

**ПРОБЛЕМЫ ОБРАЗОВАНИЯ  
В ОБЛАСТИ МЕТОДОЛОГИИ И МЕТОДИКИ  
СОЦИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЙ**

**К ВОПРОСУ О ПРИНЦИПАХ ПОСТРОЕНИЯ УЧЕБНОГО КУРСА  
"АНАЛИЗ СОЦИОЛОГИЧЕСКИХ ДАННЫХ"**

**Ю. Н. Толстова**

*(Москва)*

Ниже коротко описываются и обсуждаются принципы, заложенные в программе названного курса, разработанной автором и реализованной в процессе чтения лекций для студентов социологических факультетов ряда вузов г. Москвы. Программа опубликована Институтом социологии РАН.

Курс "Анализ социологических данных" в методическом плане является одним из самых сложных из числа преподаваемых студентам-социологам. Объясняется это прежде всего тем, что существование дисциплины с таким названием пока наукой не "узаконено"<sup>1</sup>. Имеется довольно обширная совокупность статистических методов, так или иначе используемых в социологии и, на первый взгляд, "безразличных" к происхождению тех данных, которые с их помощью изучаются. Возникает естественный вопрос: так ли это? И соответственно: должно ли содержание рассматриваемого курса сводиться к простому изложению математической сути ряда статистических методов? Должен ли он быть просто фрагментом прикладной статистики, или же мы имеем дело с некоторой специфической дисциплиной, в каком-то смысле "привязанной" к социологии? Имеет ли смысл объединить все рассматриваемые -

мые методы таким образом, чтобы сделать их описание для социологов отличным от аналогичных описаний, предназначенных для студентов других специальностей?

Конечно, говоря об особенностях построения того или иного курса лекций, т. е. об особенностях подачи рассматриваемого материала студентам, в первую очередь надо опираться на специфику самого предмета, в данном случае - процесса использования математико-статистических методов в социологии. И наша дилемма такова: является ли указанный процесс механическим применением известных статистических подходов к очередному массиву данных (природа которых, вообще-то говоря, для математики безразлична), или же он включает в себя аспекты, сопряженные именно с социологией.

Мы придерживаемся второй точки зрения. Эффективное применение любого математического аппарата в социологическом исследовании предполагает постоянное сопряжение формализма с содержательным смыслом решаемой задачи. Это требование объясняется объективным состоянием науки: отсутствием в большинстве случаев адекватных методов, предназначенных именно для решения стоящих перед социологом задач (что, в свою очередь, обусловлено известными проблемами, связанными со сложностью формализации наших представлений о социальных явлениях).

Чтобы быть успешным, использование математики не должно являться самоцелью. Непосредственное применение того или иного алгоритма должно "вплестаться" в общую канву исследования, быть лишь одним из элементов общей исследовательской стратегии. Модель, "заложная" в каждом методе (а такая модель всегда предполагается, независимо от того, дает ли себе в этом отчет исследователь или нет), должна тщательно изучаться с точки зрения того, насколько она отвечает априорным концептуальным представлениям социолога об интересующем его явлении.

Указанная специфика процесса анализа социологических данных, конечно, не может не вносить определенного своеобразия и в методику преподавания. В этой связи отметим два, как нам кажется основных, момента. Именно они дают возможность унифициро –

вать способы методического изложения материала и тем самым придать определенную целостность курсу лекций для студентов.

Во-первых, само собой разумеется, что студентам нужно освоить те или иные методы сами по себе. Они должны хорошо разбираться в существе алгоритма, условиях его применимости и т. д. Однако каждый студент должен ясно понимать, зачем алгоритм может использоваться, для решения какой социологической проблемы он приспособлен, какие элементы формализма отвечают тому или иному представлению о сути изучаемого социологического явления. Для этого в процессе преподавания необходимо четко выделять те аспекты математических алгоритмов, которые требуют сопряжения с содержанием решаемой задачи; приводить конкретные примеры, когда соответствующие предположения отвечают (или не отвечают) концепциям исследователя.

Во-вторых, чтобы органично включить математический метод в общую исследовательскую стратегию, социолог, с нашей точки зрения, должен также тщательно рассмотреть весь процесс и каждый этап исследования отдельно (начиная с формулировки основных целей и заканчивая интерпретацией результатов применения формализма и изучением соответствующих "обратных" связей) с точки зрения того, как идеи, реализуемые на каждом из них, влияют на выбор элементов формализма, интерпретацию результатов его применения и т. д. И это в той или иной мере должно присутствовать при изложении основ каждого метода.

Составляя программу, автор имела в виду указанные моменты. Но, к сожалению, специфика "жанра" полностью не позволяет их раскрыть в тексте программы. Это, конечно, должно быть учтено при ее использовании. В частности, представляется, что при описании любого метода (и при изложении его на лекции, и в ответах студентов) целесообразно: четко формулировать ту задачу, которую позволяет решать метод; характеризовать соответствующие "вход" и "выход"; указывать основные элементы алгоритма, непосредственно связанные с содержательными аспектами решаемой с его помощью задачи (и особенно с интерпретацией исходных данных и результатов применения методов).

Сложность преподавания рассматриваемого курса усугубляется тем, что многие названные аспекты использования математических методов в социологии для большого числа алгоритмов остаются слабо изученными. Поэтому от преподавателя требуется самостоятельная творческая работа. Следует отметить, что при составлении настоящей программы (и, конечно, при чтении лекций и проведении семинарских занятий) автор использовала соответствующие разработки, осуществляемые в Институте социологии РАН и не нашедшие пока достаточного освещения в литературе. Вообще говоря, все, что касается связи математического аппарата с содержанием решаемой социологической задачи, можно отнести к зарождающейся в настоящее время *новой ветви научного знания*. Поскольку соответствующие положения пока слабо известны, то, во избежание непонимания отдельных пунктов программы, они излагаются в более подробном варианте, чем это принято.

Необходимо указать также, что программа предусматривает изучение достаточно ограниченного круга алгоритмов. В основном таких, которые предназначены для анализа номинальных данных, поскольку именно такого рода данные являются наиболее важными, наиболее существенными для социолога (обоснование этого предусматривается программой). Кроме того, предполагается, что студентам уже известны некоторые основы традиционного статистического анализа данных из предваряющего курса математической статистики (в соответствии с традицией, он обычно не включает так называемую статистику объектов нечисловой природы, в том числе методы анализа номинальных данных).

Для большинства методов анализа номинальных данных используется еще одно соображение, придающее, на наш взгляд, дополнительную целостность рассматриваемому курсу. Речь идет о прослеживании определенной связи между методами, устанавливаемой через "родство" заложенных в них моделей. Практическое прослеживание подобной связи может быть осуществлено посредством выработки единого основания для классификации всех рассматриваемых алгоритмов, основания, связанного с определенной типологией социологических задач.

Предлагаемое в настоящем курсе основание классификации методов опирается на то обстоятельство, что для социолога важно выделение таких ситуаций, когда отдельному значению (градации, альтернативе) номинального признака необходимо придать статус самостоятельной переменной. Это требование можно обобщить: самостоятельной переменной может отвечать не одно значение некоторого признака, а сочетание таких значений, каждое из которых соответствует своему признаку. Другими словами, все исходные номинальные признаки как бы "рассыпаются" на отдельные градации, которые затем по-разному комбинируются, т. е. на их основе строятся новые признаки, взаимоотношения которых далее изучаются. Каждый метод анализа номинальных данных предлагается рассматривать как метод поиска либо связей между разными группами альтернатив, либо групп альтернатив, определяющих некоторое поведение респондентов (задаваемое разными способами). Методы систематизируются в зависимости от отвечающих им способов агрегирования отдельных альтернатив в новые признаки.

Такое разделение методов было положено в основу читаемых автором для студентов-социологов лекций по анализу данных. Практика показала, что применение соответствующей логики делает изложение более доходчивым, чем то, которое является традиционным для социологической литературы. Кроме того, предлагаемый подход, как представляется, побуждает исследователя не забывать о существовании многих методов, весьма адекватных социологическим задачам, но мало используемых социологами. Чтобы внести ясность, коротко опишем предлагаемую схему с перечислением вписывающихся в нее методов. При этом отметим, что автор отнюдь не претендует на полную классификацию всех важных для социолога методов анализа данных.

Итак, в соответствии с предлагаемой типологией методов, выделяются методы типа " (альтернатива) - (альтернатива) "<sup>2</sup>, т. е. такие, которые позволяют изучать связь между отдельными значениями любых рассматриваемых признаков (примером является детерминационный анализ <sup>3</sup>); типа " (группа альтернатив) - (группа альтернатив) " (анализ фрагментов таблицы сопряженности<sup>4</sup>;

алгоритмы типа "пятна" и "полосы" <sup>5</sup>). Последнюю группу методов можно обобщить, а результаты такого обобщения условно назвать методами типа " (группа альтернатив) - (поведение объектов) ", где "поведение" (подчеркнем, не одного объекта, а целой совокупности, заданной рассматриваемой группой альтернатив; такое "поведение" в определенном смысле есть описание указанной совокупности, которое, в свою очередь, можно интерпретировать как характеристику некоторого типа объектов) понимается по-разному: как определенный каким-либо образом средний уровень заранее заданного "результатирующего" признака; как справедливость для рассматриваемой совокупности некоторой логической функции от элементарных формул типа  $P(a) = i$  (так называемых логических закономерностей; приведенное выражение означает: "Значение признака P для объекта a равно i") и т. д. (методы поиска детерминирующих сочетаний значений рассматриваемых признаков, поиска логических закономерностей <sup>6</sup>).

В обосновании необходимости "склеивания" отдельных значений разных (вообще говоря) признаков просматривается актуальность решения следующей проблемы социологического измерения: чтобы отразить латентные свойства объекта, мы вынуждены прибегать к сочетанию значений разных признаков, надеясь, что какое-то сочетание хотя бы частично явится индикатором определенного "поведения" объекта. Продолжая рассуждения, нетрудно понять, что та же специфика измерительных процедур может вызвать потребность объединять не только "надерганные" из разных признаков альтернативы, но и признаки в целом. В рамках нашей классификации выделим методы типа "признак - признак" (обычные коэффициенты парной связи); "признак - (группа признаков) " (регрессионный анализ, многие методы построения индексов; отметим, что при использовании регрессионного анализа решается также задача типа " (группа альтернатив) - (поведение объекта) "; это ярко демонстрирует так называемый номинальный его вариант <sup>7</sup>); " (группа признаков) - (группа признаков) " (канонический анализ <sup>8</sup>). Обобщение рассматриваемой проблемы измерения очевидно:

новым измеряемым признаком является здесь некоторая комбинация исходных признаков.

Еще одно замечание. Рассматриваемый курс тесно связан с курсом по теории измерений в социологии<sup>9</sup>. В процессе разработки соответствующих учебных программ используемые в социологии методы сбора и анализа данных (методы, так или иначе опирающиеся на использование математического аппарата) были условно разделены на две отвечающие названным курсам части. Однако условность этого деления носит принципиальный характер. Многие известные методы измерения (факторный, латентно-структурный анализ, многомерное шкалирование, метод парных сравнений) дают возможность получить значения измеряемых характеристик лишь в процессе анализа некоторых вспомогательных (первичных) данных. Для подобного измерения "при анализе" также справедливы все высказанные выше методические соображения. Но многие методы анализа данных далеко не в последнюю очередь интересуют социолога как методы определенного рода измерения. И это должно находить отражение в процессе применения метода (на всех его этапах). Скажем, результаты классификации часто приходится рассматривать как итог измерения некоторой латентной переменной, "олицетворяющей" тип объектов. Кроме того, процесс измерения органично связан с процессом анализа собранных в результате измерения данных: метод должен формально соответствовать типу используемых шкал; быть связан с моделями порождения данных, в том числе с моделями восприятия респондентом предлагаемых ему объектов (суждений); используемая при измерении интерпретация данных должна согласовываться со способами решения стоящих перед социологом задач.

Особенно ярко связь между измерением и анализом его результатов проявляется при интерпретации исходных данных. Такая интерпретация является, с одной стороны, органическим элементом процесса измерения (именно в процессе интерпретации четко определяется измеряемая реальность), а, с другой стороны, представляет собой основной фактор, влияющий на выбор метода анализа. Весьма важным примером и с методической, и с практической точек

зрения является предположение о том, что за наблюдаемой номинальной (порядковой) переменной "стоит" некоторый количественный признак и что номинальность (порядковость) - лишь следствие неумения исследователя достаточно точно его измерить. Это подразумевается при использовании многих традиционных методов анализа номинальных и порядковых данных, начиная с критерия "Хи-квадрат" для измерения связей между признаками. И хотя указанное предположение не всегда оправдано, социолог часто не думает об этом и подсознательно фактически опирается на него, пользуясь упомянутыми методами.

Другой пример родства измерения и анализа по существу обсуждался выше, когда речь шла об актуальности для социологии предложенной классификации методов анализа данных.

В заключение отметим еще один немаловажный методический момент. Опыт показывает, что студенты-социологи часто бывают настроены на "гуманитарный лад", "отталкивают" от себя любые математические методы, глубоко не вникая в них и, по-видимому, считая соответствующие знания для себя лишними. Представляется, что сформулированные выше положения могут помочь преодолению возникающих в связи с этим методических трудностей.

### Примечания

---

<sup>1</sup> Сам термин "анализ данных" в литературе понимается по-разному. См., например: *Толстова Ю. Н. Анализ данных // Социология. Словарь-справочник. Т. 4: Методы, методика, математика и статистика. М.: Наука, 1991. С. 7-11.*

<sup>2</sup> Используются как синонимы термины: альтернатива, градация, значение признака.

- 
- <sup>3</sup> Чесноков В. С. Детерминационный анализ социально-экономических данных. М.: Наука, 1982. Здесь и ниже даются ссылки на литературу только в тех случаях, когда соответствующие методы представляются нам сравнительно слабо известными социологам.
- <sup>4</sup> Интерпретация и анализ данных в социологических исследованиях. М.: Наука, 1987. Гл. 2.
- <sup>5</sup> Ростовцев П. С. Алгоритмы анализа структуры прямоугольных матриц "пятна" и "полосы" // Анализ нечисловой информации в социологических исследованиях. М.: Наука, 1985. С. 203-214.
- <sup>6</sup> Интерпретация и анализ данных в социологических исследованиях. М.: Наука, 1987. Гл. 6; Типология и классификация в социологических исследованиях. М.: Наука, 1982. Гл. 8,9; Лбов Г. С. Методы обработки разнотипных экспериментальных данных. Новосибирск: Наука, 1981. Отметим, что Институтом социологии РАН выпущена брошюра (Ливанова Т. Н. Методическое пособие по использованию программы AID3 системы OSIRIS /анализ взаимодействия или поиск структуры качественных данных/. М.: ИС АН СССР, 1990), где подробно описан процесс реализации на ЕС ЭВМ метода последовательных разбиений, раскрытый также в указанных выше работах. Хотя в наше время увлечения персональными компьютерами такое описание не является актуальным, тем не менее на наш взгляд работа не стала бесполезной для социолога, поскольку в ней, помимо правил обращения с ЭВМ, подробно раскрывается сущность самого алгоритма.
- <sup>7</sup> О номинальном регрессионном анализе см.: Аргунова К. Д. Номинальный регрессионный анализ. М.: ИС АН СССР, 1990; Типология и классификация в социологических исследованиях. М.: Наука, 1982. Гл. 9.
- <sup>8</sup> Интерпретация и анализ данных в социологических исследованиях. М.: Наука, 1987. Гл. 2. Канонический анализ крайне редко ис –

---

пользуется социологами, считающими его типично количественным методом. В действительности же этот подход является актуальным для анализа именно номинальных данных: он дает возможность осуществлять их оцифровку, изучать связи между признаками с так называемыми "совместимыми" альтернативами, эффективно находить веса признаков при формировании из них индекса и т. д.

<sup>9</sup> См.: *Толстова Ю. Н.* Теория измерений в социологии. Учебная программа. М.: ИС РАН, 1992; предполагается, что этот курс предшествует рассматриваемому.