.

Taylor H. C., Russell J. T. The relationship of validity coefficients to the practical effectiveness of tests in selection // *Journal of Applied Psychology*. — 1939. — 23. — P. 565—578.

Textile industry // *Encyclopaedia Britannica*. — *Macropaedia*. — 1974. — 18. — P. 170—189.

Thorndike E. L. *Human Learning*. — N. Y., 1931.

Thorndike E. L. *The Fundamentals of Learning*. — N. Y., 1932.

Torgerson W. S. *Theory and Methods of Scaling*. — N. Y., 1958.

Wagenaar W.A. Note on the construction of diagram-balanced Latin squares // *Psychological Bulletin*. — 1969. — 72. — P. 384—386.

Wallis W.A., Roberts H.V. *Statistics, a New Approach*. — N.Y., 1962.

Weston H.C., Adams S. The effects of noise on the performance of weavers. Report № 65. Part II. Industrial Health Research Board. — London, 1932.

White R.W. Motivation reconsidered: The concept of competence // *Psychological Review*. — 1959. — 66. — P. 297—333.

Wise R.A., Dawson V. Diazepam-induced eating and lever pressing for food in sated rats // *Journal of Comparative and Physiological Psychology*. — 1974. — 86. — P. 930—941.

Yerkes R.M. *The Dancing Mouse*. — N.Y., 1907.

Yerkes R.M., Dodson J.D. The relation of strength of stimulus to rapidity of habit formation // *Journal of Comparative Neurology and Psychology*. — 1908. — 18. — P. 459—482.

Словарь экспериментатора¹

Настоящая книга — хорошее учебное пособие по экспериментальной психологии. Чтобы успешно использовать ее в учебном процессе (и прежде всего самому начинающему психологу), удобно иметь общую сводку определений терминов, применяемых при описании планирования и проведения психологического эксперимента. Автор назвал их «Словарем экспериментатора» — базовым условием полноценного обсуждения и усвоения материала. Перечислить основные термины (а здесь их около ста) и дать им краткую характеристику — главная цель словаря. Описание содержания используемых понятий позволяет также уточнить особенности перевода терминов. В некоторых случаях полезно знать их эквиваленты на языке оригинала. А если перевод термина не буквален, в словаре можно найти соответствующее разъяснение.

Как обычно, термины представлены отдельно, в алфавитном порядке. Однако большинство из них довольно тесно связаны между собой. В первую очередь это относится к таким фундаментальным понятиям книги, как «безупречный эксперимент», «валидность», «обобщение», «контроль», «экспериментальная схема» и др. Эти связи учтены и отражены в словаре: содержательно близкие термины сгруппированы в единые блоки (иногда довольно крупные), в тексте каждого блока другие термины, вошедшие в словарь, выделены полужирным шрифтом. Кроме того, неслучайно словарь открывается термином «валидность».

При работе со словарем хорошо пользоваться следующим за ним «Предметным указателем», где приведены все термины, выделенные курсивом в тексте книги, а также названия экспериментов-иллюстраций.

ВАЛИДНОСТЬ (*validity*) — достоверность (или степень достоверности) вывода, которую обеспечивают результаты реального эксперимента по сравнению с результатами **безупречного эксперимента**. «В.» — центральное понятие словаря экспериментатора: оно объединяет основные цели исследования с идеальными эталонами их достижения и реальными процедурами проведения эксперимента.

Глобальной целью любого экспериментального исследования является обобщение полученных результатов и вывода об экспериментальной гипотезе. Однако полное достижение этой цели возможно лишь в мысленном, безупречном эксперименте, невыполнимом на практике. Реальный эксперимент в той или иной мере репрезентирует (представляет) безупречный, и чем лучше эта репрезентативность, тем выше В. экспе-

¹ Сост. В. В. Петухов.

римента. Таким образом, повышение В., т.е. планирование проведения эксперимента в соответствии с его безупречным образцом, является конкретной задачей исследователя, успешность решения которой зависит, во-первых, от характера реальных условий и, во-вторых, от адекватности выбора средств. Так, источники нарушения В. (прежде всего ненадежность и смешение) отдалают реальный эксперимент от безупречного, а способы их контроля позволяют приблизиться к нему, т.е. обеспечить высокую В. для обобщения экспериментальных результатов. В полученных данных можно оценить статистически, например В. теста (в корреляционном исследовании) определяется степенью корреляции результатов его выполнения с изучаемым видом деятельности испытуемых. В соответствии с разными типами безупречного эксперимента различают внутреннюю и внешнюю В.

— **внутренняя В.** — достоверность выводов, которую обеспечивают результаты реального эксперимента по сравнению с результатами идеального и бесконечного экспериментов. Повышение внутренней В. связано с устранением результатов действия побочных переменных и с усреднением их изменчивости и неустойчивости. Внутренняя В. — первое и необходимое требование к экспериментальным выводам: эксперимент, не обладающий внутренней В., назван автором просто *неудачным*;

— **внешняя В.** — достоверность выводов, которую обеспечивают результаты реального эксперимента по сравнению с результатами эксперимента полного соответствия. Повышение внешней В. обеспечивается достижением соответствия уровней дополнительных переменных в эксперименте уровням этих переменных в изучаемой реальности. Эксперимент, не обладающий внешней В., является *неверным*, несоответствующим поставленной гипотезе (но, быть может, пригодным для проверки другой гипотезы).

Следует отметить, что любую *переменную (фактор)*, отличную от независимой, также оказывающую влияние на зависимую переменную, автор называет просто «*other*» — *другая*. При переводе для обозначения таких переменных используются два разных термина «*побочная*» — применительно к переменной, нарушающей внутреннюю В., и «*дополнительная*» — переменная, уровень которой входит в экспериментальную гипотезу и должен быть адекватно представлен с целью достижения внешней В. Иногда одни и те же факторы могут выступать в эксперименте как побочные (нарушающие внутреннюю В.) и как дополнительные (влияющие на внешнюю В.). Устранение и унификация побочных факторов, повышающие внутреннюю В., могут приводить к несоответствию уровней дополнительных переменных, что снижает внешнюю В. (см. *Соответствие*).

— **операциональная В.** — разновидность внешней В. для лабораторного эксперимента, соответствие (или степень соответствия) применяемых методических процедур тем теоретическим понятиям, которые входят в экспериментальную гипотезу.

ВЗАИМОДЕЙСТВИЕ — количественный результат, обусловленный соотношением между действием двух или нескольких независимых переменных, выделенный в факторном эксперименте. Вычисляется как разность между различиями значений зависимой переменной, полученных

при действии разных условий первой, второй и т.д. переменных, и иллюстрируется графически. По графическому изображению В. выделяют три его вида: *нулевое (zero)*, или отсутствие В., *расходящееся (spreading)* и *пересекающееся (crossing)* В.;

— В. 1-го порядка — В. между двумя независимыми переменными (факторами), В. 2-го порядка — В. трех независимых переменных и т.д.

ВОСПРОИЗВОДИМОСТЬ (*replicating*) эксперимента — возможность повторить эксперимент, руководствуясь описанием применяемых методических процедур в *экспериментальном отчете* (раздел «Методика»).

ВЫБОРКА (*sample*) — испытуемые, отобранные из изучаемой популяции для участия в эксперименте. В. разделяют на экспериментальную и контрольную группы. Эксперименты с привлечением В. являются, как правило, межгрупповыми, но в принципе могут проводиться и по индивидуальным схемам. Основное качество В., необходимое для полноценного эксперимента, — ее **репрезентативность** (см.), которая определяет **внешнюю валидность** эксперимента и достигается адекватными методами отбора испытуемых (см.). **Внутренняя валидность** зависит от степени взаимного соответствия экспериментальной и контрольной групп и обеспечивается с помощью различных стратегий распределения испытуемых по группам (см.).

ВЫВОД (*inference*) **СТАТИСТИЧЕСКИЙ** — нахождение такой величины различия между значениями зависимой переменной, которая в случае подтверждения нуль-гипотезы может быть превышена только с вероятностью, задаваемой *правилом статистического решения (decision)* (например, $p = 0,05$). В.С. определяет значимость полученных различий, служит условием для экспериментальных выводов (*conclusions*), зависит от числа испытуемых и величины стандартного отклонения.

ВЫДЕЛЕНИЕ (*isolation*) независимой переменной — отделение независимой переменной, входящей в экспериментальную гипотезу, от сопутствующих ее воздействию побочных влияний, достигаемое с помощью специальных методических процедур в лабораторном эксперименте. В отличие от практических экспериментов, где независимая переменная фактически всегда представляет собой группу факторов, в лабораторных (или собственно научных) ее стремятся выделить, т.е. сделать **единичной**, чтобы иметь возможность исследовать только ее воздействие на зависимую переменную. В. независимой переменной вместе с очищением ее условий позволяет ставить и проверять **точные экспериментальные гипотезы**, касающиеся механизмов изучаемого поведения.

ГИПОТЕЗА ЭКСПЕРИМЕНТАЛЬНАЯ — конкретизация некоторой догадки или идеи с целью ее проверки в эксперименте. Г.Э. включает в себя: **независимую переменную**, **зависимую переменную**, **отношение** между ними и **уровни дополнительных переменных**. Таким образом, Г.Э. — это предположение об отношении между независимой и зависимой переменными при определенном уровне дополнительной переменной. В той мере, в какой это отношение не будет зависеть от побочных и дополнительных переменных, Г.Э. может становиться все более обобщенной;

— **контргипотеза** — Г.Э., альтернативная к основному предположению; возникает автоматически;

— *третья конкурирующая Г.Э.* — Г.Э. об отсутствии влияния независимой переменной на зависимую переменную; проверяется только в лабораторном эксперименте (см. Нуль-гипотеза);

— *точная Г.Э.* — предположение об отношении между единичной независимой переменной и зависимой переменной в лабораторном эксперименте; проверка требует выделения независимой переменной и очищения ее условий;

— *Г.Э. о максимальной (или минимальной) величине* — предположение о том, при каком уровне независимой переменной зависимая переменная принимает максимальное (или минимальное) значение. Основанный на представлении о двух базисных процессах, оказывающих противоположное действие на зависимую переменную: при достижении определенного (высокого) уровня независимой переменной «негативный» процесс становится сильнее «позитивного»; проверяется только в многоуровневом эксперименте;

— *Г.Э. об абсолютных и пропорциональных отношениях* — точное предположение о характере постепенного (количественного) изменения зависимой переменной с постепенным (количественным) изменением независимой переменной (см. Отношение); проверяется в многоуровневом эксперименте;

— *Г.Э. с одним отношением* — предположение об отношении между одной независимой и одной зависимой переменными. Для проверки Г.Э. с одним отношением может быть использован и факторный эксперимент, но вторая независимая переменная является при этом контрольной;

— *комбинированная Г.Э.* — предположение об отношении между определенным сочетанием (комбинацией) двух (или нескольких) независимых переменных, с одной стороны, и зависимой переменной — с другой (см. также Взаимодействие); проверяется только в факторном эксперименте.

ДАННЫЕ — первоначальные, еще не обработанные результаты эксперимента, зафиксированные в протоколе (например, время выполнения задачи, субъективный отчет испытуемого и т. п.). Наиболее важная часть Д. служит материалом для определения значений зависимой переменной. (См. Протоколирование, Показатели измеряемые, Способ представления результатов.)

ЗНАЧЕНИЕ (*meaning*) — единица измерения зависимой переменной, конечный результат обработки полученных данных. Каждому из условий, или уровней, независимой переменной соответствует определенное З. Совокупность З. представлена обычно в таблицах или на графиках.

ЗНАЧИМОСТЬ (*significance*) результатов — статистическая достоверность полученных различий между средними значениями зависимой переменной (при воздействии разных условий независимой переменной), отвечающих экспериментальной гипотезе. Для повышения внутренней валидности эксперимента проводится проверка З. — проверка отвержения (или неотвержения) нуль-гипотезы. З. результатов — необходимое (иногда — недостаточное) условие достоверности вывода об исследуемой гипотезе.

ИЗМЕНЧИВОСТЬ (*variability*) **НЕСИСТЕМАТИЧЕСКАЯ** — одно из свойств любой переменной, проявляющееся в непостоянстве ее измене-

ний и приводящее к широкому разбросу экспериментальных данных (ср. «*variability*» и «*variable*» — переменная). Основной источник ненадежности эксперимента, нарушения внутренней валидности. В практических (естественных) индивидуальных экспериментах — это И. Н. побочных переменных с течением времени (см. Фактор времени), для усреднения которой применяются способы *первичного* контроля (в частности, увеличение числа проб). В условиях искусственных экспериментов с относительной унификацией побочных факторов И. Н. сокращается, а благодаря возможности предъявить все необходимое число проб за сравнительно короткое время уменьшается и разброс данных. Внутреннюю валидность группового эксперимента может нарушить И.Н. (здесь — изменения, или вариации) индивидуальных различий испытуемых; для контроля следует либо увеличивать количество испытуемых, либо добиваться взаимного соответствия экспериментальной и контрольной групп. Возможность И.Н. необходимо учитывать даже в лабораторных экспериментах, ведь полная стабилизация всех побочных факторов (т. е. проведение идеального эксперимента) неосуществима на практике.

ИСПЫТУЕМЫЙ (*subject*) — субъект, привлеченный к участию в эксперименте для испытания воздействия независимой переменной. Эксперименты могут быть индивидуальными (здесь — с одним И.) и групповыми. *Потенциальными* И. являются специально отобранные или *имеющиеся в наличии* (*available*) представители изучаемой популяции. **Индивидуальные различия** И. могут нарушать внутреннюю валидность эксперимента.

ИССЛЕДОВАНИЕ (*study*) **КОРРЕЛЯЦИОННОЕ** — исследование соотношения двух (или нескольких) переменных, характеризующих различные индивидуальные различия людей и их поведения. В отличие от эксперимента, т. е. *активного* управления состояниями независимой переменной, в И.К. проверяются гипотезы о взаимосвязи уже существующих характеристик. И.К. необходимо также для определения диагностической ценности *тестов*, измеряющих эти характеристики. И.К. проводятся, как правило, на большом количестве испытуемых, с применением *статистических* способов контроля (для выделения возможных детерминант изучаемых различий используется, в частности, факторный анализ). И.К. разделяют по следующим параметрам: 1) возможность интерпретации одной из переменных в качестве независимой: либо эта переменная предшествует другой во времени (что позволяет высказывать эмпирические предположения об их причинно-следственной связи), либо нет; 2) возможность вычисления коэффициента корреляции (см. **Корреляция**) и 3) цель — либо собственно познавательная, теоретическая, либо чисто прикладная.

КВАДРАТ ЛАТИНСКИЙ — наиболее распространенная схема кросс-индивидуального (многоуровневого) эксперимента, основанная на позиционном уравнивании предъявления уровней независимой переменной (см. **Уравнивание позиционное**). «К.Л.» — это набор таких последовательностей уровней независимой переменной, в которых каждый из уровней встречается в каждой позиции только один раз. Схема К.Л. — случайный выбор и использование экспериментатором одного из вариантов этих наборов, когда каждая последовательность уровней предъявляется одно-

му испытуемому (или одной группе). Способ контроля эффектов однородного и, главное, неоднородного переноса при относительно небольшом числе испытуемых. Возможность возникновения эффектов ряда и центрации сохраняется;

— К. сбалансированный — редуцированный вариант латинского К., в последовательностях которого каждому уровню независимой переменной только один раз непосредственно предшествует каждый из остальных уровней. Позволяет контролировать эффекты неоднородного переноса с несложной статистической обработкой.

КОНТРОЛЬ, СПОСОБЫ КОНТРОЛЯ (*controls*) — способы усовершенствования эксперимента, приближающие его к безупречному. При обозначении С.К. учитывается, во-первых, та общая или частная цель, которая достигается с их применением, и, во-вторых, — общий или частный фактор, подлежащий К. (т. е. устранению или усреднению). Так, с одной стороны, любые С.К. являются в итоге К. обобщения полученных результатов и выводов, а также К. внутренней или внешней валидности. С другой стороны, они выступают как С.К. основных источников нарушения валидности — ненадежности и смещения, и порождающих их конкретных факторов (например, фактора времени, фактора задачи, индивидуальных различий испытуемых, эффектов последовательности проб, предубеждений экспериментатора и др.). Конкретные С.К. приводятся в словаре вместе с описанием этих контролируемых факторов. Для общей характеристики С.К. их можно организовать так.

Во-первых, С.К. могут быть непосредственно связаны с самим типом эксперимента. Некоторые источники нарушения валидности устраняются либо в результате «улучшения» реальности (например, К. систематического смещения и ненадежности в искусственном и лабораторном экспериментах), либо благодаря другим преимуществам, которые дает каждый новый тип эксперимента. Неслучайно поэтому некоторые виды экспериментов автор рассматривает как С.К.: многоуровневый эксперимент — К. неполноты независимой переменной и др., факторный — К. сопутствующего смещения базисной переменной при проверке гипотез с одним отношением. В корреляционных исследованиях оказывается возможным только *статистический К.*

Во-вторых, С.К. побочных и дополнительных влияний, неизбежно возникающих при проведении эксперимента, является любая экспериментальная схема. В соответствии с центральным параметром классификации экспериментальных схем — основанием сравнения условий (или уровней) независимой переменной, выделяют интраиндивидуальный, межгрупповой и кроссиндивидуальный К. О возможностях применения конкретных экспериментальных схем для К. внутренней валидности см. Последовательность, Стратегия построения групп, Распределение, Отбор, Уравнивание позиционное и др. Нужно добавить, что для К. сопутствующего смещения в лабораторных экспериментах требуются специальные схемы, не входящие в основную классификацию. С.К. внешней валидности, связанные с решением проблем соответствия привлекаемых переменных, приведены в табл. 1.

И наконец, существуют универсальные меры по К. побочных факторов, не зависящие от типа эксперимента и применяемых эксперимен-

тальных схем и названные поэтому *С. первичного К.* Они могут быть общими, например выбор адекватного количества проб, испытываемых, уровнем независимой переменной, и частными — подбор задач, предварительные меры по распределению испытываемых и т. п. Сюда же относятся и сами условия проведения любого эксперимента (см. *Планирование действий, Протоколирование*), а также строгое соблюдение экспериментальных процедур.

КОРРЕЛЯЦИЯ — реально наблюдаемый факт, пример взаимосвязи того или иного состояния независимой переменной с определенным значением зависимой переменной. Без наличия и использования плана эксперимента исследователь может получить только отдельные *К.*, не позволяющие проверить гипотезу. В тех случаях, когда управлять независимой переменной невозможно, привлекается *корреляционный подход*, т. е. установление статистического соотношения исследуемых переменных на большом числе испытываемых (см. *Исследование корреляционное*). Если распределение полученных значений каждой из этих переменных имеет куполообразную форму, то можно вычислить *коэффициент К.* между ними. Коэффициент *К.* используется также для проверки надежности экспериментальных результатов, валидности тестов.

ЛИНИЯ ПРЕДСКАЗАНИЯ — линия на *диаграмме разброса* данных, полученных в *корреляционном исследовании* двух переменных (например, при апробации нового теста: ось абсцисс — тестовые оценки, ось ординат — оценки диагностируемых различий). Л.П. соединяет средние оценки исследуемой переменной и тем самым позволяет в дальнейшем предсказывать ее значения по тестовым оценкам. Достоверность предсказания определяется величиной *коэффициента корреляции* переменных, которая и задает угол наклона Л.П. (от нуля — горизонтальная линия — до единицы — 45°).

МЕТОДИКА (*method*) — совокупность конкретных операций, процедур проведения эксперимента. Одним из условий достижения внутренней валидности является точность (*precision*) соблюдения *М.*

При проведении лабораторных экспериментов особо обсуждается вопрос о соответствии компонентов *М.* теоретическим понятиям, привлекаемым для определения независимой переменной (например, может ли резкий звук зуммера расцениваться как «наказание» и т. п.) (см. *Операциональная валидность*). *М.* подробно излагается в соответствующем разделе *экспериментального отчета*.

НЕНАДЕЖНОСТЬ (*unreliability*) — одно из основных нарушений внутренней валидности эксперимента. Источником *Н.* является главным образом несистематическая изменчивость побочных переменных, прежде всего фактора времени (в индивидуальном эксперименте) и индивидуальных различий испытываемых (в групповом), а также *нестабильность* всех привлекаемых переменных. *Н.* экспериментальных данных означает, что при каждом новом повторении эксперимента результаты оказываются иными. Аналогично оценивается надежность (или *Н.*) тестов в *корреляционных исследованиях*: при каждом новом испытании одни и те же испытываемые должны получать одни и те же оценки. Способами обеспечения *надежности* результатов является увеличение числа проб и испытываемых (эталон — бесконечный эксперимент), выбор адекватных измеряе-

мых показателей изучаемого поведения, искусственная стабилизация побочных факторов и др. Одна из форм Н. — *неполнота* независимой переменной, преодолеваемая с помощью введения большого числа *уровней* в **многоуровневом** эксперименте.

НУЛЬ-ГИПОТЕЗА — гипотеза об отсутствии различия между условиями независимой переменной (по их действию на зависимую переменную). Проверка Н.-г. возможна только в **лабораторном** эксперименте и **необходима** для установления статистической **значимости** экспериментальных результатов. Вероятности отвержения и неотвержения Н.-г. задаются **альфа-** и **бета-уровнями** (см. также **Вывод статистический**).

ОБОБЩЕНИЕ (*generalization*) — главная цель любого эксперимента: на основании ограниченного числа полученных **данных** сделать **выводы**, *распространяющиеся* на более широкую область практики. В **практических** индивидуальных экспериментах О. совершается так же, как и в жизни, — это общее заключение на основании ограниченного числа опытов, справедливое только для данного конкретного испытуемого. Возможность О. результатов связана с **репрезентативностью** эксперимента и его **валидностью** и зависит от контроля побочных факторов (**внутренняя валидность**) и от соответствия (и полноты представленности) **дополнительных** переменных в эксперименте. Выделяют О. экспериментальных результатов на изучаемую реальную ситуацию (**искусственный эксперимент**), на других представителях популяции (**групповой эксперимент**), распространение теоретических представлений (**лабораторный эксперимент**). Для контроля О. привлекают также новые уровни дополнительной переменной в **факторном** эксперименте. Наиболее широкое О. исследуемой гипотезы требует проведения целого ряда экспериментов при разных уровнях **дополнительных переменных**.

ОСНОВАНИЕ СРАВНЕНИЯ — первый и определяющий параметр классификации экспериментальных схем. Выделяются три глобальных типа О. С. между **условиями независимой переменной**, необходимого для проверки экспериментальной гипотезы: *интраиндивидуальное* (*within-subject*), *межгрупповое* (*between-groups*) и *кроссиндивидуальное* (*across-subject*). В первом случае сравнение данных, полученных при предъявлении исследуемых условий, проводится по каждому испытуемому. Во втором случае условия сравниваются по данным соответствующих (экспериментальной и контрольной) групп испытуемых. А в третьем — для получения средних оценок каждого из сравниваемых условий (здесь — **уровней**) необходимы данные всех участников эксперимента, т.е. подсчет проходит по всем испытуемым. Интраиндивидуальными, межгрупповыми и кроссиндивидуальными называют **экспериментальные схемы**, а также сами типы эксперимента и используемые в них способы контроля.

ОТБОР (*selection*) — способ привлечения испытуемых, позволяющий обеспечить представленность изучаемой популяции и тем самым — высокую внешнюю валидность межгруппового эксперимента. Для О. испытуемых используются две из основных стратегий построения групп, т.е. две эффективные схемы **межгруппового сравнения** — *случайный О. групп* и *случайный О. групп с предварительным выделением слоев* (или *случайный послыйный О.*). Последняя схема (при условии адекватности принципа выделения слоев изучаемой деятельности) позволяет достичь более вы-

сокой внешней валидности при одинаковом числе испытуемых. В указанных схемах *O*. сочетается с распределением испытуемых по группам, поэтому они обеспечивают не только внешнюю, но автоматически и высокую внутреннюю валидность (см. **Распределение**).

ОТНОШЕНИЕ между независимой и зависимой переменными — центральная составная часть экспериментальной гипотезы. С достоверностью полученного *O*. связана внутренняя валидность эксперимента. В многоуровневом эксперименте становится возможным проверять точные гипотезы об определенном типе количественных *O*. — абсолютных и пропорциональных, и выражать их в виде математической зависимости. Существуют три типа *O*. между независимой и зависимой переменными:

— **абсолютно-абсолютное *O*.**: равные абсолютные изменения независимой переменной связаны с равными абсолютными изменениями зависимой переменной (математически — линейная зависимость);

— **относительно-абсолютное *O*.**: равные относительные изменения независимой переменной связаны с равными абсолютными изменениями зависимой переменной (логарифмическая зависимость);

— **относительно-относительное *O*.**: равные относительные изменения независимой переменной связаны с равными относительными изменениями зависимой переменной (степенная зависимость).

Количество входящих в гипотезу *O*. определяет ее вид: различают гипотезы с одним *O*. и комбинированные.

ОЧИЩЕНИЕ (*purifying*) условий — отделение условий независимой переменной от сопутствующих их предъявлению, но не входящих в экспериментальную гипотезу, влияний, достигаемое с помощью специальных методических процедур в лабораторном эксперименте. *O*. условий необходимо для проверки точных экспериментальных гипотез (см. также **Выделение независимой переменной**).

ПЕРЕМЕННАЯ (*variable*) — основной термин словаря экспериментатора: любая реальность, которая может изменяться, и это изменение проявляется и фиксируется в эксперименте;

— **независимая *П*.** (или **фактор**) — *П*., изменяемая экспериментатором; включает в себя два или несколько состояний (условий) или уровней;

— **зависимая *П*.** — *П*., изменяющаяся при действии независимой *П*., принимая различные значения;

— **побочная *П*.** — *П*. (или фактор), порождающая ненадежность или систематическое смещение; совокупностями побочных *П*. являются, например, фактор времени (см.), фактор задачи (см.), индивидуальные различия испытуемых (или субъективный фактор) (см.);

— **дополнительная *П*.** — одна из составных частей экспериментальной гипотезы; для адекватной проверки частной экспериментальной гипотезы необходимо, чтобы уровень дополнительной *П*. соответствовал ее уровню в изучаемой реальности, а для проверки общей экспериментальной гипотезы — проведение экспериментов при разных уровнях дополнительной *П*.;

— **ключевая *П*.** — дополнительная *П*., наиболее важная для проверки экспериментальной гипотезы и ее обобщения;

- **единичная** (*unitary*) независимая П. — независимая П. в лабораторном эксперименте (см. Выделение независимой П.);
- **вторичная** П. — любая П., привлекаемая в лабораторном эксперименте для контроля действия *первичной* (независимой) П.;
- **расширенная** (*wider*) П. — результат введения нового условия независимой П. для контроля *сопутствующего смещения*;
- **качественная** (*qualitative, ungraded*) П. — независимая П., условия (состояния) которой отличаются друг от друга качественным образом;
- **количественная** (*quantitative, graded*) — независимая П., различия между уровнями которой можно количественно измерить; привлекается в многоуровневом эксперименте;
- **базисная** (*underlying, «лежащая в основе»*) П. — П., на которую (и только на нее), согласно теоретическим предположениям, оказывает свое действие независимая П.; точное определение результата этого действия требует проведения факторного эксперимента;
- **контрольная** П. — одна из дополнительных П., ставшая второй независимой П. в факторном эксперименте, проведенном для контроля *сопутствующего смещения базисной П. или обобщения результатов*.

ПЛАНИРОВАНИЕ ДЕЙСТВИЙ (*planned manipulation* — «запланированная манипуляция») эксперимента — первое необходимое условие его проведения, возможность самостоятельного управления *независимой переменной*. На практике — заранее составленный план предъявления условий или уровней, независимой переменной, т.е. экспериментальная схема.

ПОКАЗАТЕЛИ ИЗМЕРЯЕМЫЕ (*assessment*) — количественная характеристика деятельности испытуемого в эксперименте, материал для определения значений *зависимой переменной*. Адекватный выбор П.И. повышает *надежность* и тем самым *внутреннюю валидность* эксперимента: чем больше отдельных реакций или действий испытуемого характеризуется одним П.И., тем меньше *несистематическая изменчивость* получаемых результатов. От степени соответствия П.И. изучаемой реальной деятельности испытуемого зависит *внешняя валидность*.

ПОПУЛЯЦИЯ — контингент *потенциальных испытуемых*, объединенных общей возрастной, профессиональной или другой принадлежностью, изучаемый в экспериментальном исследовании. Распространение экспериментальных результатов на П. (*обобщение*) — основное преимущество эксперимента с привлечением большого числа испытуемых. Необходимым условием такого эксперимента является *представленность* П. исследуемой выборкой испытуемых, достигаемая с помощью их правильного отбора.

ПОСЛЕДОВАТЕЛЬНОСТЬ (*order*) — порядок предъявления испытуемым условий (или уровней) *независимой переменной* в экспериментах, проводимых по индивидуальным схемам. С последовательным характером предъявления проб связаны *эффекты П.* (в частности, *эффекты переноса*). Выделяют три типа П., которым соответствуют основные схемы *интраиндивидуального эксперимента*:

- **схема случайной П.** — предъявление испытуемому условий (или уровней) независимой переменной в случайном порядке. Эффективный способ *интраиндивидуального контроля факторов времени*. Для исполь-

зования схемы (и повышения надежности) требуется большое количество проб. Систематическое смещение практически устраняется (за счет усреднения эффектов переноса), кроме случаев асимметричного взаимовлияния предъявляемых условий. Случайная П. применяется также для распределения экспериментальных задач по разным условиям независимой переменной;

— **схема регулярного чередования** — поочередное предъявление испытуемому двух условий независимой переменной; способ интраиндивидуального контроля факторов времени. При достаточно большом количестве проб повышается надежность эксперимента, усредняются эффекты однородного, симметричного и неоднородного переноса. Регулярный характер предъявления условий не позволяет устранить ряд источников систематического смещения: предубеждения экспериментатора, асимметричные взаимовлияния условий независимой переменной, а также последствия периодических событий, происходящих (по объективным причинам) во время предъявления одного из этих условий;

— **схема позиционно уравненной (counterbalanced) П.** — П. предъявления условий (или уровней) независимой переменной, в которой позиции (номера) каждого из них составляют в среднем равные числа. Применяется при относительно небольшом количестве проб (или *блоков* проб). Эффективна как способ контроля факторов времени, если его изменения линейны, а также однородных и симметричных влияний П. В случаях неоднородного и асимметричного переноса схема не обеспечивает контроля систематического смещения. Возможность усреднения последних видов эффектов П. проб возникает при использовании позиционного уравнивания сразу для нескольких П. уровней в кроссиндивидуальных экспериментах (см. *Уравнивание позиционное*).

ПРЕДУБЕЖДЕНИЯ ЭКСПЕРИМЕНТАТОРА (experimenter bias) — один из источников систематического смещения, оказывающий побочное благоприятное (или неблагоприятное) влияние на действие одного из условий независимой переменной. П.Э. (в отношении этого условия) могут передаваться испытуемым и влиять на их деятельность в эксперименте, приводить к необъективности учета экспериментальных данных и т. п. Способы контроля П.Э. связаны с повышением точности соблюдения экспериментальных процедур (в частности, с применением автоматических устройств). Если снятие показателей необходимо включает субъективный элемент, то используется схема случайной последовательности предъявления условий, привлекаются методы экспертных оценок и т. п. Эффективным средством устранения П.Э. (и испытуемых) является постепенное и точно фиксируемое изменение независимой переменной в многоуровневом эксперименте. П.Э. и испытуемых — основные (и трудноконтролируемые) побочные переменные в корреляционных исследованиях. В специальной литературе термин «bias» часто переводится буквально — «байес», причем выделяются его различные виды, каждый из которых представляет собой побочное систематическое влияние.

ПРОТОКОЛИРОВАНИЕ (documentation — «подтверждение документов») — второе необходимое условие проведения эксперимента. П. плана эксперимента и хода его проведения обеспечивает полный учет данных для составления *экспериментального отчета*.

ПРОЦЕСС БАЗИСНЫЙ — составная часть представления комплексной зависимой переменной, входящей в теоретическую модель, которая позволяет интерпретировать результаты эксперимента. Такие модели служат для постановки точных экспериментальных гипотез о механизмах изучаемого поведения. Например, модели о взаимодействии двух П.Б., на каждый из которых независимая переменная оказывает различное воздействие, являются источником гипотез о *максимальной (или минимальной) величине* (см.). При проведении эксперимента П.Б. соответствует базисная переменная и от адекватности этого соответствия зависит его внутренняя валидность.

РАЗЛИЧИЯ ИНДИВИДУАЛЬНЫЕ (*subject variable* — «субъективная переменная», или фактор) испытуемых — основной источник ненадежности (и тем самым — нарушения внутренней валидности) в групповом эксперименте. Испытуемые могут различаться как по известным признакам (возраст, пол и т.д.), так и по трудноконтролируемым. Усреднение Р.И. испытуемых экспериментальной и контрольной групп достигается с помощью различных схем межгруппового сравнения (см. Распределение, Отбор). Как в экспериментальных, так и — корреляционных исследованиях построению групп может предшествовать выявление (и точное определение) Р.И., важных для изучаемого поведения, с помощью *предварительных проверок* (см. Стратегия построения групп).

РАСПРЕДЕЛЕНИЕ (*assigning*) — способ организации экспериментальных групп из имеющихся в наличии испытуемых — либо специально привлеченных («сaptive», «заложников»), либо добровольных участников эксперимента. В соответствии со стратегиями, применяемыми для Р. испытуемых по группам, выделяют три эффективные схемы межгруппового сравнения: *случайное Р. групп, попарное Р. групп* и *случайное Р. групп с предварительным выделением слоев* (или *последовательное случайное Р.*). Если признаки *подбора пар* и *выделения слоев* тесно связаны с изучаемой деятельностью, то отвечающие им схемы обеспечивают более высокую внутреннюю валидность по сравнению со случайным Р. при одинаковом числе испытуемых. Достижение же внешней валидности эксперимента зависит не только от Р. выборки испытуемых, но и от их правильного отбора.

РЕЗУЛЬТАТ ДЕЙСТВИЯ ОСНОВНОЙ (*main effect*) — количественно выраженная сила воздействия независимой переменной на зависимую (базисную) переменную. Вычисляется как разность средних значений зависимой переменной, полученных при воздействии разных уровней второй независимой переменной в факторном эксперименте. Для полноценного анализа полученных данных необходимо учитывать не только основные Р.Д. привлекаемых независимых переменных, но и взаимодействия между этими переменными;

— *простой Р.Д.* — Р.Д. независимой переменной, не связанный с влиянием дополнительных факторов (альтернатива — действие комбинации нескольких независимых переменных). Определение простого Р.Д. требует проведения факторных экспериментов при разных уровнях дополнительных переменных.

РЕПРЕЗЕНТАТИВНОСТЬ эксперимента — степень приближения реального эксперимента (или конкретной экспериментальной схемы) к одному из видов безупречного эксперимента. Поскольку реальный экс-

перимент в принципе не может быть безупречным, постольку оценка Р. каждого данного эксперимента (или схемы) проводится обычно путем сравнения его с другим аналогичным экспериментом (или схемой) по отношению к одному из видов безупречного эксперимента, который служит *образцом* для сравнения.

Репрезентативность выборки испытуемых — степень адекватности отражения свойств и характеристик изучаемой популяции в выборке испытуемых, привлеченных для эксперимента. Высокая Р. выборки означает высокую *представленность* популяции, то и другое достигается с помощью правильного отбора испытуемых. Термин «Р.» (здесь — «соответствие») используется также по отношению к конкретному испытуемому, а иногда — к условиям проведения эксперимента.

Буквальный перевод термина «*representativeness*» принят в специальной литературе. Однако при переводе однокоренных слов мы сочли более удобным русский вариант: «*to represent*» — «*представлять*», непосредственно передающий содержание. При чтении учебного текста это позволит избежать трудностей в понимании таких терминов, как «представленность» (ср. «репрезентированность»), «представитель» (ср. «репрезентант») и т. п.

СМЕШЕНИЕ (*confounding*) СИСТЕМАТИЧЕСКОЕ — второй основной источник нарушения **внутренней валидности**.

С. С. связано с тем фактом, что действие **независимой переменной** сопровождается целым рядом других переменных, которые могут систематически различаться при предъявлении разных условий независимой переменной, и тем самым оказывать на действие одного из них благоприятное (или неблагоприятное) влияние. Предъявить же эти условия одновременно невозможно, и поэтому иногда С. С. называют «*процедурным*».

Факторы, порождающие С. С., можно разделить на три группы. Во-первых, это различия уровней необходимых дополнительных переменных: факторов задачи, индивидуальных различий испытуемых и т. п. Во-вторых, периодические или долговременные изменения побочных факторов, входящих в состав факторов времени (например, изменение погодных условий, периодическая регулировка экспериментальной аппаратуры и т. д.). И наконец, различия взаимных влияний самих условий независимой переменной, приводящие (в индивидуальных экспериментах) к эффектам последовательности.

Совокупность способов контроля С. С. можно описать следующим образом. Во-первых, с усовершенствованием экспериментов появляется возможность полного устранения некоторых (но не всех) источников С. С. Так, в искусственном эксперименте унифицируется ряд важных дополнительных переменных, в групповом — устраняются факторы задачи и эффекты последовательности, в многоуровневом — предубежденность испытуемых. Во-вторых, для усреднения результатов неустраимых влияний применяются различные экспериментальные схемы. В интраиндивидуальных экспериментах, где таких влияний больше всего, наиболее эффективной является схема случайной последовательности (и это понятно, ведь остальные последовательности — систематичны): только она позволяет усреднить эффекты неоднородного (и отчасти — асимметричного) переноса, снять влияния предубеждений экспериментатора и т. п.

Схемы межгруппового сравнения позволяют достичь соответствия групп по индивидуальным различиям испытуемых. Кроме того, привлекаются меры для контроля отдельных факторов: подбор и распределение экспериментальных задач, применение косвенных инструкций и т.д.

В специальной литературе термин «*confounding*» иногда переводят как «*смешивание*». Предложенный нами вариант выбран не только в силу чуть лучшего звучания по-русски. Ведь постоянное взаимное сопровождение, т.е. смешивание основной и дополнительной переменных происходит в реальности (где, собственно, нет никаких проблем контроля С.С.), а экспериментатор имеет дело с уже совершившимся событием — фактом С., последствия которого ему нужно предусмотреть, устранить или усреднить, т.е. проконтролировать.

СМЕШЕНИЕ СОПУТСТВУЮЩЕЕ (*associative confounding*) — источник нарушения внутренней валидности, возникающий с необходимостью оперировать единичными переменными для проверки точных экспериментальных гипотез. Напрямую это оказывается невозможным даже в лабораторном эксперименте, где независимая переменная освобождена от побочных влияний: предъявление активного условия независимой переменной (например, расщепление свода) необходимо (дословно — «*ассоциативно*») связано с воздействием, не входящим в экспериментальную гипотезу (оперирование животного). По характеру этой связи различаются два вида С.С. — искусственное и естественное. С.С. неустранимо, его последствия можно только усреднить. Проблема контроля С.С. возникает также в отношении зависимой (точнее, базисной) переменной, когда для проверки теоретической модели требуется точно определить результат действия независимой переменной (а результат, свидетельствующий, например, о нарушении памяти, может быть приписан нарушению восприятия). Кроме того, вопросы С.С. являются важными для теоретических корреляционных исследований, где необходимо различать возможные детерминанты полученных результатов (например, порядок рождения детей, величина семьи, социальное происхождение родителей и т.д.);

— искусственное (*artifactual*) С. — сопутствующее С. независимой или базисной переменной в лабораторном эксперименте, вызванное своеобразием условий и процедур его проведения. Для контроля искусственного С. независимой переменной (например, введение лекарства — факт инъекции) контрольной группе испытуемых предъявляется не просто пассивное условие этой переменной (отсутствие инъекции), но специальное контрольное условие (инъекция без лекарства), и, таким образом, при сравнении результатов экспериментальной и контрольной групп воздействия сопутствующего влияния усредняются. Контроль искусственного С. базисной переменной (например, нарушения узнавания у животных, требующего для своего исследования привлечения задач на запоминание по ассоциации) возможен только в факторном эксперименте: используется вторая, контрольная переменная (объекты, не получающие подкрепления, либо предъявляются, либо нет), при активном условии которой воздействие испытывает только сопутствующая переменная, а при пассивном — и сопутствующая, и базисная;

— естественное (*natural*) С. — натуральная, не зависящая от методических процедур связь независимой (или базисной) и сопутствующей ей

переменных, не позволяющая проверить точную гипотезу в лабораторном эксперименте (например, плач ребенка может быть вызван тем, что из комнаты вышла мать, или же тем, что он просто остался один). Способ контроля естественного С. независимой переменной аналогичен введению контрольного условия, используется расширенная переменная (уход человека): при одном из ее условий (уход матери) сопутствующая переменная смешивается с основной, а при другом (уход ассистента) — нет; при сравнении этих двух условий воздействие сопутствующей переменной можно «вынести за скобки». Естественное С. базисной переменной (например, связь памяти с восприятием) контролируется так же, как и искусственное — привлекается контрольная переменная (здесь — интервал отсрочки выполнения тестовой задачи): «чистое» воздействие основной независимой переменной (рассечение свода) на базисную (нарушение памяти) определяется как разность ее воздействий при одном из условий контрольной переменной (длинная отсрочка), затрагивающем и базисную, и сопутствующую переменные (память и восприятие), и при другом (короткая отсрочка), затрагивающем только сопутствующую (восприятие).

СООТВЕТСТВИЕ (*appropriateness*) — степень адекватности отражения изучаемой реальности или теоретического представления о ней в проводимом эксперименте. Проблемы С. связаны с обеспечением внешней валидности (и ее безупречный образец неслучайно назван экспериментом полного С.), они возникают в отношении всех составных частей экспериментальной гипотезы: независимой, зависимой и уровней дополнительных переменных. Контроль С. обычно представляет собой анализ (предварительный или констатирующий) всех последствий изменения реальности в наличной экспериментальной ситуации. Так при интерпретации результатов искусственных экспериментов рассматриваются главным образом вопросы о С. значений зависимой переменной, а также уровней ключевых и других дополнительных переменных (например, анализ действий испытуемого, выполняемых одновременно с задачей, снятия напряженности и сжатия (*compression*) проб во времени. Проблемы С., возникающие в основных видах эксперимента, а также способы контроля и обобщения результатов приведены в табл. 1.

СОСТОЯНИЕ, или **УСЛОВИЕ** (*treatment*) — форма предъявления независимой переменной. Независимая переменная имеет обычно два или несколько С. (например, два метода заучивания фортепьянных пьес — целостный и частичный), причем С. всегда различается качественно. Содержание термина «*treatment*» в данном контексте, т.е. «форма, или способ, предъявления, действия», трудно передать на русский язык одним словом. Выбор термина «С.» связан, прежде всего, с конкретной спецификой первых экспериментов, рассмотренных в книге (скажем, использование и неиспользование наушников). Среди возможных его синонимов автор указывает термин «У.» (*condition*), и именно он оказывается удобным для описания большинства приводимых экспериментов. Поэтому термин «С.» встречается главным образом в первой главе, а далее заменяется более нейтральным — «У.».

СПОСОБ ПРЕДСТАВЛЕНИЯ ПОКАЗАТЕЛЕЙ, или результатов (*combining the assessments*) — способ количественной обработки измере-

мых показателей поведения испытуемого в эксперименте, применение которого окончательно определяет конкретный вид зависимой переменной. Итоги применения С.П. (или обработки) данных — значения зависимой переменной — представляются обычно в таблицах или на графиках.

СТРАТЕГИЯ ПОСТРОЕНИЯ (*constructing*) **ГРУПП** — адекватный способ организации испытуемых для участия в групповом эксперименте, позволяющий усреднить их индивидуальные различия в различных группах (внутренняя валидность) и адекватно представить изучаемую популяцию (внешняя валидность). Выделяют три основных С.П.Г. — случайную (*random*, или *randomization*, «рандомизация»), попарную (*matching*, «подбор пар») и случайную с предварительным выделением слоев (*stratified random*) (комбинация первых двух). Каждая из них применяется для распределения испытуемых по группам, а первая и третья — для отбора испытуемых из популяции. Это дает пять эффективных схем — схем межгруппового сравнения. Отсутствие специальной С.П.Г., т.е. использование реально существующих (*existing*) групп, является неудачной схемой, неизбежно порождает систематическое смещение независимой переменной с индивидуальными различиями испытуемых. Для достижения внутренней валидности в лабораторных экспериментах (в том числе — с небольшим количеством испытуемых) применяются следующие С.П.Г.: 1) *предварительные меры* — случайное распределение или отбор; 2) *серийное решение* — достижение соответствия групп по ходу эксперимента (в частности, с помощью предварительной проверки испытуемых по значимым характеристикам) и 3) *использование максимально сходных индивидов* (например, животных одного помета). Попарная С. и выделение слоев (*stratification*, «стратификация») используются также как способы контроля в корреляционных исследованиях: это С. попарного подбора испытуемых (в разные группы) и построение однородных подгрупп, т.е. подбор пар или групп таких испытуемых, которые уравнины между собой по всем переменным, кроме интересующей исследователя.

СХЕМА (*design*) **ЭКСПЕРИМЕНТАЛЬНАЯ** — определенный план предъявления испытуемым (или их группам) разных условий независимой переменной для адекватной проверки экспериментальной гипотезы. С.Э. классифицируется по трем основным параметрам: 1) *основание сравнения условий* (или уровней); 2) *тип изменения* и 3) *число независимых переменных*;

— *интраиндивидуальная С.Э.* — предъявление одному (или нескольким — группе) испытуемому всех исследуемых условий независимой переменной. Используется главным образом в практических индивидуальных экспериментах (и поэтому иногда называется просто С. индивидуального эксперимента). Различаются по используемым последовательностям предъявления условий: 1) С. случайной последовательности; 2) С. регулярного чередования; 3) С. позиционно уравненной последовательности;

— *межгрупповая С.Э.* — предъявление каждого из исследуемых условий независимой переменной разным группам испытуемых. Сочетание двух типов набора испытуемых (распределение по группам и отбор из популяции) с тремя стратегиями построения групп дает пять возможных С. межгрупповых сравнений: 1) случайное распределение групп; 2) попарное распределение групп; 3) случайное распределение групп с

выделением слоев; 4) случайный отбор групп; 5) случайный отбор групп с выделением слоев (см. табл 4);

— **кроссиндивидуальная С.Э.** — предъявление каждому из испытуемых (или их групп) всех исследуемых уровней независимой переменной в определенной последовательности, причем позиции каждого уровня в наборе этих последовательностей численно уравниены. Применяется как наиболее оптимальная в лабораторных многоуровневых экспериментах. В соответствии с различными вариантами позиционного уравнивания выделяют: 1) реверсивное (обратное) уравнивание; 2) полное уравнивание; 3) латинский и 4) сбалансированный квадрат;

— С.Э., связанные с разными типами независимой переменной, легко различить по используемой терминологии: условия — для *качественного* и уровни — для *количественного* типа;

— **факторная С.Э.** — С. эксперимента с двумя или несколькими независимыми переменными. Применяется для проверки гипотез с *одним отношением*, обеспечивая контроль сопутствующего смещения базисной переменной и обобщения результатов, и для проверки *комбинированных* гипотез;

— **гомогенная С.Э.** — С. факторного эксперимента с одинаковыми основаниями сравнения и типами изменения всех привлекаемых независимых переменных. Примеры гомогенных С.Э. даны в табл 8.2;

— **гетерогенная С.Э.** — то же при разнородности оснований сравнения и типов изменения независимых переменных.

С.Э. иногда называют также особые способы контроля (например, контроль обоих видов сопутствующего смещения).

ТИП (*gradation*) независимой переменной — характер ее изменения, один из параметров классификации экспериментальных схем. Два Т. независимой переменной — *качественный* и *количественный* — характеризуют различия между ее условиями (в первом случае) или уровнями (во втором). Использование переменных количественного Т. позволяет вводить метрические шкалы и проверять точные экспериментальные гипотезы в многоуровневых экспериментах.

УЛУЧШЕНИЕ (*improvement*) (в узком смысле) — изменение (или имитация) реальности при проведении *практического* эксперимента с целью повышения его *внутренней валидности*. Так, «У.» реального мира в искусственных экспериментах позволяет устранить *систематическое смещение* и повысить надежность результатов за счет увеличения количества проб и сокращения *несистематической изменчивости*. В более широком смысле термин «*improvement*» переводится как *преимущество* каждого более совершенного типа эксперимента. Такими преимуществами являются, во-первых, возможность проверки новых экспериментальных гипотез и их более широкого обобщения и, во-вторых, контроль (и даже устранение) ряда источников нарушения внутренней валидности (см. табл. 2). Необходимо помнить, что практически каждое У. реальности (т.е. удаление от нее) порождает новые проблемы обеспечения внешней валидности, *соответствия* реальности (см. табл. 1).

УРАВНИВАНИЕ ПОЗИЦИОННОЕ (*counterbalancing*) — достижение численного равенства средних позиций каждого из условий (или уровней) независимой переменной в последовательности (или наборе после-

довательностей) их предъявления. У.П. применяется в индивидуальных экспериментах для контроля эффектов последовательности проб. Схема позиционно уравненной последовательности — одна из схем интраиндивидуального эксперимента (см. Последовательность). В многоуровневых экспериментах, где использование этой схемы требует слишком много времени (а межгрупповой — слишком большого числа испытуемых), применяется У.П. по определенному набору последовательностей уровней, каждая из которых предъявляется одному испытуемому (или одной группе). Таких схем У.П., или кроссиндивидуальных схем — несколько:

— реверсивное (обратное) У. (П.) — предъявление двух последовательностей уровней — прямой и обратной — двум разным испытуемым (или их группам); позволяет контролировать эффекты *однородного переноса*;

— полное У. (П.) — использование всех возможных вариантов последовательностей уровней с предъявлением каждой из них одному испытуемому (или одной группе); обеспечивает контроль эффектов *неоднородного переноса*, поскольку каждый уровень оказывается в каждой позиции одинаковое число раз; используется крайне редко из-за необходимости привлекать большое количество испытуемых;

— латинский квадрат (см. Квадрат латинский);

— сбалансированный квадрат (см. К. сбалансированный).

УРОВЕНЬ (*level*) — форма предъявления, или способ действия *независимой переменной количественного типа*. Привлечение большого количества У. в многоуровневом эксперименте позволяет преодолеть *неполноту* независимой переменной. Термин «У.» используется также применительно к разным способам действия *побочных и дополнительных переменных*. Определенный У. дополнительной переменной — необходимая составная часть *экспериментальной гипотезы*. В данном случае различия между У. являются, как правило, не количественными, а *качественными* (например, разные типы музыкальных пьес). Соответствие У. дополнительных переменных — главное условие достижения *внешней валидности*, а проведение экспериментов на разных У. этих переменных обеспечивает возможность широкого обобщения результатов. При контроле *сопутствующего смешения* в факторном эксперименте выделяют *активный* (т.е. У. действия) и *пассивный* У. (а также более и менее активные У.).

— *альфа* (α)-У.—вероятность появления случайного события (например, определенного значения зависимой переменной), при которой *нуль-гипотеза отвергается*. А.-у. иногда называют уровнем значимости статистического критерия, используемого для проверки *экспериментальной гипотезы*. Так, в гл. 6 приводятся примеры проверки гипотез с помощью *t*-критерия при А.-у., равных 0,05 и 0,01. С уменьшением величины А.-у. повышается строгость проверки исследуемого отношения между независимой и зависимой переменными, т.е. понижается *риск* ошибочного отвержения *нуль-гипотезы (ошибки I типа)*;

— *бета* (β)-У.— вероятность появления определенных значений зависимой переменной, свидетельствующих о наличии ее связи с воздействием независимой переменной, при которой *нуль-гипотеза не отвергается*. Величину, равную $1 - \beta$, называют *мощностью* используемого статистического критерия. При обработке результатов следует учитывать связь

между альфа- и Б.-у.: уменьшение альфа-у. приводит к снижению мощности критерия, т. е. к увеличению *риска* ошибочного неопровержения нуль-гипотезы (*ошибки II типа*). Выбор величин альфа- и Б.-у. должен быть адекватным общей задаче эксперимента.

УСЛОВИЕ: а) (*treatment*) — см. Состояние, или Условие; б) У. (*condition*) контрольное — способ контроля сопутствующего смещения независимой переменной. Введение У. К. на специальной группе испытуемых позволяет в итоге достичь постоянного уровня сопутствующей переменной и тем самым «очистить» активное У. независимой переменной. В данном случае (к счастью, он — единственный) термины «*condition*» и «*treatment*» приходится различать по контексту (контрольное У. и У. независимой переменной). Впрочем, идентичность русского перевода этих терминов не столь опасна: так, при контроле сопутствующего смещения базисной переменной новое, специально введенное У. (*condition*) становится одним из У. (*treatment*) второй независимой (контрольной) переменной. Добавим, что как в оригинале, так и при переводе «У.» используется и как просто слово языка, не получая терминологического смысла (например, «погодные У.» или «У. проведения эксперимента»).

ФАКТОР — любая реальность, влияющая на поведение испытуемого в эксперименте. Термин «Ф.» содержательно близок термину «переменная» и используется как его синоним при описании факторных экспериментов и для обозначения ряда побочных переменных — Ф. времени, Ф. задачи, субъективного Ф.:

— Ф. времени (*time variable*) — условное обозначение для совокупности Ф., оказывающих побочное влияние на результаты эксперимента (значения зависимой переменной) и связанных с изменениями, которые происходят с течением времени. К *изменениям во времени* автор относит, во-первых, известные побочные Ф., которые при правильном планировании эксперимента можно произвольно сохранять постоянными по своему уровню (например, время дня, погодные условия), и, во-вторых, различные виды *нестабильности во времени* побочных, независимых и зависимых переменных (причем сюда относятся изменения и в поведении испытуемого, и в измеряемых показателях). Нестабильность во времени трудно устранить непосредственно, необходимы способы *первичного* контроля, применение экспериментальных схем (здесь — интраиндивидуальных). Один из основных Ф., порождающих несистематическую изменчивость и ненадежность результатов, особо опасен для естественных и долговременных экспериментов;

— Ф. задачи (*task variable*) — совокупность побочных влияний на результаты эксперимента, связанных с различием экспериментальных *задач*, которые предъявляются испытуемым при разных условиях независимой переменной. Эти влияния (различия) усредняются с помощью правильного подбора задач. В большинстве индивидуальных практических экспериментов Ф. задачи входит в состав Ф. времени; применение межгрупповых схем позволяет полностью устранить его;

— субъективный Ф. (*subject variable*) (см. Различия индивидуальные).

ЭКСПЕРИМЕНТ — проведение исследования в условиях заранее запланированного (в частности, специально созданного) изменения реальности с целью получить результаты, которые можно обобщить: сред-

ство проверки экспериментальной гипотезы. Э. называются как *реально выполняемые (actual)* исследования, так и их мысленные образцы (*standards*).

Реальные Э., обсуждаемые в книге, подразделяются, прежде всего, на естественные (дублирующие реальный мир), искусственные (улучшающие реальный мир) и лабораторные. Цели первых двух видов Э., как правило, чисто *практические*, а в третьем исследуются сами механизмы изучаемого поведения, и поэтому он называется также собственно *научным*:

— Э., который *дублирует* реальный мир, — Э., проводящийся в естественных условиях, в которых экспериментатор изменяет только *независимую переменную*; это индивидуальный Э. в смысле распространения его результатов только на данного конкретного испытуемого;

— Э., который *«улучшает»* реальный мир, или *искусственный Э.* — Э. в условиях имитации реальности, позволяющий достичь относительной стабилизации уровней побочных и дополнительных переменных;

— лабораторный Э. — Э. в условиях специального выделения независимой переменной и очищения ее условий.

Реальные Э. различаются также по используемым в них экспериментальным схемам, получая от них свои названия:

— индивидуальный, или интраиндивидуальный Э. (см. Схема экспериментальная);

— Э. с одним испытуемым (*single-subject e.*) — частный вариант индивидуального Э.;

— групповой, или межгрупповой Э. (см. там же);

— кроссиндивидуальный Э. (см. там же);

— бивалентный Э. — Э. с двумя условиями независимой переменной;

— мультивалентный, многоуровневый Э. — Э. с несколькими (более двух) уровнями независимой переменной;

— факторный Э. (см. там же);

— многомерный (*mullivariate*) Э. — Э. с несколькими (не менее двух) независимыми и несколькими зависимыми переменными.

Мысленный образец для проведения любых возможных реальных Э. (выполнение которого невозможно или бессмысленно):

— *безупречный (perfect) Э.*, идея которого соотносится с понятием *валидности Э.* Различные типы безупречного Э. (примеры их содержательных интерпретаций приведены в табл. 3) соответствуют разделению внутренней и внешней валидности. Так, образцами для достижения высокой *внутренней валидности* являются:

— *идеальный Э.* — Э., при проведении которого изменяется только независимая переменная, а все другие факторы остаются неизменными; таким образом, исследуется только само отношение между независимой и зависимой переменными;

— *чистый (pristine) Э.* — разновидность идеального Э., при проведении которого экспериментатор оперирует *единичной независимой переменной* и ее полностью очищенными условиями, мысленный образец для лабораторного Э.;

— *бесконечный Э.* — бесконечно продолжающийся Э. (т.е. Э. с бесконечным количеством проб, испытуемых и т.п.), позволяющий усреднить результаты неизбежных изменений всех побочных факторов, влияющих на зависимую переменную.

Мысленный Э., обладающий безупречной внешней валидностью — Э. полного соответствия — Э. с привлечением таких уровней необходимых дополнительных переменных, которые совпадают с уровнями этих переменных в изучаемой реальности.

ЭФФЕКТЫ ПОСЛЕДОВАТЕЛЬНОСТИ (*sequence effect*) — побочные и, как правило, систематические влияния на результаты эксперимента (значения зависимой переменной), связанные с предъявлением испытуемому экспериментальных *проб* (задач) в определенной последовательности. В специальной литературе иногда пользуются буквальным переводом — «*сериальные Э.*». Основной источник систематического смещения в индивидуальных экспериментах. В межгрупповых экспериментах Э. П. отсутствуют. В интра- и кроссиндивидуальных экспериментах выделяют следующие разновидности Э. П.:

— Э. переноса (*transfer*) — результат взаимных влияний последовательно предъявляемых условий (уровней) независимой переменной, дающий преимущество действию одного из них. По характеру взаимовлияний различают *однородный* и *неоднородный*, *симметричный* и *асимметричный* Э. Наиболее уязвимы в отношении этих Э. регулярные последовательности предъявления условий — *регулярное чередование* и *позиционно уравненная последовательность* (индивидуальный эксперимент), а также *реверсивное уравнивание* (кроссиндивидуальный). При использовании этих схем устраняются Э. только однородного и симметричного переноса. Э. неоднородного переноса усредняется с применением *случайной последовательности* и остальных кроссиндивидуальных схем. Последствия асимметричного взаимовлияния условий зачастую не устраняются даже при случайном предъявлении каждого из них;

— Э. ряда (*range*) — благоприятный (или неблагоприятный) Э. асимметричного переноса на действие того или иного уровня независимой переменной в многоуровневом эксперименте. Ряд — это последовательность *уровней независимой переменной* от наименьшего количественного значения к наибольшему. Э. Р. зависит от удаленности позиции предъявления уровня от концов ряда и объясняется тем, что в позиционно уравненных последовательностях, используемых в кроссиндивидуальных схемах, каждому уровню ни разу не предшествуют идентичные уровни: более низким предшествуют более высокие и наоборот;

— Э. центрации (*centering*) — частный вариант Э. ряда, благоприятный для действия уровней независимой переменной, предъявляемых в середине ряда, связанный с тем, что именно этим уровням предшествовали как более низкие, так и более высокие уровни. Э. ряда и центрации определяются характером взаимовлияний между настоящим и предшествующим уровнями независимой переменной в последовательности их предъявления и могут сохраняться при использовании любых кроссиндивидуальных схем.

Проблемы соответствия (достижения внешней валидности) и возможные способы их решения (контроля) в различных видах эксперимента

Вид эксперимента	Проблемы соответствия	Способы контроля	Достижимая цель
Естественный (индивидуальный)	Соответствие уровня ключевой дополнительной переменной	Подбор задач, соответствующих изучаемой деятельности	Обобщение результатов на изучаемую деятельность и рекомендации для каждого из испытуемых
Искусственный	Соответствие всех переменных, вошедших в гипотезу, их реальным аналогам	Выбор адекватных условий независимой переменной, измеримых показателей деятельности испытуемого и способов их представления; анализ возможных влияний имитации реальности	Обобщение результатов на изучаемые практические ситуации
Групповой	Соответствие исследуемой выборки и изучаемой популяции	Схемы отбора испытуемых	Распространение результатов на популяцию
Лабораторный	Соответствие методических процедур и теоретических понятий; проблема универсальности результатов	Выбор методических процедур	Проверка и уточнение теоретических моделей
Факторный (и корреляционное исследование)	Интерпретация характера полученного результата действия независимой переменной (установленных корреляций)	Проведение серии экспериментов при разных уровнях дополнительной переменной (в корреляционных исследованиях подбор групп по этим уровням и статистический контроль)	Наиболее широкое обобщение результатов; выявление простого (или комбинированного) действия независимой переменной; (в корреляционных исследованиях выявление базовых факторов)

Основные преимущества различных видов эксперимента и некоторые факторы, нарушающие их внутреннюю валидность

Вид эксперимента (основание сравнения)	Достигаемые преимущества	Факторы, нарушающие внутреннюю валидность	Способы их контроля
Искусственный (интраиндивидуальный)	Устранение систематического смещения и сокращение ненадежности данных	Эффекты последовательности, предубеждения экспериментатора, факторы задачи и др.	Интраиндивидуальные схемы, адекватный подбор задач
Групповой	Возможность распространения результатов на популяцию и применения межгрупповых схем; устранение эффектов последовательности	Индивидуальные различия испытуемых	Схемы распределения испытуемых по группам
Лабораторный	Возможность методического обеспечения единичной независимой переменной (и очищения ее условий); возможность проверки точных экспериментальных гипотез (в том числе — нуль-гипотезы)	Сопутствующее (искусственное и естественное) смещение независимой переменной	Введение контрольного условия, или расширенной переменной
Многоуровневый (кроссиндивидуальный)	Преодоление неполноты независимой переменной, дополнительный контроль сопутствующего смещения (осведомленности испытуемых); возможность проверки точных гипотез о количественных отношениях	Эффекты неоднородного переноса, ряда, центрации; неверно установленное отношение между независимой и зависимой переменными	Кроссиндивидуальные схемы; контроль репрезентативности обобщенных данных, адекватного числа уровней, проверка правильности измерений

Факторный	Контроль сопутствующего смещения базисной переменной и обобщения результатов при проверке гипотез с одним отношением; возможность проверки комбинированных гипотез и измерения взаимодействий	Дополнительные сопутствующие переменные	Новый факторный эксперимент
Корреляционное исследование	Возможность проверки гипотез о соотношении (и взаимодействиях) факторов без их активного изменения	Все переменные, помимо интересующих исследователя	Статистический контроль, подбор групп, уравненных по побочным или дополнительным факторам

Примеры содержательной интерпретации типов безупречного эксперимента (образцов)
для различных видов реальных экспериментов

Вид эксперимента	Идеальный (чистый)	Образец Бесконечный	Полного соответствия
Практический (индивидуальный)	Одновременное предъявление испытуемому всех условий независимой переменной, использование одних и тех же задач для разных условий	Бесконечное количество проб	Совпадение уровня дополнительных переменных с реальными (решение тех же самых задач)
Лабораторный (индивидуальный или групповой)	Оперирование единичной независимой переменной (и ее чистыми условиями)	Бесконечное множество реальных экспериментов, в каждом из которых предъявляются новые пробы (или участвуют новые испытуемые)	Совпадение методических процедур с теоретическими понятиями
Групповой	Предъявление условий независимой переменной одной и той же группе испытуемых	Бесконечное количество испытуемых	Участие тех же самых испытуемых, на которых будут распространяться результаты
Многоуровневый	Предъявление каждого из уровней во всех позициях их последовательности	Бесконечное количество уровней	
Факторный	Оперирование чистой базисной переменной	Бесконечное количество независимых переменных (факторов)	Привлечение всех уровней дополнительных переменных (итоговый)

Предметный указатель

- Абсолютная ошибка 92
Алгебраическая ошибка 93
Альфа-уровень 185—190
ANOVA (Программа дисперсионного анализа на ЭВМ) 246 (см. также Дисперсионный анализ)
Асимметричные эффекты переноса 56—57
Базисные переменные, их смешение 259—260
Базисные процессы противоположных направлений 214—216, 276—277
Безупречный эксперимент:
— виды 33—35
— идеальный (см.)
— чистый (см. Идеальный эксперимент)
— бесконечный (см.)
— полного соответствия (см.)
Бесконечный эксперимент:
— количество проб 33—34
— количество испытуемых 123
— уровни независимой переменной 207—208
— нуль-гипотеза 189—190
Бета-уровень 187
Бивалентные эксперименты 205
Блоки проб 51
Валидность:
— определение 37—39
— внутренняя (см.)
— внешняя (см.)
— операциональная 166—168, 302
— проверка 317—318
Взаимодействие:
— определение 256
— виды 256—259
— измерение 255—256
— предсказание 264—265, 271—279
— в корреляционных исследованиях 309—312
Внешняя валидность:
— определение 37—39
— отношение к внутренней валидности 131—132
— препятствия адекватного соответствия
— независимой переменной 89, 166—168
— зависимой переменной 90—93, 166—168, 302
— уровни дополнительных переменных 34—35, 93—96, 264—269
— и реальный мир 88—97
— и изучаемая популяция 129—133, 166—168, 308
— и теоретические представления
Внутренняя валидность:
— определение 37—38
— отношение к внешней валидности 131—132
— надежность (см.)
— систематическое смешение (см.)
— контроль (см.)
— искусственный эксперимент 75
— возможные нарушения:
— при интраиндивидуальных сравнениях 39—46
— при межгрупповых сравнениях 120—123
— при кроссиндивидуальных сравнениях 224—225, 227—229
— выводов о нуль-гипотезе 189—190

- в корреляционных исследованиях 301—302, 307—312
- невыявленное отношение 207, 229—233
- представленность индивида 230—232
- Внутригрупповая оценка популяционной дисперсии 238
- Выборка из популяции 106, 128—129
- Выборка статистическая 31, 63
- Выборочное распределение:
 - различие между средними 197—198
 - t 200—201
 - F 238—239
- Гетерогенные схемы 280
- Гипотезы о максимальном (минимальном) значении 212
- Гистограмма 30
- Гомогенные схемы 280
- Диаграммы разброса 314, 326
 - линия предсказания, ее наклон 316—317, 326—327
 - ошибка предсказания 316—317
 - таблицы предсказаний 316
- Дисперсионный анализ:
 - одномерный 237—243
 - двумерный 285—291
- Дополнительные факторы при анализе больших массивов данных 192
- Единичная независимая переменная 148—150
- Естественное смешение 159—163, 259—261
- Зависимая переменная:
 - определение 14
 - оценка 14—15
 - компоненты:
 - ответы (поведение) испытуемого 14, 89—91
 - измеряемые показатели 14, 91
 - способ представления показателей 14—15, 91—93
 - нестабильность во времени 42—44
 - соответствие:
 - в искусственных экспериментах 90—93
 - в научных экспериментах 167—168
 - в корреляционных исследованиях 302
- Идеальный эксперимент:
 - определение 33
 - для интраиндивидуальных схем 33
 - в искусственных экспериментах 75—76
 - для межгрупповых схем 120
 - чистый эксперимент 153—154, 264
 - единичная независимая переменная 148—150, 209—211
 - очищение условий 150—154
 - для кроссиндивидуальных схем 230—231
 - для корреляционных исследований 293—294
- Изменения во времени:
 - известные побочные переменные 39—40
 - нестабильность:
 - независимой переменной 42
 - зависимой переменной 42—44
 - побочных переменных 40—42
- Измеряемые показатели ответов (поведения) испытуемого 14, 91—93
- Интраиндивидуальные схемы:
 - последовательности предъявления условий:
 - случайная 50
 - регулярного чередования 50—51
 - позиционно уравненная 51—52
 - контроль:
 - факторов времени 49—52
 - факторов задачи 52—53
 - эффектов последовательности 54—57
 - в многоуровневых экспериментах 222—223
- Искажения:
 - при психологическом шкалировании 232—233

- при измерении небольших физических величин 232—233
- Искусственное смещение 154—159, 261—264
- Испытуемые:
 - распределение и отбор 124—133, 169—172
 - имеющиеся в наличии, их привлечение к эксперименту 133
 - выбор экспериментальных животных 168—170
- Качественная независимая переменная 210—211
- Количественная независимая переменная 206—207, 210—212
- Количественные гипотезы 207
- Конкурирующая гипотеза 15
- Контроль, способы контроля:
 - определение 46
 - первичный 46—49
 - факторов времени 49—52
 - факторов задачи 52—53
 - эффектов последовательности 54—57, 223—229
 - индивидуальных различий 124—129
 - сопутствующего смещения 154—164, 208—209, 259—264, 283—302, 307—312
 - искусственного смещения 156—159, 261—264
 - естественного смещения 159—164, 259—261, 301—302, 307—312
 - контрольное условие 156—164, 263
 - контрольная группа 157
 - базисных переменных 259—264
 - контрольная переменная 262
 - в корреляционных исследованиях 301—302, 308—312
 - статистический (в отличие от организуемого в эксперименте) 301—302, 307—312
- Корреляционное исследование оптимальной приспособленности 294—302, 308—309
- Корреляционное исследование порядка рождения и интеллекта 302—308
- Корреляционное исследование с целью отбора контролеров 312—318
- Корреляционный подход 293
- Корреляционные исследования
 - определение 293—294
 - виды 318—321
 - побочные и дополнительные переменные, смещение 301—302, 307—312
 - сопутствующее смещение 301—302, 307—308
 - способы контроля:
 - попарный подбор испытуемых 308—309
 - однородные подгруппы 309—312
 - взаимодействие 310—312
- Кроссиндивидуальные схемы:
 - виды:
 - реверсивное уравнивание 224—225
 - полное уравнивание 225—226
 - латинский квадрат 226—227
 - сбалансированный квадрат 226—227
 - эффекты последовательности:
 - неоднородные 224
 - асимметричные 227—228
 - ряда 227—229
 - центрации 228—229
- Линия предсказания (регрессии), ее наклон (см. также Диаграммы разброса) 326—327
- Литература по экспериментальной психологии 164—166
- Межгрупповая оценка популяционной дисперсии 238
- Межгрупповые схемы:
 - классификация (таблица) 134
 - индивидуальные различия испытуемых 120—123
 - потенциальные испытуемые:
 - распределение 124—129
 - отбор 129—133
 - стратегии построения групп:
 - случайная 124—127, 130—131
 - попарная 127—128

- случайная с выделением слов 128—129, 132—133
- использование существующих групп 128
- предварительные меры 171
- сериальное решение 171
- в многоуровневых экспериментах 221—222, 229—232
- Многомерный эксперимент** 250
- Многоуровневый (мультивалентный) эксперимент** 205
- Мощность статистического критерия** 187—188
- Надежность:**
 - определение 47
 - несистематическая изменчивость (см.)
 - выбор адекватного количества:
 - проб 47
 - испытуемых 123
 - повышение в искусственных экспериментах 81—82, 86—87
 - и статистическая значимость 183—185
 - проверка 318
- Невыявленное отношение между независимой и зависимой переменными** 207—208, 229—234
- Независимая переменная:**
 - определение 13—14
 - условия (состояния) 14
 - нестабильность во времени 42
 - соответствие:
 - в искусственных экспериментах 89
 - в лабораторных экспериментах 166—168
 - уровни 24, 206—207
 - качественная 210—211
 - количественная 206—207, 210—212
 - неполнота 207
- Ненадежность (см. Надежность)**
- Неоднородные эффекты переноса** 54—55, 225
- Несистематическая изменчивость:**
 - независимой переменной 42, 47, 153—154
 - побочных переменных 39—42, 47—48, 153—154
 - в поведении испытуемого 42—44, 46—47, 121—122
 - измеряемых показателей 44, 47
 - неустановленного происхождения 40
 - сокращение в искусственных экспериментах 86—87
 - индивидуальных различий испытуемых 121—123
 - и статистические выводы 182—185, 189—190
- Нулевое взаимодействие** 256—259, 271—272, 277
- Нуль-гипотеза:**
 - отношение к бесконечному эксперименту 189—190
 - логика 178—181
 - отвержение или неотвержение 182—192, 200—202, 242—243
 - правило статистического решения 182, 185
 - статистический вывод 182—183
 - альфа-уровень 185—190, 242—243
 - бета-уровень 187—188
 - виды риска 186—188
 - ошибки I и II типа 186—190
 - мощность 187—188
 - надежность 183—185
 - отношение между решениями и выводами 185—190
 - валидность выводов 189—190
- Обобщение:**
 - определение 35
 - минимальное 11—12
 - в экспериментах с большим числом испытуемых 110—111
 - в экспериментах с выделением независимой переменной 173, 268
 - простые результаты действия 267—268
- Одновременные действия** 94—95
- Однородные подгруппы** 309—312
- Омега в квадрате** 139—143, 200—201

- Операциональная валидность 166—168
- Оптимальный уровень мотивации:
— закон Йеркса—Додсона 216, 269—271, 276—277
- Основание сравнения в экспериментальной схеме 279—281
- Основные результаты действия, их измерение 253—256
- Отклонение оценки 322
- Отрицательная корреляция 315
- Ошибки I типа 186—187
- Ошибки II типа 187—188
- Параметры генеральной совокупности (популяции) 61—63, 198, 285, 289
- Планируемые действия 22—23, 292—293
- Первичный контроль 46—49
- Переменная:
— определение 13
— независимая (см.)
— зависимая (см.)
— побочные и дополнительные (см.)
— контроль 262—263
- Пересекающееся взаимодействие 257—259, 275—278
- Побочные и дополнительные переменные:
— известные 39—40, 120—121
— нестабильность во времени 40—42
— факторы времени 48—52
— факторы задачи 44, 52—54
— эффекты последовательности 45, 53—57, 224—225, 227—229
— предубеждения экспериментатора 45—46, 301—302
— индивидуальные различия испытуемых 130—131
— «шум» в лабораторных экспериментах 210—211
— уровни:
— соответствия 35, 38
— индивидуальных различий 35, 38, 93—96
— и обобщение 264—268
— контроль (см.)
- Показатель ошибки 289
- Полное уравнивание 225—226
- Попарное распределение групп 127—128
- Попарный подбор испытуемых 308—309
- Популяции 106, 129—130
- Порядок взаимодействия (первый, второй) 278
- Предубеждения экспериментатора 45—46, 301—302
- Предубежденность испытуемого 302
- Представленность поведения испытуемого 229—231
- Причинность и корреляция 311—312
- Проверка валидности 318
- Проверка значимости 179 (см. также Статистическая значимость)
- Проверка надежности 318
- Протокол 18
- Протоколирование 31—33
- Процентное отношение 92—93
- Психологическое шкалирование 211—212, 232—233
- «Пучки» (неединичные независимые переменные) 150, 211—212
- Различия задач 44
- Расходящееся взаимодействие 258—259, 272—274, 277—278
- Реализм, его цена 97—98
- Результаты, явные и неявные 190—192
- Реверсивное уравнивание 224—225
- Репрезентативность (безупречного эксперимента) 35—36
- Риск, связанный с принятием нуль-гипотезы 186—188
- Секущая оценка 316—317
- Сжатие предъявляемых проб во времени 96
- Сила связи (независимой и зависимой переменных) 181—188, 263—264
- Систематическое смещение:
— определение 48—49
— контроль (см.)
— устранение с «улучшением» реальности 75—76

- процедурное:
- в интраиндивидуальных схемах 48—57
- в межгрупповых схемах 122, 128
- в кроссиндивидуальных схемах 224, 227—229
- и статистическая значимость 194—195
- Случайная стратегия (рандомизация) 48—49, 52—53, 124—127, 130
- Случайное распределение групп 124—127
- Случайное распределение групп с выделением слоев 128—129
- Случайные числа (таблица) 125—126
- Случайный отбор групп 130—131
- Случайный отбор групп с выделением слоев 132—133
- Сопутствующий процесс 215, 261—263, 276—277
- Сопутствующее смешение:
 - искусственное 154—159, 164, 261—263
 - естественное 159—164, 259—261
 - базисных переменных 259—264
 - в корреляционных исследованиях 387—389, 395—403
- Состояние (условие) 14
- Способ представления показателей 14, 91—93
- Способы контроля (см. Контроль)
- Среднее:
 - вычисление 29—30, 103
 - представление 30
- Среднее квадратичное:
 - внутригрупповое 241, 286
 - межгрупповое 241—242
 - по строкам, по столбцам 287—288
 - строки \times столбцы 288—289
- Стандартная ошибка различия между средними 199
- Стандартное отклонение:
 - вычисление:
 - по индивидуальным замерам 61—62
 - по данным интервальной классификации 103
 - и несистематическая вариация 61—62
 - как способ представления показателей 92—93
 - графическое представление 104
- Стандартные оценки (z -оценки) 323—324
- Статистическая значимость:
 - нуль-гипотеза 178—193, 242—243, 290
 - ограничения:
 - сильные эффекты 190—192
 - ограниченные эффекты 193
 - дополнительные факторы 192—193
 - внешняя валидность 194
 - систематическое смешение 194
- Статистические способы контроля 293, 308
- Статистические таблицы:
 - 1 (квадраты и корни квадратные) 64—66
 - 2 (критические значения t) 202—203
 - 3 (критические значения F) 245—246
- Статистический вывод 182
- Статистическое решение 182—185
- Статья 26
- Степени свободы 200, 241—243, 287—289
- Стресс 95—96
- Сумма квадратов:
 - внутригрупповая 240, 286
 - межгрупповая 240—241
 - по строкам, по столбцам 287—288
 - строки \times столбцы 288—289
- Схема позиционно уравненной последовательности (см. Интраиндивидуальные схемы)
- Схема регулярного чередования (см. там же)
- Схема случайной последовательности (см. там же)

- f:**
- критерий 182, 197—201
 - вычисление 200
 - выборочное распределение 197—199
 - критические значения 200
 - таблица 202—203
- Тип независимой переменной 280**
- Точность эксперимента 47**
- Уровень:**
- независимой переменной 13, 207
 - дополнительных переменных 38, 266—268
- Условие (состояние)**
- F:**
- критерий 237—240
 - описание 237—240
 - вычисление 239—242
 - выборочное распределение 238—239
 - отношение 242, 289
 - отвержение нуль-гипотезы 242, 289
 - критические значения (таблица) 245—247
- Факторные схемы:**
- определение 250
 - основные результаты действия 251—255
 - взаимодействие 255—256
 - гипотезы с одним отношением:
 - контроль 259—264
 - обобщение 264—268
 - простые гипотезы (о прямом действии) 264—265
 - комбинированные гипотезы 250, 269—279
 - и схемы с позиционным уравнением 268—269
- Факторный эксперимент 250**
- Фоновый уровень ответа 172**
- Формула предсказания (регрессии) 327**
- Частотное распределение 100—104, 119—120**
- Число независимых переменных в схеме 280**
- Чистота условий 150—151, 153—154**
- Чистый эксперимент (см. Идеальный эксперимент)**
- Экспериментальная гипотеза:**
- определение 14, 38
 - источники 164—165
- Экспериментальная группа 156—157**
- Экспериментальные схемы:**
- классификация (таблица) 281
 - интраиндивидуальные (см.)
 - межгрупповые (см.)
 - кроссиндивидуальные (см.)
 - многоуровневые 205
 - факторные (см.)
 - гомогенные 280
 - гетерогенные 280
 - многомерные 250
- Экспериментальный отчет 26—28**
- Эксперимент в ткацком цехе 8—11**
- Эксперимент на заводе «Хауторн» 41—42**
- Эксперимент по научению 167**
- Эксперимент полного соответствия:**
- определение 35
 - независимая переменная 89, 166—168
 - зависимая переменная 90—93, 167—168, 302
 - уровни дополнительных переменных 35, 38, 93—96, 129
 - и статистическая значимость 194 (см. также Внешняя валидность)
 - операциональная валидность 166—168, 302
- Эксперимент по социальной психологии 166—167**
- Эксперимент по этике труда 148—150**
- Эксперимент с введением диазепама 156, 264—267**
- Эксперимент с влиянием прошлого опыта на решение задач 271—274**
- Эксперимент с влиянием темпа предъявления стимулов на их запоминание 204—206, 212, 215**

- Эксперимент с временем реакции 222—226
- Эксперимент с выбором высотомера 83—86
- Эксперимент с двумя методами обучения испанскому языку 114—120
- Эксперимент с детским плачем 159—164
- Эксперимент с заучиванием фортепьянных пьес 15—19
- Эксперимент с исследованием памяти 216—217, 220—221, 271—272, 276—278
- Эксперимент с мысленной обработкой двигательных навыков 111—114
- Эксперимент с направленным обонянием 150—153
- Эксперимент с нерелевантной информацией 274—276
- Эксперимент с ночными посадками самолета 69—74
- Эксперимент с обработкой деталей 228—229
- Эксперимент с одним испытуемым 172
- Эксперимент со спасательным поиском на море 76—81
- Эксперимент со способами информирования о ценах на продукты 105—110
- Эксперимент со сравнением двух сортов томатного сока 20—26
- Эксперимент с оценкой тяжести 218—221
- Эксперимент с рассечением свода 154—156, 248—252
- Эксперимент с реакцией выбора 217—218, 220—221
- Эксперимент с точной гипотезой о количественных отношениях:
- абсолютно-абсолютное отношение 216—217
 - относительно-абсолютное отношение 217—218
 - относительно-относительное отношение 218—220
- Эффекты последовательности:
- описание 45, 54
 - эффекты переноса:
 - неоднородные 54—55, 224
 - асимметричные 55—57, 227—228
- Эффекты ряда 227—229
- Эффекты центрации 228—229

Оглавление

Предисловие научного редактора	3
Предисловие автора	5
Глава 1. Эксперименты, которые дублируют реальный мир	8
Шум на ткацком производстве	8
От повседневной жизни — к экспериментам	11
Планирование действий — первое условие эксперимента	12
Основные термины из словаря экспериментатора	13
Сравнение двух методов заучивания фортепьянных пьес — как протоколировать эксперимент	15
Протоколирование — второе условие эксперимента	19
Выбор сорта томатного сока — как описывать эксперимент	20
Экспериментальный отчет: заключительный этап оформления	25
Краткое изложение	27
Статистическое приложение: вычисление и представление средних значений времени реакции	29
Глава 2. Основы планирования эксперимента	31
Просто планы и планы более удачные	32
Безупречный эксперимент	33
Обобщение, репрезентативность и валидность	35
Факторы, угрожающие внутренней валидности	39
Повышение внутренней валидности с помощью первичного контроля	46
Схемы контроля за факторами времени в эксперименте с одним испытуемым	49
Как контролировать факторы задачи	52
Проблема эффектов последовательности	54
Краткое изложение	57
Статистическое приложение: определение генеральной совокупности и параметров на основе выборочных статистик; среднее и стандартное отклонение	61
Глава 3. Эксперименты, которые «улучшают» реальный мир	67
Ночные посадки самолетов под визуальным контролем	69
Первый способ улучшения реальности: устранение систематического смещения	75
Спасательный поиск на море	76

Второй способ улучшения реальности: больше данных — выше надежность	81
Выбор подходящего высотомера	83
Третий способ улучшения реальности: высокая надежность за счет сокращения несистематической изменчивости	86
Несколько способов сразу	87
Внешняя валидность: вопросы соответствия	88
Какова цена реализма!	97
Краткое изложение	98
Статистическое приложение: частотные распределения	100
Глава 4. Эксперименты на представительных выборках	105
Как сообщать покупателям о ценах на продукты?	106
Первое преимущество эксперимента с большим числом испытуемых: распространение результатов на популяцию ...	110
Мысленная отработка двигательных навыков	111
Второе преимущество эксперимента с большим числом испытуемых: возможность применения схем межгрупповых сравнений	113
Два метода обучения испанскому языку	114
Достижение внутренней валидности в экспериментах с межгрупповым сравнением	120
Три стратегии составления групп, три схемы сравнения ...	124
Внешняя валидность: представленность изучаемой популяции	129
Краткое изложение	135
Статистическое приложение: сила связи между независимой и зависимой переменными	138
Глава 5. Выделение независимой переменной	144
Изолированная независимая переменная	148
Очищение условий	150
Искусственное смещение	154
Естественное смещение	159
Источники экспериментальных гипотез	164
Операциональная валидность	166
Испытуемые	168
Итоги и перспективы	172
Краткое изложение	174
Глава 6. Значимые результаты	178
Нуль-гипотеза	180
От решений к выводам	185
Неприятные проблемы, которые остаются	190
Другие аспекты валидности	193
Краткое изложение	195
Статистическое приложение: <i>t</i> -критерий	197
Глава 7. Многоуровневые эксперименты	204
Многоуровневый эксперимент как контрольный	206

Количественные и качественные независимые переменные	210
Гипотеза о максимальной (или минимальной) величине ...	212
Гипотезы об абсолютных и относительных отношениях	216
Предыдущие экспериментальные схемы в приложении к многоуровневому эксперименту	221
Схемы с позиционным кроссиндивидуальным уравниванием	223
Можем ли мы доверять кривым?	229
Краткое изложение	234
Статистическое приложение: однофакторный дисперсионный анализ и F -критерий	237
Глава 8. Факторные эксперименты	248
Основные результаты и взаимодействия	251
Факторная схема при проверке гипотез с одним отношением	259
Проверка комбинированных гипотез	269
Классификация экспериментальных схем	279
Краткое изложение	281
Статистическое приложение: двухфакторный дисперсионный анализ	285
Глава 9. Корреляционные исследования	292
Исследование предпосылок оптимальной психологической приспособленности	294
Исследование порядка рождения и интеллекта	302
Контроль в корреляционных исследованиях	308
Исследование с целью отбора контролеров	312
Краткое изложение	321
Статистическое приложение: коэффициент корреляции	323
Библиография	328
Словарь экспериментатора	332
Предметный указатель	357

Учебное издание

Готтсданкер Роберт

Основы психологического эксперимента

Учебное пособие

Научный редактор *Ю. Б. Гиппенрейтер*

Редактор *М. И. Черкасская*

Ответственный редактор *С. А. Шаренкова*

Технический редактор *Е. Ф. Коржуева*

Компьютерная верстка: *Е. Ю. Матвеева*

Корректоры *Э. Г. Юрга, О. Н. Тетерина*

Изд № А-1146-І. Подписано в печать 28.01.2005. Формат 60×90/16.
Гарнитура «Таймс». Бумага тип. №2. Печать офсетная. Усл. печ. л. 23.
Тираж 5000 экз. Заказ № 14449.

Лицензия ИД № 02038 от 13.06.2000. Издательский центр «Академия».
Санитарно-эпидемиологическое заключение № 77.99.02.953.Д.004796.07.04 от 20.07.2004.
117342, Москва, ул. Бутлерова, 17-Б, к. 360. Тел./факс: (095)330-1092, 334-8337.

Отпечатано на Саратовском полиграфическом комбинате.
410004, г. Саратов, ул. Чернышевского, 59.

Книга английского психолога профессора Роберта Готтсданкера широко известна в России. В 1982 г. она была впервые переведена на русский язык и с тех пор стала одним из базовых учебных пособий по экспериментальной психологии в нашей стране. Это и не удивительно – книгу отличает предельно доступный стиль изложения, поэтому студенты предпочитают ее более современным, но чрезмерно сложным изданиям. Над русским переводом работали ведущие специалисты МГУ им. М.В. Ломоносова (профессор Ю.Б. Гиппенрейтер, доценты Ч.А. Измайлов и В.В. Петухов), которые не только точно передали содержание оригинала, но и подготовили дополнительные материалы, в их числе подробный «Словарь экспериментатора» В.В. Петухова.

Замечательно, что с появлением нового издания студенты-психологи получают возможность прочитать эту книгу не только в библиотеках.

ОСНОВЫ ПСИХОЛОГИЧЕСКОГО ЭКСПЕРИМЕНТА



Издательский центр «Академия»