

## **5. СПЕЦИФИКА ИСПОЛЬЗОВАНИЯ МЕТОДОВ АНАЛИЗА ДАННЫХ В СОЦИОЛОГИИ**

### **5.1. Необходимость соотнесения модели, «заложенной» в методе, с содержанием задачи**

Выше мы уже говорили о том, что любой математический метод предполагает адекватной реальности определенную модель того явления, которое с помощью этого метода изучается. Но любая модель – это лишь некоторое приближение к действительности. Рассмотрим более подробно вопрос о достаточности такого приближения для социологических задач анализа данных.

Одним из проявлений трудностей с формализацией наших представлений о социальных явлениях является то, о чем мы уже упоминали: если для решения какой-то задачи существует некоторый математический метод, то этот метод практически никогда не бывает единственным. Примером могут служить уже самые простые характеристики одномерных распределений. Так, вообще говоря, существует много мер средней тенденции (и разброса) частотного распределения значений любого признака. Выше уже говорилось о том, что для измерения связи даже между двумя номинальными признаками могут служить более сотни известных из литературы коэффициентов соответствующего плана. Еще большее разнообразие присуще сложным методам изучения многомерных распределений<sup>33</sup>. И за каждым методом «стоит» свое понимание изучаемого явления (средней тенденции, разброса, связи и т.д.).

Какой метод выбрать? Как сравнивать результаты применения разных методов? Эти и другие подобные вопросы встают практически перед каждым исследователем. И любой социолог, использующий хотя бы самые элементарные математические методы (скажем, рассчитывающий среднее арифметическое значение, моду, медиану какого-либо признака), зачастую фактически дает ответы на вопросы такого рода, даже если он об этом и не задумывается (а, к примеру, при использовании какого-либо относительно сложного метода выбирает с помощью ЭВМ вариант «по умолчанию»).

Все сказанное обуславливает особую остроту для социологии вопроса об адекватности модели, заложенной в том или ином методе, содержанию решаемой с помощью этого метода задачи (точнее, концептуальным представлениям исследователя о ее сути). Реализация процесса соответствующего соотнесения – задача социолога. И здесь вряд ли помогут советы представителей других наук. Ведь решение этой задачи требует обеспечения естественности используемого математического языка; вычленения из живой реальности моделируемых с помощью математики фрагментов; четкого выделения таких элементов используемых алгоритмов, которые имеют непосредственный «выход» на содержательные представления социолога об изучаемом явлении. Приведенное утверждение является достаточно общим и, вероятно, может показаться в какой-то степени очевидным. Однако лишь задавшись целью обязательного сопряжения формализма и содержания, можно прийти к тем многочисленным и (как нам представляется), далеко не столь тривиальным, утверждениям, которые можно считать конкретизацией высказанного положения применительно к реальным интересующим социологов методам.

Приведем несколько примеров.

Начнем, казалось бы, с самого простого – с расчета мер средней тенденции. В математике известно бесконечное количество таких мер. В руководствах, ориентированных на социолога, обычно рекомендуют три из них – те, которые были названы нами выше – среднее арифметическое, медиану, моду. Сейчас мы не будем принимать в расчет то, что, как хорошо знает каждый социолог,

далеко не для всех шкал могут быть использованы две первые меры. Рассмотрим случай, когда тип шкалы нас не ограничивает в выборе среднего (предположим, например, что мы имеем дело с интервальными шкалами). Для того, чтобы показать, что такой выбор может диктовать нам содержание задачи, позволим себе описать несколько эксцентричный пример, приведенный нами в [Толстова, 1990а, с.62-63].

Опишем некоторую задачу о моде в житейском смысле этого слова. Предположим, что модельер должен определить, какая длина должна быть у очередной модели женских юбок, выпускаемых какой-то фабрикой, и для этой цели опрашивает женщин рассматриваемого региона, просит их указать «любимую» длину. Если мы в качестве длины, рекомендуемой фабрике, укажем медиану соответствующего распределения, то тем самым окажемся перед риском выпустить неходовой товар: половина женщин решит, что юбка для них слишком коротка, а половина – что чересчур длинна. Покупать продукцию фабрики никто не захочет. А вот если в качестве меры средней тенденции мы используем моду, то удовлетворим женщин, выразивших наиболее часто встречающееся мнение.

Коротко укажем на другие известные из литературы примеры. Терстоун, предлагая свой хорошо известный (см. [Толстова, 1998]) метод построения шкалы для измерения установки, рекомендовал на последнем этапе процедуры, при расчете приписываемого каждому респонденту итогового балла, использовать медиану в качестве среднего значения весов тех суждений, с которыми этот респондент согласился (а не среднее арифметическое, хотя с формальной точки зрения его в данном случае можно было бы посчитать; правда, здесь мы используем определенный взгляд на тип получающихся шкал, который требует специального обсуждения).

В некоторых конкретных ситуациях может возникнуть потребность использования совершенно иных мер средней тенденции. Так, в [Дэйвисон, 1988] рассматривается задача изучения пространства восприятия респондентами некоторых объектов с помощью многомерного шкалирования. Предлагается способ

построения матрицы близости между объектами на основе своеобразного опроса респондентов. И для усреднения соответствующих мнений рекомендуется использовать среднее геометрическое.

Приведем еще один пример, где речь идет о более сложном (по сравнению с расчетом средних) методе анализа данных. Предположим, что мы хотим построить типологию изучаемых объектов, используя для этого какой-либо из алгоритмов многомерной классификации (напомним, что в соответствии с этими алгоритмами каждый классифицируемый объект задается как точка некоторого признакового пространства). В таком случае выбор алгоритма должен определяться нашими априорными представлениями об искомых типах. Так, если мы считаем, что каждый тип может быть представлен неким «центральным» объектом, вокруг которого «кучкуются» другие объекты того же типа (т.е. если все однотипные объекты близки друг к другу одновременно по всем рассматриваемым признакам и, вследствие этого, центральный объект может служить как бы «олицетворением» типа), то мы должны выбрать какой-либо из алгоритмов, направленных на поиск круглых «сгущений» в рассматриваемом признаковом пространстве. Если же мы отождествляем каждый искомый тип с тем, какова форма зависимости какого-либо из рассматриваемых признаков от остальных, то подобные алгоритмы в принципе становятся неприменимыми. В таких случаях надо использовать методы, позволяющие искать «длинные» скопления точек в признаковом пространстве, «олицетворяющие» упомянутые зависимости.

Более обстоятельное описание подобных ситуаций можно найти, например, в работах [Патрушев и др., 1980; Типология и классификация..., 1982; Математические методы анализа и интерпретация..., 1989, гл.1], где подробно говорится о той априорной модели, которую должен сформировать исследователь, желающий решать задачу типологии тех или иных объектов с помощью методов многомерной классификации (речь идет об априорных представлениях об искомых типах и о том, что, не имея таких представлений, исследователь рискует получить нелепые

результаты, поскольку в таком случае математика не может выполнять функции «орудия труда» социолога).

Ясно, что социолог должен уделять большое внимание анализу моделей, заложенных в используемых им методах. И это — одна из причин присутствия термина «социологический» в названии нашей работы. Но существуют и другие.

## 5.2. Связь разных этапов исследования друг с другом

Для того, чтобы использование математического языка обладало той естественностью, о которой шла речь выше, необходимо, чтобы применение математики было буквально вплетено в логическую канву исследования. Математика должна служить «орудием труда» социолога, а не играть роль инструмента «пришлепывания» к исследованию модного «бантика», не очень-то вяжущегося со всем остальным (что, к сожалению, очень часто бывает на практике). Для достижения этой цели недостаточно того сопряжения формализма и содержания, о котором мы только что говорили. Чтобы не оставлять за математикой лишь роль средства придания некоторого наукообразия работе социолога, необходимо учитывать, что корректность использования математического аппарата на любом из этапов исследования тесно связана с принципами реализации других этапов, в том числе и таких, в которых не задействованы никакие математические методы. Это требование конкретизируется в виде целого ряда положений. Из-за недостатка места мы упомянем только два, сопроводив их примерами использования в социологии сравнительно сложных методов анализа данных.

**Первое** — о связи измерения и анализа его результатов. В п.1.3 мы уже упоминали о целесообразности сопряжения самого понятия статистической закономерности не только с выбором собственно алгоритма ее нахождения, но и с тем, что такому выбору предшествует и, в первую очередь, с формированием используемых

понятий и способа их операционализации. А это – стадии процесса измерения. Там же, а также в п.2.2 речь шла о том, что выбор конкретного алгоритма анализа и интерпретация результатов измерения взаимно обуславливают друг друга. Конечно, серьезное обсуждение указанной связи невозможно без конкретизации соответствующих положений для тех или иных используемых в социологии алгоритмов, что требует рассмотрения последних и не входит в число наших задач. Тем не менее, приведем небольшой пример, чтобы пояснить, что мы имеем в виду.

Соответствующие соображения уже были описаны нами в [Типология и классификация..., 1982]. Осуществляя типологию респондентов на основе данных об их бюджетах времени, мы стоим перед выбором: можно считать, что количества минут, затраченных тем или иным респондентом на какие-то виды деятельности, могут нами восприниматься с точки зрения различий разностей между ними (например, можно считать осмысленными, естественным образом интерпретируемыми выражения типа  $120-80=50-10$ ); можно полагать, что нам важна только структура времяпрепровождения человека (и, как следствие, учитывать не указанные разности, а то, что 120 в полтора раза больше, чем 80, а 50 – в пять раз больше, чем 10), а можно «видеть» в рассматриваемых количествах минут лишь порядок их расположения по величине (в таком случае указанные выше разности и отношения для нас становятся содержательно бессмысленными; о соответствующих числах мы можем сказать только, что  $80 < 120$ ,  $10 < 50$ ). Каждый вариант означает свою интерпретацию результатов измерения. Что именно мы выберем – зависит от нашего априорного понимания типа респондента (и, значит, от реализации еще одного этапа исследования – первичного формирования проверяемых гипотез). Но наше решение определит то, какой алгоритм классификации мы выберем для построения требуемой типологии (ср. с.47–48).

Существует много других причин, обуславливающих неразрывную связь между измерением и анализом данных. В социологии практически никогда нельзя провести четкую границу между этими двумя понятиями. Так, наиболее интересные для социолога

переменные чаще всего являются латентными, их значения не поддаются непосредственному наблюдению. Такие переменные измеряются не в процессе первичного сбора (наблюдения) данных, а в процессе анализа некоторой полученной в результате непосредственного наблюдения информации (для этого используются такие методы, как факторный, латентно-структурный анализ, многомерное шкалирование, методы парных сравнений, методы одномерного шкалирования Терстоуна, Лайкерта и т.д.). Напротив, многие методы анализа интересуют исследователя, в первую очередь, как результаты определенного рода измерения некоторых переменных. К примеру, именно с соответствующей точки зрения социолог часто интерпретирует результаты многомерной классификации: номер класса рассматривается им как значение переменной, которую можно было бы назвать «тип объекта».

Неразрывность двух проблем – построения признакового пространства (т.е. выявления способа описания исходных объектов) и выбора алгоритма анализа соответствующих данных – косвенно подтверждается наличием довольно большого количества работ, посвященных предложению методов одновременного решения этих проблем для некоторых классов содержательных задач [Браверман и др., 1974; Применение факторного..., 1976; Типология и классификация..., 1982].

Сказанным мы, к сожалению, здесь вынуждены ограничить рассмотрение проблемы связи измерения и анализа данных, хотя эта проблема весьма важна и с теоретической, и с практической точки зрения, и требует более глубокой проработки.

**Второе** – о зависимости интерпретации результатов применения метода от концептуальных установок исследователя, от стоящих перед ним целей. Для примера вспомним наше обсуждение возможных подходов к построению многомерной типологии изучаемых объектов с помощью разных алгоритмов классификации (п.5.1). Если мы считаем, что каждый тип может быть представлен неким «центральным» объектом, к которому примыкают другие объекты того же типа и выбираем алгоритм, направленный на поиск круглых «сгущений» в рассматриваемом

признаковом пространстве, то для интерпретации результатов классификации можно будет рассчитывать координаты центра тяжести каждого из найденных классов. Этот центр, как мы упоминали, можно считать «олицетворением» класса. Если же мы отождествляем каждый искомый тип с тем, какова форма зависимости какого-либо из рассматриваемых признаков от остальных, то подобная интерпретация становится неприменимой. В таких случаях для интерпретации надо искать упомянутые зависимости.

Упомянем также пример, уже описанный нами в [Математические методы анализа..., 1989]. В этом примере в процессе рассмотрения той же задачи построения типологии респондентов рассказывается, каким образом представления социолога об искомых типах позволяют корректировать результаты формальной классификации с целью превращения ее в содержательно интерпретируемую типологию.

Будем считать, что приведенных примеров достаточно для того, чтобы сформировать хотя бы самые приблизительные представления о том, что мы имеем в виду, говоря о необходимости соотношения всех этапов исследования друг с другом. И представляется совершенно очевидным то, что такое соотношение может быть осуществлено только самим социологом. Ведь оно по существу означает определенную целостность, неразрывность всего социологического исследования.

### 5.3. Другие методологические принципы анализа социологических данных

Выше мы сформулировали два основных методологических принципа, соблюдение которых является необходимым для того, чтобы использование математики было эффективным: сопряжение формализма и содержания и органическая связь всех этапов исследования друг с другом. Можно было бы говорить еще о целом ряде подобных требований, носящих более частный характер:

необходимость выполнения некоторых принципов измерения интересующих социолога показателей; обеспечения определенной однородности той совокупности объектов, на которой «действует» наша предполагаемая закономерность; соблюдения некоторых принципов интерпретации результатов применения метода; выполнения определенных правил комплексного использования целой серии методов при решении практически любой социологической задачи и т.д. (некоторая «сводка» подобных принципов дана нами в [Толстова, 1991а,б]).

Раскрытие каждого из названных принципов требует серьезного рассмотрения. Все они многоаспектны, имеют сложную структуру. Их практическая реализация требует достаточно глубокого анализа концептуальных представлений социолога об изучаемом явлении, для чего требуется четкая формулировка самих этих представлений.

Так, говоря об *измерении*, мы должны давать себе отчет в том, какие именно элементы реальности собираемся отобразить в тех или иных математических конструктах (чаще всего – в числах); какова наша модель восприятия респондентом предлагаемых ему объектов (суждений и т.п.); какая именно интерпретация этих конструктов будет нами использоваться при их анализе и т.д. [Толстова, 1998].

Обеспечивая *однородность* подвергаемой анализу совокупности данных о наших объектах, необходимо задуматься о том, имеем ли мы право для всех интересующих нас респондентов использовать один и тот же инструмент измерения и одинаковым образом интерпретировать результаты последнего; можем ли мы считать, что формальный вид искомой закономерности должен быть одним и тем же для всей выборки; можем ли мы одинаковым способом интерпретировать результаты анализа и т.д. [Толстова, 1986, 1991а].

*Интерпретируя* результаты применения того или иного алгоритма анализа мы должны обеспечивать, чтобы эта интерпретация не противоречила интерпретации исходных данных; чтобы при ее осуществлении по возможности компенсировались бы те недостатки формализма, которые волей-неволей мы

вынуждены были игнорировать при измерении и выборе метода анализа («идеальная» формализация того, что интересует социолога, как правило, бывает невозможна) и т.д. [Интерпретация и анализ..., 1987, гл.1; Толстова, 1991a].

Продумывая вопрос об адекватности тех или иных методов измерения и анализа данных, понимая, что все они не в полной мере отражают то, что нужно социологу, последний часто приходит (или должен прийти) к выводу о том, что достаточно полное отражение интересующей его картины реальности требует *комплексного использования разных методов*. За каждым – свои плюсы и минусы. А будучи примененными в комплексе друг с другом, они могут дать вполне адекватное представление о действительности. Но здесь встает множество вопросов, связанных с глубоким анализом модели, заложенной в каждом методе, с разработкой принципов сравнения разных методов друг с другом и т.д. [Толстова, 1991a].

Полагаем, что сказанного достаточно для того, чтобы читателю стало ясно, почему (и в каком смысле) в заглавии нашей книги мы «привязываем» анализ данных именно к социологии.

\* \* \*

Итак, мы в самых общих чертах описали, что такое «анализ социологических данных». При этом мы не только активно использовали то, что о соответствующих вопросах говорится в литературе, но и изложили свое видение ряда положений. Последнее в особой степени касается роли термина «социологический» в интересующем нас словосочетании.

Выше коротко раскрыта роль методов анализа данных в социологии и рассмотрены основные методологические принципы их использования при изучении общественных процессов. Конечно, все изложенное раскрывает суть анализа социологических данных действительно лишь «в самых общих чертах». Поэтому, вероятно, не все сказанное выше стало читателю полностью понятно; отдельные положения, может быть, показались очевидными либо, напротив, слишком «заумными», оторванными от реальности.

Наше убеждение состоит в том, что все приведенные соображения имеют самое непосредственное отношение к практике, к обеспечению хорошего научного уровня любого эмпирического социологического исследования. И каждое сформулированное выше утверждение становится весьма нетривиальным, когда дело доходит до его воплощения в жизнь. Но показать это, равно как и разъяснить более подробно то, что, возможно, осталось неясным читателю, можно только на реальных примерах. Необходимы: рассмотрение реальных социологических задач; демонстрация того, как их решению может способствовать математический аппарат; подробный анализ процесса сопряжения каждого метода с концептуальными представлениями исследователя и т.д. В определенной мере об этом пойдет речь во второй части (особый упор будет сделан на анализ моделей, заложенных в рассматриваемых методах).