

МИНОБРНАУКИ РОССИИ

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

Кафедра социальной информатики

**КОМПЬЮТЕРНОЕ ОБЕСПЕЧЕНИЕ В СОЦИОЛОГИЧЕСКИХ  
ИССЛЕДОВАНИЯХ. ИСТОРИЯ И РАЗВИТИЕ.**

(автореферат бакалаврской работы)

студентки 4 курса 451 группы  
направления 09.03.03 - Прикладная информатика  
профиль Прикладная информатика в социологии  
Социологического факультета  
Загорновой Анны Константиновны

Научный руководитель  
Кандидат физико-математических наук

\_\_\_\_\_ М.Г.Плешаков  
подпись, дата

Зав. кафедрой  
кандидат социологических наук, доцент

\_\_\_\_\_ И.Г. Малинский  
подпись, дата

Саратов 2020

## ВВЕДЕНИЕ

**Актуальность проблемы.** В настоящее время любое социологическое исследование обращается к компьютерным технологиям практически на каждом этапе своего исследования. В процессе работы социолог применяет довольно разнообразное программное обеспечение, которое сопровождает его, начиная от формулировки темы и заканчивая оформлением результатов. Традиционные технологии не ушли в прошлое. Однако, ещё начиная с описания проблемной ситуации и постановки целей и задач, исследователь обращается не только к изученной литературе, личным наблюдениям, рассуждениям или дебатам с коллегами, но также использует тематический и словарный поиск в сетях Интернет.

Стоит отметить, что в самом процессе прикладного социологического исследования использование компьютерных технологий весьма неравномерно. Имеется в виду, что на некоторых этапах компьютер лишь вспомогательное средство, но совсем другая ситуация, например, для обработки статистических данных, собранных во время полевого этапа, или подготовки аналитического отчета. Построение распределений, вычисление корреляционных матриц, представление графиков в настоящий момент не выполняется без соответствующего пакета программ. Из-за того, что социологу в процессе исследования приходится выполнять большой спектр задач, существует огромное количество программных сред.

По официальным данным Международного статистического института, число только статистических программных продуктов приближается к тысяче. Среди них есть профессиональные статистические пакеты, предназначенные для пользователей, хорошо знакомых с методами математической статистики, и есть пакеты, с которыми могут работать специалисты, не имеющие глубокой математической подготовки.

**Степень научной разработанности проблемы.** В научной литературе широко описаны технологии и методы, которые социологи использовали в своей работе в разное время.

**Объект** – применение компьютерных технологий в социологических исследованиях.

**Предметом** исследование является историческое развитие и применение компьютерных технологий в социологическом исследовании.

**Целью** исследования является изучение эволюции ПО для прикладных социологических исследований.

Для достижения цели необходимо решить следующие задачи: – провести теоретический анализ методов социологии с момента возникновения до настоящего дня; – определить основные направления использования ЭВМ и персональных компьютеров в социологии, а так же уровни и виды технических средств, применяемых при сборе и обработке социологической информации. – сравнить существующие пакеты программ и степени их применения в исследованиях.

**Эмпирическая база ВКР:** литература и материалы сети Интернет по таким направлениям как технологии используемые для социологических исследований, развитие средств обработки и анализа данных, а так же использование компьютера на различных этапах социологического исследования.

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

## **ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ.**

**В первом разделе «Применение математических методов в социологии»** описывается совершенствование применяемых математических методов в социологии методов в социологии в связи с развитием науки.

О необходимости использования в социологии математических методов говорили многие исследователи, начиная с Конта, Кетле, Парето. Однако долгое время это сводилось к изучению разного вида вероятностных распределений и расчету простейших их параметров.

Математические методы в социологии — методы статистического анализа данных и методы математического моделирования социальных явлений и процессов.

Конкретные социологические исследования проводятся на самых различных уровнях: на уровне общей теории, специальных социологических теорий и т.д., поэтому важной задачей является изучение и выработка специфических математических теорий, средств и методов для каждого уровня в отдельности. Отмечается, что развитие теоретической и эмпирической социологии долгое время происходило отдельно друг от друга, вплоть до начала XX века, а их соотношение между собой порождало множество дискуссий.

Тема взаимодействия социологии и математики отражалась во статьях разных ученых. Учение Конта состояло из двух частей – социальной статистики, описывающей законы существования, и социальной динамики, описывающей законы и этапы изменения общества. А.А. Чупров, был активным сторонником использования достижения математической статистики для получения социального знания, в его логике метода объединялись методы статистических исследований, развитых континентальными и британскими учеными. Социолог Э. Дюркгейм в работе «Правила социологического метода» обосновал объединение подходов, описывая преимущества и ограничения каждого из них. Кетле, Парето и многие другие исследователи тоже говорили о том, что социальные факты необходимо изучать статистически.

Социологи уже тогда использовали множество методов социальной статистики, такие как методы измерения связи, регрессионный, дисперсионный и факторный анализ в их простейших вариантах. Применение приёмов статистики в социологии долгое время ограничивалось, в частности для построения статистических таблиц. Во главу угла ставились вопросы о выборе признаков для формирования таблицы, разбиения диапазонов изменения значений этих признаков на интервалы. Дело в том, что непрерывная переменная, используемая в основе математической статистики не всегда подходит под нужды социологического исследования. Переход от

непрерывности переменной к дискретности позволил перейти от математической статистики к анализу данных.

Процесс исследования состоит из непрерывной постановки и проверки разнообразных предположений. Можно выдвигать множество рабочих гипотез, но они должны быть подкреплены проверяемыми индикаторами. За время своего развития социология усовершенствовала исследовательский инструментарий и обогатилась новыми методами изучения социальной жизни. Методы исследования — это система способов сбора, обработки и анализа данных. Совершенствовались сами представления о распределениях.

Анализ данных в социологии - не просто алгоритм исследователя, а полноценная ветвь науки близкая к математической статистике появилась только к середине XX века. Порой перед социологами возникали такие задачи, когда математические теории и законы не могли корректно применяться. расширение методологии исследования было необходимо, поскольку количество данных росло огромными темпами. В своих исследованиях ученые смогли использовать не только данные переписи и опросов небольших групп, но и взяли на вооружение анкетные опросы, охватывающие огромную аудиторию. Новые возможности получения данных повлекли за собой поставку новых задач, о которых раньше исследователи не могли даже думать. Сложные измерения с использованием шкал низких типов; построение разного рода типологий объектов; “подгонка” известных методов под номинальные и порядковые шкалы. Методы математической статистики оказались недостаточно гибкими для новых условий, поэтому стали предлагаться не полностью формализованные методы.

В мировой социологии активнейший период разработки новых методов анализа период пришёлся на 20-ые – 60-ые годы XX в. В этой выделилось несколько тенденции развития.

Первое направление - измерительные процедуры в опросах, анализ особенностей социологических данных. Целью такого развития было выработка

у исследователей общего стандарта. Переосмыслялось само понятие признака, предложенное основателями социологии.

Второе направление – разработка методов анализа. Строго говоря, потребность этого направления определялась не только социологией, но и другими науками, такими как психология, медицина, геология и некоторыми другими. Важной проблемой являлось стремление формализовать методы анализа данных, но при этом учесть познаваемого объекта.

Третье направление – методы моделирования социальных явлений. Благодаря им, исследователь получает возможность построить модели с множеством исходных показателей, рассчитывать большое количество вариантов развития системы.

Основные направления развития математических методов в мировой науке 20- 60-х годов. Начало активного внедрения математики в западной социологии приходится на 20-е гг. и связано с бурным развитием опросов больших совокупностей людей. Возникла задача разработки теоретически состоятельных способов сбора и анализа информации. Но именно в эти годы свертываются эмпирические исследования в СССР, так что к периоду их возрождения в конце 50-х советским социологам пришлось поначалу осваивать зарубежный опыт. К тому времени на Западе в области развития математических методов в социологии произошли радикальные сдвиги. Волна переводов зарубежных авторов вводят в научный оборот методологический опыт европейских и американских социологических школ и в последствии позволяет начать наработку собственных методов.

В 60-ые-70-ые годы ряд методологических публикаций содержал рекомендации по применению известного опыта зарубежных коллег. Методисты формировали взгляд на систему представления математики в социологии. Появление возможности ознакомиться в этот период времени с иностранными работами привело к разработке первых советских исследований, относящиеся к эмпирическим методам социологии. Собственная научная школа была весьма

плодотворной и также осуществлялось в рамках, перечисленных трех направлений.

Например, начало современной теории выборки было положено в работе А. Г. Ковалевского. В исследованиях У. Кокрена, В.Н. Крылова к ключевым моментам относят становление использования выборочного метода в мировой и отечественной социологии, а также даются ответы на вопросы сущности и содержания понятия «репрезентативность». Можно отметить работы А.А. Гурьева, Н.К. Дружинина, Ф. Йейтса, А. Кауфмана и многих других, которые относят к «социологам первого поколения».

Разработку математических методов в социологии вели несколько институтов. Однако, стоит сказать, что в существующей плановой системе науки главным центром, на котором лежала вся ответственность, являлся Институт социологии АН СССР (ИСАИ), и в первую очередь - сотрудники отдела методики социологических исследований (отдел существовал до 1991 гг.).

Усложнение методологии исследований привело к процессу компьютеризации в социологии. Высокие научные стандарты включают статистическую репрезентативность и оперативность обработки данных. Добиться их было бы невозможно без связи социологов с математиками и программистами.

Важным моментом является то, что социологические лаборатории не могли себе позволить иметь собственную электронно-вычислительную технику. Естественно, исследователи вынуждены были обращаться в учреждения, располагавшие вычислительными центрами.

В середине 70-х, при относительно небольшом объеме до 500 анкет производился всё ещё ручной подсчёт. С увеличением объема в 500-1000 анкет рекомендовалась воспользоваться СПМ. Обращение к ЭВМ считалось целесообразным только свыше 1000 анкет.

Непростой процедурой являлась подготовка к обработке данных, она могла занимать месяцы.

Кроме предварительной обработки, существовала проблема большого разнообразия ЭВМ в СССР. В связи с этим, невозможно было создать универсальное статистическое и социологическое программное обеспечение, наподобие SPSS, Statistica и др. Институт научного атеизма вел работу по созданию программного обеспечения для обработки анкет на базе ЭВМ «Минск-22». Социологи г. Горького применяли «БЭСМ-3М» и «Минск-22», в Пензе применялся «Урал-11», в Грозном – «Минск-22», в Йошкар-Оле – «М-222».

Программа обработки данных на ЭВМ этого поколения позволяла создавать многомерные таблицы по изучаемым зависимостям, группировать данные, анализировать совокупности, но не проводить более глубокий анализ. Не применяли корреляционный и дисперсионный анализ.

Новое поколение ЭВМ в 80-ых упростило процесс обработки. Вычислительные центры всё ещё оставались очень дорогостоящим и отдаленным от рабочего места инструментом, но на рынке появились программируемые микрокалькуляторы. Они не плохо подходили для вторичных вычислений (после обработки на ЭВМ), но для них требовалась высокая квалификация как программиста и математика. Этот способ не получил широкого распространения.

Следующий этап был связан с распространением персональных компьютеров, имеющих архитектуру IBM PC и относимых к четвертому поколению ЭВМ. Применение статпакетов значительно ускорило работу с массивами данных. Стало возможно применять факторный и корреляционный анализ. Работа социолога значительно упростилась на этапах от ввода информации до получения её представления в удобной форме. Приобретение персональных компьютеров ограничивалось из-за финансов, но несмотря на это исследователь выигрывал во времени и скорости по сравнению с машинами прошлого поколения.

Распространение науки социологии сильно увеличилось уже к середине 1990-х гг. в каждом экономико-географическом регионе России было создано по одному профессионально подготовленному региональному центру,

проводящему опросы по заданию или по контрактам московских и других (включая зарубежные) центров.

В наше время разработка программы, составление анкеты, хранение, поиск, информационное обслуживание, анализ данных, представление отчётов – всё это сопровождается с помощью компьютерных технологий.

**Второй раздел работы «Сравнительный анализ пакетов программ для прикладных социологических исследований»** касается технологий компьютерного анализа социологических данных.

В зависимости от характера выполняемых работ, на разных этапах необходимо использовать различные классы программного обеспечения: статистические пакеты, текстовые процессоры, базы данных и т.д.. Универсального пакета, который мог бы сопровождать исследование на всех этапах не существует.

Автоматизация начинающегося исследования зачастую начинается ещё на моменте разработки алгоритма. Для работы с собираемой информацией и другими членами исследования чаще всего используют стандартные текстовые редакторы, либо самописные веб-инструменты.

В прошлом стадия распространения анкеты была довольно трудоемкой, вопросы писались вручную или с использованием текстовых редакторов. В наше время существуют множество различных программ опросников, позволяющих облегчить процесс разработки анкеты. Среди них можно выделить Interro-SL, iSpring QuizMaker, MS InfoPath. У этих программ схожие функции. Прибегая к ним, исследователь может получить данные опроса в формате таблицы или графика.

В последнее время всё больше исследователей прибегают к помощи онлайн-сервисов. Среди самых распространённых формы Google, сервисы Aeterna, WebAnketa, Анкетер и другие.

Для проведения телефонных исследований опросов существует система CATI (Computer Assisted Telephone Interview). Специалист при помощи телефонной гарнитур общается с респондентом, одновременно фиксирует

ответы на персональном устройстве, с которого читает вопросы. Заполненная анкета по аналогии с онлайн опросниками тут же попадает в общую базу.

Когда собраны все необходимые данные, возникает потребность их обработать и проанализировать. Активная разработка ПО в области анализа данных велась благодаря потребности в обработке больших массивов информации и в анализе всех составляющих социологического измерения: статистики случайных величин, многомерного статистического анализа, статистики временных рядов и всевозможных других. Для проведения анализа массива данных социологи используют большое число различных математических методов, позволяющих полно и всесторонне анализировать собранную информацию. В современной социологии для этой цели активно применяются компьютерные программы, как математической, так и статической обработки данных.

Стандартные статистические методы обработки данных включены в состав электронных таблиц, таких как Excel, Lotus 1-2-3, QuattroPro, и в математические пакеты общего назначения, например, Mathcad. Функционал этих программ во многом аналогичен друг друга, при помощи этих программных пакетов пользователю доступны некоторые вычисления, построение графиков, управление форматом ввода-вывода данных, сортировки данных. Однако, возможности этих программ ограничены по сравнению со специализированными пакетами, которые включают в себя множество модулей для обработки данных по последним современным стандартам.

Среди программных средств можно выделить узкоспециализированные пакеты, в первую очередь статистические — SPSS, STATISTIC A, STADIA, SAS, STATGRAPHICS, ДА-система и PULSAR. Эти пакеты имеют большой набор статистических функций: факторный анализ, регрессионный анализ, кластерный анализ, многомерный анализ, критерии согласия и т. д. Данные программные продукты содержат и средства для визуальной интерпретации полученных результатов. В работе рассмотрены различия между этими программами и проведено их сравнение по функционалу, удобству управления данными,

статистическому разнообразию (количество статистических модулей), графическими возможностями (наличие встроенного графического редактора, возможность показа отдельных элементов графика, возможности экспорта графиков).

В современном мире получило широкое применение имитационное моделирование, которое позволяет программному обеспечению создать эмуляцию требуемой системы, в виде математически точной модели, над которой можно проводить различные эксперименты. Этому помогает высокая производительность современных аппаратных средств, на базе которой можно спроектировать имитационную модель.

Многообещающей выглядит перспектива использования в социологии таких новых методов, как эволюционное моделирование (генетические алгоритмы), нейронные сети и облачные вычисления.

Особое место в работе отведено перспективам применения нейростевых технологий в исследованиях.

При увеличении объёмов выборок, повышении общей сложности объектов исследований и постоянно возрастающие требования к методикам их анализа, логично предположить, что в таких условиях потребуются гораздо более мощные и точные инструменты с функционалом уже совсем другого уровня. Для аналогии, человеческий мозг, когда воспринимает большие и сложные массивы неодинаковой информации и связывает между собой различные выборки, использует ассоциативное мышление, которое выражается в законах простой логики, самообучении на уже изученных материалах и причинно-следственных связях.

Если получится создать нейронную сеть, которая будет использовать системы, как в человеческом мозге, то такие задачи, которые не имеют простого и однотипного решения, нейросеть сможет решать намного быстрее и точнее. Также, благодаря масштабированию, нейронные сети смогут решать ещё большие по объёму выборки задачи, чем один человеческий мозг. Таким образом, у социологов появится ещё один, совершенно новый инструмент

анализа больших и сложных данных. Применение такого типа анализа подойдет для ситуаций, когда невозможно применить одинаковый шаблон для решения нескольких похожих задач. При этом, стоит учитывать, что некоторые ситуации и модели могут требовать узких и специфических знаний, поэтому анализ работы нейронных сетей здесь, будет требовать проверки точности и качества специалистом в этой области.

Если получится создать нейронную сеть, которая будет использовать системы, как в человеческом мозге, то такие задачи, которые не имеют простого и однотипного решения, нейросеть сможет решать намного быстрее и точнее. Также, благодаря масштабированию, нейронные сети смогут решать ещё большие по объёму выборки задачи, чем один человеческий мозг. Таким образом, у социологов появится ещё один, совершенно новый инструмент анализа больших и сложных данных.

Востребованность, необходимость и актуальность применения искусственных нейронных сетей для решения социальных задач несомненна. Их применение позволит вывести уровень осознания действий общества на новый уровень. Нейросеть принимает решения быстрее и на основе собственного опыта, исследователь имеет возможность наблюдать за “ходом мысли” нейросети и при необходимости вносить коррективы. Также такие методы помогут найти решения для задач, в которых аналитические способности одного человека могут оказаться недостаточными.

## **ЗАКЛЮЧЕНИЕ**

В ходе работы были рассмотрены математические методы, применяемые в социологической науке, их развитие, а также процесс компьютеризации в исследованиях.

Социологическая наука концентрированно продвигает комплексное понимание социального мира и тем самым создает этот мир в его многообразии, в различных формах и вариантах, показывает логические "связки", обладающие потенциалом устойчивого развития.

В XXI веке сообщество социологов в общей теории и в методологии эмпирического исследования должно стремиться к взаимопониманию, к объединению усилий сторонников разных теоретико-методологических парадигм и подходов к социальной реальности. Естественными видятся уточнение понятийного аппарата эмпирической социологии, места этой дисциплины и, главное, оценка ее вклада в решение современных проблем. Надо думать, о социологии как макронауке, изучающей тенденции и механизмы развития современного общества. Работы, посвященные теоретической социологии, все больше начинают оперировать информацией, которая мало или совсем не ориентируется на данные прикладных исследований, превращаясь в изложение мыслительных поисков. Что касается прикладных исследований, организаторов многих из них перестают волновать теоретико-методологические проблемы, и они ограничиваются более или менее добросовестным комментированием полученных данных.

Цель можно считать достигнутой. Была изучена эволюция применения математики в социологии в докомпьютерный этап, история постепенной компьютеризации в прикладных социологических исследованиях, а так же проведено сравнение программных средств на различных этапах прикладных социологических исследований.