Вопрос 19. Способы обобщения и отображения первичных данных в социологическом исследовании

Обобщение и отображение результатов исследования

Социологические данные, полученные в результате обработки первичной информации, представляют собой, образно говоря, различных цветов и оттенков «краски», с помощью которых предстоит нарисовать реалистическую картину знаний о предмете исследования. «Краски» эти подбирались не произвольно. Целенаправленность их поиска была предопределена логикой перехода от качественных к количественным характеристикам предмета исследования, способным дать целостное представление о состоянии и изменении его сторон и свойств. Характер отображения социологических данных предопределен, прежде всего формами обобщения первичной информации.

1. Статистическая группировка.

Наиболее простая форма обобщения первичной социологической информации - группировка. Она позволяет зачислить респондента в ту или иную группу в соответствии с выбранным признаком (или признаками) группировки. Подытоживая число ответов на вопросы анкеты в опоре на такие признаки, исследователь осуществляет не что иное, как простую группировку респондентов с учетом их социально-демографических признаков, мнений, установок, информированности, индивидуальных оценок и т. д. Выделенные таким образом однородные по составу (по признаку группировки) группы значительно легче соотносить, сравнивать, анализировать.

Выбор признака группировки - не произвольная процедура, он диктуется задачей социологического исследования, сформулированными ранее гипотезами. Ошибочный выбор признака группировки приводит к неверным выводам» при анализе характеристик объекта исследования.

В зависимости от шкалы, измерения, в соответствии с которой получены ответы на вопрос, группировка социологической информации может представлять собой:

- зачисление респондентов в номинальные группы (группировка опрошенных по полу, национальности и т. д.);
- упорядочение информации в ранжированном ряду, например, по характеру труда (выполняющие ручной труд, работающие с механизмами, выполняющие труд интеллектуальный) или по степени включенности в общественную работу (состоят членами соответствующих движений, сочувствуют им, к партиям и движениям относятся с безразличием, либо отрицательно) и т. д.;
- группировку по количественному признаку, я результате которой группы респондентов характери¬зуются числовой величиной и потому количественно сравнимы между собой. (Например, группировка по возрастным интервалам: 18-25 лет, 26-30 лет, 31-40 лет, 41-50 лет, 50 лет и старше).

Операции с номинальными и ранжированными труппа¬ми осуществляются при помощи математических прие¬мов, соответствующих номинальной и ранговой шкалам. Что касается трупп, распределенных по количественному признаку, то они могут быть подвергнуты дальнейшему изучению при помощи любых приемов математической статистики, так как такая группировка осуществляется по интервальной шкале.

Проиллюстрируем сказанное на примерах.

Предположим, что нами осуществлена группировка п = 600 респондентов по номинальному признаку «род занятий»: работник сельского хозяйства: п1=120 человек (20%); рабочие промышленных предприятий: п2=300 человек (50%); инженерно-технические работники: п3 =180 человек (30%). В этом случае «максимальная глубина» математического анализа результата группировки ограничивается вычислением процентной величины. Вычисление среднестатистических величин для таких (группировок недопустимо, так как говорить о «среднем роде занятий» - явная бессмыслица.

Процентная величина вычисляется из соотношения: п1 / п х 100 % где п - общее число респондентов, подлежащих группировке; пі - число респондентов в і-й группе (в нашем примере і изменяется от 1 до 3-х). Количественные признаки принято делить на непре-рывные (возраст, зарплата и др.) и дискретные (число детей в семье, квалификационный разряд и др.) Непрерывные признаки, как правило, при любом чис-ловом выражении имеют конкретный физический смысл. Например, если мы распределили респондентов по коли¬чественному признаку «ежечасный заработок»: 5-10\$ по¬лучают 120 человек, 15-20\$ - 300 человек и 25-30\$ - 180 . человек, то утверждение — «среднечасовой заработок в расчете на одного опрошенного составляет 18,5\$.» впол¬не правомерно и соответствует действительной «физичес¬кой» градации такого индикатора, как деньги. Что же касается групп, разделенных по дискретному количественному признаку, для них применение матема-тических операций, соответствующих интервальной шка-ле, условна и допустимо лишь в рамках межгруппового сравнения. Связано это с тем, что такие расчеты имеют сугубо познавательную функцию, не неся в себе однов-ременно конкретного «физического» смысла. Например, осуществим группировку по другому количественному признаку – «число членов семьи»: у 120 опрошенных се¬мьи состоят из двух человек, у 300 - из трех человек и у 180 - из четырех человек. В этом случае также не будет ошибочным утверждение, что «средний размер семьи опрошенных» - 3,1 человека. Однако ясно, что такое чис-ло имеет только познавательное значение. Учитывая, что в дальнейшем нам придется встречаться с вычислением статистических

Учитывая, что в дальнейшем нам придется встречаться с вычислением статистических величин и во избежание повторения в объяснении их обозначений, здесь уместна дать некоторые поясне¬ния.

В математике числовые переменные величины принято обозна¬чать буквами латинского или греческого алфавита: x, y, a, b, т, п и т. д. Если количественные величины изменяют свое значение в стро¬го заданном порядке, то этот порядок обозначается индексом. В качестве индекса, как правило, ставится буква і или j в нижней части обозначенных, величин: x i,

y i, n i.

Когда респондентов распределяют в группы по двум или более признакам, например, выделяют мужчин в возрасте до 30 лет, имеющих высшее образование (три признака - пол, возраст, образование), то говорят о комбинационной группировке. В зависимости от решаемых задач она может быть структурной, типологи¬ческой и аналитической. Когда требуется узнать возрастной или квалификаци¬онный состав респондентов, применяется структурная группировка по возрастным интервалам или квалификационным разрядам. Таким образом, при структурной груп¬пировке респондентов классифицируют по некоторому свойственному всей совокупности

объективному призна¬ку.

Если же ставится задача выделить из состава опрошен¬ных группы по признаку «удовлетворенность содержани¬ем труда», то осуществляется типологическая группи¬ровка. Типы, как правило, выделяются на основании субъективного признака, измеренного по оценочной шкале, либо на базе нескольких признаков (многомерная группировка). В связи с этим типообразующие признаки чаще всего конструируются самим исследователем (та¬кие, к примеру, признаки, как «активный – неактивный», «обладающие различным уровнем художественной куль¬туры» и др.).

Группировка, произведенная по двум и более призна¬кам и служащая выявлению их взаимосвязи, называется аналитической. Если, скажем, проверяется, имеется ли связь между удовлетворенностью содержанием труда и соблюдением трудовой дисциплины, то необходимо осу¬ществить по этим двум признакам группировку.

2. Ряды распределений.

Из проеденных выше примеров, в результате группировки обычно выделяется не одна, а несколько групп (в соответствии с числом позиций признака группировки). При этом каждой выделенной группе соответствует некоторое число, отражающее ее количественный состав. Такой ряд чисел, получаемый в результате груп¬пировки, называется рядом распределения. Ряды распределений, отражающие результат группировки респондентов по качественным признакам, называются атрибутивными, а по количественным - вариационными. В со¬ответствии с характером количественных признаков вари¬ационные ряды делятся на дискретные и непрерывные.

Непрерывные, как правило, носят интервальный характер. Это значит, что та или иная группа респондентов характеризуется по количественному признаку не одним числом, а числовым интервалом, например возрастными интервалами: 20-24 года, 25-30 лет и т.д. Это важно учитывать при вычислении среднестатистического показателя.

Выбор интервалов во многом зависит от исследователя и осуществляется в соответствии с задачами исследования. Так, изучение вопросов адаптации молодых специалистов на предприятии предполагает в показателе «стаж работы» довольно узкие интервалы (1-3 месяца, 4-6 месяцев, 7-9 месяцев, 10-12 месяцев), а вот изучение различий в уровне профессионального мастерства - более широкие (1-3 года, 4-5 лет, 6-10 лет). В соответствии с задачами группировки интервалы могут выбираться равные и неравные, с возрастающими и убывающими значениями. Числа, обозначающие интервалы, называются их границами.

Ранее были приведены примеры рядов распределения, в которых интервалы имеют установленные границы. Наряду с этим бывают интервалы с неустановленной нижней или верхней границей (например, возраст до 18 лет, 60 лет и старше; стаж работы до трех лет, 5 лет и более и т. д.).

Ряды распределения наряду с числовой имеют и тек¬стовую характеристику. Такое отображение данных с сопроводительным объясняющим текстом осуществляет¬ся при помощи таблиц.

3. Составление таблиц.

Данная процедура не представляет собой отдельный вид математической операции обобщения первичной социологической информации. Это только форма отображения рядов распределения, имеющая преимущество в том, что в ней кратко

даются пояснения числовых значений соответствующих групп.

Числовые данные в таблице объясняются заголовками, подлежащим и сказуемым. Заголовки в таблице бывают общие, выступающие в качестве названия таблицы и раскрывающие структуру группировки рассматриваемой совокупности респондентов либо связь между ря¬дами распределения. Наряду с этим в названии таблицы целесообразно указывать место и время, к которым относятся приводимые сведения, а также общие для всех показателей единицы измерения (проценты, число человек, оценочный индекс, номинал денег и т. д.). Содержание строк и столбцов раскрывается внутренними заголовками: боковыми для строк и верхними для столбцов.

Подлежащим в таблице называется объект (совокупность респондентов), характеристики которого - сказуемое - выражены в таблице в числовом виде.

Виды таблиц. Наиболее простая - перечневая таблица, составленная на основании ряда распределения по одному признаку.

Таблицы, отображающие ряды распределений по двум и более признакам называются комбинационными.

Приведем пример комбинационной таблицы, отражающей посещение населением кинотеатров в группировке по двум признакам: месту проживания и «покленческому» признаку.

4. Графики и диаграммы.

Наряду с табличными в целях наглядности широко применяется графический способ отображения социологических данных. Чаще всего он имеет вид полигона или гистограммы. Полигон преимущественно используется для графического отображения непрерывных рядов, а гистограмма - дискретных. Строятся графики в прямоугольной системе координат, в которой на оси «у» отмечается общая численность, или доля респондентов (в %) по группам, на оси «х» - значения, или порядок признака. Для примера отобразить ряд распределения при помощи полигона. Построить гистограмму для ряда распределения в таблице.

Примечание. На оси "х" в данном случае в равных интервалах откладываются единичные значения признака (число лет). При этом градации на оси могут быть выбраны по усмотрению исследователя (по 1 году, по 2 года и т. д.), а отрезок, служащий основанием прямоугольника, равен масштабу соответствующего интер¬вала. Отсюда следует, что "ширина" прямоугольников одной и той же гистограммы может быть различной, если ряд распределен по признаку с неравными интервалами. Если прибегнуть к помощи современных компьютерных программ, то "графическое" отображение социологических данных может быть более разнообразным. Например, диаграммное отображение плотности групп по одному признаку, картографическое сопоставление плотности групп по нескольким признакам, многомерное графическое отображение информации.

5. Средняя арифметическая, дисперсия.

Средняя арифметическая есть интегральная, обобщен¬ная величина, позволяющая сравнить между собой не только группы одного ряда распределения, но и сами ряды распределения в том слу¬чае, если они строятся по идентичным признакам. Общая формула для ее вычисления имеет вид:

– Σ xi

```
X = -----
```

где Х і - числовые значения вариаций признака, і - число вариаций;

Σ - сумма значений признака.

Рассмотрим пример вычисления средней арифмети¬ческой величины. Имеются три числовые величины, харак¬теризующие посещаемость занятий студентами: первое занятие посетили 80 человек (x1), второе - 70 человек (x2), третье - 90 человек (x3). Тогда средняя посещаемость в расчете на одно занятие составляет:

X = 80+70+90/3 = 80 человек.

Простая средняя арифметическая вычисляется в том случае, когда группировка осуществлена по признаку, не имеющему собственных вариаций. Но это бывает довольно редко, так как большинство признаков, для группировки по которым вычислимы средние значения, являются количественными. Поэтому в социологическом исследовании, как правило, вычисляется так называемая взвешенная средняя арифметическая.

Покажем, как она может быть рассчитана.

Предположим, что в результате опроса 200 респондентов получены следующие сведения по признакам «возраст» и «число общественных движений, вызывающих симпатию у респондента». В таблицах приведены данные по двум показателям. Вычисление средней арифметической для данных в каж¬дой таблице имеет свои особенности. Признак «число общественных движений, вызывающих симпатию у респондента» содержит позиции, выраженные однозначными числовыми величинами (одно, два движения и т. д., относятся к движениям нейтрально или отрицательно - 0). По признаку «возраст» они распределены в интервалах (например, от 20 до 25 лет). Взвешенные средние арифметические для значений признака вычисляются по однозначным величинам, поэтому для интервалов предвари¬тельно необходимо определить среднее значение по каждой позиции показателя «возраст» (данные в третьем столбце таблицы 3). Усреднение интервала происходит путем вычисления простой средней для каждой градации возраста (например: 20-25 или 31-40), то есть сумма крайних значений интервала делится на число этих значе¬ний (в нашем случае на 2).

Наша задача, - исходя из данных, содержащихся в таблицах, определить среднее число общественных движений, вызывающих симпатию у респондентов и средний возраст в расчете на одного респондента. В этих случаях взвешенную среднюю арифметическую оп¬ределяем по формуле:

$$\overline{X} = \frac{\sum xi \times Ni}{N}$$

где хі - числовое значение і-й позиции признака;

Ni - число респондентов, выделенных по i-й позиции признака;

N - обще число респондентов, подлежавших группировке ($N = \Sigma Ni$).

После подстановки в формулу данных из таблицы 3 по признаку «число общественных движений, вызывающих симпатию у респондентов» получим:

X = 1x100 + 2x50 + 3x40 + 0x10 / 200 = 1.6 движений

Это значит, что среднее число общественных движений, вызывающих симпатию и приходящихся на одного респондента составляет 1.6, или же округленно - примерно 2 движения.

Вычисление средней для признака «возраст» осуществляется при помощи той же формулы, с использованием усредненных значений по каждому интервалу. Итак средний возраст опрошенной группы:

200

Недостаток средней арифметической как характеристики опрашиваемых по некоторому признаку заключается в том, что она может скрывать за собой различную сте¬пень «разброса» значений, и тем самым качественное сравнение различных групп по данным характеристикам затрудняется. Рассмотрим это на примере.

Предположим, проанализирована посещаемость четырех занятий в двух группах студентов. В одной занимается 20, в другой - 30 студентов. В течение четырех дней занятия посетили в первой группе (соответственно) 18, 20, 20, 18; во второй - 15, 23, 10, 28 человек. Вычислим среднюю посещаемость в первой (х1) и во второй (х2) группе:

4

Среднее значение в обоих случаях одинаково. Однако видно, что во второй группе этот показатель подчинен воздействию неких специфических факторов. Для того, чтобы измерить степень равномерности или неравномер-пости распределения той или иной интересующей исследователя характеристики опрашиваемых (особенно в тех случаях, когда «на глаз» это определить нельзя), исполь-зуется формула вычисления степени разброса значений признака, называемого дисперсией и обозначаемого ог (сигма квадрат):

где N - общее число респондентов;

Ni- число респондентов, выделенных по i-й позиции;

хі-, - числовое значение і-й позиции;

Х - средняя арифметическая.

6. Коэффициенты корреляции.

Корреляция означает наличие статистической взаимосвязи признаков. Рассмотрим один из них - коэффициент ранговой корреляции - р. Он легко рассчитывается — «вручную», а применение его весьма эффективно при анализе распределений социологической информации, полученной при помощи ранговой шкалы. Формула коэффициента ранговой корреляции имеет следующий вид:

Коэффициент ранговой корреляции изменяет свою величину от - 1 до + 1.

Фактически коэффициент ранговой корреляции выявляет степень идентичности распределения установок, мнений, характеристик двух сравниваемых групп опрашиваемых при их ответах на один и тот же вопрос, либо близость распределения ответов одних и тех же групп респондентов на «смежные» вопросы, позиции которого (которых) являются показателями ранговой шкалы. При р = -1 порядок распределения ответов по двум сравниваемым группам прямо противоположен, а при р = +1 он полностью совпадает. Применение коэффициента ранговой корреляции удобно также для сравнения данных анкетного опроса и данных контент-анализа при изучении, в частности, эффективности рекламной либо политико-агитационной деятельности средств массовой информации.

Коэффициенты парной корреляции призваны измерять взаимосвязь между двумя признаками исследуе¬мого объекта. Эта взаимосвязь, при изучении социальных явлений, не обязательно носит причинно-следственный характер. Например, если, у большинства блондинов голубые глаза - это не обязательно следствие того, что они блондины, либо они блондины не обязательно по той причине, что у них голубые глаза, хотя взаимосвязь (корреляция) между этими двумя признаками может быть весьма тесной.

Если при корреляции двух признаков между ними существует причинно-следственная связь, то признак, влияющий на характер вариации другого признака принято называть факторным, а зависимый от него признак - результативным.

Наиболее простой метод выявления взаимосвязи двух признаков - группировка и сравнение средних. Например, если мы хотим оценить зависимость частоты посещения кинотеатров от возраста респондентов, то сначала группируем их по возрасту, после чего в каждой возрастной группе вычисляем среднюю посещаемость кинотеатров (в неделю, в месяц, в год). Сравнение средних покажет нам наличие или отсутствие корреляции между выбранными двумя признаками - возраст респондентов и частота по¬сещения ими кинотеатров.

Корреляционная связь двух (или нескольких) признаков социального объекта носит не функциональный, а статис¬тический характер, в связи с чем она является не строгой закономерностью, а лишь тенденцией.

В процессе измерения корреляции между признаками решаются две задачи: определение формы и тесноты связи признаков.

При определении формы связи выявляется вариация (характер изменения) средних параметров результативного, под воздействием факторного признака, при условии их независимости от других признаков.

При установлении направленности и тесноты связи возможны три случая:

- 1) прямая связь увеличение (уменьшение) параметров факторного признака приводит к увёличению (уменьшению) параметров результативного признака;
- 2) обратная связь увеличение (уменьшение) параметров факторного признака приводит к умень¬шению (увеличению) параметров результативного признака;
- 3) изменение параметров факторного признака не приводит к изменению параметров результативного признака, т. е. корреляция отсутствует.

Наиболее простой метод определения связи между признаками, измеренными при помощи номинальной шкалы - расчет коэффициента ассоциации. Он измеряет тесноту связи между альтернативными признаками и вычисляется по формуле:

Ka = ad - bc / ad + bc

Например, предположим, мы хотим выявить наличие взаимосвязи между возрастом кинозрителей и частотой посещения ими кинотеатров, имея следующие эмпирические данные.

Наиболее совершенным для измерения взаимосвязи между двумя признаками является линейный коэффициент корреляции. Он используется в том случае, когда связь между признаками (обозначим их «х» и «у») выражается прямой линией.

Расчет линейного коэффициента корреляции осуществляется по формуле:

$$r(x,y) = x x y - x x y / \sigma x \sigma y$$

где ху - средняя из произведений значений признаков х и у;

х - средняя признака х; у - средняя признака у;

ох - среднее квадратическое отклонение признака х;

оу - среднее квадратическое отклонение признака у;

 $\sigma X = X - X \quad V \quad \sigma V = V - V$

При замене в формуле ох и оу на их адекватное значение, выраженное при помощи средних величин, по¬лучим итоговую формулу линейного коэффициента корреляции,

более удобную для практического применения:

Линейный коэффициент корреляции изменяет свою величину от +1 (прямая связь), до -1 (обратная связь).

При r(x,y) = 0 связь между признаками слабая, либо отсутствует.

Рассчитывается этот коэффициент только для признаков, измеренных при помощи интервальной шкалы. Для наглядности приведем пример расчета коэффициента парной корреляции для определения зависимости интенсивности миграции населения в города от доли аграрного населения региона.

Из-за чрезмерной трудоемкости ручного расчета большинства коэффициентов корреляций, их расчёт осуществляется на ЭВМ.

7. Расчет индексов.

Наряду с приведенными математическими методами обобщения первичной информации и ее сведения в эмпи¬рический показатель путем группировки, либо интеграции в индекс (средняя арифметическая, дисперсия, коэффи¬циент корреляции), отражающий динамику показателя, исследователь может для решения отдельных задач сам сконструировать некоторый индекс.

Рассмотрим пример. Предположим, что мы провели опрос в нескольких группах и выявили тех респондентов, кто читает материалы на политические темы «регулярно» или «нерегулярно».

Далее нам желательно сравнить между собой эти группы по «уровню обращаемости к материалам на поли¬тические темы». Для этого следует сконструировать индекс. Обозначим условно буквой "А" тех респондентов, кто читает материалы на политические темы регулярно, и буквой "Б" тех, кто их читает редко. Тогда наш индекс может принять вид формулы:

то есть разность читающих материалы на политическую тему часто и читающих эти материалы редко, деленная на число всех опрошенных. Такой индекс удобен тем, что он имеет четкие границы изменения и получаемый на его основе интегральный показатель легко поддается содержательной интерпретации.

Определим границы изменения значений индекса. Для этого предположим, что все

опрошенные читают сообщения на политические темы регулярно (то есть Б=0). Тогда значение индекса равно +1 (плюс единица). Если же предположить, что все опрошенные читают эти мате¬риалы редко (то есть A=0), то значение индекса равно -1 (минус единица). Таким образом, индекс изменяет свое значение от +1 до -1 и принимает значение 0 при A=Б, то есть при условии, когда число читающих регулярно сообщения на политические темы равно числу читающих их редко.

В том же случае, когда необходимо решить какую-то сложную исследовательскую задачу, следует прибегнуть к методам многомерного статистического анализа. Дальнейший, более глубокий анализ - социологических данных уже находится «вне власти» математических приемов. На стадии конкретных, содержательных выводов приоритет имеет исследователь, его интеллектуальные, творческие способности, профессиональный и социальный опыт.