

Министерство образования и науки Российской Федерации

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»

кафедра социальной информатики

**ВОЗМОЖНОСТИ ОПТИМИЗАЦИИ СОЦИОЛОГИЧЕСКОГО
ИССЛЕДОВАНИЯ В ПОЛЕ ПРОГРАММНОГО
ИНСТРУМЕНТАРИЯ**

(автореферат бакалаврской работы)

студентки 4 курса 451 группы
направления 09.03.03 - Прикладная информатика
профиль прикладная информатика в социологии
Социологического факультета
Иванова Максима Андреевича

Научный руководитель

кандидат философских наук, доцент _____
подпись, дата

А.И. Завгородный

Зав. кафедрой

кандидат социологических наук, доцент _____
подпись, дата

И.Г.Малинский

Саратов 2017

ВВЕДЕНИЕ

Актуальность проблемы. Большой выбор исследовательских работ в разных областях науки в социологии, экономике, медицине, биологии, и др. основан на использовании методов математической статистики и компьютерных программ, объединенных единым понятием «анализ данных». Применение анализа данных в каждой из областей науки, имеет соответствующие особенности, связанные со структурой информации, содержанием задач и интерпретацией результатов.

На данный момент имеется многочисленное количество мощных, программных продуктов для проведения углубленного статистического анализа. Автоматизированная обработка экспериментальных данных с помощью компьютерных программ требует основательной подготовки: надо не только уметь работать на компьютере, но и знать основы статистики и прикладные программные средства статистического анализа

На сегодняшний момент на рынке статистических программных пакетов, существует более сотен и даже тысяч разнообразных программных пакетов решающих задачи статистического анализа социологических данных. И конечно же на таком рынке существует конкуренция.

Актуальность темы исследования для работы обусловлена тем, что при проведении любых исследований, включающих в себя анализ массивов данных, например социологические опросы и исследования, выбор инструментария для их обработки, диагностики и прогнозирования является очень важным, и на первый план выходят совокупность быстроты, точности и удобства в процессе анализа полученных результатов. В дипломной работе будет проведено сравнение потенциала трех программных пакетов для обработки статистических данных – IBM SPSS Statistics, Statistics и MINITAB.

Степень научной разработанности проблемы довольно не высокая как в России, так и за границей. Но с каждым годом их становится все больше, тем не менее, существуют некоторые статьи и исследований

затрагивающие данный вопрос. В большинстве случаев статьи сравнения статистических пакетов программ ограничиваются по таким категориям как: количество упоминаний в академических статьях за определенный год, по количеству книг написанных по каждому из пакетов и количеству ссылок на сайт в интернете. Такая статья была написана Робертом А. Мюнхеном.¹ Практически по всем пунктам программа IBM SPSS Statistics вырывается на Перове место, оставляя позади STATISTICA и MINITAB.

Объектом: является пакеты программ для статистического анализа данных, IBM SPSS Statistics, STATISTIC и MINITAB

Предметом: же будет, является оценка потенциала программ для реализации вычисления описательной статистики, факторного и кластерного анализа.

Целью работы, будет выяснить какая из программ способна точнее вычислить Описательную статистику, Факторный анализ и Кластерный анализ.

Задачи:

- 1) Сравнит функционал и основные системные требования программ
- 2) Определить поля статистики Факторного и Кластерного анализа
- 3) Выявить анализ проведения вычисления описательной статистики
- 4) Определения функционала в работе системного и сравнительного анализа программ

Эмпирическая база исследования. Для этого будет использована Анкета ОБЛАСТЬ в которой изучалось ситуация в Петровском районе, наиболее острых проблемах района. Для проведение данной работы была взята Анкету Область. Целью ее было выявить наиболее острые проблемные зоны в Петровском районе. В анкетирование проводили студенты СГТУ. С помощью этой анкеты были попытки выявить какие проблемы наиболее

¹ The Popularity of Data Science Software [Электронный ресурс]: - URL: <http://r4stats.com/articles/popularity/> (дата обращения: 10.05.2017)

остро влияют на ситуацию в Петровском районе. Так же были вопросы, которые были направлены на то что же могло улучшить ситуацию в Петровском районе. За какие партии будут голосовать респонденты, какие телеканалы они смотрят и откуда получают информацию из газет из телевиденья или из интернета. В опросе принимали участия 400 человек все старше 18 лет.

Структура бакалаврской работы. Работа состоит из введения, двух разделов, заключения, списка использованных источников и приложения

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

В первом разделе Сравнение функционала и системных требований Statistica, IBM SPSS Statistics, Minitab. Описательные способности программ. Рассматриваю способности программ описательной статистики по критериям Среднее. Программы Statistica и Minitab показали в первом вопросе одинаковый результат(2,4900), программа IBM SPSS Statistics результат был несколько точнее (2,5344). Во 2 вопросе Statistica и Minitab(3,2075), в IBM SPSS Statistics (3,2236). В 3 вопросе Statistica и Minitab (3,0450), SPSS(3,1392). В 4 вопросе Statistica и Minitab (2,7975), SPSS (2,8766). В 5 Statistica и Minitab (3,3625), IBM SPSS Statistics (3,4311). И в 6 вопросе Statistica и Minitab (2,8575), IBM SPSS Statistics (3,0892). Критерий Стандартная ошибка среднего статистические пакеты Statistica и Minitab вновь показали одинаковый (0,0444)результат, и IBM SPSS Statistics показала опять наиболее точный результат(0,04185). По 2 вопросу результаты Statistica и Minitab (0,0452) а у программы IBM SPSS Statistics (0,04394). По 3 вопросу результаты Statistica и Minitab (0,0476) у программы SPSS (0,04052), в 4 вопросе Statistica и Minitab (0,0430), IBM SPSS Statistics (,03695), в 5 вопросе Statistica и Minitab (0,0487), IBM SPSS Statistics (0,04321), И в 6 вопросе Statistica и Minitab (0,0553), IBM SPSS Statistics (,04051). По критерию стандартное отклонение статистические пакеты Statistica и Minitab показали одинаковый результат(0,8871), IBM SPSS Statistics (0,82960). Во 2 вопросе Statistica и Minitab (0,9036), IBM SPSS Statistics (0,87665). В 3 вопросе

Statistica и Minitab (0,9514), IBM SPSS Statistics (,79807). В 4 вопросе Statistica и Minitab (0,8593) IBM SPSS Statistics (0,72878). В 5 Statistica и Minitab (0,9739), IBM SPSS Statistics (0,85546). В 6 Statistica и Minitab(1,1069), IBM SPSS Statistics (0,77923). По критериям Дисперсия по 1 вопросу показали Statistica и Minitab (0,7869), IBM SPSS Statistics (0,688). По 2 вопросу Statistica и Minitab(0,8165), SPSS (0,769). По 3 вопросу Statistica и Minitab(0,9052), IBM SPSS Statistics (0,637). По 4 вопросу Statistica и Minitab (0,7383), IBM SPSS Statistics (0,531). По 5 Statistica и Minitab (0,9485), SPSS (0,732). По 6 вопросу Statistica и Minitab (1,2253) IBM SPSS Statistics (0,607).

Можно сделать вывод что программа IBM SPSS Statistics вычисляет практически все используемые описательные статистики общего характера: медиана, мода, среднее значение и стандартное отклонение, и т.д. В рамках описательной статистике программа IBM SPSS Statistics справилась лучше, чем остальные пакеты, даже при большем количестве вычислений программа сохранила свою точность и справилась с поставленной задачей.

Программа STATISTIC так же вычисляет практически все используемые описательные статистики общего характера: медиана, мода, среднее значение и стандартное отклонение, и т.д. Она справилась немного хуже данные которые были выведены с помощью Statistics отличаются от тех которые были получены с помощью пакета SPSS, на несколько тысячных что тоже является не плохим показателем.

Программа Minitab так же способна вычислять все те же характеристики что в SPSS и Statistica. Данные полученные с помощью пакета Minitab схожи с теми, что были получены с помощью Satistics. Так же отличаются от таблиц полученных с помощью SPSS на несколько сотых и тысячных.

Во втором разделе Характеристики пакетов программ IBM SPSS, Statistica, Minitab для проведения факторного и кластерного анализа имеющейся, эмпирической базы. рассматривал способности программ в кластерном и факторном анализе Проанализировав данные таблице с можно

сказать что результаты во всех программах IBM SPSS Statistics, MINITAB и STATISTICA совпадают и различаются не несколько сотых. Все программы разделили полученные данные на 3 фактора. Кто лучше справился с задачей сказать тяжело. Факторный анализ это не искусственный интеллект а метод в котором вы сами должны дать названия переменным.

Можно проанализировать полученные данные во всех трех таблицах. В программе SPSS большую значимость имеют такие варианты ответов как алкоголизм. Коррупция, взяточничество в органах власти, Неудовлетворительное состояние ЖКХ. Высокие цены на услуги ЖКХ, Низкий уровень доходов. Положение дел в сельском хозяйстве. Положение дел в промышленности. Преступность. Рост цен на товары и услуги. Проблемы здравоохранения. Проблемы образования. Безработица. Недостаточность мест культурного досуга. Ситуация с ветхим и аварийным жильем

В программе Statistics большую значимость имеют такие варианты ответов как алкоголизм, Коррупция, взяточничество в органах власти. Низкий уровень доходов. Проблемы здравоохранения, Проблемы образования. Безработица. Состояние дорог. Экология, состояние окружающей среды

В программе Minitab большую значимость имеют такие варианты ответов как алкоголизм. Коррупция, взяточничество в органах власти Неудовлетворительное состояние ЖКХ. Высокие цены на услуги ЖКХ. Низкий уровень доходов. Положение дел в сельском хозяйстве. Положение дел в промышленности. Преступность. Рост цен на товары и услуги. Проблемы здравоохранения. Проблемы образования. Безработица.

Можно сделать вывод что программа IBM SPSS Statistics оказалась более чувствительной к кластерному анализу и определила, максимально большее число вариантов ответов значимость которых была ниже 0.05. Исходя из этого можно считать что SPSS справилась с кластерным анализом лучше, чем остальные программы. SPSS потребовалось меньше итераций для

получена более точного количества кластеров и более точного уровня значимости.

Программа Statistica проявила себя чуть хуже и вывела меньше количество вариантов ответов значимость которых была ниже 0,05. Можно считать программа справилась с поставленной задачей. Избегая не нужных значений она вывела наиболее важные варианты ответов и наиболее нагруженные.

Программа Minitab оказалась наименее чувствительной чем программа SPSS и Statistics. Она не показала нужное значение значимости так где она действительно была.

Заключение: Программа IBM SPSS Statistics хорошо показала себя в рамках проведенного исследования она справилась со всеми поставленными задачами. Постараемся выделить те плюсы, которые получилось выявить в ходе проведенной работы

IBM SPSS Statistics имеет почти безграничные возможности в статобработке, предоставляемые данными инструментами, требуют от компьютера больших ресурсов. Для работы SPSS необходимо не менее 1 Гб оперативной памяти. Операционные системы, в которых можно запускать SPSS: Windows, MacOS и Linux. В максимальной конфигурации имеет методы, отсутствующие у ближайших конкурентов. Автоматически формирует отчет. Имеются русифицированные версии пакета. Выпущено достаточно литературы по работе в пакете. Реализован основной набор популярных статистических методов анализа. Более популярен среди людей с социально-экономическим уклоном. технологическим образованием. Позволяет параллельно обрабатывать несколько подвыборок. Хорошо реализовано прогнозирование временных рядов. Прост в освоении. Удобен при обработке результатов опроса. Имеется модуль для автоматизации процесса разработки анкеты и ввода результатов опросов.

Программа IBM SPSS Statistics вычисляет практически все используемые описательные статистики общего характера: медиана, мода,

среднее значение и стандартное отклонение, и т.д. В рамках описательной статистике программа IBM SPSS Statistics справилась лучше, чем остальные пакеты, даже при большем количестве вычислений программа сохранила свою точность и справилась с поставленной задачей

С факторным анализом программа SPSS тоже справилась. Она продемонстрировала хороший уровень распределения нагрузок.

Программа IBM SPSS Statistics оказалась более чувствительной к кластерному анализу и определила, максимально большее число вариантов ответов значимость которых была ниже 0.005. Исходя из этого можно считать что IBM SPSS Statistics справилась с кластерным анализом лучше, чем остальные программы. IBM SPSS Statistics потребовалось меньше итераций для получения более точного количества кластеров и более точного уровня значимости

Существенным недостатком является же цена только цена.

В Statistica Интегрированный табличный процессор позволяет оформлять таблицы с помощью привычных для каждого пользователя офисных приложений способов. Сохраненные таблицы, а также результаты расчетов, графики и отчеты в Statistica можно удобно расположить в одном файле, который называется «Рабочей книгой».

Есть специальные модули, в которых достаточно полно реализованы методы для статистического управления качеством. Автоматически формирует отчет, который можно редактировать в Word . Имеются русифицированные версии пакета. Выпущено достаточно литературы по работе в пакете. Реализован основной набор популярных статистических методов анализа. Более популярен среди людей с технологическим образованием. Стоит учесть что, Statistica, разработана только под Windows, что несколько уменьшает число ее пользователей.

Программа STATISTIC так же вычисляет практически все используемые описательные статистики общего характера: медиана, мода, среднее значение и стандартное отклонение, и т.д. Она справилась немного

хуже данные которые были выведены с помощью Statistics отличаются от тех которые были получены с помощью пакета SPSS, на несколько тысячных что тоже является не плохим показателем.

Программа Statistica проявила себя чуть хуже и вывела меньше количество вариантов ответов значимость которых была ниже 0,005. Можно считать программа справилась с поставленной задачей. Избегая не нужных значений она вывела наиболее важные варианты ответов и наиболее нагруженные.

Minitab 16 позиционируется разработчиками как статистический пакет для аналитической работы на современных предприятиях. Версия представлена на семи различных языках (английском, французском, немецком, японском, корейском, китайском и испанском). Надежности/выживаемости, анализ временных рядов и прогнозирование, многомерный анализ. Основной недостаток – несовместимость с операционными системами, отличными от Windows.

Программа Minitab так же способна вычислять все те же характеристики что в SPSS и Statistica. Данные полученные с помощью пакета Minitab схожи с теми, что были получены с помощью Statistics. Так же отличаются от таблиц полученных с помощью SPSS на несколько сотых и тысячных.

Программа Minitab оказалась наименее чувствительной чем программа SPSS и Statistics. Она не показала нужное значение значимости так где она действительно была.

В рамках данного исследования все задачи были подтверждены. Из этого можно сделать вывод что в рамках Описательной статистики лучше использовать программу SPSS. Для факторного анализа лучше использовать программу SPSS, а для кластерного лучше использовать программу Statistica.