

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

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ВОЗМОЖНОСТЕЙ ФАКТОРИЗАЦИИ В ПРОГРАММАХ SPSS И STATISTICA

(автореферат бакалаврской работы)

студентки 4 курса 451 группы
направления 09.03.03 - Прикладная информатика
профиль прикладная информатика в социологии
Социологического факультета
Алюшевой Раили Рафиковны

Научный руководитель

кандидат социологических наук, доцент _____ С.В.Ситникова
подпись, дата

Зав. кафедрой

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

Саратов 2017

Введение. Весь окружающий нас мир насыщен информацией – разнообразные потоки данных окружают нас, захватывая в поле своего действия. Поэтому без адекватных технологий анализа данных человек оказывается беспомощным в жесткой информационной среде. Статистика позволяет компактно описать данные, понять их структуру, провести классификацию, увидеть закономерности в хаосе случайных явлений. Продолжительное время анализ социологических данных был уделом специалистов, так как это требовало серьезной предварительной подготовки. Компьютеризация в существенной степени изменила исследовательский процесс. На сегодняшний день на рынке представлено около тысячи компьютерных программ для статистической обработки данных.

Все программы статистической обработки данных можно разделить на:

1. Профессиональные пакеты характеризуются большое количество методов анализа. Они позволяют работать со сверхбольшими объемами данных, применять узкоспециализированные методы анализа, создавать собственную систему обработки данных.

2. Популярные пакеты - имеет количество функций, достаточное для универсального применения. По-другому их называют пакеты общего назначения (например, SPSS, STATA, Statistica, S-PLUS, Stadia, STATGRAPHICS, SYSTAT, Minitab). Эти пакеты не ориентированы на специфическую предметную область и могут применяться для анализа данных из различных областей деятельности.

3. Специализированные же пакеты ориентированы на какую-либо узкую область анализа данных (например, BioStat, MESOSAUR, DATASCOPE).

В процессе проведения исследований социолог часто сталкивается с необходимостью структурировать и типологизировать данные. Тогда на помощь приходит метод факторного анализа, который позволяет не только выявить наличие скрытых взаимосвязей между признаками, но и определить круг наиболее важных переменных. Успешный результат факторного анализа

приближает исследователя к выявлению причинно-следственных связей и позволяет выделить более или менее однородные группы респондентов на основе выявленных факторов.

Все вышеизложенное позволяет сформулировать объект, предмет, цель и задачи данного исследования.

Объектом данной работы является факторный анализ, как основа моделирования в прикладных социологических исследованиях.

Предметом является преимущества и недостатки программных пакетов SPSS и Statistica при проведении факторного анализа на примере исследования мотивации выбора будущей профессии.

Целью, работы является рассмотрение и представление основных возможностей, недостатков и преимуществ факторного анализа в статистических пакетах SPSS и Statistica.

Задачи:

1. Представить теоретико-методологические основы функциональности статистических программных обеспечений SPSS и Statistica;
2. Показать специфику использования факторного анализа;
3. Провести сравнительный анализ методов факторизации в статистических пакетах SPSS и Statistica на примере

Эмпирическую базу данной работы составили данные авторского исследования «Исследование мотивации выбора профессии и будущей профессиональной деятельности», статистический пакет SPSS 21. Исследование было проведено в 2013 году. Объем выборочной совокупности составил 83 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. А также для дальнейшего изучения мотивационной модели абитуриентов/школьников.

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

Основное содержание работы. Раздел 1 рассматривает теоретико-методологические и практические возможности статистических пакетов SPSS и Statistica. На сегодняшний день существует много универсальных программ обработки и анализа статистической информации. Благодаря кругу охватываемых задач, они могут быть полезны не только студентам, но и научным работникам, экономистам, решающим задачи анализа и прогноза с использованием статистических данных.

IBM SPSS Statistics - интегрированный комплект продуктов, охватывающий весь процесс анализа, от планирования до сбора данных и анализа, составления отчетов и развертывания. Среди более чем десятка полностью интегрированных модулей вы можете найти специализированные функции. Сегодня SPSS - занимает ведущее положение среди программ, предназначенных для статистической обработки информации.

Основу программы SPSS составляет SPSS Base (базовый модуль), предоставляющий разнообразные возможности доступа к данным и управления данными. Он содержит методы анализа, которые применяются чаще всего.

SPSS представляет дружелюбный пользовательский интерфейс, который делает процесс ввода и статистического анализа доступным для начинающего и удобным для опытного пользователя. Редактор данных пакета позволяет удобно (табличным способом) вводить и корректировать входные данные. SPSS дает возможность получать множество высококачественных графиков и различных диаграмм. С помощью пакета, используя таблицы, простые меню и диалоговые окна, можно выполнять, во-первых, анализ огромных файлов данных с тысячами переменных, и во-вторых, делать все это без строчной записи команд в языке программирования.

Возможные области применения SPSS: хранение и анализ данных опросов, маркетинговых исследований и продаж, финансовый анализ и др. В

социологии пакет позволяет автоматизировать процесс создания баз данных социологической информации, их хранение и обработку.

Другим несомненным лидером на рынке статистических программ является система Statistica, представляющая собой интегрированную систему статистического анализа и обработки данных. Содержит многофункциональную систему для работы с данными, широкий набор статистических модулей, в которых собраны группы логически связанных между собой статистических процедур, специальный инструментарий для подготовки отчетов, мощную графическую систему для визуализации данных, систему обмена данными с другими Windows-приложениями.

Statistica включает большое количество методов статистического анализа (более 250 встроенных функций) объединенных следующими специализированными статистическими модулями: основные статистики и таблицы, непараметрическая статистика, дисперсионный анализ, множественная регрессия, нелинейное оценивание, анализ временных рядов и прогнозирование, кластерный анализ, факторный анализ, дискриминантный функциональный анализ, анализ длительностей жизни, каноническая корреляция, многомерное шкалирование, моделирование структурными уравнениями и другие. Несложный в освоении этот статистический пакет может быть рекомендован для социологических исследований любой сложности. Все известные на нынешний момент методы статистического анализа данных включены в систему Statistica, что позволяет сделать процесс анализа высокотехнологичным.

Основные преимущества программы:

- Удобный интерфейс и уникальные возможности настройки.
- Непревзойденная графика.
- Возможность работы с файлами неограниченного размера, непревзойденная скорость обработки данных.
- Повышенная точность вычислений.

- Широкие возможности интеграции и совместимости, простой импорт/экспорт данных, легкий доступ к базам данных.

- Автоматизация любых процедур с помощью
- Поддержка Web-технологий.

Пакет Statistica имеет более 300 тыс. зарегистрированных пользователей во всем мире и является наиболее динамично развивающимся пакетом на рынке статистического программного обеспечения.

Пакеты прикладных программ SPSS и Statistica являются полностью русифицированными. В них реализован основной набор популярных статистических методов анализа и выпущено достаточно литературы для удобства работы.

- Программа статистической обработки данных SPSS популярна среди специалистов с социально-экономическим уклоном. Позволяет параллельно обрабатывать несколько подвыборок. Хорошо реализовано прогнозирование временных рядов. Прост в освоении. Имеются специфические методы, нацеленные исключительно на маркетинговые и социологические исследования. Если нужен мощный, общепризнанный пакет с простым и понятным даже начинающим пользователям интерфейсом, то лучше воспользоваться SPSS.

- Программа статистической обработки данных Statistica более популярна среди специалистов с технологическим образованием. Очень здорово реализованы карты контроля качества и методы планирования экспериментов. Имеет очень мощный модуль для нейросетевого моделирования. Для начинающих и профессионалов, которым нужна подсказка и развитая документация на русском языке, можно рекомендовать Statistica. Это мощное приложение с профессиональными возможностями.

В раздел 2 рассматривается описание модели факторного анализа. Факторный анализ является той ветвью многомерного анализа, которая исследует внутреннюю структуру матриц ковариаций и [корреляций](#). Факторный анализ имеет сравнительно небольшую, но насыщенную и

непростую историю, начиная с полного отказа математиков признать ценность используемого психологами метода, до обязательного включения различных вариантов ФА в любую сколько-нибудь известную компьютерную статистическую программу.

Основная идея факторного анализа была сформулирована еще Ф. Гальтоном, сводится к тому, что если несколько признаков, измеренных на группе индивидов, изменяются согласованно, то можно предположить существование одной общей причины этой совместной изменчивости - фактора как скрытой (латентной). Можно выделить следующие задачи факторного анализа:

1. Отбор факторов определяющих исследуемые результативные показатели.
2. Классификация и систематизация факторов с целью обеспечения комплексного и системного подхода к исследованию их влияния на результаты хозяйственной деятельности.
3. Определение формы зависимости между факторами и результативными показателями.
4. Моделирование взаимосвязей между факторами и результативными показателями.
5. Расчет влияния факторов и оценка роли каждого из них в изменении результативного показателя.
6. Работа с факторной моделью. Методика факторного анализа.

Математическая модель ФА сходна с обычным уравнением множественной регрессии:

$$V_i = A_{i,1}F_1 + A_{i,2}F_2 + \dots + A_{i,k}F_k + U ,$$

где V_i — значение i -й переменной, которое выражено в виде линейной комбинации k общих факторов;

$A_{i,k}$ — регрессионные коэффициенты, показывающие вклад каждого из k факторов в данную переменную;

$F_{1...k}$ — факторы, общие для всех переменных;

U — фактор, характерный только для переменной V_i .

В ходе проведения исследования с использованием факторного анализа выделяют следующие различные этапы:

I. Сбор эмпирических данных. Для начала необходимо оценить изучаемый объект по ряду предложенных исследователем характеристик.

II. Следующий важнейший этап ФА — собственно факторизация матрицы корреляций (ковариаций) или выделение первоначальных (ортогональных) факторов.

III. Вращение факторной структуры и содержательная интерпретация результатов ФА.

Методы факторного анализа позволяют визуализировать структуру изучаемых явлений и процессов, а это значит определять их состояние и прогнозировать развитие. Данные полученные при факторном анализе дают основания для идентификации объекта, то есть решения задачи распознавания образа.

Методы факторного анализа обладают свойствами, весьма привлекательными для их использования в составе других статистических методов, наиболее часто в корреляционно-регрессионном анализе, кластерном анализе, многомерном шкалировании и другие.

В раздел 3 рассматривается сравнительный анализ факторизации в пакетах прикладных программ SPSS и Statistica на примере изучения профессиональных ориентаций студентов. В факторном анализе приходится сталкиваться с большим объемом вычислений. Появление электронных вычислительных машин в значительной степени способствовало преодолению этого затруднения.

Как оказалось, факторный анализ, проведенный в программах SPSS и Statistica, дал идентичные результаты. Процедура редукции данных позволила выделить четыре значимых фактора — «Денежный», «Внешний», «Имиджевый», «Личный» и сократить число переменных с 9 до 4. Переменные входящий в каждый фактор совпадают и в той и другой

программе.

В ходе прохождения процедуры факторного анализа были отмечены преимущества и недостатки работы в той или иной программе. Так в программе Statistica можно легко и быстро найти в главном меню факторный анализ во вкладке «Анализ» – «Многомерный анализ». В SPSS чтобы найти данный метод необходимо будет поискать его в огромной выпадающем списке во вкладке «Анализ» и далее – «Снижение размерности». В целом в программе SPSS главное диалоговое окно факторного анализ является более удобным по сравнению со второй программой, так как на нем находится сразу все необходимые функции для задания условий факторного анализа. В программе Statistica главное диалоговое окно устроенное по принципу ступенчатой процедуры, задав параметры в одном окне ты переходишь в следующий, в то время как в SPSS все находится в одном месте компактно.

В целом, программа SPSS имеет более яркий дизайн, насыщенный цветами, где разными оттенками различаются поля и кнопки, стрелочки и другие элементы. В программе Statistica оформление более аскетичное и скудное.

При интерпретации данных факторов, главное отличие программ заключается в том, как эти данные отражены в итоговых выводах. На заключительном этапе в программе Statistica по умолчанию открывается таблица факторных нагрузок, главный недостаток заключается том, что каждая новая таблица/график – это новое окно вывода. В свою очередь, в программе SPSS вывод всех результатов осуществляется разом в одном окне вывода

Также большим достоинством программы SPSS является таблица факторных нагрузок, она является наиболее гармоничной и сбалансированной. Программа оставляет только самые нужные значения, в таблице сразу отражено содержание анализируемых переменных, сразу становится понятно какая переменная и с какой коррелирует и можно сразу наметить основные факторы.

Особенность же Statistica заключается в том, что при выводе таблицы факторных нагрузок программа выдает значение по каждой переменной,

каждого фактора, а переменные, имеющие самую большую факторную нагрузку в каждом отдельном факторе выделяются красным цветом. Но, в свою очередь, в таблице отображены только имена переменных, не видна их содержательная сторона.

В обеих программах имеется возможность построить трех мерный график. Сравнивая два графика, наиболее оптимальным, по моему мнению, является график в программе Statistica. Благодаря имеющейся разметки типа линейная, пикам исходящих из каждого фактора и опорным линиям процедура интерпретация факторов и вычисление наиболее коррелирующих между собой переменных становится проще. Несмотря на, имеющиеся возможности редактирования графиков, в целом, по внешним параметрам, читабельности и практичности трехмерные графики в программе Statistica имеют преимущество.

Полученные результаты, то есть все содержимое окон вывода или его фрагменты, можно сохранить в обеих программах для дальнейшего использования окончательного результата либо, экспортировать в другой формат, перенести в документ Word или вывести на печать.

Заключение. Статистические программные средства является неотъемлемыми помощниками для научного анализа, исследования и моделирования в профессиональной деятельности исследователя.

Рассмотренные в данной работе пакеты прикладных программ SPSS, Statistica обладают удобным пользовательским интерфейсом, снимают трудности в применении указанных методов, имеющий сложный математический аппарат, опирающийся на линейную алгебру, теорию вероятностей и математическую статистику, а также громоздкие вычисления. В нынешний момент неотъемлемой частью любой серьезной статистической программы является факторный анализ.

Осуществленный в данной работе анализ позволил выделить следующие достоинства проведения процедуры факторизации в программе SPSS:

- Обладает простым, удобным и ярким графическим интерфейсом;

- Автоматически формирует отчет о проделанном анализе в отдельном окне вывода, включающий в себя все необходимые для дальнейшей работы таблицы и графики;
- Таблицы представления факторных нагрузок – лаконичны и содержат только необходимую информацию;
- В таблице факторных нагрузок отражается содержание анализируемых переменных, что значительно упрощает дальнейшую работу с данными.

В программе Statistica:

- Обладает широким и точным аппаратом построения графиков, предназначенных для визуализации данных. В частности обладает функцией построения трехмерных и двухмерных графиков;
- В таблице представления факторных нагрузок значения имеющие самые большие нагрузки выделяются красным цветом, что упрощает дальнейший анализ.

А также были отмечены следующие недостатки:

В программе SPSS:

- Отсутствует функция построения двухмерных графиков;
- Методы визуального представления данных во многом уступают методам представленных в Statistica.

В программе Statistica:

- Обилие кнопочек и лишних функций затрудняют проведение процедуры;
- Сложность в устройстве диалогового окна факторного анализа;
- Аскетичный и блеклый дизайн;
- В таблице факторных нагрузок отображается только имена переменных, не видна их содержательная сторона, что усложняет процесс анализа.

Если для вас важно удобство и простота задания параметров, необходимые для дальнейшего анализа, то стоит воспользоваться программой

SPSS. Так же несомненным преимуществом данной программы является представление таблиц факторных нагрузок. В таблице отражаются только самые нужные значения и показаны содержания анализируемых переменных за счет чего интерпретация данных получается более четкой и что важно быстрой.

Если же для вас визуально и схематично воспринимать информацию легче, то для своего исследования стоит использовать статистический пакет анализа данных Statistica. Он обладает более широким и точным аппаратом построения графиков, предназначенных для визуализации данных, графического вывода результатов и выбора последующих направлений анализа. Большой выбор двумерных и трехмерных научных и деловых графиков и диаграмм являются легкодоступными для пользователя.

В заключение отметим, что программы SPSS или Statistica можно рекомендовать для использования в социологических исследованиях, в частности при использовании методов многомерного статистического анализа, так как они обладают широкими возможностями и могут быть использованы без предварительной математической подготовки.