

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования

**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»**

Кафедра социальной информатики

**АНАЛИЗ СТАТИСТИЧЕСКИХ ДАННЫХ НА БАЗЕ ПРОГРАММЫ SPSS
В РЕЖИМЕ ПРОГРАММИРОВАНИЯ**

(автореферат бакалаврской работы)

Студента 4 курса 451 группы
направления 090303 – Прикладная информатика,
профиль – Прикладная информатика в социологии
социологического факультета
Абайдуллина Эльдара Ренатовича

Научный руководитель

Кандидат социологических наук, доцент

С.В. Курганова

Заведующий кафедрой

Кандидат социологических наук, доцент

И.Г. Малинский

Саратов 2019 год

ВВЕДЕНИЕ

Актуальность проблемы исследования. Появление доступных и массовых информационных технологий изменило многие стороны общественной жизни, в том числе и особенности реализации научного поиска. При различных институтах начали создавать и расширять базы данных, доступных для широкого круга исследователей. В свою очередь, внедрение персональных компьютеров и всевозможных программных продуктов самой различной специализации позволило исследователям стать более самостоятельными и при этом способными на глубокий, всесторонний и тщательный анализ имеющейся информации. Этому в немалой степени способствовала разработка программ с доступным интерфейсом, позволяющем с помощью нескольких кнопок запускать стандартные аналитические процедуры, и при этом было совершенно не обязательно владеть даже элементарными навыками программирования.

В то же время, обращение к аналитическим инструментам, «защитым» в стандартные алгоритмы действий, является фактором, ограничивающим исследователя в его работе с данными. Несколько смягчает ситуацию наличие довольно большого числа опций, позволяющих варьировать особенности проведения анализа, однако и здесь пользователь программы сталкивается с конечным числом выборов. Для решения этой проблемы разработчики многих программ оставляют желающим доступ к возможности самостоятельного программирования тех или иных аналитических действий с данными. Такая опция есть и у SPSS.

Степень научной разработанности проблемы. Востребованность программы вызвала публикацию многочисленных учебников, пособий и руководств по работе с SPSS. 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.

Интерес исследователей был сосредоточен не только на SPSS как одной из наиболее популярных программ для статистической обработки данных, но и на других программных продуктах. Появились попытки реализации комплексного соединения информационных технологий и социологии, а также теоретико-методологического осмысления проблем компьютерной поддержки социологического эмпирического исследования.

Постепенно стали появляться русскоязычные работы, предоставляющие помощь в пользовании программным языком SPSS, в первую очередь, учебные пособия и интернет-ресурсы. Однако таких ресурсов пока еще слишком мало, работа с SPSS в режиме программирования является скорее исключением, чем правилом, и оставляется на откуп лишь редким профессионалам, что делает нашу работу весьма актуальной.

Объектом данного исследования является программа статистической обработки данных SPSS; **предметом** выступает программирование как инструмент анализа данных.

Цель исследования – выявить потенциал программирования с точки зрения расширения аналитических возможностей программы SPSS. Исходя из цели были сформулированы следующие **задачи** исследования:

1. Описать программу SPSS как набор разных стратегий работы с данными;
2. Рассмотреть возможности программирования как инструмента анализа данных в SPSS на примере изучения нарушения трудовых прав работников.

В качестве **эмпирической базы** исследования были использованы данные массового социологического опроса, проведенного студентами

социологического факультета СГУ по теме «Специфика стандартной занятости жителей крупного города (на примере г. Саратова)»¹.

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

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

В первом разделе «**Особенности организации работы с данными в программе SPSS**» дается характеристика SPSS как специализированному программному продукту, рассматриваются особенности ее структуры, выявляются ее положительные и отрицательные стороны.

SPSS является одной из старейших систем статистического анализа и управления данными. За полувековой период времени программа SPSS претерпела множество нововведений и изменений.

Первые версии программы содержали всего одиннадцать статистических процедур, но SPSS уже нашла отклик у потребителя. К моменту основания Норманом Наем фирмы было проведено уже шестьдесят инсталляций.

В течение следующих пяти лет велась активная работа по адаптации программы под разные операционные системы и расширению числа статистических процедур. Уже в 1975 году была разработана 6 версия программы и проведено 600 инсталляций.

В 1983 году программа SPSS претерпела революционные изменения: был переработан командный язык; увеличилось число статистических процедур; была выпущена версия SPSS\PC+ для установки на персональные компьютеры; появилось европейское представительство компании. В результате SPSS стала самым популярным программным обеспечением для статистического анализа информации во всем мире.

¹ Данное социологическое исследование было проведено в 2018 году студентами социологического факультета СГУ. Метод сбора информации – стандартизированное анкетирование. Выборка строилась по принципу квотно-стратифицированной (объектом исследования выступали жители города Саратова в возрасте от 18 до 60 лет. Всего было опрошено 400 респондентов.

Следующей важной вехой на пути развития SPSS стал выпуск новой версии SPSS для операционной системы Windows (SPSSforWindows). С одной стороны, в новой версии программы сохранились исходные возможности для больших ЭВМ, с другой – теперь программой могли пользоваться исследователи, не специализирующиеся в области прикладного программирования, причем делать это без особых затруднений. С тех пор программа продолжала совершенствоваться: была разработана концепция Viewer (Окно просмотра), технология мобильных таблиц, улучшился дизайн таблиц и графиков. Вместе с тем, с усложнением программы стали проявляться и ее слабые стороны: высокие требования к возможностям компьютера.

В настоящее время новейшей версией программы является IBMSPSSStatistics25.0, которая предоставляет еще больше возможностей для работы с информацией. Эргономичный интерфейс программы содержит все функции, необходимые для управления данными, процедуры для статистического анализа высокого уровня сложности, а также широкие возможности для написания отчетов по итогам проведенной работы. Таким образом, в настоящее время IBMSPSSStatistics является полнофункциональной системой, направленной на решение задач в сфере бизнеса и науки с помощью анализа данных. Она позволяет глубоко и всесторонне исследовать данные, наглядно представлять итоговую информацию в виде таблиц и графиков.

SPSS выстроена как традиционная база данных. Сближает SPSS с рядовыми базами данных и особенности интерфейса, что также делает более простым первое знакомство с программой. При этом SPSS специализируется на обработке результатов опросов, т.е. имеет структурные особенности, заключающиеся в принципах формализации накапливаемого массива исходной информации, принципах статистической обработки и представления результатов информации.

Опытных пользователей SPSS привлекает работа с программным языком по следующим причинам:

1. Работа с повторяющимися операциями на обновленном / новом массиве данных;
2. Работа с учетом операций, требующих изменений;
3. Создание и применение новых пользовательских функций;
4. Возможность использовать циклы для больших массивов данных;
5. Наличие более широких возможностей для работы со сложными операциями, касающимися трансформации данных, по сравнению с возможностями меню;
6. Наличие легко читаемого журнала с комментариями.

Среди основных элементов программного языка SPSS можно выделить: команду (инструкцию); вспомогательную команду; спецификации; ключевые слова.

Для выполнения готовой программы в SPSS у пользователя существует две возможности:

- Загрузка программы в редактор синтаксиса, маркировка соответствующего текста и ее выполнение с помощью нажатия кнопки RunCurrent (Выполнить текущее задание).
- Запуск программы в операционном модуле SPSS. Выполнение программы происходит при этом не с помощью диалога с SPSS, а фоновом режиме, причём во время расчёта можно выполнять на компьютере и другие задачи.

Еще одной очень удобной опцией в SPSS, особенно для начинающих пользователей, интересующихся программированием, является возможность сочетать программный синтаксис с диалоговым режимом SPSS. Во всех статистических процедурах установки, произведенные в соответствующих диалоговых окнах, могут быть перенесены в редактор синтаксиса при помощи специального переключателя Paste (Вставить). При необходимости, можно дополнительно отредактировать этот синтаксис для того, чтобы добиться выполнения некоторых задач, недоступных в режиме работы через диалоговые окна.

Второй раздел «Примеры использования программного языка при решении аналитических задач с помощью программы SPSS» посвящен анализу нарушений трудовых прав жителей г. Саратова с помощью написания соответствующих программ.

На первом этапе было осуществлено перекодирование переменных из номинальной шкалы с тремя вариантами ответов в дихотомическую, что позволило упростить процедуру создания индексов.

На втором этапе были созданы четыре индекса, каждый из которых характеризует свою сторону трудовых отношений между работником и работодателем. Первый индекс – «Нарушения в сфере оплаты труда» - базируется на двух переменных «Полная официальная заработная плата» и «Сталкивались ли Вы с задержками заработной платы?». Различные сочетания вариантов ответов по этим вопросам позволили выделить три варианта нарушений в сфере оплаты труда: - отсутствуют (когда заработная плата «белая» и нет задержек); - частичные (когда заработная плата «белая», но есть задержки по ее выплате, либо наоборот, заработная плата «серая», но выплачивается без задержек); - абсолютные (когда заработная плата «серая» и есть задержки по ее выплате).

Анализ полученных данных позволил выявить, что подавляющее большинство респондентов (71,6%) не сталкиваются с нарушениями своих прав в оплате труда, около пятой части опрошенных сталкиваются с той или иной формой нарушений, но 6,1% подвергаются всем формам нарушения своих прав при оплате труда.

Второй индекс – «Нарушения трудового режима» - основывается на переменных «Отпуск в полном объеме» и «8-часовой рабочий день». В данном случае различные сочетания вариантов ответов по этим вопросам также позволили выделить три варианта нарушений в сфере режима труда: - отсутствуют (когда отпуск предоставляется в полном объеме и рабочий день длится 8 часов); - частичные (когда отпуск предоставляется в полном объеме, но есть нарушения по длительности рабочего дня, либо наоборот, отпуск в

сокращенном объеме, но длительность рабочего дня не превышает);- абсолютные (когда отпуск дается не в полном объеме и длительность рабочего дня превышает 8 часов).

Лишь чуть больше половины респондентов (56,6%) не сталкиваются с нарушениями своих прав при организации режима труда, четвертая часть опрошенных сталкиваются с той или иной формой нарушений, а у почти пятой части респондентов (18,4%) режим труда нарушается полностью.

Третий индекс – «Нарушения в сфере оплаты труда» - основывался на переменных «Выплата отпускных в срок», «Оплата больничного листа» и «Предоставление декретного отпуска». Различные сочетания вариантов ответов по этим вопросам позволили выделить четыре варианта нарушений в сфере социальных выплат: - отсутствуют (когда выплачиваются все виды социальных выплат: и больничные, и отпускные, и декретные деньги);- в меньшей степени (когда не выплачивается лишь один вид социальных выплат из трех);- в большей степени (когда не выплачиваются два вида социальных выплат из трех);- абсолютные (когда не выплачивается ни один из трех видов социальных выплат).

68,8% респондентов не сталкиваются с нарушениями своих прав в отношении социальных выплат со стороны работодателя, что примерно соответствует уровню защиты работников в сфере оплаты труда, при этом 10,9% сталкиваются с такими нарушениями в меньшей степени, 5,1% - в большей, и целых 15,2% недополучают все возможные социальные выплаты со стороны работодателя.

Четвертый индекс – «Индекс сверхэксплуатации работников» - отталкивались от значений переменных «Работодатель просит Вас работать сверхурочно?», «Работодатель просит Вас выходить на работу в официальный выходной?» и «Двойная оплата сверхурочной работы, в праздничные и выходные дни». Различные сочетания вариантов ответов по этим вопросам позволили выделить три варианта нарушений в сфере эксплуатации работников: - эксплуатация без финансовой компенсации (когда работник

работает сверхурочно и / или в свой выходной день и при этом его работа не оплачивается в нужном размере);- эксплуатация с финансовой компенсацией (когда работник работает сверхурочно и / или в свой выходной день, но при этом его работа оплачивается в нужном размере);- эксплуатация отсутствует (когда работник не работает сверхурочно и / или в выходные дни).

Не подвергаются чрезмерной трудовой эксплуатации лишь треть опрошенных или 34,2% респондентов. Перерабатывают и при этом получают финансовую компенсацию и того меньше – лишь 28,2% опрошенных. Наиболее же часто работники сталкиваются со сверхэксплуатацией без соответствующего финансового вознаграждения – 37,7%, что является наиболее распространенной формой нарушения трудового законодательства, зафиксированного нами.

На третьем этапе переменные-факторы были разделены на две группы: социально-демографические (пол, возраст, семейное положение, уровень образования, семейный доход) и сферы занятости (трудовой стаж, социальное положение, форма собственности организации). В ходе анализа полученного материала были выявлены 9 таблиц, описывающих статистически значимые связи между переменными. Из всех рассмотренных факторов наиболее сильное влияние на нарушения в трудовой сфере оказал фактор формы собственности организации. В данном случае была зафиксирована связь между переменными на уровне значимости 99,9% точности. Исключение составил лишь последний индикатор, связь с которым оказалась на уровне точности 95%.

ЗАКЛЮЧЕНИЕ

Статистический пакет IBM SPSS Statistics является одним из старейших среди программных продуктов, ориентированных на анализ информации из разных областей знания, и доступным как для специалистов в программировании и математике, так и для тех, кто специализируется в других областях. Несомненными достоинствами программы являются навигация с помощью мыши и диалоговых окон, сопряженность с другими распространенными программами, модульная структура и мн. др.

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 обществознании. Вместе с тем, длительная история развития SPSS, заключающаяся в расширении аналитических возможностей программы через добавление все новых и совершенствование уже существующих модулей, привела к тому, что в настоящее время программа предоставляет пользователю не просто широкие возможности, а чрезмерно широкие. С одной стороны, это делает более сложным процесс освоения программы для начинающих исследователей, с другой – уже продвинутый пользователь может выбрать для себя наиболее удобный метод и алгоритм действий.

Уже опытные пользователи часто выбирают работу с помощью синтаксиса в виде некоторой программы на языке SPSS. В качестве несомненных преимуществ данной стратегии работы с данными можно назвать работу с повторяющимися операциями на обновленном или новом массиве данных; работу с учетом операций, требующих изменений; создание и применение новых пользовательских функций; возможности использовать циклы для больших массивов данных; наличие более широких возможностей для работы со сложными операциями, касающимися трансформации данных, по сравнению с возможностями меню; наличие легко читаемого журнала с комментариями.

Для выполнения готовой программы в SPSSy пользователя существует две возможности: загрузка программы в редактор синтаксиса, маркировка соответствующего текста и ее выполнение с помощью нажатия кнопки Run; а также запуск программы в операционном модуле SPSS, когда выполнение программы происходит в фоновом режиме и во время расчёта можно выполнять на компьютере и другие задачи.

Удобной опцией в SPSS, особенно для начинающих пользователей, интересующихся программированием, является возможность сочетать программный синтаксис с диалоговым режимом SPSS. Во всех статистических процедурах установки, произведенные в соответствующих диалоговых окнах, могут быть перенесены в редактор синтаксиса при помощи специального переключателя Paste (Вставить). При необходимости, можно дополнительно отредактировать этот синтаксис для того, чтобы добиться выполнения некоторых задач, недоступных в режиме работы через диалоговые окна.

Для иллюстрации аналитических возможностей программирования в SPSS было реализовано исследование уровня и характера нарушений в трудовой сфере работников – жителей г. Саратова. Осуществленный анализ можно разбить на три этапа: трансформация шкал некоторых переменных; расчет индексов; поиск факторов, влияющих на уровень нарушений.

На первом этапе были перекодированы переменные с одинаковыми шкалами, что позволило с помощью программы получить результат гораздо быстрее, чем при помощи диалоговых окон. В итоге было создано 8 новых переменных с дихотомическими шкалами «1 – соблюдается; 2 – не соблюдается» с теми же названиями, что и у старых переменных.

Поскольку эмпирический материал, описывающий нарушения в трудовой сфере, отличается качественным разнообразием, было решено создать четыре индекса, каждый из которых характеризует свою сторону трудовых отношений между работником и работодателем: «Нарушения в сфере оплаты труда»; «Нарушения трудового режима»; «Нарушения в сфере социальных выплат»; «Индекс сверхэксплуатации работников».

Полученные в ходе расчета индексов данные позволили выявить, что реже всего респонденты сталкиваются с нарушениями в сферах оплаты труда и социальных выплат, от них страдают около 30% работников. Причем если от абсолютного нарушения прав в оплате труда страдает меньшинство опрошенных (6%), то от полного отказа в социальных выплатах, напротив, большинство – 15%. Наиболее же распространены нарушения, связанные с

организацией трудового режима. И если с нарушениями продолжительности отпуска и длительности трудового дня сталкиваются 45% опрошенных, то с отказом от оплаты работы в выходные дни и сверхурочные – уже 60% респондентов.

На третьем этапе было выявлено, что наиболее сильное влияние на нарушения в трудовой сфере оказал фактор формы собственности предприятия. В данном случае была зафиксирована связь между переменными на уровне значимости 99,9% точности. Исключение составил лишь последний индекс, описывающий эксплуатацию работников, связь с которым оказалась на уровне точности 95%.

Государственный сектор остается наиболее защищенным с точки зрения реализации трудовых прав работников, начиная с оплаты труда и заканчивая режимом труда, тогда как индивидуальная трудовая деятельность выступает полным его антиподом, где различные формы нарушений прав работников просто зашкаливают. В негосударственном секторе тоже не все обстоит благополучно, но он занимает промежуточное положение между государственным сектором и индивидуальной трудовой деятельностью. Интересным исключением из данного правила является право работника на компенсацию чрезмерных трудовых усилий, которое нарушается во всех сегментах занятости, что указывает на повсеместную тенденцию сверхэксплуатации работников в условиях современной российской занятости.