

АКАДЕМИЯ НАУК УКРАИНСКОЙ ССР
ИНСТИТУТ ФИЛОСОФИИ

В.И. ПАНИОТТО

КАЧЕСТВО СОЦИОЛОГИЧЕСКОЙ ИНФОРМАЦИИ

(МЕТОДЫ ОЦЕНКИ
И ПРОЦЕДУРЫ ОБЕСПЕЧЕНИЯ)

КИЕВ НАУКОВА ДУМКА 1986

ВВЕДЕНИЕ

Актуальность проблемы повышения качества социологической информации достаточно очевидна. В документах XXVII съезда КПСС поставлены задачи, требующие активного участия обществоведов, в том числе социологов. Это и вопросы совершенствования социальной политики, более полного учета интересов различных социальных групп, и задачи демократизации нашего общества, усиления роли общественного мнения, и задача всемерной активизации человеческого фактора, и разработка надежных социальных прогнозов и повышение качества всей работы¹. Требования к качеству социологической информации в ближайшие годы будут постоянно расти.

Это связано, прежде всего, с повышением роли социологических исследований в жизни нашей страны, в решении задач перехода от экстенсивного к интенсивному развитию экономики, совершенствования хозяйственного механизма, дальнейшего развития социализма. Внимание, уделяемое социологическим исследованиям в партийных документах, закономерно. От социологии как прикладной общественной науки сейчас требуется «решительный поворот к реальным, практическим задачам, которые ставит жизнь перед нашим обществом. Общественные науки в такой же мере, как и естественные, должны стать эффективным помощником партии и всего народа в решении этих задач»². В этом, как отметил М. С. Горбачев, состоит главный смысл, пафос решений Пленума по вопросу о значении теоретической работы в современных условиях [Горбачев, 1985, с. 11].

Другой причиной повышения роли социологических исследований является усиление субъективного фактора в процессах, происходящих в нашем обществе. Поэтому потребности партийного руководства и государственного управления в социологической информации постоянно растут. Социологи все в большей мере участвуют в изучении назревших проблем социального развития социалистического общества и в выработке действенных рекомендаций по его дальнейшему совершенствованию. И чем более сложные задачи ставятся перед социологами, чем более ответственные решения принимаются на основании их рекомендаций,

¹ Горбачев М. С. Политический доклад Центрального Комитета КПСС XXVII съезду Коммунистической партии Советского Союза. — М., 1986, с. 55, 69, 76, 108.

² Материалы Пленума Центрального комитета КПСС, 14—15 июня 1983 г. — М.: Политиздат, 1983, с. 6.

тем выше требования к качеству информации, на основе которой эти рекомендации разрабатываются.

Проблему оценки и обеспечения качества социологической информации нельзя считать малоработанной. Специально этой проблеме в нашей научной литературе посвящено несколько монографий [Волович, 1974; Докторов, 1979; Саганенко, 1979; Саганенко, 1983], разделы в книгах по методике социологических исследований, ряд статей. Отдельные аспекты этой проблемы затрагиваются также при описании тех или иных методов сбора социологической информации. В этих работах рассматриваются методологические проблемы формирования эмпирического социального знания, разрабатываются понятия, характеризующие определенные аспекты качества социологической информации, предлагаются методы обеспечения надежности отдельных этапов исследования и статистические показатели некоторых компонентов надежности информации.

Специфика данной книги заключается, во-первых, в том, что качество социологической информации в ней рассматривается не само по себе, а в контексте понятий, характеризующих качество социологического исследования в целом. Это позволяет определить качество информации в системе понятий, описывающих качество исследования. Основная цель здесь — показать, что проблемы, связанные с проведением социологического исследования, отнюдь не исчерпываются обеспечением высокого качества эмпирической социологической информации. Существует множество иных проблем, которые, как правило, не отражаются в методической литературе и не принимаются во внимание социологами при проведении исследований. Речь идет, например, о необходимости при определении эффективности социологических методов учитывать затраты с точки зрения не только исследователей, но и исследуемых и не только экономические, но и социальные, о необходимости контролировать побочный результат исследований (изменения, происходящие в объекте, субъекте и средствах исследования) и т. д.

Во-вторых, в этой книге сделана попытка систематизировать характеристики качества социологической информации и предложить набор статистических показателей для их измерения. Новизна здесь не столько в разработке новых показателей (так как большинство показателей взяты из социологической литературы), сколько в систематизации ситуаций, требующих разработки статистических показателей надежности информации. Поэтому те случаи, когда не удалось найти показатель в литературе или разработать

самому, формулируются как нерешенные задачи. Такое упорядочение показателей — одна из предпосылок для разработки в будущем стандарта качества социологической информации³.

В-третьих, особое внимание я уделял вопросам обеспечения обоснованности (валидности) социологической информации. Это наиболее важная, с моей точки зрения, характеристика качества информации, которая в научных исследованиях разработана меньше, чем другие его характеристики.

Если для определения репрезентативности, устойчивости, правильности информации разработаны специальные методы, то определение валидности производится преимущественно на основании интуиции исследователя. Поэтому часто оказывается, что все «здание» социологического исследования построено «на песке» из-за недоказанности валидности методических средств исследования. В этой книге сделана попытка разработать специальную методологию построения валидных методик измерения социальных переменных. С этой целью предложена типология валидности. Для одного из наиболее важных типов разработан показатель степени валидности и процедуры ее обеспечения. Конструктивность такого подхода иллюстрируется разработкой ряда социологических тестов (трудовой вклад работника, степень доброты респондента и коэффициент интеллекта).

Особенность книги заключается и в стремлении показать при изложении наиболее традиционных вопросов, связанных с репрезентативностью, ограниченность возможностей статистического подхода к пониманию репрезентативности, возникающие отсюда проблемы и связь репрезентативности с обоснованностью социологической информации.

Включение в монографию некоторых общеизвестных положений, а также попытки приблизить стиль изложения к литературно-публицистическому связаны со стремлением автора сделать книгу более доступной для социологов-практиков, аспирантов, студентов.

В работе над проблемами, поставленными в книге, мне оказали помощь Е. И. Головаха, Н. В. Панина, А. А. Кроник, принимавшие участие в «мозговом штурме» при разработке суждений для теста на доброту и трудовой вклад;

³ Как прецедент отмечу введение Государственного стандарта на экспертные методы оценки качества промышленной продукции, ГОСТ 23554.2-81, утвержден 30 сентября 1981 г., № 2079.

В. Н. Тарасенко, Ю. И. Яковенко и С. Н. Оксамитная — в исследовании бригадного подряда в городах Припять и Белая Церковь; С. В. Матяш — в исследовании трудового вклада инженеров проектных институтов; О. И. Белик и А. С. Мессин — в сборе информации, а Р. Э. Мункевица — в отборе материала для разработки теста на умственные способности; Б. П. Битинас, И. М. Власенко, Р. Б. Паулавичюс и Ю. И. Яковенко — в обработке информации⁴. Пользуюсь возможностью, чтобы еще раз выразить признательность им, а также В. Е. Хмелько, Б. З. Докторову, В. И. Воловичу, Е. И. Головахе, Е. А. Донченко, В. С. Максименко, В. А. Поддубному, А. Ю. Винокуру и сотрудникам отдела социологических исследований труда Института философии АН УССР — за ценные указания и рекомендации по подготовке рукописи к изданию.

⁴ Обработка данных осуществлялась на ЕС-1022, ЕС-1033 с помощью пакетов статистических программ «Социометрия» [Паннотто, Яковенко, 1980, с. 201—209] и «Паула» [Битинас, Паулавичюс, Поникарова, 1980; Битинас, Паулавичюс, 1981].

I. КАЧЕСТВО СОЦИОЛОГИЧЕСКОЙ ИНФОРМАЦИИ. ПОНЯТИЙНЫЙ АППАРАТ

Не то меня удивляет, что астрономы открыли (впрочем, зря) всякие планеты, но то, что они узнали их имена.

Ю. Тувим

В этой главе сделана попытка предложить систему понятий, характеризующих качество эмпирической социологической информации. Работа эта неблагодарна, так как чуть ли не каждый исследователь пользуется своим набором терминов, привыкает к нему и не испытывает особой потребности его менять. Что говорить о терминах науки, когда даже о самом определении науки трудно договориться. Уж на что древняя и строгая наука математика, но и в ней до сих пор идут споры о предмете, а определений самой математики существует не меньше полусотни. Вместе с тем такого рода работа необходима — без унификации терминологии затруднено взаимопонимание между исследователями.

Каким же образом можно было бы ввести систему понятий, характеризующих качество эмпирической социологической информации? Есть, по крайней мере, два пути. Один из них начинается с обзора работ, посвященных изучаемому вопросу, продолжается сравнением достоинств и недостатков той или иной системы понятий и завершается предложением собственной системы. Второй путь, которому я здесь отдаю предпочтение, состоит в попытке построить систему понятий, описывая содержание каждого из них, но не давая ему заранее «имя». Затем для выделенного понятия подыскивается по его содержанию «имя», наиболее употребительное в литературе. Например, можно сначала ввести понятие «горизонтальная поверхность, укрепленная, как правило, на четырех опорах и предназначенная для размещения пищи и посуды», а затем выбрать для него наиболее подходящее имя: «стол», «тэйбл» («калька» с английского — нередко встречающийся способ введения новых научных понятий), «кухонный стол».

Естественно начать с анализа содержания понятия, вынесенного на обложку книги, — т. е. с качества социологической информации. В философской литературе под качеством объекта понимают его существенную определенность, видовое отличие, то, что делает его именно этим, а не

другим объектом. В этом смысле качество соотносится с количеством, оно может быть выражено более или менее ярко, но само слово «качество» не несет оценочной нагрузки, нейтрально в ценностном отношении.

Вместе с тем, существует и другое, оценочно окрашенное понимание качества. В этом смысле о качестве говорят, что оно высокое или низкое, соответствует или не соответствует некоторым критериям, требованиям, стандартам, то есть качество понимается как «хорошесть» некоторого объекта. Такое понимание качества встречается и в научной литературе (именно в таком смысле слово «качество» используется в большинстве работ по квалиметрии) и в языке средств массовой коммуникации («знак качества»; «работы завершены с высоким качеством»), и в бытовой речи («берите, берите, гражданочка, не сомневайтесь — вещь качественная»). Обсуждая это различие, В. И. Волович [1974, с. 37] отмечает, что «философская и прикладная интерпретации качества хотя и отличаются, однако не противоречат друг другу».

В социологической литературе понятие качества употребляется обычно в прикладной интерпретации — как достоинство, ценность, «хорошесть» (высокое качество, в частности, информации), и данная книга не исключение. Поэтому в дальнейшем я буду рассматривать не все существенные характеристики социологической информации и социологического исследования, а только те, которые связаны с прикладным пониманием качества, то есть могут повысить или понизить качество исследования или информации¹.

Определение понятия может быть либо «внешним» (контекстуальным), построенным на описании связи данного понятия с контекстом, с понятиями, входящими в более широкую систему понятий [Кондаков, 1971, с. 226], либо «внутренним», построенным на описании понятий, входящих в объем данного понятия. Для внешнего определения понятия «качество социологической информации»² рас-

¹ Интересна эволюция понятия качества от нейтрального и даже отрицательного к понятию, имеющему положительный оттенок. Сравните у В. Даля: «Качество — ... свойство или принадлежность, все, что составляет сущность лица или вещи... Народ понимает *качество человека* в дурн. знач. *За ним, кажись, никаких качеств нет*» [Даль, 1955, т. 2, с. 99—100].

² В широком смысле под социологической информацией понимается любая информация, полученная в ходе социологического исследования (включая результаты исследования), но здесь и далее под социологической информацией имеется в виду первичная эмпирическая информация, определение которой будет дано позже.

смотрим более общее по отношению к нему понятие — «качество социологического исследования» — и место понятия качества информации среди понятий, описывающих качество исследования.

1. Социологическое исследование и социологическая информация

Социологическое исследование, как всякое научное исследование, — это процесс выработки новых научных знаний. Как системный процесс, социологическое исследование включает в себя в качестве процессов-компонентов взаимодействия субъекта исследования с объектом исследовательской деятельности и с окружающей их средой, а также взаимодействия самого объекта с этой же средой. Поэтому один из способов анализа качества социологического исследования заключается в изучении свойств субъекта, объекта, среды и тех звеньев, которые опосредуют их взаимодействие. Например, качество исследования зависит от таких характеристик субъекта, как знания об объекте исследования, умение с ним взаимодействовать, например, умение пользоваться существующими и создавать новые исследовательские средства; это и ценности, которыми он руководствуется, и многое другое. На качество исследования влияют также такие характеристики объекта исследования, как его сложность, социальная значимость неисследованных аспектов, выделяемых в качестве предмета исследования, и т. п.

Однако, когда перед социологом стоит задача провести исследование некоторого социального объекта, субъект, объект и среда в этой ситуации уже заданы, зафиксированы, и в этом смысле качество исследования от них не зависит (во всех возможных в этой ситуации исследованиях субъект, объект и среда одинаковы). Именно исходя из такой ситуации и рассматривается качество социологического исследования в данной книге. Поэтому я не останавливаюсь на важных и требующих специального изучения вопросах о качестве исследовательских кадров, а также о влиянии на качество исследования тех или иных характеристик объекта или среды исследования.

С этими ограничениями рассмотрим социологическое исследование как один из видов познавательной деятельности. Всякая деятельность, как известно, характеризуется целью, средством и результатом. (Качество социологического исследования зависит, таким образом, от качества цели, качества средств, качества процесса и качества результата

исследования, а также от отношений между этими составляющими, от их соответствия друг другу.

Теперь можно перейти к описанию характеристик каждой из составляющих качества исследования, причем оказывается, что одни и те же характеристики могут относиться к разным составляющим. Обратим внимание на соотношение цели и результата, а также средств и процесса исследования. Цель выступает как идеальное предвосхищение результата деятельности, как потенциальный результат, а собственно результат — как итог осуществления цели. Можно рассматривать цель и средства как потенциальные результат и процесс соответственно, а результат и процесс — как реальное воплощение цели и средств. Поэтому некоторые характеристики цели (например, теоретическая значимость) могут совпадать с характеристиками результата, а некоторые характеристики средств (например, экономичность) — с характеристиками процесса исследования.

Рассмотрим прежде всего понятие *качество цели* исследования. Обзор литературы в поисках «имени» для наиболее общей характеристики качества цели социологического исследования показывает, что за этим понятием не закрепился какой-либо термин.

В работе немецких социологов-марксистов указывается, что «цели исследований должны иметь большое общественное значение. Все крупные проекты социальных исследований должны иметь цели, изучение которых отвечает насущной общественной потребности... Недопустимо расходовать силы крупных исследователей и финансовые средства только потому, что тема кажется «очень интересной», или потому, что она до сих пор «мало разрабатывалась» [Процесс социального исследования, 1975, с. 23]. Таким образом, речь идет об общественной значимости, важности цели исследования и (неявно) о ее познавательной значимости («интересная тема») и уникальности (неработанность темы, которой посвящено исследование).

В книгах В. А. Ядова [1972, с. 50], А. Г. Здравомыслова [1969, с. 54], С. Михайлова [1975, с. 103], а также в ряде других работ [например: Рабочая книга социолога, 1983, с. 114; Лекции по методике ..., с. 11] в качестве основной характеристики цели исследования указывается на ее преимущественную ориентацию на анализ проблемы либо в теоретическом, либо в прикладном направлении, то есть на ее теоретическую важность, значимость или на практическую важность, первоочередность, актуальность. В философских работах, рассматривающих соотношение цели

и средств, в качестве одной из основных качественных характеристик цели рассматривается ее человечность [Макаров, 1977, с. 32—34], гуманистичность. Н. И. Ропак [1980, с. 44—50] выделяет ближайшие и конечные цели, эмпирические и теоретические, главные и неглавные, общественные и индивидуальные и, наконец, реальные и нереальные цели. Деление цели на ближайшие и конечные, общественные и индивидуальные не относится к качеству цели (в том смысле, в котором понятие качества используется в этой книге, — качество как достоинство, ценность). Все остальные характеристики (кроме степени реалистичности) описывают важность, значимость цели. Например, уникальность, неразработанность темы, которой посвящено исследование, характеризуют теоретическую, а гуманистичность — практическую значимость цели исследования. Такое перечисление, основанное на обзоре литературы, всегда оставляет сомнения — все ли характеристики учтены, не существует ли каких-либо других характеристик, не входящих в описанные классы?

Попытка перейти от «литературно-эмпирического» (назовем его так) к «теоретическому» подходу, сформулированному в начале этой главы, приводит к следующим рассуждениям, вытекающим из определения цели. Когда цель рассматривается в триаде «цель — средства — результат» или тетраде «цель—средства—процесс—результат», то цель понимается в самом широком смысле, то есть она включает тему, предмет, задачи исследования. В этом смысле цель — это то, ради чего проводится исследование. В конечном итоге — это его значимость, важность. В таком аспекте все характеристики цели, которые встречаются в литературе, непосредственно или опосредованно описывают ее важность, значимость³. Это как бы «внешняя» характеристика цели, связывающая исследование — через ожидаемый результат — с социальной системой, в рамках которой оно проводится. Это качество цели всегда в конечном счете обуславливается партийностью мировоззрения исследователя, классовым характером его жизненной позиции.

³ Например, степень согласованности цели исследования с целями других исследований (об этой характеристике я еще здесь не упоминал) специально контролируется или, по крайней мере, должна контролироваться различными координационными советами с тем, чтобы избежать ненужного повторения, распараллеливания работ, сосредоточить их на наиболее важных направлениях, то есть способствовать повышению теоретической и практической отдачи исследования.

Но существует и «внутренняя характеристика цели, соотносящая цель со средствами, процессом и результатом, как элементами исследования. Прежде всего, речь идет о вероятности достижения цели, о ее реалистичности, о том, позволит ли данное исследование достичь ожидаемого результата.

Таким образом, обнаруживаются две основные характеристики цели — *важность* (значимость) и *реалистичность* (вероятность ее достижения)⁴. Вторая характеристика в настоящее время практически не используется при оценке целей и задач социологического исследования, но это, как мне кажется, связано не с ее несущественностью, а с недостаточно высоким уровнем планирования исследований. Очевидно, что реалистичность целей важна для любой деятельности, а особенно для народнохозяйственного планирования, требующего увязки целей огромного количества различного рода работ, в том числе научных исследований и разработок.

Учитывая сказанное, я не буду присваивать «имя» качеству цели исследования, полагая, что качество достаточно полно описывается двумя основными характеристиками — значимостью (теоретической и практической) и реалистичностью цели исследования (рис. 1.1).

Следующий аспект качества социологического исследования — *качество средств* исследования. Под средством и в научной литературе [см., например: Трубников, 1967, с. 84] и в обиходной речи понимают «все, что служит к достижению цели...» [Даль, 1955, т. 2, с. 177]. Говоря о средстве, отмечают, что оно «занимает промежуточное положение между целью — моделью будущего результата и ее материальным воплощением» [Росаков, 1980, с. 51]. Интересно, что приведенное выше определение Даля расположено в его словаре среди производных от слова «середина». «Серединное» положение средство занимает не только между целью и результатом, но и между субъектом и объектом действия. Подобно тому как в процессе труда «сред-

⁴ Можно выделить и другие характеристики цели (например, степень чувствительности цели к выбору средств, одно- или многозначность, определенность цели и др.), но они представляются менее важными. Ведь при сопоставлении цели исследования с «внешней системой» («внешняя» характеристика) или с другими элементами исследования («внутренняя» характеристика) наиболее фундаментальной является общая оценка этого соотношения: степень соответствия. В первом случае это соответствие цели потребностям или, по крайней мере, интересам, запросам социальной системы (значимость цели), во втором — соответствие возможным средствам: процессу и результату (реалистичность).

ство труда есть вещь или комплекс вещей, которые человек помещает между собой и предметом труда и которые служат для него в качестве проводника воздействия на этот предмет» [Маркс, Энгельс, с. 190], средство исследования субъект исследования помещает между собой и объектом исследования, оно является посредником, проводником воздействия на объект исследования.

Средство исследования применительно к социологическому исследованию — это теория, методология, методы, методики, инструмент исследования, — все то, что требуется для успешного (соответствующего цели) осуществления исследования. К средствам исследования относятся как вещи, предметы (помещение для работы исследовательской группы, столы, шкафы, анкеты, вычислительная техника и т. п.), так и способы реализации цели, приемы, алгоритмы, программы. Из определения средств вытекает и основная характеристика их качества: поскольку средства — это то, что служит достижению цели (в этом их сущность, это их назначение — сами по себе те или иные объекты не являются средством, они выступают в качестве средств только по отношению к той или иной цели или совокупности целей), постольку основной их характеристикой является степень достижения цели, то есть *результативность*.

Существуют ли другие характеристики? Обзор социологических работ показывает, что упоминаемые характеристики программ (эксплицитность, последовательность, гибкость) [см., например: Ядов, 1972, с. 74, 75], методики (обоснованность, надежность и т. п.) и других средств исследования прямо или косвенно описывают результативность. Однако результативность — это, пользуясь ранее введенной терминологией, характеристика «внутренняя» описывающая соотношения между элементами исследования: в какой мере применение данных средств позволит получить в процессе исследования результат, соответствующий цели исследования.

Но можно рассматривать средства не только «изнутри», с точки зрения отношений между элементами исследования, но и «извне», с точки зрения соотношения между исследованием и тем социумом, в рамках и в интересах которого проводится исследование. С этой точки зрения важна не столько результативность средств исследования, сколько результат (планируемый, то есть цель, и реальный), а выбор средств, которые позволят получить этот результат — внутреннее дело исследовательской группы. Что касается социума, то для него прежде всего важны затраты на достижение планируемого результата. Таким образом, вто-

рой («внешней») характеристикой средств является объем и характер затрат, необходимых для получения результата исследования. Хотя и существуют некоторые расхождения в терминологии, характеристику соотношения результатов и затрат чаще всего называют *эффективностью*. «Обычно мы говорим, что эффективность — это отношение полученных результатов к произведенным затратам. Чем больше результаты, достигнутые за счет одинаковых затрат, или чем меньше затраты, произведенные при достижении одних и тех же результатов, то есть чем больше частное от деления результатов на затраты, тем выше эффективность»⁵. Эффективность требует специального рассмотрения, здесь сделаю лишь некоторые замечания.

Прежде всего, следует учитывать социальную, а не только экономическую эффективность, то есть отношение социального (в широком смысле) результата к социальным затратам. Причем к социальным затратам относятся не только *затраты субъекта исследования* — издержки исследователей, зависящие от доли непривлекательного труда, трудоемкости, экономичности, оперативности и других подобных характеристик, но и *затраты объекта исследования* — исследуемых. Все эти вопросы очень редко рассматриваются в социологических публикациях.

Итак, качество средств исследования определяется результативностью и эффективностью. Результативность зависит от качества различных составляющих средств исследования (теории, методологии, методики и т. п.). Эффективность при фиксированном результате зависит от затрат, произведенных исследователями и исследуемыми. Издержки исследователей характеризуются трудоемкостью, экономичностью, оперативностью и т. п. (экономические затраты), а также долей малоквалифицированного и непривлекательного труда (социальные затраты). Издержки исследуемых характеризуются трудоемкостью выполнения задания исследователя, этичностью способов использования средств исследования и их психологической травматогенностью. Этот перечень возможных затрат не претендует на полноту, а является скорее ориентиром при рассмотрении вопросов об эффективности тех или иных средств исследования.

Перейдем от рассмотрения качества средств к *качеству процесса* исследования. Процесс представляет собой реализацию средств исследования, средства соотносятся с про-

⁵ Дрекслер Л. Измерение и планирование эффективности общественного производства. — М.: Экономика, 1984, с. 19.

цессом как возможность и действительность. Поэтому, как уже указывалось, понятия, описывающие качество средств исследования, в полной мере относятся и к качеству исследования. Можно говорить о степени достижения цели, то есть о результативности процесса исследования, его эффективности и ее составляющих — экономичности, трудоемкости, оперативности, этичности, психологической травматичности и т. п. (рис. 1.1). Это избавляет от необходимости подробного рассмотрения этих характеристик применительно к процессу исследования и позволяет перейти к качеству результата.

Итак, *качество результата* социологического исследования. «Результат — объективно достигнутое состояние, продукт процесса или деятельности, направленной на реализацию цели и применяющей для этого определенные средства. Результат есть функция условий процесса или функция самой деятельности и применяющихся в ней средств для реализации цели»⁶. Результат социологического исследования — это итог, качество результата зависит от качества других составляющих: цели, средств и процесса исследования; их характеристики в той или иной степени отражены в результате. Но что конкретно является результатом исследования?

Субъект исследования, воздействуя с помощью средств на объект исследования, получает исходные данные, первичную эмпирическую социологическую информацию, обработка которой, интерпретация, теоретическое осмысление дает в качестве результата новое знание, новую научную (если исследование проводится в научных целях) или «инженерную» (в прикладных исследованиях) информацию. Но это не единственный результат исследования. Воздействие субъекта на объект приводит к изменениям в субъекте и объекте исследования, которые тоже можно рассматривать как «объективно достигнутое состояние, продукт процесса или деятельности», то есть результат.

Здесь тоже применима использованная уже аналогия с процессом трудовой деятельности, одним из видов которой является исследовательская деятельность. В трудовой деятельности можно выделить два вида результатов: изменения во внешней природе (вещественные результаты) и изменения в самом работнике (личностные результаты). Причем под вещественными результатами понимают не только функциональные (целеположенные) преобразования пред-

⁶ Философская энциклопедия. — М.: Советская энциклопедия, 1967, т. 4, с. 483.

мета труда в продукт, но и технически дисфункциональные изменения в средствах труда (изнашивание, разладка, поломки и т. д.), а также изменения вещественных элементов окружающей среды, вызываемые взаимодействием работника и средства труда с его предметом (экологический аспект содержания труда). По отношению к самому процессу преобразования предмета труда в продукт и к его целевой направленности, изменения в окружающей среде — это косвенные побочные результаты, изменения в средствах труда — прямые побочные результаты, а изменения в предмете труда — основные результаты трудовой деятельности как природопробразующего процесса [НТР и формирование духовного облика ... , 1982, с. 67—69]. К прямым побочным результатам относятся и изменения в самом работнике.

Таким же образом можно выделить *основной*, целеположенный результат социологического исследования (получение нового знания в соответствии с целью исследования) и *побочный*, обычно не предусмотренный целью исследования и состоящий в изменении объекта, средств и субъекта исследования.

Поскольку основной результат, предвосхищенный в цели и являющийся ее реальным воплощением, соотносится с ней как категории «действительное» и «возможное», его главные характеристики по существу совпадают с характеристиками цели. Если «внутренняя» (в рамках исследования), характеристика цели, связывающая ее с другими элементами исследования, определялась как вероятность достижения в процессе исследования с помощью данных средств совпадающего с целью результата и была названа реалистичностью цели, то «внутреннюю» характеристику основного результата можно определить как реальную, достигнутую в процессе исследования с помощью использованных средств степень достижения цели, которую можно так и назвать — степень *целедостижения*. «Внешней» относительно исследования характеристикой результата, так же, как и цели, является его значимость, важность для общества. Мне кажется, что по отношению к результату чаще употребляют прилагательное *ценный*, чем *важный* и *значимый*, поэтому указанную характеристику результата предлагаю назвать *ценностью*.

Преимущество такого названия, отличающегося от названия «важность» («значимость»), заключается также в том, что это позволит развести понятия важности исследования и ценности результата. Когда речь идет о цели, то ее важность (теоретическая и практическая) определяется

возможностью использовать полученные в результате исследования знания для дальнейшего развития науки или для внедрения вытекающих из полученного знания рекомендаций в практику. При этом при планировании исследования предполагается, что полученное знание будет соответствовать реальной действительности. Когда получен конкретный результат, то его предполагаемое соответствие действительности в той или иной степени становится реальностью. Поэтому наряду с важностью, значимостью исследования целесообразно рассматривать и характеристику соответствия результата реальной действительности. И соотношение этих характеристик может быть самым разным. Социологическое исследование может быть проведено по очень важной (в теоретическом или практическом плане) теме, но получен недостоверный результат, что существенно снижает (или вообще сводит к нулю) его ценность. И наоборот — может быть получен абсолютно достоверный результат, но по неактуальной, очень узкой, малозначимой теме, поэтому ценность его будет также невелика⁷.

Что касается «имени» для степени соответствия полученного знания действительности, то наиболее распространенным и общепринятым представляется *достоверность* [см., например: Волович, 1974, с. 59]. В философской литературе понятие достоверности используется в «качестве характеристики знания как обоснованного, доказательного, бесспорного и как синоним истины»⁸.

Таким образом, ценность знания (информации), полученного в результате социологического исследования, определяется *важностью*, значимостью (теоретической и практической) полученного результата и его *достоверностью*. Можно было бы говорить, что значимость результата зависит от значимости цели и не разграничивать понятия ценности и значимости, но ведь результат может не соответствовать цели исследования, причем он может оказаться не только менее важным (что, конечно, бывает чаще), но и более важным, чем предусмотрено целью исследования. Поэтому мне кажется более предпочтительным про-
извести это несколько искусственное разделение понятий,

⁷ Хотя социологов чаще ругают за мелкотемье и неактуальность, чем за недостоверность информации, мелкотемье мне кажется менее опасным. Недостоверность чаще всего сводит к нулю ценность результата, а малозначимое исследование все же даст какой-то результат и, кроме того, позволяет совершенствовать методику.

⁸ Философский энциклопедический словарь, — М.: Советская энциклопедия, 1983, с. 176.

понимая под ценностью результата характеристику его качества, включающую важность и достоверность результата. Под важностью (значимостью) результата исследования понимается вклад в развитие теоретического знания или практическая отдача, практический эффект, который имел бы данный результат при условии его абсолютной достоверности (ценность в данном контексте это как бы произведение значимости результата на его достоверность). По этим же соображениям степень целедостижения не является составной частью ценности результата, но влияет на его качество, так как от этой характеристики зависит планирование новых исследований (рис. 1.1).

Следующий шаг в экспликации качества основного результата социологического исследования — анализ его достоверности. Как уже указывалось, основной результат — это новое знание, полученное путем анализа собранной и обработанной первичной эмпирической социологической информации. В цепочке «сбор первичных данных — обработка — анализ» некоторые сложности вызывает определение обработки и вообще выделение этапов преобразования информации от исходной эмпирической в итоговую теоретическую. Особенно сложно отделить обработку от анализа социологической информации. Но для целей, которые стоят перед данной книгой, можно вообще не проводить этого разделения, понимая анализ широко и включая в него в качестве первой стадии обработку информации. Тогда, следуя Г. И. Саганенко [1983, с. 9, 10], в качестве первичной эмпирической информации будем рассматривать данные, подготовленные к вводу в ЭВМ, а все остальные преобразования информации назовем анализом (в широком смысле).

Если принять такой подход, то достоверность основного результата социологического исследования определяется качеством первичной эмпирической информации и качеством методов ее анализа, которые, в свою очередь, делятся на методы формального (математического, машинного) и содержательного (теоретического) анализа.

Рассмотрим теперь побочный результат социологического исследования, который заключается в изменении субъекта, средств и объекта исследования. *Изменения в субъекте исследования* могут заключаться в приобретении опыта, росте квалификации, расширении кругозора, удовлетворении потребности в знаниях, в высококвалифицированном труде, с одной стороны, и в усталости, неудовлетворенности, разочарованности — с другой. *Изменения в средствах исследования* заключаются в расходовании различ-

ных ресурсов (например, бумаги, анкет, амортизации ЭВМ и т. д.) и в совершенствовании методики исследования. Более сложные случаи, когда обучение исследователей или совершенствование исследовательских средств становятся специальной задачей исследования и относятся, таким образом, к прямому, а не побочному результату, требуют особого рассмотрения.

Изменения в объекте исследования могут заключаться в повышении социальной активности, осознании роли общественного мнения, расширении кругозора, удовлетворенности причастностью к значимой, имеющей общественное звучание работе, с одной стороны, и, при неэтичных и травматогенных средствах исследования, — в ухудшении межличностных отношений, росте недоверия к социологам (и косвенно к представителям научных организаций вообще), моральной травме и т. п. — с другой стороны. Аналогом экологических изменений, сопровождающих процесс труда, может быть распространяющееся по каналам межличностного общения и формирующееся таким образом общественное мнение о социологах, общественной науке или (в значительно меньшей мере) об исследуемых социологами проблемах⁹.

Итак, результат социологического исследования может быть основным или побочным. Качество побочного результата зависит от обусловленных проведенным исследованием изменений в объекте, субъекте или средствах исследования. Качество основного результата — от его ценности и степени целедостижения. Ценность результата определяется его теоретической и практической значимостью и достоверностью. Достоверность основного результата исследования зависит от качества содержательных и формальных методов обработки и анализа эмпирической информа-

⁹ Общественный резонанс от проводимых исследований может быть не так мал, как кажется. Так, в исследовании, проведенном отделом социально-профессиональных ориентаций Института философии АН УССР в г. Припяти в 1983 г. было опрошено более 2,1 тыс. человек. Учитывая, что в ближайшее социальное окружение входит от 6—10 до 20—40 человек [см., например: Сороко, 1978, с. 133], это означает, что каждый человек либо опрошен сам, либо имеет 1—2 человек из ближайшего социального окружения, подвергавшихся опросу. Поэтому пришлось планировавшиеся там эксперименты перенести в город Белую Церковь. Но оказалось, что и там часть респондентов уже «участвовали» в каких-то опросах и имеют по этому поводу такое мнение, что это привело к большому числу отказов от участия в опросе. Исследование, оставившее неприятные воспоминания у респондента, наносит, таким образом, ущерб не только респондентам, но и последующим исследователям.

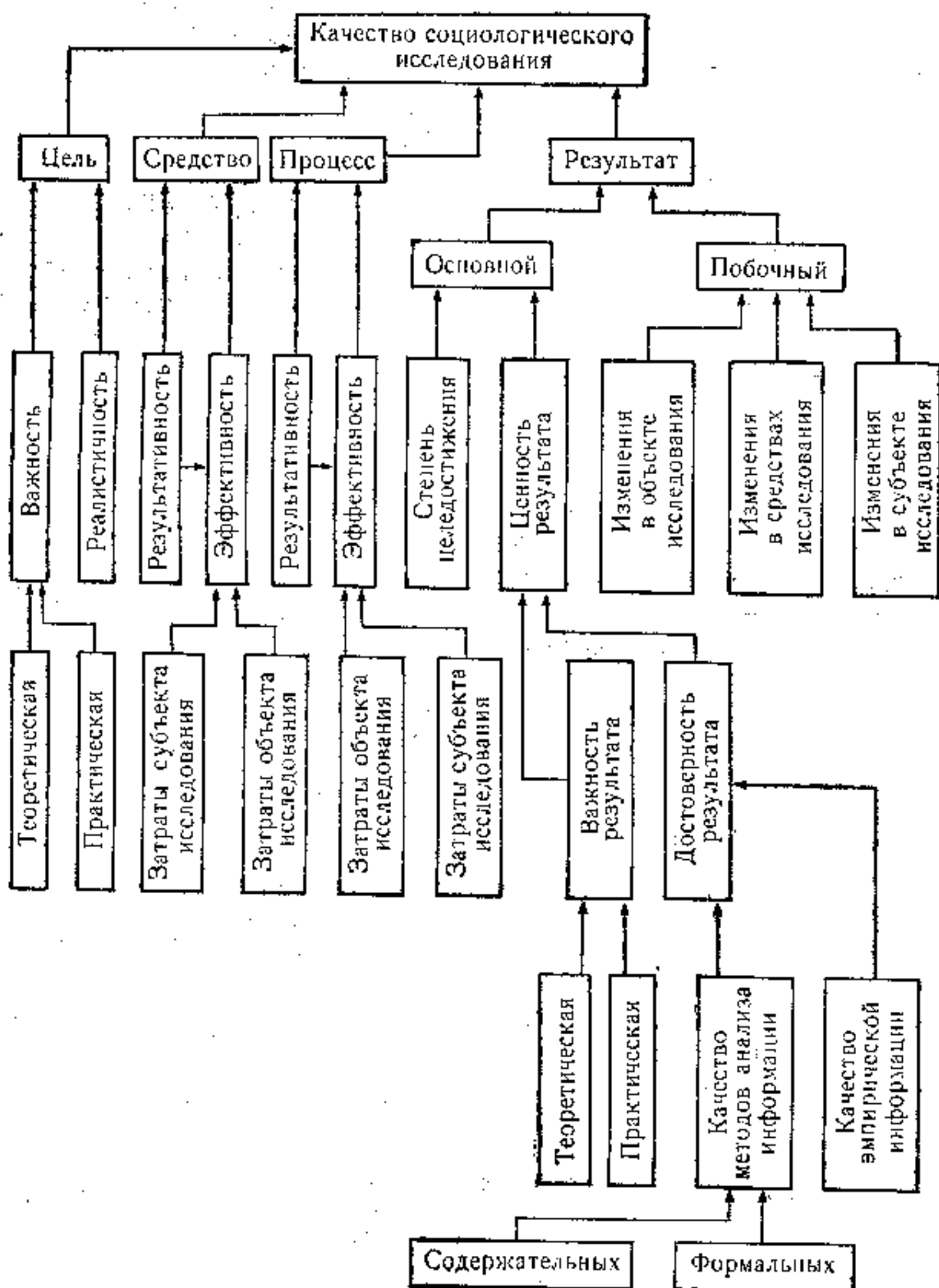


Рис. 1.1. Понятия, описывающие качество социологического исследования.

ции и от качества самой информации, которое является предметом исследования в данной монографии.

Таким образом, дано внешнее описание качества социологической информации, показывающее соотношение этого понятия с другими понятиями, описывающими качество социологического исследования (рис. 1.1), что позволяет перейти к внутреннему описанию и рассмотреть понятия, входящие в объем понятия качества первичной эмпирической социологической информации.

2. Обоснованность, репрезентативность, правильность и устойчивость социологической информации

Под *первичной эмпирической социологической информацией* будем понимать данные, полученные в ходе сбора информации (в ходе наблюдения, опроса, контент-анализа и т. п.) и подготовленные к обработке и анализу (прошедшие селекцию, уточнение, коррекцию выборки и т. п.). Этапы предварительной подготовки целесообразно отнести к сбору информации, так как они тесно сливаются непосредственно с наблюдением, опросом или регистрацией данных в контент-анализе: обнаруженные при просмотре, селекции и других операциях предварительной подготовки недостатки могут устраняться путем повторного обращения к респонденту, поиска новых документов в контент-анализе и т. п.

В литературе встречается множество понятий, описывающих качество эмпирической социологической информации: правильность, точность, устойчивость, обоснованность, валидность, надежность, воспроизводимость, достоверность, научность, истинность. Некоторые из них различные авторы предлагают как названия наиболее общей характеристики качества информации. Например, С. Михайлов [1975, с. 187] в качестве такого «имени» использует понятие научности информации, полагая, что она обеспечивается двумя составляющими — представительностью и достоверностью¹⁰: представительность зависит от того, «был ли способ подбора... единиц таким, что есть научное обоснование для обобщения полученных сводных характеристик

¹⁰ Здесь и далее в некоторых местах книги при цитировании речь идет о достоверности, валидности, надежности или других характеристиках теста или измерения, а не информации, но в большинстве случаев из контекста следует существование однозначного соответствия: информация, полученная с помощью надежного или обоснованного теста или измерения, будет надежной или обоснованной соответственно.

этих единиц и распространения этих характеристик на всю генеральную совокупность» [там же, с. 185], а достоверность — от того, «верно ли то, что зарегистрировано в отношении отдельной изучаемой единицы, точно ли оно отражает действительность... соответствуют ли ответы на вопросы вопросника реальным фактам. В конечном итоге достоверность — это адекватная регистрация» [там же, с. 186]. Достоверность же, в свою очередь, характеризуется надежностью и обоснованностью [там же, с. 189—191].

Р. Пэйто и М. Гравитц [1972, с. 479—480] используют понятия валидности (обоснованности) и достоверности; последняя «характеризует способы сбора данных, т. е. не только построение теста, но и все этапы его проведения. Прежде всего она связана с объективностью методов проведения записи и т. д.». Поскольку далее указывается, что тест может быть достоверным, не будучи валидным, но не может быть валидным, не будучи достоверным [там же, с. 479], это означает, что в качестве интегрального показателя выступает понятие «валидность».

В психологической литературе используются понятия валидности (соответствие назначения теста тому, что он в действительности измеряет) и надежности (показатель стабильности, устойчивости тестовых оценок). В качестве ведущего фактора рассматривается валидность [Психологическая диагностика ..., 1981, с. 27—33], иногда при этом отмечается, что надежность является условием валидности (но не наоборот)¹¹, то есть валидность и здесь является интегральным показателем.

Вообще валидность и надежность — наиболее употребимы в качестве интегральных характеристик социологической информации, поэтому остановимся несколько подробнее на соотношении между ними. В литературе распространены три варианта: надежность и валидность (или обоснованность) — эти понятия используются как синонимы) рассматриваются как две основные рядоположенные характеристики, определяющие качество социологической информации; валидность рассматривается как интегральная характеристика качества, надежность — как ее составляющая; надежность рассматривается как интегральная характеристика качества, а валидность — как ее составляющая.

Первый из вариантов характерен для западной социологической литературы. Надежность при этом определяется как «стабильность, объективность и отсутствие двусмы-

¹¹ Психологічний словник. — К.: Вища школа, 1982, с. 100.

сленности статистического теста или ряда измерений»¹², а валидность — как «соответствие между тем, для чего предназначен инструмент измерения, и что он действительно измеряет»¹³. Иногда надежность и обоснованность понимаются более узко — тогда они не исчерпывают содержание понятия качества социологической информации, поэтому наряду с ними вводится еще ряд характеристик: надежность, обоснованность, устойчивость, точность (как аккуратность) и точность статистическая [Ноэль, 1978, с. 293—299]¹⁴. Во всех этих случаях не вводится интегральной характеристики качества социологической информации наряду с его составляющими.

Второй вариант (валидность рассматривается в качестве интегральной характеристики качества информации) связан не столько с иным определением валидности, так как различные авторы определяют валидность более или менее одинаково, сколько с иной трактовкой ее определения. Логика рассуждения в этом случае примерно такова: если валидность показывает, действительно ли исследователь измеряет то, что он хотел измерить, а информация ненадежна, в ней присутствуют ошибки, значит, исследователь не получит то, чего он хочет, следовательно, информация не валидна. Такое понимание валидности характерно преимущественно для психологической литературы.

В отечественных работах по социологии довольно единодушно интегральную характеристику качества социологической информации называют надежностью. Такой точки зрения придерживаются В. А. Ядов [1972, с. 79, 82], В. И. Волочич [1974, с. 42], представивший подробную аргументацию в защиту своего мнения по этому поводу, Б. З. Докторов [1979, с. 27] и Г. И. Саганенко [1979, с. 30; 1983, с. 16]. Такая же точка зрения высказывается в «Рабочей книге социолога» [1983, с. 25] (что, впрочем, неудивительно, так как раздел о надежности написан Г. И. Саганенко).

Редким исключением среди работ по социологии является книга А. С. Аванесова [1982, с. 200], использующего в качестве интегрального показателя понятие валидности. Сам автор, однако, полагает, что он пользуется понятиями валидности и надежности как рядоположенными. После об-

¹² Theodorson G. A., Theodorson A. G. A modern dictionary of sociology. — London, 1970, p. 343.

¹³ Ibid., p. 455.

¹⁴ Переводчики этой книги перевели немецкое слово, соответствующее указанному Ноэль английскому варианту — validity, как достоверность, что, на мой взгляд, не отвечает сложившейся традиции переводить его как обоснованность или как валидность («калькой»).

зора работ, использующих «как основной критерий качества социологической информации надежность» [там же, с. 105], В. С. Аванесов отмечает, что другие авторы [Осипов, Андреев, 1977, с. 102] выделяют не один, а два основных критерия оценки качества информации — надежность и валидность (обоснованность). «Такой подход — более традиционный для теории социологических и психологических измерений; соответственно лучше разработан формально-логический и статистический аппарат обоснования степени соответствия измерений отмеченным критериям» [Аванесов, 1982, с. 105]. Однако если Г. В. Осипов и Э. П. Андреев, ссылаясь на Гилфорда [Guilford, 1954], действительно рассматривают надежность и обоснованность как две равноправные характеристики качества информации (по крайней мере, речь не идет о их соподчиненности), то сам В. С. Аванесов [1982, с. 135—136] далее отмечает, что понятие «валидность» относится не столько к самому тесту, сколько к цели, ситуации и к процедуре его применения: «Тест может быть надежным и совсем не валидным для той или иной цели. Обратное неверно. Если тест валидный для данной цели, то он надежен. Значение валидности теста не может быть выше значения надежности». Но сказанное означает не что иное, как понимание валидности как показателя общей пригодности теста (и, следовательно, полученной с его помощью информации) для целей исследования, а надежности — как одной из ее составляющих, то есть валидность выступает здесь интегральным показателем качества социологической информации.

Такая расширительная трактовка валидности представляется неудобной из-за того, что теряется узкое терминологическое значение валидности, которое это понятие имеет в работах большинства исследователей: информация *валидна*, если исследователь измеряет именно то, что стремился измерить. Широкая трактовка этого определения основывается на его лингвистической неопределенности, поэтому ниже будет дано более однозначное определение валидности как отсутствия в методике измерения логических ошибок. С этой точки зрения вполне возможно, что информация валидна, но не надежна. Это становится более понятным, если рассмотреть примеры, которые приводят при определении валидности. «Если кто-либо интересуется, действительно ли данное обследование предпочтений избирателей измеряет их предпочтения или действительно ли данный тест умственных способностей измеряет умственные способности, то он затрагивает вопросы валид-

ности обследования или теста»¹⁵. Вопрос здесь в том, соответствует ли в принципе измерение вербальных предпочтений реальному голосованию на выборах или соответствуют ли показатели теста умственных способностей реальным достижениям учащихся в процессе обучения. Вполне возможно представить себе ситуацию, когда в принципе вербальное поведение не рассогласовано с реальным, но из-за случайных или систематических смещений при опросах (например, респонденты боятся высказать свои предпочтения, либо на их ответы оказывает влияние пол, раса или национальность интервьюера, либо вопросник навязывает респондентам установку исследователя и т. п.) данные получаются неточными, то есть информация валидна, но неточна, неустойчива, содержит ошибки. С устранением возможных смещений она станет точной и пригодной для прогнозирования. В случае же, если она не валидна, то есть вербальное поведение в опросе в принципе не дает возможности прогнозировать реальные предпочтения на выборах, то даже самая точная, безошибочная информация не обеспечит высокого качества информации, пригодности ее для целей исследования. Вот это принципиальное соответствие измеренного тому, что должно быть измерено, соответствие при условии абсолютной безошибочности исходных данных (при отсутствии случайных и систематических ошибок) представляется чрезвычайно важной характеристикой информации. Поэтому понятие валидности кажется мне неподходящим для обозначения интегральной характеристики качества информации (в этом случае пришлось бы вводить новое понятие для узкого понимания валидности).

Из двух оставшихся вариантов — использовать как «имя» понятие «надежность» или вообще не присваивать имени качеству информации — более удобным я считаю последний вариант (использовать надежность и обоснованность как равноправные составляющие качества). Но для того, чтобы не возникло расхождений с уже устоявшимся в отечественной социологической литературе словоупотреблением (я имею в виду цитированные работы В. А. Ядова, В. И. Воловича, Б. З. Докторова, Г. И. Саганенко), в интересах ускорения процесса стандартизации социологической терминологии здесь и в дальнейшем вслед за указанными авторами я буду понимать надежность как интеграль-

¹⁵ *Theodorson G. A., Theodorson A. G. A modern dictionary of sociology, p. 455.*

ную, обобщающую характеристику качества социологической информации, как ее «имя».

Итак, высококачественная *надежная информация*. Ее определяют как «информацию, в которой относительно отсутствуют ошибки, которая не деформирует познавательного образа объекта» [Волович, 1974, с. 42]. Этот термин «отражает наше представление о качестве информации, о степени доверия к ней» [Саганенко, 1979, с. 30]. Надежность информации определяется надежностью методики, с помощью которой она получена, уровнем *квалификации* персонала, использующего эту методику, и *условиями*, в которых осуществляется сбор информации. Характеристики исследователей (исполнителей) и условий в монографии не рассматриваются, и в дальнейшем речь будет идти только о надежности методики. Под *методикой* я буду понимать технологию и инструмент сбора информации, а *надежной* будет называться такая методика, которая позволяет получить надежную эмпирическую социологическую информацию. Как показали Г. И. Саганенко [1979] и Б. З. Докторов [1979], определение надежной информации как информации, в которой отсутствуют ошибки, позволяет связать используемые в социологии характеристики информации с понятиями метрологии. Выделение тех или иных составляющих надежности информации индуцируется типологией возможных ошибок.

Первый шаг — выделение теоретических, или методологических ошибок. Теоретические погрешности, как считает Докторов Б. З. [1979, с. 11, 12], коренятся «в несовершенстве теории, положенной в основу измерительного прибора и соответствующей методики, грубости посылок, обосновывающих измерительную процедуру, неверном выборе логической модели изучаемого феномена. Погрешности подобного вида называются погрешностями метода, теоретическими погрешностями, или квазипогрешностями согласования». Поскольку измерение можно представить как создание числовой модели некоторой эмпирической системы, теоретические погрешности можно определить как несоответствие принципов построения модели, структуры модели самой моделируемой системе [Клигер, Косолапов, Толстова, 1978, с. 14—16].

Докторов убедительно показывает соответствие принятого в метрологии понятия теоретической ошибки или *квазипогрешности согласования* понятию валидности (обоснованности), используемому в социологических работах.

Таким образом, отсутствие в информации теоретических, методологических, логических ошибок, верность ис-

ходных посылок при разработке методики исследования, соответствующие числовой модели изучаемой эмпирической социальной системе целесообразно назвать валидностью или обоснованностью информации (и методики, с помощью которой эта информация получена). Именно в этом смысле будем понимать распространенное определение валидности как соответствия измеряемой информации назначению методики.

Перейдем теперь к рассмотрению других аспектов качества методики сбора социологической информации. Достаточно четко методика сбора информации разделяется на методику *отбора* единиц исследования и методику *собственно измерения* (в отличие от широкого понимания измерения, включающего и отбор), или получения единичных данных [Саганенко, 1983, с. 9], регистрации [Михайлов, 1975, с. 186] и т. п. Таким образом, качество методики сбора информации определяется качеством процедуры отбора единиц исследования и качеством процедуры и инструмента измерения характеристик отобранных единиц исследования. Найдем имена для качества отбора и измерения.

Приятно отметить, что по поводу качества процедур отбора единиц исследования существует полное единодушие — различные исследователи все проблемы, возникающие в связи с отбором респондентов (или других объектов исследования), называют проблемами *репрезентативности* или *представительности*, а информацию, в которой отсутствуют ошибки репрезентативности (и случайные, и систематические), называют репрезентативной или представительной. Поступим точно так же. Нюанс здесь заключается лишь в том, что поскольку качество методологии мы рассмотрели отдельно, то в понятие репрезентативности войдет лишь отсутствие случайных и систематических ошибок, отсутствие же методологических, теоретических ошибок процедуры отбора единиц исследования и процедуры их измерения будем называть валидностью, не вводя отдельных понятий для валидности единичного измерения и измерения массового (то есть валидности отбора).

Поясню это следующим примером. Пусть в некотором крупном городе Н. заканчивается строительство метро и существующая при горсовете социологическая группа получила задание предсказать изменение пассажиропотоков. С этой целью предполагается провести опрос и выяснить, каким транспортом будут пользоваться респонденты после введения в строй метрополитена.

Для проведения опроса по избирательным спискам с определенным шагом производится с некоторым резервом отбор квартир, а в квартирах по методике Киша [см., например: Петренко, Ярошенко, 1979, с. 101, 102] — лиц, подлежащих опросу. Резерв при этом используется для замены во время опроса лиц, не оказавшихся дома в период посещения интервьюера. Такая процедура приводит к ряду ошибок. Прежде всего, неверно определена генеральная совокупность. Планируемая выборка может обеспечить репрезентативность по отношению к населению города. Исследование же пассажиропотоков требует изучения всех лиц, пользующихся транспортом, в частности, маятниковых мигрантов, проживающих в пригороде и работающих в городе (их численность может достигать 15% от численности занятого населения), а также приезжих — командировочных, туристов, жителей сельских населенных пунктов, приезжающих за покупками, в театры, на футбольные матчи и т. п. (их численность может достигать 10—15% от общей численности лиц, пользующихся городским транспортом). Таким образом, даже при абсолютной репрезентативности выборки для жителей города, ошибка прогноза за счет неучета ряда категорий пассажиров может достигать 20—25%. Эту ошибку будем называть ошибкой валидности (обоснованности) и говорить, что выборка невалидна, необоснованна. Описанная процедура вносит также систематическую ошибку за счет использования резервного списка: интервьюер заменяет лиц, не оказавшихся дома, теми, кого он может застать, увеличивая тем самым в выборке долю малоподвижного населения и уменьшая долю подвижного населения (молодежи, студентов-вечерников и т. п.). Кроме того, в любой выборке присутствует также случайная ошибка за счет того, что опрос выборочный, а не сплошной. И систематическую и случайную ошибку выборки будем называть ошибкой репрезентативности, а полученную информацию — репрезентативной или перспре-зентативной в зависимости от того, превышает ли суммарная ошибка репрезентативности допустимый для исследователя порог.

Здесь придется сделать некоторое отступление и пояснить, что имеется в виду под случайными и систематическими ошибками, о которых только что шла речь. Тем более, что типология ошибок будет использоваться и в других главах. Под *ошибкой измерения* в статистической литературе понимают отличие результата измерения от истинного значения измеряемой величины (в тех случаях, когда измерение проводится в метрической шкале, это раз-

личие определяется как разность истинного значения и результата измерения). Различают три вида ошибок измерения. В справочнике по математической статистике они определяются следующим образом:

«1. *Промахи* — ошибки, являющиеся результатом грубых оплошностей, низкой квалификации лица, производящего измерения, небрежностей и нечеткостей в работе или неожиданных внешних воздействий на измерения. Обычно тщательные измерения и контроль за ними позволяют либо полностью избежать промахов, либо распознать измерения, являющиеся следствием таких промахов, и исключить их.

2. *Систематические ошибки* — такие ошибки, которые при данных условиях проведения измерения имеют вполне определенное значение (например, ошибки измерительного прибора). Часто систематические ошибки характеризуются постоянством знака...

3. *Случайные* — такие ошибки, которые являются результатом взаимодействия большого числа незначительных в отдельности факторов. Эти ошибки имеют в каждом отдельном случае различные значения... Случайные ошибки измерения обладают рядом свойств: при большом числе измерений ни одна из ошибок не превосходит определенного предела, крупные ошибки встречаются реже мелких и число положительных ошибок примерно равно числу отрицательных, вследствие чего сумма всех ошибок близка к нулю»¹⁶.

Аналогичным образом определяются ошибки репрезентативности. *Систематические* ошибки — это различие между характеристиками выборочной и генеральной совокупности, которые возникают из-за неправильного отбора единиц исследования, из-за нарушения принципа равновероятного отбора всех единиц генеральной совокупности. *Случайные* же ошибки существуют даже при правильном отборе в силу того, что проводится выборочный, а не сплошной опрос и судить по части о целом (в случае неоднородности целого) можно всегда лишь с некоторой вероятностью.

Вернемся к определению понятий, характеризующих качество измерения. Итак, измерение подвержено промахам, систематическим и случайным ошибкам. Отличие промахов от систематических ошибок заключается, пожалуй, в величине и частоте появления. Промахи встречаются редко и по величине значительно превосходят другие виды оши-

¹⁶ Венецкий И. Г., Венецкая В. И. Основные математико-статистические понятия и формулы в экономическом анализе: Справочник. — М.: Статистика, 1979, с. 351.

бок. Например, если при опросе в школе некоторые из учащихся в ответ на вопрос «Ваш пол» пишут «паркетный» и столь же остроумно отвечают на другие вопросы анкеты, то это промах, случайное единичное отклонение. Если же несерьезные ответы встречаются во многих анкетах, то это свидетельствует о систематических ошибках в организации процедуры опроса. Промахи обычно легко обнаруживаются и устраняются на этапе селекции собранной информации, систематические ошибки обнаружить гораздо сложнее, их обнаружение требует опыта, высокой квалификации исследователя, проведения специальных методических экспериментов. Например, при изучении школьного туризма формулировки вопросов типа «Как Вам понравилось пребывание в нашем городе?» или «Как Вам понравилось пребывание в городе-герое Киеве?» предположительно вносят систематическое смещение в сторону увеличения положительных ответов по сравнению с нейтральной формулировкой «Как Вам понравилось пребывание в Киеве?» (предпочтительна нейтральная формулировка вопроса; если же исследование уже проведено, для оценки смещения нужны специальные методические эксперименты).

Мне кажется наглядной аналогия между информацией, полученной в социологическом исследовании, и текстом, отпечатанным машинисткой. Промахи аналогичны грубым, но редким ошибкам типа пропуска страницы или строки, искажающим текст до неузнаваемости. Систематические ошибки аналогичны постоянным ошибкам в каком-нибудь слове, букве. Например, часто машинистки печатают более знакомое им слово «школа» вместо «шкала». Вспомним также придуманную И. Ильфом и Е. Петровым пишущую машинку «с турецким акцентом», который был вызван тем, что в машинке не хватало буквы «е», и текст печатался следующим образом: «Отпуститэ податэлю сэго курьэру гр. Паниковскому для Чэрноморского отдэлэния...» и т. д. Случайные ошибки аналогичны опечаткам — пропуск буквы, перестановка двух соседних букв, вставка лишней буквы. Промахи можно попытаться устранить, попросив машинистку быть внимательней, систематическую ошибку — если указать машинистке, в каких словах она делает ошибки, опечатки же можно несколько сократить, если машинистка будет очень стараться, но полное их устранение в достаточно длинном тексте практически невозможно.

Свойство процедуры и инструмента измерения (либо свойство информации, полученной с помощью данной про-

цедуры и инструмента, либо просто свойство измерения), которое заключается в отсутствии промахов, не получило какого-либо названия в литературе. Б. З. Докторов отмечает, что «промахи встречаются достаточно редко; результаты, к которым они приводят, как правило, значительно отличаются от основного массива наблюдений и могут быть изъяты из него» [Докторов, 1979, с. 10], поэтому для простоты последующих рассуждений данный вид ошибок он не рассматривает. Г. И. Саганенко [1979, с. 46—48] рассматривает грубые ошибки при определении правильности измерения, причем оба автора называют правильностью отсутствие систематических ошибок. Объединим промахи с систематическими ошибками и, следуя Б. З. Докторову и Г. И. Саганенко, будем называть отсутствие систематических ошибок и промахов *правильностью измерения* или *правильностью информации*, полученной с помощью этого измерения.

Четко разграничить правильность и обоснованность информации не всегда удастся, так как в некоторых случаях теоретические ошибки совпадают с систематическими, в других же случаях это различие выражено достаточно ярко (отделить обоснованность от правильности — это, в конечном счете, значит определить, где кончается методология и начинается методика). Например, если исследователь предлагает респондентам оценить значимость для них ряда основных сфер и условий жизнедеятельности (работа, супружество, условия для творческих занятий на досуге и т. п.) для выяснения ценностных ориентаций изучаемой совокупности респондентов, то полученная информация будет необоснованна. Респондент не в состоянии адекватно оценить значимость тех или иных сфер жизнедеятельности, его оценки являются не столько реальностью, сколько представлениями о ней, таким путем может быть получен лишь когнитивный уровень ценностных ориентаций. Как показал В. Е. Хмелько [ИТР и формирование духовного облика ..., 1982, с. 123—125, 129, 130, 300—305], для выяснения реальной значимости, изучения эмотивных аспектов ценностных ориентаций обоснованным является другой способ: определение вклада удовлетворенности отдельными элементами жизненной ситуации в удовлетворенности жизнью в целом. Различного же рода смещения, которые возникают из-за порядка предъявления респонденту разных сфер жизнедеятельности для оценки, из-за способа оценивания (парные сравнения, упорядочивание, балльная оценка) или из-за используемой шкалы (трехбалльной, семибалльной) и т. п., будут влиять на правильность инфор-

мации, но не на ее обоснованность, информация может быть правильной, без смещений, но необоснованной (возможен и другой вариант — информация может быть обоснованной и неправильной).

Б. З. Докторов [1979, с. 28—36] выделяет класс ошибок обоснованности (арифметические погрешности или квазипогрешности), которые тождественны систематическим ошибкам. Описанный ранее пример систематической ошибки, которая возникает из-за использования резервных списков при проведении опроса, тоже может быть отнесен к ошибкам обоснованности, так как эта ошибка связана с несовершенством некоторых теоретических представлений о возможности замены респондентов. Я думаю, что в пересечении понятий, характеризующих качество социологической информации, ничего страшного нет (хотя это не всегда оправдано), так как основная цель введения понятий в данном случае — не столько отличить одно от другого, сколько понять природу ошибок для того, чтобы можно было их контролировать. И главное здесь — не упустить какой-либо важной характеристики.

Образно говоря, обоснованность характеризует направление, в котором мы идем, а правильность — траекторию движения в данном направлении. И это движение нередко подчиняется следствию Вейнберга из закона Мэрфи: «Квалифицированный специалист — это человек, который удачно избегает маленьких ошибок, неуклонно двигаясь к какому-нибудь глобальному заблуждению».

Следующая характеристика качества измерения (и информации) — отсутствие случайных ошибок. Г. И. Саганенко в цитированных работах [1983, с. 21; см. также: Рабочая книга социолога, 1983, с. 251] называет эту характеристику устойчивостью, а Б. З. Докторов [1979, с. 10, 44—54] — точностью измерения (информации). Причем Г. И. Саганенко [1979, с. 21] использует как синонимы отсутствие случайных ошибок, устойчивость и воспроизводимость измерения: «Случайная ошибка устанавливается, когда имеется более одного измерения, и характеризует степень разброса данных, что относится к устойчивости, воспроизводимости измерения. Многократные, устойчивые, т. е. с малым разбросом, значения величин свидетельствуют о том, что мы получили результат, в очень узких пределах отклоняющийся от истинного значения. Здесь изучается лишь вариация повторных замеров одной и той же шкалы на одних и тех же объектах». В. А. Ядов [1972, с. 79] и В. И. Волович [1974, с. 61] тоже понимают устойчивость как сходство результатов при повторных измере-

ниях объекта, но не связывают это свойство с отсутствием случайных ошибок. Другие авторы тоже используют понятия точности, устойчивости и воспроизводимости, но в иной трактовке. Имеется, таким образом, три «претендента» на название указанной характеристики качества измерения, но выбор наиболее достойного из них сложен из-за весьма существенных расхождений между исследователями в понимании точности, устойчивости и воспроизводимости. Рассмотрим этот вопрос более подробно.

Б. З. Докторов указывает, что понятие устойчивости или точности, которым пользуются В. А. Ядов, В. И. Волович и Г. И. Саганенко, соответствует метрологическому понятию воспроизводимости как качества информации, которое отражает близость друг к другу результатов измерений, выполненных в различных условиях (в различное время, в различных местах, разными методами и средствами). С его точки зрения, это понятие является второстепенным по сравнению с точностью, которая характеризует качество информации непосредственно (как отклонение результата измерения от истинного значения признака), в то время как воспроизводимость характеризует качество информации косвенно (если близки между собой результаты повторных измерений, то можно надеяться, что они не очень существенно отклоняются и от истинного значения признака). Чтобы понять дальнейшую аргументацию, требуется ввести некоторые обозначения.

Пусть на множестве N респондентов измеряется значение некоторого количественного признака X в двух исследованиях, проведенных по одной и той же программе с использованием той же методики при сохранении условий измерения, но с некоторым интервалом времени, исключаяющим однако динамику изучаемых процессов. Получим два ряда данных: $x_1^I, x_2^I, \dots, x_N^I$ (в первом исследовании) и $x_1^{II}, x_2^{II}, \dots, x_N^{II}$ (во втором исследовании), где x_i^I и x_i^{II} — значение измеренного признака для i -го респондента в 1-м и 2-м исследованиях соответственно. Одним из показателей устойчивости информации является показатель

$$S = \sqrt{\frac{1}{2N} \sum_{i=1}^N (x_i^{II} - x_i^I)^2}. \quad (1.1)$$

Предположим, что у нас один респондент ($N=1$); на нем проведена серия k измерений некоторого признака, подверженных только случайным ошибкам: $x_1^{(1)}, x_1^{(2)}, \dots, x_1^{(k)}$. При этом истинное значение данного признака $x_1^{\text{ист}}$

отличается от среднего арифметического проведенных измерений при фиксированной вероятности на величину, зависящую от дисперсии:

$$|\bar{x}_i - x_i^{\text{ист}}| \leq t \frac{\sigma_i}{\sqrt{k-1}}, \quad (1.2)$$

где t — показатель, соответствующий фиксированной вероятности, k — число измерений, $\bar{x}_i$ и σ_i — соответственно среднее арифметическое и среднее квадратическое отклонение¹⁷ проведенных измерений. Но обычно в социологических исследованиях не проводят больше двух измерений на одном и том же объекте (потому что объект здесь — это, как правило, человек, и ему не нравится, когда его два раза об одном и том же спрашивают — какие уж тут k измерений!). Но для $k=2$ $\bar{x}_i = \frac{x_i^{(1)} + x_i^{(2)}}{2}$

а дисперсия $\sigma_i^2 = \frac{1}{4}(x_i^{(1)} - x_i^{(2)})^2$. Таким образом, формула (1.3) приобретает следующий вид:

$$|\bar{x}_i - x_i^{\text{ист}}| \leq t \cdot \frac{|x_i^{(1)} - x_i^{(2)}|}{2}. \quad (1.3)$$

Отсюда видно, что точность как близость к истинному значению признака непосредственно связана с устойчивостью как близостью оценок между собой, а формула (1.1) представляет собой среднюю квадратическую ошибку, когда у нас не один респондент, а N респондентов (Г. И. Саганенко [1979, с. 61] называет ее ошибкой по Линнику, ссылаясь на его работу¹⁸). Поэтому я не вижу оснований предпочитать ни точность, ни устойчивость как характеристику качества информации, состоящей в отсутствии в измерении случайной ошибки. Что же касается понятия *воспроизводимости*, то за ним, по-моему, целесообразно закрепить такой же смысл, как в работах по метрологии — близость друг к другу результатов измерений выполненных в различных условиях (в отличие от понимания устойчивости В. А. Ядовым, В. И. Воловичем и

¹⁷ В формуле обычно используется несмещенная оценка выборочной дисперсии, которая связана с дисперсией соотношением $S^2 = \frac{k}{k-1} \sigma^2$ поэтому в знаменателе $\sqrt{k-1}$, а не $\sqrt{k}$.

¹⁸ Линник Ю. В. Метод наименьших квадратов и основы теории обработки наблюдений. — М., 1962, с. 114.



Рис. 1.2. Понятия, описывающие качество эмпирической социологической информации (сплошная стрелка от А к Б означает, что А определяет, детерминирует Б, пунктирная — что А описывает, характеризует Б).

Г. И. Саганенко — как близости измерений, выполненных в тех же условиях).

Итак, отсутствие в информации случайных ошибок измерения мы будем называть *устойчивостью*, или *точностью* информации, используя эти понятия как синонимы¹⁹. Рассмотренная система понятий изображена на рис. 1.2.

Введенные понятия (надежность, валидность, репрезентативность, правильность, устойчивость), в определенном

¹⁹ Такой паллиатив, связанный с попыткой согласования различных определений, вызывает у меня неудовлетворенность. Быть может, следовало бы термин «точность» использовать, как в некоторых работах, в широком смысле, для обозначения близости результата измерения к истинному значению признака (отсутствие как случайных, так и систематических ошибок), то есть как интегральный показатель качества измерения. Но с другой стороны, точность ассоциируется с чувствительностью, решающей способностью инструмента, дробностью шкалы (чем дробнее шкала — тем выше точность). С этой точки зрения, «точность» следует использовать в более узком смысле.

смысле континуальны, то есть могут быть измерены с помощью некоторой, по крайней мере порядковой, шкалы. Они могут иметь высокое или низкое значение. А отсутствие ошибок того или иного типа, через которое вводились определения указанных понятий, — дискретно: ошибки или нет. Для устранения этой несогласованности следовало бы говорить о вероятности наличия или отсутствия ошибок заданной величины (например, высокая степень отсутствия ошибок репрезентативности или высокая репрезентативность — отсутствие ошибок репрезентативности, превышающих 1% при изучении доли признака).

Таким образом, надежность эмпирической социологической информации определяется ее обоснованностью (валидностью), репрезентативностью, правильностью и устойчивостью (точностью). Введем теперь количественные показатели, характеризующие надежность информации, и рассмотрим методы повышения надежности.

II. ОЦЕНКА НАДЕЖНОСТИ СОЦИОЛОГИЧЕСКОЙ ИНФОРМАЦИИ

Какое бы качество вы ни захотели оценить, всегда найдутся по меньшей мере три противоречивых критерия его оценки.

Э н о н

В предыдущей главе мы определили надежность как наиболее общее понятие, характеризующее качество социологической информации. Под надежностью понимается степень отсутствия различных ошибок (методологических и методических, в том числе ошибок отбора и ошибок измерения). Надежность характеризует пригодность информации для целей исследования. И здесь есть два аспекта, которые имеет смысл различать.

Во-первых, степень отсутствия ошибок, при которой информация пригодна для целей исследователя, можно понимать как отсутствие ошибок, превышающих пороговый уровень. Предполагается, что исследователь знает (умест измерять) величину всех ошибок в полученной информации, для каждой из ошибок он может задать пороговый уровень, и в том случае, если все ошибки лежат ниже порогового уровня, информация является надежной. В этом смысле одна и та же информация может быть ненадежной для одного исследователя и надежной для другого (если, например, другой исследователь обладает более совершенными методами анализа информации).

Во-вторых, степень отсутствия ошибок можно понимать как долю ошибок, которые измерены, долю учтенных исследователем ошибок. Наличие в информации неучтенных исследователем ошибок, то есть ошибок, величину которых он не в состоянии оценить или о которых он не подозревает, — ситуация гораздо более опасная, чем наличие больших, но измеренных, учтенных ошибок. Если, например, ошибка репрезентативности по доле некоторого признака составит 20%, то исследователь может экстраполировать данные выборки на генеральную совокупность в тех случаях, когда разность долей этого признака, полученная для каких-либо изучаемых им групп, значима (например, превышает 40%). В этом случае он сможет что-то сказать лишь о самых грубых различиях, проявляющихся даже при больших ошибках репрезентативности, но какой-то результат все же получит. Более того, если даже ошибка такова,

что исследователь вообще не получит каких-либо значимых различий, то и в этом случае ситуация менее опасна, чем при наличии неучтенных ошибок: надо продолжать работу, совершенствовать инструмент исследования, выборку и т. п. Иначе говоря, наличие больших ошибок скажется лишь на эффективности исследования. Наличие же неучтенных ошибок приводит к появлению ложных теорий, неверных рекомендаций, которые могут привести к негативным социальным последствиям. Степень учета исследователем ошибок определяет надежность как степень доверия исследователя к полученной информации. И в этом случае одна и та же информация может быть ненадежной для одного и надежной для другого исследователя (в обратной зависимости от уровня квалификации и добросовестности исследователя).

Таким образом, проблема повышения надежности социологической информации сводится к выявлению возможных ошибок, разработке методов их измерения и способов их уменьшения или устранения.

1. Методы контроля надежности социологической информации

Первый шаг в решении проблемы повышения надежности социологической информации — выявление возможных ошибок. Каким образом это осуществляется? Обычный путь — формулирование гипотез о возможных источниках смещения и проведение специальных исследований (методических экспериментов) для проверки этих гипотез. Возникающая здесь ситуация аналогична ситуации с проверкой научных теорий. Истинность теории подтверждается согласующимися с теорией эмпирическими фактами. Однако никакое множество фактов не может доказать теорию окончательно, всегда может найтись факт, который будет ей противоречить и для объяснения которого придется создавать новую теорию. По аналогии с экспериментальной проверкой теорий оказывается, что доказать эмпирически надежность полученной исследователем информации исчерпывающим образом невозможно, но чем более полная система гипотез о возможных ошибках будет выдвинута и проверена, тем больше уверенности в надежности информации.

Сложность выявления ошибок в информации связана с тем, что не существует ни метода построения гипотез, ни критериев для оценки полноты учета ошибок. Все, что пока можно предложить, — это различные классификации

ошибок, которые могут служить исследователю ориентиром при выдвижении гипотез о наличии возможных смещений в полученной информации. С этой целью и проводится классификация и выделение характеристик надежности, рассмотренных в первой главе, — обоснованности, репрезентативности, правильности и устойчивости. Это, однако, очень общие ориентиры, нужна более детальная классификация. Поэтому указанные характеристики информации конкретизируются применительно к этапам социологического исследования (пилотаж, сбор информации, кодирование, обработка и т. п.), как это сделано в книге Б. З. Докторова [1979], либо применительно к тем или иным методам сбора информации (интервью, раздаточное анкетирование, почтовый опрос, наблюдение, контент-анализ и т. п.), либо — конкретным ситуациям или проблематике исследования. Собственно, содержание учебных пособий по методике [Воронов, 1974; Здравомыслов, 1969; Лекции по методике ..., 1972; Михайлов, 1975; Процесс социального исследования, 1975; Рабочая книга социолога, 1976; Ядов, 1972; и др.] представляет собой изложение возможных ошибок и правил проведения исследований, которые позволяют избежать этих ошибок.

Такое изложение имеет обычно несистематический характер по отношению к возможным типам ошибок, то есть для каждого из рассматриваемых этапов исследования, каждого метода, для той или иной проблемы не проводится классификация возможных ошибок — рассматривается лишь перечень некоторых ошибок и даются некоторые рекомендации. Поскольку у нас практически отсутствуют стандартные методики (специально предназначенные для распространения, проверенные на наличие смещений и содержащие сведения о степени надежности — стандартные ошибки измерения, коэффициенты обоснованности и т. п.), то на основании таких, далеко не исчерпывающих рекомендаций и имеющегося опыта каждый исследователь разрабатывает свой инструмент и свои способы учета возможных ошибок для конкретных ситуаций.

В качестве примера приведу рабочую схему, которую я использую для определения возможных ошибок при оценке процедуры исследования и инструмента опроса респондентов.

Если исключить ошибки отбора респондентов и ошибки обоснованности (они будут рассмотрены в следующих главах), то для получения надежной информации необходимо, чтобы респондент, во-первых, хотел (был согласен) отвечать на вопросы анкеты и, во-вторых, мог на них отве-

тить. Рассмотрим возможные причины нежелания или неспособности респондентов отвечать на анкету.

1. Возможные причины нежелания респондента добросовестно отвечать на анкету.

1.1. Причины, связанные с ситуацией опроса. Респондент оторван опросом от своего занятия и торопится поскорее к нему вернуться (специалисты по телефонному опросу обращают внимание исследователей на тот факт, что респондент может продолжать заниматься своим делом, например смотреть по телевизору футбольный матч или детектив, отвечая так, чтобы поскорее отделаться от интервьюера); отсутствие времени, респондент торопится уходить (например, опрос школьников после уроков может привести к появлению нетерпеливых и нервных ответов или к отказу от заполнения анкет); неблагоприятная обстановка — шум, дым, движение механизмов и другие факторы, порождающие рассеянное внимание (обычно при опросе на рабочих местах); влияние окружающих (сотрудники по работе, члены семьи); события, предшествовавшие опросу и отвлекающие респондента («голова не тем занята»), и так далее.

1.2. Причины, связанные с присутствием интервьюера. Респонденту не нравится внешний вид и манеры интервьюера (в городском опросе, проведенном нашим отделом, одного из интервьюеров — человека хорошего, но на вид несколько угрюмого — респонденты дважды доставляли в милицию); респонденту нравится интервьюер, он старается лучше выглядеть в глазах интервьюера; интервьюер фактом своего присутствия, своим поведением или в силу тех или иных характеристик (пол, национальность, определенные установки и т. п.) оказывает влияние на ответы респондента.

1.3. Причины, связанные с процедурой и инструментом исследования. У респондента есть сомнения в анонимности процедуры опроса; респондента не интересует тема исследования, вопросы кажутся неинтересными или глупыми (а может быть, таковы и есть); респондент считает исследование бесполезным; слишком интимные или затрагивающие неприятные для респондента темы вопросы (например, о причинах развода); в вопросах проявляется установка исследователя, навязывающая респонденту «правильный» ответ (это вызывает смещения и в том случае, когда вопрос раздражает респондента, и он по полям анкеты пишет все, что думает по этому поводу).

2. Возможные причины неспособности респондента правильно отвечать на вопросы анкеты.

2.1. *Респондентам предлагается оценить то, что сам человек (в принципе) не может оценить.* Чаще всего это вопросы о том, что для респондента значимо, то есть «лобовая» попытка измерения ценностных ориентаций. Такая информация интересна как показатель представлений индивида (но не более чем представлений). Многочисленные эксперименты показали наличие компенсаторных механизмов (типа рационализации) и защитных механизмов, которые ставят предел осознанию, предел тому, что человек способен обсуждать сам с собой, в чем он может дать себе отчет [Шибутани, 1969, с. 71—78].

2.2. *Респондентов просят оценить то, что в принципе можно оценить самостоятельно, но для этого требуется определенный уровень рефлексии, которым данный респондент не обладает.*

2.3. *Респондентов спрашивают о том, что в принципе поддается оценке и не превышает их уровня рефлексии, но по вопросам, в которых они некомпетентны.*

2.4. *Вопросы анкеты не учитывают особенностей восприятия и памяти респондентов (у меня, например, очень плохая память, и меня легко поставить в тупик вопросом типа: «С какого года Вы регулярно читаете нашу газету?»).*

2.5. *Респондент знает и помнит то, о чем его спрашивают, но неправильно понимает вопрос анкеты.* Это может быть связано с различием «словарей» исследователя и респондента, а иногда просто с недостаточной продуманностью и неоднозначностью вопроса. Удивительный пример, показывающий, что тут можно, как говорится, «споткнуться на ровном месте», приводит Б. З. Докторов: в ответе на фактологический и очень простой вопрос «Есть ли у Вас дома телевизор?» с вариантами ответа «да» и «нет» были получены довольно существенные ошибки, связанные с неоднозначностью вопроса [Докторов, 1979, с. 97, 98]. Казалось бы, телевизор — это такая вещь, которая либо есть либо ее нет, какая здесь может быть неоднозначность? Но обнаружилось, что около 10% опрошенных рабочих проживает в общежитии и для них этот вопрос был неоднозначным. Поэтому часть из них ответили положительно, имея в виду телевизор на этаже, часть — отрицательно.

Г. И. Саганенко приводит ряд критериев для оценки трудности вопросов в анкете и предлагает начислять по каждому критерию штрафные баллы. К таким критериям относится длина вопроса, сложность его грамматической конструкции, число альтернатив (вопросы с числом аль-

тернатив больше четырех-пяти представляются сложными). Трудным считается также составной вопрос, включающий на самом деле два вопроса. Вопрос может быть сложным с точки зрения логики (например, используется двойное отрицание: «Не находите ли Вы, что Вам трудно отвечать «нет», отказывать кому-либо?»). Определенные затруднения вызывают вопросы, требующие припоминания событий, вычислений (например, среднего дохода); представления гипотетической ситуации, а также сравнения и обобщения большого количества отдельных наблюдений. Употребление в вопросе отвлеченных понятий или понятий, редко используемых данной группой респондентов и имеющих для них иной смысл, чем для исследователей, тоже усложняет понимание вопроса [Саганенко, 1983, с. 50]. К различию «словарей» исследователя и респондента можно отнести также ситуацию, в которой респондент, даже обладая требуемым уровнем рефлексии, не может выразить словами свои чувства и переживания (в социологических опросах с преимущественно закрытыми вопросами такая ситуация встречается не очень часто, но указанное обстоятельство приводит, например, к ограничениям на открытые вопросы в почтовом опросе).

Одним из подходов, позволяющих обеспечить более полный учет возможных смещений в информации, может быть разработка метрологической карты исследования [Докторов, 1984]. Предлагаемая Б. З. Докторовым карта представляет собой трехмерную таблицу. Первое измерение — этапы социологического исследования (разработка теоретической модели, создание анкеты, ее тиражирование, проектирование и реализация выборки, сбор и кодирование информации, ее обработка и интерпретация — всего 9 этапов). Второе измерение — тип ошибки (ошибки обоснованности, правильности и точности — выделяется всего три типа).

Третье измерение — метрологические задачи (всего 6 задач). Они выражаются следующими вопросами: 1) каковы причины возникновения погрешностей? 2) какова частота их возникновения? 3) какова величина погрешностей? 4) на каком этапе исследования погрешности каждого вида могут быть обнаружены впервые? 5) каким образом обнаруживаются погрешности? 6) как устранить данный вид погрешностей?

Пересечение трех измерений дает 162 клетки ($9 \times 3 \times 6$). Для оценки надежности результатов социологического исследования заполняется каждая клетка этой трехмерной таблицы. Полученная таким образом метрологическая

карта является интегральным показателем надежности результатов исследования.

Но как решить сформулированные выше метрологические задачи? Как оценить величину ошибок и какими методами их устранить? Для большинства описанных выше возможных ошибок оценка проводится исследователем на основе своего (и чужого) опыта без использования специальных процедур. Компенсация смещений проводится тоже без специального исследования ее результативности. Исследователь разрабатывает обращение к респонденту для создания у него положительного отношения к опросу, использует с этой же целью контактные вопросы, специальное оформление анкеты, старается найти нейтральные формулировки вопросов, включает «фильтры» для отбора компетентных лиц, контрольные вопросы и тесты для оценки искренности респондентов (например, тест на искренность в методике Айзенка или вымышленные названия в тестах на знания), применяет процедуры обеспечения анонимности опроса и другие методы, описанные в методической литературе. Все эти меры, однако, не избавляют исследователя от необходимости контролировать и измерять ошибки с помощью специальных исследований (пилотаж, методический эксперимент) и процедур.

Эти исследования и процедуры (назовем их методами контроля надежности социологической информации) можно разбить на две группы — методы внешнего и внутреннего контроля. Под внешним контролем будем понимать сопоставление эмпирической информации, полученной в данном исследовании, с какой-либо другой информацией, под внутренним контролем — определение ошибок только на основании информации, полученной в данном исследовании. Рассмотрим наиболее распространенные методы контроля надежности.

1. Внешний контроль.

1.1. *Контроль, внешний по отношению к моменту исследования*, или другими словами, сопоставление результатов измерения в данном исследовании с результатами повторных измерений для тех же респондентов и того же исследователя в тех же условиях с помощью того же инструмента и той же процедуры исследования. Это наиболее распространенный метод, с помощью которого изучают устойчивость (точность) информации. Его еще называют методом перепроверки или методом test-retest. Он рассматривается практически во всех работах, посвященных надежности информации [см., например: Ядов, 1972; Волович, 1974; Докторов, 1979; Саганенко, 1979, 1983; и др.]. Одна из

основных проблем при таком методе контроля — выбор интервала проведения опроса; при коротких промежутках времени влияет запоминание респондентами своих предыдущих ответов, при длинных — можно получить не характеристику надежности, а характеристику динамики исследуемых показателей.

1.2. *Контроль, внешний по отношению к ситуации (условиям) исследования.* Такой контроль имеет смысл осуществлять, если исследователь полагает, что на информацию могут влиять условия ее получения (например, опрошен респондент дома или на производстве). Для осуществления контроля могут сравниваться либо данные двух последовательных опросов на тех же респондентах (он будет внешним также и для момента исследования), или одновременного опроса на стохастически (или каким-либо другим способом) выравненных массивах. Например, формируются две случайные подвыборки рабочих завода, одна из которых опрашивается на производстве, а другая — дома по одной и той же анкете. Такое деление, возможно, несколько условно, так как изменение ситуации и условий опроса можно считать изменением процедуры, но тогда нужно расчленять отдельные элементы процедуры и вводить типологию по отношению к каждому из элементов. Данный вид контроля измеряет прежде всего правильность информации.

1.3. *Контроль, внешний по отношению к инструменту и процедуре исследования.* Можно указать несколько разновидностей такого способа контроля. Одна из них — сопоставление данных опроса с данными инструмента, который считается эквивалентным. Его называют методом эквивалентных или альтернативных форм [Ядов, 1972; Волович, 1974; и др.]. Один из вариантов — метод расщепления (split-half), который применяется преимущественно в тестовых методиках, содержащих батарею однородных вопросов и направленных на получение некоторого итогового (например, среднего или средневзвешенного) показателя. Вопросы расщепляются на две группы, как предполагается, эквивалентные друг другу, и далее проводится опрос либо одновременно на двух однородных группах респондентов либо с интервалом времени на одних и тех же респондентах. Таким способом изучается устойчивость информации.

Другая разновидность данного вида контроля — сопоставление неэквивалентных методов (например, почтового опроса и интервью). Вариант — сопоставление информации с данными реального поведения респондентов или с их реальными характеристиками (эти данные получают из

статистических источников, путем анализа документов, методом наблюдения и т. п.). Различие лишь в том, считаются ли сопоставляемые неэквивалентные методы равноценными или один предполагается более надежным, чем другой. Этим способом контроля изучается обычно обоснованность социологической информации.

1.4. Контроль, внешний по отношению к исследователю. Под исследователем в данном случае имеется в виду лицо, проводящее сбор информации, исполнитель — интервьюер, анкетер, наблюдатель, кодировщик (при проведении контент-анализа). Выделять контроль, внешний по отношению к исследователю-автору, не имеет смысла, так как его влияние осуществляется опосредованно (через методику, процедуры, выбор объекта исследования и т. п., то есть через параметры, которые отражены в данной типологии). О влиянии интервьюера и анкетера сообщается во многих работах. Очевидно, что лицо, проводящее наблюдение или осуществляющее кодирование текста, имеет свои особенности, которые могут не в меньшей степени, чем особенности интервьюера, сказываться на надежности информации. Для обозначения характера этих субъективных ошибок Б. З. Докторов [1979, с. 60—65] заимствует из астрономии понятие «личное уравнение», которое описывает особенности работы наблюдателя с измерительным прибором (например, в какую сторону направлены систематические погрешности, какие минимальные различия обнаруживает наблюдатель и т. п.).

Внешний по отношению к исследователю контроль, например, сопоставление данных опроса, полученных одной группой интервьюеров (скажем, женщинами) с данными опроса, полученными другой группой (мужчинами) на эквивалентных выборках из того же массива респондентов с использованием той же процедуры и инструмента исследования, используется обычно для оценки правильности полученной информации.

1.5. Контроль, внешний по отношению к объекту исследования. Применяется в случае, если в соответствии с целью исследования инструмент должен давать одинаковые показатели (распределения, средние и т. п.) на разных объектах. Например, при разработке теста умственных способностей, независимого от тезауруса респондента, нужно получить такой тест, чтобы распределение тестовых оценок для различных образовательных групп совпадало. В социологических исследованиях необходимость в данном виде контроля, то есть в сопоставлении информации, полученной с помощью идентичных средств на разных объ-

ектах, крайне редка (такое сопоставление проводится постоянно, но не для определения надежности инструмента, а для изучения различий объектов).

2. Внутренний контроль

2.1. *Изучение распределения данного признака по всей выборке.* В качестве примера можно привести расчет ошибок репрезентативности, основанный на вычислении дисперсии признака. Другой пример — изучение формы распределения. Это вопрос довольно обстоятельно рассмотрен Г. И. Саганенко [1979, с. 38—47]. Она, в частности, рассматривает отсутствие разброса в значениях по шкале (это может свидетельствовать о неудачном выборе градации шкалы), использование респондентами лишь части шкалы, наличие «выпадающих» пунктов шкалы, наличие грубых ошибок, то есть значений, намного превышающих среднее арифметическое и с большой вероятностью ошибочных. Такого рода контроль основывается на некоторых допущениях о «правдоподобном» распределении и свидетельствует о надежности информации в той мере, в какой верны эти допущения.

Можно выделить две ситуации. Одна из них — когда исследователь вводит в научный обиход понятие, отражающее какие-то свойства исследуемого объекта, и разрабатывает методику измерения признака, соответствующего данному свойству. Условно назовем такие признаки конструируемыми (это, например, пармия, коменция, протензия, ратимия [Мейли, 1975, с. 235], интроверсия, скорость умственных процессов и т. п.). В этом случае исследователь сам задает требуемое распределение и вопрос о верности допущений не имеет смысла. Например, при создании теста на определение скорости умственных процессов («коэффициент интеллектуальности») Г. Айзенк провел специальное преобразование числа решенных задач и отобрал задачи так, чтобы распределение введенного показателя по национальной выборке было нормальным со средним 100 и средним квадратичным отклонением 15 [Айзенк, 1972, с. 19—21]. Другого, независимого способа измерения построенного таким образом признака не существует.

Вторая ситуация — измерение некоторого ранее введенного признака, для которых существуют и другие, кроме проводимого исследователем, способы измерения. В этом случае допущения в полученном в исследовании распределении могут быть верными, а могут быть и неверными. Например, отсутствие разброса может означать, что мы выбрали неправильный масштаб шкалы (ввели интервалы до 100 руб., 101—130 руб., 131—160 руб., 161—

190 руб., 191—220 руб. и свыше 220 руб.; в то время как зарплата респондентов в данном учреждении варьирует в основном от 120 до 150 рублей), а может означать, что вариации в действительности нет (исследуемые являются преимущественно молодыми специалистами, получающими одинаковую зарплату). Или, например, оказалось, что доход какого-либо респондента по данным городского опроса в несколько раз выше, чем доход любого другого респондента. Это может объясняться грубой ошибкой, а может — тем, что в выборку попал композитор-песенник.

2.2. Сопоставление распределения различных признаков по всей выборке. Одна из наиболее распространенных задач определения надежности, попадающая в данный класс, — определение внутренней согласованности (консистентности) признаков. Эта задача решается в тех ситуациях, когда по некоторому набору признаков строится итоговый признак и исследователю требуется определить, действительно ли использованные признаки согласуются с итоговым. Один из способов решения этой задачи, основанный на том, что по консистентным признакам должны различаться крайние группы респондентов так же, как и по итоговому признаку, рассмотрен Г. И. Саганенко [1979, с. 49]. Другой способ, заключающийся в построении регрессионной модели и позволяющий отбирать не только отдельные признаки, но и наилучшее подмножество признаков, описан Б. Битинасом и Р. Паулавичюсом и реализован в пакете программ «Паула» [Битинас, Паулавичюс, 1981, с. 10—15; Битинас, Паулавичюс, Поникарова, 1980].

«Эффект личного уравнивания» может быть характерен не только для интервьюеров, наблюдателей или кодировщиков, но и для респондентов. В серии экспериментов по определению возможных ошибок при оценке школьниками престижа профессий (влияние длины списка; соседства профессий в списке; шкалы, используемой для оценки и т. п.) я получил результат, показывающий влияние оценки первой профессии списка. Все респонденты были разбиты на группы в зависимости от оценки первой профессии (всего выделено три группы: оценившие престиж первой профессии как очень высокий, просто высокий и средний; оценивших первую профессию — инженера-радиотехника — как имеющую низкий и очень низкий престиж практически не было). Для каждой из групп был рассчитан индекс престижа остальных профессий списка. Оказалось, что распределение оценок в каждой из групп приблизительно совпадало (коэффициенты корреляции: $r(x, y) = 0,92$; $r(y, z) = 0,95$; $r(x, z) = 0,89$, где x, y, z — это пер-

вая, вторая и третья группы соответственно), в то время как средние оценки сдвинуты (средняя оценка престижа для первой группы по всем профессиям равна 0,20, для второй — 0,11 и для третьей — 0,05). Построенный график (ось X — профессия, ось Y — индексы престижа каждой из групп) показал, что линии распределений всех групп идентичны, но сдвинуты одна относительно другой — чем выше оценка первой профессии, тем выше оценены все остальные.

В публикации, посвященной изложению этих результатов, я интерпретировал их как доказательство влияния оценки первой профессии [Черноволенко, Оссовский, Паниотто, 1979, с. 101—106]. Первая профессия, как мне представлялось, задаст шкалу, которой пользуется респондент для оценки других профессий: ранжировка профессий по престижу у двух респондентов может полностью совпадать, но относительность понятий «очень высокий», «высокий» и «средний» престиж может привести к тому, что они дадут разные оценки профессии и все их последующие оценки будут сдвинуты одна относительно другой. Комментируя эти результаты, Б. З. Докторов указал на то, что они могут объясняться не влиянием оценки первой профессии, а «личным уравнением» респондента, типичным для него способом оценивания: у одних оценки обычно полярны, другие избегают крайних оценок и т. д. Личное уравнение, таким образом, может определять не только оценку первой, но и вообще всех профессий, в том числе и первой.

Эта гипотеза представляется вполне правдоподобной. Возможно, впрочем, и действие обоих факторов: в основном выбор первой профессии определяется личным уравнением, оно сужает множество вариантов для выбора, но выбор внутри суженного множества вариантов тоже более или менее случаен и оказывает дополнительное влияние на оценку остальных профессий. Как бы то ни было, в тех случаях, когда с помощью одинаковых шкал проводится оценка множества объектов (табличные вопросы, батареи суждений в тестах), имеет смысл контроль смещений, задаваемых личным уравнением респондента, путем вычисления средней оценки для всего множества объектов и отклонений от нее. Этот способ контроля можно отнести к сопоставлению распределения различных признаков по выборке, с его помощью контролируется правильность информации.

Если средние для разных респондентов существенно различаются, а по смыслу изучаемых оценок они должны

совпадать (например, престиж — оценка относительная, повышение престижа одних профессий приводит к понижению престижа других, поэтому средние для всех профессий не должны существенно различаться), то может применяться коррекция оценок, нормировка средним:

$$y_j^{(i)} = \frac{x_j^{(i)}}{\bar{x}^{(i)}}. \quad (2.1)$$

Если же должны совпадать не только средние, но и дисперсии, то можно воспользоваться следующей нормировкой:

$$y_j^{(i)} = \frac{x_j^{(i)} - \bar{x}^{(i)}}{\sigma^{(i)}}, \quad (2.2)$$

где $y_j^{(i)}$ — скорректированная оценка j -го объекта i -м индивидом, $x_j^{(i)}$ — исходная оценка j -го объекта i -м индивидом, $\bar{x}^{(i)}$ — средняя арифметическая оценка для данного набора объектов, $\sigma^{(i)}$ — соответствующее данной средней среднеквадратическое отклонение.

Еще один пример метода контроля данного класса — уже упомянутый метод расщепления, если опрос проводится на одном массиве в один период времени (этот пример показывает относительность предложенной классификации — метод расщепления попадет при небольших модификациях в разные классы).

2.3. Изучение распределения признаков в различных подвыборках. Начну с задачи, которая возникла в уже упомянутых экспериментах на определение надежности оценок престижа профессий. Задача эта типична при использовании методов внешнего контроля — сопоставлении различных методов на одном и том же массиве респондентов. В упомянутых экспериментах на эквивалентных выборках из одного и того же массива (школьники выпускных классов г. Кисва) были использованы разные шкалы оценки престижа, в частности, обычная пятибалльная¹ и

¹ «Мы предлагаем Вам список профессий, прочтите его внимательно до конца. Подумайте и сообщите нам, каким авторитетом, уважением — очень высоким, высоким, средним, низким, очень низким — они пользуются у Вас. Отметьте крестиком графу, соответствующую Вашему мнению». Далее шла таблица с перечнем профессий, вариантами оценок и колонками для ответа [Черноволенко, Оссовский, Паниотто, 1979, с. 189].

«последовательная» пятибалльная² шкала. В том случае, если при использовании разных методов результаты совпадают, это свидетельствует о надежности информации. Но что делать, если результаты не совпадают, какой из методов надежнее? В нашем исследовании оказалось, что результаты, полученные с помощью обоих методов, отличаются не очень значительно, но статистически значимо. Поэтому возник вопрос о выборе более надежного способа оценки.

В данном случае для надежности данных было бы достаточно, если бы школьники отвечали на вопрос искренне и вдумчиво, проявив индивидуальные особенности в восприятии профессий. Несерьезное отношение к опросу, недостаточное знакомство с профессиями приводят к тому, что ответы становятся случайными, респондент меньше проявляет индивидуальность, на ответах меньше сказываются социальные характеристики респондентов, поэтому различия в ответах разных социальных групп уменьшаются. Таким образом (при верности этого соображения), разность оценок престижа для респондентов из различных социальных групп можно считать некоторым показателем влияния случайных факторов на ответы респондентов — чем ближе оценки, тем больше влияют на ответы случайные факторы (если бы респонденты пользовались таблицами случайных чисел для выбора ответов, то разность была бы статистически незначима). Следовательно, по разности оценок престижа можно судить, какой из инструментов надежнее, какой из них заставляет респондента более вдумчиво относиться к опросу, больше проявить индивидуальные особенности в восприятии профессии.

Были выбраны два признака, которые, как хорошо известно из многочисленных исследований, оказывают значительное влияние на престижные оценки, — пол и социальное происхождение школьников. В трех опросах для каждой профессии находилась разность средних оценок,

² «Мы предлагаем Вам список профессий. Прочитайте его внимательно до конца. Отметьте, пожалуйста, плюсом (+) те профессии, которые пользуются у Вас авторитетом, уважением, а минусом (—) те профессии, которые не пользуются у Вас авторитетом и уважением. Если Вы не знаете, как оценить — ничего не ставьте». Затем шел список профессий, а после него следующий текст: «А теперь среди профессий, пользующихся у Вас авторитетом (которые Вы отметили одним плюсом), отметьте еще одним плюсом профессии, наиболее авторитетные для Вас, а среди профессий, которые не пользуются авторитетом, отметьте еще одним минусом те, которые пользуются наименьшим авторитетом» [там же, с. 193]. Получили такую же пятибалльную шкалу: два плюса (максимальный престиж), один плюс, нет ответа, минус, два минуса (минимальный престиж).

данных школьниками и школьницами, а затем разность усреднялась по всем профессиям (всего оценивалось 40 профессий). Оказалось, что средняя разность средних оценок престижа профессий, данных школьниками и школьницами, по пятибалльной шкале составила 0,089 в опросе 1971 года (объем выборки 1364 респондента), 0,087 в опросе 1973 года (316 респондентов), а разность оценок для последовательной пятибалльной шкалы составила 0,117 (301 респондент) и статистически значимо отличается от разности для обычной пятибалльной шкалы [там же, с. 184, 187, 89, 90]. Значит, последовательная пятибалльная шкала дает более надежную информацию.

К сожалению, эти данные не были сопоставлены с результатами проверки надежности по методу перепроверки, тоже контролирующему наличие случайных ошибок, то есть устойчивость информации. Их совпадение позволило бы во многих случаях использовать более простой метод изучения оценок в подвыборках вместо метода перепроверки. Описанный метод я предлагаю назвать дисперсионным, так как он основан на изучении различий в разных подвыборках, выделенных по признакам, значимым для изучаемых в исследовании показателей.

Другой метод, который можно отнести как к данному классу, так и к предыдущему, связан с весьма любопытной техникой опроса по особо интимным и деликатным темам, что предъявляет повышенные требования к анонимности опроса. Основная идея метода, принадлежащая Уорнеру [см. об этом: Докторов, 1979, с. 72—80], заключается в стохастизации вопросов, задаваемых респонденту. Респондент отвечает на какие-то из вопросов, но на какие — интервьюеру неизвестно, а затем доля лиц, ответивших утвердительно на каждый из вопросов, приблизительно оценивается с помощью специальной обработки полученных ответов.

Идея метода хорошо видна на следующей модификации Симмонса. Респонденту вручается колода карточек, на половине из которых вопрос «Употребляете ли Вы ежедневно спиртные напитки?», а на половине — вопрос «Вы родились в апреле?». Респондент тасует карточки, затем (так, чтобы не видел интервьюер) вытаскивает одну карточку, отвечает «да» или «нет» на содержащийся в ней вопрос, вкладывает карточку обратно и опять тасует колоду. Интервьюер знает, что ответил респондент, но не знает на какой вопрос. Данные опроса всего массива опрошенных обрабатываются следующим образом. Предположим, что утвердительных ответов всего 20%, а нас инте-

ресует доля лиц, положительно ответивших на первый вопрос. Если бы все ответили только на вопрос о дне рождения, то положительных ответов было бы около 1/12 (если считать равномерной рождаемость по месяцам — можно, впрочем, уточнить эту долю по статистическим справочникам), то есть около 8%. Так как на этот вопрос отвечали около половины опрошенных, то положительных ответов на него дали около 4% от всего массива опрошенных, значит, остальные 16% положительных ответов относятся к вопросу об алкоголе. Но на этот вопрос отвечала только половина опрошенных, значит, доля ответивших утвердительно для всего массива составляет приблизительно 32% (доверительные границы этой оценки устанавливаются с помощью соответствующих статистических методов).

Модификации методов отличаются способом рандомизации (в методе Уорнера, например, респондент крутит волчок, поверхность которого разбита на два сектора, установка волчка у соответствующего сектора определяет выбор вопроса для ответа), числом оцениваемых суждений, характером суждений (существует модификация для количественного вопроса — например, о доходе) и т. д.

К сожалению, исследования показывают, что даже этот метод не позволяет добиться абсолютной искренности и не всегда полученные с его помощью результаты существенно отличаются от данных, полученных другими методами. Например, в одном из исследований, авторы которого имели возможность сравнивать данные опроса с данными, полученными из достоверных документов, оказалось, что по вопросу о банкротстве указанный метод (random response) дает самую надежную информацию, лишённую ошибки полностью, по сравнению с 32% ошибок при опросе путем интервью (face to face), 29% при телефонном интервью, 32% — при использовании раздаточной анкеты (self-administred questionnaire). По вопросу о вождении машины в пьяном виде процент ошибок равен 35% по сравнению с 47%, 46% и 54% в интервью, телефонном интервью и раздаточном опросе соответственно. На вопрос же о поведении при голосовании с помощью этого метода были получены менее надежные ответы (48% ошибок по сравнению с 39%, 31% и 36% при интервью, телефонном интервью и раздаточном опросе) [Bradburn, Sudman, 1979].

Итак, мы рассмотрели классы методов и некоторые конкретные методы контроля надежности социологической информации. Перейдем теперь к рассмотрению показателей, характеризующих надежность в каждом из указанных методов.

2. Статистические показатели надежности социологической информации

Изложение вопроса о показателях надежности социологической информации обычно начинают с основного уравнения, описывающего исходное допущение в теории измерения [см., например, Аванесов, 1982, с. 106—111; Процесс социального исследования, 1975, с. 124—125; Dick, Hagerty, 1971, р. 13]:

$$x = x_t + x_e, \quad (2.3)$$

где x — результат измерения, x_t — истинное значение измеряемой величины (индекс t — от английского true — правильный), x_e — ошибка измерения (error).

Поскольку систематическая ошибка по определению постоянна, ее дисперсия равна нулю, поэтому дисперсия ошибки определяется лишь дисперсией ее случайной компоненты. Общая дисперсия (которая получается в результате измерения) равна:

$$\sigma^2 = \sigma_t^2 + \sigma_e^2, \quad (2.4)$$

где σ^2 — общая дисперсия измеренной величины, σ_t^2 — дисперсия истинного значения признака, σ_e^2 — дисперсия ошибки.

В качестве показателя надежности принимается отношение дисперсии истинного значения признака к дисперсии измеренного значения. Обозначив коэффициент надежности через r_{tt} , получим:

$$r_{tt} = \frac{\sigma_t^2}{\sigma^2}. \quad (2.5)$$

Чем больше доля вариации истинного значения признака в вариации измеренного значения, тем выше надежность измерения. Поскольку дисперсия истинного значения признака нам неизвестна, для реальных расчетов используют другую формулу. Разделив обе части равенства (2.4) на σ^2 , получим:

$$1 = \frac{\sigma_t^2}{\sigma^2} + \frac{\sigma_e^2}{\sigma^2}, \quad (2.6)$$

откуда

$$r_{tt} = \frac{\sigma_t^2}{\sigma^2} = 1 - \frac{\sigma_e^2}{\sigma^2}. \quad (2.7)$$

Как указывает В. С. Аванесов [1982, с. 111], эта формула называется формулой Ф. Рулона, хотя предложил ее Дж. Фланаган. В этой формуле нам известна дисперсия σ^2 , а дисперсию ошибки можно определить путем специального исследования (например, методом перепроверки, параллельных форм или расщепления). В частности, как указывалось в предыдущей главе при рассмотрении формулы (1.3), дисперсия случайной ошибки для i -го респондента при двух последовательных опросах (I и II) равна

$\sigma_i^2 = \frac{1}{4} (x_i^I - x_i^{II})^2$, тогда средняя дисперсия по всем респондентам, которую можно использовать в формуле (2.7), будет равна:

$$\sigma_e^2 = \frac{1}{4N} \sum_{i=1}^N (x_i^I - x_i^{II})^2, \quad (2.8)$$

где x_i^I — значение признака X для i -го респондента в опросе I, x_i^{II} — значение признака X для i -го респондента в опросе II, N — число опрошенных.

Так как r_{ii} характеризует лишь отсутствие случайных ошибок, его можно назвать *коэффициентом устойчивости*. Что касается обоснованности, то она определяется путем сопоставления данных измерения с данными, полученными с помощью заведомо обоснованного метода (например, сопоставлением с реальным поведением), и может измеряться коэффициентом корреляции между полученными на тех же респондентах значениями измеряемого признака X и полученными с помощью обоснованного метода (критерия валидизации) значениями признака Y . Коэффициент корреляции r_{xy} можно было бы назвать коэффициентом обоснованности.

Если здесь остановиться и дальше не продолжать, то получится довольно стройная и красивая картина. Есть два основных показателя — r_{ii} , характеризующий устойчивость, и r_{xy} , характеризующий обоснованность, оба изменяются по абсолютной величине от 0 до 1 (то, что r_{xy} может иметь отрицательные значения, не важно, так как исследователя интересует лишь степень связи, а не ее направленность, $r_{xy} = -1$ позволяет получить абсолютно обоснованную методику). Существуют методы определения указанных показателей (о них шла речь в предыдущем параграфе). Что еще нужно?

К сожалению, способ измерения надежности в соответствии с формулой (2.7) имеет множество ограничений.

Во-первых, он пригоден безусловно для метрических шкал, с определенной натяжкой в некоторых случаях может быть использован для шкал порядковых и совсем не годится для шкал номинальных — для них некорректно даже исходное уравнение (2.3). Во-вторых, этот подход предполагает, что респондент в ответ на социологический вопрос указывает значение некоторого признака, и не пригоден для случая, когда ответом респондента является упорядочивание (например, методом парных сравнений), отбор и упорядочивание и другие более сложные случаи. В-третьих, формула (2.7) не предназначена для оценки надежности показателей, построенных на множестве признаков (например, тестовая оценка, полученная путем суммирования оценок по отдельным признакам). Изложенный подход, впрочем, допускает обобщение на этот случай, если использованные шкалы — метрические. Наконец, в-четвертых, этот метод не позволяет оценить надежность итоговых показателей (коэффициентов связи, результатов классификации и т. п.).

Таким образом, показатели надежности зависят от шкалы измерения (номинальная, порядковая или метрическая), от измеряемого показателя (значение признака, упорядочение, функция от значений ряда признаков и т. п.) и от способа контроля ошибок, то есть от способа получения информации об ошибках. Все это очень усложняет картину и не позволяет свести все показатели надежности в стройную систему.

Попытаюсь несколько упорядочить ситуации, требующие разработки статистических показателей надежности информации. Тогда в тех случаях, когда показатели уже разработаны, можно будет сослаться на соответствующие работы, а в тех случаях, когда показателей не существует, — сформулировать их разработку как нерешенную проблему, ждущую своих исследователей.

Для определенности рассмотрим сначала все эти вопросы применительно к ситуации измерения надежности социологического опроса. Пусть опрошено N респондентов, каждый из которых ответил на анкету, содержащую k вопросов. Рассмотрим следующие ситуации.

1. Определение надежности показателя, рассчитываемого для одного (для каждого) респондента по данным ответа на один вопрос.

1.1. Ответом на вопрос является значение некоторого признака X .

1.1.1. Признак X измерен в номинальной шкале. Здесь может быть два случая — вопрос однознач-

ный, когда респондент в ответ на вопрос может указать лишь одно значение признака X (например, пол), или многозначный³ (например, свободное время). Если вопрос однозначный, то при использовании повторного опроса тех же респондентов для оценки устойчивости информации может использоваться показатель W , предложенный Г. И. Саганенко [1979, с. 56—58]:

$$W = \frac{1}{N} \sum_{i=1}^l n_{ii}, \quad (2.9)$$

где W — показатель полной устойчивости, n_{ii} — количество респондентов, указавших и в первом и во втором опросе ответ i , l — число градаций (вариантов ответа) признака X . Обозначение n_{ii} связано с тем, что проведение расчета показателя W проводится по матрице:

		Опрос II			
		1	2	...	l
Опрос I	1	n_{11}	n_{12}	...	n_{1l}
	2	n_{21}	n_{22}	...	n_{2l}
	$\vdots$				
	l	n_{l1}	n_{l2}	...	n_{ll}

(2.10)

где n_{ij} — число респондентов, указавших в первом опросе i -е, а во втором опросе j -е значение признака X .

Если признак X — многозначный, то сумма ответов в I и во II исследованиях уже не обязательно будет совпадать, матрица (2.10) будет формироваться так же, как обычная таблица двумерного распределения для многозначных вопросов, а в формуле (2.9) в знаменателе будет уже не число респондентов, а число ответов (сумма маргиналов матрицы (2.10)). Хотя такая ситуация встречается довольно часто, этот случай почему-то никем не рассматривается.

Можно предложить также показатель, характеризующий устойчивость ответов на каждую градацию признака X (такая информация тоже представляет интерес для доработки данного вопроса анкеты):

$$w_i = \frac{\sqrt{2}n_{ii}}{n_i^I + n_i^{II}}, \quad (2.11)$$

³ Иногда для обозначения этого различия говорят также об одно- и многоальтернативных вопросах и о вопросах с несовместимыми и совместимыми альтернативами (соответственно).

где w_i — устойчивость i -й градации признака X , n_i^I и n_i^{II} — число респондентов, отметивших i -ю градацию признака X в I и во II опросе соответственно, n_{ii} — число респондентов, указавших i -ю градацию признака X в I и во II опросе.

$$n_i^I = \sum_{j=1}^l n_{ij}, \quad n_i^{II} = \sum_{j=1}^l n_{ji}. \quad (2.12)$$

При использовании методов контроля, внешних по отношению к условиям исследования и по отношению к исследователю-исполнителю, могут использоваться те же коэффициенты W и w_i .

При использовании метода контроля, внешнего по отношению к инструменту и процедуре исследования, признак X во втором опросе (или в другом методе, использованном для измерения признака X) может быть получен в другой шкале, имеющей иное число градаций (и даже не обязательно номинальной). В этом случае, уже не для определения устойчивости, а для определения правильности или обоснованности в зависимости от использованной процедуры, может использоваться один из коэффициентов корреляции для таблиц $k \times l$, например коэффициент Крамера [Паниотто, Максименко, 1982, с. 83, форм. (II, 2, 5); Статистические методы..., 1979, с. 120, форм. (8.16); и др.].

К методам внутреннего контроля надежности (правильности) выбора i -го значения признака каждым респондентом относятся выявление отсутствия разброса в распределении респондентов по градациям шкалы (это означает, что шкала неудачна и навязывает респонденту лишь одно значение, не учитывая реального распределения), а также выявление «неработающей» части шкалы или ее отдельных пунктов с использованием предложенных Г. И. Саганенко [1979, с. 38—46] критериев.

1.1.2. Признак X измерен в порядковой шкале (контроль, внешний по отношению к моменту исследования). Поскольку порядковая шкала является в то же время и номинальной (а метрическая — и порядковой и номинальной, так как непрерывный признак можно разбить на дискретные градации), то все показатели, используемые для номинальных шкал, полностью применимы и для порядковых, в частности, и показатели W и w_i , однако при этом теряется часть информации. Коэффициенты W и w_i фиксируют только совпадения, все несовпадения для них одинаковы, в то время как для ранговых шкал уже имеет значение величина несовпадений: если данные повторного опроса (при равенстве числа совпадающих зна-

чений) отличаются не более, чем на один пункт шкалы, то измерение более устойчиво, чем если данные отличаются на два, три и более пунктов. В этом случае может быть использован коэффициент ранговой корреляции τ .

Г. И. Саганенко [1979, с. 58—67] считает, что коэффициенты корреляции недостаточно чувствительны и предлагает для оценки ошибки использовать среднюю арифметическую и среднюю квадратичную ошибку

$$\bar{\Delta} = \sum_{i=1}^l \sum_{j=1}^l n_{ij} |x_i^I - x_j^{II}|, \quad (2.13)$$

где $\bar{\Delta}$ — средняя арифметическая ошибка (точнее, средняя арифметическая модулей ошибок), n_{ij} — число респондентов, указавших i -е значение признака X в I и j -е значение признака X во II опросе, x_i и x_j — ранги i -го и j -го значения признака X соответственно (если значения признака совпадают с рангами, то $x_i - x_j = i - j$).

$$S = \sqrt{\frac{1}{2N} \sum_{i=1}^l \sum_{j=1}^l n_{ij} (x_i^I - x_j^{II})^2}, \quad (2.14)$$

где S — средняя квадратическая ошибка (или ошибка по Лишнику), остальные обозначения те же. Для сравнения шкал, имеющих разное число вариантов ответа, в той же работе предлагается использовать относительный показатель:

$$\Delta_{\text{отн}} = \frac{\Delta}{\Delta_{\text{max}}}, \quad (2.15)$$

где Δ_{max} — максимально возможная ошибка. Если респонденты используют весь диапазон шкалы, то максимально возможная ошибка равна $l-1$ (l , как и раньше, число вариантов ответа, число возможных значений признака X). В качестве Δ_{max} Г. И. Саганенко предлагает использовать число градаций в работающей части шкалы.

Аналогично w_i можно было бы также ввести Δ_i и относительную ошибку $\Delta_{i\text{отн}}$. Например, ошибка устойчивости для i -го значения признака равна:

$$\Delta_i = \frac{1}{2} \sum_{j=1}^l (n_{ij} |x_i^I - x_j^{II}| + n_{ji} |x_j^I - x_i^{II}|). \quad (2.16)$$

Множитель $1/2$ введен для того, чтобы выполнялось условие:

$$\Delta = \sum_{i=1}^l \Delta_i. \quad (2.17)$$

Недостатком подхода, выраженного в формулах (2.13) — (2.17), является использование операций, некорректных для ранговых шкал, в частности вычитания рангов, то есть при строгом подходе все эти формулы некорректны. Вопрос этот дискуссионный. Существуют «пуристы», ратующие за чистоту и корректность использования математических методов, и «практики», доказывающие, что, во-первых, если так рассуждать, то вообще ничего нельзя будет делать, и во-вторых, если применять правильные методы, все равно получается то же самое. Пока что ситуация такова, что ничем не рискуя, можно объявить: «Кто использует лишь «чистые» методы, пусть первый бросит в меня камень». Однако, я думаю, что применять некорректные методы, основанные на недоказанных предположениях об уровне измерения, можно лишь в крайнем случае (когда не существует корректных методов, нет программного обеспечения и т. п.). Но каждый такой случай необходимо специально оговаривать, объяснять, с чем это связано, и пытаться оценить возможные смещения в результатах исследования.

Второй недостаток — несопоставимость границ изменения коэффициентов (2.13) — (2.17), что усложняет сравнимость показателей надежности. Если показатель обоснованности как коэффициент корреляции результатов, полученных с помощью данной и с помощью другой, обоснованной методики, изменяется по модулю от 0 до 1, то показатель Δ — от 0 до $l-1$, а показатель $\Delta_{\text{отн}}$ — от 0 до 1, но в отличие от коэффициента корреляции здесь 0 соответствует максимальной надежности, а 1 — минимальной. От этого недостатка можно избавиться, введя соответствующие преобразования. Например:

$$k_s = 1 - \frac{\sqrt{2}}{l-1} S, \quad (2.18)$$

$$k_{\Delta} = 1 - \Delta_{\text{отн}}, \quad (2.19)$$

где k_s и k_{Δ} получены из показателей S и Δ , и изменяются от 0 до 1, причем 0 соответствует минимальной, а 1 — максимальной надежности.

Г. И. Саганенко [1979, с. 54, рис. 4], доказывая низкую чувствительность коэффициентов корреляции, приводит в пример две матрицы типа (2.10): в одной устойчивость

данных равна одной градации шкалы, в другой — двум. Коэффициент даст 0,97 для первой и 0,91 для второй матрицы. Для устранения этого недостатка предлагается ввести показатели S и Δ . Действительно, используя, например, Δ , мы получим для первой матрицы значение выше, чем для второй, более чем в два раза. Однако после преобразования (2.19) получим 0,95 для первой и 0,88 для второй матрицы, то есть практически те же различия, что и у коэффициентов корреляции.

В остальном все, что было сказано в пункте 1.1.1, применимо и здесь. Кроме того, при контроле, внешнем по отношению к инструменту и процедуре исследования, для определения правильности и обоснованности может использоваться коэффициент ранговой корреляции.

1.1.3. Признак X измерен в метрической шкале (контроль, внешний по отношению к моменту исследования). Здесь как раз и может быть осуществлен этот подход, отраженный в формулах (2.3) — (2.8). Другой подход — введение коэффициента корреляции r — вероятно, менее эффективен [Саганенко, 1979, с. 53, 54]. Кроме того, могут быть использованы все те показатели, которые применялись для ранговых шкал (с той лишь разницей, что если для ранговых шкал формулы (2.13) — (2.17) могут быть использованы лишь как ориентировочные, то для метрических шкал — с полным на то правом). Те же показатели пригодны и для контроля, внешнего по отношению к условиям опроса и к исполнителю. При контроле, внешнем по отношению к инструменту и процедуре исследования, используется коэффициент корреляции r .

1.2. Ответом на вопрос является упорядочение некоторых объектов. Обычно исходной для расчета различных статистических показателей является матрица данных: в строках — объекты, в столбцах — признаки на пересечении — значение j -го признака на i -м объекте. В социологических исследованиях ситуация усложняется тем, что ответом на социологический вопрос не обязательно может быть значение признака, а и некоторое отношение между объектами. Если даже не принимать во внимание социометрических вопросов, то в обычных вопросах респондент не только указывает значения тех или иных признаков, но и иногда задаст также отношения на этих значениях (в основном — больше, меньше, равно).

Б. З. Докторов и Б. М. Фирсов выделяют следующие типы вопросов: вопрос с единственным выбором из двух градаций; вопрос с единственным выбором из числа градаций больше двух; блок-вопрос (или табличный, или матричный).

вопрос), представляющий собой перечень объектов (суждения, профессии и т. п.), каждый из которых оценивается с помощью той же шкалы, обычно компактно печатается в анкете в виде таблицы; вопрос с множественным свободным выбором (обычный номинальный вопрос, который мы называли многозначным, или же с совместимыми альтернативами); вопрос с множественным ограниченным выбором (если в инструкции указывается, какое число выборов может сделать респондент); вопрос с множественным ограниченным выбором с упорядочением градаций (например: «Выберите, пожалуйста, из предложенного ниже перечня пять наиболее привлекательных профессий и упорядочьте их по степени привлекательности»); вопрос с упорядочением всех предложенных градаций [Докторов, Фирсов, 1975, с. 67].

Некоторые из этих типов вопросов мы уже рассмотрели (номинальные вопросы с единственным и множественным выбором). Рассмотрение всех типов усложнит изложение, поэтому ограничимся тремя случаями: ответ на вопрос — значение признака, измеренное по той или иной шкале; ответ на вопрос — упорядочение всех градаций; остальные типы (куда входит все то, что не было рассмотрено раньше, то есть рассматриваются крайние случаи, а все другие отнесены в категорию «остальные»).

Итак, ответом на вопрос является упорядочивание множества объектов или градаций признака. Существует несколько основных способов получения информации о предпочтениях респондента. Один из способов — балльные оценки, которые выставляет респондент каждому объекту (этот способ можно отнести к другому типу вопроса и здесь не рассматривать). Другой способ — ранжирование: респондент непосредственно упорядочивает объекты из предложенного исследователем набора (такое упорядочивание удобно производить, если название объектов написаны на карточках). Третий способ — парные сравнения: респонденту предъявляются всевозможные сочтения по два объекта из всего множества объектов, подлежащих упорядочению, и предлагается в каждой паре указать более предпочтительный (в соответствии с заданным критерием) объект. Наконец, четвертый способ — множественные сравнения, которые можно рассматривать как обобщение методов парного сравнения и ранжирования и на которых я останавлиюсь подробнее несколько позже.

Указанные методы изучения предпочтений наиболее широко используются в экспертных опросах. Экспертные опросы в настоящее время настолько бурно и интенсивно раз-

виваются, что давно перестали быть только разновидностью социологического опроса и превратились в некоторую самостоятельную область науки, которая, как это ни странно, ближе всего к математике и кибернетике. Подавляющее большинство работ посвящены методам анализа данных экспертного опроса, значительно меньше работ посвящены организации экспертизы (причем, рассматриваются преимущественно такие аспекты организации, как отбор экспертов, определение их компетентности и процедуры согласования мнений⁴), и совсем уже мало работ посвящены надежности исходной информации. Для доказательства этого тезиса я задумал было подсчитать долю докладов и сообщений по надежности исходной информации на очень представительном совещании по экспертным оценкам [I Всесоюзное совещание по статистическому и дискретному анализу..., 1981], состоявшемся в Алма-Ате в 1981 г. Но ни в докладах, ни в кулуарном общении, ни в опубликованных тезисах (а в сборнике опубликованы тезисы 218 докладов, правда, часть из них — по дискретной оптимизации) не удалось обнаружить попыток оценить надежность исходной информации (за исключением, может быть, одной работы⁵).

Отчасти такая диспропорция компенсируется статистическим подходом к информации и изоциренной обработкой. В одной из последних работ по экспертным оценкам дается такое описание: статистический подход «основывается на предположении, что существует истинное, единственно верное, но не известное заранее суждение. А суждение каждого эксперта есть его искаженный вариант. Искажения, отклонения от истины происходят в силу случайных факторов. Появление суждений экспертов есть реализация случайного «выбора», при котором каждое суждение имеет определенную вероятность оказаться выбранным. Суждения разных экспертов понимаются как статистически независимые испытания в эксперименте по «выбору» суждений. Чем лучше эксперт, тем выше вероятность того, что его суждение близко к истинному. Случайный человек, не эксперт, с равной вероятностью может назвать любое суждение. Поэтому в экспертизе вероятность появления какого-то суждения тем ниже,

⁴ Обстоятельный обзор методов экспертных оценок см.: [Шмерлинг, Дубровский, Аржанова, Френкель, 1977, с. 290—382] (отмечу, что список литературы к обзору состоит из 313 наименований).

⁵ Выступление Казанской Т. А. «Исследование непротиворечивости данных экспертизы, проведенной методом парных сравнений» [там же, с. 122—132].

чем дальше оно от истинного суждения, которое имеет наибольшую вероятность. Поскольку все экспертные оценки отражают, хотя и в искаженном виде, одну истину, то между ними должно быть определенное сходство. Наличие его проверяется в ходе анализа суждений. Затем посредством дальнейшего анализа выявляется то истинное суждение, которое скрывается за реальными суждениями опрошенных экспертов. Такой взгляд на анализ суждений восходит к статистической обработке результатов наблюдений, и эксперт здесь трактуется фактически как измерительный прибор, работающий со случайными погрешностями» [Панкова, Петровский, Шнейдерман, 1984, с. 54].

Вместе с тем очевидно, что статистическая обработка позволяет получить достоверный результат при условии определенного уровня надежности обрабатываемой информации, и существуют ошибки, устранить которые невозможно никакими статистическими методами (они только создадут иллюзию надежности и породят артефакты). Поэтому следует развивать методы контроля надежности и вводить показатели, оценивающие надежность исходной информации. Для этих целей пригодны и многие из методов, рассмотренных в предыдущем параграфе.

Так, применимы фактически все методы внешнего контроля. В качестве показателя устойчивости при контроле, внешнем по отношению к моменту опроса, и в качестве показателя обоснованности при контроле, внешнем по отношению к инструменту и процедуре исследования, а также для других методов внешнего контроля может использоваться среднее арифметическое коэффициентов корреляции, полученных от каждого респондента:

$$T = \frac{1}{N} \sum_{i=1}^N r_i \quad (2.20)$$

где r_i — коэффициент ранговой корреляции между упорядочиванием объектов, проведенным i -м респондентом в I и во II опросе (либо с помощью I и II метода и т. п.), N — число респондентов.

Методы внутреннего контроля в этом случае специфичны, так как теряет смысл понятие распределения по данному признаку. Их разработку можно сформулировать как нерешенную задачу. При использовании метода парных сравнений и, возможно, метода множественных сравнений к методам внутреннего контроля надежности можно отнести контроль транзитивности предпочтений респондента [см., например: I Всесоюзное совещание по статистичес-

кому и дискретному анализу . . . , 1981, с. 122—123] ⁶ (выявление таких случаев, когда он предпочитает объект *a* объекту *b*, а объект *b* объекту *c*, но объект *c* объекту *a*).

Хотя я и не могу привести данные каких-либо экспериментов, в которых сравнивалась бы надежность различных методов упорядочивания объектов ⁷, из достаточно очевидных соображений, связанных с процедурой опроса, можно сделать вывод, что метод парных (или, как его иногда называют, попарных) сравнений является наиболее надежным методом выявления предпочтений. «Метод попарного сравнения, — отмечает Б. Г. Миркин [1974, с. 32], — применяется обычно, чтобы выявить предпочтения экспертов «в чистом виде». Считается, что качественное сравнение двух объектов сделать гораздо легче, чем выражать свои предпочтения в балльной или ранговой шкале, этот метод оценки не навязывает эксперту априорных условий в отличие от других». В методе парных сравнений упорядочивание сводится к самому элементарному акту выбора, поэтому маловероятно, чтобы для небольшого числа объектов можно было бы предложить более надежный метод. . .)

Вместе с тем в этом методе при увеличении числа объектов число требуемых сравнений растет почти пропорционально квадрату числа объектов, возрастает трудоемкость метода и надежность его быстро падает. Думаю, что при числе объектов большем 10—15 (что требует 45—105 сравнений) целесообразность применения этого метода становится сомнительной. Что касается метода ранжирования, то его, как мне представляется, можно использовать при ранжировании не более чем 20—25 объектов, увеличение числа объектов влечет к существенным сложностям и падению надежности. Можно предложить метод изучения предпочтений, который по надежности близок к методу парных сравнений, но годится для большего числа объектов (80—100), его можно назвать методом множественных, или *k*-арных сравнений — по числу сравниваемых (одновременно) объектов [Паниотто, 1980].

⁶ Это упоминавшееся уже выступление Казанской Т. А. «Исследование непротиворечивости данных экспертизы, проведенной методом парных сравнений», о котором я не могу однозначно сказать, посвящено ли оно надежности исходной информации, поскольку вопрос о том, чем вызвана транзитивность: влиянием ошибок или природой объектов исследования, приводящей к тому, что не все пары объектов эксперт оценивает исходя из одних и тех же критериев.

⁷ Ниже будет рассмотрена на экспериментальном материале устойчивость ранжирования профессий по престижу, но для небольшого числа объектов. Другие данные [Саганенко, 1969] не дают возможности определить, какой из сравниваемых методов надежнее.

Предлагаемый метод представляет собой экстраполяцию метода парных сравнений на большое число объектов. Как показывают эксперименты, для небольшого числа объектов метод парных сравнений эквивалентен непосредственному упорядочиванию объектов, в частности, ранжированию карточек с названиями объектов, то есть ранжировки, полученные с помощью метода парных сравнений и путем непосредственного ранжирования объектов, совпадают. Обозначим через k число объектов, подлежащих ранжированию. Если $k=2$, то метод ранжирования это и есть метод парных сравнений. С увеличением k коэффициент ранговой корреляции между ранжировкой, полученной упорядочиванием карточек с названиями объектов, и ранжировкой, полученной методом парных сравнений, сначала будет равен 1, а затем, начиная с некоторого места, начнет падать. Пока он не упал ниже 1 (или ниже заданного порога), оба метода можно считать эквивалентными.

Идея метода заключается в том, чтобы предварительно экспериментальным путем определить максимальное k , при котором ранжирование карточек эквивалентно методу парных сравнений, а затем, в эмпирическом исследовании предлагать респондентам (или экспертам) ранжировать наборы не по 2 объекта, как в методе парных сравнений, а по k объектов, выбранных определенным образом из всего множества объектов. Наборы по k объектов выбираются таким образом, чтобы после транзитивного замыкания всего множества сравнений (если в одном из наборов респондент предпочел объект a объекту b , а в другом наборе объект b — объекту c , то считаем, что он предпочитает объект a объекту c) оказалось, что нет такой пары объектов, относительно которой не было бы информации о предпочтениях экспертов (непосредственной или опосредованной, полученной из соображений транзитивности). Таким путем мы заполним матрицу парных сравнений, то есть получим такую же информацию, какую получим при использовании метода парных сравнений⁸.

⁸ Об этом методе я рассказывал на конференции по экспертным методам в изучении театра (Ленинград, 1977 г.). После этого выступления Д. С. Шмерлинг сообщил мне, что такой метод уже известен. Правда, оказалось, что в работах, посвященных методу множественных (прежде всего, 3-арных) сравнений, он рассматривается в несколько ином аспекте. В частности, в них не затрагиваются вопросы эквивалентности метода парных и множественных сравнений, а также не используются блок-схемы и соображения транзитивности для минимизации числа наборов, предъявляемых респондентам. Более того, задача минимизации числа наборов вообще не ставится авторами, что ограничивает применение описанных моделей небольшим числом объектов [см., например: Дэвид, 1978, с. 12, 13].

Что дает применение этого метода? Предположим, требуется проранжировать несколько десятков объектов (40—50). При использовании метода парных сравнений для этого понадобилось бы около 1000 сравнений, что практически неосуществимо. При использовании же описанного метода множественных сравнений может понадобиться 6—7 наборов по 10 объектов, что при корректной организации исследования занимает в среднем не более получаса. Надежность же полученных данных будет близка к надежности данных, полученных методом парных сравнений для 10 объектов, в той мере, в какой верно предположение о транзитивности предпочтений респондента. Можно исключить предположение о транзитивности предпочтений респондента, потребовав, чтобы каждая пара объектов попала хотя бы в один из наборов по k объектов (то есть всю информацию получать из непосредственных оценок экспертов, не используя процедуру транзитивного замыкания множества сравнений), но при этом существенно увеличивается необходимое число наборов по k объектов n , следовательно, время, требуемое на опрос респондента.

1.3. Вопросы других типов. Напоминаю, что мы рассматриваем определение надежности показателя, рассчитываемого для каждого респондента по данным ответа на один вопрос. Был рассмотрен случай, когда ответом на вопрос является значение некоторого признака X (измеренное в номинальной, порядковой или метрической шкале), и случай, когда ответом на вопрос является упорядочивание всегда предложенного набора объектов (альтернатив). Существуют, однако, вопросы других типов — например, вопрос, в котором респондент сначала отбирает некоторое число объектов, а затем ранжирует отобранные объекты.

Введение показателей надежности для таких типов вопросов является пока нерешенной задачей. Правда, поскольку ответы на вопросы любых типов представимы в виде матриц [Докторов, Фирсов, 1975, с. 63—73], содержащих только 0 и ± 1 , то показатель надежности, наверно, можно вводить как функцию от таких матриц. Например, можно ввести показатель устойчивости как нормированное число совпадающих единиц.

2. Определение надежности показателя, рассчитываемого для одного респондента по множеству вопросов. Социологи редко ограничиваются изучением только тех признаков, которые содержатся в анкете. Обычный прием анализа социологической информации — это своего рода конструирование новых признаков как функции от исходных.

Создать новый признак — это значит по ответам каждого респондента на ряд вопросов, содержащихся в анкете, вычислить некоторую арифметическую или логическую функцию и значение этой функции отнести к каждому респонденту. Тем самым респондент как бы отвечает на невключенный в анкету $(k+1)$ -й вопрос (k — число вопросов в анкете). Создание таких признаков предусмотрено во всех современных пакетах программ анализа социологической информации (если такая возможность не предусмотрена — пакет несовременный). Примерами могут служить «логический квадрат», построение типологий (если респондент указал 2-й и 4-й вариант ответа в вопросе 3 и 1-й или 3-й в вопросе 11 и не указал 6-й вариант в вопросе 7 — это тип 1, если указал первые три варианта ответа в вопросе 4 — это тип 2 и т. д.), определение тестовой оценки по ответам на данное множество вопросов теста, вычисление индексов и т. п. Кроме того, исследователя может интересовать среднее по некоторым признакам для данного респондента, дисперсия, коэффициент корреляции (в приведенных выше экспериментах при разработке метода множественных сравнений для каждого респондента рассчитывался коэффициент корреляции между ранжировками профессий методом парных сравнений и методом непосредственного упорядочивания).

Во всех случаях итогом является некоторая функция от исходных признаков анкеты. Значения этой функции и есть тот показатель, надежность которого требуется определить. Здесь возникает две задачи: непосредственное определение надежности показателя (с помощью описанных выше методов контроля) и определение надежности показателя через надежность исходных признаков.

2.1. *Определение надежности показателя с помощью методов внешнего и внутреннего контроля.* Поскольку значением функции является некоторое число — номинальное (если функция логическая), порядковое или метрическое, то в качестве показателей надежности пригодны все описанные для соответствующих случаев показатели.

Кроме того, в тестовых методиках, содержащих набор однородных признаков, надежность может определяться методом расщепления (иногда близость результатов, полученных по разным подгруппам тестовых вопросов, называют свойством *внутренней консистентности* теста). Набор вопросов, входящих в тест, разбивается на две группы (например, нечетные и четные вопросы). Если тест представляет собой некоторую функцию от m вопросов (например, среднее арифметическое от ответов на каждый исход-

ный вопрос) и выделены группы вопросов $(X_{l_1}, X_{l_2}, \dots, X_{l_r})$ и $(X_{l_{r+1}}, X_{l_{r+2}}, \dots, X_{l_s})$, где $l_1, l_2, \dots, l_r$ и $l_{r+1}, \dots, l_s$ некоторые подмножества из множества $(1, 2, \dots, m)$, то для каждого респондента рассчитываются значения функции $f(X_{l_1}, X_{l_2}, \dots, X_{l_r})$ и $f(X_{l_{r+1}}, X_{l_{r+2}}, \dots, X_{l_s})$, а затем рассчитывается коэффициент корреляции, который и является показателем внутренней консистенции (в нашей терминологии — показателем устойчивости, так как сходство показателей для различных подмножеств теста — косвенный показатель отсутствия случайных ошибок).

Существует много способов расщепления. Если рассматривать только расщепления пополам, то их может быть $C_m^{m/2}$. Предположим, что тест построен так, что за «правильный», релевантный ответ на вопрос присваивается 1, за «неправильный», нерелевантный — 0, и результатом является сумма показателей по всем вопросам⁹. Предположим также, что всеми возможными способами тест был расщеплен на два подтеста с равным количеством вопросов, для каждого расщепления рассчитывался коэффициент корреляции r и затем было найдено среднее арифметическое всех коэффициентов корреляции. Это среднее и используется в качестве показателя надежности информации, но вычисляется оно проще, без выполнения указанных выше операций, с помощью формулы, которая почему-то называется формулой Кудера — Ричардсона-20 [Dick, Nagerthy, 1971, p. 54]:

$$r_{tt} = \frac{m}{m-1} \cdot \left(\frac{\sigma^2 - \sum_{i=1}^m p_i q_i}{\sigma^2} \right), \quad (2.21)$$

где m — число вопросов в тесте, p_i — доля респондентов, «правильно» ответивших на вопрос i , q_i — доля респондентов, «неправильно» ответивших на вопрос i ($q_i = 1 - p_i$), σ — дисперсия теста.

Для некоторых случаев удобна модификация формулы, не требующая распределений ответов каждого респондента (формула Кудера — Ричардсона-21):

$$r_{tt} = \frac{m}{m-1} \cdot \left(\frac{\sigma^2 - m\bar{p}\bar{q}}{\sigma^2} \right), \quad (2.22)$$

⁹ Определения «правильный» и «неправильный» употреблены в переносном смысле: если измеряется некоторое свойство (скажем, интроверсия), то ответ на вопрос, свидетельствующий о более высокой выраженности свойства, обозначается как единица («правильный»), а о более низкой — как ноль («неправильный»). В дальнейшем будем говорить о релевантных и нерелевантных (в указанном смысле) вопросах.

где $\bar{p}$ — средняя доля «правильных» ответов.

$$\bar{p} = \frac{1}{N} \sum_{i=1}^N p_i = \frac{\bar{y}}{m} = \frac{1}{m} \cdot \frac{1}{N} \sum_{i=1}^N y_i, \quad (2.23)$$

где y_i — тестовый показатель i -го респондента (уже итоговый, а не исходный, как в предыдущей формуле), N — число респондентов.

2.2. *Определение надежности итогового (производного) показателя через характеристики надежности исходных признаков.* Задача эта решается лишь для ряда частных случаев. Если итоговый показатель $y = f(X_1, X_2, \dots, X_m)$ представляет собой аналитическую функцию от исходных признаков и известны среднеквадратические ошибки исходных признаков $S_{x_1}, S_{x_2}, \dots, S_{x_m}$, то среднеквадратическая ошибка итогового признака приближенно вычисляется по формуле [см., например: Саганенко, 1979, с. 81]:

$$S_y \approx \sqrt{\left(\frac{\partial f}{\partial x_1}\right)^2 S_{x_1}^2 + \dots + \left(\frac{\partial f}{\partial x_m}\right)^2 S_{x_m}^2}. \quad (2.24)$$

В том случае, когда исходные признаки независимы между собой, равенство становится точным. Если, например, $y = a_1 x_1 + a_2 x_2 + \dots + a_m x_m + b$, то

$$S_y = \sqrt{\sum_{i=1}^m a_i^2 S_{x_i}^2}. \quad (2.25)$$

Частным случаем задачи определения надежности итогового показателя является также определение надежности методики в целом. Если у нас есть показатели надежности для всех вопросов анкеты и эти показатели однородны (например, изменяются от 0 до 1), то надежность методики может определяться просто как средняя арифметическая показателей отдельных вопросов. Г. И. Саганенко [там же, с. 77] предлагает для этих же целей показатель, являющийся обобщением показателя (2.9).

Существуют также показатели, разработанные для тестовых методик и позволяющие определить надежность теста в целом, если известна надежность подтестов (при этом, разумеется, предполагается, что тест содержит однородные вопросы):

$$r_{tt} = \frac{\alpha r_{11}}{1 + (\alpha - 1) r_{11}}, \quad (2.26)$$

$$\alpha = \frac{m}{k}, \quad (2.27)$$

где r_{tt} — надежность всего теста, r_{11} — надежность подтеста, коэффициент определен для подтеста по формуле (2.7) или (2.21), k — число вопросов в тесте, m — число вопросов в подтесте.

Эта формула называется формулой Спирмена—Брауна и используется также при конструировании тестов. Если исследователь обнаружил, что создаваемый им тест имеет недостаточную надежность, то по формуле (2.26) можно определить, во сколько раз нужно увеличить число вопросов, чтобы добиться нужной надежности теста. Для этой цели удобнее, впрочем, пользоваться альтернативной формой этой формулы [см., например: Аванесов, 1982, с. 121, 122] (обозначения те же):

$$\alpha = \frac{r_{tt}(1 - r_{11})}{r_{11}(1 - r_{tt})}. \quad (2.28)$$

Формулой Спирмена — Брауна можно пользоваться также в том случае, если исследователь намерен сократить число вопросов в тесте, и его интересует, какой при этом станет надежность.

3. Определение надежности показателя, рассчитываемого для множества респондентов по ответам на один вопрос. К таким показателям относятся те или иные характеристики распределения. Если вопрос номинальный, то это может быть доля признака, мода, энтропия распределения; если вопрос порядковый — то же, что и для вопроса номинального, но также медиана и квантильный размах; если метрический — среднее арифметическое и дисперсия (реже — другие характеристики).

При методах внешнего контроля для определения смещений могут использоваться обычные статистические показатели значимости различий долей признака, значимости различий средних арифметических, значимости различий дисперсии [см., например: Статистические методы . . . , 1979; Паниотто, Максименко, 1982, с. 191—198]. Несколько хуже обстоит дело с определением значимости статистических различий моды, медианы, квантильного размаха [Закс, 1976, с. 236], но эти показатели используются значительно реже.

При методах внутреннего контроля могут использоваться упомянутые в предыдущем параграфе дисперсионные критерии надежности (хотя, вообще говоря, вызывает со-

мнение отнесение их к данному классу, поскольку речь идет не только об определении надежности показателей распределения). Введем соответствующие статистические показатели. Предположим, что измерен признак Y и требуется оценить надежность инструмента измерения для сравнения с надежностью какого-либо другого инструмента. Предположим, что одновременно с признаком Y мы измерили ряд признаков, которые оказывают значительное влияние на признак Y (это обычно такие признаки, как пол, возраст, социальное положение и т. п.). Выберем один из них X . Предположим, что, анализируя ошибки, которые могли возникнуть в процессе измерения признака Y , мы пришли к выводу, что они носят случайный характер и могут лишь уменьшить степень влияния X на Y , точнее уменьшить различия переменной y лиц, имеющих различные значения переменной. Введем показатель, характеризующий влияние случайных ошибок. Рассмотрим четыре случая.

3.1. Признак X — неметрический (номинальный или порядковый), признак Y — метрический. Данные содержатся в следующей матрице:

$$\begin{array}{c|cccc}
 & 1 & 2 & \dots & N \\
 \hline
 x_1 & y_{11} & y_{12} & \dots & y_{1N} \\
 x_2 & y_{21} & y_{22} & \dots & y_{2N} \\
 \vdots & \dots & \dots & \dots & \dots \\
 x_n & y_{n1} & y_{n2} & \dots & y_{nN}
 \end{array} \quad (2.29)$$

где x_i — варианты значений признака X , y_{ij} — значение признака Y j -го респондента, имеющего значение x_i признака X ($i=1, 2, \dots, n$; $j=1, 2, \dots, N$), n — число градаций признака X , N — число респондентов.

По данным матрицы (2.29) рассчитаем:

$$y_i = \frac{1}{m_i} \sum_{j=1}^N y_{ij}; \quad \bar{y} = \frac{1}{N} \sum_{i=1}^n y_i; \quad \sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (\bar{y}_i - \bar{y})^2}, \quad (2.30)$$

где m_i — число ненулевых элементов i -й строки.

Тогда показателем надежности будет коэффициент вариации:

$$V^{(1)} = \left| \frac{\sigma}{\bar{y}} \right|. \quad (2.31)$$

3.2. Оба признака неметрические. Данные опроса могут быть сведены в матрицу:

$$\begin{array}{c|cccc|c}
 & y_1 & y_2 & \cdots & y_k & \\
 \hline
 x_1 & n_{11} & n_{12} & \cdots & n_{1k} & n_1 \\
 x_2 & n_{21} & n_{22} & \cdots & n_{2k} & n_2 \\
 \vdots & \cdots & \cdots & \cdots & \cdots & \vdots \\
 x_n & n_{n1} & n_{n2} & \cdots & n_{nk} & n_n
 \end{array} \quad (2.32)$$

где y_j — варианты значений признака Y ($j=1, 2, \dots, k$), k — число вариантов значений признака Y , n_{ij} — число людей, имеющих значение x_i признака X и одновременно значение y_j признака Y .

$$n_i = \sum_{j=1}^k n_{ij}. \quad (2.33)$$

Преобразуем матрицу (2.32) так, чтобы в ее клетках стояли доли по отношению к суммам частот в строке:

$$\begin{array}{c|cccc|c}
 & y_1 & y_2 & \cdots & y_k & \\
 \hline
 x_1 & p_{11} & p_{12} & \cdots & p_{1k} & \\
 x_2 & p_{21} & p_{22} & \cdots & p_{2k} & \\
 \vdots & \cdots & \cdots & \cdots & \cdots & \\
 x_n & p_{n1} & p_{n2} & \cdots & p_{nk} &
 \end{array} \quad (2.34)$$

где $p_{ij} = \frac{n_{ij}}{n_i}$.

Введем показатель, характеризующий различия между строками полученной матрицы. Это и будет показатель надежности:

$$V^{(2)} = \frac{1}{k} \sum_{j=1}^k v_j, \quad (2.35)$$

где

$$v_j = \frac{\sigma_j}{p_j}; \quad \sigma_j = \sqrt{\frac{1}{n} \sum_{i=1}^n (p_{ij} - \bar{p}_j)^2}; \quad \bar{p}_j = \frac{1}{n} \sum_{i=1}^n p_{ij}.$$

3.3. Признак X — метрический, признак Y — неметрический. Исходные данные сведены в матрицу (2.36)

	y_1	y_2	$\dots$	y_k	
1	x_{11}	x_{12}	$\dots$	x_{1k}	(2.36)
2	x_{21}	x_{22}	$\dots$	x_{2k}	
$\vdots$	$\dots\dots\dots$				
m	x_{m1}	x_{m2}	$\dots$	x_{mk}	

где x_{ij} — значение признака X i -го респондента, имеющего значение y_j признака Y , m_j — число ненулевых элементов j -го столбца.

Обозначим:

$$\bar{x}_j = \frac{1}{m_j} \sum_{i=1}^m x_{ij}; \quad \sigma_y = \sqrt{\frac{1}{m} \sum_{i=1}^m (x_{ij} - \bar{x}_j)^2}; \quad w_j = \left| \frac{\sigma_j}{\bar{x}_j} \right|. \quad (2.37)$$

Тогда искомый показатель равен:

$$V^{(3)} = 1 - \frac{1}{k} \sum_{j=1}^k w_j. \quad (2.38)$$

Кроме того, можно свести признак X к качественному, разбив область его изменения на интервалы, и пользоваться формулой (2.35).

3.4. *Оба признака метрические.* В этом случае за показатель надежности можно принять коэффициент корреляции между признаками X и Y :

$$V^{(4)} = |r_{xy}|. \quad (2.39)$$

Показатели $V^{(2)}$ и $V^{(3)}$ построены, по существу, соответственно на факторной и остаточной корреляции, применяемой в дисперсионном анализе (отличаются исходные матрицы). Поэтому все четыре коэффициента можно назвать дисперсионными коэффициентами надежности.

4. **Определение надежности показателя, рассчитываемого для множества респондентов по ответам на множество вопросов.** К таким показателям относятся меры связи, группировка респондентов в те или иные классы, результаты факторного анализа и т. п. Это, фактически, вторичные итоговые показатели, результаты математического анализа исходных данных, по которым производится интерпретация. Так же, как и в предыдущем случае, задача распадается на две: непосредственное определение надежности и определение надежности по показателям надежности исходных признаков. Например, зная устойчивость измерения исходных признаков X и Y (причем устойчи-

Т а б л и ц а 2.1. Показатели надежности социологической информации

Число вопро- сов	Тип пока- зателя	Тип шкалы	Число		
			один		
			контроль		
			внешний по отношению		внутренний
			к моменту, усло- виям исследова- ния, исполнителю	к методу иссле- дования	
один	значение признака	номи- нальная	показатели (2.9), (2.11)	коэф. корреля- ции для таблиц	показатели см. в: [Са- гапенко, 1979, с. 38—46]
		порядко- вая	показатели (2.13) — (2.19)	коэффициенты	
		метриче- ская	показатели (2.7), (2.8)		
	упорядо- чивание	—	(2.20)		нерешенная задача
	другие типы	—	нерешенная задача		
множе- ство	значение признака	—	все показатели, которые используются в случае «один респондент — один вопрос», а также (2.21), (2.22), (2.24), (2.26), (2.28)		
	упорядо- чивание	—	нерешенная задача		
	другие типы	—	нерешенная задача		

вость вычислялась как коэффициент корреляции между первым и вторым опросом отдельно для признака X — r_{xx} и для признака Y — r_{yy}), можно уточнить значение коэффициента корреляции между этими признаками¹⁰:

$$r_{xy} = \frac{\tilde{r}_{xy}}{\sqrt{r_{xx}} \cdot \sqrt{r_{yy}}}, \quad (2.40)$$

¹⁰ Приводимую формулу Б. З. Докторов, ссылаясь на Спирмена, называет поправкой на аттенуацию [Докторов, 1979, с. 54]. Она, по-видимому, предназначена для компенсации уменьшения коэффициента корреляции, связанного с низкой устойчивостью признаков X и Y .

респондентов	
множество	
контроль	
внешний	внутренний
значимости различий долей, средних и т. п.	доля признака, медиана и другие показатели описательной статистики; см. также: (2.31), (2.35), (2.38), (2.39).
нерешенная задача	нерешенная задача
решены частные случаи этой задачи — см.: показатель (2.40); методы планирования экспериментов	

где r_{xy} — уточненное значение коэффициента корреляции, $\tilde{r}_{xy}$ — значение коэффициента корреляции, рассчитанное по данным опроса.

Может быть, можно найти еще несколько примеров оценки влияния надежности исходных признаков на результаты математического анализа социологической информации, но в целом это целина, которая ждет своих сеятелей. Имеет смысл для всех существующих методов анализа социологической информации (корреляционный, регрессионный, факторный, кластерный, причинный, логлинейный анализ, многомерное шкалирование и т. п.) разработать методы, оценивающие их устойчивость в зависимости от надежности исходных признаков.

Задача непосредственного определения надежности показателей, полученных в ходе математической обработки эмпирических данных, относится не только (а может быть

и не столько) к качеству социологической информации, но и к качеству методов ее анализа, и обычно она рассматривается именно в этом плане [Сагапенко, 1983, с. 116—176]. Однако несомненно она имеет отношение и к качеству социологической информации, так как существует ряд факторов, влияющих на надежность результатов анализа, но принципиально не устранимых путем совершенствования методов анализа информации, а требующих изменения способа сбора первичных данных. Эти вопросы относятся к области планирования эксперимента. Здесь получено много интересных результатов, существует обширная лите-

ратура, поэтому я отсылаю читателя к специальным работам, обобщающим полученные результаты [Налимов, 1971; Кэмпбелл, 1980; Готтсданкер, 1982]. Особенно обстоятельной является книга Д. Кэмпбелла. Как справедливо отмечает Г. М. Андреева [1980, с. 7], «широта взглядов, глубокая образованность и большой круг проблем, при разработке которых Кэмпбелл выступает как профессионал-исследователь, не только показывают его компетентность в ряде областей социальной науки, но и придают особый вес его методологическим исследованиям».

Итак, рассмотрены статистические оценки надежности информации для различных ситуаций (табл. 2.1). Следующий шаг после оценки надежности — использование методов повышения надежности в тех случаях, когда имеющийся уровень надежности не устраивает исследователя. Некоторые из этих методов были описаны при рассмотрении методов оценки надежности, но речь шла преимущественно о повышении правильности и устойчивости информации. В следующих главах мы остановимся на других составляющих надежности — репрезентативности и обоснованности.

III. ВОПРОСЫ РЕПРЕЗЕНТАТИВНОСТИ

Полное осуществление выборки... дается намного труднее, чем это можно предположить, читая статистическое пособие о формировании выборочной совокупности».

Э. Ноэль

В этой главе мы рассмотрим показатели репрезентативности социологической информации. Похоже, что вопросы репрезентативности хорошо разработаны в математической и статистической литературе. Библиография работ по теории выборки выглядит намного солиднее, чем по социологии. Казалось бы, никаких проблем — читай и пользуйся. И поначалу у меня было намерение не писать на эту тему, а ограничиться ссылками на имеющиеся работы. Но попытка дать такие ссылки привела к неожиданному выводу: показатели, используемые для описания выборки и оценки репрезентативности информации в статистической литературе, в большинстве случаев не совсем подходят (а во многих случаях — и совсем не подходят) для описания выборки реального социологического исследования. Прежде всего, это относится к многоступенчатым выборкам, осуществляемым при проведении крупномасштабных социологических исследований. Кроме того, существуют такие проблемы репрезентативности, которые вообще не относятся к статистике и требуют специфического подхода. Поэтому в данной главе будет сделана попытка показать, что различные характеристики выборки определяются не столько статистическими требованиями, сколько требованиями глубины анализа информации, затем будут рассмотрены показатели, описывающие выборку, и некоторые виды социологического исследования, требующие нестатистического подхода к репрезентативности информации.

1. Статистический подход к репрезентативности

Объектом социологических исследований обычно являются социальные группы. Если изучаются все индивиды данной совокупности, то говорят о сплошном исследовании, если же только часть, то — о выборочном. Как правило, социологические исследования (так же как и биологические, химические и т. п.) носят выборочный характер. Это связано прежде всего с экономическими и временными ог-

раничениями на проведение сплошного исследования (затраты на проведение Всесоюзной переписи, например, составляют десятки миллионов рублей и требуют более 700 тысяч интервьюеров). Но даже в тех случаях, когда сплошное исследование практически осуществимо, зачастую рентабельней проводить выборочное. Сэкономленные по сравнению со сплошным исследованием ресурсы позволяют уделить больше внимания совершенствованию инструмента исследования и тем самым не только компенсировать снижение надежности за счет того, что исследование не сплошное, а выборочное, — но и получить в итоге более полную и надежную информацию. Но для того чтобы сплошное исследование можно было заменить выборочным, требуется, чтобы интересующие нас свойства изучаемой совокупности были адекватно представлены в выборочной совокупности.

(В статистике исследуемую совокупность принято называть генеральной, а способность выборки (выборочной совокупности) отражать характеристики генеральной совокупности — репрезентативностью выборки. Можно сказать, что генеральная совокупность — это та, на которую исследователь намерен распространить выводы, сделанные при изучении выборки. Различие характеристик выборочной и генеральной совокупности называют ошибкой репрезентативности. Как уже указывалось, можно выделить два вида таких ошибок — систематические и случайные.)

Систематические ошибки представляют собой некоторое постоянное смещение, которое не уменьшается с увеличением числа опрошенных и вызвано недостатками и просчетами в системе отбора респондентов. Если, например, для изучения общественного мнения жителей города в архитектурном управлении получить сведения о жилом фонде и из всех имеющихся в городе квартир отобрать случайным образом 400 квартир, а затем предложить интервьюерам опросить всех, кого они застанут в момент посещения в этих квартирах, то полученные данные не будут репрезентативны. Допущена систематическая ошибка: более подвижная часть населения попадает в выборку в меньшей пропорции, а менее подвижная — в большей пропорции, чем в генеральной совокупности. Пенсионеров, например, можно чаще застать дома, чем студентов-вечерников. При увеличении выборки эта ошибка не устраняется: если мы проведем опрос в 800 квартирах или даже во всех квартирах города (то есть сплошной опрос), то полученные данные будут репрезентативны для населения,

находящегося дома в момент прихода интервьюера, а не для всех жителей города. Заметим, что определение генеральной совокупности не всегда очевидно. Так, при изучении общественного мнения жителей города следует определить, включать ли в изучаемую совокупность учащихся (если да, то с какого возраста, так как для многих ситуаций не имеет смысла опрашивать, скажем, школьников младших классов), маятниковых мигрантов, относить к генеральной совокупности лиц, проживающих в городе, но не прописанных в нем, лиц, находящихся в больницах, военнослужащих частей, находящихся на территории города и т. п.

Случайные ошибки — это те, которые при повторных измерениях изменяются по вероятностным законам. В частности, если мы определяем некоторую характеристику выборки, например, среднее арифметическое, то извлекая все новые и новые случайные выборки того же размера, будем получать отклонение этой характеристики то в одну, то в другую сторону от истинного значения (то есть от значения в генсовкупности) приблизительно с одинаковой частотой. При увеличении числа выборок средняя ошибка приближается к нулю. Но это лишь при условии, что выборки действительно являются случайными.

Случайность выборки вовсе не означает хаотичность отбора, опрашивая «кого попало», трудно получить случайную в математико-статистическом смысле выборку. Случайным отбором (случайной выборкой) называют такой, при котором все элементы исследуемой совокупности имеют равную вероятность попасть в выборку. Отклонение от этого принципа отбора является систематической ошибкой. Так, в приведенном выше примере студенты-вечерники, например, имели меньше вероятность попасть в выборку, чем пенсионеры, а для обеспечения случайного отбора необходимо, чтобы эти вероятности были равны (далее мы приведем пример, показывающий, как обеспечивается равенство вероятностей). Систематическую ошибку можно устранить, изменяя процедуру формирования выборки: случайная ошибка будет присутствовать всегда, при любом выборочном опросе. Тем не менее, систематическая ошибка значительно опасней, так как по выборке ее невозможно оценить. Случайная же ошибка подчиняется определенным закономерностям и поддается оценке. Выборка, при которой непосредственно из генеральной совокупности случайным образом отбираются респонденты, называется одноступенчатой случайной. Рассмотрим, как определить ошибку (или объем) выборки данного типа.

Одноступенчатая случайная выборка. Репрезентативность выборки характеризуется двумя взаимосвязанными параметрами — уровнем ошибки и вероятностью отсутствия ошибки данного уровня. Говорить о какой-либо выборке, что она репрезентативная, не совсем точно, так как любая выборка имеет определенный уровень репрезентативности (хотя этот уровень может нас совершенно не устраивать). Более точно говорить, что ошибка репрезентативности данной выборки с вероятностью P не превышает Δ (вероятность P называют доверительной). Задавшему некоторым уровнем вероятности, ошибку выборки можно определить, если известен объем генсовокупности и выборки. При планировании социологического исследования обычно решают иную задачу — задаются некоторым устраивающим исследователя уровнем точности результатов, то есть допустимой ошибкой и доверительной вероятностью и вычисляют необходимый объем выборки. В частности, объем выборки для определения доли некоторого признака X в генеральной совокупности находится по формуле [см., например: Кокрэн У., 1976, с. 90]:

$$n = \frac{1}{\frac{\Delta^2}{t^2 v (1 - v)} + \frac{1}{N}}, \quad (3.1)$$

где N — объем генеральной совокупности, n — объем выборки, t — коэффициент, соответствующий доверительной вероятности P (он определяется по так называемым таблицам Стьюдента; при $P=0,954$ $t=2$, а при $P=0,997$ $t=3$ эти два значения чаще всего используются в социологии) v — доля признака X в генсовокупности, Δ — величина допустимой ошибки¹ (в долях).

Если, например, исследователь хочет получить с вероятностью 0,95 (при этом $t \approx 2$) данные о доле признака X (скажем, доле мужчин) в генсовокупности (пусть $N=10\,000$) с ошибкой, не превышающей 5% ($\Delta=0,05$), и ему известно, что искомая доля составляет приблизительно 20% ($v=0,20$), то по формуле (3.1) получим, что требуется опросить 256 человек ($n=256$).

¹ Иногда выделяют абсолютную и относительную ошибку. Если m — число лиц с данным значением признака X , полученное с ошибкой $\pm k$ респондентов, то абсолютная ошибка в долях равна $\Delta = \frac{k}{n}$, а относительная $\delta = \frac{k}{m}$. Здесь и далее всюду имеется в виду абсолютная ошибка ($v \pm \Delta$).

Недостаток этой формулы заключается в том, что она требует хотя бы приближенной информации о доле признака в генеральной совокупности, то есть как раз о том, что исследователю требуется определить. Однако при $v = 0,5$ произведение $v(1-v)$ максимально, следовательно, n тоже максимально. Поэтому если в (3.1) вместо v подставить 0,5, получится формула, которой можно пользоваться при любых значениях доли признака в генеральной совокупности (объем выборки при этом будет получаться с некоторым запасом). Зафиксировав также значение доверительной вероятности, положив его равным 0,954 (при этом $t=2$), получим:

$$n = \frac{1}{\Delta^2 + \frac{1}{N}} \quad (3.2)$$

По этой формуле можно исследовать, как объем выборки зависит от объема генеральной совокупности и от величины допустимой в исследовании ошибки:

Объем генеральной совокупности	500	1000	2000	3000	4000	5000	10 000	100 000	∞
Объем выборки	222	286	333	350	360	370	385	398	400

$$\Delta = 0,05$$

$$P = 0,954$$

Отсюда видно, что для обеспечения заданной репрезентативности при исследовании города с населением 100 тысяч жителей надо опросить 398 человек, а при исследовании всей страны практически столько же — 400 человек. Для обеспечения одного и того же уровня репрезентативности (5%) требуется опросить такие доли генсовокупности: для $N=500$ $n=222$ человека, то есть приблизительно 44% генсовокупности, для $N=5000$ $n=370$ человек — 7,4%, а для $N=4\,000\,000$ (например, население Ленинграда) $n=400$ человек — 0,01% генсовокупности. Поэтому изредка встречающиеся в публикациях характеристики выборки типа «было опрошено 15% генсовокупности» или «опрашивался каждый двадцатый школьник» ничего не говорят о репрезентативности выборки. Начиная с некоторого момента (см. выше) увеличение объема генеральной совокупности не оказывает существенного влияния на увеличение объема выборки, поэтому при больших генеральных совокупностях (скажем, при $N > 5000$) величиной $\frac{1}{N}$ в формуле (3.2) можно пренебречь. Тогда формула (3.2) примет вид:

$$n = \frac{1}{\Delta^2}, \quad (3.3)$$

откуда
$$\Delta = \frac{1}{\sqrt{n}}. \quad (3.4)$$

Принимая решение относительно объема выборки, важно знать соотношение между ошибкой и объемом выборки (в случае, если генеральная совокупность больше 5 000 единиц, можно считать, что она бесконечная и пользоваться формулами (3.3) и (3.4), объем выборки при этом получим с некоторым запасом):

Ошибка выборки, %	20	15	10	9	8	7	6	5	4	3	2	1	0,5
Объем выборки	25	45	100	123	156	204	278	400	625	1111	2500	10 ⁴	4·10 ⁴
$N = \infty$ $P = 0,954$													

При планировании выборки следует иметь в виду следующее. Формула (3.3) и эта таблица позволяют получить объем выборки для заданной точности при анализе выборки в целом, то есть в том случае, когда мы не будем рассчитывать ее на части. Если, например, требуется определить долю лиц, состоящих в браке, для крупного города, то опросив 625 случайным образом отобранных человек мы определим искомую долю с ошибкой, не превышающей 4% (с вероятностью 0,954). Но если требуется определить эту долю с такой же точностью не для всего массива в целом, а для женщин и для мужчин, то необходимо, чтобы в выборке было 625 мужчин и 625 женщин, то есть 1250 человек. Чем больше будет дробиться массив при анализе информации, тем больший объем выборки потребуется.

Эти соображения приводят к важному выводу: объем выборки (вопреки распространенной среди многих социологов точке зрения) зависит в большей степени от задания на обработку анкет (от степени подробности анализа), чем от того, с какой ошибкой готов примириться исследователь. Дело в том, что ошибка обычно задается в соответствии с некоторым стандартом: в социологии, как правило, считают достаточно точной выборку, позволяющую с вероятностью 0,95 оценить долю признака в генеральной совокупности с ошибкой 5% (при этом объем выборки равен приблизительно 400). Таким образом, чтобы получить описание любой генеральной совокупности с помощью одномерных распределений, достаточно опросить около 400 случайно отобранных человек. Если же задание на обра-

ботку включает расчет двумерных распределений и исследователю потребуются сравнить между собой строки (или столбцы) двумерного распределения, каждая строка (столбец) должна быть достаточно наполненной. При этом можно ориентироваться на различия, которые исследователь хотел бы интерпретировать как значимые.

Для приближенной оценки требуемой наполненности двумерных таблиц распределения можно использовать данные следующей таблицы, в которой приведены значимые различия долей некоторого признака при сравнении двух равных по численности групп респондентов, выбранных случайным образом из бесконечной генеральной совокупности.

Численность сравниваемых групп	50	100	150	200	300	500	1000	5000
Какая разность долей значима, %	20	14	11,5	10	8	6,3	4,5	2

$$n_1 = n_2$$

$$P = 0,954$$

Поскольку эта разность значима или не значима в зависимости от конкретного сочетания долей, таблица составлена для случая, когда одна из долей равна 0,5 (50%), при этом численность групп максимальна, то есть расчеты сделаны с некоторым запасом.

Если исследователь считает, что ему достаточно интерпретировать различия, превышающие 10%, а среди признаков, являющихся подлежащими в таблицах двумерных распределений, максимальное число градаций имеет признак с k градациями, то объем выборки равен $200 k$. Если при этом выборка получается слишком объемной, то исследователь, естественно, может ориентироваться на признаки с меньшим числом градаций; если, например, большинство признаков имеет до 6 градаций, а несколько признаков 6 и более, то опросив около 1000 человек, исследователь сможет для большинства признаков интерпретировать различия, превышающие 10%, а для признаков с большим числом градаций — различия, превышающие 11,5%, 14% и т. п. (в зависимости от численности каждой из сравниваемых групп). Когда требуется анализ трехмерных распределений, то объем выборки увеличивается еще во столько раз, сколько градаций имеет признак максимальный из добавляемых по числу градаций.

Как видим, теория случайных выборок достаточно хорошо разработана, она позволяет определить, сколько респондентов надо опросить, чтобы результаты исследования

имели заданный уровень репрезентативности. Основные сложности здесь — организационные. Во-первых, во многих случаях, особенно при крупномасштабных исследованиях, сформировать случайную выборку очень трудно: отсутствуют списки респондентов; процедуры, моделирующие случайный отбор, вносят систематические ошибки и т. п. Во-вторых, при использовании одноступенчатого случайного отбора в крупномасштабных исследованиях респонденты, выбранные для опроса, рассеяны по всей исследуемой территории, что очень удорожает стоимость исследования.)

В качестве примера одноступенчатого отбора опишем исследование социальных источников воспроизводства трудовых ресурсов г. Киева, проведенное отделом конкретных социологических исследований Института философии АН УССР [Семья и воспроизводство..., 1984]. Генеральную совокупность составляло работающее население города, проживающее на его территории. Выборка была организована следующим образом. В каждом из избирательных списков всех избирательных участков города (всего их более 700, но списки избирательных участков района хранятся в райисполкомах 12-ти районов г. Киева, что существенно упрощает работу) с шагом 100 отбирались адреса избирателей². Такой отбор (например, 1-й в списке, затем 101-й, 201-й, и т. д.) называется систематическим и моделирует случайный отбор. Чисто случайным является отбор с помощью специальных таблиц случайных чисел или специальных процедур типа тех, что используются в спорте. Поскольку список не один, а много, могла быть допущена систематическая ошибка такого рода. Фамилии в списках расположены в алфавитном порядке. Начиная с 1-го номера, мы почти всегда включаем в выборочный список фамилии на букву «А», то есть в выборке доля лиц с фамилией на букву «А» будет выше, чем в генеральной совокупности. Если у лиц какой-либо национальности (например, армянской) фамилии чаще начинаются на «А», чем у лиц другой национальности, то их доля в выборке будет выше, чем в генеральной совокупности. Чтобы этого не случилось, выбор первой фамилии в списке был стохастизирован: отбор в каждом списке начинался с номера, равного целой части числа $k/7 + 6/7$, где k — номер избирательного участка, изменяющийся от 1 до 700. Теперь уже

² Для удобства изложения приведенные числа округлены или несколько изменены — реальное число участков больше 700, шаг отбора был несколько иным и т. п.

отбор первой фамилии равномерно распределен внутри шага выборки — в 7 избирательных участках фамилии будут отбираться с 1-го номера, в 7 — со 2-го и т. д. Следующая сложность: если опрашивать тех, кого мы отобрали, то из выборки выпадут лица, моложе 18 лет, а также лица, прописанные за пределами Киева, но проживающие в Киеве, если они внесены в избирательные списки по месту прописки. Кроме того, списки «стареют». Если они составлены, скажем, за год до проведения опроса, то из выборки выпадут уже лица до 19 лет, часть переехавших и т. п., то есть списки не соответствуют полностью генеральной совокупности и вносят систематическую ошибку. Для устранения ошибок использовалась некоторая модификация процедуры Киша [Кокрен, 1976, с. 384, 385]. По адресу, указанному в избирательном списке, интервьюер при посещении переписывал в определенном порядке всех проживающих (независимо от прописки) и работающих в Киеве: сначала мужчины по убыванию возраста, затем женщины и нумеровал их. Затем по специальной процедуре отыскивал номер члена семьи, подлежащего опросу [Семья и воспроизводство . . ., 1984, с. 219, 220]. Если он дома, интервьюер опрашивает его, если его нет — узнает, когда его можно застать, и приходит еще раз или несколько раз, пока не опросит его. Замены респондентов приводят к тому, что в выборку попадают в большей пропорции менее мобильные представители генеральной совокупности, то есть возникает систематическая ошибка, поэтому так называемые резервные списки снижают репрезентативность выборки.

Как видим, даже при проведении городских опросов использование одноступенчатых случайных выборок (несмотря на кажущееся наличие двух ступеней — сначала отбор адресов, потом отбор респондентов по каждому адресу — можно показать, что выборка одноступенчатая случайная, то есть все респонденты имеют равные всроятности попасть в выборку) сопряжено с некоторыми сложностями. А что делать в таких случаях, когда практически невозможно получить список всех членов генеральной совокупности, например, при построении выборки, репрезентативной для республики или страны в целом? В этом случае используются многоступенчатые выборки, которые, к сожалению, не имеют такого четкого статистического обоснования как одноступенчатые. Рассмотрим подробнее методы построения этого типа выборки.

2. Стандартизация описания выборки

Велико многообразие типов формирования выборочной совокупности. Возникающие в связи с этим сложности еще усугубляются различиями в терминологии авторов. Выборки называют стратифицированными, расслоенными, районированными, гнездовыми, квотными и т. д. Попытаемся унифицировать описание различных способов формирования выборки.

Каждую выборку можно рассматривать как состоящую из нескольких ступеней (этапов), например, при построении республиканской выборки, более подробно описанной дальше, на первом этапе отбирались области республики (устанавливалось, в каких областях будет проводиться исследование), на втором — в отобранных областях отбирались районы, на третьем — в отобранных районах отбирались населенные пункты, и наконец, на четвертом этапе в отобранных населенных пунктах производился отбор и опрос респондентов. Таким образом, республиканская выборка была 4-ступенчатой. Применительно к многоступенчатым выборкам целесообразно различать единицы отбора и единицы наблюдения. Единица отбора — промежуточный элемент выборочной совокупности, содержащий некоторое множество единиц наблюдения, то есть элементов выборки, с которых непосредственно ведется сбор социальной информации. В республиканской выборке на 1-й ступени единицами отбора были области, на 2-й — районы, на 3-й — населенные пункты, а на 4-й — единицы отбора совпали с единицами наблюдения: респонденты и отбирались и опрашивались (если бы в отобранных населенных пунктах проводился сплошной опрос, можно было бы сказать, что респонденты являются лишь единицами наблюдения).

В качестве единиц наблюдения выступают, как правило, респонденты, хотя могут быть и другие объекты (малые группы или, скажем, экземпляры газеты при контент-анализе прессы и т. п.). В том случае, когда единицы наблюдения — респонденты, можно выделить территориальные (республика, край, область, населенный пункт) и производственные (министерство, предприятие, цех, участок, бригада) единицы отбора. В зависимости от этого выборки называются территориальными или производственными. Могут быть, разумеется, и смешанные выборки — например, на первых ступенях отбираются республики, области, районы (территориальный принцип), а на последующих — в районах отбираются предприятия, цеха, участки, респонденты (производственный принцип).

Важно отметить, что те или иные названия типов выборки, с которыми читатель может столкнуться при изучении социологических публикаций или чтении статистической литературы, относятся, как правило, к одноступенчатой выборке. Многоступенчатые выборки практически не могут быть отнесены к какому-то одному типу, так как на разных ступенях используются разные типы отбора. Поэтому полное описание многоступенчатых выборок сводится к описанию типов выборки, использованных на каждой ступени. Все ступени выборки можно представить как имеющие одинаковую структуру, описываемую тремя параметрами: тип единицы отбора, способ группировки, способ отбора. Эти три параметра характеризуют тип выборки на данной ступени. Выборки внутри данного типа различаются числом отбираемых на каждой ступени единиц. Последняя ступень отличается от предыдущих тем, что в качестве единиц отбора выступают единицы наблюдения. Поскольку единица отбора данной ступени содержит множество единиц отбора следующей ступени, единицы отбора часто называют гнездами, а выборку, при которой сначала производится отбор гнезд, а затем изучение всех единиц, входящих в гнездо, называют гнездовой. Например, гнездовой будет выборка выпускников школ города (скажем, при изучении престижа профессий [Черноволенко, Оссовский, Паниотто, 1979, с. 184—186]), при которой отбираются случайным образом школы города, а затем в отобранных школах производят опрос всех школьников выпускных классов. Такую выборку можно описать как двухступенчатую следующим образом: на 1-й ступени единицы отбора — школа, группировка — отсутствует, отбор — случайный; на 2-й ступени единица отбора (и единица наблюдения) — выпускник школы, группировка — отсутствует, отбор (и опрос) — сплошной. Если к этому добавить объем генеральной совокупности, общее число школ, число отбираемых школ и объем выборки, то теоретическое описание выборки будет исчерпывающим. Практическое описание выборки требует указания на организационные процедуры, обеспечивающие построение выборки (например, где получен перечень школ, как осуществляется отбор, производилась ли выборка лиц, отсутствующих в школе в момент опроса и т. п.). Приведенный пример показывает также, что некоторые из параметров типа выборки на той или иной ступени могут отсутствовать.

Иногда перед тем, как производить отбор, единицы отбора объединяют в группы (слои, страты), а затем отбор производят отдельно из каждой группы. Объединение про-

водится так, чтобы в одну группу попадали однородные единицы — близкие по ряду признаков, интересующих следователя. Как следует из формулы (3.1), в этом случае для достижения той же точности объем выборки может быть существенно меньше. Например, для бесконечной нормальной совокупности при полной максимальной неоднородности по доле признака ($v=0,5$ и $1-v=0,5$), объем выборки почти в три раза больше, чем при однородной совокупности, на 90% состоящей из лиц, имеющих данное значение признака ($v=0,9$ и $1-v=0,1$). Иногда такое расслоение проводят из организационных соображений. Известно, что в одном из слоев выборку удобно организовать по одной схеме, а в других слоях — по другой (часто, например, для городского и сельского населения при исследовании регионов выборки строятся по-разному). Выборка, при которой единицы отбора разбиты на группы, а отбор проводится для каждой группы отдельно случайным методом, называется расслоенной (стратифицированной, районированной).

Способы группировки (способы выделения слоев) можно классифицировать по-разному. Например, можно выделить «естественную» группировку (уже существующую, например, городское и сельское население, студенты данного курса, рассматриваемые как один слой и т. п.) и «искусственную», полученную с помощью специальных методов разбиения объектов на группы, внутри которых объекты близки между собой по ряду признаков, интересующих следователя, а объекты из разных групп существенно отличаются по этим признакам (например, группировка с помощью методов кластерного анализа или классификации)³. Математические методы группировки можно также классифицировать по используемому алгоритму классификации. Но более важно при описании способа группировки указать критерий объединения единиц — по каким признакам и насколько близки между собой единицы одной группы и насколько отличаются единицы разных групп.

Последний из параметров, характеризующих каждую ступень выборки, — это способ отбора. Рассмотрим д

³ Воронов Ю. П. Районированные выборки в социологии. — Новосибирск: Наука, 1970; Айвазян С. А., Бежаева З. И., Староверов О. Классификация многомерных наблюдений. — М.: Статистика, 1979; Дюрин Б., Оделл П. Кластерный анализ. — М.: Статистика, 1979; Многомерный статистический анализ в социально-экономических исследованиях. — М.: Наука, 1974; Применение факторного и классификационного анализа для типологии социальных явлений. — Новосибирск: ИЭ и ОПП СО АН СССР, 1976; Распознавание образов в социальных исследованиях. Новосибирск: Наука, 1968.

случая. Первый — «вырожденный», когда отсутствует группировка единиц отбора данной ступени. Для этого случая наиболее распространенными способами отбора являются сплошной, случайный и квотный. Сплошной и случайный отборы мы рассматривали ранее. Разновидностью случайного отбора можно считать систематический (то есть отбор с определенным шагом упорядоченных некоторым образом объектов), поскольку, если порядок расположения объектов не связан с исследуемыми признаками, то систематический отбор моделирует случайный.

Квотный отбор возможен в том случае, если известно распределение некоторых признаков генеральной совокупности, которые связаны с признаками, интересующими исследователя (как правило, это социально-демографические характеристики, в частности, пол, возраст, образование и др.). По распределениям этих признаков рассчитываются квоты — определяют долю лиц с данным сочетанием признаков (например, мужчины возраста 20—25 лет с высшим образованием составляют 1% генеральной совокупности, женщины возраста 40—45 лет со средним специальным образованием — 3% и т. д.). После расчета объема выборки определяют, сколько лиц с данным сочетанием признаков требуется опросить, чтобы структура выборки совпала со структурой генеральной совокупности (если объем выборки — 1000, то из них следует опросить 10 мужчин возраста 20—25 лет с высшим образованием, 30 женщин возраста 40—45 лет со средним специальным образованием и т. д.). Затем интервьюеры производят целенаправленный поиск и опрос лиц, удовлетворяющих требованиям по данным квотам.

Второй случай — «невыврожденный», когда единицы группируются и из каждой группы (слоя) производится отбор отдельно. В этом случае тоже применимы указанные способы отбора, но задача существенно усложняется, возникает ряд новых задач. Например, при случайном отборе из каждого слоя отбор может производиться таким образом, чтобы число единиц, отбираемых из данного слоя, было пропорционально общему числу единиц в слое. При тех же условиях отбор можно осуществлять пропорционально числу единиц наблюдения (респондентов) в слое. При небольших выборках из каждого слоя может отбираться одна единица с характеристиками, близкими к средним для слоя, и т. д. В этом случае возникает задача оптимизации размещения выборки [Территориальная выборка..., 1980, с. 72—83], которая решается разными методами.

Предложить классификацию способов отбора единиц для данного случая затруднительно.

Рассмотрим в качестве примера формирование выборки для размещения сети интервьюеров Гостелерадио СССР [Рукавишников, Паниотто, Чурилов, 1984, гл. 5]. Цель формирования постоянно действующей сети интервьюеров — изучение восприятия передач радио и телевидения, генеральная совокупность — население Украинской ССР старше 7 лет. Исходя из требуемой глубины анализа объем выборки был определен в 2 000 респондентов. Организационные возможности позволяли создать сеть приблизительно из 400 интервьюеров. Интервьюеров следовало набирать из жителей населенных пунктов, в которых проводится опрос, число интервью, приходящихся на одного интервьюера, от 3 до 7 — вот основные требования, которые предъявлялись к выборке. Нужно было определить, в каких населенных пунктах и в каком количестве должны быть размещены интервьюеры, сколько респондентов должен опросить каждый из них, каким образом интервьюер должен отбирать респондентов для опроса.

На первой ступени выборки единицами отбора были области УССР. Всего в республике 25 областей. С помощью методов факторного и кластерного анализа они были разбиты на 9 групп, в каждую из которых входили области, близкие по признакам, которые могут оказывать влияние на восприятие передач средств массовой информации. Из теоретических соображений было выделено пять групп таких признаков: личностные характеристики (пол, возраст, образование, национальность и т. д.), характеристики социального окружения (прежде всего, национальные и культурные особенности населения области), уровень благосостояния для области в целом (обеспеченность жильем, бытовыми услугами и т. п.), степень доступности культурных ценностей в данной области (обеспеченность киноустановками, театрами, музеями, телевизорами и т. п.), уровень промышленного развития области. К сожалению, существующая статистика не даст полной информации по теоретически требуемым признакам, однако по каждой из областей было получено 44 показателя, в значительной степени описывающих указанные группы признаков, в пространстве этих показателей и проводилась классификация областей. Далее из каждой группы (слоя) отбиралась одна область с характеристиками наиболее близкими к средним по группе. Таким образом на 1-й ступени было отобрано 9 областей, каждая из которых была типичным предста-

вителем группы областей (например, в первую группу входила лишь Киевская область — естественно, она попала в выборку; во вторую группу входили Ворошиловградская, Донецкая и Харьковская области — в выборку была отобрана Донецкая область и т. д.).

На 2-й ступени выборки в каждой из 9 областей, отобранных на 1-й ступени, аналогичным образом производился отбор районов (всего из 187 районов 9 областей была отобрана одна треть).

На 3-й ступени отбор осуществлялся отдельно для сельского и городского населения. Для сельского населения тип единицы отбора — населенный пункт; способ группировки — выделение 5 слоев по численности для всех населенных пунктов каждого района, отобранного на предыдущей ступени; способ отбора из каждого слоя — случайный. Для городского населения отбор был аналогичный, но несколько иной была схема выделения слоев (все крупные города включались в выборку, отбор производился лишь для мелких городов и поселков городского типа). Таким образом были отобраны населенные пункты для размещения интервьюеров. Число интервьюеров определялось числом респондентов, которые должны быть опрошены в данном населенном пункте, а число респондентов вычислялось следующим образом: если в некоторой группе областей проживает 10% населения республики, то в области, отобранной в выборку, опрашивается 10% объема выборки, то есть для выборки, равной 2 000 — 200 человек. Аналогично вычисляется сколько респондентов должно быть опрошено в районах и населенных пунктах, попавших в выборку.

Наконец, на последней, 4-й ступени выборки — производится отбор и опрос респондентов. Группировка отсутствует, способ отбора — квотный. Квота рассчитывается по каждой из 9 групп областей отдельно для сельского и городского населения по полу, возрасту и образованию. Каждый интервьюер получает задание на опрос лиц с заданными полом, возрастом и образованием. На данной ступени возможен также отбор случайный (по избирательным спискам). Для этого интервьюерам, размещенным в отобранных населенных пунктах, центральная исследовательская группа рассылает специальную инструкцию.

Завершая описание многоступенчатых выборок, рассмотрим некоторые типичные вопросы, возникающие у социологов при определении объема выборки и способов отбора респондентов.

В статистической литературе приводятся методы, позволяющие определить, каким должен быть объем выборки,

если требуется с заданной исследователем доверительной вероятностью и точностью оценить среднее арифметическое, долю признака, коэффициент корреляции между двумя признаками и другие показатели. При этом получается, что для каждого показателя необходим свой объем выборки, но в реальных исследованиях бывает нужно оценить все эти показатели. Что делать в этом случае? Теоретически следует брать максимальный из объемов выборки, рассчитанных для каждого показателя, но практически это очень неудобно. Во-первых, нужно большое количество расчетов. Во-вторых, эти расчеты опираются на дисперсию изучаемого показателя в генеральной совокупности, а эта дисперсия, как правило, неизвестна. В данной ситуации, за исключением некоторых специальных случаев, целесообразно ориентироваться на долю признака (на процент), то есть расчеты производить по формуле (3.2), а при больших генеральных совокупностях по формуле (3.3) или по таблице, приведенной на с. 82 с учетом максимальной размерности признаков и глубины анализа информации.

Другой вопрос, который возникает, связан с тем, что выборка, репрезентативная для одних признаков анкеты, не обязательно будет репрезентативной для других признаков. Однако если взять максимально возможную неоднородность, как это сделано при переходе от формулы (3.1) к формуле (3.2), то есть положить $v(1-v)=0,25$, то случайная выборка будет репрезентативной для любого признака анкеты.

Гораздо серьезнее другая проблема. Дело в том, что теория, позволяющая определить объем выборки при заданной доверительной вероятности и точности, разработана лишь для случайной одноступенчатой выборки. Для более сложных случайных выборок, в частности, для гнездовой выборки, при которой сначала производится случайный отбор гнезд, а затем сплошное изучение всех элементов гнезда или опять случайная выборка элементов каждого гнезда, а также для расслоенной выборки при условии случайного отбора из каждого слоя, можно решить обратную задачу: зная объем выборки, определить ее ошибку⁴. Это хотя и не дает возможности определить заранее объем выборки, но позволяет оценить ошибку выборки после проведения исследования. Кроме того, теоретическое исследование подсказывает, какая стратегия проведения исследования предпочтительней. Так, гнездовой отбор дает большую

⁴ Статистический анализ случайных выборок см.: [Кокрен, 1976].

точность при большой неоднородности элементов внутри гнезд и большом сходстве гнезд между собой, причем, более мелкие гнезда дают большую точность, чем более крупные. При расслоенном отборе, наоборот, большая точность достигается при большой однородности единиц внутри слоя и большом различии между слоями. Так или иначе для случайных выборок ситуация контролируется исследователем. Объем выборки устанавливается более или менее произвольно (обычно определяют, каков был бы объем выборки, если бы она была одноступенчатой, а затем, поскольку гнездовой или расслоенный отбор может быть менее эффективным, увеличивают выборку в 1,3—1,5 раза), но после сбора информации ошибка выборки поддастся оценке, так как становятся известными дисперсии признаков.

Но что делать с многоступенчатыми выборками других типов, для которых не разработана соответствующая статистическая теория? В этом случае объем выборки устанавливается приблизительно так же, как и для случайных выборок. Что же касается контроля ошибки после сбора информации, то обычно в анкету включают признаки, по которым есть данные о генеральной совокупности (как правило, это социально-демографические признаки респондентов, отраженные в государственной или ведомственной статистике, а в региональных исследованиях — это данные, полученные в переписях населения). После проведения исследования данные, полученные по выборке для этих признаков, сопоставляются с данными генеральной совокупности и определяется максимальная ошибка Δ . Если в исследовании не допущены какие-либо систематические ошибки, можно предположить, что поскольку ошибка по ряду признаков не превышает Δ , следовательно, и по другим признакам анкеты она будет не выше Δ .

Наконец, последнее замечание. Любая выборка является компромиссом между желательной точностью и реальными ограничениями кадровых, временных и финансовых ресурсов. Стремясь увеличить объем выборки, социологи иногда идут на некоторые упрощения процедуры опроса. Например, в городских опросах интервьюерам понижают норму числа повторных посещений, которые они должны совершить, если не застанут респондента, выбранного для опроса, или наряду с основным вручают резервный список адресов, по которым можно производить опрос, если не застали респондента из основного списка и т. п. При этом в данные опроса вносится неконтролируемая систематическая ошибка. В то же время при тех же ресурсах может

быть получена меньшая по объему выборка, но уже без систематической ошибки. Я полагаю, что вторая альтернатива намного предпочтительней, так как уменьшение объема выборки приведет лишь к тому, что исследователь большие различия будет интерпретировать как значимые (в самом худшем случае он не получит различий там, где они есть), а при увеличении выборки за счет систематической ошибки может получить артефакты — значимые различия, когда их в действительности нет.)

3. Нестатистическое понимание репрезентативности

Существуют отдельные виды социологических исследований, а также ситуации в обычных исследованиях, в том числе в массовых опросах, когда статистическое понимание репрезентативности вообще не имеет смысла.

Начнем с примера. Предположим, что в некотором производственном объединении социологическая лаборатория исследовала психологический климат во всех бригадах одного из предприятий для выявления факторов, определяющих психологический климат, и для разработки мероприятий по его улучшению. С этой целью проведен, в частности, социометрический опрос и рассчитаны характеристики сплоченности и другие социометрические индексы, описывающие структуру межличностных отношений в каждой бригаде. Требуется оценить допущенную в исследовании ошибку репрезентативности.

Эту выборку можно рассматривать как двухступенчатую со сплошным отбором бригад на первой ступени и лиц, подлежащих опросу, на второй ступени. Поскольку на первой ступени отбор сплошной, ошибка репрезентативности нулевая. На второй ступени исследователи тоже пытались провести сплошной опрос, но это не удалось из-за болезни респондентов, поездок в колхоз, командировок или по другим причинам. Предположим, что всего нужно было опросить 40 бригад, в каждой бригаде работало по 30 человек, то есть всего 1200 человек, а удалось опросить по 25 человек, то есть — 1000. Если даже считать, что причины, по которым не опрошены 5 человек в каждой бригаде, никак не связаны с целью исследования, случайны по отношению к исследуемым характеристикам и, следовательно, процедура моделирует случайный отбор, то тем не менее будет некорректно рассчитывать ошибку как ошибку случайного отбора 1000 из 1200 респондентов (по формуле (3.2) ошибка равна 1,3%). Точнее, это будет корректно лишь для признаков, описывающих характеристики отдель-

ных индивидов — пол, возраст, специальность и т. д.⁵. Что же касается интересующей исследователей структуры межличностных отношений и ее характеристик, то статистический подход мало что даст для оценки ошибки выборки.

Действительно, пусть в качестве меры сплоченности группы используется отношение числа взаимных выборов, сделанных по некоторому критерию, к числу всех выборов в группе [Волков, 1970]. Пусть выборы, сделанные членами данной группы, варьируют от 0 до 8, среднее число выборов равно 2 и каждый третий выбор — взаимен. Тогда индекс сплоченности будет равен $20/60=0,33$. Если в число 25 опрошенных не попал лидер группы, получивший 10 выборов (причем сам он сделал бы 8 выборов, если бы был опрошен), то поскольку от распределения его ответов будет зависеть 40% числа взаимных выборов в группе, ошибка выборки может составить 40% и свести к нулю ценность информации по данной группе. А при случайном отборе вероятность отсутствия лидера среди опрошенных равна $5/30$, или $1/6$ и, значит, данные о сплоченности в $1/6$ (17%) бригад совершенно ненадежны. Таким образом, ошибка распределения бригад по сплоченности при сделанных допущениях равна не менее 17% (а не 1,3%). И это при учете отсутствия ответов лишь одного (а не 5 человек). Если же учесть отсутствие ответов 5 человек, то используя лишь статистические соображения, мы получим, что ошибка репрезентативности настолько велика, что проводить исследование бессмысленно, а полученная информация не представляет никакой ценности.

Более плодотворным в данном случае является определение репрезентативности как степени отсутствия ошибок, связанных с отбором респондентов, но основанное не на статистических, а на содержательных соображениях, связанных с характером изучаемой информации. Очевидно, что если отсутствие высокостатусных членов группы может существенно исказить картину межличностных отношений, то отсутствие изолированных, нейтральных, низкостатусных членов группы, или с небольшим стажем пребывания в ней не окажет существенного влияния на характеристики группы, полученные в результате опроса (для большинства изучаемых характеристик). Поэтому в данном случае репрезентативность может контролироваться вычислением

⁵ Я уже не говорю о том, что и допущение случайности отбора (точнее, самоотбора) в данном случае вряд ли корректно: поездки в колхоз распределяются неравномерно в зависимости от пола, возраста и должности, командировки — в зависимости от пола и специальности и т. п.

статуса отсутствующих членов группы по ответам присутствующих или введением специальных вопросов, с тем чтобы опрошенные дали оценку включенности отсутствующих в структуру межличностных отношений. В принципе, для каждого из социометрических показателей могут быть разработаны методы оценки смещений, возникающих за счет неполного охвата группы. Такая оценка позволит решить, репрезентативны ли полученные данные или требуется дополнительный опрос.

Другой пример нестатистического подхода к пониманию репрезентативности — отбор экспертов при проведении экспертного опроса. Репрезентативность как степень отсутствия ошибок отбора означает в данном случае правильный, лишенный ошибок отбор лиц, наиболее компетентных в области, содержащей проблему, подвергаемую экспертизе. Предположим, что мы умеем измерять степень компетентности экспертов в 10-балльной шкале, составили перечень всех экспертов и построили их распределение по степени компетентности: столько-то экспертов компетентны на 1 балл, столько-то — на 2, столько-то — на 10. Возникает вопрос — что здесь генеральная совокупность? Все эксперты? Но ведь включение малокомпетентных экспертов может внести шум в информацию и дать менее надежный результат, чем в случае опроса более компетентных лиц. Возможно, под генеральной совокупностью имел бы смысл понимать такое множество экспертов, опрос которых даст результат наиболее близкий к истине. Но поскольку истина нам неизвестна (иначе зачем нужно было бы проводить экспертный опрос), определить генеральную совокупность мы не в состоянии. Естественно, что в этом случае невозможно определить и выборочную совокупность и неприменимы рассмотренные в предыдущих параграфах способы расчета ошибки выборки.

Реально отбор проводится следующим образом. Исследователь задастся некоторым уровнем компетентности (зависящим от использованного им способа определения компетентности экспертов) и спрашивает всех экспертов, имеющих компетентность выше заданного уровня. Судя по всему, наиболее обоснованным методом отбора экспертов является метод «снежного кома», построенный на экспертных межличностных оценках (как единственный метод заслуживающий внимания, его рассматривают, например Л. А. Панкова, А. М. Петровский, М. В. Шнейдерман [1984, с. 8]). Метод заключается в том, что исследователь, проводящий экспертизу, заранее знает хотя бы несколько лиц, которые определенно являются специалистами в об-

ласти исследуемой проблемы. Он обращается к ним с просьбой указать, кто может быть экспертом по изучаемому вопросу, затем обращается к названным лицам и просит их в свою очередь указать экспертов в данной области и т. д. Если бы каждый раз в ответах появлялись только новые фамилии, то число экспертов росло бы как снежный ком (при наличии на первом шаге 2-х экспертов и при 10 ответах в среднем — 200 фамилий на 3-м шаге, 2000 — на 4-м, 20 000 — на 5-м и т. д.). Однако поскольку реально круг экспертов ограничен, доля новых фамилий, появляющихся на каждом следующем шаге, быстро падает и становится близкой к нулю или равной нулю. Исследователь затем опрашивает либо всех лиц, попавших в сформированный таким образом список, либо учитывает частоту встречаемости экспертов в ответах других и, используя эту частоту как показатель компетентности, «обрезает» список ниже определенного порога и опрашивает только по верхней части списка. После выявления первоначального круга лиц, названных экспертами, можно использовать также другие процедуры определения их компетентности и вычисления порогового уровня [там же, с. 6—26; Бешелев, Гурвич, 1980, с. 96—100].

Эта процедура весьма далека от рассмотренных в предыдущих параграфах. Более того, неприменимость понятий генеральная (следовательно, и выборочная) совокупность заставляет задуматься над вопросом, отражается ли данная ситуация понятием репрезентативности. Хотя считается, что отбор экспертов осуществляется с помощью разновидности выборочного метода [Рабочая книга социолога, 1983, с. 201], а качество выборки описывается понятием репрезентативности, возможен и другой, более адекватный данному случаю подход. Отбор экспертов в отличие от отбора респондентов направлен не на то, чтобы представить реально отразить распределение мнений некоторой «генеральной» группы экспертов, а на то, чтобы получить оценку каких-то аспектов объективной реальности (прогноз какого-то события, актуальность проблемы, качество продукции и т. п.) Само по себе распределение мнений экспертов (в отличие от мнений респондентов) нас не интересует, это лишь промежуточный результат, средство для выявления или оценки некоторой объективной реальности, существующей независимо от экспертов. Эксперты оценивают одну и ту же величину, и разброс их оценок связан с различиями в их компетентности. В этом смысле эксперты выступают в качестве своего рода средств (приборов) для измерения этой величины. Если нужно точно определить

время, а у нас дома несколько часов, то мы их используем аналогично тому, как «используем» экспертов: одни вообще не принимаем во внимание — они давно стоят (никак не купим батарейку), на других пять минут девятого, но они всегда спешат, на третьих одна минута девятого, но они обычно отстают, следовательно, сейчас две или три минуты девятого.

Эксперты, таким образом, являются не элементами объекта (как респонденты), а элементами средств исследования, отбор экспертов можно рассматривать как выбор методики, «инструмента» исследования, а ошибки выбора — как теоретические ошибки. Отсутствие такого рода ошибок (рис. 1.2) в информации определялось как обоснованность, поэтому проблемы отбора экспертов являются проблемами обеспечения обоснованности, а не репрезентативности информации.

Теперь очевидно, что для оценки качества отбора экспертов требуются специальные показатели. Такие показатели, насколько мне известно, не разработаны. Для их разработки нужны специальные эксперименты, моделирующие процедуру экспертного опроса, направленные на определение оптимального порогового уровня компетентности. Можно предположить, что функция зависимости качества экспертизы от суммы показателей компетентности (предположим, что эксперты проранжированы по убыванию степени компетентности и найдены накопленные суммы путем прибавления по одному показателю компетентности каждого следующего эксперта) имеет максимум: сначала качество растет, так как увеличение числа экспертов повышает качество прогноза, а затем, начиная с некоторого момента, падает, так как включение в опрос малокомпетентных экспертов ухудшает качество прогноза. Положение конкретной экспертизы на этом графике и было бы показателем степени обоснованности процедуры отбора экспертов в данной экспертизе.

Метод «снежного кома» используется, однако, не только при отборе экспертов, но и при отборе представителей редких групп респондентов. Если, например, требуется опросить коллекционеров грампластинок, мигрантов из мест, не являющихся типичными для миграции жителей данного населенного пункта, представителей редких национальностей, верующих, принадлежащих к той или иной религиозной секте и т. п. Это несомненно проблема репрезентативности, хотя формирование выборки здесь весьма специфично. Генеральную совокупность здесь тоже определить практически невозможно, хотя в отличие от предыдущего

случая она существует. Кроме того, в данном случае не выполняются исходные предпосылки для формирования случайной выборки: для случайного отбора необходимо, чтобы события отбора респондентов не были связаны между собой — в данном же случае сам принцип отбора основывается на связи респондентов, иначе метод «снежного кома» не будет «работать». Данный случай требует специальных показателей, характеризующих качество отбора.

Есть некоторые особенности и в определении репрезентативности выборки при проведении контент-анализа тестов [Бухалов, Чурилов, 1980, с. 182], но в целом описание такой выборки соответствует изложенному в предыдущем параграфе.

Приведу еще один пример, показывающий, что проблема отбора объектов в социологических исследованиях не описывается исчерпывающе статистической теорией репрезентативности. Уже более двадцати лет различными исследователями в разных регионах страны проводятся исследования престижа и привлекательности профессий для школьников [Водзинская, 1969; Константиновский, 1977; Матуленис, 1983; Титма, 1975; Черноволенко, Оссовский, Паниотто, 1979; Шубкин, 1970; и др.]. Все эти исследования включают оценку респондентами профессий из некоторого составленного исследователями списка, и качество результатов исследования во многом зависит от того, какие профессии отобраны в список.

Как оценить качество отбора профессий в данном случае? Если считать, что профессии списка являются выборкой из генеральной совокупности всех профессий (по некоторым оценкам их около 40 тысяч), то во всех названных исследованиях, включавших от нескольких до восьми десятков профессий, списки профессий нерепрезентативны со статистической точки зрения. И здесь, как при отборе экспертов, нужно говорить об обоснованности, а не о репрезентативности отбора. Ведь список профессий является элементом инструмента, а не объекта исследования. Но вопрос о том, к какому классу ошибок — ошибкам репрезентативности или обоснованности — отнести ошибки отбора профессий, интересует нас не сам по себе, а для того, чтобы оценить и уменьшить эти ошибки до требуемой величины. А эта оценка существенно зависит от цели исследования.

Пусть, например, целью является изучение престижа профессий для выяснения влияния престижа на привлекательность и выбор профессий. В этом случае ситуация становится более определенной.

Во-первых, поскольку изучаются школьники, а у каждого школьника формируются представления не обо всех профессиях, а только об ограниченном их круге, то при статистическом подходе генеральная совокупность — это профессии, входящие в круг представлений школьников данного региона. Их уже может быть намного меньше, например, 1 000. Со статистической точки зрения это нам мало помогает, так как выборка 40 профессий из 1 000 не намного репрезентативней, чем из 40 000 (по формуле (3.2) получается, что при генеральной совокупности, равной 40 000, ошибка по доле профессий, обладающих той или иной характеристикой, с вероятностью 0,954 не превышает 15,8%, а при уменьшении генеральной совокупности до 1 000 — не превышает 15,3%). Можно, однако, считать, что генеральная совокупность — это представления о профессиях, образы профессий в сознании школьников, а число единиц генеральной совокупности, связанных с данной профессией, равно числу школьников, имеющих образ этой профессии. Если, например, мы проводим городской опрос школьников 10-х классов (предположим, что всего их 50 тысяч) и каждый второй имеет образ профессии повара (реальный или иллюзорный), то образ данной профессии в генеральную совокупность и в выборку (в список профессий) попадает как 25 тысяч условных единиц.

Во-вторых, следует учесть, что близкие по характеру и содержанию труда, а также по оплате и другим признакам профессии имеют сходные оценки престижа и привлекательности, поэтому каждая профессия является представителем некоторой группы профессий. Возможно, что после некоторой предварительной работы, состоящей в исследовании близости представлений школьников о тех или иных профессиях и в выделении типов профессий по этим представлениям дальнейшая работа по созданию обоснованного списка могла бы строиться на статистическом подходе аналогичном формированию многоступенчатых выборок. Кроме того, объектом оценки престижа может быть не профессия, а социальная или социально-профессиональная группа, а профессия является для респондента лишь индикатором этой группы и средством вхождения в нее [Матусевич, Оссовский, 1982; Матусевич, Паниотто, 1978].

Как видим, реальные проблемы, связанные с обеспечением высокого качества отбора единиц наблюдения в социологическом исследовании, могут существенно отличаться от проблем, рассматриваемых в статистической литературе. Причем иногда эти проблемы сводятся к обеспечению обоснованности, а не репрезентативности информации. По

этому требуется дальнейшая работа по созданию показателей, характеризующих качество отбора единиц наблюдения в социологических исследованиях. Хочу обратить внимание на то, что в описанных примерах сложности возникали из-за того, что единицы наблюдения включали некоторые характеристики, входящие в сами условия отбора (степень включенности респондента в структуру межличностных отношений в группе, степень компетентности эксперта, степень распространенности образа профессии в сознании школьников). В этом, по-видимому, проявляется определенная закономерность, которую можно использовать при создании более общей схемы описания выборки в социологических исследованиях.

И =

△

IV. ОБОСНОВАННОСТЬ СОЦИОЛОГИЧЕСКОЙ ИНФОРМАЦИИ

Мадам, Ваши претензии необоснованны.

Из разговора

Купить всех знакомых за столько, сколько они в действительности стоят, а продать за столько, во сколько они оценивают себя сами, — вот бешеные деньги.

Ю. Тувим

В первой главе обоснованность, или валидность¹ информации определяется как степень отсутствия в информации искажений, вытекающих из теоретических, методологических ошибок — неверности исходных посылок при разработке методики исследования, несоответствия числовой модели изучаемой эмпирической системе и т. п. Именно так интерпретируется распространенное определение валидности как соответствия получаемой при измерении информации назначению методики («информация валидна, если исследователь измеряет именно то, что хотел измерить»).

Обоснованность представляется мне наиболее важной характеристикой надежности информации. Конечно, репрезентативность, правильность и устойчивость тоже являются важными компонентами надежности, но ошибки обоснованности в большей степени обесценивают информацию, делают ее трудноинтерпретируемой и менее пригодной для исследователя. Если, скажем, неустойчивую, но обоснованную методику можно спасти, незначительно меняя формулировки вопросов, укрупняя пункты шкалы, используя контрольные признаки и методы агрегирования данных (многое из этого можно осуществлять, анализируя уже собранную информацию без проведения нового исследования), то необоснованная методика методикой, собственно,

¹В монографии всюду я употребляю как синонимы то понятие «обоснованность», то — «валидность». С одной стороны, хочется, следуя Б. З. Докторову, Г. И. Саганенко и др., употреблять «обоснованность», с другой стороны, если уж речь идет о валидации, то было бы естественно, чтобы ее результатом была «валидность», а не «обоснованность». Кроме того, понятие «валидность» используется в психологической литературе. Чтобы не было «чересполосицы», в этой главе я буду пользоваться преимущественно понятием «валидность».

и не является, ее надо переделывать кардинально, косметические средства здесь не помогут².

Между тем, этой характеристике социологической информации в нашей литературе уделяется меньше внимания, чем другим характеристикам надежности. В социологических публикациях практически отсутствуют сведения о валидности использованной методики (и другим характеристикам, исключая, репрезентативность, уделяется не намного больше внимания). Даже в специальных работах по надежности информации разделы, посвященные валидности, выглядят скромнее других разделов и по объему и по степени разработанности, и по аргументации. Поэтому я рассматриваю обоснованность более подробно, чем были рассмотрены другие характеристики качества информации.

1. Типы валидности

Контроль ошибок обоснованности может осуществляться разными способами. В литературе сам процесс определения обоснованности методики называется проверкой валидности, валидацией или валидизацией (иногда валидизацией называют процесс создания обоснованной методики, например, исследование, направленное на отбор суждений, имеющих высокую корреляцию с реальным поведением респондентов). В зависимости от способа валидизации, назначения методики и т. п. выделяют типы валидности: валидность по критерию, или эмпирическая (в том числе — предсказательная и актуальная), содержательная, конструктивная, дифференциальная валидность и др.³

Валидность по критерию, или эмпирическая валидность, заключается в соответствии информации, полученной с помощью данной методики, и информации, полученной с помощью методики, которая является более обоснованной и называется критерием валидности (то есть валидность по критерию — это валидность, обоснованная контролем, внешним по отношению к процедуре и инструменту исследования). В свою очередь, она делится на типы в зависимости от того, какой критерий валидности используется.

Предсказательная (predictive), или прогностическая валидность — как разновидность эмпирической — заключа-

² Напомню, что обоснованной, или валидной, является методика, которая позволяет получить обоснованную (или валидную) информацию. Аналогично определяется надежная, репрезентативная, правильная и устойчивая методика.

³ При изложении типов валидности я ориентируюсь преимущественно на работы: [Аванесов, 1982; Процесс социального исследования, 1975; Standarts..., 1974; Dick, Hegerty, 1971].

ется в пригодности информации как основания для прогноза поведения респондентов⁴. Дальнейшая классификация этого вида валидности может осуществляться по периоду времени, в течение которого осуществляется прогноз. Прогностическая валидность определяется путем соотнесения данных, полученных с помощью валидируемой методики, и данных, полученных с помощью методики-критерия через определенный период времени (это и есть интервал прогнозирования, его можно было бы назвать прогностической мощностью методики).

При определенных условиях оценка актуальной валидности может экстраполироваться на предсказательную валидность методики. Достаточно очевидно, однако, что, эти понятия существенно различаются и актуальная валидность не может использоваться взамен прогностической без соответствующего логического (а при возможности — и эмпирического) обоснования.

а Актуальная (concurrent) валидность — другая разновидность эмпирической валидности — состоит в способности методики измерять актуальный уровень характеристики индивида, значение этой характеристики в данный фиксированный момент времени. Например, экзамены в школе являются своего рода тестом с довольно высокой актуальной валидностью для оценки уровня знаний учащегося в данный момент (предсказательная валидность при этом может быть низкой — бывает, что на следующее утро ученик уже мало что помнит из пройденного курса). Другой пример — тест для определения психических отклонений, который валидируется путем сравнения с диагнозом психиатров. Актуальная валидность определяется путем соотнесения данных, полученных с помощью валидируемой методики, с данными, полученными с помощью методики-критерия в тот же период времени.

В. С. Аванесов в качестве третьей разновидности эмпирической валидности выделяет *валидность по известным группам* (еще одно название — *валидность по контрастным группам*) [Аванесов, 1931, с. 141; см. также: Ядов, 1972, с. 84]. Суть метода заключается в том, что каким-либо

⁴ Не хочу усложнять изложение, пытаюсь давать определения, которые годились бы для всех методов сбора данных. Условно речь идет об информации, полученной с помощью метода опроса. Все сказанное, однако, легко применить и для других методов. Если, например, речь идет о контент-анализе, то прогностическая валидность полученной с его помощью информации может означать его пригодность для прогноза поведения коммуникатора или социальной группы, установки которых отражены в анализируемом тексте [Баранов, 1973, с. 38, 39].

способом (например, по экспертным оценкам) выделяются две группы лиц, которые отличаются по измеряемому свойству, и проводится их опрос с помощью валидизируемой методики. Затем по результатам опроса выделяются такие вопросы (суждения, задания), на которые лица из известных групп отвечают по-разному. Отобранные вопросы и составляют инструментарий валидной методики. Мне представляется, что этот метод полностью попадает под определение актуальной валидности и выделять его в качестве самостоятельной разновидности эмпирической валидности нецелесообразно.

Рассмотрим пример использования метода известных групп. Для исследования особенностей функционирования и тенденций развития заводского сектора науки на ряде предприятий г. Киева было проведено социологическое исследование. Для исследования противоречий между заданными и реальными (выполняемыми) функциями работников, обусловленных двойственным положением заводского сектора науки (необходимость непосредственно участвовать в производстве) требовалось разработать методику измерения степени научного содержания труда. Из определенных теоретических соображений было выделено семь признаков, характеризующих труд как научный (исследовательский характер труда, его творческое содержание, использование теоретических знаний и т. п.), для каждого из них были разработаны эмпирические референты и проведен опрос респондентов (например: «Какую преимущественно работу Вы ведете?»; варианты ответа: «теоретическую», «экспериментальную», «экспериментально-теоретическую», «исследовательскую», «опытно-конструкторскую», «работу по внедрению новой техники», «проектно-конструкторскую», «работу по обслуживанию производственных процессов»). Каждому из ответов в соответствии с некоторыми теоретическими соображениями был присвоен балл и возникла задача объединения всех показателей в один индекс⁵ [Белосвет, Папиотто, 1976].

Сначала мы пытались вывести этот индекс путем экспертного опроса (эксперты оценивали вес каждого признака, а затем вес использовался для построения итогового индекса), но разброс экспертных оценок был настолько велик, что стала ясна ненадежность такого способа измерения интересующего нас показателя. Тогда мы приме-

⁵ Конечно, приписывание баллов вносит определенный произвол в процедуру измерения, но последующее «взвешивание» компенсирует внесенные ошибки: если какие-то баллы завышены, то весь этот признак получит более низкий вес, и наоборот.

нили метод известных групп. Предварительное наблюдение и консультации с экспертами показали, что основная нагрузка в выполнении научных исследований ложится на лиц, явно ориентированных на науку. В эту группу включили лиц, имеющих научную степень, и лиц, проявивших стремление ее получить (показатели: обучение в заочной аспирантуре, работа над диссертацией, сдачи кандидатских минимумов). Поэтому в дальнейшем мы сравнивали результаты ответа на анкету двух известных групп: лиц, явно ориентированных на науку, и всех остальных. Была найдена средняя разность баллов для первой и второй группы по ответам на вопросы анкеты о характере выполняемого труда по каждой из 7 составляющих. Оказалось, например, что разность средних баллов для «получения знаний» равна 0,23, для «потребности в теоретических знаниях» — 0,07, а для «использования существующих научных знаний» — 0. Полученные показатели нормировались и использовались как веса при построении индекса научности труда:

$$y_i = a_1 x_{i1} + a_2 x_{i2} + \dots + a_7 x_{i7}, \quad (4.1)$$

где y_i — индекс степени научности труда i -го респондента x_{ij} — балл каждого из 7 признаков степени научности труда, полученный i -м респондентом ($j=1, 2, \dots, 7$), a_j — вес, равный нормированной разности средних баллов, полученных каждой из двух «известных» групп.

На этом примере хорошо видно, что построенная шкала (способ ее валидации представляет одну из распространенных модификаций теста по известным группам) имеет актуальную валидность; она дает результат, который может быть получен с помощью метода-критерия в тот же период времени. Интересно отметить, что экспертный опрос в этом случае (а это типичная процедура для оценки суждений, широко используемая при построении шкал [см., например: Зайцева, 1970, с. 223—226]) дал ненадежные и плохоинтерпретируемые результаты.

Содержательная (content) валидность, или *валидность по содержанию*, характеризует, насколько ситуации (суждения, задания), содержащиеся в методике, репрезентативны для всех ситуаций, которые должен описывать тест. Хотя это достаточно общая формулировка, но фактически речь идет лишь о тестах, направленных на определение некоторых знаний и умений (типа экзаменов). Например, нужно определить, насколько материал контрольной работы отражает программу, содержание курса. Обычно при рассмотрении этого вида валидности формулируют пере-

чень тем и подсчитывают, сколько суждений теста приходится на каждую тему, равномерно ли они распределены и т. п.

Конструктивная (construct), или концептуальная валидность состоит в соответствии измеряемого с помощью данной методики признака теоретической концепции исследователя, введенному им понятию, конструкту. Другое название — *понятийная обоснованность, или обоснованность свойствами*. Доказательство концептуальной валидности проходит в несколько этапов. Исследователь, исходя из теоретических соображений, вводит некоторое понятие, конструкт, который играет важную роль в его теории, например, тесно связано с поведением респондента в некотором классе ситуаций. Такими понятиями могут быть, например, «степень интеллектуальности», «интроверсия», «тревожность», и т. п. Некоторые из них могут иметь аналоги в обыденном языке (интеллектуальность, тревожность), но абстрагируясь от их многозначности, исследователь вкладывает в них свое, более определенное значение. Смысл введения этих понятий заключается в их конструктивной роли при объяснении тех или иных явлений. Оценка степени интеллектуальности позволяет предсказывать успехи в различных областях деятельности, обнаружение интро- или экстраверсии — поведение при общении и в ряде других ситуаций, тревожности — поведение в стрессовой ситуации и т. п. Исходя из содержания понятия и его роли в теории, проводится его экспликация, анализ, конкретизация, в нем выделяются определенные аспекты, оно определяется через другие понятия и так далее до тех пор, пока не будет построена структура, вычленяющая совокупность эмпирических референтов, индикаторов данного понятия. Теоретическое обоснование перехода от понятия к эмпирическим референтам и есть концептуальной валидизацией. Полученная с помощью построенной таким образом методики информация — концептуально валидна.

Однако процесс на этом не заканчивается. Теперь надо показать, что введенное понятие действительно полезно, что измеренный показатель, соответствующий данному понятию, позволяет прогнозировать поведение, что в результате исследований получается предсказываемая теорией корреляция с другими переменными и т. д. Фактически, речь идет о проверке определенного фрагмента теории. Если результат не совпадет с предсказанием теории, то это означает либо неверность теории либо неверность перехода от понятия к эмпирическим референтам. Доказательство концептуальной валидности нельзя получить в одном отдель-

ном исследовании, обычно требуются множество исследовательских результатов (иногда проведенных разными исследователями в течении многих лет). Совокупность ситуаций, в которых применение данной методики (теста) обоснованно, называется *областью валидности методики* (теста).

В качестве примера можно привести один из наиболее известных тестов — определение коэффициента интеллектуальности (КИ) Гансом Ю. Айзенком [1972], валидность которого исследуется уже несколько десятилетий, но до сих пор вызывает споры. Показательным является интервью Г. Айзенка «Литературной газете» [1975, 12 нояб.] и дискуссия по поводу этого интервью [там же, 1976, 4 февр.]. С точки зрения Айзенка, разработанный им тест измеряет врожденное свойство, — «скорость протекания умственных процессов», — не зависящее от уровня образования, с широко интерпретируемой валидностью: «...действительно ли наши тесты измеряют «ум» в обычном смысле этого слова? Отвечаю: да, это так. В основе тестов интеллектуального развития лежит фундаментальный научный факт, который доказан экспериментально. Попросту говоря, не пользуясь устрашающим научным жаргоном, у каждого из нас есть собственная мера способности решать интеллектуальные задачи без различия того, будут ли это задачи, выдвигаемые обыденной жизнью, профессией, или задачи, специально подобранные экспериментатором. Поэтому можно разработать задачи, которые измеряют такую способность» [там же, 1975, 12 нояб.]. «Недавно в глухой африканской деревне был проведен опыт: жителей проверяли с помощью тестов, а затем попросили их назвать «умных» односельчан. Совпадение этого списка с результатами тестирования было почти полным» [там же]. Этот пример приведен Г. Айзенком для доказательства аспекта валидности, связанного с независимостью результатов теста от уровня образования и степени совпадения его понятия интеллекта с житейским понятием ума. Требовалось доказать также прогностическую валидность теста — об этом тоже упоминается в интервью Г. Айзенка. Существует огромное количество фактов, свидетельствующих о том, что «коэффициент интеллектуальности» человека отчетливо (но неоднозначно) связан с успешностью его работы, с достижениями в различных областях жизни. Пожалуй, самый впечатляющий факт — знаменитое исследование одаренных детей. В течение 35 лет исследователи наблюдали за группой из 857 мальчиков и 671 девочки, имевших в детстве высокий «коэффициент интеллекта». Вы-

сокое значение коэффициента сохранялось у них на протяжении всех этих лет. Они великолепно учились, более 85% из них поступили в университеты, более 70% получили ученые степени. Им принадлежит большое количество научных работ, литературных произведений и т. д.

Несмотря на большое количество эмпирических данных, подтверждающих интерпретацию теста, предлагаемую Айзенком, были получены также и данные, противоречащие этой интерпретации. И в упомянутой дискуссии и в других работах [см., например: Блейхер, Бурлачук, 1978, с. 25—46; Психологическая диагностика . . . , 1981, с. 13; и др.] указывается на недопустимость интерпретации различных результатов тестирования, полученных для разных рас, как показателя отсталости той или иной расы в интеллектуальном отношении, на связь тестов интеллекта с уровнем образования и т. п. Все это говорит о том, что даже сейчас, спустя десятилетия после создания теста, область его валидности продолжает уточняться.

Рассмотренные типы валидности (предсказательная, актуальная, содержательная и концептуальная) упоминаются чаще, чем другие. В отдельных работах идет речь об *очевидной* (face), или *логической* валидности, или *представительности*, — когда валидность процедуры очевидна, (умение стенографии, машинописи, знание орфографии достаточно очевидны даже при небольших тестах и не требуют специального обоснования валидности); *дифференциальной* валидности, определяемой разностью между корреляциями классификационного теста с двумя отдельными критериями, которые нужно предсказать (если коэффициент корреляции тестовых оценок качеств респондента с нормативными показателями, необходимыми для работы садовником, равен — 0,80, а для работы механиком — равен 0,95, то дифференциальная валидность равна $0,95 - (-0,80) = 1,75$) и др.

Такое деление валидности на типы неудачно — нет исчерпывающего деления объема понятия валидности, смешаны разные уровни, частный случай — содержательная валидность — рассматривается рядоположенно с концептуальной валидностью.

Поэтому я попытаюсь предложить несколько иной подход к типологии валидности.

Предлагаемая типология. Прежде всего, целесообразно разделить валидность на *теоретическую* (конструктивную, концептуальную) и *эмпирическую* (валидность по критерию). С теоретической валидностью исследователь чаще всего имеет дело тогда, когда его методика предназначена

для измерения признака: соответствующего введенному им теоретическому понятию (такие признаки мы называли конструируемыми, см.: гл. 2). В этом случае валидность определяется теоретическим путем (анализ процесса конструирования методики, корректности редукции содержания понятия, адекватности эмпирических референтов и т. п.). Потом уже проводят эмпирические исследования для определения других видов валидности, например, предсказательной, но эти исследования проводят не для разработки методики, а для выяснения области валидности уже разработанной методики. При обнаружении невалидности исследователь возвращается к теории, к процессу операционализации понятия, а не использует полученный коэффициент корреляции для отбора вопросов или заданий методики. В этом отличие эмпирического этапа валидации теоретической валидности от валидности эмпирической. Для проверки теоретической валидности нужно получить пусть небольшие, но значимые коэффициенты корреляции по широкому классу ситуаций, предсказываемых теорией. Для достижения эмпирической валидности нужно получить высокие коэффициенты корреляции с критерием валидации, но для одной или нескольких ситуаций. Если валидность эмпирическую через измерение обеспечивают, то валидность теоретическую — только проверяют, то есть уточняют область валидности методики, границы интерпретации получаемых результатов.

Конечно, под приведенное описание подпадают «чистые» типы, и можно, наверное, придумать ситуацию валидации, лежащую между ними. Но так дело обстоит практически с любой типологией. В большинстве же случаев различие между теоретической и эмпирической валидностью можно провести очень четко.

Следующий шаг — деление теоретической и эмпирической валидности на типы. Например, рассмотренная ранее содержательная валидность и упомянутая очевидная валидность являются разновидностями теоретической валидности, так как в обоих случаях метод обоснования — теоретический, логический. Вопрос о классификации способов теоретического обоснования методик, о типологии теоретической валидности, кажется мне очень сложным; к сожалению, я не могу предложить ничего, что заслуживало бы внимания, и формулирую эту проблему как нерешенную. Что же касается эмпирической валидности (валидности по критерию), то здесь у меня есть конструктивное, как мне кажется, предложение.

Остановимся на этом вопросе более подробно.

Естественно валидность по критерию типологизировать в зависимости от типа критерия, используемого для валидации. Естественно также, что «логика валидности по критерию предполагает валидность самого критерия», как записано в американских стандартах на психологические тесты [Standarts . . . , 1974, p. 36]. Далее читаем: «Очень часто тесты валидизируются относительно доступного критерия вне соответствующего исследования самого критерия. Ценность исследования валидности по критерию зависит от пригодности и качества измерения по данному критерию» [Ibid.].

Но как определить валидность самого критерия валидации? Может быть, нужен какой-то другой критерий? Как полагают социологи из ГДР, «проверка критерия на обоснованность предполагает наличие критерия для критерия, и было бы логично далее доказывать обоснованность второго критерия по третьему, третьего по следующему, что привело бы к бесконечному процессу определения обоснованности, означаящему крах практических усилий обеспечить надежность процедуры» [Процесс социального исследования, 1975, с. 140].

Эта мрачная картина далека от практики социологических исследований не столько потому, что социологов мало волнует валидность самого критерия валидации, сколько потому, что социологам не требуется абсолютная надежность. Существует определенный уровень надежности, который полностью приемлем при проведении любых социологических исследований. Например, валидизацию вопроса о поле респондента достаточно проводить, сопоставляя ответы респондентов с данными наблюдения (критерий валидности — внешний вид респондента). Для социологов это предел требующейся надежности, в рамках социологии этот вопрос можно считать абсолютно обоснованным (однако для медиков, например, это далеко не так).

Поэтому процесс валидации методики критерием валидности, а этого критерия — другим и т. д. не длится бесконечно, а заканчивается критерием, валидность которого достаточно очевидна и не требует проверки другим критерием («последняя инстанция»). В практике социологических исследований эта цепочка, как правило, состоит не более чем из двух-трех звеньев, причем звенья могут быть связаны с разными исследованиями: один исследователь провел валидизацию методики А критерием В, а другой усомнился в валидности этого критерия и провел валидизацию критерия В критерием С и т. д. Последний критерий

в этой цепочке, валидность которого априорно достаточна для всех социологов и не требует доказательств, назовем *эталонным критерием* или *эталоном*. Это как бы принятые в социологии аксиомы, к которым сводятся другие утверждения, подлежащие доказательству. Например, при проведении опроса на предприятии валидизация вопроса анкеты о среднемесячном заработке респондентов может осуществляться с помощью ведомостей уплаты профсоюзных взносов. Если исследователь считает и эти данные недостаточно валидными, предполагая, что не все респонденты аккуратно и точно платят взносы, то он может сопоставить профсоюзные ведомости с документами бухгалтерии. Последние выступают для социологов *эталонным критерием*, они не требуют дальнейшего обоснования, для социологов не существует более валидной информации о зарплате (хотя органы ОБХСС, например, могли бы заинтересоваться, действительно ли суммы, зафиксированные бухгалтерией, соответствуют реально выплаченному).

Слово «эталон» в названии критерия употреблено не столько в прямом (образцовая мера, образцовый измерительный прибор, служащие для воспроизведения, хранения и передачи единиц измерения с наивысшей достижимой при данном состоянии науки и техники точностью), сколько в переносном смысле (мерило, образец для сравнения).

Суть моего предложения заключается в том, чтобы проводить типологию методов валидизации по критериям последней инстанции, эталонам. Конструктивность такого подхода я вижу в том, что он задает некоторый алгоритм построения валидной методики: если мы установили, что для измерения данного признака наиболее адекватным является такой-то тип валидизации, то это обязывает нас либо использовать для валидизации *эталонный критерий* либо показать, как соотносится другой критерий, который мы намерены использовать, с *эталонным критерием*, то есть построить всю цепочку доказательства валидности данной методики. Если мы, скажем, пришли к выводу, что для валидизации методики измерения установки *эталонным критерием* является реальное поведение, а нашей методикой предусмотрено построение шкалы установок типа Терстоуна [Зайцева, 1970, с. 220—243; Ядов, 1972, с. 106—112] и валидизация проходит путем оценки каждого суждения экспертами (судьями), то надо показать, как оценки судей связаны с реальным поведением. Впрочем, введение реального поведения как *эталонного критерия* тоже неудачно, так как неясно, каким путем получены данные о поведении: по ответам респондентов, путем наблюдения

или путем анализа документов. Выделение критерия в качестве эталона должны быть более определенным, чтобы его валидность либо была очевидной либо теоретически доказуемой (если валидность требует эмпирического обоснования, то это уже будет не эталонный, а промежуточный критерий).

В первом приближении можно выделить *объективные* и *субъективные* эталонные критерии. Это деление касается всех критериев и не специфично именно для критериев эталонных. «Объективные критерии охватывают реальное или «овеществленное» поведение обследуемого, воспринимаемое непосредственно, то есть не из сообщений, суждений и объяснений данного лица или других лиц [Процесс социального исследования, 1975, с. 141]. Соответственно, субъективные критерии состоят в сообщениях, суждениях и объяснениях данного лица и других лиц.

Далее объективные критерии могут разделяться по способу фиксации поведения (либо тех или иных характеристик респондентов) на критерии непосредственного наблюдения явления и предметной фиксации.

Критерий непосредственного наблюдения в качестве эталонного может использоваться при простых видах поведения, имеющих однозначную трактовку. Например, исследователь для контроля правильности ответов может визуально определить, кто с кем из школьников сидит за одной партой, кто сколько раз обращается к бригадиру в строительной бригаде и т. д., но уже определение наблюдателем возраста, эмоционального состояния респондента, характера обращения (просьба, совет, приказ) и т. п. само нуждается в валидации и не может быть эталоном.

Критерий предметной фиксации состоит в использовании для валидации методики характеристик респондента и его поведения, зафиксированных в некоторых материальных (в том числе — знаковых) системах. Дальнейшую классификацию этих критериев можно проводить по способу фиксации, выделив документальные и приборные критерии.

Документальные критерии — фиксация характеристик респондента, его поведения и результатов деятельности с помощью документов. Документ при этом понимается широко — наряду с письменными в это понятие включают иконографические документы (картины, кино-, видео- и фотодокументы и т. п.), фонетические документы (магнитофонные записи, грампластинки) и документы в машиночитаемой форме [Рабочая книга социолога, 1983, с. 277]. Надо сказать, что определение документа, данное в «Рабочей

книге социолога» («специально созданный человеком предмет, предназначенный для передачи и хранения информации» [там же, с. 276]), включает в число документов также и приборы, что представляется мне нецелесообразным. Например, радиопередатчик или телефон специально созданы для передачи информации, но стоит ли включать их в число документов? Удобнее под документом понимать лишь носитель информации (киноплёнки, фотографии, магнитофонные ленты, перфокарты), а не всю аппаратуру для записи, хранения и передачи информации. Если же принять такое деление, то можно выделить *приборные критерии* — фиксацию характеристик респондента, его поведения и результатов деятельности с помощью приборов.

Если использование документов в социологических исследованиях достаточно распространено, то приборы используются крайне редко (чаще их используют психологи). В качестве примера можно привести способ валидации ответов респондентов на вопрос о том, какие телевизионные программы они смотрят. С этой целью на репрезентативной выборке телевизоров устанавливалось специальное устройство, фиксирующее время и длительность включения того или иного телевизионного канала. Кроме этого случая, с помощью приборов делались попытки определить сплоченность малой социальной группы⁶. Но в общем приборы в социологических исследованиях практически не используются.

Как и в ранее проведенных классификациях, существуют случаи, которые трудно отнести к тому или иному классу. Если исследователь с помощью телекамеры записал на видеомagneфон фильм о поведении респондентов и затем анализирует его, то это и использование приборов (телекамера) и документов (видеофильм) и наблюдение за поведением (хотя и опосредованное). Поскольку многие приборы имеют регистрирующую часть и их показания так или иначе фиксируются на каких-либо носителях (документах), то могут возникнуть сложности с различием доку-

⁶ Сплоченность определялась опосредованно, по уровню содержания в моче членов экипажей бомбардировщиков 17-кетостероидов [Шовен, 1972, с. 160—164]. Приводятся также данные о связи социометрической сплоченности с содержанием жирных кислот в крови у членов экипажа и связи ритма сердца и кожно-гальванической реакции с иерархией социометрических статусов в группе. «Быть может эти исследования, — задается вопросом автор, — представляют собой первые шаги истинно научной экспериментальной социологии?» [там же]. «Святая вера» французского биолога в приборы демонстрирует абсолютизацию методов своей науки и непонимание специфики социальных наук.

ментальных и приборных критериев. В большинстве же случаев показания приборов, дающие четко фиксируемую однозначную информацию существенно отличаются от обычных социологических документов (никто ведь не станет относить видеофильм к показаниям прибора «телекамера») ⁷.

Предметно фиксируемым является также поведение респондента, овеществленное им в создании некоторых предметов (например, булочки, выпеченные поваром), но исследователь не использует их непосредственно (по крайней мере для научных целей), а использует лишь информацию о них (например, для определения производительности труда подсчитывается число булочек, выпеченных за смену). Информацию же он получает либо непосредственным наблюдением (считает булочки), либо путем анализа документов, либо по данным автоматического счетчика, либо путем опроса (мастера, респондента), поэтому выделять фиксацию результатов труда как один из видов критерия нецелесообразно, так как этот критерий сводится к уже рассмотренным.

Субъективные эталонные критерии можно разбить на два вида — информация, полученная от самого респондента, и информация о респонденте или о тех или иных объектах, получаемая от других лиц (экспертов, судей, информантов, резидентов и т. п.). Информацию, полученную от респондента, назовем *самооценочной*. Это название не совсем точно, так как информация респондента включает не только самооценки, но и другие высказывания: сообщения, суждения, объяснения и т. п., но, во-первых, в качестве эталона выступают преимущественно самооценки (сообщения, суждения, высказывания сами нуждаются в валидации), во-вторых, не хочется вводить неологизмы типа *самосуждения* и *самовысказывания*. Информацию, полученную от других лиц, как и самих этих лиц, в социологии и в смежных с ней областях называют по-разному (этносоциологи, например, пользуются понятием «информанты»). Мы назовем эту информацию *экспертной*, понимая под ней любые оценки кроме самооценок.

Каждый тип критерия задает свой тип валидности. Таким образом мы рассмотрели объективную валидность, включающую наблюдаемую, документальную и приборную валидность, и субъективную, включающую самооценочную и экспертную валидность.

⁷ Можно было бы считать приборно-фиксируемую информацию одним из видов документов, а предметно-фиксируемый критерий — одним из видов документального критерия.

Это деление индуцирует типологию социальных признаков по тому, какой критерий валидации является эталонным для данного признака. Например, для большинства социально-демографических характеристик респондентов (пол, возраст, образование, партийность и т. н.) эталонным критерием являются документы. Как правило, это постоянные либо устойчивые характеристики респондентов, не меняющиеся в течение нескольких месяцев, лет, а иногда и десятилетий, а также характеристики его поведения, отражающие основные жизненные выборы и события в долгосрочных процессах (выбор учебного заведения, работы, рост квалификации, разряда, переход с работы на работу, вступление в брак, развод, рождение ребенка и т. п.). Реже это документы, фиксирующие ситуативные факты поведения (процент выполнения нормы за смену, хулиганский поступок, зафиксированный милицией, получение книги в библиотеке и т. п.).

Для ситуативных фактов поведения, не фиксируемых документами, эталонным может быть критерий непосредственного наблюдения (наблюдать можно те или иные действия респондента, их последовательность, время их выполнения, формы общения и кооперации между членами группы при выполнении групповых действий и т. п.). Для некоторых видов ситуативного поведения, не фиксируемых документами, может в качестве эталонного критерия использоваться информация, полученная с помощью прибора. Думаю, что это направление очень перспективно для социологии, особенно с развитием микропроцессорной техники и возможности снабжать приборы устройствами, более полно фиксирующими информацию.

Для удовлетворенности и других признаков, связанных с самоощущением человека, эталонным критерием являются самооценки. Хочу еще раз подчеркнуть отличие валидности от правильности и устойчивости. Валидность как отсутствие методологических ошибок при выборе методики означает, что сама методика, основные принципы ее построения выбраны правильно. Тем не менее, при более детальной разработке методики и при проведении исследования могут быть допущены случайные и систематические ошибки, полученная информация станет неустойчивой и неправильной. Эти ошибки могут быть настолько велики, что несмотря на валидность информация будет совершенно непригодной для исследователя.

Рассмотрим, например, удовлетворенность респондента работой. Говоря, что эталонным критерием для нее является самооценка, я тем самым утверждаю, что информация

об удовлетворенности респондента работой, полученная от него самого, максимально валидна (при данном уровне развития социологии). Значит, при полном отсутствии случайных и систематических ошибок мы получим именно то, что требовалось измерить, — точную и правильную оценку удовлетворенности респондентов работой, и не существует другого более надежного способа измерения удовлетворенности.

Никакие объективные показатели, — как физиологические (пульс, дыхание, кожно-гальваническая реакция и т. п.), так и социальные (высокая зарплата, хорошие жилищные условия и т. п.), — свидетельствующие, с точки зрения исследователя, о полном благополучии респондента, никакие наблюдения или суждения других людей не являются основанием для утверждения, что данный человек удовлетворен, если он отрицает это. Мы можем подозревать его в неискренности, в нежелании по каким-то причинам говорить о своей удовлетворенности, можем думать, что под удовлетворенностью он понимает нечто иное, чем мы и т. п., — то есть что в информации содержатся случайные и систематические ошибки. В этом случае можно пытаться устранить возможные смещения — изменить методику опроса, ситуацию, исполнителя, но в любом случае аутентичную информацию об удовлетворенности человека мы можем получить только от него самого. Любые другие способы (например, приборные) придется валидизировать свидетельствами самого респондента, в то время как самооценка валидизируется самооценкой же, но очищенной от возможных смещений. В этом состоит специфика данного критерия как эталонного.

Для широкого класса характеристик личности (прежде всего, стабильных, проявляющихся в межличностном общении) и ее поведения (прежде всего, ситуативного, проявляющегося в течение достаточно длительного времени, в частности, в результатах деятельности) эталонным критерием являются экспертные оценки. Если вспомнить вынесенное в эпиграф изречение Юлиана Тувима, то можно сказать, что стоимость знакомых «в действительности» — это то, во сколько их оценивают другие. Не существует другого более обоснованного способа оценки.

Специфика экспертных оценок как эталонных заключается в следующем. Если валидизируя один критерий другим, мы дошли в конце концов до экспертных оценок, то их валидизация может осуществляться только другими экспертными оценками, в которых устранены возможные ошибки измерения, то есть речь идет, фактически, не о ва-

лидизации, а об устранении случайных и систематических ошибок. Более подробно на экспертных (прежде всего, межличностных) оценках я останавлиюсь в следующем параграфе.

Деление валидности на актуальную и прогностическую проведено на другом основании и может быть отнесено к каждому из выделенных мною типов валидности. Например, экспертная валидность может быть актуальной, если экспертиза проводится параллельно с валидируемым тестом, и прогностической, если экспертиза проводится через некоторый отрезок времени.

Введенная здесь типология валидности дает возможность разрабатывать конкретные способы валидизации методик измерения социальных признаков.

Но для изложения методов валидизации необходимо рассмотреть вопрос о вычислении показателей эмпирической валидности.

2. Показатели валидности

При определении степени эмпирической валидности в отличие от других компонентов надежности используются не специальные меры, а коэффициент корреляции между значениями данного признака, полученными для каждого респондента с помощью валидируемой методики и с помощью критерия валидности.

Вид коэффициента корреляции определяется соотношением типов шкалы, применяемых в валидируемой методике и в методике-критерии. Если, например, оба измерения метрические, то используется коэффициент Пирсона r , если одно — дихотомическое, а другое — метрическое, то используется точечно-бисериальный коэффициент и т. д.⁸ Что же касается теоретической валидности, то ее, как указывалось, не измеряют, ее только оценивают. Далее, когда выясняется область валидности и теоретически обоснованный тест сопоставляют с другими признаками, то используются коэффициенты корреляции. Теоретически валидная методика должна быть и эмпирически валидной. Но для теоретически валидных тестов достаточно, чтобы были значимыми коэффициенты корреляции, которые в соответствии с теорией исследователя должны быть связаны с тестом. В этом случае задачей эмпирического исследования является лишь проверка гипотез о связи измеряемого с други-

⁸ Таблицу соответствия уровня измерения виду коэффициента и формулы расчета коэффициентов можно найти в: [Гласс, Стэнли, 1976, с. 144; Паниотто, Максименко, 1982, с. 166, табл. 36].

ми признаками. При построении же эмпирически валидных тестов необходимо провести отбор вопросов (суждений и т. д.). Поэтому коэффициенты должны быть не просто значимыми, но и достаточно большими. Но что значит «достаточно», то есть каким образом оценить степень валидности?

Рассмотрим этот вопрос на следующем примере. Предположим, что мы построили тестовую методику для оценки

трудового вклада работников и на некотором массиве (285 человек) определили трудовой вклад каждого работника с помощью теста и с помощью эталонного критерия валидизации.

Сопоставим теперь тестовую оценку с критерием валидизации. Для этого разобьем интервал изменения тестовой оценки

на два участка («высокий» и «средний» трудовой вклад) и найдем таблицу двумерного распределения с критерием валидизации, эталоном (табл. 4.1). Принятые для таблиц 2×2 коэффициенты связи составляют: коэффициент Юла $Q=0,57$; коэффициент контингенции $\Phi=0,31$ (он, собственно, совпадает с коэффициентом Чупрова и Крамера); критерий $\chi^2=26,9$, в то время как для 5 %-й значимости достаточно, чтобы значение критерия было 3,84, для 1 %-й — 6,63, для 0,1 %-й — 10,83, то есть связь значима на более чем 0,1 %-м уровне. Среди лиц, имеющих высокий трудовой вклад по тестовой оценке, приблизительно в 3 раза больше лиц с высоким трудовым вкладом, чем со средним. Хорошо это или плохо? Много или мало? Что можно сказать о степени валидности построенной тестовой шкалы?

Иногда формулируют некоторые критерии для оценки степени валидности. «Для определения индивидуальной выраженности признака со статистической точки зрения требуется $r_{tc}=0,7$ (r_{tc} — коэффициент корреляции между тестом и критерием валидизации. — В.П.). В практических же целях считается: высокая обоснованность $r_{tc} \geq 0,6$; средняя обоснованность $0,6 > r_{tc} \geq 0,4$; низкая обоснованность $r_{tc} < 0,4$. Пригодные процедуры должны обладать средней степенью обоснованности. Но и при r_{tc} порядка 0,3 можно производить отбор, если этот показатель значимо

Таблица 4.1.

Трудовой вклад по оценке			ИТОГО
тестовой	Эталонной		
	высокий	средний	
высокий			
абс.	74	60	134
%	55,2	44,8	100
средний			
абс.	38	113	151
%	25,2	74,8	100

отличается от 0» [Процесс социального исследования, 1975, с. 148]. Такие критерии, хотя и являются некоторым ориентиром для социологов, но скорее могут ввести их в заблуждение, чем оказаться полезными. Здесь такая же ситуация, как с объемом выборки: если на вопрос, сколько людей нужно опросить, ответить конкретной цифрой, то не для всех случаев это будет справедливо. Конструктивный ответ должен учитывать, для чего проводится опрос, каковы цели исследования. Иначе говоря, при выборке определенного объема ошибка по доле любого признака с какой-то вероятностью не будет превышать такой-то величины. Если же исследователя интересует сравнение различных групп путем расчета таблиц двухмерного распределения, как мы видели в предыдущей главе, объем выборки определяется иначе.

Аналогично показатель степени валидности должен давать исследователю информацию о том, каким образом может использоваться тест данной валидности. Попытаемся ввести такой показатель.

Коэффициент валидности α . Что можно сказать о тесте, описываемом таблицей 4.1? Прежде всего, ясно, что тест непригоден для индивидуальной диагностики трудового вклада — мы будем ошибаться в одном случае из четырех, утверждая, что данное лицо имеет высокий трудовой вклад. Но социологам это и не нужно, они делают выводы о тех или иных социальных общностях или категориях. В данном случае нас интересует лишь доля лиц с высоким трудовым вкладом для изучения различий между какими-либо группами по этому признаку. Например, может представлять интерес, какая подготовка — ИТУ или десятилетка с обучением на рабочем месте — более эффективна с экономической точки зрения. Рассмотрим вопрос о сравнении двух групп по тестовому показателю более подробно.

Предположим, что мы выделили две группы — A и B и изучали выборку респондентов n_a из первой и n_b из второй группы, определив долю лиц с высоким трудовым вкладом в каждой группе (d'_a и d'_b) с помощью теста, описываемого таблицей 4.2.

Очевидно, что показатели, приведенные в табл. 4.2, удовлетворяют следующим соотношениям:

$$n_1 = n_{11} + n_{12}, \quad n_2 = n_{21} + n_{22}; \quad (4.2)$$

$$N = n_1 + n_2, \quad \bar{n}_1 = n_{11} + n_{21}, \quad \bar{n}_2 = n_{12} + n_{22}; \quad (4.3)$$

$$r = \frac{n_{11}}{n_1}, \quad s = \frac{n_{22}}{n_2}. \quad (4.4)$$

Значимость различий между двумя группами по доле лиц, имеющих данное значение признака (в частности, лиц с высоким трудовым вкладом), определяется по формуле:

$$z = \frac{|d_a - d_b|}{\sqrt{d(1-d)\left(\frac{1}{n_a} + \frac{1}{n_b}\right)}}, \quad (4.5)$$

где $d = \frac{d_a n_a + d_b n_b}{n_a + n_b}$ (формулой можно пользоваться при $n_a, n_b \geq 50, n_a d_a > 5, n_b d_b > 5, n_a(1-d_a) > 5, n_b(1-d_b) > 5$).

Таблица 4.2*

Трудовой вклад по оценке					ИТОГО
тестовой	эталонной				
	высокий		средний		
Высокий	n_{11}	r	n_{12}	$1-r$	n_1
Средний	n_{21}	$1-s$	n_{22}	s	n_2
Итого	$\tilde{n}_1$		$\tilde{n}_2$		N

* n_{ij} — число лиц, имеющих i -е значение тестового и j -е значение эталонного теста в валидационном исследовании (то есть при разработке теста), 1 — высокий, 2 — средний, $i, j = 1, 2$.

Истинная доля лиц с высоким трудовым вкладом в группе А (обозначим эту долю через d_a) и в группе В (d_b), которую можно узнать с помощью эталонного критерия, неизвестна. Известны лишь доли, определенные с помощью теста, — d_a^t и d_b^t (все показатели, относящиеся к данным, полученным с помощью теста, обозначим индексом t). Но так как есть характеристика теста, заданная таблицей 4.2, то истинную долю можно вычислить:

$$\begin{aligned} d_a &= r d_a^t + (1-s)(1-d_a^t) = d_a^t(r+s-1) + (1-s), \\ d_b &= d_b^t(r+s-1) + (1-s). \end{aligned} \quad (4.6)$$

Тогда

$$d_a - d_b = (d_a^t - d_b^t)(r+s-1). \quad (4.7)$$

Чтобы воспользоваться формулой (4.5) необходимо найти d :

$$d = \frac{1}{n_a + n_b} \cdot \{n_a [d_a^t(r+s-1) + (1-s)] +$$

$$+ n_b [d_b^t (r + s - 1) + (1 - s)] \} = \frac{1}{n_a + n_b} [(n_a d_a^t + n_b d_b^t) \times \\ \times (r + s - 1) + (1 - s)(n_a + n_b)] = d^t (r + s - 1) + (1 - s), \quad (4.8)$$

где
$$d^t = \frac{d_a^t n_a + d_b^t n_b}{n_a + n_b}.$$

Тогда

$$1 - d = s - (r + s - 1)d^t, \\ d(1 - d) = [d^t (r + s - 1) + (1 - s)] \cdot [s - d^t (r + s - 1)]. \quad (4.9)$$

Наконец, подставив всё это в формулу (4.5), получим:

$$z = \frac{|d_a^t - d_b^t| \cdot (r + s - 1)}{\sqrt{[d^t (r + s - 1) + (1 - s)] \cdot [s - d^t (r + s - 1)] \left(\frac{1}{n_a} + \frac{1}{n_b} \right)}}. \quad (4.10)$$

После некоторых простых преобразований:

$$z = z^t \cdot \frac{(r + s - 1) \cdot \sqrt{d^t (1 - d^t)}}{\sqrt{[d^t (r + s - 1) + (1 - s)] \cdot [s - d^t (r + s - 1)]}} = z^t \cdot k, \quad (4.11)$$

где в соответствии с формулой (4.5):

$$z^t = \frac{|d_a^t - d_b^t|}{\sqrt{d^t (1 - d^t) \left(\frac{1}{n_a} + \frac{1}{n_b} \right)}}. \quad (4.12)$$

Итак, при использовании тестовой методики при определении значимости различий показатель, рассчитанный по формуле (4.12), домножается на какое-то устрашающего вида выражение, обозначенное через k (4.11). Чтобы лучше представить себе смысл этого выражения, рассмотрим частный случай.

Пусть $d^t = 0,5$. Тогда (4.11) преобразуется следующим образом:

$$z = z^t \cdot \frac{r + s - 1}{\sqrt{1 - (r - s)^2}} \geq z^t (r + s - 1). \quad (4.13)$$

Неравенство связано с тем, что $(r - s)^2$ может принимать лишь положительные значения.

Что же выражает формула (4.13)? Что при использовании тестовой методики значимость различий изменяется в $(r+s-1)$ раз по сравнению с ситуацией, когда мы используем эталонный критерий. Это «плата» за использование менее точного метода. Например, для разработанного теста (таблица 4.1) $(r+s-1)=0,3$, то есть показатель уменьшается более чем в три раза.

При использовании тестовой методики мы должны либо примириться с тем, что сможем интерпретировать только самые явные различия, либо увеличивать выборку. Например, если мы хотим интерпретировать как значимые различия не менее чем на 15% (при 5 %-й значимости для этого нужно, чтобы объем выборки в каждой из групп был равен 85 при использовании эталонного критерия), то при использовании данного теста объем выборки надо увеличить приблизительно в 11 раз, то есть в каждой группе должно быть около 945 человек. Чем лучше тест, чем больше он совпадает с эталонным критерием, тем меньше увеличение выборки требуется. Если бы, например, мы определяли лиц с высоким и средним трудовым вкладом не с вероятностью 0,75 и 0,55 соответственно, а с вероятностью 0,85 для тех и других, то объем выборки надо было бы увеличить лишь в два раза, — до 190 респондентов в группе. Как видим, разработка хорошего теста снижает трудоемкость исследования.

Выражение $(r+s-1)$ изменяется от 0 до 1 по модулю. Если тест не прогнозирует значение эталона, то есть предсказывает его с вероятностью 0,5 (либо да, либо нет), $r=s=0,5$, а $(r+s-1)=0$, если предсказывает с вероятностью 1 — то $(r+s-1)=1$. Мне кажется, это выражение удобно использовать для характеристики качества теста. Обозначим его α и назовем *коэффициентом валидности*:

$$\alpha = r + s - 1. \quad (4.14)$$

Таким образом, коэффициент валидности разработанного теста равен 0,3. Коэффициент не очень высокий, требует увеличения выборки в 11 раз, поэтому тест пригоден лишь для крупномасштабных исследований при объемах выборки более 2—3 тысяч и при не очень большом дроблении массива.

Формула (4.13) применима лишь при d^t близких к 0,5. При других значениях d^t следует использовать формулу (4.11). Для упрощения расчетов я табулировал значения формулы (4.11) (табл. 4.3). Расчеты приведены в предположении, что $r=s$. Если $r \neq s$, то данные таблицы 4.3 дадут результат с некоторым запасом. Достаточно рассчитать

Таблица 4.3*

Коэффициент валидности α		Значение d^t				
		0,1 или 0,9	0,2 или 0,8	0,3 или 0,7	0,4 или 0,6	0,5
0,1	k	0,06	0,08	0,09	0,10	0,10
	l	278	156	123	100	100
0,2	k	0,12	0,16	0,18	0,20	0,20
	l	69	39	31	25	25
0,3	k	0,18	0,24	0,28	0,29	0,30
	l	31	17	13	12	11
0,4	k	0,25	0,33	0,37	0,39	0,40
	l	16	9,2	7,3	6,6	6,3
0,5	k	0,33	0,42	0,47	0,49	0,50
	l	9,2	5,7	4,5	4,2	4,0
0,6	k	0,41	0,51	0,57	0,59	0,60
	l	5,9	3,8	3,1	2,9	2,8
0,7	k	0,51	0,62	0,67	0,69	0,70
	l	3,8	2,6	2,2	2,1	2,0
0,8	k	0,62	0,73	0,77	0,79	0,80
	l	2,6	1,9	1,7	1,6	1,6
0,9	k	0,78	0,85	0,88	0,90	0,90
	l	1,6	1,4	1,3	1,2	1,2

* k — коэффициент уменьшения значимости различий, l — коэффициент компенсирующего увеличения объема выборки.

значимость различий тестовых показателей — формулы (4.5), (4.8), (4.12)⁹ — и домножить на коэффициент k из таблицы 4.3. Коэффициент увеличения объема выборки l (табл. 4.3) необходим для полной компенсации уменьшения значимости различий, связанного с использованием теста вместо эталонного критерия. Как видим, с увеличением d^t значимость различий существенно уменьшается, а объем выборки увеличивается. К счастью, значения d^t , равные 0,1; 0,2; 0,8 и 0,9, встречаются редко, а при d^t больше чем 0,2 и меньше чем 0,8 показатели k и l мало зависят от d^t . Из таблицы видно также, что тест с $\alpha=0,1$ практически непригоден, тест с $\alpha=0,2$ может использоваться при больших выборках, для установления явных различий, тест с $\alpha=0,3$ — либо при больших объемах выборки, либо для установления наиболее явных различий. Тесты с $\alpha=0,5$ или $\alpha=0,6$ могут использоваться при выборках объема

⁹ Подробнее см.: [Палиотто, Максименко, 1982, с. 191—195] (либо любой другой учебник по статистике).

1—1,5 тысячи. Тесты же с более высоким коэффициентом валидности не очень существенно отличаются от эталонных критериев. Таблица 4.3 позволяет, таким образом, конкретизировать представление о степени валидности теста¹⁰.

Введенный коэффициент валидности α дает, как мне кажется, достаточно полное представление о валидности разработанного теста. Однако он пригоден лишь для дихотомических признаков или для признаков, сведенных к дихотомическим. Практически любые признаки могут быть сведены к дихотомическим, но при этом теряется часть информации. Поэтому актуальной задачей является разработка хорошо интерпретируемых показателей для оценки степени валидности методик измерения социальных признаков без дихотомической редукции. Это не значит, что необходимо разработать обязательно новые коэффициенты валидности — такими показателями могут быть и традиционные коэффициенты корреляции. Требуется, однако, разработать критерии, позволяющие определить, для каких целей пригодна методика, позволяющая получить то или иное значение коэффициента корреляции тестовой и эталонной методик.

Важно отметить, что валидизация методики осуществляется всегда на некоторой выборке, которая называется *группой валидизации*. Коэффициенты валидности, полученные на данной группе валидизации, с полной уверенностью можно использовать лишь для выборки из той же генеральной совокупности. Для использования методики по отношению к иной генеральной совокупности необходимо оценить, насколько изучаемая выборка отличается от груп-

¹⁰ Возможен и другой подход к компенсации уменьшения различий, связанных с заменой эталонного критерия тестом. Пусть, как и раньше, нас интересует, отличаются ли две группы респондентов A и B по значению некоторого признака, измеренного с помощью теста t , то есть интересует значимость различий d_a^t и d_b^t . Предлагаемый мною подход заключается в увеличении выборки в l раз (в соответствии с табл. 4.3), с тем чтобы различия, уменьшенные при использовании теста вместо эталона, стали значимыми. Еще один подход заключается в том, чтобы компенсировать это уменьшение, домножив показатель значимости различий на некоторый коэффициент, зависящий от характеристики теста (например, $1/k$). Этот подход аналогичен поправке на аттенюацию (форм. (2.45)) и, видимо, предъявляет более жесткие требования к объему выборки, на которой разрабатывался тест. Вопрос о соотношении этих двух подходов требует дальнейших теоретических и экспериментальных исследований и может быть сформулирован как нерешенная задача. Кроме того, дальнейших исследований требуют надежность характеристики теста (табл. 4.2) и связь надежности теста со значимостью различий между группами, установленных с помощью этого теста.

пы валидации и как это может сказаться на валидности.

Таким образом, просто утверждение о том, что данная методика (или информация) валидна, слишком абстрактно. Для того, чтобы оно имело практический смысл, необходимо указать при эмпирической валидности — группу валидации, тип валидности и степень валидности (коэффициент корреляции), при теоретической валидности — способ построения методики и область валидности (эмпирические данные — если они есть — о связи измеренного с другими признаками).

Рассмотрим следующий пример.

Измерение коэффициента интеллектуальности. Задача эта типична для психологии, реже она встречается в социальной психологии и уже совсем редко в социологии. Отчасти это связано с характером задач, решаемых в социологии, отчасти — со сложностями измерения этой характеристики в массовых исследованиях. Если типичными для психологии являются методики с высокими показателями надежности, но требующие относительно длительных сроков на выполнение заданий методики, то в социологии используются менее надежные методики, но более оперативные. Кроме того, для социологических исследований требуются, как правило, достаточно большие массивы респондентов. Наконец, психологические методики требуют, как правило, непосредственного участия экспериментатора, а социологические — не обязательно.

Анализ истории развития социологических методов показывает тенденцию переориентации социологов от очных к заочным методам опроса [Яковенко, 1984] (раздаточное анкетирование, почтовый опрос, телефонный опрос [Рукавишников, Панютто, Чурилов, 1984, гл. 4], использование технических устройств и т. п.). Эти особенности приводят к необходимости разрабатывать для социологии инструментарий (преимущественно анкеты), задания которого не требуют много времени и могут выполняться респондентом в отсутствие экспериментатора, что накладывает существенные ограничения на сложность методики. Тесты же на определение умственных способностей (типа теста Айзенка) требуют решения ряда тестовых задач, инструкции к которым уже требуют определенных умственных усилий для их понимания. Кроме того, на их решение отводится фиксированное время, поэтому их нельзя использовать в заочных методах опроса. Наконец, не очень простой задачей является создание мотивации для решения задач теста — далеко не всем респондентам нравится положение экзаменуемых. Все это отчасти и объясняет, почему тесты

на определение умственных способностей (коэффициента интеллектуальности) редко используются в массовых исследованиях. На примере разработки оперативного (и сравнительно несложного) теста на определение коэффициента интеллектуальности я хочу продемонстрировать возможный путь перехода от психологических к социологическим тестам.

В качестве критерия валидизации был принят тест Айзенка, имеющий теоретическую валидность¹¹. Исходным материалом для создания нового теста служили юмористические картинки. Я обратил внимание на то, что понимание некоторых юмористических картинок требует определенных умственных усилий и может быть связано с умственными способностями. Когда о человеке говорят, что «юмор до него не доходит», это вызывает подозрение, что у него не все в порядке и с умственными способностями. Во всяком случае, антонимом «остроумия» является «тупоумие», которое интерпретируется как низкий уровень умственного развития.

Разумеется, восприятие юмористических картинок может определяться не только логическими («левополушарными») способностями индивида, которые преимущественно и измеряются тестом Айзенка, но и его художественными («правополушарными») способностями, тезаурусом, вкусом. Все же думаю, что можно отобрать такие картинки, которые высоко коррелируют с тестом Айзенка. Кроме того, ценностные ориентации, вкус, чувство юмора определяют лишь оценку степени остроумия рисунка, а не степень его понимания. Поэтому предлагаемые респондентам картинки (прил. 3) требовалось оценить не только по степени остроумия, но и по степени понятности.

Поскольку критерий валидизации измеряет врожденные способности, распределение которых в различных социальных группах одинаково, группа валидизации не должна обеспечивать репрезентативность по социальным признакам¹². Поэтому использовалась такая процедура опроса: интервьюеры опрашивали лиц из своего социального окружения, стараясь обеспечить разнообразие опрошенных по полу и возрасту. Респондентам сообщали, что цель исследования им разъяснят после проведения опроса, и предлагали анкету «Юмористические картинки» (прил. 3) с просьбой оценить остроумие каждой картинки, ее понят-

¹¹ Использован 4-й тест из книги [Айзенк, 1972, с. 67—79].

¹² Обычное для подавляющего большинства психологических исследований (хотя и сомнительное) предположение.

ность и кратко описать, что на ней изображено. По этому описанию интервьюер оценивал степень понимания картинки и указывал в соответствующей графе анкеты. Таким образом, с каждой картинкой были связаны три оценки: степень остроумия, степень понятности по самооценке респондента, степень понятности по оценке интервьюера (последняя оценка давалась исходя из специально разработанного перечня «канонических» интерпретаций). После опроса по анкете «Юмористические картинки» респонденту с интервалом в несколько дней предлагался тест Айзенка (это, впрочем, не принципиально и связано с большими затратами времени для респондента — опрос с помощью обеих анкет требует от полутора до двух часов интенсивной работы). Затем интервьюер (для желающих) сообщал цель исследования и результат респондента по тесту Айзенка.

Объем выборки предполагалось определить с помощью процедуры типа последовательного анализа Вальда — проводить опрос до тех пор, пока не будут получены значимые коэффициенты. Оказалось, однако, что уже первый «останов» — после опроса 40 респондентов — дал результаты, значимые на 0,1 %-м уровне. Для каждого респондента определялась сумма оценок остроумия, самооценки понятности и оценки интервьюера по всем картинкам, а также рассчитывался коэффициент интеллектуальности по тесту Айзенка. Тем самым для каждого респондента были заданы четыре признака. Затем рассчитывались коэффициенты корреляции r между первыми тремя и четвертым признаком — коэффициентом интеллектуальности по Айзенку. Для оценки остроумия картинок: $r=0,42$, $t=2,82$; для самооценки понятности: $r=0,52$, $t=3,74$; для оценки понятности, данной интервьюером: $r=0,52$, $t=3,71$.

Затем из 29 картинок были отобраны 11, дающие наибольшую корреляцию оценки понятности с тестом Айзенка. Их номера (в соответствии с прил. 3): 2, 7, 11, 12, 13, 14, 16, 17, 18, 26, 27. Затем опять рассчитывались коэффициенты корреляции, но для сумм, полученных не по всем 29, а только по этим 11 картинкам. Коэффициенты корреляции теста Айзенка с оценкой остроумия: $r=0,49$, $t=3,46$; с оценкой понятности, данной респондентом: $r=0,58$, $t=4,33$; с оценкой понятности, данной интервьюером: $r=0,62$, $t=4,91$. Все эти коэффициенты значимы на 0,1 %-м уровне (для данного уровня значимости при 40 опрошенных, то есть при 38 степенях свободы, $t=3,57$, следовательно, оценка остроумия близка к 0,1 %-му уровню значимости, хотя и несколько «не дотягивает»).

Здесь, как и в предыдущих случаях, удобно построить таблицу 2×2 и рассчитать коэффициент валидности. Разобьем интервал изменения показателя теста Айзенка на два интервала: до 105 (средний коэффициент), 105 и выше (высокий коэффициент). Аналогично разобьем на два интервала оценки остроумия и понятности и построим соответствующие таблицы, а затем рассчитаем по ним значение α . Оказывается, что для оценки остроумия $\alpha=0,25$, для самооценки понятности $\alpha=0,58$, а для оценки понятности, данной интервьюером, $\alpha=0,61$.

Таблица 4.4.

Умственные способности выражены			
по тесту «Юмористические картинки»	по тесту Айзенка		итого
	в высокой степени	в средней степени	
в высокой степени			
абс.	17	7	24
%	70,8	29,2	100
в средней степени			
абс.	2	14	16
%	12,5	87,5	100

Полученные данные говорят о том, что самооценка понятности хорошо согласуется с оценкой по тесту Айзенка. Построенный тест оказался достаточно эффективным. Он дает те же результаты, что и тест Айзенка, при троекратном увеличении объема выборки, то есть 20 %-е различия на 5 %-м уровне можно интерпретировать как значимые при 150 респондентах в группе, а 14 %-е — при 300 респондентах. Характеристика теста приведена в таблице 4.4 (сумма оценок понятности разбита на интервалы «до 17» и «17 и более»).

Как видим, тест вполне пригоден для массовых социологических опросов. Предполагается включить его в анкету для почтового опроса жителей Киева, чтобы «ненавязчиво» получить данные о коэффициенте интеллекта респондентов (к тому же юмористические картинки дают разрядку и облегчают заполнение анкеты).

3. Методы валидизации

Рассмотренная типология валидности задает определенный алгоритм построения надежной методики измерения социальных признаков, который можно представить состоящим из следующих этапов.

1. Решение вопроса о том, каким способом будет валидизироваться методика — теоретическим или эмпирическим.

При теоретической валидизации:

2. Операционализация введенного исследователем понятия.

3. Поиск эмпирических референтов, способов их измерения и интеграции в показатель (признак), соответствующий данному понятию.

4. Выдвижение гипотез о связи признака, соответствующего данному понятию, с другими признаками.

5. Проведение эмпирического исследования для проверки гипотез.

6. При подтверждении гипотез методика готова, при неподтверждении — возврат к пунктам 3, 2, 1 или к теории, положенной в основу измерения. В дальнейших исследованиях область валидности методики может уточняться.

При эмпирической валидизации:

2. Подбор материала (суждения, задачи, картинки и т. п.), предположительно дифференцирующего респондентов по валидизируемому признаку.

3. Определение типа валидизации (по эталонному критерию).

4. Выбор критерия валидизации и доказательство его связи с эталонным критерием.

5. Проведение эмпирического исследования для сопоставления методики с критерием и отбор дифференцирующего материала.

6. При высокой корреляции отобранного материала с критерием методика готова (один из возможных показателей корреляции — коэффициент α). При низкой — либо возврат к пункту 2, либо поиск возможных ошибок измерения критерия, их устранение и повторение исследования (либо то и другое).

При теоретической и при эмпирической валидизации.

7. Проведение исследований по определению правильности и устойчивости методики и (если необходимо) ее доработка.

Рассмотрим процесс валидизации на примере метода взаимных (экспертных) оценок. Почему для специального рассмотрения я выбрал именно этот тип валидности? На это есть некоторые принципиальные соображения.

В предыдущем параграфе было введено несколько типов объективной и субъективной валидности. Когда речь идет об объективном и субъективном, то у большинства людей объективное обычно ассоциируется с устойчивостью,

надежностью, правильностью, а субъективное — с ненадежностью, неправильностью. Словари дают одно из значений объективности — непредубежденность, соответствие объективной реальности в противовес субъективности как оценке чего-либо в зависимости от личных симпатий, настроения, склонностей. Поэтому объективные способы фиксации характеристик и поведения респондентов кажутся нам более надежными, чем субъективные. Но так ли это в действительности?

Думаю, что нет. По крайней мере, когда речь идет о валидации социальных признаков, то объективные критерии не более обоснованы, чем субъективные. Более того, если не ограничиваться первым приближением, то за объективными критериями часто стоят субъективные.

Действительно, рассмотрим введенные ранее объективные критерии.

Критерий непосредственного наблюдения надежен лишь при фиксации простых видов поведения. При наблюдении более сложных видов поведения (например, эмоциональных взаимоотношений между членами группы) возникает неоднозначность интерпретации увиденного, оценки исследователя подвержены всем тем смещениям, что и другие экспертные оценки, и в этом смысле являются субъективными или, точнее субъективными.

Документальная валидность означает сопоставление измеряемых признаков с содержанием документов. Но как возникли эти документы? Как правило, в них зафиксированы экспертные оценки людей, составляющих документы, либо людей, с чьих слов составляются документы. Если задать себе вопрос о валидности документа, то можно будет построить цепочку, которая закончится экспертной оценкой. Некоторые из документов, к которым доверчиво относятся социологи, скажем, для криминалистов являются лишь отправной точкой для начала работы, которая заканчивается свидетельскими показаниями или экспертизой.

Все это говорит о том, что субъективные критерии шире распространены и чаще используются как эталонные, чем это кажется на первый взгляд. В социологии особенно часто встречаются признаки, для валидации которых эталонным критерием являются субъективные оценки. К таким признакам относятся, в частности, многие характеристики личности исследуемых: нет более обоснованных оценок личности человека, чем оценки других людей. Именно поэтому в этом параграфе на примере нескольких исследований мы подробно рассмотрим использование взаимных

межличностных оценок для валидации социальных признаков.

Определение величины трудового вклада работника. Подчеркну важность этого конкретного вопроса и его роль в совершенствовании хозяйственного механизма. Главная народнохозяйственная задача, стоящая перед нашей страной, — переход на преимущественно интенсивный путь развития. Одним из основных направлений совершенствования хозяйственного механизма и интенсификации общественного производства является повышение производительности труда и качества работы. По оценке специалистов, совершенствование системы стимулирования труда, повышение трудовой активности, максимальное использование трудового потенциала личности и коллектива может дать наибольший эффект для повышения производительности труда по сравнению с затратами во всех других звеньях хозяйственного механизма [Буняч, 1981, с. 253]. Здесь возникает вопрос об адекватных количественных методах в исследовании трудовой деятельности. Чтобы ответить на этот вопрос, выделим сначала две группы стимулов труда. Первая — характеристики самого труда (прежде всего, его содержания), вторая — материальные и моральные вознаграждения.

Стимулирование через изменение характера и содержания труда намечено в Основных направлениях экономического и социального развития СССР на 1986—1990 гг. и на период до 2000 г. и предполагает повышение технической вооруженности труда, внедрение комплексной механизации и автоматизации производственных процессов. Партия и государство прилагают много усилий, чтобы сделать труд человека не только более производительным, но и содержательным, интересным, творческим. Важную роль в этом призвана сыграть ликвидация малоквалифицированного и тяжелого физического труда, предусмотренная взятым партией курсом на всемерную интенсификацию производства на основе научно-технического прогресса, структурной перестройки экономики, организации и стимулирования труда, повышения его качества и эффективности.

Но механизация и автоматизация производственных процессов — стратегическое направление развития народного хозяйства, они требуют больших затрат и не могут быть завершены в течение одной-двух пятилеток. Поэтому в ближайшие годы основными формами стимулирования остаются моральные и материальные вознаграждения за труд.

При таком стимулировании наиболее актуальна проблема эффективного применения социалистического принципа распределения по количеству и качеству труда, что требует ликвидации уравниловки, совершенствования системы стимулирования на основе справедливой и объективной оценки трудового вклада каждого работника.

Оказывается, однако, что оценка трудового **вклада** работников сталкивается с определенными трудностями. Наиболее распространенной в настоящее время формой определения трудового вклада является степень выполнения нормы (планов, обязательств, определенных объемов работы), но сама задача разработки таких норм является чрезвычайно сложной. Даже на рабочих местах с легко фиксируемыми результатами труда (например, количество деталей, произведенных станочником) при одном и том же показателе выполнения нормы трудовые затраты работников могут значительно отличаться в зависимости от сложности работы, от состояния оборудования, качества сырья, качества выполненной работы и т. п. Для многих других рабочих мест (например, научных работников, врачей, учителей, большинства научно-технических работников) возможность установления такой нормы вообще проблематична¹³. Такой подход — определение трудового вклада через нормы — критиковался в научной литературе также за то, что делает выгодным для работников (и для предприятия в целом) занижение нормы и побуждает к сокрытию резервов.

В качестве альтернативы указанному способу оценки труда неоднократно предлагалось считать критерием трудового вклада как отдельных работников, так и коллективов, не степень выполнения нормы, а результат сопоставления итогов деятельности работников, то есть оценка работников (и коллективов) должна исходить из того, как они выглядят на фоне других, а не на фоне своих собственных прошлых достижений, как это зачастую бывает сейчас.

Экспериментальная проверка систем стимулирования, основанных на этих принципах [Якушев, 1977, с. 150—161; 1978, с. 94—106; Кара-Мурза, 1981, с. 85—91] (например, система «Пульсар» или «карповская» система оценки научных работников), выявила как значительные резервы

¹³ Некоторые из существующих норм такого рода явно не выдерживают критики. Например, во многих НИИ деятельность научных работников оценивается числом и объемом публикаций. Это все равно, что оценивать писателя по толщине книг, художника — по площади картин, танцора — по длительности танца.

роста производительности труда, свидетельствующие о перспективности этих систем, так и весьма существенные недостатки самой системы стимулирования. Суть такого рода систем заключается в некоторой формализованной процедуре оценки деятельности каждого работника: деятельность расчленяется на составляющие — отдельные виды работы (выполнение норм, планов выступает лишь одним из ряда других, а не единственным критерием), каждая составляющая получает балльную оценку, затем эти баллы суммируются и получается итоговая оценка деятельности работника. Все работники ранжируются в соответствии с этой оценкой, а затем материальные и моральные вознаграждения (а также негативные санкции) распределяются между ними в соответствии с местом, которое они занимают в выявленной иерархии.

Отмечу две основные особенности таких систем стимулирования труда по сравнению с существующими методами. Первая особенность — распределение вознаграждений в соответствии с оценкой деятельности работника, соотношенной с оценкой деятельности других работников — заслуживает одобрения. Вторая — формализованная оценка различных составляющих деятельности и объединение их затем в один показатель — обладает целым рядом недостатков и подвергается поэтому критике в научной литературе. Получить такой показатель не проще, чем разработать обоснованную норму: возникают задачи, во-первых, выделения различных видов или аспектов деятельности, во-вторых, их соотношения, то есть определения удельного веса баллов по каждой из составляющих, в-третьих, определения круга экспертов или руководителей, которые будут выставять эти баллы. Сложность системы делает ее недостаточно гибкой, бюрократичной, поэтому результаты таких оценок часто воспринимаются членами коллектива как несправедливые.

Каков же выход? Думаю, что оценку трудового вклада следует проводить не опосредованно — через степень выполнения нормы или число набранных баллов, а через прямые сравнительные оценки трудового вклада каждого работника лицами, связанными с ним по работе (прежде всего, сотрудниками по трудовому коллективу, непосредственными руководителями и подчиненными). Дело в том, что эффективность тех или иных методов оценки трудового вклада работников определяют, в конечном итоге, по тому, воспринимаются ли они как справедливые членами коллектива, — без этого любая методика не будет стимулировать трудовую активность. Именно к этому апеллируют авторы

методик оценки трудового вклада, доказывая объективность своих разработок¹⁴.

Действительно, поскольку оценка трудового вклада требуется для стимулирования трудовой активности членов коллектива, необходимо, чтобы эти оценки воспринимались как справедливые, каждый сотрудник должен быть уверен, что оценка его труда адекватна и повышение его трудовой активности приведет к пропорциональному повышению оценки его вклада, — лишь в этом случае стимулирующая роль таких оценок будет максимальна. Возникает вопрос, зачем разрабатывать сложные методы оценки трудового вклада работников кем-то, доказывая их обоснованность ссылкой на мнение членов трудового коллектива, когда проще получить прямые оценки трудового вклада каждого работника через взаимные оценки членов коллектива?

На мой взгляд, наиболее обоснованными методами определения трудового вклада работников являются взаимные оценки членов трудового коллектива, а наиболее веским критерием совершенства системы стимулирования — общественное мнение членов коллектива о справедливости данной системы. Более того, для целого ряда видов труда взаимные оценки являются не только наиболее обоснованными, но и вообще единственным методом оценки трудового вклада. Другими словами, эталонным критерием для валидации методов измерения трудового вклада являются взаимные оценки членов трудового коллектива. Методы оценки трудового вклада, основанные на учете взаимных оценок, могут использоваться для совершенствования показателей трудового участия при внедрении коллективных форм организации и оплаты труда (замечу, что для рабочих мест с трудно измеримыми результатами труда — инженеры, учителя, научные работники и т. п. — указанный способ является едва ли не единственным возможным путем внедрения коллективных форм оплаты труда), для улучшения организации социалистического соревнования, усиления его связи с мерами по совершенствованию хозяйственного механизма.

Отмечу еще одну, тесно связанную с рассмотренными выше область использования взаимных оценок — это под-

¹⁴ «Вместе с тем после подведения итогов работы за первый месяц эксперимента стало очевидно, что в его условиях существует большая недоработка. Некоторые инженеры, выполнявшие наиболее сложную и ответственную работу и справедливо считавшиеся наиболее сильными (разрядка моя. — В. П.), попали в число отстающих... Это вызвало недоумение и обиды. Потребовалось срочное принятие мер» [Якушев, 1977, с. 154].

бор и расстановка кадров. Успех внедрения новых форм хозяйствования, их действенность во многом зависит от степени привлечения трудящихся к управлению производством. Так, работа совсов бригадиров, участие рабочих в совете бригады, решающем не только вопросы распределения вознаграждений, но и связанные с производственными заданиями вопросы — все это развивает ответственность за свою бригаду, участок, предприятие, способствует формированию активной жизненной позиции личности¹⁵. Действенность участия трудящихся в управлении производством, на мой взгляд, может существенно повыситься при широком использовании коллективных взаимных оценок при формировании резерва руководящих работников, более полного учета общественного мнения в вопросах подбора и расстановки кадров.

Учитывая адекватность метода взаимных оценок трудового вклада задачам, решаемым в социо-инженерных разработках, этот метод можно считать наиболее обоснованным для данной области приложения¹⁶.

Перейдем теперь к особенностям измерения трудового вклада при проведении массовых социологических исследований. Необходимость разработки методики таких измерений возникла в ходе выполнения плановой темы отдела социологических исследований труда Института философии АН УССР. Одной из задач темы было выяснение соответствия трудового вклада и удовлетворенности трудом у выходцев из разных социальных типов семей, занятых на различных типах рабочих мест¹⁷. Эмпирический этап исследования предусматривал проведение репрезентативного для Киева опроса занятого населения.

Сложность заключалась в том, что для получения оценки трудового вклада необходим опрос по месту работы респондента (поскольку нужны межличностные взаимные оценки членов трудового коллектива), то есть исследование

¹⁵ Чувство хозяина страны. — Коммунист, 1981, № 3, с. 3—13.

¹⁶ Методика оценки трудового вклада инженеров для подведения итогов социалистического соревнования изложена в отчете о научно-исследовательской работе, выполненной в соответствии с договором о творческом сотрудничестве между Институтом философии АН УССР и Киевским технологическим институтом легкой промышленности (Тема № 846: «Совершенствование социально-психологического климата трудового коллектива УкрПТКИгидроспецстрой на основе повышения эффективности соцсоревнования», номер государственной регистрации 0181.4014845, 1983 г.).

¹⁷ Тема «Влияние социального типа семьи на трудовое самоопределение молодого поколения» выполняется с 1983 по 1985 г., научный руководитель темы — В. Ф. Черноводенко.

должно было опираться на производственную выборку. Но построение репрезентативной для Киева производственной выборки чрезвычайно сложно в организационном отношении: нужно получить сведения по ряду признаков обо всех предприятиях и учреждениях города, провести их типологию, осуществить отбор предприятий и учреждений, на которых будет проводиться опрос, а затем обеспечить доступ социологов на каждое из отобранных предприятий, вступая в длительные переговоры с руководством предприятий или министерств и ведомств, которым они подчинены.

Территориальная же выборка, предполагающая отбор респондентов (точнее их адресов) по избирательным спискам и опрос по месту жительства, во-первых, достаточно проста в организационном отношении, во-вторых, основана на случайном, а не на районированном отборе, что позволяет более точно оценить ошибку выборки, в-третьих, дает возможность сократить затраты путем комбинирования интервью с почтовым и телефонным опросом [Рукавишников, Паниотто, Чурилов, 1984, гл. 4, с. 109—130]. Для того, чтобы заменить производственную выборку территориальной, нужно получить адекватную информацию о трудовом вкладе от самого респондента, разработать тестовую методику определения трудового вклада.

Понятие трудового вклада широко употребимо и в практике и в научной литературе, поэтому валидизация методики осуществлялась эмпирическим способом. Методом конечной инстанции для определения трудового вклада являются взаимные межличностные оценки, они непосредственно и использовались в качестве критерия валидизации. Поскольку методика предназначена для занятого населения Киева, то группа валидизации тоже должна была бы представлять собой выборку из генеральной совокупности занятого населения, но построение такой выборки связано с теми же трудностями, что и построение выборки для всего исследования (только объем меньше). Если идти таким путем, то не нужна и тестовая методика, ее разработка обойдется не намного дешевле, чем проведение всего исследования. Но есть еще один способ — метод контрастных групп, то есть выделение групп, находящихся в некотором смысле на полюсах воображаемого континуума, либо просто двух или больше групп, существенно различающихся между собой. Если удастся разработать методику, которая даст сходные результаты на контрастных группах, то можно ожидать, что приблизительно тот же результат будет получен и для всех других групп.

При разработке нашей методики использовались две группы валидизации, то есть было проведено два исследования — трудового вклада инженеров проектного института и рабочих строительных бригад. Рассмотрим одно из них.

Исследование труда инженеров проектного института осуществлялось в 1981 г. путем сплошного опроса сотрудников 28 отделов Украинского проектно-технологического института «Гидроспецстрой», расположенных в Киеве, Одессе, Днепропетровске и Харькове (всего опрошено 285 человек)¹⁸. Использовался групповой социометрический опрос. Анкетер выдавал каждому респонденту из данного отдела (в случае, если отдел размещался в нескольких комнатах, сотрудники собирались в одной из них) анкету и список членов отдела без номеров. Затем он просил кого-либо из членов отдела продиктовать остальным произвольные номера, а сам выходил на это время. Таким путем обеспечивалась анонимность опроса по отношению к исследователю. Анонимность ответов данного респондента по отношению к коллегам обеспечивалась тем, что анкеты вручались непосредственно анкетеру, который давал соответствующие гарантии. Анкета содержала ряд социометрических вопросов, а также ряд суждений и незаконченных предложений, на которые, предположительно, лица с высоким и низким трудовым вкладом будут отвечать по-разному, это и есть исходный материал для создания теста (прил. 1).

Основной социометрической вопрос, использованный для расчета эталонной оценки трудового вклада, был сформулирован как просьба оценить работу каждого сотрудника, поместив его номер в одну из клеток таблицы, состоящей из колонок для указания номеров лиц, работающих с очень высокой, высокой, средней, низкой и очень низкой эффективностью (анкета прил. 1, вопр. 42—46). Фактически этот вопрос объединяет пять социометрических вопросов, результатом являются пять социоматриц. По каждому из вопросов для каждого респондента рассчитывался социометрический статус¹⁹ St_{ij} , где i — номер респондента, j —

¹⁸ Руководители исследования — С. А. Матяш и В. И. Паниотто. Наряду с задачей разработки методики оценки трудового вклада изучалась эффективность социалистического соревнования и вопросы совершенствования социально-психологического климата. Результаты исследования изложены в упомянутом отчете.

¹⁹ Социометрический статус респондента по данному критерию (вопросу) определяется как количество полученных им выборов, разделенное на максимально возможное число полученных выборов [см., например: Словарь прикладной социологии, 1984, с. 268].

номер вопроса ($j=1, 2, 3, 4, 5$). На основании полученных данных каждому респонденту присваивался индекс, характеризующий его трудовой вклад, вычисляемый по формуле:

$$St_i = \frac{1}{5} (St_{i_1} + 0,5 St_{i_2} - 0,5 St_{i_4} - St_{i_5}), \quad (4.15)$$

где St_{i_1} — статус i -го респондента по вопросу «работает с очень высокой эффективностью», St_{i_2} — «с высокой», St_{i_3} — «со средней», St_{i_4} — «с низкой», St_{i_5} — «с очень низкой эффективностью». Коэффициент при St_{i_3} равен 0, поэтому St_{i_3} отсутствует в формуле.

Введенный индекс, изменяющийся от -1 до $+1$, использовался для валидации тестовой методики как эталонная оценка трудового вклада.

Суждения, являющиеся материалом для создания теста, включают незаконченные предложения (анкета прил. I, вопр. 64—76; ответы кодировались в соответствии с прил. IБ), а также некоторые утверждения (вопр. 95—117).

Следующий шаг — поиск суждений, высоко коррелирующих с показателем (4.15). Поскольку степень согласия с суждениями измерялась по ранговой шкале и ответы на вопросы, сформулированные в виде незаконченных предложений, тоже представляли определенные ранги (почти для всех предложений), использовался ранговый коэффициент корреляции²⁰.

Полученные результаты меня крайне разочаровали: все коэффициенты либо были незначимыми либо значимыми, но очень небольшими (0,1 либо 0,2 — не выше). Допуская некоторую нестрогость, провожу регрессионный анализ — с тем же успехом: коэффициент множественной корреляции всех суждений с индексом трудового вклада близок к нулю, коэффициент детерминации (доля объясненной дисперсии в общей дисперсии) равен 4%. Использование алгоритма выбора наилучшего подмножества показывает, что можно оставить лишь одно-два суждения и в результате иметь почти ту же долю объясненной дисперсии. Пытаюсь провести факторный анализ — получаю совершенно неинтерпретируемую картину. Выделялось четыре фактора, объясняющие 23%, 15%, 10% и 9% дисперсии соответственно, но признаки объединялись в факторы каким-то нелепым образом. Впрочем, и регрессионный и факторный анализ основаны на корреляциях, а поскольку

²⁰ Обработка данных осуществлялась на ЕС-1022, ЕС-1033 с помощью пакетов статистических программ «Соцнометрия» [Паниотто, Яковенко, 1980, с. 201—209] и «Паула» [Битинас, Паулавичюс, Пони-карова, 1980].

они преимущественно незначимы, то можно было ожидать, что таким путем выделить суждения, позволяющие «восстановить» эталонную оценку, не удастся.

Таким образом, что в целом между суждениями и критерием валидации нет линейной связи (использованные коэффициенты корреляции измеряют степень линейности связи). Это означает, что либо неудачен выбор суждений, либо связь нелинейна, либо в целом связи нет и только

Таблица 4.5

Трудовой вклад	Отношения с руководителем					всего опрошено
	очень хорошие	хорошие	средние	плохие	очень плохие	
высокий						
абс.	7	67	13	8	2	112
%	6,2	59,8	11,6	7,1	1,8	100
средний						
абс.	4	80	33	10	4	173
%	2,3	46,2	19,1	5,8	2,3	100

отдельные варианты ответа лица с высоким и низким трудовым вкладом выбирают по-разному.

Дальнейший анализ показал, что справедливо последнее предположение. Интервал изменения индекса трудового вклада (4.15) был разбит на два участка: выше 0,01 («хорошие работники») и ниже 0,01 («средние работники» — в эту группу попали работники, имеющие значение индекса близкое к 0, так как отрицательные значения индекса почти не встречались). Таким образом был как бы создан новый вопрос «трудовой вклад» с двумя значениями — высоким и средним. Затем были найдены таблицы двухмерного распределения данного признака со всеми остальными признаками анкеты, в том числе с тестовыми суждениями. Оказалось, что для некоторых суждений по отдельным пунктам были получены значимые различия. Приведу пример того, как при отсутствии связи в целом существуют значимые различия между отдельными вариантами ответа (табл. 4.5).

Расчеты показывают, что на 5 %-м уровне значимости существенно лишь различие ответов тех, кто расценивает свои отношения с руководителем как хорошие: среди лиц с высоким трудовым вкладом их приблизительно на 14% больше. Из таблицы видно также, что шкала отношений с руководителями используется неравномерно, большинство ответов приходится на 2-ю и 3-ю градации («хорошие» и «средние»), что может быть причиной низкого коэффи-

ента корреляции. Всего суждений, по отдельным пунктам которых были значимые различия, оказалось 17. По этим суждениям и строился тест на трудовой вклад.

По каждому из 17 суждений каждый респондент получал балл (1 или 0) в зависимости от того, соответствуют ли его ответы значимым различиям. Например, если респондент оценивал отношения с руководителем как хорошие, то его ответу присваивается 1 балл, так как такой ответ характерен для лиц с высоким трудовым вкладом (табл. 4.5), во всех остальных случаях — 0 баллов. Как мы условились, будем называть такой ответ релевантным высокому значению измеряемого признака или просто *релевантным*; *нерелевантным* назовем ответ, соответствующий низкому значению измеряемого признака, то есть ответ, за который присваивается балл 0 (при этом все остальные ответы на данный вопрос получают балл 1). Приведем перечень релевантных и нерелевантных ответов для данного теста:

Релевантные ответы													
Номер вопроса	5	22	64	65	66	68	69	73	74	77	97	106	109
Номер ответа	4	5	4	4	5	3	4	3	2	6	4	1	3

Нерелевантные ответы				
Номер вопроса	1	9	21	66
Номер ответа	2	3	1	4

Сумма баллов, полученных данным респондентом и является тестовой оценкой его трудового вклада.

Сопоставление тестовой оценки с критерием валидизации дает $\alpha=0,3$ (табл. 4.1; этот пример был использован при выведении показателя α). Коэффициент невысокий, требует увеличения выборки в 11 раз, но, тем не менее, для поставленных целей пригоден. Планируемая при городских опросах выборка составляет около 3—4 тысяч респондентов, поэтому при не очень большом дроблении массива сравнение осуществимо.

Следующий этап, который может понадобиться при разработке теста, — его редуцирование. Может оказаться, что отбросив значительное число суждений, мы не очень существенно ухудшим характеристику теста либо что условия опроса требуют существенного уменьшения числа суждений. Обычно полный тест сопровождается одним или несколькими сокращенными, редуцированными вариантами, которые дают показатель, близкий к полному тесту. Для этого желательно иметь программу для ЭВМ, выбирающую наилучшее подмножество суждений. Поскольку в на-

шем программном обеспечении такие программы есть лишь для коэффициентов регрессии, а не коэффициентов валидации, пришлось рассчитывать связь каждого вопроса с критерием валидации и отбирать их по одному. При этом итогом является не оптимальное, а просто достаточно хорошее решение. Для теста на трудовой вклад редуцирование привело к выделению 6 суждений:

	Релевантные ответы			Нереле- вантные ответы		
Номер вопроса	5	65	74	1	9	21
Номер ответа	4	4	2	2	3	1

Сумма баллов, полученных во всех 6 вопросах каждым респондентом, разбивалась затем на два участка (меньше или равна 3; больше 3), принадлежность к которым и является показателем высокого или среднего трудового вклада. Характеристика полученного теста дана в таблице 4.6.

Таблица 4.6

Трудовой вклад по оценке			
тестовой	эталонной		итого
	высокий	средний	
высокий %	52,1	47,9	100
средний %	26,9	73,1	100

Коэффициент валидности уменьшился в редуцированном тесте до 0,25. По формуле (4.11) исследователь может оценить, устраивает ли его редуцированный вариант или ему нужен полный. Экстраполируя данные таблицы 4.3, получаем, что при использовании редуциро-

ванного теста значимость различий уменьшается в 0,23—0,25 раза и для полной компенсации выборку нужно увеличить в 16—17 раз. В реальных исследованиях возможна, вероятно, лишь частичная компенсация, поэтому применение теста даст возможность выявить лишь наиболее значимые различия.

Рассмотрим теперь второе исследование — трудового вклада рабочих строительных бригад. Исследование проводилось в 1984 г. в городе Белая Церковь в рамках изучения эффективности бригадного подряда²¹ в тресте «Хим-

²¹ Исследование проводилось по заданию Киевского обкома Компартии Украины сотрудниками отдела социологических исследований труда Института философии АН УССР Оксамитной С., Тарасенко В., Яковенко Ю. (руководитель — В. Паниотто).

строй» и Домостроительном комбинате. Предполагалось провести сплошной опрос, но в силу ряда организационных причин (в частности, отсутствия значительного числа бригад в городе — временно они находились в других городах области) были опрошены преимущественно находящиеся в Белой Церкви в момент опроса бригады треста «Химстрой» и несколько бригад Домостроительного комбината — всего 32 бригады, 348 респондентов (впрочем, репрезентативность, как уже указывалось, не имеет значения для наших целей).

Специфика объекта исследования внесла некоторые изменения в процедуру опроса. Пилотаж показал, что рабочие-строители заполняют анкету значительно дольше, чем инженеры. Причем условия опроса не располагали к длительным беседам — опрос проходил зимой, рабочих собирали в «бытовках» во время работы или перерыва — и тесно, и времени нет, и большой перепад между тем, чем они только что занимались, и тем, что от них требовалось во время опроса, а тут нужно собраться, подумать. Поэтому, во-первых, анкета была сокращена на треть, во-вторых, после инструкции, ответа на вопросы респондентов и заполнения некоторых сложных пунктов в присутствии анкетера анкету (вместе со списком членов группы) отдавали заполнять дома, в-третьих, обеспечение анонимности социологической части опросов проводилось только в тех бригадах, в которых кто-то из респондентов выразил такое желание (предварительно объяснили, зачем это нужно). Упрощена была также социометрическая часть: основными вопросами, по которым рассчитывался эталонный показатель, были 48-й и 49-й вопросы (прил. 2): «Кто из членов бригады работает, на Ваш взгляд, лучше всех? А кто хуже всех?» В остальном исследование и его обработка осуществлялось таким же образом, как предыдущие.

Корреляционный, регрессионный и факторный анализ не дал нужных результатов. Как и в предыдущем случае рассматривались таблицы двумерного распределения, отбирались вопросы, по отдельным пунктам которых были значимые расхождения, и по ним строился тест. При релевантных, соответствующих значимым расхождениям ответах, типичных для лиц с высоким значением измеряемого признака, респондентам присваивался 1 балл, в остальных случаях — 0 баллов; в некоторых случаях выделялись нерелевантные ответы, типичные для лиц с низким трудовым вкладом, за которые присваивалось 0 баллов, а в остальных случаях — 1 балл.

В конечном итоге было отобрано 5 вопросов (прил. 2)

Релевантные ответы

Номер вопроса	42	43	74	91	93
Номер ответа	5	5	4	5	3

Для каждого респондента находится сумма баллов (она может варьировать от 0 до 5): если сумма больше или равна 3 — трудовой вклад респондента высокий, если меньше 3 — средний. Полученный тест описывается таблицей 4.7.

Таблица 4.7

Трудовой вклад по оценке			Итого
тестовой	эталонной		
	высокий	средний	
высокий %	56,3	43,7	100
средний %	14,3	87,5	100

Коэффициент валидности теста, $\alpha=0,42$, довольно высокий (полная компенсация тестовой ошибки достигается приблизительно 6-кратным увеличением объема сравниваемых групп; интерпретировать 15%-ные различия как значимые мы смо-

жем при 500 респондентах в группе, а 20%-ные различия — при 300 респондентах в каждой группе). Достаточно высоки и показатели связи ($Q=0,77$; $\Phi=0,35$; $\chi^2=31,1$), хотя в данном случае они нас не интересуют.

Итак, разработаны два теста — один для инженеров (точнее, в группе валидизации было около 70% инженеров, остальные — техники и др.), другой для рабочих-строителей. Эти тесты могут использоваться лишь для соответствующих категорий респондентов. Следующий этап — разработка теста, пригодного для обеих групп. Для этого были выделены суждения, общие в обоих исследованиях, и такие, по которым были получены значимые различия в ответах по отдельным пунктам между группами с высоким и средним трудовым вкладом, выделенным с помощью эталонного критерия. Таких вопросов оказалось 8 (прил. 3):

Исследование труда инженеров	137	146	146	149	42	46	50	125
Исследование труда рабочих	74	75	75	77	10	15	18	66
Если респондент выбрал вариант ответа	4	4	5	5	5	4	5	5
то присваивается балл	1	0	0	1	1	1	1	1
в противном случае присваивается балл	0	1	1	0	0	0	0	0

По выделенным вопросам, как и раньше, находились суммы баллов для каждого респондента, разбивались на

два интервала и затем рассчитывалась таблица связи показателя теста с эталонной оценкой. Были получены следующие результаты.

Исследование трудового вклада инженеров. Тестовая сумма баллов разбита на два интервала — больше или равно 4 («высокий трудовой вклад») и меньше 4 («средний трудовой вклад»). Характеристика полученного теста приведена в таблице 4.8. Показатели связи — $Q=0,41$; $\Phi=$

Таблица 4.8

Трудовой вклад по оценке			
тестовой	эталонной		итого
	высо- кий	сред- ний	
Высокий %	53,0	47,0	100
Средний %	51,8	68,3	100

Таблица 4.9

Трудовой вклад по оценке			
тестовой	эталонной		Итого
	высо- кий	сред- ний	
Высокий %	38,3	61,7	100
Средний %	13,8	86,2	100

$=0,21$; $\kappa^2=12,4$. Коэффициент валидности $\alpha=0,21$, то есть невысок, нужны большие выборки.

Исследование трудового вклада рабочих-строителей. Тестовая сумма баллов разбита на два интервала — больше или равна 4 («высокий трудовой вклад») и меньше 4 («средний трудовой вклад»). Характеристика теста приведена в таблице 4.9. Коэффициент валидности $\alpha=0,245$, несколько выше, чем предыдущий, но тоже низкий.

Итак, разработано три теста на трудовой вклад — валидный при исследовании инженеров, валидный при исследовании рабочих-строителей и общий тест. Тест, валидный для рабочих-строителей, имеет сравнительно высокий коэффициент валидности 0,42, что позволяет интерпретировать 20 %-ные различия как значимые при 300 респондентах в каждой группе. Тест, валидный для инженеров, имеет коэффициент валидности 0,30, что позволяет интерпретировать 20 %-ные различия как значимые при 550 респондентах в каждой из сравниваемых групп. Наконец, тест, валидный для обеих групп, имеет низкий коэффициент валидности, но не ниже, чем 0,21, что позволяет интерпретировать 20 %-ные различия как значимые при 550 респондентах в каждой из сравниваемых групп. Поскольку группы инженеров и рабочих-строителей существенно различаются и по содержанию и условиям труда, и по уровню образования, их можно рассматривать как контрастные группы (служащие низкой квалификации, вероятно, меньше отличаются от инженеров, чем рабочие-строители, а рабочие про-

мышленных предприятий меньше отличаются от рабочих-строителей, чем инженеры проектных институтов). Разработанный тест может, таким образом, использоваться для всех групп работающего населения с коэффициентом валидности не ниже чем 0,21. Хотя использование теста с низкой валидностью требует существенного увеличения массива опрошенных (в исследовании, проводимом нашим отделом, необходимо обеспечить репрезентативность для 6 типов семьи, поэтому объем выборки должен быть не менее чем 750×6 , то есть 3—4 тысячи), принятая нами технология опроса путем сочетания почтового опроса и интервью [Семья и воспроизводство..., 1984] позволяет это осуществить без особых сложностей.

Приведенные результаты показывают принципиальную возможность оценить трудовой вклад по суждениям респондентов. До проведения исследования это было далеко не очевидно: во-первых, механизмы психологической защиты препятствуют правильной самооценке (в частности для лиц с низким трудовым вкладом), во-вторых, процедура и инструмент опроса вносят ошибки в измерение самооценки даже в том случае, если она правильная. Оказалось, однако, что хотя прямая самооценка (анкета, прил. 1А, вопр. 73) не очень тесно связана с трудовым вкладом, по ряду других суждений можно различить лиц с высоким и низким трудовым вкладом. Это говорит либо о том, что трудовой вклад осознанно или подсознательно все же оценивается респондентом достаточно адекватно, либо о том, что величина трудового вклада связана, например, со структурой ценностей личности, что и приводит к зафиксированным выше различиям в ответах лиц с разным по величине трудовым вкладом.

Актуальна задача дальнейшего совершенствования теста, ведь увеличение коэффициента валидности на 0,1 позволяет в некоторых случаях вдвое уменьшить объем выборки. Требуется проанализировать удачные и неудачные суждения, включенные в анкету, и использовать полученные результаты для разработки новых суждений. Проведение серии не очень больших методических экспериментов позволяет разработать тест, который потом можно будет использовать во многих исследованиях. Создание хороших тестов с высоким значением α , очень эффективно: затраты на создание теста должны с лихвой окупиться при проведении массовых исследований. Рассмотрим еще один пример обеспечения валидности теста.

Тест для измерения доброты. Доброта, чуткость, отзывчивость, способность к эмпатии, альтруизм и другие поня-

тия описывают близкие по смыслу характеристики личности, а возможно различные проявления одной характеристики, которая входит в ядро простых норм человеческой нравственности. По частоте употребления характеристик личности в русском языке слово «добродетель» занимает одно из первых мест [Хмелько, 1981 (6), с. 124—134] (первые три слова русского языка, используемые для характеристики положительных качеств людей: ум — 371; добродетель — 286, героизм — 241). Данная характеристика является основой для формирования целого ряда норм коммунистической морали, коллективизма и товарищеской взаимопомощи, гуманных отношений и взаимного уважения между людьми.

Разработка теста производилась в тех же исследованиях, что и разработка теста на трудовой вклад. Наряду с социометрическими вопросами о трудовом вкладе задавались социометрические вопросы о добродетели (прил. 1, вопр. 30, 31; прил. 2, вопр. 50). Например: «Кому из членов Вашей бригады в наибольшей степени присущи добродетель, чуткость, отзывчивость?» В работе В. Е. Хмелько [1981 (6)] эти три качества были доминантами в констелляции качеств, включающих добродетель. Далее вся процедура была такой же, как при разработке теста на трудовой вклад, — коэффициенты корреляции ответов на суждения и социометрического статуса по эталонному вопросу оказались низкими, поэтому рассматривались таблицы двумерного распределения, выделялись значимые по отдельным пунктам различия и т. д. В итоге было получено два теста — для инженеров и для рабочих-строителей.

Характеристики теста для инженеров приведены в таблице 4.10. При построении теста были отобраны следующие вопросы (прил. 1А):

Т а б л и ц а 4.10

Доброта по оценке			итого
тестовой	эталонной		
выражена в степени	выражена в высокой степени	выражена в средней степени	
Высокой абс. %	166 67,5	80 32,5	246 100
Средней абс. %	10 25,6	29 74,4	39 100

Релевантные
ответы

Нерелевант-
ные ответы

Номер
вопроса
Номер
ответа

67 74
3 2

69 71 74
0 0 0

Как видим, 74-й вопрос используется дважды. Здесь возникает следующая сложность: что делать, если ответ не является ни релевантным ни нерелевантным? Если ответ не равен 2, нужно присваивать 0, а если не равен 0 — нужно присваивать 1. Поэтому в случае использования одного и того же вопроса дважды ответам, не являющимся ни релевантными, ни нерелевантными, присваивается балл 0,5. В качестве тестовой оценки доброты была принята сумма баллов, полученная за ответы на приведенные вопросы. Интервал изменения суммы разбит на два участка: выше 1 («высокая степень доброты») и ниже или равно 1 («средняя степень доброты»). Эталонный показатель — социометрический статус — тоже разбит на два участка: ниже 0,05 («средняя степень доброты») и выше или

Таблица 4.11

Доброта по оценке			итого
тестовой (выражена в степени)	эталонной (выражена в степени)		
	высокой	средней	
высокой			
абс.	40	22	62
%	64,5	35,5	100
средней			
абс.	65	119	184
%	35,3	64,7	100

равно 0,05 («высокая степень доброты»). Коэффициент валидности — $\alpha=0,42$ — сравнительно высокий.

Характеристика теста для рабочих-строителей приведена в таблице 4.11. При разработке теста были отобраны вопросы:

	Релевантные ответы							Нерелевантные ответы	
Номер вопроса	74	87	91	95	96	121		80	95
Номер ответа	4	5	5	1	1	2		5	3

Тестовая оценка доброты — сумма баллов, полученных за ответы на эти вопросы. Интервал изменения разбит на участки: меньше или равно 4 («средняя степень доброты») и больше 4 («высокая степень доброты»). Коэффициент валидности — $\alpha=0,29$, — средний.

Общий тест построить не удалось, так как оказалось, что в обеих анкетах мало «перескакающих» суждений. Это связано с тем, что получив негативные результаты по пунктам первой анкеты, относящимся к доброте (незначимые и малозначимые коэффициенты корреляции), я решил, — как оказалось, ошибочно, — что эти суждения вообще не годятся и поэтому не включал их в анкету для рабочих.

Приведенные примеры позволяют уточнить интерпретацию тестовых показателей — прежде всего, различие теоретически и эмпирически обоснованных тестов. Предположим, исследователь разработал теоретически обоснованный тест на доброту, ввел теоретическое понятие, дав соответствующее определение (например: «доброта — это готовность пожертвовать своим интересом в пользу интересов других»), провел редукцию понятия (например: определил понятие интереса, разработал типологию интересов и т. д.), нашел эмпирические референты (например: «готовность пожертвовать» эмпирически выражается в том, согласен ли человек отдать больше, чем получает; затем в соответствии с типологией интересов определяется, что именно готов отдать — время, материальные ценности и т. п.) и наконец, провел исследование и получил данные о доброте, точнее о распределении тестового показателя доброты в исследуемой совокупности.

Как их можно интерпретировать? Что, например, означает высокое значение показателя доброты у данного индивида или большая доля лиц, имеющих высокое значение показателя, в данной группе?

Высокое значение полученного показателя при корректном построении методики означает, что данный индивид (группа) имеет высокое значение какой-то характеристики, которую исследователь назвал добротой, то есть измерена доброта в понимании исследователя. В рамках его теории этот показатель может быть интерпретирован, если исследователь докажет его связь с другими понятиями этой теории. Если же, например, исследователь полагает, что доброта является фактором стабилизации трудового коллектива, улучшения психологического климата, то он должен показать, что доля лиц с высоким показателем доброты в коллективе положительно скоррелирована со средним стажем работы членов коллектива, отрицательно — с показателями текучести, числом конфликтов и т. п.

В отличие от теоретически валидного теста, высокое значение показателя в эмпирически валидном тесте, например показателя доброты в построенном тесте для инженеров, означает, что измерена доброта в понимании инженеров. Если, например, при приеме на работу респондент получил высокое значение по этому тесту, то с вероятностью, близкой к 0,7, то есть приблизительно в 70% случаях (табл. 4.10), будущие коллеги сочтут, что доброта выражена у него в высокой степени. Это, разумеется, лишь пояснение интерпретации теста, так как тест не предназначен для оценки индивида.

Преимущество теоретически валидного теста — в том, что он дает именно ту информацию, которая нужна для теории исследования, они как бы валидны по определению (при корректной процедуре перехода от понятия к эмпирическим референтам). Недостаток же заключается в фактическом отсутствии методов контроля перехода от теоретического к эмпирическому уровню.

Преимущество эмпирически валидного теста — в контролируемости всей процедуры разработки теста, сближении языка исследователя и респондента, в четкой интерпретации результатов тестирования. Недостаток — в сложности определения эталонного критерия и контроля возникающих смещений.

Коррекция смещений эталонных оценок. Эталонные оценки, хотя и являются абсолютно валидными (по определению), могут быть ненадежными, поэтому требуется контроль и коррекция возможных случайных и систематических ошибок. Рассмотрим возможные смещения взаимных оценок в малой социальной группе.

Сначала об устойчивости информации. Обзор 53 исследований, включавших проверку социометрических данных на устойчивость (отсутствие случайных ошибок) показал, что межличностные оценки довольно стабильны²². Еще более стабильны обобщенные характеристики, в частности показатели социометрического статуса, которые нас интересуют прежде всего. Коэффициенты корреляции между статусами, полученными членами группы в двух исследованиях, проведенных с интервалом во времени, и в исследованиях, проведенных методом расщепления (рассматривались отдельно оценки, данные одной и другой половиной группы) в большинстве были не ниже, чем 0,85—0,90 и достигали 0,95—0,97.

Что же касается правильности информации, то здесь дело обстоит сложнее. На оценки, которые один человек дает другому, оказывает влияние эмоциональное отношение оценивающего к оцениваемому. Здесь проявляется то, что называется субъективизмом и вызывает недоверие к такого рода оценкам. Многочисленные эксперименты, проведенные у нас в стране и за рубежом, показывают, что тем людям, которые нам нравятся, мы приписываем положительные черты, а тем, которые не нравятся, — отрицательные. Например, красивой женщине мы приписываем

²² Moulton J. S., Blake R. R., Fruchter B. The reliability of sociometric measures. — In: The Sociometry Reader. New York, 1960, p. 320—362.

просоциальные, а некрасивой — асоциальные черты²³. В том случае, когда поступки оцениваемого не соответствуют нашим представлениям и нашим эмоциональным отношениям с оцениваемым, «услужливый» рассудок помогает устранить противоречия: хорошие действия тех, кто нам нравится, мы объясняем их диспозициями, а плохие — обстановкой, ситуацией, выпудившей к таким поступкам. Те, кто нам не нравится, оцениваются иначе: их хорошие поступки объясняются ситуацией, а плохие — диспозициями [Regan, Shaus, Fasio, 1974, p. 385—387].

Данные опроса инженеров-проектировщиков дают возможность проверить существенность влияния симпатии на оценки доброты, трудового вклада и другие межличностные оценки членов малой группы. В анкете, с помощью которой проводился опрос, кроме всего прочего были вопросы, по которым можно судить об интенсивности симпатии одних членов группы к другим (прил. 1, вопр. 48, 50, 51, 54, 55, 56, 60, 61). Это обычные социометрические вопросы, в которых респонденту в проективной ситуации предлагается отобрать несколько членов группы по заданному критерию. Например: «Предположим, что судьба забросила Вас на необитаемый остров. Кого бы Вы взяли с собой, если бы можно было взять несколько человек и если — одного?»

Рассмотрим связь вопросов, характеризующих симпатию, с оценками различных характеристик членов группы (прил. 1, вопр. 24—37). Поскольку и те и другие вопросы являются социометрическими, корреляционный анализ здесь неприменим. Для этого случая я предлагаю использовать аналог коэффициента корреляции — коэффициент импlications r_{AB} , который можно интерпретировать как вероятность того, что наугад взятый член группы назовет члена группы в ответе на вопрос А, если он назвал его в ответе на вопрос В [Паниотто, 1975, с. 34]. Для каждого из 8 вопросов, характеризующих симпатию, были найдены коэффициенты импlications, показывающие определенную несимметричную связь этих процессов с вопросами о качествах личности, а затем рассчитывался средний коэффициент для всех вопросов о симпатии (табл. 4.12).

Как видно из таблицы, показатели достаточно высокие, что говорит о явной связи интенсивности симпатии и суждений о качествах личности. Наиболее высокая степень

²³ Панферов В. Н. Восприятие человека в межличностном общении. — В кн.: Проблемы психологического воздействия. — Иваново, 1978, с. 67—74.

Таблица 4.12

Средний по вопросам о симпатии коэффициент импликации, характеризующий связь этих вопросов с вопросами о качестве личности:

профессиональный уровень	культурный уровень	ум, понятливость	добродота	трудолюбие	порядочность	личное обаяние, привлекательность
0,36	0,29	0,40	0,47	0,40	0,40	0,34

связи с оценками доброты, наиболее низкая — с оценкой культурного уровня. Это, однако, лишь косвенное свидетельство возможных смещений оценок, так как возможно, что эти характеристики личности являются факторами, определяющими интенсивность симпатий. Поэтому целесообразна прямая проверка. Рассмотрим, как это осуществить.

В соответствии с разработанной ранее методикой [Паниотто, 1975, с. 54, форм. (4)] для каждого члена группы (i -го и j -го) рассчитывались интенсивности симпатий i -го к j -му члену группы, то есть строилась матрица симпатий S с элементами s_{ij} ($i, j = 1, 2, \dots, N$; N — число членов группы). Отмечу, что шкала для измерения интенсивности симпатий является теоретически валидной. Уровень измерения — интервальный, то есть получается своего рода единица измерения симпатии (назвать ее можно, например, Дон Жуаном или просто Жуаном). Полученная матрица сопоставлялась с каждой из матриц оценок $\|d_{ij}\|$ ($d_{ij} = 1$, если i -й член группы считает, что данная характеристика личности выражена у j -го члена группы в высокой степени и $d_{ij} = 0$ в том случае, если это не так). Интервал изменения интенсивности симпатий (от 0 до 20 для всего массива опрошенных) был разбит на 6 участков: $[0; 0,01[$, $[0,01; 0,33[$, $[0,33; 0,66[$, $[0,66; 2,0[$, $[2,0; 5,0[$, $[5,0; 20]$ ²⁴. Далее оценки, полученные j -м респондентом, были разбиты на 6 групп так, чтобы соответствующие им элементы s_{ij} матрицы симпатий принадлежали к выделенным участкам, то есть к k -й группе относились такие d_{ij} , для которых элементы s_{ij} с соответствующими индексами

²⁴ Выбран вариант разбиения, при котором каждый из участков достаточно наполнен. В 1-й участок попадает 1120 элементов всех матриц симпатий s_{ij} , в последующие 348, 107, 134, 130, и 75 элементов соответственно. Здесь, конечно, допущен некоторый произвол, так как перебирались лишь несколько вариантов и критерии отбора вариантов разбиения не были четко сформулированы.

i, j принадлежат участку $[a_k; b_k[$ (a_k, b_k — границы k -го участка; $k=1, \dots, 6$). Эти элементы матриц S и D обозначим через $s_{ij}^{(k)}$ и $d_{ij}^{(k)}$.

Таким образом, для каждого респондента рассматриваются отдельно оценки, которые он получил от тех, кто испытывает к нему симпатию интенсивности 0; от 0 до 0,33 и т. д. Следовательно, каждый респондент получит в итоге 6 оценок:

$$d_j^{(k)} = \frac{1}{n_j^{(k)}} \sum_i d_{ij}^{(k)}, \quad (4.16)$$

где $n_j^{(k)}$ — число элементов $s_{ij}^{(k)}$, принадлежащих k -му интервалу; $s_{ij}^{(k)} \in [a_k, b_k[$, $k=1, 2, \dots, 6$.

Но все сказанное относится лишь к одной группе (одному отделу), а всего исследовалось 28 отделов. Если обозначить номер отдела через l , то $d_{jl}^{(k)}$ будет k -я оценка j -го респондента из l -го отдела, а $n_{jl}^{(k)}$ — число элементов, принадлежащих k -му участку в матрице l -го отдела. Усреднив данные по всем респондентам и всем отделам, получим:

$$d^{(k)} = \frac{1}{N_k} \sum_l \sum_j d_{jl}^{(k)}, \quad (4.17)$$

где $N_k = \sum_l \sum_j n_{jl}^{(k)}$.

Данные расчета для каждой характеристики личности и для каждого из k интервалов приведены в таблице 4.13. Как видим; с увеличением интенсивности симпатии оценки всех положительных характеристик личности существенно возрастают (особенно оценки доброты; в наименьшей степени это относится к оценкам культурного уровня, что соответствует данным таблицы 4.12, но и для этой характеристики рост оценки весьма существенный). Расчет среднеквадратического отклонения показывает, что различия в оценках одной и той же характеристики лицами, испытывающими симпатию разной интенсивности к оцениваемому, даже выше, чем различия в оценках разных характеристик оцениваемого лицами, испытывающими к нему симпатию одного уровня. Другими словами, похоже, что неважно, какими характеристиками ты обладаешь, а важно, как к тебе относятся. В действительности это не совсем так. Здесь все-таки существенно влияет усреднение по лицам, имеющим разные значения оцениваемой характеристики. Кро-

Таблица 4.13 *

Характеристика личности	Средние оценки, полученные респондентами от лиц, испытывающих к ним симпатию интенсивности:								
	0—0,01	0,01—0,33	0,33—0,66	0,66—2,0	2,0—5,0	5,0—20,0	для массива в целом	среднее арифметическое	среднее квадратическое отклонение
профессиональный уровень	0,06	0,22	0,32	0,28	0,40	0,37	0,15	0,27	0,11
культурный уровень	0,04	0,16	0,19	0,29	0,31	0,25	0,11	0,21	0,09
ум	0,06	0,22	0,25	0,26	0,44	0,40	0,15	0,27	0,12
доброта	0,04	0,15	0,22	0,36	0,42	0,57	0,14	0,29	0,18
трудолюбие	0,05	0,22	0,34	0,34	0,38	0,44	0,16	0,30	0,12
порядочность	0,05	0,22	0,21	0,31	0,37	0,42	0,14	0,26	0,12
привлекательность	0,02	0,12	0,23	0,31	0,30	0,37	0,10	0,22	0,12
среднее арифметическое	0,05	0,19	0,25	0,31	0,37	0,40	0,14	0,26	0,12
среднее квадратическое отклонение	0,01	0,04	0,05	0,03	0,05	0,09	0,05	0,03	—

* В колонке «для массива в целом» приведены значения $\frac{1}{N} \sum_k a^{(k)} N_k$, а в следующих колонках — $\frac{1}{k} \sum_k a^{(k)}$ и соответствующее среднее квадратическое отклонение.

ме того, при некоторых других вариантах разбиения интервала изменения симпатии было видно, что при дальнейшем увеличении интенсивности симпатии намечается уменьшение величины оценок положительных характеристик личности, то есть возможно, что зависимость носит $\cap$ -образный характер (эти результаты незначимы, хотя и близки к значимым, поэтому я их не привожу).

Но как бы то ни было, интенсивность симпатии оказывает очень большое влияние на оценку характеристик личности. Возникает вопрос, как же устранить это смещение и кто, собственно, «прав» в своей оценке — те, кто испытывает симпатию нулевой, низкой, средней или высокой интенсивности. Здравый смысл подсказывает, что более правильна оценка тех, кто испытывает симпатию нулевой интенсивности, что они беспристрастны. Но так ли это в действительности? Во-первых, интенсивность симпатии измерена в интервальной шкале, а не в шкале отношений и

поэтому неизвестно, что собой представляет симпатия нулевой интенсивности: действительно ли это полная индифферентность к данному лицу или же, скажем, слабая степень отрицательного отношения. Во-вторых, симпатия означает большее внимание к объекту симпатии, предполагает более близкое с ним знакомство и, возможно, более благоприятные условия для «изучения» тех или иных характеристик личности объекта симпатии. Наконец, если для измеряемых характеристик межличностные оценки действительно являются эталоном, то есть нет более обоснованного способа измерения, то какие основания могут быть у нас, чтобы предпочесть одни оценки другим?

На последнем вопросе хочу остановиться особо, так как он относится не только к данной ситуации, но и вообще к широкому классу измеряемых признаков: ко всем тем случаям, когда трудно отделить изучаемое свойство от процедуры измерения. Например, исследование, проведенное в городе Припяти отделом социально-профессиональной ориентации Института философии АН УССР, показало существенное влияние пола, возраста и некоторых других характеристик интервьюеров на ответы респондентов. Возникает вопрос, какие данные являются более надежными — полученные, скажем, мужчинами или же женщинами? Какие есть основания, чтобы предпочесть те или иные результаты? Еще один пример. Изучая удовлетворенность различными сторонами работы на двух репрезентативных для Киева выборках [Панниотто, Яковенко, 1981] занятого населения с помощью интервью и с помощью почтового опроса, мы получили, что удовлетворенность в почтовом опросе во всех случаях ниже на 10—15%. Мы можем предполагать, что данные почтового опроса более надежны, так как здесь не сказывается присутствие интервьюера, побуждающее завышать удовлетворенность. Однако можно ли это доказать?

Я думаю, что во всех этих случаях есть, по крайней мере, два выхода. Во-первых, попытаться определить, какие данные более надежны при измерении других характеристик, причем таких, для которых есть независимые от примененных эталонные методы (для измеряемых нами характеристик это невозможно, так как по определению наши методы являются эталонами). Например, мы можем определить, как влияет интенсивность симпатий на оценку социометрического статуса, затем определить социометрический статус и установить, при какой интенсивности симпатий члены группы дают наиболее точную оценку. Или же мы можем определить, какие из интервьюеров — муж-

чины или женщины — позволяют получить более надежные оценки зарплаты, стажа и любых других характеристик, которые можно проверить независимым методом. Если данные таких экспериментов дадут убедительные результаты (скажем, во всех случаях женщины получают более надежные результаты), то можно предположить, что и при изучении других характеристик в качестве интервьюеров лучше привлекать женщин (хотя такая экстраполяция тоже может привести к артефактам).

Другой выход — ввести *конвенциональный эталон*, условное начало отсчета. Например, можно считать эталонными данные интервью. Когда же потребуются сравнения, — пересчитывать данные почтового опроса, опираясь на результаты методических экспериментов, так, чтобы результаты были такими же, как в случае, если бы опрос проводился путем интервью (это требование можно было бы ввести как обязательное при публикациях). Можно, наоборот, принять за эталонные данные почтового опроса и пересчитывать данные интервью. Точно так же можно принять за эталонные данные, полученные мужчинами, и пересчитывать данные, полученные женщинами (и наоборот). Аналогично тому, как физики вводят условные эталонные единицы длины, массы, времени.

Для случая, представленного в таблице 4.13, или для «нейтрализации» влияния характеристик интервьюера на ответы респондентов в качестве эталонного условия измерения удобно принять равномерное распределение фактора, вызывающего смещение. Например, эталонным считать случай, когда среди интервьюеров 50% мужчин и 50% женщин, и все реальные ситуации пересчитывать относительно этого эталона.

Пусть фактор смещения имеет k градаций и при каждой градации получает среднее значение измеряемого признака $\bar{x}_k$ (либо доли p_k), рассчитанных для n_k респондентов. Итоговые показатели для всего массива равны:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^k \bar{x}_k n_k, \quad p = \frac{1}{N} \sum_{i=1}^k p_k n_k, \quad N = \sum_{i=1}^k n_k. \quad (4.18)$$

В случае равномерного распределения респондентов по градациям фактора $n_1 = n_2 = \dots = n_k = \frac{N}{k}$, поэтому:

$$\bar{x}' = \frac{1}{N} \sum_{i=1}^k \bar{x}_k \frac{N}{k} = \frac{1}{k} \sum_{i=1}^k \bar{x}_k, \quad (4.19)$$

$$p' = \frac{1}{k} \sum_{k=1}^k p_k. \quad (4.20)$$

Это не что иное, как средние арифметические. Таким образом, для учета смещения доли или среднего значения признака под влиянием некоторого фактора требуется рассчитать среднее значение признака или долю для респондентов, находящихся под влиянием фактора со значением $1, 2, \dots, k$ (даным значением фактора либо обладает сам респондент, либо он находится под воздействием этого фактора), а затем найти среднее арифметическое от полученных долей или средних. В таблице 4.13 сначала приведены данные для массива в целом, а затем в колонке «среднее арифметическое» — скорректированные указанным способом средние оценки. Замечу, что подобную коррекцию целесообразно проводить в том случае, если все группы, выделенные по значениям фактора, достаточно наполнены, то есть превышают некоторый, зависящий от заданного уровня значимости порог.

Полученные результаты позволяют теперь провести пересчет индивидуальных оценок, «навязывая» оценкам каждого индивида коэффициенты, полученные на множестве всех опрошенных

$$\beta_k = \frac{x_k}{x'}. \quad (4.21)$$

Как указывалось ранее (4.15), каждый респондент получает k оценок $d_j^{(k)}$ (в некоторых случаях n_k могут быть равны 0 и соответствующие оценки отсутствуют). Тогда скорректированная оценка j -го респондента d'_j равна:

$$d'_j = \sum_k \beta_k d_j^{(k)}. \quad (4.22)$$

Аналогично рассчитывается скорректированная оценка для p_j .

Другие возможные смещения межличностных оценок связаны с самим механизмом оценивания. Некоторые эксперименты дают основание предположить, что при оценивании респонденты исходят из некоторых критериев-эталонов и оценка может зависеть от того, что принимает за эталон данный респондент. А. А. Кроник [1982, с. 40—42] в зависимости от эталона, используемого индивидом, выделяет, например, 4 типа оценивания: нормативное оценивание, в котором эталоном, выполняющим функцию уста-

новки, являются ожидаемые оценки и мнения со стороны «значимых других»; ценностное, базирующееся на идеалах — образцах желаемого — и анти-идеалах — образцах нежелаемого; социоцентрическое, базирующееся на предыдущих оценках данного лица, и эгоцентрическое, базирующееся на самооценках личности, использующее их как точку отсчета. Например, эксперименты показывают, что испытуемые обращают внимание при оценке других на такие характеристики личности, которыми у себя они больше удовлетворены [Shranger, Patterson, 1974, p. 569—585]. В работе А. А. Кроника приведено определенное эмпирическое подтверждение наличия разных типов оценивания.

Кроме различий в типе оценивания на адекватности оценки может сказываться также ассимилятивно-контрастное восприятие отклонений значения оцениваемой характеристики от точки отсчета. Эта закономерность проявляется в том, что небольшие отклонения недооцениваются (оценка получается заниженной), а большие — переоцениваются (оценка завышается) [Кроник, 1982, с. 42—55].

Существующие пока методы не позволяют надежно определить тип оценивания и ассимилятивно-контрастные отклонения, поэтому задачу коррекции такого рода смещений можно сформулировать как нерешенную. Можно предположить, что эти смещения намного меньше, чем смещения, вызванные симпатиями и антипатиями.

В заключение подчеркну, что претензии социологии на участие в решении реальных практических задач, стоящих перед нашим обществом, могут быть обоснованными лишь в том случае, если рекомендации социологов будут базироваться на надежной (прежде всего — валидной) социологической информации. Повышение надежности социологической информации — важный фактор повышения эффективности социологической науки и увеличения ее практической отдачи.

ЭФФЕКТИВНОСТЬ СОЦИОЛОГИЧЕСКОГО ИССЛЕДОВАНИЯ

[вместо заключения]

В области исследований и разработок из трех параметров лишь два можно определить одновременно.

1. Если заданы цель и время для ее достижения, то нельзя угадать, сколько это будет стоить.

2. Если ограничены время и ресурсы, невозможно предсказать, какая часть задания будет выполнена.

3. Если четко ставится цель исследования и выделяется конкретная сумма денег, то нельзя предсказать, когда эта цель будет достигнута.

Если же Вам повезет и Вы сможете определить все три параметра, значит, Вы имеете дело не с исследованиями и разработками.

*Расширенный принцип
Эпштейна—Гейзенберга*

Мы рассмотрели понятия, описывающие качество социологической информации, методы оценки составляющих качества и некоторые процедуры его повышения. Завершая книгу, я хочу вернуться к вопросу о соотношении качества первичной социологической информации с качеством социологического исследования. Мы остановимся на проблемах, на которые (в отличие от проблемы повышения качества информации) редко обращают внимание не только социологи-практики, но и специалисты по методике социологических исследований. Прежде всего, речь идет об эффективности социологического исследования.

Эффективность — это отношение полученных результатов к произведенным затратам. Здесь требуются определенные комментарии. Говоря об эффективности, я имею в виду общественную эффективность в широком смысле, а не только экономическую. Ведь не секрет, что очень часто в теоретической работе, а еще чаще — в практической деятельности критерием является именно экономическая эффективность, причем иногда социальный (правда, в узком

смысле) эффект рассматривается как средство достижения эффекта экономического. Не раз я слышал, например, такой вопрос: «Скажите, а что нам даст улучшение межличностных отношений в коллективе, это повысит производительность труда?»

Между тем усиление социальной направленности планов, ориентация на достижение конечного общественного результата требуют рассмотрения экономических показателей как средства для достижения стратегических (в конечном итоге — социальных) целей нашей партии — неуклонного подъема материального и культурного уровня жизни народа, формирования гармонично развитой, общественно активной личности [Материалы XXVII съезда КПСС, 1986, с. 163, 171]. Поэтому основным критерием оценки деятельности (в том числе деятельности исследовательской) должна быть общественная эффективность, включающая экономическую в качестве одной из своих составляющих.

Замечу, что даже в тех случаях, когда в работах экономистов речь идет об оценке социальной эффективности на поверку оказывается, что измеряется экономическая эффективность, детерминируемая социальными факторами. Например, социальную эффективность мероприятий по профилактике заболеваний оценивают в рублях, вычисляя число рабочих дней, «сэкономленных» благодаря этим мероприятиям и умножая их на выраженный в рублях средний объем продукции, производимой за день¹. В 1983 г. участвуя в разработке методики определения социальной эффективности РАСУ СССР (республиканская автоматизированная система управления), я просмотрел много отраслевых методик, в которых были сделаны попытки определить социальную эффективность отраслевых АСУ (Минвуза, Минфина, Минздрава СССР и т. п.). Оказалось, что практически все эти методики измеряют не социальную эффективность, а экономическую эффективность социальных мероприятий.

Г. Н. Черкасов [1978, с. 41] справедливо отмечает, что «социальные явления по своей внутренней сущности не обладают стоимостью, хотя вследствие широкого и длительного распространения в обществе товарно-денежных отношений и могут выражаться в форме цены. Но эта форма, как указывал К. Маркс, будет мнимой, иррациональной, т. е. не имеющей объективного содержания в виде об-

¹ Если пользоваться таким показателем, то окажется, например, что лечение пенсионеров пагубно сказывается на социальной эффективности здравоохранения — затраты есть, а результат нулевой.

ущественно необходимых затрат абстрактного труда». Это означает, что для определения социальной эффективности социальный результат и затраты должны выражаться не в рублях, а в адекватных сути дела социальных показателях (например, степень удовлетворения потребностей населения в тех или иных продуктах по отношению к рациональной норме).

Для социологических исследований данное замечание неактуально в той части, в которой речь идет о результатах, поскольку, как правило, эти исследования направлены на изучение именно социальных процессов и результаты исследований выражаются в виде соответствующих социальных показателей. Однако все, кто касается затрат на само исследование, в полной мере относится и к области социологии, поскольку обычно учитываются лишь экономические, а не социальные затраты на проведение исследования.

Рассмотрим этот вопрос подробнее. Прежде всего, затраты можно разделить на издержки исследователей и исследуемых.

Специфика социологии заключается в том, что здесь объект исследования — это сами люди и отношения между исследователем и исследуемым имеют характер взаимоотношений, а не только одностороннего воздействия². Очень часто социологи учитывают это только при определении надежности собранной информации, то есть рассматривают исследуемых как датчиков информации. Тем самым не учитывается нравственный аспект взаимоотношений с исследуемыми — они выступают лишь как средство. Можно присоединиться к беспокойству, высказываемому по этому поводу в научных и публицистических работах В. Я. Шубкиным. Исследователь, «как врач, обязан неуклонно выполнять требование — не повредить. Социолога должна отличать гуманная и высоконравственная позиция, основанная на признании безусловной суверенности личности человека, которая сама по себе обладает высшей ценностью» [Шубкин, 1980, с. 9].

Я позволю себе привести здесь ту же цитату из Л. Н. Толстого, которую приводил В. Н. Шубкин. Речь

² Эффект взаимодействия средств наблюдения и объекта наблюдения характерен не только для социологии, и внимание к такого рода эффектам связано больше всего с развитием ядерной физики, но для социологии, (социальной психологии, психологии) взаимодействие исследователя, средств исследования и объекта исследования имеет свое особенное значение [Алексеев, Бородкин, 1970, с. 37—48].

идет о переписи 1882 г. в Москве. «Цель переписи — научная. Перепись есть социологическое исследование. Цель же науки социологии — счастье людей. Наука эта и ее приемы резко отличаются от всех других наук. Особенность в том, что социологическое исследование не проводится учеными по своим кабинетам, обсерваториям и лабораториям, а двумя тысячами из общества. Другая особенность та, что исследования других наук производятся не над живыми людьми, а здесь над живыми людьми. Третья особенность та, что цель всякой другой науки есть только знания, а здесь — благо людей. Туманные пятна можно исследовать одному, а для исследования Москвы нужно 2000 людей. Цель исследования туманных пятен только та, чтобы узнать все про туманные пятна; цель исследования жителей та, чтобы вывести законы социологии и на основании этих законов учредить лучше жизнь людей. Туманным пятнам все равно, исследовать их или нет, и они ждали и еще долго готовы ждать; но жителям Москвы не все равно, особенно тем несчастным, которые составляют самый интересный предмет науки социологии... Будем записывать, считать, но не будем забывать, что если нам встретится человек раздетый и голодный, то помочь ему важнее всех возможных исследований, открытий всех возможных наук» [Толстой, 1937, с. 173—175].

С чем-то в этом утверждении Л. Н. Толстого можно не соглашаться, наука за эти сто лет приобрела коллективный характер, для исследования туманных пятен нужна сложная аппаратура и много обслуживающего персонала, туманные пятна исследуют тоже для того, чтобы в конечном итоге «учредить лучше жизнь людей», но одно несомненно: интересы исследуемых при проведении социологических исследований должны приниматься во внимание не в меньшей мере, чем интересы исследователей. Поэтому при определении эффективности средств исследования издержки исследуемых на исследование должны учитываться особенно тщательно.

Таким образом, качество средств исследования определяется результативностью и эффективностью, причем последняя определяется отношением результата (прямого и побочного) к издержкам исследователей и исследуемых.

Результативность — это степень достижения цели. Но определить степень достижения цели можно только после проведения исследования, я же имею в виду потенциал, возможность достижения цели данными средствами (в каком-то смысле можно говорить о вероятности достижения цели с помощью данных средств). Степень достижения

результата зависит от реального процесса, то есть не только от средств, но и от условий исследования (другое дело, что различные средства в различной мере чувствительны к этим условиям). Но в значительной степени результативность средств может определяться до проведения исследования. Особенно это относится к тем случаям, когда до проведения исследования ясно, что с помощью этих средств цели не достичь, но и что касается положительной оценки средств исследования, она тоже может быть дана заранее — при наличии апробированной теории, исходя из предыдущих исследований, опираясь на результаты методических экспериментов и т. п. Дальнейшее развертывание понятия результативности может идти по пути анализа различных составляющих средств исследования: результативность зависит от качества специальной методологии — теорий, на основе которых разработана методика, качества инструментария и т. д.

Эффективность средств исследования тоже определяется в процессе исследования и после его завершения, но еще на начальном его этапе при выборе средств нужна предварительная оценка их эффективности. При одном и том же результате (до окончания исследования он задается в виде цели и предвидимых побочных следствий) эффективность зависит от издержек исследователей и исследуемых.

Типологию естественно начать с исследовательских затрат и проводить различие по расходуемым ресурсам. Например, можно выделить затраты труда — численность людей, необходимых для проведения исследования и объем работы каждого из сотрудников. Это трудоемкость исследования. Финансовые затраты на оплату работников, тиражирование документов исследования, обработку данных на ЭВМ и т. п. — экономичность исследования. Затраты времени на проведение исследования — это его оперативность (имеются в виду не суммарные затраты времени исследователей — это входит в понятие трудоемкости, — а время, затрачиваемое на проведение всего исследования от начала до конца).

Понятно, что методы сбора информации отличаются по оперативности — например, проведение почтового опроса требует, как правило, больше времени, чем проведение телефонного интервью (но менее трудоемко). Различные по целям исследования предъявляют различные требования к указанным характеристикам средств: например, изучение отношения аудитории к той или иной передаче телевидения требует большей оперативности, чем изучение об-

раза жизни советских людей. Значение тех или иных характеристик средств исследования зависит от имеющихся у исследователей ресурсов. Так, вузовские социологические лаборатории, привлекающие студентов для проведения опросов, меньше обеспокоены трудоемкостью исследования, лаборатории крупных предприятий — экономичностью, академические исследовательские группы, проводящие фундаментальные исследования, — оперативностью исследования.

В зависимости от ситуации исследовательская группа может быть лимитирована не только общим объемом финансовых или трудовых затрат, но и каким-либо специфическим ресурсом: бумагой для тиражирования документов, кадрами интервьюеров или программистов, доступом к ЭВМ и т. п.

(Наконец, дальнейшую типологию тех или иных видов затрат можно проводить, относя расходование тех или иных видов ресурсов к различным этапам проведения социологического исследования: например, разработка концепции исследования, программы, методики, сбор информации, ее обработка и анализ, интерпретация полученных результатов. Каждый из этапов имеет характерный для него способ расходования ресурсов.)

Однако все названные показатели исследовательских затрат не выходят за рамки сугубо экономических. Для определения же общей, в конечном итоге, социальной эффективности требуется учет издержек исследователей, сугубо социальных по своему характеру. Изучение с этой точки зрения социологической литературы еще раз доказывает мудрость народной пословицы о сапожниках, которые ходят без сапог. Хотя изучение человека и его работы является одной из центральных тем в нашей социологии, сами социологи и их работа еще не стали объектом таких исследований.

Комплексным социальным показателем значения для работника его трудовой деятельности может выступать степень удовлетворения потребностей работника, в той мере, в какой это обусловлено выполнением данной деятельности. Обычно выделяют удовлетворение потребности личности в самом труде (при этом рассматривается характер и содержание, условия, иногда смысл труда) и вознаграждение за труд (моральное, например, престиж и материальное).

Вознаграждение за труд социологов мало отличается от вознаграждения экономистов, историков, биологов и представителей других специальностей, требующих выс-

шего образования (разве что гораздо уже спектр рабочих мест и чаще возникает проблема оплаты труда временных работников, привлекаемых в качестве интервьюеров). А характер и содержание труда социологов имеют определенную специфику. С точки зрения социальных затрат она заключается в наличии трудоемких этапов сбора и подготовки больших объемов первичной социологической информации, требующих ручного малоквалифицированного и непривлекательного труда. Доля такого труда по сравнению с общими трудовыми затратами на проведение исследования должна служить одним из существенных показателей при оценке эффективности исследовательских средств. Важным в этом отношении социальным показателем является несоответствие квалификации исследователя и сложности выполняемой работы. Для оценки этого несоответствия могут использоваться методы, аналогичные разработанным для других видов деятельности в связи с решением проблемы ликвидации малоквалифицированного труда [см., например: Хмелько, 1981].

Есть здесь еще одна особенность, на которую хочется обратить внимание. Если работа по подготовке к обработке больших объемов информации (кодировка, перфорация, проверка и исправление ошибок) безусловно малоквалифицированная и требует ликвидации, то работа по сбору информации является, вообще говоря, высококвалифицированной (это относится и к контент-анализу, и к наблюдению, и к опросу). Это хорошо видно из требований предъявляемых к профессиональным качествам интервьюера в методических работах по социологии. К ним относят сознательность, гражданскую убежденность, чувство политической ответственности, общую культуру, способность принимать решения в нестандартных ситуациях, способность четко формулировать свои мысли, контактность, общительность, физическое здоровье, психологическую устойчивость, дисциплинированность, собранность, бодрость, серьезность, опрятность в одежде, способность вызывать симпатию, вежливость, отсутствие самодовольства, умение наблюдать и понимать поведение других и т. д. и т. п. И все это качества, выписанные только из одной книги по методике социологических исследований [Процесс социологического исследования, 1975, с. 251, 252]. А если добавить к ним перечень качеств, встречающихся в других книгах?

Казалось бы, работа, требующая такой высокой квалификации и такого «букета» личностных качеств, должна быть необычайно привлекательной и престижной — не на-

много меньше, чем работа космонавта, артиста, полярника.

Реально же работа по проведению социологического опроса является не только одним из самых трудоемких, но и одним из самых неприятных этапов проведения социологического исследования. Объясняется это несоответствием на этот раз уже не между квалификацией работника и выполняемой работой, а между типом личности и требованиями, предъявляемыми к качествам личности характером и содержанием работы. Как справедливо отмечают авторы цитируемой книги, требуемые интервьюеру качества «в известной степени могут быть приобретены в результате обучения, однако только в известной степени» [Процесс социального исследования, 1975, с. 251]. Во многом же наличие указанных качеств связано с неизменяющимися в результате обучения чертами личности (например, экстраверсией, общительностью). Более определенно высказывается Э. Ноэль: «Так или иначе важные качества хорошего интервьюера невозможно привить не подходящему для этой роли человеку. Вместо этого следует разрабатывать тесты, чтобы найти людей, обладающих нужными качествами» [Ноэль, 1978, с. 193]. Для тех же, кто такими качествами не обладает, работа интервьюера является морально (а иногда и физически) тяжелой³, она требует преодоления скованности, застенчивости и воспринимается как обуза, а не как возможность раскрыть свои способности и проявить необходимые качества.

Сказанное означает, что необходимо предусматривать подбор интервьюеров не только в соответствии с теми требованиями, которые, с точки зрения организаторов, значимы для получения надежной информации, но и в соответствии с интересами тех лиц, которые будут выступать в качестве интервьюеров. Работа должна нравиться — в этом состоит социальный подход к определению эффективности при подборе интервьюеров. Однако во многих случаях исследовательской организации не приходится выбирать — интервьюерами назначаются работники этой организации, и как правило, большинство (например, в академическом институте могут «мобилизовать» аспирантов, лаборантов и младших научных сотрудников без степени). В этом случае, несмотря на то, что работа интервьюера

³ В капиталистических странах из-за высокого уровня преступности работа интервьюера, которому приходится проводить опрос в вечернее время в разных частях города, сопряжена к тому же с риском для жизни [Dillman, 1978].

потенциально является творческой, к социальным издержкам исследователей должна быть отнесена несбалансированность типов личности и требований к интервьюерам (в качестве показателя может выступать доля лиц, неудовлетворенных работой интервьюера).

Рассмотрим теперь издержки исследуемых, так как этот вопрос почти не рассматривается в социологической литературе (точнее, рассматривается обычно лишь с такой точки зрения — сможем ли мы с помощью длинной анкеты получить надежные данные, не приведет ли усталость к понижению качества информации). Кстати сказать, длина анкеты и вообще трудоемкость ее заполнения в аспекте взаимоотношений исследователя и исследуемого — это наименее болезненная проблема, которая решается «автоматически».

Тут существует обратная связь, поскольку при превышении допустимого уровня трудоемкости падает качество информации, что побуждает исследователя (но, к сожалению, не всегда) выбирать методы, не выходящие за приемлемые для исследуемого границы степени трудоемкости. Исследователь и исследуемый находятся в отношении равной зависимости и эта ситуация почти не порождает морально-этических проблем.

Нередко, однако, они находятся в отношениях односторонней зависимости (иногда, как это бывает при использовании метода наблюдения, не требуется даже согласия исследуемого), что предъявляет повышенные требования к нравственности исследователя и связано с серьезными проблемами.

В качестве примера неконтролируемой нравственным чувством экспансии исследователей приведу эксперимент, проведенный тремя американскими учеными в мужском туалете университета для изучения некоторых социально-психологических особенностей протекания физиологических процессов, в частности влияния на эти процессы присутствия других лиц [Middlemist, Knowles, Matter, 1976]. Публикация этой работы стала затем предметом обсуждения — действительно ли цели исследования (и полученный результат) настолько важны, что ради них нужно было ставить эксперименты, оскорбляющие человеческое достоинство? [Koosher, 1977].

Конечно, такого рода примеры являются крайним случаем неэтичности способов исследования. Более распространенными являются неделикатность, навязчивость, «давление» на респондента. Иногда можно столкнуться и с попыткой «насильственного» проведения опроса — путем при-

нуждения к ответам административными указаниями⁴. Как справедливо отмечает В. Н. Шубкин, «неделикатность в области человеческих отношений, общение с «позиции силы», давление креслом, удары авторитетом — вещи столь же недопустимые для социолога, как для хирурга пользование топором. Это любезность, что незнакомый человек согласился затратить свое время и ответить на заданные ему вопросы. Он имеет полное право не отвечать. Если же у него возникает хотя бы малейшее подозрение, что информация, которую он предоставляет, может повредить ему, то это ошибка исследователя. После такого сигнала он должен немедленно прекратить анкетирование и заново пересмотреть всю методику» [Шубкин, 1980, с. 10]. И далее: «Разумеется, в исследовании могут использоваться самые тонкие методы: тесты, вопросы, значение которых понятно лишь профессионалам. Поэтому нельзя по отдельным, выдернутым из контекста пунктам анкеты пытаться судить о работе в целом. Однако, два условия должны быть учтены непременно: первое — добровольность представления информации и второе — честность во взаимоотношениях между социологом и тем, кого он опрашивает» [там же, с. 12].

К сожалению, эти требования выполняются не всегда даже квалифицированными специалистами. Кроме того, честность в отношениях с исследуемым понимается по-разному. Сама процедура исследования, необходимость устранения влияния исследователя и возможных смещений, связанных с его присутствием, вынуждают скрывать либо сам факт исследования (как, например, в методе включенного наблюдения), либо цели исследования, либо некоторые особенности средств и процедуры.

Приведу следующий пример. В одном исследовании, проведенном весьма уважаемыми мною коллегами, требовалось сопоставить структуру межличностных отношений в студенческой группе с характеристиками личности,

⁴ В этом случае информация часто бывает ненадежной, а иногда и фальсифицированной.

Моя первая встреча с методами эмпирической «социологии» состоялась, когда я учился на втором курсе университета. Однажды наш преподаватель за двадцать минут до конца занятий раздал нам анкеты для строителей кооператива искомандовал: «Представьте себя мужчинами и женщинами от 25—30 до 50—60 лет, которые намерены вступить в кооператив или уже строят его. Войдите в роль и заполните анкету от имени выбранного вами лица. Следите за логикой! Сдать до конца занятий!» Деканат поручил ему провести этот опрос среди преподавателей факультета, но поскольку это очень хлопотно, он нашел более простой выход.

полученными с помощью некоторых тестовых методик. Для обеспечения искренности ответов социометрический опрос, выясняющий структуру межличностных отношений, должен был проводиться анонимно. Но тогда нельзя было сопоставить данные опроса с результатами тестов. Поэтому исследователи, объявив опрос анонимным, проставляли между тем номер респондента на второй странице под скрепкой, скрепляющей страницы анкеты. При обсуждении этичности этой процедуры, они указывали на то, что фактически опрос был анонимным, так как их не интересовали ответы каждого отдельного студента (хотя они и могли их установить), данные использовались лишь для расчета коэффициентов корреляции тестовых показателей с межличностными оценками. Считать ли это обманом?

Еще один пример — с включенным наблюдением. После одной из конференций гостеприимные хозяева устроили экскурсию на завод, на котором их лаборатория проводила исследование. В числе других процедур использовалось включенное наблюдение: сотрудница лаборатории устроилась работницей в бригаду. Когда ее разыскали, она, испуганно озираясь, чтобы подруги по работе не увидели ее с гостями и администрацией предприятия, рассказала о сложностях работы, о том, что она чувствовала себя предательницей, отказывалась от этого задания, первое время даже плакала и едва освоилась со своей ролью («Меня даже как-то раз предупредили — с Тamarой будь осторожна, она все начальству докладывает; а каково мне это слушать, если я раз в неделю в оперативках участвую»). Социологи же пожаловались, что она настолько вошла в новый коллектив, что на дирекции несправедливо отстаивает интересы бригады, в которой работает. Не помню уже, какой результат был получен этим методом, но девушку-наблюдателя было искренне жаль. Возникает вопрос, допустимы ли такие ситуации?

Естественно, напрашивается отрицательный ответ. Принцип «цель оправдывает средства», выдвинутый иезуитами, глубоко чужд марксистско-ленинскому обществоведению. Однако сказать так и этим ограничиться значило бы упрощать диалектику цели и средств, лишь создавая иллюзию решения проблемы, которая в действительности очень сложна. Социология (а еще в большей степени — психология) вынуждены вторгаться в сферу интимных человеческих чувств, переживаний, отношений. Здесь нельзя забывать, что использование аморальных методов для достижения благой цели не позволяет ее достичь, так как в этом случае достижение цели перестает быть благом, —

средства должны соответствовать цели. Но, с другой стороны, отказ от средств, которым трудно дать однозначную нравственную оценку, может привести к отказу от цели, если другие средства недоступны. Вопрос всегда стоит конкретно и конкретно должен решаться.

Например, проблема службы знакомств, «электронной свахи», неоднократно подвергавшейся критике (раньше — чаще, сейчас — реже) средствами массовой информации. Связанные со службой знакомств процедуры (постановка на учет, оплата денег, картотека, способы знакомства) многим представляются унижительными, низводящими до чиновничьих будней таинство зарождения любви и т. п.

Если даже предположить, что это так (хотя я не считаю, что этот способ знакомства чем-то хуже, чем, например, случайное знакомство), то спросим, все-таки: к каким последствиям ведет бездействие? Только в Киеве 200 тыс. лиц, не состоящих в браке, многие из них лишены радости любви, материнства и отцовства. Удовлетворенность браком занимает одно из первых мест в иерархии факторов, определяющих удовлетворенность человека своей жизнью в целом [ИТР и формирование духовного облика ..., 1982, с. 139]. Речь идет об удовлетворении одной из базовых потребностей, о потребности, которая может придать смысл жизни. Поэтому самая несовершенная служба, которая осуществляет подбор пар лишь по двум признакам — полу и возрасту — лучше, чем бездействие⁵. В данном случае средства (хоть они и не всем нравятся) соответствуют целям — по крайней мере нет других, более адекватных средств. Неконструктивная критика, направленная не на совершенствование, а на ликвидацию службы знакомств, представляется ханжеской и бесчеловечной. Более того,

⁵ Особенно удивило меня высказывание известного математика, автора очень интересных книг по планированию эксперимента, вероятностей теории языка и др. В. В. Палимова: «Еще один пример узидка культуры — брачный клиринг, когда для облегчения встречи будущих супругов используется ЭВМ, в память которой заносятся дискретно задаваемые характеристики человека. Нам представляется, что в действительности каждый человек — это некий текст, написанный на несловесном уровне. Влюбленность создаст возможность чтения этого текста..., любовь открывает возможности узнавания, слияния, сопереживания. В физическом плане внезапное чтение происходит на биохимическом — гормонально обусловленном уровне» (В. В. Палимов. Непрерывность против дискретности в языке и мышлении. — Тбилиси, 1978, с. 66). Зачем же форму знакомства смешивать с рождением любви? Если предположить, что человек — это несловесный текст, то нужна библиотека, где можно было бы познакомиться с ним для дальнейшего «внезапного чтения». Чем это хуже, чем брать «тексты» у знакомых?

развитие системы таких служб в рамках службы семьи должно рассматриваться не как одна из бытовых услуг населению (наряду с циклевкой полов, доставкой подарков, мытьем окон и обивкой дверей дерматином, как это имеет место сейчас, например, в Киеве), а как важный рычаг нашей демографической политики, средство решения важной социальной проблемы, повышения качества жизни советских людей.

Приведу еще один пример, показывающий, что вопрос о соответствии цели и средств далеко не тривиален. Пример этот, правда, не из социологии, а из области знания находящейся на стыке с социологией — сексологии, но по степени этичности средств он, на первый взгляд, близок к эксперименту американских социальных психологов по изучению влияния вторжения посторонних в личное пространство человека на протекание его физиологических процессов и поэтому нагляден⁶. Сексологами была подобрана группа мужчин, сексуальные проявления которых соответствовали принятым в научной сексологии представлениям о безупречном здоровье, и проведено исследование для определения некоторых норм половой жизни⁷. При известном сходстве между двумя этими исследованиями есть коренное различие. Исследование Г. С. Васильченко необходимо для выяснения вопросов, от которых зависит успех лечения (в том числе терапевтического) и, следовательно, здоровья и счастья людей и их семей (до выяснения нормы вообще не было известно, кто из обратившихся к врачу болен, а кто — здоров). Исследование было предпринято из-за «его особой важности для практической работы врача-сексопатолога»⁸, отказ от проведения этого исследования нанес бы значительно больший ущерб, чем моральный ущерб, нанесенный участникам эксперимента (большинство из них, кстати, добровольцы): на другой чаше весов лежит разрушенная семейная жизнь тех людей, которые не были бы вылечены врачами-сексопатологами. В американских же экспериментах был получен достаточно тривиальный результат, ценность которого весьма сомни-

⁶ Замечу, кстати, что при разработке методик, определении надежности измерения социологи вынуждены вторгаться в область общей и социальной психологии. При взаимодействии респондента с интервьюером или анкетой проявляются закономерности, изучаемые именно психологией, может быть целесообразно даже выделить специальную область — социологическую психологию — на стыке психологии и методики социологических исследований.

⁷ Общая сексопатология: Руководство для врача / Под ред. Г. С. Васильченко. — М., 1977, с. 188.

⁸ Там же, с. 185.

тельна. Как показала дискуссия, возникшая в связи с этим исследованием, авторы и не пытались сопоставить стоящие перед ними цели со средствами исследования, оценить, существуют ли альтернативные способы исследования, не следует ли отказаться от его проведения. Диалектика цели и средств — это не абстрактный вопрос, представляющий теоретический интерес, а насущная проблема практики научных исследований, не имеющая однозначного решения и требующая тщательного рассмотрения в каждом конкретном случае. Ситуации небезупречного в нравственном отношении использования социологических методик представляют, на мой взгляд, бóльшую опасность, чем ситуации исследовательского бездействия из-за моратория на конкретные способы проведения социологического исследования. Поэтому целесообразно уделять особое внимание этому вопросу в социологических публикациях, практиковать обсуждение с этой точки зрения процедур социологического исследования, стимулировать разработку новых способов сбора информации с учетом нравственного аспекта методики.

Например, включенное наблюдение во многих случаях может быть заменено процедурой «наблюдающего участия», которая отличается тем, что исследователь становится участником коллектива, не скрывая своего «происхождения», своих целей и задач [Алексеев, 1982, с. 161—170; Ряжских, 1975, с. 91—99]. В этом случае наблюдение (точнее, участие) проводится дольше — с тем чтобы члены коллектива успели адаптироваться и перестали обращать внимание на исследователя, исследователь по-другому строит свое отношение с членами коллектива и т. п. По другому может быть также достигнута анонимность по отношению к исследователю⁹ социометрической процедуры. Например, так, как в описанной ранее процедуре определения трудового вклада инженеров. Аргумент же, что допустима «военная хитрость», так как обмана нет и опрос, фактически, анонимен, не очень убедителен. В проведении исследования участвует, по крайней мере, несколько человек, и контролировать эту ситуацию трудно. Вообще для каждого метода сбора социологической информации в рам-

⁹ В некоторых случаях целесообразно различать анонимность: по отношению к исследователю, к администрации, к другим исследуемым и т. п. В разных исследованиях более важно обеспечить тот или иной вид анонимности, для каждого из видов может использоваться своя процедура. Например, для обеспечения анонимности по отношению к другим респондентам в устной инструкции исследователь указывает, что анкеты нужно взять домой, заполнить и на следующий день сдать лично исследователю.

ках разработки профессиональной этики должны быть разработаны свои правила нравственного поведения, своего рода «кодекс чести», входящий как составная часть в методику исследования.

Рассмотрим некоторые этические проблемы, связанные, например, с проведением почтового опроса. В 1979 г. мне довелось руководить проведением эмпирического этапа исследования воспроизводства социальных источников трудовых ресурсов¹⁰. Этот этап исследования включал почтовый опрос занятого населения г. Кисва. Кроме 3400 вернувшихся анкет, мы получили также письма в адрес исследовательской группы. Все эти письма наряду с чисто эмоциональными комментариями (например: «Гвоздюк здесь не проживает, уехал к чертовой матери») содержали жалобы и просьбы о помощи, вопросы юридического характера и замечания к анкете. В силу загруженности работой по проведению исследования мы не ответили на эти письма. Думаю, что это нарушение профессиональной этики. Люди писали с доверием к исследовательской группе, нельзя было их разочаровывать.

Наверное, требование Л. Н. Толстого о помощи лицам, нуждающимся в ней, при проведении социологического исследования вряд ли может стать правилом профессиональной этики, это дело совести исследователей. Но безусловным правилом должен быть ответ на все письма, полученные исследовательской группой. В тех случаях, когда исследователи не могут дать индивидуальный ответ, целесообразно предусмотреть стандартный ответ с объяснением, оправдывающим такую форму ответа и, по мере возможности, с советом, куда лучше обратиться по тому или иному вопросу. Например: «Уважаемый товарищ! Благодарим Вас за добросовестное заполнение анкеты. К сожалению, мы не можем подробно ответить на Ваше письмо, так как исследовательская группа не очень большая, а получаем мы сотни писем. По интересующему Вас вопросу можете обратиться... Исследовательская группа»¹¹.

¹⁰ Исследование проведено отделом конкретных социологических исследований Института философии АН УССР (руководитель исследования — В. Ф. Черноволенко) [Семья и трудовое самоопределение, 1984; Паниотто, Яковенко, 1981, с. 201—208].

¹¹ Изящный стандартный ответ придумал английский актер Питер О'Тул, известный у нас по фильму «Как украсть миллион» и др. В ответ на письма почитателей он высылает свою фотографию с факсимиле дарственной надписи на обратной стороне и письмо такого содержания: «Благодарю Вас за Ваше письмо, тронут Вашим вниманием к моей игре. Прошу простить, что не могу ответить на него, так как если бы я отвечал на все письма, которые получаю, то мне пришлось бы нанять человека, который играл бы за меня в театре».

Некоторые респонденты интересуются результатами исследования. Поэтому может быть предусмотрена подготовка экспресс-отчета, коротко излагающего результаты исследования (либо буклета с результатами предыдущего или аналогичного исследования). Замечу также, что вообще участие в социологическом исследовании может (при правильной организации) иметь и положительный эффект для респондентов (до сих пор мы вели речь лишь о затратах). Это — повышение социальной активности респондента, сознание причастности к важному исследованию, а иногда — и осознание значимости своего мнения, своей роли в подготовке рекомендаций как участия в управлении общественными делами.

Некоторые авторы даже считают, что «социальный эффект многих социологических исследований пока еще в большей степени проявляется именно в процессе опроса, во время непосредственного общения с людьми, чем после подведения итогов» [Погосян, 1985, с. 36]. Автор приводит классический пример Хотторнских экспериментов Мэйо, показавших, что сам факт исследования, внимания к работницам оказал намного большее влияние на исследуемых (в том числе и на интересовавшую исследователей производительность труда), чем все другие факторы, которыми манипулировали исследователи в эксперименте.

Целесообразно выделить еще одну характеристику издержек исследуемых — возможный материальный или моральный ущерб, который может быть связан с проведением исследования. Назовем эту характеристику психологической травматогенностью методики исследования. Хотя этические средства исследования не могут быть травматогенными в указанном смысле, понятия травматогенности и неэтичности не совпадают, поскольку неэтичные средства исследования, связанные, например, с обманом респондента, могут и не нанести ему ущерба (если он не обнаружил обмана). Вероятно, психологическая травматогенность — это наиболее важное из непосредственных последствий исследования для исследуемых, и поэтому при разработке методики исследования следует уделять ей особое внимание.

Степень травматогенности может варьировать от едва заметной до способной поставить под сомнение само существование исследуемого объекта (например, разрушить исследуемую структуру межличностных отношений). Мои коллеги Е. И. Головаха и А. А. Кроник, исследуя психологическое время личности [Головаха, Кроник, 1984; см также: Знание — сила, 1983, № 9, № 11; 1984, № 2, № 3], обрати-

лись ко мне с просьбой выступить в качестве испытуемого и ответить на вопросы специальной анкеты. Одним из первых был вопрос: «Как Вы думаете, в каком году Вы умрете?» Мне этот вопрос почему-то сразу не понравился. Я высказал предположение, что для мнительных или больных людей этот вопрос может оказаться психологически травматогенным. В более позднем варианте методики он был заменен вопросом: «До какого возраста Вы доживете?» и в таком виде был опубликован в журнале «Знание — сила» [Кроник, Головаха, 1984, с. 22]. Поскольку в многочисленных откликах на статью (было получено более полутора тысяч писем) не содержалось критики вопроса, можно предположить, что травматогенность его невелика.

В. Н. Шубкин приводит пример социологического исследования по определению трудового вклада и личностных характеристик сотрудников НИИ (исследования проводились при подготовке к переходу на Карповскую систему оплаты труда научных работников), которое привело к созданию конфликтной ситуации и резко ухудшило психологический климат в коллективе [Шубкин, 1978, с. 194—196]. Правда, в данном случае, как мне представляется, вред нанесло не само исследование, а способ внедрения его результатов. Но несомненно, что некорректно проведенное исследование может принести существенный ущерб изучаемому социальному объекту или отдельным испытуемым.

Рассмотрев затраты на проведение социологического исследования с точки зрения его непосредственных социальных последствий для исследователей и исследуемых, обратимся теперь к экономическим затратам. Как это ни странно, возможные способы уменьшения такого рода затрат на проведение социологических исследований крайне редко рассматриваются в нашей литературе. Потребители социологической информации, а также многие из социологов склонны недооценивать материальные, кадровые и временные расходы, требующиеся для получения этой информации. Не всегда осознается тот факт, что имеющиеся в распоряжении исследователя ресурсы оказывают очень существенное влияние на качество получаемой информации. Если учесть, что социологические исследования приобретают все более широкий размах и масштабность, ясно, что важность экономии имеющихся ресурсов возрастает. Разработка новых, более эффективных методов исследования стимулируется, таким образом, и необходимостью экономии наличных ресурсов и повышения качества социологической информации.

Важным этапом, во многом определяющим весь дальнейший ход работ, является планирование исследования. В этих вопросах социология явно отстает от многих других областей науки и техники, не использует существующих методов научного планирования работы. Перспективными представляются методы планирования эксперимента [см., например: Кэмпбелл, 1980; Налимов, 1971; и др.] и сетевого планирования (попытки использования сетевых графиков описаны в: [Рукавишников, Паниотто, Чурилов, 1984, с. 47—49; Семья и воспроизводство..., 1984, с. 161—165]). Существенный эффект могут дать совершенствование методов проектирования выборки, использование математических методов и вычислительной техники для построения многоступенчатых выборок [гнездовых, стратифицированных], включающих автоматическую классификацию единиц отбора, факторный анализ признаков, положенных в основу стратификации, оптимальное размещение выборки и т. п. [Проектирование и организация..., 1977; Территориальная выборка..., 1980]. Это позволяет повысить надежность информации и снизить затраты на одном из самых трудоемких этапов исследования — сборе информации.

Но наиболее существенного повышения эффективности исследований можно ожидать при переходе от разовых к регулярным социологическим исследованиям, что позволяет изменить технологию исследования. Большинство затрат при этом носит единоразовый характер: построение выборки, формирование опросной сети, подбор и обучение интервьюеров и т. п. Создание единой методики сбора и обработки социологической информации позволяет сделать процесс рутинным, придать ему производственный характер [Рукавишников, Паниотто, Чурилов, 1984, гл. 5]. Замена разовых исследований регулярными возможна, конечно, лишь при некотором постоянстве исследовательских целей и задач, когда требуется не столько углубленное изучение взаимосвязей, латентных факторов, проверки гипотез, сколько описание, обзор, фиксация состояния некоторого процесса — то, что часто называют обследованием, а не исследованием¹².

Следующий этап — разработка методики исследования. Основным путем повышения эффективности на этом этапе является разработка типовых стандартных методик,

¹² В английском языке эти понятия обозначаются даже разными терминами *research* (собственно исследование, то есть научное — в узком смысле — изучение) и *survey* (описание, обозрение, обследование).

тестов для измерения тех или иных социальных признаков. Стандартизированные методики или тесты, предназначенные для широкого использования, должны удовлетворять определенным требованиям: содержать инструкцию для пользователей, описание способа разработки методики, данные методических экспериментов по определению различных составляющих надежности методики и т. п. [см., например: Standarts ..., 1974]. Такие тесты позволяют не только уменьшить затраты, но и повысить качество результатов исследования.

К сожалению, в нашей социологии нет традиции разработки типовых или стандартных методик. Думаю, что это связано с отсутствием заинтересованности разработчиков в широком распространении созданных ими средств исследования. Созданию такой заинтересованности может способствовать обеспечение авторских прав на методику, материальное и моральное стимулирование разработчиков, публикация или другие способы распространения методик и т. п. Возможным путем является приравнивание методик, удовлетворяющих определенным требованиям, к изобретениям или рационализаторским предложениям, а также стимулирование подготовки стандартных методик в качестве диссертационных работ. Целесообразно наряду с банками данных создать всесоюзный и региональные банки методик социологического исследования.

Тиражирование анкет и других документов исследования не является трудоемким этапом, но может занимать много времени и требует большого количества бумаги. Существуют некоторые методы проведения исследований, которые позволяют избавиться от тиражирования (телефонный опрос, использование технических средств при сборе информации), но их разработка была связана с попытками повысить эффективность других этапов исследования и о них пойдет речь позже, в связи с рассмотрением этих этапов.

Наиболее трудоемкими в социологическом исследовании являются этапы сбора информации и ввода ее в ЭВМ для обработки. Возможности уменьшения затрат при сборе информации особенно интенсивно исследуются социологами капиталистических стран в связи с ростом стоимости интервью. Были проведены исследования, посвященные рациональному использованию времени интервьюерами, различным формам оплаты, связью оплаты и качества работы интервьюеров. Выяснилось, например, что наиболее удовлетворительной формой оплаты (при средней оплате 7 долларов за интервью) является оплата интервью в соответ-

ствии с формулой, учитывающей дорогу, длину интервью и другие факторы [Sudman, 1967, p. 89—100]. Оказалось также, что к уменьшению стоимости опросов приводит совершенствование методов отбора интервьюеров [там же, с. 100—153].

(Но наиболее перспективным направлением повышения эффективности опросов является использование почтового и телефонного опросов [Рукавишников, Паниотто, Чурилов, 1984, гл. 4]. Длительное время почтовый и телефонный опросы — как у нас в стране, так и за рубежом — считались ненадежными и не использовались в научных исследованиях. Отдельные опросы, проводимые, например, журналами, давали нерепрезентативную информацию, непригодную для использования в исследовательских целях. Однако в последние годы интерес к этим методам возрос. В конце 70-х — начале 80-х годов появились первые примеры удачного использования почтового опроса советскими социологами и экономистами для изучения общественного мнения, определения потребности населения в товарах длительного пользования, изучения воспроизводства структуры занятости и для других целей. Появились первые, пока еще экспериментальные, попытки использовать телефонный опрос.

Негативное отношение к почтовому опросу связано с двумя основными недостатками этого метода: низкий возврат (20—40% от числа разосланных анкет, если не применять специальных методов стимулирования возврата) и нерепрезентативность, связанная с тем, что отвечает более активная часть населения. В последние годы в результате методических экспериментов удалось разработать методы, компенсирующие эти недостатки. К ним относятся методы стимулирования возврата, в частности, напоминание, правильно выбранное время рассылки, специальные вложения в конверт и т. п. [Докторов, 1981; Докторов, Паниотто, 1982], а также методы сочетания почтового опроса и интервью. Например, применяется изучение методом интервью небольшой подвыборки из тех, кто не ответил на почтовую анкету [Паниотто, Яковенко, 1981; Яковенко, 1982; Яковенко, 1984]. Разработанные методы позволяют повысить возврат до 60—80% и скорректировать полученный массив по данным дополнительного интервью, чтобы устранить ошибку репрезентативности.

Таким образом, почтовый опрос позволяет получить информацию того же уровня надежности, что и анкетирование-интервьюирование, но при значительно (в несколько раз) меньших затратах.

Важно отметить, что почтовый опрос легко поддается механизации. Так, например, отечественной промышленностью выпускаются листоподборочные машины, осуществляющие подборку листов в блок, фальцевальные машины, складывающие листы перед укладкой в конверт, конвертозаклеивающие машины, адресовальные и маркировальные машины, конвертовскрыватели, а также автоматические линии, включающие в себя указанные виды машин и осуществляющие комплексную обработку исходящей и получаемой корреспонденции. Производительность автоматических линий достигает 6 000 конвертов в час. Это означает, что их использование дает возможность в течение часа подготовить к отправке все документы, необходимые для проведения крупномасштабного опроса.

Перспективным является телефонное интервью, хотя в настоящее время этот метод может использоваться лишь для опроса городских жителей. К достоинствам телефонного опроса относится прежде всего оперативность, он позволяет сэкономить один из важных ресурсов — время опроса. Кроме того, телефонный опрос избавляет от необходимости тиражировать анкету (экономия и времени и бумаги). Нерепрезентативность телефонного опроса, связанная с отсутствием телефонов у значительной части населения компенсируется соответствующим ремонтом (корректировкой) выборки.

Поскольку при телефонном опросе все вопросы анкеты должны восприниматься на слух, существуют ограничения на сложность анкеты. Этот недостаток компенсируется специальной методикой построения анкеты. Кроме того, в последнее время появился эффективный метод проведения телефонного интервью с использованием ЭВМ: анкета появляется перед интервьюером на экране дисплея и он по мере опроса вводит ответы респондента в ЭВМ с клавиатуры дисплея или с экрана — с помощью светового пера. (Этот метод получил специальное название CATI (computer-assisted telephone interviews) и в настоящее время вызывает все больший интерес (один из номеров журнала «Sociological methods and research» целиком посвящен CATI [см., например: Shanks, 1983]). Телефонное интервью с использованием ЭВМ, во-первых, позволяет устранить этап перфорирования и подготовки данных для ввода в ЭВМ, во-вторых, повышает качество и сложность данных, которые можно получить путем телефонного интервью. Дело в том, что компьютер может выполнять те функции, которые в ходе интервью очень сложно выполнить интервьюеру (нужно обладать способностями Юлия Цезаря);

менять последовательность вопросов в зависимости от ответов респондента, вводить сложные многоуровневые системы фильтров, персонализировать вопросы, запомнив имя и отчество респондента или названных им членов семьи и т. д. Кроме того, появляется возможность оперативно вычислять статистические характеристики распределения опрошенных по ходу опроса и управлять процессом опроса. Таким путем, в частности, можно определить, какие категории респондентов должны быть дополнительно опрошены, а какие уже не следует опрашивать, чтобы обеспечить репрезентативность выборки, а также определить достаточное число интервью для получения значимых показателей и прекратить опрос, когда достигнут требуемый порог значимости.

В зависимости от целей и задач исследования, от категорий населения, которые следует опросить, от требуемой оперативности опроса и от наличия тех или иных ресурсов целесообразно сочетание различных методов опроса. Это может быть сочетание почтового опроса с интервью, телефонного опроса с интервью (например, использование телефона, чтобы договориться с респондентом о времени визита интервьюера), телефонного и почтового опроса (например, использование напоминания по телефону для стимулирования возврата), а также сочетание всех трех методов опроса. Думается, что сочетание взаимодополняющих методов — наиболее перспективное направление, так как позволяет добиться гибкости в выборе стратегии проведения опроса.

Более широкое использование ЭВМ, появление персональных ЭВМ, развитие микропроцессорной техники создают новые возможности для повышения эффективности социологического исследования на этапе сбора информации и ввода ее в ЭВМ. К ним относится опрос непосредственно на ЭВМ с помощью телстайпов [Прилуков, 1977] или видеотерминалов [Каганов, Шумляев, 1980]: респондент с клавиатуры (после предварительной инструкции) вводит в ЭВМ ответы на задаваемые ею вопросы. Однако сфера использования этого метода ограничена — проводить такие опросы можно только среди студентов, военнослужащих, посетителей выставки и т. п., то есть в тех случаях, когда социолог может привести респондента к терминалу.

Существуют также автономные устройства, включающие электронно-вычислительные блоки с теми же функциями, которые выполняет ЭВМ при опросе с терминалов. К такому типу приборов можно отнести «Социометр-2» [Ковалев, Чепуров, 1976], предназначенный для одновре-

менного опроса 30 человек и позволяющий получить распределение их ответов. Этот прибор имеет, однако, довольно узкую специализацию и может использоваться для изучения межличностных отношений в малых группах. Прототипом универсального прибора для проведения опросов по любой тематике является электронный интервьюер РИСЭП [Социологическая информация в ЭВМ, 1983, с. 169—226], однако габариты, вес и стоимость не позволяют использовать его в массовых опросах. В перспективе возможно создание переносного электронного интервьюера (дисплей, на котором появляется инструкция респонденту и последовательно вопросы анкеты, и клавиатура для ответа) весом не более 1—2 кг [там же, с. 221—223]. Интервьюер может производить опрос с помощью прибора по месту жительства респондента, ответы респондента через радиопередатчик непосредственно вводятся в ЭВМ или записываются на магнитную ленту.

Такой способ опроса дает возможность применять более сложную анкету, чем при обычном анкетировании и сократить этапы тиражирования анкеты и перфорирования результатов опроса, но не устраняет затраты, связанные с необходимостью встретиться с респондентом (дорога, поиск квартиры, повторные визиты, если респондента нет дома и т. п.).

Радикальные изменения в технологии опросов связаны с развитием кабельного телевидения и (или) информационных сетей общего пользования, объединяющих домашние персональные компьютеры. Создание таких сетей планируется в некоторых странах в течение 5—10 лет. Проведение опросов станет возможным тогда, когда персональные компьютеры станут таким же распространенным предметом домашнего обихода как, например, телефон. Передача вопросов анкеты на экраны телевизоров или персональных компьютеров позволит полностью устранить все трудоемкие этапы социологического исследования.

Спустимся, однако, на грешную землю и посмотрим, какие средства минимизации затрат на подготовку данных и обработку их на ЭВМ существуют в настоящее время. Перфорация данных опроса, исправление ошибок и ввод в ЭВМ занимает 90—95% времени, затрачиваемого на обработку. Один из путей сокращения этого времени — подготовка данных непосредственно на магнитной ленте на специальном устройстве или на терминалах ЭВМ. Поскольку такие устройства обладают некоторыми вычислительными возможностями, контроль подготовки данных производится одновременно с переносом анкеты на магнит-

ную ленту. Обнаружив ошибку, оператор тут же вносит необходимые исправления, что позволяет сэкономить время, требуемое на поиск неправильно отперфорированной анкеты и на перебивку перфокарт, а также сэкономить сами перфокарты. Другой путь минимизации затрат на перфорацию — использование устройств, которые позволяют считывать рукописный текст или специальные знаки (например, почтовый индекс). Наличие такого устройства позволяет конструировать анкету так, что информацию с нее можно непосредственно вводить в ЭВМ.

Существенного повышения эффективности обработки информации можно добиться путем использования современных пакетов программ статистической обработки социологической информации и сочетанием различных средств вычислительной техники: существуют такие классы задач, для решения которых микро-ЭВМ и даже программируемые калькуляторы эффективнее, чем большие ЭВМ [Рукавишников, Паниотто, Чурилов, 1984, с. 152—174; Паниотто, Максименко, 1982, с. 220—225].

(Наконец, рассмотрим последний этап исследования — анализ и интерпретацию данных обработки. Повышение эффективности исследования на этом этапе может быть достигнуто за счет развития современных средств анализа социологической информации на ЭВМ (методы многомерного шкалирования, логлинейного анализа, детерминационного анализа и т. п. [Статистические методы анализа ..., 1979; Миркин, 1980; и др.]).

Но наиболее существенного влияния на повышение эффективности социологического исследования можно ожидать от развития методов имитационного моделирования или вычислительного эксперимента на ЭВМ. Причем речь идет не только об этапе анализа социологических данных, но о социологическом исследовании в целом. Практика социологических исследований в настоящее время ориентирована не столько на комплексное, системное (но пока поверхностное) описание социального объекта, сколько на углубленное изучение причинно-следственных зависимостей между отдельными интересующими исследователя переменными. Еще в 60-х годах такой подход критиковался Н. М. Амосовым, который одним из первых в нашей стране вместе с сотрудниками руководимого им отдела Института кибернетики АН УССР начал разработку имитационных моделей поведения личности и социальных групп [Амосов, 1968; Амосов, 1969]. Суть критики сводилась к тому, что сбор отдельных фактов и зависимостей не рационален и не позволяет свести полученные результаты в единое целое. Имитационное же

моделирование (называемое в работах Н. М. Амосова эвристическим моделированием или моделированием гипотез) отражает представление автора модели об изучаемом объекте и есть по сути системой гипотез. Вычислительный эксперимент позволяет выявить, какие из параметров модели в наибольшей степени детерминируют ее поведение, и указывает на недостающие данные, наиболее необходимые для описания объекта. Затем для сбора этих данных уже проводятся целенаправленные исследования. Полученные данные вносят коррективы в модель, опять проводятся вычислительные эксперименты, опять выявляются отсутствующие данные и т. д. Такой подход позволяет построить модель, все более согласующуюся с реальной действительностью, с тем чтобы в дальнейшем на модели можно было проигрывать различные варианты принимаемых решений и прогнозировать их социальные последствия.

Рискну завершить монографию следующим пророчеством. Существуют нововведения, на первый взгляд не очень заметные, которые впоследствии оказывают огромное влияние на всю человеческую культуру. К ним можно отнести появление книгопечатания (а казалось бы — просто несколько увеличилась скорость тиражирования книг). Таковы и перспективы ЭВМ, в том числе персональных, которые на наших глазах производят изменения в характере человеческой деятельности, последствия которых трудно оценить. Я думаю, что моделирование социальных процессов на ЭВМ окажет такое же глубокое влияние на развитие социологических исследований, на социологию в целом и на управление социальными процессами, как появление книгопечатания — на развитие цивилизации. Вычислительный эксперимент на модели дает возможность определить, какие ошибки в исходной информации допустимы при прогнозировании социальных последствий принимаемых решений. Таким образом, развитие моделирования предъявляет свои требования к качеству социологической информации. Можно предположить, что через 10—15 лет социологи будут пользоваться четкими критериями качества социологической информации и ни одно уважающее себя научное подразделение не сможет обойтись без создания модели изучаемого социального объекта, как сейчас не может обойтись без обработки на ЭВМ первичной социологической информации.

СПИСОК ЛИТЕРАТУРЫ

*Хорошая мысль, откуда бы она ни была
взята, гораздо лучше, чем собственная глупая.*

Ф. Ламот-Левайе

Маркс К., Энгельс Ф. Сочинения. 2-е изд., Т. 23. — М.: Госполитиздат, 1960. — 907 с.

Материалы XXVII съезда КПСС. — М.: Политиздат, 1986. — 352 с.

Материалы Пленума Центрального Комитета КПСС, 23 апреля 1985 г. — М.: Политиздат, 1985. — 31 с.

Горбачев М. С. Живое творчество народа. — М.: Политиздат, 1985. — 46 с.

Аванесов В. С. Проблема психологических тестов. — *Вопр. психологии*, 1978, № 5, с. 102—111.

Аванесов В. С. Тесты в социологическом исследовании. — М.: Наука, 1982. — 200 с.

Айзенк Г. Проверьте свои способности. — М.: Мир, 1972. — 176 с.

Алексеев А. Н. Социальные нормы производственной организации и жизненная позиция личности (из опыта «экспериментальной социологии»). — В кн.: Проблемы социального познания и управления. Томск, 1982, с. 161—170.

Алексеев И. С., Бородкин Ф. М. Принцип дополнителности в социологии. — В кн.: Моделирование социальных процессов. М.: Наука, 1970, с. 37—48.

Амосов Н. М. Моделирование сложных систем. Киев: Наук. думка, 1968. — 88 с.

Амосов Н. М. Метод моделирования социальных систем: Вопросы эвристического моделирования. — Киев: Ин-т кибернетики АН УССР, 1969. — Вып. 2. 124 с.

Андреева Г. М. Методологические проблемы экспериментального метода в социальных науках: Вступительная статья. — В кн.: Кэмпбелл Д. Т. Модели экспериментов в социальной психологии и прикладных исследованиях. М.: Прогресс, 1980, с. 5—26.

Баранов А. В. Контент-анализ и возможности прогнозирования. — В кн.: Методологические и методические проблемы контент-анализа. М.; Л.: ИСИ АН СССР, 1973, вып. 1, с. 38—40.

Белосвет К. Д., Паниотто В. И. Измерение в конкретно-социологическом исследовании. — *Филос. думка*, 1976, № 5, с. 83—93.

Бешелев С. Д., Гурвич Ф. Г. Математико-статистические методы экспертных оценок. — 2-е изд. — М.: Статистика, 1980. — 264 с.

Битинас Б. П., Паулавичюс Р. Б., Поникарова Т. М. Система статистической обработки социальных данных. — М.: Акад. нар. хоз-ва СССР, 1980. — 128 с.

Битинас Б. П., Паулавичюс Р. Б. Статистические методы эмпирического предсказания. — *Шяуляй: Шяуляйск. пед. ин-т*, 1981. — 52 с.

Блейхер В. М., Бурлачук Л. Ф. Психологическая диагностика интеллекта и личности. — Киев: Вища шк., 1978. — 142 с.

Бунич П. Г. Стимулирование труда в развитом социалистическом обществе. — *Социол. исслед.*, 1981, № 2, с. 25—33.

Бухалов А. Ю., Чурилов Н. Н. Некоторые вопросы расчета выборки в контент-анализе. — В кн.: Методологические проблемы использова-

ния математических методов в социологии. М.: ИСИ АН СССР, 1980, с. 182—192.

Водзинская В. В. О социальной обусловленности выбора профессий. — В кн.: Социальные проблемы труда и производства: Советско-польские сравнительные исследования. Москва; Варшава, 1969, с. 43—51.

Волович В. И. Надежность информации в социологическом исследовании: Проблемы методологии и методики. — Киев: Наук. думка, 1974. — 134 с.

Волков И. П. Социометрические методы в социально-психологических исследованиях. — Л.: Изд-во Ленингр. ун-та, 1970. — 88 с.

Воронов Ю. П. Методы сбора информации в социологических исследованиях. — М.: Статистика, 1974. — 160 с.

Гласс Дж., Стэнли Дж. Статистические методы в педагогике и психологии. — М.: Прогресс, 1976. — 496 с.

Головаха Е. И., Кроник А. А. Психологическое время личности. — Киев: Наук. думка, 1984. — 208 с.

ГОСТ 23554. 2-81. Система управления качеством продукции. Экспертные методы оценки качества продукции: Обраб. значений эксперт. оценок качества продукции. — Утверждено 30. IX. 1981, № 2079. — 82 с.

Гостковский З. О повышении достоверности опроса в социологических исследованиях. — Социол. исслед., 1978, № 4, с. 160—165.

Готтсданкер Р. Основы психологического эксперимента. М.: Изд-во Моск. ун-та, 1982. — 464 с.

Даль В. Толковый словарь живого великорусского языка. М.: ГИС, 1955, т. 2, — 780 с.

Докторов Б. З. О надежности измерения в социологическом исследовании. — Л.: Наука, 1979. — 128 с.

Докторов Б. З. Повышение возврата анкет при почтовом опросе. — Социол. исслед., 1981, № 3, с. 127—133.

Докторов Б. З. Метрологическая карта исследования общественного мнения. — Там же, 1984, № 1, с. 115—123.

Докторов Б. З., Паниотто В. И. Почтовый опрос в социологических исследованиях. — Філосо. думка, 1982, № 4, с. 42—51.

Докторов Б. З., Фирсов Б. М. Методические вопросы формализации социологических анкет. — Социол. исслед., 1975, № 3, с. 63—73.

Дэвид Г. Метод парных сравнений. — М.: Статистика, 1978. — 142 с.

Жабский М. И. Надежность социологических измерений. — Социол. исслед., 1982, № 2, с. 176—185.

Зайцева М. И. Методы шкалирования при измерении установки. — В кн.: Социальные исследования: Теория и методы. М.: Наука, 1970, вып. 5, с. 220—242.

Заке Л. Статистическое оценивание. — М.: Статистика, 1976. — 600 с.

Здравомыслов А. Г. Методология и процедура социологических исследований. — М.: Мысль, 1969. — 206 с.

Каганов А. Б., Шумяев В. С. Опыт проведения опросов с использованием видсотерминалов. — В кн.: Математико-статистические методы анализа данных в социологических исследованиях. М.: ИСИ АН СССР, 1980, с. 193—200.

Кара-Мурза С. Г. Формализованные системы оценки научных кадров. — Социол. исслед., 1981, № 2, с. 85—91.

Клигер С. А., Косолапов М. С., Толстова Ю. Н. Шкалирование при сборе и анализе социологической информации. — М.: Наука, 1978. — 112 с.

Ковалев В. Н., Чепуров В. М. Сбор и первичная обработка социологической информации. — М.: Наука, 1978. — 112 с.

логической информации с помощью прибора «Социометр-2». — Социол. исслед., 1976, № 4, с. 109—119.

Кокрен У. Методы выборочного исследования. — М.: Статистика, 1976. — 440 с.

Константиновский Д. Л. Динамика профессиональных ориентаций молодежи Сибири. — Новосибирск: Наука, 1977. — 140 с.

Кроник А. А. Межличностное оценивание в малых группах. — Киев: Наук. думка, 1982. — 160 с.

Кроник А. А., Головаха Е. И. Психологическое время: чувство возраста. — Знание — сила, 1984, № 3, с. 20—22.

Кэмпбелл Д. Модели экспериментов в социальной психологии и прикладных исследованиях. — М.: Прогресс, 1980. — 392 с.

Лекции по методике конкретных социальных исследований. — М.: Изд-во Моск. ун-та, 1972. — 202 с.

Матусевич В. А., Оссовский В. Л. Социальная микросреда и выбор профессий. — Киев: Наук. думка, 1982. — 140 с.

Матусевич В. А., Паниотто В. И. Типы профессиональных намерений. — Социол. исслед., 1978, № 3, с. 136—141.

Матулёнис А. Л. Включение молодежи в социальную структуру. — Вильнюс: Минтис, 1983. — 208 с.

Макаров М. Г. Категория «цель» в марксистской философии и критика теологии. — Л.: Наука, 1977. — 188 с.

Миркин Б. Г. Проблема группового выбора. — М.: Наука, 1974. — 256 с.

Миркин Б. Г. Анализ качественных признаков и структур. — М.: Статистика, 1980. — 320 с.

Михайлов С. Эмпирическое социологическое исследование. — М.: Прогресс, 1975. — 384 с.

Мейли Р. Структура личности. — В кн.: Фресс П., Пиаже Ж. Экспериментальная психология. М.: Прогресс, 1975, с. 196—277.

Налимов В. В. Теория эксперимента. — М.: Наука, 1971. — 208 с.

Налимов В. В. Непрерывность против дискретности в языке и мышлении. — Тбилиси: Изд-во Тбил. ун-та, 1978. — 110 с.

Ноэль Э. Массовые опросы: введение в методику демоскопии. — М.: Прогресс, 1978. — 380 с.

НТР и формирование духовного облика советского рабочего. — Киев: Политиздат Украины, 1982. — 328 с.

Осипов Г. В., Андреев Э. П. Методы измерения в социологии. — М.: Наука, 1982. — 184 с.

Паниотто В. И. Метод множественных сравнений. — Социол. исслед., 1980, № 1, с. 140—149.

Паниотто В. И. Структура межличностных отношений. — Киев: Наук. думка, 1975. — 128 с.

Паниотто В. И., Максименко В. С. Количественные методы в социологических исследованиях. — Киев: Наук. думка, 1982. — 272 с.

Паниотто В. И., Яковенко Ю. И. Некоторые способы совершенствования почтового опроса. — Там же, 1981, № 3, с. 134—138.

Паниотто В. И., Яковенко Ю. И. Разработка пакета программ для обработки данных социометрических исследований на ЭВМ. — В кн.: Математико-статистические методы анализа данных в социологических исследованиях. М.: ИСИ АН СССР, 1980, с. 201—208.

Панкова Л. А., Петровский А. М., Шнейдерман М. В. Организация экспертизы и анализ экспертной информации. — М.: Наука, 1984. — 120 с.

Петренко Е. С., Ярошенко Т. М. Социально-демографические показатели в социологических исследованиях. — М.: Статистика, 1979. — 166 с.

I Всесоюзное совещание по статистическому и дискретному анализу нечисловой информации, экспертным оценкам и дискретной оптимизации: Тез. докл. — Москва; Алма-Ата, 1981. — 440 с.

Погосян Г. А. Методы сбора информации в социологии. — Ереван: Изд-во АН Арм. ССР, 1985. — 214 с.

Прилуков Н. Н. О возможности использования ЭВМ на первых этапах социологического исследования. — В кн.: Опыт применения ЭВМ в социологических исследованиях. М.: ИСИ АН СССР, 1977, с. 154—158.

Проектирование и организация выборочного социологического исследования. — М.: ИСИ АН СССР, 1977. — 168 с.

Процесс социального исследования. — М.: Прогресс, 1975. — 576 с.

Психологическая диагностика: проблемы исследования. — М.: Педагогика, 1981. — 232 с.

Пэнто Р., Гравитц М. Методы социальных наук. — М.: Прогресс, 1972. — 608 с.

Рабочая книга социолога. — 2-е изд. — М.: Наука, 1983. — 478 с.

Ропак Н. И. Категория цели: проблемы исследования. — М.: Мысль, 1980. — 126 с.

Рукавишников В. О., Паниотто В. И., Чурилов Н. Н. Массовые опросы: (Методический опыт). — М.: Финансы и статистика, 1984. — 208 с.

Ряжских И. А. Опыт использования включенного наблюдения для изучения жизни производственного коллектива. — Социол. исслед., 1975, № 3, с. 91—99.

Саганенко Г. И. Социологическая информация: статистическая оценка надежности исходных данных социологического исследования. — Л.: Наука, 1979. — 144 с.

Саганенко Г. И. Надежность результатов социологического исследования. — Л.: Наука, 1983. — 142 с.

Саганенко Г. И. Эксперимент на сопоставление различных методов ранжирования качественных признаков. — Личность и ее ценностные ориентации: Информ. бюл. ИКСИ АН СССР, 1969, вып. 1, № 4 (19), с. 42—47.

Семья и воспроизводство структуры трудовой занятости. — Киев: Наук. думка, 1984. — 240 с.

Словарь прикладной социологии. — Минск: Изд-во Белорус. ун-та, 1984. — 268 с.

Сороко Э. М. Концепция уровней, отношение, структура. — Минск: Наука и техника, 1978. — 158 с.

Социологическая информация в ЭВМ. — Л.: Наука, 1983. — 236 с.

Статистические методы анализа информации в социологическом исследовании. — М.: Наука, 1979. — 320 с.

Территориальная выборка в социологических исследованиях. — М.: Наука, 1980. — 216 с.

Титма М. Х. Выбор профессии как социальная проблема. — М.: Мысль, 1975. — 118 с.

Толстой Л. Н. О переписи в Москве. — Полн. собр. соч. М.: Гос. изд-во худож. лит., 1937, с. 170—177.

Трубников Н. Н. О категориях «цель», «средство», «результат». — М.: Высшая школа, 1968. — 148 с.

Хмелько В. Е. Методика соціально-інформаційного забезпечення збалансованості робочих місць на промисловому підприємстві. — Філософ. думка, 1981 (а), № 6, с. 91—101.

Хмелько В. Е. Номинальная шкала личностных качеств: (Опыт построения базового тезауруса). — Социол. исслед., 1981 (6), № 2, с. 124—134.

- Хмелько В. Е. Паниотто В. И. Методика социально-информационного обеспечения сбалансированного совершенствования структур рабочих мест и кадров промышленного предприятия. — Деп. в ИНИОН АН СССР № 10747 от 11.08.82. — 168 с.
- Черкасов Г. И. Социальная эффективность: сущность и критерии. — Социол. исслед., 1978, № 1, с. 39—46.
- Черноволенко В. Ф., Оссовский В. Л., Паниотто В. И. Престиж профессий и проблемы социально-профессиональной ориентации молодежи. — Киев: Наук. думка, 1979. — 214 с.
- Чурилов Н. Н. Планирование выборки в социологических исследованиях (некоторые методологические и методические проблемы): Автореф. дис. ... канд. филос. наук. — М., 1980. — 24 с.
- Шибутани Т. Социальная психология. — М.: Прогресс, 1969. — 536 с.
- Шмерлинг Д. С., Дубровский С. А., Аржанова Т. Д., Френкель А. А. Экспертные оценки: Методы и применение: (Обзор). — В кн.: Статистические методы анализа экспертных оценок. М.: Наука, 1977, с. 290—382.
- Шубкин В. Н. Социологические опыты. — М.: Мысль, 1970. — 288 с.
- Шубкин В. Н. Пределы. — Новый мир, 1978, № 2, с. 194.
- Шубкин В. Н. Социология и искусство (в самопознании общества). — М.: Знание, 1980. — 56 с.
- Ядов В. А. Социологическое исследование: Методология, программа, методы. — М.: Наука, 1972. — 240 с.
- Яковенко Ю. И. Методика подготовки и проведения почтового опроса. — Киев, 1982. — 36 с. Деп. в ИНИОН АН СССР 04.05.1983, № 12878.
- Яковенко Ю. И. Почтовый опрос: применение в социологических исследованиях и пути совершенствования: Автореф. дис. ... канд. филос. наук. — Киев, 1984. — 24 с.
- Якушев В. М. О путях соединения хозяйственного механизма с соревнованием. — Социол. исслед., 1978, № 1, с. 94—106.
- Якушев В. М. «Пульсар»: Опыт социального эксперимента. — Там же, 1977, № 3, с. 150—161.
- Bradburn N. M., Sudman S. and associates. Improving interview method and questionnaire design. — San Francisco, London, 1979. — 260 p.
- Dick W., Hagerty N. Topics in measurement: Reliability and validity. — New York, 1971. — 190 p.
- Dillman A. Mail and telephone surveys. — New York, 1978. — 480 p.
- Guilford J. P. Psychometric methods. — New York, 1954. — 210 p.
- Koocher G. P. Bathroom behaviour and human dignity. — J. Person. Soc. Psychol., 1977, 35, N 2, p. 113—117.
- Middlemist D. R., Knowles E. S., Matter C. F. Personal space invasion in the lavatory: suggestive evidence for arousal. — Ibid., 1976, 33, p. 87—94.
- Mouton J. S., Blake R. R., Fruchter B. The reliability of sociometric measures. — In: The sociometry Reader. New York, 1960, p. 673—677.
- Regan D. T., Shaus E., Fasio R. Liking and the attribution process. — J. exp. soc. psychol., 1974, 10, p. 114—119.
- Shanks J. M. The current status of computer-assisted telephone interviewing. — Soc. Meth. Res., 1983, 12, N 2, p. 126—134.
- Shranger J. S., Patterson M. B. Self-evaluation and the relation of dimensions for evaluating others. — J. person. 1974, 2, p. 64—73.
- Standards for educational and psychological tests. — Washington, 1974. — 264 p.
- Sudman S. Reducing the cost of surveys. — Chicago, 1967. — 194 p.

ПРИЛОЖЕНИЯ

1А. Социологическая анкета

Уважаемый товарищ!

В Вашем институте проводится социологическое исследование, цель которого — совершенствование социально-психологического климата и системы материального и морального поощрения работников. Мы надеемся, что учет Вашего мнения и мнения Ваших товарищей при принятии управленческих решений позволит улучшить отношения у Вас в коллективе, повысить удовлетворенность работой, сделать более адекватной систему подведения итогов социалистического соревнования.

Чтобы работа эта удалась, очень важно, чтобы каждый сотрудник отвечал с максимальной добросовестностью и искренностью.

Анкета анонимная. Все ответы будут использоваться лишь в обобщенном виде. Свою фамилию указывать не нужно.

Наш интервьюер тов. _____ гарантирует тайну ваших ответов.

В какой мере Вас удовлетворяют следующие стороны вашей работы?

Обведите кружком номер выбранного Вами варианта ответа, пользуясь следующими обозначениями:

- 5 — вполне удовлетворяет
- 4 — больше удовлетворяет, чем нет
- 3 — насколько удовлетворяет, настолько же не удовлетворяет
- 2 — больше не удовлетворяет, чем удовлетворяет
- 1 — совсем не удовлетворяет
- 0 — для меня это не имеет значения

1 Условия труда

5 4 3 2 1 0*

- 2 Возможность применять свои знания, опыт, творческие способности
- 3 Степень уважения, которым пользуется в обществе Ваша профессия
- 4 Размер материального вознаграждения (зарплата, премия...)
- 5 Польза, которую Ваша работа приносит людям, обществу
- 6 Система оплаты труда
- 7 Система морального стимулирования
- 8 Организация труда
- 9 Выполнение производственных задач коллективом отдела
- 10 Как Вы лично справляетесь со своей работой
- 11 Степень самостоятельности при выполнении работы
- 12 Возможность участвовать в управлении работой отдела
- 13 Работа общественных организаций
- 14 Личные качества Ваших сотрудников
- 15 Отношения с товарищами по работе
- 16 Отношения с руководителем отдела
- 17 Возможность повышения квалификации
- 18 Перспектива повышения заработка

* Такие же варианты ответа у вопросов 2—22, 78—94, 96—118 данной анкеты, а также вопросов 74—93 анкеты строителей (Прил. 2), поэтому я их далее не привожу.

- 19 Перспектива улучшения жилищных условий
- 20 Порядок распределения жилплощади
- 21 Предоставляемые институтом условия для отдыха
- 22 Затраты времени и способы передвижения между местом жительства и местом работы
- 23 Какую работу Вы предпочитаете?
 - 1 — с относительно невысоким, но постоянным заработком
 - 2 — работу, на которой можно много заработать, но для этого надо работать с постоянным напряжением

Ниже приведен перечень некоторых качеств личности. Напишите, пожалуйста, против каждого качества номера тех сотрудников Вашего отдела (включая себя, а также руководителя отдела), у которых это качество выражено в наиболее высокой и в наиболее низкой степени (желательно указать по каждому качеству не более 5 лиц с высокой и 5 лиц с низкой выраженностью качества).

Качества личности	Лица, у которых данное качество выражено	
	в высокой степени	в низкой степени
24 Высокий профессиональ- 25 ный уровень		
26 Высокий культурный 27 уровень		
28 Ум, понятливость, сооб- 29 разительность		
30 Доброта, чуткость, 31 отзывчивость		
32 Трудлюбие, усердие, 33 добросовестность		
34 Порядочность 35		
36 Личное обаяние, 37 привлекательность		

Внимание! Просим Вас и в последующих вопросах при оценке Ваших сотрудников оценивать также себя и руководителя отдела.

Эффективность работы зависит как от творческих способностей человека, его знаний и опыта, так и от интереса к работе, от желания сделать работу как можно лучше. Попробуйте отнести Ваших коллег к одной из групп по сочетанию этих качеств (проставьте номера сотрудников в оставленной для этого строке):

- 38 Обладает необходимыми способностями и знаниями, работает с полной отдачей, с интересом _____

- 39 Стараются, работает с полной отдачей, но нет необходимых способностей или знаний, опыта _____
- 40 Есть и способности и знания, но нет желания работать _____
- 41 Нет ни необходимых знаний, ни способностей, ни желания работать _____

Как правило, товарищи по работе лучше, чем руководство, могут оценить работу своих коллег. В коллективе обычно знают, «кто чего стоит», тут не играют роли ни хорошие отношения с начальством, ни формальные звания и награды. Никто лучше Вас и Ваших товарищей не может определить «вклад», эффективность работы каждого члена коллектива.

Просим Вас оценить работу каждого из сотрудников, поместив его номер в одну из клеток следующей таблицы:

Номер членов коллектива, которые работают:				
42	43	44	45	46
с очень высокой эффективностью	с высокой	со средней эффективностью	с низкой	с очень низкой эффективностью

- 47 С кем из Ваших коллег Вы чаще всего советуется по производственным вопросам? _____
- 48 С кем из Ваших коллег Вы чаще всего советуется по личным вопросам? _____
- 49 Кто, по Вашему мнению, пользуется наибольшим авторитетом в коллективе? _____
- 50 Предположим, что Ваш коллектив формируется заново. Кого из Ваших коллег Вы бы оставили в нем? _____
- 51 Кого из Ваших коллег Вы наиболее охотно пригласили бы в гости? _____
- 52 Занимаемая работником должность не всегда соответствует его возможностям. Кого из Ваших коллег следовало бы в первую очередь повысить в должности? _____
- 53 Кого следовало бы понизить? _____
- 54 Представьте себе, что по тем или иным причинам (например, капитальный ремонт дома) Вам предстоит несколько лет жить

в коммунальной квартире. С кем из Ваших коллег Вы согласились бы жить в одной квартире в первую очередь? _____

55 Во вторую очередь? _____

56 В третью очередь? _____

57 Как Вы думаете, кто из Ваших коллег выберет Вас в качестве соседа в ответ на предыдущие вопросы? _____

58 С кем Вам меньше всего хотелось бы жить в одной квартире? _____

59 А кто укажет, что ему не хотелось бы жить с Вами в одной квартире? _____

60 Предположим, что судьба забросила Вас на необитаемый остров. Кого бы Вы взяли с собой, если бы можно было взять несколько человек? _____

61 А если одного? _____

62 Кого из Ваших коллег Вам меньше всего хотелось бы взять с собой? _____

63 Укажите, пожалуйста, Ваш номер _____

Ниже приведен ряд незаконченных предложений. Допишите их, пожалуйста, таким образом, чтобы они отражали Ваше мнение.

64 Я думаю, что товарищи по работе считают меня человеком...

(Другие незаконченные предложения — вопр. 65—76 — приведены вместе с кодировкой в анкете и прил. 1 Б)

77 Если взвесить все «за» и «против», то в какой мере Вас удовлетворяет Ваша теперешняя работа?

7 — полностью удовлетворяет

6 — в основном удовлетворяет

5 — больше удовлетворяет, чем нет

4 — насколько удовлетворяет, настолько же и не удовлетворяет

3 — больше не удовлетворяет, чем удовлетворяет

2 — в основном не удовлетворяет

1 — полностью не удовлетворяет

В какой мере Вас удовлетворяют следующие стороны Вашей жизни?

78 Состояние Вашего здоровья

79 Состояние здоровья Ваших родных и близких

80 Жилищные условия, домашний быт

81 Уровень материальной обеспеченности

82 Супружество, личная жизнь

83 Поведение Ваших детей, развитие их личных качеств (если у Вас нет детей, на этот вопрос не отвечайте)

84 Наличие и личные качества друзей

85 Условия для творческих занятий на досуге

- 86 Условия для отдыха, развлечения на досуге
- 87 Возможность быть независимым в мыслях и поступках
- 88 Возможности участия в общественной жизни, общественной работе
- 89 Возможности учиться, повышать образование
- 90 Условия для разностороннего развития своих сил, задатков, способностей
- 91 Возможности стать руководящим работником
- 92 Общественное признание, общественная оценка Ваших заслуг
- 93 Отношение к Вам окружающих Вас людей
- 94 Ваше собственное поведение, поступки по отношению к окружающим Вас людям
- 95 Как Вы считаете, достаточно ли у Вас опыта, знаний, квалификации для той работы, которую Вы выполняете?
 - 1 — мои возможности значительно выше, чем требуется для работы
 - 2 — возможности несколько выше
 - 3 — работа соответствует моим возможностям
 - 4 — возможности несколько ниже
 - 5 — мне существенно недостает опыта, знаний, квалификации

Укажите, пожалуйста, степень Вашего согласия с перечисленными ниже суждениями (см. сноску на с. 189).

- 96 Я считаю, что трудно найти работу, более интересную, чем моя
- 97 Если бы зарплата действительно соответствовала той работе, которую мне приходится выполнять, то я получал(а) бы значительно больше
- 98 Плохо разбираться во всем, что не касается твоей непосредственной работы, — явный признак ограниченности
- 99 Я никогда не сплетничаю
- 100 При разумной организации труда на нашей работе и при соответствующем стимулировании я бы мог (могла) выполнять свою работу вдвое быстрее, чем сейчас
- 101 Начальник, никогда не использующий свою власть, теряет уважение подчиненных
- 102 Я работаю не так хорошо, как мог (могла) бы, из-за того, что эта работа меня не очень интересует
- 103 Все-таки главное — это не деловые качества, а порядочность человека
- 104 Я никогда не рассказываю анекдотов
- 105 Работа — это, конечно, важно, но основную часть жизни человек проводит все-таки вне работы
- 106 Работа — это в первую очередь средство обеспечить себя и семью, сфера моих основных жизненных интересов вне работы
- 107 Я никогда не грублю близким людям
- 108 Я так устаю на работе, что уже нет сил заниматься чем-нибудь в свободное время
- 109 После работы — в транспорте по дороге домой, в очередях в магазине — мне тяжело отключиться и я продолжаю думать о работе
- 110 Я бы всегда платил(а) за проезд в транспорте, даже если бы не было проверок
- 111 Если я стою в очереди у кассы кинотеатра, а билеты кончаются, и ко мне подходит кто-нибудь и просит взять билет, то я никогда не отказываю — ведь если человек станет в очередь, то билеты ему не достанутся

- 112 Когда я покупаю яблоки или апельсины, то обычно прошу, чтобы продавщица выбрала получше
- 113 Идти вегетарианцев — питаться лишь одними растениями, чтобы спасти животных, — крайне наивны
- 114 Я никогда не говорю о людях плохо за глаза
- 115 Когда я говорю по телефону-автомату, а у автомата очередь, то так нервничаю, что едва соображаю, о чем говорю
- 116 Я всегда говорю только правду
- 117 Убивать зверей на охоте, где у них есть шансы спастись, гораздо благороднее, чем осуждать охотников и есть при этом мясо специально откормленных, обреченных на смерть животных
- 118 Лица, подбирающие бездомных кошек и собак, делают это, как правило, из чувства одиночества (вопр. 119—137 о социалистическом соревновании и о социально-демографических характеристиках респондента опущены)
- Наконец, еще один вопрос.
- 138 В какой мере Вас удовлетворяет то, как складывается Ваша жизнь? (варианты ответа такие же, как и вопр. 77)

Большое Вам спасибо!

І Б. Кодирование неоконченных предложений и открытых вопросов

- | | | |
|----|--|---|
| 64 | Я думаю, что товарищи по работе считают меня человеком ... | 4 |
| | положительные черты как работника | 3 |
| | положительные черты как личности | 2 |
| | отрицательные черты как работника | 1 |
| | отрицательные черты как личности | |
| 65 | Отношения с руководителем у меня, | 5 |
| | очень хорошие («отличные», «лучше быть не может» и т. п.) | 4 |
| | хорошие («нормальные», «хорошие», «удовлетворительные», «они меня устраивают» и т. п.) | 3 |
| | и да, и нет («хотелось, чтобы были лучше», «неплохие, но...» и т. п.) | 2 |
| | плохие («предвзято относится» и т. п.) | 1 |
| | очень плохие («хуже нискуда» и т. п.) | |
| 66 | Я мог (могла) бы работать значительно эффективнее, если бы... | 1 |
| | личностные характеристики, способности, потенциал | 2 |
| | желание, интерес к работе | 3 |
| | условия и характер труда: | |
| | содержание | 4 |
| | условия | 5 |
| | организация | 6 |
| | отношения с товарищами | 7 |
| | отношения с руководством | 8 |
| | зарплата | 9 |
| | жилье | |
| 67 | Сотрудники оценивают мою работу... | 4 |
| | очень хорошо | 3 |
| | хорошо, удовлетворительно | 2 |
| | плохо (посредственно, неудовлетворительно, отрицательно) | 1 |
| | очень плохо | |
| 68 | Когда я показываю сделанную работу руководителю, то он... | 4 |
| | очень доволен | 3 |
| | доволен, удовлетворен | |

	не очень доволен, находит ошибки	2
	очень недоволен («хватается за голову»)	1
69	Условия труда у нас...	
	очень хорошие (отличные)	5
	хорошие (вполне удовлетворительные)	4
	удовлетворительные, средние, нормальные	3
	плохие	2
	очень плохие	1
70	Отношения в коллективе сложились...	
	очень хорошие	5
	хорошие (благоприятные, дружеские, вполне удовлетворительные)	4
	удовлетворительные	3
	плохие («могли быть лучше», неровные)	2
	очень плохие	1
71	Успешная работа отдела зависит прежде всего от...	
	внешних, не зависящих от отдела причин (например, планирование работ дирекцией)	1
	личностных качеств руководителя	2
	психологического климата,	
	качеств сотрудников	3
	содержания труда	4
	условий труда	5
	организации труда	6
	вознаграждения за труд (зарплата, жилье)	7
72	Что касается возможности проявить себя в работе...	
	такая возможность есть	1
	ее нет (без указания причин)	2
	ее нет из-за:	
	содержания труда	3
	условий труда	4
	организации труда	5
	отношений с руководством	6
	отношений с товарищами	7
	низкой зарплаты	8
	необеспеченности жильем	9
73	Оценивая свой вклад в работу отдела, я могу сказать, что по сравнению с другими сотрудниками мой вклад...	
	выше	3
	такой же	2
	ниже	1
74	Когда я иду на работу, то обычно настроение у меня...	
	хорошее	3
	обычное	2
	плохое	1
75	Думаю, что по складу характера мне больше всего подходит работа...	
	такая же, какую выполняет	1
	другая	2
76	Честно говоря, мое отношение к работе связано прежде всего с тем, что...	
	положительное отношение к работе без указания причин	1
	интерес к содержанию труда	2
	отношения с коллегами, руководством	3
	интерес к вознаграждению за труд (зарплата, жилье)	4
	отрицательное отношение к работе без указания причин	

отсутствие интереса к содержанию труда	6
плохие отношения с коллективом, руководством	7
неудовлетворенность вознаграждением за труд	8

2 А. Социологическая анкета
«Ваше мнение о бригадном подряде».

Уважаемый товарищ!

На Вашем предприятии внедряются бригадные формы организации и оплаты труда. Для того, чтобы сделать эту работу более эффективной, целесообразно учесть общественное мнение коллектива. Поэтому мы и обращаемся к Вам с просьбой ответить на вопросы этой анкеты, касающейся бригадного подряда и Вашей работы на предприятии. От Вашей искренности и добросовестности при заполнении анкеты зависит обоснованность рекомендаций, которые будут разработаны на основе данных опроса.

Анкета анонимная. Все ответы будут использованы лишь в обобщенном виде. Свою фамилию указывать не нужно.

Просим прощения, что отрываем Вас от дел и заранее благодарим за сотрудничество!

С уважением, сотрудники исследовательской группы.

(Вопросы 1—41 об отношении к бригадному подряду и удовлетворенности работой опущены).

42 Можете ли Вы сказать, что Вы лично оказываете влияние на решение вопросов производственной жизни в своей бригаде?

определенно да	5
пожалуй, да	4
пожалуй, нет	2
определенно нет	1
затрудняюсь ответить	3

43 Как бы Вы определили отношение к бригадному подряду? Изменилось ли оно за последнее время?

с самого начала считал бригады делом стоящим и вижу, что не ошибся	5
раньше сомневался, а теперь вижу — дело стоящее	4
ожидал многого, но мои надежды не оправдались	2
с самого начала был против и убедился, что прав	1
затрудняюсь ответить	3

В ответах на приведенные ниже вопросы просим Вас указать номера членов бригады, включая себя и бригадира.

44 С кем из членов бригады Вы чаще всего советуетесь по производственным вопросам? _____

45 А по личным вопросам? _____

46 Предположим, что бригада формируется заново. Кого из членов бригады Вы бы оставили в ней? _____

47 Кого из членов бригады Вы наиболее охотно пригласили бы в гости? _____

48 Кто из членов бригады работает, на Ваш взгляд, лучше всех? _____

49 А кто хуже всех? _____

50 Кому из членов бригады присущи в наибольшей степени доброта, чуткость, отзывчивость? _____

- 51 Укажите, пожалуйста, Ваш номер _____
- 52 (совпадает с вопросом 77 анкеты инженеров — прил. I)
- 53 В какой мере Вас удовлетворяет Ваша жизнь вне работы?
(варианты ответа такие же, как в вопросе 77 анкеты инженеров)
- В какой мере Вас удовлетворяют следующие стороны Вашей жизни? (вопросы 54—73 в основном совпадают с вопросами 78—95 анкеты для инженеров — прил. I)
- Укажите, пожалуйста, степень Вашего согласия с перечисленными ниже суждениями (вопрос 74 совпадает с вопросом 97 анкеты для инженеров, вопрос 75 — с вопросами 106, 76 — с вопросом 100, 77 — с 109, 78 — с 112, 80 — с 113, 81 — с 114, 83 — с 115, 85 — с вопросом 116 той же анкеты, поэтому я эти суждения не привожу)
- 79 Важно формировать в детях умение жить, иначе они ничего в жизни не добьются (варианты ответа такие же, как в вопросе 1 анкеты для инженеров, см. сноску на с. 189)
- 82 Добрые люди встречаются часто
- 84 Когда нужно сделать что-нибудь хорошее, то все средства хороши
- 86 Когда возникают разногласия в чем-то не очень важном, то я обычно уступаю
- 87 Я не могу смотреть, когда в фильмах показывают физические страдания, убийства
- 88 Для того, чтобы достичь успеха в работе, нужно все время проявлять инициативу
- 89 Я работаю добросовестно, но общественная работа доставляет мне больше удовольствия
- 90 Проявлять инициативу и активность в работе нужно, но главное — хорошая организация труда и ритмичность поставки материалов
- 91 Я за новые формы оплаты труда, ибо старые плохо учитывают трудовой вклад каждого
- 92 Кто хорошо работает, у того много и общественных поручений, а кто плохо — тот не проявляет активности и в общественной работе
- 93 При современном уровне организации производства главное — дисциплинированность и исполнительность, личная инициатива и самостоятельность часто вредят работе
- 94 Чьи интересы, как подсказывает Вам жизненный опыт, чаще преследовали люди, с которыми Вы сталкивались в жизни?
- 1 — только свои личные, индивидуальные, эгоистические
 - 2 — в какой-то мере и интересы других людей, не совпадающие с собственными
 - 3 — в равной степени и свои интересы и интересы других людей
 - 4 — преимущественно интересы других людей
- 95 Когда Вам нужна была помощь и кто-то мог помочь, что чаще получалось?
- 1 — помогали
 - 2 — не помогали
 - 3 — и то и другое случалось одинаково часто
- 96 Многие ли у Вас берут в долг (неважно какую сумму):
- 4 — очень многие
 - 3 — многие
 - 2 — немногие
 - 1 — никто
- 97 Больше всего в людях я ценю такие качества как (укажите не более трех из перечисленных ниже качеств):
- | | |
|---|---|
| красота, обаяние, привлекательность | 1 |
| ум, сообразительность | 2 |

справедливость, объективность	3
мастерство, квалифицированность	4
чуткость, отзывчивость, готовность помочь	5
жизненный опыт, практичность	6
трудолюбие	7
другие качества (напишите, какие)	

(вопр. 98—118 (в основном, социально-демографические) опущены).
Наконец, еще несколько вопросов.

119 В какой мере Вас удовлетворяет то, как складывается Ваша жизнь? (варианты ответа такие же, как в вопросе 77 анкеты для инженеров — прил. 1)

120 Согласились бы Вы еще раз принять участие в опросе?

- 1 — да
- 2 — скорее да, чем нет
- 3 — трудно сказать
- 4 — скорее нет, чем да
- 5 — нет

121 Если да, то по какой причине?

- 1 — интересно отвечать на анкету
 - 2 — думаю, что результаты исследования помогут улучшить положение дел на предприятии
 - 3 — хочу помочь исследователям
- Другие причины (допишите)

Если у Вас есть какие-либо соображения о недостатках внедрения бригадного подряда и вообще о положении дел на строительстве, замечания и предложения, то, пожалуйста, напишите ниже. Мы их учтем при разработке рекомендаций для руководства предприятия.

Большое Вам спасибо!

2 Б. Вопросы, кодируемые исследователем

Вопрос 122.

Члены бригады оценивают трудовой вклад данного респондента как

очень высокий	4
высокий	3
средний	2
низкий	1

Вопрос 123.

Члены бригады оценивают степень доброты, чуткости, отзывчивости данного респондента как:

высокую	3
среднюю	2
низкую	1

3. Анкета «Юмористические картинки»

Уважаемый товарищ!

Ниже помещены юмористические рисунки *. Просим Вас оценить каждый из них, ответив на следующие вопросы.

1. Остроумный ли рисунок?

(укажите код одного из перечисленных вариантов ответа)

очень остроумный	4
остроумный	3
скорее да, чем нет	2
не остроумный	1

2. Понятен ли замысел художника?

полностью понятен	3
в основном понятен, но есть некоторые сомнения, все ли я правильно понял(а)	2
непонятен	1

3. Напишите, пожалуйста, фразу, несколько фраз или ключевые слова, объясняющие замысел художника.

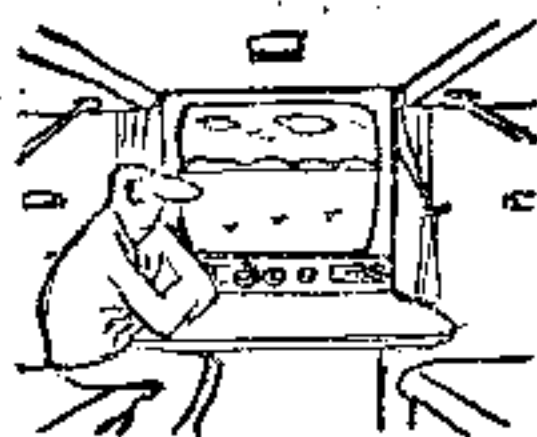
Пример:

№ рисунка	Остроумно ли	Понятно ли	Фразы или ключевые слова
А	1	3	Капля никотина убивает лошадь
Б	3	2	Переоборудовал квартиру под вагон или Цивилизация, урбанизация... уже и природа воспринимается как экран телевизора

* Использованы рисунки А. Бильжо (рис. 6 — Лит. газ., 1980, 5 ноябр.; рис. 21 — там же, 1981, 15 апр.), Д. Брейкша (рис. 27 — Звайзгне, 1969, № 7; рис. 23 — там же, 1971, № 23), Ю. Кособукина (рис. 17 — Лит. газ., 1980, 16 апр.; рис. 3 — там же, 29 окт.), И. Мелли-гайлиса (рис. 12 — Звайзгне, 1974, № 2; рис. 25 — там же, 1975, № 13), В. Пескова (рис. 19 — Лит. газ., 1980, 17 дек.; рис. 5 — там же, 19 ноябр.; рис. А — там же, 1981, 11 марта; рис. Б — там же, 25 марта; рис. 7 — там же, 15 апр.), Е. Милутки (рис. 2 — Лит. газ., 1980, 5 нояб.; рис. 10 — там же, 22 окт.), а также перепечатки из «Агеп», «Kvety», «Vozdi», «Szpilky», «Przyiazń», «Zeit und Bild», опубликованные в журнале «Звайзгне» (рис. 11 — Звайзгне, 1972, № 15; рис. 13 — там же, № 6; рис. 15 — там же, № 23; рис. 16 — там же, 1973, № 10; рис. 20 — там же, 1971, № 8; рис. 22 — там же, 1973, № 20; рис. 24 — там же, 1972, № 17; рис. 29 — там же, № 6).



A



B



1

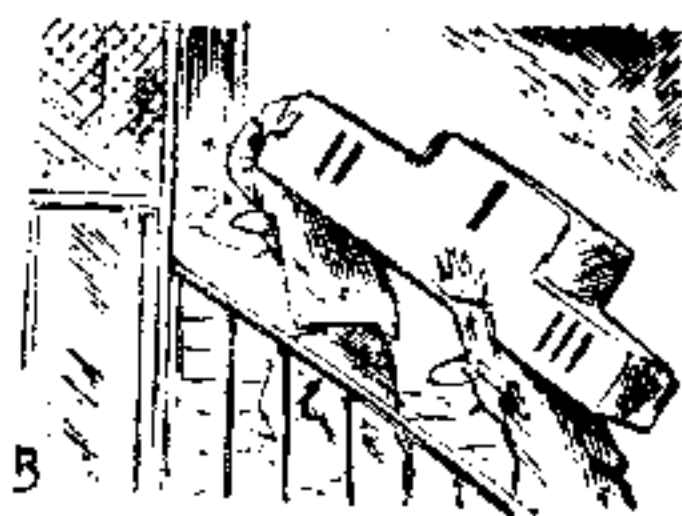


2

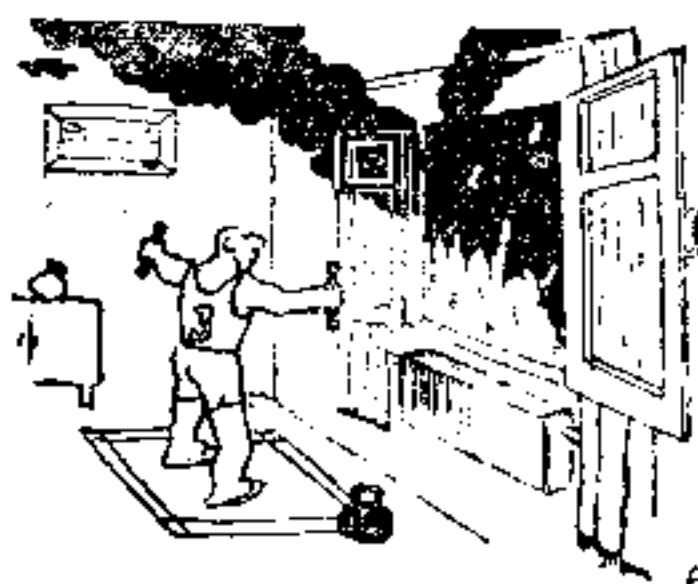


3

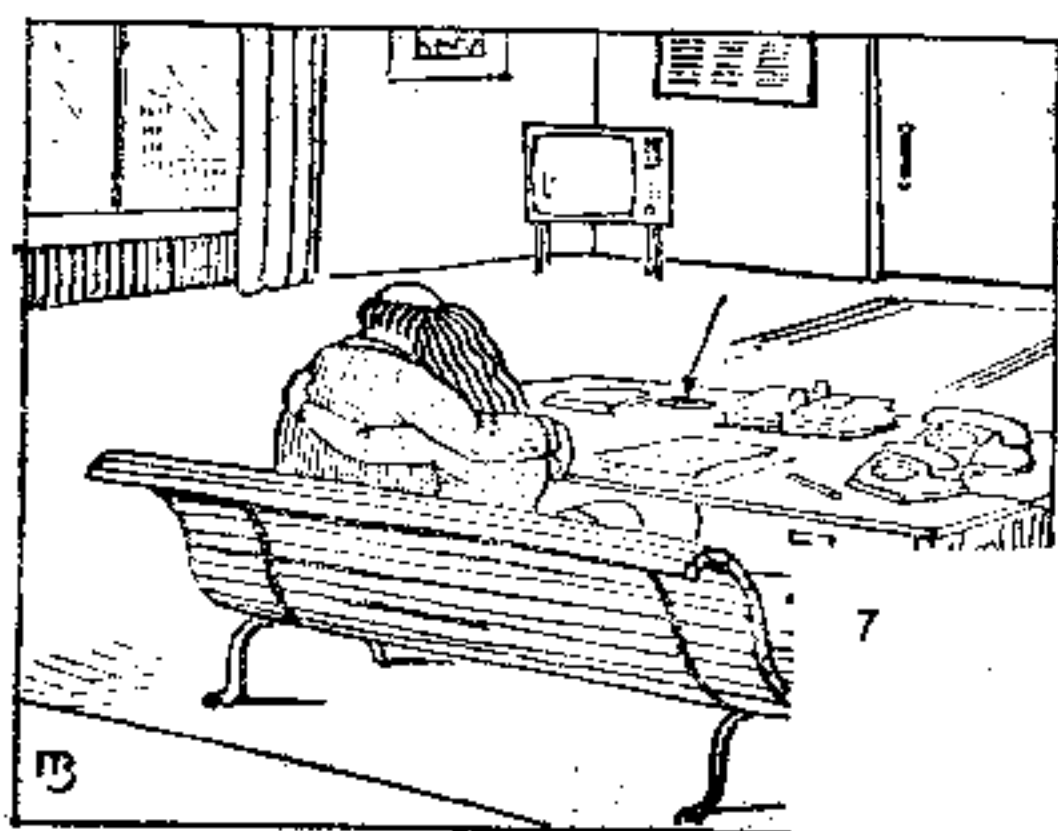




5



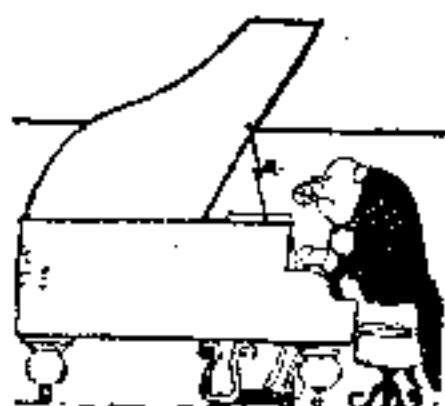
6



7



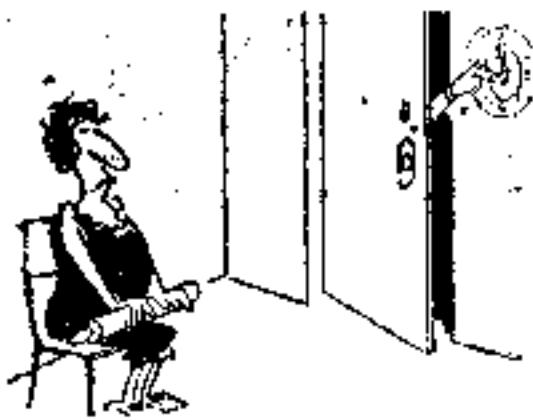
8



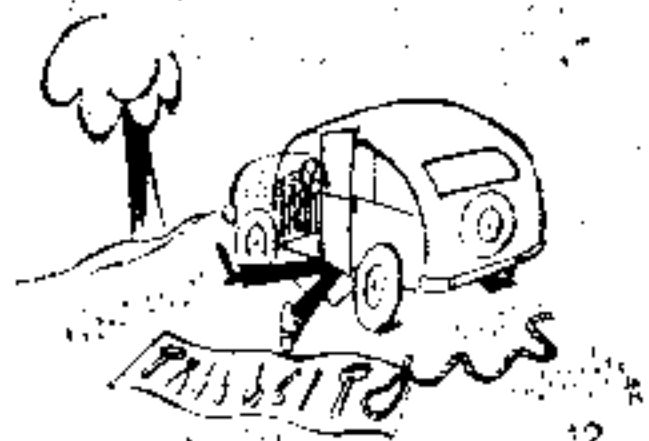
9



10



11



12



13



14



15



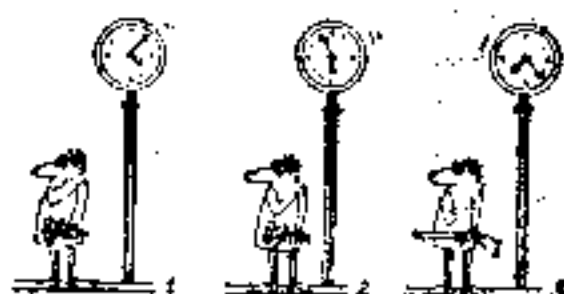
16



17



18



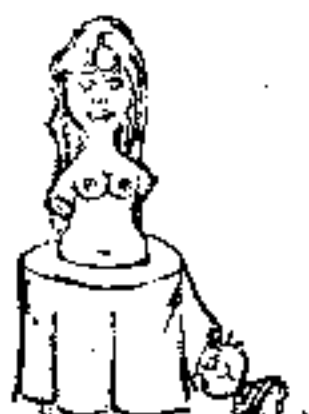
19



20



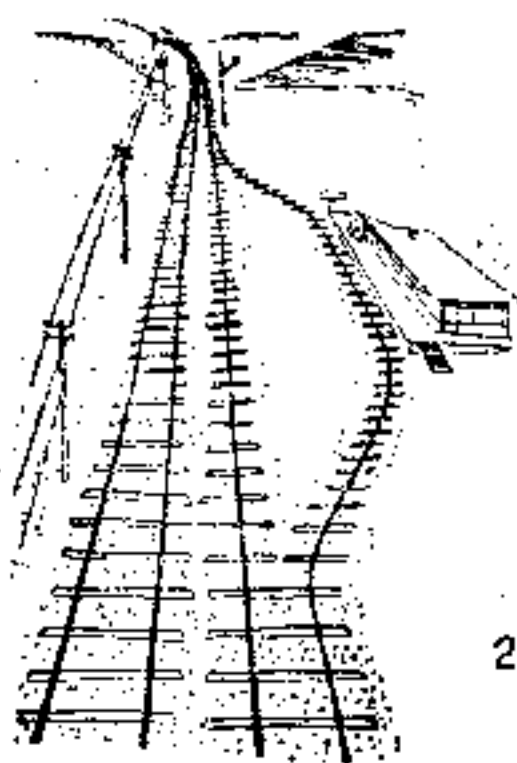
21



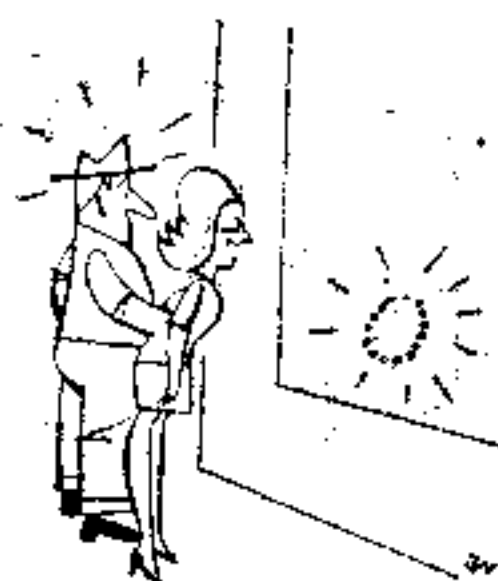
22



23



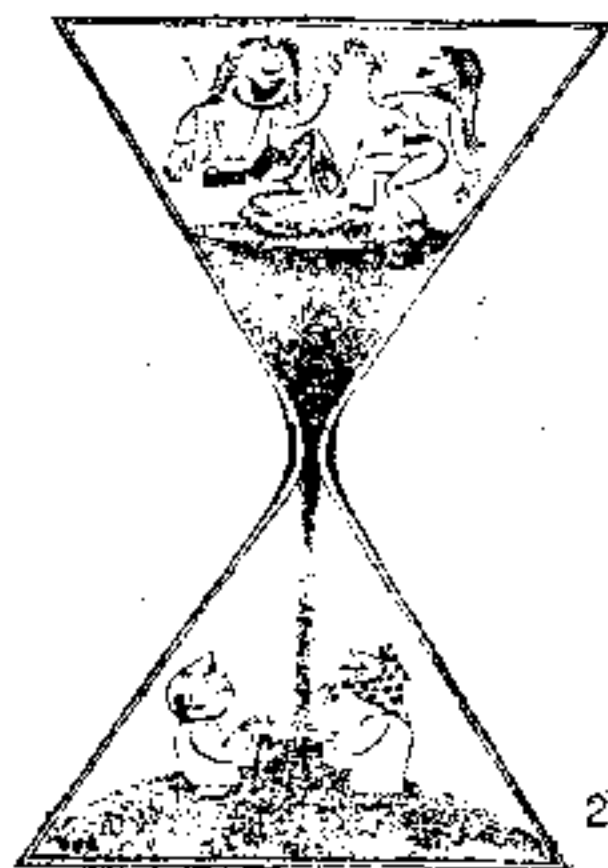
24



25



26



27



28



29

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

- Анализ эмпирических данных 18, 19, 20
- Важность результата исследования 17, 19, 20
- практическая 18, 19, 20
 - теоретическая 18, 19, 20
- Важность цели исследования 10, 12, 19, 20
- Валидизация
- группа 125, 126
 - критерий 111
 - документальный 113, 116, 131
 - конвенциональный 156
 - непосредственного наблюдения 113, 116, 131
 - объективный 113
 - предметно-фиксируемый 113, 115
 - приборный 114
 - самооценки 115-117
 - субъективный (субъектный) 113
 - экспертная оценка 115, 117
 - эталонный («последняя инстанция») 111, 112
- Валидность 22
- актуальная 104, 118
 - дифференциальная 109
 - документальная 113, 116, 131
 - конструктивная (концептуальная, или обоснованность свойствами) 107
 - наблюдаемая 113, 116, 131
 - объективная 115
 - область 108, 109, 126
 - определение 24-27, 102
 - очевидная (логическая) 109
 - по известным (контрастным) группам 104, 105
 - по критерию 103, 109-111
 - показатели 118, 119
 - α (коэффициент валидности) 120-124 - предсказательная (прогностическая) 103, 104, 118
 - приборная 114
 - самооценочная 115-117
 - содержательная 106
 - субъективная 113
 - теоретическая 109, 110, 126, 130, 149, 150
 - тип 110, 126
 - экспертная 115, 117
 - эмпирическая 103, 109-111, 126, 130, 149, 150
- Воспроизводимость 14
- Выборки (выборочная совокупность)
- гнездовая 87, 92
 - животная 89
 - многоступенчатая 85, 86
 - одноступенчатая 80, 87
 - производственная 86
 - районированная 88, 93
 - расслоенная 88, 93
 - систематическая 84, 89
 - случайная 79, 83, 84
 - стратифицированная 88, 93
 - территориальная 86
- Выборочное исследование 77, 78
- Вычислительный эксперимент 182, 183
- Генеральная совокупность 78
- Гнездовая выборка 87, 92
- Группа валидизации 125
- Гуманистичность цели социологического исследования 11, 161, 162
- Доброта (см. тест доброты)
- Достоверность результатов исследования 17
- Единица наблюдения 86
- Единица отбора 86
- группировка 87, 88
- Затраты на проведение исследования
- исследователя 14, 19, 20, 161-175
- Значимость результата исследования (см. важность результата)
- Значимость цели исследования (см. важность цели)
- Издержки респондентов (см. затраты исследуемого) изменения
- в объекте исследования 18, 20
 - в средствах исследования 18, 20
 - в субъекте исследования 19, 20
- Информация социологическая 8, 21
- Качество 7, 8
- внешнее определение 8, 11, 13, 14, 16
 - внутреннее определение 8, 12, 13, 16
 - процесса исследования 14, 15, 19, 20
 - результата исследования 15-21

социологической информации 7, 19, 20
 социологического исследования 9, 10, 19, 20
 средств 12-15, 19, 20, 162, 163
 цели 10-12, 19, 20
 Квазипогрешность согласования 26
 Квотная выборка 89
 Коэффициент интеллектуальности 126-129
 Критерий валидности (см. валидности критерий)
 Межличностные оценки (см. экспертные взаимные оценки)
 Многоступенчатая выборка (см. выборка многоступенчатая)
 Моделирование имитационное 182, 183
 Надежность 22, 35
 информация 21, 26
 методики 21, 26
 определение 25, 26, 37
 показатели 73, 74
 Нерелевантный ответ 68, 141
 Область валидности (см. валидности область)
 Обоснованность (см. валидность)
 Обработка социологической информации 18
 Одноступенчатая выборка (см. выборка одноступенчатая)
 Оперативность 163, 164
 Отбор (см. выборка)
 Ошибки
 намерения 27, 28, 35
 выборка 78, 79
 абсолютная 80
 относительная 80
 формулы расчета 80-82
 отбора 27, 35
 промах 29
 репрезентативности 27, 35
 систематические 29-32, 35, 78, 79
 случайные 29-32, 35, 78, 79
 теоретические 26, 27, 35
 Погрешности (см. ошибка)
 Почтовый опрос 178, 179

Правильность 31
 Промах 29
 Процедура Киша 85
 Процесс социологического исследования (см. качество процесса)
 Реалистичность цели исследования 12
 Результат социологического исследования
 основной 16, 19, 20
 побочный 16, 18, 19, 20
 Результативность средств исследования 13, 19, 20, 162, 163
 Релевантный ответ 68, 141
 Репрезентативность 78
 Сплошное исследование 78, 89
 Средство социологического исследования (см. качество средств)
 Телефонный опрос 178-180
 Тест
 доброты 146-149
 интеллектуальности 126-129
 редуцированный 141-142
 трудового вклада 132-146
 Точность (устойчивость) 32-35
 формулы оценки 53, 54, 56-60, 68, 69, 71-73
 Травматогенность 14, 174
 Трудоемкость 14, 175-182
 Устойчивость (см. точность)
 Целедостижения степень 16
 Ценность 16
 Чувствительность 35
 Экспертные (взаимные) оценки 130, 131, 134-136
 смещения 150, 157, 158
 Эмпирическая информация (см. информация социологическая)
 Этичность средств исследования 14, 174
 Эталон (см. валидность, критерий эталонный)
 Эффективность социологического исследования 14, 19, 20, 159
 социальная 14, 159-161
 экономическая 14, 159, 160

ОГЛАВЛЕНИЕ

Введение	3
I. Качество социологической информации. Понятийный аппарат	7
1. Социологическое исследование и социологическая информация	9
2. Обоснованность, репрезентативность, правильность и устойчивость социологической информации	21
II. Оценка надежности социологической информации	37
1. Методы контроля надежности социологической информации	38
2. Статистические показатели надежности социологической информации	53
III. Вопросы репрезентативности	77
1. Статистический подход к репрезентативности	77
2. Стандартизация описания выборки	86
3. Нестатистическое понимание репрезентативности	94
IV. Обоснованность социологической информации	102
1. Типы валидности	103
2. Показатели валидности	118
3. Методы валидизации	129
Эффективность социологического исследования (вместо заключения)	159
Список литературы	184
Приложения	189
Предметный указатель	205

ВЛАДИМИР ИЛЬИЧ ПАНИОТТО

**КАЧЕСТВО
СОЦИОЛОГИЧЕСКОЙ
ИНФОРМАЦИИ**

*(методы оценки
и процедуры обеспечения)*

*Утверждено к печати ученым советом
Института философии АН УССР*

Редактор *С. Н. Иващенко*

Художественный редактор *А. В. Косяк*

Технический редактор *А. М. Капустина*

Корректоры *Д. Я. Кашпер, И. В. Тоганенко*

ИБ № 7820

Сдано в набор 12.12.85. Подп. в печ. 29.05.86. БФ 06318.

Формат 84×108/32. Бум. тип. № 1.

Лит. гарн. Вых. печ. Усл. печ. л. 10,92.

Усл. кр.-отт. 11,24. Уч.-изд. л. 12,3.

Тираж 1450 экз. Заказ 3974. Цена 1 р. 90 к.

Издательство «Наукова думка».

252601 Киев 4, ул. Репина, 3.

Львовская областная книжная типография.

290000 г. Львов, ул. Стефаника, 11.

00363

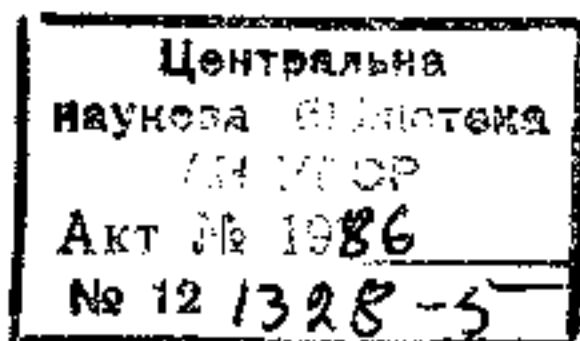
Монография посвящена разработке методов повышения качества социологической информации. Детально анализируется понятие «качество эмпирической социологической информации», рассматривается место этого понятия в системе понятий, характеризующих качество социологического исследования. Сделана попытка систематизировать характеристики качества социологической информации и предложить набор статистических показателей для их измерения.

Особое внимание уделяется обоснованности социологической информации, вводится понятие «эталона» как метода, который используется для проверки обоснованности других методов измерения и проводится типология социальных признаков по методам обеспечения обоснованности их измерения. Предложенные методы использованы для разработки тестовых методик определения трудового вклада работников, степени доброты и коэффициента интеллекта личности.

Для социологов, психологов, демографов, экономистов и других специалистов, которые имеют дело с измерением социальных признаков.

Ответственный редактор
В. Е. ХМЕЛЬКО

Рецензенты
В. И. ВОЛОВИЧ, Е. И. ГОЛОВАХА, Б. З. ДОКТОРОВ



Редакция философской и правовой литературы