



И. Ф. Девятко

МЕТОДЫ СОЦИОЛОГИЧЕСКОГО ИССЛЕДОВАНИЯ



УЧЕБНОЕ
ПОСОБИЕ



И. Ф. Девятко

МЕТОДЫ СОЦИОЛОГИЧЕСКОГО ИССЛЕДОВАНИЯ

УЧЕБНОЕ
ПОСОБИЕ

Допущено Учебно-методическим объединением
по классическому университетскому образованию
к изданию в качестве учебного пособия для студентов
высших учебных заведений, обучающихся по специальности
040201 Социология и 040102 Социальная антропология



УНИВЕРСИТЕТ
КНИЖНЫЙ ДОМ

Москва
2009

УДК 30
ББК 60.5в6
Д25

Девятко И. Ф.

Д25 Методы социологического исследования : учебное пособие /
И. Ф. Девятко. – 5-е изд. – М. : КДУ, 2009. — 296 с., ил.

ISBN 978-5-98227-606-3

В данной книге рассматриваются ведущие методы социологического исследования — опрос, эксперимент, включенное наблюдение, биографический метод, а также специальные процедуры, применяемые для сбора, анализа и оценки качества социологических данных. Логика и методы социологического исследования рассматриваются в книге в контексте более широких исследовательских перспектив и моделей теоретического объяснения, предопределяющих выбор конкретных методических решений.

Книга может использоваться для подготовки и проведения социологических и маркетинговых исследований, а также в качестве пособия по курсу «Методология и методы социологического исследования». Книга предназначена для преподавателей, аспирантов и студентов факультетов социальных наук.

УДК 30
ББК 60.5в6

ISBN 978-5-98227-606-3

© Девятко И. Ф., 2009
© Издательство «КДУ», 2009

ПРЕДИСЛОВИЕ

Тема этой книги — логика и методы социологического исследования. Хотя основное внимание в ней уделено ведущим методам социологического исследования — опросу, эксперименту, включенному наблюдению, биографическому методу, а также детальному описанию специальных процедур, применяемых для сбора, анализа и оценки качества социологических данных, — многие разделы книги рассматривают процесс исследования в контексте более широких исследовательских перспектив и моделей теоретического объяснения, используемых социологами и влияющих на выбор конкретных методических решений. Взаимосвязь социологической теории и метода прослеживается здесь на разных уровнях — от выбора способа фиксации данных до идентификации используемой модели измерения.

В первой главе рассмотрены проблемы философии и логики научного исследования, взаимосвязи теоретического и методического знания в социальных науках. Главы 2 и 3 дают достаточно детальное представление об истоках, идейных предпосылках и конкретных применениях включенного наблюдения и биографического метода в социологии. Эти методы сравнительно недавно вошли в учебные планы и явно нуждаются в более полном и систематическом представлении отечественному читателю. Главы 4 и 5 рассказывают о классических и самых современных подходах к эксперименту и массовому опросу в социальных науках. Наконец, главы 6, 7 и 8 посвящены взаимосвязи теории, метода и измерения в социологических исследованиях, а также возможностям выборочного метода и статистического анализа данных. Этот круг вопросов нередко получал весьма упрощенную трактовку не только в учебных текстах, но и в обзорных монографиях. В главе 8, посвященной анализу данных, рассматриваются некоторые довольно сложные и продвинутое аналитические техники, однако ее чтение не требует подготовки, выходящей за пределы самых основных статистических представлений.

Книга может использоваться в качестве пособия по вводному курсу «Методология и методы социологического исследования» для студентов-социологов, а также служить вспомогательным источником по общеобразовательным и специальным курсам, затрагивающим вопросы планирования и проведения социологических исследований, измерения и анализа данных в социальных науках. Кроме того, она может стать своеобразным путеводителем для всех тех, кто намеревается выступить в роли квалифицированного «портретиста» или заказчика социологических исследований.

Здесь мне хотелось бы выразить признательность всем коллегам, читавшим книгу в целом, а также отдельные ее разделы, в частности профессору В. А. Ядову, высказавшему ряд полезных замечаний. Список тех, кто оказывал мне такого рода помощь и поддержку, слишком велик, чтобы приводить его в предисловии, и включает всех моих учителей и студентов. Мои основные литературные «долги» указаны в библиографических сносках и списках дополнительной литературы к отдельным главам. Я также признательна моим екатеринбургским коллегам из Издательства Уральского государственного университета, которые способствовали выходу в свет этой книги, довольно продолжительное время ожидавшей публикации. Особая благодарность — Л. Е. Петровой, оказывавшей неоценимую интеллектуальную, техническую и моральную поддержку на всех стадиях подготовки рукописи к печати, а также помогавшей мне в проведении библиографических разысканий.

Наконец, мне хотелось бы выразить благодарность моим родителям, которые ввели меня в самую суть методологической проблематики, показав, что *как* часто бывает важнее, чем *что*.

Москва, 1997

ГЛАВА 1. ЛОГИКА СОЦИОЛОГИЧЕСКОГО ИССЛЕДОВАНИЯ

Методология и логика социологического исследования. Возможно ли объективное и научное социальное знание?

Эта книга посвящена методам социологического исследования и рассчитана прежде всего на тех, кто проводит или собирается проводить такие исследования — в практических, прикладных целях либо ради удовлетворения так называемого академического интереса (последний, как известно, в более долговременной перспективе тоже служит практическим целям). Кроме того, она может быть полезна тем читателям, которые хотят стать грамотными потребителями результатов социологических исследований, в частности лицам, принимающим политические и социальные решения на основании социологических данных, заказчикам прикладных исследований (предвыборных, маркетинговых и т. п.), а также тем, кто формирует собственное либо даже общественное мнение, основываясь на цифрах и фактах, приводимых в исследовательских отчетах или газетных публикациях.

Совершенно очевидно, что достоверность фактов и выводов, полученных исследователем, зависит от того, каким способом последний пришел к данным фактам и выводам, то есть от использованного им *метода*. В повседневной жизни мы также описываем факты, оцениваем их правдоподобие, выводим гипотетические закономерности или опровергаем выводы других людей. Однако в науке все эти повседневные методы получения нового знания подвергаются куда более тщательной разработке. В повседневной жизни, например, мы легко используем понятия «всегда», «никогда» или «очень часто», но такого рода оценки остаются субъективными и относительными, пока они полностью зависят от того, *кто*

оценивает и *какие* события подвергаются оценке. Частота «один случай из двадцати» будет оцениваться как «очень часто», если речь идет о неудачных исходах хирургической операции (особенно если оценку производит пациент), и как «почти никогда», если эта частота соответствует шансам конкретного абитуриента поступить в университет. Чтобы сделать такого рода оценки более сравнимыми и объективными, в науке используются статистические критерии и методы статистического оценивания, позволяющие судить о вероятности реализации определенного события, о сугубо случайном или, наоборот, закономерном характере полученного результата (хотя, как мы увидим в дальнейшем, статистическая значимость научного факта не обязательно свидетельствует о его социальной и личностной значимости).

Современная наука — это коллективное предприятие, требующее к тому же немалых затрат. В силу этих обстоятельств, а также в силу присущих человеку *стремления к истине* (даже несовершенной и подверженной постоянным пересмотрам) и *способности к критической оценке* существующих точек зрения (большей или меньшей, но в общем поддающейся тренировке) проблема «Научного Метода», позволяющего получить достоверные и надежные результаты и на их основе прийти к объективным и хорошо обоснованным выводам, неизменно пользуется вниманием ученых. Хотя по мере развития науки у ученых оставалось все меньше оснований верить в существование единого и универсального «Метода», пригодного для решения всех без исключения научных задач, они тем не менее стремились разработать все более точные и надежные методы опытной проверки теоретических гипотез, а также определить критерии для оценки объективности и обоснованности нового знания, получаемого в результате проводимых ими исследований.

Методология науки — это дисциплина, изучающая и технические, «процедурные» вопросы организации исследования, и более общие вопросы обоснованности используемых методов, достоверности наблюдений, критериев подтверждения или опровержения научных теорий.

Такая «всеядность» методологии связана с тем, что большинство сугубо технических вопросов, возникающих в реальной исслед-

довательской практике, не могут быть решены без обращения к более широким представлениям о том, *что* в данном случае может служить критерием истинности или обоснованности, *какие* факты следует считать релевантными, то есть относящимися к проверяемой теории, а какие — случайными ошибками наблюдения и т. п. С другой стороны, непосредственным источником самих нормативных стандартов и критериев, используемых при оценке исследовательских процедур и результатов, служат не только абстрактные и универсальные конструкции формальной логики, но и реальная, «ситуативная» логика конкретного исследования, принимающая во внимание его цели, происхождение стоящей перед исследователем научной проблемы (теоретическое затруднение, практическая задача), доступные теоретические и технические ресурсы и, наконец, научный «фон» исследования — ту теоретическую перспективу, или, если воспользоваться популярным обозначением, *исследовательскую программу*, в рамках которой и замысел, и методика, и результаты исследования приобретают свой подлинный смысл. Чтобы понять смысл и результаты, например, экспериментов физика А. Майкельсона по измерению относительной скорости света (вдоль и поперек потока «эфирного ветра»), нужно представлять себе суть спора между сторонниками различных теорий «светоносного эфира» и сторонниками теории относительности Эйнштейна¹. Точно так же трудно оценить процедуру и результаты эксперимента, в котором психолог сначала помещает крыс в лабиринт, оставляя их там на некоторое время, а потом, установив в конце лабиринта кормушку с приманкой, измеряет с помощью секундомера время, затраченное крысами на прохождение лабиринта. Чтобы понять методику последнего эксперимента, следует представлять

¹ Интересно отметить, что такое понимание не всегда доступно самому исследователю в момент осуществления исследования. Так, Майкельсон до опубликования основных положений теории относительности Эйнштейна, породившей абсолютно новую исследовательскую программу, воспринимал свои опыты не как решающее опровержение теории эфира, а скорее как подтверждение того, что при движении Земли в эфире не возникает так называемый «эфирный ветер» (либо, если возникает, то очень маленький). Подробнее см.: *Лакатос И.* Фальсификация и методология научно-исследовательских программ. М.: Московский философский фонд «Медиум», 1995.

себе основные положения *исследовательской программы бихевиоризма*, весьма влиятельного направления в науках о поведении. (Только что описанный эксперимент², в частности, играет важную роль в споре о том, способны ли внутренние образы окружения, «когнитивные карты», оказывать реальное влияние на скорость выработки поведенческих навыков у животных. Для человека, не имеющего представления об этом споре, предварительное помещение крыс в лабиринт выглядит ничего не значащим фактом или, в худшем случае, методическим промахом экспериментатора; для сторонника же бихевиористской исследовательской программы — это логически необходимое условие для любых выводов относительно влияния предварительной ориентировки на обучение.)

До недавнего времени в философии и методологии науки господствовал упрощенный взгляд на логику и процедуру научного исследования, который можно обозначить как *«традиционный образ науки»*. Этот взгляд игнорировал только что описанные сложные взаимосвязи между самыми общими философскими и теоретическими представлениями, входящими в определенную исследовательскую программу, и более частными предположениями, из которых явно или неявно исходит ученый, планирующий конкретное исследование. «Традиционный образ науки» представлял процесс исследования как простую линейную последовательность:

проверяемая общая теория, → из которой выводится *основная теоретическая гипотеза*, → определение основных теоретических понятий в терминах конкретных измерительных операций, то есть их *операционализация*, → *решающий эксперимент*, ведущий к → однозначному *принятию или отвержению гипотезы*, а заодно и общей теории, из которой гипотеза была выведена.

Предполагалось, что отрицательный результат «решающего эксперимента», то есть эмпирическая демонстрация ложности предсказаний, выведенных путем логической дедукции из теоретических предпосылок, свидетельствует о ложности этих предпосылок. На практике, однако, всегда можно приписать «неудачное» наблю-

² Данный эксперимент представляет собой упрощенную версию экспериментов, реально проводившихся психологом Э. Толменом.

дение либо неучтенным особенностям исходных условий конкретного эксперимента, либо ложности множества вспомогательных гипотез и предположений, используемых при проверке основной теоретической гипотезы. Так, в частности, негативный результат всегда может быть приписан не проверяемой гипотезе, а артефактам используемого метода или погрешностям в операционализации и измерении отдельных показателей³. Кроме того, в теоретическую гипотезу, вспомогательные предположения или, наконец, в самые общие теоретические предпосылки исследовательской программы, включающей в себя спорную теорию, почти всегда можно внести разного рода небольшие поправки или дополнения (так негативный результат упоминавшихся выше опытов Майкельсона с измерением «эфирного ветра» поначалу объяснялся тем, что движение эфира непосредственно у поверхности Земли столь незначительно, что использованная в опыте аппаратура не обладала достаточной точностью для его фиксации). Иными словами, «решающие эксперименты» чрезвычайно редки даже в точных науках: влиятельная и пользующаяся поддержкой многих ученых теория имеет большие шансы устоять даже при наличии нескольких контрпримеров — до тех пор, пока более широкая исследовательская программа, в которую включена теория, не уступит место новой, более плодотворной и привлекательной. Кроме того, «традиционный образ науки» игнорировал то обстоятельство, что далеко не всегда исходной точкой исследовательского процесса служит вполне сложившаяся абстрактная теория. Источником исследовательских гипотез и в социальных науках, и, например, в биохимии или физике часто становится случайное наблюдение, личный «внеаучный» опыт ученого, неожиданные результаты, полученные при использовании новой методики или технического прибора. В попытках объяснить необычный факт исследователь формулирует пробные теоретические предположения, которые пока не могут быть привязаны ни к какой из существующих абстрактных теорий (и тем

³ Примеры влияния так называемых *моделей измерения* — вспомогательных гипотез, касающихся используемых социологами шкал и индикаторов, — на теоретическую интерпретацию результатов исследования подробнее рассмотрены в гл. 5 и 6.

более не могут быть выведены из нее дедуктивно). Некоторые из возникших таким образом догадок повышают «теоретическую чувствительность» исследователя и ведут к формулировке новых рабочих гипотез, введению новых описательных понятий или даже, в конечном счете, к возникновению новых исследовательских программ⁴. Так, например, непосредственно в процессе включенного наблюдения за взаимодействием членов некоторой организации, проводимого социологом, у последнего могут возникнуть весьма полезные догадки, касающиеся существующей в этой организации статусной иерархии. (Подробнее о возможностях теоретического обобщения полевых данных говорится в гл. 2⁵). Таким образом, налицо значительное несоответствие между вышеописанным «традиционным образом науки» и реальной логикой исследования. Это несоответствие становится особенно очевидным, когда мы обращаемся к области социальных наук:

1) где существует сравнительно мало развитых формальных теорий, из которых можно было бы строгим образом вывести проверяемые гипотезы, а для каждой из таких гипотез уже в момент ее выдвижения можно найти множество контрпримеров;

2) где возможности экспериментального метода заведомо ограничены, а имеющиеся данные о естественно случающихся событиях либо о результатах специальных опросов редко позволяют разделить главные и побочные эффекты;

3) где, наконец, одновременно существует несколько очень влиятельных исследовательских программ (например, бихевиористская, интерпретативная и структуралистская), каждая из которых обладает собственным набором методологических норм,

⁴ О том, как происходят такого рода научные революции, можно подробнее узнать из книги Т. Куна «Структура научных революций». См.: Кун Т. Структура научных революций. М.: Прогресс, 1977.

⁵ Важно помнить, что за подобными догадками стоит не столько мистическая интуиция или «внутреннее знание», сколько целая система «обыденных социологических теорий», на которой основывают свои повседневные решения и интерпретации и профессиональные социологи, и обычные люди. «Обыденные теории», подобно научным, вполне могут оказаться как полезными, так и бесполезными, как верными, так и ошибочными — поэтому они также нуждаются в формализации, операционализации и проверке.

излюбленных исследовательских техник и образцовых теоретических интерпретаций⁶.

Воспользуемся наглядным примером, взятым из области социальных наук. В социальной психологии, социологии и экономике существуют довольно убедительные теории, относящиеся к единой исследовательской программе и описывающие поведение человека как эгоистическое, направленное на достижение максимального удовлетворения потребностей индивида. С точки зрения этих теорий, в частности, даже самые идеальные мотивы скрывают за собой рациональный расчет, направленный на увеличение личных «активов», в том числе материальных благ, социального престижа, политического влияния. Самым рациональным, таким образом, будет поведение, которое позволяет максимизировать доступные индивиду блага — материальные и нематериальные — для данного набора индивидуальных возможностей, то есть ресурсов, которыми этот индивид располагает. Соответственно, деятельность ученого, проводящего трудоемкие изыскания без какой-либо материальной поддержки, или поступок филантропа, жертвующего деньги на благотворительность, также рассматриваются с точки зрения скрытых вознаграждений, удовлетворяющих эгоистические потребности: ученый рассчитывает завоевать признание коллег, получить хорошо оплачиваемую работу в престижном университете; филантроп заинтересован в удовлетворении собственного тщеславия и создании благоприятного общественного мнения о себе. В конкретном исследовании данная *исследовательская программа* (назовем ее «концепцией экономического человека») принимает форму *гипотез*, описывающих ожидаемое поведение людей в ситуациях определенного типа (например, «если в ситуации *S* для достижения максимального вознаграждения *A* нужно сделать *B*, то испытуемый *X* сделает *B*»). Получаемые в результате исследования *данные* могут соответствовать либо не соответствовать тому, что предсказывает гипотеза. Следовательно, результаты будут рассматриваться

⁶ Подробнее о соотношении исследовательских программ, теоретических моделей и методов исследования в социальных науках см.: *Девятко И.* Модели объяснения и логика социологического исследования. М.: ИСО РЦГО-TEMPUS/TACIS, 1996.

как подтверждающие или не подтверждающие выдвинутую первоначально гипотезу. Всякое удачное предсказание будет вести к относительно подтверждению гипотезы (для данного типа ситуаций). Интереснее, однако, то, что происходит, когда теоретические предсказания не сбываются или сталкиваются с контрпримерами⁷. «Концепция экономического человека», в частности, сталкивается с трудностями, если требуется предсказать поведение в ситуациях неопределенности, когда выбор основан на недостаточной информации и окончательная «цена» того или иного решения в принципе не может быть определена заранее. В таких ситуациях самой разумной с точки зрения «экономического» подхода стратегией обычно является случайное угадывание. Однако эксперименты и наблюдения за поведением отдельных людей или групп показывают, что они практически никогда не основывают свой выбор на простом подбрасывании монетки. Напротив, они всячески стремятся обосновать свой выбор и придумать какую-то «стратегию принятия решений», исходя при этом из ложных или заведомо непроверяемых представлений. Еще сложнее сторонникам описанного подхода справиться с другого рода контрпримерами, демонстрирующими слепое подчинение разума вполне неразумным целям либо рациональное стремление увеличить или уменьшить не собственное, а чье-либо еще благо. Человек, совершающий самоубийство, вполне способен осуществить задуманное самым экономичным и рациональным способом. Возможно, конечно, утверждать, что при этом он достигает максимального блага, или, если воспользоваться специальным жаргоном, реализует свое «наибольшее предпочтение», но в таком случае нам придется признать свою неспособность дать независимое определение того, что следует считать «разумным желанием» — рациональной будет выглядеть любая цель, мотивирующая реальное поведение. Более того, так как люди часто стре-

⁷ Полезно различать *негативный результат* исследования, специально предназначенного для проверки предсказаний теории, и *контрпример* — наблюдение или исследовательский результат, который противники научной теории предлагают в качестве критического, предположительно не имеющего объяснения с точки зрения этой теории (то есть контрпример — это такой факт, для которого из теории не удастся дедуктивно вывести гипотезу, его предсказывающую).

мятся представить любые свои поступки в качестве разумных, «рационализировать» их, то нам будет весьма трудно отличить истинно расчетливые поступки от «рационализаций», изобретаемых для оправдания уже свершенного. Трудности в определении ключевых понятий приведут к трудностям в их операционализации, и в результате эмпирическое подтверждение или отвержение отдельных гипотез, выводимых из «концепции экономического человека», будет полностью зависеть от используемых показателей или методов измерения рациональности поступков. Кроме того, результат, противоречащий теоретическим предсказаниям, всегда может быть отнесен на счет несовершенства этих методов. Существуют и такие контрпримеры, которым обсуждаемая концепция (являющаяся, напомним, не отдельной теорией, а обширной исследовательской программой) не может дать сколько-нибудь удовлетворительного объяснения. Так, иногда люди совершают добрые поступки анонимно (вносят благотворительные пожертвования, возвращают потерянные кошельки и т. п.). Ученый из вышеприведенного примера может тратить значительные силы не на собственные труды, а на то, чтобы любой ценой воспрепятствовать публикации трудов коллеги. «Концепция экономического человека» может многое объяснить в том, как люди расчетливо используют наличные средства для достижения максимально возможного счастья (так определял рациональное поведение Гоббс), но она малопригодна для объяснения того, как человек может руководствоваться стремлением к счастью или несчастью другого человека⁸. Такая «ограниченная подтверждаемость» (обратной стороной которой является «ограниченная опровергаемость») скорее типична для социальных наук. Сильная взаимосвязь между теорией, методом и получаемыми «на выходе» эмпирическими данными, характерная для реальной логики всякого научного исследования, здесь становится еще более очевидной в силу наличия конкурирующих теоретических перспектив, каждая

⁸ Заметим, не вдаваясь в детали, что выход из такого теоретического тупика можно найти в принятии базисных предположений других исследовательских программ, например психоаналитической, но за это обычно приходится платить отказом от собственных базисных предположений.

из которых располагает большим количеством контрпримеров, ставящих под сомнение базисные предположения остальных⁹.

Приведенный пример показывает, что чрезмерно упрощенный «традиционный образ науки» и основанные на нем методологические рекомендации едва ли могут оказаться полезными при разработке и оценке исследовательских методов для социальных наук. Это, однако, не означает, что нам следует отказаться от разработки критериев *рационального* и *объективного* научного исследования, основанных на более адекватном представлении о научной практике.

Во-первых, всякое исследование в социальных науках направлено на поиски объяснения человеческого поведения и, следовательно, ориентировано на поиск некоторой *гипотетической закономерности*, обладающей большей или меньшей степенью общности, но всегда требующей эмпирического подтверждения и критического сопоставления с другими альтернативными гипотезами. Источником таких гипотез, как мы уже видели, могут быть и сложившиеся научные концепции, и — реже — «обыденные теории». Далее, возможность объяснения и предсказания в общественных науках основана на признании *причинной обусловленности* объясняемых событий. Даже объясняя поведение людей их целями, представлениями и убеждениями, социолог стремится продемонстрировать работу некоторого причинного механизма, обеспечивающего взаимосвязь целей и идей (рациональных или иррациональных, истинных или ложных) с поведением. Изучение чисто логических отношений *согласованности* между целями и средствами деятельности, элементами системы верований или, например, брачными правилами, существующими в некотором сообществе, само по себе еще не позволяет объяснить, почему произошли или не произошли некоторые события. Логические отношения между идеями или высказываниями (отношения импликации) позволяют осуществлять логический вывод — от одного формально истинного высказывания к другому, но не позволяют предположить, что произойдет (или произошло) в

⁹ Так, например, *холистские* исследовательские программы, объясняющие поведение отдельных действующих надындивидуальными, структурными факторами, противостоят *индивидуалистским* программам, сводящим любое социальное целое к мотивам и поступкам отдельных людей.

действительности. Из совокупности высказываний можно логически вывести лишь другую совокупность высказываний. (Человек, логично рассуждающий о преимуществах любви к ближнему, как известно, не обязательно следует собственным рассуждениям на практике¹⁰.) Поэтому, вопреки рецептам некоторых сторонников «гуманистической модели» социальных наук, социологи не ограничиваются интерпретацией того, что люди говорят, либо того, во что они верят.

Оценка существующих теорий и гипотез в социальных науках, как и в науках естественных, предполагает введение определенных критериев эмпирической проверяемости и истинности теоретических высказываний, а также разработку и применение соответствующих этим критериям методов исследования.

Таким образом, процесс социологического исследования неизбежно включает в себя:

1) стадию осознания теоретической или практической недостаточности существующего знания (источники такого осознания, как говорилось выше, могут лежать и в области теории, и в области повседневного опыта или социальной практики);

2) стадию формулировки проблемы и выдвижения гипотетического объяснения, а также

3) стадию эмпирической проверки сформулированной гипотезы, за которой нередко следует

4) стадия переопределения и уточнения проблемы или гипотезы, дающая начало новому исследовательскому циклу.

Разнообразие существующих в социологии исследовательских программ, а также реальных *контекстов исследования*, то есть теоретических и практических целей исследования, возможностей, которыми располагает исследователь, а также возникающих в исследовательской практике технических и этических ограничений, ведет к тому, что конкретные реализации описанного процесса исследования могут существенно различаться. *Ведущие методы социологического исследования и представляют собой такие конкретные реализации, или стратегии, процесса социологического исследования.*

¹⁰ Впрочем, для всякого «проступка» обычно находится *другое объяснение*.

Методы социологического исследования: общий обзор

В этой книге рассматриваются основные методы социологического исследования — *эксперимент*, *метод включенного наблюдения*, *биографический метод*, *массовый опрос*, а также конкретные методики, используемые на разных стадиях исследовательского процесса (в частности, методики построения выборки, измерения и анализа данных, в силу своей относительной сложности и значимости выделенные в отдельные главы). Описания специфических процедур сбора, анализа и интерпретации данных, характерных для каждого из рассматриваемых методов, как и подробный анализ преимуществ и недостатков последних, будут представлены в соответствующих главах, здесь же имеет смысл ограничиться кратким обзором, позволяющим прежде всего проследить взаимосвязь основных социологических методов с теми исследовательскими программами, в рамках которых они первоначально формировались, а также с теми контекстами исследования, в которых они чаще всего используются.

Эксперимент — это метод, обеспечивающий наилучшие эмпирические данные для проверки гипотез о наличии причинной связи между явлениями, а также самое надежное средство решения многих практических задач, связанных с оценкой эффективности социальных и политических программ. Многомерный контролируемый эксперимент, как мы увидим в дальнейшем, соответствует самым строгим стандартам научного вывода и незаменим при сравнении объяснительных возможностей разных теорий. В некоторых отношениях процедура экспериментальной проверки гипотез даже превосходит эталоны вышеописанного «традиционного образа науки», так как возникающая при планировании эксперимента необходимость в формализации теоретической модели, операционализации переменных, определяющих «главный эффект», а также в нахождении инструментов контроля посторонних, смешивающих влияний, ведет не только к прояснению основной гипотезы, но и к анализу всех тех внешних условий и факторов окружения, для которых соблюдаются постулируемые теорией соотношения (такой анализ, как будет показано в гл. 4, призван гарантировать *внешнюю валидность* эксперимента). Недостатки экспериментального мето-

да являются продолжением его достоинств (что, впрочем, верно и применительно ко всем остальным методам). Возникнув в натуралистической традиции социологического исследования, экспериментальный метод был изначально ориентирован на лабораторный или квазилабораторный исследовательский контекст, высокий уровень формализации проверяемых теорий и максимальные возможности измерения и контроля всех существенных переменных. Кроме того, сторонники экспериментального метода с самого начала отдавали предпочтение скорее абстрактным и общим понятиям научной теории в ущерб специфическим и уникальным понятиям, используемым при описании социального взаимодействия его непосредственными участниками или «непрофессиональными» наблюдателями. Иными словами, эксперимент оказался методом, пригодным скорее для проверки наиболее «сложившихся» и развитых социологических и социально-психологических теорий, чем для поисковых исследований, направленных на выработку адекватного теоретического языка и формулировку пробных гипотез, описывающих закономерности естественного протекания социальных процессов. Кроме того, следует помнить об этических проблемах, иногда возникающих при экспериментальном манипулировании переменными социального окружения. Эти проблемы могут касаться не столько гипотетического влияния нежелательных факторов, сколько возможного социального неравенства, возникающего в крупномасштабных полевых экспериментах при распределении участников по экспериментальным и контрольным группам, так как в результате члены контрольных групп не получают «позитивного» экспериментального воздействия (на оценку эффективности которого и направлен эксперимент), например социального пособия, нового прогрессивного метода обучения и т. п. Наконец, экспериментальный метод мало пригоден для получения результатов, которые можно было бы распространить на общество в целом или на большие социальные группы, он не позволяет увидеть «срез» широкомасштабных социальных процессов. Результаты хороших лабораторных экспериментов обладают высокой надежностью, однако они довольно далеки от «реального мира» (справедливости ради нужно отметить, что социальным наукам далеко не всегда следует стремиться к отражению многообразия «живой жизни»). Результаты

полевых экспериментов в целом характеризуются большей близостью к «реальному миру», однако это преимущество достигается ценой несколько меньшей надежности и большей подверженности всяческому смещению. Качество данных, получаемых в широкомасштабных социальных экспериментах, далеко не всегда оправдывает их чрезвычайно высокую стоимость.

Массовый опрос является, пожалуй, самым популярным социологическим методом. Он превосходит эксперимент с точки зрения дескриптивных возможностей и служит не только сугубо академическим целям, являясь наилучшим средством получения *социальной статистики*. Именно опросы общественного мнения используются при изучении мнений и установок широких слоев общества, обеспечивая, при корректном применении, возможность «отслеживания» даже небольших изменений в самых разнообразных сферах общественной жизни — от распределения семейных бюджетов до динамики избирательских предпочтений. Современные подходы к построению выборки и анализу данных, о которых рассказывается в гл. 7 и 8, позволяют максимально приблизить возможности проверки причинных гипотез, предоставляемые методом массового опроса, к возможностям экспериментального метода. Недостатки опросного метода отчасти также совпадают с недостатками последнего. Речь идет прежде всего о низкой чувствительности этого метода к уникальным чертам исследуемой социальной ситуации, об относительно меньшем внимании к субъективным и индивидуальным характеристикам опыта исследуемых людей и групп, к их самоописаниям, интерпретациям и «обыденным теориям». Описанные недостатки, в свою очередь, являются обратной стороной стремления к теоретическому обобщению результатов и концептуальной строгости.

Преимущества *включенного наблюдения* и *биографического метода* заключены, прежде всего, в возможности получения детальной «дотеоретической» информации об изучаемых социальных явлениях. Непосредственная включенность исследователя в изучаемую социальную ситуацию, группу или культуру нередко позволяет получить уникальные сведения о используемых самими участниками значениях и символах, о локальных или субкультурных «языках взаимодействия», знакомство с которыми, как будет показано далее, является само собой разумеющимся условием их дальнейшего

теоретического анализа. Хотя ученый не может «влезть в шкуру» других людей, особенно принадлежащих к чужой культуре или другой исторической эпохе, он может попытаться упорядочить и подвергнуть более глубокому и систематическому рассмотрению те слова, символы и культурные формы, посредством которых изучаемые им люди описывают и передают свой опыт, делая это зачастую непоследовательно, случайно или не вполне осознанно. Сравнительно абстрактные и высокосодержательные термины научного описания, в свою очередь, позволяют социологу или этнографу превратить спонтанное переживание и изменчивые культурные формы в предмет собственно теоретического анализа, сделать еще один шаг к увеличению достоверного, доступного коллективному пониманию и проверяемого научного знания. Наиболее очевидные недостатки включенного наблюдения и, в несколько большей степени, биографического метода связаны с излишне дескриптивным характером получаемых данных, опасностью подмены научных объяснений высокохудожественными и вполне субъективными повествованиями, в которых на смену внятным теоретическим представлениям и эмпирическим доказательствам приходят риторические фигуры и суггестивные авторские интонации.

Дополнительная литература

- Батыгин Г. С.* Обоснование научного вывода в прикладной социологии. М.: Наука, 1986.
- Девятко И. Ф.* Модели объяснения и логика социологического исследования. М.: ИСО РЦГО-TEMPUS/TACIS, 1996.
- Лакатос И.* Фальсификация и методология научно-исследовательских программ. М.: Московский философский фонд «Медиум», 1995.
- Кун Т.* Структура научных революций. М.: Прогресс, 1977.
- Уинч П.* Идея социальной науки и ее отношение к философии. М.: Русское феноменологическое общество, 1996.
- Ядов В. А.* Социологическое исследование: методология, программа, методы. 2-е изд., перераб. и доп. М.: Наука, 1987. Гл. 1.
- Ядов В. А.* Стратегии и методы качественного анализа данных // Социология: 4М. 1991. № 1.

ГЛАВА 2. ВКЛЮЧЕННОЕ НАБЛЮДЕНИЕ

Что такое включенное наблюдение?

Включенное наблюдение и этнографический метод: определение и исторические истоки

Под «включенным наблюдением» в социологии обычно подразумевают либо особый *метод сбора социологических данных* (1), либо *целостную исследовательскую стратегию*, то есть методологию социальных и гуманитарных исследований, качественно отличную от методологии естественных наук (2). «Включенное наблюдение-2», таким образом, шире по содержанию, чем «включенное наблюдение-1». Причина заключается в том, что за более широкой трактовкой наблюдения-2 стоит значительно большее количество явных и неявных *теоретических предпосылок и предположений*. Некоторые из этих предпосылок относятся к возможностям и ограничениям наших методов познания социального мира, некоторые же связаны с общими представлениями о том, как устроен сам этот мир.

Разумеется, выделенные нами две трактовки включенного наблюдения — это заведомая идеализация. Даже среди тех, кто убежден в том, что включенное наблюдение — ведущая методология социальных наук, нет полного единства мнений. Некоторые исследователи полагают, что преимущества включенного наблюдения связаны с возможностью уточнения и усовершенствования теоретических понятий в ходе непосредственного взаимодействия исследователя с описываемой им реальностью, что особенно существенно в том случае, когда сам исследователь исходно не принадлежит к изучаемой культуре или сообществу. В этом случае социолог-«аутсайдер» получает значительную часть своих теоретических представлений в прямом сотрудничестве с хорошо осведомленным информатором-«туземцем». Информатор здесь становится прямым источником содержательных представлений и понятий, которые социолог в дальнейшем подтверждает, уточняет или опровергает

(например, это могут быть сведения о структуре взаимодействия в уличной шайке либо об отношениях родства или нормах этикета в индейском племени). Описанной исследовательской установке часто соответствует широкое понимание включенного наблюдения как *этнографического метода* описания человеческих сообществ¹.

Несколько иной тип исследовательской установки при использовании метода включенного наблюдения имеет место в том случае, когда социолог стремится понять и принять точку зрения тех, кого он исследует, реконструировать субъективный смысл, который первые вкладывают в свои поступки, проникнуть в изменчивый символический мир социальных субъектов. Этот тип исследовательской установки особенно характерен для работ, выполненных в традиции *символического интеракционизма*, понимающего под включенным наблюдением прежде всего «отношение, которого не может избежать человек, ведущий наблюдение за другими человеческими существами, а именно — необходимость каким-то образом соучаствовать в опыте и поступках тех, кого он наблюдает»².

Наконец, возвращаясь к тому, что мы обозначили как «включенное наблюдение-1», исследователь может придерживаться весьма жестких стандартов научного вывода, стремясь к построению обобщенных причинных объяснений и предсказаний, и вместе с тем использовать включенное наблюдение как метод сбора данных и эвристическую процедуру для формирования предварительных теоретических гипотез и понятий на ранних стадиях исследования³. По мере уточнения сферы теоретических интересов, сравнительной роли различных теоретических понятий и переопределения исследовательской проблемы социолог может постепенно перейти к использованию результатов включенного наблюдения для измерения, проверки более строгих гипотез или построения причинных моделей происходящего⁴.

¹ См.: *Hammersley M., Atkinson P. Ethnography: Principles in Practice. L.: Tavistock, 1983.*

² *Blumer H. Foreword // Severyn T. Bruyn. The Human Perspective in Sociology: The Methodology of Participant Observation. Englewood Cliffs (N. J.), 1966. P. VI.*

³ См.: *Lazarsfeld P. F. Qualitative Analysis. Boston: Allen and Bacon, 1972.*

⁴ См.: *Lofland J., Lofland L. H. Analyzing Social Settings. Belmont (Ca.): Wadsworth, 1984.*

Метод включенного (или полевого) наблюдения, таким образом, получает несколько различающееся толкование в различных теоретических перспективах, в зависимости от того, как понимаются природа и цели социологического исследования. Не менее разнообразны и сферы использования полевых наблюдений в социологии, те «жизненные миры», области социального опыта, которые могут стать предметом этнографического описания.

Классическим примером стало исследование У. Ф. Уайта, описавшего структуры взаимодействия и неявные статусные иерархии, организующие повседневную жизнь бедного итало-американского района в большом городе на восточном побережье США, и его обитателей — местных политиков, членов молодежных банд и ориентированных на карьеру способных студентов⁵.

Другой классический пример — проведенное Л. Фестингером и соавторами исследование небольшой секты, пророчившей скорое наступление конца света⁶. Заметим, что в исследовании Уайта социолог явно выступал в роли наблюдателя и не скрывал целей своего присутствия «в поле». Фестингер и соавторы полагали, что единственный способ проникнуть в замкнутую общину верующих, воспринимавших внешний мир как враждебный и нечувствительный к тайному откровению, заключался в том, чтобы стать полноправными членами секты и скрыть свою профессиональную роль социолога. В этом случае наблюдатели получили доступ в «поле», представившись путешествующими бизнесменами, слышавшими прежде о существовании группы, сочувствующими ее целям и желающими узнать о ней больше. Однако вскоре социологи, стремившиеся наблюдать сектантов в естественных условиях (важная предпосылка этнографического метода), столкнулись с тем, что их собственное присутствие интерпретировалось верующими как прямое подтверждение подлинности их откровения. Ситуация усугублялась тем, что один из социологов, стремясь придать большее правдоподобие своей «легенде», рассказал членам секты о якобы

⁵ *Whyte W. F. Street Corner Society*. 2nd ed. Chicago: University of Chicago Press, [1943] 1955.

⁶ *Festinger L., Riecken H., Schachter S. When Prophecy Fails*. N. Y.: Harper & Row, 1956.

имевшемся у него опыте оккультной практики и сверхчувственного познания. Члены секты восприняли его «обращение» как важное событие. Таким образом, значительная часть «естественного» хода событий оказалась вольно или невольно сфабрикованной социологами.

К другим, сравнительно недавним образцам успешного использования метода включенного наблюдения в социологии можно отнести — ограничившись лишь несколькими примерами — исследования повседневной жизни «внутри» лабораторной науки⁷, исследования профессиональной социализации в хирургическом отделении больницы и способов оценки и контроля медицинских ошибок, совершаемых молодыми врачами⁸, изучение роли «кокаиновой экономики» в жизни маленьких сельскохозяйственных общин в перуанских Андах⁹, анализ процесса старения и способов, с помощью которых обитатели еврейского центра для престарелых в Калифорнии осмысливают и организуют свою жизнь¹⁰.

Мы будем рассматривать включенное наблюдение не как альтернативу другим исследовательским подходам, а как один из важных методов социальных наук, часто использующий элементы других методов и техник (например, анализ документов, клиническое интервью, квазиэксперимент) и, в свою очередь, позволяющий расширить содержательную интерпретацию результатов, полученных другими, более формализованными методами.

Включенное наблюдение основывается на широком круге источников информации. Наблюдатель «явно или неявно соучаствует в повседневной жизни людей в течение достаточно продолжительного времени, наблюдая за происходящим, прислушиваясь к сказанному, задавая вопросы. В сущности, он собирает любые доступные

⁷ См.: *Latourel B., Woolgar S. Laboratory Life*. Beverly Hills (Ca.): Sage, 1979, а также *Lynch M. Art and Artifact in Laboratory Science*. L.: Routledge and Kegan Paul, 1985.

⁸ См.: *Bosk Ch. L. Forgive and Remember: Managing Medical Failure*. Chicago: The University of Chicago Press, 1979.

⁹ См.: *Morales E. Cocaine: White Gold Rush in Peru*. Tuscon: University of Arizona Press, 1989.

¹⁰ См.: *Myerhoff B. Number Our Days*. N. Y.: Simon and Schuster, 1978.

данные, которые могут пролить свет на интересующие его (или ее) проблемы»¹¹.

В общем случае, включенное наблюдение чаще основано на неформализованных интервью, менее репрезентативных данных, нестатистическом подходе к обоснованию выводов и причинных моделей. Из сказанного, однако, не следует, что при использовании этнографических методов в социологии «все позволено» и исследователь может с легкостью отказаться от любой теоретической логики, стандартов репрезентативности или от обоснования своих выводов. Включенное наблюдение основывается на некоторых теоретических предпосылках и абстрактных идеях, понимание которых существенно для осмысленного использования этого метода.

Методология включенного наблюдения подчеркивает важность «логики открытия»¹², не проводящей жесткой границы между формализованной теорией и эмпирическим знанием, между формально-логическим рассуждением и здравым смыслом в процессе поиска новых понятий, обобщений и теорий. Предполагается, что более гибкие способы определения исследовательской проблемы и теоретических гипотез и соответствующие методы сбора и оценки эмпирической информации создают предпосылки для построения теорий, укорененных в реальности конкретного социального опыта, в повседневных словах и поступках людей¹³.

Планирование исследования: определение проблемы, отбор случаев, ситуаций и групп

На той стадии работы, которая предшествует полевому наблюдению, исследователь обычно определяет для себя ключевую проблему и соответствующую совокупность ключевых вопросов, на которые можно получить теоретически осмысленные, то есть подлежащие и поддающиеся теоретическому осмыслению, ответы.

¹¹ *Hammersley M., Atkinson P.* Ethnography: Principles in Practice. P. 2.

¹² См.: *Kaplan A.* The Conduct of Inquiry. San Francisco: Chandler, 1964.

¹³ См.: *Glazer B., Strauss A.* The Discovery of Grounded Theory. Chicago: Adline, 1967, а также *Agar M. H.* Speaking of Ethnography. Beverly Hills et al.: Sage, 1986. (Qualitative Research Methods Series. Vol. 2.)

Социологи или этнологи (в отличие, например, от психологов или экономистов) довольно редко сталкиваются с ситуацией, когда их предварительные гипотезы можно строго вывести из развернутой и логически согласованной теории. Значительно типичнее ситуация, когда выбор проблемы определяется совокупностью более или менее отчетливых теоретических понятий и идей, совокупностью интересных и требующих объяснения фактов и, наконец, различными политическими, практическими и этическими соображениями. Включенное наблюдение, как и другие типы социологического исследования, часто начинается в ситуации, где в той или иной мере присутствуют все перечисленные компоненты: идеи, факты, политика, этический и практический интерес. Однако в случае этнографического исследования (мы будем иногда использовать последний термин как синоним «включенного наблюдения») исследователь обычно ставит своей целью не столько проверку гипотез, выводимых из существующей теории, сколько развитие новых теоретических представлений.

Хорошим примером поиска теоретического объяснения уникального факта может служить известная работа М. Фрейлиха, посвященная роли социокультурных факторов в объяснении такого необычного (хотя и достаточно известного) явления, как «сверхпредставленность» индейцев-мохавков среди нью-йоркских монтажников-высотников¹⁴.

Личный опыт и интересы также нередко становятся предпосылками исследования. Так, например, А. Стросс и Б. Глезер незадолго до начала своего знаменитого исследования процесса умирания в больничных условиях оба пережили потерю близких. Обстоятельства смерти их близких были таковы, что уже на предварительной стадии полевого исследования их внимание было приковано к рутинным процедурам обращения медиков с неизлечимыми пациентами и влиянию знания о безнадежном прогнозе на социальное взаимодействие больных, их родственников и персонала¹⁵. В. Богтз, исследовавший мир латиноамериканского джаза, и особенно той его разновидности, которую иногда называют «сальса» (своеобразный

¹⁴ *Freilich M.* Mohawk Heroes and Trinidad Peasants // *Freilich M.*, (ed.) *Marginal Natives: Anthropologists at Work*. N. Y., 1970.

¹⁵ *Glazer B.*, *Strauss A.* *Awareness of Dying*. Chicago: Adline, 1965.

синтез латиноамериканского бита и негритянского джаза), смог постепенно перейти от многолетнего увлечения и непосредственной включенности в это музыкальное движение к формулировке социологического проекта, направленного на анализ расовых и классовых аспектов музыки. Однако переход этот был нелегким и потребовал изрядных усилий: не существует легкого способа превращения личного интереса в профессиональную вовлеченность ученого¹⁶.

Уже на ранних стадиях исследования — при изучении литературы, анализе доступных документальных источников, ознакомлении с ситуацией «в поле» — социолог часто меняет или уточняет исходную формулировку проблемы, некоторые теоретические предпосылки или рабочие понятия. Вполне может оказаться, что проблема в исходной своей формулировке пока — или в принципе — неразрешима либо в исходных теоретических представлениях содержались существенные пробелы. В этой ситуации нет ничего необычного — ученому часто приходится вспоминать старую истину: задать правильный вопрос труднее, чем найти на него ответ¹⁷.

Особое значение неожиданностям, радикальным изменениям точки зрения исследователя и ощущению неадекватного понимания ситуации придает *герменевтическая традиция*. Не вдаваясь здесь в обсуждение сложных философских вопросов, мы лишь кратко опишем, как трактуется в этой традиции собственно «этнографическое понимание»¹⁸. Здесь особое значение придается тому обстоятельству, что социолог или культурный антрополог, по сути, сталкивается с чуждым, иным жизненным миром. (Даже если это мир его собственной культурной традиции, ученый стремится сделать его понятным для мира рационального научного знания, как если бы это был чужой мир.) Задача ученого, ведущего включенное наблюдение, — «показать, как социальное действие в одном мире может быть понято (осмыслено) с точки зрения другого мира»¹⁹.

¹⁶ Boggs V. Finding Your Spot // Smith C. D., Kornblum W. (eds.) In the Field: Readings on the Field Research Experience. N. Y., 1989. P. 147–152.

¹⁷ См.: Merton R. K. Introduction: Notes on Problems-Finding in Sociology // Merton R. K., Broom L., Cottrell L. S., Jr. (eds.) Sociology Today. N. Y., 1959. Vol. 1.

¹⁸ Более подробное и вполне ясное изложение можно найти в книге: Agar M. H. Speaking of Ethnography.

¹⁹ Ibid. P. 12.

Отсюда ясно, что недоумение, неясность, несоответствие теоретическим ожиданиям, иными словами разрывы и «неисправности» в знании и взаимопонимании, и являются исходным материалом для ученого. Задача этнографического понимания — обнаружить и зафиксировать разрыв, чтобы в дальнейшем дать объяснение, этот разрыв исключающее. Как только объяснение «странному обычаю» получено, разрыв перестанет восприниматься как таковой. И ученый, и читатели, к которым первый адресует свое изложение открывшегося ему смысла действия, перестанут воспринимать обычай как «странный», то есть, например, не будут больше расценивать как необычное то обстоятельство, что цыганки не гадают друг другу, что отец в традиционной кавказской семье никогда не берет ребенка на руки (хотя, по всей видимости, любит его) или что профессиональные ученые часто придают большее значение публикации результатов, чем собственно их получению.

Однако движение от разрыва и недоумения к пониманию — это не только коррекция исходной формулировки теоретической проблемы. Конечным результатом этнографического понимания является *слияние двух или более культурных традиций — этнографа, изучаемого им сообщества, аудитории*²⁰. Наблюдатель становится посредником между различными социальными мирами, расширяющим горизонты культурных традиций и способствующим их коммуникации²¹.

Социолог в такой трактовке самым очевидным образом оказывается в одной из главных своих профессиональных ролей — посредника между социальными сообществами и культурами²².

²⁰ Agar M. H. Op. cit. P. 20.

²¹ Подробнее об этом см.: Geertz C. From the Native's Point of View: On the Nature of Anthropological Understanding // Rabinow P., Sullivan W. M. Interpretive Social Scena: A Reader. Berkeley, 1979.

²² В одной из работ З. Баумана сделана попытка показать, что эта роль посредника и переводчика в наше время вытесняет традиционную роль социолога-эксперта, дающего советы просвещенным правителям. Бауман подчеркивает, однако, следующее важное обстоятельство: посредническая роль социолога и — шире — интеллектуала, его открытость к пониманию разных «способов жизни» не должны вести к отказу от его собственной традиции рационального объяснения и интеллектуальной честности. См.: Bauman Z. Legislators and Interpreters. Cambridge: Polity Press, 1987.

Возвращаясь к обсуждению выбора теоретической проблемы и предмета включенного наблюдения, заметим, что описанные различия между группами, культурами и системами значений делают особенно важной проблему *сравнения*, то есть *выбора групп, ситуаций и условий* для проведения этнографического исследования.

Какие «случаи» считать релевантными, значимыми, существенными для данной исследовательской проблемы? Прежде чем ответить на эти вопросы, отметим, что включенное наблюдение можно рассматривать как некую разновидность (возможно, самую распространенную) *метода монографического «анализа случая» (case-study)*. Под последним принято понимать детальное, целостное описание индивидуального случая, включенного в более широкий социальный и культурный контекст. В качестве «случая» могут рассматриваться культура, сообщество, субкультура, организация, социальная группа, а также такие явления, как верования, практики, формы взаимодействия — иными словами, почти все аспекты человеческого существования²³. Анализ случая может включать в себя интервьюирование, включенное наблюдение, анализ личных документов, литературных источников. Весь этот широкий круг методов объединяет идея максимально полного описания критически важного для прояснения данной исследовательской проблемы случая (или) нескольких случаев. В отличие от массовых опросов, ориентированных на сбор данных о больших популяциях, методология анализа случая не придает большого значения статистической репрезентативности полученных данных. Возможность обобщения и переноса выводов исследования в более широкий контекст здесь обосновывается через «типичность» случая, через возможность теоретического объяснения выбора данного объекта, места и времени его изучения.

Критики методологии «анализа случая» и, соответственно, включенного наблюдения часто (и справедливо) подчеркивают возможность систематических смещений и необоснованных обобщений, выводимых из исследования единичного явления. Особую остроту,

²³ См.: *Forgensen D. L. Participant Observation: A Methodology for Human Studies*. Newbury Park et al.: Sage, 1989. P. 19–20. (Applied Social Research Methods Series. Vol. 15.)

таким образом, приобретает *проблема отбора* — случаев, ситуаций, групп — и обоснования переносимости результатов включенного наблюдения в более широкий контекст.

Самой успешной попыткой справиться с проблемой отбора в этнографическом методе стала последовательная разработка понятия *теоретической выборки*, впервые предпринятая Б. Глезером и А. Строссом (1967)²⁴. Выбор исследуемого явления здесь обосновывается через логику проверяемой теории, определяющей, какие особенности данного явления (случая, группы и т. п.) существенны с точки зрения содержательных, теоретических соображений. Хотя основанный на включенном наблюдении анализ случая обычно не подразумевает использования статистических процедур репрезентативного отбора, эти процедуры могут использоваться для селекции наблюдений «внутри» данного случая, построения сравнительных групп и т. п.

Чтобы проиллюстрировать эти несколько абстрактные соображения, обратимся к уже упоминавшемуся исследованию социальных контекстов умирания, проведенному Глезером и Строссом²⁵.

Исследование Глезера и Стросса проводилось в шести больницах, расположенных в прибрежном районе Сан-Франциско. В самой общей форме исследовательская проблема была следующей: какого рода события происходят вокруг пациентов, умирающих в американских больницах? На предварительной стадии исследования эта общая проблема сузилась до нескольких вопросов: каковы устойчивые типы *взаимодействия* между умирающим пациентом и персоналом больницы; какого рода тактики использует медицинский персонал в отношении пациента; в каких организационных условиях внутри больницы эти типы взаимодействия и тактики имеют место и как они влияют на пациента, его семью, медиков, больницу как целое, всех тех, кто вовлечен в ситуацию, окружающую процесс смерти?²⁶ В поиске ответов на эти вопросы исследователи пришли к формулировке следующей теоретической гипотезы: все происходящее может быть объяснено тем, как и в какой

²⁴ Glazer B., Strauss A. The Discovery of Grounded Theory.

²⁵ Glazer B., Strauss A. L. Awareness of Dying.

²⁶ Ibid. P. 8.

мере *осознается* судьба пациента каждой из взаимодействующих сторон в ситуации умирания. Можно сформулировать эту гипотезу еще проще: важно «кто — в ситуации умирания — что знает о вероятности фатального исхода для умирающего пациента»²⁷. Ключевой теоретической переменной в исследовании Глезера и Стросса стало, таким образом, понятие *контекста осознания* (или «контекста знания») о приближающейся смерти.

Чтобы сделать яснее теоретическую логику этого подхода, заметим, что он основан на интерпретативной, интеракционистской традиции социологического мышления. Напомним, что с точки зрения этой традиции за любым социальным взаимодействием стоит постоянное осознание и осмысление людьми их повседневной жизни. Люди не просто пытаются осмыслить и осознать причины и последствия поступков и событий, они взаимодействуют и совершают поступки, основываясь на тех смыслах, которые они приписывают событиям повседневной жизни²⁸.

Верно или неверно люди определяют ситуацию и толкуют события и намерения других людей — что во многих случаях в принципе не поддается оценке, — они реально руководствуются своими мнениями и убеждениями в своих поступках. Если воспользоваться общеизвестной формулировкой: если люди определяют ситуацию как реальную, она реальна по своим последствиям²⁹.

Глезер и Стросс, основываясь на своих представлениях о возможных «контекстах осознания» приближающейся смерти, осуществили *теоретическую выборку* мест, условий и ситуаций внутри

²⁷ Ibid. P. IX.

²⁸ См., в частности: *Schutz A. The Phenomenology of the Social World / Transl. by G. Walsh, F. Lehnert. Evanston: Northwestern University Press, 1967; Blumer H. Symbolic Interactionism. Englewood Cliffs: Prentice Hall, 1969*, а также применительно к методам социального исследования: *Denzin N. Interpretive Interactionism. L. et al.: Sage, 1989*.

²⁹ *Thomas W. I., Thomas D. S. The Child in America. N. Y.: Khopf, 1928*. Заметим, однако, что эта теоретически продуктивная позиция становится крайне спорной, как только начинает восприниматься как утверждение о возможности полного, безостаточного сведения (редукции) мира поступков, наблюдаемого поведения к миру сознания и смыслов. В этом случае знаменитая фраза становится столь же плоской и бессодержательной, как и противоположная крайняя позиция: поступки, которые люди действительно совершают, реальны по своим последствиям.

больницы, относительно которых можно было предположить, что они представляют все типичные «контексты осознания» смерти и все типы взаимодействия, происходящие в этих контекстах. Например: «Когда пациента доставляют в больницу в необратимой коме и ставят диагноз неизбежной смерти, никто не предполагает, что пациент может когда-нибудь узнать о своем диагнозе. Не исключена возможность того, что врачи и медсестры могут по-разному определить статус пациента, но такое расхождение маловероятно. С другой стороны, когда пациент поступает в сознательном состоянии и нельзя с полной определенностью сказать, умирает ли он, то, конечно, его собственное определение (ситуации) может резко расходиться с определением, даваемым медиками, которые, в свою очередь, тоже могут разойтись во мнениях. То, что *каждый* из участников взаимодействия знает о том, как было определено состояние пациента, наряду с признанием каждым участником того, что другие участники знают о его собственном определении ситуации, — всю картину, как ее увидел бы социолог, — мы будем называть *контекстом осознания*. Это тот контекст, внутри которого люди взаимодействуют, в то же самое время осознавая его»³⁰.

Соответственно исследователи осуществили отбор мест и условий наблюдения — от реанимационной палаты и онкологического отделения, где смерть является частым и в разной степени осознаваемым событием, до акушерского отделения, где неизбежность летального исхода обычно оказывается драматической неожиданностью не только для пациента и его близких, но и для медицинского персонала.

Еще раз повторим, что целью отбора (теоретической выборки) субъектов, групп, места, условий и времени наблюдения было нахождение всех теоретически сконструированных возможных значений главной объяснительной переменной — «контекста осознания». Возможные значения этой переменной (типы «контекстов осознания») были получены в результате комбинирования знания/незнания о состоянии пациента для каждой из сторон (медперсонал, пациент, родные) и учета возможного «притворства» — стремления скрыть от другой стороны свое осознание того, как эта другая

³⁰ Glazer B., Strauss A. L. Awareness of Dying. P. 9–10.

сторона определяет ситуацию. Например, онкологический пациент может знать о том, что врачи считают его инкурабельным, однако вести себя так, чтобы не позволить врачам явным образом определить сложившуюся ситуацию; еще очевиднее случай, когда врач манипулирует «закрытым контекстом осознания» и скрывает от подозревающего правду пациента не только то, что «ничего больше нельзя сделать», но и свое осознание наличия подозрений у пациента. Построение такой своеобразной теоретической выборки помогло Глезеру и Строссу не только отобрать случаи и условия наблюдения, но и провести анализ полученных таким образом «сравнительных групп»³¹.

Идея теоретической выборки — или, выражаясь точнее, теоретического отбора — относится, как мы увидели на примере работы Глезера и Стросса, не только к отбору случаев для изучения, но и к *отбору внутри случаев*, то есть к отбору *времени, места и людей* для наблюдения. В приведенном нами примере отбор места для полевой работы — это отбор палат, отделений и служб внутри медицинской организации. Отбор времени для наблюдения также предполагает введение каких-то разумных ограничений: исследователь не может быть круглосуточным наблюдателем «в поле», даже если он все время находится там. (Впрочем, М. Агар вспоминает о своем опыте полевой работы: «После нескольких месяцев, проведенных в Индии, я как-то сидел в своей хижине, читая книгу при свете фонарика и расслабленно прислушиваясь к невнятным шумам сумерек в “танде”. Внезапно дверь отворилась, и я услышал — если перевести это очень приблизительно — следующее: “Где твоя записная книжка? У нас здесь как раз происходит важная церемония. Что случилось, ты не работаешь нынче вечером?”»³².) Такие разумные ограничения обычно также связаны с содержательными представлениями о том, что и в каком порядке обычно происходит в наблюдаемом сообществе, какие *временные рутины* и *расписания* определяют последовательность значимых и незначимых событий. Социолог, наблюдающий школьный класс и взаимоотношения де-

³¹ Glazer B., Strauss A. L. Awareness of Dying. P. 286–293. (Appendix.)

³² Agar M. H. The Professional Stranger: An Informal Introduction to Ethnography. N. Y. et al.: Academic Press, 1980. P. 61.

тей во время уроков, по всей вероятности, будет вести свои наблюдения в дообеденное время; если же он изучает, скажем, изменения стилей управления в производственном подразделении, он постарается понаблюдать попеременно и за утренними, и за вечерними сменами. То же относится к большим временным циклам: сезонам, годам, смене поколений. Отбор людей — это обычно отбор интервьюируемых или информантов. Мы вернемся к этой теме, ограничившись пока замечанием, что и здесь определяющую роль играет теоретический контекст исследования. Существенными параметрами отбора могут быть такие категории, как пол, возраст, ранг в групповой иерархии, уровень осведомленности и т. п. Принято различать «исследовательскую категоризацию», конструируемую наблюдателем на основании признаков, существенных для принятой им теоретической перспективы, и «членскую категоризацию», то есть ту классификацию, которую сами члены группы считают существенной в повседневной жизни³³. Однако польза этого различия относительна: если считать, что преимуществом этнографического метода действительно является гарантируемая им особая близость точки зрения исследователя тому, что сами члены группы считают существенным и определяющим в своей жизни, исследовательская и членская категоризации не должны резко противостоять друг другу (не считая возможных терминологических и лексических расхождений).

Иногда говорят об отборе *контекстов наблюдения*. Контекст в данном случае представляет собой несколько абстрактное понятие, включающее в себя не только время, место и общую структуру взаимодействия, но и некую — обычно неявную — совокупность норм (нормативную структуру), регулирующих поведение людей в данных обстоятельствах места и времени. В этом смысле можно говорить о различных контекстах наблюдения во время рабочего совещания, или в ходе неформального празднования какого-то события внутри одной и той же организации, или, скажем, о различии контекста семейного взаимодействия в присутствии гостей и на кухне. В еще более обобщенной форме различие контекстов социального

³³ См.: Lofland J. *Doing Social Life: The Qualitative Study of Human Interaction in Natural Settings*. N. Y.: Wiley, 1976.

взаимодействия может быть описано с помощью введенного И. Гоффманом противопоставления *сценических* и *закулисных областей*³⁴. Именно «за кулисами» (хотя и не обязательно на кухне в физическом смысле) супруги выясняют отношения, не руководствуясь более нормативной структурой публичного поведения, и там же — не обязательно в географически определенном месте — выясняется, скажем, ранг ученого среди коллег. Разнообразие контекстов, которое следует принимать во внимание, — подчеркнем это еще раз — это разнообразие социально сконструированных³⁵, а не физически заданных мест взаимодействия.

Выбор исследовательской проблемы и ситуации наблюдения — это результат предварительной стадии исследования. Завершение этой стадии ставит социолога перед другой совокупностью теоретических и практических вопросов, связанных с получением доступа в полевую ситуацию и непосредственным вовлечением во взаимоотношения с интересующими его группами.

Вхождение в ситуацию наблюдения, роли наблюдателя, взаимоотношения «в поле»

Проблема получения доступа к полевым данным — на первый взгляд, сугубо практическая — играет ключевую роль в этнографическом методе.

Постольку, поскольку эта проблема может быть разрешена за счет личных психологических и социальных ресурсов и практических стратегий, которыми располагает социолог-наблюдатель, можно говорить о значении здравого смысла и знания повседневной жизни в использовании этнографического метода. С другой стороны, более или менее эффективные попытки включиться в ситуацию наблюдения, в том числе трудности, с которыми социолог сталкивается на этом пути, часто оказывают существенное влияние на теоретическую логику и степень понимания ученым того, что

³⁴ См.: *Goffman E. The Presentation of Self in Everyday Life. N. Y.: Doubleday, 1959.*

³⁵ Под «социальной сконструированностью» здесь достаточно пока понимать смысловую и ценностно-нормативную определенность, то символическое «поле» смыслов, в котором происходит взаимодействие.

он наблюдает. Некоторые примеры позволяют прояснить эти соображения.

Один из этих примеров относится не столько к социологии или этнографии, сколько к тому, что часто называют документальной журналистикой. Журналистское расследование ситуации нередко принимает форму включенного наблюдения, и неудивительно, что у истоков использования методов «анализа случая» в американской социологии стояла, помимо культурно-антропологической (этнографической) традиции, так называемая журналистика факта. Наш пример относится к 1960-м годам, когда молодой и честолюбивый журналист Том Вулф предпринял квазиэтнографическое исследование сообщества хиппи, называвших себя «веселыми шалунами». Ядро этой коммуны составляли Кен Кизи (автор знаменитого романа «Кто-то пролетел над гнездом кукушки») и его друзья-хиппи.

Позднее Вулф написал книгу о своем опыте общения с хиппи³⁶. Однако в начале своего исследования Вулф демонстрировал определенную дистанцию по отношению к мало озабоченным благопристойностью, карьерой и «традиционными американскими ценностями» хиппи. Как-то раз он беседовал со своими новыми знакомыми в комнате, где Кизи красил потолок. На безукоризненный белый льняной костюм журналиста упала изрядная капля желтой краски, и, хотя последний вытер пятно, сохраняя невозмутимость истинного джентльмена, он не смог скрыть некоторого раздражения. Кизи философски изрек: «Уж так это устроено, Том. Если ты хочешь войти в это дело, тебе приходится немного в него вляпаться».

Иногда для того, чтобы получить доступ в ситуацию включенного наблюдения, достаточно просто «слоняться поблизости». Эллиот Лайбоу, участвовавший в большом исследовательском проекте по изучению практики воспитания детей в семьях с низким доходом, проводил включенное наблюдение за мужчинами из этих семей, чтобы дополнить данные семейного интервьюирования³⁷. Первый день исследования оказался не очень продуктивным. Хотя

³⁶ *Wolfe T.* The Electric Kool-Aid Acid Test. N. Y.: Bantam Books, 1983. Цит. по: *Smith C. D., Kornblum W. (eds)* In the Field: Readings on the Field Research Experience. N. Y.: Praeger, 1989. P. 2–4.

³⁷ *Liebow E.* Tally's Corner. L.: Routledge and Kegan, 1967.

Лайбоу и познакомился с одним из зевак, наблюдавших сцену препровождения шумно сопротивлявшейся женщины в полицейский участок, он не выполнил разработанного им заранее плана — приступить к сбору материала для выполнения трех или четырех исследовательских задач со сравнительно четкими границами между ними. «Завтра, — решил я, — я вернусь к моему исходному плану, еще ничего не потеряно. Но завтра никогда не наступило...»³⁸. На следующий день, беседуя с тремя пьянчужками об уходе за щенком, которого один из них держал за пазухой, Лайбоу опять оказался у угловой «точки», торговавшей навынос. В этом угловом магазинчике, ближайшие окрестности которого стали неизбежным стратегическим центром всех его этнографических изысканий, Лайбоу познакомился с хорошо одетым чернокожим молодым человеком Толли Джексоном, ставшим его попечителем, доверенным лицом и другом и открывшим ему доступ в относительно закрытые области своего социального окружения. Книга, написанная Лайбоу на полученном материале, стала одной из ключевых работ по этнографии города.

Другой пример значимости неформального «попечительства» — взаимоотношения Уильяма Ф. Уайта с его ключевым информатором — лидером местной «братвы» Доком, сыгравшие решающую роль в проведенном Уайтом исследовании, упоминавшемся нами ранее³⁹.

Иногда доступ к «попечителям» и ключевым информаторам открывается не в результате каких-то полевых импровизаций, а в ходе использования уже существующих социальных связей — профессиональных, дружеских, родственных и т. п., — а также через использование собственной идентичности исследователя⁴⁰. Если

³⁸ Ibid.

³⁹ Whyte W. F. Op. cit.

⁴⁰ Следует упомянуть также более редкий случай, когда ученый, вынужденно и по далеком от науки причинам оказавшийся в некоторой ситуации, решает использовать эту ситуацию в исследовательских целях. В сущности, именно таким образом оказался на Тробриандовых островах выдающийся антрополог Б. Малиновский, интернированный как подданный враждебной державы («Argonauts of the Western Pacific», 1922). Отечественный исследователь Ю. Д. Карпов, находившийся в административной ссылке в Амурской области в 1969–1972 гг., осуществил комплексный анализ жизни сельской общины (устное сообщение канд. экон. наук, доц. НГУ Е. Е. Горяченко, участвующей в подготовке материалов исследования Ю. Д. Карпова к изданию).

вернуться к уже анализировавшимся нами примерам, то можно отметить, что Э. Моралес, изучавший «кокаиновую экономику» в Перу, проводил свое исследование в местах, где когда-то родился и рос, что — наряду со знанием местного диалекта и обычаев — обеспечило его проникновение в мир спрятанных в горах кокаиновых лабораторий и тайных троп, по которым местные крестьяне перевозили готовый продукт. Однако даже в этом случае личностная идентичность исследователя всегда недостаточна для того, чтобы получить автоматический доступ ко всем аспектам изучаемой ситуации. В частности, Моралес пишет: «Во время моего приезда в родной городок, весной 1980 года, я решил выполнить совет, данный мне профессором в колледже: “Езжай домой и посмотри, какие изменения произошли в общине”. Ни в детстве, ни позднее, в ходе многочисленных приездов домой, я не отваживался выезжать за границы моей родной общины, Лламеллик. В юности я принадлежал к коренной культуре, но как взрослый визитер уже не располагал необходимыми навыками, чтобы понимать очевидное. Хотя я и был местным уроженцем, мое превращение в исследователя этнографии Анд было долгим. Я обнаружил, что очень трудно наблюдать за людьми, культура и общество которых представляют и твои собственные корни, трудно учиться у этих людей. Это оказалось возможным только тогда, когда я включил в свою повседневную жизнь более дисциплинированный подход»⁴¹.

Двойственность позиции исследователя, чья личностная идентичность в существенных чертах совпадает с идентичностью тех, кого он изучает, подчеркивает и Б. Майерхоф, проводившая исследование в еврейском доме для престарелых в Калифорнии. С одной стороны, ее собственные еврейские корни и предшествовавший научной карьере некоторый опыт социальной работы с пожилыми людьми облегчали ей доступ «в поле». С другой стороны, эти же факторы создавали определенные трудности в *сохранении объективности* и в *построении роли исследователя*, отличной от роли члена изучаемой общности. К тому же, как уже отмечалось, частичное совпадение личностной идентичности исследователя и

⁴¹ Morales E. Cocain. Цит. по: Smith C. D., Kornblum W. (eds) In the Field: Readings on the Field Research Experience. P. 116.

исследуемых по определению не может быть полным: в случае Майерхоф эта неполнота проявляется в том, что она была молодой, имевшей мужа и детей, сравнительно здоровой и сделавшей довольно успешную научную карьеру американкой во втором поколении. При этом она изучала процесс старения и способы преодоления возрастных проблем среди одиноких стариков, перебравшихся в Америку, спасаясь от нацизма, часто говоривших между собой на идише, воспитанных в иудейской традиции и сохранивших специфическое мировоззрение восточноевропейского «галута»⁴². Майерхоф так описывает исходную ситуацию: «Я не принимала никаких сознательных решений исследовать свои корни или сделать яснее смысл моего происхождения. Я была одним из нескольких антропологов из университета Южной Калифорнии, вовлеченных в анализ “Этничности и старения”. Сначала я планировала изучать пожилых “чиканос”⁴³, поскольку прежде я уже занималась полевой работой в Мексике. Однако в начале 1970-х в городской Америке этнические группы не очень приветствовали любознательных посторонних, и люди, к которым я обращалась, постоянно спрашивали меня: “Зачем вам работать с нами? Почему бы вам не изучать своих?” Эта идея была для меня новой. Я не была подготовлена к такому проекту. Антропологи обычно исследуют экзотические, отдаленные, дописьменные общества. Но такого рода группы становятся во все большей степени недоступны и часто — негостеприимны. В результате все больше антропологов вынуждены нынче работать у себя дома. Это неизбежно создает проблемы с объективностью и идентификацией, и я предвидела, что тоже получу причитающуюся мне долю этих проблем, работая с людьми, населявшими Центр. Но, возможно, будут и какие-то преимущества...»⁴⁴.

Проблемы, с которыми столкнулась Майерхоф, были связаны с необходимостью соблюдения баланса между исследовательскими интересами, интересами тех, кого изучает исследователь, этическими проблемами (зачем, например, изучать людей, нуждающихся

⁴² Галут — изгнание, рассеяние (*иврит*).

⁴³ Чиканос — потомки мексиканских иммигрантов.

⁴⁴ *Myerhoff B.* Op. cit.

в поддержке и участии, вместо того чтобы просто помочь им?) и задачей сконструировать свою личностную тождественность собственно исследовательской роли. Однако прежде чем осознать и отчасти решить эти проблемы, исследовательница прошла через ряд довольно драматических событий и изменений: «В начальных фазах моей работы с пожилыми я испытывала острое чувство вины. Оно периодически всплывало наружу, принимая самые разные формы в разное время. Поначалу оно фокусировалось на моей компетентности в решении задачи, за которую я взялась. Достаточно ли я знаю иудаику? Достаточно ли я знаю идиш? Не слишком ли я молода? Не слишком ли я эмоционально вовлечена в ситуацию? Не следует ли мне работать ради благополучия стариков, вместо того чтобы изучать их? И так далее. В ходе разговора со Шмуэлем — очень ученым человеком, ставшим в дальнейшем одним из главных моих информантов, — я призналась, что испытываю страх не разобраться должным образом с теми материалами, которые он мне дает. Было так много вещей, которых я не понимала. Как всегда, он ответил сурово, но справедливо: “Ты не понимаешь. Как ты могла ожидать, что поймешь? Ты спрашиваешь меня обо всех этих вещах, но сама ты ничего не знаешь. Ты не знаешь идиша. Ты не знаешь иврита. Ты не знаешь арамейского. Ты не знаешь ни русского, ни польского. Ты не присмотрелась внимательно ни к какой части того места, где мы живем. Как ты можешь ожидать, что поймешь?” Я согласилась с ним и чувствовала себя чудовищно растерянной... Я подумывала о том, чтобы бросить все. Казалось невыносимым безропотно принимать все те бесчисленные способы использования чувства вины, к которым прибегали обитатели Центра — часто несознательно, намереваясь не причинить вред, а лишь создать у себя ощущение могущества. Но после некоторого времени я приняла как факт, что никого нельзя “сделать” виновным, заставить испытывать чувство вины. Каждый соглашается на это добровольно. Пробуждение вины — это то, что я называю “стратегией интимности”, одной из многих, используемых стариками из Центра. Бесплезная по отношению к чужаку, эта стратегия основывается на взаимосвязанности и взаимозависимости. Светлая сторона вины заключается в том, что она выражает чувство ответственности за благополучие другого человека. Когда я осознала это, я стала более

отстраненно и даже с благодарностью воспринимать эти свои чувства»⁴⁵.

В ситуации, когда исследователю требуется доступ в формальную организацию, ключевую роль начинают играть не столько добровольные «спонсоры», сколько обладающие высоким статусом и контролирурующие «входы» в эту организацию влиятельные фигуры. Иногда таких лиц называют «стражниками»: именно от них зависит удачный исход переговоров о формальном доступе в организацию.

Иногда довольно трудно решить, кто в данной ситуации является «стражником» — например, оперирующий главный врач в больнице может в действительности меньше участвовать в принятии административных решений, чем контролирующая все перемещения персонала, больных и оборудования главная медсестра.

Получение разрешения со стороны «стражников» и поддержки «попечителей» создает лишь минимальные предпосылки доступа к полевым данным. Однако цена, которую «платит» исследователь за получение исходного доступа, — это стремление ключевых фигур оказывать влияние на ход и результаты исследования. Первая возникающая здесь проблема носит скорее этический характер: насколько позволителен прямой обман или манипулирование собственным «имиджем» со стороны исследователя? Как правило, социологи легче идут на обман и полное или частичное сокрытие целей исследования в том случае, когда доступ жестко контролируется и «стражники» по каким-либо причинам не заинтересованы в том, чтобы открыть посторонним какие-то, обычно скрываемые, сферы. Еще одним оправданием для полного или частичного сокрытия целей исследования в процессе ведения переговоров о доступе может служить уверенность исследователя в том, что на более поздней стадии, когда между социологом и ключевыми фигурами установится межличностное доверие, можно будет обсудить все открыто. В этом случае заверения в соблюдении анонимности и в исключении из публикации потенциально опасных для информантов материалов попадут на более благоприятную почву и будут встречены с большим доверием.

⁴⁵ Myerhoff B. Op. cit. P. 89.

Вообще же, как справедливо замечают М. Хаммерсли и П. Аткинсон, «...даже если сообщение “всей правды” в ходе переговоров о доступе — как и в большинстве прочих социальных ситуаций — может оказаться не самой умной и даже не самой окупающей себя стратегией, обмана следует по возможности избегать. И дело здесь не только в сугубо этических причинах, но и в том, что обман может иметь неблагоприятные последствия позднее, в ходе полевой работы»⁴⁶.

Наблюдатель, успешно включившийся в полевую работу, оказывается в определенный момент в положении «новичка» («салаги» на армейском жаргоне). Он точно так же располагает статусно-детерминированным правом проявлять невежество, задавать вопросы, обнаруживать некомпетентность в простых, повседневных вещах. Однако социолог, в отличие от традиционного этнографа, редко может использовать до конца все преимущества роли «новичка». Если его профессиональная идентичность известна, его «попечители», информанты и прочие могут успешно навязывать ему роль знатока, эксперта, беспристрастного судьи и т. п.

Еще более важное разграничение исследовательских ролей во включенном наблюдении предложил Р. Гоулд⁴⁷. Это разграничение основано на степени *вовлеченности–отстраненности* исследователя в ситуации наблюдения и, соответственно, на степени *закрытости–открытости* его собственно этнографической, научной деятельности. В этом случае принято выделять следующие роли:

- 1) полный участник;
- 2) участник как наблюдатель;
- 3) наблюдатель как участник;
- 4) полный наблюдатель.

В роли «полного участника» цели и статус исследователя остаются тайной для всех остальных, поэтому эту ситуацию нередко называют ситуацией скрытого наблюдения. Выше мы уже приводили пример исследования Л. Фестингера и его соавторов.

⁴⁶ Hammersley M., Atkinson P. Op. cit. P. 71–72.

⁴⁷ Gold R. L. Roles in Sociological Fieldwork // Social Forces. 1958. Vol. 36. P. 217–223. См. также: Junker B. Field Work. Chicago: University of Chicago Press, 1960.

В отечественной социологии одним из самых известных примеров скрытого наблюдения является исследование В. Б. Ольшанского⁴⁸. Этот подход, наряду с очевидными достоинствами, имеет и существенные недостатки: дело здесь не только в неразрешимости некоторых этических проблем, но и в невозможности гарантированного контроля над использованием наблюдателем своих актерских способностей, что, наряду с риском утраты профессиональной идентичности и необходимой исследователю дистанции, может вести к самым неожиданным результатам.

Находящаяся на противоположном полюсе роль «полного наблюдателя» подразумевает полное исключение реакций изучаемых людей на исследователя. Последний, как это бывает в некоторых психологических экспериментах, смотрит на обследуемых через одностороннее зеркало, ведет наблюдение скрытой камерой либо использует иные приемы оперативной разведработы. Как отмечают М. Хаммерсли и П. Аткинсон, «парадоксальным образом полное наблюдение разделяет многие преимущества и недостатки полного участия. В их пользу говорит то, что оба метода минимизируют проблемы реактивности: ни в одном из случаев этнограф не будет взаимодействовать с теми, кого он изучает, *как исследователь*. С другой стороны, исследователь может столкнуться с сильными ограничениями на то, что подлежит или не подлежит наблюдению, при том что расспросить участников будет невозможно. Принятие какой-либо одной из этих ролей может вести к большим сложностям в построении и строгой проверке теории, хотя обе могут оказаться полезными и приемлемыми стратегиями на отдельных стадиях полевой работы, а в некоторых ситуациях эти роли оказываются неизбежными»⁴⁹.

Чаще же всего наблюдатель принимает роль, находящуюся между описанными крайними позициями. При этом, заметим, ситуация *скрытого наблюдения* чаще всего ведет к принятию *уже существующей* в наблюдаемой группе *роли*, например рабочего или пациента психиатрической клиники. В случае *открытого наблюдения*

⁴⁸ Ольшанский В. Б. Личность и социальные ценности // Социология в СССР. М., 1966. Т. 1.

⁴⁹ Hammersley M., Atkinson P. Op. cit. P. 96.

часто — хотя и не обязательно — исследователь участвует в *конструировании* своей роли в процессе явных и неявных переговоров с участниками.

Степень вовлеченности социолога в наблюдаемую ситуацию, играющая ключевую роль в описанной нами типологии ролей, тесно связана с другим противопоставлением: описание с точки зрения наблюдающего «чужака» — описание с точки зрения участвующего в событиях «своего» члена группы. Мы уже затрагивали этот вопрос раньше, отмечая, что чрезмерная включенность в происходящее может не только стимулировать, но и, напротив, сильно ограничивать возможности исследователя. Необходимость постоянно балансировать между доверительными отношениями и возможностью отстранения от происходящего, между «близостью» и «чуждостью», между следованием «туземным» правилам и собственным комфортом делает результаты применения этнографических методов чрезвычайно зависимыми от способности исследователя придерживаться пограничной, сбалансированной и, в сущности, маргинальной позиции. Эта особенность метода включенного наблюдения делает его весьма уязвимым для острой и отчасти обоснованной критики, о чем мы еще будем говорить позднее. Пока же обратимся к примеру.

Классик американской культурной антропологии Маргарет Мид во второй половине 1920-х годов проводила включенное наблюдение процесса взросления на Самоа. Материалы этого исследования стали основой для самой знаменитой и спорной книги Мид⁵⁰, ставившей целью доказать «сильную» версию культурного детерминизма и идеи «суверенитета культуры» по отношению к биологическим предпосылкам человеческого существования. В этой книге, рассматривавшейся современниками как блестящее эмпирическое подтверждение взглядов учителя М. Мид и основателя американской традиции культурной антропологии Ф. Боаса, Мид интерпретировала самоанскую культуру как нерепрессивную, несоревновательную, поощряющую скорее кооперацию, чем агрессию. Многие фактические утверждения Мид были позднее оспорены другими исследователями. Особую известность приобрела

⁵⁰ Mead M. *Coming of Age in Samoa*. N. Y.: Morrow, 1923.

книга Д. Фримена⁵¹, где на материале многолетних наблюдений опровергаются ключевые положения М. Мид и детально реконструируется исторический фон написания ее труда. Фримен полагает, что основной причиной неточностей и прямых ошибок в интерпретации самоанской культуры, предложенной М. Мид, была предвзятая и жесткая теоретическая схема. Немалую роль, однако, сыграла и социальная и психологическая неискушенность двадцатитрехлетней аспирантки, неопределенность ее исследовательской роли. В частности, Мид, получившая изначально хорошие шансы доступа в поле (она была даже наделена почетным статусом «церемониальной девственницы», *taipou*), не решилась жить в самоанской хижине с туземной семьей и испросила у Боаса разрешения поселиться в единственной белой семье на острове (мотивируя это возможным снижением продуктивности исследовательской работы в результате непривычной диеты)⁵².

Приведенный пример еще не доказывает, что сохранение дистанции между социологом и теми, кого он наблюдает, заведомо хуже, чем полное участие. В действительности, как мы уже отмечали, полная включенность подразумевает принятие исследователем всех нормативных ограничений, которым следуют настоящие участники. Например, социолог, поступивший на военную службу для того, чтобы узнать, как новобранцы приспосабливаются к новой социальной роли, будет всегда оставаться «по одну сторону занавеса»: сложившаяся рутина военной службы не подразумевает проявления особой любознательности со стороны рядового и таким образом он не может исследовать, как офицеры и старшины принимают решения, общаются между собой, определяют правила и нормы и т. п.

Иными словами, социологу, оказавшемуся в реальном полевом контексте, приходится использовать самые разные роли в поисках компромисса между объективностью, профессиональной автономией, эмпатической вовлеченностью и интеллектуальной честностью. Достигнуть такого компромисса удастся далеко не всегда.

⁵¹ Freeman D. Margaret Mead and Samoa: The Making and Unmaking of an Anthropological Myth. Cambridge: Harvard University Press, 1983.

⁵² Ibid. P. 66–68.

Процесс анализа и описания результатов

В результате включенного наблюдения, интервьюирования, неформализованных бесед с информаторами социолог получает изрядное количество данных, которые следует каким-то образом организовать, описать и проанализировать.

Было бы ошибкой полагать, что анализ и описание в этнографии — это искусство загонять «сырые» факты в упорядоченные аналитические категории. Скорее речь идет об искусстве создавать некий метатекст, интерпретирующий тексты интервью и наблюдений. Это искусство ничуть не менее важно, чем умение добывать сведения в полевых условиях, завоевывать доверие, интервьюировать и т. п.

Более того, именно в процессе анализа данных и написания отчета или статьи результаты этнографического исследования приобретают общезначимость и валидность, то есть становятся чем-то большим, чем совокупность «здесь-и-сейчас» сделанных наблюдений. Эксперименты, массовые опросы — благодаря статистическому методу — позволяют получать результаты, которые уже являются некоторыми эмпирическими обобщениями, тогда как уникальность записей бесед и других данных, полученных в ходе наблюдения, очевидна. Возможность их дальнейшего использования прямо зависит от наличия теоретической категоризации, интерпретации и т. п.

Известный культурный антрополог К. Гирц пишет: «Способность убедить читателей (большинство из которых — это люди науки, и практически все участвуют хотя бы на полставки в той особой форме существования, которую мы уклончиво называем “современной”) в том, что читаемый ими текст — это подлинное объяснение, данное кем-то, знающим, как устроена жизнь в некотором месте, в некоторое время, в какой-то группе, — это фундамент, на котором в действительности держится все остальное, чем стремится заниматься этнография — анализировать, объяснять, развлекать, смущать, прославлять, наставлять, оправдывать, удивлять, ниспровергать»⁵³.

⁵³ Geertz C. *Works and Lives: the Anthropologist as Author*. Stanford (Ca.): Stanford University Press, 1988. P. 143–144.

Анализу и описанию данных обычно предшествует их фиксация. Практически невозможно записать все данные, полученные в ходе полевой работы. Помимо возникающих здесь сугубо технических проблем, неизбежно существуют ограничения на способность исследователя осознать и словесно оформить свои «смутные впечатления», неявные догадки и подсознательные озарения. Однако существование таких ограничений не избавляет нас от необходимости максимально полно фиксировать и описывать все данные, на основании которых мы делаем определенные теоретические утверждения.

Нередко в ходе включенного наблюдения социологи используют видео- и аудиотехнику, все большую популярность приобретает метод фотодокументального исследования⁵⁴. Но основным источником эмпирического материала по-прежнему остаются полевые заметки. Их полнота и качество решающим образом зависят от использования некоторой заранее разработанной *системы организации полевых заметок (записей)*. Важность использования такой заранее разработанной системы фиксации наблюдений усиливается тем обстоятельством, что при использовании этнографических методов исследования ученый, как правило, располагает значительно большими возможностями для «корректировки» полученных данных — с целью увеличения степени их соответствия теоретическим положениям, — чем при использовании более «жестких» методов.

Первая проблема организации данных — это *проблема отбора*. Если на предварительной стадии исследования ученый решает, что, где, когда и как наблюдать, то на полевой стадии ему нужно определить, *что* записывать, *как* записывать и *когда* записывать.

Методом записи чаще всего бывает письменный отчет, хотя во многих ситуациях предварительная запись осуществляется с помощью диктофона. Иногда используют также видеозапись и фотосъемку.

Не так просто принять решение о том, когда делать записи. Конечно, немедленная фиксация всех наблюдений предпочтитель-

⁵⁴ См., в частности: *Becker H. S. Doing Things Together: Selected Papers. Evanston (Ill.): Northwestern University Press, 1986. Part 4.*

нее всего. На практике, однако, даже в случае открытого принятия исследовательской роли («наблюдатель-как-участник» либо «полный наблюдатель») социолог не решается использовать метод прямой фиксации наблюдений, чтобы не нарушить естественный ход событий и не подорвать доверие к себе со стороны участников. Человек, который постоянно что-то записывает, действует на нервы окружающим, даже если он получил предварительное согласие с их стороны. Таким образом, большая часть полевых заметок, которые наблюдатель делает на месте, носит очень сжатый и фрагментарный характер и служит, по сути, лишь неким каркасом для описания беседы или ситуации, делаемого исследователем позднее «по памяти». Желательно, чтобы, воссоздавая по памяти некоторую сцену или разговор, ученый не только стремился к максимальной полноте и достоверности, но и использовал *специальные отметки*, указывающие, в частности, на пропуск некоторых слов или событий в записи, либо на то, что какая-то часть описания воссоздана с чьих-то слов, а не фиксирует его личные наблюдения. Полученные записи могут быть более или менее детализированными, но почти никогда не бывают дословными. Как правило, степень детализации и точности обратно пропорциональна охвату, широте воссоздаваемой картины событий.

Очень важной проблемой является *сохранение «естественного словаря»* участников. Эта проблема тем острее, чем больше разница между жизненными мирами самого исследователя и тех, кого он изучает. Реальный, ситуативно-привязанный язык, используемый участниками взаимодействия, позволяет узнать, как члены данной группы или данной культуры категоризируют, описывают и воспринимают события повседневной жизни. Определяя социальную реальность, люди «конструируют» ее: используемые классификации, «ярлыки» и способы означивания задают смысловые горизонты действия для членов данной социальной группы или культуры. Ведя полевые записи, исследователь нередко «подчищает», исправляет или дополняет реплики участников. Это происходит даже в случае расшифровки фонограммы, сделанной с помощью диктофона. По сути, такая первичная интерпретация сказанного не только неизбежна, но и полезна, так как именно пояснения и дополнения, внесенные исследователем, делают более понятными контекст и

эмоциональную тональность происходившего. Важно, однако, организовать записи таким образом, чтобы исходный текст и более поздние комментарии можно было различить.

При ведении полевых заметок важную роль играет также фиксация *невербального поведения*. Мимика, позы, жесты часто обнажают коммуникативные намерения участников взаимодействия даже лучше, чем слова. Во многих культурах значимым и подлежащим интерпретации является *пространственное расположение* участников взаимодействия. Речь здесь идет не только о том, что, скажем, для большинства традиционных культур порядок расположения гостей за столом отражает социальный статус последних. Пространственная организация взаимодействия встроена во многие институты и культурные коды современных обществ: достаточно упомянуть, например, школьный класс, университетскую аудиторию или приемную высокопоставленного чиновника — где секретарь обычно организует разметку открытых, полуоткрытых и конфиденциальных зон с помощью журнала записи посетителей, табличек, стульев, цветочных горшков и реплик типа «туда нельзя!», — чтобы увидеть, как много существенных вещей в повседневной жизни «говорится» без помощи слов и жестов.

Отметим, что многие современные компьютерные программы, предназначенные для организации и первичного анализа качественных данных («Ethnograph», «Ethno» и др.) позволяют вести дневник полевых наблюдений, в котором прямые цитаты отделены от сделанных исследователем описаний, аналитических ремарок и проясняющих дополнений. В этом случае при обращении к полевым данным реже возникают вопросы: «Кто из присутствовавших это сказал?» или «Действительно ли она явно признала это?»

Обычно исследователи фиксируют слова и действия таким образом, чтобы не оставалось никаких сомнений относительно того, где разворачивались события, *в какое время, кто и при каких обстоятельствах* в них участвовал⁵⁵. Своего рода *контрольный лист для фиксации полевых наблюдений* был предложен Дж. Спрэдли. Он включает девять ключевых признаков — «маркеров», фиксация

⁵⁵ Hammersley M., Atkinson P. Op. cit. P. 156–157.

которых дает возможность достаточно полно описать контекст происходящего⁵⁶:

- 1) пространство: физическое местоположение или местоположения;
- 2) актер (деятель): люди, участвующие в происходящем;
- 3) деятельность: совокупность взаимосвязанных действий, совершаемых людьми;
- 4) объект: наличествующие физические предметы;
- 5) акт: отдельные действия, совершаемые людьми;
- 6) событие: множество взаимосвязанных деятельностей, ведущихся людьми;
- 7) время: временное упорядочивание происходящего;
- 8) цель: то, чего люди стремятся достичь;
- 9) чувства: ощущаемые и выражаемые эмоции.

Записывая или воссоздавая по памяти полевые наблюдения, исследователь нередко придумывает и тут же теряет гениальные гипотезы, объяснения и догадки. Поэтому ведение полевых записей требует отчетливой фиксации и выделения по меньшей мере графического — не только реплик участников и сделанных наблюдателем описаний, но и *аналитических примечаний*. Аналитические примечания могут выделяться скобками либо выноситься на поля в качестве маргиналий.

Так как исследователь должен стремиться к сохранению и «сырых» данных полевого наблюдения, и аналитических примечаний, и смысловых категорий, использованных им при первичной фиксации и сортировке материала, возникает необходимость поиска компромисса между *эпизодической* и *семантической* организацией полевых данных. В первом случае доступ к полевым заметкам обеспечивается через указание места и времени наблюдения, тогда как сами первичные описания остаются «нетронутыми» и неклассифицированными. Во втором случае поиск нужных материалов в дальнейшем осуществляется через созданную исследователем систему смысловых категорий. Все эти категории становятся элементарными единицами *аналитического указателя*. Скажем, обратившись

⁵⁶ Spradley J. P. Participant Observation. N. Y.: Holt, Rinehart & Winston, 1980. P. 78.

к аналитическому указателю, социолог может найти все случаи (а при более дробной категоризации — даже отдельные реплики), относящиеся, например, к теме «*Стратегии обмена комплиментами в женском коллективе*». Каждое из найденных таким способом «сырых» описаний содержит какой-то материал, касающийся отношений обмена комплиментами между женщинами. При последующем анализе каждая категория может быть разбита на некоторое количество более дробных субкатегорий посредством введения дополнительных «измерений», то есть дополнительных оснований для классификации⁵⁷. Скажем, стратегии обмена комплиментами могут быть *эквивалентными* (то есть рассчитанными на ответный комплимент или сходное вознаграждение) либо *неэквивалентными* и основанными на отношениях доминирования и зависимости. Обмен комплиментами может быть *детерминирован* стандартной социальной ситуацией или наоборот — *спровоцирован* нормативной неопределенностью и неясностью намерений и ожиданий другой участницы и т. д.

В принципе следует стремиться к максимальному числу первичных категорий. Даже самые странные и необычные способы категоризации могут оказаться неожиданно продуктивными при более глубоком анализе. Наименования категорий могут изобретаться исследователем, заимствоваться из специальной литературы или иметь своим источником ту систему обозначений, которую используют сами информанты.

Если исследователь предпочел использование аналитического указателя, он помечает каждую полевую запись словами или значками, позволяющими определить, какие из категорий имеют отношение к данному событию, наблюдению, разговору.

Другой способ заключается в *физической сортировке* первичных записей по разным аналитическим категориям. В последнем случае каждая полевая запись может быть неоднократно скопирована, разрезана на более мелкие фрагменты, включена в разные разделы отчета и т. п. Иными словами, физическая сортировка требует почти неограниченных возможностей копирования. Наилучшим

⁵⁷ *Straus A., Corbin J. Basics of Qualitative Research: Grounded Theory Procedures and Techniques. Newbury Park et al.: Sage Publications, 1990. P. 69–72.*

3) если реплики участников следуют друг за другом непрерывно — но не накладываясь друг на друга, — их связывают в записи знаком равенства:

К. Л.: Я подожду, пока прозвонит звонок =
Учительница: = Только не стойте у двери.

Знак равенства используется также при продолжении записи высказывания на следующей строке либо при соединении непрерывного речевого потока, перебитого репликой другого участника:

Том: Я раньше курил { больше =
Боб: { Ты раньше курил
Том: = Но я никогда не затягивался;

4) паузы, возникающие в ходе разговора, измеряются в секундах и десятых долях секунд и отмечаются в круглых скобках внутри высказывания одного участника или между высказываниями. Например:

Ольга: Со стороны можно было
= подумать (0,7), что они незнакомы (1,2)
Алла: Так ты была там;

5) знаки пунктуации обычно используются не столько для разделения синтаксических единиц, сколько для характеристики интонационно-мелодического рисунка речи. В частности:

- двоеточие (:) характеризует удлинение звука или слога, после которого оно стоит;
- точка (.) отмечает завершающее понижение тона, не всегда совпадающее с окончанием предложения;
- запятая (,) указывает на продолжающую, «соединяющую» интонацию (не обязательно между подчиненными и сочиненными частями сложного предложения);
- вопросительный знак (?) характеризует восходящую интонацию, в том числе и вопросительную;
- восклицательный знак (!) обозначает эмоционально насыщенный тон, не обязательно собственно восклицание;
- тире (—) указывает на разрывы, запинки, внезапные обрывы и т. п.

Для обозначения восходящей и нисходящей интонаций используются направленные вверх либо вниз стрелки (↑ ↓), смысловое

ударение передается подчеркиванием, громко произнесенные слова выделяются прописными буквами.

Отчетливо слышимые вдохи и выдохи также фиксируются ('ххху' — выдох, 'уххх — вдох):

Алла: 'уххх О! спасибо!

Неанализируемые внеречевые звуки заключаются в двойные круглые скобки: ((чихает)).

Одинарными скобками отмечают сомнительные места в транскрипции, в том числе альтернативные версии расшифровки.

Иногда используют также обозначения направления взгляда (подчеркивание над строкой говорит о том, что взгляд говорящего направлен на слушающего, под строкой — взгляд слушающего направлен на говорящего). Аплодисменты обозначают последовательностью латинских букв «Х»: строчные указывают на умеренные аплодисменты, прописные — на громкие. Их длительность также может быть указана в круглых скобках.

Знакомство с этой системой обозначений необходимо нам для того, чтобы проиллюстрировать ранее высказанное положение: *теоретические представления, используемые при анализе данных качественного исследования, не только определяют процесс фиксации данных и их истолкование, но и «производятся», явно или неявно оформляются в этом процессе.* Так, этнометодологический анализ разговора направлен на выявление тех методов, которые обычные люди (люди-с-улицы) систематически используют для организации и совместного упорядочения своего общения. Эти методы, с точки зрения этнометодологов, не просто могут быть описаны формально с какой-то внешней точки зрения. Они сами представляют собой множество формальных процедур, устойчиво используемых обычными людьми для достижения взаимопонимания и создания возможности для взаимодействия⁵⁹. Существуют, в частности, устойчивые способы (методы) для взаимного согласования предмета беседы, для гарантированного получения ответа,

⁵⁹ С более точным — хотя и более сложным — описанием теоретических предпосылок анализа разговора лучше познакомиться по записи лекций Х. Сакса. См.: *Atkinson J. M., Heritage J. (eds.) Structures of Social Action: Studies in Conversation Analysis.* P. 21–27. Ch. 2 («Notes on Methodology»).

способы неявного выражения несогласия через смену предмета обсуждения, способы провоцирования аплодисментов при публичном выступлении и т. п. Вот, например, некоторые образцы данных, используемых для анализа «устройств» и техник для постепенного выхода из разговора о неприятностях⁶⁰. Помимо резкого прекращения беседы как радикального средства выхода из неприятного или неловкого разговора люди часто используют менее радикальные средства. Например, *переход к завершающим фразам* нередко позволяет не только постепенно уйти от неприятной темы, но может служить и для подкрепления ранее достигнутых соглашений, и для отсылок к будущей совместной деятельности:

1. «П.: 'уххх Но я надеюсь это само сгладится,
 2. М.: Я почти уверена
 3. П.: _____
- } в этом
} Увидимся во вторник»⁶¹
1. «Дж.: Так что: м-м-все будет хорошо и — =
 2. П.: = Хорош [о
 3. Дж.: [М:ожет 'ухх может если в =
 4. = следующий выходной вы с Фредди зайдете»⁶².

Иной прием — это «*перезапуск разговора*», когда разговор каким-то образом начинается заново либо происходит переключение на новую тему:

1. «Дж.: Да, не очень
 2. М.:
 3. Дж.: Нет : : .
 4. М.: 'ухх Так что ты подельывал на прошлой =
 5. = неделе?»⁶³
- } Хорошо. () —
} 'ухх Не очень.

Наш пример показывает, что для иллюстрации некоторых теоретических представлений в анализе разговора используются лишь определенные «куски» данных, отпрепарированные в согласии с некими правилами. Этнометодолог, стремящийся доказать, что существуют *устойчивые* методы выхода из неприятного разговора, в

⁶⁰ Ibid. P. 191–222.

⁶¹ Ibid. P. 191.

⁶² Ibid. P. 192.

⁶³ Ibid. P. 193.

действительности сам использует такие способы фиксации и анализа данных, которые предполагают и привносят искомую характеристику в его данные. Он игнорирует какие-то обстоятельства контекста (скажем, присутствие наблюдателя), осуществляет скорее произвольную временную разметку (как «обычные люди» мы знаем, что к прежнему разговору можно вернуться и на следующий день, и через год), не производит систематических сравнений между *разными* фрагментами *разных* разговоров о *разных* неприятностях. Эмпирический материал, с которым он имеет дело, безусловно обнаруживает некую упорядоченность. Однако эта упорядоченность не существует сама по себе, как некий «факт».

Важный вывод, следующий отсюда, таков: данные естественного наблюдения, результаты использования этнографического метода не следует считать «более естественными», натуральными, чем данные, получаемые в эксперименте или опросе. Они в той же мере подвержены влиянию теоретических (и не только теоретических) представлений социолога. Столь же сильно зависит от теоретических убеждений и истолкование «фактов» в ходе анализа. Применительно к культурной антропологии эта мысль точно формулируется Н. Томасом: «Нельзя предполагать, что существует какая-то определенная вещь типа “человеческой культуры”, содержащая головоломки, которые попросту составлялись различными антропологическими школами по-разному. Та вещь, о которой идет речь, и те виды головоломок, которые имеют значение, производятся и производились интеллектуальными и политическими интересами. Вновь возникающая совокупность интересов будет переопределять объект и устранять некоторые старые головоломки. От споров чаще отказываются, чем разрешают их. Те же, кто затевают пересмотры или более радикальные концептуальные изменения, предположительно верят, что это приведет к более правильным объяснениям либо истолкованиям — в зависимости от их интересов, — ведущим к более совершенной науке или к более совершенному пониманию, если их субъективные предубеждения являются соответственно позитивистскими или релятивистскими»⁶⁴.

⁶⁴ Thomas N. Out of Time: History and Evolution in Anthropological Discourse. Cambridge: Cambridge University Press, 1989. P. 9.

Возможности *теоретического анализа* в этнографическом исследовании определяются двумя крайними подходами:

1) проверка определенных гипотез о причинных связях и влияниях, поиск подтверждения неких априорных представлений о существовании устойчивых закономерностей;

2) поиск интерпретации, стремление понять смысл событий с точки зрения их участников.

Реальная исследовательская практика, как уже говорилось, обычно находится между этими крайними позициями.

Первый из очерченных подходов (иногда называемый «априорным», так как гипотезы и предположения формулируются заранее) ориентирован на максимальное приближение к идеальной модели каузального вывода. Эта идеальная модель причинного вывода, как мы увидим позднее, полнее всего воплощена в эксперименте. Выборочные обследования достигают приближения к экспериментальной модели с помощью принципа рандомизации и статистического вывода.

При использовании данных включенного наблюдения для создания таких условий используют метод *аналитической индукции*, позволяющий исследователю формулировать обобщения, применимые ко всем эмпирическим случаям, относящимся к данной теоретической проблеме. Аналитическая индукция предполагает систематические сравнения между группами, подвергавшимися и не подвергавшимися воздействию интересующих исследователя причинных факторов. У. Робинсон описал последовательность процедур, необходимых для осуществления аналитической индукции⁶⁵:

«1. Формулировка приблизительного определения того явления, которое надлежит объяснить.

2. Формулировка гипотетического объяснения этого явления.

3. Изучение одного конкретного случая с целью проверки соответствия гипотезы фактам.

4. Если гипотеза для данного случая не соответствует фактам, то следует либо переформулировать гипотезу, либо так переопре-

⁶⁵ *Robinson W. S. The Logical Structure of Analytic Induction // Amer. Sociological Review. 1951. Vol. 16. P. 813.*

делить изучаемое явление, чтобы опровергающий случай оказался исключенным из теоретической выборки.

5. Достаточная для практических целей степень подтверждения может быть достигнута в результате проверки небольшого числа случаев, однако обнаружение негативных (опровергающих гипотезу) случаев доказывает ошибочность данного объяснения и требует его переформулировки.

6. Эта процедура проверки случаев, переопределения изучаемого явления и переформулировки гипотез продолжается до тех пор, пока универсальное отношение не установлено, при том что каждый негативный пример требует дальнейшего переопределения или переформулировки».

Легко видеть, что в обосновании метода аналитической индукции решающую роль играют именно негативные случаи, не подтверждающие исследовательскую гипотезу. А. Линдсмит, одним из первых использовавший этот метод в изучении наркомании, развил эту идею: «Руководящим принципом отбора случаев для проверки теории должна быть максимизация шансов для обнаружения решающего негативного примера. Исследователь, располагающий рабочей гипотезой, которая относится к его данным, начинает осознавать, какие конкретные области критически важны. Если его теория ошибочна или неточна, он знает, что ее слабые места будут выявлены отчетливее и скорее при переходе к исследованию этих критических областей»⁶⁶.

Исследуя механизмы наркотической зависимости от опиатов, сам Линдсмит поначалу выдвинул гипотезу о том, что эта зависимость формируется как собственно психологическая, и те пациенты, которые не знают о том, что получают наркотикосодержащие препараты, не станут от них зависимы. И наоборот, люди становятся зависимыми, если знают, какие препараты принимали, и принимают их достаточно долго, чтобы ощутить признаки «синдрома отмены» (абстиненции) при лишении их наркотиков⁶⁷. Однако почти сразу были обнаружены случаи довольно длительного осознанного

⁶⁶ *Lindesmith A.* Comment on W. S. Robinson's «The Logical Structure of Analytic Induction» // *Amer. Sociological Review.* 1952. Vol. 17. P. 492.

⁶⁷ *Lindesmith A.* *Opiate Addiction.* Bloomington: Principia Press, 1947.

приема морфиносодержащих препаратов, не приведшие к возникновению зависимости.

Линдсмит уточнил исходную гипотезу, предположив, что для формирования зависимости важно также осознание дискомфорта, связанного с отменой наркотикосодержащих препаратов, как собственно «синдрома отмены». И здесь были обнаружены негативные случаи — люди, переживавшие «синдром отмены» вполне осознанно, однако не использовавшие наркотики для снятия физического дискомфорта. В конечном счете Линдсмит пришел к следующему объяснению: 1) зависимость возникает в результате «синдрома отмены», а не в результате физико-химических изменений, вызванных приемом наркотика; 2) условием формирования зависимости является использование наркотикосодержащих препаратов для снятия абстинентного синдрома, причем последний осознается и явно интерпретируется как таковой и связывается с культурными образцами наркозависимого поведения.

В некоторых отношениях метод аналитической индукции дает приемлемое приближение к идеальной модели причинного вывода. Однако систематическая оценка метода аналитической индукции с точки зрения этой модели обнаруживает его существенные недостатки.

Во-первых, предлагаемая методом аналитической индукции стратегия подразумевает, строго говоря, введение все новых объясняющих переменных для каждого негативного наблюдения. Как мы увидим далее при обсуждении логики статистического вывода (гл. 8), такой подход может, конечно, объяснить «свыше 100%» разброса наблюдений, однако он делает бессмысленным само понятие «опровержение». Даже самая нелепая первоначальная гипотеза имеет хорошие шансы уцелеть в результате исследования, если для того, чтобы подогнуть ее к уже добытым данным, будет принято как угодно много вспомогательных гипотез — более или менее правдоподобных, — которые, собственно, не были сформулированы в качестве априорных и подлежащих самостоятельной проверке. Так как изначальные предположения могут совершенствоваться и дополняться как угодно долго, исследователь просто не имеет никаких оснований — кроме здравого смысла — после какого-то случая счесть рабочую гипотезу окончательно опровергнутой и принять другую.

Второй из недостатков аналитической индукции связан с первым.

Модель причинной взаимосвязи, полученная в результате исследования, должна содержать в себе некие предсказания относительно того, когда найденное отношение будет выполняться либо не выполняться. Но процедура аналитической индукции, как мы уже видели, не требует явной формулировки такого рода теоретических предсказаний (хотя и не исключает их). Так, в исследовании Линдсмита не проверялись теоретические прогнозы о том, кто станет и кто не станет принимать наркотики, или о том, как «синдром отмены» будет переживаться разными группами пациентов. Поэтому социолог, использующий аналитическую индукцию, должен особое внимание уделить разработке теоретических прогнозов, которые подлежат проверке.

Наконец, третья проблема, возникающая при использовании аналитической индукции, — это преимущественная ориентация на описание качественных механизмов взаимосвязи, создающая, по выражению Н. Дензина, «проблемы, когда изучаемые процессы включают в себя непрерывные переменные, проявляющиеся лишь до какой-то степени»⁶⁸. Указание на то, какие количественные значения теоретических переменных являются «пороговыми» или «достаточными», чтобы изучаемые отношения наблюдались, обычно отсутствуют.

Включенное наблюдение, как и другие методы социологического исследования, сталкивается с проблемами *внешней* и *внутренней валидности*, то есть, как и в случае эксперимента или выборочного обследования, социолог должен каким-то образом определить, могут ли его результаты быть обобщены и перенесены на другие подобные группы (внешняя валидность). Он должен также убедиться в том, что его результаты не являются побочным продуктом самого процесса наблюдения (внутренняя валидность), то есть следует учесть возможные источники смещений, детально обсуждаемые в последующих главах: историческое развитие, «созревание» респондентов в ходе исследования, реактивные эффекты и т. п. (Здесь в значительной мере применимы те подходы к валидации, о которых говорится в гл. 4 и 6).

⁶⁸ Denzin N. The Research Act. 3rd ed. New Jersey: Prentice Hall, 1989. P. 170.

Выше мы упомянули другой, интерпретативный, подход к теоретическому анализу результатов включенного наблюдения. Этот подход ориентирует исследователя на поиск смысла социального поведения с точки зрения самих деятелей, на создание наблюдателем теории, отражающей собственные «теории» наблюдаемых. Если воспользоваться формулировкой К. Гирца, целью здесь является «не экспериментальная наука, ищущая закон, а интерпретативная наука, нацеленная на поиск смысла»⁶⁹. Обсуждая подходы к определению теоретической проблемы включенного наблюдения, мы уже говорили о том, какой смысл здесь придется этнографическому пониманию. Теперь нам достаточно будет рассмотреть, каким способом понимание смысла человеческого действия может достигаться в ходе *теоретического анализа* результатов включенного наблюдения.

Отказ от абстрактных, формальных теорий не означает отказа от теоретических понятий и перехода к прямой фиксации обыденного опыта. Интерпретация осмысленных социальных действий также не требует от исследователя какой-то сверхъестественной способности сопереживания и «вчувствования» в чужой опыт. Смыслы и нормы социального действия по своей сути *интерсубъективны*, то есть не могут быть сведены к неповторимым индивидуальным состояниям, переживаниям или мнениям. Они изначально ориентированы на возможность понимания, коммуникации и сотрудничества и неотделимы от языка, используемого для их описания. Исследователь имеет дело с совокупностью значений и символов, используемых людьми для самоописания и самоанализа своих поступков, и обладает возможностью их понимания, поскольку сам постоянно включен в этот процесс производства смыслов. Выбирая из нескольких возможных интерпретаций верную, он анализирует наблюдаемую социальную практику приблизительно так же, как комментатор анализирует текст. Альтернативных интерпретаций текста много, но не бесконечно много. Последнее верно и для социальной практики, и для социальных институтов, порождаемых этой практикой и ее воплощающих. Они открыты для понимания, поскольку с самого начала нацелены на понимание, ориентированы на других людей.

⁶⁹ Geertz C. The Interpretation of Cultures. N. Y.: Basic Books, 1973. P. 5.

Если представить данное исследователем описание символической формы и смысла каких-то событий как «внешнее» описание, а то описание своего видения ситуации, которое дают участники, как «внутреннее», возникает вопрос о соотношении теоретических и обыденных понятий, используемых в этих двух типах описания. Для решения этого вопроса сторонники интерпретативного подхода обычно используют разграничение «отдаленных-от-опыта» и «близких-к-опыту» понятий.

«Близкими-к-опыту», в формулировке К. Гирца, можно назвать те понятия, которые сам исследуемый (субъект, информант) мог бы «естественно и без специальных усилий использовать для определения того, что он сам или его ближние видят, чувствуют, думают, представляют себе и т. п., и которые он мог бы без труда понимать, когда эти понятия сходным образом применяются другими»⁷⁰. «Отдаленными-от-опыта» являются те понятия, которые «разного рода специалисты — психоаналитик, экспериментатор, этнограф и даже священник либо идеолог — используют в достижении своих научных, философских или практических целей. “Любовь” — это близкое-к-опыту понятие; “фиксированность либидо на определенном объекте” — отдаленное-от-опыта. “Социальная стратификация” и, возможно, для большинства людей даже “религия” (не говоря уж о “религиозной системе”) являются отдаленными-от-опыта, тогда как “каста” или “нирвана” — по крайней мере для буддистов и индуистов — это близкие-к-опыту понятия»⁷¹.

Исследовательская роль имеет определенные преимущества: хотя ученый и не может «влезть в шкуру» других людей, он может попытаться упорядочить и подвергнуть более глубокому и систематическому анализу те слова, символы и культурные формы, посредством которых изучаемые им люди описывают и передают свой опыт, делая это зачастую непоследовательно, случайно или не вполне осознанно. Сравнительно абстрактные, отдаленные-от-опыта понятийные конструкции позволяют ученому превратить живой

⁷⁰ Geertz C. From the Native's Point of View: On the Nature of Anthropological Understanding // Rabinow P., Sullivan W. M., (eds.) Interpretative Social Science: a Reader. P. 227.

⁷¹ Ibid. P. 227.

опыт и изменчивые культурные формы в предмет собственно теоретического анализа, сделать еще один шаг к увеличению достоверного, доступного коллективному пониманию и проверяемого научного знания. Возникающие здесь возможности и границы теоретического понимания в науках о человеке могут быть очерчены одной меткой фразой: «В стране слепых, которые не так ненаблюдательны, как кажутся, одноглазый — не король, а зритель»⁷².

Дополнительная литература

Девятко И. Ф. Модели объяснения и логика социологического исследования. М.: ИСО РЦГО-TEMPUS/TACIS, 1996.

Ионин Л. Г. Основания социокультурного анализа. М.: Изд-во РГГУ, 1995.

Козина И. М. Особенности применения стратегии исследования случая (case study) при изучении производственных отношений на промышленном предприятии // Социология: 4М. 1995. № 5–6.

Маслова О. М. Качественная и количественная социология: методология и методы (по материалам круглого стола) // Социология: 4М. 1995. № 5–6.

Никишенков А. А. Из истории английской этнографии: Критика функционализма. М.: Изд-во МГУ, 1986. С. 48–90. Гл. 2.

Ольшанский В. Б. Личность и социальные ценности // Социология в СССР. М.: Наука, 1966. Т. 1. С. 242–251.

Ряжский И. А. Опыт использования включенного наблюдения для изучения жизни производственного коллектива // Социологические исследования. 1975. № 3. С. 92–99.

Ядов В. А. Стратегии и методы качественного анализа данных // Социология: 4М. 1991. № 1. С. 14–31.

⁷² Ibid. P. 228.

ГЛАВА 3. БИОГРАФИЧЕСКИЙ МЕТОД

Определение и истоки биографического метода в социологии

Биографические данные в социологии — это основной источник детальных и мотивированных описаний «истории» отдельной личности. И значимые социальные связи, и мотивы действий получают здесь убедительное освещение «с точки зрения деятеля». Чаще всего источником биографических данных становятся личные документы (мемуары, записки, дневники и т. п.) либо материалы интервью и бесед.

Лишь в очень редких случаях исследователь имеет дело с жизнеописанием, включающим в себя все события «от первого крика до последнего вздоха». Обычно основное внимание уделяется конкретным аспектам или стадиям жизни — карьере, межличностным отношениям и т. п. Некоторые авторы даже предлагают взамен широко употребляемых терминов «биографический метод» или «история жизни» использовать термин «история отдельного случая» («individual case history»), подчеркивающий избирательный, селективный характер жизнеописания¹.

В социологии «истории жизни» чаще всего использовались для изучения социальных меньшинств — тех групп, которые довольно трудно поддаются пространственной и временной локализации (и, следовательно, менее доступны для масштабных выборочных обследований).

В 1920–40-х гг. биографический метод широко применялся представителями Чикагской школы. Так, например, в 1920-е гг. чикагский социолог К. Шоу изучал подростковую преступность, используя написанные по его просьбе автобиографические заметки

¹ См.: *Hakim C. Research Design: Strategies and Choices in the Design of social Research*. L.: Allen & Unwin, 1987. P. 65–66.

юного правонарушителя, дополненные полицейскими и судебными документами, результатами медицинских освидетельствований и т. п. Всю совокупность этих данных он рассматривал как «историю случая»².

Биографический метод имеет очень много общего с методом включенного наблюдения и по сути является еще одной разновидностью этнографического подхода к «анализу случая». Отличием биографического метода можно считать большую сфокусированность на уникальных аспектах истории жизни человека (иногда — группы, организации) и на субъективном, личностном подходе к описанию человеческой жизни, карьеры, истории любви и т. п. В центре внимания социолога здесь оказывается *документальное или устное описание событий с точки зрения самого «случая»*, то есть те сведения, которые в медицине называют субъективным анамнезом. Как и метод включенного наблюдения, биографический метод имеет «этнографические» корни. Культурные антропологи и историки часто опирались (и опираются) на «устные истории» или дневниковые записи и мемуары, когда им приходится изучать соответственно «доисторические», не имеющие письменной традиции сообщества либо «закулисные» политические механизмы. Еще очевиднее тот вклад, который внесли в развитие биографического метода документальная журналистика и мемуаристика. (Достаточно вспомнить о столь раннем примере использования сравнительно-биографического метода, как «Жизнеописания» Плутарха.)

Первой собственно социологической работой, «узаконившей» использование личных документов, писем и автобиографий в анализе социальных процессов, стала опубликованная в 1918–1920 гг. книга У. Томаса и Ф. Знанецкого «Польский крестьянин в Европе и в Америке». Один из томов этой книги составила автобиография польского эмигранта Владека, описавшего свой путь из провинциального Копина в Чикаго. Этот путь включил в себя и учебу в деревенской школе, и работу помощником в лавке, и выезд в Германию в поисках заработка, предшествовавшие эмиграции в США. Томас

² Shaw C. R. The Jack-Roller. A Delinquent Boy's Own Story. Chicago: The University of Chicago Press, [1930] 1966; Shaw C. R. The Natural History of a Delinquent Career. Chicago: The University of Chicago Press, 1931.

и Знанецки первыми выступили с обоснованием использования биографического метода в рамках интерпретативного подхода в социологии (о чем уже говорилось в гл. 2, посвященной методу включенного наблюдения). Они полагали, что социальные процессы нужно рассматривать как *результат постоянного взаимодействия сознания личности и объективной социальной реальности*. В этом взаимодействии личность и «ее» определения реальности выступают и как постоянно действующий детерминант, и как продукт социального взаимодействия. Следовательно, изучение сознания и самосознания — необходимое условие анализа социального мира. Кроме того, Томас и Знанецки предполагали, что исследование, базирующееся на «историях жизни», позволит выйти к более широким обобщениям, касающимся социальных групп, субкультур, классов и т. п.

Н. Дензин дал одно из самых популярных *определений биографического метода* (метода «историй жизни», «жизнеописаний»): «...биографический метод представляет переживания и определения одного лица, одной группы или одной организации в той форме, в которой это лицо, группа или организация интерпретируют эти переживания. К материалам жизненной истории относятся любые записи или документы, включая “истории случая” социальных организаций, которые проливают свет на субъективное поведение индивидов и групп. Такие материалы могут варьировать от писем до автобиографий, от газетных сообщений до протоколов судебных заседаний»³.

Предположение о необходимости учета «перспективы деятеля», его смыслового горизонта и определения ситуации играет ведущую роль при использовании биографического метода. Так как целью здесь в конечном счете оказывается понимание тех или иных аспектов «внутреннего мира» субъекта, необходимым становится и предположение о том, что исследуемые располагают достаточно сложной структурой субъективного опыта и способны отделить собственный «образ Я» от образа окружающего мира, способны

³ *Denzin N. The Research Act. A Theoretical Introduction to Sociological Methods. 3rd ed. Englewood Cliffs: Prentice Hall. 1989. P. 183.*

«воспринять себя в качестве активного субъекта своей собственной истории жизни, отличного от социального мира»⁴.

Следующая фундаментальная особенность биографического метода — его направленность на воссоздание исторической, развернутой во времени перспективы событий. Используя биографический метод, социолог становится в некотором роде социальным историком. История социальных институтов и социальных изменений здесь раскрывает себя через рассказы людей об их собственной жизни. Это открывает дополнительные возможности для пересмотра «официальных» версий истории, написанных с позиций властвующих классов и групп, и сопоставления этих версий с основанным на повседневном опыте знанием социальной жизни, которым располагают непривилегированные и «безгласные» социальные группы. «*Кто* говорит и *кого* слушают — это политические вопросы; факт, становящийся особенно очевидным, когда голоса получают люди, обладающие низким статусом и властью»⁵.

Вот, например, как описывает свои отношения с начальством женщина-работница, проинтервьюированная Дж. Уитнер в ходе исследования «биографии» игрушечной фабрики в Чикаго (эта работа, кстати, может служить примером использования биографического метода в исследовании организаций):

«Один мастер — он прежде служил лейтенантом или еще кем-то там в армии — все время доводил работавших на его участке контролеров качества до слез, потому что он кричал на них, и они расстраивались. (Вопрос: почему он кричал на них?) Потому что они отказывались что-нибудь делать, а он не терпел, чтобы кто-то отказывался на его участке, и мы тусовались. Он накричал на меня. Я сказала ему, что мне наплевать. Тогда он побежал к начальству, чтобы пожаловаться, что я вела себя непочтительно. Он хотел от меня объяснительной. Но разве кто-то не покрикивал на меня? Я не собака»⁶.

⁴ См.: Kohli M. *Biography: Account, Text, Method* // *Biography and Society*. Beverly Hills, 1981. P. 164.

⁵ См.: McCall M. M., Wittner J. *The Good News about Life History* // Becker H. S., McCall M. M., (eds.) *Symbolic Interaction and Cultural Studies*. Chicago: The University of Chicago Press, 1990. P. 47.

⁶ Ibid. P. 66.

Особое внимание проблеме предоставления права голоса «безгласным» уделяет традиция *символического интеракционизма*. Здесь эта проблема рассматривается не столько в политическом, сколько в теоретическом аспекте. Предполагается, что во всяком обществе существует определенная «иерархия правдоподобия» в производстве и распространении значений и социального знания. Те, кто находятся на «верхнем этаже» этой иерархии, имеют преимущество в формулировке правил, используемых для приписывания смысла действиям и определения ситуации. В результате кто-то диктует правила и нормы в соответствии со своими интересами, а кто-то, *также следуя своим интересам*, нарушает эти правила и нормы, оказываясь в положении аутсайдера, маргинала или преступника. Если социолог принимает одну, господствующую точку зрения при описании фрагмента социальной реальности, он заведомо игнорирует те интересы, знания и смыслы, которые определяют поступки другой стороны. Биографический подход с точки зрения символического интеракционизма увеличивает шансы исследователя в понимании нестандартных или «отклоняющихся» от общепринятого смысловых перспектив, хотя именно в этом случае его нередко обвиняют в одностороннем или тенденциозном анализе:

«Когда мы обвиняем себя или коллег-социологов в необъективности? Я думаю, что рассмотрение типичных примеров показало бы, что эти обвинения возникают — если обратиться к одному важному классу таких случаев, — когда исследователь оказывает сколько-нибудь серьезное доверие перспективе подчиненной группы в некоем иерархическом отношении. В случае девиантов таким иерархическим отношением оказывается отношение морали. Здесь в положении превосходства оказываются те участники отношения, которые представляют силы официальной и одобряемой морали, а подчиненными становятся те, кто якобы нарушил эту мораль... (другими словами) обвинения в предвзятости, относящиеся к нам или к другим, провоцируются отказом проявлять доверие и почтение к сложившемуся статусному порядку, где право быть услышанным и доступ к истине распределены неравномерно»⁷.

⁷ Becker H. S. Sociological Work. Chicago: Aldine, 1970. P. 125–126.

Естественно предположить, что направленность биографического метода на то, чтобы представить субъективный опыт деятеля через его собственные категории и определения, требует какого-то переосмысления *критериев объективности исследования*. Действительно, социолог здесь должен прежде всего определить, какова «собственная история», личная трактовка субъекта. То, как субъект сам определяет ситуацию, в данном случае важнее, чем то, какова ситуация «сама по себе» (мы уже говорили об этом в гл. 2). Эта «собственная история» может и должна быть дополнена сведениями о том, как определяют ситуацию другие участники. Сопоставление точек зрения и сведений, полученных с помощью разных методов и (или) из разных источников, позволяет полно и достаточно объективно воссоздать не только внешнюю картину событий, но и их субъективный смысл для участников. Такой тип исследовательской стратегии в социологии принято обозначать как *множественную триангуляцию*⁸. (Термин «триангуляция» призван подчеркнуть сходство со способом определения удаленности или месторасположения некоторого объекта, используемым в геодезии или радиопеленгации.) Множественная триангуляция помогает в анализе различающихся определений ситуации, относящихся к одним и тем же элементам опыта.

Сбор биографического материала

Любой устный или письменный рассказ субъекта о событиях его жизни может рассматриваться в качестве биографического материала. При определенных условиях для воссоздания «истории жизни» могут использоваться и вторичные источники — мемуары других лиц, письма, официальные документы и т. п. Например, если обратиться к исследованиям «истории жизни» организаций, К. Литлер изучал трудовые отношения в двух британских компаниях в 1930-е гг. Интервьюирование профсоюзных активистов, участвовавших в событиях того времени, было дополнено материалами ар-

⁸ См.: *Webb E. J., Campbell D. T., Swartz R. D., Sechrest L.* Unobtrusive Measures: Nonreactive Measure in the Social Sciences. Chicago: Rand McNally, 1966, а также *Denzin N.* Op. cit. Ch. 10.

живов этих компаний и газетными сообщениями, касавшимися трудовых конфликтов, которые там происходили⁹. В недавнем совместном исследовании британских и российских ученых¹⁰ изучалась, в частности, кадровая политика на предприятиях разного типа («неблагополучных», «благополучных», «новых»). Помимо анализа документов кадровой и экономической статистики использовались полуструктурированные интервью работников предприятий, основной темой которых стали трудовые биографии респондентов (всего было проведено 260 интервью на 12 предприятиях).

Важно, однако, различать биографические (автобиографические) истории и так называемые устные истории. *Устная история* — это фактуально точное воссоздание определенных исторических событий. В ее фокусе — не субъективный опыт деятеля, а *историческое знание о событиях, процессах, движущих силах и причинах*. Устные истории, рассказанные участниками событий, используются для накопления такого исторического и фактического знания. Историческое знание «с точки зрения очевидца» необходимо, например, антропологу, стремящемуся воссоздать историю разделения труда между соседними племенами или историю вражды между кланами. Историк — представитель школы «Новой социальной истории» также сможет использовать устные истории, например, описывая бурные политические изменения «снизу» как изменения повседневной жизни простых людей¹¹.

В социологии принято различать три основных типа «историй жизни»: *полные, тематические и отредактированные*.

Полная «история жизни» в идеале очерчивает весь жизненный опыт субъекта — от колыбели до могилы (что само по себе не требует большого объема и степени детализации).

⁹ *Little C. R.* The Development of Labour Process in Capitalist Societies: A Comparative Study of the Transformation of Work Organization in Britain, Japan and the USA. L.: Heinemann, 1982. P. 117–145.

¹⁰ Реструктурирование занятости и формирование локальных рынков труда в России: Научный доклад. М.: ИСИТО, 1996.

¹¹ Подробнее см., например: *Thompson P.* The Voice of the Past: Oral History. Oxford: Oxford University Press, 1978; *Gardner G. B., Adams G. R.* Ordinary People and Everyday Life: Perspectives on the New Social History. Nashville: American Association for State and Local History, 1983.

Тематическая «история жизни» отличается от полной тем, что она относится преимущественно к одной стороне или фазе жизненного цикла субъекта. Например, Э. Сазерленд написал книгу о профессиональной карьере «вора в законе», который выступил в качестве соавтора произведения¹². Сазерленд подготовил опросник, позволивший структурировать письменный рассказ своего соавтора, провел ряд дополнительных интервью и прокомментировал получившуюся «историю жизни». Однако он не использовал никакие дополнительные источники.

Отредактированная «история жизни» может, вообще говоря, быть и полной, и тематической. Ее основная особенность — *ведущая роль социолога-интерпретатора*, явно организующего биографический материал в соответствии с теоретической логикой, избирательно редактирующего и интерпретирующего исходный рассказ (или рассказы) субъектов для того, чтобы ответить на поставленные в исследовании вопросы. Нередко множество отредактированных «историй жизни» становятся иллюстративным или доказательным материалом в теоретическом по сути исследовании. Примером может служить знаменитая работа И. Гофмана «Стигма». Под «стигмой» здесь понимается свойство (атрибут), рассматриваемое как порочащее, неуместное для представителя определенной социальной категории и отличающее его от социально определяемой «нормы» (например, значительный физический дефект, моральное «уродство» наподобие алкоголизма, принадлежность к «не той» расе и т. п.). Так, в главе, посвященной типам «духовной карьеры» стигматизированных людей, Гофман использует десяток различных автобиографических источников, мемуаров, «жизненных историй», чтобы показать, как влияют на личностную идентичность время и обстоятельства осознания субъектом своей стигмы, ее очевидность для окружающих и т. п. Субъектами «жизненных историй» здесь оказываются и человек, заболевший в юности полиомиелитом, и профессиональная проститутка, и слепая девушка, и гомосексуалист.

¹² См.: Conwell C., Sutherland E. H. The Professional Thief. Chicago: The University of Chicago Press, 1937.

Основными *источниками биографических данных*, как уже говорилось, служат, помимо опросов и интервью, публичные и частные архивные материалы.

Интервью, опросники и дословные записи устных сообщений неизменно играют ведущую роль в получении значимых для социологии «историй жизни». Их применение гарантирует релевантность получаемых сведений той теоретической проблеме, которая стоит перед социологом (хотя эта «социологическая релевантность», по мнению некоторых, достигается ценой меньшей спонтанности и непосредственности изложения). Процедуры интервьюирования и опросники, используемые в этом случае, по сути, отличаются от традиционных для социологии лишь тем, что они отчетливо структурированы временной перспективой человеческой жизни как целого. Опросник, или «биографический путеводитель», используемый при интервьюировании, позволяет субъекту упорядочить свой рассказ и уделить достаточное внимание всем фазам жизненного цикла (детство, юность и т. п.) и всем сферам жизненного опыта (семья, карьера и т. п.), которые значимы для него и (или) интересуют социолога. Помимо того, что биографический опросник или краткий тематический путеводитель позволяют не позабыть или не пропустить существенные сведения, они полезны и самому исследователю как средство отчетливой и явной операционализации тех понятий, которые он собирается использовать в теоретическом анализе.

Заметим здесь, что нередко биографический материал собирается в ходе вполне традиционного выборочного обследования. В большинстве случаев выборка такого исследования представляет какую-то возрастную когорту или профессиональную группу. Разумеется, исходя из практических соображений стоимости широкомасштабного интервьюирования и доступности «редких» совокупностей (см. гл. 6), исследователи чаще всего ограничиваются *квотной выборкой*. Например, в осуществленном в 1970-е гг. исследовании социальных изменений в канадской провинции Квебек было собрано 150 биографических интервью тех, кто начинал свою профессиональную карьеру в 1940-е гг.¹³

¹³ См.: *Gagnon N. On the Life Accounts // Bertaux D. (ed.) Biography and Society. Beverly Hills, 1981.*

Даже в тех случаях, когда социолог проводит серию глубинных («клинических») интервью без использования жесткого плана беседы или «путеводителя», он ориентируется на какую-то совокупность теоретически значимых тем, пунктов беседы и постоянно возвращается к их обсуждению. В качестве примера мы можем использовать известную работу «отца-основателя» этнометодологии Г. Гарфинкеля, посвященную анализу «индивидуального случая» изменения полового статуса¹⁴. Основной эмпирический материал здесь — это многочисленные интервью Агнессы, девятнадцатилетней девушки, рожденной и воспитывавшейся до 17 лет как мальчик и сознательно решившей сменить пол. Агнесса к моменту поступления в университетскую клинику уже два года жила в облике девушки и успешно скрывала от окружающих свой секрет. По ее собственным словам и некоторым косвенным данным, она всегда хотела стать нормальной женщиной и ощущала себя девушкой, рассматривая свои нормальные мужские гениталии как «злую шутку природы», превратность судьбы. С точки зрения генетики, анатомии и эндокринного статуса, Агнесса представляла собой редкий случай «чисто гормональной» (тестикулярной) феминизации в подростковом возрасте: физиологически и анатомически нормальные мужские органы соседствовали с вполне отчетливыми женскими вторичными половыми признаками, и внешне, для неосведомленных наблюдателей, Агнесса выглядела как привлекательная юная девушка. Конечно, Гарфинкеля интересовал не сам по себе «медицинский случай». Его интересовала та тонкая социальная «работа», направленная на достижение и сохранение избранного сексуального статуса, которую приходилось осуществлять Агнессе. Любая ошибка, нарушение нормативных ожиданий окружающих, отклонение от «социально-понятных» ролевых моделей привели бы Агнессу к краху ее идентичности и к полной маргинализации. Однако Агнесса не только «управилась» с необходимостью вести обычный, социально-принятый образ жизни молоденькой девушки, иметь подруг и поклонников и т. п., но и добилась сложной хирургичес-

¹⁴ *Garfinkel H. Passing and the Managed Achievement of Sex Status in an «Intersexed» Person. Part 1 (in collab. with R. J. Stoller) // Studies in Ethnomethodology. Cambridge, [1967] 1984. P. 116–185.*

кой операции, которая позволила избавиться от мужских гениталий и обрести — средствами пластической хирургии — «минимальный анатомический набор», необходимый, чтобы стать «обычной женщиной» (конечно, лишенной собственно репродуктивной функции). Именно сложная «жизненная история» Агнессы дала возможность проанализировать те механизмы конструирования и поддержания «правильного», рационального и «объяснимого-с-точки-зрения-других-людей» статусно-ролевого поведения, которые в повседневной жизни «нормальных» мужчин и «нормальных» женщин не осознаются и действуют автоматически.

Гарфинкель в беседах с Агнессой постоянно обращался к тем темам («пунктам») ее биографии, которые позволяли пролить свет на определенные теоретические проблемы: «наивное» восприятие разделения полов как однозначного, абсолютного и морально-нагруженного порядка вещей; идентификация сексуального статуса посредством культурно-детерминированных знаков отличия, воспринимаемых в обыденном сознании как естественные, и т. п. Одной из тем, интересовавших исследователя, была тема ретроспективного конструирования личностью согласованной с избранным статусом автобиографии: все события, поступки, атрибуты, отношения прежней жизни Агнессы, воспитывавшейся в качестве мальчика, последовательно интерпретировались ею как история «ошибочно воспринимавшейся окружающими в качестве мальчика» девочки. Разумеется, такая автобиография была не лишена каких-то пропусков и труднообъяснимых фактов, но в главном отличалась незаурядной согласованностью: «Уже сама выраженность преувеличений в ее женской биографии, в описании маскулинности ее друга (за которого Агнесса собиралась выйти замуж), “бесчувственности” ее мужских гениталий и т. п. представляет постоянно подчеркиваемую черту: последовательно женскую идентификацию»¹⁵.

Случай Агнессы, проанализированный Гарфинкелем, еще раз демонстрирует те трудности в оценке объективности данных, которые возникают при использовании биографического метода: любые искажения фактов здесь могут оказаться и результатом их намеренного сокрытия, и вполне «искренним» механизмом защиты личной

¹⁵ *Garfinkel H. Op. cit. P. 130.*

самотождественности (то есть неотъемлемой частью реального «образа Я»), и результатом простой неосведомленности.

Так, Гарфинкель отметил, что Агнесса поразительно мало знала о мужской гомосексуальности и при неоднократных попытках обсуждения этой темы, несмотря на явный интерес и эмоциональность восприятия разговора, просто не могла объяснить, как она воспринимала признаки гомосексуальных интересов у других мальчиков. Она также отказывалась провести какие-либо сравнения между собой и гомосексуальными мужчинами либо трансвеститами, хотя легко и охотно сопоставляла свой статус со статусами нормального мужчины или нормальной женщины. Исследователь был лишен возможности услышать рассказы других участников событий, однако добросовестно зафиксировал особое мнение одного из урологов, не участвовавших непосредственно в лечении Агнессы. Этот человек полагал, что решение об операции было медицинской и этической ошибкой, результатом мистификации, ссылаясь на весьма неоднозначные медицинские признаки и даже на то, что у случайно встреченного им жениха Агнессы была отнюдь не мужественная внешность.

Изрядно времени спустя, когда исследовательский проект был успешно завершен, книга Гарфинкеля находилась в печати, а бывшая пациентка уже более пяти лет вела активную жизнь молодой, привлекательной и сексуально благополучной женщины, Агнесса посетила ученых и сообщила, что никогда не имела абсолютно никаких биологических дефектов, которые вели бы к феминизации в подростковом возрасте. Просто с 12 лет она тайно принимала эстрогены (женские половые гормоны), прописанные ее матери после серьезной хирургической операции.

К *частным архивным материалам*, используемым при изучении «истории жизни», относят преимущественно личные записи и документы. Основной тип частного документа — это автобиография. (К автобиографиям относятся и те детальные жизнеописания, которые создаются по просьбе исследователя.) Существуют заметные различия между автобиографией, написанной в расчете на дальнейшую публикацию, и автобиографией, обращенной лишь к узкому кругу близких. Если в первом случае преимуществом является большая фактическая достоверность и «читабельность» изложения, то во втором обычно имеет место высокая степень раскры-

тия личного отношения к пережитому, особое стремление мотивировать совершенные выборы и поступки.

Как и основанные на автобиографических сведениях «истории жизни», сами автобиографии могут быть разделены на *полные*, *тематические* и *отредактированные*. Тематические автобиографии, в отличие от полных, ориентированы на определенную сферу личного опыта или период жизни (ср., например: «Моя жизнь в искусстве» и «Подлинная история моей жизни»).

Достоинство автобиографий — большая достоверность в описании личностной «подкладки» событий. Однако нужно всегда помнить о том, что автобиография — это *реконструированная субъектом в определенный момент жизни история*. Здесь особенно вероятны смещения и ошибки, вызванные и стремлением рационально мотивировать любой поступок с точки зрения «сегодняшнего» мировосприятия, и необходимостью придать повествованию некоторую литературную форму. Методологическая триангуляция, о которой говорилось выше, становится единственным средством достижения достоверности и объективности при анализе «историй жизни», основанных на автобиографических данных. Иными словами, *автобиографические данные должны интерпретироваться в контексте сведений, полученных из иных источников*.

К частной архивной документации относятся также *дневники*, *частные записи*, *мемуары*, *личные письма*, *записи разговоров* и т. п. Дневник и мемуарные записки иногда трудно различимы: можно считать, что мемуары в целом отличает более безличный стиль изложения и необязательность линейного и упорядоченного описания сменяющих друг друга во времени событий. Повышение достоверности «историй жизни», основанных на такого рода личных (иногда говорят «экспрессивных») документах, как дневниковые и мемуарные записи, требует, как и в ранее описанных случаях, привлечения дополнительных источников, использования специальных приемов критического анализа (в том числе критической оценки экспрессивного документа как исторического источника, как литературного текста и т. п.). Важным подспорьем здесь могут оказаться и не столько экспрессивные, сколько *функциональные личные документы* — расписания, черновики, планы работы, записи финансовых поступлений и расходов.

Личные письма также могут рассматриваться как важный источник биографических данных. Письмо может рассказать достаточно важные вещи не только об его авторе, но и о получателе и взаимоотношениях между первым и вторым. И стиль, и способ изложения, и частота переписки могут быть столь же информативны, как и собственно содержание письма. К сожалению, современная социология довольно мало внимания уделяет этому типу личной документации, хотя литературоведение и история дают немало примеров использования личной переписки в качестве полезного источника данных.

Важным *дополнительным источником биографических данных* являются также официальные архивные документы: записи актов гражданского состояния (рождения, смерти, браки), правительственные документы, данные социальной статистики, архивы политических, общественных организаций и административных органов. В ведомственных архивах могут быть обнаружены важные биографические документы, связанные, в первую очередь, с профессиональной карьерой: личные листки по учету кадров, сведения о наградах и взысканиях, характеристики. Большой интерес представляет документация медицинских учреждений, органов юриспруденции, однако в этом случае необходимо принимать во внимание и существующие обычно жесткие ограничения на доступ к таким источникам, и этические соображения.

Анализ и интерпретация биографического материала

В начале этой главы мы говорили о том, что «истории жизни», биографический метод — это, по сути, разновидность этнографического метода, имеющая дело с анализом «индивидуального случая». Поэтому нам нет нужды детально обсуждать возможности анализа и интерпретации этнографических данных, рассмотренные в главе, посвященной включенному наблюдению. Все, что было сказано об интерпретативном подходе, аналитической индукции, типах понятий и требованиях к валидности, *в полной мере* применимо и к биографическому методу. Здесь мы остановимся лишь на тех проблемах, которые возникают в связи с «индивидуальной» природой биографических данных.

Применение *причинных моделей* к анализу «историй жизни» требует использования процедур аналитической индукции (см. с. 56–57). Роль негативных, опровергающих примеров в этом случае особенно существенна: обобщения, построенные на нескольких «историях жизни», могут быть уточнены, дополнены или опровергнуты лишь при сопоставлении с новыми, отобранными по теоретически-релевантным признакам, случаями. Излишне говорить о необходимости обоснования «типичности», репрезентативности отобранных для изучения индивидуальных случаев. Здесь применимы идеи *теоретической выборки*, рассмотренные в главе о включенном наблюдении. Например, в исследовании изменения семейных взаимоотношений и циклов семейной жизни¹⁶ сбору «историй жизни» предшествовал детальный анализ доступных демографических данных о межклассовых и поколенческих различиях по таким параметрам, как размер семьи, время рождения самого младшего ребенка и его отделения от родительской семьи и т. п. В результате исследователи сочли возможным ограничиться 130 биографическими интервью с мужчинами и женщинами, рожденными в конце 1890 — начале 1900-х гг. в канадском городке Гамильтон (Онтарио) и его окрестностях. Квотная выборка репрезентировала три типичные социальные группы — городской средний класс, городских рабочих и фермеров.

Те соображения, которые ранее были высказаны применительно к *внешней и внутренней валидности* этнографических данных, применимы и к «историям жизни». В целом биографический метод особенно уязвим для критики, указывающей на наличие таких угроз внутренней валидности, как субъективные смещения и историческая эволюция субъектов. Все респонденты, рассказывающие свои «жизненные истории», анализируют свое прошлое (и предугадывают будущее) с точки зрения конкретного, «вот этого», момента своего личностного развития, обычно стремясь дать социально-одобряемую и согласованную картину жизни как целого. К тому же социолог должен помнить о том, что сама форма

¹⁶ См.: *Synge J. Cohort Analysis in the Planning and Interpretation of Research Using Life History // Bertaux D. (ed.) Biography and Society. Beverly Hills, 1981. P. 235–249.*

биографического повествования — литературная по сути и корням — подталкивает субъекта к использованию популярных биографических канонов, расхожих «сценариев» (например, «история успеха», «рассказ о поиске личностной идентичности», «жизнь прирожденного неудачника» и т. п.). С этой точки зрения «хорошая» биография не должна быть излишне согласованной во всех деталях.

Использование *интерпретативных моделей* в анализе биографических данных, как мы неоднократно отмечали выше, ориентировано не столько на выведение общих объяснений и причинных закономерностей, сколько на понимание субъективного смысла событий с точки зрения деятеля. Однако и в этом случае достоверность интерпретации зависит от сопоставления сведений, полученных из разных источников, и критической оценки личных сообщений. Фактически биографический метод ведет исследователя к тем же проблемам, что и метод *историографический*. Здесь часто необходимы и оценка достоверности и подлинности личного документа, и соотнесение с другими свидетельствами, а иногда — и установление авторства. Биографический метод по определению *историчен* — используя документы прошлого, он стремится к созданию убедительного исторического объяснения полученных сведений. Поскольку историографией называют всякую попытку реконструкции прошлого на основе документальных данных, «история жизни» — тоже форма историографии¹⁷. Источники данных в историографии принято делить на *первичные* и *вторичные*. К первичным относят те источники, которые содержат непосредственные свидетельства очевидцев или прямых участников событий, а ко вторичным — свидетельства или рассказы тех, кто не присутствовал при описываемых событиях. В историографии принято считать более надежными те документы, автор которых ближе включен в описываемую ситуацию и дает описание «из первых рук». Кроме того, выше ценятся свидетельства более опытного и искушенного наблюдателя, иными словами, — эксперта. Многие авторы полагают, что достоверность и надежность документов тем выше, чем уже аудитория,

¹⁷ Классический анализ этого вопроса см.: *Gottschalk L., Kluckhohn C., Angell R. The Use of Personal Documents in History, Anthropology, and Sociology*. N. Y.: Social Science Research Council, 1945.

к которой адресуется автор¹⁸, то есть по мере роста предполагаемой аудитории автор все больше оказывается под влиянием тенденции описывать события в апологетическом и драматическом ракурсе: интимная исповедь постепенно превращается в пропаганду.

Для социолога, использующего личные документы, определенный интерес представляют и те приемы критики источников и установления их подлинности, которые традиционно применяются в историографии¹⁹. Во-первых, речь идет о *проверке подлинности* (несфальсифицированности) *текста* и *установлении его авторства*. Если для социолога, имеющего дело с «живым» рассказом, эти проблемы сравнительно малозначимы, то использование личных документов «в отсутствие» субъекта выдвигает их на первый план. Исследователь должен убедиться в том, что документ является именно тем, за что его принимают (например, предсмертной запиской, а не наброском поэмы), а также определить принадлежность документа данному автору. Для такой проверки используются и внешние материальные признаки — почерк, бумага, место хранения, и формальные особенности текста — стиль изложения, лексические характеристики, отсутствие анахронизмов.

Немаловажное значение имеет *обоснованность интерпретации* текста с точки зрения его характера, целей написания, предполагаемой аудитории и шире — его социального контекста.

Наконец, даже последовательно интерпретативная трактовка биографического метода не избавляет от необходимости проверить *фактическую правдивость* содержащихся в биографических документах сведений. Как известно, даже один из основателей интерпретативного подхода в социологии (У. Томас) полагал, что самые радужные перспективы для социологии откроются по мере развития надежной государственной системы учета личных сведений о гражданах.

Конечно, и расшифровка смысла документа, и установление его подлинности никогда не бывают окончательными. Наша способность к пониманию биографических и — шире — исторических

¹⁸ Denzin N. Op. cit. P. 202–203.

¹⁹ См., в частности: Блок М. Апология истории, или Ремесло историка. 2-е изд., доп. М.: Наука, 1986. Гл. 3 («Критика»).

событий всегда ограничена и доступным нам смысловым горизонтом социального действия, и принимаемыми теоретическими схемами. Один из подходов к объективному анализу исторических данных и поступков деятеля — это известная концепция «идеальных типов».

М. Вебер понимал под идеальным типом некую социокультурную модель, служащую орудием теоретического понимания. Идеальный тип — это не гипотеза и не исторически конкретное описание фактов, а сугубо теоретическая, абстрактная конструкция, которая может и не существовать в реальности, но позволяет ученому понять и объяснить реальность. Идеальный тип — это отнюдь не что-то более совершенное и идеально соответствующее норме. Это скорее намеренно преувеличенное и одностороннее описание собственной точки зрения социолога, его видения смысла поступков деятелей. «Этот мысленный образ сочетает определенные связи и процессы исторической жизни в некий лишенный внутренних противоречий космос мысленных связей. По своему содержанию данная конструкция носит характер утопии, полученной посредством мысленного усиления определенных элементов действительности... Задача исторического исследования состоит в том, чтобы в каждом отдельном случае установить, насколько действительность близка такому мысленному образу или далека от него...»²⁰. Примерами идеальных типов могут служить «нуклеарная семья», «капитализм», «целерациональное действие» и т. п.

Конструирование «идеально-типических» понятий (всегда «далеких-от-опыта», см. предыдущую главу) может стать шагом к построению собственно эмпирически проверяемых гипотез.

Н. Дензин предложил общую схему анализа и описания «исторической жизни»:

«Шаг 1. Отберите исследовательские проблемы и гипотезы, которые могут быть исследованы и проверены с помощью истории жизни.

Шаг 2. Отберите субъекта или субъектов и определите, в какой форме будут собраны биографические данные.

²⁰ Вебер М. Избранные произведения: Пер. с нем. / Под ред. Ю. Н. Давыдова. М.: Прогресс, 1990. С. 389–390.

Шаг 3. Опишите объективные события и переживания из жизни субъекта, имеющие отношение к интересующей вас проблеме. Эти события подлежат оценке с точки зрения различных источников и перспектив (триангуляция) таким образом, чтобы противоречия, непоследовательность и нерегулярность стали очевидны.

Шаг 4. Получите от субъекта его интерпретации этих событий, следуя естественному, или хронологическому, порядку.

Шаг 5. Проанализируйте все утверждения и сообщения с точки зрения их внутренней и внешней валидности... (Проверьте достоверность источников.)

Шаг 6. Примите окончательное решение о достоверности вышеупомянутых источников и установите приоритетные источники для последующей проверки гипотез.

Шаг 7. Начните проверку предварительно сформулированных гипотез, поиск опровергающих примеров. Продолжайте модифицировать эти гипотезы, выдвигать новые и проверять их.

Шаг 8. Составьте черновой набросок всей “истории жизни” и ознакомьте с ним исследуемых, чтобы узнать их реакцию.

Шаг 9. Переработайте исследовательский отчет, изложив события в их естественной последовательности и учтя замечания исследуемых субъектов. Представьте в отчете те гипотезы и предположения, которые получили подтверждение. В заключении остановитесь на теоретической значимости ваших выводов и перспективах дальнейшего исследования»²¹.

Эта схема может служить ориентиром в работе с биографическими данными.

Дополнительная литература

Альмодавар Ж.-П. Рассказ о жизни и индивидуальная траектория // Вопросы социологии. 1992. Т. 1. № 2.

Бургос М. История жизни. Рассказывание и поиск себя // Вопросы социологии. 1992. Т. 2. № 2.

Блок М. Апология истории, или Ремесло историка, 2-е изд., доп. М.: Наука, 1986. Гл. 3.

²¹ *Denzin N.* Op. cit. P. 205.

- Журавлев В. Ф.* Нарративное интервью в биографических исследованиях // Социология: 4М. 1993–1994. № 3–4.
- Знаецкий Ф.* Мемуары как объект исследования // Социологические исследования. 1989. № 1.
- Козина И. М.* Поведение работников на рынке труда. Способы трудоустройства и личные стратегии занятости // Реструктурирование занятости и формирование локальных рынков труда в России. М. 1996. С. 84–107.
- Козлова Н. Н.* Крестьянский сын: Опыт биографического исследования // Социологические исследования. 1994. № 6. С. 112–123.
- Оболенская С. В.* «История повседневности» в историографии ФРГ // Одиссей. Человек в истории. М., 1990. С. 182–197.
- Рождественский С.* Подходы к формализации жизненных историй качественными методами // Судьбы людей: Россия XX век. Биографии людей как объект социологического исследования / Отв. ред. В. Семенова, Е. Фотеева. М., 1996. С. 412–422.
- Тернер Р.* Сравнительный контент-анализ биографий // Вопросы социологии. 1992. Т. 1. № 1.
- Томпсон П.* Гуманистическая традиция и жизненные истории в Польше // Биографический метод в социологии: история, методология, практика / Ред. колл.: В. В. Семенова, Е. Ю. Мещеркина. М., 1993. С. 51–62.
- Томпсон П.* История жизни и анализ социальных изменений // Вопросы социологии. 1993. № 1–2. С. 129–138.
- Фукс-Хайнриц В.* Биографический метод // Биографический метод в социологии: история, методология, практика / Ред. колл.: В. В. Семенова, Е. Ю. Мещеркина. М., 1993. С. 11–41.
- Хоффман А.* Достоверность и надежность в устной истории // Биографический метод в социологии: история, методология, практика / Ред. колл.: В. В. Семенова, Е. Ю. Мещеркина. М., 1993. С. 42–50.

ГЛАВА 4. ЭКСПЕРИМЕНТ В СОЦИАЛЬНЫХ НАУКАХ

Эксперимент — это опытное исследование воздействия отдельного фактора (или нескольких факторов) на интересующую исследователя переменную. Экспериментальное исследование строится в соответствии с правилами индуктивного вывода о наличии причинно-следственной связи между событиями, во-первых, демонстрируя регулярный характер появления события-«отклика» после предшествующего по времени события-воздействия и, во-вторых, исключая посредством особых приемов экспериментальной изоляции и контроля альтернативные объяснения появления «отклика» с помощью посторонних влияний и конкурирующих каузальных гипотез¹. Соответственно, данные экспериментального исследования представляют собой наилучшее приближение к модели статистического вывода о наличии причинной взаимосвязи между воздействием и «откликом» или, в более привычных терминах, между независимой и зависимой переменными.

В главах, посвященных массовым опросам и статистическому анализу результатов социологического исследования, обсуждаются возможности и ограничения неэкспериментальных, сугубо статистических методов анализа связи между переменными. В частности, речь идет о случаях ложной корреляции между переменными, а также о том, что в неэкспериментальных исследованиях часто невозможно однозначно упорядочить изучаемые переменные во времени и, следовательно, задать направление причинной связи².

¹ Эти общие правила индуктивного вывода были сформулированы еще Дж. С. Миллем в «Системе логики» (1843).

² При отсутствии четких представлений о механизме причинной связи переменных (или отдельных событий) *A* и *B* единственным разумным основанием для построенного на опытных данных вывода о направлении причинной зависимости служит их последовательность во времени. Предшествование во времени — необходимое (но не достаточное) условие причинного воздействия *A* на *B*.

Основанная на эксперименте модель статистического вывода в значительной степени *лишена этих недостатков*, хотя ее использование в социальных науках во многих случаях также сталкивается с существенными техническими, этическими и прочими ограничениями.

Определение и виды эксперимента. Основные принципы экспериментирования в социальных науках

Экспериментальный метод пришел в социальные науки из естественных наук, где он примерно с XVII века стал основным способом опытной проверки научных теорий³. Самым популярным видом эксперимента в точных науках был и остается лабораторный эксперимент, получивший распространение также в науках о человеческом поведении.

Лабораторный, или истинный, эксперимент нацелен на проверку теоретической гипотезы и осуществляется в условиях максимального контроля над уровнем воздействия независимой переменной и очищения (изоляции) этого воздействия от посторонних влияний, оказываемых внешними, то есть иррелевантными с точки зрения проверяемой гипотезы, переменными. Экспериментальный контроль и изоляция позволяют отвергнуть иные возможные объяснения наблюдаемого эффекта — *конкурентные гипотезы*. Немаловажным условием обоснованности, *валидности* результатов, получаемых в лабораторном эксперименте, является возможность достаточно *надежного* измерения зависимой переменной. В этом случае при бесконечном количестве испытаний результаты неизбежных случайных возмущений в зависимой переменной «погасят» друг друга и исследователь получит точную оценку интересующего его воздействия.

³ Справедливости ради следует отметить, что торжеству экспериментального метода в естественных науках немало способствовало совершенство некоторых технических приемов и устройств, позаимствованных из донаучной традиции алхимиков. Последняя также уделяла большое внимание опытному знанию (как манифестации сверхчувственного знания) и активно использовала эксперименты-демонстрации или эксперименты, ориентированные на практические цели, если можно считать практической целью получение гомункулуса или философского камня.

На практике описанные требования к истинному эксперименту могут быть полностью воплощены лишь в *бесконечном идеальном эксперименте*, при проведении которого внешние, так называемые экзогенные, переменные остаются неизменными и изменяется лишь независимая переменная, что обеспечивает полную валидность выводов об изучаемом соотношении между независимой и зависимой переменными⁴. Идеальный эксперимент — это эталон, относительно которого могут оцениваться и сопоставляться эксперименты реальные, однако буквальное выполнение всех его требований обычно невозможно или даже бессмысленно с точки зрения конкретной научной задачи, стоящей перед исследователем.

Дональд Кэмпбелл приводит в качестве примера эксперимент Николсона и Карлслайла, которые продемонстрировали явление гидролиза, всего лишь «взяв в мае 1880 г. образец воды в районе Сохо (образец очень частный, характерный для местных условий и данного периода) и опустив в него весьма специфичный кусок медной проволоки, по которой пропускался электрический ток...». Влиянием посторонних факторов (прежде всего примесей) на протекание гидролиза в данном эксперименте можно было пренебречь, так как целью было подтверждение теоретически предсказанной закономерности, а не измерение скорости гидролиза или изучение особенностей его протекания в различных средах⁵.

В социальных науках от лабораторного эксперимента принято отличать *полевой* эксперимент, проводимый в естественных условиях и в большинстве случаев имеющий своей целью не столько проверку научной гипотезы о причинной связи между переменными, сколько оценку эффективности различных программ или методов воздействия.

Для многих прикладных социологических исследований, ориентированных на разработку и оценку социальных программ,

⁴ См.: Петухов В. В. Словарь экспериментатора // Готтсданкер Р. Основы психологического эксперимента. М.: Изд-во МГУ, 1982. С. 454.

⁵ Кэмпбелл Д. Модели экспериментов в социальной психологии и прикладных исследованиях: Пер. с англ. / Сост. и общ. ред. М. И. Бобневой. М.: Прогресс, 1980. С. 207–208.

социальное экспериментирование столь же типично, сколь типичны лабораторные эксперименты для социальной психологии или социологии малых групп. Социальное экспериментирование позволяет ответить на самые разнообразные вопросы, относящиеся преимущественно к сфере практической политики и администрирования, например, как влияет отмена смертной казни на показатели преступности, растет ли посещаемость музеев при снижении цен на входные билеты, во всех ли случаях повышение размера вознаграждения ведет к росту производительности труда и т. п.

Например, в исследовании воздействия детского сериала «Улица Сезам» на культурное и интеллектуальное развитие американских дошкольников⁶ полевой эксперимент охватил детей и родителей, проживавших в городах (Бостон, Дарэм, Феникс), а также в сельских районах Калифорнии и Филадельфии. В ходе эксперимента детей и их родителей побуждали к просмотру сериала (независимая переменная), фиксируя изменения в когнитивном развитии дошкольников с помощью тестов достижений и тестов общего развития (зависимые переменные). Двухлетний полевой эксперимент позволил продемонстрировать заметный обучающий эффект, связанный с просмотром сериала, особенно очевидный в группе детей из неблагополучных семей.

Полевой эксперимент — ведущий метод ориентированных на практику *оценочных исследований* (evaluative research)⁷. Однако далеко не всегда оценка эффективности новой, компьютеризованной системы обучения или, скажем, нового танкового прицела происходит в реальных полевых условиях. Иногда исследователи проводят эксперимент в условиях, имитирующих реальность или даже представляющих некоторые особенности реальной ситуации — обучения, вождения танка и т. п. — в преувеличенном, «очищен-

⁶ Цит. по: Bogatz G. A., Ball S. (eds.). The Second Year of Sesam Street: A Continuing Evaluation. Princeton (N. J.): Educational Testing Service, 1971. Vol. 1–2. // Riecken H. W., Boruch R. F. (eds.) Social Experimentation: A Method for Planning and Evaluating Social Intervention. N. Y., 1974. P. 306–307.

⁷ Об оценочных исследованиях см., в частности: Стародубцев С. П. Оценочные исследования: первое знакомство // Социологические исследования. 1992. № 7. С. 60–62.

ном» виде. Р. Готтсданкер предложил различать два типа полевых экспериментов — *эксперименты, дублирующие реальный мир* (то есть уже описанные «натурные» эксперименты), и *эксперименты, улучшающие реальный мир*⁸. Эксперименты, улучшающие реальный мир, прежде всего позволяют повысить валидность и надежность данных. Так, данные «натурного» эксперимента по эффективности нового метода обучения вождению автомобиля будут подвержены влиянию множества трудноконтролируемых различий в условиях видимости, ландшафте, состоянии дорог и используемых автомобилей, тогда как данные тренажерных испытаний будут меньше подвержены такого рода смещениям. Кроме того, надежность эксперимента в «улучшенных» условиях также повысится за счет возможности контролировать частоту «встречного движения» на экране тренажера, соблюдать четкий временной режим, исключая утомление оператора, и т. д.

Для того чтобы понять, чем руководствуются социологи, осуществляя выбор одного из перечисленных видов эксперимента и конкретного плана построения экспериментального исследования, нам следует сначала рассмотреть ключевые понятия и принципы, лежащие в основе современного подхода к социальному экспериментированию.

Суммируя вышеизложенное, можно сказать, что в лабораторном эксперименте обоснованность выводов экспериментального исследования, то есть их *валидность* и *надежность*, обеспечиваются благодаря трем принципам планирования эксперимента: 1) *контролю* над уровнем независимой переменной; 2) *изоляции основного эффекта* (то есть собственно воздействия независимой переменной на зависимую переменную) от влияния посторонних, смешивающих факторов, а также 3) *многократному воспроизведению* полученных результатов, которое позволяет нивелировать случайные изменения результата⁹ отдельных испытаний, связанные с несистематическими

⁸ Готтсданкер Р. Основы психологического эксперимента. М.: Изд-во МГУ, 1982.

⁹ Случайные изменения результата, то есть фиксируемое в конкретном испытании значение *зависимой переменной*, которое, собственно, и характеризует основной эффект — воздействие независимой переменной на зависимую (или отсутствие такового).

колебаниями фона, случайными ошибками, усталостью и т. п. При этом первые два принципа планирования лабораторного эксперимента позволяют обеспечить *валидность как соответствие эксперимента его цели*, измерение именно того эффекта, который предполагалось измерить. Идеальный, то есть совершенно валидный, эксперимент фиксирует лишь то отношение между переменными, которое и планирует изучить экспериментатор, и «отсекает» любые другие источники систематической вариации результатов. Валидность эксперимента, следовательно, определяет достоверность выводов о наличии либо отсутствии предполагаемой причинной связи и о подтверждении либо неподтверждении проверяемой в эксперименте теоретической гипотезы (рис. 4.1). Третий принцип обеспечивает надежность результатов — защиту от случайной ошибки, являющейся, как говорилось выше, необходимым условием валидности.

Однако большая часть экспериментов в социальных науках (как, впрочем, и в ряде инженерных дисциплин или агробиологии) происходит в условиях, когда перечисленные принципы не могут быть реализованы полностью. Возникающие здесь ограничения имеют технический, а иногда — скорее принципиальный характер. Если, например, в социологическом эксперименте независимыми переменными служат раса или социальное происхождение, то мы даже технически — не говоря уже о соображениях морали — не можем осуществлять полный контроль над этими переменными, то есть совершенно произвольно определять их значение для каждого отдельного случая (субъекта). И даже если бы эту трудность можно было бы каким-то образом преодолеть на время проведения испытаний, нам было бы трудно изолировать интересующий нас основной эффект от влияния сопутствующих, «закоррелированных» с независимой переменной факторов — связанного с расой или происхождением специфического социального опыта, соответствующих социальных навыков и т. п. (Со схожими трудностями сталкивается и агробиолог, оценивающий урожайность нового сорта пшеницы и пытающийся отделить главный фактор от прочих, также влияющих на урожайность: различий в освещенности опытных участков, в составе почвы и т. п.)

Знаменитый английский статистик Р. Фишер первым обосновал возможность применения несколько иного подхода к планированию

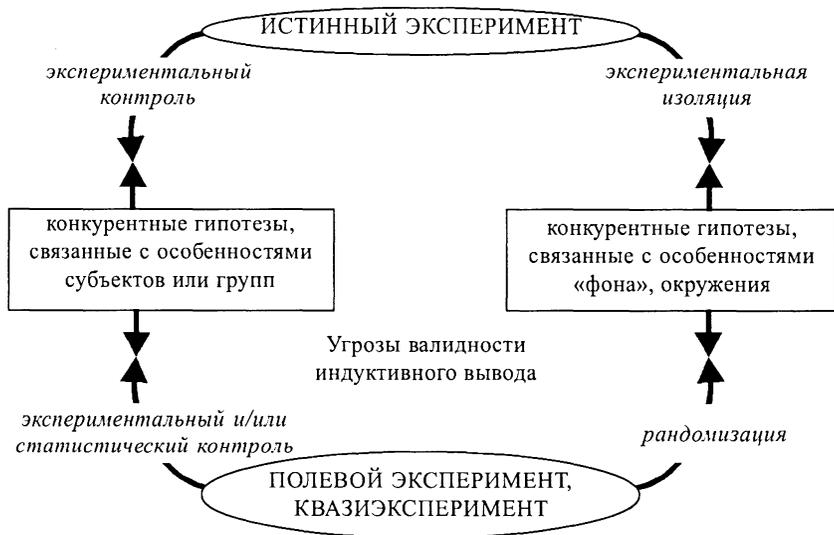


Рис. 4.1. Типы экспериментов

полевых экспериментов, лабораторных экспериментов с неполным контролем, а также квазиэкспериментов (о последних речь пойдет позднее). Этот подход основан на целенаправленном использовании законов случая и теории вероятности. Он требует введения в планирование эксперимента *принципа рандомизации*.

Рандомизация (от англ. random — случайный, выбранный наугад) — это стратегия случайного распределения субъектов по различным условиям (режимам) эксперимента и экспериментальным группам.

Вспользуемся в качестве иллюстрации тем же простым примером психофизического опыта, которым пользуется сам Фишер при обсуждении роли принципа рандомизации в планировании эксперимента¹⁰.

Представим себе, что некая леди взялась угадать, в какие четыре из восьми чашек чая с молоком сначала налили молоко, а в

¹⁰ Fisher R. A. The Design of Experiment. 3rd ed. L.: Oliver and Boyd, 1942. P. 17–19.

какие — чай. Экспериментальная гипотеза состоит, таким образом, в том, что испытуемая обладает описанной способностью различения. Если воспользоваться простейшей формулой из комбинаторики, то общее число способов выбрать четыре чашки из восьми равно числу размещений A :

$$A_8^4 = \frac{8 \times 7 \times 6 \times 5}{4!}.$$

Иными словами, если бы испытуемая не обладала способностью отличать «молочно-чайные» чашки от «чайно-молочных» и прибегла к случайному угадыванию, то вероятность правильно определить 4 чашки составила бы $\frac{1}{70}$ (из 70 существующих различных способов выбрать 4 чашки из 8 лишь один является правильным). Говоря точнее, при достаточно длинной серии испытаний частота правильных ответов приближалась бы к $\frac{1}{70}$. Если же частота правильных ответов существенно, с точки зрения избранного статистического критерия, превышает частоту случайного угадывания, то мы вправе сделать вывод о том, что интересующая нас специальная способность действительно существует.

Однако описанная схема испытания явно не соответствует тем требованиям, которые предъявляются к уровню контроля над переменными и изоляции основного эффекта в лабораторном эксперименте. Возможные угрозы *валидности* наших статистических выводов носят довольно очевидный характер. Во-первых, мы не обладаем никакими средствами прямого измерения зависимой переменной — способности вкусового различения. Правильные угадывания — это лишь косвенные индикаторы такой способности и могут отражать влияние «посторонних» переменных, даже превосходящее основной эффект. Если, например, во все чашки, в которые сначала было налито молоко, был добавлен сахар, ясно, что все они будут безошибочно опознаны. Статистические выводы о значимости окажутся, таким образом, невалидными, то есть не имеющими отношения к нашей экспериментальной гипотезе (хотя и подтверждающими конкурентную гипотезу о способности распознавать сладкий вкус). Систематическое постороннее влияние, составляющее угрозу валидности статистического вывода, может

носить и другой характер: «молочные» и «чайные» чашки могут отличаться друг от друга толщиной, температурой, цветом. Чтобы устранить перечисленные угрозы валидности и нивелировать возникающие систематические смещения, экспериментатор должен использовать принцип рандомизации, то есть обеспечить случайный порядок предъявления стимулов-чашек и их оценки, определяемый, например, с помощью таблицы случайных чисел или жребия. В идеале и отбор испытуемых для каждой серии опытов, и распределение уровней независимой переменной между чашками («молоко-чай» или «чай-молоко») должны основываться на законах случая.

При планировании конкретного экспериментального исследования описанные принципы находят воплощение при разработке *плана, или схемы, эксперимента*, определяющего порядок предъявления испытуемым (или их группам) различных уровней (условий) независимой переменной для адекватной проверки экспериментальной гипотезы¹¹.

Основные экспериментальные планы с контрольной группой и рандомизацией

В социологии, психологии и других поведенческих науках особую роль играет использование принципа рандомизации *при распределении испытуемых по группам*. В эксперименте с неполным контролем или в полевом эксперименте, происходящем в естественных условиях (то есть в условиях школьного класса, промышленной организации, городского района и т. д.) часто нужно доказать не только наличие ожидаемого эффекта в результате некоторого воздействия *X*, но и *отсутствие того же эффекта в тех случаях, когда воздействия не было*. Например, исследователь, изучающий воздействие просмотра антивоенных фильмов на изменения установок студентов, случайным образом отбирает из некоторой совокупности студентов *экспериментальную группу*, которой будет показан антивоенный фильм, а также *контрольную группу*, которой он продемонстрирует нейтральный фильм, никак не связанный с

¹¹ Петухов В. В. Указ. соч. С. 446.

изучаемыми установками. План этого простейшего *рандомизированного эксперимента с предварительным и итоговым тестированием и контрольной группой* ($RT_{1-2}C$) будет выглядеть таким образом:

$$\begin{array}{cccc} R & O_1 & X & O_2 \\ R & O_3 & & O_4 \end{array}$$

где R — процедура рандомизации (случайного распределения по группам), $O_{1,2}$ — уровни установок в экспериментальной группе до и после просмотра фильма X , $O_{3,4}$ — уровни установок в контрольной группе, не смотревшей фильма. Использование контрольной группы позволяет устранить некоторые важнейшие угрозы валидности эксперимента. Во-первых, если бы исследователь отказался от использования контрольной группы и ограничился тестированием, то есть измерением установок «до-после» просмотра, то обнаруженные изменения в уровне установок можно было бы приписать влиянию на испытуемых *самого по себе факта участия* в эксперименте. Испытуемые, возможно, осознавали, что они отобраны для важного исследования и стремились некоторым образом соответствовать своей роли и оправдать некие гипотетические «ожидания» экспериментатора. Описанная угроза валидности широко известна, и для нее существует несколько обозначений. В психологическом тестировании и экспериментальной психологии это называют «эффектом морской свинки» или «мотивом экспертизы». Иногда применяют термин, возникший в медицине, где при клинических испытаниях новых фармакологических средств и методов лечения часто наблюдают «эффект плацебо», то есть заметное улучшение статуса у многих участников контрольной группы, в которой вместо реального воздействия использовались индифферентные средства и нейтральные врачебные манипуляции. В социологии самое популярное обозначение систематического смещения, возникающего из-за реакции испытуемых на ситуацию эксперимента — это «хоуторнский эффект».

В так называемых «хоуторнских экспериментах» (по названию промышленного предприятия в Чикаго) исследовались организационные и социально-психологические факторы, влияющие на производительность труда. Исследователи обна-

ружили, что эффект роста производительности труда в бригадах сохранялся даже при отсутствии собственно экспериментального воздействия. Предположительной причиной этого явления был рост группового самосознания у участников эксперимента¹².

«Хоуторнский эффект»¹³ — это угроза валидности, связанная с особенностями экспериментальной группы.

Еще одна угроза валидности выводов, которую позволяет преодолеть описываемый план эксперимента, также связана с особенностями групп, а именно — с *процессом отбора для участия в эксперименте*. Если бы мы отказались от предварительного тестирования и случайного распределения испытуемых по экспериментальной и контрольной группам, мы совершили бы методическую ошибку, весьма характерную для любых экспериментов с добровольцами. Очень часто исследователи отбирают испытуемых для участия в специальной программе обучения или в новаторском организационном проекте, основываясь на изъявленном ими желании, а затем сравнивают результаты, показанные участниками оцениваемой программы, с результатами какой-либо другой доступной группы (или даже случайной выборки из соответствующей генеральной совокупности), не участвовавшей в такого рода программе. Однако такое сравнение некорректно: само по себе желание участвовать в эксперименте часто свидетельствует об изначально более высокой мотивации, осведомленности или интеллекте. Эти факторы сами по себе, или *взаимодействуя* с главной независимой переменной *X*, могут объяснить значимые различия в результатах, показанных экспериментальной группой. В только что описанном примере сравнение группы добровольцев, пожелавших посмотреть антивоенный фильм, с прочими студентами может вести к завышенной оценке воздействия просмотра, если добровольцы изначально проявляют больший интерес к политическим проблемам.

¹² См.: *Roethlisberger F. G., Dickson W. J. Management and the Worker*. Cambridge: Harvard University Press, 1939.

¹³ «Хоуторнский эффект», который, следуя выработанному в естественных науках образцу разрешения споров о приоритете, можно было бы именовать «хоуторнским эффектом плацебо, данного морской свинке при проведении эксперименты»...

Лишь случайное распределение добровольцев по контрольной и экспериментальной группам при соблюдении «непрозрачности» такого распределения для всех испытуемых (участники обеих групп должны считать, что они подвергаются некоторому экспериментальному воздействию) позволяет судить о роли независимой переменной X в возникновении межгрупповых различий¹⁴.

Вышеописанные угрозы валидности связаны преимущественно с особенностями групп, проявляющимися на стадии отбора или в ходе эксперимента. Однако рандомизация позволяет справиться и с некоторыми угрозами валидности, исходящими от переменных внешнего окружения, *фона*. К фоновым относятся, в частности, «возможные влияния времени года или событий, возникающих на институциональном уровне», а также факторы *естественного развития* — «все те биологические или психологические процессы, которые независимо от конкретных внешних событий систематически изменяются с течением времени»¹⁵. Испытуемые взрослеют, обучаются, устают, улучшают свои результаты при повторных тестированиях и т. п., что может сказываться на их результатах. Однако если такие посторонние влияния не оказывают избирательного воздействия только на членов экспериментальной группы, они будут вносить вклад лишь в случайную ошибку, а не в систематическое смещение. Иными словами, они будут с равной вероятностью распределены между случайным образом отобранными участниками контрольной и экспериментальной групп. План типа $RT_{1-2}C$ позволяет обеспечить случайный и равновероятный характер внешних, фоновых воздействий на контрольную и экспериментальную группу. Более того, он позволяет «вычесть» величину фонового влияния и оценить чистый основной эффект: если внешнее влияние все же имело место, оно в равной степени действовало на

¹⁴ Сама процедура случайного распределения может быть осуществлена аналогично процедуре построения простой вероятностной выборки с использованием таблицы случайных чисел (см. гл. 7). Если единицы отбора — отдельные испытуемые, классы, городские районы — имеют тенденцию к естественной группировке либо экспериментатор особенно заинтересован в сравнении малочисленных подгрупп, то можно использовать вероятностные процедуры с кластеризацией и стратификацией.

¹⁵ Кэмпбелл Д. Указ. соч. С. 51.

показатели и экспериментальной, и контрольной групп; следовательно, разность между средними значениями первого и второго замеров уровня зависимой переменной в контрольной группе ($O_4 - O_3$) нужно вычесть из аналогичной разности значений зависимой переменной, зафиксированной в экспериментальной группе ($O_2 - O_1$), то есть:

$$|X| = (O_2 - O_1) - (O_4 - O_3),$$

при этом предварительное и послезэкспериментальное тестирование в экспериментальной и контрольной группах должны проводиться практически одновременно.

Систематическая угроза валидности, связанная с фоновыми факторами, может все же возникнуть и при использовании рандомизации и контрольной группы. Это происходит тогда, когда фоновые факторы *взаимодействуют* с независимой переменной (или некоторыми ее уровнями). Природу такого взаимодействия легко понять на примере исследования, в котором изучается влияние тревожности, возникающей в ситуации неопределенности, на успешность решения сложных задач. В такого рода экспериментах для создания ситуации неопределенности и повышения реактивной тревожности (независимые переменные) часто используют неясные инструкции, косвенные негативные оценки действий испытуемого, высказываемые лицами, проводящими эксперимент (типа «ну-ну, посмотрим, как вы сможете это использовать»), а также предварительные серии, где испытуемому приходится решать заведомо неразрешимые задачи. Разумеется, всем этим воздействиям (X) подвергаются лишь члены экспериментальной группы. Если испытуемые — это студенты, которым в силу случайного стечения обстоятельств через неделю предстоит сдавать экзамены, или сотрудники подразделения фирмы, ожидающие скорой переаттестации, то эти факторы «фона» будут взаимодействовать с независимыми переменными, не только суммируясь с ними, но и *усиливая* их эффект. Негативные подкрепления, получаемые в ходе эксперимента, будут восприниматься значительно острее накануне экзаменационной сессии или переаттестации, а связанные с этими событиями фоновая тревожность и неопределенность будут взаимодействовать с тревожностью и неопределенностью, создаваемыми преднамеренно.

Оценить чистый эффект взаимодействия с помощью сравнения с результатами контрольной группы вышеописанным способом в данном случае невозможно, так как *взаимодействие* фоновых смещений с основным эффектом происходит лишь там, где имеет место экспериментальное воздействие X . В контрольной группе приближающееся неприятное событие также может повлиять на результаты итогового замера или даже обоих замеров, но оно не будет взаимодействовать, «перемножаться» с воздействием независимой переменной (такое взаимодействие можно описать в более точных терминах, однако это требует введения некоторых статистических понятий, см., в частности, с. 107–110). Для оценки величины взаимодействия фоновых смешивающих переменных с основным эффектом нужны более сложные экспериментальные планы. Некоторые из них будут обсуждаться ниже.

Иногда имеет смысл воспользоваться упрощенным вариантом описанного плана с рандомизацией и контрольной группой, а именно **планом с рандомизацией без предварительного тестирования (RTC)**, который схематически выглядит следующим образом:

R	X	O_1
R		O_2

Привлекательность этого плана заключается прежде всего в его экономичности. Если при распределении испытуемых между группами и уровнями воздействия использовалась истинная вероятностная процедура¹⁶, то проведение предварительного тестирования — как в плане $RT_{1-2}C$ — лишь увеличивает стоимость эксперимента,

¹⁶ Самая распространенная ошибка, совершаемая исследователями при проведении рандомизации в эксперименте, — замена случайного распределения *попарным уравниванием*, когда, скажем, к двум пожилым домохозяйкам со средним образованием в экспериментальной группе подбирают двух пожилых домохозяек в контрольной группе и т. д. Попарное уравнивание может вести к таким же неконтролируемым смещениям, как и использование квотной выборки (см. гл. 7). Иногда в случаях, когда отбор производится внутри команд, школьных классов и других естественных группировок, попарное уравнивание после разбиения группы пополам допустимо, если приписывание групп к контрольным или экспериментальным условиям будет проводиться *случайно*. Однако в двумерном эксперименте (типа «есть воздействие — нет воздействия») такой подход неприемлем, так как ведет к резкому снижению статистических свойств получаемых оценок.

не оказывая существенного влияния на качество получаемых данных. По сути, правильно осуществленная рандомизация (R) — это наилучшая гарантия отсутствия изначального смещения между группами (то есть равенства исходных среднегрупповых значений зависимой переменной O в контрольной и экспериментальной группах). Кроме того, в широкомасштабных социологических исследованиях, а также в прикладных исследованиях в области педагогики и социальной работы проведение *и предварительных, и повторных* измерений зависимой переменной (политических установок, криминального поведения и т. п.) часто неосуществимо либо ведет к возникновению реакции на саму процедуру тестирования. Если, к примеру, мы изучаем влияние участия в предвыборной групповой дискуссии на последующую поддержку политических партий, то предварительное измерение политических установок может повлиять на активность опрошенных в ходе самой дискуссии и их последующие установки.

Можно заключить, что план RTC предпочтителен всегда, когда нежелательно повторное предъявление весьма специфичных по содержанию и схожих по форме тестов, вопросов, измерительных процедур. Иными словами, этот план заслуживает широкого применения в социологических и оценочных исследованиях. Кроме того, он является самым приемлемым средством проведения так называемых *методических экспериментов*, в которых оценивается эффективность различных видов опроса, способов заполнения и форматов анкет и т. п. (примером такого методического эксперимента является описанное в главе «Массовые опросы» исследование Бредберна и Судмана, в котором эффективность метода «случайного ответа» сравнивалась с эффективностью использования данных административной статистики). План RTC использовался, в частности, в одном из практически ориентированных социальных экспериментов по совершенствованию системы правоохранительных и пенитенциарных учреждений, проводившихся в США в 1960 — 70-е гг.

В 1961 г. Федеральное бюро пенитенциарных учреждений США начало трехлетний эксперимент, целью которого было изучить воздействие поддерживающего консультирования и специальных реабилитационных программ на поведение

мальчиков-подростков, находящихся в исправительных учреждениях. Местом проведения эксперимента стала специальная школа-колония для несовершеннолетних правонарушителей.

При проведении рандомизации мальчиков случайным образом распределяли между отрядами, где проводился эксперимент, и отрядами, служившими контрольными (каждый отряд жил в отдельно расположенном корпусе). В экспериментальных отрядах было увеличено количество воспитателей, проводились специальные индивидуальные и групповые консультации, использовалась система вознаграждений за хорошее поведение. В контрольных группах применялись обычные методы воспитания и обучения, а также традиционные наказания за нарушение внутреннего распорядка. Результаты эксперимента показали, что мальчики из экспериментальных групп раньше покидали спецшколу, лучше успевали в учебе, вели себя адаптивнее. Не было обнаружено значимых различий в показателях рецидивной преступности для подростков из экспериментальных и контрольных групп, освобожденных из школы-колонии, однако ребята из экспериментальных групп значительно отличались от ребят из контрольных групп по показателю тяжести вновь совершенных преступлений (первые в случае рецидива совершали менее тяжкие преступления). По результатам эксперимента было принято решение о внедрении экспериментальной коррекционной программы во всех подразделениях школы¹⁷.

Еще один популярный план с рандомизацией и контрольной группой — это план Соломона¹⁸. *План Соломона* — это расширенный вариант плана $RT_{1-2}C$, позволяющий проконтролировать и оценить эффекты естественного развития и фона, а также определить взаимодействие эффекта тестирования с основным воздействием X . Здесь наряду с экспериментальной и контрольной группами с предварительным тестированием используются экспе-

¹⁷ *Federal Bureau of Prisons*. Rational Innovation: An Account of Changes in the Program of the National Training School for Boys from 1961–1964. Washington, (D. C.), 1964. Цит. по: Riecken H. W., Boruch R. F. (eds.) Social Experimentation. N. Y.: Academic Press, 1974.

¹⁸ Впервые описан в статье Р. Соломона. См.: Solomon R. L. An Extension of Control Group Design // Psychological Bulletin. 1949. Vol. 46. № 1. P. 137–150.

риментальная и контрольная группы без предварительного тестирования (как в плане RTC). Схематически это выглядит следующим образом:

R	O_1	X	O_2
R	O_3		O_4
R		X	O_5
R			O_6

Очевидно, что если главный эффект X реален, то даже при наличии существенного эффекта тестирования («хоуторнского эффекта») будут выполняться четыре неравенства: $O_2 > O_1$; $O_2 > O_4$; $O_5 > O_6$; $O_5 > O_3$. Оценкой сравнительной величины эффекта предварительного тестирования (без взаимодействия с X) может служить величина разности $O_6 - O_3$. Сравнение O_6 с O_1 и O_3 позволяет оценить влияние фоновых факторов и факторов естественного развития¹⁹.

До сих пор мы обсуждали содержательные аспекты проверки экспериментальной гипотезы о наличии главного эффекта X с помощью различных планов эксперимента, а также преимущества разных планов для обеспечения разных аспектов валидности. Очевидно, однако, что в каждом конкретном эксперименте величина главного эффекта, то есть наблюдаемого различия результатов экспериментальной и контрольной группы, будет варьировать не только под воздействием независимой переменной, но и просто в результате действия различных случайных возмущений. Конечно, если бы наш эксперимент был идеален и абсолютно надежен (см. выше), то при каком угодно числе повторений мы бы всегда получали одну и ту же истинную оценку величины воздействия (при отсутствии или контроле смешивающего влияния дополнительных переменных). Однако реальные эксперименты — особенно, как уже говорилось, эксперименты в социальных науках — не бывают и не могут быть идеальными и безупречно надежными. Следовательно, перед исследователем всегда стоит задача *статистической оценки значимости полученных результатов*.

Вероятностным «воплощением» содержательной экспериментальной гипотезы является *статистическая гипотеза*. Принятие или непринятие статистической гипотезы — необходимое, но не

¹⁹ См.: Кэмпбелл Д. Указ. соч. С. 88–89.

достаточное условие принятия или отвержения содержательной гипотезы, проверяемой в эксперименте. Проверяемая в конкретном эксперименте статистическая гипотеза всегда формулируется как гипотеза о том, что при бесконечном количестве повторений этого эксперимента среднее различие между экспериментальной и контрольной группами (или между воздействием разных уровней независимой переменной) равнялось бы нулю. Такую статистическую гипотезу, фактически сводящуюся к утверждению о случайном характере наблюдаемых в реальном эксперименте различий, называют *нулевой гипотезой*, или *нуль-гипотезой* (H_0). Отвержение или неотвержение нуль-гипотезы позволяет говорить о том, что в данном эксперименте содержательная гипотеза подтвердилась, либо подтвердилась противоположная ей альтернативная гипотеза, либо не было получено подтверждения ни одной из них.

Вспомним воображаемый эксперимент Фишера с чаем и молоком, описанный в начале главы. Мы отмечали, что при использовании тактики случайного угадывания испытуемая смогла бы правильно определить последовательность наполнения чашки примерно в 1 случае из 70. Следовательно, 2 «попадания» из 100 даже при очень большом количестве испытаний едва ли могут считаться значимым результатом. Куда достовернее выглядят 90 или даже 95 «попаданий» из 100 (оставшиеся «ошибки» можно отнести на счет действия случайных факторов). Чтобы определить уровень значимости и построить доверительный интервал, в этом случае достаточно метода, описанного в гл. 8.

Вообще, критерии значимости и статистические методы, используемые при проверке статистической гипотезы для конкретного плана эксперимента, называют *статистическими моделями*. Для планов с контрольной группой основная статистическая модель — это использование *t*-критерия, о чем подробнее говорится чуть ниже. Для более сложных планов многомерных или факторных экспериментов, общий обзор которых дается в следующем разделе главы, ведущие статистические модели — это дисперсионный анализ и использование *F*-критерия Фишера.

Итак, для того чтобы оценить статистическую значимость в элементарных рандомизированных планах, описанных выше, нам необходимо проверить *статистическую гипотезу о разности сред-*

них значений зависимой переменной в контрольной и экспериментальной группах. Конкретное значение разности средних значений зависимой переменной в экспериментальной и контрольной группе, обнаруженное в отдельном эксперименте (скажем, 4 балла по некоторой «шкале пацифизма»), нужно соотнести с определенным интервалом, в который это значение «укладывается» с заданной (доверительной) вероятностью. Иными словами, нужно решить задачу интервального оценивания, подобную задаче оценки отдельного параметра совокупности в выборочном обследовании (эта задача описана в гл. 8). Разница в том, что, проводя эксперимент, мы интересуемся не вероятными пределами, в которых лежит некая характеристика выборки из реально существующей совокупности, а пределами, в которых лежит полученный нами в эксперименте результат относительно результата воображаемой бесконечной совокупности идентичных экспериментов. Нулевая гипотеза утверждает, что истинное значение различия средних равно нулю, варьируя в каких-то пределах от эксперимента к эксперименту (то есть $H_0 : (\bar{X}_{\text{эксперимент.}} - \bar{X}_{\text{контрольн.}}) = 0$). Если удастся показать, что полученное в эксперименте значение разности групповых средних не позволяет принять нулевую гипотезу, то делается вывод о подтверждении гипотезы, противоположной нулевой (то есть $H_1 : (\bar{X}_{\text{эксперимент.}} - \bar{X}_{\text{контрольн.}}) \neq 0$) — о статистической значимости различий между группами — и, значит, о подтверждении экспериментальной гипотезы (или о подтверждении гипотезы, противоположной экспериментальной — если различие между экспериментальной и контрольной группой оказалось с обратным знаком). Заметьте, что нулевая гипотеза всегда формулируется как гипотеза о том, что истинное значение разности средних (или, скажем, величины взаимосвязи между двумя переменными) равно нулю, а полученные в эксперименте величины отличаются от нуля исключительно из-за случайной ошибки выборки. Чем дальше от нуля — в ту или другую сторону — расположено наблюдаемое значение, тем больше его статистическая значимость и меньше вероятность того, что оно явилось результатом ошибки выборки.

Для того чтобы сравнить полученное в эксперименте с контрольной и экспериментальной группами значение разности между средними с гипотетическим *выборочным распределением этой*

величины для бесконечного числа испытаний (такие распределения имеются не только для разности средних, но и для средних величин, стандартных отклонений и т. д. — см. с. 247–249), нужно вычислить стандартную ошибку разности между средними. Формула для стандартной ошибки разности между средними — $S_{M_3-M_k}$ — немного отличается от формулы стандартной ошибки средней S_M^{20} , приведенной в гл. 8. Тем не менее, она весьма проста:

$$S_{M_3-M_k} = \sqrt{\frac{S_3^2}{n_3} + \frac{S_k^2}{n_k}},$$

где S_3 и S_k — величины стандартного отклонения, рассчитанные для экспериментальной и контрольной групп, n_3 и n_k — число наблюдений (испытуемых) в экспериментальной и контрольной группах.

После этого нужно определить, на сколько единиц стандартной ошибки отстоит полученная разность средних ($\bar{X}_{\text{эксперимент.}} - \bar{X}_{\text{контрольн.}}$) от нуля, представляющего собой — в согласии с нуль-гипотезой — среднее гипотетического распределения разностей средних, t-распределения. Для этого полученную в эксперименте разность групповых средних нужно перевести в t-единицы (то есть единицы стандартного отклонения для t-распределения). Для данной разности средних величину t можно вычислить по формуле:

$$t = \frac{(\bar{X}_{\text{эксперимент.}} - \bar{X}_{\text{контрольн.}})}{S_{M_3-M_k}}.$$

²⁰ Напомним, что стандартная ошибка среднего равна стандартному отклонению теоретического распределения выборочных средних. Эта величина используется как мера ошибки выборки (с. 250–251). Выборочные средние (то есть средние значения множества выборок из гипотетической генеральной совокупности) распределены приблизительно нормально. А вот распределение *разностей выборочных средних* (t-распределение) выглядит как «уплощенное» нормальное распределение, причем чем меньше размер выборки, тем более плоским и «размазанным» выглядит t-распределение. Это распределение было впервые описано У. Госсетом (опубликовавшим свои результаты под псевдонимом Стьюдент). Госсет показал, что для *малых выборок* вероятностное оценивание выборочных средних дает надежные результаты лишь в том случае, если вместо самого распределения средних мы рассматриваем распределение их разностей.

Полученное значение t нужно сравнить с соответствующим значением из таблицы t -распределения для избранного уровня значимости ($p = 0,05$ или $0,01$) и числа *степеней свободы*, соответствующего количеству наблюдений в каждой группе (или подвыборке). Число степеней свободы — довольно сложное статистическое понятие, анализ которого выходит за пределы этого учебника (в самом общем виде оно обсуждается в гл. 7). На практике число степеней свободы можно рассматривать как величину, равную числу наблюдений (испытуемых, опрошенных, баллов и т. п.) минус число оцениваемых параметров. Для разности средних двух групп это составит число наблюдений в экспериментальной группе минус один ($n_э - 1$) плюс число наблюдений в контрольной группе минус один ($n_к - 1$):

$$N_{ст. своб.} = (n_э - 1) + (n_к - 1).$$

Таблицы t -распределения можно найти в любом учебнике или справочнике по статистике (см. список дополнительной литературы к данной главе, а также к гл. 8). Здесь мы приводим лишь фрагмент такой таблицы.

Рассмотрим пример вычисления t для описанного выше эксперимента, в котором изучалось воздействие антивоенного фильма на изменение установок студентов. Пусть для контрольной и экспериментальной групп при итоговом тестировании по шкале пацифистских установок были получены следующие результаты:

Контрольная группа

$$n_к = 28 \text{ чел.}$$

$$\bar{X}_{\text{контрольн.}} = 15,3$$

$$S_к = 5,6$$

Экспериментальная группа

$$n_э = 34 \text{ чел.}$$

$$\bar{X}_{\text{эксперимент.}} = 19,5$$

$$S_э = 3,4$$

Наша статистическая задача заключается в том, чтобы определить, отличаются ли средние двух групп настолько, чтобы можно было отвергнуть нулевую гипотезу о том, что эти средние взяты из

Т а б л и ц а 4.1

Сокращенная таблица t-распределения Стьюдента (W. Gosset, 1908)

Число степеней свободы	$P = 0,05$	$P = 0,01$
1	12,706	63,657
2	4,303	9,925
5	2,571	4,032
8	2,306	3,355
10	2,228	3,169
14	2,145	2,977
16	2,120	2,921
20	2,086	2,845
30	2,042	2,750
60	2,000	2,660
120	1,980	2,617
∞	1,960	2,576

одной генеральной совокупности. Воспользуемся приведенной выше формулой для вычисления значения t^{21} :

$$1) S_{M\bar{D}-M\bar{K}} = \sqrt{\frac{5,6^2}{28} + \frac{3,4^2}{34}} \cong 1,2083;$$

$$2) t = \frac{(19,5 - 15,3)}{1,2083} \cong 3,4760.$$

²¹ Используемая нами формула основана на некоторых важных предположениях: о том, что группы отбирались независимо и случайно; что дисперсии соответствующих генеральных совокупностей неравны; что параметры совокупностей распределены нормально. Существуют и иные, несколько отличные формулы для расчета t , которые применяются в тех случаях, когда сравниваемые подвыборки-группы невелики и получаемые для них данные каким-то образом «связаны», скоррелированы (например, если мы сравниваем групповые средние братьев и сестер или средние оценки одних и тех же школьников в первом классе и на выпускных экзаменах). Соответствующие формулы и пояснения можно найти в любом статистическом руководстве, а также в книгах, включенных в список дополнительной литературы по теме.

Число степеней свободы в приведенном примере: $(28 - 1) + (34 - 1) = 60$.

Полученное значение $t = 3,4760$ заведомо превосходит табличные значения и для $p < 0,05$, и для $p < 0,01$ (на 5%-м уровне значение t для 60 степеней свободы составит 2,00, а на 1%-м — 2,660). Следовательно, мы можем отклонить нулевую гипотезу и сделать вывод, что существует статистически значимая разница между средними уровнями пацифизма в группе студентов, посмотревших антивоенный фильм, и в контрольной группе.

Важно, однако, всегда помнить о том, что статистическая значимость результатов совершенно отлична от их *содержательной* значимости! Даже высокая статистическая значимость результатов эксперимента не гарантирует, что эти результаты будут иметь сколько-нибудь интересную интерпретацию и повлияют на состояние современного социологического знания. Содержательная значимость зависит прежде всего от нашей способности увязать экспериментальную гипотезу с существующими социологическими теориями.

Многомерные и факторные эксперименты: общий обзор

В описанных выше экспериментах с контрольной группой каждый раз используются лишь два типа условий — «есть воздействие» либо «нет воздействия». Эти два типа условий, по сути, можно рассматривать как два уровня независимой переменной, которым можно присвоить условные числовые значения, например «1» и «0». Иными словами, с точки зрения уровня измерения независимая переменная является номинальной, качественной. В контрольной группе ее значение равно нулю, в экспериментальной — единице. Однако исследователь часто располагает значительно большей информацией о независимой переменной и способен измерить и проконтролировать ее по крайней мере на трех-четырёх уровнях значений. Соответственно, экспериментальная гипотеза может быть сформулирована в терминах более или менее интенсивного воздействия либо наличия–отсутствия «отклика» зависимой переменной при конкретных уровнях независимой переменной.

В психологии хорошо известен закон «оптимума мотивации», так называемый закон Йеркса–Додсона.

В начале XX века Р. Йеркс изучал, как влияет негативное подкрепление в форме удара электрическим током на выработку элементарных навыков у животных. В частности, в опытах с «танцующими мышами» (разновидность домашней мыши, имеющая генетический дефект, который заставляет ее двигаться по кругу или по восьмерке) он использовал три уровня силы тока — «сильный» (500 усл. ед.), «средний» (300 усл. ед.) и «слабый» (125 усл. ед.). Мышь должна была научиться выбирать один из двух туннелей. В конце туннеля ее в любом случае ожидало «вознаграждение» — мышь противоположного пола. При ошибочном выборе (белый туннель) мышь испытывала удар током, при правильном выборе (черный туннель) негативное подкрепление отсутствовало. Местоположение туннелей (слева–справа) менялось случайным образом от пробы к пробе. Выяснилось, что быстрее всего обучение происходит при «средней» величине стимуляции. Обнаруженный в этом эксперименте нелинейный характер связи между величиной стимула к решению определенной задачи и успешностью решения был затем неоднократно подтвержден и во многих других экспериментах, в том числе с испытуемыми-людьми и с позитивной стимуляцией. Чрезмерная мотивация и чрезмерная величина подкрепления, как и слабая мотивация, всякий раз оказывали меньшее воздействие на успешность выполнения различных задач.

Эксперименты, в которых используется несколько (более двух) уровней независимой переменной, называются многоуровневыми. Схема вышеописанного эксперимента с рандомизацией и тремя уровнями независимой переменной (X_1, X_2, X_3) такова:

R	X_1	O_1
R	X_2	O_2
R	X_3	O_3

Экспериментальная гипотеза в этом случае формулируется как гипотеза об отношениях значений O_1, O_2 и O_3 (в рассмотренном примере $O_1 < O_2$ и $O_2 > O_3$). Независимая переменная в многомерном эксперименте может иметь и более трех уровней. Иначе говоря, она может быть «нормальной» количественной переменной, из-

меренной на интервальном или абсолютном уровне. Соответственно, гипотеза многомерного эксперимента может формулироваться в более точных терминах — как гипотеза об «относительно-абсолютных» или даже «абсолютно-абсолютных» отношениях переменных. Например, в эксперименте может изучаться влияние привлекательности лектора на частоту посещения занятий студентами, воздействие количества доступных источников информации о продукте на формирование потребительских предпочтений либо характер взаимосвязи между размером денежного вознаграждения испытуемых и успешностью решения ими однотипных задач. Таким образом, многомерные эксперименты позволяют *проверять более тонкие и точные содержательные гипотезы о механизмах индивидуального и группового поведения.*

Статистические гипотезы, проверяемые в многомерных экспериментах, — это гипотезы о различиях между значениями зависимой переменной для разных уровней независимой переменной. Нулевая гипотеза формулируется как гипотеза о том, что разброс индивидуальных значений внутри одного уровня независимой переменной (внутри соответствующей экспериментальной группы) идентичен разбросу индивидуальных значений между различными уровнями (группами), то есть отношение дисперсии межгрупповых оценок к дисперсии внутригрупповых оценок равно 1. Последнее отношение обозначается как F -критерий. Для того чтобы определить, не превышает ли полученная в конкретном эксперименте величина F пороговое значение статистического F -распределения для заданного уровня значимости, используют статистическую технику *однофакторного дисперсионного анализа*. Термин «однофакторный» в данном случае означает, что в эксперименте использовалась лишь *одна* независимая переменная (фактор воздействия). Рассмотрение техники дисперсионного анализа и статистического оценивания получаемой в результате величины F выходит за пределы данного обзора (детальные описания и рекомендации при необходимости можно найти в книгах из списка дополнительной литературы к главе).

В тех областях социологии и социальной психологии, которые имеют сравнительно развитую традицию экспериментальных исследований (межличностное и межгрупповое восприятие, исследования

динамики установок, социальные процессы в малых группах, оценивание эффективности образовательных программ и т. д.), часто используют более сложные схемы экспериментирования, объединяемые термином «факторные эксперименты».

Факторный экспериментальный план включает в себя две и более независимые переменные (именуемые также «факторами»), каждая из которых имеет несколько уровней воздействия. Так как при увеличении числа независимых переменных очень быстро возрастает число групп, в каждой из которых применяется одна из возможных комбинаций этих переменных и их уровней²² (в полном факторном плане число групп равно произведению числа уровней, задаваемых для каждой независимой переменной), в целях экономии ресурсов и рационального распределения исследовательских усилий были разработаны многочисленные планы, где каждый из «уровней» переменных реализуется один раз, а обобщение и статистический анализ взаимодействия различных факторов и их изолированного и совместного влияния на зависимую переменную проводится на групповом уровне²³.

Всякий факторный эксперимент — это, в сущности, несколько экспериментов, объединенных в одном плане. Обобщенные данные факторного эксперимента позволяют ответить на два типа вопросов: 1) имеется ли эффект воздействия для каждой отдельно взятой независимой переменной; 2) зависит ли величина этого эффекта воздействия от величины значений других независимых переменных? Изолированный эффект воздействия одной независимой переменной называют *главным эффектом*, а изменение величины этого эффекта под влиянием другой независимой переменной называют *взаимодействием*.

²² Такие комбинации называют еще «обработками». Источник последнего термина — сельскохозяйственные опыты, для которых Р. Фишер разработал первые факторные планы, сочетавшие различные способы ухода за растениями, условия освещенности, типы почвы и режимы полива.

²³ Многочисленные примеры таких планов и описания соответствующих методов анализа результатов см. в: Дружинин Н. К. Выборочное наблюдение и эксперимент. М.: Статистика, 1977; Готтсданкер Р. Основы психологического эксперимента. М.: Изд-во МГУ, 1982; Вознесенский В. А. Статистические методы планирования эксперимента в технико-экономических исследованиях. 2-е изд., испр. и доп. М.: Финансы и статистика, 1981. Гл. 2, 3.

Т а б л и ц а 4.2

Факторный эксперимент 2×2

Новизна изображения	Тип изображения		Средняя длительность разглядывания, сек
	лицо	геометрическое	
Новое	60	50	55
Старое	30	10	20
Средняя длительность, сек	45	30	

В таблице 4.2. представлен план простейшего факторного эксперимента «два на два» (2×2), в котором изучалось влияние новизны и типа изображения на интерес, проявляемый к этому изображению четырехмесячными младенцами. В качестве индикатора интереса использовалась длительность разглядывания. Каждая из независимых переменных была представлена только двумя уровнями: для новизны — новое или старое, предъядлявшееся в предыдущих сериях изображение; для типа изображения — геометрический контур либо схематическое изображение человеческого лица (схематические рисунки использовались для уравнивания изображений по визуальной сложности, так как время фиксации взгляд обычно зависит от сложности и количества деталей). Как видно из приведенных в таблице 4.2 данных, налицо оба главных эффекта. Влияние новизны на интерес становится очевидным при сравнении средних по строкам — средняя длительность разглядывания изображений (и геометрических, и «физиономий») заметно выше в случае предъядления новых рисунков (55 сек против 20). Сравнение по столбцам показывает, что при усреднении данных по двум группам (новые и старые рисунки) изображения человеческого лица вызывают значительно больший интерес, проявляющийся в более длительном разглядывании (45 сек). Налицо также *взаимодействие* между типом изображения и новизной. Результаты предъядления разных типов изображений различны для «старой» и «новой» группы. Различаются и значения разностей по столбцам для каждой строки ($60 - 50 = 10$ сравнительно с $30 - 10 = 20$), и соответствующие показатели по строкам ($60 - 30 = 30$ сравнительно с $50 - 10 = 40$). Иными словами, большая привлекательность

человеческих лиц сильнее проявляется при предъявлении старых рисунков (различие в 10 сек при предъявлении новых картинок увеличивается до 20 для старых изображений), а различие между предъявлением старых и новых рисунков при использовании геометрических контуров возрастало до 40 сек.

При обработке результатов многофакторных экспериментов основной статистической моделью является *многофакторный дисперсионный анализ*.

Многофакторные эксперименты в социологии — это очень часто полевые эксперименты, моделирующие сложные взаимосвязи реального мира. Преимущество полевых многофакторных экспериментов — в их «жизнеподобии», то есть внешней, лицевой валидности. Но здесь же кроется и главный недостаток таких экспериментов — более низкие надежность и внутренняя валидность. Критики полевых многофакторных экспериментов часто отмечают, что приближение эксперимента к реальному миру здесь нередко достигается за счет замены экспериментального контроля чисто статистическим. В последнем случае возрастают угрозы валидности, связанные с неправильной спецификацией модели измерения, с «закоррелированностью» отдельных уровней независимых переменных с неконтролируемыми внешними переменными (см. гл. 5, 6). Кроме того, в многофакторных экспериментах острее, чем в индивидуальных и межгрупповых, стоит проблема агрегирования данных — практически всегда существует вероятность того, что отношения, выявленные при анализе сводных групповых данных, в точности не соблюдаются ни для одного отдельно взятого испытуемого (так же, как среднее некоторой выборки может не относиться ни к одному конкретному выборочному наблюдению). К неоспоримым достоинствам факторных экспериментов следует отнести значительно большие возможности статистического анализа, в том числе анализа различных эффектов взаимодействия переменных-«факторов».

В социальных науках часто употребляют также понятие *квазиэксперимента*, или *квазиэкспериментального исследовательского плана*. Речь идет о панельных, трендовых и т. п. планах выборочных обследований (гл. 5). Выборочные обследования, особенно продолжающиеся или проводимые как сравнительные «срезовые»

исследования для подвыборок, испытавших либо не испытавших определенное, локализованное во времени воздействие (например, социальную революцию, реформу образования или крах фондового рынка), действительно позволяют делать выводы о взаимоотношениях между интересующими исследователя независимыми и зависимыми переменными, а значит — проверять гипотезы о предполагаемых причинно-следственных связях, однако экспериментальную рандомизацию и контроль в выборочных исследованиях, как показано в соответствующих главах, здесь заменяет использование случайных выборок и специальных методов статистического анализа данных.

Дополнительная литература

Вознесенский В. А. Статистические методы планирования эксперимента в технико-экономических исследованиях. 2-е изд., перераб. и доп. М.: Финансы и статистика, 1981. Гл. 2, 3.

Гласс Дж., Стэнли Дж. Статистические методы в педагогике и психологии: Пер. с англ. / Общ. ред. Ю. П. Адлера. М.: Прогресс, 1976. Гл. 15–19.

Дружинин Н. К. Выборочное наблюдение и эксперимент. М.: Статистика, 1977.

Кэмбелл Д. Модели экспериментов в социальной психологии и прикладных исследованиях: Пер. с англ. / Сост. и общ. ред. М. И. Бобневой. М.: Прогресс, 1980.

Готтсданкер Р. Основы психологического эксперимента: Пер. с англ. М.: Изд-во МГУ, 1982.

Ядов В. А. Социологическое исследование: методология, программа, методы. М.: Наука, 1987. Гл. 5.

ГЛАВА 5. МАССОВЫЕ ОПРОСЫ В СОЦИОЛОГИИ

Определение и истоки

Метод опроса — самый распространенный из социологических методов, определяющий «образ» социологии в глазах непосвященных и к тому же имеющий самую богатую и давнюю историю. Утверждение о том, что почти невозможно дать строгое и исчерпывающее определение того, что такое опрос, на первый взгляд кажется нелепостью. Однако в действительности представления о том, каким должен быть хороший социологический опрос, менялись так часто, что любая попытка свести определение опроса к конкретной технике сбора информации, плану исследования, типу анализа данных или характеру использования полученных сведений наверняка столкнется с трудностями. Трудности эти так существенны, что один известнейший специалист в этой области в монографии, посвященной анализу истории и перспектив опросного метода, предложил говорить о некотором «базовом типе» опроса, по отношению к которому можно было бы упорядочить все многообразие реальных опросных исследований¹. Идеальной моделью он предложил считать «модель Гэллапа», то есть тот тип опроса общественного мнения, который сложился в 1930–40-х гг. в результате сотрудничества (и конкуренции) между основанным Дж. Гэллапом в 1935 году Американским институтом общественного мнения и другими исследовательскими фирмами. Для типичного «гэллаповского» опроса характерны следующие признаки:

- 1) общенациональный характер;
- 2) отбор из генеральной совокупности всех лиц, достигших избирательного возраста;

¹ *Miller W. L. The Survey Method in the Social and Political Science: Achievements, Failures, Prospects. L.: Frances Printer Publ., 1983. Part 1.*

- 3) максимальная приближенность времени проведения опроса ко времени выборов или референдумов;
- 4) среднее число респондентов в выборке — 2000 человек;
- 5) случайный или квотный характер выборки;
- 6) использование стандартных вопросников и личное интервьюирование каждого респондента по месту жительства;
- 7) «закрытый» характер вопросов;
- 8) сбор индивидуальных, неагрегированных данных (каждое наблюдение может быть соотнесено с конкретным индивидуумом в выборке)².

Широко распространенные отклонения от описанной «гэлловской» нормы все же столь существенны, что нам следует рассмотреть и другие подходы к определению сути опросного метода. Во-первых, следует вспомнить о том, что для социологии как науки главной функцией опроса является все же не предсказание результатов завтрашних выборов, а проверка гипотез о характере связей между различными переменными. (Переменная-признак задается как операционализация неких содержательных представлений о существенном для социологической теории качестве, свойстве: «социально-экономическом статусе», «отчуждении», «расовой сегрегации» и т. п.) Во-вторых, использование выборочного обследования, как говорится в гл. 7 и 8, как раз и имеет основной целью либо оценку значения определенного параметра в совокупности, либо — в большинстве случаев — проверку статистической гипотезы о связи между переменными. Эксперимент — это идеальная модель исследовательского плана для анализа причинных связей. Выборочное обследование (опрос) — хорошее приближение к идеальной модели. Для идеального эксперимента, напомним, характерны:

- 1) контроль условий, то есть возможность варьирования независимых переменных и измерения зависимых;
- 2) использование экспериментальной и контрольной групп для проведения повторных сравнений;
- 3) рандомизация, то есть случайный отбор испытуемых в контрольную и экспериментальную группы.

² Ibid. P. 6–7.

В выборочном исследовании, строго говоря, отсутствует возможность контроля, так как исследователь лишен возможности манипулировать независимыми переменными, произвольно задавать их значение. Однако с помощью *количественных методов измерения* и *статистического анализа связи* между переменными выборочный опрос может максимально приблизиться к той модели причинного вывода, которая лежит в основе экспериментального метода.

В целом анализ связи между переменными — и экспериментальный, и сугубо статистический, основанный на опросных данных, — подразумевает перекрестную группировку данных по двум переменным (независимой и зависимой), обнаружение связи между ними и введение третьей, контрольной переменной для оценки ее влияния на изучаемую связь. (Кстати, те возможности для контроля влияния «посторонних» факторов на исследуемую взаимосвязь, которые возникают при анализе связи в выборочных обследованиях, обычно даже превосходят возможности эксперимента.) В последнем случае набор контрольных переменных, «изолируемых» с помощью эксперимента, обычно ограничен. В выборочном обследовании список переменных чаще всего значительно обширнее и к тому же включает в себя такие переменные, которые в принципе не могут использоваться в эксперименте из практических или этических соображений: нельзя, например, произвольно назначить испытуемому экспериментальное условие «родился чернокожим» или «часто подвергался жестокому обращению». Однако заметим сразу, что последнее обстоятельство все чаще используется не столько для восхваления, сколько для критики — во многих отношениях справедливой — применимости выборочных опросов для анализа причинных связей (о чем еще будет сказано ниже).

Случайный отбор, используемый на том или ином этапе как основа построения выборки для массового опроса может рассматриваться как подобие рандомизации в эксперименте. В идеальном случае, почти не встречающемся на практике, любая единица генеральной совокупности имеет равные шансы попасть в выборку. Поэтому влияние внешних, «посторонних» факторов нейтрализуется, и систематическое смещение отсутствует. В реальности, как показано в обсуждении выборочного метода, мы редко можем реа-

лизовать простую вероятностную выборку, довольствуясь каким-то приемлемым и экономичным компромиссом между случайным отбором, стратификацией и квотированием.

Контрольная и экспериментальная группы, используемые в экспериментальных планах для сравнения и выявления эффекта некоего причинного фактора, «отбираются» в выборочных обследованиях на стадии анализа, апостериорно. Фактически они «конструируются» исследователем *ad hoc* в ходе сравнения подвыборок, выделенных с помощью фиксации разных уровней одной (или нескольких) объяснительных переменных.

В целом опросные методы обладают рядом существенных достоинств:

1) позволяют достаточно быстро получить большой массив наблюдений, причем каждый индивидуальный «случай» (отдельное наблюдение) описывается с помощью целого набора теоретически релевантных переменных-признаков;

2) стоимость выборочного опроса оказывается сравнительно небольшой, если принять во внимание объем получаемой информации;

3) использование стандартных опросных процедур и однородных количественных показателей при соблюдении определенных условий позволяет не только проверять гипотезы о причинных зависимостях, но и проводить вторичный и сравнительный анализ результатов.

Недостатки, также присущие этому методу, мы анализируем в следующих разделах.

Выбор исследовательского плана

Даже в том случае, когда исследователь четко осознал, в чем заключаются содержательные вопросы, на которые он хочет получить ответ в ходе выборочного обследования, ему не стоит торопиться составлять анкету и нанимать интервьюеров. Прежде ему нужно поразмыслить над тем, какого рода логику анализа данных он собирается использовать, *после того* как эмпирические данные будут получены. Для того чтобы сведения о людях, группах или обществах (об их поведении, установках или других чертах) можно

было рассматривать в качестве доказательства каких-то теоретических гипотез, следует сначала решить, *что* именно можно считать доказательством в данном случае, по каким *правилам* будут строиться логические сопоставления и статистические выводы, иными словами, необходимо выбрать принципиальный *исследовательский план*.

В гл. 4 довольно подробно говорится о том, как различия в логике и целях анализа влияют на выбор плана эксперимента. В планировании выборочного опроса исследователи исходят приблизительно из тех же соображений: сравнение «случаев», подгрупп, сравнение типа «до–после». Здесь мы рассмотрим лишь самые общие типы исследовательских планов, используемых в выборочных опросах (другие проблемы планирования детально анализируются в гл. 7, посвященной построению выборки).

Первый шаг в планировании опроса — это принятие решения о том, что считать *единицей анализа*. В простейшем случае мы стремимся приписать каждому индивиду (респонденту) определенное значение по каждой переменной. Предположим, наша цель заключается в том, чтобы на основании опроса 2000 респондентов узнать, как распределены в генеральной совокупности «партийная принадлежность», «судимость» и некоторые другие переменные, а кроме того, мы собираемся проанализировать связь этих переменных с полом, возрастом и семейным статусом. Некоторые из переменных будут строго количественными, другие будут описываться как качественные признаки. В любом случае нам нужно будет охарактеризовать каждого респондента по каждой переменной. В результате мы сможем построить *структурированную матрицу данных*, подобную той, что изображена в таблице 5.1. В столбцах этой матрицы содержится вся информация о респондентах, которые здесь и являются *единицами анализа (или «случаями»)*. Именно их свойства нам предстоит оценивать, сравнивать в поисках взаимосвязей и т. п.

Обычно единицами анализа, то есть теми, *кого* исследуют, бывают именно люди. *Однако единицами анализа могут быть и семьи, и организации, и регионы, и государства*. Например, в матрице данных столбцы могли бы соответствовать городам, а строки — переменным типа «уровень преступности», «население», «число

Таблица 5.1

Пример матрицы данных типа «респондент × переменные»

Переменная	«Случай»			
	1-й респондент	2-й респондент	...	2000-й респондент
Пол	Мужской	Женский	...	Мужской
Возраст	38 лет	23 года	...	62 года
Семейный статус	Разведен	Замужем	...	Вдовец
Судимость	Отсутствует	Отсутствует	...	2 судимости
Партийная принадлежность	Конституционный демократ	Беспартийная	...	Христианский социалист

безработных» и т. п. Некоторые из переменных были бы получены путем агрегирования, «объединения», индивидуальных данных (например, о наличии дополнительных источников дохода), другие характеризовали бы город как целое (наличие аэропортов, доля прямых налоговых поступлений в бюджете). *В любом случае исследователю нужно заранее представить себе, как будет выглядеть матрица данных и какие приемы анализа он собирается к ней применить.*

Любое конкретное исследование может предполагать и использование различных единиц анализа, то есть полученная в нем эмпирическая информация может характеризовать и отдельных индивидов, и семьи, и — в результате использования агрегированных показателей — регионы или государства. Важно лишь, чтобы все единицы анализа, которые вы намерены использовать, были определены заранее. В ином случае в матрице данных «единица анализа × переменная» неизбежно возникнут пропуски или дублирование одной и той же информации. Так как количество матриц данных равно количеству предполагаемых единиц анализа (хотя размерность их будет разной³), можно заранее создать соответствующее количество

³ Например, размерность матрицы «респонденты × переменные» может быть 2000 (респондентов) × 32 (переменных), а размерность матрицы «города × переменные» — 6 (городов) × 4 (агрегированных показателей).

отдельных массивов данных (файлов), содержащих те данные, которые относятся к данной единице анализа. Скажем, сведения о возрасте попадут в массив «респонденты», а сведения о составе семьи — в массив «семьи» (даже если последние и были получены в результате беседы с одним из членов семьи).

Описанная выше двумерная матрица данных типична для одномоментного, *«срезового»* исследования, характеризующего ситуацию в момент опроса. Целью такого исследования может быть, во-первых, описание распределения каких-то переменных в совокупности. Например, мы можем узнать, сколько человек собираются проголосовать за демократов при условии, что выборы будут проведены тотчас же (типичный «гэллаповский» опрос). Во-вторых, мы можем попытаться использовать «срезовые» данные для характеристики отдельных подвыборок — например, «работающих пенсионеров», «высококвалифицированных рабочих в возрасте от 30 до 45 лет» и т. п. Далее, применяя различные методы статистического анализа, можно проверить какие-то гипотезы о взаимосвязи переменных (в данный момент времени). В последнем случае исследование становится *объяснительным*. Однако даже в чисто описательном исследовании мы столкнемся с необходимостью каких-то сравнений, делающих полученные нами оценки осмысленными. Если, например, мы узнаем, что 15% подростков читают медицинские журналы не реже 1 раза в месяц, то для того, чтобы понять, много это или мало, нам нужно будет с чем-то сопоставить этот показатель. Скажем, мы можем сравнить подростков 1994 года с подростками 1954 года. (Конечно, нам предварительно придется найти данные соответствующего опроса 40-летней давности.)

Изменения во времени подвержены не только отдельные показатели, но и взаимоотношения между переменными. Так, глобальные социально-экономические изменения — экономический кризис, сдвиг в социально-классовой структуре — могут привести к тому, что высокая зависимость дохода от продолжительности образования станет незначимой. Следовательно, изучение сложного причинного механизма воздействия образовательного уровня на доходы требует какой-то серии разделенных во времени обследований, позволяющих проследить динамику интересующего нас отношения под влиянием существенных внешних переменных.

Исследовательские планы, позволяющие анализировать данные во временной перспективе, называются лонгитюдными. Данные получают многократно, в разные моменты времени, причем цели исследования могут быть сугубо дескриптивными (доля голосующих за коммунистов, распределение положительных и отрицательных установок по отношению к «мыльным операм») и объяснительными.

Принято выделять основные виды лонгитюдных планов, каждый из которых имеет множество модификаций и «переходных» форм. Это *трендовые, когортные и панельные* исследования.

Трендовые обследования ближе всего к уже описанным однократным, «срезовым» опросам. Некоторые авторы даже предлагают обозначать их просто как *регулярные опросы*, то есть проводимые через более или менее равные промежутки времени⁴. В *трендовом опросе одна и та же генеральная совокупность изучается в разные моменты времени, причем каждый раз выборка строится заново. Иными словами, анализируются последовательные выборки из одной и той же совокупности.* Например, опрос Института Гэллапа, проводимый ежемесячно в ходе избирательной кампании, является трендовым обследованием, показывающим динамику установок населения по отношению к кандидатам или партиям. Строго говоря, если количество тех, кто собирается голосовать за кандидата *X*, за месяц увеличилось на 16%, мы можем лишь зафиксировать изменение картины предпочтений избирателей, но не можем наверняка утверждать, что определенная группа избирателей изменила свои предпочтения, так как в двух последовательных опросах мы имеем дело с разными респондентами. Преимуществом оперативных трендовых исследований является возможность «привязки» наблюдаемых изменений к текущим событиям — политическим скандалам, решениям правительственных органов, изменениям в финансово-экономической ситуации, — что облегчает их интерпретацию.

Однако, например, ежегодные исследования занятости и безработицы, проводимые по этому плану, могут привести к трудноин-

⁴ См.: *Hakim C. Research Design: Strategies and Choices in the Design of Social Research.* L.: Allen & Unwin, 1987. P. 76–77.

терпретируемым результатам. Если в результате двух таких исследований окажется, что социально-демографические характеристики людей, получающих пособие, почти не изменились, будет большой неосторожностью утверждать, что существует какая-то «типичная» группа людей, постоянно живущая на средства налогоплательщиков. Вполне вероятно, что большинство респондентов, охваченных первым опросом, уже нашли работу.

В качестве особого исследовательского плана иногда рассматривают *когортные исследования*. Основания для выделения этого плана несколько условны и связаны скорее с теоретической логикой интерпретации (а не сбора) данных. Если в трендовых исследованиях отбор каждый раз производится из общей совокупности — всех избирателей, всех семей и т. п., — то, исследуя «когорты» (от лат. *cohors* (*cogortis*) — подразделение, видовая группа), мы каждый раз производим отбор из *одной специфической совокупности*, стремясь проследить перемены в ее поведении, установках и т. п. Пусть, например, мы изучали ценностные ориентации десятиклассников в 1985 году, а в 1995 году нам захотелось снова опросить бывших десятиклассников, так как мы предполагаем, что их ценностные ориентации изменились с переходом в иную стадию жизненного цикла (создание собственной семьи, формирование профессиональной идентичности и т. п.). В этом случае мы будем работать с новой выборкой из прежней специфической совокупности, сравнивая представителей одной и той же «когорты» с десятилетним интервалом, а не десятиклассников 1985 года с десятиклассниками 1995-го (в последнем случае можно было бы говорить о трендовом исследовании десятиклассников).

Самым совершенным воплощением идеи введения временной перспективы в исследовательский план является *панельное исследование*. Если вернуться к нашей структурированной матрице данных (см. табл. 5.1), то можно сказать, что панель — это прибавление к двумерной матрице еще одного измерения, превращающего ее в пределе в некий «параллелепипед» данных. Панельные исследования позволяют не только зафиксировать какие-то социальные изменения в установках, поведении и т. п., но и выявить причины и последствия этих изменений на микроуровне, то есть на уровне отдельных индивидов. Если трендовое исследование показывает,

что десятая часть потребителей, предпочитавших отечественные макароны, «переметнулась» к поклонникам спагетти, мы не можем точно определить, кто из респондентов изменил свои предпочтения и, следовательно, каковы общие характеристики «перебежчиков». Таким образом, мы лишены возможности проверить, какие объяснительные переменные позволяют предсказывать динамику предпочтений на микроуровне.

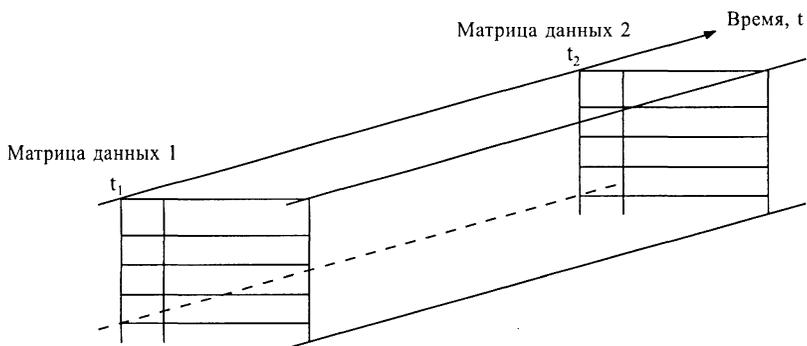


Рис. 5.1. Элементарный план панельного исследования
(два замера — две матрицы данных)

Панельный исследовательский план — это многократное обследование одной и той же выборки из генеральной совокупности в разные моменты времени. Эту многократно используемую выборку и называют *панелью*. Исследовательский план, использующий панель респондентов, — весьма дорогостоящее предприятие, требующее к тому же очень тщательной проработки всех деталей до начала опроса. В трендовом и когортном исследовании данные нередко сравниваются с данными других опросов, проводившихся ранее иными исследовательскими группами. Этот путь проще и дешевле, однако сравнимость результатов обследований, планировавшихся разными исследовательскими командами и — чаще всего — для разных целей, всегда проблематична. Возможность оценки «чистого эффекта» и величины наблюдаемых изменений — большое преимущество панельного плана. Однако эта возможность

прямо зависит от величины усилий, предпринятых социологами для сохранения неизменности самой панели и инструментов сбора данных. Если, например, в первой волне панели (*волной* обычно называют один полный цикл опроса панели, один «замер») социолог позабыл включить в список для ранжирования тяжести преступлений квартирные кражи, то использование дополненного списка во второй и третьей волнах не поправит дела: сопоставимость полученных в панели ранжировок будет ничуть не выше, чем в случае обычных «срезовых» обследований, при более высоких затратах. Поэтому панельные исследования чаще всего используют как очень точное средство проверки конкретных гипотез в отчетливо очерченной предметной области. Выбор панельного плана в случае пилотажных или поисковых исследований совершенно неоправдан.

Панельные исследования незаменимы в проверке причинных гипотез, особенно в тех случаях, когда отсутствует «естественный» критерий для разделения независимой и зависимой переменных во времени. Например, множество «срезовых» исследований может демонстрировать устойчивую высокую корреляцию между систематическим потреблением алкоголя и проявлениями социальной дезадаптации (развод, потеря статуса и т. п.), однако лишь длительное панельное исследование может дать необходимый материал для того, чтобы решить, ведет ли алкоголизм к дезадаптации либо, наоборот, является ее следствием. Панельные исследования незаменимы и для анализа более сложных причинных моделей с отсроченными эффектами (лагами), петлями «обратной связи» и т. п.

Основным преимуществом панельного плана с сугубо статистической точки зрения является возможность отделить реальные изменения показателей от разброса, связанного с ошибкой выборки.

В случае «непанельного» опроса (трендовый опрос, сравнение данных двух независимо проведенных «срезовых» опросов) какое-то различие между двумя последовательно опрошенными выборками, значимое на 5%-м уровне, скажем, различие между 49 и соответственно 54% предпочитающих «сильную руку» институтам представительной демократии с вероятностью, превосходящей 1 : 20, будет связано с выборочной ошибкой, а не с радикальными переменами в политической атмосфере. Аналогичные данные па-

нельного исследования позволяют говорить о наличии реальных изменений.

Однако не следует считать, что *любое* реальное изменение, фиксируемое в панельном опросе, также подлежит содержательной интерпретации с использованием теоретически «привлекательных» переменных. Во-первых, наша панель — это всего лишь выборка среди других возможных выборок (других возможных панелей). Используя панельный план, мы уменьшаем влияние ошибки выборки на значимость различий между двумя волнами, но не исключаем ошибку выборки полностью: результаты сравнений для второй (девятнадцатой, двадцатой...) панели могли оказаться иными. Далее, фиксируемые изменения могут быть связаны с низкой надежностью нашего измерительного инструмента, о чем мы еще будем говорить при обсуждении проблем измерения. Наконец, наша интерпретация результатов может оказаться необоснованной *из-за изменений в самой панели*. Полезно помнить, что панельное исследование по логике анализа результатов ближе всего стоит к простейшему экспериментальному плану типа «до–после». Малоприятным продолжением этого достоинства является подверженность панельного плана тем же угрозам систематического смещения (см. гл. 4). В частности, *эффекты «созревания»* участников панельного опроса были неоднократно продемонстрированы даже в таких нейтральных сферах, как изучение семейного бюджета или чтение газет и журналов. После двух-трех волн члены панели достоверно чаще фиксируют свои ежедневные расходы и начинают тратить на чтение газет на 10–15% больше времени. Перемены в политических установках и поведении обычно носят еще более драматический характер: политические симпатии смещаются к крайним полюсам, доля активно участвующих в выборах возрастает. Основным механизмом, отвечающим за этот эффект, является стремление индивидуума к когнитивному балансу, к поддержанию высокой степени согласованности между собственными высказываниями и действиями. Соответственно, описываемый тип смещения резче выражен в панелях с маленькими интервалами между волнами. По мере увеличения промежутков между последовательными опросами — по крайней мере, до 1–2 лет — эффект «созревания» уменьшается, так как все сильнее становится влияние

направленных в противоположную сторону эффектов «памяти» (вернее сказать, «забывания»): респонденты просто плохо помнят, что они говорили год или десять лет тому назад.

К сожалению, именно в тех случаях, когда панельный план социологического исследования более всего осуществим и его применение возможно и обоснованно — и с точки зрения логики анализа, и по реальным возможностям внеакадемического финансирования, — тактика увеличения интервалов между циклами панели может оказаться неосуществимой. Пример тому — предвыборные опросы, где интервалы бывают равны 1–2 неделям и редко превышают 1–1,5 месяца.

Так, интерпретация классического «Народного выбора», проведенного П. Лазарсфельдом и его соавторами исследования президентских выборов в США в 1940 году⁵, остается неоднозначной, хотя его основные результаты были много раз воспроизведены другими исследователями. Панель Лазарсфельда состояла из семи волн, разделенных месячным интервалом. Столь сложный план требовался для того, чтобы проследить, как меняются предпочтения американского электората в ходе выборной кампании и какие факторы влияют на изменение решений отдельных избирателей. Самым поразительным результатом исследования оказалось то, что почти половина опрошенных ни разу не меняла свои политические предпочтения на протяжении полугода. Вероятно, немалую роль в формировании столь обширной группы «непоколебимых» сыграли описанные эффекты «созревания» в результате участия в панели.

Самый серьезный и распространенный тип смещения связан, однако, с другой постоянной проблемой всех панельных планов — проблемой «выбывания» из панели (или, что звучит несколько мрачно, со «смертностью», или «истощением», панели). Истощение панели проявляется в увеличении неучастия и «не-ответов» респондентов от первой волны к последующим. Некоторые респонден-

⁵ См.: *Lazarsfeld P. F., Berelson B., Gaudet H. The People's Choice: How the Voter Makes Up His Mind in a Presidential Campaign.* N. Y.: Columbia University Press, 1944.

ты оказываются недоступными для контактов: они меняют место жительства, болеют, умирают. Другие участники панели просто отказываются от следующего интервью. В результате и репрезентативность панели, и эффективный объем параллелепипеда данных, то есть реальная возможность сравнивать ответы одного респондента в разные моменты времени, резко снижаются от волны к волне (хотя расходы на поддержание панели продолжают расти). Особенно неприятна ситуация, когда «вымирают» определенные социально-демографические группы респондентов, что приводит к непоправимым систематическим смещениям. Эта ситуация возникает не так уж редко. Исследователи, работающие в коммерческих опросных фирмах, неоднократно замечали, что в рыночных исследованиях и исследованиях аудитории газет и журналов самой высокой «смертностью» отличаются молодые участники панели, особенно учащиеся—юноши в возрасте 18–25 лет. Иногда даже увеличение платы за участие в панельном опросе с каждой последующей волной не влияет на выбывание (это должно служить слабым утешением академическим исследователям, лишенным возможности платить респондентам).

В больших общенациональных панелях, приближающихся к «микрореперсиям», для борьбы с выбыванием иногда используют *метод самовосстановления*, особенно в случаях, когда выборочной единицей является семья, домовладение, организация и т. п. Например, в проводимом с середины 1960-х гг. Мичиганским университетом исследовании бюджета американских семей (PSID) ежегодно опрашивается более 5000 семей. Каждый отделившийся член семьи (например, взрослый сын, решивший жить отдельно от родителей) остается в выборке в качестве новой единицы наблюдения, так что выборка остается репрезентативной по типам семей, возрасту их членов и т. п. Выбывание из этой панели за первые десять лет составило 28% исходной выборки (кстати, это совсем немного для панельного опроса), однако за счет самовозобновления, то есть включения в выборку «отселившихся» членов семей, абсолютный размер панели за это же время даже вырос с 5000 до 5860 семей⁶.

⁶ Hakim C. Op. cit. P. 91–92.

Очевидно, что панельные исследования — очень сложное, хотя и эффективное средство проверки социологических гипотез. Вышеприведенные соображения вполне объясняют, почему панельный план используется реже других типов исследовательского плана. Панельный план практически доступен лишь для достаточно крупных исследовательских организаций и требует привлечения значительных материальных и финансовых ресурсов⁷, однако он абсолютно незаменим при исследовании социальных эффектов исторических изменений, сложных причинных моделей индивидуального выбора, процессов социализации и т. п. Многие социологи полагают, что оптимальным решением является использование *комбинированных исследовательских планов*, сочетающих в себе некоторые черты «срезовых», трендовых и панельных опросов. Самый простой из таких планов — *это ретроспективное панельное исследование*, когда опрос проводится однократно, однако включает большое количество вопросов о прошлом респондента. Например, в исследованиях профессиональной мобильности респондентов спрашивают о деталях их карьеры, периодах безработицы, причинах изменения места работы и т. п. Реконструированные таким образом «профессиональные биографии» анализируют так, как если бы они были получены в лонгитюдном обследовании. Возникающие здесь проблемы связаны, в первую очередь, с субъективными погрешностями припоминания, с изменением точки зрения на события прошлого, иногда — с намеренным искажением информации. Так, использование ретроспективного плана в изучении зависимости социально-экономического статуса от образования может вести к неверным выводам: доказано, что большинство людей имеет склонность задним числом «завышать» свои успехи в обучении. Однако этот тип плана может оказаться достаточно эффективным, например, при сравнительном изучении динамики занятости замужних и незамужних женщин. Основное достоинство ретроспективного плана — радикальное решение проблемы выбывания.

⁷ Достаточно сказать, что проведение панели требует постоянного отслеживания адресов участников, поддержания контактов с ними. С этой целью используют и поздравительные открытки, и рассылку отчетов, и даже местные собрания респондентов.

Более сложные типы комбинированных планов используют в микропереписях, общенациональных обследованиях занятости и безработицы, преступности и т. п. Очень эффективны *циклические планы с замещением*, где в каждой последующей волне какая-то доля исходной выборки «отдыхает», будучи замещенной новой эквивалентной подвыборкой. Скажем, если в ежегодном опросе треть панели каждый раз замещается, то каждая из исходных «третей» будет опрошена от одного до трех раз, прежде чем состав участников полностью обновится. «Поперечный» и «продольный» анализ позволит и учесть эффекты участия (при сравнении результатов «кратковременных» и «длительных» респондентов), и дать текущую картину распределения переменных по социальным группам, и зафиксировать резкие изменения. Иногда часть вопросов предъявляется лишь сравнительно небольшой подвыборке, имеющей характеристики «фокусной» группы (например, только матерям-одиночкам, получающим социальные пособия), что позволяет проанализировать динамику поведения и мнений «труднодоступных» популяций. Нередко общую базу данных поддерживает и анализирует одна исследовательская группа, а для анализа «периферийных» тем и специфических подвыборок привлекаются эксперты из других институций. Объективная логика развития регулярных опросов, основанных на комбинированных исследовательских планах, явно ведет к созданию междисциплинарных, многоцелевых проектов и баз данных, имеющих множество источников финансирования (таковы, например, некоторые общенациональные лонгитюдные исследования преступности, здоровья населения). Соответственно, все выше ценятся услуги методологов, специализирующихся в планировании исследований, стандартизации показателей, социологическом измерении.

Концептуализация и измерение: общий обзор

Избрав определенный исследовательский план, социолог может сказать, *что* он будет рассматривать в качестве «случаев» в структурированной матрице данных (см. табл. 5.1) и какой будет логика сравнений между случаями на стадии анализа. Теперь ему предстоит решить, какими будут его *исследовательские переменные* —

строки матрицы данных — и как будет осуществлен *переход от теоретического понятия к измеряемому показателю*. Решение этих двух взаимосвязанных проблем — *концептуализации и измерения* — необходимое условие перехода к разработке анкеты, плана интервью, схемы эксперимента и к сбору данных. Отметим сразу, что речь идет лишь о предварительном решении, так как многие исследовательские задачи, связанные с измерением и истолкованием теоретических конструктов, возникают позднее, на стадии анализа данных (и будут рассмотрены нами в соответствующих разделах).

Понятия социологической теории — скажем, «отчуждение», «культура бедности», «социальный статус» или «коронарный тип личности»⁸ — используются в качестве элементов для построения неких *теоретических моделей*, описывающих отношения между понятиями. Предположения о характере таких отношений — это собственно *исследовательские гипотезы*.

Сложная структура социологических теорий не позволяет говорить о простой и однозначной их проверяемости. Как говорилось ранее (см. гл. 1), правдоподобие гипотез оценивается не в каком-то абсолютном смысле, а лишь относительно целой совокупности других вспомогательных гипотез, явно или неявно связанных с теоретическим «ядром». По этой причине сколько-нибудь разработанные теоретические модели оказываются довольно сложными, и их предварительное описание (спецификация) — это необходимое условие любой эмпирической проверки.

Эмпирическое «истолкование» теоретических понятий в качестве переменных в матрице данных (их *концептуализация*) и перевод этих понятий на язык наблюдаемых признаков, то есть измерение, могут оказаться довольно сложными процедурами, в чем-то сходными с процедурами построения теоретической модели. На первый взгляд, некоторые типы переменных не создают вовсе никаких проблем для измерения, так как они очень близки к тем способам категоризации, которые мы употребляем в повседневной жизни (например, пол, возраст). Другие же, более абстрактные

⁸ То есть тип личности, особенно подверженный сердечно-сосудистым заболеваниям.

теоретические конструкты — отчуждение, социально-экономический статус или расовая сегрегация, — явно требуют большего, чем формулировка одного показателя или одного вопроса анкеты. Ясно, что уточнение теоретического понятия и поиск соответствующих индикаторов в этом случае может быть только результатом специальной аналитической работы. Конечная цель такой работы — создание *модели измерения*, в которой будут определены (специфицированы) все предполагаемые связи между теоретическим конструктом (понятием) и теми эмпирическими показателями, которые мы намерены использовать для его измерения. В этой модели нам придется также сделать некоторые предположения о возможных *ошибках измерения* (их случайном или систематическом характере). Ведь в действительности даже сравнительно простые и очевидные показатели, фиксируемые с помощью одного стандартного вопроса, могут быть подвержены влиянию не только случайных ошибок, связанных с невнимательностью или погрешностями выборочной процедуры. Может быть, например, незамужние женщины склонны систематически занижать свой возраст? Если обратиться к «случаю Агнессы», описанному в гл. 3 (см. с. 72–74), можно увидеть, что даже биологический пол в некоторых случаях трудно определить однозначно. Пример столь простого признака, как «пол», позволяет увидеть и другую сторону проблемы: прежде чем искать подходящий показатель, нужно решить, как мы намерены интерпретировать соответствующее понятие в нашей теории. Если мы, к примеру, собираемся проверить гипотезу о влиянии половой идентичности на социальные достижения, то нам недостаточно просто разбить наших респондентов на «муж.» и «жен.»: внутригрупповой разброс показателей успешности наверняка окажется очень велик, и вся наша объяснительная схема «поплывет». В действительности нам лучше интерпретировать «половую идентичность» как некий континуум, плавный переход от одного жесткого полоролевого стандарта к другому, от крайней «маскулинности» к «фемининности». Используя соответствующие показатели и шкалы, мы, скорее всего, обнаружим, что большего социального успеха добиваются люди, не следующие жестким предписаниям традиционной половой роли.

Итак, первый шаг в поиске индикаторов для теоретических понятий — это прояснение самих понятий. Теоретические переменные,

в отличие от платоновских идей, не существуют «сами по себе», ожидая, когда мы наткнемся на них. Они не имеют какого-то абсолютного, раз и навсегда определенного значения. Их значение определяется контекстом употребления, концептуальной схемой, которую мы используем. Например, если мы используем «религиозность» как понятие, характеризующее роль некой конфессии в политическом укладе национального государства, наибольший интерес для нашего исследования могут представлять агрегированные (то есть относящиеся к надындивидуальному уровню) переменные, показывающие роль церкви в поддержании нормативной системы общества, в принятии политических решений. Показателями здесь могут быть количество церковных приходов, наличие обязательных уроков закона Божьего в государственных школах, участие церковных иерархов в работе законодательной власти и т. п. Если целью нашего анализа является индивидуальная «религиозность», то нас скорее заинтересует широкий спектр поведения и установок от институциональной религиозности, связанной с участием в церковных обрядах, верой в спасение души и т. п., до расплывчатой убежденности в том, что «существуют некие сверхъестественные силы», или даже до устойчивого интереса к астрологическим прогнозам.

Конечно, мы можем заключить, что отсутствие абсолютного самоочевидного смысла в теоретических понятиях дает нам полную свободу в их определении. Но такое заключение будет ошибочным. Во-первых, теоретическое понятие, неповторимый смысл которого известен только самому теоретику, обладает всеми достоинствами, кроме одного — оно больше не может служить средством коммуникации. Дабы этого не случилось, лучше всего давать определения, понятные не только вам, но и другим: все же наука — это коллективное предприятие. Более того, нужно соотносить собственные определения понятий с теми, которые использовались вашими предшественниками, в том числе и теми, чьи теоретические взгляды противоположны вашим. Ценность теоретического понятия — в его включенности в более широкую сеть теоретических представлений, во множестве связей с другими понятиями. Попытки начать с «нулевой ступени» ни к чему хорошему не ведут. Даже если вам отвратителен марксизм как идеология тоталитаризма, не-

возможно сказать что-то содержательное и интересное о таких вещах, как «классы» или «отчуждение» без учета того, что сказал о них Маркс.

В работе по уточнению теоретических понятий можно выделить три стадии. *На первой стадии* нужно составить по возможности полный список существующих определений интересующего нас понятия. Основной путь здесь — анализ литературы. Часто приходится анализировать и те смыслы, которые придаются какому-то понятию в обыденной речи: понятия повседневного языка редко обладают достаточной степенью формальной строгости, но их многозначность иногда позволяет выразить неочевидный и нетривиальный взгляд на вещи. Например, анализ того контекста, в котором употребляются понятия «стресс» или «психологическая травма», открывает широкий диапазон жизненных событий — от развода до потери работы. Если мы изучаем влияние травмирующих жизненных событий на рост хронической заболеваемости, нам не обойтись без анализа субъективного смысла различных событий для разных людей. На этом этапе могут оказаться полезными неформальные глубинные интервью, групповая дискуссия, анализ доступных биографических материалов и т. п. В результате описанной работы по обобщению существующих определений (научных и обыденных) мы получаем возможность исходить из достаточно общего и разделяемого большинством исследователей определения. Так, в работе Дж. Хиллери⁹ перечислено 94 определения понятия «сообщество» (community), большая часть которых включает три основных признака: локальная область расселения; общие связи, основанные на чувстве идентичности с группой; социальное взаимодействие.

На *второй стадии* мы осуществляем и обосновываем свой выбор трактовки понятия. Обоснование необходимо и в том случае, если мы решили использовать общепринятое определение, и тогда, когда нами предложено нечто абсолютно новое. Позднее, в ходе анализа данных, наша теоретическая модель, скорее всего, будет уточняться, но и в сборе, и в анализе данных мы будем

⁹ См.: Hillery G. A. Communal Organizations. Chicago: Chicago University Press, 1955.

руководствоваться принятым рабочим определением. Так, если мы решим, что социологический смысл понятия «профессия» заключается в способе регуляции рыночных условий в пользу определенной группы, ограничивающей и контролирующей доступ новых членов в свои ряды, мы, скорее всего, сосредоточим свое внимание на таких аспектах профессионализма, как автономия, контроль над процессом определения «внештатных» ситуаций и приписыванием ответственности, обучение новичков и управление «публичным» образом профессиональной группы. При этом мы, возможно, не уделим того же внимания таким аспектам профессионализма, как отношения с потребителями товаров или услуг, контроль над определенными ресурсами и т. п.

Большинство полученных нами определений будут многомерными, то есть они будут включать в себя более одного аспекта или измерения. Поэтому на *третьей стадии* следует отчетливо очертить существующие аспекты понятия и, возможно, выбрать те из них, с которыми мы собираемся работать. Во-первых, выделение отдельных измерений в многомерном теоретическом понятии необходимо для того, чтобы найти соответствующие индикаторы для каждого из измерений. Во-вторых, в социологии мы часто используем *категориальные переменные*, состоящие из множества взаимосвязанных признаков, то есть двух, трех или более качественных категорий. Примерами здесь могут служить пол, профессия, семейный статус, религиозная конфессия и т. д. Нередко признаки, составляющие категориальную переменную, могут быть упорядочены по какой-то ординальной шкале. Скажем, социальный статус может быть низким, средним или высоким. Анализ размерности теоретического понятия, представляемого с помощью такой категориальной переменной, позволяет выявить различия между упорядочениями категорий по разным измерениям. Упорядочение религиозных конфессий по престижности будет отличаться от их упорядочения по степени религиозного фундаментализма. Сделав явным это различие между смысловыми измерениями теоретического понятия, мы обезопасим себя от ошибочных выводов о характере взаимосвязей данной переменной с другими, то есть от ошибок на стадии анализа данных.

Прояснив теоретические понятия, используемые в нашем исследовании, мы переходим к следующей важной задаче — поиску *конкретных индикаторов* для этих понятий. Нередко эту стадию работы называют стадией *операционализации понятий* (о том, почему это обозначение является не вполне точным, будет сказано чуть ниже). Если, скажем, в исследовании профессиональной мобильности ученых мы используем понятие «престижность университета», нам необходимо решить, в чем, собственно, выражается престижность: в высоком проходном балле на вступительных экзаменах, в количестве ежегодно проводимых международных конференций, в среднем индексе цитирования для профессоров и преподавателей? Возможно, полезной для определения престижности будет экспертная процедура — например, престижность американских университетов определяется в ходе регулярных опросов ведущих специалистов в разных областях знания. *Под операционализацией, таким образом, понимают процесс связывания теоретического понятия с эмпирическими наблюдениями, где последние выступают индикаторами, показателями каких-то свойств, относящихся к данному понятию.* Предполагается, что, скажем, результаты оценивания респондентами престижности университетов показывают высокий или низкий престиж данного рода заведений приблизительно так же, как показания стрелки манометра показывают давление. Однако аналогия здесь весьма условна. Измерение в социологии обычно носит не прямой характер: *отдельный индикатор может отражать влияние более чем одной переменной, а каждая переменная может иметь множество индикаторов*, то есть операциональные определения теоретических понятий в социологии отличаются от таковых, скажем, в физике.

Многие эмпирические индикаторы могут рассматриваться как взаимозаменяемые. Идея *взаимозаменяемости индикаторов* была впервые проанализирована П. Ф. Лазарсфельдом. Так как измерение носит не прямой характер, ни один из существующих индикаторов не будет совершенным или безупречным. Хотя в определенной исследовательской ситуации можно указать причины, по которым один индикатор лучше другого, в сущности, они взаимозаменяемы. «Истинное значение» переменной — это какая-то функция значений показателя и ошибки измерения. Поэтому измерение значения

переменной и проверка гипотез о связях между индикаторами требуют использования множества показателей (более детальное обсуждение этого вопроса содержится в гл. 6). На практике социологи чаще всего используют несколько индикаторов для каждой существенной теоретической переменной, объединяя их на стадии анализа в некоторый суммарный показатель (индекс), или строя шкалу. То, как соотносятся индикаторы и теоретическая переменная, описывается с помощью *модели измерения*. В простейшем случае, когда все индикаторы (обозначаемые прописными латинскими буквами — X_1, X_2, X_3, X_4) являются следствиями, результатами действия *латентной*, то есть не наблюдаемой непосредственно переменной X , модель измерения будет выглядеть, как на рис. 5.2.

Обозначения a, b, c, d относятся к коэффициентам, показывающим влияние латентной переменной на конкретный индикатор (они, как мы увидим позднее, выражают надежность этого индикатора), а e_i (то есть e_1, e_2, e_3, \dots и т. д.) — это ошибка измерения i -го индикатора. Для ошибок в этой модели предполагается, что они не скоррелированы друг с другом ($cov(e_i, e_j) = 0$) и с истинным значением X , а их средняя равна 0. В модели, представленной на рис. 5.2, все индикаторы — это так называемые *эффект-индикаторы*, все они находятся под влиянием X , и сила связей a, b, c, d соответствует «силе» этого влияния.

Модели измерения с латентной переменной и эффект-индикаторами очень популярны в социальных науках. Причина этой популярности — в нашей склонности объяснять явные поступки людей

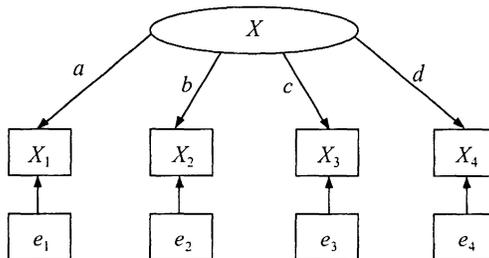


Рис. 5.2. Модель измерения латентной переменной с четырьмя эффект-индикаторами

в частности ответы на вопросы анкеты или выполнение тестовых заданий, неким внутренним свойством, качеством, навыком или предрасположенностью. Латентная переменная может быть, например, интеллектом, измеряемым с помощью индикаторов-тестов. Другой пример: мы можем полагать, что участие в выборах и ежедневное чтение политических новостей в газете — это индикаторы латентной «политической активности» или «вовлеченности в политику».

Однако использование эффект-индикаторов — это не единственная возможность. Например, мы можем использовать такие индикаторы, как потеря работы, развод, болезнь для измерения латентной переменной «жизненный стресс». В этом случае мы *не предполагаем*, что латентная переменная является причиной своих индикаторов¹⁰, скорее, травмирующие жизненные события могут быть причиной стресса. Если мы имеем дело с какой-то из распространенных моделей социально-экономического статуса, в ней тоже будут присутствовать не эффект-индикаторы, а *причинные* (или *формативные*) индикаторы, то есть индикаторы, значения которых детерминируют, определяют значение латентной переменной. На рис. 5.3 изображена элементарная модель латентной переменной с причинными индикаторами ($Y_1 - Y_4$ — это индикаторы, Y — латентная переменная).

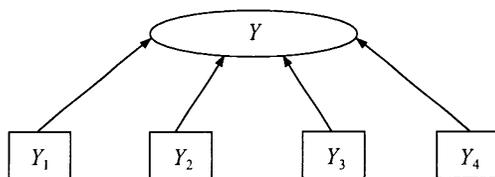


Рис. 5.3. Модель измерения с латентной переменной и причинными индикаторами

¹⁰ Здесь и далее мы говорим о «причине» лишь в том смысле, что значение латентной переменной детерминирует, определяет значения индикаторов (или наоборот).

Если Y — это социально-экономический статус (СЭС), то $Y_1 - Y_4$ могут представлять собой доход, образование, престиж профессии данного человека и «качество» его жилья (стоимость, престижность района и т. п.). Ясно, что скорее доход является причиной СЭС, чем наоборот. Несмотря на кажущееся сходство моделей измерения, изображенных на рисунках 5.2 и 5.3, *их «поведение» на стадии анализа будет очень разным*. Разными могут оказаться и методы оценки качества индикаторов для этих моделей. Даже без специального анализа можно сказать, что в модели с эффект-индикаторами (рис. 5.2) всякий «хороший» индикатор должен чутко реагировать на рост или убывание латентной переменной и изменяться «в согласии» с остальными. В модели, изображенной на рис. 5.3, дело обстоит не так просто: если, скажем, возрастет доход — возрастет и статус, но образование или профессиональный престиж вполне могут не измениться, остаться на прежнем уровне. Другое очевидное отличие связано собственно с отбором индикаторов: для модели на рис. 5.2 *любой «хороший»* (то есть надежный и валидный, см. гл. 6) *эффект-индикатор может заменить любой другой*, и их общее число вполне можно сократить: скажем, высокие результаты выполнения одного «хорошего» теста интеллекта будут достаточно надежно предсказывать результаты бесчисленного множества других тестов. Если же мы попытаемся убрать какой-то причинный, формативный индикатор, то изменится не только объем нашей анкеты — *изменится сама латентная переменная*, которую эти индикаторы собственно и составляют: так, стоит «убрать» доход из числа индикаторов СЭС, как мы уже будем изучать что-то вроде социального, но уж никак не экономического статуса. Приведенные примеры позволяют понять, почему так важно явно задать модель измерения, связывающую индикаторы, которые мы собираемся отобразить, с теоретическими понятиями.

Многие реальные модели измерения еще сложнее только что описанных. Индикаторы могут быть скоррелированы между собой и, что хуже, с ошибками измерения, в число индикаторов могут одновременно входить и эффект-индикаторы, и индикаторы-причины. Часто разработка модели измерения ведет к радикальному прояснению теоретических гипотез и понятий, которые на предыдущих стадиях исследования носили чрезмерно абстрактный и общий

характер Так, социолог, стремящийся найти индикаторы, скажем, «межэтнической напряженности», попытается по меньшей мере разделить «причины» и «эффекты» среди таких показателей напряженности, как *поселенческая сегрегация* (склонность представителей этнических групп к компактному и раздельному поселению), *отсутствие семейных и дружеских связей* с представителями «чужого» этноса, *число столкновений и вооруженных конфликтов*, *недоброжелательное освещение «чужаков»* в местной прессе и т. п. В ходе такой работы он наверняка сделает более ясными и отчетливо сформулированными свои представления о механизмах возникновения межэтнической напряженности и ее последствиях.

В целом при поиске и отборе индикаторов полезно руководствоваться некоторыми общепринятыми правилами:

1. Используйте индикаторы, применявшиеся в более ранних исследованиях. Существует множество устоявшихся и проверенных индексов (то есть суммарных показателей) и шкал, свойства которых достаточно известны. При возможности проверьте, насколько хорошо «работают» эти показатели в вашем случае, проведя небольшое разведочное (пилотажное) исследование. Сориентироваться в многообразии существующих показателей и шкал помогают соответствующие справочные издания и тематические обзоры¹¹.

2. Если общепринятого способа измерения для какого-то понятия не существует, попытайтесь разработать множество индикаторов для различных определений понятия и проверьте, как различия индикаторов будут влиять на различия в интерпретации результатов. Имея дело с многомерным понятием, стоит подумать, какие именно измерения, аспекты понятия существенны в рамках вашей исследовательской гипотезы.

3. Обычно установки и мнения имеют более сложную структуру и требуют использования большего количества индикаторов, чем, например, поведенческие события. Конечно, решающее слово

¹¹ Среди справочных изданий общего характера следует в первую очередь указать на: *Robinson J. P. et al. Measures of Political Attitudes*. Ann Arbor: ISR, 1968; *Shaw M. E., Wright J. M. Scales for the Measurement of Attitudes*. N. Y.: McGray-Hill, 1967; *Miller D. E. Handbook of Research Design and Social Measurement*. N. Y.: McKay D., 1970.

в определении количества индикаторов (количества вопросов в анкете) принадлежит практическим соображениям. Пилотажные исследования, интервьюирование «фокусных» групп могут оказаться полезными в отборе индикаторов и исключении лишних вопросов. Они также важны для оценки надежности и валидности показателей (см. гл. 6).

Прежде чем перейти к практическим проблемам конструирования вопросов для анкет и интервью, мы коротко рассмотрим уровни измерения, так как общее представление об уровнях измерения понадобится нам при обсуждении логики построения вопросов и ответов и использования шкал.

Уровни измерения

Существует несколько концепций измерения, по-разному определяющих, что может быть названо *операцией измерения*. В гуманитарных науках — и социология не является исключением — наибольшее влияние имеет *репрезентационная концепция измерения*, впервые детально обоснованная психофизиком С. С. Стивенсом. В этой концепции всякая операция измерения в конечном счете определяется как приписывание чисел вещам (свойствам, событиям) в соответствии с определенными правилами, так что отношения между числами отражают (или представляют, репрезентируют) отношения между вещами. Таким образом, измерение представляет определенные свойства в виде чисел, поддающихся суммированию, сравнению и т. п. Однако наша возможность измерить какие-то эмпирически наблюдаемые свойства, представить отношения между вещами в виде чисел редко носит абсолютный характер. О некоторых эмпирических свойствах мы можем сказать, что они выражены «больше» или «меньше» для каждого конкретного наблюдения, но не можем указать случаи, когда это свойство абсолютно отсутствует: так, даже если испытуемый не решил ни одной задачи, мы едва ли осмелимся утверждать, что он полностью лишен «интеллекта». Иногда наша способность измерять ограничена лишь возможностью отнести какую-то вещь (наблюдение) к определенному классу, причем между разными классами нельзя задать отношение порядка (больше–меньше). Иными словами, при измерении *отно-*

шения между числами как-то зависят от отношений между вещами, и, следовательно, существуют ограничения для возможных преобразований чисел: игнорируя эти ограничения, мы теряем право утверждать, что наши числа что-то представляют, репрезентируют. Правила приписывания чисел вещам, используемые нами в каждом конкретном случае, воплощают в себе эти ограничения и определяют достигнутый *уровень измерения* (номинальный, порядковый, интервальный, абсолютный).

Номинальные измерения

Номинальным измерением называют процесс отнесения объектов в классы. Все, что мы можем сказать об объектах, сгруппированных в один класс, — это то, что они идентичны в отношении некоторого свойства или признака, то есть фактическое отношение между объектами — это отношение тождества (или различия). Для обозначения полученных классов могут использоваться и названия свойств, и числовые символы. Скажем, мы можем обозначать символом «0» мужчин, а символом «1» — женщин. Однако нельзя сказать, что признак «является мужчиной» в каком-то отношении меньше признака «является женщиной» или что «сумма одного мужчины и одной женщины равна единице». Хотя номинальные измерения довольно примитивны, они отнюдь не бесполезны, в чем мы убедимся при обсуждении методов анализа данных. Другими примерами номинального измерения могут служить национальность или место жительства.

Порядковые измерения

Измерение на порядковом (ординальном) уровне предполагает, что мы способны упорядочить объекты по степени выраженности свойства или признака, то есть определить для них отношение «больше–меньше». Например, мы можем говорить о низком, среднем или высоком социальном статусе или низкой, умеренной или высокой коммуникабельности. Однако в случае порядкового измерения мы не можем определить точно, насколько велико расстояние между соседними категориями. Иными словами, мы не можем утверждать, что человек, получивший оценку «3» по шкале популярности, в три раза более популярен, чем получивший оценку «1»,

или что расстояние между категориями «48» и «45» по порядковой (ординальной) шкале равно расстоянию между категориями «22» и «19». Иными словами, ординальное измерение задает отношение порядка между категориями какого-то свойства, но не позволяет говорить о том, «на сколько» или «во сколько раз» одна категория больше другой, то есть ни точка отсчета (абсолютный ноль), ни единица измерения здесь не могут быть определены.

Интервальный уровень измерения

Об интервальном уровне измерения можно говорить тогда, когда мы способны не только определить количество интересующего нас свойства в эмпирических наблюдениях, но также определить равные расстояния между категориями, то есть ввести *единицу измерения*. Соответственно числовое приписывание становится здесь менее произвольным: объекту (наблюдению) присваивается число, соответствующее количеству измеряемого свойства, то есть мы можем установить отношения равенства уже не между самими объектами, а между интервалами числовой шкалы: равные разности чисел соответствуют равным разностям значений измеряемого свойства или признака. Классический пример интервального измерения в физических науках — это измерение температуры по шкале Цельсия (или Фаренгейта). Единицы измерения — градусы — равны, однако «0» — это произвольная точка. При 0 °С вода замерзает, однако свойство «иметь температуру» отнюдь не исчезает. Если нулевая точка неабсолютна, то бессмысленно утверждать, что 30 °С предполагают в три раза больше свойства «температура», чем 10 °С.

Измерение отношений

Шкала температуры Кельвина, как известно, начинается с абсолютного нуля, и этот абсолютный ноль имеет определенный физический смысл (вспомните термодинамику), так что можно даже сказать, что здесь «температура кончается». Шкала Кельвина — это шкала отношений. То же можно сказать и о физическом измерении расстояний, в частности об измерении роста. Человек, имеющий рост в 2 метра, в два раза выше ребенка, чей рост 1 метр. Возраст человека, доход — другие примеры шкалы отношений.

Зачем учитывать уровень измерения?

Во-первых, отметим, что наше изложение существующих представлений об уровнях измерения — пусть оно и было далеко не полным¹² — позволило заметить, что хотя приписывание чисел объектам возможно практически всегда, далеко не все операции над полученными числами будут иметь какой-то смысл. Соответственно, далеко не все методы группировки и статистического анализа данных уместны для номинального или, скажем, интервального уровня измерения (с ними социологам чаще всего приходится иметь дело). Существуют различные техники анализа для разных уровней измерения переменных. Специальные методы построения социологических шкал, о которых будет говориться далее, также основаны на определенных представлениях о *метрике* переменных, то есть об уровне их измерения. Все эти соображения должны быть приняты во внимание и при конструировании инструмента сбора данных, например вопросника. Если мы хотим анализировать переменную «образование» по крайней мере на интервальном уровне, нам, вероятно, лучше использовать показатель «количество лет, затраченных на получение образования», и включить в анкету соответствующие вопросы. Однако если наша цель — всего лишь показать, что лица с высшим образованием или ученой степенью чаще выписывают научно-популярные журналы, достаточно будет использовать привычные «ординальные» категории — неполное среднее, среднее, высшее и т. п. (кстати, при анализе они, возможно, будут рассматриваться как номинальные).

Важно помнить, что каждая переменная может быть измерена на разных уровнях. Выбор определяется практическими соображениями, требованиями к качеству измерения (как правило, существует обратная зависимость между уровнем и качеством измерения, о чем еще будет говориться дальше), предполагаемой стратегией анализа данных. Практически всегда данные, позволяющие получить

¹² Более строгое и систематическое изложение этой темы см.: *Стивенс С. С.* Математика, измерение и психофизика // *Экспериментальная психология*: Пер. с англ., М.: Изд-во иностр. лит., 1960. Т. 1. С. 19–89. См. также: *Клигер С. А., Колосанов М. С., Толстова Ю. Н.* Шкалирование при сборе и анализе социологической информации. М.: Наука, 1978. С. 7–39.

высокий уровень измерения, могут быть перегруппированы так, что уровень измерения станет ниже (обратное утверждение, к сожалению, неверно). Например, при анализе мы можем разбить наших респондентов на три возрастные категории, хотя в опросе использовали семь. Важно, однако, и то обстоятельство, что исследователь, использующий наши данные для вторичного или сравнительного анализа (возможно, мы и сами захотим к ним вернуться), сможет пользоваться «сырыми», более детальными категориями.

Общие правила конструирования опросников

Исследователь может использовать различные техники сбора данных: наблюдение, контент-анализ, анкетный опрос, интервьюирование и т. п. Самой распространенной техникой все же является опрос. Используемые в ходе опроса анкеты могут заполняться самими респондентами или специально обученными интервьюерами. В любом случае каждый из респондентов отвечает на фиксированные вопросы.

Используя стандартный инструмент сбора данных — анкету, опросный лист, — социолог получает те сведения, которые позволяют заполнить матрицу данных «респонденты × переменные» (см. с. 117). Те пропуски в данных, которые возникли из-за неясности вопросов, неопределенной интерпретации ответов или нежелания людей отвечать на предложенный вопрос, обычно нельзя восстановить, вернувшись домой к респонденту. Поэтому так важно продумать заранее, какие вопросы следует задать.

Решающим соображением в выборе вопросов, которые будут заданы респонденту, является осуществленный исследователем выбор индикаторов теоретических понятий (см. выше): например, изучая влияние успехов в учебе на коммуникабельность студентов, мы задаем конкретные вопросы об экзаменационных оценках, числе друзей среди сокурсников, участии в самодеятельности, посещении студенческого бара и т. п. Если наше исследование является скорее объяснительным, чем сугубо описательным, полезно представить себе схематически ту теоретическую модель, которую мы намерены проверить, снабдив каждый теоретический конструкт «его» индикаторами.

Рассмотрим это на примере. Предположим, наша теоретическая модель сводится к простой гипотезе: *люди, подвергавшиеся преступным посягательствам или бывшие свидетелями преступлений, то есть имеющие «опыт жертвы», в большей степени поддерживают применение высшей меры наказания.* Схематически наша гипотеза представлена на рис. 5.4.



Рис. 5.4. Схема, иллюстрирующая зависимость поддержки смертной казни (ПСК) от «опыта жертвы» (ОЖ)

Естественно, мы постараемся найти достаточное число индикаторов и к независимой («опыт жертвы»), и к зависимой (поддержка смертной казни) переменным. Разрабатывая понятие «опыт жертвы», мы можем решить, например, что здесь существенны не только реальные биографические факты, но и информированность человека об уголовных происшествиях, основанная на общении с друзьями и близкими, устойчивый интерес к соответствующим сообщениям в газетах или теленовостях, а может быть, и некоторые психологические факторы — уровень тревожности, идентификация с жертвой и т. п. Выбирая индикаторы для зависимой переменной, мы должны будем по крайней мере учесть возможные различия в диапазоне и интенсивности выражаемой респондентами поддержки смертной казни, что также приведет нас к целому «вееру» прямых и косвенных показателей. Нельзя не признать, в частности, что существует какое-то различие между людьми, требующими расстрела на месте за порчу телефонного автомата, и теми, кто полагает, что единственным «показанием» к смертной казни может быть убийство при отягчающих обстоятельствах. Возможно, следует также признать существенным различие в силе убеждений между теми, кто поддерживает высшую меру «вообще», и теми, кто при необходимости сам готов пристрелить преступника.

Каркас теоретической модели, изображенной на рис. 5.4, начнет, таким образом, обростать какими-то операциональными

индикаторами независимой и зависимой переменных, постепенно превращаясь в конкретную модель измерения (см. рис. 5.5).



Рис. 5.5. Дополненная схема для примера поддержки смертной казни

(Отметим, что индикаторы независимой переменной в данном случае — это формативные индикаторы, то есть индикаторы-причины). Пропуски в схеме, представленной на рис. 5.5, подразумевают наличие других, не рассмотренных нами показателей. Отсутствие стрелки и вопросительный знак в связи «личностная тревожность — опыт жертвы» отражают не столько неясность направления этой причинной связи, сколько другое важное обстоятельство: в нашей теоретической схеме мы забыли учесть контрольные переменные, которые могут влиять на отношение между зависимой и независимой переменными или даже полностью определять это отношение. В разделе, посвященном анализу данных, мы увидим, как учет контрольной переменной может полностью менять характер наблюдаемой связи. Пока же достаточно заметить, что личностная тревожность может быть опосредующей перемен-

ной, то есть может оказаться, что связь «опыта жертвы» и «поддержки смертной казни» очень высока для высокотревожных опрошенных и совершенно незначима в других группах: действительно, люди с высокой личностной тревожностью¹³ склонны к аффективной переоценке даже незначительных происшествий, поэтому при прочих равных их «опыт» всегда будет обладать большей субъективной значимостью. Среди других возможных контрольных переменных почти всегда будут фигурировать фоновые социально-демографические факторы, подобные возрасту, образованию, социальному классу и т. д.

В последнем утверждении нет ничего загадочного: принадлежность человека к устойчивой социальной группе — к тому же воспринимаемой другими людьми как таковая — в немалой степени «формирует» его поведение и установки, определяет горизонт нормативных ожиданий и т. п.

Достигнутая ясность теоретической схемы исследования (в нашем вымышленном примере с поддержкой смертной казни — скорее недостаточная) дает нам вполне практический ориентир для *отбора анкетных вопросов*. Составляя анкету, мы, во-первых, включим в нее вопросы, позволяющие измерить зависимую переменную (переменные) и, во-вторых, постараемся убедиться в том, что все объяснительные, независимые переменные также переведены на язык соответствующих вопросов. В-третьих, мы осуществим поиск вопросов, относящихся к возможным контрольным переменным, не забыв о стандартных «паспортных» вопросах, которые будут касаться пола, возраста, рода занятий, образования, семейного статуса и т. п., то есть будут измерять фоновые переменные.

До сих пор мы говорили лишь об отборе вопросов, которые составят «ядро» нашего инструмента сбора данных. Теперь нам предстоит обсудить, как эти вопросы могут формулироваться и оцениваться и, кроме того, как из множества вопросов может быть составлен макет анкеты.

¹³ Личностная тревожность, в отличие от реактивной, мало зависит от внешних травмирующих или угрожающих факторов и может рассматриваться как устойчивая индивидуальная черта.

Специальные методические исследования показывают, что *формулировка вопроса* имеет решающее значение для качества данных. Но даже основываясь исключительно на здравом смысле, можно предположить, что вопросы и предлагаемые подсказки (альтернативы ответов) должны быть ясными, недвусмысленными и удобочитаемыми. Каким бы ни был *формат вопроса* — открытым, полузакрытым или закрытым¹⁴, — его словесная форма должна гарантировать возможность двусторонней коммуникации между исследователем и респондентом.

Д. де Вос свел те *требования*, которые обычно предъявляют к *словесной формулировке вопроса*, в удобный контрольный список, которым можно руководствоваться в практической работе¹⁵. Мы изложим основные правила формулировки вопросов, руководствуясь этим контрольным списком:

1. В словесной формулировке вопроса следует избегать использования специальных терминов или сленга. Чаще всего специальные термины в вопросах — результат того, что социолог не смог достаточно отчетливо операционализировать исходное понятие, найти его эмпирические эквиваленты в повседневном поведении или высказываниях людей. Вопросы типа «является ли ваша семья нуклеарной?» или «поддерживаете ли вы либералистскую концепцию роли государства в экономике?» свидетельствуют о том, что ученый пытается решить свою теоретическую проблему посредством прямого опроса общественного мнения. Смешение собственно исследовательского вопроса с вопросом к респонденту допустимо лишь в одном случае — в опросе экспертов, когда целью как раз и является расширение компетентности исследователя путем учета мнений высококвалифицированных специалистов. Использование жаргонных выражений обычно свидетельствует о стремлении социолога быть понятным, «своим», разделяющим проблемы опера-

¹⁴ Открытым обычно называют вопрос, ответ на который респондент дает в свободной форме, закрытым — вопрос с жестко фиксированными альтернативами ответа. Открытые вопросы, в принципе, позволяют респонденту точнее выразить свое мнение, однако к их недостаткам можно отнести сложность кодирования, ограничения на сопоставимость данных и возможности анализа.

¹⁵ *Vaus D. A. de, Survey in Social Research. L.: Allen & Unwin, 1986. P. 71–74.*

шиваемых. Эти похвальные усилия нередко, однако, оказываются бесплодными, так как далеко не все респонденты принадлежат к одной и той же субкультуре, да и не всем придется по душе предложенный неформальный тон. Конечно, часть подростков сразу поймет, что имеется в виду, когда их спросят: «Доводилось ли тебе баловаться “травкой”?», но это не обязательно сделает их ответы более откровенными. Кроме того, многие могут просто не понять точный смысл вопроса.

2. Стремитесь к коротким формулировкам. При прочих равных, чем меньше слов в вопросе и предложенных альтернативах ответа, тем меньше шансов, что вас неправильно поймут. (Разумеется, и эту рекомендацию не стоит доводить до абсурда: вопросы не должны превращаться в намеки.)

3. Проверьте, не является ли вопрос многозначным, то есть не содержит ли он в себе двух или более различных по смыслу вопросов, на каждый из которых можно получить независимый ответ. Простой, на первый взгляд, вопрос «когда вы в последний раз читали газету “Известия”?» в действительности требует двух разных вопросов, первый из которых должен касаться того, читает ли человек данную газету вообще. Вопрос «как часто вы и ваша жена (ваш муж) посещаете парфюмерный магазин?» следует разделить по крайней мере на два вопроса, относящиеся к самому респонденту и его супруге (супругу).

4. Избегайте «подталкивающих» (или наводящих) вопросов, неявно указывающих респонденту, какой ответ желателен. «Подталкивающий» вопрос заставляет респондента выбирать «правильный» или социально-желательный ответ. Безусловно «подталкивающими» будут, например, вопросы «примете ли вы участие в выборах, если отказ людей участвовать в голосовании приведет к установлению диктатуры?» или «поддержите ли вы движение за равные права для сексуальных меньшинств даже в том случае, если будете опасаться, что кто-нибудь сможет развратить вашего ребенка?». В менее очевидных случаях к «подталкиванию» могут вести прямые ссылки на мнение авторитетных или влиятельных людей (скажем, «согласны ли вы с мнением премьер-министра X...?»), использование слов, имеющих явную эмоционально-оценочную нагрузку (например, «безответственные политики» или «рискованные

средства»). Еще один способ навязать респондентам собственное мнение — это ограничение числа альтернатив ответа в закрытом вопросе или исключение позиций «другой ответ», «затрудняюсь ответить», «не знаю». Конечно, это значительно облегчает анализ данных, но исследование в этом случае становится просто дорогостоящим средством демонстрации вашей собственной точки зрения.

5. Без крайней необходимости не используйте выражения, содержащие в себе отрицание. Например, спрашивая респондента о согласии или несогласии с утверждением «нельзя делать профилактические прививки взрослым без их добровольного согласия», мы не сможем уверенно утверждать, что означает ответ «нет» — несогласие с суждением или подтверждение согласия. Подобной путаницы не возникнет, если использовать утвердительную формулировку («согласны ли вы с тем, что нужно делать прививки взрослым даже в принудительном порядке?»). Если по каким-то причинам нужно все же сохранить форму отрицания, то выходом становится использование развернутых ответов (например, «нет, принудительные прививки делать нельзя» и т. п.).

6. Вопросы, требующие особой компетенции или осведомленности о чем-то, нужно задавать лишь тем, кто может на них ответить. Если есть основания считать, что не все респонденты могут ответить на вопрос из-за отсутствия каких-то знаний или опыта, нужно использовать предварительный вопрос-фильтр, чтобы отсеять тех респондентов, которые могут иметь квалифицированное мнение. Бессмысленно спрашивать о святом причастии у мусульманина или о мажоритарной системе голосования — у человека, которой не интересуется политикой и не участвует в выборах. Задавая специальные вопросы без предварительного «просеивания» респондентов, исследователь рискует принять искусственно созданное мнение неосведомленных людей за реальное: люди могут весьма уверенно высказываться не только о малознакомых, но даже и о вымышленных предметах, если будут полагать, что это доставит удовольствие социологу.

7. Избегайте любых многозначных или двусмысленных слов и фраз. Конечно, любое слово в некоторых контекстах может выглядеть двусмысленно, так что предыдущая фраза, скорее всего, выражает благое пожелание. И все же нужно пытаться находить замену

для каждого слова, которое по-разному понимается в разных субкультурных группах (например, выражения «быть безразличным к чему-то», «относиться к чему-то безразлично» могут восприниматься как абсолютно нейтральные рабочими и как слегка негативные оценки — школьными учителями).

8. Учитывайте возможное влияние фактора социальной желательности. Фактор социальной желательности — одна из основных угроз валидности измерения (см. гл. 6): стремление людей к социальному одобрению, к «престижному» поведению и образу жизни, к самопрезентации может влиять на их ответы на самые разные вопросы: о предпочитаемой марке автомобиля, о сексуальной активности, даже о доходе или образовании. Поскольку фактор социальной желательности начинает определять ответы в той же мере, что и интересующая нас переменная, наше измерение становится невалидным, возникает систематическое смещение, которое в общем случае обозначается как «установка на ответ» (англ. response set). Борьба с этим видом смещения очень сложно. В некоторых случаях «установку на ответ» можно учесть и оценить ее величину на стадии анализа (для этого применяют модели измерения со множественными индикаторами). Иногда для выявления респондентов, склонных давать социально-желательные ответы, используют специальные шкалы (в психометрике их называют «шкалами лжи»). Такие шкалы состоят из вопросов, провоцирующих конформистские или установочные ответы. Например, если человек утверждает, что ни разу в жизни не солгал или что он никогда не чувствует раздражения, когда кто-нибудь указывает на его промахи, можно предположить, что он испытывает весьма сильную потребность в одобрении. К сожалению, специальные методические исследования показали, что такого рода шкалы в действительности не очень эффективны для идентификации респондентов, особенно подверженных влиянию фактора социальной желательности¹⁶. Идеального решения этой проблемы просто не существует. Важно, однако, осознавать возможность таких смещений, избегать «подталкивающих» вопросов и провоцирующих «социальное тщеславие» формулировок,

¹⁶ Bradburn N. M., Sudman S. et al. Improving Interview Method and Questionnaire Design. San Francisco: Jossey-Bass Publ., 1979. Ch. 6. P. 85–106.

а также уделять особое внимание этой проблеме на стадии анализа и интерпретации.

9. В вопросах, касающихся фактического положения дел или поведения людей, следует достаточно конкретно определять временные и пространственные координаты интересующих вас событий. Трудно ответить на вопросы «пользуетесь ли вы общественным транспортом, находясь вдали от дома?» или «часто ли вы читаете детективы?». Нужно конкретизировать понятие «вдали» (другой город, другая страна, другой микрорайон) и указать, какой период времени имеется в виду (например, «как часто в течение последнего года вы читали...?»).

10. Не стремитесь к излишней детализации вопросов. Во-первых, респонденту в большинстве случаев легче указать некоторый числовой интервал, чем оценить точное значение признака. Даже такой явно «числовой» признак, как доход, может оцениваться по-разному, в зависимости от того, какие источники или временные рамки принимаются в расчет. К тому же вы, скорее всего, не сможете полностью использовать полученные точные оценки — даже если допустить, что они абсолютно надежны, — так как другие переменные будут измерены на номинальном или ординальном уровнях.

«Сензитивные» вопросы

Эту проблему стоит обсудить отдельно, так как необходимость задать «сензитивные» (иногда — «угрожающие», деликатные) вопросы возникает не так уж редко. «Сензитивными» могут считаться любые вопросы, направленные на получение сведений, которые люди обычно предпочитают утаивать. Ответы на личностные или деликатные вопросы чаще бывают неискренними и соответственно ведут к не связанным с выборкой систематическим ошибкам в данных. Влияние «установки на ответ» на качество таких данных изучалось в целом ряде специальных методических исследований¹⁷. В результате удалось показать, что помимо собственно содержания

¹⁷ См., в частности: *Barton, A. Asking the Embarrassing Questions // Public Opinion Quarterly. 1958. Vol. 22. № 1. P. 67–68; Bradburn N. M., Sudman S. Op. cit.*

вопроса на величину смещения влияет его форма. Задавая вопросы, относящиеся к «сензитивным» сферам поведения людей, лучше всего избегать прямых формулировок, подобных вопросу: «Случалось ли вам попадать в вытрезвитель?». Косвенные формулировки обычно используют прием *проекции* нестандартного поведения — на «других людей», «всех людей» — и его *рутинизации*, то есть подчеркивания его обыденности. Примерами косвенных формулировок могут служить вопросы: «Известно, что каждый взрослый мужчина хотя бы раз в жизни может “упиться в стельку” и попасть в медвытрезвитель. Случалось ли что-нибудь подобное с вами?», «Есть ли среди ваших знакомых люди, задушившие своих жен? (Ответ) А вам самому доводилось это делать?».

В некоторых — скорее редких — случаях уместно применение приема, позаимствованного из психологического тестирования — метода *«вынужденного выбора»*. Опрашиваемый должен выбрать одно из 4–5 суждений, каждое из которых имеет одинаковую негативную «нагрузку» по факторам моральной оценки или социальной желательности. Его заранее предупреждают о необходимости делать выбор быстро и о невозможности другого ответа («не знаю», «другое» и т. п.). Задачей является выбор суждения, «наиболее близкого» к точке зрения или «лучше характеризующего» респондента. Разумеется, использование этого приема требует специального обучения интервьюеров и очень высокой мотивации к сотрудничеству со стороны опрашиваемых. Оба эти условия довольно трудно реализовать в практике массовых опросов. Ниже приведены примеры:

«Выберите один из ответов, в наибольшей мере соответствующий вашему поведению:

А. Из наркотических средств я предпочитаю:

- кокаин;
- гашиш;
- морфий;
- героин;

Б. Иногда я практикую сексуальные отношения:

- с лицами своего пола;
- с насекомыми;
- с домашними животными;
- с собственной бабушкой».

Этические проблемы, связанные с использованием метода «вынужденного выбора», достаточно очевидны и не требуют комментариев.

Американский социолог С. Уорнер в конце 1950-х гг. предложил использовать для уменьшения смещения в ответах на «сензитивные» вопросы *модель случайного ответа*. В этом случае респондент отвечает на один из двух случайно выбранных вопросов, а интервьюер не знает, на какой из этих вопросов ответил респондент. Эта модель позднее подвергалась модификациям, в частности было предложено использовать один «сензитивный» и один нейтральный вопрос. В простейшем случае респондент выбирает случайным образом карточку из колоды (статистическое распределение вопросов на карточках известно исследователю) и, не зачитывая вопрос, отвечает на него. Если предположить, что исследователь знает, как распределяется в генеральной совокупности нейтральный признак — скажем, дата рождения, — он может оценить статистически выборочное значение «сензитивного» признака. Фактически в этой модели заданы следующие параметры:

ρ_s = истинное значение доли имеющих «сензитивный» признак s ;

P_1 = вероятность предъявления респонденту карточки с вопросом о наличии у него признака s ;

$1 - P_1$ = вероятность предъявления карточки с вопросом о нейтральном признаке n ;

ρ_n = доля имеющих признак n ;

λ_1 = доля ответов «да».

В этом случае выборочная оценка может быть определена следующим образом:

$$\rho_s = \frac{\lambda_1 - \rho_n(1 - P_1)}{P_1}.$$

Техническое воплощение модели случайного ответа может выглядеть, например, так же, как в специальном методическом исследовании, осуществленном Н. Бредберном и С. Судманом¹⁸.

¹⁸ См.: Bradburn N. M., Sudman, S. et al. Op. cit. P. 1–13, 175–184.

Респондент пользовался ящичком, содержащим 70% красных и 30% синих пластиковых шариков. Ящичек был устроен таким образом, что при потряхивании в маленьком окошке, видимом только респонденту, появлялся в случайном порядке синий либо красный шарик. Опрашивающий в соответствующей части интервью предъявлял респонденту пару вопросов («сензитивный» и нейтральный), напечатанных соответственно на «красной» и «синей» карточках. Респондент отвечал на тот вопрос, которому соответствовал цвет выпавшего шарика. Разумеется, респонденты были предупреждены, что шарики выпадают случайно и, следовательно, интервьюер не знает в каждом конкретном случае, на какой из двух вопросов дается ответ. Кроме того, опрашиваемым объясняли, что механизм случайного выбора используется для того, чтобы облегчить ответы на некоторые трудные вопросы. В описанном исследовании использовались, например, такие карточки¹⁹:

«26... Пожалуйста, встряхните ящичек снова и ответьте на вопрос, цвет которого совпадает с цветом появившегося в окошке шарика. Ответьте “да” или “нет”».

Карточка G.	
(Красный)	Штрафовали ли вас за езду на красный свет в течение последних 12 месяцев?
(Синий)	Родились ли вы в июле?

Исследование Бредберна и Судмана показало, что использование модели случайного ответа несколько уменьшает смещение в ответах о социально-осуждаемых поступках (в этом специальном исследовании оценки, полученные в результате опроса, сопоставлялись с объективными данными, которыми располагали полиция и другие учреждения). Однако для многих «угрожающих» вопросов оно все же оставалось значительным: оценка частоты задержаний за управление машиной в состоянии алкогольного опьянения была на 35% ниже объективного показателя, которым располагала

¹⁹ Ibid. P. 179.

дорожная полиция. Мера уверенности респондентов в анонимности, как и их способность попросту понять «правила игры», также варьировали в достаточно широких пределах. Учитывая технические трудности процедуры и необходимость использования больших выборок, следует признать, что модель случайного ответа может иметь довольно ограниченное применение.

Определенное значение для решения проблемы «сензитивности» имеет и избранная исследователем процедура сбора данных: анкеты, заполняемые самим респондентом, несколько лучше, чем телефонные или личные интервью в тех случаях, когда необходимо предотвратить завышение частоты социально-одобряемого поведения. Однако анкетная процедура ведет к более значительным смещениям в оценке неодобряемого поведения, к тому же доля ответивших в анкетном опросе обычно ниже, чем в случае личного или телефонного интервью.

Кроме того, для «сензитивных» вопросов, требующих достаточно детального (не сводящегося к «да» или «нет») ответа, предпочтительно использовать достаточно длинные и состоящие из привычных респонденту слов формулировки, а также открытый или полужакрытый формат вопросов. Важно также спокойное, нейтральное поведение интервьюера (в случае личного или телефонного интервью) и заверения в конфиденциальности, данные в начале беседы.

Довольно полезным приемом является *использование вопросов о друзьях или знакомых*. Конечно, полученные таким способом данные не могут быть «приписаны» к конкретным выборочным единицам, но они дают довольно правдоподобные оценки для некоторых типов поведения (обратите внимание на то, что речь идет не об использовании проективных вопросов о «других людях», а о вопросах, касающихся реального окружения респондента). Видимо, такие вопросы особенно уместны, когда необходимо оценить распространенность или частоту каких-то поведенческих актов, и сравнительно бесполезны для точной оценки интенсивности поведения или установок. Такой подход может быть проиллюстрирован следующим вопросом: «Представьте себе троих своих ближайших друзей. (Нет нужды называть их имена.) Как много из них, по вашему мнению, когда-либо употребляли наркотики?»

Наконец, не исключен и подход, описанный в пионерской статье А. Бартона (см. сноску на с. 150) как «метод Кинзи» (имеется в виду известный исследователь сексуального поведения): посмотрите прямо в глаза респонденту и спросите его прямо и спокойно: «Вы когда-нибудь убивали свою жену?»

Выбор формата для ответов

Когда исследователь избрал по тем или иным соображениям *закрытый* или *полузакрытый* тип вопроса, он должен решить, *в какой форме будут представлены категории ответа*. Формат ответа — это, собственно, та задача, которую должен выполнить респондент, отвечая на вопрос.

Преимуществами закрытого вопроса являются простота кодирования и возможность облегчения задачи для тех респондентов, которые затрудняются сформулировать ответ самостоятельно. Кроме того, на закрытые вопросы люди обычно отвечают значительно быстрее. Основной недостаток закрытых вопросов заключается в том, что в ряде случаев они могут вести к созданию ложных, не существующих в действительности мнений либо из-за недостаточного диапазона категорий ответа, либо из-за неявного подталкивания респондентов к выбору «приемлемых» ответов. Поэтому, используя закрытые вопросы, социолог должен уделить особое внимание разработке списка категорий ответа. Этот список должен быть исчерпывающим, согласованным с логической структурой вопроса и содержать в себе категорию ответа, соответствующую отсутствию мнения у респондента («не знаю»). При разработке списка уместны и небольшие пилотажные исследования на группе респондентов, и экспертиза коллег. Дж. Гэллап предложил специальную *пятишаговую процедуру для разработки «хороших» закрытых вопросов*²⁰. Эта процедура отличается трудоемкостью, но ее целесообразно использовать по крайней мере для ключевых исследовательских переменных.

Предлагается включать в опрос последовательно:

²⁰ Gallup G. H. The Quintamensional Plan of Question Design // Public Opinion Quat. 1947. Vol. 11. P. 385–393.

- 1) закрытый вопрос для выявления осведомленности респондента или наличия у него какого-то отношения к данной проблеме;
- 2) открытый вопрос для оценки общего отношения, обобщенной установки;
- 3) закрытый вопрос, содержащий описание конкретных аспектов, сторон проблемы;
- 4) открытый или полузакрытый вопрос, позволяющий респондентам обосновать свое мнение, объяснить, почему они так считают;
- 5) закрытый вопрос на интенсивность мнения, показывающий, насколько респондент убежден в своей точке зрения.

Наиболее распространенными форматами ответа на простые закрытые вопросы являются: *оценочная шкала; контрольный список (меню); выбор между суждениями, выражающими установку или мнение; ранжирование.*

В случаях, когда закрытый вопрос является в строгом смысле шкалой или пунктом шкалы (индекса), к его формату обычно предъявляют особые требования, которые мы рассмотрим в разделе, посвященном конструированию шкал (см. гл. 6). Пока же достаточно будет дать общую характеристику форматов ответа.

Типичная *оценочная шкала* — это шкала согласия с суждением (суждениями). Респонденту предлагается оценить степень своего согласия с некоторым мнением или высказыванием, то есть расположить **самого себя** на одномерном континууме, крайними точками которого являются полное согласие и полное несогласие с высказыванием. Обычно предполагается также, что респондент способен осуществить такую оценку по крайней мере на порядковом (ординальном) уровне, то есть отнести себя к какой-то градации оценочной шкалы («скорее согласен», «совершенно согласен» и т. п.). Оценки могут быть представлены в вербальной или графической форме. Примеры оценочных шкал представлены на рис. 5.6.

К оценочным шкалам можно отнести также те, которые выражают оценку респондентом своего положения на континуумах удовлетворенности чем-либо, значимости какого-то мнения, интенсивности установки. На рис. 5.7 приведен пример графической оценочной шкалы для оценки значимости высказывания (проблемы) для респондента. Самому большому кружку соответствует макси-

Вербальная

«Большинство людей пользуются правилами вежливости, чтобы прикрыть жестокость конкурентной борьбы.»

- совершенно согласен;
- согласен;
- затрудняюсь сказать;
- не согласен;
- совершенно несогласен.

Графическая

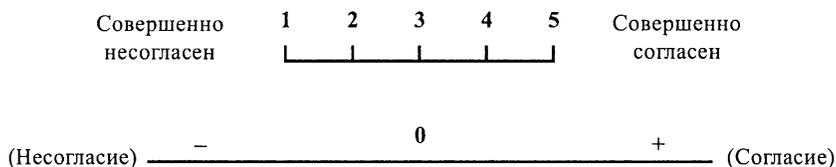


Рис. 5.6. Примеры оценочных шкал

мальная субъективная значимость, наименьшему — минимальная (респонденту даются соответствующие инструкции по заполнению опросника).

Особый тип оценочных шкал — это шкалы категориальной оценки, используемые в методах, подобных «семантическому дифференциалу» (СД). Подробный анализ процедур семантического шкалирования выходит за пределы данного раздела, но общий подход таков: респондент должен многократно оценить некий объект

Оцените, насколько для вас серьезны, важны следующие проблемы (отметьте соответствующий кружок):

Снижение уровня преступности в городе	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
Снижение таможенных пошлин на ввоз оборудования	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>

Рис. 5.7. Пример графической оценочной шкалы для оценки значимости высказывания для респондента

или несколько объектов (личность, товар, профессию и т. п.) по шкалам, каждая из которых задана двумя полюсами, точнее — парой противоположных по смыслу прилагательных (антонимов). Мы назвали СД-подобные шкалы особым типом оценочных шкал, имея в виду их использование для специальных исследовательских задач и довольно сложные методы анализа результатов. Важно, однако, обратить внимание на то обстоятельство, что описанные нами выше оценочные шкалы с большими основаниями могут рассматриваться в качестве частных, «вырожденных» случаев семантического шкалирования — с одним объектом оценивания (например, «Я сам») либо с одной шкалой (скажем, шкалой удовлетворенности). Процедуры, в которых используются СД-подобные шкалы, приводят к заполнению параллелепипеда данных типа «субъект × объект оценивания × переменная-признак», что значительно расширяет возможности анализа²¹. Пример СД-подобного формата ответов приведен ниже.

«Как бы вы описали психологическую атмосферу в вашем коллективе?»

теплая	1	2	3	4	5	холодная
активная	1	2	3	4	5	пассивная
соревновательная	1	2	3	4	5	дружелюбная»

Вопросы-«меню» требуют от респондента выбора одного или более пунктов из представленного списка. Например, одна из шкал близости межличностных отношений предлагает респонденту отметить в контрольном списке только те виды деятельности, которые обычно осуществляются им совместно с близким человеком. Приведем фрагмент списка²²:

²¹ Для более детального ознакомления с этим классом шкал см.: Шмелев А. Г. Введение в экспериментальную психосемантику. М.: Изд-во МГУ, 1983.

²² Berscheid E., Shyder M., Omoto A. The Relationship Closeness Inventory: Assessing the Closeness of Interpersonal Relationships // Journal of Personality & Soc. Psychology. 1989. Vol. 57. P. 806.

- 1) стирка [];
- 2) приготовление еды [];
- 3) просмотр телепередач [];
-
- 8) прогулки пешком, поездки [];
- 9) обсуждение личных проблем [];
- 10) посещение музеев, выставок [].

Выбор между суждениями, выражающими установку по отношению к некоторому объекту, предполагает, что респондент выберет то суждение, которое отражает степень его благожелательности по отношению к объекту. Упорядоченность суждений вдоль некоторого «установочного» континуума обычно очевидна и для исследователя, и для самого опрашиваемого (о более сложных методах измерения установок вы узнаете из раздела гл. 6, посвященного шкалам).

Приведем пример.

«Существуют разные подходы к наказанию за особо тяжкие преступления. Какое из приведенных ниже суждений отражает вашу точку зрения на смертную казнь как меру наказания:

- это наименее приемлемый из всех способов наказания преступников;
- это не очень приемлемый метод наказания, однако он не хуже других;
- это вполне приемлемое наказание за некоторые виды преступлений;
- это вполне приемлемая мера наказания преступников;
- другой ответ (напишите, пожалуйста)

_____».

Ранжирование как формат ответа подразумевает, что респондент упорядочивает весь представленный ему список — пунктов, суждений, объектов и т. п. — вдоль одномерного континуума какого-то качества (при этом не отмечая своего собственного положения на континууме). Чаще всего производится ранжирование по предпочтительности, важности, значимости, популярности. В массовых опросах нередко используются упрощенные подходы к

ранжированию (скажем, «назовите трех самых популярных и трех наименее популярных политиков из перечисленных...»). Это позволяет экономить время участников опроса, но значительно сильнее, чем принято думать, снижает качество и сопоставимость данных.

Многие реальные форматы ответов сочетают в себе черты двух или трех описанных типов. Важно, однако, чтобы задача респондента, определяемая избранным форматом ответа, была однозначной и логически непротиворечивой.

Макет анкеты (опросника)

Общий вид и структура анкеты (опросника) не менее важны, чем содержание и словесное оформление вопросов. Плохое оформление или неясные инструкции часто ведут к невосполнимым пропускам в данных.

Общая рекомендация, равно применимая и к вопросам, и к анкетам, заключается в том, что поля и интервалы между вопросами (или блоками вопросов) должны быть достаточно велики. Стремление «ужать» вопросник, сделать его объем небольшим, чтобы не отпугнуть респондентов, на практике приводит к появлению «слепого», путаного и визуально непривлекательного инструмента сбора данных. Вопросы должны печататься только на одной стороне листа. Поля должны быть не меньше 2,5 см, что необходимо для кодирования данных. Альтернативы ответа лучше располагать вертикально, а не горизонтально. (Когда они перечисляются в одной строке, респондент чаще пропускает средние позиции.) Излишне говорить, что наличие нескольких незаполненных строк для ответа положительно сказывается на желании опрашиваемых подробно ответить на открытый вопрос.

Большинство социологов-практиков убеждены, что наиболее удобной формой для фиксации ответа на закрытый вопрос является перечеркивание «квадратика» рядом с избранным ответом (или проставление в нем «галочки»), например:

- да
- нет
- не знаю

Возможно также использование квадратных [] или круглых () скобок. Если каждая из позиций ответа имеет заранее присвоенный цифровой код, то можно попросить респондента обвести соответствующие цифры кружком:

- 05 да
- 06 нет
- 07 не знаю

Последний вариант, однако, менее предпочтителен: во-первых, не все смогут обвести цифры достаточно аккуратно (что приведет к некоторым проблемам при вводе данных), и, во-вторых, обильно усеянная цифрами анкета некоторыми респондентами воспринимается как математическая головоломка. Разумеется, последние два замечания относятся лишь к анкетам, заполняемым респондентами: интервьюеров можно обучить безошибочной работе с цифровыми кодами.

Важную роль в структуре анкеты играют *вопросы-фильтры* (или «отсекающие» вопросы). Они позволяют отобрать группу респондентов, которые могут (в силу своего статуса, опыта, заинтересованности или знаний) ответить на следующие за вопросом-фильтром *зависимые вопросы*. Для остальных респондентов зависимые вопросы будут попросту нерелевантны. Скажем, спрашивать о том, какое средство контрацепции они предпочитают (зависимый вопрос), имеет смысл у тех респондентов, которые пользовались или пользуются контрацептивами (вопрос-фильтр). Естественно, вопрос-фильтр всегда предваряет зависимые вопросы, которые респонденты, давшие «не тот» ответ, пропускают. Недопустимо, как уже говорилось, сочетать вопрос-фильтр и зависимый вопрос в одном «условном» вопросе: «Если у вас есть собака, кормите ли вы ее сухим кормом “Педигри пал”?». Ответ «нет» дадут и те люди, которые не держат собак, и те, кто не кормят своих собак сухим кормом, и те, кому особенно не нравится «Педигри пал». (Конечно, использование развернутых ответов для каждой из логически возможных альтернатив может помочь в решении проблемы интерпретации, но в этом случае ваш список «подсказок» станет длиннее, чем инструкция по использованию кофемолки.) Очень удобно, когда вопросы-фильтры и зависимые вопросы выделены с помощью

графических средств: стрелок, рамок, цвета и т. п. Примером может служить рис. 5.8.

1. Есть ли у вас дети?
- Да
 Нет (переходите к вопросу 4)
-
- При ответе «да» на первый вопрос:
2. Сколько у вас детей? (Напишите) _____
3. Каков возраст самого старшего из ваших детей? _____ лет

Рис. 5.8. Пример зависимых вопросов

Иногда социологу нужно задать несколько вопросов, имеющих одинаковые категории ответа (в частности, это могут быть вопросы-пункты, входящие в одну шкалу). В этом случае целесообразно использовать *табличный вопрос*. Использование табличных вопросов делает анкету более компактной и, кроме того, позволяет респонденту увидеть смысловую и логическую организацию однотипных вопросов. Пример табличного вопроса с вербальной оценочной шкалой приведен на рис. 5.9.

Порядок вопросов в анкете обычно определяется несколькими простыми правилами:

1) начинайте с простых фактических вопросов, на которые респондент сможет легко и быстро ответить (желательно, по крайней мере для самозаполняемой анкеты, чтобы это были не «демографические» вопросы о возрасте, национальности и т. п., сразу придающие анкете несколько официальный характер);

2) постарайтесь обеспечить какое-то смысловое соответствие между начальными вопросами и заявленными целями опроса;

3) группируйте вопросы в смысловые блоки, по возможности разделяя их графически: это не только облегчит задачу составления вопросника, но и придаст последнему в глазах опрашиваемых вид упорядоченного целого, а не хаотического и бессвязного нагромождения вопросов;

4) стремитесь всегда двигаться от более простых вопросов к более сложным и от более конкретных — к более абстрактным;

5) избегайте однообразия в конструкции вопросов и выборе формата для ответов, если вопросы следуют друг за другом;

Вопрос анкеты	Ответ				
	Совершенно согласен	Согласен	Не согласен	Совершенно не согласен	Не знаю
Принимая решение, лучше всего опираться на мнение старших, более опытных людей	[]	[]	[]	[]	[]
Обществу нужна «сильная рука», насаждающая закон и порядок	[]	[]	[]	[]	[]
Адвокаты помогают преступникам и мошенникам «выйти сухими из воды»	[]	[]	[]	[]	[]

Рис 5.9. Пример табличного представления вопроса

6) не забывайте о возможном влиянии «согласительской» установки на ответ: меняйте «направленность» вопросов, измеряющих одну и ту же установочную переменную;

7) избегайте слишком резких переходов от конкретных вопросов, относящихся к повседневной жизни и опыту, к мировоззренческим, установочным или «сензитивным» вопросам: трудно предсказать, как отреагирует человек, которого вы внезапно спросите: «Доводилось ли вам лечиться от венерических болезней?» или «Веруете ли вы в Бога?» (иногда в практике массовых опросов используют малоинформативные или даже вовсе неинформативные «буферные» вопросы, сглаживающие такого рода переход);

8) в некоторых случаях есть смысл использовать небольшие «введения» к каждому блоку вопросов (например: «В следующем разделе нашей анкеты мы хотели бы узнать, что москвичи думают о городском транспорте»);

9) небольшие подсказки и конкретные инструкции по заполнению должны содержаться не только в общем введении к анкете, но и везде, где это необходимо для ясного понимания респондентом сути стоящей перед ним задачи (скажем, «в случае, если ни один из

предложенных ответов не соответствует вашему мнению, впишите ваш ответ...» или «напишите “1” рядом с фамилией самого любимого телеведущего...»).

Во *введении к анкете* должны содержаться сведения о целях исследования, о том, кто проводит данный опрос. Кроме того, уместны заверения в конфиденциальности (анонимности), какие-то объяснения по поводу того, как был выбран данный респондент (скажем, «случайным образом»). Если анкета заполняется самим респондентом, нужно указать, каким образом, когда и кому (по какому адресу) она должна быть возвращена. Практика показывает, что чрезмерно длинные и обстоятельные интродукции не приводят к сколько-нибудь существенному улучшению качества ответов или увеличению мотивации к сотрудничеству. И в случае интервью, и в ситуации, когда респондент сам отвечает на вопросы, лучше перейти к сути дела сразу же, после небольшого введения: все равно опрошенные не уделят должного внимания касающимся вас и вашего проекта деталям, если они (детали) будут представлены до начала опроса. Намного лучше, если, завершив интервью, вы вручите респонденту визитную карточку, памятный значок с символикой института, который вы представляете, и т. п. Желательно также поблагодарить респондента за сотрудничество.

Если составленный вами вопросник будет заполняться в ходе *личного интервью*, нужно сразу же позаботиться о составлении памятки для интервьюеров. Полезно также использовать внутри вопросника шрифтовые выделения, позволяющие отделить те инструкции и замечания, которые адресованы только интервьюеру, от предназначенных респонденту и зачитываемых интервьюером вслух. Лучше, если все уточняющие реплики, инструкции и вводные фразы, которые интервьюеру предстоит использовать, будут включены непосредственно в текст вопросника (в скобках, с помощью курсивных выделений и т. п.). В ином случае интервьюеру придется заучивать памятку наизусть — от вводных фраз до «до свидания»²³.

²³ Подробнее об этом лучше прочитать дополнительно. См., например: *Ноль Э. Массовые опросы: Введение в методику демоскопии*. М.: Прогресс, 1978. Гл. 1, 2; *Babbie E. Survey Research Methods*. 2nd ed. Belmont: Walworth Publ., 1990. Ch. 7.

Дополнительная литература

- Клюшина Н. А.* Причины, вызывающие отказ от ответа // Социологические исследования. 1990. № 1. С. 98–104.
- Методы сбора информации в социологических исследованиях. Кн. 1. Социологический опрос / Отв. ред. В. Г. Андреенков, О. М. Маслова. М.: Наука, 1992.
- Маслова О. М.* Познавательные возможности открытых и закрытых вопросов // Социологические исследования. 1984. № 2. С. 129–138.
- Ноэль Э.* Массовые опросы. М.: Прогресс, 1978. С. 76–120.
- Рукавишников В. О., Паниотто В. И., Чурилов Н. Н.* Опросы населения (методический опыт). М.: Финансы и статистика, 1984.
- Шуман Г., Прессер С.* Открытый и закрытый вопрос // Социологические исследования. 1982. № 3.
- Ядов В. А.* Социологическое исследование: методология, программа, методы. 2-е изд., перераб. и доп. М.: Наука. С. 130–166.

ГЛАВА 6. КАЧЕСТВО ИЗМЕРЕНИЯ. СОЦИОЛОГИЧЕСКИЕ ИНДЕКСЫ И ШКАЛЫ

Проблема качества социологического измерения

В предыдущей главе мы сравнительно много говорили о переходе от теоретических конструкторов к конкретным эмпирическим показателям и анкетным вопросам и сравнительно мало — о том, что обеспечивает законность такого перехода и возможность правильной и осмысленной интерпретации «сырых» ответов. Сейчас нам предстоит восполнить этот пробел, рассмотрев возникающие здесь проблемы *качества социологического измерения и построения социологических индексов и шкал*.

Как уже говорилось ранее, переменная-признак, включенная в модель измерения, может задаваться совокупностью эмпирических индикаторов (в предельном случае — одним индикатором), наблюдаемые значения которых содержат, помимо истинного, ошибочный компонент. Задача оценки качества измерения — это, говоря обобщенно, задача разделения истинного значения и ошибки. Еще одна проблема, стоящая перед исследователем, связана с необходимостью использовать несколько показателей (вопросов анкеты, данных наблюдения и т. п.) для построения суммарного индекса или шкалы, позволяющих получать более точные, то есть менее подверженные смещению, оценки интересующей исследователя переменной. Индексы и шкалы позволяют перейти от многообразия конкретных, наблюдаемых переменных-индикаторов, отражающих лишь отдельные аспекты теоретического понятия, к более абстрактным переменным теоретической модели. Так, например, конкретный вопрос об участии в голосовании позволяет сделать определенные выводы о политической активности людей, однако ограничившись только этим вопросом, вы ничего не узнаете о других аспектах и формах политической активности. Кроме того, использование не-

скольких показателей позволяет уменьшить влияние посторонних факторов на наши оценки величины или разброса значений переменной и сделать получаемые результаты более устойчивыми и воспроизводимыми. Иными словами, использование индексов и шкал ведет к повышению *надежности* и *валидности измерения*. С рассмотрения этих понятий мы и начнем данную главу.

Надежность и валидность измерения

Качество отдельных индикаторов, суммарных показателей (индексов) и шкал определяется их надежностью и валидностью.

Надежность измерения связана с его устойчивостью и воспроизводимостью. Показатель может считаться надежным в той мере, в которой полученные оценки могут быть воспроизведены на данной совокупности объектов измерения. Основной вид надежности — это *надежность-повторяемость* (или *диахронная, ретестовая надежность*). Оценка ретестовой надежности отражает результаты повторного применения одного и того же показателя (вопроса, теста) для одной и той же выборки случаев (респондентов) в разные моменты времени. Если люди отвечают на вопрос одинаково и в первом, и во втором, и во всяком последующем случае, то этот вопрос надежен. Если тест умственных способностей дает те же результаты при повторной проверке на одной и той же группе старшеклассников, то это надежный тест.

Важно отметить, что надежность инструмента измерения не отражает его точность или правильность. Скажем, если на вопрос о доходе респонденты дважды дадут совершенно идентичные ответы, завышенные на какие-нибудь 100%, мы имеем дело с абсолютно надежным, но неточным показателем. Если, в другом случае, термометр с безукоризненной надежностью показывает электрическое сопротивление кожи, у нас нет оснований говорить о правильности, адекватности измерения. *Оценка надежности-повторяемости — это корреляция между результатами повторных измерений.* В случае единичного вопроса-индикатора достаточно просто сопоставить ответы одной и той же группы людей, полученные в первом опросе, с ответами, данными двумя неделями или месяцем позднее, и вычислить коэффициент корреляции

(см. гл. 8). На практике хорошей можно считать корреляцию 0,8 и выше.

К недостаткам оценок надежности-повторяемости следует отнести, во-первых, сложность проведения повторных замеров на больших выборках, типичных для социологии. Во-вторых, истинное значение переменной также может изменяться с течением времени, например, политические симпатии могут измениться за неделю в результате каких-то скандальных разоблачений, а зарплата — существенно возрасти за месяц из-за введения обязательного индексирования в условиях инфляции. В последнем случае перед исследователем встанет трудноразрешимая задача отделить колебания, вызванные изменением истинного значения переменной, от колебаний, связанных с надежностью показателя. Поэтому так важен еще один вид надежности — *надежность-согласованность*. Оценить надежность-согласованность можно лишь в том случае, если для измерения одной и той же переменной используют *множество индикаторов*. Зато для такой оценки достаточно однократного измерения. В психологическом тестировании, например, используют батареи тестов, нацеленных на измерение одной способности или одного личностного качества. Социологи — если они располагают достаточными средствами и техническими возможностями — используют индексы и шкалы, состоящие из множества отдельных вопросов-индикаторов.

Самый простой способ оценки надежности-согласованности — это «расщепление пополам». Если, например, у нас есть 12 вопросов, предположительно измеряющих политическую активность, где каждому вопросу присвоен 1 балл, а максимальной политической активности соответствует суммарный индекс 12 баллов, то применяется следующая процедура:

1. Вопросы в случайном порядке разбиваются на два равных списка (по 6 вопросов в каждом) и предъявляются один раз группе испытуемых.

2. Далее подсчитывается *коэффициент корреляции между результатами одних и тех же испытуемых по разным «половинам» теста*: чем выше его значение, тем согласованнее оценки истинного значения политической активности, получаемые с помощью данного набора индикаторов.

Не исключено, однако, что полученная таким способом оценка надежности¹ окажется весьма чувствительной к способу «расщепления пополам»: коэффициент корреляции будет заметно меняться в зависимости от способа составления двух списков.

Еще одна элементарная процедура, позволяющая оценить надежность отдельного вопроса (высказывания, пункта шкалы), — *это проверка его корреляции с суммарным баллом*, то есть с суммарным значением индекса. Если для данной группы опрошенных коэффициент корреляции между отдельным вопросом о частоте зарубежных поездок и суммарным «индексом ксенофобии» оказался равен 0,3, то можно предположить, что названный вопрос не отражает истинного значения переменной «уровень ксенофобии» и может быть исключен из опросника². Ведь строго определенная надежность — это та *доля измеренного разброса оценок, которая относится к истинному разбросу значений измеряемой переменной* (мы пользуемся здесь менее строгим и скорее содержательным определением, поскольку пока не обсуждали необходимые статистические понятия). Очевидно, что коррелирование с суммарным баллом — это процедура, применимая для имеющих довольно простую структуру суммарных индексов и шкал (примеры вы найдете дальше в этой главе).

В любом случае важно располагать явной *моделью измерения* теоретической переменной (см. с. 128–129), так как лишь она позволит предсказать, каковы ожидаемые отношения между отдельными индикаторами и насколько применимы описанные простые методы оценки надежности. Чтобы убедиться в этом, достаточно сравнить модель с множеством эффект-индикаторов латентной переменной с моделью, включающей только причинные индикаторы (см. рис. 5.2 и 5.3). Очевидно, что эффект-индикаторы должны быть высоко согласованны и с латентной переменной-свойством, которую они призваны измерять, и друг с другом. Однако это не

¹ Мы говорим именно об *оценке* надежности, так как строго определенная надежность равна коэффициенту детерминации измеренных значений истинными значениями переменной, то есть *квадрату коэффициента корреляции*.

² Намного более подробное и снабженное соответствующими статистическими деталями описание методов оценки надежности можно найти в книге: *Аванесов В. С. Тесты в социологическом исследовании*. М.: Наука, 1982.

так уж очевидно для причинных индикаторов: скажем, и образование, и доход — важные компоненты понятия «социально-экономический статус». Однако даже если образование растёт, доход имеет право вести себя как угодно, то есть он вовсе не должен показывать непременно высокую корреляцию с образованием. Иными словами, если от надёжных эффект-индикаторов следует ожидать высокой скоррелированности друг с другом (при использовании методов «расщепления пополам» или корреляции с суммарным баллом), то для причинных индикаторов столь простой подход к оценке надёжности неприменим. Разработать подходящий метод оценки надёжности здесь можно, лишь анализируя взаимосвязи разных индикаторов и разных теоретических переменных в модели измерения. Прогнозируя ожидаемую направленность и величину этих связей, исследователь может оценить степень соответствия своих предсказаний наблюдаемым данным и сделать вывод о качестве индикатора. Самыми универсальными методами оценки надёжности эмпирических индикаторов являются *факторный анализ* и *путевой анализ*. В идеале для оценки надёжности используют несколько индикаторов (два-три) и по крайней мере две волны панели (см. с. 121–122)³.

Очень важно помнить, что понятие надёжности связано *со случайными ошибками измерения*, то есть с ошибками, которые каким систематическим образом не связаны друг с другом или какими-то систематически действующими внешними переменными (скажем, полом или возрастом респондентов). Типичные **источники ненадёжности** — это случайные несистематические факторы, связанные с колебаниями внимания респондентов; неоднозначностью формулировки вопроса, ведущей к различию в его восприятии в разных случаях; несистематическими различиями в проведении интервью; различиями в кодировании открытых вопросов или с ошибками при вводе данных. Например, если предложить даже очень опытным специалистам классифицировать сотню населённых пунктов по заранее разработанной схеме кодирования типов поселений, то, вероятнее всего, можно будет найти по крайней мере

³ Подробнее о многоиндикаторном подходе к оценке качества измерения см.: *Деятко И. Ф.* Диагностическая процедура в социологии: очерк истории и теории. М.: Наука, 1993.

несколько расхождений в получившихся классификациях. Некоторые расхождения будут связаны с наличием «предельных» случаев, не поддающихся однозначной классификации по предложенным правилам, некоторые — с механическими ошибками записи или невнимательностью. Предварительная оценка надежности вопросов социологической анкеты требует прежде всего «отбраковки» неясно сформулированных вопросов, на которые люди часто отвечают случайным образом. Столь же низка надежность вопросов, на которые респонденты попросту не способны ответить, так как не имеют никакого мнения по затронутой проблеме или ничего не знают о ней. Далеко не все опрошенные, столкнувшись с иррелевантным вопросом, честно ответят «не знаю» или «не помню». Многие дадут наугад выбранный ответ из вежливости или нежелания продемонстрировать свою неосведомленность.

Методы *увеличения надежности* нами уже обсуждались (см. гл. 4, 5). Во-первых, нужно стремиться к использованию множественных индикаторов. Когда же это невозможно, то есть существуют теоретические или практические трудности в измерении одной переменной разными способами, то следует использовать самые устоявшиеся и общепринятые показатели (например, если можно лишь один раз спросить респондента о его возрасте, то лучше всего использовать в точности такую же формулировку вопроса и те же категории ответа, какие используются в общенациональных переписях, масштабных панельных исследованиях и т. п.).

К другим методам увеличения надежности можно отнести «отсев» иррелевантных вопросов, анализ словесной формулировки вопроса, обучение и контроль интервьюеров, совершенствование методов кодирования данных и процедур ввода.

Валидность измерения, в самом общем смысле, характеризует соответствие измерения его цели. Эмпирический показатель валиден (обоснован, правилен) в той мере, в какой он действительно отражает значение той теоретической переменной, которую предполагалось измерить. Очевидно, что нет смысла говорить о валидности какого-то индикатора самого по себе. *Валидность инструмента измерения состоит в однозначности и правильности получаемых результатов относительно измеряемого свойства объектов, то есть относительно предмета измерения.* Можно сказать, что

валидность определяет «чистоту» измерения теоретического конструкта. Когда измерение является непосредственным, то есть мы можем прямо подсчитать количество эталонных единиц измеряемого свойства, и на результаты измерения влияют только случайные ошибки, надежность и валидность неразличимы, валидность инструмента измерения равна его надежности⁴. Если мы измеряем интересующее нас свойство лишь косвенно, используя какой-то индикатор, возникает различие между надежностью и валидностью. Индикатор может обладать высокой надежностью (воспроизводимостью), но при этом измерять интересующий нас социологический конструкт недостаточно «чисто». Уже на интуитивном уровне очевидно, что вполне надежный инструмент может измерять нечто другое, помимо интересующего исследователя качества (например, не столько политическую активность, сколько конформизм). Косвенное измерение обычно содержит и случайный, и неслучайный ошибочный компонент. Именно *неслучайный компонент*, включающий в себя *систематическую* (скоррелированную) *ошибку* и, реже, имеющую одну и ту же величину для каждого случая *постоянную ошибку измерения*, определяет валидность показателя. Характерными примерами систематической ошибки измерения в социологическом опросе или эксперименте являются уже упоминавшиеся эффекты «памяти», социальной желательности, установки на позитивный или негативный ответ. Они влияют не только на правильность, валидность индикатора интересующей исследователя переменной, но и на правильность и обоснованность результатов анализа данных: скоррелированная ошибка измерения может воздействовать на любые статистические показатели, в том числе на показатели взаимосвязи между переменными и на оценки значимости различий между подгруппами. Иными словами, конечным итогом использования невалидных индикаторов могут оказаться неверные содержательные выводы.

⁴ Валидность связана с надежностью так называемым *основным психометрическим соотношением*: валидность теста не превышает его надежности, то есть надежность является необходимым условием валидности и задает верхний предел ее значения (ненадежный тест не может быть валиден, а валидный тест всегда надежен). Это легко понять интуитивно: если стрелка ненадежного, испорченного спидометра вращается случайным образом, нет смысла обсуждать, насколько «чисто» он измеряет скорость.

Проблема валидности измерения — сложнейшая проблема социологической методологии. Валидное измерение — это прежде всего результат валидной *модели измерения*, то есть результат обоснованной и ясной концептуализации теоретических представлений. Здесь мы опишем лишь основные *виды валидности* и традиционные *методы валидации*, то есть установления валидности измерений.

Валидность по содержанию показывает, в какой мере избранные исследователем индикаторы отражают различные аспекты теоретического понятия. Иными словами, речь идет о *представительности* данной совокупности измерений по отношению к концептуальной структуре переменной-признака, о полноте операционализации теоретических понятий. Например, экзамен по статистике может рассматриваться как валидный инструмент измерения статистических знаний студентов, так как экзаменационные вопросы отражают содержание лекций и учебников. Однако если все вопросы относятся лишь к одному разделу прочитанного курса — скажем, к нормальному распределению, — то результаты экзамена будут отражать, например, умение студентов переводить «сырые» баллы в стандартные оценки, но ничего не скажут о знании корреляции и регрессии.

Основная процедура оценки валидности по содержанию — это *суждение эксперта*. В некоторых случаях связь между теоретическими понятиями и измеряющими их индикаторами столь ясна, что никакие специальные обоснования попросту не требуются: понятно, что термометр измеряет температуру. Здесь можно говорить об *очевидной* (иногда — *лицевой*, от англ. *face validity*) валидности показателя. Очевидная валидность тем выше, чем тождественнее понимание цели вопроса, теста или иного показателя профессионалом-социологом и неискушенным респондентом. Вопрос о частоте покупки шампуня, по всей вероятности, не содержит в себе никаких подвохов и позволяет судить именно о том типе потребительского поведения, который описан в вопросе. Однако в более сложных случаях содержательная валидность отнюдь не сводится к очевидной. Набор простых вопросов о излюбленном способе проведения досуга, предпочитаемой марке автомобиля, частоте чтения престижного журнала и т. п. может быть нацелен на измерение

«стиля жизни» респондента (в данном случае измерение позволяет отнести человека к одной из номинальных категорий стиля жизни — «выживающий», «достиженческий», «экзистенциальный», «социальный» и т. п.⁵). Судить о полноте этого набора и относительной значимости вопросов для измерения понятия «стиль жизни» могут только специалисты. Основой такого экспертного суждения является теоретическое определение, концептуализация исследовательской переменной. Обычно экспертное суждение о валидности по содержанию выносится более или менее стихийно, после публикации результатов исследования. Иногда все же удается использовать более организованные процедуры — *метод параллельных панелей* или *метод нескольких судей*⁶. В первом случае две или три последовательные панели специалистов проводят всю процедуру валидации по содержанию, то есть сравнивают существующие дефиниции, составляют список возможных индикаторов и оценивают их репрезентативность по отношению к исследуемой концептуальной области. Если сравнение индикаторов, независимо отобранных в двух и более панелях, обнаруживает множество совпадений, можно говорить о высокой содержательной валидности.

Метод нескольких судей (экспертов) полезен в тех случаях, когда переменная-признак, которую предполагается измерить, имеет многомерную структуру. Если, например, социолог разрабатывает воображаемую шкалу социально-экономического благополучия регионов, то полезно обратиться к специалистам в таких областях, как демография, социальная политика, занятость, налоговая система и т. п. Опрос экспертов позволит выявить существенные факторы, входящие в шкалу, оценить их сравнительную значимость и найти релевантные эмпирические индикаторы.

Критериальная валидность (или валидность по критерию) показывает, насколько хорошо результаты по данному тесту или индикатору согласуются с результатами измерения другого показателя, называемого *критерием*. Чаще всего критерий — это та переменная,

⁵ См., например: *Mitchell A.* The Nine American Life-Styles. N. Y.: Warner Books, 1983.

⁶ *Ghiselli E. E., Campbell J. P., Zedeck. Sh.* Measurement Theory for the Behavioral Sciences. San Francisco: W. H. Freeman and Co, 1981. P. 277–279.

которая и представляет практический интерес для исследователя, но не может быть измерена в данный момент. Например, критериальная валидность вступительных экзаменов определяется той академической успеваемостью, которую в дальнейшем продемонстрируют студенты (то есть академическая успеваемость является в данном случае критерием). Можно также предположить, что тест моральной оценки девиантного поведения для подростков обладает критериальной валидностью по отношению к реальному отклоняющемуся поведению. Индикатор, обладающий доказанной критериальной валидностью, может рассматриваться как *переменная-преддиктор*, позволяющая предсказывать индивидуальные значения переменной-критерия. Конечно, нужда в предикторе, замещающем собственно критерий, возникает лишь в тех случаях, когда оценки по критерию получить трудно, то есть речь идет о давно прошедших или еще не наступивших событиях, либо переменную-критерий трудно измерить из-за практических или этических соображений. Валидность преддиктора обычно тем выше, чем ближе он к критерию. Скажем, идеальным методом отбора курсантов авиационного училища мог бы стать пробный краткосрочный курс обучения с проверкой практических навыков управления самолетом в финале: прошедшие проверку претенденты имели бы все шансы стать настоящими профессионалами. Однако такой метод слишком дорогостоящ, и на практике используют тесты интеллекта, испытания визуально-моторной координации и другие показатели, установив предварительно их критериальную валидность. В другом случае проективный личностный тест (типа ТАТ — теста тематической апперцепции, подразумевающего составление рассказов по фотографиям с неопределенным сюжетом) позволит выявить признаки психоза либо травмирующего сексуального опыта в прошлом пациента. Полное психиатрическое обследование могло бы занять очень много времени, да и данные о плохом обращении в детстве получить довольно трудно.

К основным типам критериальной валидности относят *прогностическую, конкурентную и постдиктивную* («предсказывающую-назад») валидности.

Прогностическая критериальная валидность описывает точность, с которой значения данной переменной — обычно характеризующей

отдельного индивида или группу — могут быть предсказаны на основании текущих значений какой-то другой переменной (предиктора). Очевидно, что наилучшим показателем такой прогностической точности будет корреляция между значениями переменной-предиктора и значениями переменной-критерия для *одной и той же* выборки. Тогда — в пределах ошибки выборки — коэффициент корреляции будет равен коэффициенту прогностической валидности.

Конкурентная валидность по критерию — это степень соответствия между текущими значениями переменной-критерия и переменной-предиктора. Попросту говоря, исследователь использует результаты измерения по одному признаку для того, чтобы оценить значение другой переменной. Причина может заключаться в том, что измерение непосредственно переменной-критерия трудноосуществимо, занимает слишком много времени и т. п. Проблема конкурентной валидности весьма значима для эмпирической социологии, где многие поведенческие или установочные переменные измеряются не непосредственно, а через самоописания, ответы на анкетные вопросы, иными словами, через вербальное поведение. Например, мы можем оценить конкурентную валидность анкетного вопроса о частоте посещения дискотеки студентами-отличниками с помощью серии включенных наблюдений за реальным поведением данной группы. Уже ранние исследования конкурентной валидности фактографических вопросов в социологии показали, что даже для относительно «безобидных», несензитивных индикаторов конкурентная валидность может изменяться в самых широких пределах.

В так называемом Денверском исследовании валидности, проведенном С. Стауффером и соавторами в 1947 г., сопоставлялись данные ответов респондентов на фактографические вопросы и данные официальной статистики местных организаций. Речь шла о регистрации избирателей, участии в голосовании, взносах в городскую казну, о наличии водительских прав и читательского билета и т. п. Сопоставляя данные официальной статистики (критерий) с ответами респондентов (предиктор), исследователи обнаружили, что величина расхождений составляла от нескольких до почти 50%, в зависимости от содержания вопроса.

Следует, однако, помнить и об ограничениях, присущих объективным показателям-критериям: данные официальных документов также нередко основаны на самоотчетах и нередко подвержены ошибкам измерения.

Все же в некоторых случаях процедура конкурентной валидации имеет преимущество перед предиктивной, так как первая не требует от исследователя длительного ожидания того момента, когда можно будет измерить значения критерия. Если, например, исследователь хочет оценить критериальную валидность теста профессиональных интересов, разработанного для студентов, как предиктора успешной профессиональной карьеры, то ему не обязательно ждать десять лет, чтобы измерить значение переменной-критерия. Достаточно провести тестирование профессиональных интересов для двух «крайних» групп уже работающих специалистов — преуспевших и наименее преуспевших в профессии в данный момент времени. Высокая корреляция между тестовым баллом и успешностью работы (или статистически значимое различие тестовых баллов «крайних» групп) будет свидетельствовать о конкурентной валидности теста. Условиями, при которых выводы о конкурентной валидности индикатора могут все же оказаться ошибочными, являются *избирательное выбывание* из выборки (самоотбор) и *реактивность переменной-предиктора*. Самоотбор в нашем примере может иметь место в том случае, если среди выбывших из выборки (бывших студентов, отказавшихся от профессиональной карьеры в данной области и не охваченных по этой причине проводимым тестированием) будут сверхпредставлены высоко- либо низкомотивированные, то есть естественное выбывание будет носить неслучайной характер. В случае реактивности индикатора наши испытуемые будут отвечать на вопросы теста мотивации не так, как они отвечали будучи студентами (из-за повлиявших на них профессионального опыта, изменения социального статуса и т. п.). И все же существуют нереактивные переменные-предикторы, конкурентная валидность которых вполне поддается обоснованию. Примером могут служить такие стабильные характеристики, как коэффициент интеллекта или «фоновые» переменные (социальное происхождение, национальная принадлежность и т. п.).

Наконец, в некоторых случаях мы заинтересованы в том, чтобы установить точность, с которой мы можем оценить наличие какого-то критерия-признака или черты, присущей индивиду (группе) в прошлом. Пример использования проективного психологического теста для постдиктивного «прогноза» детских травм приведен выше. Эта ситуация обозначается как оценка *постдиктивной* валидности.

Описанные нами виды валидности существенны в тех ситуациях, когда перед исследователем стоит задача сравнить некий показатель с уже существующими или с используемым в практике критерием. Иными словами, критериальная валидность показателя — это корреляция с другим, предположительно «чистым» эмпирическим показателем. Однако существует и другой подход к валидности, где оценка индикатора основана на том, насколько хорошо его «поведение» соответствует *теоретическим ожиданиям*. Такая оценка может быть проведена лишь в рамках целостной теоретической модели, описывающей отношения между теоретическими переменными, их индикаторами, случайными и неслучайными ошибками измерения. Предположим, мы хотим проверить валидность новой шкалы групповой сплоченности. Основываясь на существующих теоретических представлениях, мы можем предположить, что большей групповой сплоченности соответствует меньшая частота открытых конфликтов и большая интенсивность коммуникаций. Сравнив различные по уровню групповой сплоченности группы и определив для них значения других двух переменных (частота конфликтов и интенсивность коммуникаций), мы увидим, насколько хорошо «ведет» себя разработанный нами показатель групповой сплоченности. Если паттерн его отношений с двумя другими переменными соответствует предсказаниям теории, то мы можем заключить, что новая шкала валидна, то есть измеряет именно ту теоретическую переменную, которая нас интересует. Этот вид валидности обычно обозначают термином «*конструктивная валидность*»⁷. (Иногда используют также обозначение «*концептуальная валидность*».)

Со статистической точки зрения абсолютная конструктивная валидность предполагает, что весь наблюдаемый разброс в значениях

⁷ См.: Cronbach L. J., Meehl P. E. Construct Validity in Psychological Tests // Psychological Bulletin. 1955. Vol. 52. № 3. P. 281–302.

показателя связан исключительно с измеряемым теоретическим конструктом. Если же часть вариации индикатора связана с другой переменной — будь то другой теоретический конструкт или систематическая ошибка измерения, — конструктивная валидность окажется меньше.

Легко заметить, что оценка конструктивной валидности предполагает какую-то связь между проверкой содержательных теоретических гипотез и проверкой качества измерения. Действительно, оценка конструктивной валидности посредством сопоставления теоретической модели «поведения» изучаемой переменной с реальными отношениями индикаторов требует включения *модели измерения* (см. выше) в более широкую *теоретическую модель*.

Предположим, мы используем некоторый суммарный показатель — индекс «демократизма политической системы», состоящий из ряда индикаторов (X_1, X_2, X_3), каждый из которых имеет свой собственный «вес» в индексе «демократизма». К таким индикаторам могут относиться наличие парламента (номинальная дихотомическая⁸ переменная), количество независимых телерадиокомпаний и т. п. Исходя из теоретической модели, мы ожидаем, что степень «демократизма» находится в обратной связи с долей ВВП, расходуемой на модернизацию вооружений. Основываясь на этой модели (см. рис. 6.1), можно проверить конструктивную валидность изобретенного нами индекса «демократизма».

Собрав необходимые данные для 10–15 национальных государств, мы можем обнаружить, что наш индекс «демократизма» невалиден, так как ожидаемое отношение между теоретическими переменными (c) не выполняется, их корреляция равна нулю. Просмотрев наши данные, мы, например, обнаружим, что в некоторых странах, считаемых за образец демократического общественного устройства, изрядную часть бюджета составляют военные расходы, тогда как некоторые деспоты из «банановых республик» вполне обходятся кремневыми ружьями. Однако вывод о невалидности нашего измерения «демократизма» верен лишь в том случае, если верны наши теоретические представления о связи демократии и

⁸ То есть принимающая лишь два возможных значения, в данном случае — «да» или «нет».

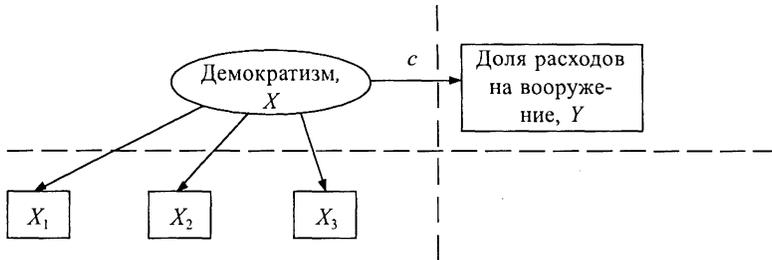


Рис. 6.1. Модель взаимосвязи для переменных «демократизм» и «доля расходов на вооружении»

пацифизма. Если же демократия и пацифизм отнюдь не связаны друг с другом, наши результаты вовсе не доказывают низкую валидность индикатора: вполне возможно, что как раз «демократию» мы измеряли правильно, но неверна была наша теоретическая гипотеза. Существует своеобразное *отношение дополнителности между собственно теоретическими моделями и моделями измерения*. Оценить качество показателей в модели измерения (см. левую часть рис. 6.1) можно, лишь приняв теоретическую модель как безусловно верную. Для оценки справедливости собственно теоретической модели (см. верхнюю часть рисунка) нужно принять предположение о конструктивной валидности индикаторов и провести **новое** исследование с **новыми** данными.

Существуют сложные статистические методы, позволяющие одновременно оценивать модель измерения и теоретическую модель (часто их называют «LISREL-методы»). Они применимы лишь к моделям с несколькими индикаторами для каждой переменной. Однако некоторые методологи полагают — на наш взгляд, справедливо, — что попытки проверить модель измерения и совокупность теоретических гипотез **на одних и тех же данных** чреваты возможностью ошибочных выводов. Если теория, предсказания которой мы используем для проверки конструктивной валидности, относительно нова и не стала еще общепринятой истиной, мы просто не сможем определить, связан ли отрицательный результат исследования с невалидностью показателя, или причиной всему — ложные теоретические представления. Кроме того, может оказаться, что мы

отберем худший из показателей, ибо именно он поддерживает неверную теорию. Поэтому проверка конструктивной валидности индикаторов и проверка теорий требуют от нас *разных исследований, множественных показателей и разных матриц данных*.

Конструирование индексов и шкал

Использование нескольких индикаторов, как было показано выше, увеличивает валидность и надежность измерения переменных. Здесь, однако, возникает новая проблема: как использовать полученные значения индикаторов для того, чтобы охарактеризовать каждый «случай» (каждого респондента, группу, страну и т. п.) *одним числовым значением*, однозначно определяющим его положение на одномерном континууме переменной-признака, для измерения которой мы использовали данный набор индикаторов. Иными словами, нужно осуществить обратный переход от набора значений эмпирических индикаторов, описывающих каждую конкретную единицу анализа, к *упорядочению всех единиц анализа по оси интересующей нас переменной*. Такое упорядочение и называется собственно *шкалой*, мерой выраженности переменной-признака, а логика перехода от набора наблюдаемых значений к шкальным значениям называется *моделью шкалирования*. Заметим сразу, что некий набор индикаторов — например, набор оценочных шкал (см. гл. 5) — может использоваться для измерения более чем одной переменной, и, следовательно, данные о наблюдаемых значениях этих индикаторов в принципе позволяют упорядочить «случай» по нескольким переменным, то есть по нескольким шкалам. Однако это уже задача *многомерного шкалирования*, мы же пока ограничимся обсуждением *одномерных шкал и индексов*.

Если вернуться к структурированной матрице данных «переменная × случай» (см. с. 117), то можно увидеть, что процедура конструирования шкалы может быть описана и как процедура «сжатия» матрицы данных, уменьшения ее размерности. Предположим, три строки нашей матрицы соответствуют переменным-индикаторам «доход», «род занятий» и «образование». Мы включили эти индикаторы в наше исследование ради того, чтобы охарактеризовать социально-экономический статус каждого респондента, то

есть расположить их от низкого статуса к высокому. Если мы вместо трех строк, соответствующих доходу, образованию и профессии, введем в нашу матрицу данных *одну* строку, отражающую положение каждого респондента на сконструированной нами шкале СЭС, размерность матрицы уменьшится. Однако сначала нам нужно решить, как объединить три значения — три строки матрицы — в одно, то есть нам нужно избрать модель шкалирования.

Пусть, скажем, три строки нашей матрицы данных — это полученные каким-то образом (тестирование, опрос экспертов и т. п.) оценки «жизнерадостности», «энергичности» и «независимости». Исследователь предполагает, что эти три индикатора могут быть использованы для измерения важной для его теории переменной «сила Я». Все, что ему нужно сделать, — это решить, как перевести оценки в строках 1–3 в оценки «силы Я» (см. рис. 6.2).

Самый простой и очевидный способ — это суммировать для каждого индивида оценки по каждому индикатору. Получившийся суммарный балл будет отражать индивидуальные различия в «силе Я», так как позволит упорядочить всех респондентов от минимального к максимальному значению этой переменной (в нашем примере — от 0 до 6 баллов). Еще одно преимущество суммирования — увеличение разброса индивидуальных значений. Действительно, максимально возможное различие по первичным индикаторам составляло 2 балла (от 0 до 2). В суммарном показателе разница между индивидуальными значениями может составить 6 баллов. Сле-

Переменные-индикаторы	Субъект («случай»)			
	Л. М.	Ф. Ж.	К. П.	...
1. Жизнерадостность	2	0	2	...
2. Энергичность	2	2	2	...
3. Независимость	0	2	2	...



«Сила Я»?

Рис. 6.2. Фрагмент матрицы данных «переменные × субъекты»

довательно, суммарный балл — это более «чуткий» и надежный инструмент для упорядочения и может быть назван шкалой в смысле определения, данного нами выше. Однако в социологии суммарные показатели чаще называют *индексами*, чтобы подчеркнуть их единственное важное отличие от «больших» шкал. Индекс позволяет эффективно «свернуть» информацию, содержащуюся в исходных индикаторах (вопросах, пунктах, тестах), однако от суммарного балла нельзя вернуться к исходной матрице, точнее, к тому *паттерну ответов*, который стоит за данным значением индекса. Если сформулировать это корректнее, *индекс не позволяет учитывать различия в структуре ответов респондентов*. Если снова обратиться к рис. 6.2, то можно заметить, что субъекты Л. М. и Ф. Ж. имеют одинаковый суммарный балл, равный 4 (достаточно высокое значение!). Но можно ли считать несущественным то обстоятельство, что у Л. М. нулевой уровень независимости, а Ф. Ж. получил тот же суммарный балл из-за недостатка оптимизма? Предположим, даже довольно мрачный человек может обладать значительной «силой Я», но следует ли считать столь же «сильным» того, кто легко поддается давлению окружения?

В принципе, индексы безусловно применимы в тех случаях, когда модель измерения (см. выше) предполагает, что некая латентная, то есть не измеряемая непосредственно переменная, может быть измерена с помощью *совокупности качественно однородных показателей*. Во многих случаях различия в значимости, важности отдельных индикаторов можно учесть с помощью «взвешивания», пересчета значений с учетом «веса» каждого индикатора в латентной переменной. Так, в примере с «силой Я» можно домножить все индивидуальные значения в строке «независимость» на 2, если принять предположение о том, что независимость влияет на латентную переменную с двукратным эффектом.

Экономисты часто используют индексы розничных цен, отражающие динамику стоимости жизни. При этом разные товарные группы, например имеющие неодинаковое значение в потребительском бюджете, — как, скажем, хлеб и деликатесы — учитываются с разными весовыми коэффициентами. Но и в этом случае индекс остается несовершенным типом шкалы: эмпирическая информация здесь используется лишь для шкалирования различий между

субъектами (или другими единицами анализа), но не для шкалирования различий между пунктами-ответами (эмпирическими индикаторами). Используя «взвешивание», мы вводим *априорные* ограничения на упорядочение входящих в индекс индикаторов, не зависящие от данных наблюдения.

Своеобразным переходом между моделью суммарного балла (индекса) и основными моделями шкалирования является *шкала Р. Ликерта* (Лайкерта). Исходным материалом для ее построения служат оценочные шкалы согласия–несогласия с суждениями, которые выражают более или менее «благожелательную» установку (см. раздел «Выбор формата для ответов» в гл. 5). Количество категорий ответа — «согласен», «совершенно согласен» и т. п. — обычно варьирует от двух до семи. Респондент получает балл по каждому суждению в зависимости от избранного им ответа. Присуждаемый данному ответу балл, в свою очередь, определяется «благожелательностью» ответа по отношению к измеряемой установке (интенсивностью согласия с суждением), то есть ответы *также упорядочены* на одномерном континууме (от крайне негативной установки к крайне позитивной). Баллы, полученные за каждый ответ, суммируются. Суммарный балл, полученный индивидуумом, характеризует уже его собственное положение на установочном континууме (например, «консерватор», «умеренный консерватор», «умеренный либерал», «либерал»). Отметим сразу, что эта же модель шкалирования может использоваться и для измерения мотивации или осведомленности (соответственно, респондента просят оценить степень важности какого-то объекта или сказать, верно или неверно определенное утверждение). Для отбора списка суждений, составляющих шкалу Ликерта, исходный список высказываний предъявляют репрезентативной выборке респондентов (так называемой «выборке стандартизации»). В окончательный список попадают те высказывания, для которых были получены высокие оценки надежности-согласованности и валидности. Обычно используют описанные нами ранее методы оценки надежности и валидности (коррелирование с суммарным баллом, сравнение «крайних групп» и т. п.).

Приведем в качестве примера некоторые высказывания «Теста для измерения художественно-эстетической потребности молоде-

жи»⁹ (в скобках дан ключ к каждому высказыванию, показывающий, за какой ответ присуждается балл):

1. Думаю, что вполне можно обойтись без общения с произведениями искусства (неверно).
2. Я не люблю стихов (неверно).
3. Я коллекционирую записи классической музыки (верно).
4.

Шкалирование по описанной модели дает *ординальный* уровень измерения.

Шкалы социальной дистанции Э. Богардуса — старейшая модель социологического шкалирования, не утратившая, однако, своей популярности. Исследователь разрабатывает совокупность вопросов, отражающих различную степень близости отношений с определенной социальной или этнической группой, например:

1. Согласны ли вы, чтобы хорваты жили с вами в одном городе?
2. Согласны ли вы жить по соседству с хорватами?
3. Согласны ли вы работать в одном отделе (учреждении) с хорватом?
4. Позволите ли вы своей дочери выйти замуж за хорвата?

Предполагается, что согласие с каждым последующим утверждением отражает переход к очередной градации ординальной шкалы установок — от меньшей близости к большей. Существенным требованием к избранной совокупности вопросов является их содержательная валидность, иными словами, здесь необходимы экспертные процедуры, описанные выше (см. с. 173–174). Важно также убедиться в обоснованности предположения об одномерности шкалируемой переменной. Если в данных, полученных при использовании шкал социальной дистанции, встречаются «нелогичные» (так называемые нешкалируемые) индивидуальные паттерны ответов, причиной чаще всего бывает влияние другой переменной. Примером нешкалируемого паттерна ответов может служить ситуация, когда респондент, отрицательно ответивший на «слабые» вопросы,

⁹ См.: Аванесов В. С. Указ. соч. М.: Наука, 1982. С. 57–60.

неожиданно соглашается с более «сильными», предполагающими высокую степень близости (среди специалистов по социологическим методам имеет хождение соответствующая шутка: если человек, не желающий жить в одном городе с черными, согласен выдать свою дочь замуж за черного, это не ошибка измерения — просто он одинаково ненавидит негров и собственную дочь).

Шкала равнокажущихся интервалов Л. Терстоуна позволяет достичь более высокого уровня измерения установок, чем ординальный. Она представляет собой целый класс методов интервального шкалирования и будет рассмотрена здесь в качестве наиболее простого примера¹⁰.

Первая шкала равнокажущихся интервалов была описана в работе 1929 года и предназначалась для измерения установок по отношению к церкви как социальному институту¹¹. Этой работой мы воспользуемся для того, чтобы проиллюстрировать основные этапы предложенной Терстоуном процедуры.

Шкала Терстоуна позволяет расположить и суждения, и индивидов вдоль одномерного континуума установки, полюсам которого соответствует крайне благожелательное и крайне негативное отношение к объекту установки (церкви, партии, прогрессивному налогообложению или чему-либо еще). Шкальный балл суждения или индивида отражает *степень* этой благожелательности или неблагоприятности.

На первом этапе исследователь составляет максимально широкий список суждений (высказываний), выражающих интересующую его установку. Так, Терстоун собирал мнения коллег, студентов, высказывания из публикаций, касающихся церкви. Здесь уместны также интервьюирование, использование открытых вопросов («Что вы

¹⁰ Тех, кто хочет узнать больше о разных методах шкалирования и готов преодолеть трудности, связанные с использованием некоторых статистических понятий, мы можем отослать к обзорным работам, содержащим также необходимую библиографию: *Грин Б. Ф.* Измерение установки // Математические методы в современной буржуазной социологии. М.: Прогресс, 1966. С. 227–228; *Девятко И. Ф.* Указ. соч.

¹¹ *Thurstone L. L., Chave E. F.* The Measurement of Attitudes. A Psychophysical Method and Some Experiments with a Scale for Measuring Attitude toward Church. 7th ed. Chicago: University of Chicago Press, [1929] 1964.

думаете о...?»), групповая дискуссия и т. п. Собранные суждения были подвергнуты *первичному отбору*. Исследователи отсеяли те высказывания, которые не удовлетворяли обычным требованиям к конструированию вопросов — двусмысленные, слишком длинные, содержащие специальные термины и т. п. (см. гл. 5). При первичном отборе суждений для шкалы Терстоуна используют и некоторые специальные критерии:

1. Исключаются суждения, относящиеся скорее к прошлому, чем к настоящему (например, «в средневековой церкви играла важную роль в общественной жизни»).

2. Исключаются суждения, описывающие факты, а не мнения и отношения. Конечно, далеко не всегда можно отделить высказывания, описывающие фактическое положение дел, от прочих. Скажем, слова «Бог любит нас всех» — факт для верующего, хотя другие люди могут усмотреть в них определенное отношение к религии. В практических целях вполне достаточно руководствоваться следующим критерием для выявления фактических суждений, подлежащих устранению из шкалы Терстоуна: фактом является любое высказывание, для установления истинности которого могут быть использованы какие-то «посюсторонние» процедуры верификации.

3. Исключаются также суждения, содержащие слова «все», «всегда», «никто», «никогда», так как этим словам люди обычно придают различный смысл, что затрудняет интерпретацию.

В результате исходный список из 350–400 суждений сокращается до 100–120. Следующим этапом является «судейская» процедура, позволяющая определить шкальное значение для каждого суждения и провести среди них окончательный отбор. Терстоун предложил разделить гипотетический континуум благожелательно–неблагожелательного отношения к церкви на 11 категорий (от «А» до «К»), разделенных субъективно равными интервалами. Требование субъективного равенства интервалов между градациями весьма существенно для построения шкалы Терстоуна, и обычно его специально подчеркивают в инструкции для «судей» (например: «Представьте, что карточки с буквами от “А” до “К” представляют расположенные на равном расстоянии градации шкалы, так что градация “А” соответствует максимально благожелательное отношение

к X (объекту установки), а «К» — максимально неблагоприятное, негативное отношение». Каждое из утверждений списка печатается на отдельной карточке, которые и раздаются «судьям» (в конструировании шкалы установок по отношению к церкви участвовало 300 таких экспертов). Задача «судей» заключается в том, чтобы разложить все 100–120 суждений по 11 рубрикам *соответственно степени выраженного в них благожелательного или неблагоприятного отношения к объекту установки*.

Подчеркнем, что «судей» не просят высказать их собственное мнение, они должны лишь рассортировать высказывания.

Шкальное значение (балл) каждого из высказываний определяется распределением оценок «судей», поэтому началом следующего этапа (собственно *построения шкалы*) является подсчет процента экспертов, положивших высказывание в определенную стопку. Далее подсчитывается суммарный (кумулятивный) процент «судей», отнесших суждение к данной градации и предшествующим градациям. Терстоун присваивал использовавшимся градациям числовые значения от 1 (градация «А», максимально благожелательное отношение к церкви) до 11 (градация «К»). Проиллюстрируем дальнейшее на примере гипотетического суждения N , данные для которого представлены в таблице 6.1.

Распределение кумулятивных (накопленных) процентов позволяет вычислить значения медианы и междуквартильного размаха. Медиана, или процентиль 50 в распределении накопленных частот, — это такое значение на шкале «А»–«К», относительно которого половина «судей» дала большие, а другая половина — мень-

Таблица 6.1

Распределение «судейских» оценок для суждения N

Распределение ответов	Градация (числовое значение)										
	«А» (1)	«В» (2)	«С» (3)	«D» (4)	«Е» (5)	«F» (6)	«G» (7)	«H» (8)	«I» (9)	«J» (10)	«K» (11)
Процент судей, отнесших суждение к данной градации	1	2	2	1	3	33	34	12	7	3	2
Кумулятивный процент	1	3	5	6	9	42	76	88	95	98	100

шие оценки данного утверждения¹². Медиана, таким образом, делит пополам упорядоченное множество значений признака. Вычислить медиану мы можем по следующей формуле:

$$Md = \left(\begin{array}{c} \text{фактическая} \\ \text{нижняя граница} \\ \text{интервала медианы} \end{array} \right) + \left(\begin{array}{c} \text{ширина} \\ \text{интервала} \end{array} \right) \times \frac{50 - \left(\begin{array}{c} \text{кумулятивный процент} \\ \text{для нижней границы интервала} \\ \text{медианы} \end{array} \right)}{\left(\begin{array}{c} \text{процент, соответствующий} \\ \text{интервалу медианы} \end{array} \right)}.$$

В методе Терстоуна ширина интервала между соседними численными градациями по определению равна 1 (равнокажущиеся интервалы). В используемом нами примере границами интервала, где расположена медиана (перцентиль 50), являются градации «F» и «G» (см. табл. 6.1). Фактической нижней границей интервала медианы будет значение 6,5¹³. Отсюда:

$$Md_n = 6,5 + 1 \times \frac{50 - 42}{34} \approx 6,7.$$

Значение медианы и принимается за шкальный балл («цену») суждения. Для гипотетического суждения *N* в нашем примере он оказался равен 6,7. (В принципе более простым является графический метод нахождения медианы. Для этого на миллиметровой бумаге строится кривая накопленных процентов — огива,

¹² Более полное представление о медиане как мере центральной тенденции и межквартильном размахе как мере разброса численных значений признака при необходимости можно получить из любого учебника по основам прикладной статистики. См., например: *Гласс Дж., Стэнли Дж.* Статистические методы в педагогике и психологии. М.: Прогресс, 1976. Гл. 4, 5. См. также гл. 8 настоящего издания.

¹³ Фактические границы интервалов при построении гистограмм или полигонов распределения частот задают таким образом, чтобы численное значение выпадающей на данный интервал градации шкалы оказалось — с учетом принятого способа округления — в середине интервала (так, для градации 3 фактические границы могут составить 2,5 и 3,5). Более подробные сведения о правилах группирования значений переменной и графического представления полученного распределения частот можно найти в соответствующих разделах любого учебника по основам прикладной статистики.

позволяющая легко найти числовое значение, соответствующее процентилю 50.)

Ясно, однако, что не все суждения, получившие оценку «судей», в равной мере пригодны для шкалы: некоторые из суждений получают весьма согласованные и единодушные оценки экспертов, тогда как другие вызовут разнобой во мнениях. Для оценки *внутренней согласованности* отдельных высказываний шкалы Терстоун применил меру разброса судейских оценок — междуквартильный размах. (Здесь снова вместо распределения абсолютных частот экспертных оценок используется распределение процентилей, то есть накопленные частоты выражают в кумулятивных процентах, что позволяет сравнивать выборки разного объема.) Междуквартильный размах — это расстояние между первым и третьим квартилем распределения. Первый квартиль (Q_1) задается точкой на оси, до которой лежит 25% полученных оценок суждения, а третий (Q_3) — точкой, выше которой лежит 25% оценок. (Легко видеть, что второму квартилю соответствует медиана.) Для вычисления междуквартильного размаха ($Q_3 - Q_1$) сначала устанавливаются значения, соответствующие первому и третьему квартилям распределения. Для этого используются формулы, аналогичные формуле для медианы, с соответствующими поправками: берется фактическая нижняя граница интервала соответствующего квартиля, кумулятивный процент для нижней границы интервала данного квартиля и т. д. Так, для первого квартиля формула подсчета такова:

$$Q_1 = \left(\begin{array}{l} \text{фактическая нижняя} \\ \text{граница интервала} \\ \text{первого квартиля} \end{array} \right) + \left(\begin{array}{l} \text{ширина} \\ \text{интервала} \end{array} \right) \times \left(\begin{array}{l} 50 - \left(\begin{array}{l} \text{кумулятивный процент} \\ \text{для нижней границы} \\ \text{интервала первого квартиля} \end{array} \right) \\ \left(\begin{array}{l} \text{процент, соответствующий} \\ \text{интервалу в котором находится} \\ \text{первый квартиль} \end{array} \right) \end{array} \right).$$

Для нашего примера с суждением N :

$$Q_1 = 5,5 + 1 \times \frac{25 - 9}{33} \approx 6.$$

Читатель может самостоятельно выписать аналогичную формулу для третьего квартиля (75 перцентиль), произвести необходи-

мые подсчеты и убедиться, что для вымышленного суждения N междуквартильный размах ($Q_3 - Q_1$) составит 1,7.

Те суждения, для которых разброс оценок, измеренный через междуквартильный размах, оказывается слишком велик, исключаются из шкалы Терстоуна. Предполагается, что высказывание, получившее столь разные оценки, воспринимается слишком неоднозначно. Так, Терстоун исключил из первоначально предъявленного «судьям» списка 90 высказываний из 130. В результирующей шкале оставляют одно-два высказывания для каждого деления шкалы, чтобы все градации предполагаемого установочного континуума оказались в равной мере представлены.

Если получившуюся шкалу предъявить теперь группе респондентов, то *индивидуальным баллом* каждого субъекта, выражающим меру «благожелательности» его установки, можно считать медиану (или средний балл) всех суждений, с которыми он согласился.

Многие критики шкалы Терстоуна указывали на возможность влияния на получаемые результаты характеристик «судейской» группы и широты диапазона предлагаемых суждений¹⁴. Все же существуют веские основания считать, что такая шкала обладает вполне удовлетворительной воспроизводимостью и в среднем диапазоне дает уровень измерения, превосходящий ординальный (является так называемой шкалой разностей). Удаление или прибавление пункта шкалы не меняет шкальных значений других пунктов-высказываний. Приведем некоторые примеры высказываний, включенных Терстоуном в окончательный вариант шкалы установок по отношению к церкви (в скобках указан шкальный балл суждения):

- «Я думаю, что церковь — это наиважнейший социальный институт в современной Америке» (0,2);
- «Когда я нахожусь в храме, мне доставляет удовольствие наблюдать за службой, особенно, если при этом звучит хорошая музыка» (4,0);

¹⁴ Дополнительные сведения о критике шкалы Терстоуна, а также о других методах отсева иррелевантных данных см., в частности: *Клигер С. А., Косолапов М. С., Толстова Ю. Н.* Шкалирование при сборе и анализе социологической информации. М.: Наука, 1978. Гл. 3; *Девятко И. Ф.* Указ. соч.

- «Я ощущаю потребность в какой-то религии, но не нахожу того, что мне нужно, ни в одной из существующих церквей» (6,1);
- «С моей точки зрения, церковь безнадежно устарела» (9,1).

В основе шкалы Гутмана также лежит идея кумулятивности: одни высказывания-пункты имеют большую «нагрузку» на одномерном континууме шкалируемой переменной-признака, другие — меньшую. Модель шкалирования, предложенная Гутманом, подразумевает, что в идеальном случае респондент, согласившийся с более «нагруженным» пунктом, согласится и со всеми менее «нагруженными». Таким образом, знание максимального шкального балла респондента позволяет полностью воссоздать паттерн его ответов. Шкалируемая переменная-признак не обязательно является установкой, она может характеризовать поведение: одна из первых гутмановских шкал, например, содержала пункты, описывающие симптомы реактивного невроза, — расстройства сна, тошнота, страх и т. п.¹⁵ В предложенной Гутманом процедуре обычно используется совокупность дихотомических вопросов, то есть вопросов, на которые могут быть даны лишь два ответа: «да» или «нет», «согласен» или «не согласен». Совокупность вопросов или утверждений, составляющих гутмановскую шкалу, должна обладать одним существенным свойством: *вероятность положительного ответа монотонно возрастает с ростом значения шкалируемой (латентной) переменной*. Отсюда следует, что субъекты, имеющие больший шкальный балл, то есть большее значение латентной переменной, с большей вероятностью дают положительный ответ на каждый отдельный вопрос¹⁶. Это условие легко проиллюстрировать на примере следующих вопросов о росте (для простоты изложения предположим, что ошибка измерения отсутствует, то есть все респонденты знают и точно сообщают свой истинный рост):

¹⁵ Примеры шкал, разработанных Л. Гутманом и его сотрудниками, можно найти в классической работе: *Stouffer S. A. et al. Measurement and Prediction*. N. Y.: John Wiley & Sons, [1950] 1966.

¹⁶ В более строгой формулировке: логическая форма вопроса (суждения) должна предполагать, что вероятность принятия суждения является монотонно возрастающей (или убывающей) функцией шкальной позиции респондента.

1. Верно ли, что ваш рост превышает 1,50 м?
2. Верно ли, что ваш рост превышает 1,60 м?
3. Верно ли, что ваш рост превышает 1,70 м?

Эти вопросы образуют идеальную гутмановскую шкалу: если ошибка измерения отсутствует, любой респондент, ответивший положительно на вопрос 3, даст положительный ответ и на вопросы 2 и 1. Вообще, зная максимальный балл респондента, мы можем полностью воссоздать его паттерн ответов. Для вопросов о росте все возможные паттерны ответов (*шкальные типы*) приведены в таблице 6.2.

Если респондентов и вопросы расположить на одной шкале латентной переменной (в данном случае «роста»), то станет очевидным, что респонденты реагируют на вопросы в зависимости от своего ранга (положения) на этой же шкале: респондент данного роста, позитивно прореагировавший на некий вопрос-пункт, будет также позитивно реагировать на все пункты, имеющие более низкий ранг (в нашем примере — на все более «низкорослые» вопросы-пункты). Скажем, для примера с ростом совместное упорядочение вопросов и респондентов на гутмановской шкале могло бы выглядеть, как на рис. 6.3.

Легко видеть, что в нашем примере самыми «популярными» (имеющими наибольшую частоту положительных ответов) будут менее «нагруженные» по шкалируемой переменной вопросы. С другой стороны, большее число положительных ответов будет у тех респондентов, для которых значение шкалируемой переменной выше.

Т а б л и ц а 6.2

Ответы на вопросы о росте для четырех гипотетических респондентов

Респондент	Номер вопроса		
	1	2	3
А	+	+	+
Б	+	+	–
В	+	–	–
Г	–	–	–

Для того чтобы определить, является ли данная совокупность вопросов шкалируемой, то есть дает ли она хорошее приближение к идеальной модели шкалирования по Гутману, нужно, во-первых, убедиться, что вопросы по своему смыслу могут соответствовать вышеописанному условию монотонного возрастания вероятности положительного ответа. Так, например, рассмотренные выше вопросы шкалы социальной дистанции Богардуса теоретически могут составить гутмановскую шкалу. Однако так называемые «точечные» вопросы — например, «верно ли, что ваш рост равен 1,65 м?» — наверняка не могут быть использованы в рамках рассматриваемой шкальной модели. Дальнейшая процедура построения гутмановской шкалы заключается в *проверке соответствия* реальных данных модели совершенной шкалы Гутмана, и ее лучше рассмотреть на конкретном примере.

Допустим, некий социолог пытается измерить переменную «благопристойность», которую он почему-то связывает со стремлением следовать нормативным предписаниям, регулирующим публичное (то есть безличное) социальное общение. Среди использованных им вопросов есть три следующих:

1. Вы обычно стучите в дверь кабинета или комнаты, прежде чем войти (верно, неверно).
2. В публичных обсуждениях или дискуссиях вы высказываете свое мнение лишь тогда, когда вас прямо об этом попросят (верно, неверно).
3. Вы предпочтете промолчать, если вам покажется, что ваше справедливое и существенное замечание может вызвать раздражение у окружающих (верно, неверно).

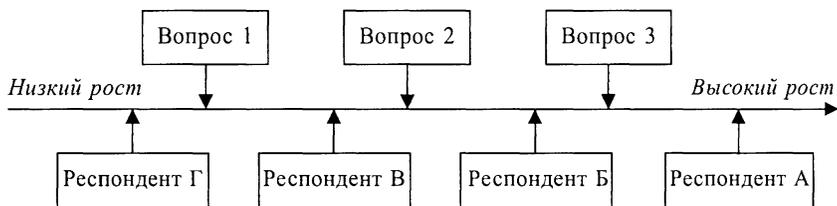


Рис. 6.3. Шкала Гутмана для трех вопросов о росте и четырех респондентов

Конечно, можно предположить, что некоторые из изобретенных исследователем вопросов отражают скорее «социальную тревожность» или «конформизм», но в данном случае мы будем руководствоваться тем, что идеальных индикаторов не бывает: бывают индикаторы, дающие лучшее или худшее приближение к идеальной шкале для реальных данных.

Итак, социологу предстоит проверить, насколько полученное им эмпирическое распределение ответов соответствует тем теоретически возможным паттернам ответов на три вопроса, которые в табл. 6.2 образуют безупречный «параллелограмм», характерный для идеальной гутмановской шкалы. Для случая трех вопросов возможны четыре «правильных» паттерна ответов, обозначаемых обычно как *шкальные типы ответов*:

1. + + +
2. + + -
3. + - -
4. - - -

Предположим, что наш социолог получил следующую картину распределения шкальных типов (см. табл. 6.3).

Судя по таблице 6.3, априорное упорядочение вопросов совпало с реальным: самый «легкий» первый вопрос оказался и самым популярным (см. выше), тогда как на самый «тяжелый» вопрос шкалы положительно ответили лишь 30 опрошенных: нежелание высказывать свою точку зрения требует значительно большего количества «благопристойности», чем привычка стучать в дверь. Если

Т а б л и ц а 6.3

Распределение ответов для шкальных типов

«+» — «верно», «-» — «неверно»

Паттерны ответов	Вопрос 1 («стук в дверь»)	Вопрос 2 («публичная дискуссия»)	Вопрос 3 («опасение вызвать раздражение»)	Число случаев, <i>N</i>
1	+	+	+	30
2	+	+	-	50
3	+	-	-	45
4	-	-	-	10
				Всего 135

бы использованный нами исходный порядок вопросов не совпал с их реальным ранжированием по числу позитивных ответов, это само по себе не доказывало бы «нешкалируемости» данной совокупности пунктов: для того чтобы получить столь же красивую «гутмановскую» картину распределения ответов, как в таблице 6.2, было бы достаточно просто переставить столбцы таблицы так, чтобы первым оказался самый популярный вопрос с наибольшим числом положительных ответов и т. д. (Упорядоченную таким образом таблицу обычно называют *шкалограммной матрицей*, или *шкалограммой*.)

Реальной проблемой в нашем примере, как и в большинстве случаев построения гутмановской шкалы, стало наличие так называемых *нешкальных типов*, то есть таких паттернов ответа, которые попросту не укладываются в логику одномерной модели с монотонно возрастающей вероятностью ответа. Примером «нешкального» паттерна мог бы быть положительный ответ на третий вопрос при отрицательных ответах на первые два вопроса (– – +). То обстоятельство, что некий респондент, бесцеремонно входящий в чужую дверь без стука, боится открыто выразить свое мнение, может быть и случайной ошибкой, и результатом влияния какой-то посторонней переменной: возможно, отвечая на третий вопрос, этот человек думал не о хороших манерах, а о том, что высказывать свое мнение открыто в его привычной среде «невыгодно», недальновидно и т. п. Для того чтобы проверить шкальную гипотезу о том, что данная совокупность вопросов дает хорошее приближение к гутмановской шкале, нам следует трактовать «нешкальные» типы ответа как ошибки и оценить, насколько велико отклонение от идеальной модели. Пусть наш исследователь получил следующее распределение «нешкальных» типов (см. табл. 6.4).

Разумно предположить, что «нешкальный» тип – – + можно отнести к шкальному типу – – – с одной ошибкой. Второй «нешкальный» паттерн ответа – + + можно рассматривать как отклонение от шкального типа + + + также с одной ошибкой (если бы мы отнесли этот «нешкальный» паттерн к типу – – –, то ошибок было бы две, а не одна). Существуют разные способы оценки приемлемости наблюдаемых отклонений от *совершенной шкалы*, содержащей лишь шкальные паттерны ответа. Здесь мы воспользуемся самым про-

Таблица 6.4

Распределение ответов для «нешкальных» типов

Паттерны ответов	Вопрос 1 («стук в дверь»)	Вопрос 2 («публичная дискуссия»)	Вопрос 3 («опасение вызвать раздражение»)	Число случаев, <i>N</i>
1	–	–	+	3
2	–	+	–	5
				Всего 8

стым и грубым, рассчитав коэффициент воспроизводимости шкалы *Rep* (от англ. reproducibility) по следующей формуле:

$$Rep = 1 - \frac{\text{число ошибок}}{\text{общее число ответов}}.$$

В нашем примере мы, основываясь на идеальной модели шкалы, можем воспроизвести (предсказать) по три ответа для 143 респондентов. Всего мы сделаем 429 предсказаний для отдельных ответов. Из них 8 ответов окажутся ошибочными (каждая ошибка будет отличаться от ожидаемого ответа только на 1 балл). Коэффициент воспроизводимости составит, таким образом, 0,98 (или 98%).

На практике принято считать приемлемым любое значение коэффициента воспроизводимости, которое превышает 0,90 (90%). Очевидно, что 100%-й воспроизводимостью может обладать лишь совершенная гутмановская шкала.

Если полученное значение коэффициента воспроизводимости превосходит заданный порог, данная совокупность вопросов может использоваться в качестве шкалы Гутмана. При этом вопросам присваиваются шкальные значения, отражающие их ранжирование по шкале (скажем, 1, 2 и 3), так что самый «легкий» вопрос получает самый низкий балл. Респонденты получают индивидуальный балл, соответствующий их шкальным типам (число положительных ответов либо суммарный балл).

Следует помнить о том, что полученная шкала отражает наличие определенной упорядоченности в той матрице реальных данных, для которых проверялась гутмановская модель. Иными словами, вывод о том, что данная совокупность вопросов составляет

шкалу Гутмана, верен для данной выборки и для данной серии наблюдений. Перенос шкалы с одной популяции на другую требует новых данных и нового обоснования.

Мы рассмотрели лишь некоторые, относительно простые методы конструирования индексов и шкал в социологии. Проанализированные нами примеры подтверждают полезность шкал для повышения качества социологического измерения (то есть его надежности и валидности) и для экономного представления эмпирической информации, получаемой в ходе исследования. Наконец, анализ моделей измерения, лежащих в основании любой шкалы, часто помогает прояснить природу теоретических понятий и взаимосвязей между ними. Еще одним шагом к содержательным и основанным на реальных эмпирических наблюдениях выводам является анализ данных.

Основам анализа данных посвящена гл. 8.

Дополнительная литература

Аванесов В. С. Тесты в социологическом исследовании. М.: Наука, 1982.

Грин Б. Ф. Измерение установки // Математические методы в современной буржуазной социологии. М.: Прогресс, 1966.

Девятко И. Ф. Диагностическая процедура в социологии: очерк истории и теории. М.: Наука, 1993.

Клигер С. А., Косолапов М. С., Толстова Ю. Н. Шкалирование при сборе и анализе социологической информации. М.: Наука, 1978.

Осипов Г. В., Андреев Э. П. Методы измерения в социологии. М.: Наука, 1977.

Толстова Ю. Н. Логика математического анализа социологических данных. М.: Наука, 1991.

Ядов В. А. Социологическое исследование: методология, программа, методы. 2-е изд. М.: Наука, 1987. Гл. 3.

ГЛАВА 7. ПОСТРОЕНИЕ ВЫБОРКИ СОЦИОЛОГИЧЕСКОГО ИССЛЕДОВАНИЯ

Выборочный метод: определение и истоки

Задача построения выборки возникает всякий раз, когда необходимо собрать информацию о некоторой группе или большой совокупности людей. Выборку в той или иной форме используют в ориентированных на «жесткие» статистические методы опросах, в исследованиях политических и культурных элит и даже при отборе «случаев» для включенного наблюдения и качественного анализа.

Статистические (или квазистатистические) обследования населения и ресурсов, судя по всему, зародились одновременно с первыми формами централизованной социальной и политической организации: развитые аграрные общества и древние города-государства нуждались в такой информации и использовали ее при решении разнообразнейших управленческих задач — от фискальной политики до строительства общественных бань. Эти обследования иногда принимали форму сплошных переписей населения. (Об одной такой переписи, имевшей, правда, самые печальные последствия, рассказывает нам книга пророка Самуила: когда царь Давид (X в. до н. э.) осуществил перепись населения древнего Израиля, в стране разразилась страшная эпидемия (2 Цар. 24). Однако значительно чаще приходилось довольствоваться сведениями о какой-то части совокупности: об урожайности судили по пробному обмолоту, о партии товара — по образцу, а о прихожанах — по их духовному наставнику.

Выборка — это подмножество заданной совокупности (популяции), позволяющее делать более или менее точные выводы относительно совокупности в целом. Зачем нужно строить выборки? Прежде всего, из практических соображений, так как выборка *экономит силы и средства* исследователей. Проведение полномасштабной переписи или сплошного опроса населения требует значительных

финансовых и трудовых затрат, которые к тому же могут пропасть впустую в случае, если в разработке методики исследования были допущены принципиальные просчеты.

Другая причина заинтересованности в выборках связана с тем, что выборочная процедура *представляет собой удобную и экономичную форму индуктивного вывода*¹. Третья причина заключается в том, что эта процедура реализует фундаментальный *принцип рандомизации*, то есть случайного отбора.

Представление о том, что отбор наблюдений должен носить случайный, непредумышленный характер, в общем, соответствует нашему интуитивному знанию об условиях вынесения объективного и непредвзятого суждения. Однако строгая, то есть математико-статистическая, теория случайной выборки вплоть до конца XIX — начала XX в. не пользовалась популярностью в среде профессиональных статистиков. Многим исследователям казалось, что в основе отбора должна лежать не «игра случая», а поиск типичных, характерных наблюдений. Это убеждение препятствовало применению в массовых обследованиях методов теории вероятности, достигшей высочайшего уровня развития уже в XVIII — первой половине XIX в. Применимость выборочного метода для изучения случайно распределенных признаков, например дохода или размера семьи, была впервые обоснована в работах норвежца А. Киэра, англичан А. Боули и К. Пирсона, а также русского статистика А. И. Чупрова².

Следующим принципиально важным шагом в развитии выборочного метода стала осуществленная Р. Фишером разработка *техники рандомизации* в эксперименте и выборочном наблюдении³. О роли рандомизации в планировании эксперимента говорится в гл. 4. Что же касается выборочного обследования, то оно часто используется как «замена» экспериментального метода. Нельзя про-

¹ Напомним, что под индуктивным выводом обычно понимают рассуждение по схеме «от частных наблюдений — к общей эмпирической закономерности».

² Более детальные сведения о развитии выборочного метода можно найти, в частности, в интересной и доступной книге: Дружинин Н. К. Выборочное наблюдение и эксперимент. М.: Статистика, 1979.

³ См.: Fisher R. A. The Design of Experiment. 3rd ed. L.: Oliver & Boyd, 1942.

вести эксперимент, в котором людям в случайном порядке присваиваются определенные значения переменных «пол» или «цвет кожи». Однако применение выборочного метода и статистического анализа, как мы увидим в дальнейшем, позволяет справляться с этими ограничениями и делать выводы о взаимосвязях между самыми разными переменными, включая вышеупомянутые. Но для того чтобы такие выводы были обоснованы, нужно устранить любое систематическое влияние «посторонних», смешивающих факторов на изучаемые переменные. Единственным средством для достижения этой цели является абсолютно случайный характер отбора наблюдений. Лишь *равенство шансов попадания в выборку для каждого наблюдения*, то есть отбор «наугад», гарантирует отсутствие намеренных или ненамеренных искажений. Пусть, например, в ходе опроса мы изучаем влияние пола и рода занятий респондента на его отношение к планированию семьи и ограничению рождаемости. Если используемая нами выборочная процедура ведет к тому, что работающие женщины имеют несколько меньшие шансы стать респондентами, чем домохозяйки и пенсионерки (последних, как известно, проще застать дома), наши результаты наверняка окажутся смещенными.

Поэтому *наилучшей моделью отбора считается вероятностная, или случайная, выборка*⁴, в которой строго соблюдается *принцип равенства шансов попадания в выборку и для всех единиц изучаемой совокупности, и для любых последовательностей таких единиц*.

Именно с рассмотрения разных подходов к построению вероятностной выборки мы и начнем наше обсуждение, чтобы в дальнейшем перейти к не столь совершенным видам *целевого*, то есть не основанного на вероятностях, отбора и их роли в практике социологических исследований.

Выше мы определили, что такое выборка. Сейчас нам необходимо строго определить еще несколько элементарных понятий. *Переписью* называют процедуру сбора информации о каждом члене изучаемой группы или популяции. Все члены интересующей исследователя группы (популяции) составляют *генеральную совокупность*.

⁴ В дальнейшем мы будем использовать термины «случайная выборка» и «вероятностная выборка» как взаимозаменяемые.

Выборочная процедура обеспечивает обоснованность и «законность» выводов о генеральной совокупности, сделанных на основании небольшой выборки.

Типы вероятностных выборок и их реализация

Первым шагом в построении любой модели отбора, включая вероятностную, является *определение генеральной совокупности*. Решение этой задачи далеко не всегда бывает очевидным. Прежде всего, генеральная совокупность, то есть множество интересующих социолога объектов исследования, может быть задана и описана лишь на основе каких-то содержательных представлений. Если, например, нас интересуют политические пристрастия избирателей, естественно включить в генеральную совокупность лишь тех, кто уже достиг 18-летнего возраста. Изучение факторов, влияющих на формирование семейного бюджета горожан, потребует иного определения генеральной совокупности: интересующая исследователя популяция в данном случае будет состоять из городских семей.

Полезно также помнить о том, что *идеальная генеральная совокупность*, задаваемая теоретическим описанием предмета исследования, почти никогда не будет полностью совпадать с *реальной совокупностью*. Реальная генеральная совокупность подвержена постоянным колебаниям: «взрослое население города Воронежа на 00 ч 15 ноября 1996 года» будет отличаться от «взрослого населения города Воронежа на 00 ч 16 ноября 1996 года». Некоторые люди за день уедут из города, попадут в больницу, некоторые — вернуться домой из командировки и т. п. Поэтому столь важно при описании изучавшейся в исследовании генеральной совокупности указывать время и место проведения исследования. Следует также помнить, что идеальная генеральная совокупность — это теоретическая абстракция, более или менее совпадающая с реальной совокупностью. Выборка осуществляется из реальной популяции, переход от которой к идеальной совокупности обеспечивается не только правилами статистического вывода, но и некоторой долей теоретического воображения.

Если исследователь построил выборку, которая представляет интересующую его совокупность с приемлемой степенью точнос-

ти, то полученная выборка является *репрезентативной* (представительной). В противоположном случае можно говорить о наличии существенной *выборочной ошибки*. Более строго выборочную ошибку определяют как *расхождение между оценкой некоторого показателя, получаемой на основании исследования выборки, и истинным значением этого показателя в генеральной совокупности*.

К счастью, существуют точные методы для учета и оценки случайной выборочной ошибки, связанной с не носящими систематического характера колебаниями изучаемой переменной в разных подвыборках из одной и той же генеральной совокупности. Подробнее эти методы мы будем обсуждать ниже (в частности, формулы для расчета случайной ошибки выборки будут рассмотрены в гл. 8). Значительно более серьезную проблему создает наличие *систематических смещений*, возникающих в результате *нарушения случайного характера выборочной процедуры*. Результаты такого «не вполне случайного» отбора могут выглядеть более или менее правдоподобно, однако сами по себе они никогда не позволят обнаружить смещение или оценить его величину.

Последнее утверждение можно проиллюстрировать на примере классического опыта с рулеткой. Если нам скажут, что вчера десять раз подряд выпало «красное», мы сможем назвать такую серию событий крайне маловероятной. Однако этот субъективно подозрительный результат сам по себе не дает оснований для каких-то суждений о величине и характере ошибок, порождаемых выборочной процедурой, то есть об исправности механизма самой рулетки.

Систематическая ошибка выборки не обязательно является результатом злого умысла. Например, в США во время войны во Вьетнаме (до введения контрактной системы набора на армейскую службу) правительство проводило специальные лотереи для отбора призывников. Фактически случайно отбирались даты рождения: все годные к несению строевой службы юноши, родившиеся в день, который определялся в ходе такого «розыгрыша», призывались в армию. В 1970 г. результаты отбора были подвергнуты острой критике. Проведенное специальной комиссией расследование показало, что в выборочной процедуре действительно присутствовало смещение. Билетики с напечатанными датами были заключены в

специальные капсулы, которые затем опускали в лотерейный барабан в порядке следования месяцев начиная с января. Из-за недостаточного перемешивания капсул внутри барабана капсулы с ноябрьскими и декабрьскими датами концентрировались в верхней части и, естественно, выпадали с заметно большей частотой⁵.

Самым знаменитым примером смещенной выборочной процедуры в истории социологии стал предвыборный опрос, проведенный американским журналом «The Literary Digest» в 1936 г. Результаты опроса показывали, что Ф. Д. Рузвельт получит 40,9% голосов и уступит президентское кресло республиканцу А. Ф. Лэндону. В действительности Рузвельт получил 60,2% голосов избирателей. Расхождение в 19,3% в значительной степени объяснялось характером выборочной процедуры. Дело в том, что на практике *для построения любой выборки используют какой-то список всех членов изучаемой совокупности, называемый основой выборки*. В опросе, проведенном «The Literary Digest», в качестве основы выборки использовались телефонные справочники, а также регистрационные списки владельцев автомобилей⁶. Во второй половине 1930-х гг. такие списки включали в себя почти исключительно представителей экономически благополучных классов. Беднейшие слои населения, избирательная активность которых, кстати, существенно увеличилась в годы Великой депрессии, оказались недостаточно представлены в выборке, что и послужило причиной столь значительной ошибки. (Интересно отметить, что объем выборки в описываемом случае был просто огромным — свыше двух миллионов человек!)

Существует несколько типов вероятностной выборки, различающихся характером выборочной процедуры. Мы рассмотрим лишь пять: **простую случайную, систематическую, стратифицированную, кластерную и многоступенчатую**.

Процедура построения *простой случайной выборки* включает в себя следующие шаги.

⁵ Rouncefield M., Holmes P. Practical Statistics. Basingstoke: Macmillan Education Ltd, 1989. P. 122.

⁶ Gallup G. A. Guide to Public Opinion Polls. Princeton: Princeton University Press, 1948.

Во-первых, нужно получить полный список членов генеральной совокупности и пронумеровать этот список. Такой список, напомним, называется *основой выборки*.

Во-вторых, следует определить предполагаемый *объем выборки*, то есть ожидаемое число опрошенных.

В-третьих, нужно извлечь из *таблицы случайных чисел* (см. табл. 7.1) столько чисел, сколько нам требуется выборочных единиц. Если в выборке должно оказаться 100 человек, из таблицы берут 100 случайных чисел.

Таблица 7.1

Таблица случайных чисел⁷

Номер строки	Номер столбца									
	1	2	3	4	5	6	7	8	9	10
1	98	08	62	48	26	45	24	02	84	04
2	33	18	51	62	32	41	94	15	09	49
3	80	95	10	04	06	96	38	27	07	74
4	79	75	24	91	40	71	96	12	82	96
5	18	63	33	25	37	98	14	50	65	71
6	74	02	94	39	02	77	55	73	22	70
7	54	17	84	56	11	80	99	33	71	43
8	11	66	44	98	83	52	07	98	48	27
9	48	32	47	79	28	31	24	96	47	10
10	69	07	49	41	38	87	63	79	19	76
11	09	18	82	00	97	32	82	53	95	27
12	90	04	58	54	97	51	98	15	06	54
13	73	18	95	02	07	47	67	72	52	69
14	75	76	87	64	90	20	97	18	17	49
15	67	35	86	33	26	50	10	39	42	61

⁷ Составлено на основе таблицы: Appendix C: Random Numbers // Zeller R. A., Carmines E.G. Statistical Analysis of Social Data. Chicago: Rand McNally, 1978. P. 364–367.

В-четвертых, нужно выбрать из списка-основы (см. выше) те наблюдения, номера которых соответствуют выписанным случайным числам⁸.

Прежде чем мы перейдем к обсуждению возникающих на этом пути практических затруднений, рассмотрим упрощенный пример реализации описанной процедуры.

Пусть нам предстоит построить случайную выборку объемом в 12 человек из совокупности, содержащей 60 членов. Можно предположить, что мы хотим оценить калорийность ежедневного рациона питания 60 студентов-социологов, обучающихся на втором курсе университета, чтобы исследовать возможное влияние энергетической ценности рациона на академическую успеваемость. Для этого можно пронаблюдать за питанием небольшой выборки, состоящей из двенадцати студентов. В качестве основы выборки мы используем список всех 60 студентов. Присвоим всем студентам в списке двузначные номера — от «01» до «60» (если бы максимальный номер в списке был трехзначным, мы бы присваивали трехзначные номера, используя нули в отсутствующих разрядах, например, «067», «003»). Далее нам предстоит последовательно выписать двенадцать двузначных чисел из таблицы случайных чисел (см. табл. 7.1). Отметим, что таблицы случайных чисел фактически состоят из *случайных цифр*, которые обычно сгруппированы для удобства в блоки, состоящие из двузначных либо пятизначных чисел. Объединение цифр в последовательности и блоки условно и не имеет особого статистического смысла. Поэтому в случаях, когда нужны, например, трехзначные числа, а таблица состоит из пятизначных, пользуются каким-то несложным правилом, скажем, используют только три первые цифры каждого пятизначного числа, а оставшиеся две — игнорируют. Соответственно, двузначные числа можно объединять.

Чтобы решить, с какого места в таблице начинать отсчет номеров, достаточно задаться произвольными номерами строки и стол-

⁸ Здесь и далее речь идет о случайной *безвозвратной* выборке, так как выборка с *возвращением* отобранной единицы в совокупность на каждом шаге отбора не очень удобна практически (хотя и обладает рядом статистических преимуществ).

бца. В нашем примере мы начнем с пересечения второй строки и третьего столбца. Первым номером в нашем списке окажется 51. Далее можно двигаться по любому правилу: подряд, через строку, через два столбца и т. п. Мы будем выписывать нужные нам двенадцать двузначных номеров подряд по строке, двигаясь по горизонтали и переходя при необходимости на следующую строку.

Если при этом будут попадаться числа, превосходящие по величине самый большой номер в нашем списке (60), мы будем их пропускать. То же относится и к повторяющимся числам. В результате мы получим последовательность:

51, 32, 41, 15, 09, 49, 10, 04, 06, 38, 27, 07.

Нам остается выписать из списка-основы фамилии, стоящие под этими номерами. Если вы располагаете персональным компьютером, вместо таблицы можно воспользоваться «генератором случайных чисел», имеющимся в большинстве статистических программ.

Простая случайная выборка — это не только наглядное воплощение идеи случайного отбора, но и своего рода *эталон*, с которым сравниваются другие вероятностные процедуры. Здесь необходимо заметить, что вопреки часто высказываемому и неверному мнению простую случайную выборку не следует рассматривать как самую примитивную форму вероятностного отбора. Напротив, более сложные модели случайных выборок используют в тех случаях, когда простую нельзя применить из-за практических или финансовых ограничений. О качестве этих более сложных процедур отбора также судят посредством сравнения с простой случайной выборкой.

Самые очевидные ограничения для использования простой выборки возникают в случае большого объема генеральной совокупности. Прежде всего исследователь сталкивается с проблемами поиска полной и несмещенной *основы выборки*. При обследовании небольших групп и первичных коллективов эти проблемы обычно легко решаются: достаточно воспользоваться членскими списками, списками личного состава и т. п., внося в них необходимые уточнения. В широкомасштабных опросах общественного мнения и социологических обследованиях чаще применяют другие основы: переписные листы, списки избирателей, домовые книги, карточки

паспортных столов милиции (а также картотеки РЭУ, ДЭЗ и т. п.), похозяйственные книги сельских советов. Все эти «готовые» основы выборки обладают определенными преимуществами и недостатками⁹. Решая практическую задачу планирования выборочного исследования, социолог обычно оценивает возможные основы по нескольким параметрам.

Во-первых, списки, пригодные для составления основы выборки, могут храниться либо централизованно, либо децентрализованно, «вразброс», в различных территориальных органах власти, статистических учреждениях и т. п. Естественно, что в первом случае затраты на получение доступа к основе будут значительно ниже, чем во втором. Фактически при децентрализованном хранении исследователь должен самостоятельно составить единый список-основу, собрав необходимые данные в результате обхода (или объезда) всех соответствующих институций.

Во-вторых, используемые в качестве основы выборки списки могут обладать различной степенью точности. Точность списка, в свою очередь, зависит от его полноты, частоты его обновления. Эти качества (полнота списка и высокая частота его пересмотра) редко встречаются одновременно. Как правило, *самыми полными оказываются именно те основы, которые реже всего обновляются*. Таковы, конечно, данные переписей или эпизодически составляемые именные распределительные списки (типа списков на получение приватизационных чеков). К сожалению, чем больше времени отделяет планируемое вами исследование от последней переписи, тем больше вероятность возникновения ошибок и смещений в основе выборки.

Очень существенными достоинствами обладают списки паспортных столов милиции, жилищно-эксплуатационных контор и других местных административных органов.

Качество основы выборки оценивают уже на стадии планирования исследования. Особое внимание уделяют таким потенциальным угрозам валидности, как *неполнота выборочной основы*,

⁹ В отечественной литературе сравнительный анализ разных основ и их применения в конкретных исследованиях осуществлен, например, в книге: Арутюнян Ю. В., Дробизева Л. М., Кондратьев В. С., Сусоколов А. А. Этносоциология: цели, методы и некоторые результаты исследования. М.: Наука, 1984. Гл. IV.

«склеивание» единиц отбора, «пустые» элементы в списке. О неполноте говорят в тех случаях, когда список, используемый для построения выборки, не содержит в себе некоторые единицы, безусловно относящиеся к целевой совокупности. Например, списки жильцов могут не содержать сведений о тех жильцах, которые еще не зарегистрировались по новому месту жительства. В некоторых случаях проблему неполной основы можно решить за счет использования *дополнительных основ*. В нашем примере со списками жильцов такой дополнительной основой могут стать «листки прибытия-убытия», которые хранятся в паспортных столах отделений милиции (с помощью последних ведется учет прописки граждан). Примером «склеивания» может служить ситуация, когда генеральная совокупность, определяемая объектом исследования, состоит из индивидов, а реальной основой отбора служит список квартир или домовладений, содержащий лишь сведения об ответственных квартиросъемщиках либо о собственниках недвижимости. «Пустые» элементы в основе выборки встречаются в тех случаях, когда исходный список содержит имена или адреса, за которыми не стоят реально существующие (или практически доступные) выборочные единицы. Эта проблема часто возникает при использовании устаревших списков, содержащих информацию о временно уехавших, выбывших, умерших и т. п.¹⁰

Описанные выше трудности составления *валидной*, то есть соответствующей объекту исследования (целевой совокупности), основы выборки носят и статистический, и «экономический» характер. Довольно часто исследователь сталкивается с ситуацией, когда временные и финансовые затраты на осуществление простой случайной выборки становятся неприемлемо высокими. Наиболее разумным выходом здесь является использование других, «компромиссных» процедур случайного отбора.

Систематическая выборка по качеству часто приближается к простой случайной. Систематическая выборка, как и простая случайная, требует полного списка или заданного упорядочения

¹⁰ Подробнее об источниках смещений в основе выборки и некоторых способах борьбы со смещениями см.: Kish L. Survey sampling. N. Y.: J. Wiley, 1965. P. 53–59.

совокупности (см. ниже). Техника осуществления систематического отбора элементарна: сначала случайным образом отбирается первая единица, затем отбору подлежит каждый k -й элемент. Число k в данном случае называют *шагом отбора*. Можно, например, отбирать каждый 25-й или каждый 200-й элемент. Чтобы определить шаг отбора, нужно поделить известный объем генеральной совокупности (N) на предполагаемый объем выборки (n).

Пусть, например, нужно отобрать 200 человек из 20 000 владельцев телефонов:

- 1) определим шаг отбора: $N/n = 20\,000 : 200 = 100$;
- 2) с помощью таблицы случайных чисел найдем первую выборочную единицу. Если, скажем, выпал номер «053», то из списка владельцев телефонов выпишем того, кто значится под этим номером;
- 3) с установленным шагом отбираем номера: 153, 253, 353, 453 и т. д. до исчерпания списка.

Иногда генеральная совокупность (и соответственно основа выборки) слишком велика либо исследователю известен не полный список, а лишь *правило упорядочения элементов* в генеральной совокупности. Предположим, что мы хотим составить представление о весе и формате книг, содержащихся в некой библиотеке, при том что мы не располагаем полным каталогом, а лишь видим, как книги расставлены на стеллажах. При условии, что объем библиотечного собрания нам приблизительно известен, мы можем воспользоваться процедурой систематического отбора и отобрать, скажем, каждую 55-ю книгу. Очень важно отобрать «стартовую» единицу сугубо случайным образом. Именно в этом пункте кроется основная слабость систематического отбора. Если в способе упорядочения единиц совокупности имеет место некая цикличность, то есть *неизвестная нам «система»* (систематический паттерн), а случайность в выборе «старта» должным образом не обеспечена, то полученная выборка может также оказаться *смещенной* (если о систематическом паттерне мы знаем заранее, то он не представляет собой угрозы валидности и может быть учтен в ходе отбора). Если воспользоваться примером с отбором книг в библиотеке, то легко представить себе такую гипотетическую ситуацию: исследователь выбирает в качестве стартовой **первую** книгу на нижней полке ближайшего стеллажа и далее двигается с шагом 250 единиц. Если на

каждом стеллаже размещается около 500 книг, то приблизительно половина его выборки будет взята с нижних полок. Однако известно, что на нижних полках многих библиотек нередко размещают книги больших форматов — художественные альбомы, атласы и т. п. Если в нашем примере это правило упорядочения будет соблюдено хотя бы в половине случаев (то есть половина нижних полок будет отведена под «неформатные» издания, под так называемые фолио), любые выборочные оценки «направленности» библиотечного собрания или формата представленных в нем книг окажутся **невалидными**. Аналогией примеру с библиотечными книгами может служить случай систематической выборки городских квартир. Если в результате осуществляемого непосредственно «в поле» интервьюерами систематического отбора в выборке будут сверхпредставлены квартиры, расположенные на первых и последних этажах, возникнет систематическая выборочная ошибка. На первых и последних этажах в российских городах часто живут люди из групп, имеющих более низкий социально-экономический статус и, соответственно, ограниченные финансовые ресурсы: квартиры, расположенные на «крайних» этажах и соприкасающиеся с системами коммунального водо- и теплоснабжения, обычно стоят дешевле, так как названные системы в России традиционно являются источником неприятностей и дисфункций в структуре жизнеобеспечения.

Стратифицированный отбор и, соответственно, стратифицированная выборка, используются в тех случаях, когда из каких-то содержательных соображений важно обеспечить представительность вероятностной выборки по каким-то конкретным *важным для исследовательских целей критериям*. В литературе существует определенная путаница вокруг проблемы стратификации («страта» — это социальная, возрастная или иная группа, буквально «слой»).

Применительно к стратифицированному отбору часто высказывают все те неверные и предрассудочные мнения, которые в начале XX века высказывались относительно квотной выборки (см. ниже) и ее воображаемых преимуществ перед случайным отбором. В действительности стратифицированный отбор имеет определенные практические преимущества до тех пор, пока сохраняется его вероятностный, случайный характер. Как только стратифицированная

выборка превращается в более или менее специально отобранную квотную выборку, воспроизводящую некоторые известные пропорции генеральной совокупности (например, 51% женщин, 30% горожан и т. п.), любые статистические, то есть строгие, оценки параметров генеральной совокупности становятся невозможными.

Стратификацией, строго говоря, называют процедуру, при которой отбор осуществляют как бы из *нескольких «параллельных» подсовокупностей*, заданных на одной и той же генеральной совокупности. Это абстрактное определение можно прояснить с помощью примера. Пусть у нас есть генеральная совокупность взрослых горожан, относительно которой мы располагаем какой-то существенной с точки зрения исследовательских гипотез информацией. Наличие такой *предварительной информации* — необходимое условие стратифицированного отбора. Предположим, мы знаем, что в генеральной совокупности 60% рабочих и 40% служащих. Это соотношение может оказаться весьма существенным с точки зрения наших исследовательских гипотез, если оно задает одну из *независимых переменных*, как, например, при изучении влияния рода занятий на частоту посещения футбольных матчей. Даже при отсутствии значительной систематической погрешности небольшие смещения в реализации случайной выборочной процедуры могут привести к ситуации, когда в нашей конкретной выборке соотношение рабочих и служащих будет существенно (на 5–7%) отклоняться от ожидаемой «правильной» пропорции, имеющей место в генеральной совокупности (см. обсуждение нормальной кривой и индуктивного статистического вывода в гл. 8). Соответственно, под угрозой окажется точность наших оценок взаимосвязи между главной независимой переменной (профессиональным статусом) и интересом к футболу. Такого рода неточность может быть устранена при использовании еще одной случайной выборки из генеральной совокупности, но здесь вступают в силу экономические соображения, так как исследовательский бюджет обычно ограничен. В описанной ситуации желательно заранее обеспечить представленность обеих интересующих нас групп, то есть *страт*, сохранив вероятностный характер отбора. Этого можно добиться, если осуществить некую независимую процедуру случайного отбора для каждой социальной группы в отдельности (в нашем примере для рабочих и служа-

щих) и затем **объединить** полученные случайные подвыборки в одну (заметьте, что для нашего примера объем подвыборки рабочих, в согласии с заранее известной пропорцией, будет в 1,5 раза больше объема подвыборки служащих). Полученная в результате выборка будет *стратифицированной* (по профессиональному статусу), и *вероятностной*.

На практике две случайные процедуры отбора в подвыборки-страты можно технически объединить в одну, если мы располагаем априорной информацией о принадлежности каждой выборочной единицы к той или иной страте. Для этого достаточно вести параллельный отбор из списка-основы в несколько подвыборок (по числу страт). Собственно выборочная процедура может быть *простой случайной*, и *систематической* (соответственно мы получим либо простую, либо систематическую стратифицированную выборку).

Рассмотрим эту процедуру на примере составления систематической выборки населения, стратифицированной по этнической принадлежности. Пусть мы осуществляем выборку взрослых жителей небольшого промышленного центра, при этом полученная выборка должна отражать существующую этнодемографическую ситуацию: 80% русских, 10% украинцев и 10% представителей других национальностей. Основываясь на информации, хранящейся в паспортных столах милиции (или на избирательных списках), мы в идеальном случае можем составить полный список-основу, включающий 100 000 известных административным органам постоянных жителей. Если предварительно мы предполагаем включить в нашу выборку около 1000 человек, нам нужно отобрать из карточек паспортных столов (или избирательных списков) каждого сотого. То есть доля генеральной совокупности f , включенная в выборку, составит $\frac{1}{100}$:

$$f = \frac{\text{объем выборки } (n)}{\text{объем целевой совокупности } (N)} .$$

Выборка объемом в 1000 человек будет включать в себя 800 русских, 100 украинцев и 100 представителей других национальностей. Причем шаг систематического отбора (K) для всех трех подсовкупностей будет равен 100.

Определение шага отбора (K)

80 000 человек в «русской» страте : 800 русских в выборке = 100;

10 000 человек в «украинской» страте : 100 украинцев в выборке = 100;

10 000 человек в страте «другие национальности» : 100 представителей других национальностей в выборке = 100.

Таким образом, мы будем выписывать из реальных карточек (списков) каждого сотого русского, каждого сотого украинца и т. п. (естественно, украинцы и представители других национальностей будут встречаться в списках в среднем в 10 раз реже русских)¹¹.

Выборка в описанном нами примере является *пропорциональной*, так как она представляет все страты в той пропорции, в которой они содержатся в генеральной совокупности. Пропорциональный стратифицированный отбор особенно важен для целей *дескриптивной, описательной статистики*, то есть когда перед исследователем стоит задача, основываясь на выборке, описать, как распределены те или иные параметры в разных группах генеральной совокупности. Именно так обычно можно сформулировать цель предвыборного опроса, маркетингового исследования покупательских предпочтений и т. п. Еще одним преимуществом стратифицированного вероятностного отбора является уменьшение такого источника общей ошибки измерения, как *дисперсия выборки*. Не вдаваясь здесь в статистические тонкости, заметим, что стратификация уменьшает так называемую *стандартную ошибку* (определение и формулу для стандартной ошибки см. в гл. 8) **лишь в том случае**, если интересующая исследователя переменная значительно варьирует **между стратами**, то есть когда заранее выделен-

¹¹ В действительности нам понадобится как минимум 20%-й запас карточек с именами и адресами для замещения тех респондентов, которые окажутся недоступными даже после 2–3 посещения. Доля «недоступных» в исследовании специфических популяций (например, зубных врачей или читателей «Вопросов литературы») может составить 40–50%, включая и длительно отсутствующих, и отказавшихся от сотрудничества, и т. п. Соответственно, в последнем случае «запас» должен составлять 40–50% от первоначально запланированного объема выборки.

ные страты (например, возрастные группы) сильно отличаются по уровню измеряемой переменной (например, по частоте посещения дискотек). При этом различия **внутри страт** должны быть относительно невелики, то есть межгрупповой разброс значений переменной должен значительно превосходить внутригрупповой.

Иногда, однако, основной задачей исследования является сравнение различных, обычно важных с точки зрения некоторой теории, групп внутри выборки с целью описания некоторого *соотношения*, имеющего место в генеральной совокупности. Некоторые из таких «теоретически релевантных» групп могут быть весьма малочисленными. Для того чтобы сделать такие малочисленные группы-субпопуляции статистически сопоставимыми с другими группами и, следовательно, получить статистически значимые выводы о существующих (несуществующих) межгрупповых различиях, можно использовать два метода.

Первый метод заключается в увеличении объема выборки. В этом случае пропорционально возрастает объем «редкой» страты, но столь же быстро (а иногда и быстрее) растут расходы на проведение исследования. Если, например, пожилые люди старше 85 лет составляют лишь $\frac{1}{20}$ часть целевой совокупности горожан-пенсионеров, то в исследовании эффективности социальной работы с пожилыми людьми нам понадобится выборка объемом 4000 пенсионеров, чтобы получить 200 наблюдений, относящихся к редкой подсовокупности тех, кто старше 85.

Другой, более дешевый метод заключается в *непропорциональной стратификации*, то есть в непропорциональном отборе из различных подсовокупностей. Нередко возникает необходимость сделать «распространенные» и «редкие» страты *равно представленными* в выборке. Если вернуться к обсуждавшемуся выше примеру исследования городского населения, можно, в частности, представить ситуацию, когда необходимо сравнить кулинарные предпочтения русских и украинцев. Очевидно, не вполне корректно сравнивать 800 русских и 100 украинцев. В этом случае можно прибегнуть к непропорциональному систематическому отбору из названных страт: если отбирать каждого 200-го русского и каждого 25-го украинца, мы получим две вполне сопоставимые, равные по объему — 400 и 400 человек — подвыборки (однако эти равные подвыборки будут

непропорционально репрезентировать доли соответствующих подсовокупностей, в чем можно убедиться, самостоятельно произведя подсчеты по описанным выше формулам).

Выбор между пропорциональной и непропорциональной стратификацией исследователь осуществляет, исходя из содержательных и экономических соображений. Нужно, однако, иметь в виду некоторые «послевыборочные» последствия непропорционального отбора, с которыми социологи сталкиваются на стадии анализа¹². В частности, для получения более точных оценок распределения исследуемых переменных иногда приходится применять так называемое *взвешивание* (иногда употребляют термин «перевзвешивание»). Взвешивание используют также для того, чтобы исключить влияние некоторых типов *систематического смещения* в основе выборки и других типов *систематической ошибки измерения* (см. гл. 6). Например, взвешивание полезно для исключения смещений, возникающих из-за дублирования в списке-основе или, наоборот, из-за наличия систематических «пропусков» для какой-то одной группы (скажем, если в списке пропущено много пожилых людей, постоянно проживающих с детьми, но прописанных по другому адресу). Так как необходимость взвешивания чаще всего вызвана нарушением исходных соотношений, пропорций между входящими в целевую совокупность группами, мы опишем общую идею этой процедуры на примере непропорционального стратифицированного отбора.

Напомним, что к непропорциональной стратифицированной выборке прибегают в тех случаях, когда точность оценок для выборки в целом или для отдельных подгрупп (субпопуляций) внутри выборки оказывается недостаточной. В этом случае доли генеральной совокупности (f) будут различны для разных страт. Последнее утверждение равносильно признанию *разной вероятности* попадания в выборку для единиц, принадлежащих к разным стратам. Как совместить неравные вероятности отбора с данным нами выше определением вероятностной (случайной) выборки, в котором подчер-

¹² Обсуждение «послевыборочных» последствий различных процедур отбора можно найти, в частности, в книге: *Henry G. T. Practical sampling (Appl. Research Methods Series. Vol. 21). Newbury Park etc.: Sage, 1990. Ch. 8.*

кивалось равенство шансов попадания в выборку для всех входящих в генеральную совокупность единиц-«случаев»? Некоторые статистики считают предложенное нами выше определение не вполне точным и предпочитают говорить о вероятностной выборке как о выборке, где каждая единица отбора имеет «известную, ненулевую вероятность быть включенной в выборку»¹³, хотя шансы для различных единиц не обязательно равны. Существующее многообразие определений вероятностной выборки восходит к давней дискуссии о правомерности выводов, основанных на априорных («до») и апостериорных («после испытания») вероятностях. Мы, однако, сохраним наше определение случайной выборки, внося в него некоторое уточнение: когда шансы попадания в выборку неравны, как при непропорциональном отборе из страт, они могут быть *выравнены при помощи взвешивания на стадии анализа*, то есть на собственно послевыборочной стадии исследования (конечно, если отбор *внутри* страт сохраняет свой случайный и равновероятный характер). Для этого нужно внести определенные поправки в полученные данные, а именно — приписать некоторым наблюдениям (классам наблюдений) больший «вес», компенсирующий меньшие шансы попадания в выборку (и наоборот).

Результатом приписывания веса каждому наблюдению является увеличение точности оценок для исследуемых параметров. Вес каждой единицы (респондента) в k -й страте равен отношению числа таких элементов в генеральной совокупности к объему выборки для k -й страты¹⁴, то есть:

$$w = \frac{N_k}{n_k}.$$

При расчете среднего или других параметров (см. гл. 8) каждое наблюдавшееся значение просто умножается на весовой коэффициент «своей» страты.

В частности, среднее значение какого-то параметра совокупности (например, средний доход или среднее количество хронических

¹³ Henry G. T. Op. cit. P. 25.

¹⁴ Подробнее см.: Sudman S. Applied sampling. N. Y.: Academic Press, 1975. P. 126–130.

заболеваний) будет равняться просто взвешенной сумме средних значений для отдельных страт:

$$\bar{x} = \frac{\sum_k N_k x_k}{N} = \frac{\sum_k w_k \bar{x}_k}{N}.$$

Формула расчета стандартной ошибки (см. гл. 8) для стратифицированной выборки также включает в себя весовые коэффициенты, w :

$$S_{\bar{x}} = \sqrt{\sum w_k^2 S_{\bar{x}_k}^2} = \sqrt{w_1^2 S_{\bar{x}_1}^2 + w_2^2 S_{\bar{x}_2}^2 + \dots + w_k^2 S_{\bar{x}_k}^2}.$$

Стандартные компьютерные программы, используемые при статистическом анализе данных, всегда содержат элементарные процедуры взвешивания.

Вернемся к нашему примеру с непропорциональным стратифицированным отбором русского и украинского населения. Предположим, мы выяснили, что в среднем каждая украинская семья заготавливает на зиму 50 кг варенья, тогда как среднее значение для русской страты составило 40 кг (для простоты расчетов примем маловероятное допущение, что остальные 10 000 опрошенных представителей других национальностей вообще не варят варенье, а используют другие технологии хранения ягод, то есть для этой страты среднее равно нулю). Для украинской страты весовой коэффициент составит:

$$w_{укр.} = 10\,000 : 400 = 25.$$

Соответственно для русского населения:

$$w_{русс.} = 80\,000 : 400 = 200.$$

С учетом этих весовых коэффициентов уточненная оценка среднего запаса варенья в выборке составит:

$$x = 50 \times 0 \times 200 + 25 \times 50 \times 400 + 200 \times 40 \times 400 / 100\,000 = 37 \text{ кг.}$$

Если бы мы не учли в своих расчетах сверхпредставленность украинцев в нашей непропорциональной стратифицированной выборке, то оценка среднего запаса варенья для всей совокупности оказалась бы завышенной (45 кг).

Четвертый тип вероятностной выборки, используемой социологами, — это *кластерная выборка*. «Кластеры» (дословно с англ. — гроздь) — это естественные группировки единиц наблюдения.

Например, популяция избирателей имеет тенденцию жить в городах и деревнях, генеральная совокупность военнослужащих естественным образом группируется по воинским частям и подразделениям, а совокупность студентов — по университетам, институтам и колледжам. Способность к образованию локальных группировок, которую обнаруживают генеральные совокупности, изучаемые социологами, при соблюдении ряда условий позволяет уменьшить расходы на получение единицы информации.

Цель использования кластерной выборки, таким образом, заключается в повышении эффективности затрат на проведение исследования. При фиксированном бюджете и объеме выборки социолог получает возможность снизить общие расходы на проведение личных интервью преимущественно за счет уменьшения транспортных расходов¹⁵.

В общем случае кластерная выборка основана на первоначальном отборе группировок (кластеров) и затем — на изучении всех единиц внутри кластеров. Возможными примерами кластеров, используемых в больших общенациональных опросах, являются сельские районы, городские квартиры, избирательные участки. При изучении специфических популяций используются иные кластеры: больницы — при изучении пациентов, школы — при изучении школьников и т. п.

Корректное применение кластерной процедуры основано на неукоснительном соблюдении четырех необходимых условий¹⁶:

- 1) кластеры должны быть однозначно и явно заданы: каждый член генеральной совокупности должен принадлежать к **одному** (и только одному) кластеру;
- 2) число членов генеральной совокупности, входящих в каждый кластер, должно быть известно или поддаваться оценке с приемлемой степенью точности;
- 3) кластеры должны быть не слишком велики и географически компактны, иначе кластерная выборка теряет всякий финансовый смысл;

¹⁵ Соответственно, использование кластерной процедуры отбора лишено смысла при проведении почтовых опросов, централизованных телефонных интервью и локальных обследований.

¹⁶ *Sudman S. Op. cit. P. 70.*

4) выбор кластеров должен быть осуществлен таким способом, который минимизирует рост выборочной ошибки (последний процесс, в свою очередь, является неизбежным следствием кластеризации).

Для того чтобы уяснить, как именно кластерная процедура влияет на рост выборочной ошибки, рассмотрим ее на простейшем примере. Допустим, мы изучаем труд и занятость жителей небольшого сельского района. Для того чтобы составить полный список-основу для случайной выборки, нам пришлось бы предварительно посетить все сельские советы, а в некоторых случаях — и весьма отдаленные деревни. Располагая ограниченными ресурсами, мы решаем использовать имеющуюся в нашем распоряжении карту района, на которой отмечены все населенные пункты, включая самые небольшие хутора. Известна и численность населения для каждого пункта. Естественными границами кластеров-поселений являются шоссе и проселочные дороги. Составив список всех 40 деревень и хуторов, мы можем теперь без труда осуществить *простую случайную выборку кластеров*. Для отдельного поселения вероятность попадания в выборку составит $\frac{1}{40}$. Если, например, мы собираемся опросить 200 человек, нам, скорее всего, потребуется отобрать 1–2 кластера-поселения. Отметим здесь, что естественные различия в величине кластеров¹⁷ никак не влияют на процедуру кластерного отбора.

Что при этом происходит с выборочной ошибкой и, следовательно, с получаемыми в нашем исследовании статистическими параметрами генеральной совокупности сельского населения района (то есть с оценками возраста, дохода и т. п.)? Чтобы ответить на этот вопрос, мы должны ввести еще одно статистическое понятие «*независимых наблюдений*» (степеней свободы).

Предположим, мы хотим оценить соотношение работающих и пенсионеров в обследуемом нами районе. Мы отобрали, условно, три деревни по 30 домовладений каждая (итого 90 домовладений).

¹⁷ В нашем случае так называемой *территориальной* кластерной выборки таковыми являются различия в численности населения отдельных деревень и хуторов.

Однако в ходе опроса выясняется, что в двух деревнях, не входящих ни в одно сельхозобъединение или кооператив, живут исключительно старики-пенсионеры, а в одной, построенной недавно для переселенцев из Средней Азии, живут только молодые семьи с детьми. Таким образом, каждая деревня является населенной либо только работающими семейными парами, либо исключительно «пенсионерской». В результате мы можем заранее предсказать результат обследования каждой деревни (кластера), посетив лишь один дом. Если в первом доме интервьюер обнаружит чету пенсионеров, во всех остальных домах тоже будут жить пенсионеры. Если в первом доме живут люди трудоспособного возраста, посещение остальных 29 домовладений приведет к тому же результату. Фактически для каждой деревни мы будем располагать **одним** независимым наблюдением и, посетив 90 семей в трех деревнях, получим лишь **три** независимых, информативных наблюдения относительно распределения работающих и пенсионеров в выборке. Соответственно, наши оценки величины данного соотношения в генеральной совокупности окажутся более неточными, чем в случае 90 независимых наблюдений. Причина возникающей ошибки заключается в том, что использованные нами кластеры (деревни) оказались *гомогенными*, однородными по исследуемому признаку трудовой занятости, хотя по другим признакам, например по политической активности, они вполне могут быть *гетерогенными*, неоднородными. В принципе можно показать, что рост выборочной ошибки для кластерной выборки (в сравнении с простой случайной) является функцией двух переменных — *величины кластеров* и *гомогенности* исследуемого признака внутри каждого кластера¹⁸.

Ясно, что оценка гомогенности часто становится важной практической задачей в планировании кластерной выборки. Основная проблема здесь заключается в том, что соответствующими данными о распределении признаков внутри кластеров исследователь располагает после завершения собственно полевой стадии. Практически при проектировании выборки обычно основываются на *уже существующих* данных предыдущих исследований, переписей и

¹⁸ См.: *Sudman S.* Op. cit. P. 73–78.

т. п. Мера гомогенности ρ ведет себя так же, как соответствующий коэффициент корреляции. Величина ρ — это корреляция между значениями признака для *всех возможных парных сочетаний* элементов, входящих в кластер. Эта величина обычно положительна и возрастает с ростом гомогенности элементов внутри кластера. Если наблюдения внутри кластера абсолютно независимы (как в примере случайного распределения между разными кластерами), то $\rho = 0$. При использовании территориальной кластерной выборки городского населения, например при отборе кварталов или многоэтажных домов, ρ для признаков экономического статуса может быть весьма высоким из-за «пороговых» эффектов: в престижном кооперативном доме маловероятно встретить семьи с очень низкими доходами (верхний порог) и, наоборот, лишь немногие состоятельные люди обитают в коммуналках, подобно герою «Золотого теленка» Александру Ивановичу Корейко (нижний порог).

Ориентировочное представление о типичных значениях r и их изменении для кластеров разной величины для общенационального выборочного исследования дает таблице 7.2. В ней показаны величины ρ для имеющих разные размеры кластеров, составленных из соседних городских домовладений (квартир и домов). Данные таблицы основаны на выборке городского населения США ($N > 100\,000$)¹⁹:

Еще одной немаловажной практической проблемой в планировании кластерной либо стратифицированной выборки является сравнение эффективности затрат на исследование при разных *среднем размере кластера и количестве кластеров* (заметим, что и кластеры, и страты часто обозначают общим термином — «первичные единицы отбора»). Функция, описывающая зависимость расходов от вышеперечисленных двух переменных, выглядит так:

$$C_t = ac_1 + nc_2,$$

где C_t — общая стоимость исследования, a — количество «первичных единиц отбора», c_1 — средние затраты на обследование «первичной единицы отбора», планируемые для данного исследования,

¹⁹ Источник: Hansen M., Hurwitz W. N., Madow W. G. Sample Survey Methods and Theory. N. Y.: Wiley and Sons, 1953. 2 vols. (Vol. 1. P. 264. Table 3).

Т а б л и ц а 7.2

Значения мер гомогенности ρ для кластеров, состоящих из домовладений (для основных социально-демографических параметров)

Параметр	Значение ρ для кластера, имеющего средний размер n			
	$n = 3$	$n = 9$	$n = 27$	$n = 62$
Доля домовладений: находящихся в личной собственности	0,170	0,171	0,161	0,096
наемных, с низкой квартплатой	0,235	0,169	0,107	0,062
наемных, с высокой квартплатой	0,430	0,349	0,243	0,112
Среднее количество жильцов	0,230	0,186	0,142	0,066
Доля среди жильцов: белых мужчин	0,100	0,088	0,077	0,058
безработных мужчин	0,060	0,070	0,045	0,034
мужчин в возрасте 25– 34 лет	0,045	0,026	0,018	0,008

n — общий размер планируемой выборки, c_2 — средние затраты на проведение одного интервью²⁰.

Дальнейшим обобщением идей случайного отбора из субпопуляций и естественных группировок, лежащих в основе, соответственно, стратифицированной и кластерной выборок, является *многофазная (многоступенчатая) выборка*. Построение такой выборки представляет собой довольно сложную статистическую задачу, подходы к решению которой мы рассмотрим лишь в самом обобщенном виде.

В простейшем случае многофазная выборка состоит из двух фаз случайного отбора. На первой — как при кластерном отборе — отбираются «первичные единицы отбора», например районы, избирательные участки, предприятия. На второй фазе производится

²⁰ См.: Sudman S. Op. cit. P. 78–79; Hansen M., Hurwitz W. N., Madow W. G. Op. cit.

случайный отбор единичных членов генеральной совокупности — отдельных респондентов, семей и т. п. Так как «первичные единицы отбора» могут существенно отличаться по величине (как, например, отличаются друг от друга городские квартиры или дома с разной численностью проживающих), то результатом первой фазы может стать неравная вероятность попадания в выборку для членов генеральной совокупности, относящихся к разным «первичным единицам отбора». В этом случае исследователь имеет возможность выравнивания вероятностей на последующих фазах (например, из «первичной единицы отбора», где проживает 1000 семей, он выберет 10, а из «первичной единицы», где живет 500 семей, будет отобрано 20).

Рассмотрим многофазную процедуру на простейшем примере с равной вероятностью отбора.

Пусть нам необходимо осуществить выборку размером 2000 человек из генеральной совокупности населения крупного города, где проживает 4 млн человек. Каждая «первичная единица отбора» — городской квартал — содержит 1000 единиц (то есть отдельных респондентов). На первой фазе мы отберем из 100 000 кварталов («первичных единиц отбора») 400, так что для каждого квартала вероятность попадания в выборку составит:

$$400 : 100\,000 = 0,004$$

На следующей стадии из 1000 жителей каждого квартала мы отберем 50, так что для каждого респондента суммарная накопленная вероятность попадания в двухфазную выборку составит:

$$0,004 \times (50 : 1000) = 0,0002.$$

Решение об использовании многофазной выборки обычно принимается после анализа «баланса» затрат и приобретений. Снижение затрат на сбор данных, достигаемое в этом случае, сопровождается увеличением сложности выборочной процедуры. С ростом числа фаз (в больших общенациональных обследованиях нередко используют 4 или 5 «ступенек» отбора — от области до квартала) точность получаемых оценок имеет тенденцию снижаться. Поэтому исследователям нередко приходится *сочетать многофазный отбор со стратификацией* на завершающих стадиях выборочной процедуры, что обычно ведет к улучшению характеристик выбор-

ки²¹. Отсюда понятно, почему многофазная выборка в значительной мере остается «прерогативой» крупных исследовательских организаций, которые обладают значительными финансовыми ресурсами и могут воспользоваться услугами профессионалов-статистиков при проектировании выборки.

Размер вероятностной выборки

Вопрос об оптимальном размере вероятностной выборки всегда был спорным и в значительной мере остается таковым. Мы обсудим лишь основные принципы, лежащие в основе современного подхода к оптимизации размера выборки.

Решение относительно размера выборки принимают с учетом целого ряда факторов, среди которых самую существенную роль играют два: 1) ценность и новизна получаемой в результате опроса информации и 2) затраты на проведение опроса (включая временные) при заданном размере выборки.

Некоторые исследователи полагают, что принятие решения о размере выборки может основываться на сугубо статистическом подходе²². При этом в расчет принимают допустимую величину ошибки в оценке исследуемого параметра (например, дохода). Существуют статистические формулы, связывающие размер выборки с *вероятностью ошибки* и величиной *доверительного интервала*, задающего пределы этой ошибки (два последних понятия подробнее обсуждаются в гл. 8). Так как использование этих формул требует принятия определенных предположений о том, как распределена интересующая исследователя величина, возникает необходимость в *предварительной информации*, относящейся к тому самому параметру, который мы решили изучить. Трудности, возникающие при

²¹ Примером многофазной (многоступенчатой) стратифицированной выборки может служить выборка «Всесоюзного этносоциологического исследования» (рук. Ю. В. Арутюнян, 1971–1976 гг.). См. подробнее: Арутюнян Ю. В., Дробижева Л. М., Кондратьев В. С., Сусоколов А. А. Цит. соч. С. 111–123. Отметим также, что впервые в отечественной социологии многоступенчатая территориальная вероятностная выборка использовалась в исследовании читателей газеты «Правда», проводившемся В. Э. Шляпентохом в 1970-е гг.

²² См.: Кокрен У. Методы выборочного исследования. М.: Статистика, 1976.

использовании классического статистического подхода к определению размера вероятностной выборки, можно описать одной фразой, принадлежащей известному специалисту по массовым опросам С. Судману: «Очевидно, что формула, описывающая зависимость размера выборки от предполагаемой ширины доверительного интервала и приемлемой вероятности ошибки, попросту заменяет проблему определения размера выборки другой, не менее трудной проблемой — определения ширины доверительного интервала»²³.

Во многих важных случаях можно руководствоваться сложившейся практикой, то есть *размером выборки, использовавшейся в аналогичных исследованиях*. Кроме того, нужно помнить о простейших «**правилах левой руки**» для определения размера выборки.

Размер выборки растет

— при необходимости опубликовать данные для отдельных подгрупп (размеры подвыборок при этом суммируются, и выборка в целом растет пропорционально числу подгрупп);

— при проведении общенациональных обследований, когда велика генеральная совокупность (заданная доля генеральной совокупности f будет определять тем больший объем выборки, чем больше генеральная совокупность);

— если уже имеющаяся информация по ключевым вопросам (например, о намерениях избирателей голосовать за ту или иную партию) явно недостаточна и степень неопределенности значительна.

Размер выборки уменьшается

— при исследовании организаций, институтов и прочих «первичных единиц отбора», если сравнительно невелика величина генеральной совокупности, из которой производится отбор (например, совокупности сотрудников рекламных агентств, школьников, пациентов и т. п.);

— при проведении локальных и региональных исследований;

— если уже существующая информация относительно полна и все еще остающаяся степень неопределенности незначительна.

²³ *Sudman S. Op. cit. P. 89.*

«Типичные» размеры выборок для общенациональных опросов варьируют в пределах 1000–2500 респондентов (в зависимости от числа анализируемых подгрупп), для региональных опросов и опросов специальных популяций — от 200 до 500 (при анализе многочисленных подгрупп размер региональной или специальной выборки обычно возрастает как минимум до 1000 человек). Указанные значения, разумеется, могут служить лишь самым общим ориентиром для определения оптимального размера выборки.

Целевой отбор

Иногда социологи вынуждены применять *не основанные на вероятностях выборки*. Отбор в этом случае базируется не на принципе рандомизации, а на следовании тем или иным *субъективным критериям* — доступности, типичности, равного представительства и т. п. Многие из этих критериев при систематическом использовании позволяют добиться достаточно высокого качества социологических данных. Часто такой отбор называют *целевым*, так как он в большой степени определяется целями исследования. Кроме того, в конкретной исследовательской ситуации может оказаться, что осуществление случайной выборки — это практически невыполнимое или экономически неэффективное мероприятие (затраты на построение выборки превышают ценность получаемой в результате исследования информации). Наконец, использование вероятностного отбора лишено всякого смысла, если речь идет об исследовании уникальных событий, групп или ситуаций — полетов на Луну, войн или любовных историй (об этнографическом методе, применяемом в такого рода исследованиях, говорится в гл. 2).

Основной недостаток неслучайных процедур отбора связан с тем, что не существует строгих статистических методов, позволяющих обобщить результаты, полученные в ходе исследования выборки. Оценка точности и валидности этих результатов (и основанных на них выводов) остается делом субъективного суждения, опыта, теоретических предпочтений.

Самый распространенный тип не основанной на вероятности выборки — это *выборка доступных случаев*. Такого рода выборка

может считаться корректной *лишь тогда, когда используется в экспериментальном (или квазиэкспериментальном) исследовании*. Так, в большинстве психологических экспериментов испытуемыми являются студенты. Это позволяет экономить скудные финансовые ресурсы, отпускаемые на сугубо академические изыскания. Для того чтобы исключить влияние посторонних, смешивающих факторов, экспериментатор в *случайном порядке* распределяет выборку доступных случаев (то есть доступных испытуемых) по двум группам — экспериментальной и контрольной. В нашем обсуждении роли рандомизации в эксперименте (гл. 4) подчеркивалось ее значение для получения точных и обоснованных выводов. Однако случайное приписывание испытуемых-добровольцев к экспериментальной и контрольной группам, строго говоря, не является достаточным основанием для обобщения результатов эксперимента для всей *генеральной совокупности*, из которой осуществлялась выборка доступных случаев. Точнее, в ситуации отбора доступных случаев невозможно с полной уверенностью сказать, что, собственно, являлось генеральной совокупностью в процессе исследования, так как последняя не была определена с самого начала. Поэтому, в частности, шутливое определение предмета психологии гласит, что это наука, изучающая студентов-второкурсников гуманитарных факультетов. В социологии выборкой доступных случаев чаще всего приходится довольствоваться при изучении таких специальных популяций, которые практически не поддаются локализации. Речь идет, прежде всего, об относительно *малочисленных группах, находящихся вне сферы институционального* (например, административного) *контроля*. Для таких групп трудно найти какую-то основу выборки — скажем, посетители стрелковых тиров едва ли состоят на каком-нибудь государственном учете. «Просеивание» большой случайной выборки из генеральной совокупности с целью рекрутирования сколько-нибудь значительного числа респондентов в специальную выборку требует непомерных затрат. Поэтому социологам иногда приходится уподобляться орнитологам и отбирать членов экзотических популяций в местах их «естественного обитания» или вероятного скопления. Многие исследования посетителей массовых библиотек проводятся в библиотеках, посетителей выставок — в музеях, ветеранов войны — в клубах ветеранов и т. п.

В этой ситуации исследователю приходится прилагать дополнительные усилия для получения высококачественной информации. Следует заметить, что некоторая статистическая «небезупречность» получаемых таким образом результатов при должной методической культуре исследователей иногда окупается, и мы узнаем нечто принципиально новое об относительно «закрытых» областях человеческого поведения²⁴. Однако если целью исследования является *описание распределения признаков во вполне определенной генеральной совокупности* (покупателей зубной пасты, избирателей, читателей газет), то социолог, использующий выборку доступных случаев, понапрасну тратит деньги заказчика (и пренебрегает профессиональной этикой). Квалифицированному заказчику в этом случае также не стоит принимать всерьез рассуждения о принципиально новых, нестатистических и даже «мягких» методах проведения массовых опросов.

Значительно реже социологи используют две другие разновидности целевого отбора — *отбор «критических случаев»* и *отбор «типичных случаев»*. В обоих случаях исследователь полагается на какие-то теоретические представления или предыдущий опыт, чтобы отобрать ограниченное число «симптоматических», характерных наблюдений, позволяющих сделать более широкие обобщения и предсказания. Иногда это удается, но следует помнить о том, что опыт и теоретические суждения обычно бывают субъективны. В печально знаменитых президентских выборах 1948 г. в Америке (Г. Трумэн против Т. Дьюи) ошибочные прогнозы сделали все знаменитые институты опросов общественного мнения. При этом некоторые из них избрали в качестве «типичного» случая население

²⁴ В отечественной литературе примеры очень интересных исследований, основанных на целевом отборе, особенно многочисленны (причиной чему, очевидно, является хроническая недостаточность финансирования социологических исследований). Общее представление об используемых в них методах повышения качества информации можно составить, ознакомившись с несколькими хорошими работами, например: 47 пятниц. Функционирование общественного мнения в условиях города (программы и документы исследования). М.: ССА, 1969. Вып. 1; Шубкин В. Н. Начало пути. М.: Молодая гвардия, 1979; Клявина Т. А., Хршановская С. П. В поисках зрителя (итоги опроса руководителей театров РСФСР) // Социологические исследования. 1988. № 3. С. 47–53.

штата Мэн, так как прежде жители этого штата всегда «угадывали» будущего президента. В описываемом случае «нетипично» (то есть за проигравшего выборы Дьюи) проголосовали только два штата — Мэн и Вермонт. Поэтому поговорку «как голосует Мэн, голосует вся Америка» пришлось перефразировать: «как голосует Мэн, так голосует Вермонт»²⁵.

Метод «снежного кома» — это еще один (наряду с выборкой доступных случаев) интересный подход к отбору из «редких» совокупностей. Его идея такова: первоначально идентифицированная небольшая группа членов интересующей социолога совокупности служит источником сведений о других членах этой совокупности, так что выборка постепенно разрастается вширь, подобно снежному кому, катящемуся с горы. Этот метод использовал, например, П. Лазарсфельд с коллегами в исследовании «влиятельных людей» и неформальных связей. Помимо властвующих элит данный метод применяют в изучении других групп, также избегающих широкой известности, например наркоманов или коллекционеров антиквариата. Для этого метода существуют определенные приемы оценки систематической ошибки, однако они слишком сложны, чтобы обсуждаться здесь.

К выборкам, не основанным на случайном отборе, относится и *квотная выборка*, когда-то чрезвычайно популярная даже среди профессиональных статистиков и практически не используемая сейчас. Идея квотной выборки проста: изучаемая совокупность разбивается на такие социально-демографические группы, которые исследователь почему-либо считает важными. Обычно критериями разбивки становятся пол, возраст, национальная принадлежность, место жительства и т. п. Далее, основываясь на уже известных (обычно из официальной статистики) пропорциях этих групп в генеральной совокупности, социолог составляет полевые задания для интервьюеров, указывая, сколько женщин, мужчин, лиц с высшим образованием и т. п. нужно опросить. Например, интервьюер получает задание опросить 10 женщин старше 50 лет, 8 мужчин 35–45 лет и трех восемнадцатилетних девушек, проживающих в Санкт-Петербурге. В результате должна получиться

²⁵ Henry G. T. Op. cit. P.21.

выборка, представляющая все заданные пропорции групп в генеральной совокупности.

Основная проблема квотного отбора заключается в том, что он *носит неслучайный характер и осуществляется лично интервьюером*. Последний выбирает респондентов, в конечном счете, по собственному усмотрению. Хотя число мужчин или женщин, рабочих или пенсионеров, которых следует опросить в данном районе или местности, задано заранее, интервьюер решает, в какую квартиру ему удобнее позвонить, с кем из членов семьи провести интервью, куда вернуться вторично, если на звонок никто не ответил и т. п. Это неизбежно ведет к систематическим смещениям в процессе отбора, причем не существует никаких методов для оценки величины возникающей систематической ошибки.

Еще один очевидный недостаток квотного отбора связан с тем, что обычно невозможно даже приблизительно оценить количество отказов от участия в опросе. Если интервьюер сталкивается с человеком, не желающим отвечать на вопросы, или просто недоброжелательным, или вызывающим у него антипатию, интервьюер всегда волен попрощаться и попытаться счастья в соседней квартире.

По указанным причинам квотные выборки «вышли из моды» среди социологов, несмотря на свою относительную дешевизну.

Оценивая полезность и применимость вышеописанных «неслучайных» методов отбора в исследовательской практике, следует прежде всего сказать, что в определенных обстоятельствах никакой другой альтернативы просто не существует. В ситуации нехватки денег, персонала, времени либо первичной информации о генеральной совокупности социологи использовали и будут использовать впредь выборки доступных случаев, метод «снежного кома» и даже (к сожалению) квотную выборку. При этом профессиональный долг социолога заключается в том, чтобы оценить, пусть даже очень приблизительно, величину и источники возникающей выборочной ошибки.

Безусловно, разумно использовать целевые выборки в пилотажных исследованиях, в экспериментах, в том числе методических (то есть нацеленных на проверку и отработку анкет, опросников, шкал и т. п.).

Однако всегда следует помнить о том, что возможность обобщения любых оценок, полученных на целевой выборке, для

генеральной совокупности в целом, то есть *внешняя валидность* результатов исследования, чаще всего оказывается сомнительной²⁶.

Дополнительная литература

Кокрен У. Методы выборочного обследования. М.: Статистика, 1976.

Петренко Е. С., Ярошенко Т. М. Социально-демографические показатели в социологических исследованиях. М.: Статистика, 1979.

Территориальная выборка в социологических исследованиях. М.: Наука, 1980.

Чурилов Н. Н. Проектирование выборочного социального исследования. Киев: Наукова думка, 1986.

²⁶ Предвыборные опросы общественного мнения, проводившиеся различными российскими исследовательскими центрами в первой половине 1990-х гг., изобилуют столь многочисленными подтверждениями этой истины, что трудно выбрать один «негативный пример» для критического рассмотрения. Систематический анализ просчетов в организации выборки таких опросов содержится в работах: *Шляпентох В. Э.* Предвыборные опросы 1993 г. в России (критический анализ) // Социологические исследования. 1995. № 10. С. 3–10; *Мансуров В. А., Петренко Е. С.* Изучение общественного мнения в России и СССР // Социология в России. М.: На Воробьевых, 1996. Богатый эмпирический материал, относящийся к ошибочным прогнозам итогов выборов в Думу 1995 г., см. в статье: *Рубинов А.* Социология сказала... // Лит. газета. 1995. 13 дек.

ГЛАВА 8. АНАЛИЗ ДАННЫХ

Виды анализа данных

Методы, применяемые социологами для анализа данных, многообразны. Выбор конкретного метода зависит, в первую очередь, от характера исследовательских гипотез, то есть от того, на какие вопросы мы хотим получить ответ. Если целью является описание одной характеристики выборки в определенный момент времени, разумно ограничиться *одномерным анализом*, то есть описанием распределения наблюдений («случаев») вдоль оси интересующего нас признака. Разнообразные техники *многомерно-го анализа* позволяют одновременно исследовать взаимоотношения двух и более переменных и в той или иной форме проверять гипотезы о причинных связях между ними. Различия между этими методами — точнее, классами методов — неабсолютны. В реальном исследовании каждое уточнение исходных гипотез или выдвижение новой гипотезы в ходе анализа результатов приводит к необходимости выбора новой техники анализа данных. Так, если изначальная модель взаимоотношения двух переменных (скажем, профессии и дохода) не позволяет выявить определенную закономерность в собранных данных, исследователь выбирает одну из статистических техник, позволяющих контролировать влияние какой-то третьей переменной, например пола, на интересующее его отношение.

Помимо характера исследовательских гипотез на выбор методов статистического анализа влияет и природа полученных социологом данных. Мы уже говорили о том, что разные *уровни измерения* социологических переменных определяют возможности и ограничения анализа. Для того чтобы охарактеризовать распределение в выборке такого номинального признака, как «пол», мы не

можем воспользоваться его среднеарифметическим значением и, следовательно, нам потребуются какие-то другие приемы компактного и точного представления полученной информации.

Методы, используемые для анализа связи между двумя номинальными переменными, также будут отличаться от методов анализа связи между номинальной переменной и переменной, измеренной на интервальном уровне. Таким образом, выбор той или иной статистики будет зависеть и от целей анализа, и от уровня измерения исследуемых переменных.

Существует два основных класса задач, решаемых с помощью статистических методов анализа. Задачей *дескриптивной (описательной) статистики* является описание распределения переменной-признака в конкретной выборке. Методы дескриптивной статистики позволяют также анализировать взаимосвязь между различными переменными. Другой класс задач, связанный с необходимостью вывести свойства большой совокупности, основываясь на имеющейся информации о свойствах выборки из этой совокупности, решается с помощью методов *индуктивной статистики*, или теории статистического вывода, основанной на вероятностном подходе к принятию решений. Воспользовавшись какой-то моделью для анализа полученных выборочных данных, социолог обычно также применяет некоторые методы статистического вывода, позволяющие определить, выполняются ли обнаруженные им при анализе данных отношения на уровне большой совокупности, из которой была извлечена выборка.

В этой главе мы уделим основное внимание использованию дескриптивной статистики в анализе социологических данных. Нашей целью здесь будет скорее качественное, содержательное понимание сути этих методов, основанное лишь на самых элементарных математических представлениях и, в некоторых случаях, на интуитивном понимании «физического смысла» статистических моделей. Такое понимание может служить определенным фундаментом для более глубокого изучения прикладной статистики. Кроме того, оно совершенно необходимо для того, чтобы самостоятельно формулировать задачи анализа данных и ориентироваться в существующем разнообразии методов и техник, используемых другими исследователями при решении этих задач.

Одномерный анализ: табулирование и представление данных

Результаты измерения любой переменной могут быть представлены с помощью распределения наблюдений («случаев») по отдельным категориям данной переменной. Категория, в которую попадают одинаковые наблюдения, может быть номинальной («православный», «протестант» и т. п.) либо иметь числовое значение. В любом случае результатом такого упорядочения наблюдений будет их *группировка*. Работать с упорядоченными данными значительно проще, чем с исходным «сырым» массивом: в «сырых» данных, конечно, содержатся сведения о том, как много в выборке, например, пенсионеров, однако для получения нужной цифры придется перебрать все наблюдения «случай» за «случаем». Если данные сгруппированы, достаточно посмотреть, какова *абсолютная частота*, то есть число наблюдений в данной выборке, попадающих в интересующую нас категорию. Для переменных, имеющих произвольную метрику, то есть измеренных на ординальном или интервальном уровне (см. гл. 6), нередко используется еще одна процедура, делающая представление данных более компактным и удобным в работе при сохранении заданного уровня точности. Предположим, что в каком-то исследовании 22,0782% опрошенных поддержали государственную программу приватизации, а исследование, проведенное месяц спустя, дало иное значение — 22,1327%. Даже если теоретический конструкт «поддержка программы приватизации» можно представить как непрерывный ряд числовых значений, на практике исследовательской переменной будет соответствовать некоторый набор дискретных числовых величин (категорий). Кроме того, тысячные или сотые доли процента едва ли будут существенны для интерпретации полученных результатов. Поэтому в представлении данных обычно используют процедуру *округления*. Определив необходимую степень точности — и, соответственно, приемлемый уровень неточности, — исследователь может округлить все полученные числовые значения до десятых долей или, скажем, до целых процентов. Так, в нашем примере округление до целого числа даст цифру 22%. В дальнейшем каждое последующее наблюдение, дающее числовое значение в интервале между 21,5%

и 22,5% будет попадать в класс «22% поддержки приватизации». В результате процедуры округления исследователь фактически устанавливает *границы классов*, объединяющих значения переменной в заданном интервале, и *середины (центры) классов*, то есть усредненные значения для каждого интервала.

Необходимость объединить значения переменной в 10–15 крупных классов-категорий часто возникает и при работе со «слишком хорошо измеренными» признаками, соответствующими шкалам интервалов или отношений (возраст, доход и т. п.). Во-первых, чрезмерное количество градаций переменной препятствует ее компактному представлению — табличному или графическому. Во-вторых, для конечной выборки обычно соблюдается следующая закономерность: число градаций (категорий) признака обратно пропорционально их заполненности. Переменная с огромным числом градаций, содержащих по 2–3 наблюдения, часто создает серьезные проблемы в статистическом анализе и оценивании (хотя для некоторых методов анализа — корреляция, регрессия и т. п. — эти проблемы, как мы увидим дальше, несущественны). Самым целесообразным выходом обычно оказывается перекодирование, «сжатие» исследовательской переменной. Здесь существует два основных подхода:

1) исходные градации объединяются в более крупные классы на основании каких-то содержательных соображений, причем полученные классы имеют приблизительно равную ширину (например, данные о возрасте часто перекодируют в более широкие «десятилетние» категории — 20–29, 30–39 лет и т. п.);

2) решение о способе «сжатия» переменной принимают, основываясь на распределении наблюдений («случаев») по оси переменной, например, границы между «низким», «средним» и «высоким» доходом устанавливают так, чтобы в каждую категорию попало 33% наблюдений.

Стремление к компактности и «читабельности» данных не должно вести к крайностям. Руководствуясь соображениями здравого смысла, исследователь должен избегать ситуаций, когда перегруппировка ведет к тому, что полученная переменная оказывается слишком грубым средством классификации наблюдений, не позволяющим выявить существенные для анализа различия. Важно так-

Таблица 8.1

**Частотное распределение ежемесячных расходов
на международные телефонные переговоры**

Интервал класса (расходы, руб.)	Абсолютная частота, чел.	Относительная частота, %
До 3000	51	11,0
3000–5999	40	8,6
6000–8999	135	29,0
9000–11 999	80	17,2
12 000–14 999	65	14,0
15 000–19 999	49	10,5
20 000–23 999	37	8,0
Свыше 24 000	8	1,7
Всего	N = 465	100% (= 465)
Не ответили	35	(35)

же следить за тем, чтобы объединение категорий или числовых градаций переменной-признака не привело к искусственному созданию отношений и взаимосвязей, которые в действительности отсутствуют в данных.

Независимо от того, какие статистические методы и модели собирается использовать исследователь, первым шагом в анализе данных всегда является построение *частотных распределений* для каждой изучавшейся переменной. Полученные результаты принято представлять в виде таблицы частотного распределения (или просто — таблицы распределения) для каждой существенной переменной. Примером табличного представления может служить приведенная ниже таблица 8.1, в которой представлены гипотетические данные выборочного опроса 500 владельцев домашних телефонов.

Иногда в таблице распределения указывают лишь относительные частоты, опуская абсолютные. Но и в этом случае в правом нижнем углу таблицы должны быть указаны абсолютное число ответивших (база для вычисления процентов) и число неответивших.

Помимо табличного представления частотных распределений обычно используют и различные методы *графического представления*. Самый распространенный метод графического представления

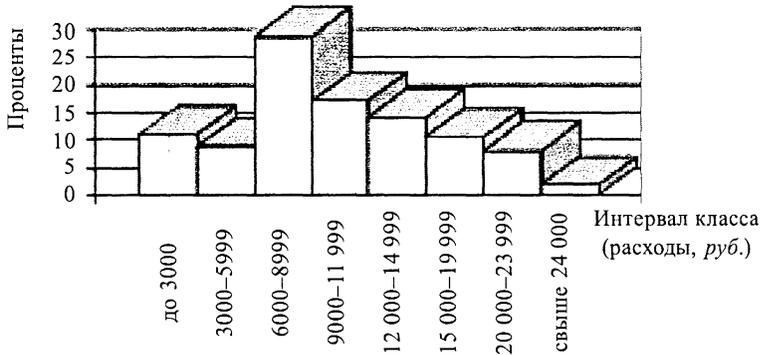


Рис. 8.1. Гистограмма для данных о расходах на телефонные переговоры

одномерных распределений — это *гистограмма*, или столбиковая диаграмма. Каждый столбик соответствует интервалу значений переменной, причем его середина совмещается с серединой данного интервала. Высота столбика отражает частоту (абсолютную или относительную) попадания наблюдавшихся значений переменной в определенный интервал. При построении гистограмм часто приходится использовать некоторые конвенции, основанные на сугубо практических соображениях. Так, используя при группировке значений переменной неравные интервалы либо оставляя крайние градации открытыми («старше 65 лет», «свыше 24 000 рублей» и т. д.), мы все же отображаем эти интервалы на гистограмме с помощью столбиков, имеющих одинаковую ширину. Другое практическое правило позволяет сделать гистограмму визуально уравновешенной, то есть более привлекательной: масштаб шкалы обычно выбирают так, чтобы общая высота гистограммы составляла приблизительно 40–60% ее ширины. Пример гистограммы для данных из таблицы 8.1 приведен на рис. 8.1.

Если просто соединить между собой точки, соответствующие абсолютным или относительным частотам (ось ординат) для середин интервалов, мы получим так называемый *полигон распределения*. Эта операция, разумеется, будет иметь какой-то смысл лишь для количественных переменных, которые мы в принципе можем представить себе как непрерывные. На рис. 8.2 изображен полигон распределения для экспертных оценок телегеничности политичес-

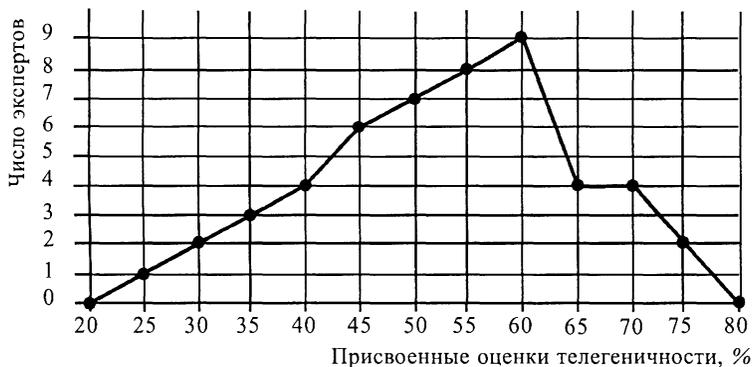


Рис. 8.2. Полигон распределения для оценок телегенности политического лидера

кого лидера (50 экспертов оценивали политика в процентах по отношению к некоторому абсолютному эталону телегенности).

Еще один популярный способ графического представления, обычно используемый для качественных данных (то есть для номинальных или ординальных измерений), — это *круговая диаграмма*. Каждый сектор круговой диаграммы представляет дискретную категорию переменной. Величина сектора пропорциональна частоте категории для данной выборки. На рис. 8.3 приведена круговая диаграмма, иллюстрирующая распределение подростков, страдающих вялотекущей формой шизофрении, по возрасту на момент начала («дебюта») заболевания¹.

Какую бы форму представления данных мы ни избрали, полученное частотное распределение все еще содержит «слишком много» деталей, не отвечая при этом на весьма важные для содержательного анализа вопросы о самых типичных значениях признака и диапазоне разброса отдельных наблюдений. Для облегчения работы с частотными распределениями, а также для обобщенного представления их характеристик обычно используют определенные числовые значения — *статистики*. Дело в том, что специалисты по статистике используют последний термин в двух значениях: как

¹ Источник данных: Личко А. Е. Шизофрения у подростков. М.: Медицина, 1989. С. 6.

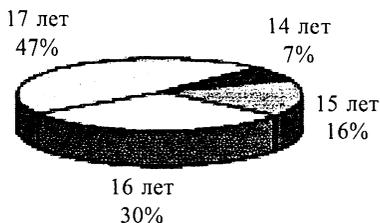


Рис. 8.3. Заболеваемость вялотекущей формой шизофрении у подростков мужского пола по возрастам, %

название своей дисциплины и как обозначение какой-либо числовой функции, описывающей результаты наблюдений. Наибольшее практическое значение имеют две группы статистик: *меры центральной тенденции* и *меры изменчивости (разброса)*.

Меры центральной тенденции указывают на расположение среднего, или типичного, значения признака, вокруг которого сгруппированы остальные наблюдения. Понятие среднего, центрального, значения в статистике, как и в повседневной жизни, подразумевает нечто «ожидаемое», «обычное», «типичное». Способность среднего значения давать некую обобщенную информацию о распределении вытекает из того соотношения, которое связывает среднее значение с другими «особыми» точками распределения — минимумом и максимумом: зная среднее значение, мы можем утверждать, что наименьшее наблюдаемое значение полученного распределения — например, распределения веса или интеллекта — было не больше среднего, а наибольшее зафиксированное значение — не меньше среднего.

Отличие статистической трактовки среднего значения (или, точнее, мер центральной тенденции) от его «житейской» трактовки заключается прежде всего в том, что в статистике, в отличие от повседневной жизни, понятие среднего значения может быть строго задано лишь для *одномерного распределения* переменной-признака. Мы можем, например, указать на семью со средним душевым доходом, но при этом не следует ожидать, что данная семья будет средней или типичной в каких-то других отношениях, то есть будет иметь средний размер, среднюю жилплощадь и т. п. В повседневном общении мы приписываем понятию среднего куда бо-

лее широкий и менее точный смысл. В этом нет большой беды, пока мы не смешиваем «житейскую» и «статистическую» интерпретации. Мы действительно получаем полезную информацию, узнав, что окружающие говорят о ком-то как о «человеке средних способностей», но будет ошибкой заключить, что некто X, имеющий средний показатель интеллекта, наверняка имеет средние успехи в учебе или посредством сочиняет стихи. Именно поэтому популярные газетные образы «среднего российского подростка» или «среднего читателя», в сущности, лежат за пределами корректного использования статистики.

Самой простой из мер центральной тенденции является *мода* (*Mo*). Для номинальных переменных мода — это единственный способ указать наиболее типичное, распространенное значение. Разумеется, исследователь может пользоваться модальным значением и для характеристики распределения переменных, измеренных на более высоком уровне, если для этого существуют содержательные основания (например, описывая распределение ответов на вопрос о количестве выписываемых журналов). *Мода — это такое значение в совокупности наблюдений, которое встречается чаще всего.* Например, если в выборке содержится 60% православных, 30% мусульман и 10% представителей других конфессий, то модальным значением будет «православный». У моды как меры центральной тенденции есть определенные недостатки, ограничивающие ее интерпретацию. Во-первых, в распределении могут быть две и более моды (соответственно оно является бимодальным или мультимодальным). Скажем, если в группе из десяти человек четверо не имеют автомобиля (0), четверо имеют один автомобиль, один человек имеет две машины и еще один — три, то нам придется указать два модальных значения — 0 и 1. Кроме того, мода чрезвычайно чувствительна к избранному способу группировки значений переменной. Объединяя категории ответа, мы резко увеличиваем число наблюдений в отдельных категориях. Это открывает широкий простор для манипулирования данными (не всегда добросовестного). Поэтому «правилом хорошего тона» при вычислении модального значения для сгруппированных количественных данных является *выравнивание ширины для всех интервалов класса*. Еще одно важное правило касается случаев, когда частоты для всех

наблюдаемых значений почти равны. Здесь лучше воздержаться от вычисления моды, так как в этом случае она просто не может быть интерпретирована как мера центральной тенденции. Если, скажем, 48% болельщиков поддерживают сборную Италии, а 49% — сборную Бразилии, модальное значение «поддерживает бразильцев» будет не очень модальным. И все же во многих случаях вычисление моды и необходимо, и полезно. Например, для архитектора, занимающегося планированием жилых домов, знание модального значения для размера семьи в данной местности может оказаться весьма важным.

Другая мера центральной тенденции — *медиана* — обычно используется для ординальных переменных, то есть таких переменных, значения которых могут быть упорядочены от меньших к большим. Пример вычисления медианы рассматривался нами в гл. 6. Напомним, что медиана (Md) — это значение, которое делит упорядоченное множество данных пополам, так что одна половина наблюдений оказывается меньше медианы, а другая — больше. Иными словами, медиана — это 50-й процентиль распределения. Как мы уже видели, при работе с большим массивом данных удобнее всего искать медиану, построив на основании частотного распределения распределение *накопленных частот* (или построив распределение накопленных процентов на основании распределения процентов). Для того чтобы найти медианное значение для маленького массива наблюдений, достаточно упорядочить наблюдения от меньших значений переменной к большим: то значение, которое окажется в середине, и будет медианным. Например, для ряда: 17 баллов, 18 баллов, 20 баллов, 21 балл, 22 балла медианой будет значение 20 баллов. Если число значений в группе наблюдений четное, то медианой будет среднее двух центральных значений. Медиану иногда называют «позиционным средним», так как она указывает именно среднюю позицию в упорядоченном ряду наблюдений. Медиана может совпадать или не совпадать с модой. При этом медиана лучше всего соответствует нашему интуитивному представлению о середине упорядоченной последовательности чисел. Некоторые исследователи даже полагают, что медиана — лучше и «справедливее» среднеарифметического при описании таких величин, как, скажем, доход семьи. Ведь семьи, имеющие доход

ниже *среднего*, могут составить и 60, и 70% населения. Когда же мы говорим, например, что *медианный* доход составил 10 млн рублей в год, то не более 50% семей окажутся «ниже среднего уровня». На медиану не влияют величины «крайних», очень больших или малых значений.

И все же для количественных переменных самой важной и распространенной является другая мера центральной тенденции — *среднее арифметическое*, которое чаще всего называют просто *средним* (и обозначают как \bar{X}). Процедура определения среднего общеизвестна: нужно просуммировать все значения наблюдений и разделить полученную сумму на число наблюдений. В общем случае:

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n},$$

$$\text{то есть } \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i,$$

где $X_1 \dots X_i$ — наблюдаемые значения, n — число наблюдений, Σ — знак арифметической суммы.

Таблица 8.2

Вычисление среднего возраста посетителей библиотеки

Возраст	Абсолютная частота, f_i	$X_i \times f_i$	\bar{X}
18	5	90	$\bar{X} = \frac{\sum_{i=1}^7 X_i f_i}{N} = \frac{489}{20} = 24,45,$ <p>где $i = 1 \dots 7$ — число различных значений</p>
19	2	38	
21	4	84	
22	6	132	
30	1	30	
35	1	35	
80	1	80	
	Всего $N = \sum_{i=1}^7 f_i = 20$	$\sum X_i f_i = 489$	

В таблице 8.2 показано, как вычислить средний возраст для выборки из 20 посетителей библиотеки. Заметьте, что каждое значение просто умножается на свою абсолютную частоту.

Приведенный нами пример (см. табл. 8.2) показывает, насколько среднее уязвимо для «крайних» значений. Фактически для нашей небольшой выборки молодых людей прибавление одного — восьмидесятилетнего — читателя заметно увеличило средний возраст. Следует, однако, помнить о том, что степень «возмущения» среднего под влиянием единичных очень больших или малых значений уменьшается в прямом соответствии с ростом объема выборки. Заметим также, что при расчете среднего для сгруппированных данных частоты умножаются на значение, соответствующее *сердине интервала группировки*.

Среднее обладает рядом важных свойств. В частности, если сложить все значения отклонений от среднего значения, то есть разности между \bar{X} и X_1, X_2, \dots, X_i (которые могут быть и положительными, и отрицательными), то *сумма отклонений* будет равна нулю. Кроме того, *сумма квадратов отклонений* наблюдаемых значений от их арифметического среднего меньше суммы квадратов отклонений от любой другой точки². Эти свойства среднего определяют его уникальную роль в решении ряда статистических задач, о которых мы будем говорить ниже. Сейчас достаточно отметить то обстоятельство, что при использовании среднего в качестве «представителя» (то есть статистической оценки) каждого из наблюдаемых значений ошибка, определяемая как сумма квадратов отклонений, будет минимальной. Не стоит, однако, забывать о том, что и минимальная ошибка может быть достаточно большой. Так, для малых выборок, имеющих более чем одну моду, любая мера центральной тенденции, включая среднее, будет недостаточно хороша. Центральной тенденции в таком распределении просто не существует.

Выбирая меру центральной тенденции, нужно руководствоваться знанием ее свойств, общей формой распределения и, наконец, здравым смыслом. Если при взгляде на гистограмму исследователь

² Доказательства этих свойств см. в книге: Гласс Дж., Стэнли Дж. Статистические методы в педагогике и психологии. М.: Прогресс, 1976. С. 64–65.

обнаруживает, что имеет дело с унимодальным симметричным распределением (половины гистограммы слева и справа от модального значения зеркально совпадают), то среднее, медиана и мода будут равны между собой. Если речь идет о выборке из большой совокупности, где интересующая исследователя переменная-признак распределена нормально (то есть большие и малые крайние значения встречаются редко, а средние — часто), наилучшим показателем будет среднее. Если в унимодальном распределении встречаются крайние значения, могущие значительно повлиять на среднее (см. пример с возрастом, табл. 8.2), нужно отдать предпочтение медиане.

Вопрос о *сравнимости* средних значений не так тривиален, как это может показаться. Сравнение значений средних показателей для различных выборок или для одной и той же выборки в разные моменты времени — весьма распространенный способ анализа результатов. Не только в научных журналах, но и в газетах мы постоянно находим сведения о сравнительной величине душевого дохода в разных регионах, о различиях в среднем числе автомобилей, приходящихся на одну семью и т. п. Следует, однако, помнить о том, что заведомо некорректны сравнения *различных* мер центральной тенденции, например медианы и среднего. Причина здесь в том, что различные меры описывают *разные* характеристики распределения: медиана — среднее положение, мода — самое часто встречающееся значение и т. д. Кроме того, даже две одинаковые меры центральной тенденции не всегда сравнимы. Средние двух распределений имеет смысл сравнивать лишь в том случае, если во всех других отношениях распределения одинаковы, имеют сходную форму. Если исследователь говорит о равенстве средних значений, забыв упомянуть о том, что одно распределение симметрично, а другое — скошено вправо или влево из-за присутствия очень больших либо очень малых значений в его «хвостовых» частях, то он подталкивает читателя к заведомо неверному выводу о том, что анализируемая переменная распределена в двух выборках совершенно одинаково. Среднее распределения с очень длинным правым «хвостом» может оказаться равным среднему распределения, скошенного влево, где встречаются крайне малые значения признака. Но этим сходство будет исчерпываться: что общего (кроме величины

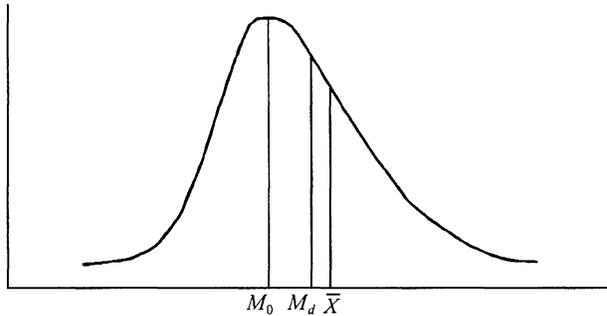


Рис. 8.4. Распределение, скошенное вправо

среднего) у группы, включающей много людей с очень низким доходом, коэффициентом интеллекта и т. п., с другой группой, включающей много наблюдений с очень высокими значениями переменной-признака?

Очевидно, важно не только знать, что *типично* для выборки наблюдений, но и установить, насколько выражены *отклонения* от типичных значений. Чтобы определить, насколько хорошо та или иная мера центральной тенденции описывает распределение, нужно воспользоваться какой-либо *мерой изменчивости, разброса*.

Самая грубая мера изменчивости — *размах* (диапазон) значений. Эта мера не учитывает индивидуальные отклонения значений, описывая лишь диапазон их изменчивости. Под размахом понимают разность между максимальным и минимальным наблюдаемым значением. Если количество карманных денег в группе из десяти субъектов варьирует от 100 рублей (1 человек) до 100 000 рублей (2 человека), размах будет равен $100\,000 - 100 = 99\,900$.

Еще одна грубая мера разброса значений — это *коэффициент вариации* (V), который определяется просто как процент наблюдений, лежащих вне модального интервала, то есть процент (доля) наблюдений, не совпадающих с модальным значением. Если от модального отличаются 60% значений, то $V = 60\%$ (или $V = 0,6$).

Рассказывая о процедуре построения шкалы Терстоуна (см. с. 186–191), мы описали, как вычислить *междуквартильный размах* — очень удобный показатель разброса значений для ординальной переменной. Напомним, что нижний, первый, квартиль (Q_1) от-

секает 25% наблюдений, а ниже третьего квартиля (Q_3) лежат уже 75% случаев. *Полумеждуквартильный размах* равен половине расстояния между третьим и первым квартилями:

$$Q = \frac{Q_3 - Q_1}{2}.$$

Если распределение приблизительно симметрично, то можно считать, что полумеждуквартильный размах указывает границы, в которых лежит 50% данных по обе стороны медианы или среднего.

Все эти меры изменчивости, как уже говорилось, можно считать скорее грубыми и приблизительными. Ни одна из них не уделяет должного внимания информации об отклонениях каждого отдельного наблюдаемого значения от среднего, хотя эта информация в большинстве случаев может быть получена из анализа распределения. Информацию о вариации некоторой совокупности значений относительно среднего несут значения отклонений от среднего, о которых мы уже говорили. Однако просуммировав все значения отклонения ($\bar{X} - X_i$), мы получим нуль. Положительные и отрицательные отклонения будут взаимоуничтожаться. Если же мы *возведем в квадрат* каждое отклонение и *просуммируем квадраты отклонений*, то мы получим хорошую меру рассеяния, которая будет маленькой, когда данные однородны, и большой, когда данные неоднородны. Чтобы суммы квадратов отклонений для выборок разного размера можно было сравнивать, нужно поделить каждую из них на N , где N — объем выборки³.

Именно так и получают важнейшую меру рассеяния — дисперсию (s^2). Если \bar{X} — среднее, X_1, X_2, \dots, X_n — индивидуальные значения измеряемой переменной X в данной совокупности, а N — объем выборки⁴:

³ Для небольших выборок ($N < 100$) лучше делить на $(N - 1)$.

⁴ Для сгруппированных данных более точной формулой дисперсии будет:

$$S_x^2 = \frac{1}{N} \sum_{i=1}^k (X_i - \bar{X})^2 n_i = \sum_{i=1}^k (X_i - \bar{X})^2 f_i,$$

где k — количество *разных* значений ($k < n$), а $\sum_{i=1}^k f_i = N$.

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{N}.$$

Для того чтобы вычислить значение дисперсии, нужно вычесть из каждого наблюдаемого значения среднее, возвести в квадрат все полученные отклонения, сложить квадраты отклонений и разделить полученную сумму на объем выборки.

Величина, равная квадратному корню из дисперсии, называется *стандартным отклонением* (s_x), то есть:

$$s_x = \sqrt{s^2}.$$

Совершенно очевидной интерпретацией стандартного отклонения является его способность оценивать «типичность» среднего: стандартное отклонение тем меньше, чем лучше среднее суммирует, «представляет» данную совокупность наблюдений.

Еще одно важное применение стандартного отклонения связано с тем, что оно, наряду со средним арифметическим, позволяет определить самые существенные характеристики *нормального распределения*. Графически нормальному распределению частот наблюдений соответствует, как известно, симметричная колоколооб-

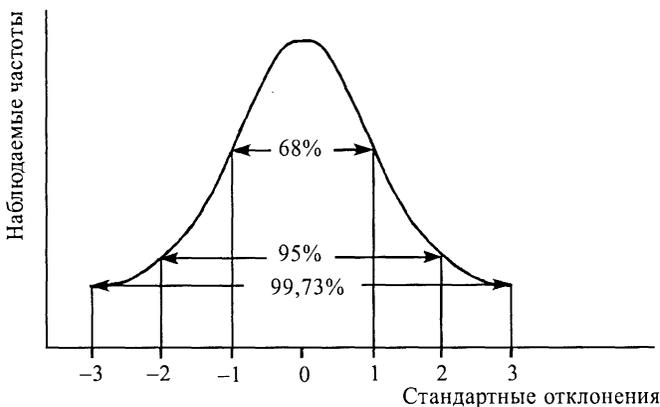


Рис. 8.5. Определение площади нормальной кривой для разных значений стандартного отклонения

разная кривая. Свойства нормального распределения прекрасно изучены, что позволяет делать важные выводы относительно самых разных распределений, не обязательно нормальных. В частности, известно, что 68% наблюдений (точнее, 68% общей площади) будет заключено в пределах ± 1 стандартное отклонение от среднего значения. Если, скажем, среднее нормального распределения равно 200, а стандартное отклонение — 4, то можно заключить, что не менее 68% наблюдений лежит между значениями 196 и 204 (то есть 200 ± 4). Соответственно, не менее 32% случаев будут лежать за этими пределами, в левом и правом «хвостах» распределения. Из теории вероятности известно также, что в пределах ± 3 стандартных отклонений окажется около 99,73% общего числа наблюдений (см. рис. 8.5).

Для любого унимодального симметричного распределения, даже если оно отличается от нормального, не менее 56% наблюдений будут попадать в промежуток ± 1 стандартное отклонение от среднего арифметического значения, для ± 3 стандартных отклонений внутри указанного интервала окажутся не менее 95% наблюдений.

Очевидно, что стандартное отклонение — это прекрасный показатель положения любого конкретного значения относительно среднего, поэтому часто возникает необходимость выразить «сырые» оценки — баллы теста, величины дохода и т. п. — в единицах стандартного отклонения от среднего. Получаемые в результате оценки называют *стандартными*, или *Z-оценками*. Для любой совокупности из N наблюдений распределение со средним \bar{X} и стандартным отклонением s_x можно преобразовать в распределение со средним, равным 0, и стандартным отклонением, равным 1. Преобразованные таким образом индивидуальные значения будут непосредственно выражаться в отклонениях «сырых» значений от среднего, измеренных в единицах стандартного отклонения. Чтобы осуществить такое преобразование, нужно из каждого значения X вычесть среднее и разделить полученную величину на стандартное отклонение, то есть Z-оценки получают по простой формуле:

$$Z_i = \frac{X_i - \bar{X}}{s_x}.$$

Использование Z -оценок не сводится к описанию положения некоторого значения относительно среднего в масштабе единиц стандартного отклонения. Стандартные оценки позволяют перейти от множества «сырых» значений к произвольной шкале с удобными для расчетов характеристиками среднего и стандартного отклонения. Домножая Z на константу c , мы можем получить распределение со стандартным отклонением (s_x). Множество данных можно расположить на любой шкале с удобным средним (например, равным 100, как во многих тестах интеллекта) и стандартным отклонением. Другие применения Z -оценок связаны со сложными методами анализа данных, о которых мы будем говорить в дальнейшем.

Описанные процедуры анализа одномерного распределения относятся к дескриптивной статистике. Если мы стремимся обобщить данные, полученные на отдельных выборках, чтобы описать свойства исходной генеральной совокупности, необходимо, как уже говорилось, обратиться к методам *индуктивной статистики*, к теории статистического вывода. *Переход от числовых характеристик выборки к числовым характеристикам генеральной совокупности называется оцениванием*. При одномерном анализе данных чаще всего решают задачу *интервального оценивания*.

Если переменная измерена на уровне не ниже интервального (доход, продолжительность образования и т. п.), мы легко можем получить выборочную оценку среднего. Но как узнать, насколько близка наша выборочная оценка, например, дохода к истинному значению этого параметра, которое мы получили бы, располагая возможностью обследовать всю совокупность? Если наша выборка была случайной, на этот вопрос можно ответить. Чтобы перейти от выборочной оценки (статистики) к характеристике генеральной совокупности (параметру), можно, в частности, определить числовой интервал, в который с заданной вероятностью «укладывается» интересующий нас параметр. Чтобы понять идею интервального оценивания, достаточно вспомнить о том, что оценки, получаемые для множества выборок из одной совокупности, будут также распределены нормально, то есть большая их часть будет попадать в область, близкую к истинному среднему, и лишь немногие окажутся в «хвостах» распределения, отклоняясь от этого значения. Для любой отдельно взятой выборки шансы оказаться близко к пара-

метру совокупности значительно выше вероятности оказаться в «хвосте». Чтобы оценить степень этой близости, используют очень важную величину — *стандартную ошибку средней*. Стандартную ошибку обозначают как s_M .

$$s_M = \frac{s_x}{\sqrt{N}},$$

где s_x — это стандартное отклонение, а N — объем выборки.

Подсчитав эту величину для наших данных, мы всегда можем определить с заданной вероятностью, в каких пределах будет лежать среднее совокупности. Совершенно аналогично приведенным выше рассуждениям для среднего отклонения можно сказать, что 95% выборочных средних будет лежать в пределах ± 2 стандартные ошибки среднего генеральной совокупности (то есть для 95 выборок из 100 выборочное среднее попадет в указанный интервал). Следовательно, любая конкретная единичная выборка, использованная в данном исследовании, с 95%-ой вероятностью даст оценку, лежащую в интервале ± 2 стандартных ошибок среднего совокупности. Заданный таким образом интервал для выборочных оценок называется *доверительным интервалом*, а та вероятность, с которой мы «попадаем» в этот интервал (например, 95% или 99%), называется *доверительной вероятностью*. Если, например, мы рассчитали, что для случайной выборки горожан средняя квартирная плата составляет 20 000 рублей, а стандартная ошибка — 500 рублей, то можно с 95-процентной уверенностью утверждать, что для всех горожан средняя квартплата окажется в интервале 19 000–21 000 рублей. Задав интервал в 3 стандартные ошибки, мы сможем достичь уровня доверительной вероятности, равного 99,73% (см. рис. 8.5). Полезно помнить о том, что чем больше используемая выборка (чем больше N), тем меньше будет s_M (см. формулу) и, следовательно, тем уже будет доверительный интервал.

Задачу интервального оценивания можно решить и для тех переменных, уровень измерения которых ниже интервального. Для этого в статистике используют свойства другого распределения — *биномиального*. Здесь мы не будем анализировать эти свойства. Достаточно отметить, что биномиальным называют распределение исхода событий, которые могут случиться или не случиться, то есть

в общей форме могут быть классифицированы как положительные или отрицательные. При этом наступление одного события автоматически означает, что другое не случилось. Степень интенсивности события (признака) просто не принимается в расчет. Классический пример — бросание монеты, которая может выпасть «орлом» или «решкой». Чтобы использовать это распределение для интервального оценивания, нужно превратить анализируемую переменную в *дихотомическую*, имеющую две категории (если, конечно, она таковой не являлась с самого начала). Примеры дихотомических переменных: пол, голосование «за» или «против» и т. п. Для дихотомической переменной стандартную ошибку можно вычислить по формуле:

$$s_{bin} = \sqrt{\frac{PQ}{N}},$$

где s_{bin} — стандартная ошибка для биномиального распределения, P — процент наблюдений в первой категории, Q — процент наблюдений во второй категории, N — объем выборки.

Если, например, нас интересует, насколько близок к истинному значению для генеральной совокупности тот процент ответов, который мы получили при опросе некоторой выборки, мы снова можем использовать интервальную оценку. Пусть, например, в выборке объемом 1000 человек 60% высказались против призыва студентов на воинскую службу, а 40% — за. Стандартная ошибка составит:

$$s_{bin} = \sqrt{\frac{60 \times 40}{1000}} = \sqrt{2,40} = 1,55.$$

Если добавить (и отнять) 2 стандартные ошибки⁵ к полученной выборочной оценке, можно построить доверительный интервал, в который интересующая нас величина попадет с 95%-й вероятностью (то есть вероятность ошибки не превысит 5%). С вероятнос-

⁵ Для больших выборок биномиальное распределение практически не отличается от нормального. Если P и Q не слишком различны по величине, достаточно и не очень большой выборки.

тью 95% доля противников обязательного призыва студентов составит $60 \pm 3,1\%$.

Анализ связи между двумя переменными

Хотя результаты одномерного анализа данных часто имеют самостоятельное значение, большинство исследователей уделяют основное внимание анализу связей между переменными. Самым простым и типичным является случай анализа взаимосвязи (сопряженности) двух переменных. Используемые здесь методы задают некоторый логический каркас, остающийся почти неизменным и при рассмотрении более сложных моделей, включающих множество переменных. Устойчивый интерес социологов к двумерному и многомерному анализу данных объясняется вполне понятным желанием проверить гипотезы о причинной зависимости двух и более переменных. Ведь утверждения о причинных взаимосвязях составляют фундамент не только социальной теории, но и социальной политики (по крайней мере, так принято считать). Так как возможности социологов проверять причинные гипотезы с помощью эксперимента, как уже говорилось, ограничены, основной альтернативой является статистический анализ неэкспериментальных данных.

В общем случае, для демонстрации причинно-следственного отношения между двумя переменными, скажем X и Y , необходимо выполнить следующие требования:

- 1) показать, что существует эмпирическая взаимосвязь между переменными;
- 2) исключить возможность обратного влияния Y на X ;
- 3) убедиться, что взаимосвязь между переменными не может быть объяснена зависимостью этих переменных от какой-то дополнительной переменной (или переменных).

Первым шагом к анализу взаимоотношений двух переменных является их *перекрестная классификация*, или построение *таблицы сопряженности*. Речь идет о таблице, содержащей информацию о совместном распределении переменных. Допустим, в результате одномерного анализа данных мы установили, что люди сильно различаются по уровню заботы о своем здоровье: некоторые люди

регулярно делают физические упражнения, другие — полностью пренебрегают зарядкой. Мы можем предположить, что причина этих различий — какая-то другая переменная, например пол, образование, род занятий, доход и т. п.

Пусть мы располагаем совокупностью данных о занятиях физзарядкой и образовании для выборки горожан. Для простоты мы предположим, что обе переменные имеют лишь два уровня: высокий и низкий. Так как данные об образовании исходно разбиты на большее количество категорий, нам придется их перегруппировать, разбив весь диапазон значений на два класса. Предположим, мы выберем в качестве граничного значения 10 лет обучения, так что люди, получившие неполное среднее и среднее образование, попадут в «низкую» градацию, а остальные — в «высокую». (Это, конечно, большое огрубление, но мы используем его из соображений простоты.) Для занятий физическими упражнениями мы, соответственно, воспользуемся двумя категориями — «делают физзарядку» и «не делают физзарядку». Таблица 8.3 показывает, как могло бы выглядеть совместное распределение этих двух переменных.

Т а б л и ц а 8.3

Взаимосвязь между уровнем образования и занятиями физкультурой

Занятия физкультурой	Уровень образования		Всего
	низкий	высокий	
Делают зарядку	50	200	250
Не делают зарядку	205	45	250
Всего	255	245	500

В таблице 8.3 два столбца (для образования) и две строки (для занятий физкультурой), следовательно, размерность этой таблицы 2×2 . Кроме того, имеются дополнительные крайний столбец (справа) и нижняя строка (маргиналы таблицы), указывающие общее количество наблюдений в данной строке или в столбце. В правом нижнем углу указана общая сумма, то есть общее число наблюдений в выборке. Не давшие ответа уже исключены (для реальных данных их число также стоит указать, но не в таблице, а в подтаб-

личной сноске). Заметим здесь, что многие исследователи при построении таких таблиц пользуются неписанным правилом: для той переменной, которую полагают независимой, отводится верхняя строка (горизонталь), а зависимую располагают «сбоку», по вертикали (разумеется, соблюдение этого правила не является обязательным и ничего с точки зрения анализа не меняет).

Обычно характер взаимоотношений между переменными в небольшой таблице можно определить даже «на глазок», сравнивая числа в столбцах или строках. Еще легче это сделать, если вместо абсолютных значений стоят проценты. Чтобы перевести абсолютные частоты, указанные в клетках таблицы, в проценты, нужно разделить их на маргинальные частоты и умножить на 100. Если делить на маргинал столбца, мы получим *процент по столбцу*. Например, $\frac{50}{255} \times 100 = 19,6\%$, то есть 19,6% имеющих низкий уровень образования делают зарядку (но не наоборот!). Если делить на маргинал строки, то мы получим другую величину — *процент по строке*. В частности, можно заметить, что 80% делающих зарядку составляют люди с высоким уровнем образования ($\frac{200}{250} \times 100$). Деление на общую численность выборки дает общий процент. Так, всего в выборке 50% людей, делающих зарядку.

Так как вывод о наличии взаимосвязи между переменными требует демонстрации различий между подгруппами *по уровню зависимой переменной*, при анализе таблицы сопряженности можно руководствоваться простыми правилами. Во-первых, нужно определить независимую переменную и, в соответствии с принятым определением, пересчитать абсолютные частоты в проценты. Если независимая переменная расположена по горизонтали таблицы, мы считаем проценты по столбцу; если независимая переменная расположена по вертикали, проценты берутся от сумм по строке. Далее сравниваются процентные показатели, полученные для подгрупп с разным уровнем независимой переменной, каждый раз *внутри* одной категории зависимой переменной (например, внутри категории делающих зарядку). Обнаруженные различия свидетельствуют о существовании взаимосвязи между двумя переменными. (В качестве упражнения примените описанную процедуру

к таблице 8.3, чтобы убедиться в наличии связи между уровнем образования и занятиями физкультурой.)

Отметим специально, что элементарная таблица сопряженности размерности 2×2 — это минимально необходимое условие для вывода о наличии взаимосвязи двух переменных. Знания о распределении зависимой переменной недостаточно. Нельзя, например, утверждать, будто из того, что 75% детей-первенцев имеют интеллект выше среднего, а 25% — средний и более низкий, следует зависимость между порядком рождения и интеллектом. Необходимо проанализировать и распределение показателей интеллекта для детей-непервенцев. *Варьировать должна не только зависимая, но и независимая переменная.*

Для таблиц размерности 2×2 и более можно рассчитать специальные показатели (статистики), дающие суммарное выражение степени взаимосвязи, ассоциации между двумя переменными. Таких мер связи довольно много. Для случая двух номинальных переменных существуют два основных подхода к подсчету коэффициентов взаимосвязи. Проанализировав их общую логику, мы получим возможность ориентироваться в многообразии конкретных показателей, предлагаемых прикладными программами анализа данных. Первый подход базируется на статистике, называемой «хи-квадрат». На ее основе можно рассчитать несколько коэффициентов взаимосвязи. Рассмотрим в качестве примера коэффициент «фи» (греч. ϕ), формула для которого была впервые предложена сэром Карлом Пирсоном в 1901 году специально для того, чтобы сделать возможным анализ взаимосвязи между двумя переменными, измеренными на неколичественном уровне.

Предположим, мы располагаем таблицей сопряженности для двух переменных-признаков X и Y , каждая из которых принимает лишь два значения, которые мы условно обозначим как «0» и «1». В каждой из четырех клеток таблицы содержатся абсолютные частоты, то есть число случаев для каждого из возможных сочетаний значений признаков (то есть для сочетаний «0–1», «1–1», «0–0», «1–0»). Обозначим частоты в каждой из клеток таблицы латинскими буквами a , b , c и d . В такой общей форме таблица сопряженности для двух дихотомических признаков будет выглядеть как на таблице 8.4.

Т а б л и ц а 8.4

Общая форма таблицы сопряженности размерности 2×2

Переменная Y	Переменная X		
	0	1	Всего
1	a	b	$a + b$
0	c	d	$c + d$
Всего	$a + c$	$b + d$	n

Для расчета коэффициента сопряженности «фи» используют формулу:

$$\varphi = \frac{bc - ad}{\sqrt{(a+c)(b+d)(a+b)(c+d)}}.$$

Эта простая в вычислительном отношении формула получается в результате ряда преобразований исходной формулы для вычисления величины «хи-квадрат» (χ^2). Эта исходная формула позволяет лучше понять общую идею оценки связи качественных признаков, которую мы опишем, не вдаваясь в статистические детали. Исходная формула для величины «хи-квадрат» выглядит так:

$$\chi^2 = \sum \frac{(\text{наблюдаемые частоты} - \text{ожидаемые частоты})^2}{\text{ожидаемые частоты}}.$$

Понятно, что наблюдаемые частоты мы можем найти в клетках таблицы сопряженности. Но что понимается под ожидаемыми, точнее, теоретически ожидаемыми частотами? Ожидаемые частоты — это те частоты, которые должны были бы стоять в клетках той же таблицы сопряженности, если бы две интересующие нас переменные были независимы, то есть расслоение наблюдений по одному признаку оставалось бы пропорциональным для разных подгрупп, выделенных по другому признаку.

Пусть, например, данные относительно участия в парламентских выборах для 1000 опрошенных позволили построить таблицу 8.5

Для приведенных в таблице 8.5 данных гипотеза (или модель) независимого поведения признаков предполагала бы, что

Таблица 8.5

Участие в выборах и пол

Участие в выборах	Женщины	Мужчины	Всего
Участвовали	200	500	700 (70%)
Не участвовали	200	100	300 (30%)
Всего	400	600	1000 (100%)

в мужской и женской подгруппах пропорция участия и неучастия в выборах должна была бы сохраняться такой же, как и для всей выборки в целом (разумеется, в пределах выборочной ошибки). Например, для женщин число участвовавших в выборах с учетом их доли в выборке (равной $\frac{400}{1000}$) составило бы $\frac{400 \times 700}{1000}$, то есть 280 проголосовавших. Отсюда автоматически следует, что до избирательных участков не дошли бы 120 дам (то есть $400 - 280$). Ожидаемая частота голосования для мужчин составила бы $\frac{600 \times 700}{1000} = 420$. Соответственно, не проголосовали бы 180 мужчин. Для модели независимости признаков таблица сопряженности выглядела бы так:

Таблица 8.6

Ожидаемые частоты для распределения участия в выборах по полу (рассчитанные в соответствии с моделью независимости признаков)

Участие в выборах	Женщины	Мужчины	Всего
Участвовали	280	420	700
Не участвовали	120	180	300
Всего	400	600	1000

Сравнив таблицы 8.5 и 8.6, мы видим, что многое во второй из них «осталось как было». Маргиналы таблицы, то есть общее ко-

личество мужчин и женщин, проголосовавших и не проголосовавших, остались, естественно, неизменными. Отличается лишь теоретически ожидаемые частоты в клетках таблицы 8.6. «Хи-квадрат» как раз и оценивает суммарную величину отклонения наблюдаемых значений от ожидаемых («взвешенную» относительно ожидаемых частот). Для данных таблицы 8.5 величина «хи-квадрат» составит 136,128 (проверьте самостоятельно, используя данные таблицы 8.6). Это явно много, но чтобы оценить существенность, значимость полученной величины, следует воспользоваться специальными таблицами⁶. Отметим, что для того, чтобы найти табличное значение, нужно определить так называемое *число степеней свободы*. В рассматриваемом примере оно равно единице, так как все теоретически ожидаемые частоты в таблице 8.6 — при заданных маргиналах — можно получить, вычислив лишь одну из них. Если бы размерность таблицы была 4×4 (по четыре номинальные градации для каждого признака), то оценка «хи-квадрат» производилась бы для $(4 - 1) \times (4 - 1) = 9$, то есть 9 степеней свободы. Обсуждавшийся выше коэффициент ϕ — это просто квадратный корень нормированного относительно численности выборки «хи-квадрата». Удобства коэффициента ϕ очевидны: его легче вычислить, не прибегая к расчету ожидаемых частот, к тому же его величина меняется в пределах от 0 до 1. (Попробуйте рассчитать значение для данных таблицы 8.5.) Существуют и другие коэффициенты взаимосвязи (сопряженности) признаков, основанные на величине «хи-квадрат», например V Крамера, T Чупрова.

Другой тип коэффициентов взаимосвязи номинальных (и не только номинальных) переменных называют *мерами «пропорционального уменьшения ошибки»*. Все они основаны на следующем предположении (или модели): если две переменные взаимосвязаны, мы можем предсказать значение одной переменной для данного наблюдения (случая), зная, какое значение принимает другая переменная. Степень соответствия такого предсказания действительности и используется в качестве коэффициента взаимосвязи. Любой коэффициент взаимосвязи, основанный на модели «пропорционального

⁶ См., например, приложение 5 в книге: Дружинин Н. К. Логика оценки статистических гипотез. М.: Статистика, 1973.

уменьшения ошибки» («ПУО»), имеет общую структуру, задаваемую формулой:

$$\text{Мера «ПУО»} = \frac{E_1 - E_2}{E_1},$$

где E_1 — количество ошибок в предсказаниях значений зависимой переменной, сделанных *без учета* распределения по второй, независимой переменной, а E_2 — количество ошибок в предсказаниях значений зависимой переменной, сделанных *на основе значений независимой* переменной. Конкретные коэффициенты, основанные на «ПУО», будут различаться в зависимости от того, что мы считаем ошибкой и как подсчитывается количество ошибок. В качестве примера можно рассмотреть «*тау-коэффициент*» Гудмана–Краскела⁷. Ошибкой в данном случае считается просто ошибочная классификация наблюдения, отнесение его в «неправильную» категорию. Рассмотрим таблицу сопряженности для приводимого Мюллером и соавторами примера⁸ гипотетических данных о влиянии типа правонарушения на характер решения суда (табл. 8.7).

Ошибка предсказания зависимой переменной (приговор), сделанного исключительно на основе ее собственного распределения, то есть без учета распределения независимой переменной, опреде-

Т а б л и ц а 8.7

Взаимосвязь характера правонарушения и решения суда

Правонарушение	Приговор			Всего
	Штраф	Условный приговор	Тюремное заключение	
Автомобильная кража	5	30	5	40
Кража со взломом	0	30	20	50
Подделка денег	5	0	5	10
Всего	10	60	30	100

⁷ См.: Mueller J. H., Schuessler K. F., Costner H. L. Statistical Reasoning in Sociology. 3rd ed. Boston: Houghton Mifflin Co., 1977. P. 196–205.

⁸ Ibid. P. 197.

ляется следующим образом. Мы знаем (см. маргиналы столбцов в нижней строчке таблицы), что в 60 случаях из 100 приговор был условным, но нам неизвестно, *в каких именно* шестидесяти случаях он был условным. Точно так же мы знаем, что в десяти случаях судья ограничился денежным штрафом, но мы наверняка неоднократно ошибемся, наугад определяя для каждого случая из 100, считать ли его одним из десяти «штрафных». Если бы каждому случаю соответствовала карточка с надлежащей надписью, которую мы с завязанными глазами помещали бы в одну из трех стопок, то при угадывании мы могли бы руководствоваться лишь значениями маргиналов по столбцам: в конечном счете в первой стопке должно оказаться 10 карточек, во второй — 60, а в третьей — 30.

Если мы наугад поместим во вторую стопку «условных приговоров» 60 карточек, то для каждой отдельной карточки (для каждого наблюдения) вероятность ошибки будет равна вероятности попадания туда карточки «штраф» или «тюремное заключение»,

то есть $\frac{10}{100} + \frac{30}{100} = \frac{40}{100}$. Иными словами, в среднем мы сделаем

$\frac{60 \times 40}{100} = 24$ ошибки для категории «условный приговор». Для пер-

вой категории («штраф») мы в среднем сделаем $10 \times \left(\frac{60}{100} + \frac{30}{100} \right) = 9$

ошибок. Для категории «тюремное заключение» (30 карточек) мы можем ожидать, что сделаем 21 ошибку. Суммарное значение числа ошибок предсказания E_1 (если в расчет принимается только распределение зависимой переменной) составит сумму этих трех значений:

$$E_1 = 24 + 9 + 21 = 54 \text{ ошибки.}$$

Представим теперь, что, распределяя карточки по трем категориям приговора, мы располагаем сведениями о том, каково значение второй переменной — «характер преступления» — для каждой карточки, то есть для каждого наблюдения. Пусть, например, кто-нибудь каждый раз сообщает нам, каким было в данном случае правонарушение, предоставляя нам возможность самостоятельно предсказать приговор суда. Мы также знаем заранее, что 5 (12,5%)

автомобильных краж из 40 повлекли за собой штраф, 30 (75%) — условный срок, а еще 5 (12,5%) — тюремное заключение.

Нам, однако, предстоит угадать, какие именно из этих 40 случаев автомобильных краж попали в каждую из трех описанных категорий приговора. Процесс подсчета числа ошибок при таком угадывании сходен с вышеописанным. Зная, каково распределение наблюдений в строке «автомобильные кражи», мы можем оценить ожидаемые ошибки. Ожидаемая ошибка при случайном помещении 5 карточек с автомобильными кражами (из 40) в категорию

«штраф» составит $\frac{35}{40} \times 5 = 4,375$ ошибки; при случайном размещении 30 карточек с автомобильными кражами в категорию «условный приговор» мы ожидаем, что ошибок предсказания в среднем будет $\frac{10}{40} \times 30 = 7,5$ ошибки и т. д. Размещая 5 фальшивомонетчи-

ков из 10 в стопку «штрафов», мы сделаем $\frac{5}{10} \times 5 = 2,5$ ошибки.

Проведя аналогичные подсчеты для всех трех строк таблицы 8.7 и просуммировав все ожидаемые ошибки, мы получим величину E_2 , то есть ожидаемое число ошибок в предсказаниях приговора суда, сделанных с учетом информации о характере преступления (независимой переменной). Для данных, приведенных в таблице 8.7, величина E_2 составит 45,25. Отсюда,

$$\tau = \frac{E_1 - E_2}{E_1} = \frac{54 - 45,25}{54} = 0,16.$$

Для простейшего случая таблицы сопряженности 2×2 существует более простая в вычислительном отношении формула:

$$\tau = \frac{(ab - bc)^2}{(a + b)(c + d)(a + c)(b + d)},$$

где a, b, c, d — частоты в клетках таблицы (см. таблицу 8.4)⁹.

⁹ Желательно не путать обсуждаемый здесь коэффициент сопряженности «тау» Гудмана–Краскела с коэффициентом ранговой корреляции «тау», предложенным Кендаллом. Отметим также, что $\tau = \varphi^2$.

Отметим здесь, что направление связи далеко не всегда очевидно, то есть не всегда можно уверенно утверждать, какая из переменных является зависимой. Если исследователь решит, что независимой является переменная, расположенная по горизонтали (а не по вертикали, как в нашем примере), он сможет подсчитать другую величину «тау-коэффициента», на этот раз идя «от строк» и выполнив все операции в обратном порядке. (Для четырехклеточных таблиц величины «тау» по строкам и по столбцам будут равны.)

Примером ПУО-коэффициента, специально предназначенного для измерения связи двух *ординальных* (то есть измеренных на порядковом уровне) переменных, может служить *коэффициент «гамма»*. «Гамма» измеряет относительное уменьшение ошибки предсказания ранга конкретного наблюдения по зависимой переменной. Для того чтобы вручную рассчитать значение «гаммы» для небольшой выборки, нужно упорядочить наблюдения по независимой и зависимой переменным, как показано в таблице 8.8 для данных о внешней привлекательности (экспертные оценки) и популярности школьников (данные опроса одноклассников).

Далее нужно сравнивать случаи (то есть школьников) попарно, определяя, сходятся или расходятся порядок расположения двух этих случаев по двум переменным. Если упорядочения сходятся, пара называется *согласованной*, если они не сходятся, то пару нужно считать *несогласованной*. Результаты анализа для данных таблицы 8.8 представлены в таблице 8.9.

Т а б л и ц а 8.8

Ранги четырех школьников по привлекательности (X)
и популярности (Y)

Случай	Переменная X (ранг по привлекательности)	Переменная Y (ранг по популярности)
Ольга	1	1
Светлана	2	3
Марьяна	3	2
Натasha	4	4

Таблица 8.9

Попарные сравнения рангов по переменным X и Y

Пара	Порядок по X*	Порядок по Y*	Знак пары («+» — согласованная, «-» — несогласованная)
Ольга — Светлана	O > C	O > C	+
Ольга — Марьяна	O > M	O > M	+
Ольга — Наташа	O > H	O > H	+
Светлана — Марьяна	C > M	M > C	-
Светлана — Наташа	C > H	C > H	+
Марьяна — Наташа	M > H	M > H	+

Здесь использованы лишь начальные буквы имен, то есть «O > C» означает, что ранг Ольги выше ранга Светланы.

Предполагается, что если согласованных (то есть правильно предсказывающих порядок по зависимой переменной) пар больше, чем несогласованных, связь между переменными велика. Если несогласованных пар больше, то связь отрицательна (чем выше ранг по одной переменной, тем ниже ранг по другой). Если же различие между числом согласованных и несогласованных пар невелико, то связь между переменными просто отсутствует. Поэтому формула для «гаммы» такова:

$$\gamma = \frac{N_s - N_r}{N_s + N_r},$$

где N_s — число согласованных пар, N_r — число несогласованных пар.

Для данных, используемых в нашем примере:

$$\gamma = \frac{5-1}{5+1} = 0,67.$$

О том, как измерить связь (корреляцию) количественных переменных, мы поговорим немного позже, сделав одно важное отступление.

Метод уточнения в анализе связи между переменными

Обнаружив наличие взаимосвязи между двумя переменными и оценив интенсивность этой связи с помощью какого-либо коэффициента, исследователь стремится проинтерпретировать эту взаимосвязь в терминах причин и следствий. Иными словами, конечной целью измерения взаимосвязи между переменными является подтверждение (или опровержение) каких-то содержательных предположений, касающихся *причинного механизма*, порождающего найденную взаимосвязь. Однако, как уже говорилось, само по себе наличие связи между двумя переменными еще не доказывает, что эта связь может быть описана моделью «причина—следствие». (А нулевой коэффициент сопряженности — еще не свидетельство отсутствия всякой причинной зависимости.)

Необходимо, во-первых, найти подтверждения того, что связь не является обратной. Если, например, мы обнаружили высокую корреляцию между полученным образованием и престижностью профессии или между алкоголизмом у родителей и алкоголизмом у детей, то таким подтверждением служит естественная упорядоченность событий: обучение обычно предшествует работе, а проблемы родителей — проблемам детей. Во-вторых, нужно исключить альтернативные объяснения обнаруженной взаимосвязи. Во многих случаях существуют вполне правдоподобные гипотезы, объясняющие найденную зависимость *воздействием третьей переменной* (или нескольких переменных). Возможно, например, что на избирательную активность (см. с. 258) влияет не столько пол избирателя, сколько его доход. Так как оплата труда женщин в среднем ниже, чем мужчин, женщины реже проявляют политическую активность. Соответственно, сравнение женщин, имеющих высокооплачиваемую работу, и мужчин в этом случае не выявит никаких различий в отношении к выборам.

Возьмем другой пример. В исследовании было показано, что существует сильная взаимосвязь между престижностью учебного заведения, где было получено высшее образование, и престижностью работы. Значит ли это, что при найме на работу потенциальные работодатели принимают во внимание рейтинг вуза, в котором

проходил обучение соискатель? Вполне возможно. Но даже основываясь исключительно на здравом смысле, легко найти и другие объяснения обнаруженному факту. Может быть, шансы окончить престижное учебное заведение во многом зависят от социально-экономического статуса родителей? Не исключено также, что при устройстве на работу «папины связи» играют столь же существенную роль. В этом случае исходная простая модель «престижное образование → престижная работа» требует уточнения и дополнения: и качество образования, и успешность карьеры зависят от социально-экономического статуса родителей. Заметьте, что такое уточнение вовсе не отменяет исходного факта — эмпирической взаимосвязи между образованием и карьерой, — оно лишь вводит более сложную модель причинной связи, показывая механизм воздействия *третьей переменной* (статуса родителей).

Классический подход к анализу взаимосвязи с введением *дополнительных, контрольных переменных* в социологии и сопредельных дисциплинах получил название *метода уточнения*. Метод уточнения был детально разработан в 1940–50-е гг. П. Лазарсфельдом, С. Стауффером, П. Кендалл и их сотрудниками для анализа элементарных таблиц сопряженности и взаимосвязей номинальных признаков¹⁰. Однако общая логика этого подхода используется, как мы увидим позднее, и в более сложных техниках статистического анализа, и при изучении количественных данных.

Для того чтобы произвести уточнение причинной модели, нужно сделать какие-то содержательные предположения о том, является ли контрольная (третья) переменная *предшествующей* либо *опосредующей*. Если контрольная переменная предшествует во времени и независимой, и зависимой переменным, то она воздействует на них как общая причина, порождая эмпирическую взаимосвязь между переменными. Эта взаимосвязь, однако, не является причинной свя-

¹⁰ См.: *Stouffer S. A. et al. The American Soldier*. Princeton: Princeton University Press, 1949. Vol. 1; *Kendall P. L., Lazarsfeld P. F. Problems of Survey Analysis // Merton R. K., Lazarsfeld P. F. (eds.) Continuities in Social Research: Studies in the Scope and Method of the «American Soldier»*. N. Y.: Free Press, 1950. P. 133–196. Существенные дополнения см. в: *Rosenberg M. The Logic of Survey Analysis*. N. Y.: Basic Books, 1968; *Ядов В. А. Социологическое исследование: методология, программа, методы*. М.: Наука, 1987. С. 190–195.

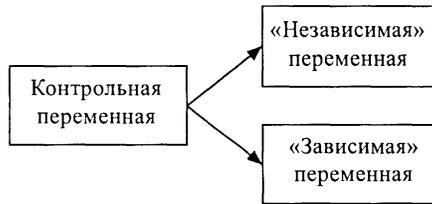


Рис. 8.6. Модель «ложной взаимосвязи»

зью, так как объясняется влиянием третьей, контрольной переменной. Причинная модель для этого случая, часто обозначаемого как «ложная взаимосвязь», приведена на рис. 8.6.

Предположим, что нам удалось установить, что 79% людей, регулярно посещающих врача, оценивают свое самочувствие как «плохое», тогда как среди людей, посещающих врача реже одного раза в год, доля оценивших таким образом свое самочувствие составила 15%. Если принять установленную взаимосвязь за собственно причинную, мы придем к несколько необычному выводу: чем чаще человек посещает докторов, тем хуже он себя чувствует. Предположим, однако, что мы имеем возможность проверить альтернативную гипотезу: люди, страдающие хроническими болезнями, и чаще обращаются за медицинской помощью, и больше подвержены плохому самочувствию. Для того чтобы узнать, сохранится ли исходная взаимосвязь «регулярные посещения врача → плохое самочувствие» при введении контрольной переменной, нам нужно построить так называемые *условные* (иногда — *частные*) *таблицы сопряженности*, где разные группы сравнивались бы при одном (постоянном) уровне объясняющего фактора. Иными словами, нужно построить одну условную таблицу «посещение × самочувствие» для людей, страдающих хроническими болезнями, и другую таблицу — для здоровых. В каждой из этих таблиц объясняющая переменная будет *поддерживаться на постоянном уровне*. Пусть, например, мы получим две частные таблицы, объединенные в таблице 8.10.

Анализ этих двух частных таблиц показывает, что частота посещений врача не оказывает сколько-нибудь заметного влияния на общую оценку самочувствия. Иными словами, метод уточнения в

Т а б л и ц а 8.10

Зависимость общего самочувствия от лечения
при контроле хронической заболеваемости, % ($N = 1000$ чел.)

Самооценка общего самочувствия	Больные		Здоровые	
	регулярно посещающие врача	редко посещающие врача	регулярно посещающие врача	редко посещающие врача
Хорошее	20	18	88	87
Плохое	80	82	12	13

данном примере позволил продемонстрировать, что исходно установленная эмпирическая сопряженность признаков является ложной и может получить *объяснение* при введении контрольной переменной.

В том случае, когда контрольная переменная *опосредует* исходное взаимоотношение¹¹ двух переменных, метод уточнения позволяет выявить собственно механизм влияния независимой переменной на зависимую (см. рис. 8.7).

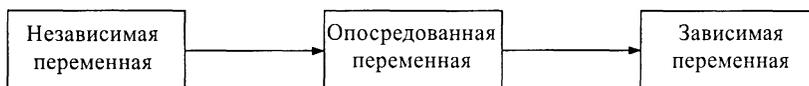


Рис. 8.7. Модель с опосредующей переменной

В таких случаях говорят о том, что контрольная переменная *интерпретирует* исходную взаимосвязь. Например, исследователь, установивший влияние образовательного уровня родителей на успехи детей в учебе, должен показать, каков механизм такого влияния. В частности, он может предположить, что образованные родители внимательнее следят за интеллектуальным развитием своих детей, активнее стимулируют любые успехи в этой сфере. Если же сравнить учебные успехи тех детей, родители которых занимают «активно-стимулирующую» позицию, то различия в успеваемости

¹¹ Это исходное взаимоотношение иногда называют *отношением нулевого порядка*, а модели, получаемые при введении второй, третьей и т. д. контрольных переменных, — отношениями второго, третьего и т. д. порядка.

между детьми более образованных и менее образованных родителей будут несущественными. Заметим, однако, что здесь исходное отношение не исчезает (как в случае ложной взаимосвязи), а лишь проясняется, получает дополнительную *интерпретацию* в терминах опосредующей переменной. Для «стимулируемых» детей учебные успехи не зависят от уровня образования родителей. То же отношение верно и для «нестимулируемых» детей.

Иногда в результате уточнения исходной модели в одной из частных таблиц сохраняется высокий уровень взаимосвязи двух переменных, а в другой таблице взаимосвязь уменьшается или исчезает, то есть коэффициент сопряженности приближается к нулю. В этом случае говорят о *спецификации* исходной модели: введение третьей переменной позволяет определить специфические условия, при которых наблюдается установленное ранее отношение двух переменных. Например, исследователь может обнаружить, что в центральноафриканских деревенских общинах частота коллективных жертвоприношений местным духам зависит от среднемесячного количества осадков. Очевидное объяснение заключается в том, что люди тем чаще обращаются за помощью к сверхъестественным силам, чем больше они нуждаются в дожде. Можно также предположить, что исходная взаимосвязь «засуха — коллективные жертвоприношения» будет менее значительной для тех традиционных сообществ, которые располагают устойчивыми ресурсами пресной воды (например, водой из близлежащей реки или озера) и, следовательно, не испытывают столь сильной зависимости от атмосферных осадков. В этом случае частная таблица сопряженности для деревенских общин, живущих вдали от постоянных источников пресной воды, покажет исходный или более высокий уровень взаимосвязи между засухами и жертвоприношениями, тогда как во второй частной таблице, построенной для речных или приозерных деревень, эта взаимосвязь окажется нулевой.

Анализ таблиц сопряженности и метод уточнения — это наглядные и достаточно эффективные средства, используемые в проверке гипотез о взаимозависимости переменных. Однако этим подходам присущи определенные ограничения. Самые существенные из таких ограничений связаны, во-первых, с тем, что, проводя перегруппировку количественных переменных в номинальные или

ординальные (то есть разбивая доход на «высокий» и «низкий», а интеллект — на «средний» и «выше среднего»), мы теряем существенную информацию о вариации признака *внутри* качественных градаций, *внутри* клеточек таблицы сопряженности, хотя эта информация содержится в «сырых» данных. Кроме того, для уточнения исходной причинной модели нам может потребоваться не одна, а две или четыре дополнительные переменные. Однако с введением новых контрольных переменных число частных таблиц сопряженности будет возрастать по степенному закону. Даже если все наши переменные будут иметь лишь две градации, общее количество клеток в частных таблицах сопряженности будет возрастать как степень двух, то есть, скажем, при четырех контрольных дихотомических переменных нам придется иметь дело с 64-клеточной общей таблицей сопряженности. Соответственно, число наблюдений, «случаев», приходящихся на каждую клетку таблицы, будет уменьшаться, а получаемые нами результаты окажутся более подверженными влиянию случайной ошибки выборки.

По этим причинам многие исследователи используют несколько более сложные статистические методы анализа, свободные от описанных ограничений.

Корреляция, частная корреляция, регрессия

При анализе связи между переменными, измеренными на интервальном уровне, часто используют графическое представление такой связи, называемое *диаграммой рассеивания*. На диаграмме рассеивания каждое наблюдение, то есть каждый «случай», изображается точкой в двумерной системе координат. Значение независимой переменной для данного наблюдения определяет положение соответствующей точки относительно оси X , а значение зависимой переменной задает вторую координату точки — по оси Y . Иными словами, перпендикуляр, опущенный из точки-«случая» на ось X , соответствует измеренному уровню независимой переменной, тогда как перпендикуляр, опущенный на ось Y , будет точно соответствовать наблюдавшемуся уровню зависимой переменной.

Пусть, например, мы располагаем данными о бюджетах 10 партий и о числе полученных этими партиями мест в парламенте.

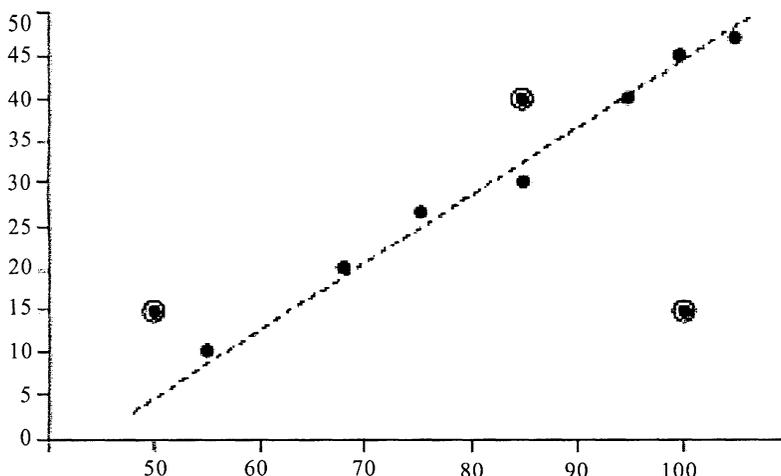


Рис. 8.8. Диаграмма рассеивания, отражающая связь величины партийного бюджета в млн руб. (X) с количеством мест в парламенте (Y) для 10 политических партий

Исходя из гипотезы о влиянии размера партийного бюджета (X) на полученное в результате выборов число депутатских мандатов (Y), мы можем построить диаграмму рассеивания, подобную изображенной на рис. 8.8.

Каждая точка на рисунке соответствует одной из десяти партий. Невзирая на некоторые «аномальные» случаи, подобные обведенным кружками, диаграмма довольно ясно показывает, что всякое приращение в размерах партийной кассы (то есть сдвиг вправо по оси X) влечет за собой увеличение парламентского представительства (сдвиг вверх по оси Y). Между переменными X и Y существует *линейное отношение*: если одна переменная возрастает по величине, то это же происходит и с другой. Помимо указания на природу отношения двух переменных, диаграмма на рис. 8.8 позволяет также сделать некоторые предположения об *интенсивности*, силе этого отношения. Очевидно, что чем более компактно, «сгруппировано» располагаются точки-наблюдения вокруг пунктирной прямой линии (описывающей *идеальное линейное отношение* X и Y), тем сильнее зависимость. На рис. 8.9 приведены еще три диаграммы рассеивания.

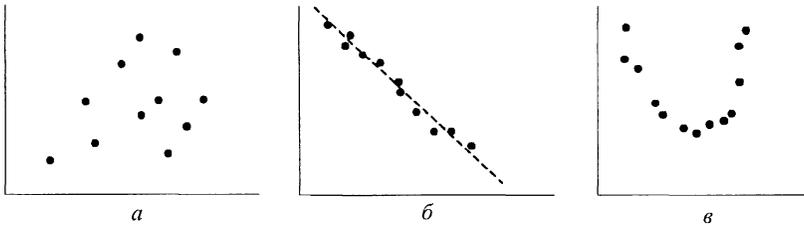


Рис. 8.9. Диаграммы рассеивания для гипотетических данных

Очевидно, что на рис. 8.9 *а* какая-либо связь между X и Y попросту отсутствует. На рис. 8.9 *б* воображаемая прямая линия (отмечена пунктиром) пересекла бы диаграмму сверху вниз, из левого верхнего в правый нижний угол. Иными словами, линейная связь в этом случае имеет обратное направление: чем больше X , тем меньше зависимая переменная Y . Заметим также, что «кучность» расположения точек вдоль воображаемой прямой на рис. 8.9 *б* не очень велика, а значит, и связь (корреляция) между переменными не только обратная, отрицательная, но еще и не очень сильная, умеренная. Наконец, на рис. 8.9 *в* зависимую и независимую переменную связывает явно нелинейное отношение: воображаемый график нисколько не похож на прямую линию и напоминает скорее параболу¹². Отметим, что методы анализа, о которых сейчас пойдет речь, не годятся для этого нелинейного случая, так как обычная формула для подсчета коэффициента корреляции даст нулевое значение, хотя связь между переменными существует.

Существует обобщенный показатель, позволяющий оценить, насколько связь между переменными приближается к линейному функциональному отношению, которое на диаграмме рассеивания выглядит как прямая линия. Это *коэффициент корреляции*, измеря-

¹² Именно так обычно выглядит зависимость между благожелательностью установки по отношению к некоторому объекту (X) и интенсивностью установки (Y): люди, занимающие крайне благожелательную или крайне неблагоприятную позицию в каком-то вопросе, обычно оценивают свои убеждения как более выраженные и интенсивные, чем те люди, чьи установки лежат в области середины, «нейтральных» значений шкалы.

ющий тесноту связи между переменными, то есть их тенденцию изменяться совместно. Как и в рассмотренных выше мерах связи качественных признаков, коэффициент корреляции позволяет оценить возможность предсказания значений зависимой переменной по значениям независимой. Общая формула для вычисления *коэффициента корреляции Пирсона* включает в себя величину *ковариации* значений X и Y . Эта величина (s_{xy}) характеризует совместное изменение значений двух переменных. Она задается как сумма произведений отклонений наблюдаемых значений X и Y от средних

\bar{X} и \bar{Y} соответственно, то есть $\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$, деленная на количество наблюдений. Чтобы понять «физический смысл» ковариации, достаточно обратить внимание на следующее ее свойство: если для какого-то объекта i в выборке оба значения — X_i и Y_i окажутся высокими, то и произведение $(X_i - \bar{X})$ на $(Y_i - \bar{Y})$ будет большим и положительным. Если оба значения (по X и по Y) низки, то произведение двух отклонений, то есть двух отрицательных чисел, также будет положительным. Таким образом, если линейная связь X и Y положительна и велика, сумма таких произведений для всех наблюдений также будет положительна. Если связь между X и Y обратная, то многим положительным отклонениям по X будут соответствовать отрицательные отклонения по Y , то есть сумма отрицательных произведений отклонений будет отрицательной.

Наконец, при отсутствии систематической связи произведения будут иногда положительными, иногда отрицательными, а их сумма (и, следовательно, ковариация X и Y) будет, в пределе, равна нулю. Таким образом, ковариация показывает величину и направление связи, совместного изменения X и Y . Если разделить ковариацию s_{xy} на стандартные отклонения s_x и s_y (чтобы избавиться от влияния масштаба шкал, в которых измеряются X и Y), то мы получим искомую формулу *коэффициента корреляции Пирсона* (r_{xy}):

$$r_{xy} = \frac{s_{xy}}{s_x s_y}.$$

Более удобная для практических вычислений расчетная формула выглядит так:

$$r_{xy} = \frac{n \sum X_i Y_i - (\sum X_i)(\sum Y_i)}{\sqrt{(n \sum X_i^2 - (\sum X_i)^2)(n \sum Y_i^2 - (\sum Y_i)^2)}}.$$

Несмотря на несколько устрашающий вид, расчетная формула очень проста. Для «ручного» вычисления r_{xy} вам понадобятся лишь пять величин: суммы значений по X и Y ($\sum_{i=1}^n X_i$ и $\sum_{i=1}^n Y_i$), суммы квадратов значений по X и Y ($\sum_{i=1}^n X_i^2$ и $\sum_{i=1}^n Y_i^2$), суммы произведений X и Y по всем объектам-«случаям» ($\sum_{i=1}^n X_i Y_i$). В таблице 8.11 приведены данные о максимальных дневных и ночных температурах, зарегистрированных в 10 городах¹³.

Просуммировав значения в столбцах, мы получим $\sum_{i=1}^{10} X_i = 258$ и $\sum_{i=1}^{10} Y_i = 156$. Возведя каждое из значений X и Y в квадрат и просуммировав, мы найдем, что $\sum_{i=1}^{10} X_i^2 = 7180$ и $\sum_{i=1}^{10} Y_i^2 = 2962$. Сумма попарных произведений X_i и Y_i ($\sum_{i=1}^{10} X_i Y_i$) составит 4359. Вы можете самостоятельно убедиться в том, что подстановка всех значений в расчетную формулу даст (надеюсь) величину $r_{xy} = 0,91$. Иными словами, корреляция между дневными и ночными температурами воздуха очень высока, но все же отлична от 1,0 (коэффициент корреляции может меняться в пределах от $-1,0$ до $+1,0$). Это отличие, вероятно, объясняется влиянием других факторов (продолжительность дня и ночи, облачность, географическое положение и т. п.). Судя по полученной величине корреляции, знание дневных температур позволяет *предсказывать* ночные температуры с очень высокой точностью, но не безошибочно.

¹³ Погода (Гидрометцентр РФ) // Сегодня. 1994. 23 авг.

Т а б л и ц а 8.11

**Максимальные дневные и ночные температуры воздуха
в некоторых городах**

Город	Дневная температура (X)	Ночная температура (Y)
Лондон	16	11
Париж	21	12
Стокгольм	20	12
Варшава	25	14
Бонн	25	16
Рим	36	23
Тель-Авив	31	23
Анкара	32	15
Каир	36	22
Москва	16	8
$N = 10$		

Величина, которая равна квадрату коэффициента корреляции Пирсона, то есть r^2 , имеет ряд интересных статистических свойств. Отметим сейчас, что r^2 является ПУО-мерой связи, подобной обсуждавшимся выше (см. с. 259–264). Можно показать, что она характеризует ту долю дисперсии значений Y , которая объясняется наличием корреляции между X и Y . (Естественно, величина r^2 будет всегда положительной и не может превзойти по абсолютной величине коэффициент корреляции.)¹⁴ Та часть разброса в значениях Y , которая не может быть предсказана по значениям X , — это *дисперсия ошибки нашего прогноза*, то есть $1 - r^2$. *Необъясненный разброс в значениях Y присутствует в том случае, когда при равных уровнях X (ср., например, дневные температуры в Варшаве и Бонне из таблицы 8.11) сохраняются различия в значениях Y .*

Коэффициент корреляции позволяет оценить степень связи между переменными. Однако этого недостаточно для того, чтобы

¹⁴ Подробный анализ можно найти в большинстве руководств по прикладной статистике. Здесь мы ограничимся обсуждением общей логики оценки *объясненной дисперсии*.

непосредственно преобразовывать информацию, относящуюся к одной переменной, в оценки другой переменной. Допустим, мы выяснили, что коэффициент корреляции между переменными «величина партийного бюджета» и «число мест в парламенте» равен 0,8. Можем ли мы теперь предсказать, сколько мест в парламенте получит партия, годовой бюджет которой равен 100 млн рублей? Похоже, что знание величины коэффициента корреляции нам здесь не поможет. Однако мы можем вспомнить, что коэффициент корреляции — это еще и оценка соответствия разброса наших наблюдений той идеальной модели линейного функционального отношения, которое на рассмотренных выше диаграммах рассеивания (см. рис. 8.8–8.9) представлено пунктирными прямыми. Эти линии называют *линиями регрессии*.

Если бы все наблюдения аккуратно «укладывались» на линию регрессии, то для предсказания значения зависимой переменной достаточно было бы восстановить перпендикуляр к оси Y из той точки прямой, которая соответствует известному значению X .

На рис. 8.10 показано, как можно было бы графически определить ожидаемые значения Y для гипотетического примера с партийной кассой и местами в парламенте. (Разумеется, найти искомое значение Y можно и без линейки, с помощью вычислений, если известен угол наклона регрессионной прямой и точка пересечения с осью ординат.)

Как говорилось выше, линия регрессии не обязательно должна быть прямой, но мы ограничимся рассмотрением самого простого

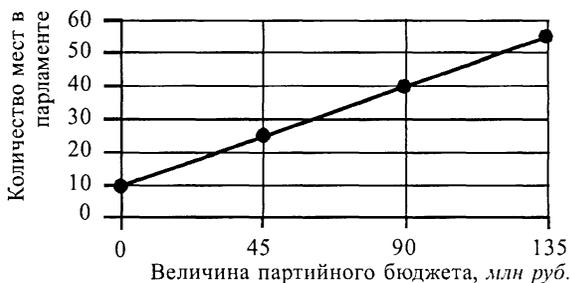


Рис. 8.10. Предсказание значения Y по значению X для гипотетических данных

случая линейной зависимости (нелинейные связи во многих случаях также могут быть приближенно описаны линейными отношениями).

Существуют специальные статистические процедуры, которые позволяют найти регрессионную прямую, максимально соответствующую реальным данным. Регрессионный анализ, таким образом, дает возможность *предсказывать значения Y по значениям X с минимальным количеством ошибок*. В общем виде уравнение, описывающее прямую линию регрессии Y по X , выглядит так:

$$\hat{Y} = a_{yx} + b_{yx}X,$$

где \hat{Y} — это предсказываемое значение по переменной Y (в только что рассмотренном примере — количество мест в парламенте), a — это точка, в которой прямая пересекает ось Y (то есть значение Y для случая, когда $X = 0$), и b — коэффициент регрессии, то есть наклон прямой. Часто удобно измерять обе переменные не в «сырых» шкалах, а в единицах отклонения от среднего. Процедура *стандартизации*, то есть перевода исходной шкалы в стандартные Z -оценки, вам уже известна (см. с. 249). Преимущество использования стандартизованных переменных в регрессионном анализе заключается в том, что линия регрессии в этом случае проходит через начало координат. Стандартизованный коэффициент регрессии (наклон прямой) обозначают обычно греческой буквой β (либо b^*). Правда, в отличие от b -коэффициента, β не позволяет прямо заключить, на какое количество исходных единиц возрастет Y при увеличении X на одну единицу (например, насколько увеличится число депутатских мандатов при увеличении бюджета на 1 млн рублей или насколько увеличивается заработная плата при увеличении стажа работы на один год). С другой стороны, β позволяет сопоставить влияние на независимую переменную контрольных переменных, измеренных в разных шкалах.

Социологи обычно осуществляют регрессионный анализ, используя возможности распространенных прикладных пакетов компьютерных программ (например, SPSS). В этом случае для нахождения линии регрессии, лучше всего соответствующей данной выборке наблюдений, которая представлена точками на диаграмме

рассеивания, используют метод минимизации взвешенной суммы квадратов расстояний между этими точками и искомой прямой¹⁵.

Хотя здесь не место для обсуждения статистических деталей, мы все же сделаем несколько замечаний, относящихся к осмысленному (или бессмысленному) использованию техники линейной регрессии.

Во-первых, как и в ранее обсуждавшихся примерах анализа связи, наличие координации, согласованности в изменениях двух переменных еще не доказывает, что обнаруженное отношение носит собственно каузальный характер. Проверка альтернативных причинных моделей, иначе объясняющих эмпирическую сопряженность переменных-признаков, может основываться только на содержательных теоретических представлениях.

Далее, нужно помнить о том, что регрессионные коэффициенты в общем случае *асимметричны*. Если мы решим, что это Y , а не X является независимой переменной, то вполне можем рассчитать другую по величине пару коэффициентов — a_{xy} и b_{xy} . (Заметьте, что порядок букв в подстрочном индексе значим: первой всегда идет предсказываемая переменная, а второй — предсказывающая.) Разумеется, при выборе кандидатов в зависимые и независимые переменные также важны не статистические, а содержательные соображения.

Если вернуться к затронутой выше *взаимосвязи между линейной регрессией и корреляцией*, то здесь мы можем сделать следующие дополнения. Пусть все точки-наблюдения аккуратно размещены на регрессионной прямой. Перед нами почти невероятный случай абсолютной линейной зависимости. Зная, например, что коэффициент b (нестандартизированный) равен 313, мы можем ут-

¹⁵ Более детальные сведения можно найти в статистической литературе. Очень доступно проблема излагается, в частности, в кн.: *Гласс Дж., Стенли Дж.* Указ. соч. С. 123–141. Для тех же, кто захочет осуществить «ручную» регрессию для какого-либо из использованных примеров, просто приведем формулы для вычисления нестандартизированных коэффициентов (обозначения те же, что и выше):

$$b_{yx} = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sum (X - \bar{X})^2} = \frac{n \sum XY - (\sum X)(\sum Y)}{n \sum X^2 - (\sum X)^2}; \quad a_{yx} = \bar{Y} - b_{yx} \bar{X}.$$

верждать, что именно такова величина воздействия переменной X на зависимую переменную Y . Кроме того, мы можем точно сказать, что единичная прибавка в величине X вызовет увеличение Y на ту же величину, 313 (если, допустим, X — стаж работы, а Y — зарплата, то с увеличением стажа на год зарплата растет на 313 рублей). В этом случае коэффициент корреляции будет равен в точности 1,0, что свидетельствует о сильном, «абсолютном» характере связи переменных. Различие между предсказанными и наблюдаемыми значениями в этом случае отсутствует. Корреляция как мера точности прогноза показывает, что ошибок предсказания просто нет.

В действительности, однако, из-за влияния других переменных и случайной выборочной ошибки точки-наблюдения обычно лежат выше или ниже прямой, которая, как говорилось, является лишь наилучшим приближением реальных данных. Коэффициент корреляции Пирсона r и величина r^2 по-прежнему служат оценкой точности прогноза, основанного на линии регрессии. Вполне возможны ситуации, когда коэффициент регрессии очень велик, воздействие X на Y просто громадно, но корреляция низка и, следовательно, точность прогноза невелика. Нет ничего необычного и в обратной ситуации: воздействие X на Y относительно мало, а коэффициент корреляции и объясненная дисперсия очень велики. Посмотрев на приведенные выше диаграммы рассеивания, можно легко уяснить себе *смысл отношения между корреляцией и регрессией*: первая имеет прямое отношение к «разбросанности» точек наблюдения (чем выше «разбросанность», тем ниже r^2 и ненадежнее прогноз), тогда как коэффициент регрессии описывает наклон, «крутизну» линии. Однако существующее здесь различие не стоит и преувеличивать: регрессионный коэффициент (наклон прямой) для стандартизованных данных в точности равен коэффициенту корреляции Пирсона r ¹⁶.

¹⁶ Легко понять, что при измерении в единицах стандартного отклонения максимальная связь ($\beta = 1,0$) соответствует ситуации, когда сдвигу от начала координат в 1 ед. стандартного отклонения по X соответствует увеличение Y также на 1 ед. стандартного отклонения. Важно заметить, что в случае стандартизованных переменных (и только в этом случае) коэффициенты регрессии Y по X и X по Y будут совпадать.

Предположим, что исследователь изучает зависимость между образованием матери (X) и образованием детей (Y). Обе переменные измерены как количество лет, затраченных на получение образования. Найдя достаточно высокую корреляцию между X и Y — скажем, равную 0,71, — он также находит коэффициенты регрессии a и b и устанавливает, что r^2 (называемый также коэффициентом детерминации) в данном случае приближенно равен 0,5. Это значит, что доля вариации в значениях переменной Y (образование детей), объясненная воздействием переменной-предиктора X (материнское образование), составляет около 50% общей дисперсии предсказываемой переменной. Коэффициент корреляции между переменными достаточно велик и статистически значим даже для не очень большой выборки. Следовательно, обнаруженная взаимосвязь переменных не может быть объяснена случайными погрешностями выборки. В пользу предложенной исследователем причинной гипотезы говорит и то обстоятельство, что альтернативная гипотеза — образование детей влияет на образовательный статус родителей — крайне неправдоподобна и может быть отвергнута на основании содержательных представлений о временной упорядоченности событий. Однако все еще не исключены те возможности, которые мы обсуждали в параграфе, посвященном методу уточнения. Иными словами, нам следует считаться с вероятностью того, что какая-то другая переменная (или несколько переменных) определяет и образование родителей, и образование детей (например, финансовые возможности либо интеллект). Чтобы проверить такую конкурирующую гипотезу, следует рассчитать так называемую *частную корреляцию*. Логика расчета частной корреляции совпадает с логикой построения частных таблиц сопряженности при использовании метода уточнения. Построить частные таблицы сопряженности для различных уровней контрольной переменной в случае, когда переменные измерены на интервальном уровне, — это практически неразрешимая задача. Чтобы убедиться в этом, достаточно подсчитать, каким должно быть количество таблиц уже при десяти-двенадцати категориях каждой переменной. Расчет коэффициента частной корреляции — это простейшее средство уточнения исходной причинной модели при введении дополнительной переменной. Интерпретация коэффициента частной корреляции не от-

личается от интерпретации частных таблиц сопряженности: *частной корреляцией называют корреляцию между двумя переменными, когда статистически контролируется, или «поддерживается на постоянном уровне», третья переменная (набор переменных).*

Если, предположим, при изучении корреляции между образованием и доходом нам понадобится «вычестъ» из полученной величины эффект интеллекта, предположительно влияющего и на образование, и на доход, достаточно воспользоваться процедурой вычисления частной корреляции. Полученная величина будет свидетельствовать о чистом влиянии образования на доход, из которого «вычтена» линейная зависимость образования от интеллекта.

Мюллер и соавторы¹⁷ приводят интересный пример использования коэффициента частной корреляции. В исследовании П. Риттербэнда и Р. Силберстайна изучались студенческие беспорядки 1968–1969 гг. Одна из гипотез заключалась в том, что число нарушений дисциплины и демонстраций протеста в старших классах учебных заведений связано с различиями показателями академической успеваемости учащихся. Корреляция между частотой «политических» беспорядков и средней успеваемостью оказалась отрицательной (хуже успеваемость — больше беспорядков) и статистически значимой ($r = -0,36$). Однако еще более высокой была корреляция между частотой беспорядков и долей чернокожих учащихся ($r = 0,54$). Исследователи решили проверить, сохранится ли связь между беспорядками и успеваемостью, если статистически проконтролировать влияние расового состава учащихся. Коэффициент частной корреляции частоты беспорядков и успеваемости *при контроле* расового состава учащихся оказался равным нулю. Исходная корреляция между беспорядками и успеваемостью в данном случае может быть описана причинной моделью «ложной взаимосвязи» (см. рис. 8.6): наблюдаемые значения этих двух переменных скоррелированы лишь потому, что обе они зависят от третьей переменной, доли чернокожих в общем количестве учащихся. Чернокожие студенты, как заметили исследователи, оказались восприимчивее к предложенным самыми активными «политиканами» образцам участия в

¹⁷ Mueller J., Schuessler K., Costner H. Statistical Reasoning in Sociology. 3rd ed. Boston: Houghton Mifflin Co, 1977. P. 279–281.

политических беспорядках. Кроме того, их успеваемость, помимо всяких политических событий, была устойчиво ниже, чем средняя успеваемость белых.

Коэффициент частной корреляции между переменными X и Y при контроле дополнительной переменной Z (то есть при поддержании Z «на постоянном уровне») обозначают как $r_{xy.z}$. Для его вычисления достаточно знать величины наблюдаемых попарных корреляций между переменными X , Y и Z (то есть простых корреляций — r_{xy} , r_{yz} , r_{xz}):

$$r_{xy.z} = \frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{(1 - r_{xz}^2)(1 - r_{yz}^2)}}.$$

Как всякая выборочная статистика, коэффициент корреляции подвержен выборочному разбросу. Существует некоторая вероятность того, что для данной выборки будет получено ненулевое значение коэффициента корреляции, тогда как истинное его значение для генеральной совокупности равно нулю. Иными словами, существует задача *оценки значимости* полученных значений корреляций и коэффициентов регрессии, относящаяся к области теории статистического вывода. Описание соответствующих статистических методов выходит за рамки этой книги, поэтому мы рассмотрим лишь самые общие принципы, позволяющие решать описанную задачу в простых случаях и интерпретировать соответствующие показатели при использовании стандартных компьютерных программ.

Прежде всего вероятностная оценка коэффициента корреляции подразумевает оценку отношения к его случайной ошибке. Удобная, хотя и не вполне надежная формула для вычисления ошибки коэффициента корреляции (m_r), выглядит так¹⁸:

$$m_r = \frac{1 - r^2}{\sqrt{n - 1}}.$$

¹⁸ См.: Дружинин Н. К. Логика оценки статистических гипотез. М.: Статистика, 1973. С. 112–114.

Всегда полезно вычислить отношение полученной величины r к его ошибке (то есть r/m_r). В использовавшемся нами примере данных о погоде коэффициент корреляции оказался равен 0,91, а его выборочная ошибка составляет:

$$m_r = \frac{1-0,91^2}{\sqrt{10-1}} = \frac{1-0,8281}{\sqrt{9}} = 0,0573.$$

Отношение r к m_r , обозначаемое как t , составит $\frac{0,91}{0,0573} \approx 15,88$. Разумеется, коэффициент, превосходящий свою случайную ошибку почти в 16 раз, может быть признан значимым даже без построения доверительных интервалов.

Когда значение r не столь близко к единице и выборка невелика, нужно все же проверить статистическую гипотезу о равенстве r нулю в генеральной совокупности. Для этого нужно определить t по формуле:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}},$$

где t — это величина так называемого t -критерия Стьюдента (см. также гл. 4), r — выборочный коэффициент корреляции, n — объем выборки. Для установления значимости вычисленной величины t -критерия пользуются таблицами t -распределения для $(n-2)$ степеней свободы (см. табл. 4.1). Во многих пособиях по статистике можно найти и готовые таблицы критических значений коэффициента корреляции r для данного уровня значимости α . В этом случае отпадает необходимость в каких-либо вычислениях t : достаточно сравнить полученную величину коэффициента корреляции с табличным значением¹⁹. (Например, величина коэффициента корреляции $r = 0,55$ будет существенной на уровне значимости $p = 0,01$ даже для выборки объемом 105, так как критическое значение составляет 0,254.)

¹⁹ См., в частности: Ликеш И., Ляга Й. Основные таблицы математической статистики. М.: Финансы и статистика, 1985. (Табл. 14).

Множественная регрессия и путьевой анализ

Выше описывалась модель линейной регрессии для двух переменных. В действительности социолог довольно редко сталкивается со столь простыми моделями данных. Влияние одного фактора обычно может объяснить лишь часть разброса наблюдаемых значений независимой переменной. Метод частной корреляции позволяет нам проконтролировать эффекты воздействия любых других контрольных переменных, которые мы в состоянии измерить. (Стоит снова подчеркнуть здесь, что статистические методы изучения причинных взаимосвязей, *в отличие от экспериментальных*, позволяют нам контролировать лишь те источники вариации, которые мы способны концептуализировать и измерить.) Однако еще более интересной задачей является контроль одновременного воздействия *нескольких* независимых на *одну* зависимую переменную, а также сравнение эффекта воздействия разных независимых переменных и предсказание «отклика» независимой переменной. Именно эти задачи решают методы анализа, о которых пойдет речь в данном параграфе. Наше изложение будет неполным, так как более детальное обсуждение требует дополнительной математической подготовки. Мы будем ориентироваться на сравнительно скромные цели понимания общей логики и интерпретации результатов соответствующих статистических процедур.

Уравнение множественной регрессии — это определенная *модель порождения данных*. Важные допущения, принимаемые в этой модели, касаются уже известного вам требования *линейности*, а также *аддитивности* суммарного эффекта независимых переменных. Последнее означает, что воздействия разных независимых переменных просто суммируются, а не, скажем, перемножаются (мультипликативный эффект, в отличие от аддитивного, имеет место тогда, когда величина воздействия одной независимой переменной на зависимую, в свою очередь, находится под влиянием другой независимой переменной, то есть независимые переменные взаимодействуют друг с другом).

Множественная регрессия во многом аналогична простой (бивариантной) регрессии. Отличие состоит в том, что регрессия осуществляется по двум и более независимым переменным *одновременно*, причем каждая из них входит в регрессионное уравнение с

коэффициентом, позволяющим предсказать значения зависимой переменной с минимальным количеством ошибок (критерием здесь снова является метод наименьших квадратов). Частные коэффициенты в уравнении множественной регрессии показывают, какой будет величина воздействия соответствующей независимой переменной на зависимую *при контроле* влияния других независимых переменных. Если воспользоваться простейшей системой обозначений, то уравнение множественной регрессии для трех независимых переменных можно записать как:

$$\hat{Y} = a + b_1X_1 + b_2X_2 + b_3X_3,$$

где \hat{Y} — это предсказываемое значение зависимой переменной, $X_1 \dots X_3$ — независимые переменные, а $b_1 \dots b_3$ — частные коэффициенты регрессии для каждой из зависимых переменных.

Коэффициенты b могут быть интерпретированы как показатели влияния каждой из независимых переменных на зависимую при контроле всех других независимых переменных в уравнении. В отличие от коэффициентов частной корреляции, коэффициенты регрессии обладают размерностью. Они показывают, на сколько единиц изменится зависимая переменная при увеличении независимой на одну единицу (при контроле всех остальных переменных модели). Пусть, например, мы построили уравнение множественной регрессии, описывающее зависимость дохода от интеллекта (X_1) и стажа работы (X_2). Если величина b_1 оказалась равной 100, это означает, что каждый дополнительный балл по шкале интеллекта увеличивает доход на 100 рублей. Значение $b_2 = 950$ говорит нам, что год стажа прибавляет 950 рублей. Однако «сырые» оценки интеллекта и стажа измерены в разных единицах. Для определения *сравнительной значимости* независимых переменных, входящих в уравнение множественной регрессии, мы должны подвергнуть все переменные *стандартизации* (то есть перевести их в Z-оценки, см. выше). Стандартизированные коэффициенты множественной регрессии, которые удобнее всего обозначать как b^* (либо греч. «бета» — β), меняются в пределах от $-1,0$ до $+1,0$. Они сохраняют свою величину при изменении масштаба шкалы: переход от измерения возраста в годах к измерению в днях не изменит соответствующий b^* .

Стандартизированные коэффициенты позволяют оценить «вклад» каждой из переменных-предикторов в предсказание значений независимой переменной. Если в примере с влиянием интеллекта и стажа работы на доход окажется, что $b_1^* = 0,25$, а $b_2^* = 0,30$, то можно заключить, что сравнительная значимость «веса» интеллекта и стажа в предсказании дохода различаются незначительно. Если же для одной переменной $b_1^* = 0,80$, тогда как $b_2^* = 0,40$, мы можем сказать, что эффект воздействия второй переменной в два раза меньше эффекта первой.

Чтобы определить ожидаемые значения зависимой переменной для отдельных индивидов, достаточно подставить в уравнение множественной регрессии соответствующие значения переменных-предикторов и вычисленных коэффициентов b . Пусть, например, мы хотим рассчитать прогнозное значение величины дохода для человека, чей коэффициент интеллекта равен 110, а стаж работы — 20 годам. Если b_1 , как в вышеприведенном примере, составляет 100, $b_2 = 950$, а слагаемое $a = 50\,000$, то мы получим:

$$\text{ожидаемый доход} = 50\,000 + 100 \times 110 + 950 \times 20 = 80\,000 \text{ руб.}$$

Множественную регрессию можно использовать и для предсказания *средних групповых значений*, например среднего дохода мужчин-врачей. Единственное различие в данном случае заключается в использовании *средних значений независимых переменных* для подстановки в уравнение множественной регрессии.

В качестве независимой переменной множественной регрессии могут использоваться и дихотомические переменные, которым приписывают значения 0 и 1 (например, пол). Для того чтобы включить в уравнение номинальную переменную с более чем двумя категориями, нужно создать соответствующее число новых, «*фиктивных*» переменных, каждая из которых будет кодироваться как 0 или 1 в зависимости от наличия или отсутствия категории-признака. Скажем, состоящую из трех категорий переменную «цвет глаз» можно представить с помощью трех переменных: X_1 — «голубые глаза», X_2 — «карие глаза», X_3 — «зеленые глаза». (Человек с голубыми глазами получит 1 по X_1 и 0 по двум другим переменным.)

Метод множественной регрессии очень популярен среди социологов. Вот, например, как выглядели результаты его применения в

исследовании Л. Бэрона и М. Строса, изучавших факторы, влияющие на статистику изнасилований²⁰. Используемая в планировании этого исследования матрица данных включала в себя в качестве объектов («случаев») различные штаты США. Признаками, по которым описывались штаты, служили около десятка независимых и собственно контрольных переменных, предположительно воздействующих на зависимую переменную, — количество зарегистрированных полицией изнасилований на 100 000 населения в год для данного штата (по данным ежегодных статистических отчетов ФБР). Предполагалось, что существующие различия между штатами в уровне изнасилований можно будет объяснить различиями в уровнях независимых переменных. Нужно отметить, что разброс «случаев» по зависимой переменной был весьма велик — от 71,9 на Аляске до 8,2 в Северной Дакоте (1979). Из десятка переменных, включенных в уравнение множественной регрессии, девять оказались статистически значимы. Основные результаты регрессионного анализа для семи переменных представлены в таблице 8.12.

Из таблицы видно, что индекс совокупного тиража порнографических журналов (интегральный показатель, учитывающий уровни продаж восьми популярных изданий) имеет коэффициент регрессии 6,99. Это означает, что рост индекса на единицу в среднем увеличивает количество изнасилований почти на 7 случаев (в расчете на 100 000 населения). Весьма значительно и влияние числа убийств, что особенно заметно при сравнении стандартизированных коэффициентов (b^*), не зависящих от шкалы измерения признака. Фактически количество убийств вносит самый значительный «вклад» в предсказание значений зависимой переменной ($b^* = 0,55$). Интересно отметить, что одна из независимых переменных в описываемом исследовании — индекс положения женщин, рассчитанный на основании 22 политических, экономических и социальных индикаторов, — при анализе простых взаимосвязей продемонстрировала практически нулевую корреляцию с количеством изнасилований ($r = 0,17$), причем результаты анализа

²⁰ Baron L., Strauss M. A. Sexual Stratification, Pornography, and Rape in the United States // Malamuth N., Donnerstein E. (eds.) Pornography and Sexual Aggression. Orlando et al.: Academic Press, 1984. P. 185–209.

Таблица 8.12
 Множественный регрессионный анализ статистики изнасилований, 1979 г.²¹

Независимая переменная	Коэффициент <i>b</i>	Коэффициент <i>b*</i>	<i>p</i> <
Индекс совокупного тиража порно-графических журналов (SMCX)	6,99	0,52	0,001
Показатель числа убийств, в т. ч. непредумышленных	1,70	0,55	0,001
Показатель числа публичных оскорблений с угрозой применения физической силы	0,04	0,32	0,001
Индекс положения женщин (SWX)	0,43	0,27	0,014
Число грабежей	-0,03	-0,25	0,052
Процент черного населения	-0,41	-0,38	0,001
Процент живущих ниже федерального уровня бедности	1,11	0,29	0,011

диаграммы рассеивания также не дали никаких свидетельств в пользу гипотезы о нелинейной связи. Множественная регрессия позволила уточнить первоначальные выводы: при контроле прочих переменных модели, чем *выше* статус женщин, тем *выше* уровень изнасилований (результат, которому довольно трудно найти теоретическое объяснение). Использование девяти независимых переменных позволило объяснить 83% дисперсии в показателях количества изнасилований (квадрат коэффициента множественной корреляции r^2 составил 0,83).

При интерпретации результатов множественной регрессии стандартизированные коэффициенты, как уже говорилось, используются в качестве показателей значимости, «вклада» соответствующих переменных. Эта трактовка верна лишь в определенных пределах. При нарушении некоторых условий сравнение абсолютных величин стандартизованных коэффициентов может вести к неверным выводам. Дело в том, что коэффициенты регрессии подверже-

²¹ Таблица приводится в сокращении по источнику: *Baron L., Strauss V. A. Sexual Stratification, Pornography, and Rape...*

ны влиянию *случайных ошибок измерения* (см.: с. 170–171). Использование *ненадежных индикаторов* «сдвигает» регрессионные коэффициенты к нулю²². Иными словами, *более надежные* индикаторы дают *более высокие* оценки коэффициентов. Пусть, например, для предсказания риска сердечно-сосудистых заболеваний использовались две независимые переменные индивидуального уровня — «ориентация на достижения» и «склонность подавлять агрессию», — причем шкала для измерения первой обладала *более высоким* коэффициентом надежности. Если стандартизованный коэффициент регрессии для достиженческой мотивации окажется *выше*, чем для подавления агрессии, это может рассматриваться как следствие таких содержательных различий между переменными, которые важны с точки зрения теории психосоциальных факторов заболеваемости. Но нельзя исключить и альтернативное объяснение, связывающее более высокий регрессионный коэффициент первой переменной с побочными эффектами методов измерения: влияние ориентации на достижения не превосходит влияния, оказываемого на риск инфаркта склонностью подавлять агрессию, а наблюдаемые различия регрессионных коэффициентов связаны лишь с ненадежностью использованных индикаторов склонности к подавлению.

Другая проблема, требующая некоторой осторожности в интерпретации коэффициентов регрессии, возникает вследствие того, что модель множественной регрессии не обязывает нас ни к каким строгим предположениям о причинных связях *между независимыми переменными*. Регрессионное уравнение, образно говоря, не делает никаких различий между собственно независимыми, то есть теоретически специфицированными, переменными и дополнительными — контрольными, опосредующими и т. п. — факторами, вводимыми в модель с целью уточнения. В тех случаях, когда теоретическая гипотеза, проверяемая в ходе исследования, допускает: 1) существование взаимосвязей между независимыми переменными; 2) наличие прямых и косвенных (опосредованных) влияний, а также 3) использование нескольких индикаторов для каждого

²² Это явление называют *аттенуацией*. Существуют специальные методы внесения поправок на аттенуацию, но здесь они обсуждаться не будут.

латентного фактора, могут понадобиться более совершенные статистические методы. Одна из возможностей здесь — это использование *путевого анализа*.

Путевой анализ — один из основных способов построения и проверки причинных моделей в социологии. Многие более продвинутые статистические техники основаны на сходной исследовательской методологии.

Важным достоинством путевого анализа является то, что он позволяет оценить параметры каузальных моделей, причем в расчет принимаются не только *прямые*, но и *непрямые (опосредованные)* влияния. Если, например, в результате корреляционного или регрессионного анализа мы обнаружили, что интеллект (измеренный как IQ) лишь умеренно влияет на доход, нам не следует торопиться с общими выводами. Мы оставили неучтенной возможность того, что интеллект может иметь существенное влияние на образование, которое, в свою очередь, воздействует на последующий доход. Таким образом, нам нужно принять во внимание то, что интеллект — помимо прямого эффекта — может иметь еще и опосредованное, непрямое влияние на доход посредством влияния на образование. Методы, рассматривавшиеся нами до сих пор, описывали только прямые эффекты.

Путевой анализ включает в себя технику представления прямых и косвенных причинных влияний при помощи специальных диаграмм (поточковых графов). Эти диаграммы часто называют просто *причинными (структурными) моделями*.

Последовательно «считывая» такую модель, можно легко определить все пути влияния одной переменной на другую и, соответственно, оценить величину чистого эффекта. Во многих разделах этой книги причинные модели уже использовались для представления сравнительно сложных причинных гипотез (см., например, с. 180), поэтому общая логика их построения не требует детального обсуждения. Порядок представления переменных на диаграмме отражает предполагаемое направление причинной связи, а диапазон включенных в диаграмму переменных и отношения между ними зависят от принятых исследователем теоретических гипотез. Так называемые *путевые коэффициенты*, описывающие связи между переменными (связям соответствуют стрелочки на диаг-

рамме), равны стандартизированным коэффициентам множественной регрессии (b^*)²³.

Обычно путевую диаграмму рисуют слева направо — от самых «ранних» по порядку следования независимых переменных до зависимой. Путевые коэффициенты часто обозначают латинскими « p » с подстрочными индексами (p_{21} — это путевой коэффициент для связи между переменными $X_1 \rightarrow X_2$). На рис. 8.11 в качестве примера изображена путевая диаграмма, отражающая гипотетические отношения между интеллектом (X_1), образованием (X_2), социально-экономическим статусом (X_3), доходом (X_4) и размерами сбережений (X_5).

Специальные правила позволяют перевести отношения, изображенные на диаграмме, в совокупность *структурных уравнений*, описывающих механизмы прямого и опосредованного воздействия одних переменных на другие. На рис. 8.11, в частности, видно, что не существует пути для *прямого воздействия* интеллекта на размеры сбережений, однако общий эффект воздействия интеллекта будет включать в себя *совокупность непрямых эффектов*: X_1 воздействует на X_5 и через образование (X_2), и через достигнутый статус

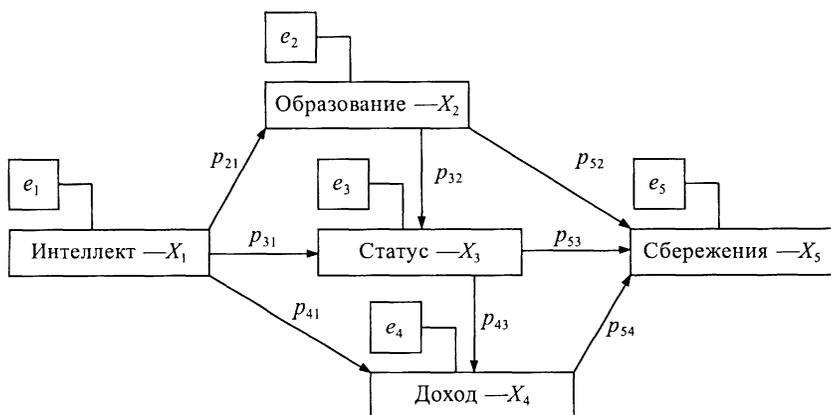


Рис. 8.11. Путевая диаграмма для примера со сбережениями

²³ В оценивании также используется метод наименьших квадратов.

(X_3), и через доход (X_4). Иными словами, хотя и нельзя утверждать, что склонность откладывать деньги «в кубышку» зависит от умственных способностей, последние влияют и на возможность получения образования, и на статус, и на доход. В свою очередь, люди с определенным социальным и экономическим статусом обнаруживают склонность иметь сбережения.

В общем случае, *полный эффект влияния переменной равен сумме ее непосредственного эффекта и всех косвенных эффектов влияния*. Величины возмущений (e_1 – e_4) на рисунке позволяют оценить, насколько хорошо работает модель, показывая, какая часть дисперсии соответствующей переменной осталась необъясненной. В результате путевой анализ позволяет пересматривать и уточнять исходную теоретическую модель, сравнивать «эффективность» нескольких конкурирующих теорий для объяснения существующей совокупности эмпирических наблюдений. Существуют даже компьютерные программы, осуществляющие автоматический поиск наилучшей структурной модели, то есть процедуру, сходную с отбором из нескольких существующих теорий такой, которая максимально соответствовала бы полученным в исследовании данным²⁴. Важно, однако, осознавать, что сами по себе результаты применения регрессионных методов и причинных моделей (регрессионные коэффициенты, линии регрессии, путевые диаграммы) решают прежде всего задачу *обобщенного описания* уже полученных эмпирических данных. Они могут служить надежной основой для интерполяции, оценки положения гипотетических «точек» в пределах ряда наблюдавшихся значений, однако их использование в целях *экстраполяции и прогноза* может вести к существенным ошибкам в тех случаях, когда такой прогноз не подкреплен более широкой теорией, не сводимой к отдельной модели для конечной совокупности данных. (Достаточно указать в качестве примера на многочисленные ошибочные прогнозы в экономике — науке, где количество эмпирических данных и описывающих их структурных моделей многократно превзошло количество существующих теорий.)

²⁴ Подробнее см.: *Девятко И. Ф.* Диагностическая процедура в социологии: очерк истории и теории. М.: Наука, 1993. С. 121–136.

Путевой анализ, как и множественная регрессия, сегодня является частью большинства стандартных статистических программ для компьютера. Не стоит, однако, забывать о том, что при любом уровне прогресса в компьютерном обеспечении задать причинную модель, то есть *совокупность содержательных гипотез, подлежащих статистическому оцениванию*, может только сам исследователь.

Дополнительная литература

Вайнберг Дж., Шумекер Дж. Статистика. М.: Финансы и статистика, 1979.

Гласс Дж., Стэнли Дж. Статистические методы в педагогике и психологии. М.: Прогресс, 1976.

Интерпретация и анализ данных в социологическом исследовании. М.: Наука, 1987.

Татарова Г. Г. Типологический анализ в социологии. М.: Наука, 1993.
Типология и классификация в социологических исследованиях. М.: Наука, 1982.

Толстова Ю. Н. Логика математического анализа социологических данных. М.: Наука, 1991.

Хейс Д. Причинный анализ в статистических исследованиях. М.: Финансы и статистика, 1981.

Флейс Дж. Статистические методы для изучения таблиц долей и пропорций. М.: Финансы и статистика, 1989.

Ядов В. А. Социологическое исследование: методология, программа, методы. 2-е изд. М.: Наука, 1987. Гл. 5.

ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ	3
ГЛАВА 1. ЛОГИКА СОЦИОЛОГИЧЕСКОГО ИССЛЕДОВАНИЯ	5
Методология и логика социологического исследования. Возможно ли объективное и научное социальное знание?	5
Методы социологического исследования: общий обзор	16
ГЛАВА 2. ВКЛЮЧЕННОЕ НАБЛЮДЕНИЕ	20
Что такое включенное наблюдение? Включенное наблюдение и этнографический метод: определение и исторические истоки	20
Планирование исследования: определение проблемы, отбор случаев, ситуаций и групп	24
Вхождение в ситуацию наблюдения, роли наблюдателя, взаимоотношения «в поле»	34
Процесс анализа и описания результатов	45
ГЛАВА 3. БИОГРАФИЧЕСКИЙ МЕТОД	63
Определение и истоки биографического метода в социологии	63
Сбор биографического материала	68
Анализ и интерпретация биографического материала	76
ГЛАВА 4. ЭКСПЕРИМЕНТ В СОЦИАЛЬНЫХ НАУКАХ	83
Определение и виды эксперимента. Основные принципы экспериментирования в социальных науках	84
Основные экспериментальные планы с контрольной группой и рандомизацией	91
Многомерные и факторные эксперименты: общий обзор	105
ГЛАВА 5. МАССОВЫЕ ОПРОСЫ В СОЦИОЛОГИИ	112
Определение и истоки	112
Выбор исследовательского плана	115
Концептуализация и измерение: общий обзор	127

Уровни измерения	138
Общие правила конструирования опросников	142
«Сензитивные» вопросы	150
Выбор формата для ответов	155
Макет анкеты (опросника)	160
ГЛАВА 6. КАЧЕСТВО ИЗМЕРЕНИЯ.	
СОЦИОЛОГИЧЕСКИЕ ИНДЕКСЫ И ШКАЛЫ	166
Проблема качества социологического измерения	166
Надежность и валидность измерения	167
Конструирование индексов и шкал	181
ГЛАВА 7. ПОСТРОЕНИЕ ВЫБОРКИ СОЦИОЛОГИЧЕСКОГО	
ИССЛЕДОВАНИЯ	199
Выборочный метод: определение и истоки	199
Типы вероятностных выборок и их реализация	202
Размер вероятностной выборки	225
Целевой отбор	227
ГЛАВА 8. АНАЛИЗ ДАННЫХ	233
Виды анализа данных	233
Одномерный анализ: табулирование и представление данных	235
Анализ связи между двумя переменными	253
Метод уточнения в анализе связи между переменными	265
Корреляция, частная корреляция, регрессия	270
Множественная регрессия и путевой анализ	284

Учебное издание

Девятко Инна Феликсовна

**МЕТОДЫ СОЦИОЛОГИЧЕСКОГО
ИССЛЕДОВАНИЯ**

Учебное пособие

Зав. редакцией *Игнатова Е. С.*

Редактор *Тюлькина М. Г.*

Корректор *Савичева О. А.*

Художник *Макух Т. А.*

Компьютерная верстка *Каган Т. М.*

Подписано в печать 13.01.09.

Формат 60×84/16. Бумага офсетная.

Гарнитура «TimesNewRoman Cyr». Печать цифровая.

Усл. печ. л. 17,21. Тираж 1000 экз. Заказ № Т-006

ООО «Издательство «КДУ». 119234, Москва, а/я 587

Тел./факс: (495) 939-40-51, 939-57-32

E-mail: kdu@kdu.ru; <http://www.kdu.ru>

Отпечатано в типографии КДУ

Тел./факс (495) 939-40-36. E-mail: press@kdu.ru