

Глава 4. ЭКСПЕРИМЕНТЫ НА ПРЕДСТАВИТЕЛЬНЫХ ВЫБОРКАХ

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

Вы скажете, что здесь должны быть какие-то правила. Правила, конечно, есть. Они сформулированы в “Акте об упаковке и обозначении товаров”. Но, как видите, эти правила не очень помогают покупателям. Двое психологов (Гейтвуд и Перлофф, 1973) решили провести эксперимент и найти способ, который действительно помогал бы при выборе покупок. По поводу упомянутого “Акта” они цитируют Кохана (1969, с. 10):

“Правила расфасовки и определения стоимости продуктов ... это один из ярких примеров полного отсутствия всяких правил”. В эксперименте, который мы обсудим в этой главе, Гейтвуд и Перлофф сравнивали существующий способ информирования покупателей о ценах на продукты с двумя другими, усовершенствованными. Один из них был предложен продавцами бакалейных товаров, а другой—самими покупателями.

В отличие от предыдущих экспериментов, рассмотренных в нашей книге, данный эксперимент проводился не на одном или нескольких испытуемых и не с целью найти наиболее удобный способ подачи информации именно для них. В эксперименте приняли участие 75 испытуемых, а его целью было распространить полученные результаты на всех посетителей универсамов в Соединенных Штатах. Любой тип индивидуального эксперимента для этой цели явно не пригоден. И конечно, возможность получить результаты, применимые не только к одному человеку, но и к целой популяции, — большое преимущество нового эксперимента. Но 75 человек — это далеко не все покупатели. Достаточно ли успешно представляет популяцию выборка испытуемых, чтобы считать эксперимент обладающим внешней валидностью? Мы постараемся ответить на этот вопрос.

Но для начала укажем, что новый тип эксперимента поднимает некоторые новые вопросы в отношении внутренней валидности. В упомянутом эксперименте каждым из трех способов получения информации о ценах на продукты пользовалась отдельная группа испытуемых. С одной стороны, при такой схеме эксперимента снимаются влияния последовательности проб, которые так досаждали нам в главе 2. Но с другой стороны, на каком основании мы будем сравнивать результаты этих трех групп испытуемых? Ведь люди разные, и группы людей—тоже.

Вы увидите, что существует немало способов проведения групповых экспериментов. Кроме отказа от совершенно неудачных схем, в этой главе не будет каких-либо универсальных рекомендаций. Наше обсуждение будет строиться так, чтобы в результате

вы смогли сами принимать решение в каждом конкретном случае. Таким образом, когда вы прочтете эту главу, вы научитесь выбирать наиболее подходящую схему эксперимента, зная о тех трудностях, которые можно преодолеть, и о тех, которые, все же остаются. Знакомясь с экспериментальными работами, вы сумеете оценить, насколько успешно справлялись с этими трудностями другие исследователи.

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

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

Эксперимент 1: КАК СООБЩАТЬ ПОКУПАТЕЛЯМ О ЦЕНАХ НА ПРОДУКТЫ!

Было предложено два новых способа, помогающих выбирать покупки из упаковок разного размера. Один из них—воспользоваться круглой счетной линейкой, показанной на рис. 4.1. Покупатель поворачивает внутреннюю шкалу этого устройства до тех пор, пока вес пакета не соединится с его ценой, обозначенной на внешней шкале. Тогда в первом окошке будет показана цена продукта в центах за унцию, а во втором — количество унций, которое можно купить на доллар. Как вы видите (см. рис. 4.1), если пакет с 9 унциями маисовых лепешек стоит 49 центов, то их цена за унцию—меньше $5 \frac{1}{2}$ цента. Этот пакет и будет самой дешевой покупкой. Другой предложенный способ — просто поставить на пакете цену продукта за единицу веса. То есть на пакетах в 13, 10 и 9 унций вместе с их общей ценой стоял бы штамп: 5,85, 5,90 и 5,44 цента за унцию. Продавцы согласны, что столь очевидное решение довольно удачно, но отмечают, что для них оно может обойтись слишком дорого. Гейтвуд и Перлофф (1973, с. 81) пишут:

“Большинство продавцов были против последнего предложения”. “Управляющие и поставщики универсамов недовольны,—читаем мы в статье “Заботы продавца” (“Век рекламы”, 1969, с. 3),—ведь, с финансовой стороны подобное нововведение очень невыгодно: оно требует увеличения в два раза рабочей силы, повышения цен, и это может провалить все дело”. Они утверждают, что использование дешевого (а возможно,

полезного и для покупателей) вычислительного устройства даст более удачные результаты.

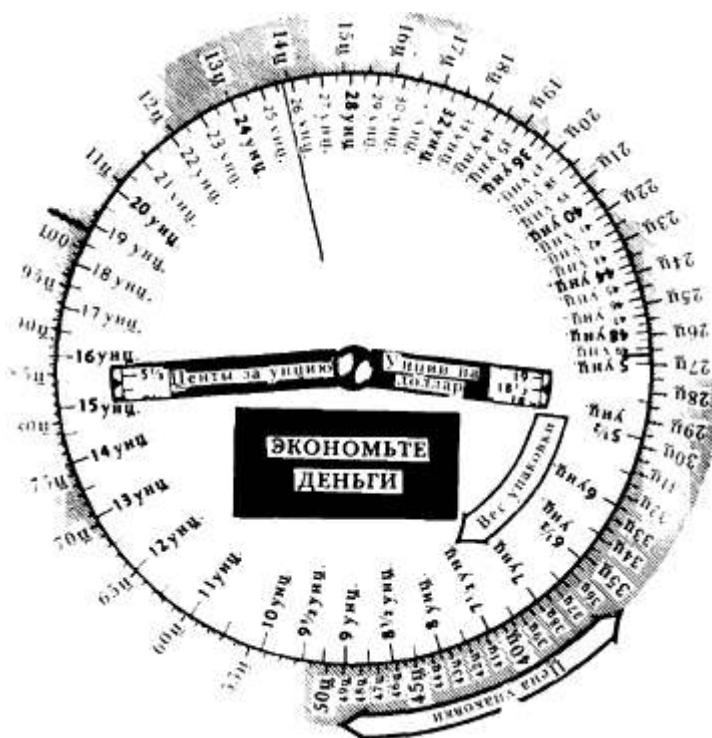


Рис. 4.1. Вычислительное устройство для определения стоимости одной унции маисовых лепешек в разных упаковках. Внутренняя шкала повернута так, что отметка “9 унций” (вес пакета) совмещена с отметкой “49 центов” (цена пакета) на внешней шкале. В левом окошке показана цена за единицу веса — 5,5 цента за унцию

Экспериментальная гипотеза

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

Методика

Испытуемые. “В исследовании приняли участие 75 испытуемых—добровольцев, 64 из них—женщины. 60 из 75 закончили одногодичный курс в колледже. Возраст испытуемых: от 20 до 29 лет—48 человек, от 30 до 39— 17, от 40 и старше—10” (Гейтвуд и Перлофф, 1973, с. 82).

Испытуемых случайным образом распределили по трем группам — по группе на каждое из трех экспериментальных условий. Те, кому нужно было пользоваться вычислительным устройством, “обучались работе с ним и допускались к эксперименту после трех успешных определений цены продукта за единицу веса” (с. 82).

Задача. В опытах использовались наборы девяти разных пищевых продуктов. Пакеты каждого набора продуктов были пронумерованы. Испытуемого просили выбрать из каждого набора “самый дешевый пакет” (с. 82). Обстановка эксперимента напоминала универсам. Цены на продукты фактически повторяли цены одного из магазинов данной фирмы. Испытуемому давали листок с вопросами, на котором он должен был написать номер выбранного им пакета для каждого”

Результаты

Каждый из девяти выборов, произведенных испытуемым, характеризовали по двум параметрам. Это—правильность выбора и затраченное время (в минутах). В левой части табл. 4.1 показаны среднее количество правильных выборов и стандартные отклонения для трех групп испытуемых. Видно, что по средним оценкам лучшим оказалось условие В—около 8 правильных выборов (а условия А и Б—несколько ниже б). При условии В наименьшим является и значение стандартного отклонения, поскольку большинство испытуемых верно выполнили все 9 выборов, а выше этой точки оценки располагаться не могли.

Таблица 4.1

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

(А)—сообщение об общей цене и весе упаковки;

(Б) — то же с использованием вычислительного устройства;

(В) — о цене за унцию.

	Условия			Затраченное время		
	А	Б	В	А	Б	В
Среднее	5,78	5,96	8,04	23,93	31,72	3,60
Стандартное отклонение	1,31	0,45	1,57	10,00	9,57	1.11

Еще большим оказалось различие во времени, затраченном на каждый из 9 выборов. Это видно по правой части таблицы. В условии В для выбора требовалось в среднем 4 минуты, а в условии А—почти 24. Понятно, что работа с вычислительным устройством требовала дополнительного времени, и среднее время для условия Б составило почти 32 минуты. Таким образом, для выбора покупки с помощью вычислительного устройства требуется в 8 раз больше времени, чем в том случае, когда цена продукта за унцию проставлена на пакете.

Выводы

Экспериментальная гипотеза подтвердилась. Непосредственное сообщение о цене за единицу веса позволяло выбирать покупки более точно и быстро, чем два другие способа. Использование вычислительного устройства для более точного выбора оказывало сомнительную помощь, кроме того, его применение увеличивало время покупки.

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

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

Если в эксперименте привлекается большое число испытуемых, то полученные результаты можно переносить на более широкую популяцию. В предыдущих экспериментах приложение результатов ограничивалось одним человеком — испытуемым и касалось его деятельности в дальнейшем. Но некоторые исследования просто неинтересны, если они не позволяют сделать достаточно широких обобщений. Для того чтобы после проведения эксперимента обозначение цены продукта за единицу веса было введено повсеместно, его результаты должны быть справедливы для всех покупателей. Кроме того, преимущество эксперимента с большим числом испытуемых — в его экономичности. Можно, конечно, тщательно исследовать каждого испытуемого и провести много индивидуальных экспериментов. Но если вы хотите показать справедливость полученных результатов (скажем того, что спасательный поиск на море лучше осуществлять без бинокля) сразу для целой выборки испытуемых, то каждому из них вовсе не обязательно участвовать во всех экспериментальных условиях (как было с Дион и Хоки). Хотя, разумеется, кто-то всегда будет продолжать настаивать на проведении индивидуальных экспериментов, поскольку Дион и Хоки в чем-то различны. Но у эксперимента с большим числом испытуемых есть и еще одно, третье, преимущество. Благодаря количеству собранных данных и возможности применять более эффективные экспериментальные схемы, которые будут описаны ниже, исследования нового типа становятся более качественными, чем эксперименты с отдельными испытуемыми. Короче говоря, возможность распространения результатов эксперимента на популяцию дает ему тройное преимущество. Во-первых, он и ставится для изучения популяции в целом. Во-вторых, групповой эксперимент более экономичен, и, в-третьих, это более удачный эксперимент по сравнению с группой индивидуальных экспериментов.

Эксперимент 2: МЫСЛЕННАЯ ОТРАБОТКА ДВИГАТЕЛЬНЫХ НАВЫКОВ

Мы вновь поднимаемся в воздух! Вспомните тренажер, о котором мы рассказывали в предыдущей главе, в эксперименте с ночными посадками самолета. Такие тренажеры очень дороги. Экспериментальный психолог Дирк Пратер (1973) из Академии воздушных сил попытался найти более дешевый способ для обучения начинающих пилотов. Он имел дело с самолетом Т-37. А научить курсантов сажать такой самолет без тренажера, да еще за короткий срок практически невозможно — не хватит ни самолетов, ни инструкторов. И вот при изучении литературы по формированию навыков у Пратера возникла идея о самом дешевом тренажере—тренировке мысленной. На самом деле, большинство предшествующих исследователей отмечали высокую эффективность мысленной тренировки для совершенствования таких двигательных навыков, как, например, игра в волейбол или точное выполнение штрафных бросков в баскетболе.

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

Экспериментальная гипотеза

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

Методика

Испытуемые. “Испытуемые были случайно отобраны среди курсантов, проходящих обучение по тренировочной программе “Т-37” на воздушной базе в Уильямсе. В экспериментальную группу вошли 13 человек. Еще 10 человек отобрали для контрольной группы (также случайным образом). Все курсанты были не очень опытные: 20 часов тренировок на самолете Т-41 и 4 часа— на Т-37” (Пратер, 1973, с. 353).

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

Условием А для испытуемых экспериментальной группы была дополнительная программа специальных мысленных тренировок. Четыре раза за период обучения приземлению испытуемые помещались в макете кабины самолета Т-37. Макет состоял из подвижных рычагов управления и контрольной рукоятки. Через наушники испытуемым подавалась специальная магнитная запись. На ней рассказывалось о выполнении всех необходимых операций, сначала очень подробно, позднее—в самых общих чертах, т. е. “подсказки” постепенно убирались. “Испытуемых экспериментальной группы просили как можно более отчетливо представлять себе ситуацию посадки и выполнять те же действия, что и в реальном полете, вплоть до движения глаз” (Пратер, 1973, с. 354).

Процедура. Испытуемым экспериментальной группы (условие А) предлагали осуществлять мысленную отработку после четвертого, пятого, шестого и седьмого взлетов. Каждая такая серия продолжалась примерно 11—15 минут. Испытуемые контрольной группы (условие Б) выполняли в течение этого периода времени обычные тренировки (Пратер, 1973, с. 354).

Результаты

Самая высокая оценка работы испытуемых равнялась 7 (лучший курсант), самая низкая— 1 (неудачник). Оценки выставлялись отдельно за знание операций (насколько хорошо понимал курсант, что ему полагается делать) и за технику их выполнения (насколько успешно операции осуществлялись). При условии А (экспериментальная группа) средняя оценка за знание операций составила 4,53, а при условии Б (контрольная группа)—4,26. Оценки за технику выполнения были соответственно 4,21 и 3,89.

Выводы

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

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

Возможность проведения новых экспериментов

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

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

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

Устранение некоторых факторов, мешающих достижению внутренней валидности

Предположим, что Гейтвуд и Перлофф применили схему эксперимента с одним испытуемым, используя позиционно уравненную последовательность АБВВБА. Тогда, во-

первых, возникли бы эффекты последовательности, которые, как мы знаем по главе 2, зачастую просто неустранимы. Вполне возможно, что испытуемый, осознав все выгоды условия В — непосредственного сообщения о цене за единицу веса, просто не захочет пользоваться вычислительным устройством в условии Б. Во-вторых, для каждого из трех исследуемых способов понадобятся разные задачи. Один и тот же список продуктов использовать нельзя. Кроме того, нужно было бы как-то уравнивать отдельные пункты в списках, применяемых для разных экспериментальных условий.

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

Эксперимент 3: ДВА МЕТОДА ОБУЧЕНИЯ ИСПАНСКОМУ ЯЗЫКУ

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

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

Планирование эксперимента

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

Для начала экспериментаторы изложили в своем лабораторном дневнике каждый из методов обучения. Это было несложно. Для описания старого метода они просто

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

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

Разумеется, эксперимент проводился со студентами, изучавшими испанский впервые. Сначала возникла идея распределить их по группам в случайном порядке: по каждому из преподавателей и затем по каждому методу обучения, которые будут называться теперь письменным и разговорным. Но тут есть одна сложность: некоторые студенты уже немного знали язык—они жили в районах, где говорят по-испански. Вполне возможно, что при случайном распределении в какие-то из четырех групп попадет слишком много таких студентов, а в другие—очень мало. Поэтому экспериментаторы решили провести предварительную проверку.

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

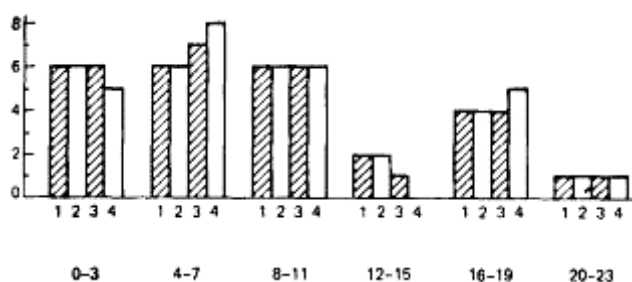


Рис. 4.2. Диаграммы результатов распределения студентов по 4-м группам согласно

оценкам, полученным в предварительной проверке. Ось абсцисс: первая — номера групп; вторая — оценки студентов в предварительной проверке. Ось ординат — количество студентов

Таблица 4.2

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

Оценка	Группа	Фамилия	Оценка	Группа	Фамилия	Оценка	Группа	Фамилия	Оценка	Группа	Фамилия
23	1	Мендоса	8	4	Грэхам	12	2	Ламб	6	4	Уи
22	2	Лира	8	4	Хавкинс	11	3	Сохен	5	3	Кронхайт
22	3	Сарагоса	8	3	Келли	11	4	Ховертон	5	2	Форрестер
26	4	Оливера	8	2	О Брайен	11	4	Смит	5	1	Фуимото
19	4	Диас	8	1	Миллер	10	3	Самри	5	1	Портер
19	3	Руис	8	1	Шапиро	10	2	Джонс	5	2	Янг
18	2	Авила	8	2	Уотсон	10	1	Ким	4	3	Гарнер
18	1	Бесерра	7	3	Александр	10	1	Валентин	4	4	Мецгер
18	1	Крус	7	4	Д Амбуаз	10	2	Вондерски	4	4	Зилиотто
18	2	Гутьерес	7	4	Фукимото	9	3	Брэг	3	3	Андерсон
18	3	Миллер	7	3	Хаскел	9	4	Дуглас	3	2	Джонсон
18	4	Нунес	7	2	Джонс	9	4	Гилберт	3	1	Петрини
18	4	Зандовал	7	1	Леви	9	3	Келлер	3	1	Вудс
17	3	Аргуэллес	7	1	Мазетти	9	2	Келли	2	2	Девис
17	2	Крус	7	2	Оизен	9	1	Левин	2	3	Кац
17	1	Гузман	7	3	Смит Б.	9	1	МсДональд	2	4	Джонсон
17	1	Жорес	7	4	Смит Г.	8	2	Барнес	2	4	Мартин
17	2	Робинсон	7	4	Циммерман	8	3	Кровел	2	3	Вохан
16	3	Якобе	6	3	Христенсен	1	2	Купер	0	3	Карпенгер
16	4	Морено	6	2	Даниелс	1	1	Айленд	0	2	Дейч
16	4	Санчер	6	1	Фишер	1	1	Мак-Килош	0	1	Кеннеди
15	3	Инграм	6	1	Грехем	1	2	Мейер	0	1	Кирк
15	2	Торрес	6	2	Джонс	1	3	Шварц	0	2	Пинеро
14	1	Рос	6	3	Макинтош	1	4	Ван Дейк	0	3	Рейес
13	1	Хирасима	6	4	Сандерсон	0	4	Бакер	0	4	Вильямс

Проведение эксперимента

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

После окончания занятий был проведен итоговый экзамен. Материал для него заготовили заранее. На магнитофон записали вопросы по-испански. Чтобы исключить влияние дополнительных факторов (например, знакомый или незнакомый голос), вопросы студентам задавал тот преподаватель, у которого они занимались. Экзамен проходил в лингафонном кабинете. Каждый студент отвечал на 100 вопросов в два приема. После каждого вопроса студенту давалось 30 секунд, чтобы записать ответ на листочке бумаги, где стоял номер) вопроса и кодовый номер испытуемого. Каждый преподаватель проверял работы студентов другого преподавателя и поэтому узнавание почерка, а возможно, и других индивидуальных черт также было исключено. Работы студентов, занимавшихся по разным методам, складывались в одну пачку и перемешивались. Ответы на каждый вопрос оценивались по шкале от 5 до 0.. О содержании каждой оценки преподаватели договорились заранее: “5”—безупречно; “4”—одна незначительная ошибка; “3”—одна

грубая ошибка или две незначительных; “2” — несколько грубых ошибок; “1” — в целом перевод неверен, но есть некоторые совпадения; “0” — неверно вообще. Наибольшее количество баллов, которые мог набрать студент, ответив на 100 вопросов, — 500. Ради удобства общая оценка делилась на 5 и округлялась до ближайшего целого числа. Таким образом, наивысший возможный общий балл — 100. Это — критерий оценки работы студента.

Анализ результатов

Было установлено, что у студентов м-с Кроутерс и м-ра Рамоса, обучавшихся по одному и тому же методу, оценки в среднем примерно одинаковы. Это дало возможность представить результаты всего эксперимента в двух частотных распределениях (рис. 4.3): одно — для разговорного метода, другое — для письменного. Фактически здесь просто показано, сколько студентов получили оценки от 52 до 99. Для более ясной картины оценки сгруппированы в интервалы по четыре. Так, во второй по возрастанию интервал включены студенты, получившие оценки 56, 57, 58 и 59. За ним идут студенты с оценками 60, 61, 62 и 63; в этот интервал попали 3 студента, обучавшихся по разговорному методу, и только один — обучавшийся по письменному. Можно убедиться, что у разговорного метода есть небольшое преимущество: его частотное распределение несколько выше — смещено вправо — по сравнению с письменным.

Правда, средняя оценка для разговорного метода составила 78 баллов, в то время как для письменного — 75. Это различие не очень показательное, и, уж конечно, оно вряд ли убедит школьный совет тратить деньги на нововведения. Но Кроутерс и Рамос по-прежнему считают, что в новом методе обучения что-то есть. Быть может, им не хватало опыта для его правильного применения. Но вот однажды в школьном коридоре их приветствовали словами: “Благодарим вас за наши успехи в испанском, новый метод нам очень понравился. Грация!”

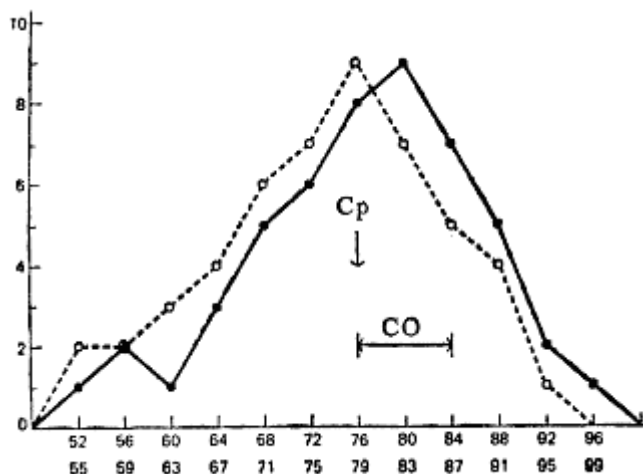


Рис. 4.3. Результаты эксперимента с методами обучения испанскому языку: частотное распределение (показаны средняя оценка (Ср) и величина стандартного отклонения (СО) для разговорного метода). Ось абсцисс — экзаменационные оценки. Ось ординат — количество студентов

ДОСТИЖЕНИЕ ВНУТРЕННЕЙ ВАЛИДНОСТИ В ЭКСПЕРИМЕНТАХ С МЕЖГРУППОВЫМ СРАВНЕНИЕМ

Новая угроза для внутренней валидности: индивидуальные различия испытуемых

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

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

Известные характеристики

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

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

Несистематические вариации

Дайте любой группе людей какой-нибудь тест — и вы получите определенное распределение ответов. Чаще всего оно будет иметь куполообразную форму. Так распределились оценки студентов, прошедших предварительную проверку на знание слов испанского языка. То же было справедливо и для покупателей, пользовавшихся вычислительным устройством. Такую же форму имело распределение времени научения, необходимого для правильного выполнения задач на различение в опытах с крысами. В

большинстве случаев нельзя заранее выделить те индивидуальные различия, которые объясняли бы подобный разброс. Эти различия непостоянны и изменчивы.

Какую-то часть несистематических вариаций можно отнести за счет изменчивости поведения каждого испытуемого во времени или нестабильности в оценке этого поведения, о чем мы говорили в главе 2. Во-первых, в любой конкретной пробе одни испытуемые будут более работоспособны, другие—менее и т. д. Во-вторых, в каждый конкретный замер обычно попадает лишь какая-то часть изучаемого поведения, и поэтому данная оценка работы одних испытуемых может оказаться удачной, а других—неудачной. В отличие от индивидуальных экспериментов, в экспериментах с межгрупповым сравнением подобная изменчивость сглаживается за счет увеличения количества отдельных действий, оцениваемых в одном замере. В эксперименте с испанским это достигалось двумя путями. Во-первых, каждый студент участвовал в опытах на протяжении двух лет (при одном из условий независимой переменной). Таким образом, обучение происходило как в благоприятные, так и в неблагоприятные периоды. Во-вторых, они проводились через длительный тест. Возможно, что для еще большей минимизации факторов во времени было бы лучше повторно проводить тесты по 2—3 в месяц.

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

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

Еще раз о систематическом смещении и ненадежности

Итак, основным препятствием для достижения внутренней валидности в эксперименте с межгрупповым сравнением является потенциальная возможность влияния на полученные результаты индивидуальных различий испытуемых в каждой из групп. Из-за этих влияний подлинные воздействия экспериментальных условий могут и вовсе не проявиться. Если группы составлены так, что между ними сохраняется устойчивое различие (например, в одной группе юноши, в другой—девушки), то систематическое смещение независимой переменной с индивидуальными особенностями испытуемых неизбежно. А случайные различия между испытуемыми в группе, обусловленные ограниченным размером выборки, будут снижать надежность значений тех зависимых переменных, которые исследуются в эксперименте.

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

Для проведения бесконечного эксперимента с межгрупповым сравнением нужно, чтобы обе группы состояли из бесконечного числа испытуемых (причем способ составления групп не должен порождать систематического смещения). Мы уже советовали вам привлекать для эксперимента большое число испытуемых. Как вы помните, в эксперименте с одним испытуемым для достижения высокого уровня надежности с увеличением разброса данных требуется больше проб. А в эксперименте с межгрупповым сравнением соответствующим требованием будет увеличение количества испытуемых.

Кроме того, надежность индивидуального эксперимента можно повысить путем усреднения изменчивости во времени. Поскольку какая-то часть различий между испытуемыми целиком зависит от субъективных особенностей, этот метод применяется и при межгрупповом сравнении. Эти различия уменьшатся также в том случае, если для всех испытуемых будут обеспечены примерно равные условия проведения эксперимента. Если в эксперименте с испанским языком применение того или другого метода обучения к разным студентам как-то различалось, то нужно ожидать подобных различий и в успешности обучения. В индивидуальных экспериментах соблюдение строгости в отношении всех необходимых аспектов—независимой, зависимой и дополнительных переменных — будет увеличивать надежность, а в новом случае ее можно повысить путем сокращения межгрупповых различий, уравнивая условия работы испытуемых. И если добиться этого, то высокая надежность эксперимента может быть достигнута и не на столь большом количестве испытуемых.

ТРИ СТРАТЕГИИ СОСТАВЛЕНИЯ ГРУПП, ТРИ СХЕМЫ СРАВНЕНИЯ

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

Случайная стратегия

В двух экспериментах, описанных в этой главе, применялась случайная стратегия. В эксперименте с сообщением о ценах на продукты участвовали 75 испытуемых-добровольцев. Случайным образом их распределили в соответствии с тремя

экспериментальными условиями, т. е. каждый участник эксперимента мог с одинаковой вероятностью попасть в любую из экспериментальных групп.

Для такого распределения можно использовать таблицу случайных чисел (табл. 4.3). Фамилии 75 испытуемых записывают в алфавитном порядке (чтобы упростить их учет). Затем берут пачку карточек, пишут фамилию первого человека на верхней карточке и т. д. Потом решают, с какой колонки таблицы начать распределять испытуемых по группам. Бросают ту же игральную кость, и если она выпадает тремя очками вверх, значит, нужно начать с 3-й колонки таблицы (колонка 9—12). Так, Ааронсон (первая карточка) помечается верхним числом в колонке 5901. Адаме (следующая карточка) помечается числом 4310, стоящим ниже. Когда эта колонка закончится, переходят к следующей, и так до тех пор, пока всем 75 испытуемым не будут присвоены номера из таблицы случайных чисел.

Таблица 4.3

Таблица случайных чисел

Ряды	Колонки																			
	1-4	5-8	9-12	13-16	17-20	21-24	25-28	29-32	33-36	37-40										
1	23	15	75	48	59	01	83	72	59	93	76	24	97	08	86	95	23	03	67	44
2	05	54	55	50	43	10	53	74	35	08	96	61	18	37	44	10	96	22	13	43
3	14	87	16	03	50	32	40	43	62	23	50	05	10	03	22	11	54	38	08	34
4	38	97	67	49	51	94	05	17	58	53	78	80	59	01	94	32	42	87	16	95
5	97	31	26	17	18	99	75	53	08	70	94	25	12	58	41	54	88	21	05	13
6	11	74	26	93	81	44	33	93	08	72	32	79	73	31	18	22	64	70	68	50
7	43	36	12	88	59	11	01	64	56	23	93	00	90	04	99	43	64	07	40	36
8	93	80	62	04	78	38	26	80	44	91	55	75	11	89	32	58	47	55	25	71
9	49	54	01	31	81	08	42	98	41	87	69	53	82	96	61	77	73	80	95	27
10	36	76	87	26	33	37	94	82	15	69	41	95	96	86	70	45	27	48	38	80
11	07	09	25	23	92	24	62	71	26	07	06	55	84	53	44	67	33	84	53	20
12	43	31	00	10	81	44	86	38	03	07	52	55	51	61	48	89	74	29	46	47
13	61	57	00	63	60	06	17	36	37	75	63	14	89	51	23	35	01	74	69	93
14	31	35	28	37	99	10	77	91	89	41	31	57	97	64	48	62	58	48	69	19
15	57	04	88	65	26	27	79	59	36	82	90	52	95	65	46	35	06	53	22	54
16	09	24	34	42	00	68	72	10	71	37	30	72	97	57	56	09	29	82	76	50
17	97	95	53	50	18	40	89	48	83	29	52	23	08	25	21	22	53	26	15	87
18	93	73	25	95	70	43	78	19	88	85	56	67	16	68	26	95	99	64	45	69
19	72	62	11	12	25	00	92	26	82	64	35	66	65	94	34	71	68	75	18	67
20	61	02	07	44	18	45	37	12	07	94	95	91	73	78	66	99	53	61	93	78
21	97	83	98	54	74	33	05	59	17	18	45	47	35	41	44	22	03	42	30	00
22	89	16	09	71	92	22	23	29	06	37	35	05	54	54	89	88	43	81	63	61
23	25	96	68	82	20	62	87	17	92	65	02	82	35	28	62	84	91	95	48	83
24	81	44	33	17	19	05	04	95	48	06	74	69	00	75	67	65	01	71	65	45
25	11	32	25	49	31	42	36	23	43	86	08	62	49	76	67	42	24	52	32	45

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

В преимуществах случайной стратегии вы сможете убедиться на следующем примере. Первых 25 человек, пожелавших участвовать в эксперименте, исследователи могли бы включить в группу условия А, вторых 25— в условие Б, а третьих—в В. Тогда между тремя группами испытуемых было бы очевидное и устойчивое различие, связанное с очередностью их записи на участие в эксперименте. Таким образом, неудачный способ составления групп привел бы к систематическому смещению независимой переменной — способа сообщения о ценах — с другой переменной — индивидуальными различиями испытуемых. Применение же случайной стратегии позволяет сделать влияние этих различий не столь систематичным. Но если количество испытуемых в каждой группе не будет достаточно большим, то применение случайной стратегии не обеспечит эквивалентности групп. Предположим, что в эксперименте с сообщением о ценах участвовало 15 испытуемых, разделенных на 3 группы.

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

Стратегия подбора пар

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

Ничто не помешало бы исследователям использовать в эксперименте с испанским случайную стратегию. Какая же стратегия лучше? Обе равно хороши для устранения систематического смещения независимой переменной с индивидуальными различиями испытуемых. Ни в том, ни в другом случае индивидуальные характеристики членов каждой группы (занимающихся по одному из методов) не будут отличаться от характеристик другой группы систематически. Стратегия подбора пар позволит достичь большого подобия групп с одним и тем же числом испытуемых, если основание попарного сравнения (здесь — предварительная проверка) действительно связано с изучаемым видом деятельности (здесь — обучение испанскому языку). Но даже если предположения экспериментатора неверны и выделенные им характеристики не соответствуют виду деятельности, изучаемому в эксперименте,—ничего страшного не произойдет. Ведь в отношении всех остальных индивидуальных характеристик группы набирались случайно, и их подобие будет не хуже, чем при простом использовании случайной стратегии. Опасность возникает тогда, когда экспериментатор слишком полагается на подбор пар, привлекая небольшое число испытуемых, а связь сравнительных характеристик с изучаемым видом деятельности является при этом *недостаточной*.

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

Стратегия случайного распределения слоев

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

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

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

ВНЕШНЯЯ ВАЛИДНОСТЬ: ПРЕДСТАВЛЕННОСТЬ ИЗУЧАЕМОЙ ПОПУЛЯЦИИ

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

насколько мы приближаемся к столь недостижимой цели благодаря своим способам отбора испытуемых.

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

Отбор из популяции

Популяций существует очень много. Многие будущие покупатели, летчики и студенты, которые будут изучать испанский, пока еще дети или даже не родились. С другой стороны, более и менее обширные популяции, из которых можно выбирать испытуемых, уже существуют, например все студенты Соединенных Штатов, начавшие изучать курс испанского языка. В одном из трех исследований этой главы — в эксперименте с мысленной тренировкой — использовалась относительно небольшая популяция. Но, конечно, курсантов на воздушной базе в Уильямсе гораздо больше, чем можно было привлечь в качестве испытуемых для эксперимента. Поэтому случайная стратегия применялась в этом случае не для распределения всех имеющихся в распоряжении курсантов, а для *отбора* экспериментальных групп из данной популяции.

Случайный отбор групп

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

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

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

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

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

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

Связь внешней и внутренней валидности при отборе из популяции

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

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

Схема интраиндивидуального сравнения. В эксперименте с сообщениями о ценах можно было бы использовать интраиндивидуальную схему. Каждый испытуемый мог бы пройти все три экспериментальных условия. Но между внешней и внутренней валидностью такого эксперимента не будет абсолютно никакой связи. Если выборка

испытуемых хорошо представляет популяцию, то достигается высокая внешняя валидность. Однако внутренняя валидность индивидуального эксперимента нарушается главным образом за счет влияний предшествующих проб, различия задач и изменчивости поведения во времени. На представительность выборки эти факторы не влияют. Так что даже если тщательно контролировать их, это все равно не позволит улучшить репрезентативность выборки, т. е. внешнюю валидность эксперимента.

Послойный случайный отбор

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

Если основание выделения слоев (здесь — оценка в тесте на пространственное воображение) связано с изучаемым видом деятельности (здесь — обучение сажать самолет Т-37), то данная процедура дает крупное преимущество. При одном и том же количестве испытуемых она обеспечит лучшую представленность популяции, и это повысит внешнюю валидность эксперимента. А поскольку применяется схема эксперимента с межгрупповым сравнением, то внутренняя валидность тоже будет выше.

Привлечение испытуемых, имеющих в наличии

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

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

Обзор схем межгрупповых сравнений

В этой главе мы встретили так много различных схем, что запомнить их все сразу довольно трудно. Вам поможет табл. 4.4. В ней представлена классификация этих схем

согласно типам стратегий построения экспериментальных групп и типам набора испытуемых. Приводится также общая оценка каждой из шести схем: насколько успешно решаются в них проблемы внутренней и внешней валидности. Две ячейки в таблице не заполнены, так как соответствующие комбинации типа стратегий и типа набора испытуемых просто невозможны. Подбор пар и привлечение реальных групп имеют смысл только в случае распределения испытуемых, но не их отбора.

Вы видите, что стратегия подбора пар, а также случайного распределения и отбора с выделением слоев обеспечивает более высокую внутреннюю валидность, чем случайная стратегия. Это объясняется тем, что зачастую при одном и том же числе испытуемых можно сделать группы более эквивалентными. Хуже всего— сравнивать реально существующие группы, в их подобии никогда нельзя быть уверенным из-за множества побочных влияний. При отборе испытуемых из популяции стратегия послойного случайного отбора может

Таблица 4.4

Сводка и оценка межгрупповых схем для представления популяции

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

обеспечить также более высокую внешнюю валидность по сравнению с обычной случайной стратегией. Нередко она позволяет лучше представить популяцию.

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

При оценке *каждой* из приведенных схем мы предполагали наличие достаточно большого числа испытуемых. Если число испытуемых слишком мало, то мала и надежда на хорошую представленность популяций (внешнюю валидность), да и достичь подобия групп (внутренней валидности) становится труднее.

КРАТКОЕ ИЗЛОЖЕНИЕ

В этой главе мы обсудили три новых эксперимента, для каждого из которых, в отличие от ранее описанных, привлекалось большое число испытуемых. Первый эксперимент посвящен способам информирования о ценах на продукты. Экспериментальная гипотеза состояла в том, что прямое сообщение о цене за единицу веса (путем маркировки¹ пакетов) позволит покупателям выбрать более дешевую покупку точнее и быстрее, чем два другие способа: (1) сообщение о весе пакета и его общей цене (текущая практика) и (2) то же плюс использование специального вычислительного устройства. Гипотеза подтвердилась. Первое преимущество эксперимента данного типа по сравнению с индивидуальным экспериментом заключается в том, что его результаты, полученные на выборке испытуемых, распространяются на более обширную популяцию, а не только на поведение отдельного человека.

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

Детально описан третий эксперимент — с двумя методами обучения испанскому языку. Для его проведения было очень важно подобрать две группы испытуемых, сходных по способности изучать испанский. И тогда стало ясно, что эксперимент с межгрупповым сравнением, устраняя некоторые источники нарушения внутренней валидности, порождает еще один. Это — индивидуальные различия испытуемых. Описаны два вида таких различий. Во-первых, это такие очевидные характеристики, как пол, возраст испытуемых, их образование. Сюда же относятся различия, выявляемые с помощью тестов, скажем, на знание слов испанского языка. Кроме того, существуют несистематические различия испытуемых, которые включают как изменчивость поведения каждого человека, так и отличия в выполнении задания разными людьми. Во всех трех экспериментах с межгрупповым сравнением использовались схемы, устраняющие систематическое смещение независимой переменной с индивидуальными различиями испытуемых. Но если группы, работающие в разных экспериментальных условиях, имеют явные различия, скажем, взяты из разных школ, то такое смещение неизбежно. Надежность эксперимента повышается за счет привлечения большого числа испытуемых и сокращения несистематических вариаций. Последнее обеспечивается тем,

что условия работы каждого испытуемого в группе одинаковы, а также путем сокращения разброса данных каждого испытуемого, как описано в главе 2.

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

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

Хотя для эксперимента с мысленной тренировкой было отобрано не очень много испытуемых, выборка успешно представляла изучаемую субпопуляцию, так как индивидуальные различия между курсантами были не слишком велики. Если же исследователь хочет распространить результаты эксперимента за пределы данной популяции, скажем, на курсантов следующих лет, то проблемы внешней валидности станут более серьезными. Возникнет вопрос о соответствии “уровней” субъектных характеристик изучаемой популяции и той популяции, к которой прилагаются полученные результаты.

Связь внешней и внутренней валидности была описана для экспериментов, в которых испытуемые отбираются из популяции. Если способ отбора испытуемых обеспечивает высокую внешнюю валидность (в отношении индивидуальных различий), то тем самым достигается достаточное подобие групп, т. е. повышается внутренняя валидность.

Напротив, группы, хорошо уравненные между собой по индивидуальным характеристикам, могут не быть достаточно представительными для исследуемой популяции. Иначе говоря, высокая внутренняя валидность эксперимента не гарантирует его высокой внешней валидности. При использовании схемы индивидуального эксперимента связь между внешней и внутренней валидностью отсутствует, поскольку та и другая зависят от разных факторов.

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

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

ВОПРОСЫ

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

СТАТИСТИЧЕСКОЕ ПРИЛОЖЕНИЕ: СИЛА СВЯЗИ МЕЖДУ НЕЗАВИСИМОЙ И ЗАВИСИМОЙ ПЕРЕМЕННЫМИ

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

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

2. Время реакции испытуемого на световой сигнал равно 185 мс; время реакции на звуковой сигнал равнялось 162 мс. Разница составляет 23 мс.
3. С умственной тренировкой испытуемые получили среднюю балльную оценку 4,21; «контрольная» группа получила среднюю балльную оценку 3,89. Разница составляет 0,32 деления на шкале балльных оценок.

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

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

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

ИСПОЛЬЗОВАНИЕ ЧАСТОТНЫХ РАСПРЕДЕЛЕНИЙ

Для того чтобы понять, насколько различаются условия, мы можем нанести частотные распределения для этих условий на один и тот же график. Сравнение становится более понятным, если не только отмечать высоту колонок, но соединить их вершины линией. (Это называется *полигоном частот.*) Данный метод уже был показан на рис. 4.3 для эксперимента по обучению испанскому языку.

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

Рассмотрим теперь распределения для эксперимента по измерению времени реакции, описанного в статистическом приложении к главе 1. Они показаны на рис. 4.4. Напомним, что это вымышленные данные. Предположим теперь, что они были получены в *межгрупповом* эксперименте. Тогда каждый из показателей времени реакции представляет среднее для одного из испытуемых, где 17 испытуемым предьявлялось данное условие. Этот пример может быть с тем же успехом представлен в терминах интраиндивидуального эксперимента, как он первоначально излагался. Рассмотрение эксперимента как межгруппового мы делаем только для того, чтобы связать наш анализ с тематикой данной главы. Видно, что в данном случае различия между условиями более отчетливы, чем в эксперименте с испанским языком, т. е. *перекрытие* между распределениями *меньше*. Было бы хорошо иметь количественную меру различия вместо таких неопределенных терминов, как «кажется», «очевидно» и т. д. Такая количественная мера давала бы информацию, насколько *сильна* связь между независимой и зависимой переменными.

ВЫЧИСЛЕНИЕ ω^2

Мы можем получить численную величину силы связи, вычислив ω^2 (ω — малая греческая буква омега; мы называем ω^2 омегой в квадрате). По существу, ω^2 — один из параметров

генеральной совокупности, о которых рассказывалось в статистическом приложении к главе 1. Его полное описание можно найти в работе Хейса (1973).

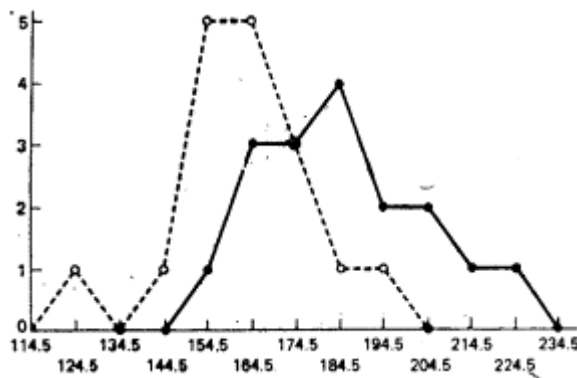


Рис. 4.4. Частотное распределение средних значений времени реакции на световой (условие А) и звуковой (условие Б) сигналы. Ось абсцисс — средние значения времени реакции (в мс), ось ординат — частота. Сплошная линия — световой сигнал, пунктирная — звуковой

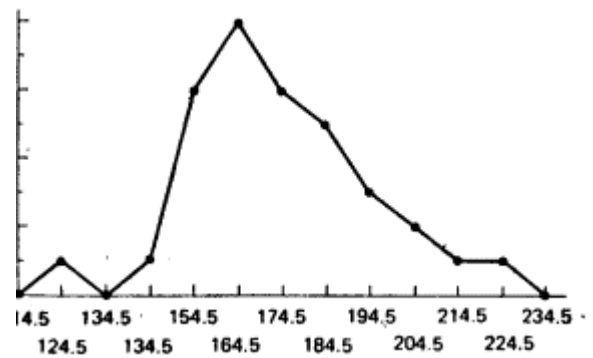


Рис. 4.5. Объединенное частотное распределение средних значений времени реакции. Ось абсцисс — средние значения времени реакции (в мс), ось ординат — частота

Наше вычисление с использованием данных по выборке испытуемых дает оценку ω^2 . Мы будем называть ее *est* ω^2 . Давайте построим новый график результатов эксперимента по измерению времени реакции. Однако теперь мы не будем делать различий между тем, какое из двух условий — А (свет) или Б (тон) — было использовано. Как видно на рис. 4.5, это комбинированное распределение несколько более растянуто, чем каждое из отдельных распределений для света и тона. Чем больше исходные распределения отличаются друг от друга, тем больше будет растянуто комбинированное распределение. Если бы мы смогли провести бесконечный эксперимент и при этом получили бы распределения, показанные на рис. 4.4 и 4.5, мы могли бы вычислить ω^2 прямо из параметра σ^2_x следующим образом:

$$\omega^2 = \frac{\bar{\sigma}_{\text{комб}}^2 - \bar{\sigma}_{\text{отд}}^2}{\bar{\sigma}_{\text{комб}}^2}.$$

(4.1) Однако поскольку наши данные получены только по одной выборке испытуемых, а не в бесконечном эксперименте, мы должны *оценивать* ω^2 по статистике S^2_x :

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

(4.2) Квадрат стандартного отклонения распределения называется *дисперсией*. Числитель этой формулы дает разность между дисперсиями комбинированного распределения и отдельного распределения, в нашем случае любого из условий А или Б. Делением этой разности на дисперсию комбинированного распределения мы придаем ей форму пропорции. Она отвечает на вопрос, на какую часть уменьшается дисперсия показателей при переходе от комбинированного распределения к отдельному. Производя вычисления, нет необходимости сначала вычислять S_x и затем возводить его в квадрат, чтобы получить S^2_x . Вспомните (из формулы 2.2):

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

Поэтому

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

(4.3)

В статистическом приложении к главе 3 мы вычислили $\sum x^2$ для условия А (свет) и условия В (тон):

$$S_A^2 = \frac{5894,25}{16} = 368,$$

$$S_B^2 = \frac{4048,25}{16} = 253.$$

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

$$S_{отд}^2 = \frac{S_A^2 + S_B^2}{2}.$$

(4.4)

поэтому

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

Такое же вычисление производится для комбинированного распределения:

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

Подставляя эти величины в формулу 4.2, получаем

$$est \omega^2 = \frac{S_{комб}^2 - S_{отд}^2}{S_{комб}^2},$$

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

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

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

Это очень слабая связь между независимой и зависимой переменными.

ПРИМЕНЕНИЯ ω^2

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

ИНТЕРПРЕТАЦИЯ ω^2 Мы можем рассматривать ω^2 как индикатор величины, на которую уменьшается *неопределенность* за счет того, что нам известно экспериментальное условие. В эксперименте по измерению времени реакции мы кое-что знаем о среднем показателе для каждого отдельного испытуемого благодаря тому, что знаем экспериментальное условие. Но в пределах каждого условия показатели варьируют, т.е. как-то распределены. Наша неопределенность измеряется дисперсией этого распределения. Если мы не знаем условия, предъявлявшегося испытуемому, наша неопределенность увеличивается: дисперсия комбинированного распределения больше, чем дисперсия для отдельного условия. Таким образом, зная, какое из условий предъявляется испытуемому, мы уменьшаем *неопределенность*. Как уже говорилось, деление этого уменьшения на уменьшаемую дисперсию ($S^2_{\text{комб}}$) превращает ответ в отношение. Тем самым ω^2 сообщает нам часть, на которую уменьшается неопределенность при знании экспериментального условия. Это и есть мера воздействия на поведение независимой переменной. *Задача:* Вычислите *est* ω^2 для эксперимента, сравнивающего условие В и Г, с 18 испытуемыми в каждой группе:

$$\sum x^2_{\text{в}}=4700; \sum x^2_{\text{г}}=4900; \sum x^2_{\text{комб}}=15,000.$$

Ответ: *est* $\omega^2=0,34$.
