

КЛАССИФИКАЦИЯ ОБЪЕКТОВ (ТАКСОНОМИЯ), КЛАССИФИКАЦИЯ ПРИЗНАКОВ (ФАКТОРНЫЙ АНАЛИЗ) И НЕКОТОРЫЕ ДРУГИЕ МЕТОДЫ АНАЛИЗА ИНФОРМАЦИИ

Кроме описанных, в социологических исследованиях используются и другие методы анализа информации, обзору которых и посвящена настоящая глава.

В §9 главы 1, описывая матрицу данных (табл. 1), мы дали эмпирическую информацию компактной, удобной для анализа. Рассмотренный нами ранее путь заключается в расчете характеристик, описывающих распределение опрошенных по каждому признаку. Например, использование средних позволяет «свернуть» матрицу данных в одну строку, состоящую из средних характеристик всего массива по каждому из изучаемых признаков. Такое представление, однако, во-первых, эффективно при достаточной однородности объектов по изучаемым признакам, во-вторых, не решает полностью задачу конденсации информации при большом числе признаков. Рассматриваемые ниже методы позволяют, с одной стороны, «сжать» матрицу данных, классифицируя опрошенных¹ и объединяя их в небольшое число однородных групп (таксономия), с другой стороны, позволяют объединить признаки в небольшое число групп (факторный анализ).

Таксономия. В качестве синонимов для обозначения этой группы методов используют также термины «кластерный анализ», «авто классификация» или (более широко) говорят об использовании методов «распознавания образов». Пусть матрица данных включает характеристики N объектов по двум количественным признакам (например, стаж работы и зарплата). Откладывая признаки по осям координат, мы можем изобразить все объекты на плоскости в виде N точек: абсцисса – значение стажа, ордината – значение зарплаты данного объекта. В этом случае говорят, что N объектов

[207]

расположены в двухмерном признаковом пространстве; (по сути, это один из способов изображения двухмерного распределения признаков). Как видно из рисунка, все объекты можно разбить на три группы таким образом, что объекты внутри групп близки между собой (это означает, что они имеют близкие характеристики и по X и по Y), а объекты из разных групп – далеки.

Множество близких между собой точек называется *таксоном* и при интерпретации результатов рассматривается

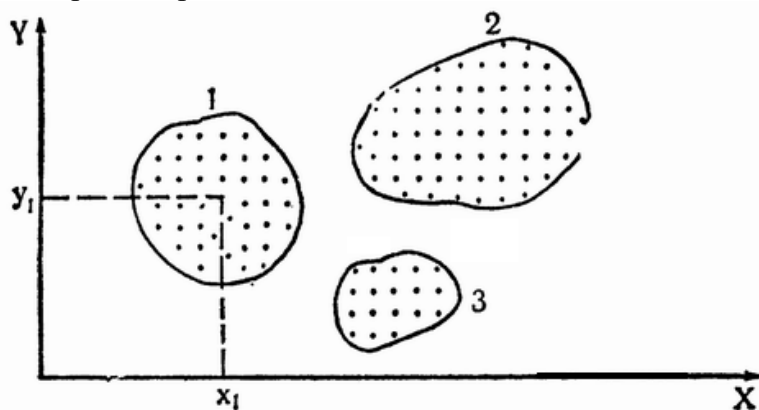


Рис. 28. Изображение объектов в пространстве двух признаков (1, 2, 3 – таксоны)

¹ Поскольку в матрице данных могут быть не только индивиды, но и бригады, предприятия, населенные пункты и т. п., мы будем далее говорить об «объектах», а не об «опрошенных».

как некоторый социальный тип. Если имеется k признаков, то говорят, что объекты расположены в k -мерном признаковом пространстве, Если признаков более чем два, то точки уже невозможно изобразить на плоскости. В этом случае группировку можно осуществить с помощью формальных методов, которые и называются методами таксономии.

Результатом работы алгоритмов таксономии обычно является разбиение множества объектов на группы (таксоны) в пространстве признаков, заданных исследователем, а также расчет некоторых обобщенных характеристик каждого из таксонов (центр таксона, средние, меры вариации). Существуют алгоритмы, позволяющие проводить классификацию не только в пространстве признаков, измеренных с помощью метрических шкал, но и для шкал номинальных и порядковых.

В качестве примера рассмотрим применение таксономии для классификации сельских мигрантов². Задача заключалась в проверке гипотезы о том, что направление мигра-

[208]

ции зависит от пола, возраста, образования, семейного положения, числа детей и типа профессии, который характеризовался числом месяцев обучения. Выделение типов мигрантов дало возможность разрабатывать мероприятия, направленные на регулирование миграции дифференцированно и, следовательно, более эффективно. С помощью методов

Таблица 40

Характеристика групп, полученных методом таксономии

№№ группы	Общая характеристика	Средние показатели по группе			
		Возраст лет	Число детей	Общее образование, лет	Специальное образование, месяцев
1	Семейные мужчины и женщины	30,6	1,7	5,4	2,0
2	Неженатые молодые мужчины	22,1	0,2	6,5	2,4
3	Молодые незамужние женщины и	19,7	0,1	7,1	0,8
4	Престарелые женщины без мужей	61,0	0,9	1,3	0,1
5	Одинокие женщины среднего возраста без специальности	32,1	0,1	1,4	0,9
6	Одинокие женщины специалисты	24,8	0,2	11,6	41,0

² Распознавание образов в социологических исследованиях. Новосибирск, 1968.

таксономии все мигранты, или «точки» в шестимерном пространстве из перечисленных признаков, были разбиты на 6 групп (таксонов): в один таксон попадали лица, близкие по приведенным в табл. 40 (совпадение числа признаков пространства с числом полученных групп, разумеется, случайно).

Чтобы проверить, действительно ли эти группы представляют разные типы, сравним характер их миграции (табл. 41). Мы видим, что группы существенно различны. Например, первая – семейные мигранты с детьми - дает наибольшую долю движения в пределах сельской местности, к ней приближается 4-я группа - «бабушки», но «бабушки» явно отличаются от группы! по направлению миграции. С помощью методов таксономии решались также задачи блокировки профессий, выявления групп рабочих по степени устойчивости на предприятии и др.,

[209]

Другим направлением конденсации информации является *факторный анализ* признаков. Как уже неоднократно отмечалось, индивиды обладают самыми разнообразными признаками, которые не являются независимыми. Связи между ними, как мы видели, изучаются с помощью методов корреляционного анализа. Можно предположить, что некоторые признаки образуют группы, каждая из которых

Таблица 41

Распределение мигрантов, вошедших в разные группы, по характеру движения между городом и деревней

Характер движения	Номера групп					
	1	2	3	4	5	6
1. Из крупного или среднего города в село	7,3	0,8	0	1,2	0	3,0
2. Из малого города в село	8,0	6,6	3,2	10,6	0	21,2
3. Из села в село	47,4	14,6	13,8	41,7	22,8	15,2
4. Из села в малый город	16,5	47,0	35,0	4,1	31,0	22,7
5. Из села в средний или крупный город	20,8	31,0	48,0	42,4	46,2	37,9

отражает определенный аспект сложного явления. При анализе системы признаков мы сталкиваемся не с классификацией объектов, а с классификацией признаков, т.е. с выявлением групп признаков, имеющих сходный характер изменения при переходе от одного объекта к другому. В частности, ставится задача найти максимально взаимосвязанные группы признаков. Выделяемые группы – это новые, комплексные переменные, называемые *факторами*.

Факторный анализ позволяет не только выделить группы наиболее взаимосвязанных признаков, но и отделить несущественные признаки от существенных, оценить их информативность.

Обоснованная замена большого числа признаков, описывающих объекты наблюдения, меньшим числом комплексных характеристик (факторов) составляет сущность факторного анализа.

Подчеркнем, что факторы не сводятся к некоторым, пусть главным, основным признакам исходного набора, Каждый фактор - это группа взаимосвязанных признаков из упомянутого набора, и вся совокупность входящих в него признаков определяет содержательную интерпретацию этого фактора.

[210]

Выделение групп признаков, подобно выделению таксонов, означает «конденсацию» информации, построение более простого описания, которое помогает вскрыть логическую структуру изучаемого явления, выделить наиболее характерные связи в системе признаков, проверить гипотезы о взаимосвязях, выдвинуть новые и т.д.

Попутно отметим, что выделение факторов упрощает решение задачи многомерной классификации объектов наблюдения, т.е. группировки объектов со сходными значениями признаков (задача таксономии). Здесь факторный анализ выступает как предклассификация, предварительный этап классификации объектов, Переход к небольшому числу комплексных переменных (факторов) упрощает применение графического анализа, интерпретацию результатов,

Рассмотрим конкретный пример³ применения факторного анализа к изучению природы стимулирующего воздействия на трудовую деятельность,

Общая схема стимулирующего акта может быть представлена следующим образом: создаются условия для реализации целей, формируется сознательная ориентация работников на выполнение цели. Необходимым условием достижения целей являются высокие показатели в работе (объективное отношение к труду, или фактическое поведение в сфере трудовой деятельности). Фактическое поведение фиксировалось как выполнение норм выработки, качество работы, дисциплинированность, участие в рационализации и изобретательстве.

При изучении субъективного отношения к труду, или ориентации на трудовую деятельность, рассматривались: отношение к работе, к специальности, к различным элементам производственной ситуации,

В исследовании фиксировались также демографические и функциональные признаки работников, которые можно рассматривать как референты социальных условий жизнедеятельности, как признаки, свидетельствующие о диапазоне реальных возможностей той или иной группы работников в определенной системе общественных отношений.

Таким образом, члены производственного коллектива являются носителями ряда признаков, а именно: демогра-

[211]

фические и функциональные признаки, оценки-ориентации разного рода, различные качества работника и т.д.,

Естественно предположить (и это в известной мере заложено в приведенной выше интуитивной априорной классификации, вытекающей из предварительного теоретического анализа), что некоторые признаки могут быть объединены в группы, т.е. возникает задача группировки признаков. Особый интерес при этом представляет характер связи различных групп признаков с конечным результатом стимулирования – фактическим поведением, конечно, если образуется группа признаков, описывающая это поведение,

Для анализа была выбрана следующая система признаков работающих:

- 1) квалификация;
- 2) стаж работы на заводе;
- 3) стаж работы по данной специальности;
- 4) образование;
- 5) возраст;
- 6) величина заработной платы;
- 7) выполнение норм выработки;
- 8) состояние трудовой дисциплины;
- 9) качество работы;

³ Исследование проводилось социологическим отделом Одесского отделения ИЭ АН УССР в 1971-75 гг. под руководством И. М. Поповой. Разработка методики сбора, обработки и анализа информации осуществлялась В. С. Максименко.

- 10) участие в рационализации и изобретательстве;
- 11) удовлетворенность работой (предприятием);
- 12) оценка степени физической нагрузки (тяжела ли работа физически?);
- 13) удовлетворенность содержанием труда интересна ли работа?);
- 14) оценка организации труда (простой, "штурмовщина");
- 15) удовлетворенность заработной платой;
- 16) удовлетворенность отношениями с администрацией;
- 17) мнение о справедливости распределения премий;
- 18) удовлетворенность специальностью.

Собор признаков, с одной стороны, диктовался стремлением учесть социально-демографические характеристики работников, объективное отношение к труду и субъективное (удовлетворенности оценки как интегральные – работой в целом, специальностью, так и частные – отдельными элементами рабочей ситуации). А с другой стороны, лимитировался техническими возможностями расчета корреляций между признаками, число которых M , как известно, является квадратичной функцией числа

$$\text{признаков } n: M = \frac{n(n-1)}{2}$$

Информация о системе отобранных признаков содержится в матрице корреляций, которая была построена на основе коэффициента Чупрова⁴.

Рассмотрим основные результаты применения факторно-

[212]

го анализа. При выделении двух факторов в одну группу попадают все признаки, характеризующие субъективное отношение к трудовой деятельности (11 - 13), во вторую - остальные, описывающие и социально-демографические характеристики работников и объективное отношение к трудовой деятельности.

Все удовлетворенности-оценки (первый фактор) тесно взаимосвязаны, хотя они связаны и с социально-демографическими признаками работающих, и с объективным отношением к труду (второй фактор), эта связь меньше, чем взаимосвязь; корреляция факторов 0,269, а факторные нагрузки⁵, описывающие корреляцию признаков, характеризующих субъективное отношение к труду с фактором, заключены между 0,601 и 0,417. Можно предположить, что эмпирический материал свидетельствует об относительной самостоятельности сферы сознания. Отметим, что социально-демографические признаки, попадающие во второй фактор, в большей степени связаны с объективным, чем субъективным отношением к трудовой деятельности.

В группе признаков, описывающих субъективное отношение к труду, максимальная факторная нагрузка у признака «отношение к специальности» (0,601), далее идут «отношение к содержанию труда» (0,524), «отношение к работе в целом» (0,520) и т.д.

Во второй группе признаков на первое место по величине факторной нагрузки выходит возраст (0,582), на второе – квалификация (0,551), на третье – качество работы (0,520) – первый из признаков, описывающих объективное отношение к труду – и т.д.

Обратим внимание на то, что внутри второго фактора социально-демографические признаки не локализованы, а чередуются с признаками, описывающими объективное отношение к труду.

При выделении трех факторов образуются группы, описывающие:

⁴ Обсчет информации осуществлялся в Институте проблем управления АН СССР И. Б. Мучником и Н. Е. Киселевой по алгоритму, изложенному в статье Э. Бравермана, А. Дорофеева, М. Луганского, И. Мучника «Методы диагонализации матриц связи» (Проблемы расширения возможностей автоматов. Труды Ин-та проблем управления, вып. 1, 1973).

⁵ Коэффициенты, характеризующие связь признаков с фактором; их можно интерпретировать как коэффициенты корреляции фактора с признаками.

- 1) f_1 социальные условия жизнедеятельности (1 - 6);
- 2) f_2 объективное отношение к труду, показатели в работе (7 - 10);
- 3) f_3 субъективное отношение к труду, ориентацию на трудовую деятельность (11 - 18).

[213]

Таким образом, как бы распадается на две части группа признаков, составлявшая ранее второй фактор. Теперь все факторы состоят из *сходных* признаков и тем самым могут быть естественно интерпретированы. Оказывается, что максимально взаимосвязаны f_1 и f_2 , минимально f_2 и f_3 , т.е. подтверждаются и детализируются выводы, сделанные ранее при рассмотрении двух факторов.

Внутри признаков, описывающих объективное отношение к труду, максимальная факторная нагрузка у такого признака, как качество работы, на втором месте - выполнение норм, далее идут дисциплинированность, участие в рационализации и изобретательстве. (Эта последовательность сходна с ранее полученной. Заметим, что она сохраняется в дальнейшем при переходе к большему числу факторов).

Как мы видим, два последних признака относительно менее информативны. Это связано с тем, что: 1) практически все работники дисциплинированы и 2) большая часть их в рационализации не участвует.

В группе социально-демографических признаков по-прежнему на первом месте – возраст, на последнем – образование, а стаж по специальности «опережает» стаж на заводе. Последовательность признаков – возраст, стаж по специальности, квалификация, заработная плата, стаж на заводе, образование – сохраняется и при переходе и большему числу факторов.

То обстоятельство, что образование менее тесно связано с фактором, чем стаж, по-видимому, отражает специфику объекта – судоремонтные предприятия. Как было выяснено в ходе исследований, для судоремонтных профессий при прочих равных условиях - стаж в большей мере определяет результаты трудовой деятельности, а также квалификацию рабочих, чем образование.

То, что стаж по специальности в большей степени связан с фактором, чем стаж работы на предприятии, можно также рассматривать как следствие специфики судоремонта, где рабочим приходится сталкиваться с самыми разнообразными типами судов, и профессиональные навыки, референтом которых является стаж по специальности, в результатах трудовой деятельности играют более важную роль, чем адаптация к условиям данного предприятия, референтом которой можно считать стаж работы на предприятии.

В случае четырех факторов группа признаков субъективного отношения распадается на две. В первую входят 4 при-

[214]

знака, находившихся на более высоких местах: удовлетворенность специальностью, содержанием труда, предприятием в целом, отношением с администрацией. Во вторую - остальные. Внутри каждой из этих групп последовательность признаков практически такая же, как и в исходной группе, из которой они образовались.

С дальнейшим увеличением числа факторов (5, 6...) результаты становятся менее надежными, и мы не станем их приводить.

В заключение рассмотрим основные содержательные результаты применения факторного анализа к проблеме изучения стимулирующего воздействия на трудовую деятельность.

Все выделенные признаки оказываются практически взаимосвязанными, это можно рассматривать как свидетельство сложной природы стимулирующего воздействия.

Предложенная исходная группировка признаков целесообразна. Различные группы признаков характеризуют разные, относительно самостоятельные уровни регулирования трудовой деятельности: связь между признаками одного и того же уровня более тесная, чем между признаками разных уровней.

При исследовании взаимосвязей различных групп признаков обращает на себя внимание относительная самостоятельность сферы сознания (оценки, удовлетворенности), ее относительная ограниченность от сферы фактического поведения и - в меньшей мере - от признаков, характеризующих условия жизнедеятельности.

Фактическое поведение в сфере трудовой деятельности в большей степени определяется социально-демографическими признаками, чем субъективным отношением и труду.

Выше рассмотрены результаты факторного анализа признаков, корреляции между которыми описывались с помощью коэффициента Чупрова. В дальнейшем факторному анализу была подвергнута матрица корреляции, построенная на основе коэффициента Крамера, теоретически более предпочтительного. В принципе результаты получились близкие. Отметим только, что такой признак, как заработная плата, который в первом случае попадал в группу социально-демографических, во втором вошел в группу признаков, описывающих фактическое поведение в сфере трудовой деятельности. При выделении трех и четырех факторов эта группа остается компактной; последовательность признаков - качество работы, выполнение норм выработки, участие в

[215]

рационализации и изобретательстве, заработная плата, дисциплинированность – сохраняется.

«Переход» признака заработная плата в фактор, описывающий фактическое поведение работников, предоставляется теоретически оправданным. Все остальные содержательные выводы подтверждаются.

Отметим, что логическая непротиворечивость и естественность интерпретации полученных результатов могут свидетельствовать о возможности применения коэффициентов для факторного анализа социальных признаков. Аналогом факторного анализа является *латентно-структурный анализ*⁶.

Для углубленного изучения связей между признаками используется также причинный и дисперсионный анализ.

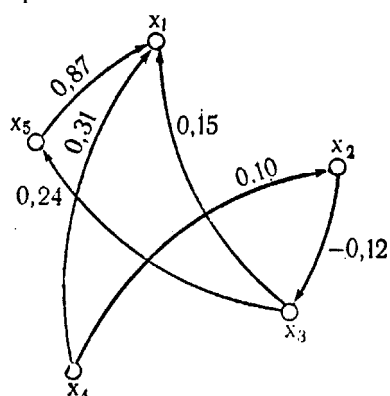


Рис. 29 Пример графа причинных связей для пяти признаков.

Причинный (или *путевой*) анализ используется для выявления *непосредственных* влияний одних признаков на другие⁷. Можно рассматривать *путевой анализ*, с одной

⁶ Статистические методы анализа информации в социологических исследованиях. М., 1979, гл. 14; Математические методы в современной буржуазной социологии. М., 1966.

⁷ Математика в социологии. Моделирование и обработка информации. М., 1977. ч. 1; Статистические методы анализа информации в социологических исследованиях. М., 1979, гл. 15. Хейс Д. Причинный анализ в

стороны, как своеобразное развитие методов частной корреляции и, с другой – как поиск направленных мер, характеризующих влияние одного признака на другой. Результаты причинного анализа часто представляют как в виде ориентированного графа⁸, вершинами которого обозначаются признаки, а дугами – влияния одних признаков на другие. Степень влияния i -го признака на j -й характеризуется специальным показателем⁹ P_{ij} . На рис. 29 изобра-

[216]

жен пример графа причинных связей для пяти признаков. Видно, например, что 3-й признак влияет на 1-й признак как непосредственно (0,15), так и через 5-й признак, наибольшее непосредственное влияние на 1-й признак оказывает 5-й (0,87) и т.д.

Рассматривая свойства дисперсии (§4, гл. 1), мы вывели формулу (148), показывающую, что общая дисперсия состоит из межгрупповой и внутригрупповой. Это равенство лежит в основе другого подхода к изучению влияния одних признаков на другие, который называется *дисперсионным анализом*. При изучении влияния набора признаков на вариацию некоторого результирующего признака (например, различных стимулов на повышение производительности труда) дисперсионный анализ позволяет вычленить влияние каждого из признаков. Это дает возможность отойти от традиционного метода планирования эксперимента (поддерживание стабильными всех переменных и вычленение влияния одного признака на результирующий) и перейти к экспериментам, в которых одновременно изменяются все признаки. Поэтому дисперсионный анализ очень тесно связан с *планированием эксперимента*¹⁰ (так называется один из разделов математической статистики) и может широко использоваться в планировании социологического исследования.

Чрезвычайно перспективным направлением статистики, развитым специально для нужд психологии, социальной психологии и социологии, является *многомерное шкалирование*¹¹. Методы многомерного шкалирования позволяют продуцировать гипотезы о критериях, которыми пользуются индивиды для оценки различных объектов. Исходной информацией для использования этих методов являются эмпирические данные либо о ранжировании индивидами некоторого набора объектов (например, ранжировка профессий по привлекательности), либо о сходстве объектов между собой (например, респондентам предъявляют всевозможные сочетания по две профессии из всего множества профессий и просят оценить сходство каждой пары профессий с помощью балльной оценки).

По этим данным находят минимальное признаковое пространство (т.е. пространство с минимальным числом осей), в котором можно так разместить оцениваемые объекты, чтобы

[217]

сохранились такими же, как в эмпирических данных, порядок или показатели близости объектов. Как правило, размерность этого пространства невелика, и его можно наглядно изобразить. Если, например, это пространство размерности 2, то получим картину, аналогичную изображенной на рис. 28. Далее, каждая из осей интерпретируется как неявная

статистических исследованиях, М., 1981.

⁸ Под ориентированным графом можно понимать некоторое множество точек (называемых вершинами графа), соединенных стрелками (называемых дугами).

⁹ Интерпретация этих показателей напоминает интерпретацию коэффициентов регрессии, хотя и не совпадает с ней. (Хейс Д. Причинный анализ в статистических исследованиях, с. 37, 104).

¹⁰ Гласс Дж., Стэнли Дж. Статистические методы в педагогике и психологии. М., 1979; Налимов В. В. Теория эксперимента. М., 1971.

¹¹ Клигер С. А., Косолапов М. С., Толстова Ю. Н. Шкалирование при сборе и анализе социологической информации. М., 1978.

шкала, которой пользуются респонденты для оценки объектов. Например, исследователь, обнаруживший, что проекции точек, изображающих профессии, на ось X легли в таком порядке - продавец-лоточник, водитель трамвая, корректор, монтажник радиоаппаратуры, техник связи, инженер, научный работник - может интерпретировать ось X как уровень образования, требуемый для данной профессии. Следующим шагом может быть проверка этого предположения в эмпирическом исследовании.

Важно отметить, что таким способом производится измерение по шкале, которую исследователь не задает априорно и, следовательно, не навязывает ее респонденту (обычный путь предполагает определение признаков, влияющих, например, на привлекательность профессии, до исследования, разработку шкал и включение их в анкету - при этом исследователь может пропустить важные для респондентов критерии оценки).

Отметим, наконец, две группы методов, отличающихся от изложенных не столько по возможностям, сколько по сфере приложения. Первая группа - статистический анализ *экспертных оценок*¹². Многие изложенные выше методы могут использоваться для анализа экспертных оценок, а методы, развитые для анализа экспертных оценок, - к анализу других видов социологической информации. Тем не менее целесообразно выделить эту группу методов, так как специфика экспертных опросов (небольшие выборки при больших объемах информации, полученных от экспертов, и сложных процедурах опроса, наличие специальных процедур согласования мнения и т.п.) все же приводят к определенной специализации методов. С точки зрения математического обоснования экспертных оценок и обработки результатов экспертизы выделяют следующие виды задач¹³ 1) построение моделей, описывающих поведение эксперта (модель поведения экспертов при ранжировании объектов,

[218]

модель парных или множественных сравнений¹⁴ и т.п.); 2) проверка адекватности модели данным, полученным от экспертов; 3) оценка степени согласованности мнений экспертов; 4) получение коллективного мнения группы экспертов.

Вторая группа методов - анализ социометрических данных¹⁵. Под социометрическими обычно понимают методы исследования структуры межличностных отношений в малой социальной группе путем изучения выборов, сделанных членами группы по тому или иному критерию. Социометрические данные можно представить в виде графа, вершинами которого являются члены группы, а дугами - сделанные ими выборы. Более широко социометрические методы можно определить как методы сбора и анализа информации, представленной в виде графа, вершинами которого являются члены группы. Специфика здесь связана с тем, что результатом социометрического опроса является не значение признака, а выявление взаимоотношений индивидов между собой. При этом возникают задачи описания характеристик вершин графа (т.е. расчет так называемых индивидуальных социометрических индексов), характеристик структуры графа (групповые социометрические индексы и методы выделения подструктур - лидера, подгруппы, ослабляющих и укрепляющих членов группы и т.п.), описания связи между графами, построенными по разным критериям. Весьма специфичны методы проверки гипотез, основанные на проверке соответствия случайных графов с графами, полученными в исследовании.

¹² Статистические методы анализа экспертных оценок. М., 1977.

¹³ Шмерлинг Д. С. и др. Экспертные оценки. Методы и применение (обзор). - В кн.: Статистические методы анализа экспертных оценок. М., 1977, с. 307.

¹⁴ Дзвид Г. Метод парных сравнений. М., 1978; Паниотто В. И. Метод множественных сравнений. - Социологические исследования, 1980, №1.

¹⁵ Паниотто В. И. Структура межличностных отношений. Методика и математические методы исследования. Киев, 1975.

Наконец, статистические методы используются также при моделировании социальных объектов, но рассмотрение этого вопроса выходит за рамки книги.

[219]