

Глава третья

Психологические тесты и социологические шкалы

1. Математические проблемы теории тестов

Впервые идею тестов выдвинул английский биолог Ф. Гальтон. Ему же принадлежит заслуга создания техники изучения индивидуальных различий на основе использования статистического метода. Гальтон искал способ математического описания тех закономерностей, которым подчиняются индивидуальные различия. В качестве метода он применил вариационную статистику.

В своих исканиях он опирался на работы А. Кетле — одного из создателей современной статистики. В своей работе «Социальная физика» (1835) Кетле показал, что формулы теории вероятности позволяют обнаружить факт подчинения поведения людей некоторым закономерностям. При анализе статистического материала он получил постоянные величины, дающие количественное описание таких актов человеческого поведения, как вступление в брак или самоубийство.

В интерпретации полученных данных Кетле исходил из идеи «среднего человека» — некоего идеала, отклонение которого изображается нормальным законом распределения. Если средняя является постоянной, то за ней должна стоять некая реальность, что позволяет предсказывать явления на основе статистических законов. Чтобы найти эти законы, безнадежно изучать каждого индивида в отдельности. Следует изучать поведение большой массы людей, используя метод вариационной статистики.

Давая статистический анализ биографических данных, Гальтон предлагал использовать законы Кетле для распределения человеческих способностей. Он считал, что существует средняя величина умственных способностей, так же как и средний рост.

Испытания, проводившиеся в антропологической лаборатории, организованной Гальтоном на Международной выставке в Лондоне в 1884 г., он назвал психологическими тестами. В широкий оборот этот термин вошел после выхода в свет в 1890 г. статьи его ученика — американского психолога Дж. М. Кеттелла—«Психологические тесты и измерения». В этой статье Кеттелл выдвинул требование статистического подхода — применения серии тестов к большому числу индивидов, полагая, что психология не сможет стать такой же точной наукой, как науки физические, если не будет базироваться на эксперименте и измерении.

Определенный сдвиг в этом направлении, по его мнению, может быть сделан путем применения серии психологических тестов к большому числу индивидов.

Разработанные Гальтоном приемы вариационной статистика вооружили психологию важным методическим средством. Среди них наиболее перспективным оказался метод исчисления коэффициента корреляции между переменными. Этот метод, усовершенствованный английским математиком К. Пирсоном и другими последователями Гальтона, внес в психологическую науку ценные математические методики, в результате использования которых возник в дальнейшем факторный анализ (работы Ч. Спирмена, Э. Л. Торндайка, Л. Терстона и др.). Постепенно совершенствовались методы статистической обработки результатов, контроля их надежности, обоснованности.

Начиная с 1905 г. французский психолог Л. Бине в сотрудничестве с врачом Т. Симоном выполнили серию экспериментов по изучению умственных способностей людей различных возрастов. Проведенные на многих испытуемых (и тем самым подчиненные статистическим критериям), эти эксперименты превратились в тесты для определения уровня умственного развития. Индивидуальные различия испытывались уже не сами по себе, а в их отношении к возрастному ряду. Так возникла метрическая шкала интеллекта Бине—Симона, которая явилась первым стандартизированным тестом не только по унификации заданий и процедур их выполнения, но и по оценке полученных показателей. Это побудило ввести понятие умственного возраста — МА (Mental Age) в отличие от хронологического возраста — СА (Chronological Age). Их несовпадение считалось показателем либо умственной отсталости (МА ниже СА), либо одаренности (МА выше СА).

В 1912 г. немецкий психолог В. Штерн ввел понятие «коэффициент интеллекта» Intelligence Quotient — IQ как показатель темпа умственного развития, свойственного данному ребенку:

$$IQ = \frac{MA}{CA} \cdot 100.$$

Коэффициент указывает на опережение или отставание умственного возраста относительно хронологического.

Л. Терман, профессор психологии Стэнфордского университета в США, вместе со своими сотрудниками дважды (в 1916 и 1937 гг.) осуществил переработку шкалы Бине—Симона. Так возникла известная шкала Стэнфорд—Бине.

В связи с тем, что интеллект, как полагали, есть нечто увеличивающееся вместе с возрастом ребенка, и так как согласно всем предварительным стандартным оценкам чем старше ребенок, тем лучше его характеристики почти в любом единичном умственном тесте, то вполне естественно возраст был воспринят в качестве психологической меры интеллекта.

Кроме индивидуальных тестов, подобных тестам Бине, применяющих возраст в качестве измеряющей шкалы, существуют также групповые тесты, которые применяются более широко и используют в качестве единиц измерения тестовые вопросы. Тестовый балл — это взвешенное и невзвешенное число вопросов, на которые правильно ответил индивид с некоторой поправкой на случайность правильного ответа, если это необходимо. Тесты обычно состоят из большого числа разнородных вопросов, меняющихся по трудности от очень легких до очень трудных.

В настоящее время теория тестов представляет собой хорошо разработанную дисциплину, включающую в себя целый ряд проблем. При составлении тестов, анализе и интерпретации результатов учитываются пять основных требований, по Галликсену¹: 1) составление и отбор тестовых вопросов; 2) присвоение балла каждому индивиду; 3) определение точности (надежности или ошибки измерения) тестовых баллов; 4) определение предсказывающего значения тестовых баллов (обоснованности или ошибки оценивания); 5) сравнение этих результатов с результатами, полученными при использовании других тестов или других групп испытуемых.

Нас в первую очередь будут интересовать вопросы, связанные с математической стороной техники тестирования.

Наиболее общим видом теста является тест, в котором каждый вопрос представляет определенную задачу. Балл — это число «правильных» ответов в ограниченное время. Вопросы могут быть неодинаковыми по трудности или же концентрироваться на одном уровне трудности. В последнем случае одни испытуемые имеют более высокую вероятность ответов, чем другие и, следо-

¹ Gulliksen H. Theory of Mental Tests. N. Y., 1950, p. 2.

вательно, более высокий уровень способности. Если вопросы имеют разные уровни трудности, то некоторые испытуемые могут ответить на более трудные вопросы и тем самым получить более высокий балл. Временной предел является еще одним источником разброса баллов, так как разные испытуемые работают с разной скоростью.

Самый обычный тип тестового балла — суммарный балл. Он основан на суммировании ответов определенной категории — правильных или неправильных.

Если мы предложим каждому из испытуемых тест из l вопросов, то результаты сможем представить в матрице вопросных баллов, в которой каждый столбец представляет соответствующий вопрос i , а каждая строка соответствующего испытуемого j .

Вопросы													
	a	b	c	i						$\sum_{i=1}^n sj = xt$			
1	1	0	1	0	0	0	0	0	0	0	0	0	2
2	1	1	1	0	0	1	0	0	0	0	0	0	4
3	1	1	1	1	0	0	0	0	0	0	0	0	4
4	1	1	0	1	1	0	0	1	0	0	0	0	5
5	1	1	1	1	1	0	0	0	0	0	0	0	5
•	1	1	1	0	1	1	1	0	0	0	0	0	6
•	1	1	1	1	1	1	1	0	0	0	0	0	7
j	1	1	1	1	0	1	1	1	1	1	0	0	9
•	1	1	1	1	1	1	1	1	1	1	1	0	11
N	1	1	1	1	1	1	1	1	1	1	1	1	12
$\sum_{j=1}^N$	10	9	9	7	6	6	5	4	3	3	2	1	$65 = \sum X_t$
	1,0	0,9	0,9	0,7	0,6	0,6	0,5	0,4	0,3	0,3	0,2	0,1	$6,5 = M_x$

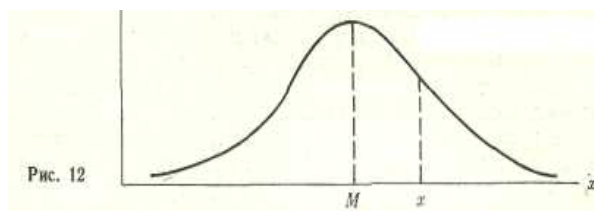
Каждый испытуемый пытается ответить на каждый вопрос, и поэтому каждый вопросный балл a_{ji} будет либо 0, либо 1. Сумма вопросных баллов в строке — это общий балл испытуемого, сумма вопросных баллов в столбце — это число испытуемых, правильно ответивших на вопрос этого столбца. Таким образом, матрица является источником информации как относительно вопросов, так и относительно испытуемых. Поделив каждую сумму вопросных баллов в столбцах на N , мы получим долю испытуемых p_i , правильно ответивших на соответствующий вопрос.

Доля тех, кто правильно ответил на определенный вопрос, является простым способом описания трудности этого вопроса. Фактически это есть средняя баллов, и чем она ниже, тем труднее вопрос.

Необходимо отметить, что сумма вопросных средних M_i равна средней общих баллов испытуемых, т. е.

$$M_x = \sum M_i = \sum p_i$$

где M_x — средняя общих баллов в тесте x ; M_i — средняя баллов в каждом вопросе; p_i — доля испытуемых, правильно ответивших на вопрос.



Дисперсия вопроса i определяется из уравнения

$$\sigma_i^2 = p_i q_i$$

где p_i — число испытуемых, правильно ответивших на вопрос, и $q_i = 1 - p_i$. Стандартное отклонение вопроса таким образом будет следующим:

$$\sigma_i = \sqrt{p_i q_i}$$

Предположим, что эмпирические баллы и исследуемая характеристика распределены нормально (рис. 12).

Пусть M и σ^2 — средняя и дисперсия баллов; каждый индивид имеет свой балл как сумму правильных ответов. Этот эмпирический балл соответствует площади под кривой слева от некоторого значения x , которое определяется из таблиц нормального распределения. Каждый индивид получит характеристику в виде некоторой величины, измеренной в единицах a . Это так называемая σ -шкала. Иногда делают преобразование, с тем чтобы нормальное распределение имело характеристики $M = 50$, $\sigma = 10$. Это так называемая T -шкала.

Выделяют два вида тестов: скорости и мощности.

В тестах скорости испытуемый не имеет времени ответить на все вопросы. В тестах мощности каждый испытуемый имеет возможность ответить на каждый вопрос. Но большинство тестов

находится между этими крайностями. Возникает серьезный вопрос, что измеряет тест, когда придается особое значение скорости в сравнении с мощностью.

В практике тестирования имеет существенное значение предоставление каждому испытуемому одного и того же ограниченного времени. Оптимальный предел времени определяется на основе некоторого ценностного суждения в зависимости от цели измерения. Важно, какие именно психологические свойства измеряются, когда времени много в сравнении с тем, когда времени мало. Как мы увидим, нельзя отделить проблему времени от проблемы трудности.

Существует предположение, согласно которому время и трудность в качестве переменных относительно взаимозаменяемы, т. е. мы можем измерить одну и ту же способность, либо определяя, сколько единиц работы можно произвести за единицу времени, либо устанавливая уровень трудности, который можно измерить в неограниченное время. Это значит, что индивид получит один и тот же балл (число успешных действий — ответы на вопросы) при различных комбинациях времени и уровней трудностей, поскольку произведение времени на трудность будет постоянным.

Терстон сделал первую попытку решить эту проблему. Его рассуждения заключались в следующем. Он определил мощность индивида в качестве такого уровня трудности задач, при котором вероятность успеха равна 0,5 при условии, если дано бесконечное время. Бесконечное время есть все то время, которое практически затрачивает испытуемый для решения задачи. Вместе с этим определением мощности принимается предположение, что вероятность успеха представляется огиной трудности (рис. 13).

Диаграмма на рис. 13 представляет собой предполагаемую зависимость вероятности успеха от уровня трудности для разных отрезков времени: $T_1, T_2 \dots T_\infty$. В бесконечное время середина кривой приходится на уровень трудности A , который, по определению, является мерой мощности данного индивида при данном типе задачи. С уменьшением отрезков времени огиба перемещается влево по шкале трудности.

Успех в качестве функции времени (для того же индивида при том же типе задачи) изображается на второй диаграмме при разных уровнях трудности: $D_1, D_2 \dots$. Можно предположить, что вероятность успеха по времени имеет вид восходящей огины. В нулевое время вероятность успеха равна нулю. Задачи большей трудности решаются с вероятностью меньше, чем 0,5, даже после продолжительного интервала времени.

Будут ли обычные суммарные баллы давать соответствующие

меры отдельным характеристикам индивидов» когда тест ограничен во времени? Измерение мощности индивида при выполнении задачи, как было определено, — это измерение его порогового балла в бесконечное время.

Гулликсен дал более строгое определение скорости и мощности тестов на основе статистических критериев. Сначала он определил следующие символы; w — число неверных ответов; u — число вопросов без ответа; $x=w+u$ — общая ошибка балла (вопросы с неправильными ответами).

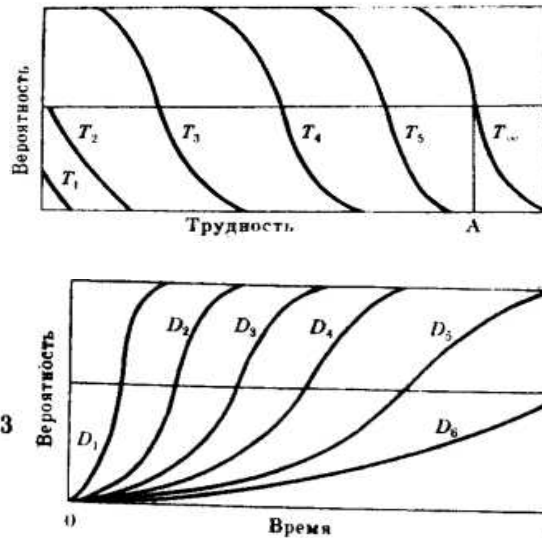


Рис. 13

В чистом тесте скорости $w = 0$, так что $x = u$, $M_x = M_u$, а $\sigma_x = \sigma_u$. Любой тест становится тестом скорости в той мере» в какой M_w и σ_w достигают нуля, а M_u и σ_u достигают M_x и σ_x соответственно.

В чистом тесте мощности M_u и σ_u равна нулю» а M_w и σ_w равны M_x и σ_x соответственно.

Гулликсен² предложил критерий» согласно которому, если отношение σ_w/σ_x очень мало, то тест по существу является тестом скорости, а если отношение σ_u/σ_x очень мало, то это — тест мощности.

² См.: Guilford / Я. Psychometric Methods N. Y., 1954» p. 233.

2. Классификация методов шкалирования

Более чем за 50-летнюю историю эмпирических социологических исследований было разработано большое количество шкал, иначе говоря, процедуры количественного представления социальных явлений. В настоящее время социологические шкалы различаются прежде всего по трем моментам: природе шкалируемого социального явления, уровню числового представления, способам получения числового представления.

Первый момент относится к природе явлений. Одно дело, когда изучается социальная установка группы лиц, другое — когда измеряется степень спаянности этой группы. Исторически так сложилось, что социологами наиболее разработаны шкалы измерения социальных установок, а также техника измерения структурных особенностей группы, так называемая социометрия, хотя это разграничение весьма относительно, поскольку социометрия основана на изучении предпочтения, а последнее связано с установкой.

Вместе с тем, изучая группу лиц по социальным установкам, в определенной степени изучают и структуру группы. Однако социологическое исследование не исчерпывается изучением структуры группы. Как измерить микрокультуру группы, уровень социализации индивида, степень влияния друг на друга членов определенной неформальной группы? Можно поставить множество подобных вопросов. Измерить субкультуру группы значительно сложнее, чем установку отдельного индивида, поскольку субкультура — это сложное явление, которое складывается из сочетания отдельных установок. Говоря математическим языком, субкультура — это многомерное явление, тогда как установка — одномерное. Последнее утверждение тоже весьма условно, так как установка — синтез сознательных и бессознательных ориентации и подсознательных влечений; мы полагаем, что этот синтез реализуется на каком-то одномерном непрерывном континууме. Измерить социальную установку индивида — означает определить его положение на этом гипотетическом континууме. Все развитые социологические шкалы установок и предлагают различные методы такого расположения.

Естественно, что при решении проблемы измерения и шкалирования в каждой отрасли социологии имеется своя специфика, что следует относить к содержательной стороне социального явления. Но в то же время природа социального явления обладает формой существования, которая включает три составляющие, а именно: континуум, связь, структуру.

Для измерения структуры группы была разработана социометрическая техника и различные варианты многомерного анализа (факторный анализ, латентно-структурный анализ, причинный анализ, многомерное шкалирование, таксономия). При измерении связи использовалось большое многообразие коэффициентов корреляции.

Классическим примером измерения континуума является определение социально-экономического статуса как измерение некоторой гипотетической непрерывной величины. Статус складывается из образования и материального дохода. Ясно, что статус больше у того, у кого больше образование и доход. А как быть с теми, у кого образование меньше, а доход больше, или наоборот?

Предположим, что мы оцениваем качество лекции по трем факторам — содержанию (С), интересу (И), технике (Т). Могут быть следующие варианты оценок:

Строка	С	И	Т	Строка	С	И	Т
I	+	+	+	V	-	-	+
II	+	+	-	VI	-	+	-
III	+	-	+	VII	+	-	-
IV	-	+	+	VIII	-	-	-

Знаком «плюс» обозначаем присутствие данного фактора в лекции, знаком «минус» — его отсутствие.

Сразу определяются две полярности качества лекции: одна полярность характеризуется всеми тремя положительными значениями факторов, другая — тремя отрицательными значениями этих факторов. Все прочие возможности находятся где-то между этими двумя полярностями. Качество лекции представляет собой как бы некий континуум, обладающий нижней и верхней границей. Эти границы являются граничными, предельными показателями. Эмпирически, операционально они фиксируются в трех одновременно положительных и одновременно отрицательных значениях факторов. Любой другой эмпирический вариант сочетания факторов является показателем промежуточного положения качества соответствующей лекции на этом континууме. При предположении равноправности рассматриваемых факторов естественно полагать, что два положительных ответа являются показателем более высокого качества лекции, чем при одном положительном ответе. В этом случае восемь вариантов ответов конструируют нам четыре показателя качества лекции: строка I — отличное; строки II—IV — хорошее; строки V—VII — удовлетворительное; строка VIII — неудовлетворительное. Но не яв-

ляется ли это слишком большим упрощением? На самом деле, можно ли считать лекцию удовлетворительного качества, если она малосодержательна? Приходится строки IV—V отнести к неудовлетворительному уровню и, следовательно, конструирование меры качества становится иным. Наличие двух полярностей означает существование между ними континуума и дает основание для построения шкалы.

Второй момент в классификации методов шкалирования связан с уровнями (шкалами) числового представления, предложенными Стивенсом. Существуют два фундаментальных аспекта теории измерения — определение информации, содержащейся в наблюдении на фенотипическом уровне, и определение отношения между генотипическим и фенотипическим уровнями. Эти оба аспекта обеспечивают базис для осуществления генотипических выводов из наблюдений.

Наблюдаемые данные (ответы на вопросы) — это фенотипический (наблюдаемый, эмпирический) уровень. Внутренний механизм, обуславливающий эти ответы, — генотипический уровень. К генотипическому уровню относятся установки, убеждения, способности и т. п. Проблема двоякая: во-первых, каковы измерительные свойства шкалы; во-вторых, к какому уровню — фенотипическому или генотипическому — относится шкала³. Соответственно возникает разделение шкал на две группы — фенотипические и генотипические. Если в психологии, психометрике и психофизике можно выделить пять групп шкал — фенотип-ординальные, фенотип-интервальные, фенотип-пропорциональные, генотип-интервальные, генотип-пропорциональные, то в социологии при такой логике классификации все методы шкалирования образуют только две группы — фенотип-ординальные и генотип-интервальные. Фенотип-ординальные методы дают ординальную шкалу на фенотипическом уровне. Это методы ранжирования, ранговых категорий, шкалограммный анализ Гутмана, техника развертывания Кумбса. Генотип-интервальные методы дают интервальную шкалу измерения на генотипическом уровне. Это метод парных сравнений, основывающийся на законе сравнительного суждения Терстона⁴. Существо метода парных сравнений может быть изложено следующим образом.

Закон сравнительного суждения имеет вид

$$S_1 - S_2 = x_{12} \sqrt{\sigma_1^2 + \sigma_2^2} - 2r \sigma_1 \sigma_2$$

³ Garner W., Creetman C. Problems and methods of psychological scaling.— In: Attitude Measurement. G. Summers (Ed.). Chicago, 1970.

⁴ Thurstone L. L. A law of comparative judgement.— «Psychol. Rev.», 1927, v. 34, p. 273.

где S_1 и S_2 — шкальные значения сравниваемых стимулов, x_{12} — величина, соответствующая вероятности того, что стимул 1 предпочтительнее стимула 2; σ_1 и σ_2 — дисперсии стимулов; r — корреляции между стимулами.

В зависимости от того, какие величины даны в правой части, различают пять случаев закона Терстона. Простейший вариант имеет вид:

$$S_1 - S_2 = x_{12} \sigma = \sqrt{2} z_{12}$$

z_{12} — это вероятности, но практически приходится принимать частоты за вероятности. Чаще всего производится многократное сравнение, и за вероятности принимаются усредненные частоты. По Терстону, эксперты оценивают каждую пару установок в отношении предпочтения. В квадратной таблице ставят процент экспертов, оценивших данную пару. Если все различия подчиняются нормальному закону, то эти проценты рассматривают как площади под нормальной кривой и в таблицах нормального распределения находят соответствующие абсциссы. По каждому столбцу они складываются, что дает интервальный балл каждой установке по данной группе экспертов.

Третий момент связан с тем, что шкала получается путем оценки, самооценки, сравнения, экспертного суждения и т. п.

Это последнее направление в классификации методов шкалирования включает две основные проблемы — характер инструментария измерения и процедуру приписывания баллов (мер, индексов). Первая проблема дает следующий перечень альтернатив (далеко не полный): отдельный выбор, вынужденный выбор, множественный выбор, косвенные меры, физиологические меры и т. п. Вторая проблема может быть разрешена посредством экспертных оценок, суммарных шкал, кумулятивных шкал и т. д.⁵

Проблема измерения, как мы уже отмечали, включает два аспекта — логико-теоретический и эмпирический, операциональный. Логико-теоретический аспект связан с характеристикой числовой системы. Эмпирический аспект означает множество операций над самими объектами, посредством которых можно наблюдать некоторые свойства этих объектов таким образом, чтобы удовлетворить определенным аксиомам числовой системы. Если отвлечься от специальных математических вопросов, то для социологии актуальны два момента современной теории измере-

⁵ Scott W. A. Attitude measurement.— In: Handbook of Social Psychology, v. 2. G. Lindzey, E. Aronson (Eds). N. Y., 1968.

ния: во-первых, способ выявления выполнимости аксиом порядка в эмпирической системе и, во-вторых, выявление и характеристика промежуточных шкал, например таких, как частично упорядоченная шкала, т. е. выявление математических свойств эмпирических систем, часть членов которых удовлетворяет аксиомам порядка, а часть — аксиомам равенства.

Вопрос о специфике именно социологического измерения имеет давнюю историю, хотя по сегодняшний день он остается открытым. Еще Г. Морфи в середине 30-х годов различал психологическое измерение как получение индивидуального балла в отношении рассматриваемой установки, а социологическое измерение — как измерение установки на основе групповых данных, полученных на основе вопросника.

В процедуре измерения в соответствии с его двумя аспектами выделяются два подхода. Первый связан с характеристикой и особенностями числовых систем, на которых оно реализуется, и определяет типы или уровни измерения (номинальный, порядковый, интервальный, числовой). Второй подход акцентирован на выявлении путей, посредством которых результируется измерение. Этот подход приводит к двум видам измерения — производным и фундаментальным. Квадранты Кумбса — основа для классификации методов измерения, однако полная их систематизация в социологии еще не осуществлена.

Мы можем задать себе вопрос в отношении некоторого человека, с одной стороны, нравится ли он вам или нет, с другой — занимается, например, он спортом или нет. В первом случае эмпирические данные будут отражать отношение этого человека к нашему предпочтению, идеалу. Во втором — данные отражают отношение этого человека к определенному атрибуту (хороший). Известный психолог и социолог Кумбс это различие обозначил как две задачи: *A* и *B*⁶. Задача *A* — это оценка суждения объекта, признака, вопроса, т. е. реализации предпочтения (например, я выбираю таких друзей в силу того, что их взгляды близки моим взглядам, привязанностям, наши мнения совпадают. Задача *B* — оценка суждения относительно соответствующего атрибута, т. е. выявление содержания (продолжая идею предыдущего примера, можно сказать, что в этом случае выбор таких друзей происходит в силу того, что они принадлежат к одной партии, одному поколению, что они имеют одинаковое семейное и общественное положение, и т. п.). Задача *A* — это соотношение вопроса и субъекта, задача *B* — соотношение вопроса анкеты и атрибута, объекта.

⁶ *Coombs C. A Theory of Data. N. Y., 1964.*

Специфика природы данных и эмпирических показателей также связана с относительностью и безотносительностью поведения при ответах на вопрос. При относительном поведении ответ выражает отношение двух или более вопросов, т. е. возможен выбор: в случае задачи *A* спрашивается, какой из данных пунктов предпочитается, в случае задачи *B* — какой из данных пунктов обладает большей характеристикой. При безотносительном поведении мы рассматриваем ответ на отдельный вопрос, не связанный со сравнением: в случае задачи *A* — это пример с арифметическим тестом, в случае задачи *B* — это шкала оценки (например, требуется оценить по шкале «знает — не знает» знания данного студента).

Эти два подхода — задачи *A* и *B* и относительное и безотносительное поведение — в характеристике данных эмпирического социологического исследования наглядно представлены с помощью квадрантов Кумбса ⁷:

	Безотносительное поведение	Относительное поведение
Задача А	II	I
Задача В	III	IV

Это можно представить на более формальном языке с помощью трех альтернатив: 1) отношения на паре точек или паре расстояний; 2) элементы пары — из одного множества или из разных множеств; 3) отношения — или отношение — порядка или близости (рис. 14).

Четыре класса, образованные перекрестным делением первых двух дихотомий, обозначены квадрантами от I до IV, а дополнительное деление дается отношениями порядка (>) и близости (0), которое делит квадранты на две части — *a* и *b* соответственно.

Наблюдаемое поведение в квадранте I — на паре расстояний, где каждое расстояние — между парой точек из различных множеств (обычно это индивиды и стимулы). Контекст реального мира, в котором наиболее часто можно получить этот вид

⁷Coombs C. A Theory of Data, — «Psychol. Rev.», 1960, N 67, p. 143.

данных, состоит в наблюдении предпочтительного выбора индивида среди множества стимулов. Данные могут рассматриваться как сравнение различий индивид — стимул или как так называемые данные *предпочтительного выбора*.

Наблюдаемое отношение в квадранте II — на паре точек, которые взяты из различных множеств (обычно это индивиды и стимулы). Такие данные называются данными *простого стимула*. Важно отметить, что они включают в себя не только данные

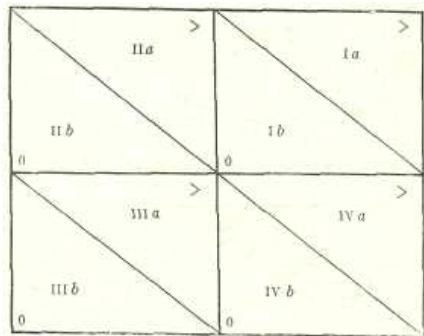


Рис. 14

психологического тестирования, выявления мнения и психофизических порогов, но также рейтинг-шкалы, метод последовательных интервалов и то, что называется данными *абсолютного суждения*.

Наблюдаемое поведение в квадранте III — на паре точек, которые взяты из одного и того же множества (они суть стимулы). Это данные *сравнения стимулов*. С таким видом данных мы имеем дело в социометрических матрицах.

Наблюдаемое поведение в квадранте IV — на паре расстояний, где каждое расстояние — между парой стимулов. Это данные *сравнения, различий стимулов*. Ввиду того что это поведение типично при ответе на относительное подобие стимулов, то они еще называются *данными подобия*. Следовательно, содержание квадрантов можно представить таким образом:

Данные простого стимула	Данные предпочтительного выбора или сравнения различий индивид — стимул
Данные сравнения стимулов	Данные подобия или сравнения различий стимулов

Кумбс различает методы сбора и методы анализа данных, хотя некоторые методы, например метод последовательных интервалов Терстона, могут выступать и в качестве первых, и в качестве вторых. Методы сбора классифицируются по указанным квадрантам следующим образом:

I квадрант включает методику развертывания, предложенную Кумбсом;

II—III квадранты — методы тестирования и метод шкалирования по Лайкерту;

IV квадрант — метод последовательных интервалов по Терстону.

Методы анализа можно сгруппировать так:

I квадрант включает шкалограммный анализ Гутмана;

II—III квадранты — латентно-структурный анализ Лазарсфельда;

IV квадрант — метод последовательных интервалов Терстона.

3. Основные социологические шкалы

Первыми в социологической практике стали применяться так называемые оценочные шкалы⁸. Простейшей оценочной шкалой является обычная школьная система баллов — для каждого ученика учитель определяет оценку как один из пяти возможных баллов, и ученики ранжируются по этим баллам. Вне школьных стен эта методика впервые стала применяться в середине XIX в. на английском флоте для оценки условий погоды. В начале XX в. известный статистик К. Пирсон применил эту шкалу для изучения способностей, после чего оценочные шкалы получили широкое распространение в практике оценки ситуаций.

Оценочная шкала включает в себя три компонента: того, кто оценивает, то, что оценивается, и то, по отношению к чему оценивается. Были развиты два вида оценочных шкал — графические и дескриптивные. При графическом способе каждый эксперт должен оценить явление на графике. График представляет линию, разделенную на 5, 7, 9 или 11 интервалов. Нечетное число интервалов выбрано для того, чтобы можно было найти нейтральную позицию. Желательно, конечно, иметь как можно больше градаций для большей точности и детализации оценки, но оказывается, что психологические возможности человека не позволяют ему производить классификацию больше чем по 11—13 позициям.

⁸ Downie N. M. *Foundamentals of Measurement*. N. Y., 1967.

Это как бы разрешающая способность его психологических возможностей. Левая сторона графика — положительная оценка, правая — отрицательная. Эксперт ставит штрих на том интервале, который соответствует его оценке данного явления или объекта. Затем все оценки экспертом складываются и делятся на их число. Полученное среднее дает оценку данного явления, объекта. При дескриптивном способе в зависимости от содержания проблемы даются словесные характеристики одиннадцати подразделений между крайними отношениями. Оценочным шкалам присущи два вида ошибок — так называемый гало-эффект и ошибка щедрости. Если индивид оценивается по нескольким чертам, то, с одной стороны, под влиянием его наиболее сильной черты эксперт оценивает более высоко и другие его черты, и, с другой — черты индивида получают более высокую оценку, если он обладает какой-либо чертой, которая особенно нравится эксперту. Впервые оценочная шкала была применена для собственно-социологического исследования в 1925 г. Богардусом⁹.

Наибольшее развитие оценочные шкалы получили при использовании метода парных сравнений, который основывается на законе сравнительного суждения Терстона.

Терстон предложил метод, который получил широкое распространение в социологической практике и в настоящее время не потерял своего значения. По исследуемой установке отбираются суждения — из литературы, предыдущих исследований, а также придумываются новые. Эти отобранные суждения оцениваются экспертами, которые каждому суждению присваивают балл. Затем с этими суждениями знакомят лиц, чьи установки предполагается изучить. По тому, какие суждения выбрал каждый индивид, ему присваивается балл, соответствующий баллу выбранных суждений.

Все трудности — в технике процедур каждого шага. Терстон предложил следующую методику. Прежде всего отбираются вопросы (суждения). Они должны быть по возможности краткими; их следует сформулировать так, чтобы они могли быть приняты или отвергнуты в соответствии с их предпочтением в отношении исследуемой установки. Желательно избегать двусмысленных суждений. Всего отбирается для экспертной оценки примерно 200—300 суждений, относящихся к данной установке. Затем отбираются эксперты 100—150 человек.

⁹ *Bogardus E. S. Immigration and Race Attitudes. Boston, 1928; Bogardus E. S. The Development of Social Thought. N. Y., 1960; Social Problems and Social Processes. E. S. Bogardus (Ed.). N. Y., 1967.*

Желательно, чтобы и экспертов, и оцениваемых суждений было как можно больше, чтобы больше была бы гарантия от систематических ошибок. Практически трудно придумать больше 300 неповторяющихся суждений по какому-то вполне конкретному явлению, и потому ограничиваются именно этим числом¹⁰. Эксперты же в идеале должны представлять собой некоторую модель исследуемых лиц, т. е. и числом, и своими установками они должны представлять некоторое подобие исследуемой совокупности. Практически уже 100 экспертов дают довольно устойчивую картину оценки, но все же вопрос об их числе не столь прост. Проблема статистической оценки этого числа еще не решена в социологическом шкалировании.

Каждый эксперт дает оценку по одиннадцати градациям (группам). Градация, или группа, 1 означает сильное предпочтение, градация 2—не очень сильное предпочтение, 3—умеренное предпочтение, 4—слабое предпочтение, 5—очень слабое предпочтение, 6—нейтральное отношение, 7—очень слабое противодействие, 8—слабое, 9—умеренное, 10—сильное, 11 — очень сильное противодействие. Считается, что интервалы предпочтения между градациями примерно равны. В силу этого метод называется методом кажущихся равными интервалов.

Технически процедура оценки следующая. Отобранные суждения записываются на карточки с указанием номера суждения. Каждому эксперту дается полный набор этих карточек в случайном порядке. Эксперт записывает номер градации, в которую он помещает данное суждение. Теперь для каждого суждения по экспертным оценкам можно построить кривую накопленных частот — кумуляту. Все отобранные суждения будут характеризоваться своими собственными кумулятами (у самого Терстона используются огивы).

Возьмем произвольное суждение. Каждое суждение имеет определенное эмпирическое частотное статистическое распределение. Как всякое распределение, оно характеризуется мерой центральной тенденции (мерой положения) и мерой рассеивания. В качестве меры положения принимается медиана, поскольку, строго говоря, нельзя использовать номера градаций как числа, следовательно, их складывать нельзя и потому нельзя найти среднюю арифметическую. По этой же причине в качестве меры рассеивания берут не дисперсию, а квартильное отклонение.

¹⁰ Ядов В. А. Социологическое исследование. Методология, программа, методы. М., 1972, с. 109.

Таблица частот

Градации	1	2	3	4	5	6	7	8	9	10	11
Частота	0,16	0,41	0,28	0,10	0,04	0,0	0	0	0	0	0

Таблица накопленных частот

Градации	1	2	3	4	5	6	7	8	9	10
Частота	16	57	85	95	99	99	99	99	99	99

Кумулята строится обычным способом. По оси x откладываются значения признака (в данном случае 11 точек по числу градаций), по оси y — накопленные частоты. Поскольку определяются медиана и квартильное отклонение, постольку строится именно кумулята, а не полигон распределения. Дело в том, что на кумуляте очень просто определяются указанные величины. Медиана — такое значение признака M , что половина генеральной совокупности имеет значения меньше M , а половина — больше M . На графике кумуляты берем ординату, равную 0,5, и находим соответствующую ей абсциссу. Через точку 0,5 на оси ординат проводим прямую параллельно оси абсцисс до пересечения с графиком. Из точки пересечения опускаем перпендикуляр на ось абсцисс. Основание этого перпендикуляра и будет искомая медиана. Если основание попадает между точками, то проводим обычную интерполяцию. В этом случае, конечно, обнаруживается некоторая некорректность, поскольку имеются порядковые величины, а при интерполировании приходится обращаться с ними, как с обычными числами. Полученное значение медианы и является баллом данного суждения по экспертной оценке.

Квартильное отклонение задается формулой

$$Q = \frac{Q_3 - Q_1}{2}$$

где Q_3 — верхний квартиль, Q_1 — нижний квартиль. Квартили — обобщения понятия медианы. Если медиана — это такое значение признака, когда половина популяции имеет значения большие, половина — меньшие, то нижний квартиль — это такое значение признака, когда $\frac{1}{4}$ популяции имеет значения, мень-

шие этого квартиля. Верхний квартиль — такое значение признака, когда $\frac{1}{4}$ популяции имеет значения признака, большие его. По графику кумуляты квартили определяются аналогично вычислению медианы. По формуле находится квартильное отклонение.

В результате всего этого процесса каждое суждение будет характеризоваться двумя числовыми мерами. По этим мерам следует из всех 200—300 суждений отобрать наиболее подходящие для анализа.

Можно рассуждать следующим образом. С одной стороны, нужно отбросить суждения, имеющие большие квартильные отклонения, и оставить те квартильные отклонения, которые сравнительно малы. Последние будут соответствовать более единодушному мнению экспертов в отношении этого суждения. С другой стороны, оставляют и те суждения, медианы которых образуют наиболее компактную и однородную совокупность. В результате такого «двойного прочесывания» обычно остается 20—25 суждений. Ими и измеряется установка.

Полный набор отобранных суждений предлагается каждому исследуемому лицу: оно должно выбрать те, которые, по его мнению, наиболее полно и точно характеризуют исследуемую установку. Причем число выбираемых каждым лицом суждений не ограничивается. После того как отобрано определенное число наиболее важных суждений, выводится их средний балл (средняя арифметическая), который и является баллом данного лица по данной установке. Затем лица ранжируются по этим баллам.

Шкала Терстона весьма трудоемка. Главная ее трудность в том, что приходится использовать экспертов, которые могут внести и вносят существенное искажение в окончательный результат измерения установки исследуемых лиц.

В 1932 г. Лайкерт предложил метод измерения установки без использования экспертной оценки, который получил название шкалы Лайкерта. Ее идея довольно проста. Даны вопросы (суждения), которых следует оценить по пятибалльной системе:

5—«полностью согласен»,

4 — «согласен»,

3—«нейтрален»,

2—«не согласен»,

1—«полностью не согласен».

Сумма баллов по всем вопросам и будет баллом данного лица. Затем лица ранжируются по баллам. Главная заслуга Лайкерта состоит в том, что он предложил внутренний критерий вопросов, который позволяет отбрасывать вопросы, не впи-

сывающиеся в программу исследования. Определяется корреляция между баллами вопроса и общим баллом.

Это можно проиллюстрировать следующим примером. Десять лиц (*A, B, C* и т.д.) должны оценить предложенные им десять вопросов по пятибалльной системе. Каждый получает такой бланк:

Ответ

	Полностью согласен	Согласен	Нейтрален	Не согласен	Полностью
1-й	+				
2-й	+				
3-й		+			
4-й			+		
5-й				+	
6-й			+		
7-й			+		
8-й					+
9-й					+
10-й				+	

Индивид делает отметку по каждому вопросу в соответствии со своим отношением к нему. Далее составляется следующая таблица по каждому вопросу (в примере даны вычисления для 5-го вопроса).

Лицо	Общий (S ₀)	Балл 5-го вопроса (S ₅)	Разность (S ₀ -S ₅)	Лицо	Общий (S ₀)	Балл 5-го вопроса (S ₅)	Разность (S ₀ -S ₅)
A	45	5	40	F	39	4	35
B	42	5	37	G	33	3	30
C	35	4	31	H	40	4	36
D	35	4	31	I	22	1	21
E	20	1	19	J	27	2	25

Затем определяется коэффициент парной корреляции между баллами лиц по пятому вопросу (S₅ в третьем столбце) и значениями разности между общим баллом S₀ и баллом по пятому вопросу S₅ (S₀-S₅ — в четвертом столбце).

Для вычисления этого коэффициента корреляции строим корреляционную таблицу.

S5	S ₀ —S ₅			
	10-19	20—29	30-39	40—50
1	1	1	—	—
2	—	1	—	—
3	—	—	1	—
4	—	—	4	—
5	—	—	1	1

Простым вычислением находится коэффициент корреляции между S₅ и S₀—S₅. Эта процедура осуществляется для каждого вопроса. Вопросы с малыми корреляциями выбрасываются.

Можно действовать и по терстоновской методике, используя квартальные отклонения. Нерешенным остается вопрос о значимости квартальных отклонений и коэффициентов корреляции вопросов.

Вплоть до настоящего времени идет дискуссия о преимуществах подходов Терстона и Лайкерта. В 30-х годах Г. Морфи защищал шкалу Лайкерта, отдавая ей предпочтение по сравнению со шкалой Терстона ¹¹, в 40-х Л. Фергюсон выступил в защиту шкалы Терстона ¹², в 50-х Л. Эдвардс попробовал найти баланс между этими шкалами ¹³.

Основной недостаток этих шкал заключается в том, что лица могут получить одинаковый балл, отвечая по-разному на разные вопросы. Кроме того, вопросам приписывается (например, по Лайкерту) балл по пятибалльной системе. Это вполне законная операция, как нумерация футболистов, т. е. измерение по номинальной шкале. Но затем эти баллы складываются, что делать, строго говоря, нельзя на этом уровне измерения; метод

¹¹ *Likert R., Roslow S., Murphy G.* A simple and reliable method of scoring the thurstone attitude scales.— «J. Social Psychol.», 1934, v. 5.

¹² *Ferguson L. W.* The measurement of primary social attitudes.— «J. Psychol.», 1940, v. 10, p. 199; *Ferguson L. W.* A study of the Likert technique of attitude scale construction.—«J. Social Psychol.», 1941, v. 13, p. 51.

¹³ *Edwards A. L., Kenney K. C.* A comparison of the Thurstone and Likert techniques of attitude scale construction.—«J. of Applied Psychol.», 1946, v. 30, p. 72; *Edwards A. L., Kilpatrick F. P.* A technique for the construction of attitude scales.— «J. Psychol.», 1948, v. 32, p. 374.

Терстона зависит от оценок экспертов, которые могут вносить систематическую ошибку в измерение.

Одним из путей преодоления трудностей шкалирования по Терстону и Лайкерту является метод Гутмана — так называемый шкальный (или шкалограммный) анализ.

Метод Гутмана не требует экспертов и приписывания баллов, но, к сожалению, здесь имеются свои трудности, в силу которых проблема измерения установок так и осталась переменной. Существо этого метода заключается в следующем: N лицам задаем n так называемых дихотомических вопросов (т. е. вопросов, которые допускают только два альтернативных ответа, которые мы обозначим « + » и «—»). Следовательно, первоначально получим таблицу плюсов и минусов. Каждый столбец представляет некоторый вопрос, каждая строка — некоторое лицо. Метод сводится к построению шкалограммы. Это такая картина, когда лица и вопросы ранжированы, т. е. расположены так, что первое лицо отвечает положительно на все вопросы (первая строка состоит из одних плюсов), второе лицо отвечает положительно на все вопросы, кроме последнего, третье — на все, кроме последних двух и т. д. В таблице в этом случае можно провести наклонную прямую, отсекающую слева минусы от плюсов. Но практически так не получается. Переставляя строки и столбцы исходной эмпирической таблицы, можно составить только приближенную шкалограмму, где наклонная прямая не рассекает точно таблицу на часть с единицами и часть с нулями. Приводим шкалограмму в идеальном случае.

Лица	Вопросы				
	1	2	3	4	5
1	+	+	+	+	+
2	—	+	+	+	+
3	—	—	+	+	+
4	—	—	—	+	+
5	—	—	—	—	+

Рассмотрим простой пример: 15 лицам задано 8 дихотомических вопросов. В I матрице представлена эмпирическая картина ответов. Проранжируем лица по их баллам. Для этого просто переставим строки таблицы в порядке убывания чисел

Таблицы данных (I)									Ранжирование лиц (II)								Ранжирование лиц и вопросов (III)											
Вопрос									Балл	Вопрос								Балл	Вопросы							Балл		
Лица	1	2	3	4	5	6	7	8	Лица	1	2	3	4	5	6	7	8	Лица	7	5	1	7	2	4	5	3		
1	+	+	+	+	+	-	+		6	7	+	+	-	+	+	+	+	7	7	+	+	+	+	+	+	-	7	
2	+	2	2	2	+	2		+	4	9	+	+	-	+	+	+	+	7	9	+	+	+	+	+	+	+	-	7
3	+	+	-	-	+	-	+	+	5	10	+	+	-	+	+	-	+	6	10	+	+	+	+	+	+	-	-	6
4	-	-	-	-	+	-	+		2	1	+	+	+	+	+	-	+	6	1	+	+	+	-	+	+	-	+	6
5	+		-	-	+	-	+		3	13	+	+	-	+	+	-	+	6	13	+	+	+	+	+	+	-	-	6
6	+	-	-	-		-	+	+	4	3	+	+	-	-	+	-	+	5	3	+	+	+	+	+	-	-	-	5
7	+	+	-	+	+	+	+	+	7	2	+	-	-	+	-	+	+	4	2	+	+	+	+	-	-	-	-	4
8	+	-	-	+	+	-	+		4	6	+	-	-	+	-	+	+	4	6	+	т	+	+	-	-	-	-	4
9	+	+	-	+	+	+	+	+	7	8	+	-	-	+	+	-	+	4	8	+	+	+	-	-	+	-	-	4
10	+	+	-	+	+	-	+	+	6	14	+	-	-	-	+	-	+	4	14	+	+	+	+	-	-	-	-	4
11	-	-	-	-	-	-	-	+	1	5	+	-	-	-	+	-	+	3	5	+	+	+	-	-	-	-	-	3
12	-	-	-	-	-	-	+		1	15	+	-	-	-	+	-	+	3	15	+	+	+	-	-	-	-	-	3
13	+	+	-	+	+	-	+	+	6	4	-	-	-	+	-	+	-	2	4	+	+	-	-	-	-	-	-	2
14	+	-	-	-	+	-	+	+	4	11	-	-	-	-	-	-	+	1	11	-	-	-	+	-	-	-	-	2
15	+	-	-	-	+	-	+	-	3	12	-	-	-	-	-	+	-	1	12	+	-	-	-	-	-	-	-	1
	12	6	1	6	13	2	14	9			12	6	1	6	13	2	14	9		14	13	12	9	6	6	2	1	

в последнем столбце (II матрица). В качестве последнего шага проранжируем вопросы по числу положительных ответов. Для этого переставим столбцы в порядке убывания чисел последней строки (III матрица). В результате получается некоторая шкалограммная картина (шкалограмма), в общем случае не идеальная, так как нельзя одной прямой отсечь все плюсы от минусов, а приходится проводить ломаную.

Вводится коэффициент воспроизводимости, определяющий возможное число отклонений (ошибок) от идеальной шкалы:

$$r = 1 - \frac{n_e}{Nn}$$

где n_e – общее число ошибок; N – число лиц, n – число вопросов.

Гутман полагает, что $r \geq 90\%$.

В случае большого числа лиц определяются так называемые шкальные типы лиц. N испытуемых на четыре вопроса дали такие варианты ответов, причем всего 2^n дали 16 различных вариантов ответов:

	2	3	4	5	
V	+	+	+	+	0
	-	-	-	+	1
IV	+	+	+	-	1
	-	+	+	+	0
	-	+	-	+	1
III	-	+	+	-	1
	-	-	+	+	0
	-	-	-	-	1
	+	-	+	+	1
II	+	-	+	-	+
	-	-	-	+	0
	+	-	-	-	1
I	-	-	-	-	0
	-	+	-	-	1
	+	-	-	-	1
	+	+	-	-	2

В таблице пять столбцов;
 1—номера (римские буквы) классов;
 2—варианты ответов;
 3—число ошибок в каждом варианте ответов;
 4—число лиц данного варианта ответа;
 5—число ошибок (произведения числа лиц на число ошибок в варианте).
 Существуют пять шкальных типов, соответствующих идеальной шкалограмме:

V	+	+	+	+
IV	-	+	+	+
III	-	-	+	+
II	-	-	-	+
I	-	-	-	-

Следовательно, оставшиеся 11 вариантов ответов являются нешкальными типами. Нужно попытаться их как-то разбросать по шкальным типам так, чтобы коэффициент воспроизводимости был наименьший. Естественно, например, что вариант ответа:

- + - +

следует отнести к типу IV, а не к V, поскольку в первом случае будет одна ошибка (отклонение от идеального IV типа), а во втором — две ошибки (отклонения от идеального V типа). Затем определяется коэффициент воспроизводимости, и если он значим, то возможна классификация лиц по шкальным типам на основе их вариантов ответов.

4. Социометрическая техника и семантический дифференциал

Понятие социометрии получило широкую известность с 1934 г. после выхода книги Дж. Морено «Кто выживет» («Who survive»). Хотя Морено вкладывал в это понятие весьма широкий смысл, впоследствии термин «социометрия» приобрел более ограниченный смысл, связанный с определенной техникой изучения человеческих предпочтений. Надо отметить, что такая техника использовалась уже до Морено, но ее широкое распространение и дальнейшее развитие связано с его именем.

Социометрия или (что то же самое) социометрическая техника, социометрическая мера — это некоторый инструмент оценки предпочтения и неpreferенция индивидов в группе. Как простейший пример можно взять квадратную таблицу.

	В	С	Д	Е	
А	1	1	1	0	0
В	0	1	0	1	0
С	0	1	1	0	1
Д	1	0	1	1	0
Е	1	1	0	1	1

Вдоль левого столбца и верхней строки записаны обозначения N индивидов группы. В клетке по строке пишем 1, если индивид i предпочитает индивида j , и 0, если не предпочитает. Клетки по диагонали можно заполнить единицами.

Таким образом, мы имеем матрицу D порядка $N \times N$:

$$D = \|d_{ij}\|, \quad i, j = 1, \dots, N,$$

где

$$d_{ij} = \begin{cases} 1, & \text{если } i \text{ предпочитает } j, \\ 0, & \text{если } i \text{ не предпочитает } j. \end{cases}$$

Если мы суммируем столбцы, то получим величины, определяющие число предпочтений данного индивида. Можно считать их баллами индивидов по предпочтению и проранжировать индивидов по этим баллам.

Это можно изобразить графически в виде так называемой социограммы. Индивида обозначаем точкой. Если он предпочитает некоторого другого индивида, то от него идет стрелка к этому второму индивиду. Очевидно, социограмма сразу показывает наглядно, какие индивиды – наиболее предпочитаемые лидеры, и какие – наименее предпочитаемые.

В дальнейшем при использовании социометрической таблицы начали применяться модификации. Можно, например, ввести три показателя – предпочтение, нейтральность, не предпочтение и соответственно в клетках будет стоять одно из трех чисел – 1,2,3 (или 1,0–1) и т. д.

С помощью социометрической таблицы можно ввести разнообразные индексы, характеризующие структуру всей группы". Таких индексов множество. Каждый исследователь вводил свой. Некоторые из них представляют интерес. Можно ввести отношение предпочтения как число всех предпочтений и неpreferтений, а также индекс совместимости группы как число взаимных предпочтений, деленное на объем группы.

В настоящее время начинают интенсивно развиваться статистические методы в отношении социометрических данных.

Наконец, дают интересные результаты матричные методы¹⁵. Вводится понятие доминирования, которое определяется таким образом:

$$d_{ij} = \begin{cases} 0, & \text{если } i = j, \\ 1, & \text{если } i \neq j \text{ и } d_{ji} = 0. \end{cases}$$

Матрица $D = \|d_{ij}\|$ определяет так называемое одночленное доминирование, а именно: если $d_{ik} = 1$, то $A_i \gg A_k$ (где доминирование отмечается знаком \gg). D^2 – двучленное доминирование, а именно: если $d_{ik} d_{ki} = 1$, то $A_i \gg A_k \gg A_j$.

Число всех одночленных доминирований i -го индивида равно сумме элементов i -й строки матрицы D . Число всех двучленных доминирований равно сумме элементов i -й строки матрицы D^2 . Можно ввести понятие ранга индивида как числа всех его одночленных и двучленных доминирований, т. е. ранг 1-го индивида равен сумме элементов 1-й строки матрицы:

$$S = D + D^2$$

Социометрическая техника в настоящее время применяется в педагогике, индустриальной социологии, военном деле и во многих других областях.

Уже давно было известно в психологии явление синестезиса – возникновение одних чувственных восприятий под воздействием других. Например, под влиянием различных музыкальных звуков рождаются некоторые зрительные представления. Антропологам было известно, что у индейцев центральной Америки, племен юго-восточной Азии, негров центральной Африки существуют определенные ассоциации в языке: хорошие вещи и люди сопоставляются с понятиями «высокий» и «светлый», а плохие – с понятиями «низкий» и «темный»

¹⁴ Волков И. П. Социометрические методы в социально-психологических исследованиях. Л., 1970.

¹⁵ Кемени Дж., Снелл Дж., Томпсон Дж. Введение в конечную математику М, 1963, с. 415.

В 40-х годах Осгуд показал наличие связи значения одних понятий с другими понятиями¹⁶. Были развиты понятия семантической шкалы и семантического пространства.

Возьмем два прилагательных с противоположным значением: сильный – слабый. Расстояние между ними разделим на семь, интервалов:

Очень сильный	3	Не очень слабый	-1
Сильный	2	Слабый	-2
Не очень сильный, ни сильный, ни слабый	1	Очень слабый	-3
	0		

Каждого человека, естественно, можно оценить по этой шкале, поставив ему балл от +3 до -3. Формально можно оценить любое понятие по этой семантической шкале (сильный – слабый). Можно взять шкалу: красивый – некрасивый, тонкий – толстый и т. д.

Эта совокупность шкал образует так называемое семантическое пространство. Его можно представить как обычное n-мерное евклидово пространство. На каждой оси отложены три единицы в положительную сторону и три единицы в отрицательную сторону. Любое понятие, по нашей оценке, займет некоторую точку в этом семантическом пространстве.

Была проделана следующая процедура (рис. 15). Отобрали 50 семантических шкал (50 пар противоположных прилагательных):

верхний – нижний

легкий – тяжелый

светлый – темный

и т. д. (всего 50).

Выбрали 20 понятий. Каждое понятие нужно было оценить по каждой шкале. Если оказывалось, что понятие совершенно не имеет смысла в данной шкале, то его значение имеет оценку «0». Было выбрано 100 экспертов. Каждый получил по 20 карточек, по одной карточке на понятие. Каждая карточка имела 50 шкал в одном и том же порядке, в соответствии с выбранными прилагательными. Очевидно, что был получен параллелепипед эмпирических данных 100X50X20.

Просуммировали результаты по экспертам и затем получили 50X50 матрицу корреляций семантических шкал. Применили

¹⁶ Semantic differential technique. J. G. Snider, C. E. Oseood (Eds). Chicago, 1969, p. 26.

центроидный метод факторного анализа и получили три фактора. Первый фактор объединял шкалы «красивый – некрасивый», «хороший – плохой» и т. п. и был обозначен как оценка; второй – как сила, поскольку объединял шкалы «сильный – слабый», «большой – маленький» и т. д.; третий – активность – объединял шкалы «активный – пассивный», «быстрый – медленный» и т. д. Это исследование было повторено для многих экспертов, понятий и шкал. Иногда получалось больше факторов, но самыми главными факторами оставались указанные

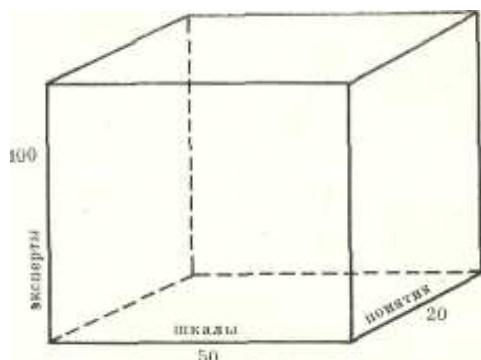


Рис. 15

три – оценка, сила, активность. Из этого был сделан вывод, что значение понятий можно рассматривать в трехмерном семантическом пространстве.

Осгуд в книге «Измерение значения» предложил метод для измерения значения понятий – так называемый семантический дифференциал¹⁷. Его идея состоит в следующем. Берется минимальное число шкал для каждого из трех измерений, обычно четыре шкалы – для первого измерения, три – для второго и три – для третьего измерения.

Предположим, мы избрали понятие, оценили его по всем четырем шкалам первого измерения, определили среднюю, которая и будет значением понятия по первому измерению. Аналогично для второго и третьего измерений. В результате исходное понятие можно представить некоторой точкой в трехмерном пространстве. Можно дать оценить значение этого понятия другому лицу – получим другую точку и увидим, насколько разнятся значения этого понятия у этих лиц и тем самым различаются лица по этому понятию.

¹⁷ Osgood C. E., Suci G. J., Tannenbaum P. H. The measurement of meaning. Urbana, 1957.

Метод семантического дифференциала получил широкое развитие в психологии, антропологии, социологии. Графически семантический дифференциал проще представлять в виде десяти параллельных линий. Каждое понятие на каждой шкале будет определяться некоторой точкой. Соединяя эти точки у всех шкал, получим профиль. В зависимости от того, как близки профили, можно судить о близости понятий и (соответственно) установок лиц.

5. Проблема надежности и обоснованности тестирования и шкалирования

Проблема обоснованности обусловлена тем, что исследователь не имеет гарантии уверенности в том, что он измеряет именно то, что стремился измерить. В конечном счете обоснованность полностью не гарантируется. Даже если провести тонкий анализ соответствия поставленных вопросов содержанию проблемы, он может быть элиминирован выборочной ошибкой, которую, как мы говорили, трудно оценить. Обоснованность социологической шкалы зависит от соотношения эмпирического и теоретического в социальном исследовании, от того, насколько правомерен эмпирический вывод и на какую теоретическую концепцию он опирается. Эмпирический факт получает достоверность на базе определенной теории, а определенная теория основывается на некоторых эмпирических данных. Надежность означает то, как измерено то, что было намечено измерить. Проверка надежности может быть осуществлена повторным измерением или на другой выборке. Однако в первом случае искажение может быть нейтрализовано динамикой явления, а во втором – искажение может быть обусловлено ошибкой выборочного измерения. Но всегда его границы весьма условны и подвижны. Вне зависимости от трудности решения этих проблем или даже невозможности решения; они не могут быть обойдены при построении социологических шкал.

Обоснованность связана с вопросом о том, что измеряет тест. Когда тест применяется для предсказания действия в некоторой жизненной ситуации, то обоснованность часто определяют на основе корреляции между тестом и некоторой мерой критерия.

Рассмотрим количественные подходы к объяснению тестовых баллов. Математические модели, выбранные для этого, дают возможность применять различные статистические операции.

Предположим, что тесты составлены из вопросов и что вероятность случайного успеха в ответе на вопрос крайне мала. Понятие надежности относится к точности, с которой балл пред-

ставляет состояние индивида в любом аспекте, измеряемом при помощи теста. Общеизвестно, что чаще всего баллы являются ненадежными, что они не свободны от ошибок. Наиболее важный шаг в понимании надежности был сделан, когда полученный балл представили в виде простой комбинации истинного балла и ошибки¹⁸.

$$X_t = X_\infty + X_e,$$

где X_t = полученный тестовый балл, X_∞ = истинный балл, X_e = ошибка.

Предполагается, что истинный балл X_∞ и ошибка X_e находятся на той же самой шкале, что и полученный балл. Существуют различные способы определения истинного балла. В одном случае утверждается, что X_∞ суть балл, полученный данным индивидом в идеальных условиях или благодаря идеальным средствам измерения, в другом – что X_∞ суть среднее полученных баллов из очень большого числа независимых тестирований одним и тем же тестом отдельного индивида. Ошибка есть приращение (положительное или отрицательное), являющееся функцией условий отдельного случая тестирования отдельного индивида.

Существуют и некоторые дополнительные предположения. Первое: ошибки могут быть как положительными, так и отрицательными, и их средняя равна нулю. Второе: в больших популяциях ошибки не коррелируют с истинными баллами, т. е. у индивидов с высокими истинными баллами нет тенденции к большим положительным или к большим отрицательным ошибкам. Третье: нет корреляции между ошибками одного типа теста и ошибками параллельного типа того же самого теста. Эти три вида предположений могут быть изложены в форме уравнений:

Предположение I: $M_e = 0$,

где M_e – средняя ошибок.

Предположение II: $r_{\infty e} = 0$,

где $r_{\infty e}$ – корреляция истинного балла и ошибки.

Предположение III: $r_{e_1 e_2} = 0$,

где e_1 и e_2 – ошибки в типах I и II одного и того же теста.

Полагаем, что $M_t = M_{\infty o} + M_e = M_\infty$.

Известно, что дисперсия суммы невзвешенных величин равна сумме дисперсий:

$$\sigma_t^2 = \sigma_\infty^2 + \sigma_e^2$$

¹⁸ Guilford J. P. Psychometric Methods. N. Y., 1954.

Понятие надежности вводится как доля дисперсии истинных баллов в дисперсии полученных баллов: $r_{tt} = \sigma_{\infty}^2 / \sigma_t^2$.

Исходя из представлений о параллельных формах теста, у которых равны средние, дисперсии и надежности, легко получить формулу для удвоенного теста:

$$\sigma_{2t}^2 = 2\sigma_t^2 (1 + r_{tt}); \sigma_{2\infty}^2 = 4 \sigma_{\infty}^2.$$

Преобразуя выражение надежности для удвоенного теста, получим

$$r_{22} = \frac{\sigma_{2\infty}^2}{\sigma_{2t}^2} = \frac{4\sigma_{\infty}^2}{2\sigma_t^2(1+r_{tt})} = \frac{2r_{tt}}{1+r_{tt}}.$$

Формула $r_{22} = 2r_{tt} / (1 + r_{tt})$ названа формулой Спирмена – Брауна для надежности удвоенного теста r_{22} . Эта формула применяется для определения надежности произвольного теста¹⁹. Тест подразделяется на две равные половины (split-half method), например берутся только нечетные и только четные вопросы. Зная надежность половины теста по формуле Спирмена – Брауна, находим надежность всего теста. Известна также формула Кудера – Ричардсона.

$$r_{tt} = \left(\frac{n}{n-1} \right) \left(\frac{\sigma_t^2 - \sum pq}{\sigma_t^2} \right),$$

где n – число вопросов; p – относительное число правильных ответов на вопрос; $q=1-p$ ²⁰.

¹⁹ Guilford I. P. Psychometric methods, p. 353.

²⁰ Ibid., p. 380.