

МИНОБРНАУКИ РОССИИ

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

кафедра социальной информатики

**ПРАКТИКА ПРИМЕНЕНИЯ ПРОГРАММНЫХ ПАКЕТОВ
SPSS И STATISTICA В АНАЛИЗЕ СОЦИОЛОГИЧЕСКИХ ДАННЫХ:
СРАВНИТЕЛЬНЫЙ АНАЛИЗ**

(автореферат бакалаврской работы)

студента 5 курса 531 группы
направления 09.03.03 - Прикладная информатика
профиль подготовки - Прикладная информатика в социологии
Социологического факультета
Воеводина Дмитрия Алексеевича

Научный руководитель

кандидат социологических наук, доцент _____ И.Г. Малинский

подпись, дата

Зав. кафедрой

кандидат социологических наук, доцент _____ И.Г. Малинский

подпись, дата

Саратов 2021

ВВЕДЕНИЕ

Актуальность проблемы. Статистические программы помогают людям получить полную информацию из данных, поскольку разум человека не имеет способностей к непосредственному и глубокому пониманию массивов чисел. Это стало одной из причин развития рынка компьютерных программ для статистической обработки данных. Все пакеты различаются между собой. Многие пользователи используют стандартные статистические методы обработки информации. Такие методы находятся в пакетах прикладных программ, как Excel, QuattroPro, Scilab и другие. В настоящее время существует больше сотни статистических пакетов прикладных программ.

Все пакеты различаются между собой. Нередко перед пользователями разных областей появляется вопрос выбора более подходящего им статистического пакета, который поможет найти ответы на представленные вопросы. Естественно самым наилучшим выбором для всех людей будет статистический пакет, имеющий в себе все необходимые средства для обработки данных, хорошее качество работы даже на слабых компьютерах и достаточно приемлемую цену для пользователя.

При приобретении статистического пакета пользователь должен учитывать четыре параметра:

1. количество обрабатываемых данных;
 2. сможет ли пользователь решить поставленные задачи перед ним с помощью этого пакета;
 3. требования, которые предъявляются к знаниям пользователя в области статистики;
- имеет ли он в наличии персональный компьютер, и какой мощности;

Развитие компьютерных технологий и статистического программного обеспечения помогло снять часть ограничений с использования статистических методов в анализе социологических данных. Первой группой является большой объем данных. Второй группой является сложность и объем вычислительных процессов. Третьей группой является статистическая компетентность

пользователь, которые занимаются обработкой и анализом статистической информации. Сильно увеличенная скорость компьютерных вычислительных операций, помогла современным социологам использовать сложные и большие алгоритмы. Увеличение возможностей компьютерных технологий стимулировало социологов на развитие принципиально новых аналитических методов и средств обработки статистической информации. Раньше эти методы были невозможны из-за больших вычислений. Примером таких средств является анализ не числовых данных, графическое изображение результатов статистического анализа, развитие и использование регрессионных моделей, линейно-структурных моделей.

Краткая характеристика материалов исследования.

Информация о последних версиях программ регулярно помещается в популярных компьютерных журналах и газетах типа «PC Magazine», «PC World», «BYTE», «PC Week» и др. Известны аналогичные отечественные публикации. Они представлены, в основном, в журнале «Мир ПК». Примером являются статьи: Векслер Л. С. «Статистический анализ на персональном компьютере», №2, 1992, Кулаичев А. П. «Средства и программные системы анализа данных», №10, 1994, Макаров А. А. «STADIA против STATGRAPHICS, или кто ваш лоцман в море статистических данных», №3, 1992, Кулаичев А. П. «Пакеты для анализа данных», №1, 1995, Дюк В. А., Мирошников А. И. «Эволюция STATGRAPHICS», №12, 1995 и другие. По мнению М. Митчелла, который имеет двадцать лет опыта работы со статистическими пакетами и одиннадцать лет опыта работы в качестве консультанта по статистике в Калифорнийском университете в Лос-Анжелесе, говорит, что «статистический пакет – всего лишь инструмент в руках мастера». Использование статистических методов анализа данных нужно людям в различных областях деятельности. Одной из таких областей является социология. С проблемой исследовательских задач социолог сталкивается в самом начале подготовки исследования, на этапе разработки программы социологического исследования. Обычно социолог имеет множество различных

задач в своем исследовании. Сначала он собирает данные, а затем данные надо обработать, используя методы статистики. Статистические методы в социологии значительно изменились. А.И. Орлов назвал основным событием последних 30 лет «становление научно-практической дисциплины “прикладная статистика”, посвященной разработке и применению статистических методов и моделей». Примером является двух выборочный критерий Стьюдента, который вызывает плохие последствия, если его использовать в не области применения. На опасность бездумного применения статистических пакетов обращал Налимов В.В. еще сорок лет назад. Он говорил, что некоторые социологии имеют склонность проводить расчеты, не имея понятий о применяемых методах.

Цель: Рассмотреть и представить основные возможности, недостатки и преимущества статистических пакетов SPSSи STATISTICA.

Задачи:

1. Представить теоретико-методологические основы функциональности статистического программного обеспечения SPSS.
2. Представить теоретико-методологические основы функциональности статистического программного обеспечения STATISTICA.
3. Показать возможности практического применения программных статистических пакетов SPSSи STATISTICA.
4. Провести сравнительный анализ преимуществ и недостатков использования статистических пакетов SPSSи STATISTICA.

Объект: Программы обработки статистической информации SPSSи STATISTICA.

Предмет: Преимущества и недостатки программных пакетов SPSSи STATISTICAпри анализе социологической информации на примере изучения отношения населения г. Саратова к однополым бракам.

Эмпирическую базу данной работы составили данные авторского исследования «Отношение жителей города Саратова к однополым бракам»,

статистический пакет SPSS. Исследование проведено в 2019 году. Объем выборочной совокупности составил 200 человек.

Практическая значимость работы заключается в возможности использования основных положений и выводов квалификационной работы в области прикладной информатики в социологии, среди студентов других различных областей деятельности, среди людей, работающих в области статистики.

Структура выпускной квалификационной работы. Бакалаврская работа состоит из введения, трех разделов, заключения, списка литературы и приложений.

Основное содержание работы. В разделе 1 рассматривается теоретическое и методологическое описание возможностей статистического пакета SPSS. Статистический пакет SPSS (Statistical Package for the Social Sciences – статистический пакет для социальных наук) – универсальный пакет компании SPSS Inc. Первая версия статистического пакета SPSS вышла в 1968 году. В настоящее время, по мнению компании IBM, SPSS занимает одно из ведущих мест программного обеспечения в области статистического анализа данных. Хорошо разработанный графический интерфейс помогает пользователю в простоте и эффективности использования статистического пакета. Статистический пакет SPSS является модульной системой. В новой версии статистического пакета IBM Statistics 22 установлено шестнадцать разных модулей. Состав используемых пользователем модулей будет зависеть от варианта поставки. Поставка продуктов пакета IBM SPSS имеет несколько возможностей: IBM SPSS Statistics Standard, IBM SPSS Statistics Premium, IBM SPSS Statistics for Educators (US).

Командный язык Syntax - в SPSS многие задачи пользователь может выполнить с помощью мыши и диалоговых окон, однако в программе имеется мощный командный язык. Он позволяет сохранить и автоматизировать множество повторяющихся задач.

Как и любое программного обеспечения SPSS имеет свои достоинства и недостатки.

Достоинствами SPSS является:

- развитый аппарат статистического анализа;
- универсальность (может применяться для решения широкого круга вопросов из различных предметных областей, требующих проведения статистического анализа данных);
- широкий набор статистических и графических процедур (около 50 типов диаграмм) анализа данных, а также процедур создания отчетов;
- высокая скорость вычислений, простой и удобный интерфейс;
- детальная контекстноориентированная справочная система, позволяющая неопытному пользователю с большей легкостью ориентироваться в программе;
- возможность свободного скачивания демонстрационной версии продукта на официальном сайте компании, наличие версий продукта на различных языках;
- совместимость с операционными системами Windows, Mac, Linux;
- наличие значительного количества литературы по работе с пакетом.

Недостатками SPSS является:

- высокие требования к системе компьютера (требуется 1GB оперативной памяти, 800MB памяти на жестком диске и процессор с частотой 1GHz и выше);
- высокая цена по сравнению со статистическими пакетами аналогичного уровня (стоимость покупки для индивидуального пользования сроком на год составляет около 1000 долларов).

На сегодняшний день компаний IBM всего было разработано двадцать две версии статистического пакета SPSS.

Представление данных в пакете SPSS. В программе имеется редактор данных, который состоит из двух таблиц. Первая таблица содержит собственные данные, а вторая содержит переменные. В SPSS имеется матрица,

где каждая строка имеет информацию по одной единице анализа. Например, единицей анализа является анкета, которая содержит ответы одного респондента. Каждая единица анализа имеет автоматический порядковый номер. В первом столбце матрицы располагаются номера, в других оставшихся колонках соответствующие значения переменных;

Программа в настоящее время считается одной из самых распространенных статистических пакетов, для обработки статистической информации. Опыт работы, полученный в SPSS, может использоваться в различных областях знания. Программу в основном соотносят с областью социологии, однако это несколько не уменьшает большого разнообразия аналитических возможностей статистического пакета прикладных программ.

В разделе 2 рассматривается теоретическое и методологическое описание статистического пакета STATISTICA. STATISTICA–программный пакет используемый для статистического анализа. Статистическое программное обеспечение STATISTICA было разработано компанией StatSoftInc. На сегодняшний день разработана 10-я версия пакета (STATISTICA 10). Русифицирована только 6-я версия (STATISTICA 6). Программа имеет двести пятьдесят тысяч пользователей, которые используют программу. Пакет является самым развивающимся на рынке среди других статистических пакетов. Обычный статистический пакет STATISTICA имеет три модуля. Данные модули могут быть приобретены пользователем одним пакетом, либо отдельно. Программное обеспечение STATISTICA имеет множество различных вариантов статистических пакетов. Каждый пакет пользователь может выбрать в зависимости от задач и целей поставленных перед ним. STATISTICA, как и любой статистический пакет имеет свои преимущества и недостатки в работе.

Преимуществами программы являются:

- разработан обмен информацией между программами STATISTICA и приложениями программного обеспечения Windows;
- результаты анализа в виде графиков, таблиц и текста могут быть сохранены пользователем в файле с форматом RTF.

- использование встроенного программного языка системы STATISTICA увеличивает возможности пакета;
- пользователь легко может перенести данные из MSExcel в статистический пакет STATISTICA;
- главным свойством статистических пакетов STATISTICA определяется их большая скорость при работе с большим количеством информации и мощность вычисления приложений, которые требуют постоянного построения запросов к базе данных и комплексного управления базами данных. Компанией был проведен эксперимент сравнения пакета STATISTICA с другими статистическими пакетами. Сравнения проводились на четырёх ядерной 64 битной машине с использованием шесть десять четырёх разрядной системы Windows. По результатам система STATISTICA превзошла другие пакеты с огромным отрывом. Базовые описательные статистики для тридцати переменных и 9000000 строк рассчитались программой за три секунды, другим конкурирующим статистическим пакетам потребовалось время от четырех с половиной секунд до тридцати семи секунд. Матрица корреляции, включающая пятьсот столбцов и 1000000 строк, STATISTICA вычислила примерно за пять секунд. Конкурирующие статистические программы для решения использовали около двадцати секунд. Основные функции разбиения данных, используемые для прогнозирующих моделей, выполняются в системе STATISTICA в три, четыре, раза быстрее, если сравнивать с другими статистическими пакетами. Статистическая программа имеет способность обрабатывать большие массивы данных. База данных может составлять около тридцати двух тысяч переменных и неограниченное число наблюдений.

Основными недостатками пакета STATISTICA являются:

- применение пакета требует от пользователя большой теоретической подготовки в ТВМС (теория вероятности и математической статистики).

- не имеет функции реализации некоторых важных тестов временных рядов (теста на стационарность).

Огромное количество пользователей использует язык программирования для статистической обработки данных. Разработчики STATISTICA пытаются интегрировать различные языки со STATISTICA, для удобства использования и увеличения функциональности. Одним из таких языков стал R.

В настоящее время одним из передовых статистических пакетов считается программа STATISTICA. Она основывается на самых ведущих современных технологиях, которые используются в современном мире. STATISTICA соответствует последним достижениям в области IT технологии, анализа и обработки данных. Система может применяться в различных областях: финансах, психологии, социологии, экономике, бизнесе, промышленности, маркетинге, медицине и другие.

В разделе 3 рассматривается практическое применение статистических программ STATISTICA и SPSS в анализе данных по результатам опроса на тему: «отношение жителей города Саратова к однополым бракам». Использование частотного анализа в программах SPSS и STATISTICA позволяет пользователю описать частотные таблицы и можно легко вывести различные графики. Все полученные таблицы и графики можно перенести в Microsoft Word. В программах STATISTICA и SPSS пользователь может использовать таблицы сопряженности, для определения зависимости одних переменных от других. Как и в частотном анализе таблицы сопряженности, перенесенные из программы STATISTICA в пакет прикладных программ Microsoft Word нельзя отредактировать. Для таблиц сопряженности можно выводить диаграммы, которые наглядно показывают зависимость одной переменной от другой. Таблицы, перенесенные из программы SPSS в Microsoft Word, можно легко отредактировать. Большое количество различных графических визуализаций данных с помощью пакета STATISTICA по сравнению с конкурирующими статистическими пакетами, позволяет более наглядно показать информацию о полученных результатах.

Заключение. Данная работа показывает высокую значимость использования статистических пакетов в обработке данных. Сравнивая различные статистические программы, следует учитывать, что практически все они обладают набором стандартных процедур. Алгоритмы, используемые программами, по большей части, стандартные и различий при использовании той или иной программы нет. Поэтому на первое место выходят различия в пользовательском интерфейсе, полнота охвата современных статистических методов, программируемость, наличие дополнительных модулей расширения и легкость использования полученных графиков и таблиц в других программах. При сравнении двух статистических программ STATISTICA и SPSS были выявлены достоинства и недостатки каждой программы. Статистический пакет STATISTICA имеет английский интерфейс, что плохо сказывается в работе для русских пользователей. Почти все средства обработки информации являются интерактивными, поэтому при выборе переменных из меню пользователь должен отвечать на вопросы системы, что является удобным для начинающего пользователя. Профессионального пользователя такая процедура замедляет работу, это большой минус, не позволяя эффективно повторить одну и ту же процедуру несколько раз. Пользователь должен включить сохранения результатов иначе они пропадут. Жесткая структура встроенного языка не позволяет пользователю использовать дополнительными модулями. При использовании статистического программного обеспечения STATISTICA пользователь должен иметь мощный компьютер с высокими характеристиками виртуальной памяти. Интеграция статистического пакета не может происходить с пакетами прикладных программ устаревшей версии (MSWord 2010), поэтому все графики и таблицы передаются как рисунки. Имеет широкий выбор графических инструментов в отличие от статистического пакета SPSS. Данный пакет является одним из старейших среди пакетов статистической обработки данных. Программа была разработана в основном для непрофессиональных или начинающих пользователей. SPSS используется на платформе Windows, что позволяет более эффективно обучать новых пользователей. Имеет большое

количество статистических процедур, но в сравнение с программой STATISTICA маленькое количество графиков. Система подсказки SPSS пользователям довольно рудиментарная, чем отличается в худшую сторону от системы STATISTICA. Большое количество опций доступно из меню программы, что выгодно отличает программу от большинства других конкурирующих статистических программ. Прекрасная интеграция SPSS с любыми версиями пакетов прикладных программ и русифицированный интерфейс пользователя любой версии статистического пакета помогают статисту качественно использовать данную программу. В целом SPSS может быть рекомендована пользователям, которые хотят иметь систему с простым, интуитивным интерфейсом, относительно развитой графикой и периодически использующих язык программирования для автоматизации более сложных заданий. Вообще ряд возможностей, предлагаемых SPSS, особенно в области факторного анализа, являются самыми широкими среди всех разработанных систем. Для людей, относительно ориентирующихся в статистических методиках или начинающих изучение статистики, наиболее адекватной будет использование системы STATISTICA. Развитая система подсказки и полнота представленных статистических процедур в табличном виде и графическом позволяют рекомендовать эту систему непрофессиональным пользователям, которые хотят использовать множество различных графиков для наглядности, презентаций другим людям и другое.