

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

кафедра социальной информатики

**ОСОБЕННОСТИ ИСПОЛЬЗОВАНИЯ НОМИНАЛЬНЫХ ШКАЛ В  
ПРОЦЕССЕ ОБРАБОТКИ СОЦИОЛОГИЧЕСКОЙ ИНФОРМАЦИИ НА  
БАЗЕ ПРОГРАММЫ SPSS**

(автореферат бакалаврской работы)

студентки 5 курса 531 группы  
направления подготовки 09.03.03 – «Прикладная информатика»  
профиль «Прикладная информатика в социологии»  
Социологического факультета  
Лариной Екатерины Владимировны

Научный руководитель  
старший преподаватель

\_\_\_\_\_ М.В. Колесниченко

подпись, дата

Зав. кафедрой  
Кандидат социологических наук, доцент

\_\_\_\_\_ И.Г. Малинский

подпись, дата

Саратов 2024

## ВВЕДЕНИЕ

**Актуальность бакалаврской работы.** На рубеже XX века эмпирическая социология начала активно развиваться, открывая новые горизонты в изучении общественных явлений. Важно отметить, что в середине XX столетия произошел настоящий расцвет массовых социологических исследований. Этот период ознаменовался значительными достижениями в области эмпирической социологии, которые были невозможны без предшествующей истории развития этой науки.

Выбор программы для обработки статистических данных зависит от конкретной задачи и требований исследования. Некоторые программы ориентированы на универсальное применение, другие специализируются на определенных областях знаний, что позволяет исследователям выбирать оптимальный инструмент в зависимости от поставленных задач.

Изучение социальных характеристик играет ключевую роль в понимании общества. Оно позволяет не только теоретически оценить различные аспекты жизни людей, но и применить полученные данные на практике.

Существует несколько причин, по которым измерение социальных характеристик является необходимым и важным процессом. Важнейшим практическим преимуществом измерения социальных характеристик является возможность использования математических методов для более глубокого анализа социальных явлений.

Однако стоит отметить, что измерение социальных характеристик обладает своими особенностями. Одной из них является качественный характер данных, что влияет на выбор и использование категориальных шкал для оценки.

Исследование социальных характеристик является сложным и многогранным процессом, который требует внимательного и глубокого анализа. Однако благодаря правильному измерению и интерпретации данных, мы можем получить ценные знания о социуме и его участниках. Исходные данные, важные для проведения исследований, часто зависят от характеристик

социальной среды, в которой они собраны. Например, такие факторы, как возраст, образование, социальное положение и другие, не всегда могут быть количественно измерены, что ограничивает возможности анализа. Важно понимать, что качество данных напрямую связано с качеством исследования. Одним из ключевых ограничений является неспособность измерить определенные социальные характеристики иначе как на номинальной шкале. Это ограничение затрудняет использование стандартных математических методов анализа данных и требует разработки альтернативных подходов. Важно помнить, что широкое применение номинальных шкал ограничивает исследователя в выборе методов анализа информации. Это требует поиска новых подходов к обработке данных и развития специализированных методов для работы с качественными характеристиками социальных данных.

Исследование социологических процессов требует применения теории измерений, которая позволяет моделировать их с помощью чисел. Этот подход помогает выделить важные аспекты реальности и абстрагироваться от ненужных моментов. Построение шкалы в соответствии с этой теорией отражает диалектическое единство качественных и количественных аспектов исследуемых явлений.

Эффективное изучение интересующих социологов процессов требует учета различных аспектов их характеристик. Понимание и использование теории измерений помогает выделить ключевые моменты и анализировать их с точки зрения числовых моделей. Это позволяет более глубоко понять и объяснить сложные социальные явления.

Решение проблемы безработицы может быть рассмотрено с использованием теории измерений. Анализируя данную проблему с учетом количественных и качественных аспектов, исследователи могут прийти к новым выводам и рекомендациям. В данной работе мы постараемся продемонстрировать, как использование теории измерений может помочь в поиске решений для социальных проблем.

**Степень изученности.** Такая популярность программы не могла не вызвать публикацию многочисленных учебников, пособий и руководств по работе с SPSS. Конечно, первые работы, посвященные особенностям анализа социологических данных при помощи компьютерных технологий, были переводными. Среди авторов таких работ можно назвать А. Бююля, П. Цефеля, Дж. Хили. Работы отечественных исследователей появились вскоре после публикации первых пособий зарубежных авторов по этой тематике и вскоре стали довольно разнообразными. Областью статистической — обработки информации заинтересовались социологи, психологи, экономисты и мн.др.

Социологические — аспекты проведения эмпирического исследования изучаются социологами довольно пристально в течение длительного времени. Еще основатель социологии О. Конт заявил о необходимости эмпирической проверки теорий, возникающих в новой науке об обществе. Известнейшие социологи классического периода Э. Дюркгейм и М. Вебер следовали тем же принципам, в том числе и на практике, написав свои уже всемирно известные работы.

Ирландские статистики Дж. Граунт и У. Петтипо праву считаются основателями так называемой политической арифметики — направления социальных исследований официальных данных для целей оптимизации государственного управления. Фундамент социальной статистики был заложен в XX в. франко-бельгийским ученым Адольфом Кетле, открывшим ряд статистических закономерностей общественной жизни. Он одним из первых стал использовать в статистике достижения математики и ее методы, прежде всего теорию вероятности, которую применил в процессе анализа социальных явлений. Не отставали и отечественные исследователи. Так, в работе А.Г. Ковалевского были изложены основные идеи, положившие начало современной теории выборки.

Мировую известность имел А.А. Чупров, фамилия которого дала название одному из наиболее часто используемых в наше время коэффициентов парной связи между номинальными признаками”. С тех пор

интерес исследователей к использованию математического аппарата в социологии не снижается. Только среди отечественных социологов разработкой общеметодологических вопросов проведения социологического исследования занимаются Ядов В.А., Здравомыслов А.Г., Батыгин Г.С., Шубкин В.Н., анализом конкретных методов сбора эмпирической информации Белановский С.А., Исупова О., Козина И.Г., проблемой сопряжения социологической информации и компьютерных технологий - Крыштановский А.О., Божков О.Б.З, проблемами измерения в социологии - Девятко И.Ф, Маслова О.М., Паниотто В.И., математическими методами обработки социологических данных — Андреенков В.Г., Толстова Ю.Н., Татарова Г.Г.

**Объектом** выступает использование номинальных шкал, как важнейшего инструмента обработки социологических данных.

**Предметом** данного исследования являются возможности программы обработки статистических данных SPSS, представленных в формате номинальных шкал

**Цель исследования** — раскрыть возможность взаимодополнения аналитических операций, имеющихся в программе SPSS, обусловленного, с одной стороны, разными возможностями отдельных методов, а с другой - особенностями исторического развития самой программы SPSS (на примере использования номинальных шкал).

Постановка цели определила формулирование следующих задач исследования:

1. описать аналитические возможности программы SPSS, а также выявить особенности ее структуры, обусловленные, в том числе, историей развития данной программы;

2. определить специфику номинальных шкал как инструмента измерения социальных признаков;

3. проанализировать особенности инструментов, предлагаемых в программе SPSS, для работы с переменными, имеющими номинальные шкалы.

Методологической базой исследования выступает принцип позитивистского подхода к изучению социальной реальности, впервые сформулированный основателем социологии Огюстом Контом и заключающийся в необходимости количественного представления социальных явлений и процессов<sup>1</sup>.

Кроме того, в работе используются элементы генетического, сравнительного и системного подходов.

В качестве эмпирической базы исследования были использованы данные массового социологического опроса, проведенного студентами социологического факультета СГУ по теме «Безработица малого города»<sup>2</sup>.

**Структура бакалаврской работы.** Данная работа состоит из введения, трех разделов, заключения, списка использованных источников и приложений.

## **ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ**

**Во введении** обосновывается актуальность темы исследования, раскрывается степень её разработанности, определяются цель, задачи, объект и предмет, методологическая основа и теоретические источники.

В первом разделе «**Программа SPSS как инструмент статистического анализа социологической информации**» представлен общий обзор возможностей, предлагаемых статистическим пакетом SPSS, а также предпринята попытка выявить особенности данной программы.

Статистический пакет IBM SPSS Statistics является одним из старейших среди программных продуктов, ориентированных на анализ информации из разных областей знания (в том числе и в социологии), и доступным как для специалистов программирования и математики, так и для тех, кто специализируется в других областях. Доступность SPSS для широкого круга пользователей обеспечивается тем, что вызов необходимых процедур

---

<sup>1</sup> Радугин А.А., Радугин К.А. Социология: курс лекций. М., 1996. С.8

<sup>2</sup> База данных содержит информацию по социологическому исследованию, проведенному в 2022-2023 г методом опроса, в ходе которого были проанкетированы безработные г. Татищево Саратовской области. Выборочная совокупность построена по целевой выборке, насчитывает 87 респондентов

статистического анализа происходит при помощи стандартной техники, применяемой в Windows, с помощью мыши и диалоговых окон. При этом предусмотрены и дополнительные возможности программирования при помощи Синтаксиса.

К несомненным достоинствам пакета SPSS относится и его оконная структура, которая напоминает принцип организации Windows и требует минимальных затрат энергии для освоения. Также удобство работы с SPSS обеспечивается его сопряженностью с другими распространенными программами, среди которых можно назвать текущие версии MSOffice: Excel/Word/PowerPoint, а также в PDE и HTML формат. Располагает к этому статистическому пакету его модульная система, которая делает программу доступной для пользователей с разными финансовыми возможностями.

Особая привлекательность этого пакета для социологов связана с тем, что SPSS предлагает широкий выбор методов анализа переменных с разными шкалами. В том числе здесь имеются большие возможности анализа номинальных и порядковых шкал, которые оказались «крепкими орешками» для статистики, но являются очень распространенными в общественности.

Вместе с тем, нацеленность программы на широкую аудиторию пользователей, среди которых могут быть и обществоведы, и медики, и биологи, и математики, и программисты-информатики, обусловила предоставление очень подробной итоговой информации в выводах. Данная информация детально описывает практически каждый этап анализа включая промежуточные автоматические расчеты и команды. В результате простым исследователям, не специализирующимся в точных науках, приходится прилагать дополнительные усилия для фильтрации нужной информации. К счастью, большое число специализированных пособий, содержащих пошаговые инструкции, облегчают выполнение данной задачи.

Наконец, отметим последний момент. Длительная история развития SPSS, включающаяся в расширение аналитических возможностей программы через добавление все новых и совершенствование уже существующих модулей,

привела довольно неожиданному результату. В настоящее время SPSS предоставляет Пользователю не просто широкие возможности, а чрезмерно широкие. Для обращения к некоторым методам анализа в программе существует несколько алгоритмов действий. С одной стороны, это делает более сложным процесс освоения программы для начинающих исследователей, с другой — уже продвинутый пользователь может выбрать для себя наиболее удобный алгоритм действий.

Второй раздел **«Номинальные шкалы как инструмент измерения социальных характеристик»**, построен на анализе особенностей применения номинальных шкал в процессе анализа социальных характеристик.

Формализация социальных явлений есть сложный многоэтапный, субъективный процесс. Из этого следует необходимость постоянного контроля за тем, какой моделью реальности пользуется исследователь. В целом процесс конструирования модели включает 3 этапа: формирование априорной содержательной модели, построение концептуальной модели и конструирование формальной (математической) модели. Особым моментом в конструировании модели является выбор не только способов сбора данных и объектов измерения, но и способов их измерения, что напрямую будет определять выбор исследователем методов анализа данных. Измерение в социологической практике осуществляется при помощи трех типов шкал: номинальной, порядковой и количественной.

Номинальная шкала является наиболее простой и очень распространенной в социальных исследованиях. Это шкала, допустимыми преобразованиями которой являются произвольные взаимно однозначные преобразования, т.е. преобразования, сохраняющие отношения равенства и неравенства между числами.

В то же время существуют критерии, которых надо придерживаться, чтобы обеспечить адекватность измерения при помощи номинальных шкал: взаимоисключаемость, — исчерпываемость и однородность — категорий



переменных, а также запрет на непосредственную математическую обработку данных, полученных при помощи номинальных шкал.

Перечисленные особенности номинальных шкал накладывают серьезные ограничения на возможности их математической обработки. В распоряжении исследователя, работающего с номинальными шкалами, остается относительно небольшой перечень допустимых операций: подсчет частоты встречаемости каждой категории, поиск средней тенденции по модальной частоте, установление взаимосвязи между признаками с помощью перекрестных таблиц и статистических показателей (критерия независимости хи-квадрат, коэффициентов корреляции  $\Phi$  Фишера,  $V$  Крамера), логлинейный анализ таблиц сопряженности, логистическая регрессия и некоторые другие.

Третий раздел **«Прикладной потенциал номинальных шкал как инструмента измерения социальных признаков (на примере анализа безработицы в малом городе)»** посвящен представлению результатов анализа, направленного на построение модели безработицы в условиях малого города.

В качестве методов анализа данных нами были выбраны построение таблиц сопряженности, коэффициенты Хи-квадрат,  $V$  Крамера и  $\Phi$  Фишера, а также логистическая регрессия, что было обусловлено особенностями построения вопросов в анкете.

Во-первых, переменные в опроснике (и' все переменные, отобранные нами для анализа) имеют номинальные шкалы. Это позволяет высокоэффективно использовать перекрестные таблицы и рассчитанный на их основе критерий независимости Хи-квадрат с его производными с учетом всех отмеченных ограничений. Во-вторых, около половины вопросов анкеты имеют дихотомическую шкалу. Кроме того, часть шкал мы трансформировали в дихотомические. Это дает нам возможность построения модели безработицы при помощи логистической регрессии.

В ходе исследования было выявлено много зависимостей, причем во многих случаях были высказаны предположения о латентном влиянии других переменных. Если критерий независимости Хи-квадрат позволяет выявлять

связи между исследуемыми переменными первого уровня (между двумя переменными), то для описания связей между тремя и более переменными более эффективным — аналитическим инструментом является метод логлинейного анализа.

Проведение логлинейного анализа подтвердило статистическую значимость главных эффектов переменных возраст, наличие детей, уровень образования, уровень потребления семьи, длительность поиска работы, Трудовой стаж и причины безработицы. Кроме отдельных переменных, значимым оказалось одно трехфакторное (семейное положение\*уровень потребления\*причины безработицы) и двухфакторных взаимодействий (возраст\*наличие детей, семейное положение\*наличие детей и др.).

**В заключении** содержатся основные выводы, соответствующие решению поставленных задач для реализации цели бакалаврской работы. Измерение признаков социальных явлений с использованием номинальных шкал является одним из наиболее распространенных методов. Например, признаки, такие как возраст, образование, семейное положение, несут важную информацию, которую можно представить именно через номинальные шкалы. Однако, следует отметить, что номинальные шкалы ограничивают возможности применения классических математических методов из-за своей базовой структуры.

Номинальная шкала, будучи самым "примитивным" инструментом измерения, требует особого подхода к анализу данных. Современная социальная статистика, однако, предлагает широкий спектр эффективных методов и техник для работы с номинальными данными. Кластерный анализ, ассоциативные правила, методы машинного обучения - все это позволяет извлекать ценные знания из номинальных шкал и применять их в различных социологических исследованиях.

Изучение номинальных шкал важно для понимания основ измерений. Номинальная шкала определяется как шкала, где допустимы произвольные взаимно однозначные преобразования, сохраняющие отношения равенства и

неравенства между числами. Важно отметить, что критерии адекватности измерений при использовании номинальных шкал включают в себя взаимоисключаемость, исчерпываемость и однородность категорий переменных.

Подробнее рассмотрим эти критерии. Взаимоисключаемость означает, что каждый объект может быть отнесен только к одной категории без пересечения с другими. Исчерпываемость подразумевает, что все объекты должны быть охвачены категориями без пропусков. Однородность категорий переменных требует, чтобы все категории были одинаково значимы и не имели внутренних иерархий.

Кроме того, номинальные шкалы не позволяют проводить непосредственную математическую обработку данных. Это означает, что операции типа сложения или умножения не имеют смысла в контексте номинальных шкал. Вместо этого для анализа данных с номинальными шкалами используются частоты, моды и другие статистические методы, не требующие арифметических операций.

При работе с номинальными шкалами исследователю доступны различные методы обработки данных, которые позволяют проводить более глубокий анализ и получать ценные выводы. Одним из ключевых ограничений номинальных шкал является ограниченный набор операций, которые можно применять. Среди этих операций можно выделить подсчет частоты встречаемости каждой категории, анализ модальной частоты для определения средней тенденции, установление взаимосвязей с помощью перекрестных таблиц и статистических показателей, таких как критерий независимости "хи-квадрат", коэффициенты корреляции  $\Phi$  Фишера и  $U$  Крамера, а также применение логистической регрессии и логлинейного анализа таблиц сопряженности. Эти методы помогают исследователю извлечь максимум информации из данных на номинальной шкале и сделать обоснованные выводы.

В городе Татищево было проведено социологическое исследование, результаты которого оказались несколько неожиданными. Оказалось, что среди долгосрочных безработных две трети составляют мужчины. Это может быть связано с различными факторами, такими как отсутствие квалификации или недостаток рабочих мест в сферах, где преобладает мужское население.

Однако не только мужчины сталкиваются с проблемой долгосрочной безработицы. Молодежь также имеет высокие шансы оказаться без работы на длительное время. Это может быть связано с отсутствием опыта работы или недостатком специализированного образования.

Интересно, что среди краткосрочных безработных преобладают семейные и детные респонденты. Возможно, это связано с тем, что они более мотивированы найти работу для обеспечения своей семьи и детей. С другой стороны, среди долгосрочных безработных преобладают холостые и бездетные люди. Возможно, у них отсутствует поддержка и мотивация для активного поиска работы.

Также было обнаружено, что респонденты без профессионального образования имеют гораздо больше шансов оказаться в долгосрочной безработице и тратить на поиски работы более 6 месяцев, чем опрошенные с профессиональным образованием. Вероятно, это связано с тем, что профессиональное образование дает людям больше возможностей и конкурентных преимуществ на рынке труда.

В целом, результаты данного исследования позволяют лучше понять причины и особенности безработицы в г. Татищево. Они подчеркивают важность образования и опыта работы для успешного трудоустройства, а также необходимость создания равных возможностей для всех категорий населения.

В современном обществе проблема безработицы становится все более актуальной. При длительных сроках безработицы хотя бы одного члена семьи ее ресурсы оскудевают и уже не позволяют в полной мере удовлетворять основные потребности семьи. Это оказывает серьезное влияние на социальное благополучие и экономическую устойчивость общества.

Наличие трудового стажа более 3 лет в значительной степени сокращает длительность поиска работы безработными. Это свидетельствует о важности опыта работы при поиске нового трудового места и его влиянии на процесс трудоустройства.

Для безработных, ищущих работу менее 6 месяцев, характерны личные причины потери работы, тогда как для безработных, находящихся в поиске работы более 6 месяцев, более свойственны проблемы с образованием. Это подчеркивает важность психологической и образовательной поддержки для людей, находящихся в длительном поиске работы.

Таким образом, было выявлено много зависимостей, причем во многих случаях были высказаны предположения о латентном влиянии других переменных. Это указывает на необходимость более глубокого исследования причин безработицы и разработки комплексных мер, направленных на поддержку людей в поиске работы и повышение их трудовой активности.

Исследование взаимосвязей между переменными на различных уровнях - это важный этап анализа данных. При обнаружении связей между тремя и более переменными, метод логлинейного анализа становится предпочтительным средством для более глубокого понимания взаимосвязей.

Одним из ключевых результатов логлинейного анализа было подтверждение статистической значимости главных эффектов различных переменных, таких как возраст, наличие детей, уровень образования, уровень потребления семьи, длительность поиска работы, трудовой стаж и причины безработицы.

Важно отметить, что помимо отдельных переменных, значимыми также оказались взаимодействия между некоторыми факторами, такие как трехфакторное взаимодействие между семейным положением, уровнем потребления и причинами безработицы, а также несколько двухфакторных взаимодействий, например, между возрастом и наличием детей или семейным положением и наличием детей.

При анализе трехфакторной модели безработицы, логлинейный подход позволяет выявить интересные закономерности. Например, исследование показало, что среди семей, где есть безработные и достаточный уровень потребления, чаще всего причиной безработицы являются проблемы с образованием. В то же время, в семьях с низким уровнем потребления безработица чаще связана с экономическим кризисом. Интересно, что личные причины безработицы в семьях не зависят от уровня потребления.