

I. Программа SPSS

- [История SPSS](#)
- [Новое в версии 10.0](#)
- [Модули SPSS](#)

1.1. История SPSS

Два студента Норман Най (Norman Nie) и Дейл Вент (Dale Bent), специализировавшиеся в области политологии в 1965 году пытались отыскать в Стенфордском университете Сан-Франциско компьютерную программу, подходящую для анализа статистической информации. Вскоре они разочаровались в своих попытках, так как имеющиеся программы оказывались более или менее непригодными, неудачно построенными или не обеспечивали наглядность представления обработанной информации. К тому же принципы пользования менялись от программы к программе.

Так, не долго думая, они решили разработать собственную программу, со своей концепцией и единым синтаксисом. В их распоряжении тогда был язык программирования FORTRAN и вычислительная машина типа IBM 7090. Уже через год была разработана первая версия программы, которая, еще через год, в 1967, могла работать на IBM 360. К этому времени к группе разработчиков присоединился Хэдлай Халл (Hadlai Hull).

Как известно из истории развития информатики, программы тогда представляли собой пакеты перфокарт. Как раз на это указывает и исходное название программы, которое авторы дали своему продукту: SPSS — это аббревиатура от Statistical Package for the Social Science.

В 1970 году работа над программой была продолжена в Чикагском университете, а Норман Най основал соответствующую фирму — к тому моменту уже было произведено шестьдесят инсталляций. Первое руководство для пользователей описывало одиннадцать различных процедур.

Спустя пять лет SPSS была уже инсталлирована шестьсот раз, причём под разными операционными системами. С самого начала версиям программы присваивали соответствующие порядковые номера. В 1975 была разработана уже шестая версия (SPSS6). До 1981 последовали версии 7, 8 и 9.

Командный язык (синтаксис) SPSS в то время был ещё не так хорошо развит, как сейчас, и естественно ориентирован на перфокарты. Поэтому так называемые управляющие карты SPSS состояли из идентификационного поля (столбцы 1-15) и из поля параметров (столбцы 16-80).

В 1983 году командный язык SPSS был полностью переработан, синтаксис стал значительно удобней. Что бы отметить этот факт, программа была переименована в SPSSX, где буква X должна была служить как номером версии в римскими числами, так и сокращением для extended (расширенный).

Так как применение перфокарт к этому моменту уже стало историей, то программа SPSS и информация, подлежащая обработке, сохранялись в отдельных файлах на винчестерах больших ЭВМ, которые тогда использовались повсеместно. Год от года постоянно увеличивалось и количество процедур.

С появлением персональных компьютеров была разработана также и PC-версия SPSS, с 1983 года появилась PC-версия SPSS\PC+. рассчитанная на MS-DOS. Позже, с момента основания в 1984 году европейского торгового представительства в Горинхеме в Нидерландах, SPSS стал широко применяться и в Европе. В настоящее время это самое распространённое программное обеспечение для статистического анализа во всём мире.

Для того, чтобы отразить возможность использования программы во всех областях, имеющих отношение к статистическому анализу, буква X вновь была удалена из названия марки, а исходной аббревиатуре присвоено новое значение: Superior Performance Software System (система программного обеспечения высшей производительности).

Если PC версия SPSS/PC+ была чуть усовершенствованной версией для больших ЭВМ, то SPSS для операционной системой Windows (SPSS for Windows) стала большим шагом вперёд. Во первых эта версия SPSS обладает всеми возможностями версии для больших ЭВМ, во вторых, за некоторыми немногочисленными исключениями, программой можно пользоваться без особых знаний в области прикладного программирования. Вызов необходимых процедур статистического анализа происходит при помощи стандартной техники, применяемой в Windows, то есть с помощью мыши и соответствующих диалоговых окон.

Первая версия SPSS для Windows имела порядковый номер 5. Затем последовали версии 6.0 и 6.1 с некоторыми нововведениями в статистической и графической областях; версия 6.1 была первой статистической программой для Windows, которая использовала 32 битную архитектуру Windows 3.1. Это можно было заметить по более высокой скорости выполнения вычислений. Усовершенствования коснулись также и интерфейса пользователя. В конце концов, была выпущена версия 6.1.3, которая уже могла работать и под Windows 95 и под NT.

В начале 1996 года появилась 7-я версия SPSS, сначала как версия 7.0, а затем 7.5. Наряду с расширением возможностей в сфере статистики, разница между этими двумя версиями заключалась в том, что в версии 7.5 как меню, так и интерфейс программы были выполнены уже не только на английском, но и на других наиболее распространенных языках.

Самым весомым отличием версии 7 по отношению к предыдущим версиям, был абсолютно новый подход к выводу информации на экран. Так, во первых, получил новые очертания так называемый Viewer (Окно просмотра), и, во вторых, более приятный внешний вид приобрели таблицы результатов расчётов (мобильные таблицы). Появившаяся технология мобильных таблиц позволяет перестраивать полученные таблицы различным способами.

Если предшественница данной версии — версия 6.1.3 могла работать как под старой Windows 3.1 так и под новой Windows 95 (NT), то SPSS версии 7 могла работать только при наличии Windows 95 (NT).

За версией 7.5 последовала версия 8.0, прогресс которой заключался в усовершенствовании графической оболочки. Возможность составления интерактивных графиков предоставляет ряд преимуществ по сравнению с традиционными графиками, которые являются стандартом для многих других пакетов.

Версия 9.0 включала в себя несколько новых статистических методов, в т.ч. многозначную логистическую регрессию, и несколько новых графических возможностей, расширяющих область интерактивных графиков.

Версия, описываемая в этой книге имеет порядковый номер 10.0. Ниже изложены важнейшие нововведения, относящиеся к этой версии.

1.2. Новое в версии 10.0

Версия 10.0 SPSS имеет два самых существенных отличия по сравнению с предыдущей версией 9.0. Они будут рассмотрены в этой книге:

Было изменено строение Редактора данных. Благодаря закладкам Данные и Переменные облегчён переход между областями ввода данных и описания переменных. Таким образом, форма описания переменных была упрощена и соответствует теперь общепринятым стандартам, применяемым в сфере табличных расчётов.

В области статистики был добавлен регрессионный анализ с категориальной целевой переменной.

1.3. Модули SPSS

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.

Традиционно вместе с SPSS Base (базовым модулем) поставляются ещё два модуля: Advanced Models (продвинутые модели) и Regression Models (регрессионные модели). Эти три модуля охватывают тот спектр методов анализа, который входил в раннюю версию программы для больших ЭВМ.

В приложении А Вы сможете найти информацию о том, какие методы анализа относятся к тому или иному модулю. Пользователь, который приобрёл все эти три модуля, может не обращать внимания на данное приложение.

Наряду с тремя упомянутыми, существует еще ряд специальных дополнительных модулей и самостоятельных программ, число которых постоянно растёт, так что пользователям следует постоянно знакомиться с информацией о нововведениях в SPSS.

В этой книге описываются базисный модуль, а также модули Regression Models, Advanced Models и Tables. Назначением последнего модуля является составление презентационных таблиц. В книге не рассматриваются логлинейные модели, анализ выживания и многомерное шкалирование, а также процедура составления презентаций.

SPSS Base (Базовый модуль)

SPSS Base входит в базовую поставку. Он включает все процедуры ввода, отбора и корректировки данных, а также большинство предлагаемых в SPSS статистических методов. Наряду с простыми методиками статистического анализа, такими как частотный анализ, расчет статистических характеристик, таблиц сопряженности, корреляций, построения графиков, этот модуль включает t-тесты и большое количество других непараметрических тестов, а также усложненные методы, такие как многомерный линейный регрессионный анализ, дискриминантный анализ, факторный анализ, кластерный анализ, дисперсионный анализ, анализ пригодности (анализ надежности) и многомерное шкалирование.

Regression Models

Данный модуль включает в себя различные методы регрессионного анализа, такие как: бинарная и мультиномиальная логистическая регрессия, нелинейная регрессия и пробит-анализ.

Advanced Models

В этот модуль входят различные методы дисперсионного анализа (многомерный, с учетом повторных измерений), общая линейная модель, анализ выживания, включая метод Каплана-Майера и регрессию Кокса, логлинейные, а также логитлоглинейные модели.

Tables

Модуль Tables служит для создания презентационных таблиц. Здесь предоставляются более широкие возможности по сравнению со упрощенными частотными таблицами и таблицами сопряженности, которые строятся в SPSS Base (базовом модуле).

Ниже в алфавитном порядке приведен список остальных модулей и программ предлагаемых для расширения SPSS.

Amos

Amos (Analysis of moment structures — анализ моментных структур) включает методы анализа с помощью линейных структурных уравнений. Целью программы является проверка сложных теоретических связей между различными признаками случайного процесса и их описание при помощи подходящих коэффициентов. Проверка проводится в форме причинного анализа и анализа траектории. При этом пользователь в графическом виде должен задать теоретическую модель, в которую вместе с данными непосредственных наблюдений могут быть включены и так называемые скрытые элементы. Программа Amos включена в состав модулей расширения SPSS, как преемник LISREL (Linear Structural RELationships — линейные структурные взаимоотношения).

AnswerTree

AnswerTree (дерево решений) включает четыре различных метода автоматизированного деления данных на отдельные группы (сегменты). Деление проводится таким образом, что частотные распределения целевой (зависимой) переменной в различных сегментах значительно различаются. Типичным примером применения данной метода является создание характерных профилей покупателей при исследовании потребительского рынка. AnswerTree является преемницей программы CHAI (Chi squared interaction Detector — детектор взаимодействий на основе хи-квадрата).

Categories

Модуль содержит различные методы для анализа категориальных данных, а именно: анализ соответствий и три различных метода оптимального шкалирования (анализ однородности, нелинейный анализ главных компонент, нелинейный канонический корреляционный анализ).

Clementine

Clementine — это программа для data mining (добычи знаний), в которой пользователю предлагаются многочисленные подходы к построению моделей, к примеру, нейронные сети, деревья решений, различные виды регрессионного анализа. Clementine представляет собой "верстак" аналитика, при помощи которого можно визуализировать процесс моделирования, перепроверять модели, сравнивать их между собой. Для удобства пользования программой имеется вспомогательная среда внедрения результатов.

Conjoint (совместный анализ)

Совместный анализ применяется при исследовании рынка для изучения потребительских свойств продуктов на предмет их привлекательности. При этом опрашиваемые респонденты по своему усмотрению должны расположить предлагаемые наборы потребительских свойств продуктов в порядке предпочтения, на основании которого можно затем вывести так называемые детализированные показатели полезности отдельных категорий каждого потребительских свойства.

Data Entry (ввод данных)

Программа Data Entry предназначена для быстрого составления вопросников, а также ввода и очистки данных. Заданные на этапе создания вопросника вопросы и категории ответов потом используются в качестве меток переменных и значений.

Exact Tests (Точные тесты)

Данный модуль служит для вычисления точного значения вероятности ошибки (величины p) в условиях ограниченности данных при проверке по критерию χ^2 (Chi-Quadrat-Test) и при непараметрических тестах. В случае необходимости для этого также может быть применён метод Монте-Карло (Monte-Carlo).

GOLDMineR

Программа содержит специальную регрессионную модель для регрессионного анализа упорядоченных зависимых и независимых переменных.

SamplePower

При помощи SamplePower может быть определён оптимальный размер выборки для большинства методов статистического анализа, реализованных в SPSS.

SPSS Missing Value Analysis

Данный модуль служит для анализа и восстановления закономерностей, которым подчиняются пропущенные значения. Он предоставляет различные варианты замены недостающих значений.

Trends

Модуль Trends содержит различные методы для анализа временных рядов, такие как: модели ARIMA, экспоненциальное сглаживание, сезонная декомпозиция и спектральный анализ.

Модули Amos, AnswerTree, Categories, Conjoint, LISREL и Trends описаны в книге этих же авторов: "SPSS. Методы исследования рынка и мнений".