



Статистическая обработка данных основывается на первичных и вторичных количественных методах. Цель первичной обработки статистических данных является структурирование полученных сведений, подразумевающее группировку данных в сводные таблицы по различным параметрам. Первичные данные должны быть представлены в таком формате, чтобы человек смог провести приближенную оценку полученной совокупности данных и выявить информацию о распределении данных полученной выборки данных, например, однородность или компактность данных. После первичного анализа данных применяются методы вторичной статистической обработки данных, на основании которых определяются статистические закономерности в имеющемся наборе данных.

Полученная в ходе полевых работ информация не случайно называется первичной. Исходя из нее, невозможно установить те зависимости, которые кладутся в основу социологических выводов и рекомендаций. Поэтому она нуждается в трансформировании во вторичную информацию, представленную в виде таблиц, графиков, уравнений, коэффициентов и т.п. показателей. Суть этого трансформирования - обобщение и свертка первичной информации, ее превращение в удобную для последующего анализа.

Современное среднее по масштабам эмпирическое исследование, проведенное на основе количественных методов, дает около ста тысяч байт информации, а большое исследование - около миллиона байт. Понятно, что обработка столь объемной информации ручным способом невозможна. Машинная обработка первичных данных предусматривает комплекс следующих операций и процедур:

1. Выбраковку, т.е. отсеивание из общего массива собранных анкет, которые заполнены не полностью, небрежно, с поправками, исключающими возможность четкого определения ответа респондента, или другими нарушениями инструкций их заполнения. Такого рода исследовательские документы признаются браком и удаляются из дальнейшего использования.
2. Ремонт выборки, необходимость которого обусловлена выбраковкой. Отсеивание ряда документов ведет к нарушению параметров ранее определенной выборки. Поэтому возникает потребность возмещения забракованных документов новыми, т.е. проведения дополнительного сбора информации. Эта операция, как и первая,

делается и при ручной обработке.

3. Докодирование полученной информации - ее унификация и формализация до степени, при которой она становится пригодной к вводу в ЭВМ. Это трудоемкая техническая операция начинается в период полевых работ, но может быть завершена только после получения всего массива первичных данных. Уже хотя бы потому, что пронумеровать все полученные документы, точнее, присвоить каждому из них свой код, можно только имея их все. Кроме того, все варианты мнений респондентов, подлежащие формализации, выявятся только по завершению полевого этапа.

4. Ввод данных в ЭВМ и организация введенной информации в соответствии с используемым пакетом прикладных программ машинной обработки.

5. Проверка качества введенной информации и коррекция обнаруженных ошибок. Для этого используются специальные процедуры обнаружения искаженной информации и ее исправления, а также программы корректировки и преобразования данных, входящих в общий пакет программ машинной обработки.

6. Уточнение системы количественных показателей, на основе которых будет производиться весь последующий анализ информации, и формулирование заданий на их машинное установление. При этом учитывается, с одной стороны, содержащаяся в исследовательской программе операционализация понятий, а с другой, планируемое использование методов математического анализа полученной информации.

7. Исполнение вышеназванных заданий, получение искомых цифр, таблиц, шкал, графиков, индексов, коэффициентов и других количественных характеристик изучаемых явлений и процессов.

Особую значимость для последующего анализа информации имеют матрицы данных, без которых невозможно применение большинства математических методов. В социологии используют пять основных разновидностей матриц: а) матрица «объект-признак»;

б) таблица сопряженности признаков;

в) матрица коэффициентов связей между признаками;

г) матрица расстояний между объектами;

д) матрица близостей между объектами или признаками.

Получение этих матриц - главный критерий завершения этапа компьютерной обработки первичной социологической информации количественного свойства.

Обработка первичной социологической информации может проводиться вручную, с помощью средств малой механизации, с использованием компьютерной техники. Основные этапы обработки первичной информации следующие:

1) Разрабатываются логическая схема обработки и анализа получаемых данных. В ходе этого этапа устанавливаются формы документов для сбора информации, методы её кодирования для ручной и машинной обработки, методы контроля данных и устранения ошибок. Определяются порядок и методы обработки данных, алгоритм расчётов, разрабатываются система анализа полученных в ходе обработки данных, основные направления анализа.

2) В случае обработки данных компьютерными методами осуществляется разработка математического обеспечения, выясняется, какие необходимы программы для обработки материалов исследования, иногда разрабатывается новое программное обеспечение.

3) Подготовка данных первичной социологической информации к обработке. Работа эта очень трудоёмка. Так, при обработке данных анкетирования открытые вопросы анкет «закрывают» – классифицируют по определённым признакам, систематизируют и кодируют в соответствии с классификацией. Осуществляют проверку анкет на качество заполнения. Эта проверка включает три момента:

- на полноту заполнения (ответ может отсутствовать из-за нежелания респондента отвечать, непонимания им вопроса, небрежности в заполнении анкеты). При невозможности устранить ошибку отдельные вопросы или вся анкета изымается из обработки. Часто устанавливают «критерий полноты заполнения», например, процент незаполненных вопросов анкеты, при превышении которого она будет изъята из обработки. В случае большого изъятия, ставящего под сомнение репрезентативность выборки, возможно возникновение необходимости дополнительного сбора исходных данных;

- на надежность (определяется отклонение от репрезентативной выборки, с помощью контрольных и фильтрующих вопросов проверяется качество информации, устраняются противоречивые ответы, умышленно недостоверные, отфильтровываются ответы или анкеты лиц, некомпетентных в исследуемых

вопросах т.д.);

- на технологичность. Все ответы необходимо привести к виду, дающему возможность легко перенести информацию на машинный носитель для обработки. Из анкет убирают все пометки, которые можно неоднозначно трактовать, номера выбранных ответов чётко обводятся ручкой.

Далее подсчитывают все документы, входящие в обрабатываемый массив информации, каждому присваивается порядковый номер. Информация кодируется, т.е. категориям документа присваиваются условные обозначения (шифр, код). Если все вопросы анкеты закрыты, кодирование может осуществляться в ходе её разработки. При наличии полузакрытых и открытых вопросов кодировать информацию возможно только после их «закрытия». Закодированную информацию переносят на машинные носители, контролируют качество переноса и устраняют ошибки.

4) Обработка информации (расчёт средних величин, установление корреляционных связей, составление группировок, таблиц, графиков и пр.)

Источники:

https://studwood.ru/640554/sotsiologiya/obrabotka_pervichnoy_informatsii

https://studopedia.ru/9_110234_obrabotka-pervichnoy-sotsiologicheskoy-informatsii-analiz-rezultatov-issledovaniya.html