

Роберт Готтсданкер. Глава 7. МНОГОУРОВНЕВЫЕ ЭКСПЕРИМЕНТЫ

Роберт Готтсданкер: Основы Психологического Эксперимента

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 секунд. 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 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 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

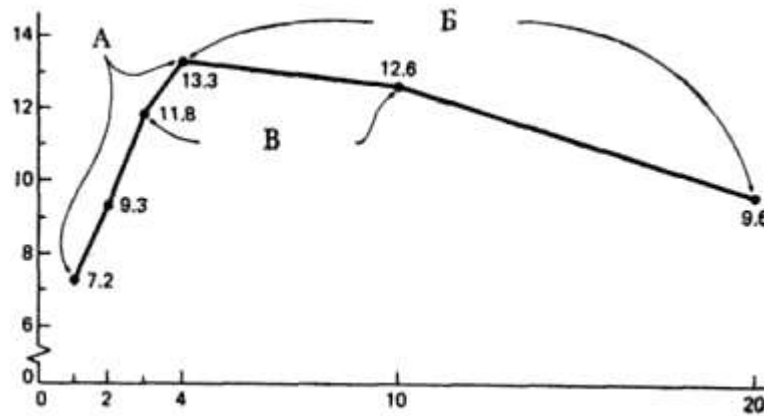


Рис. 7.1. Воображаемые двухуровневые эксперименты А, Б и В на примере результатов исследования Калфи и Андерсон (1971). Ось абсцисс — временные интервалы между предъявляемыми знаке-ми. Ось ординат — среднее количество правильных ответов (правильных ответов — заученных элементов списка), как это показано 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 (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 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 — *кросс-индивидуального* 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 нажатия 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 Сингх и Квери (1971) просто решили, что одного нажатия на шарик будет слишком мало. Точно так же нельзя предлагать очень много нажатий за один шарик, ибо в этой 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 Сингха и 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 смешения с другими переменными (см. гл. 5, с. 193). 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 смешения. К показу этого мы сейчас и перейдем.

Лучший контроль над сопутствующим смешением

Доказательство действия независимой переменной неубедительно, если явно возможно сопутствующее смешение, 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 использовалось только два условия, существовала бы ясная опасность обсуждаемого смешения. Пусть, например, в качестве активного уровня выбрано 3 миллиграмма лекарства на каждый килограмм веса 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 на лицо 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 смешение. Активный уровень независимой переменной (3 мг кофеина на 1 кг веса) 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 миллиграммов кофеина на килограмм веса, можно получить кривую, показанную на рис. 7.2. Уровень нулевой дозировки, конечно, обозначает плацебо.

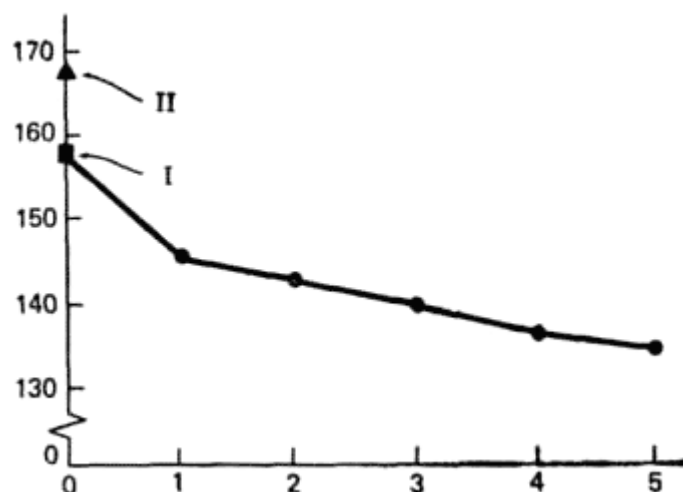


Рис. 7.2. Результаты воображаемого многоуровневого эксперимента с влиянием кофеина на время реакции. Ось абсцисс — дозы кофеина (мг/кг). Ось ординат — время реакции (мс). I — плацебо, II — без таблетки

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. Мало вероятно, что эта кривая может быть результатом нарастающего осознания увеличивающихся доз вещества. 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 1 шарик за 10 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 1 шарик за каждое нажатие) выбиралось также охотно, как и большее количество работы (например 1 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

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 Гейтвуд и Перлофф (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 отличаются по многим параметрам. Один из них состоит в том, что при чтении человек может забегать 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: вмещается 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 большую 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 Калфи и Андерсон имели «высокую степень осмысленности». Большая группа испытуемых предварительно оценила множество триграмм по степени их осмысленности. Все триграммы, использовавшиеся Калфи и 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, Ранквист, 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; 3-я — 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 Калфи и Андерсон (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 секунд экспериментального времени. Этот результат был как раз таким, который ожидали Калфи и 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

Калфи и 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 Роберт Йеркс и Джон Додсон. Йеркс, который интересовался эволюционным развитием, в то время тщательно изучал любопытное маленькое животное под названием танцующая мышь (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 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» (с. 467 — 468). *Средний* уровень составлял 300 единиц. Действие каждого уровня электроудара исследовалось на четырех мышах — двух самках и двух самцах.

В качестве зависимой переменной было выбрано среднее число проб, необходимых для достижения критерия научения всеми четырьмя мышами. Результаты показаны на рис. 7.3. Видно, что *минимальное* значение зависимой переменной, 80 проб, было получено при *промежуточном* уровне независимой переменной, 300 единиц стимуляции. На самом деле научение при самой слабой стимуляции было даже хуже, чем показано, поскольку одна 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

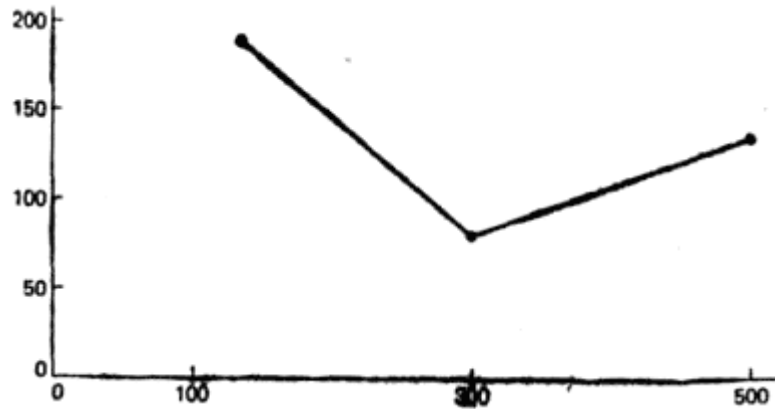


Рис. 7.3. Влияние силы электроудара на успешность научения (опыты на танцующих мышах). Ось абсцисс — сила электроудара (в условных единицах). Ось ординат — среднее количество проб, необходимых для достижения критерия успешного решения задачи Йеркса и Додсона: «Следует иметь в виду, что поскольку использовалось только три интенсивности стимула... возможно, наиболее благоприятная сила стимуляции обнаружена не была» (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. Бели бы в эксперименте Йеркса и Додсона мерой научения служило число правильных ответов, то при 300 единицах стимуляции достигался бы *максимум*. Если бы Калфи и 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 нажатия 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 282 «негативный» процесс при достижении независимой переменной высокого уровня становится сильнее «позитивного».

Эксперименты Калфи и Андерсон (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 Йеркса и Додсона (1908, с. 476), 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». 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 получили исследователи в своей дальнейшей работе. Вот их заключение: «По мере увеличения сложности различения интенсивность стимула, оптимальная для формирования навыка, приближается к порогу» (1908, 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 процентов больше, чем 1).

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

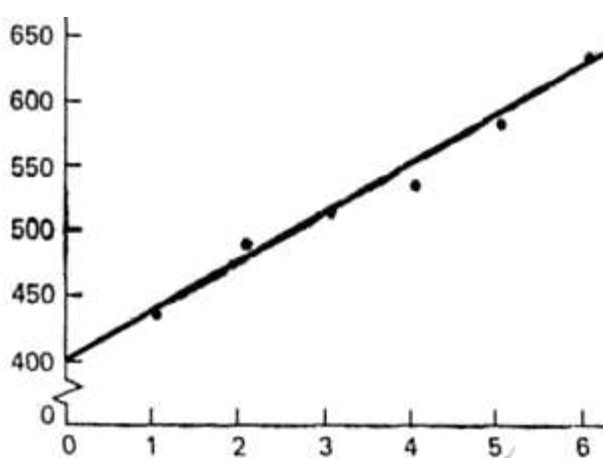


Рис. 7.4. Зависимость времени реакции опознания тестового стимула от количества знаков в ранее предъявленном наборе (эксперимент Стернберга, 1972). Ось абсцисс — размер предъявленного набора. Ось ординат — время реакции (мс).

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 (Стернберг, 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).

Эксперимент с гипотезой «относительно-абсолютного» отношения: реакция выбора

Два экспериментатора, У. Е. Хик в Англии и Рэй Хаймен в США, провели почти в одно и то же время сходные эксперименты в русле так называемого *информационного подхода* (Хик, 1952; Хаймен, 1953). Для позитивного набора стимулов они нашли иную закономерность между числом альтернатив и временем реакции, чем Стернберг.

Методика 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 либо 8, либо 6, либо 5, либо 4, либо 3, либо 2, либо даже 1 из 10 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 среднее 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 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

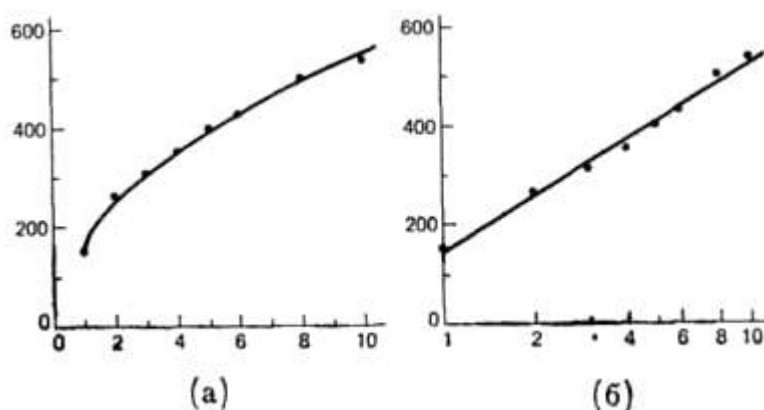


Рис. 7.5. Зависимость времени реакции от количества альтернатив (эксперимент Хика, 1952). Ось абсцисс — количество альтернатив. Ось ординат — время реакции (мс)

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 граммов как половину веса в 400 граммов? Этот вопрос задал себе С. 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 Стивенсом (1948) очень просто. Испытуемый стоял возле стола, на котором лежали 7 одинаковых закрытых контейнеров. Один из них откладывали в сторону и называли «эталоном». Задача состояла в том, чтобы, взвешивая в руке эталон и каждый из 16 других весов, выбрать тот, который воспринимался как половина веса эталона» (с. 344). Заметьте: испытуемые должны были найти не тот вес, который действительно составлял половину веса эталона, а только ощущался таковым.

Было найдено, чего вес, который воспринимается как в 2 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 «*veg*» Харпер и Стивене образовали от старого норвежского слова *veg* — поднимать (с. 345).

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 прошкалированы

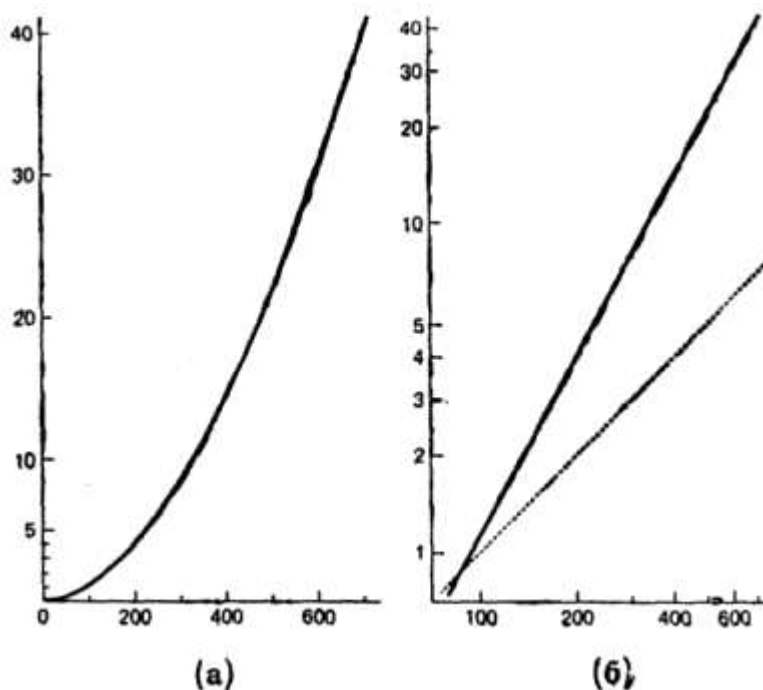


Рис. 7.6. Изменения в ощущении тяжести с увеличением поднимаемого груза (эксперимент Харпера и Стивенса, 1948). Ось абсцисс -- вес груза (г). Ось ординат — ощущение тяжести (в вегах)

в относительных единицах (например расстояние между 100 и 200 г такое же, как между 400 и 800 г, 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 единицу (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 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 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 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 внутрииндивидуальных 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

Калфи и Андерсон (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.

Схемы 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 (Готтсданкер и Уэй, 1966) варьировался в случайном порядке интервал между двумя сигналами, на которые 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, кросс-индивидуальным 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 надежности, а не для контроля систематического смещения. Если применяется кросс-индивидуальный 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 о кросс-индивидуальное 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 используется 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 Готтсданкера и 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 296 последовательности уровней, вместо двух. Такая схема с полным уравниванием для трехуровневого эксперимента выглядит следующим образом:

Группы испытуемых	Последовательности
-------------------	--------------------

1	АБВ
2	АВБ
3	БАВ
4	БВА
5	ВАБ
6	ВБА

Так, если бы в исследовании Готтсданкера и 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, 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 Готтсданкера и Уэй 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 кросс-индивидуальное 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, необходимом 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 мало вероятно, 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:

Группы испытуемых	Последовательности
1	АБВГДЕ
2	ВДГАЕБ
3	ДВАЕБГ
4	БГЕВАД
5	ГЕБДВА
6	ЕАДБГВ

Поскольку в латинском квадрате каждый уровень оказывается в каждой позиции последовательности, естественно, требуется столько групп испытуемых, сколько уровней независимой переменной. Если бы Готтсданкер и 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).

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 Уагенаара (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 кросс-индивидуальной — 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. Таково, например, влияние предшествующего опыта А на Б, но не наоборот. Эта идея может быть распространена на многоуровневые эксперименты с использованием кросс-индивидуального 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 (1963). Они провели два эксперимента, каждый по схеме латинского квадрата с двумя группами испытуемых. Задачей была токарная обработка деталей,

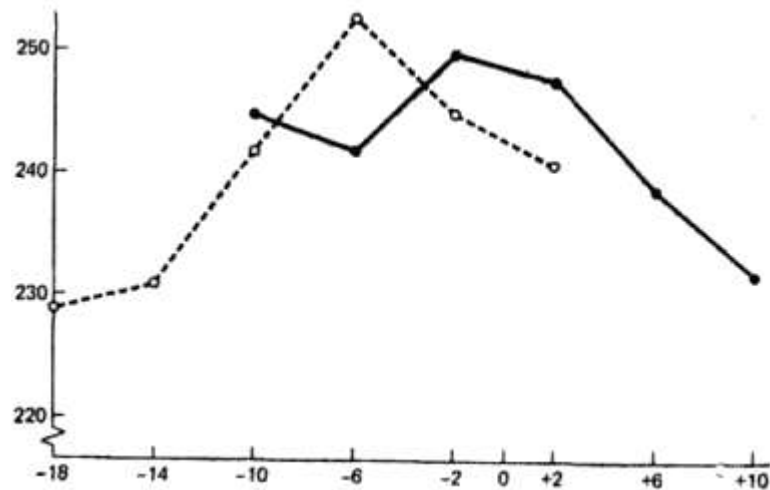


Рис. 7.7. Отношение между высотой рабочей поверхности и количеством обработанных деталей (Кеннеди и Ландесман, 1963). Ось абсцисс — высота рабочей поверхности (в дюймах, ниже (—) или выше (+) локтя). Ось ординат — среднее количество обработанных деталей. Пунктирная линия — условие А, сплошная — условие Б

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. 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 я 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 (Поултон, 1973). Оказывается, важно, какие уровни в основном предшествуют: более низкие, более высокие или смешанные.

Как быть?

При использовании кросс-индивидуального уравнивания прежде всего стоит избегать реверсивного уравнивания. Поскольку полное уравнивание, как правило, оказывается 301непрактичным, 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 следующем параграфе, в многоуровневых экспериментах кросс-индивидуальное 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 этого 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 Стернберга (1969), который проверял гипотезу «абсолютно-абсолютного» отношения между объемом позитивного набора и временем, мнемонического поиска, а также экспериментам Хича (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 полученная 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

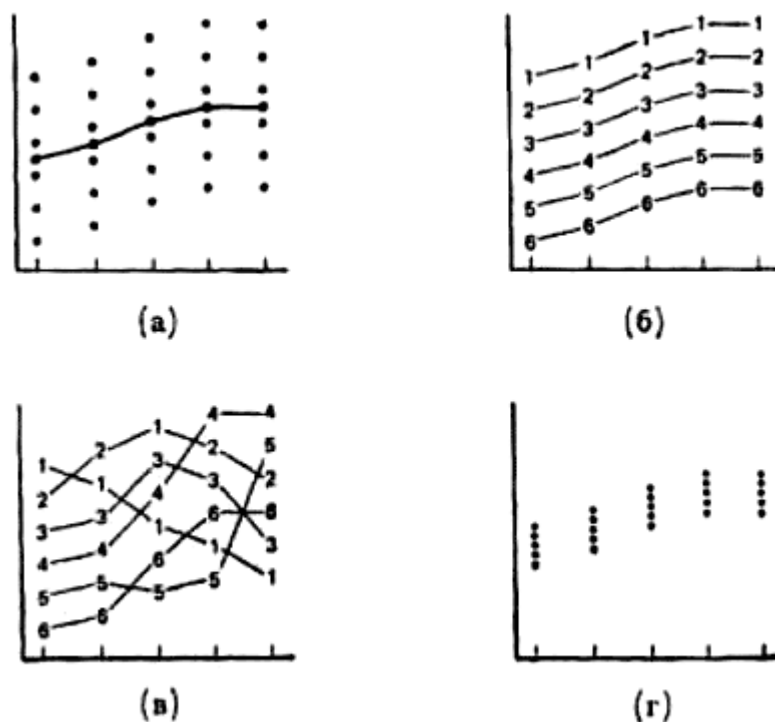


Рис. 7.8. Возможные соотношения усредненной кривой с данными идеального многоуровневого эксперимента, в котором каждому испытуемому одновременно предъявляются все уровни независимой переменной: (а) индивидуальные данные и усредненная кривая; (б) кривые по каждому испытуемому аналогичны усредненной кривой; (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.

при различных уровнях независимой переменной. Линии, соединяющие ответы «одного испытуемого», по форме очень похожи на линию, соединяющую средние в (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.

Существует два способа уменьшения этой трудности при использовании межгрупповой схемы: подбор сходных испытуемых и использование однородных групп. Если испытуемых провести через предварительные испытания, подобрать испытуемых по одинаковым уровням показанных результатов и затем предъявить так уравненным испытуемым различные уровни экспериментальной переменной, то вымышленные данные в виде наборов одинаковых цифр на рис. 7.8 (б) или (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 (б), или довольно хаотическая, как на (в).

Второй способ основан на использовании одной, но очень однородной группы испытуемых, 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 беспокоить возможное пристрастие экспериментатора при отборе в испытуемые одних индивидов и отвержении других. Однако 304 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 нескольким испытуемым, то можно будет получить более ясную картину. Хотя кривую для каждой определенной последовательности нельзя будет «очистить» от зашумляющих влияний последовательности, эти влияния будут одинаковыми для всех испытуемых, которым будет предъявлена эта последовательность. Если, говоря в общем, 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 вкрались 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 весе. 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 «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 о 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. Торгерсон, 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 смещение, сколько неполнота независимой переменной. Было показано, что, во-первых, групповая кривая может не представлять индивидуальные 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

Было рассмотрено три возможных двухуровневых эксперимента, которые оказались совершенно неадекватными по сравнению с аналогичным реально проверенным 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 максимальной (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 различием 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 переменной.

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 кросс-индивидуального 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 необходимо использовать перекрывающиеся диапазоны уровней независимой переменной.

Независимо от экспериментальной схемы 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, затем — применения слишком малого числа уровней независимой переменной, наконец, проведения неверных измерений.

Вопросы

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 Хика (1922) о времени реакции и Харпера и Стивенса (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

t -критерий нельзя 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 единственной независимой переменной, мы называем анализ *однофакторным*. В статистическом приложении к следующей главе, где будут рассматриваться эксперименты с двумя независимыми переменными, будет описана техника двухфакторного дисперсионного анализа.

Две оценки σ^2_x

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 децибел (дБ), 30 дБ, 50 дБ 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 M_1 , M_2 , M_3 и 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 — σ^2_x по данным нашего эксперимента, снова допуская нуль-гипотезу $M_1 = M_2 = M_3 = M_4$. 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 σ^2_x и *межгрупповая* оценка σ^2_x .

Выборочное распределение F -критерия

Если верна нуль-гипотеза, то при *достаточно длинной выборке* оценки σ^2_x 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 те 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:

$$\frac{\text{Межгрупповая оценка } \bar{\sigma}_X^2}{\text{Внутригрупповая оценка } \bar{\sigma}_X^2} = F$$

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

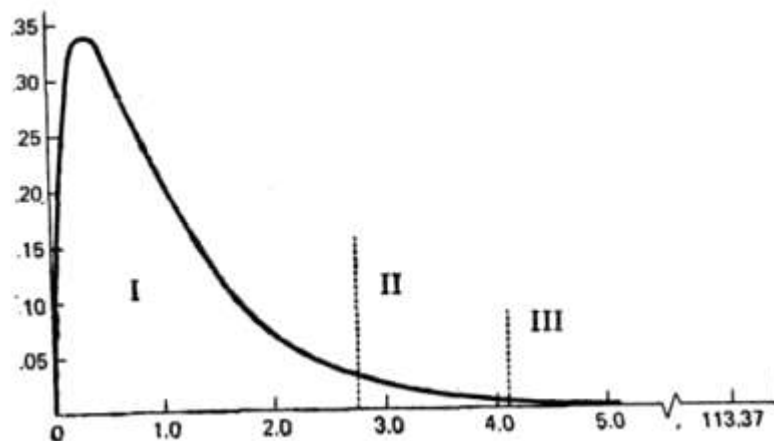


Рис. 7.9. Ось абсцисс — F -отношение. Ось ординат — относительная частота. I — область принятия нуль-гипотезы; II — область отвержения с $p=0,05$; III — область отвержения с $p=0,01$

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 σ^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

$$СК_{ВГ} = 4673 + 5391 + 5808 + 4306 = 20\,178.$$

Сумма квадратов между группами. Межгрупповая сумма квадратов будет использована при определении оценки σ^2 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,

$$d_1 = M_1 - M_{\text{общ}}, \quad d_2 = M_2 - M_{\text{общ}} \dots \quad (7.3)$$

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:

$$СК_{МГ} = n(d_1^2 + d_2^2 + d_3^2 + d_4^2). \quad (7.4)$$

Для числовых данных:

$$СК_{МГ} = 17(2450,25 + 1190,25 + 930,25 + 2862,25) = 17(7433) = 126361.$$

Внутригрупповое среднее квадратичное (СКВ_{ВГ}).

Оценка σ^2x , 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,

$$df_{ВГ} = N - k. \quad (7.5)$$

Для нашего эксперимента

$$df_{ВГ} = 68 - 4 = 64.$$

Как уже говорилось,

$$СКВ_{ВГ} = \frac{СК_{ВГ}}{N - k}. \quad (7.6)$$

Для наших данных

$$СКВ_{ВГ} = \frac{20178}{64} = 315.$$

Межгрупповое среднее квадратичное. Оценка σ^2x , 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:

$$df_{МГ} = k - 1 \quad (7.7)$$

А для числовых данных

$$df_{МГ} = 4 - 1 = 3.$$

Как уже говорилось,

$$СКВ_{МГ} = \frac{СК_{МГ}}{df_{МГ}}. \quad (7.8)$$

Или:

$$СКВ_M = \frac{126361}{3} = 42120$$

F-отношение. Последний шаг в вычислении F-деление межгруппового среднего квадратичного на внутри-групповое среднее квадратичное. Вспомните, что чем больше это отношение, тем более вероятно, что нуль-гипотеза может быть отвергнута:

$$F = \frac{СКВ_M}{СКВ_{вг}} \quad (7.9)$$

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.

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	df	СКВ	F	p
Между уровнями	590,8	5	118,16	12,64	<0,01
Внутри уровней	224,4	24	9,35		
Общая	815	29			

Статистическая таблица 3

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

	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,54
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74
	16,26	13,27	12,06	11,39	10,97	10,67	10,45	10,27	10,15	10,05
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06
	13,74	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63
	12,25	9,55	8,45	7,85	7,46	7,19	7,00	6,84	6,71	6,62
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34
	11,26	8,65	7,59	7,01	6,63	6,37	6,19	6,03	5,91	5,82
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13
	10,56	8,02	6,99	6,42	6,06	5,80	5,62	5,47	5,35	5,26
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97
	10,04	7,56	6,55	5,99	5,64	5,39	5,21	5,06	4,95	4,85
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86
	9,65	7,20	6,22	5,67	5,32	5,07	4,88	4,74	4,63	4,54
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76
	9,33	6,93	5,95	5,41	5,06	4,82	4,65	4,50	4,39	4,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67
	9,97	6,70	5,74	5,20	4,86	4,62	4,44	4,30	4,19	4,10
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60
	8,86	6,51	5,56	5,03	4,69	4,46	4,28	4,14	4,03	3,94
15	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55
	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80
16	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
	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41
	8,28	6,01	5,09	4,58	4,25	4,01	3,85	3,71	3,60	3,51
19	4,38	3,52	3,13	2,90	2,74	2,63	2,55	2,48	2,43	2,38
	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43
20	4,35	3,49	3,10	2,87	2,71	2,60	2,52	2,45	2,40	2,35
	8,10	5,85	4,94	4,43	4,10	3,87	3,71	3,56	3,45	3,37
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32
	8,02	5,78	4,87	4,37	4,04	3,81	3,65	3,51	3,40	3,31
22	4,30	3,44	3,05	2,82	2,66	2,55	2,47	2,40	2,35	2,30
	7,94	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26
23	4,28	3,42	3,03	2,80	2,64	2,53	2,45	2,38	2,32	2,28
	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21
24	4,26	3,40	3,01	2,78	2,62	2,51	2,43	2,36	2,30	2,26
	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,25	3,17
25	4,24	3,38	2,99	2,76	2,60	2,49	2,41	2,34	2,28	2,24
	7,77	5,57	4,68	4,18	3,86	3,63	3,46	3,32	3,21	3,13
26	4,22	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22
	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,17	3,09