МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования
**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»**

Кафедра социальной информатики

ОБЕСПЕЧЕНИЕ КАЧЕСТВА СОЦИОЛОГИЧЕСКОЙ ИНФОРМАЦИИ НА ЭТАПЕ СОЗДАНИЯ ЭЛЕКТРОННЫХ БАЗ ДАННЫХ

(автореферат бакалаврской работы)

студентки 5 курса 531 группы
направления 09.03.03 - Прикладная информатика
профиль Прикладная информатика в социологии
Социологического факультета
Цыра Юлии Сергеевны

Научный руководитель
кандидат социологических наук, доцент _____ К.В. Мохнаткина
подпись, дата

Зав. кафедрой
кандидат социологических наук, доцент _____ И.Г. Малинский
подпись, дата

Саратов 2021

ВВЕДЕНИЕ

Актуальность темы исследования. Высокие темпы развития социологии должны сочетаться с повышением качества результатов социологических исследований и ростом ценности социологической продукции.

Особое значение имеет повышение качества результатов социологических исследований, которых с каждым годом проводится всё больше. Реализация социологического исследования требует значительных финансовых затрат. Очевидно, что эффективность этих затрат напрямую зависит от качества полученной информации. Кроме экономических соображений, не меньшую роль играет и другой фактор. Прикладные социологические исследования чаще всего ориентированы на получение результатов, которые впоследствии будут применяться для решения тех или иных задач на практике. Это особенно касается исследований, финансируемых учреждениями, предприятиями и организациями. Недостоверные результаты исследований, использованные на практике, могут нанести обществу ущерб, намного превосходящий потери, связанные с материальными затратами. Вот почему вопрос качества социологической информации столь важен.

Разумно говорить, что совершенствование методов сбора социологической информации отстает от прогресса в методах обработки информации. В связи с этим все чаще наблюдается ситуация, при которой информация сомнительного качества подвергается обработке и анализу с помощью тончайших математических методов и электронно-вычислительных машин.[1]

Контроль за качеством социологической информации вызывает трудности. Использование надежного метода оценки качества полученных результатов — повторное осуществление другими учеными того же

[1] Методы сбора и анализа информации в социологических исследованиях: Кн. 1и2. — М., 1990.

исследовательского проекта — в социологии чаще всего невозможно. Это, помимо других причин, является результатом динамичности социальных явлений. Объект исследований социологов находится в состоянии постоянных изменений. Помимо этого в естественных науках (физике, биологии, химии) воспроизводство эксперимента является предварительным условием для того, чтобы полученный результат был введен в оборот науки.

Проблемы в оценке качества социологической информации увеличивают моральную ответственность социолога за публикуемые результаты его исследования. Особенно велика моральная ответственность социологов, осуществляющих исследования на основе контрактов с заинтересованными организациями. Эти организации, как правило, лишены специалистов-социологов, способных оценить обоснованность представленных результатов.[1]

Качество социологической информации формируется под влиянием многих факторов. Оно зависит от уровня работы коллектива социологов на каждой стадии исследовательского проекта. В дальнейших разделах книги основное внимание будет уделено работе социологов на стадии сбора социологической информации, и прежде всего качеству социальной статистики, получаемой с помощью опроса.

Несмотря на то, что в последние годы часто подчеркивается, что этот метод сбора информации не является единственным, опрос, тем не менее, по степени применения социологами уверенно держит первое место среди других методов сбора информации. Между тем качество информации, получаемой с помощью опроса, вызывает особенно большие нарекания: Такое положение не является случайным. Ведь использование опроса для получения информации предполагает, что интересующее социолога явление изучается не прямо, а косвенным путем.

[1] Ядов В.А. Стратегия социологического исследования: описание, объяснение, понимание социальной реальности. — М., 2000.

Между социологом и объектом исследования стоит респондент, который далеко не всегда является источником получения доброкачественной информации. В известном смысле респондента можно рассматривать, используя понятия инженерной психологии, как оператора — участника социологического исследования.

Повышение уровня социологических опросов требует тщательного изучения возможностей респондентов выдавать информацию высокого качества. Решение этой задачи предполагает, чтобы процесс получения информации во время опросов был сам объектом изучения социологов, психологов, статистиков.

Социологи недостаточно внимания уделяют систематическим ошибкам в своих исследованиях. В их публикациях редко можно встретить соображения о том, какова степень точности полученных ими результатов с учетом возможностей появления различных систематических ошибок на всех стадиях социологического исследования.[1] В то же время широко приводятся результаты социологических исследований с точностью до десятых, а иногда и сотых долей процента.

При принятии решений в исследовательском проекте приходится в максимальной степени учитывать конкретные особенности объекта исследования, размер выделяемых ресурсов, характер требований заказчиков. Поэтому успех прикладных социологических исследований зависит не только от того, в какой мере социолог овладел общими методологическими и методическими принципами своей науки, но и от его эрудиции, от опыта, от знания многообразных «казусов», когда та или иная процедура оказывалась удачной, а другая, наоборот, была несовместима с условиями исследования.[2]

В этом отношении исследователь должен в чем-то походить на хорошего врача, сильного не только знанием теоретической медицины, но и своим богатым

[1] ЭБС «Znanium. com.» Добреньков В.И. Кравченко А.И. Методы социологического исследования: учебник. — М., 2013.
[2] Американская социология: Перспективы, проблемы, методы. — М., 1972.

клиническим опытом, интуицией, опираясь на которые он принимает решения, максимально учитывающие индивидуальные особенности пациента.

Повышение качества социологической информации невозможно без разработки таких методик, которые бы в максимальной степени учитывали специфику объекта исследования. Поэтому обратим основное внимание на разнообразные технические особенности социологического исследования, порожденные природой конкретных объектов изучения. При этом следует иметь в виду, что массовый опрос, даже в том случае, если он организован в соответствии с современными научными требованиями, чаще всего не может служить основным, а тем более единственным источником сведений об изучаемых социальных явлениях. При проектировании социологических исследований следует, учитывая уязвимые места массового опроса, использовать и другие, более надежные способы получения информации.[1]

Внедрение компьютеров буквально во все сферы человеческой деятельности является на сегодняшний день, наверное, самым очевидным итогом научного прогресса. И как следовало ожидать, в существенной степени компьютеризация изменила характер самих научных исследований, в том числе в психологии и социальных науках.[2] Компьютер обычно применяется исследователями для выполнения такой работы, которая считается самой скучной и утомительной: учет и организация исходных данных, вычисления различных показателей и пр.

Анализ данных с применением компьютера включает выполнение ряда необходимых шагов.

1. Определение структуры данных.

2. Ввод данных в компьютер в соответствии с их структурой и требованиями программы.

---

[1] Анализ данных [Электронный ресурс] : учеб. для акад. бакалавриата / под ред. В.С. Мхитаряна. - М. : Юрайт, 2017.

[2] Социологическая энциклопедия: В 2 т. — М., 2003.

3. Задание метода обработки данных в соответствии с задачами исследования.

4. Получение результата обработки данных.

5. Интерпретация результата обработки.

Последовательность шагов, требуемая для решения задач социологического исследования, продолжает оставаться жестко заданной. Каждый шаг по-своему важен, его практически нельзя исключить или выполнить в другом порядке. Например, нельзя вводить информацию, предварительно не закодировав ее, или пытаться выполнить статистический анализ, не проведя контроля введенных данных. Обработку собранных в поле данных лучше всего выполнять в приведенной ниже последовательности, поэтапно: подготовительный этап; ввод и корректировка данных; контроль данных; получение результатов статистических процедур; анализ данных и подготовка отчета.

Основной смысл подготовительного этапа состоит в выполнении работ, обеспечивающих адаптацию анкеты к виду, позволяющему использовать средства автоматизации при ее обработке и выполнении расчетов. Еще на этапе разработки инструментария в бланке формализованного интервью во всех закрытых вопросах было выполнено кодирование ответов опрашиваемых числами. Эти числа и использовались интервьюерами при фиксации ответов респондентов. Следующий важный шаг – присвоение каждому вопросу анкеты восьмисимвольного смыслового имени. Причем первым символом имени должна быть буква.[1] Эти имена и становятся именами переменных с момента их введения в систему. Переменная – это вопрос анкеты и набор ответов (их кодов) к нему. В результате выполнения рассматриваемого шага к массиву

[1] Егорова, У.Г. Статистический анализ данных как критерий качества психологических и педагогических исследований / У.Г. Егорова // Педагогическая наука сегодня: философско-методологические проблемы : сб. науч. ст. [содержит материалы Всерос. метод. семинара, состоявшегося 22-23 апр. 2011 г.] / РАО, Науч. совет по философии и проблемам методологии исслед. в образовании, Моск. ин-т открыт. образования, Центр инновац. моделей образования, Моск. гор. пед. ун-т, Лаб. философии образования ; [науч. ред. Е.В. Бережнова ; сост. Н.В. Малкова]. - М., 2011. - С. 89-93.

анкет с первичной информацией добавляется еще один бланк со всем индикаторами, расписанными по смысловым именам – переменным – со всеми возможными в данном исследовании кодами индикатора, а также с указанием размера ячейки (ширины переменной). Форма такого бланка называется «Макет ввода данных в ЭВМ». В других работах сходный по назначению документ называется «кодировочной таблицей». [1]Подытожим задачи, которые решаются при составлении кодировочной таблицы:

1. Кодировочная таблица устанавливает соответствие между отдельным вопросам анкеты и переменными.

2. Кодировочная таблица устанавливает соответствие между возможными кодовыми числами и значениями переменных. Перед вводом данных выполняется визуальный контроль правильности и полноты заполнения анкеты и кодировки. Этот контроль позволяет выявить ошибки в заполнении анкет, которые возникают в результате неправильных записей, произведенных в анкете интервьюером, найти логические несоответствия (перепутан принятый в анкете порядок записи членов семьи, что в дальнейшем при панельном обследовании делало некорректным проводимый анализ), обнаружить ошибки в расчетах, выполняемых внутри анкеты. Результатом указанных работ оказывается массив полевой документации, который теперь уже подготовлен к вводу данных. Основной смысл подготовки базовой таблицы к вводу данных как раз и состоит в выполнении предварительных работ по созданию электронной версии макета ввода данных. Формирование электронного макета ввода данных выполняется в специальном режиме Переменные редактора данных. Именно для этой цели на стадии подготовки инструментария выполняется работа по построению макета - присвоению уникального имени каждой переменной и заданию ее ширины. Выполнение последовательности действий по формированию таблицы [2]- вводу имен переменных и их описания, предполагает знание следующих важных особенностей структуры окна

[1] Прангишвили И.В. Системный подход и общесистемные закономерности. — М., 2001.
[2] Афанасьев В.Г. Социальная информация. — М., 1994.

редактора данных. Каждая строка таблицы представляет собой место для записи случая или наблюдения. Любая анкета вводимого массива данных в полевых условиях называется «наблюдением», а в электронном формате наблюдение принято именовать «случаем». При вводе данных число наблюдений (случаев) равно числу анкет. Каждая колонка представляет собой место для записи одной переменной. Любой вопрос анкеты имеет как минимум один индикатор и, следовательно, должен характеризоваться как минимум одной переменной. Соответственно, столько же колонок и переменных должно быть явлено в таблице окна редактора данных. Колонки и строки состоят из ячеек. Каждая ячейка представляют собой пересечение случая и переменной. Значение одной переменной записывается в одну ячейку.[1]

Специфика информационных процессов в социологии, в первую очередь, определяется особенностью социологической информации и спецификой используемых исследовательских методов. Основными источниками социологической информации являются: данные эмпирических исследований, главным образом, данные массовых опросов населения; программы и описания исследований; отчеты по проведенным исследованиям и законченным научно-исследовательским работам; сериальные издания по социальным проблемам. Система информационного обслуживания социологов должна организовываться с учетом комплексного, многопланового, системного характера социологической информации, с учетом изменения потребностей в различных видах и формах информации в зависимости от этапов социологического исследования.

Степень научной разработанности проблемы.

Не смотря на несомненную важность и значимость данной проблемы, изучается она не в плотную и упоминается только вскользь. Недостаточное внимание со стороны социологов к ошибкам прикладных исследований в известной степени является результатом довольно распространенного мнения, что встречающиеся ошибки носят в основном случайный характер и в силу

[1] Факторный, дискриминантный и кластерный анализ. — М., 1989.

действия закона больших чисел погашают друг друга. Для многих ситуаций, возникающих в социологических исследованиях, такое предположение является неверным. Поэтому детальное изучение факторов, от которых зависит возникновение систематических ошибок, представляется важной проблемой.

Объектом данного исследования является социологическая информация. **Предметом** – обеспечение качества социологической информации на этапе создания электронных баз данных. **Целью исследования** является выявление факторов, влияющих на качество социологической информации на этапе создания электронных баз данных, и обнаружение способов, устраняющих данные факторы.

Структура бакалаврской работы представлена введением, двумя разделами, заключение, приложением и списком использованных источников.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

В первом роазделе **«Качественные характеристики социологической информации»** приведены и рассмотрены основные этапы социологических исследований и описаны проблемы, возникающие у исследователя на том или ином этапе, приводящие к меньшей репрезентативности исследования впоследствии.

Подробнее всего я рассматривала момент, когда эмпирическая информация уже собрана и:

1. Подготавливается к обработке.

2. Производится контроль полноты и правильности первичных документов.

3. Осуществляется сквозная нумерация всех документов, прошедших содержательный контроль и включенных в обработку.

4. Подготавливается электронный макет для введения эмпирической фактуры в память компьютера. Производится перевод ее в электронную форму из физической.

5. Перенос (набивка) всей эмпирической фактуры с полевых документов (анкет, бланков и т.п.) в память компьютера (на жесткие диски) и образование так называемых файлов данных.

6. Получение линейных распределений частот значений всех переменных, включенных в полевой документ, а также необходимых или заданных таблиц сопряженности между ними.

7. Используются стандартные или специализированные пакеты прикладных компьютерных программ в диалоговом режиме для проведения оригинальных статистических или математических преобразований эмпирической информации и представления получаемых результатов в удобной (табличной, графической, диаграммной и т.п.) форме на экране монитора и/или (через принтер) на бумаге.

8. Анализ выходных форм.

Довольно существенную роль в комплексе мер по минимизации величины ошибок играет форма компьютерного макета полевого документа, а также удобство[1] правил переноса информации на машинные носители и возможности используемого программного обеспечения. Компьютерный макет полевого документа представляет собой отображение на экране монитора его формальной структуры. Правила работы с ним и его форма определяются техническими возможностями и требованиями конкретной компьютерной системы. Кроме прямого назначения компьютерный макет выполняет еще важную функцию: осуществляет некоторые виды логического контроля и контроля правильности ввода и содержания первичной информации. В частности, программно исключаются ошибки по максимальному коду. Например, в какой-то переменной (вопросе) всего пять возможных значений. Поэтому введение в поле значений данной переменной любого числа, большего пяти, вызывает автоматическую остановку ввода и появление соответствующего сообщения на экране. Поскольку курсор компьютера после

[1] Гилберт Дж., Малкей М. Открывая ящик Пандоры: Социологический анализ высказываний ученых. — М., 1987.

ввода значения каждой переменной автоматически смещается к следующей, то этим самым обеспечивается высокая вероятность обнаружения довольно часто встречающейся ошибки сдвига (пропуска какого-либо кода или введения лишнего). Иключительный случай, когда система оказывается «нечувствительной» к ошибке сдвига - это моменты, когда число пропусков равно числу лишних и повторных вводов. Самый надежный и радикальный способ обеспечения качества информации (который, надо заметить, одновременно и самый трудоемкий и дорогой) — «набивка в две руки». Всё множество документов вводят два оператора отдельно друг от друга. Когда информация введена полностью, производится сравнение двух полученных массивов. Обнаруженные различия [1]сверяются с оригиналами и устраняются. Цель предварительного контроля — исключить из обработки и последующего содержательного анализа документы, не удовлетворяющие определенным, задаваемым исследователем критериям качества. Результат процедуры контроля двоякий. С одной стороны, повышается содержательная достоверность массива документов. С другой стороны, общий объем массива документов сокращается, вследствие чего нарушаются критериальные квоты, заложенные при проектировании выборки в процедуру отбора респондентов, и снижается репрезентативность информации в целом. В файлы социологической информации помимо собственно эмпирической фактуры включают и файлы вспомогательной информации, [2] которая состоит из макета документа, текста документа и паспорта исследования.

Существует мнение, что методы преобразования социологической эмпирии, рассмотренные выше, слишком простые и слабые, недостаточно эффективные и по этой причине не могут выявить всей многомерности заключенного в ней содержания. В результате, опускается чрезвычайно важное обстоятельство: использование математического метода анализа совокупности формализованных эмпирических данных возможно только если на этой

[1] Кун Т. Структура научных революций. — М., 1975.
[2] Чесноков С.В. Детерминационный анализ социально-экономических данных. — М., 1982.

совокупности работают требования системы аксиом, которая определяет возможности в решении конкретной задачи для него. Одной из задач социологического анализа эмпирической информации является поиск причин, объясняющих тот или иной характер обнаруженных социальных процессов. Решение такой задачи в статистике производится при помощи целого списка подходящих мер, которые предъявляют определенные требования к анализируемой эмпирической фактуре. Самые важные из этих требований - случайность и независимость результатов испытаний. Социологи легко впадают в соблазн использования в своей предметной сфере готового и, как им представляется, весьма «эффективного» арсенала средств.

Как известно, явления социального мира, которые изучает социология, чаще всего хорошо не отображаются с помощью формальных моделей. По сути дела, вся математизация социологической науки до недавнего времени развивалась по пути заимствования и внедрения ею математического аппарата из других областей знания. Наиболее активно внедрялись методы математической статистики и теории вероятностей. Среди них в первую очередь активно используются те, с помощью которых реализуется так называемый многомерный статистический анализ. Это дискриминантный, факторный, канонический, корреляционный, кластерный и логлинейный анализ. Каждый из этих методов может также использоваться автономно для решения соответствующих частных задач. В многочисленной и разнообразной литературе, посвященной описанию сущности этих методов, каждый из них преподносится как «самый эффективный метод статистического анализа социологической информации». Даже в «родных» для упомянутых методов — теории вероятностей и математической статистике — применение каждого из них в каждом конкретном случае требует тщательного обоснования. [1]

Во втором разделе «**Способы повышения уровня качества социологической информации на этапе создания электронных баз данных»** я рассмотрела на примере своего же исследования проблемы, качественного

[1] Берка К. Измерения. Понятия, теории, проблемы. — М., 1987.

характера на этапе внесения информации в программный пакет SPSS Statistics 22. Чтобы рассмотреть качество социологической информации и понять, с какими трудностями можно столкнуться на этапе создания электронных баз данных, мною было проведено авторское социологическое исследование, цель которого - выявить политические предпочтения молодёжи. Выборочная совокупность составила 100 человек из которых 51% мужчины и 49% женщины. Опрос был проведён путём анкетирования.

Отслеживание качества вводимых данных я начала с этапа просмотра и анализа опросников. Один опрошенный не ответил на вопрос «Участвовали ли вы в выборах президента РФ 18 марта 2018 года?», но при этом на вопрос «Участвуете ли вы в выборах?» дал ответ «да, во всех проходящих», что даёт мне возможность логически восстановить ответ.

Опросников, в которых респонденты не ответили более чем на двадцать процентов вопросов либо на 2–3 вопроса в социально-демографическом блоке, у меня не оказалось. В противном случае я была бы вынуждена их исключить из основного массива как некачественные, способные внести искажение в социологическую информацию.

При контроле анкет также проверила их на предмет наличия противоречивых ответов. В ходе проверки выявила, что у трёх респондентов в вариантах ответов прослеживаются противоречия: в начале анкеты на вопрос «Интересуетесь ли Вы политическими событиями, происходящими в стране?» респонденты ответили отрицательно, а в конце анкеты на вопрос «Участвуете ли Вы в политических мероприятиях (митинги, собрания)» ответили положительно. Такого рода противоречие относительно просто было бы снять, скорректировав в первом вопросе варианты ответа, таким образом: «интересуюсь всегда», «интересуюсь иногда», «не интересуюсь совсем». Ну а в моём случае, я скорректировала анкеты данных респондентов, исходя из ответа на второй вопрос.

В более сложных ситуациях подобные вопросы из компьютерной обработки следовало бы исключить. А если из анкеты исключено более 20 % вопросов, она подлежит выбраковке.

Нередко респондент, несмотря на имеющееся указание выбрать один, два или три варианта ответа, обводит (подчеркивает) на несколько кодов больше или меньше, что затрудняет коррекцию ответа. На практике иногда сохраняют первые отмеченные коды, однако это чревато существенными погрешностями. Такие вопросы целесообразно не вводить в компьютер. В моём опросе с данной проблемой я не столкнулась.

После выбраковки непригодных для компьютерной обработки анкет составляется бланк кодировки ответов респондента на открытые вопросы. После ввода в компьютер текстовой информации ответы перечневого характера автоматически группируются и обрабатываются. Ответы текстового характера в вопросе № 14 я предварительно выписала вручную, отмечая частоту повторяемости утверждений. Такая процедура кодировки открытых вопросов значительно ускоряет обработку, особенно в тех случаях, когда ответы по смыслу разноплановые. После подсчета частоты идентичных по смыслу суждений близкие по содержанию объединяются в одну группу. В результате этого ответы свелись к двум индикаторам и двум системным ответам, которым я присвоила числовой код и ввела в компьютер.

В результате компьютерной обработки первичной социологической информации получаются табуляграммы, содержащие сгруппированные данные в форме: линейного распределения ответов на вопросы в абсолютных числах и процентах; парное, тройное и иные распределения информации при сочетании вариантов ответов на два-три вопроса анкеты и более; взаимозависимое распределение некоторой группы ответов; средние значения, дисперсии, коэффициенты корреляции и другие статистические величины для информации, собранной на основе интервальной шкалы.

При вводе первичной социологической информации в компьютер возможны случайные ошибки: в результате нажатия не той клавиши ввода,

пропуска какого-либо кода, особенно в вопросах табличной формы. Поэтому после ввода информации целесообразно проконтролировать ее на экране компьютера. Кроме того, в программе SPSS предусмотрен контрольный ограничитель, сигнализирующий оператору о недопустимой операции (например, о вводе отсутствующего в вопросе кода либо числа, превышающего общее число вариантов ответа на вопрос).

Число не ответивших на тот или иной вопрос следует обязательно ввести в компьютер в виде отдельного кода (в качестве такового можно выбрать код «0»).

Для альтернативных вопросов, сумма ответов на которые в обязательном порядке равна 100% (например, выбор из 10 политиков при условии голосования только за одного), дополнение до 100% (т.е. учет не ответивших) компьютер может осуществлять автоматически. Следовательно, если в вопросе имеется позиция «затруднились ответить», предназначенная для селекции нефункциональных ответов, которые не будут использоваться при анализе информации, к ним можно присоединить и отсутствие ответов (компьютер это сделает автоматически, если не ответивших просто не кодировать).

Для ускорения ввода первичной информации в компьютер массивы анкет можно разбить на подмассивы в соответствии с числом операторов и вводить одновременно в несколько компьютеров.

По окончании ввода информации файлы с данными в различных компьютерах объединяются в один, после чего необходимо подсчитать линейное распределение данных с тем, чтобы по каждому вопросу проверить наличие ошибочных кодов, случайно введенных операторами при перфорировании анкет.

Массив информации с исправленными ошибочными кодами может быть подвергнут полной компьютерной обработке.

Для решения таких методических задач, как отработка модели выборки, оценка вариации рядов распределения, устойчивости показателей, полный массив

анкет может быть разбит на основании выбранных исследователем принципов на подмассивы, которые обрабатываются автономно.

ЗАКЛЮЧЕНИЕ

В результате дипломного проектирования были рассмотрены ошибки, влияющие на качество социологической информации и способы их устранения на этапе создания электронных баз данных

В ходе выполнения ВКР достигнуты следующие результаты:

1. Сделана выборка соответствующих материалов из литературы по заданной тематике.
2. Изучены существующие в настоящее время методы повышения качества социологической информации.
3. Изучена эффективность действия самых популярных методов.
4. Указаны преимущества и недостатки при применении популярных методов.
5. Проанализировано имеющееся исследование с помощью популярных методов решения проблемы качества социологической информации.