

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 predetermined 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 predetermined, 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, в результате которой группы респондентов характеризуются 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 группами осуществляются при помощи математических приемов, соответствующих номинальной и ранговой шкалам. Что касается групп, 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 $n = 600$ респондентов по номинальному признаку «род занятий»: работник сельского хозяйства: $n_1 = 120$ человек (20%); рабочие промышленных предприятий: $n_2 = 300$ человек (50%); инженерно-технические работники: $n_3 = 180$ 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: $n_i / n \times 100 \%$

где n - общее число респондентов, подлежащих группировке; n_i - число респондентов в i -й группе (в нашем примере i изменяется от 1 до 3-х).

Количественные признаки принято делить на непрерывные (возраст, зарплата и др.) и дискретные (число детей в семье, квалификационный разряд и др.)

Непрерывные признаки, как правило, при любом числовом выражении имеют конкретный физический смысл. Например, если мы распределили респондентов по количественному признаку «ежечасный заработок»: 5-10\$ получают 120 человек, 15-20\$ - 300 человек и 25-30\$ - 180 человек, то утверждение – «среднечасовой заработок в расчете на одного опрошенного составляет 18,5\$» вполне правомерно и соответствует действительной «физической» градации такого индикатора, как деньги. Что же касается групп, разделенных по дискретному количественному признаку, для них применение математических операций, соответствующих интервальной шкале, условно и допустимо лишь в рамках межгруппового сравнения. Связано это с тем, что такие расчеты имеют сугубо познавательную функцию, не неся в себе одновременно конкретного «физического» смысла. Например, осуществим группировку по другому количественному признаку – «число членов семьи»: у 120 опрошенных семьи 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 число имеет только познавательное значение.

Учитывая, что в дальнейшем нам придется встречаться с вычислением статистических величин и во избежание повторения в объяснении их обозначений, здесь уместно дать некоторые пояснения.

В математике числовые переменные величины принято обозначать буквами латинского или греческого алфавита: x, y, a, b, t, p и т. д. Если количественные величины изменяют свое значение в строго заданном порядке, то этот порядок обозначается индексом. В качестве индекса, как правило, ставится буква i 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, типологической и аналитической.

Когда требуется узнать возрастной или квалификационный состав респондентов, применяется структурная группировка по возрастным интервалам или квалификационным разрядам. Таким образом, при структурной группировке респондентов классифицируют по некоторому свойственному всей совокупности

объективному признаку.

Если же ставится задача выделить из состава опрошенных группы по признаку «удовлетворенность содержанием труда», то осуществляется типологическая группировка. Типы, как правило, выделяются на основании субъективного признака, измеренного по оценочной шкале, либо на базе нескольких признаков (многомерная группировка). В связи с этим типобразующие признаки чаще всего конструируются самим исследователем (такие, к примеру, признаки, как «активный – неактивный», «обладающие различным уровнем художественной культуры» и др.).

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

2. Ряды распределений.

Из приведенных 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 группировки, называется рядом распределения. Ряды распределений, отражающие результат группировки респондентов по качественным признакам, называются атрибутивными, а по количественным – вариационными. В соответствии с характером количественных признаков вариационные 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 текстовую характеристику. Такое отображение данных с сопроводительным объясняющим текстом осуществляется 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 рядами 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 «поклонческому» 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 интервала. 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, обобщенная величина, позволяющая сравнить между собой не только группы одного ряда распределения, но и сами ряды распределения в том случае, если они строятся по идентичным признакам.

Общая формула для ее вычисления имеет вид:

$$- \quad \Sigma x_i$$

Добавил(а) Социология
04.02.11 03:21 -

$$\bar{X} = \frac{\sum x_i}{n}$$

где x_i - числовые значения вариаций признака, i - число вариаций;

$\sum$ - сумма значений признака.

Рассмотрим пример вычисления средней арифметической величины. Имеются три числовые величины, характеризующие посещаемость занятий студентами: первое занятие посетили 80 человек (x_1), второе - 70 человек (x_2), третье - 90 человек (x_3). Тогда средняя посещаемость в расчете на одно занятие составляет:

$$\bar{X} = 80+70+90 / 3 = 80 \text{ человек.}$$

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 каждой 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 предварительно 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 значений (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 определяем по формуле:

$$\bar{X} = \frac{\sum x_i \times N_i}{N}$$

где x_i - числовое значение i -й позиции признака;

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

N - общее число респондентов, подлежавших группировке ($N = \sum N_i$).

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

$$\bar{X} = 1 \times 100 + 2 \times 50 + 3 \times 40 + 0 \times 10 / 200 = 1.6 \text{ движений}$$

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

Добавил(а) Социология
04.02.11 03:21 -

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

$$— \quad 22.5 \cdot 40 + 28 \cdot 120 + 35.5 \cdot 30 + 45.5 \cdot 10$$

$$X = \frac{22.5 \cdot 40 + 28 \cdot 120 + 35.5 \cdot 30 + 45.5 \cdot 10}{200} = 28.9 \text{ года}$$

200

Недостаток средней арифметической как характеристики опрашиваемых по некоторому признаку заключается в том, что она может скрывать за собой различную степень «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 (x_1) и во второй (x_2) группе:

$$— \quad 18+20+20 + 18$$

$$X_1 = \frac{18+20+20 + 18}{4} = 19;$$

$$— \quad 15+23 + 10 + 28$$

$$X_2 = \frac{15+23 + 10 + 28}{30} = 19$$

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 неравномерности распределения той или иной интересующей исследователя характеристики опрашиваемых (особенно в тех случаях, когда «на глаз» это определить нельзя), используется формула вычисления степени разброса значений признака, называемого дисперсией и обозначаемого σ^2 (сигма квадрат):

$$\sigma^2 = \frac{\sum N_i \cdot x_i^2 - \frac{(\sum N_i \cdot x_i)^2}{N}}{N}$$

где N - общее число респондентов;

Добавил(а) Социология
04.02.11 03:21 -

N_i - число респондентов, выделенных по i -й позиции;
 x_i - числовое значение i -й позиции;
 $\bar{X}$ - средняя арифметическая.

6. Коэффициенты корреляции.

Корреляция означает наличие статистической взаимосвязи признаков. Рассмотрим один из них - коэффициент ранговой корреляции - r . 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:

$$r = 1 - \frac{\sum d^2}{n(n^2 - 1)}$$

где d - разность рангов;
 n - общее число рангов (или же вариантов ответов);

$\sum d^2$ - сумма квадратов разности рангов.

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 $r = -1$ порядок распределения ответов по двум сравниваемым группам прямо противоположен, а при $r = +1$ 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 исследуемого 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.

Добавил(а) Социология
04.02.11 03:21 -

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 посещения ими кинотеатров.

Корреляционная связь двух (или нескольких) признаков социального объекта носит не функциональный, а статистический 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 увеличению (уменьшению) параметров результативного признака;
- 2) обратная связь - увеличение (уменьшение) параметров факторного признака приводит к уменьшению (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:

$$K_a = \frac{ad - bc}{ad + bc}$$

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:

$$r(x, y) = \frac{\sum x_i y_i - \bar{x} \bar{y}}{\sigma_x \sigma_y}$$

где $\sum x_i y_i$ - средняя из произведений значений признаков х и у;

$\bar{x}$ - средняя признака х; $\bar{y}$ - средняя признака у;

σ_x - среднее квадратическое отклонение признака х;

σ_y - среднее квадратическое отклонение признака у;

$$\sigma_x = \sqrt{\sum x^2 - \bar{x}^2} \quad \text{и} \quad \sigma_y = \sqrt{\sum y^2 - \bar{y}^2}$$

При замене в формуле σ_x и σ_y на их адекватное значение, выраженное при помощи средних величин, получим итоговую формулу линейного коэффициента корреляции,

Добавил(а) Социология
04.02.11 03:21 -

более удобную для практического применения:

$$r(x,y) = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}$$

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 эмпирический показатель путем группировки, либо интеграции в индекс (средняя арифметическая, дисперсия, коэффициент 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 политические 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:

$$I = \frac{A - B}{A + B},$$

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

Добавил(а) Социология
04.02.11 03:21 -

опрошенные читают сообщения на политические темы регулярно (то есть $B = 0$). Тогда значение индекса равно $+1$ (плюс единица). Если же предположить, что все опрошенные читают эти материалы 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 $A = B$, 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.