**Лекция №7 Тестовая традиция в социологическом шкалировании**

1. Общее представление о тестах.
2. Основная идея психологического тестирования.
3. Социологические индексы, основные проблемы, возникающие при их построении.

Известно, что понятие теста принадлежит области психологии. Тестирование осуществляют для измерения глубинных, не поддающихся непосредственному измерению, психологических характеристик человека. На первый взгляд кажется, что социолог сталкивается со сходной задачей и поэтому использование теории тестов представляется для него весьма заманчивым. Однако сделать это непросто. Способы общения с респондентами, направленность получаемых результатов у социолога и психолога различны. В отличие от психолога социолог не может позволить себе достаточно тщательное изучение мнения каждого отдельного человека; ему, как правило, требуется опросить большое количество респондентов; его чаще всего интересуют характеристики не столько отдельных людей, сколько разных их совокупностей, и т.д. (о специфике использования тестов в социологии см., например, [Аванесов, 1982]; о постепенном становлении соответствующей социологической парадигмы см. [Девятко, 1991 а]).

Но несмотря на различие задач, стоящих перед социологом и психологом, тестовая традиция активно внедряется в социологию, хотя здесь множество специфических проблем. Для практического освоения соответствующих подходов необходимо эти проблемы глубоко проанализировать. Начнем с более подробного рассмотрения того, что из себя представляет тестовая традиция.

Cовокупность разработанных в психологии тестов весьма разнообразна. Они отличаются и по предмету исследования, и по используемым техническим приемам (см., например, [Ядов, 1995, с. 187]). Мы коснемся принципиальной схемы, лежащей в основе разработки большинства психологических тестов - схемы, хорошо изученной психологами, описанной математическим языком и давшей толчок развитию мощной ветви прикладной статистики, - факторного анализа.

Опишем интересующие нас соображения на примере, заимствованном из психологии, - области, где "родились" тестовый подход и факторный анализ.

Предположим, имеются две группы вопросов-задач (наблюдаемых переменных), требующих от отвечающего на них человека способностей соответственно к логическому мышлению и к художественному воображению. Подсчитав корреляции между нашими вопросами, мы, вероятно, придем к выводу, что результаты ответов на вопросы каждой из этих групп коррелируют друг с другом. Человек, получивший высокую оценку по одному из "логических" вопросов, наверное, получит такую же оценку и по второму, и по третьему. Человек, проявивший высокие способности при решении одной из задач "на воображение", вероятно, не менее успешно решит и другие задачи подобного рода. То же будет иметь место и для низких оценок (напомним, что наличие корреляции между двумя признаками, грубо говоря, означает, что с ростом значений одного признака растут либо убывают значения другого).

Для объяснения описанных корреляций можно выдвинуть гипотезу, состоящую в том, что имеются два латентных фактора, которые условно можно назвать "логические способности" и "художественное воображение", принимающие разные значения для разных людей. И корреляции между нашими наблюдаемыми переменными объясняются действием именно этих латентных факторов: человек с высоким уровнем логических способностей будет, как правило, хорошо отвечать на вопросы первой группы, с низким уровнем таких способностей - плохо. Аналогичное утверждение будет справедливо и для второго фактора.

Таким образом, исходя из сформулированной гипотезы, мы предполагаем, что все наблюдаемые нами изменения значений эмпирических признаков обусловлены изменением некоторых внутренних свойств этих объектов - значений латентных факторов. Предполагается, что совокупность этих факторов едина для всех наблюдаемых признаков. **Такие факторы** назовем **общими.** Измерить их непосредственно мы не можем. Более того, мы не знаем заранее в точности, что из себя эти факторы представляют, сколько их. Однако предполагаем, что в принципе они существуют и что респонденты могут быть сопоставлены друг с другом по их значениям этих свойств (подчеркнем, что сказанным мы утверждаем существование латентных переменных; напомним читателю, что для социолога подобные утверждения далеко не всегда очевидны).

Общие факторы имеют разное влияние на изменение того или иного наблюдаемого признака. Вес общего фактора, определяющий степень его влияния на изменение данного наблюдаемого признака, будем называть факторной **нагрузкой фактора на признак.**

Естественно предположить, что кроме тех изменений наблюдаемых признаков, которые вызваны действием общих латентных факторов, существуют индивидуальные изменения каждого наблюдаемого признака, вызываемые, например, случайными ошибками при их измерении. Причины, вызывающие невзаимосвязанные изменения исходных признаков, называются **специфическими,** или **характерными, факторами.** Таким образом, все причины изменений наблюдаемых признаков могут быть разделены на две составляющие: группу общих факторов и специфический фактор для каждого признака. [Статистические методы..., 1979, с. 213]

Итак, значения общих латентных факторов для какого-либо человека определяют ответы этого человека на рассматриваемые вопросы, или, как мы будем говорить, поведение этого человека. Именно действием указанных латентных факторов определяются все корреляции между нашими наблюдаемыми переменными. Это означает, что фиксация значений латентных переменных должна привести к ликвидации связи между наблюдаемыми признаками.

Другими словами, если мы зафиксируем, скажем, значение фактора "логические способности", то связи между отвечающими этому фактору наблюдаемыми переменными исчезнут. Возьмем только тех респондентов, которые имеют блестящие логические способности. Конечно, они в основном будут хорошо отвечать на наши логические тесты. Но могут встретиться и плохие ответы:

скажем, кто-то из отвечающих слишком переволновался и забыл какую-то элементарную формулу, знание которой предполагалось тестом, и т.д. Однако связи между ответами на разные логические тесты уже не будет, поскольку сбои в ответах будут определяться не логическими способностями респондентов, а случайными по отношению к таким способностям обстоятельствами.

Факторный анализ родился в психологии как способ поиска латентных факторов, подобных описанным. Именно соображения, лежащие в основе тестовой традиции, привели психологов к разработке специального аппарата, позволяющего проверять гипотезы о наличии подобных факторов: выделять во множестве наблюдаемых признаков "пучки" связанных друг с другом, усматривать за каждым таким "пучком" некоторый латентный фактор и находить его значения для каждого респондента. Собственно, факторный анализ - это просто четкое выражение идей тестового подхода (рождение ФА является хорошей иллюстрацией роли математического языка в развитии науки).

В связи со сказанным целесообразно заметить, что, хотя ФА - статистический метод и как таковой в принципе не может доказать наличие или отсутствие каких бы то ни было причинно-следственных отношений, тем не менее при его использовании мы часто имеем основания полагать, что латентная переменная олицетворяет собой причину, обусловливающую тот или иной уровень относящихся к ней наблюдаемых характеристик (хотя в практических задачах далеко не всегда бывает очевидным, что является причиной, что - следствием).

В социологии рассматриваемая традиция нередко проявляется в виде стремления социолога к построению так называемых индексов для измерения латентной установочной переменной. Соответствующая процедура сводится к следующему.

Социолог, понимая, что "лобовой" вопрос в анкете не работает (что и означает латентность переменной), но что в то же время соответствующее состояние респондента может выражаться в разных аспектах его вербального поведения, задает респонденту серию косвенных вопросов, "вращающихся" как бы "вокруг да около" того, что исследователя в действительности интересует. Каждому из этих вопросов отвечает своя наблюдаемая переменная. Значение латентного признака для конкретного респондента обычно получается в результате суммирования ответов этого респондента на указанные вопросы, т.е. суммирования значений наблюдаемых переменных.

Например, применительно к уже рассматриваемой нами латентной переменной "удовлетворенность работой" описанная процедура будет означать обращение к респонденту с просьбой сказать, устраивает ли его зарплата, симпатичны ли ему товарищи по работе, авторитетен ли для него непосредственный начальник и т.д. Другими словами, одну "большую" удовлетворенность мы как бы "разлагаем" на много "маленьких". Каждый вопрос в таких случаях чаще всего сопровождается веером возможных ответов, соответствующих, скажем, традиционной пятибалльной шкале от "полностью устраивает" до "совершенно не устраивает" и т.д. (вместо баллов от 1 до 5 могут использоваться баллы от 5 до 1, от - 2 до +2, от 1 до 3, от 1 до 7 и т.д.). Баллы, соответствующие ответам одного респондента, суммируются. Считается, что полученное число можно интерпретировать как результат измерения "общей" удовлетворенности этого респондента. Далее мы полагаем, что максимальной удовлетворенности работой отвечает совокупность максимальных баллов-ответов по всем вопросам, минимальной удовлетворенности - совокупность минимальных баллов-ответов, а в промежуточном случае - удовлетворенность тем больше, чем больше суммарный балл. Сумма "маленьких" удовлетворенностей составляет одну "большую".

(Отметим очевидный, но иногда не замечаемый исследователем момент: используя обсуждаемый способ шкалирования, мы тем самым полагаем, что, скажем, максимальные значения ответов на все рассматриваемые вопросы анкеты говорят о состоянии удовлетворенности работой, а минимальные - о состоянии неудовлетворенности; так что если в анкету включены одновременно вопросы типа: "Часто ли Вам задерживают зарплату?" и "Часто ли Вы получаете премию?", оба - с веером ответов от "Очень часто" до "Крайне редко", то в первом случае мы должны приписать перечисленным вариантам ответов баллы от 1 до 5, а во втором - от 5 до 1.)

Лайкерт первым предложил измерять латентную переменную путем построения индекса такого типа, о котором шла речь выше [Likert, 1932]. Он же предложил строить фрагмент анкеты, направленный на измерение латентной переменной, в виде так называемого кафетерия - таблицы, строкам которой отвечают наблюдаемые переменные, а столбцам - значения этих переменных.

Метод иногда называют методом суммарных оценок. Он широко известен. Его описание можно найти, например, в [Грин, 1966; Ядов, 1995; Осипов, Андреев, 1977]. Кратко охарактеризуем его суть и поясним, в чем состоит связь способа построения шкалы Лайкерта с тестовой традицией.

Приведем пример кафетерия. Предположим, что нас интересует удовлетворенность респондентов своей работой. Соответствующий фрагмент анкеты будет выглядеть следующим образом.

**Фрагмент гипотетической анкеты со шкалой Лайкерта**

##### Пожалуйста, выразите степень своего согласия со следующими суждениями:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Суждение | Степень согласия и отвечающий ей балл | | | | |
| Вполне согласен | Согласен | Затрудняюсь ответить | Не согласен | Совершенно не согласен |
| 5 | 4 | 3 | 2 | 1 |

Я с удовольствием хожу на работу +

Я уважаю своего начальника +

Мне нравятся товарищи по работе +

Меня устраивает зарплата +

При разных формулировках суждений и обращениях к респонденту наборы предлагаемых ответов тоже могут быть разными. Вместо набора ответов от <вполне согласен> до <совершенно не согласен>, конечно, могут фигурировать другие наборы: от <полностью одобряю> до <совершенно не одобряю>; от <часто посещаю> до <никогда не посещаю> и т.д.

Плюсы, проставленные в табл., означают ответы гипотетического респондента. Значение латентной переменной для каждого респондента будет равно сумме баллов, отвечающих степеням его согласия с рассматриваемыми суждениями, для нашего респондента - сумме (3 + 4 + 5 + 1 + :). Если количество суждений равно, например, 10, то возможные значения нашей латентной переменной будут варьировать от 10 (наименее удовлетворенный человек) до 50 (наиболее удовлетворенный).

Предложение организации опроса с помощью включения в анкету "кафетерия" само по себе вряд ли могло получить имя автора. Соответствующая идея как бы "носилась в воздухе". Но заслугой Лайкерта явилось то, что он: а) предложил некий критерий, который, во-первых, показывает, насколько правдоподобно предположение о самом существовании измеряемой одномерной латентной переменной, и, во-вторых, дает основания отобрать именно те наблюдаемые признаки (суждения), которые имеют отношение к тому, что мы измеряем (в том числе показал, что пятибалльная шкала приемлема для измерения этих признаков); б) дал некоторое "оправдание" тому, что в качестве значения латентной переменной берется именно сумма значений наблюдаемых и что получающуюся шкалу можно считать порядковой. Необходимо отметить, что аргументированные ответы оказалось возможным дать только в результате глубоких разработок, в том числе и математического плана. И в полной мере они были осуществлены усилиями ряда ученых, а не одним Лайкертом (с начала 30-х годов интерес к соответствующей проблематике был проявлен многими учеными, перечень работ см. в [Грин, 1966]).

Алгоритм построения шкалы Лайкерта предусматривает проведение некоторого пилотажного исследования, цель которого - отбор таких признаков, значения которых коррелируют с суммой значений всех остальных. Именно такие признаки предлагается включать в анкету, предназначенную для проведения основного исследования. Упомянутая корреляция и позволяет обеспечить положительные ответы на три первых упомянутых выше вопроса. Прежде чем показать это, остановимся на вопросе о том, что значит "признаки коррелируют".

Вычислив коэффициент корреляции между рангами проверяемого признака и суммой рангов всех остальных признаков, оценим, является ли он достаточно большим для того, чтобы можно было говорить о наличии соответствующей связи. Для этого зададимся каким-то пороговым значением: будем считать, что если этот коэффициент больше 0,8, то связь есть, если меньше - то ее нет. Подчеркнем, что здесь мы имеем дело с довольно типичной для социологии ситуацией, когда задание порогового критерия является чисто субъективным делом исследователя и обоснование соответствующего выбора может опираться только на эмпирический опыт социолога. Перейдем к обещанному рассмотрению того, как в рассматриваемом случае реализуется тестовая традиция.

Первый вопрос - о существовании одномерной латентной переменной. После указанного отбора останутся только такие наблюдаемые признаки, каждый из которых коррелирует с суммой остальных. Это означает, что для измерения латентной переменной будут использованы такие наблюдаемые, которые образуют связанный "пучок".

Можно показать, что при фиксации значения латентной переменной (т.е. при рассмотрении только таких респондентов, для которых сумма баллов, приписанных ими рассматриваемым суждениям, будет одна и та же), связь между наблюдаемыми переменными пропадает. Таким образом, мы можем считать, что гипотетический латентный фактор действительно обусловливает наблюдаемые связи.

Второй вопрос - о выборе адекватных наблюдаемых признаков. Та же связь, о которой мы только что говорили, свидетельствует и о том, что наши наблюдаемые переменные имеют отношение к одной и той же латентной.

Как мы уже отмечали, решая вопрос о том, те ли наблюдаемые переменные мы берем для того, чтобы через них выражать интересующий нас латентный фактор, необходимо убедиться не только в том, что эти переменные отвечают задаче по самой своей сути (по своему наименованию), но и том, что они измерены именно так, как надо. Лайкерт задумывался о том, корректна ли традиционная пятичленная шкала. Сначала он пытался использовать подход, в наше время называемый методом оцифровки, респондентам, давшим ответы <вполне согласен>, <согласен> и т.д., приписывают не баллы 5,4 и т.д., а некоторые числа, подобранные таким образом, чтобы результирующее частотное распределение было нормальным. Но потом удалось экспериментально доказать, что две результирующие шкалы (обе измеряющие нашу латентную переменную), одна - полученная на основе суммирования описанных выше пятибалльных оценок, а другая - на основе суммирования оценок-результатов описанной оцифровки, очень сильно коррелируют друг с другом. Если считать, что наши измерения латентной переменной отвечает порядковой шкале, то указанного обстоятельства оказывается достаточно, для того чтобы считать такие шкалы идентичными (большая корреляция говорит о сходстве порядков шкальных значений, полученных по нашим шкалам). Значит, имеет смысл пользоваться пятибалльной - более простой. К определению типа шкалы Лайкерта вернемся позже, пока будем считать ее порядковой.

Третий вопрос - о форме выражения латентной переменной через наблюдаемую. То, что о суммарной связи в рассматриваемом <пучке> наблюдаемых признаков мы судили по наличию корреляции между каждым признаком и суммой всех остальных, косвенно свидетельствует о пригодности именно суммы значений наблюдаемых признаков в качестве значения латентной переменной.

Имеется и более серьезное обоснование целесообразности суммирования результатов измерения наблюдаемых переменных. Оно базируется на изучении однофакторной модели ФА: на анализе тенденций изменения корреляции между латентным общим фактором и суммой баллов наблюдаемых признаков при стремлении количества таких признаков к бесконечности (соответствующие ссылки можно найти в названной выше работе Грина). Мы не будем рассматривать этот вопрос более подробно, поскольку он требует достаточного погружения в математику. Констатируем только, что это лишний раз демонстрирует нам роль математики в эмпирической социологии.

Ради объективности следует также заметить, что имеются работы, в которых высказываются серьезные сомнения в правомерности обсуждаемой аддитивной модели по отношению к конкретным латентным переменным (так, в [Сознание и трудовая:, 1985] именно в таком ракурсе рассматривается проблема измерения удовлетворенности человека своим трудом).

Перейдем к рассмотрению нашего четвертого вопроса - о типе получающейся шкалы. Представляется очевидной ее порядковость. Однако нередко имеется возможность полагать, что она интервальна. Попытаемся это обосновать. Соответствующие рассуждения близки к тем, с помощью которых мы доказывали интервальность установочной шкалы Терстоуна.

Наш порядковый признак может принимать большое количество значений (если, скажем, у нас 10 суждений, то суммарный балл изменяется от 10 до 50). Человеку трудно дифференцировать свои представления о таком количестве качественно различных состояний латентной переменной. И даже если расстояния между соседними баллами не равны, этим можно пренебречь, поскольку соответствующие различия будут очень малы с точки зрения возможности их четкой содержательной интерпретации. Будем поэтому считать их одинаковыми. Тем самым будем воспринимать эту шкалу как интервальную.

Описанная идея Лайкерта очень схожа с идеями, заложенными в ФА. Отличие состоит в том, что 1) здесь заведомо предполагается, что фактор только один (в ФА количество факторов не задается априори, а определяется характером статистических данных); 2) исходные признаки измеряются по порядковой шкале, соответствующая информация легко может быть получена от респондента (ФА, как мы говорили, предполагает интервальность исходных шкал); 3) анализ корреляционной матрицы (анализ совокупной корреляции всех признаков друг с другом) заменяется оценкой силы корреляции каждого из них с суммой значений всех остальных; 4) значение фактора определяется как сумма значений наблюдаемых переменных (в линейном ФА задействована взвешенная сумма; веса определяются характером данных и несут содержательный смысл, помогают интерпретировать найденные факторы). Можно сказать, что шкала Лайкерта в описанном варианте представляет собой эвристический, легко реализуемый "вручную" (без использования ЭВМ) и опирающийся на сравнительно легко получаемую от респондента информацию, подход, который в более серьезном, опирающемся на строгие математические гипотезы, виде заложен в ФА.

**Вопросы домашнего задания:**

1. В чем состоит сущность тестовой традиции?
2. В каком смысле теорию тестов и метод факторного анализа можно воспринимать как синонимы?
3. Почему тестовый подход в "чистом" виде неприменим в социологии?
4. Какова роль априорных исследовательских предположений при использовании тестового подхода?
5. В чем сходство и различие процесса факторного анализа и процесса рождения понятий?

**Рекомендуемая литература:**

1. Толстова Ю.Н. Измерение в социологии. М., 2007 С.90-127