

# Измерения, выборка и обработка данных

### Обзор задач главы

В этой главе мы начнем подробное рассмотрение процедуры сбора данных в ходе исследования. Мы будем говорить о разнообразных параметрах, измеряемых в психологических исследованиях, о факторах, определяющих, имеют ли смысл проводимые измерения; рассмотрим различные типы так называемой шкалы измерений и методы отбора участников исследования. В этой главе также даются начальные сведения о различиях между описательной статистикой и статистикой вывода (некоторые из вас, я надеюсь, смогут пополнить уже имеющиеся знания) и рассматривается логика проверки гипотез. Изучив данную главу, вы сможете:

- понять взаимоотношения между конструктом (например, зрительным образом) и измеряемым параметром действия (например, скоростью реакции);
- различать надежность и различные виды валидности измерений;
- описать особенности номинальной, порядковой, интервальной шкал и шкалы отношений, а также указать, когда каждая из них должна использоваться;
- описать три вида вероятностной выборки (простую случайную, расслоенную и кластерную), а также указать, когда каждая из них должна использоваться;
- понять, когда нужно использовать вероятностную выборку и почему в большинстве психологических исследований используются простые (невероятностные) выборки;
- обобщать данные, определяя общую тенденцию (например, вычисляя среднее арифметическое), изменчивость (например, находя стандартное отклонение) и делая визуальные представления данных (например, методом стебля и листа);
- понимать суть процедуры проверки гипотезы и знать составляющие процесса заключительной обработки данных.

Из главы 1 вы узнали, что психологи-исследователи применяют информационный подход: выводы о поведении должны основываться на данных, собранных научными методами. Решить, параметры каких именно действий измерять, как прово-

дить измерения и как организовать полученный набор чисел, весьма нелегко. Данная глава открывает обсуждение взаимоотношений между собираемыми данными и психологическим знанием, которое мы продолжим в следующих главах.

## Что измерять: разнообразие поведенческих актов

Разнообразие поведенческих актов, параметры которых измеряются психологами-экспериментаторами, практически не ограничено. Поддающиеся измерению феномены варьируются от внешних форм поведения, как передвижение по лабиринту, до получаемых с помощью опросников вербальных отчетов и записей физиологической активности, проявляемой при выполнении некоторых заданий. Для иллюстрации многообразия измерений в ходе психологических исследований рассмотрим следующие примеры:

1. В работе «Зависимость объема восприятия больных шизофренией от отвлекающих воздействий и латеральности» (Elkins, Cromwell & Asarnow, 1992) изучались ограничения объема восприятия у больных шизофренией. В сравнении с контрольной группой эти испытуемые хуже справились с заданием на распознавание отдельных букв в массиве других, отвлекающих внимание. Оценивалось, могут ли испытуемые точно указать нужные буквы в различных ситуациях.
2. В работе «Способности к математическим вычислениям у детей младшего возраста из семей со средним и низким достатком» (Jordan, Huttenlocher & Levine, 1992) исследовалась способность детей к решению различных математических задач. Было обнаружено, что дети из семей со средним достатком лучше выполняли задания, если их формулировали устно, при этом обе группы одинаково успешно справлялись с невербальными задачами. Оценивались две характеристики: ответ на вопрос задачи и используемая стратегия. Детей наблюдали во время выполнения задания, следя за тем, используют ли они счет на пальцах.
3. В исследовании «Восприятие социальной поддержки, социальные навыки и особенности отношений с окружающими у женщин, страдающих булимией» (Grissett & Norvell, 1992) женщины с булимией и контрольная группа выполняли батарею тестов самоочета. Оценивались ответы на поставленные вопросы (например, по шкале восприятия поддержки). Кроме того, с участницами проводились краткие беседы, записываемые ассистентами и оцениваемые наблюдателями. Речевое и неречевое поведение женщин в процессе беседы оценивалось по показателю «социальная эффективность». Женщины с булимией считали, что не получают большой социальной поддержки, их социальная эффективность была невысока.
4. В работе «Определяется ли выбор направления в радиальном лабиринте с помощью когнитивной карты?» (Brown, 1992) на этот вопрос был получен ответ: «Не всегда». Радиальный лабиринт, состоящий из центральной площадки и дорожек, расходящихся от нее во всех направлениях, — один из наиболее популярных видов современных лабиринтов. Его часто используют при изучении пространственной памяти (когнитивных карт) у крыс. В ис-

следовании Брауна изучалась способность крыс не пользоваться повторно одними и теми же дорожками. Оценивались ситуации «макровыбора», операционально определенные как возникающие, когда крыса заглядывает за поворот дорожки, и «микровыбора», обозначенные как ситуации, когда крыса поворачивается к дорожке, но не идет по ней (все это фиксировалось с помощью видеокамеры, направленной вниз на лабиринт).

### **Разработка измерений на основе конструкторов**

Из приведенных примеров видно, что исследователи оценивают параметры поведения различными способами. Но как они решают, что необходимо измерять; что можно определять вычислительные стратегии, наблюдая счет на пальцах, фиксировать микровыборы, наблюдая положение тела, а степень внимания — через выбор букв из массива?

Ученые знают что надо измерять, в частности, потому, что они знакомы с литературой по теме своего исследования и знают, какие измерения в этих случаях обычно проводятся. Новые измерения возникают также из модификации общеизвестных в психологии вообще. Кроме того, разработка измерений может проводиться в процессе усовершенствования исследуемых конструкторов. Рассмотрим эти вопросы более подробно.

При планировании исследования одна из основных задач — как можно более точно определить конструкторы, используемые в проекте. Звучит знакомо? Так и должно быть, ведь мы опять говорим об операциональных определениях. В любой исследовательский план входит установление границ интересующих ученого конструкторов, которые по определению не наблюдаются непосредственно, и выявление типов поведения, адекватно отражающих эти конструкторы. В приведенных выше примерах все исследователи ставили перед собой одну и ту же задачу: на основе определенного явления построить управляемый эксперимент, определив конструкторы в терминах измеряемых характеристик поведения. В табл. 4.1 представлена классификация четырех исследований по изучаемым конструкторам и способам их операционального определения на основе конкретных видов поведения.

**Таблица 4.1**

#### **Примеры конструкторов и способы их измерений**

<b>Конструктор</b>	<b>Параметр поведения, на основании которого измерялся конструктор</b>
Объем внимания	Точность узнавания букв
Способность к вычислениям	Количество правильно решенных задач
Воспринимаемая социальная поддержка	Результаты теста самоотчета
Социальная эффективность	Оценка наблюдателем социальных взаимодействий
Макровыбор у крыс	Заходит или нет крыса на дорожку, ведущую от центра в радиальном лабиринте
Микровыбор у крыс	Ориентация тела крысы в направлении дорожки, ведущей от центра в радиальном лабиринте

Легко заметить, что ни один из этих конструкторов (внимание, способность к вычислениям) нельзя наблюдать непосредственно — выводы о них делаются на основании проведенных измерений. Психологи обращаются к этому методу снова и снова, что позволяет исследователям ставить такие эмпирические вопросы, ответы на которые на первый взгляд невозможно получить. Рассмотрим более подробно пример двух процедур, которые часто используются при изучении вопросов, получение эмпирических ответов на которые может представляться трудным или даже невозможным.

Понимают ли младенцы, что существует сила тяжести?

Можно ли продемонстрировать, что люди используют зрительные образы в процессе обработки информации?

Измерения, необходимые для изучения этих, на первый взгляд неэмпирических вопросов, так же просты, как в случае исследований вопросов: а) как долго младенец может смотреть на один объект или б) сколько времени требуется людям, чтобы принять решение.

### **Пример 1. Привыкание**

Знают ли младенцы о существовании силы тяжести? Как вообще возможно это определить? Конечно, их нельзя спросить прямо, но вопрос можно задать опосредованно, с помощью методики измерения количества времени, которое затрачивает ребенок, глядя на различные объекты. Эта так называемая процедура привыкания заключается в том, что ребенку много раз показывают один и те же стимульный материал, а затем переключаются на новый. Из исследований известно, что младенцы любят новизну (Spelke, 1985), так что если им многократно демонстрируется один и тот же объект, они теряют к нему интерес (т. е. прекращают смотреть на него). Термин «привыкание» определяется как постепенное ослабление ответной реакции на повторяющиеся стимулы. Когда показывают новый объект и младенец воспринимает его как нечто новое или необычное, то он будет смотреть на него дольше. Таким образом, если время разглядывания объекта уменьшается, а затем внезапно увеличивается, то можно предположить, что младенец заметил что-то новое.

Имея это в виду, рассмотрим восхитительный эксперимент, проведенный Кином и Спилке в 1992 г. Они сравнили 5- и 7-месячных детей и выяснили, что представление о силе тяжести формируется именно в этот период жизни. Сначала детям многократно показывали кадры, на которых шары катились вверх или вниз по наклонным плоскостям, как показано на верхних двух изображениях на рис. 4.1. Во время процедуры привыкания одним младенцам показывали шар, катящийся с ускорением вниз по плоскости (условие 1), а другим — шар, катящийся вверх по плоскости, замедляющий свое движение (условие 2). Эти явления отражают естественное воздействие силы тяжести на шары, катящиеся вверх или вниз по склону. Когда привыкание было сформировано (т. е. время рассматривания после повторных попыток заметно снизилось), детям показали либо «естественное тестовое явление» (изображения в середине), либо «неестественное тестовое явление» (нижние изображения). Обратите внимание, что естественное тестовое явление отличается от явления, формирующего привыкание, в двух аспектах: в аспектах направления и скорости. При этом если младенец при формировании привыкания



### Пример 2. Скорость реакции

Используем ли мы зрительные образы в процессе обработки информации? Как это можно выяснить? Конечно, об этом можно спросить, но если человек ответит «Да, я использую образы», то как узнать что именно он делает, используя их? Таким образом, перед вами встанет проблема, которая привела к отказу от интроспекции как метода, — отсутствие объективности. Однако можно попросить людей выполнить определенное задание, решая которое они будут вести себя одним образом, используя образы, и другим — не используя. Именно такую стратегию использовали Шепард и его коллеги в серии хорошо известных экспериментов по изучению так называемого «мысленного вращения».

Рассмотрите две пары геометрических фигур на рис. 4.2. Являются ли правые фигуры в каждой паре теми же самыми, что слева, только перевернутыми, или это другие фигуры? Как вы это определили? В 1971 г. Шепард и Мецлер в ходе эксперимента попросили участников не просто ответить на этот вопрос, но также зафиксировать *время*, за которое был найден ответ. Их идея заключалась в том, что если участники решают задачу, мысленно переворачивая левую фигуру так, чтобы она совпала с правой, то этот процесс займет определенное количество времени. Более того, чем сильнее должен быть поворот, чтобы фигуры совпали, тем больше времени на это уйдет. Я полагаю, вы понимаете, как это происходит. Шепард и Мецлер последовательно изменяли градус поворота и обнаружили, что с его увеличением растет время, необходимое для ответа на вопрос. Таким образом, измерив скорость реакции, они заключили, что участники используют мысленные образы.

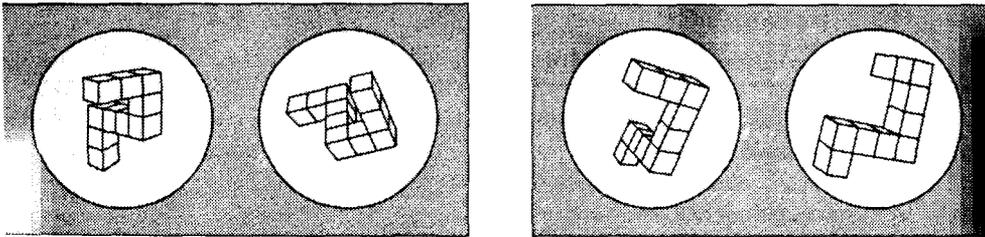


Рис. 4.2. Стимульные материалы для исследования мысленного вращения, проведенного Шепардом и Мецлером в 1971 г.

Измерение скорости реакции — один из старейших и наиболее устойчивых методов в психологии, но причины, по которым к нему прибегают, с течением времени изменяются. Дополнительные сведения о происхождении и развитии этого весьма эффективного метода, применяемого в экспериментальной психологии, вы получите, прочитав вставку 4.1.

## Оценочные измерения

Сак можно определить правильность измерения поведения? Чем объясняется уверенность ученых в оценке таких вещей, как время наблюдения события, скорость реакции, выполнение теста интеллекта, определение восприятия социальной поддержки и т. д.? Чтобы ответить на этот вопрос, необходимо обсудить два ключевых фактора: надежность и валидность.

## ВСТАВКА 4.1

**История: скорость реакции — от хронометрических экспериментов до мысленного вращения**

В психологической литературе понятие скорости реакции впервые упоминается в работе Ф. К. Дондерса, датского физиолога, который установил, что продолжительность психического явления можно определить, найдя разницу между скоростями реакций, получаемых при выполнении различных видов заданий (Boeing, 1950). Его идея породила водопад исследований, посвященных явлению, сегодня известному как «психическая хронометрия». Сначала исследователи измеряли скорость простой сенсомоторной реакции (ПСР) - скорость однократной реакции, полученной сразу за предъявлением однократного стимула, например сигнала красного цвета. Далее они усложняли задание, показывая поочередно два стимула и предлагая испытуемому отвечать только на один из них. Это получило название скорости реакции выбора (РВ), так как испытуемому требовалось сначала дифференцировать стимулы, например красный и зеленый сигналы, а затем отвечать на один из них. СДР включает СПР плюс время психического явления «дифференциации», поэтому было решено, что найдя разность СДР и СПР можно получить продолжительность этого психического явления:

$$РВ = ПСР + \text{дифференциация};$$

$$\text{дифференциация} = РВ - ПСР.$$

Процедуру можно разрабатывать и дальше, получая более дробные разности психических явлений - этот процесс очень воодушевлял исследователей конца XIX в., так как в то время они пытались заявить психологию как науку, а что может быть более научным, чем измерение психических явлений с точностью до долей секунд? Особенно высокой популярностью эта процедура пользовалась в лаборатории Вундта в Лейпциге, и как видно из рис. 4.3, она была быстро импортирована в Соединенные Штаты. На рисунке показан ход эксперимента по измерению скорости реакции, проводящийся в Университете Кларк в начале 1890-х гг.

К сожалению, вскоре обнаружили серьезные проблемы. В частности, в одних случаях скорость реакции была ниже, чем предсказывалось, а в других выше. Освальд Кюльпе, один из студентов Вундта, отметил главнейшую ошибку метода - более сложные психические явления не образуются из простых в результате их простого суммирования. Напротив, сложные психические явления обладают присущими только им качествами, которые не являются суммой более простых.

Хотя доводы Кюльпе быстро положили конец психической хронометрии и дискредитировали измерение скорости реакции как метод еще в период расцвета бихевиоризма (приблизительно 1930-1950 гг.), впоследствии она пережила возрождение в некоторых областях когнитивной психологии. Возникла идея применять психическую хронометрию не для измерения точной продолжительности психических явлений, а для проверки предположений, сделанных на основе когнитивных теорий. Хороший пример тому - феномен мысленного вращения. Шепард предположил, что если участники применяют мысленное вращение, то эта умственная активность должна занимать определенное количество времени (Shepard & Metzler, 1971). Поворот на больший угол должен занимать больше времени, и как вы могли убедиться, это действительно так.

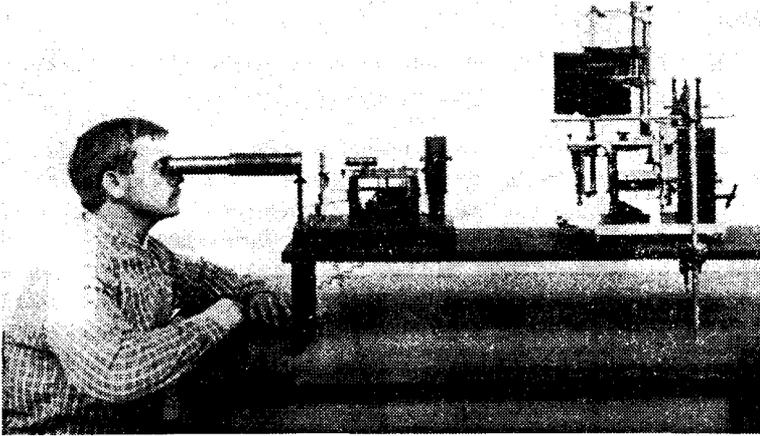


Рис. 4.3. Эксперимент по изучению скорости реакции, проводившийся в Университете Кларк около 1892 г. Ответная реакция выражалась в том, что участник, видя в трубе предъявляемый стимул, правой рукой переставал нажимать на телеграфный ключ

### Надежность

В целом измерение характеристик поведения считается надежным, если его результаты повторяются при повторном измерении. Скорость реакции является хорошим примером — высокая надежность была одной из причин многолетней популярности таких измерений. Человек, при первой попытке реагирующий на красный сигнал через 0,18 секунды, наверняка покажет ту же скорость при следующих попытках, и практически все попытки будут давать результаты, близкие к 0,18 секунды. К примеру, оценки теста *GRE* (*Graduate Record Exam*) также относительно надежны. Человек, получивший 850 за общий тест *GRE*, во второй раз, скорее всего, получит сходную оценку, и очень маловероятно, что его балл будет 1350.

Из этих двух примеров можно понять, почему надежность измерений настолько важна. Без нее невозможно было бы определить, что означают оценки, полученные при конкретном измерении. Предположим, измеряя скорость реакции, вы хотите узнать, насколько быстро реагирует конкретный человек. Если это время станет сильно изменяться от попытки к попытке, то на данный вопрос ответить будет невозможно. Аналогично, если разница в оценках теста *GRE* для одного человека может составлять 400 или 500 баллов, вузы не смогут использовать результаты этого теста, так как по ним нельзя будет предположить действительный уровень знаний студента.

Надежность измерения характеристик поведения — это функция от количества **ошибок измерения**. Если ошибок много — надежность низкая, и наоборот. Ни одно подобное измерение не является абсолютно надежным, так как все они содержат некоторое количество ошибок. Это означает, что результат измерения представляет собой гипотетическую истинную оценку плюс ошибку измерения. В идеале ошибка измерения настолько мала, что полученная оценка близка к истинной.

Процедура измерения скорости реакции хорошо иллюстрирует, что такое ошибка измерения и как она влияет на надежность. Как и в предыдущем примере, предположим, что скорость реакции человека составила 0,18 секунды. Является ли это истинным значением измеренной скорости? Нет, как несложно заключить из следующих результатов, полученных при измерении скорости реакции у того же человека:

0,16с, 0,15 с, 0,19 с, 0,17 с, 0,19 с.

Результаты различаются, потому что каждая из попыток не обошлась без ошибки измерения. Эта ошибка вызвана различными факторами — некоторые из них действуют случайным образом от попытки к попытке. Например, при одной из попыток человек может реагировать быстрее, чем предполагает истинное значение, если будет ожидать появления стимула, или медленнее из-за кратковременного снижения внимания. Также может возникнуть систематическая ошибка, если, к примеру, перед тем, как предъявить стимул, экспериментатор дает участнику сигнал приготовиться, а время между этим сигналом и стимулом остается постоянным. В этом случае участник начнет предугадывать стимул и скорость реакции будет систематически выше, чем истинная оценка.

Несмотря на небольшую ошибку измерений, приведенные выше оценки неплохо соответствуют друг другу и измерение скорости реакции в этом случае будет считаться более надежным, чем при следующем наборе оценок, полученном вслед за результатом 0,18 секунды:

0,11с, 0,25 с, 0,19 с, 0,09 с, 0,31с.

При результатах, варьирующихся от одной десятой секунды до одной третьей, очень трудно установить действительную скорость реакции человека.

Если оценки надежны, то исследователь может в зависимости от их значений приписать им некоторый смысл. Надежность также позволяет исследователю делать более осмысленные сравнения одних оценок с другими. Например, сравнение первого набора оценок (0,16 с, 0,15 с и т. д.) с нижеследующим обнаруживает разницу в общей скорости реакции:

0,23 с, 0,26 с, 0,21с, 0,22 с, 0,24 с.

Очевидно, что второй испытуемый реагировал медленнее, чем первый.

Есть разные способы вычисления надежности, но в экспериментальных исследованиях это редко проделывают. Уверенность в надежности измерений вырабатывается со временем в результате процесса их повторения. Например, процедура выработки привыкания и измерение скорости реакции проводились достаточно часто и приносили весьма непротиворечивые результаты, так что исследователи убедились в их надежности.

Надежность вычисляется более строго в исследованиях, посвященных оценке адекватности всевозможных психологических тестов, разработанных для измерения таких конструктов, как личностные факторы, способности (например, *IQ*) и установки. Это обычно письменные тесты, в которых требуется отвечать на вопросы или что-либо утверждать/опровергать. В описанном выше исследовании были-

мии участницы заполняли различные тесты, в том числе так называемую шкалу восприятия поддержки. Анализ, призванный установить надежность такого вида тестов, требует применения корреляционных процедур. Например, участники могут выполнить тест в двух различных ситуациях, а далее будет определена степень сходства двух наборов оценок. Если не повлияют серьезные изменения, произошедшие в жизни участников, результаты двух измерений по шкале восприятия поддержки должны быть похожи. Степень сходства выражается в терминах корреляции (высокое сходство = высокая корреляция). В главе 9 будут более подробно объяснены особенности этого вида анализа и его значение для психологического тестирования в целом.

### **Валидность**

Измерение характеристик поведения считается валидным, если измеряют именно то, что собирались. В ходе оценки восприятия социальной поддержки должно действительно измеряться количество поддержки, которую (по их мнению) получают испытуемые, а не какой-либо другой конструкт. Тест на интеллект должен измерять интеллект, а не что-то иное.

Иногда считается, что измерение обладает определенной валидностью просто потому, что оно представляется разумным. При измерении интеллекта тест с задачами, требующими размышлений, представляется более подходящим, чем тест, в ходе которого участники должны аккуратно проехать на велосипеде между двумя белыми линиями. Это значит, что решение задач и рассуждения имеют большую очевидную валидность для измерения интеллекта, чем езда на велосипеде (которая обладает очевидной валидностью для определения способности к удержанию равновесия). Но конечно, только очевидной валидности недостаточно — тест может выглядеть весьма разумным, но при этом не быть валидным. Большинство опросов, которые публикуются в популярных журналах, попадают именно в эту категорию.

Более строгий вид валидности называется критериальной валидностью. Она показывает: а) можно ли на основании измерения предсказать будущее поведение и б) связано ли это измерение с другими измерениями поведения. Например, чтобы тест на уровень интеллекта был валидным: а) должно быть возможным на его основании предсказать успеваемость ребенка в школе и б) его результаты должны быть сходны с результатами других известных тестов по измерению уровня интеллекта. Валидность названа «критериальной», потому что результаты рассматриваемых измерений сопоставляются с некоторым значением или критерием. В данном примере критериальными переменными являются: а) будущие оценки в школе и б) оценки по существующим тестам на интеллект. Так же как оценка надежности, исследование критериальной валидности корреляционное по своей природе и проводится в первую очередь для оценки психологических тестов. В главе 9 вы снова встретитесь с этим вопросом.

Еще один вид валидности — конструктивная валидность — особенно подходит для экспериментальных исследований. Она имеет отношение сразу к двум вопросам: является ли оцениваемый конструкт валидным, а инструмент, с помощью которого производится оценка, — наилучшим. Конструктивная валидность тесно свя-

зана с сущностью теории, построением гипотезы на основании теории и оценкой теорий, выведенных из результатов исследования. Конструктивная валидность как таковая никогда не подтверждается и не разрушается одним исследованием, а также ее невозможно доказать по тем же причинам, но каким невозможно доказать теорию. Уверенность в конструктивной валидности растет постепенно по мере получения исследованиями подтверждающих результатов.

Как мы видели, аппарат психологии насыщен конструктами, которые не могут наблюдаться непосредственно, — такими понятиями, как голод, тревожность, интеллект, депрессия, воспринимаемая социальная поддержка, объем внимания и т. д. Так как эти конструкты являются гипотетическими, их наличие может быть выведено только из специально проведенных измерений. Исследование, в котором проводится измерение конструкта определенным образом и которое дает ранее предсказанные результаты, подтверждает не только сам конструкт, но и инструмент, использованный для его измерения.

### Пример 3. Конструктивная валидность

Для примера конструктивной валидности рассмотрим серию исследований, проведенных Уолтером Мишелем — ученым, занимающимся психологией личности, — и его коллегами. Мишель интересовался вопросом, почему дети бывают нетерпеливы, почему они иногда хотят чего-то «прямо сейчас» и почему им трудно ждать. Мишель придумал конструкт, который назвал «задержка удовольствия» и попытался разработать для него подходящий способ измерения. Его исследовательская программа показала, что задержка удовольствия валидна как конструкт и соответствует общей когнитивно-социальной теории личности, а также валидность разработанных им измерений.

Одно простое измерение, придуманное Мишелем, состояло в том, чтобы попросить детей выбрать между маленькой наградой, доступной в этот же момент, и большей, но при условии, что ее выдадут спустя некоторое время. Мишель предположил, что если неспособность к задержке удовольствия является неотъемлемой особенностью маленьких детей, то старшие дети должны охотнее ждать большую награду. Эти рассуждения привели к очевидному эксперименту (процитировано в Mischel, 1981), в ходе которого дети выполнили небольшое задание, а затем им сказали:

Я бы хотел раздать каждому из вас по конфете, но сегодня я взял с собой очень мало вот таких (показывая большую конфету). Поэтому вы можете либо взять такую (показывая маленькую конфету) прямо сейчас, или, если хотите, можете подождать такую (показывая), которую я принесу в следующий четверг (неделей позже).

Mischel, 1981, p. 164-165

Результаты подтвердили предположение Мишеля: немедленное (но маленькое) вознаграждение было выбрано 81% 7-летних, 48% 8-летних и 20% 9-летних детей.

Естественно, одного этого эксперимента недостаточно для того, чтобы установить валидность задержки удовольствия как конструкта или процедуры измерений, предложенной Мишелем, поэтому он продолжил разработку серии экспериментов, исследуя возможные взаимосвязи между изучаемым конструктом и другими признанными конструктами. Например, он обнаружил, что дети, решившие

отложить удовольствие, были также эмоционально более зрелыми, сильнее нацелены на успех, вероятность делинквентности для них была ниже, а вероятность стать социально ответственными — выше (Mischel, 1981). Таким образом, на основании ряда исследований, результаты которых подтвердили сделанные предположения, задержка удовольствия была признана валидным конструктом.

### **Надежность и валидность**

Чтобы измерение имело ценность для психологического исследования, оно должно обладать одновременно и надежностью, и валидностью. Надежность важна для формирования уверенности в том, что проводимое измерение приближает к истинному значению, а валидность важна, так как она указывает, что измеряется именно то, что вы предполагаете. Обратите внимание, что валидность предполагает надежность, а обратное не верно. Надежные измерения могут не быть валидными, но валидные должны быть надежными.

Проиллюстрировать это можно на простом примере. Из главы 1 вы получили некоторые сведения о френологии — популярной в XIX в. теории, утверждающей, что можно определить «способности» человека по форме его черепа. После обсуждения проблемы надежности вы можете сделать вывод, что френологические измерения черепа действительно обладали высокой надежностью — расстояние между точкой, расположенной пятью сантиметрами выше вашего левого уха, и другой, находящейся пятью сантиметрами выше вашего правого уха, не сильно изменится от измерения к измерению. Однако утверждение, что при измерении определяется «деструктивная» способность человека, — это совсем другое дело. Мы уверены, что измерение черепа не является валидным измерением деструктивности, поскольку оно не кажется нам разумным (очевидная валидность), на основании его невозможно предсказать проявление агрессивности (критериальная валидность), а также потому, что оно не согласуется с другими исследованиями конструктов, связанных с деструктивностью, таких как импульсивность, или исследованиями функций мозга (конструктивная валидность).

Вопрос надежности и валидности имеет также этическую сторону, особенно если результаты измерений используются для принятия решений, касающихся жизни людей. Студентов принимают или нет в учебные заведения, людей берут или нет на работу, кому-то ставят психиатрический диагноз и предоставляют лечение — и все это на основании измерений способностей или оценки особенностей поведения. Если вы ищете работу и ваши оценки по некоторому тесту являются решающим показателем, то вы испытаете законное разочарование, узнав, что тест не был достаточно надежным и валидным.

И последнее замечание. В этой главе понятие валидности обсуждалось в контексте темы измерений. Как вы узнаете из следующей главы, валидными могут быть не только измерения психологических конструктов. Валидность также в более общем смысле характеризует весь исследовательский проект. По отношению к измерениям валидность означает, что используемый инструмент измеряет именно то, что должен. В более широком смысле валидность характеризует правильность проведения экспериментов и проверки гипотезы.

## Виды шкал измерений

Результаты измерения характеристик поведения представляются в виде набора чисел. Мы говорим, что кто-то среагировал через 3,5 секунды, получил 120 баллов за /Q-тест или нашел выход из лабиринта третьим по счету. Мы также относим X человек к определенной категории в зависимости от их поведения или обладания некоторыми особенностями. Данные примеры иллюстрируют четыре числовых способа описания явлений, а следовательно, четыре **шкалы измерений**. Ясное понимание этих способов необходимо для дальнейшего обсуждения статистических методов, так как от вида используемой шкалы зависит вариант соответствующего статистического анализа данных. Именно неразбериха со шкалами измерений оказалась столь затруднительна для Дилберта, история которого изображена на рис. 4.4.

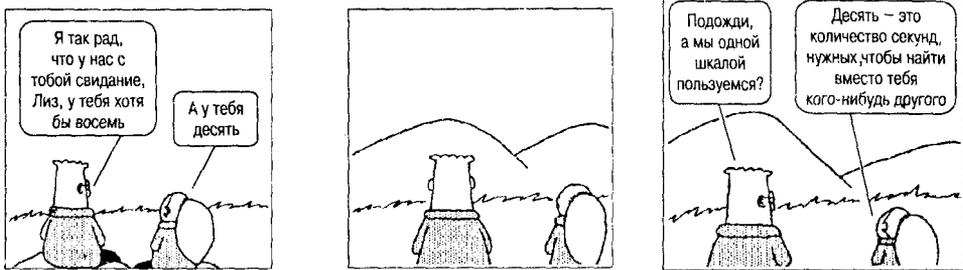


Рис. 4.4. Проблема со шкалой измерений

### Номинальная шкала

Иногда численные характеристики, которые мы приписываем явлениям, служат только для того, чтобы разделить эти явления по группам. В этом случае мы используем так называемую **номинальную шкалу** измерений и числа являются не более чем ярлыками категорий (например, 15 мужчин, 25 женщин). В исследованиях с использованием этой шкалы люди обычно разделяются по категориям и далее подсчитывается количество людей, попадающих в каждую из них. Номинальную шкалу используют, если поставлены, например, такие эмпирические вопросы:

- Среди людей, занимающихся бегом, кто чаще бежит по утрам, а кто по вечерам, мужчины или женщины?
- Если разделить психически неуравновешенных людей на замкнутых и общительных, будут ли интроверты чаще испытывать состояние тревоги, чем экстраверты?
- Если разделить людей на ожидающих сильный электрический шок и на не ожидающих, повлияет ли это их ожидание на тенденцию к аффилиации?

Последний эмпирический вопрос был поставлен в хорошо известном эксперименте по социальной психологии, проведенном Стэнли Шехтером в 1959 г. Половине женщин, участвующих в исследовании, он сказал что в эксперименте будет изучаться психологическая реакция на сильный электрический шок, а остальным — что разряд будет очень слабым и станет щекотно. После того как женщин в обеих

группах проинформировали о предстоящем испытании, их попросили подождать 10 минут, пока настраивается оборудование, и спросили, как они предпочитают ждать, вместе с другими, в одиночестве, или же не имеют в этом отношении предпочтений. В действительности Шехтера интересовал ответ именно на этот вопрос, и эксперимент был окончен после того, как участницы высказались.

Шехтер изучал влияние состояния тревожности на конструкт, названный аффилиацией (желание находиться вместе с другими). Каковы же были результаты? Те, кто ожидал сильного шока, предпочитали ждать вместе с другими, а те, кто ожидал слабого — в основном не имели предпочтений. Этот эксперимент иллюстрирует применение номинальной шкалы, поскольку Шехтер разделил все данные на шесть категорий и подсчитал количество людей, попадающих в каждую из них в зависимости от того: а) какой силы электрический разряд они ожидали получить и б) как они ответили на ключевой вопрос. Количество людей («подсчет частости») в каждой категории показано в табл. 4.2. Информация также переведена в проценты, что является обычным при использовании номинальных данных.

Таблица 4,2

#### Использование номинальной шкалы: исследование Шехтером аффилиации

	Количество людей, выразивших следующие предпочтения		
	Ждать в одиночестве	Ждать вместе с другими	без предпочтений
Ожидающие сильный шок	3 (9%)	20 (63%)	9 (32%)
Ожидающие слабый шок	2(7%)	10(33%)	18(60%)

### Порядковая шкала

**Порядковая шкала** измерений представляет собой набор порядковых мест. Например, в списках успеваемости в колледжах обычно отражаются места студентов в группе: 1е, 2е, 3е, 50е и т. д. Из такого списка можно узнать, что один студент получает более высокие оценки, чем другой. Однако относительная позиция — это все, что можно узнать. Эд, Тед и Фред будут занимать соответственно 1-е, 2-е и 3-е места в каждом из двух приведенных ниже случаев, несмотря на то что оценки Эда заметно выше оценок Фреда и Теда только во втором случае:

Случай 1

Средний балл Эда = 4,0

Средний балл Фреда = 3,9

Средний балл Теда = 3,8

Случай 2

Средний балл Эда = 4,0

Средний балл Фреда = 3,6

Средний балл Теда = 3,5

Исследования, в которых используется порядковая шкала, пригодны для решения таких эмпирических вопросов:

- Если ребенок расположит пять игрушек в ряд в порядке убывания привлекательности, а затем ему дадут игрушку, занимающую третье место в ряду,

то увеличится или уменьшится порядковый номер этой игрушки в ряду спустя неделю игры с нею?

- Изменяется ли ранжирование студентами авторов учебников по естественным и гуманитарным дисциплинам, если они знают пол авторов?
- Как молодые и пожилые люди оценят и расположат в ряд 10 кинофильмов с разным количеством сексуальных сцен и сцен насилия?

Хороший пример использования порядковой шкалы — исследование Корна, Девиса и Девиса (Korn, Davis, and Davis), проведенное в 1991 г. Историков психологии и заведующих кафедр попросили назвать 10 психологов, чьи работы, по их мнению, наиболее важны, и расположить их в ряд. Требовалось составить два ряда: один из десяти психологов, лучших «во все времена», и другой — из современников. Затем результаты обобщили. Кто же был первым в списке? И историками, и заведующими кафедр самым выдающимся современным психологом был признан Б. Ф. Скиннер. При этом заведующие кафедр назвали Скиннера и лучшим психологом в истории, а историки, склонные предпочитать психологов более ранних периодов, поместили Скиннера на девятое место, а на первое — Вундта.

### Интервальная шкала

В большинстве психологических исследований используется либо интервальная шкала измерений, либо шкала отношений. Интервальная шкала дополняет идею ранжирования принципом равных интервалов между ранжируемыми явлениями. Наиболее распространенный пример использования интервальной шкалы — психологические тесты личности, установок и способностей. Например, результаты теста интеллекта обычно представляются подобным образом. Некто, имеющий  $IQ$  120, предполагается более умным (предположим, что  $IQ$  определяет умственные способности), чем тот, чей  $IQ$  равен 100. Более того — и в этом заключается отличительная особенность интервальной шкалы — разница в интеллектуальных способностях двух людей с  $IQ$  120 и 110 предполагается такой же, как разница между людьми с  $IQ$  110 и 100. Другими словами, предполагается, что каждый шаг увеличения значения  $IQ$  отражает одинаковую степень увеличения интеллекта (интервалы равные). Обратите внимание на слово «предполагается» — некоторые психологи считают, что  $IQ$  (а также результаты большинства тестов личности) строятся по принципу шкалы отношений; они утверждают, что довольно трудно, если вообще возможно, в этом случае предполагать равные интервалы. Но большинство ученых считают, что  $IQ$  относится к интервальной шкале, хотя говорят так отчасти по прагматическим причинам: психологи любят использовать интервальную шкалу и шкалу отношений в основном потому, что структура данных этих шкал позволяет проводить более сложный статистический анализ.

В описанном ранее исследовании булимии использовались различные виды измерений (например, по шкале восприятия социальной поддержки), являющихся примером интервальной шкалы. Во вставке 4.2 рассматривается классическая серия экспериментов, в которых интервальная шкала использовалась при попытке показать влияние телосложения на тип личности.

#### ВСТАВКА 4.2

### **Классические исследования — определение соматотипа: когда 7 - 1 - 1 встречается 1 - 1 - 7**

Вы уже знаете, что френологи считали, что существует взаимосвязь между определенными физическими характеристиками (форма черепа) и свойствами личности. Френология кажется нам сегодня нелепой, но идея наличия взаимосвязи между физическими особенностями и личностью имеет долгую историю. Более систематическую попытку изучения такой взаимосвязи предпринял Вильям Шелдон в 1940 и 1942 гг.

В работе Шелдона впечатляет его попытка определить телосложение при помощи точной шкалы измерений. Изучив 4000 фотографий обнаженных мужчин (студентов колледжа), он и его исследовательская группа разработали классификацию типов телосложения с помощью трех интервальных шкал, имеющих по 7 делений. Каждая шкала отражала степень выраженности у мужчины одного из идеальных антропометрических типов: эндоморфного (тучный), мезоморфного (мускулистый) и эктоморфного (худой). Предполагалось, что каждый человек несет в себе все три типа, выраженные в разной степени, обычно с преобладанием одного из них. Так, очень полный человек обозначался как 7-1-1, очень худой - как 1-1-7, а Арнольд Шварценеггер - как 1-7-1. 4-4-4 обозначало гармоничное телосложение. Набор чисел, соответствующих конкретному человеку, назывался его «соматотипом».

После измерений соматотипа Шелдон занялся определением типов личности, которые также, по его мнению, можно разделить на 3 категории и оценить по 7-балльной интервальной шкале, обобщающей результаты различных тестов личности. Он обозначил категории как «висцеротоническую», «соматотоническую» и «церебротоническую». Висцеротоники были в основном общительными, уравновешенными, двигались медленно, любили развлекаться и очень любили поесть, соматотоники были агрессивными, эгоистичными и любили риск, а церебротоники были замкнутыми, застенчивыми, предпочитали одиночество и любили решать умственные задачи.

Заключительным шагом Шелдона была проверка взаимосвязи соматотипа и типа личности. Вас не удивит, что чаще всего вместе появлялись следующие пары:

эндоморф-висцеротоник;  
мезоморф- соматотоник;  
эктоморф- церебротоник.

Шелдон считал, что тип телосложения является причиной формирования у человека определенного типа личности, но критики его теории указали, что взаимосвязь здесь не такая прямая и может проявляться по-разному. Эндоморфное телосложение может явиться причиной любви к еде, но разве не может сильное пристрастие к еде сформировать эндоморфа?

Вокруг работы Шелдона велись бурные споры, его теория была в целом опровергнута, но для нас эта теория важна тем, что она - классический пример попытки количественно определить человеческий характер и связать его с внешним видом человека, определяемым по шкале телосложений, в данном случае по интервальной шкале. Возможно, вы обратили внимание на то, что в описании работы Шелдона речь идет только о мужчинах. Это потому, что Шелдон только их и изучал. Но почему это так? Вспомните, что его процедура включала изучение фотографий 4000 обнаженных мужчин. По-видимому, он хотел повторить исследование и изучить также и женщин, но засомневался из-за «санкций против фотографирования обнаженных женщин (студенток колледжа)» (Bavelas, 1978). Далее в этой главе вы узнаете об «удобной» выборке и о том, что одни выборки могут быть более пригодны, чем другие.

Необходимо отметить, что на интервальной шкале отметка ноль — это рядовая отметка, не означающая отсутствия измеряемого признака. Стандартным примером является температура. Ноль градусов не означает отсутствия температуры, а лишь является делением шкалы, которое означает «наденьте свитер». Аналогично в тесте на застенчивость, оценки которого варьируются от 0 до 20, значение 0 — это просто нижнее деление на шкале и не означает полного отсутствия застенчивости.

### Шкала отношений

В случае **шкалы отношений** принципы упорядоченности и равных интервалов перенесены из структур порядковой и интервальных шкал, но в отличие от них шкала отношений имеет точку *истинного нуля*. На шкале отношений отметка ноль означает полное отсутствие измеряемого признака. Например, если количество ошибок для крысы, пробежавшей лабиринт, имеет значение ноль, то это означает отсутствие неправильных поворотов. Шкала отношений обычно используется при исследованиях, в которых проводятся измерения физических величин, таких как расстояние, вес или время. Примеры, посвященные процедуре привыкания и оценке скорости реакции, приведенные ранее в этой главе, иллюстрируют использование шкалы отношений.

## «Кого» измерять: выборки

Кроме решения о том, что измерять при проведении психологических исследований, необходимо также решить, кого просить участвовать в исследовании и чье поведение будет оцениваться. Для этого есть два подхода: вероятностная и простая (невероятностная) выборки.

### Вероятностная выборка

Эта стратегия используется, если поставлена задача узнать что-либо конкретное об определенной группе людей. Группа в целом называется популяцией, а любая из ее подгрупп называется **выборкой**. Иногда бывает возможно изучать всех членов популяции. Например, если вы хотите изучить отношение всех студентов вашей группы по экспериментальной психологии к экспериментам с участием животных и не хотите делать выводы для находящихся за пределами группы, вы можете опросить всех ее членов. В этом случае размер популяции будет равен размеру вашей группы. Однако, как вы можете догадаться, интересующие исследователей популяции обычно слишком велики для того, чтобы изучать каждого их члена. Поэтому из этой популяции необходимо отбирать подмножество, или выборку.

Хотя вся популяция может и не рассматриваться в исследовании, ученым требуется делать выводы именно о популяции, а не только о выборке. Поэтому важно, чтобы выборка отражала особенности популяции в целом. Если это так, выборка считается **репрезентативной**, если нет — **нерепрезентативной**. Если вы хотите изучить восприятие студентами учебы в колледже, то будет серьезной ошибкой выбрать для исследования только тех, кто живет в студенческом общежитии. Из-за того что отношение прочих студентов может отличаться от отношения живущих в общежитии, результаты вашего опроса могут оказаться необъективными и сместиться в пользу последних (более подробно о процедуре опроса вы узнаете в главе 12).

Одним из наиболее известных исторических примеров нерепрезентативной выборки является случай, произошедший во время президентских выборов в 1936 г. Успешно прогнозируя события нескольких предыдущих выборов, журнал «Литрери Дайджест» попытался предсказать результаты выборов, послав 10 миллионов пробных бюллетеней своим подписчикам, людям, выбранным по телефонным книгам всей страны, и людям из регистрационных списков автомобилей (Sprinthall, 2000). Около 25% бюллетеней (почти 2,5 миллиона) вернулись в журнал заполненными. Из людей, заполнивших бюллетень, 57% выбрали кандидата от республиканской партии Альфа Лэндона, а 40% выбрали действующего президента Франклина Рузвельта. На действительных выборах победил Рузвельт, набрав более 60% голосов. Как вы думаете, почему выборка была нерепрезентативной?

Хотя редакторы журнала знали, что их подписчики в основном принадлежат к верхней прослойке среднего класса и одновременно являются республиканцами, они подумали, что, расширив выборку за счет людей, выбранных из телефонных книг и регистрационных списков, они сделают ее более репрезентативной. В действительности они просто набрали еще больше республиканцев. В разгар Великой депрессии позволить себе иметь телефоны и автомобили могли лишь принадлежащие к верхней прослойке среднего класса и верхнему классу люди, а они в большинстве своем были республиканцами, а не демократами. Поэтому на самом деле в своем опросе журнал спрашивал республиканцев о том, кому они отдадут свои голоса.

Возможно, вы заметили еще один недостаток опроса в «Литрери Дайджест». Большое количество бюллетеней вернулось, и составители журнала были вполне уверены в правильности предсказания о победе Лэндона, так как полученные данные отражали взгляды значительного числа людей — около 2,5 миллионов. Но заметьте, что помимо того, что полученные бюллетени составили только одну четверть от посланных, все эти бюллетени были отправлены теми людьми, которые *сами решили* это сделать. Так что люди, приславшие бюллетени обратно в журнал, были не просто республиканцами, но республиканцами, пожелавшими сделать свои взгляды общеизвестными (не кажется ли вам, что в свете этого 57% выбравших Лэндона — это довольно мало?).

**Проблема самовыбора** нередко возникает при опросах, проводимых популярными журналами, а также в случае вопросов, задаваемых своим читателям такими людьми, как Энн Лэндер, чтобы узнать их мнение по поводу отдельных тем. Проводится опрос, затем спустя месяц или два результаты, полученные от тех, кто прислал обратно заполненный опросник, представляются в виде отчета, подразумевающего их валидность. При этом человек, делающий отчет об исследовании, старается произвести впечатление числом ответов, а не репрезентативностью выборки. Пример такого явления — отчет о женской сексуальности (Hite, 1987), в котором кроме прочего утверждалось, что более 90% женщин, состоящих в браке, чувствуют себя эмоционально ущемленными. В ответ на замечание, что опрос был послан лишь отдельным женским организациям и что заполненные опросники прислали обратно только 4,5% из 100 000 женщин, автор просто заявила, что 4500 человек для нее достаточно (как 2,5 миллиона человек было достаточно для «Литрери Дайджест»).

Обладая научным мышлением, вы должны весьма скептически относиться к утверждениям, сделанным на основании такого рода выборок. Если вы хотите еде-

лать правильный вывод о конкретной популяции, необходимо использовать выборку, отражающую популяцию в целом, и самостоятельно осуществлять отбор, не полагаясь на тех, кто решил вам ответить самостоятельно.

### Случайная выборка

Основной вид вероятностной выборки — **простая случайная выборка**. По сути, она означает, что каждый член популяции с равной вероятностью может попасть в выборку. Например, чтобы сделать случайную выборку из 100 студентов вашего учебного заведения, вы можете сложить бумажки с именами всех студентов в большую шляпу, а затем достать из нее 100 бумажек. В действительности процедура немного более сложнее и обычно проводится с использованием таблицы случайных чисел, образец которой представлен в приложении D. Чтобы узнать, в чем заключается эта процедура, изучите пример, приведенный в табл. 4.3, — в нем описывается, как использовать случайные числа для выбора 5 человек из популяции в 20 человек.

Простая случайная выборка — эффективный практический способ построения репрезентативной выборки. Иногда его также используют по этическим причинам. В случае если только небольшая группа людей может получить выгоду или понести издержки и нет других разумных оснований для принятия решения, лучше всего использовать метод случайной выборки. Известный пример такой выборки — лотерея для отбора людей, призывавшихся на военную службу, проводившаяся в 1969 г. в разгар Вьетнамской войны. Чтобы все было честно, 365 дней года должны были иметь равную вероятность стать первым, вторым, третьим и т. д. К сожалению, действительная процедура имела некоторые недостатки (Kolata, 1986). В большой барабан поместили капсулы, соответствующие каждому из дней года, при этом одновременно опускали все капсулы одного месяца. Первыми опустили январские капсулы, затем февральские и т. д. Барабан прокрутили, чтобы перемешать капсулы, но, по-видимому, не очень тщательно, потому что когда стали выбирать дни, капсулы, опущенные последними, в большинстве своем выбирались первыми. В то время лучше было не отмечать свой день рождения в декабре.

Простая случайная выборка имеет два недостатка. Во-первых, вам может понадобиться отразить в выборке определенные особенности популяции, а во-вторых, процедуру может быть невозможно применить на практике в случае, если популяция очень велика. Сможете ли вы получить список всех людей, проживающих в Соединенных Штатах, если вам потребуется сделать простую случайную выборку североамериканцев? Первую проблему решает расслоенная выборка, а вторую — гнездовая.

**Таблица 4.3**

#### Создание случайной выборки с помощью таблицы случайных чисел

- Задача: Сделайте случайную выборку в 5 человек из популяции в 20 человек  
 Шаг 1: Обозначьте номерами от 01 до 20 всех членов, составляющих популяцию  
 Шаг 2: Перейдите к таблице случайных чисел. Ниже приведена часть такой таблицы:

2	2	1	7	6	8	6	5	8	4	6	8	9	5
1	9	3	6	1	7	5	9	4	6	1	3	7	9
1	6	7	7	2	3	0	2	7	7	0	9	6	1
7	8	0	3	7	6	7	1	6	1	2	0	4	4
0	3	2	8	1	2	2	6	0	8	7	3	3	7

- Шаг 3 Выберите число из таблицы, с которого вы начнете поиск. Оно может быть любым, единственное, чего нельзя делать, — это начинать два отдельных поиска с одного и того же числа. Предположим, вы начинаете с верхнего числа в третьем столбце слева. Это число 1. Вам потребуется выбирать пары чисел, образующие двузначные числа, так как популяция состоит из людей, пронумерованных от 01 до 20. Поэтому в качестве стартовой точки можно использовать число 1 и стоящее за ним число 7
- Шаг 4 Поиск выборки начинается с числа 17, попадающего в промежуток от 01 до 20. Таким образом, первый человек в вашей выборке имеет номер 17
- Шаг 5 Продолжайте поиск вниз по двум столбцам пока не найдете пять человек с номерами от 01 до 20. Ниже опять приведена таблица с разбивкой столбцов для удобства подбора числовых пар. Пять выбранных пар подчеркнуты, а участок таблицы, в котором нужно производить поиск, выделен жирным шрифтом. Стрелки указывают направление поиска:

2	2	<u>1</u>	<u>7</u>	<b>6</b>	<b>8</b>	<b>6</b>	<b>5</b>	<b>8</b>	4	6	8	9	5
1	9	<b>3</b>	<b>6</b>	<b>1</b>	<b>7</b>	<b>5</b>	<b>9</b>	4	6	1	3	7	9
1	6	7	7	2	3	<u>0</u>	<u>2</u>	7	7	0	9	6	1
7	8	<u>0</u>	<u>3</u>	7	6	7	1	<b>6</b>	<b>1</b>	2	0	4	4
0	3	2	8	<b>1</b>	<b>2</b>	2	6	<u>0</u>	<u>8</u>	7	3	3	7

Таким образом, выборка состоит из членов популяции с номерами 17, 3, 12, 2 и 8. Обратите внимание, что номера, превышающие 20 (например, 36), пропускаются и что если номер повторяется дважды, его повторно не выбирают

### Расслоенная выборка

Предположим, вы хотите узнать отношение к абортам в вашем кампусе, а популяция вашего учебного заведения состоит из 5000 человек, из которых 4000 — женщины. Для этого вы решаете отобрать 100 студентов. Если вы используете простую случайную выборку, то, вероятно, в вашей выборке будет больше женщин, чем мужчин, но их соотношение не будет точно таким же, как в популяции. Вашей задачей является создание выборки, действительно отражающей состав популяции, а для такого вопроса, как аборт могут быть важны различия во мнениях между мужчинами и женщинами. Поэтому если в выборке окажется больше мужчин, то она может неверно отобразить отношение к этому вопросу в вашем кампусе. В подобном случае неплохо было бы заранее решить, что если 80% популяции — женщины, то они должны составлять ровно 80% выборки.

**В расслоенной выборке** точно отображается *соотношение* важных подгрупп популяции. В приведенном выше примере из списка женщин случайным образом должны были быть отобраны 80 человек, а из списка мужчин — 20.

Обратите внимание, что в таком случае исследователю требуется решить, на сколько уровней делить популяцию. В случае опроса по поводу аборта были отобраны в соответствии с их реальным соотношением мужчины и женщины. Должны ли быть пропорционально представлены в выборке также все четыре студенческих потока? А как насчет католиков и протестантов? Право- и леворуких? Оче-

видно, что исследователь должен где-то остановиться. Для расслоения выборки одни характеристики (религия) являются более важными, чем другие (право- или леворукость). Исследователь должен разумно подойти к решению этого вопроса в соответствии с условиями исследования или целями экспериментов.

### Кластерная выборка

Метод расслоенной выборки весьма эффективен, но он не решает проблему построения выборки из большой популяции, для которой обычно невозможно получить полный список членов. Эту проблему решает **кластерная выборка**, которую часто используют национальные службы опроса общественного мнения. С помощью этого подхода исследователи случайным образом выбирают группу людей (кластер), имеющих определенную особенность. Используя этот метод, можно проводить опрос в кампусе большого университета. Если исследователь хочет получить репрезентативную выборку студентов, а применить расслоенную выборку невозможно, то альтернативой будет список базовых предметов данного учебного заведения. Студенты разных специальностей, посещающие занятия по одному предмету, образуют группу. Если в вузе преподают 40 базовых предметов, исследователь может выбрать 10 из них, а затем провести опрос всех студентов, посещающих эти курсы.

Если выбранные группы слишком велики, исследователь может выбрать из них более мелкие подгруппы. Предположим, вы хотите узнать, нравится ли студентам жить в многоэтажных общежитиях вашего кампуса, которые вы операционально определили как любые общежития, имеющие восемь этажей и более. Допустим, в вашем кампусе есть 15 таких зданий, в которых проживают в общей сложности **6000** студентов. Воспользовавшись методом гнездовой выборки, сначала можно выбрать шесть зданий (каждое здание = группа), а затем для каждого здания случайным образом выбрать три этажа и включить в выборку всех проживающих на выбранных этажах в выбранных общежитиях (предположим, около 40 человек на этаже). Таким образом, общее число выборки будет 720 человек ( $40 \times 3 \times 6$ ). Обратите внимание, что в данном случае можно также использовать некоторые элементы расслоенной выборки. Если из 15 общежитий 10 женских и 5 мужских, то можно составить группу, отражающую это соотношение: 4 женских общежития и 2 мужских.

### Простая выборка

Прочитав приведенную выше информацию, вы можете решить, что без вероятностной выборки качество исследования снизится. Это действительно так, если целью проекта является точное описание конкретных особенностей некоторой популяции на основании изучения ее части. Но большинство исследований по экспериментальной психологии не таковы и обычно цель исследования заключается в изучении отношений между переменными, например: использование изображений улучшает запоминание; наблюдение агрессии приводит к агрессивному поведению; с увеличением количества свидетелей происшествия снижается число желающих оказать помощь и т. д. Конечно, можно надеяться на вероятность того, что результаты таких исследований будут пригодны не только для людей, участвующих в них, но исследователи просто предполагают, что если выявлена сильная взаимосвязь, то она обязательно проявится у большинства людей, принадлежащих к определенной популяции, вне зависимости от способа их отбора. Будет ли это предположение

истинным, естественно, зависит от стандартных процессов повторения и дополнения эксперимента, рассмотренных в главе 3. Поэтому, например, при изучении границ кратковременной памяти у взрослых нет необходимости делать случайную выборку — подойдет практически любая группа взрослых людей.

### Удобная выборка

Это чаще всего используемый (и действительно наиболее удобный) вид простой выборки. Чтобы создать **удобную выборку**, исследователь просто использует добровольцев из людей, соответствующих основным требованиям исследования. Обычно это первокурсники или второкурсники из «фонда испытуемых» — студенты, занимающиеся общей психологией, которых попросили поучаствовать в одном или нескольких исследованиях (возможно, вам тоже приходилось участвовать в такой работе). Более подробно об этических аспектах «фонда испытуемых» см. вставку 5.2.

Иногда при этом требуется участие определенного типа людей (иногда такая выборка называется «целевой»). Например, когда Стэнли Милгрэм в первый раз набирал участников для исследования явления подчинения, он дал в местную газету объявление о наборе добровольцев. Он намеренно не стал использовать студентов, так как считал, что они могут оказаться «слишком однородной группой... (он) хотел исследовать разных людей, принадлежащих к различным слоям общества» (Milgram, 1974, p. 14). Он мог бы попробовать более сложный подход и сделать расслоенную или кластерную выборку, но не было гарантии, что выбранные им люди согласятся прийти в лабораторию. Кроме факта, что не для всех исследований требуется делать выборки, точно отражающие популяцию, способ, выбранный Милграмом, показывает, что часто имеются прагматические основания для использования простой выборки.

## Статистический анализ

Первое положение хорошо известного каждому гражданину Соединенных Штатов руководства «Неисхоженная дорога» гласит: «Жизнь трудна» (*The Road Less Traveled*, Реск, 1978, p. 15). Верить в это начинают многие студенты, приступающие к изучению статистики (они легко узнают себя на рис. 4.5). Я не стану убеждать вас в том, что занятия статистикой подобны отдыху на пляже во Флориде, но, надеюсь, вы увидите, что часть удовольствия от психологического исследования приносят анализ информации, которую вы так усердно собирали, и выяснение, *действительно* ли исследование показало то, что, как вы думаете, оно показало. Я видел, как взрослые и ответственные люди затаив дыхание ожидали результатов статистического анализа, а затем поникали головой или прыгали от восторга, едва взглянув на волшебные числа. Так что если у вас появилась страсть к психологическим исследованиям, тонкости статистического анализа вызовут у вас немало эмоций.

Моя задача состоит в том, чтобы познакомить вас со статистическим образом мысли и теми видами статистического анализа, с которыми вы встретитесь, проводя психологические исследования. Хорошо, если вы уже изучили курс статистики, но если нет, то вам необходимо сделать это как можно быстрее, особенно если вы хотите поступать в аспирантуру на связанную с психологией специальность. Хотите верить, хотите нет, но в списке называемых аспирантурами дисциплин, ко-

торые обязательно должны изучить поступающие, статистика занимает первое место, а второе — методы исследований (вспомните описанное в главе 1 исследование, проведенное Норкроссом, Ханичем и Терранова в 1996 г.). Я расскажу вам об основных статистических методах, но это не заменит полного курса.



Рис. 4.5. Распространенное представление о статистике

Информацию о том, как выбрать подходящий метод статистического анализа и как проводить некоторые обычные процедуры обработки данных, вы найдете в приложении С. По ходу книги время от времени вам будут также встречаться темы, касающиеся статистики, поскольку разработку психологического исследования невозможно отделить от анализа этого исследования. В этой главе вы узнаете о различиях между описательной статистикой и статистикой вывода, а также о логике проверки гипотез.

### Описательная статистика и статистика вывода

Два самых общих вида статистической обработки данных — это описательная статистика и статистика вывода. Различия между ними соответствуют различию между выборкой и популяцией. Говоря просто, **описательная статистика** обобщает данные, собранные на выборке участников занятых в вашем исследовании, а **статистика вывода** позволяет вам делать такие выводы об этих данных, которые могут быть применены к популяции в целом.

#### Описательная статистика

По сути, методы описательной статистики позволяют вам свести огромное количество чисел, смысл которых невозможно охватить сразу, к очень небольшому набору, значение которого понять гораздо легче. Описательная статистика включает оценку общей тенденции, изменчивости и взаимосвязей, представленных как численно, так и наглядно (в виде графиков). В этой главе мы рассмотрим основные процедуры оценки общей тенденции и изменчивости. Оценка взаимосвязей (вычисление коэффициентов корреляции) будет описана в главе 9.

Для иллюстрации оценки общей тенденции и изменчивости рассмотрим данные гипотетического исследования памяти, в котором 20 человек запоминали, а затем пытались воспроизвести список из 25 слов. Каждое представленное ниже число соответствует количеству слов, запомненных каждым из 20 участников:

16	18	19	19
17	19	15	21
14	16	15	17
17	20	17	15
18	17	18	18

Сразу видно, что обобщение результатов этого исследования требует чего-то большего, чем простой демонстрации набора из 20 чисел. Например, можно попытаться вычислить *типичную* оценку, или так называемую «общую тенденцию». Чаще всего психологи-исследователи определяют общую тенденцию вычисляя среднее арифметическое. Для этого складывают все оценки и делят полученную сумму на общее количество оценок:

$$\bar{X} = \frac{\sum x}{n};$$

где  $X$  = среднее арифметическое;  $\sum X$  = сумма отдельных оценок;  $n$  = количество оценок в примере.

В случае данных, собранных при исследовании памяти, получаем:

$$\bar{X} = \frac{\sum x}{n} = (16 + 17 + 14 + \dots + 18)/20 = 346/20 = 17,30.$$

Два других способа нахождения общей тенденции — это вычисление медианы и моды. **Медиана** представляет собой оценку, находящуюся строго в середине набора оценок. Одна половина оценок выше, а другая — ниже значения медианы. Для определения медианы в первую очередь нужно составить последовательность оценок, от наименьших к наибольшим. В случае данных, собранных при исследовании памяти, последовательность будет следующая:

14 15 15 15 16 16 17 17 17 17 17 18 18 18 18 19 19 19 20 21  
 $\uparrow\uparrow$

Далее нужно определить **местоположение медианы** — позицию в последовательности оценок, где проходит медиана (Howell, 1997). Это вычисляется по формуле:

$$\text{местоположение медианы} = \frac{n+1}{2}.$$

Для данных из исследования памяти местоположение медианы следующее:  $(20+1)/2 = 10,5$ . Это означает, что она лежит посередине (0,5) между 10-ми и 11-ми номерами в последовательности. Считая слева направо, видим, что и 10-й, и 11-й номера — это число 17 (я отметил это место в показанной выше последовательности знаком  $\uparrow\uparrow$ ). Медиана является точной серединой набора оценок: с каждой стороны от нее лежит по 10 чисел.

Иногда медиану используют, если набор оценок содержит одну или две, сильно отличающихся от остальных. В такой ситуации среднее арифметическое дает искаженное представление о типичной оценке. Предположим, к примеру, что пять преподавателей с вашего факультета психологии получили следующие оценки IQ. 93, 81, 81, 95 и 200 (последняя оценка вероятно принадлежит преподавателю методов исследований). Среднее арифметическое оценок IQ, равное 110 (вы можете проверить), дает ложное представление о том, что в целом преподаватели психологического факультета имеют умственные способности заметно выше среднего. Медиана в данном случае позволяет лучше оценить типичную IQ-оценку. Местоположение медианы равно  $(5+1)/2 = 3$ , а в последовательности оценок третье число равно 93:

81 81 93 95 200  
          1\

Очевидно, что медиана оценок IQ, равная 93, гораздо лучше отражает обычный уровень интеллектуальных способностей на данном гипотетическом факультете психологии.

**Мода** — это значение, чаще всего встречающееся в наборе оценок. В приведенном выше примере значение моды равно 81. Мода гипотетических оценок теста памяти равна медиане: число 17 встречается 5 раз, т. е. чаще всех других чисел. Так как в данных теста памяти отсутствуют необычно высокие или низкие оценки, значения среднего арифметического (17,3), медианы (17) и моды (17) довольно близки друг другу, и каждое из них дает верное представление об общей тенденции.

Очевидно, что оценка общей тенденции требует суммирования данных. Менее очевидна, но не менее важна необходимость анализа изменчивости набора оценок. Предположим, вы — гольфер-профессионал и собираетесь вести занятия в местном клубе для двух групп: в 8:00 и 9:00. Вы измерили их способности, определив среднюю оценку для 9 лунок. Ниже приведены полученные вами данные:

Группа, занимающаяся в 8:00:	50	52	58	46	54
Группа, занимающаяся в 9:00:	36	62	50	72	40

Обратите внимание, что среднее арифметическое для каждого набора оценок гольферов равняется  $260/5 = 52$  ударам. Профессионалу будет о чем поговорить с каждым членом обеих групп. В группе, занимающейся в 8:00, оценки близки друг к другу и все ее участники имеют примерно одинаковый уровень способностей, однако вторая группа не настолько благополучна — оценки в ней варьируются от 36 (довольно хорошо) до 72 (ай-ай-ай!). Понятно, что перед началом занятий гольфер-профессионал предпочел бы знать не только среднюю оценку группы.

Самый простой и весьма приблизительный способ оценить изменчивость — это найти **разброс** — разницу между наибольшей и наименьшей оценками в группе. Диапазон данных для приведенного ранее теста памяти равен 7 (21-14). Разброс оценок 8-часовой группы в примере с занятиями гольфом равен 12 (58 - 46), а разброс оценок 9-часовой — 36 (72 - 36). Разброс дает грубую оценку изменчивости и показывает лишь разницу между крайними значениями. Более сложный способ измерения изменчивости — нахождение стандартного отклонения. Этот способ чаще всего применяется при создании сводного отчета о собранных данных.

**Стандартное отклонение** для выбранного набора оценок — это среднее значение, на которое оценки данного распределения отклоняются от среднего арифметического этих оценок. В табл. 4.4 показаны два способа вычисления стандартного отклонения. Первый способ вытекает непосредственно из определения и позволяет лучше понять сущность стандартного отклонения. Второй — это вычислительная формула, использовать которую удобнее при работе с калькулятором. Для примера с гипотетическим исследованием памяти стандартное отклонение равняется 1,81 словам. Для примера с занятиями гольфом стандартное отклонение для 8-часовой группы равняется 4,47, а для 9-часовой — 15,03 ударам.

Таблица 4,4

### Вычисление стандартного отклонения

Если вы пользуетесь пакетом статистических программ *SPSS* или *SAS*, практически любой анализ будет включать нахождение стандартного отклонения. Кроме того, большинство калькуляторов способны к вычислению основных статистических функций, в том числе стандартного отклонения. Поэтому вам, может быть, не придется подсчитывать на бумаге. Однако у вас может оказаться устаревший калькулятор или увлеченный преподаватель, который будет разделять мои представления о том, что подсчет вручную дает более глубокое понимание происходящего. Есть два способа вычислить стандартное отклонение. Первый — использовать так называемую «формулу отклонения». Внимательно изучив ее, вы лучше поймете суть стандартного отклонения, которое по определению является приблизительной характеристикой средней величины отклонения каждой оценки от среднего арифметического. Ниже показано, как найти стандартное отклонение для 20 оценок, полученных при исследовании памяти.

Шаг 1 Вычислите среднее арифметическое:

$$\bar{X} = \frac{\sum X}{n} = 17,3.$$

Шаг 2 Вычислите оценки отклонения, каждую возведите в квадрат и найдите их сумму. Оценки отклонения ( $x$  малое) находятся вычитанием среднего арифметического из каждой оценки ( $X$  большое). Таким образом,  $x = X - \bar{X}$ . Возведение в квадрат предотвращает появление отрицательных чисел:

$X$	$\bar{X}$	$x$	$x^2$
16	17,3	-1,3	1,69
17	17,3	-0,3	0,09
14	17,3	-3,3	10,89
...	...	...	...
18	17,3	-0,7	0,49
			$\sum x^2 = 62,20$

Шаг 3 Вычислите стандартное отклонение (CO):

$$CO = \sqrt{\frac{\sum x^2}{n-1}} = \sqrt{\frac{62,2}{20-1}} = \sqrt{3,27} = 1,81.$$

По формуле отклонения найти стандартное отклонение довольно просто, но для калькулятора она не совсем подходит. Более простой способ — использовать так называемую формулу для вычислений, которая математически равнозначна формуле отклонения. Она имеет следующий вид]\_\_\_\_\_

$$CO = \sqrt{\frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n-1}},$$

а вычисления проходят следующим образом:

Шаг 1 Вычислите  $\sum X^2$  и  $(\sum X)^2$ ;

$$\sum X^2 = 16^2 + 17^2 + 14^2 + \dots + 18^2 = 256 + 289 + 196 + \dots + 324 = \mathbf{6048};$$

$$(\sum X)^2 = (16+17+14+\dots+18)^2 = 346^2 = \mathbf{119716}.$$

Шаг 2  $\sqrt{\frac{\sum X^2 - (\sum X)^2}{n}} = 6048 - \frac{119716}{20} = 6048 - 5985,8 = \mathbf{62,2}.$

Шаг 3 Разделите значение, полученное на шаге 2, на  $(n-1)$ :

$$\frac{62,2}{19} = \mathbf{3,27}.$$

Шаг 4 Чтобы получить стандартное отклонение, найдите квадратный корень значения, полученного на шаге 3:

$$CO = \sqrt{3,27};$$

$$CO = \mathbf{1,81}.$$

Одной из характеристик изменчивости является **дисперсия**. Дисперсия представляет собой число, получаемое в ходе вычисления стандартного отклонения, сразу перед нахождением квадратного корня (3,27 для оценок исследования памяти). Это число редко попадает в отчеты, включающие описание данных, так как оно отражает измеряемую величину, возведенную в квадрат (например, «количество запомненных слов в квадрате»). Однако оно находится в центре вероятно самой известной в психологии процедуры статистики вывода — «дисперсионного анализа». О нем рассказывается в главах 7 и 8, а также более подробно в приложении С.

Общая тенденция и изменчивость — это универсальные характеристики, используемые при любом описании данных, но исследователи также изучают и весь набор оценок в целом. Простой просмотр данных малоэффективен, но есть и другие способы организации оценок, с помощью которых можно получить значимую картину результатов. Один из способов представления данных — это **гистограмма**. Гистограмма представляет собой график, показывающий, сколько раз встречается каждая оценка в данном наборе, или, при большом количестве оценок, частоту появления оценок в пределах определенного интервала. Чтобы построить гистограмму, необходимо предварительно построить **частотное распределение** — таблицу, в которой указывается, сколько раз встречается каждая оценка. Частотное распределение оценок, полученных при исследовании памяти, имеет следующий вид:

Оценка	Частота	Частота, обозначенная звездочками
14	1	*
15	3	***
16	2	**

17	5	*****
18	4	****
19	3	***
20	1	*
21	1	*

Построив таблицу частотного распределения, несложно начертить гистограмму. На оси  $X$  графика отметьте сами оценки, а на оси  $Y$ — частоту их появления, а затем постройте соответствующие столбцы графика. Результат должен выглядеть, как показано на рис. 4.6. Обратите внимание, что если взять столбец со звездочками из частотного распределения и повернуть его на  $90^\circ$ , результат будет такой же, как на рис. 4.6.

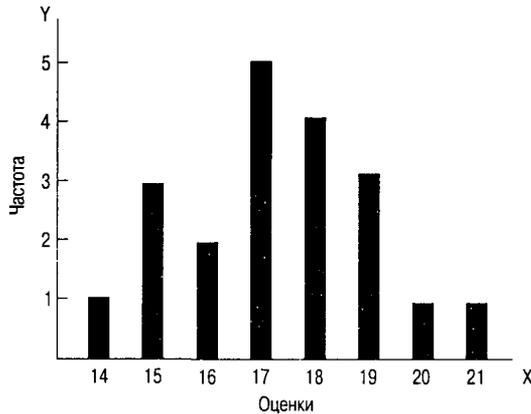


Рис. 4.6. Гистограмма оценок, полученных по тесту памяти

Также следует отметить, что гистограмма выступает вверх в районе середины и уплощается по краям, что приблизительно соответствует распределению оценок для целой популяции, а не только для 20 человек из описанного выше примера. Распределение оценок для популяции представляет собой известную колокообразную кривую, называемую **нормальной кривой**, или нормальным распределением. Вы уже встречались с ней; она представлена на рис. 4.7.

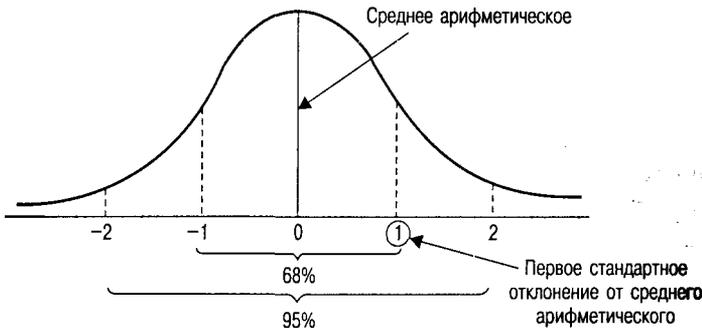


Рис. 4.7. Нормальная кривая

Так же как кривая, построенная для оценок исследования памяти, нормальная кривая представляет собой частотное распределение. Но в отличие от первой она является *нереальным* (или «эмпирическим») распределением оценок конкретной выборки, а *гипотетическим* (или «теоретическим») распределением оценок, которые могут получить члены популяции, если все они примут участие в исследовании. Среднее арифметическое, медиана и мода находятся точно в центре нормального распределения. Важнейшая особенность статистического анализа частотного распределения заключается в том, что если эмпирическое распределение оценок сходно с нормальным распределением, то математические характеристики последнего можно использовать для построения выводов о первом.

Обратите внимание, что на нормальной кривой, показанной на рис. 4.7, я отметил по два стандартных отклонения с обеих сторон от среднего арифметического. Математические характеристики кривой таковы, что около 68% всех оценок для популяции лежат в интервале между двумя первыми стандартными отклонениями, а около 95% — между вторыми. Очевидно, что оценок, попавших за пределы вторых стандартных отклонений, немного — всего 5% от общего количества. Все эти явления можно назвать «статистически значимыми». Запомните данные характеристики распределения, мы к ним очень скоро вернемся.

Кроме частотного распределения и гистограммы есть еще один способ отображения набора данных, который позволяет выявить их особенности. Это **метод стебля и листа** (Turkey, 1977). Чаще всего его используют, когда набор оценок так велик, что частотное распределение или гистограмма были бы очень громоздкими. Например, если вы протестировали 20 испытуемых на застенчивость и полученные ими оценки варьируются от 10 до 70, простое частотное распределение, подобное построенному для данных исследования памяти, будет огромным, а ось  $X$  гистограммы будет в милю длиной. Проблему можно решить сгруппировав данные по интервалам (10-19, 20-29, 30-39 и т. д.). Каждый столбец диаграммы будет отразит количество оценок в пределах определенного интервала. Обратите внимание, что подобная группировка данных приводит к потере некоторой информации. Если шесть человек при тестировании на застенчивость получают оценки между 30 и 39, то все, что вы увидите после такого обобщения, — это один столбец, отображающий частоту, равную шести, и вы не будете знать, какую оценку получил каждый из шести участников. Организовав данные методом стебля и листа, вы сможете получить эту информацию. Метод состоит в следующем. Предположим, что при тестировании на застенчивость 20 человек получены следующие оценки (я выделил жирным шрифтом шесть оценок в пределах от 30 до 39):

49	22	33	46
36	64	39	41
41	68	43	61
36	47	32	49
43	67	37	43

В методе стебля и листа с двухзначными числами «листом» будет наименьший разряд (разряд единиц), а «стеблем» — наибольший (разряд десятков). Таким образом, для первого числа (49), стеблем будет число 4, а листом число 9. Для числа 36 стебель равен 3, а лист — 6. Для организации стеблей и листов по одноименному методу сначала требуется расположить числа в порядке возрастания, как вы делали при нахождении медианы (числа от 30 до 39 выделены жирным шрифтом). Получаем:

22 32 33 36 36 37 39 41 41 43 43 43 46 47 49 49 61 64 67 68

Далее поместите стебли в левый столбец таблицы, а листья в соответствующие ряды правого столбца, как показано ниже:

Стебли	Листы
2	2
3	236679
4	113336799
5	
6	1478

Повернув таблицу влево на  $90^\circ$  и представив, как заполняются цветом цифры листов, образуя столбцы, вы получите аналог гистограммы для сгруппированных данных. Но обратите внимание, что по сравнению с обычной гистограммой метод стебля и листа обладает заметным преимуществом. На гистограмме, к примеру, в интервале 30-39 будет изображен один столбец, достигающий по шкале У отметки 6. В таблице, построенной методом стебля и листа, вы не только увидите «высоту» оценок в интервале, но также сможете изучать сами оценки. Кроме того, метод стебля и листа позволяет обнаружить оценки, относительно далеко отстоящие от остальных. В приведенном выше примере отсутствие оценок в интервале 50-59 сразу заметно, а четыре оценки в интервале 60-69 выделяются и несколько отстоят от остальных.

В статьях, посвященных результатам исследований, полученных с помощью описательной статистики, встречается три способа представления данных. Во-первых, если необходимо представить лишь несколько чисел (например, значения среднего арифметического и стандартного отклонения для двух экспериментальных групп), можно использовать повествовательное изложение результатов. Во-вторых, значения среднего арифметического и стандартного отклонения можно представить в виде таблицы, а в третьих — наглядно в виде графика. Как строить таблицы и графики, соответствующие стандартам *APA*, вы узнаете из приложения *A*, в котором приведен пример отчета об исследовании. Также, некоторую информацию о построении графиков можно найти в главах 7 и 8. Этический аспект статистического анализа и построения графиков освещается во вставке 4.3.

## ВСТАВКА 4.3

**Этика — обман с помощью статистики**

Все мы были жертвами использования статистической информации для обмана. При этом первыми в ряду жуликов идут политики, а авторы рекламы аспирина занимают близкое к ним второе место. Но и вообще статистикой злоупотребляли достаточно часто для того, чтобы большинство людей стали к ней относиться весьма скептически. Часто можно услышать: «Статистика покажет все, что вы захотите». Действительно ли это так?

Конечно, в каждом случае приходится решать специально, каким образом представлять данные, какие выводы из них сделать. Можно быть вполне уверенным в том, что статьи, которые публикуются в солидных журналах, проверяются и рецензируются специалистами и содержат выводы, основанные на надежных статистических данных. Внимательно следует относиться к статистике, если она используется в более массовых источниках или если ее применяют люди, очевидно стремящиеся убедить вас в истинности собственных убеждений.

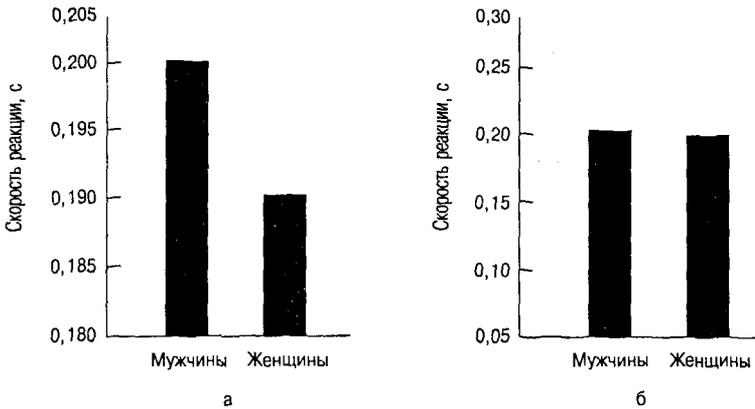
Знание о том, как правильно применять статистику, поможет вам распознать сомнительные случаи ее использования. Впервые некоторые моменты, заслуживающие особого внимания, были описаны в прекрасной работе Даррелла Хаффа «Как обмануть с помощью статистики», опубликованной в 1954 г. Работа начинается с цитаты, приписываемой британскому политику Бенджамину Дизраели и выражающей общее мнение по поводу статистики: «Существуют ложь, чудовищная ложь и статистика» (р. 2).

В данной книге мы приведем пример обмана с помощью статистики, в котором был использован способ наглядного представления данных- графики. Предположим, вы читаете об исследовании, в котором сравнивается скорость реакции у мужчин и женщин, и встречаете следующий график (рис. 4.8).



**Рис. 4.8.** Гистограмма с неразмеченной осью Y, отображающая предполагаемое различие скоростей реакции у мужчин и женщин

Такой вид графика Хафф назвал «Ого-графиком» по той очевидной причине, что, когда вы смотрите на него в первый раз, вы восклицаете «Ого! Какая огромная разница!» Однако дело в том, что это различие сильно преувеличено манипуляциями с осью Y. Обратите внимание, что она не размечена. Если бы она была размечена, деления выглядели бы, как показано слева на рис. 4.9. Разметка оси Y, приведенная на графике справа, гораздо реалистичнее - она дает более верную картину отсутствия значимых различий. Это, скорее всего, подтвердит и статистический анализ.



**Рис. 4.9.** Гистограмма, показывающая (а) повторение рис. 4.8, но с размеченной осью /и (б) те же данные, что и на рис. 4.8, но с иначе размеченной осью У

Мораль ясна: следите за осью У. Особенно скептически следует относиться к графикам, показывающим большое различие по оси У, при этом не размеченной, размеченной неясно или с очень малой ценой деления.

### Статистика вывода

Как и большинству людей, исследователям необходимо, чтобы их любили и ими восхищались окружающие. Добиться этого можно, получив интересные результаты исследования, которые можно применить не только к данным, собранным в ходе экспериментов. Это означает, что хотя в исследовании рассматривается лишь небольшая *выборка* данных, которые возможно собрать, исследователи надеются, что сделанные ими выводы можно будет применить к более широкой *популяции*. Ведь задача исследовательской работы в целом и заключается в открытии общих законов поведения.

Чтобы проиллюстрировать применение статистики вывода, рассмотрим гипотетическое исследование, в котором сравнивается поведение в лабиринте двух групп крыс: одних кормят сразу после прохождения лабиринта, а других через 10 секунд. Эмпирический вопрос: способствует ли немедленное подкрепление запоминанию лабиринта? Предположим, что при исследовании 10 крыс (по 5 в каждой группе) получены следующие результаты. Каждая оценка означает количество попыток, потребовавшихся для запоминания лабиринта. Запоминание операционально определено как безошибочное прохождение лабиринта.

№ крысы	Немедленное кормление	№ крысы	Задержанное кормление
1	12	6	19
2	13	7	20
3	16	8	16
4	11	9	18
5	14	10	15

Обратите внимание, что оценки в каждой колонке не совсем одинаковы — это результат небольших различий между пятью крысами из каждой группы и, возможно, других случайных факторов. Однако, несмотря на отсутствие полного сходства, видно, что получавшие немедленное подкрепление крысы запомнили лабиринт быстрее (т. е. за меньшее количество попыток).

Но общего впечатления от чисел недостаточно. Первый шаг анализа — вычисление таких величин описательной статистики, как среднее арифметическое и стандартное отклонение:

	Немедленное подкрепление	Отсроченное подкрепление
Среднее арифметическое	13,2	17,6
Стандартное отклонение	1,9	2,1

В среднем, по крайней мере в этом примере, на запоминание лабиринта требуется больше попыток, если пищевое вознаграждение задерживается. Так же как показывает стандартное отклонение, изменчивость оценок в каждом наборе довольно низкая и практически одинакова для обеих групп. Можем ли мы заключить, что немедленное подкрепление *в целом* ускоряет запоминание лабиринта? Пока нет. Необходим заключительный анализ данных, в данном случае включающий проверку гипотезы<sup>1</sup>.

### Проверка гипотезы

Для проверки гипотезы сначала необходимо сделать предположение о том, что разница между двумя изучаемыми условиями не сказывается на выполнении заданий, в данном случае разница между немедленным и отсроченным вознаграждением. Это предположение называется **нулевой гипотезой** (нуль = ничто), обозначается  $H_0$  (читается «аш нулевое»). Гипотеза исследования (меньшее количество попыток у крыс, получающих немедленное подкрепление), тот результат, который вы *надеетесь* получить, чтобы завоевать друзей и оказывать влияние на людей, называется **альтернативной гипотезой**, или  $H_1$ . Проводя исследование, вы будете стараться опровергнуть  $H_0$  и соответственно подтвердить (но не доказать)  $H_1$  — близкую вашему сердцу гипотезу.

Если эти слова звучат для вас странно, представьте, что дело происходит в суде. Обвиняемый человек считается невиновным, т. е. сделано предположение, что он ничего не совершал (нуль). Задача обвинения состоит в том, чтобы убедить суд в альтернативной гипотезе, а именно в том, что обвиняемый совершил преступление. Как и прокурор, исследователь должен показать, что нечто в действительности имеет место, а именно что в рассматриваемом случае задержка подкрепления влияет на запоминание лабиринта.

Заключительный анализ может привести к одному из двух результатов. Обнаруженные вами различия в поведении двух групп крыс могут быть вызваны настоящими, действительными и несомненными причинами или быть игрой случая. Это

<sup>1</sup> Вторая категория заключительного анализа называется «оценка». Она представляет собой оценивание показателей популяции на основании оценок отдельной выборки.

означает, что различия в пределах выборки могут отражать действительные явления, но могут и не делать этого, и соответственно результатом заключительного статистического анализа будет либо отказ от гипотезы  $H_0$ , либо ее подтверждение. Невозможность отвергнуть гипотезу  $H_0$ , означает, что все найденные вами различия (а разница между группами почти всегда обнаруживается в ходе исследования), вероятнее всего, вызваны случайностью, ведь вам не удалось найти действительной закономерности, которую можно обобщить для случаев, выходящих за пределы выборки. Отказ от  $H_0$  значит, что вы считаете, что некое явление действительно имело место в вашем исследовании и его результаты можно обобщить. В примере с лабиринтом отказаться от  $H_0$  значит найти статистически значимые различия, свидетельствующие о существовании общей закономерности: немедленное подкрепление способствует запоминанию лабиринта.

Абсолютную истинность гипотезы исследования ( $H_1$ ) доказать невозможно, так же как нельзя *целиком и полностью* доказать виновность подсудимого: считается, что вина доказана тогда, когда не возникает обоснованных сомнений. Таким образом,  $H_0$  может быть отвергнута (и в то же время может быть подтверждена  $H_1$ ) с определенной степенью уверенности, описываемой как значение альфа ( $\alpha$ ). Формально альфа означает вероятность получения определенных результатов, если  $H_0$  истинна. Условно альфа принимается равной 0,05 ( $\alpha = 0,05$ ), но можно задать также и другие значения альфа (например,  $\alpha = 0,01$ ). Отказ от  $H_0$  при значении альфа, равном 0,05, означает, что вы считаете вероятность того, что результаты вашего исследования вызваны случайными факторами, очень низкой (5 из 100). Если они не являются игрой случая, то должна быть какая-то другая причина, а именно (как вы надеетесь) изучаемое вами явление, в данном случае задержка подкрепления.

Выбор значения 0,05 связан с особенностями нормальной кривой, которые рассматривались выше. Вспомните, что для нормального распределения оценок вероятность того, что конкретная оценка будет лежать на расстоянии, превышающем два стандартных отклонения от среднего арифметического по оси  $X$ , довольно низкая — 5% или меньше. Такое случается редко. Аналогично при сравнении двух наборов оценок, как в случае с лабиринтом, нас интересует вероятность обнаружения различий между значениями среднего арифметического, если в действительности никаких различий не существует (т. е. если гипотеза  $H_0$  истинна). Если вероятность достаточно низкая, мы отвергаем  $H_0$  и считаем, что обнаружены действительные различия. «Достаточно низкая» — это вероятность, равная 5% ,или 0,05. Другими словами, если  $H_0$  истинна, то обнаружение различий между значениями среднего арифметического настолько маловероятно (редкое событие), что мы просто не можем поверить, что  $H_0$  истинна. Мы считаем, что произошло что-то другое (т. е. задержка подкрепления действительно снижает скорость запоминания лабиринта), а поэтому отвергаем  $H_0$  и заключаем, что между группами существуют «статистически значимые» различия.

### Ошибки 1 -го и 2-го рода

Из предыдущего примера ясно, что мы можем неверно решить, нужно отвергать или нет  $H_0$ . В действительности есть два вида таких ошибок. Во-первых, можно отвергнуть  $H_0$  и считать подтвержденным  $H_1$ , радуясь тому, что сделано новое

значительное открытие, и ошибиться. Отказ от  $H_0$ , когда она по сути истинна, называется **ошибкой 1-го рода**. Вероятность такого события равна значению альфа, т. е. обычно 0,05. Таким образом, задание альфа значения 0,05 означает, что вероятность сделать ошибку 1-го рода, т. е. решить, что мы наблюдали некоторое явление и при этом ошибиться, — 5%. Ошибку 1-го рода можно заподозрить, если несколько раз подряд не удастся получить нужные результаты при воспроизведении исследования.

Другой вид ошибки называется **ошибкой 2-го рода**. Она возникает, если вы не отвергаете  $H_0$  и ошибаетесь. Это значит, что в ходе исследования вы не обнаружили то, что ожидали, расстроились из-за этого, но при этом ошиблись, поскольку в действительности данное явление наблюдается в популяции, но вы просто не смогли обнаружить его на изученной выборке. Ошибка 2-го рода нередко совершается, если проводимые измерения недостаточно надежны или точны для того, чтобы показать различия между выполнением задания при разных условиях. Как вы узнаете из главы 10, это часто случается в исследованиях по оценке программ. Проводимая программа может иметь значимое, но небольшое влияние на испытуемых, и поэтому измерения будут недостаточно тонкими, чтобы показать его.

В табл. 4.5 показано четыре возможных результата заключительного статистического анализа данных исследования, в котором сравниваются два экспериментальных условия. Несложно заметить, что правильное решение принимается, если  $H_0$  отвергается, когда она ложна, и не отвергается, когда она истинна. Если  $H_0$  ошибочно отвергается, то происходит ошибка 1-го рода, а если ошибочно не отвергается — ошибка 2-го рода. Если это поможет вам понять терминологию, в табл. 4.5 можно сделать следующие замены:

« $H_0$  не отвергнута» означает:

- вы провели исследование, сделали все необходимые анализы и получили ноль, ничто, отсутствие значимых различий. У вас действительно есть повод для расстройства, особенно если это ваш дипломный проект!

« $H_0$  отвергнута» означает:

- вы провели исследование, сделали все необходимые анализы и обнаружили значимые различия при значении альфа, равном 0,05. Жизнь наполнилась смыслом, вы удивите ваших друзей, а особенно вашего научного руководителя тем, что выполнили работу и *действительно что-то открыли!*

« $H_0$  истинна» означает:

- «несмотря на результаты вашего исследования, действительных различий не существует.

« $H_0$  ложна» означает:

- несмотря на результаты вашего исследования, действительные различия существуют.

Таблица 4.5

Получение статистического ответа: четыре возможных результата исследования со сравнением двух условий,  $X$  и  $Y$

Действительное положение дел

		Действительное положение дел	
		$H_0$ истинна: различий между $X$ и $Y$ не существует	$H_0$ ложна: различия между $X$ и $Y$ действительно существуют
Ваше статистическое решение	$H_0$ не отвергнута: в своем исследовании я не обнаружил значимых различий между $X$ и $Y$ , поэтому не отвергаю $H_0$	Правильное решение	Ошибка 2-го рода
	$H_0$ отвергнута: в своем исследовании я обнаружил значимые различия между $X$ и $Y$ , поэтому отвергаю $H_0$	Ошибка 1-го рода	Правильное решение  (успех экспериментатора)

С учетом замен, названных выше, правильное решение означает, что а) действительных различий не существует и вы их не обнаружили или б) действительные различия существуют и вы их обнаружили (успех экспериментатора). Ошибка 1-го рода значит, что действительных различий нет, но на основании результатов исследования вы считаете, что они есть. Ошибка 2-го рода означает, что различия есть, но, проводя исследование, вы их не обнаружили.

### Заключительный анализ

Принятие решения о том, отвергать ли нулевую гипотезу в исследовании, подобном эксперименту с научением в лабиринте, зависит от анализа двух основных видов изменчивости данных. Первый из них связан с различиями между «количеством попыток, требуемых для достижения критериальных» оценок для двух групп крыс. Эти различия вызываются комбинацией а) систематической дисперсии и б) дисперсии ошибки. **Систематическая дисперсия** является результатом влияния конкретных факторов — либо изучаемой переменной (задержка подкрепления), либо факторов, которые не удастся правильно проконтролировать<sup>1</sup>. **Дисперсия ошибки** — это несистематическая изменчивость, вызванная индивидуальными различиями между крысами в двух группах и любыми случайными, непредсказуемыми факторами, которые могут появиться в ходе исследования. Также индивидуальные различия и другие случайные факторы вызывают дисперсию ошибки в каждой группе и объясняют обнаруженные различия. Многие виды заключительного анализа данных включают вычисление следующего математического отношения:

<sup>1</sup> Эти неконтролируемые факторы называются «осложнителями». Более подробно они будут рассматриваться в следующей главе.

Статистика вывода = Изменчивость оценок при нескольких данных условиях (систематическая дисперсия + дисперсия ошибки)/Изменчивость оценок в пределах одного условия (дисперсия ошибки).

В идеале изменчивость, относящаяся к нескольким условиям, должна быть огромна, а изменчивость по отношению к одному условию относительно небольшой.

Очевидно, что именно такой результат получен для гипотетических данных из эксперимента с лабиринтом. Различия, обнаруженные между результатом двух разных условий, весьма основательны — крысам с отсроченным подкреплением требуется больше времени на изучение лабиринта (17,6 > 13,2), а оценки крыс в каждой группе довольно близки друг другу, что отражается небольшими значениями стандартного отклонения (2,1 и 1,9). В таком случае вы, вероятно, воспользуетесь особым видом анализа, называемым «проверка по критерию Стьюдента для независимых групп», — процедурой, известной вам из вводного курса статистики. В приложении С описано, как проводить такую проверку.

### Мощность и сила эффекта

Все исследователи стремятся опровергнуть нулевую гипотезу, если она действительно ложна («успех экспериментатора» в табл. 4.5). Вероятность такого события описывается как **мощность** статистической проверки. На мощность влияет значение альфа, величина эффекта экспериментального воздействия, и в особенности величина выборки. Последняя характеристика находится под непосредственным контролем экспериментатора, и поэтому исследователи иногда проводят «анализ мощности» в самом начале исследования, чтобы определить подходящий размер выборки. Студенты нередко очень расстраиваются из-за того, что у них «не получается» исследование (т. е. не найдено значимых различий). Такой результат нередко является следствием небольшого размера выборки. В таком случае мощность невелика и, возможно, это спровоцирует появление ошибки 2-го рода, когда нечто действительно могло произойти в ходе исследования, но обнаружить это не удалось. В целом, ошибка 2-го рода и мощность обратно пропорциональны. Чем выше мощность, тем меньше вероятность, что будет сделана ошибка 2-го рода, и наоборот. С другой стороны, огромный размер выборки может привести к получению результата статистически значимого, но относительно бесполезного. Так, очень небольшие, но статистически значимые различия между группами могут почти не иметь практического значения для исследования с большим количеством участников<sup>1</sup>. Например, при достаточно большом количестве крыс в каждой группе можно провести исследование с лабиринтом, обнаружить разницу в скорости его прохождения, равную 1 секунде, и считать, что эта разница значима при альфе, равной 0,05. Однако различие это не будет иметь практически никакого значения. С другой стороны, если величина эффекта, который имеет экспериментальное воздей-

<sup>1</sup> Обратите внимание на слово «могут». Также возможно, что небольшие различия будут иметь большое практическое значение. Например, Розенталь и Росноу (Rosenthal and Rosnow, 1991) обнаружили взаимосвязь между приемом аспирина и возможностью избежать инфаркта. Влияние весьма слабое — прием аспирина снижает вероятность инфаркта приблизительно на 3%, но для группы риска, которая составляет около 1 миллиона людей, 3% означает снижение общего числа инфарктов в год на 30 000.

ствие, при небольшом размере выборки невелика, то значение этого факта может оказаться заметным.

Эта закономерность привела к тому, что исследователи стали уделять больше внимания **величине эффекта**, т. е. изменчивости зависимой переменной, чем независимой переменной (Cohen, 1988). Величина эффекта связана с относительной величиной различий между условиями эксперимента и размером выборки. В последние годы было разработано несколько «показателей величины эффекта», один из них (*d* по Коэну) представлен в приложении С. Важно отметить, что проверку гипотезы и определение величины эффекта можно рассматривать как взаимно дополнительные процедуры. Проверка гипотезы показывает, что экспериментальное воздействие имело определенный надежный эффект, а величина эффекта говорит о его величине. Другими словами, проверка гипотезы отвечает на вопрос, требующий ответа «да» или «нет», а определение величины эффекта — на вопрос «сколько?»

Познакомившись с основными инструментами работы с данными, вы готовы начать изучение трех первых глав об экспериментальном методе — самом мощном из орудий, помогающем понять сложные механизмы поведения и тонкости психических процессов. Мы начнем с введения в экспериментальный метод, рассмотрим проблемы, связанные с контролем за таким исследованием, и изучим особенности наиболее распространенных экспериментальных планов.

## Резюме

### Что измерять: разнообразие поведенческих актов

В психологии изучаются различные стороны поведения — от непосредственно наблюдаемых действий до результатов самоотчетов и показателей физиологической активности. Какие измерения проводить в ходе конкретного исследования, зависит от способа операционального определения конструктов, используемых в данном исследовании. Во многих областях психологии со временем были разработаны стандартные измерения (например, скорость реакции).

### Оценочные измерения

Качественные измерения характеристик поведения являются одновременно надежными и валидными. Надежность означает возможность воспроизведения и малое количество ошибок измерений. Измерения валидны, если действительно измеряется то, что предполагается. Уверенность в валидности растет, если измерение разумно (очевидная валидность) и на его основании можно предсказать будущие события (критериальная валидность). Конструктивной валидностью обладают исследования, в ходе которых изучаются отношения между измеряемым конструктом и связанным с ним явлением и которые дают стойкие, предсказуемые результаты.

### Виды шкал измерений

Данные, собираемые в ходе психологических исследований, можно разделить на четыре группы в зависимости от используемой шкалы измерений: номинальной, порядковой, интервальной и шкалы отношений. В случае номинальной шкалы деления представляют собой ярлыки категорий, а изучается частотность оценок в

пределах каждой категории. Порядковая шкала используется, если изучаемые явления располагаются в ряд в определенном порядке. Интервальная шкала и шкала отношений имеют равные промежутки между делениями, но точку истинного нуля имеет только шкала отношений. Традиционно психологи предпочитают использовать интервальную шкалу и шкалу отношений, так как они позволяют применить большое количество видов статистического анализа.

### **«Кого» измерять: выборки**

Если задача исследования состоит в изучении особенностей конкретной популяции, то сделанная для этого выборка должна быть репрезентативной и строиться как один из вариантов вероятностной выборки. При простой случайной выборке каждый член популяции имеет равные шансы быть выбранным. При расслоенной выборке важные подгруппы популяции представлены пропорционально. Кластерная выборка используется, когда нет возможности получить информацию о всех членах популяции. Большинство психологических исследований имеет своей целью не описание особенностей популяции, а выявление отношений между переменными, поэтому нередко наиболее подходящим является использование простой выборки. В большинстве психологических исследований используется удобная выборка, представляющая собой разновидность простой выборки.

### **Статистический анализ**

В психологии статистический анализ является важнейшим инструментом, позволяющим понять значение результатов исследования. Описательная статистика применяется для обобщения результатов, полученных участниками конкретного исследования, и включает оценку общей тенденции (например, нахождение среднего арифметического, медианы и моды) и изменчивости (нахождение разброса, стандартного отклонения и дисперсии). Данные могут быть наглядно представлены в виде графиков (например, гистограмм или методом стебля и листа). Статистика вывода позволяет решить, являются ли результаты исследования случайностью или же они отражают действительную взаимосвязь явлений, которую можно распространить на более широкую популяцию. Целью заключительного анализа является отклонение гипотезы об отсутствии различий (т. е. нулевой гипотезы), если они в действительности существуют. Ошибка 1-го рода происходит, если нулевая гипотеза не должна быть отвергнута, но отвергается, а ошибка 2-го рода — когда экспериментальное воздействие дает определенный эффект, но в ходе исследования его не обнаруживают. Значение результатов исследования оценивают на основе информации о величине эффекта.

## **Задания для повторения**

### **Выбор ответа**

1. В XIX в. под влиянием экспериментов, посвященных психической хронометрии, было разработано учение о скорости реакции. В одном из таких исследований скорость простой сенсомоторной реакции (ПСР) составила

0,24 с, а скорость реакции выбора (РВ) — 0,29 с. Какой вывод был сделан о психическом явлении дифференциации?

- 1) Ее длительность составила 0,29 с;
  - 2) она равна  $0,29 + 0,24$  с, или 0,53 с;
  - 3) она составляет  $0,29 - 0,24$  с, или 0,05 с;
  - 4) его длительность нельзя вычислить, так как  $СДР > СПР$ .
2. Какое из следующих утверждений о надежности измерений верно?
- 1) измерение надежно, если оно представляется разумным по отношению к определенному конструкту (например, решение задач кажется разумным способом измерения уровня интеллекта);
  - 2) если измерение конструкта надежно, то при повторном измерении будут получены аналогичные результаты;
  - 3) чем больше ошибка измерений, тем выше надежность;
  - 4) результаты теста на надежность френологических измерений черепа будут низкими.
3. Исследователь провел измерения застенчивости и обнаружил, что студенты, получившие высокие оценки, также имеют высокий показатель интроверсии по известной шкале интроверсии—экстраверсии. Такие результаты *лучше всего* иллюстрируют:
- 1) очевидную валидность;
  - 2) критериальную валидность;
  - 3) надежность;
  - 4) конструктивную валидность.
4. Каковы преимущества кластерной выборки по сравнению с простой случайной и расслоенной выборками?
- 1) в отличие от остальных кластерная выборка является разновидностью вероятностной выборки;
  - 2) кластерная выборка не требует наличия полного списка популяции;
  - 3) кластерная выборка позволяет правильно отобразить подгруппы популяции (например, мужчин и женщин);
  - 4) только кластерную выборку можно назвать «удобной» выборкой.
5. Для какой из следующих шкал измерений приведен правильный пример?
- 1) интервальная шкала — номера домов на улицах;
  - 2) шкала отношений — температура в градусах Цельсия;
  - 3) номинальная — как долго младенцы смотрят на новый стимул;
  - 4) порядковая — студенты колледжа решили, что профессор Х самый строгий, профессор Y второй после него, и т. д.

### Короткие эссе

1. В чем заключается суть исследования Дондерса, посвященного «психической хронометрии»? Каков его основной недостаток?

2. Дайте определение надежности и объясните, почему френологические измерения имели высокую надежность?
3. Опишите, как бы вы определили а) очевидную валидность и б) критериальную валидность теста.
4. На примере задержки вознаграждения покажите, как определяется конструктивная валидность.
5. Охарактеризуйте основные различия между описательной статистикой и статистикой вывода.
6. Опишите различия между средним арифметическим, медианой и модой и покажите, в каких случаях медиана лучше описывает общую тенденцию, чем среднее арифметическое.
7. Опишите суть метода стебля и листа и объясните, почему он иногда больше подходит для отображения набора оценок, чем гистограмма.
8. В терминах шкал измерений объясните историю Дилберта, изображенную на рис. 4.4.
9. Опишите суть проверки гипотезы и объясните различия между ошибками 1-го и 2-го рода.

## Упражнения

### Упражнение 4.1. Выборки

1. С помощью таблицы случайных чисел, приведенной в приложении *D*, сделайте случайную выборку из вашей группы по экспериментальной психологии. Преподаватель предоставит вам список группы и задаст размер выборки.
2. Повторите данное упражнение, сделав расслоенную выборку, отражающую количество мужчин и женщин в вашей группе.

### Упражнение 4.2. Шкалы измерений

Для каждого из приведенных ниже исследований определите, какая шкала измерений была использована при измерении характеристик поведения.

1. Салли хочет выяснить, по каким предметам дети республиканцев и дети демократов больше успевают, точным, гуманитарным или экономическим.
2. Фред решил исследовать, действительно ли крысы, изучившие один лабиринт, изучат второй быстрее, чем необученные.
3. Джим предполагает, что дети оценят цветные телевизионные программы выше, чем черно-белые, а у взрослых цвет не повлияет на оценку.
4. Нэнси считает, что соматотип изменяется с возрастом, и предлагает определять соматотипы у группы людей в 10, 15 и 20 лет по шкале Шелдона.
5. Сюзан изучает готовность людей помогать окружающим и считает, что она зависит от погоды — вероятность оказания помощи в солнечный день выше, чем в пасмурный.

6. Джон хочет узнать, какой из пяти новых сортов пива больше понравится (т. е. будет оценен как № 1) постоянным посетителям его бара.
7. Элен изучает, как студенты оценивают безопасность различных зданий студенческого городка. Она попросила нескольких студентов сложить карточки с написанными на них названиями зданий в стопку, в которой наиболее безопасные здания располагались бы сверху, а наименее безопасные — снизу.
8. Пэт считает, что люди с синдромом навязчивых состояний сделают меньше ошибок в составлении лабораторных отчетов по стандарту *APA*, чем здоровые люди.

### Упражнение 4.3. $H_0$ , $H_1$ , ошибки 1-го и 2-го рода

Для каждого из следующих исследований а) определите нулевую гипотезу, б) сделайте предположение об альтернативной гипотезе, т. е. изложите возможный ход исследования, в) опишите результаты исследования при ошибке 1-го рода и г) охарактеризуйте результаты при ошибке 2-го рода.

1. В исследовании способности людей опознать ложь женщины и мужчины — участники исследования пытаются обнаружить обман в записанных на видео высказываниях женщин (в одних случаях они говорят правду, а в других — обманывают).
2. В исследовании восприятия младенцам дают привыкнуть к обычным изображениям человеческих лиц, а затем им показывают несколько неправильные лица, чтобы определить, видят ли они разницу.
3. Пациентов с депрессией и без нее просят высказать предположение о том, смогут ли они преодолеть лабиринт в человеческий рост.
4. Несколько спортсменов проходят тренинг формирования зрительных образов по новой методике непосредственно перед тем, как бить пенальти. Их результаты сравниваются с результатами других спортсменов, не проходивших тренинг.

### Упражнение 4.4. Описательная статистика

Для гипотетического исследования по изучению лабиринта, в котором сравниваются немедленное и отсроченное вознаграждение, для каждой группы вычислите среднее арифметическое, медиану, разброс, дисперсию и стандартное отклонение.

### Упражнение 4.5. Статистика вывода

Воспользуйтесь примером, приведенным в приложении С, и проведите проверку по критерию Стьюдента для описанного выше исследования по изучению лабиринта. Какие выводы вы сделаете? Будут ли крысы, получающие немедленное подкрепление, запоминать лабиринт быстрее?