МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ

«ХАРКІВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ»

Звіт з лабораторної роботи № 1

з предмету «Математична статистика»

Виконав:

Студент групи КН-36а

Кулик В.В.

Перевірив:

Голоскоков О.Є.

Харків 2018

**Лабораторная работа 1.**

ОЗНАКОМЛЕНИЕ С ПРОГРАММНЫМ ПАКЕТОМ ДЛЯ СТАТИСТИЧЕСКОГО АНАЛИЗА

1. **История компании**

Система STATISTICA разработана компанией StatSoft Inc., основанной в 1984 г. в городе Тулса (Tulsa, Oklahoma, USA). Первые программные продукты - PsyhoStat-2,3 - были ориентированы на статистический анализ социологических данных. Первый коммерческий продукт - Statistical Supplement for Lotus 1-2-3, появился в 1985 г.

С 1985 г. начался быстрый рост фирмы. StatSoft выпускает первую систему статистического анализа для компьютеров Apple Macintosh под названием StatFast и статистический пакет для IBM PC под названием STATS+. В 1986 г. начинается работа над основной линией программных продуктов фирмы - интегрированных статистических пакетов для комплексной обработки данных.

В 1991 г. выходит первая версия системы STATISTICA for DOS, которая представляет собой новое направление развития статистического программного обеспечения. В ней реализован так называемый графически-ориентированный подход к анализу данных. Этот пакет имел ряд существенных преимуществ перед другими статистическими пакетами (за счет оптимизации удалось добиться повышения скорости обработки данных более чем в 10 раз по сравнению с другими пакетами; пакет мог работать фактически с неограниченным объемом данных). В 1992 г. вышла версия STATISTICA for Macintosh, которая быстро приобрела заслуженную популярность среди пользователей.

В 1994 г. выходит версия STATISTICA for Windows. Она сразу же занимает лидирующее положение среди статистических пакетов. В результате сравнительного тестирования она получает первое место в следующих научных и компьютерных изданиях: Medical Software Reviews (2/9), INSIGHT (Май, 1994), MacWELT (немецкое издание журнала MacWORLD, Октябрь 1994), C’T Magazine (немецкий аналог журнала BYTE, Март 1994), WINDOWS Magazine (Ноябрь 1994), где STATISTICA 4.5 сравнивалась со статистическими пакетами BMDP 1.0, SPSS 6.1, Statgraphics 1.0, Systat 5.01. В 1995 г. STATISTICA была включена в число 100 лучших программных продуктов (WINDOWS Magazine, февраль 1995 г.).

В конце 1995 г. вышла версия STATISTICA 5.0, совместимая с Windows 95/NT. В этой версии было реализовано большое количество новых статистических процедур для работы с данными и графикой, а также удобный многофункциональный текстовый редактор отчетов. В отличие от предыдущих версий в STATISTICA 5.0 включен внутренний язык программирования STATISTICA Basic, который значительно расширил функциональность системы. Возможность достраивать (наращивать) систему при помощи встроенного языка программирования (из которого можно вызвать и любую внешнюю библиотеку DLL) является несомненным достоинством пакета.

В 1996-98 годах появились новые выпуски пакета – STATISTICA 5.1, 5.1-97 и 5.1-98. В систему были добавлены новые специализированные модули, учтены все новые форматы Windows и MS Office, сделаны различные дополнения и улучшения.

В 1999 г. StatSoft вышла новая, улучшенная и оптимизированная версия STATISTICA 5.5, в которой были реализованы алгоритмы построения сложных моделей (GLM - General Linear Models, GSR - General Stepwise Regression, GLZ - Generalized Linear Models и PLS - Partial Least Squares). В это же время был выпущен ряд корпоративных продуктов серии STATISTICA. Система STATISTICA 5.5. была переведена на русский язык силами российского офиса StatSoft.

2001 г. ознаменовался выходом нового поколения программных продуктов серии STATISTICA 6, совместимых с Windows NT/2000/XP. STATISTICA 6 основана на самых современных технологиях и является реализацией более чем 10 000 различных статистических процедур и методов исследования, а также более 100 различных типов графиков. В этой версии реализован язык программирования STATISTICA Visual Basic, позволяющий настраивать систему под специлизированные задачи, начиная от записи простых макросов и заканчивая разработкой сложных приложений на базе STATISTICA.

В 2004-2009 годах были выпущены 7 и 8 версии системы STATISTICA, удовлетворяющие конкретным потребностям мировой экономики. STATISTICA 8 стала одним из наиболее производительных приложений на рынке статистического программного обеспечения, была проведена большая работа по низкоуровневому ускорению работы процедур анализа, что позволило значительно увеличить общую производительность новой версии STATISTICA.

В 2009 году обновляется вся линейка программных продуктов STATISTICA. Выходит 9 версия системы, работающая как в 64-битном, так и 32-битном варианте. 64-битные приложения используют все преимущества 64-битных операционных систем, такие как большая производительность и более эффективная работа с оперативной памятью. Появилась более удобная навигация поопциям приложения – с использованием меню в виде ленты (аналогично Office 2007). Улучшены многие графические и аналитические возможности модулей системы.

В 2010 году вышла англоязычная версия системы STATISTICA 10. На сегодняшний день она является последней версией системы. STATISTICA 10 использует преимущества многоядерной архитектуры Intel, что позволяет достичь лучшей в своем классе производительности при экспоненциальном росте объемов обрабатываемых данных.

В 2011 году сотрудниками российского офиса StatSoft 10 версия полностью переведена на русский язык, включая интерфейс, документацию и справочное руководство.

В 2016 году вышла обновленная версия Statistica 13, о новых возможностях которой можно узнать из видео или материала. В данный момент для приобретения доступна версия 13.3.

1. **Общие сведения о пакете STATISTICA**

Statistica представляет собой интегрированную систему статистического анализа и обработки данных.

Она состоит из следующих основных компонент, которые объеди­нены в рамках одной системы:

* электронных таблиц для ввода и задания исходных данных, а также специальных таблиц для вывода численных результатов анализа;
* мощной графической системы для визуализации данных и результатов статис­тического анализа;
* набора специализированных статистических модулей, в которых собраны группы логически связанных между собой статистических процедур;
* специального инструментария для подготовки отчётов;
* встроенных языков программирования SCL (Statistica Command Language*)*и Statistica BASIC,которые позволяют пользователю расширить стандартные воз­можности системы.

Процедуры системы Statistica имеют высокую скорость и точность вычислений.

Гибкая и мощная технология доступа к данным позволяет эффективно работать как с таблицами данных на локальном диске, так и с удаленными хранилищами данных.

1. **Возможности пакета STATISTICA**

Стандартную комплектацию пакета STATISTICA составляют три модуля, которые могут приобретаться как единым пакетом, так и отдельно.

* Базовый пакет STATISTICA Base предоставляет обширные возможности выбора основных типов статистического анализа. Для эффективной работы базового пакета требуется как минимум 256 MB оперативной памяти. Минимальные требования к скорости процессора – 500 MHz.
* Модуль Линейные и Нелинейные Модели (Advanced Linear/NonLinearModels) содержит большой набор инструментов для моделирования и прогнозирования, включая возможности автоматического выбора модели и расширенные интерактивные средства визуализации.
* Модуль Многомерные разведочные технологии анализа (Multivariate Exploratory Techniques) служит для применения разведочного анализа различных типов данных в сочетании с интерактивными средствами визуализации.

1. **Преимущества и недостатки пакета STATISTICA**

*Преимущества пакета STATISTICA:*

* Удобный интерфейс и уникальные возможности настройки – интуитивно понятны, аналогичны интерфейсу Windows-приложений.
* Непревзойденная графика – поддержка более 10 000 графиков различных типов, удобных для редактирования и обработки с использованием операций вращения, масштабирования, изменения прозрачности и др.
* Высокая скорость обработки данных при работе с файлами большого объема – используются собственные технологии оптимизации, а также ключевые алгоритмы интеллектуального моделирования.
* Повышенная точность вычислений – обеспечивается проведение расчетов для переменных с небольшой относительной дисперсией.
* Широкие возможности интеграции и совместимости, простой импорт/экспорт данных, легкий доступ к БД – поддержка технологий связи, обработки и динамического обмена объектов и документов ActiveX, OLE, DDE.
* Автоматизация любых процедур с помощью STATISTICA Visual Basic – в дополнение к стандартным доступны более 13 000 новых функций; создание макросов, интеграция с MS Excel, языками программирования C++, Java и др.
* Поддержка Web-технологий – используется многоуровневая архитектура клиент-сервер, обеспечивающая проведение анализа и обновление содержания HTML-страниц на Web-сервере.
* Содержит полный набор классических методов анализа данных: от основных методов статистики до продвинутых методов, что позволяет гибко организовать анализ;
* Является средством построения приложений в конкретных областях;
* В комплект поставки входят специально подобранные примеры, позволяющие систематически осваивать методы анализа;
* Отвечает всем стандартам Windows, что позволяет сделать анализ высокоинтерактивным;
* Система может быть интегрирована в Интернет;
* Поддерживает web-форматы: HTML, JPEG, PNG;
* Легка в освоении, и как показывает опыт, пользователи из всех областей применения быстро осваивают систему;
* Данные системы Statistica легко конвертировать в различные базы данных и электронные таблицы;
* Поддерживает высококачественную графику, позволяющую эффектно визуализировать данные и проводить графический анализ;
* Является открытой системой: содержит языки программирования, которые позволяют расширять систему, запускать ее из других Windows-приложений, например, из Excel.

*Недостатки пакета STATISTICA:*

* Применение пакета требует высокой теоретической подготовки в ТВиМС. статистический обработка данные пакет
* Отсутствие реализации некоторых важных тестов временных рядов (в частности - тестов на стационарность).

1. **Построение графиков в системе STATISTICA**

Некий официальный интернет-магазин BOSCH по продаже различных автомобильных запчастей вел учет количества проданных аккумуляторов и общую прибыль, начиная с 2000 по 2010 года. Мы оперируем следующими данными: год, количество и прибыль.

Для создания нового проекта необходимо выбрать соответствующий пункт в меню или нажать комбинацию горячих клавиш Ctrl + N. После откроется меню настроек нового проекта (рисунок. 1.1)

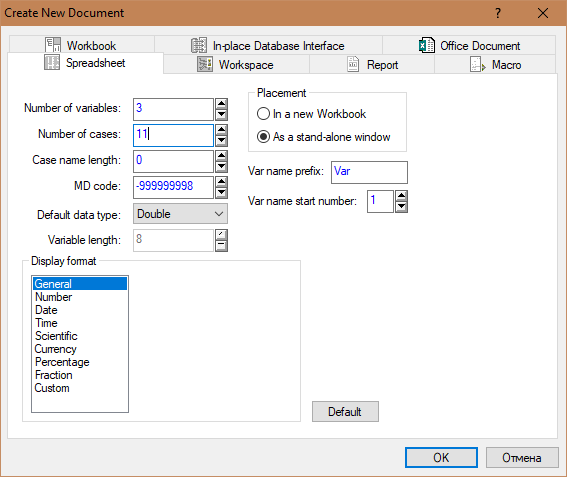


Рисунок 1.1 – Меню настроек нового проекта

Создадим новую таблицу с данными, поэтому остановимся на Spreadsheet. В которой создадим 3 переменных, и 11 строк для данных. Вид созданной таблице приведены на рисунке 1.2

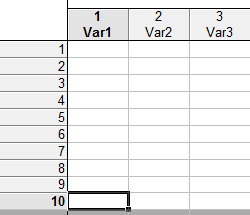


Рисунок 1.2 – Созданная таблица

Изменяем названия переменных. Для этого дважды кликаем по Var 1 и вводим желаемое название в поле Name (рисунок 1.3)

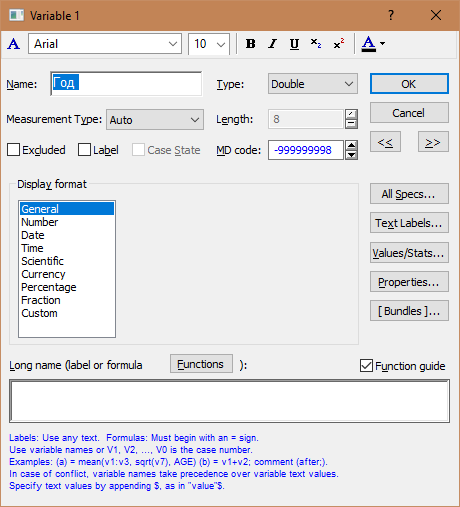


Рисунок 1.3 – Меню настроек переменной

Важно отметить, что программа предоставляет возможность сопоставления значений, и кодов, с помощью встроенной функции Text Labels. Вид таблицы с изменениями приведен на рисунке 1.4.



Рисунок 1.4 – Итоговый вид таблицы

Так как сильной стороной приложения Statistica есть графика, построим гистограммы распределения по году, количеству, и прибыли. Также приведем пример круговой диаграммы.

Для построения графиков переходим во вкладку «Graphs». Затем кликаем по кнопке 2D и выбираем график. Параметры «Graphs type:» и «Fit» оставить без изменений. Для определения переменных, по которым будет построен график нажать кнопку «Variables:». В появившемся окне (рисунок 2.2) выбрать переменную «sinus» и нажать кнопку «OK».

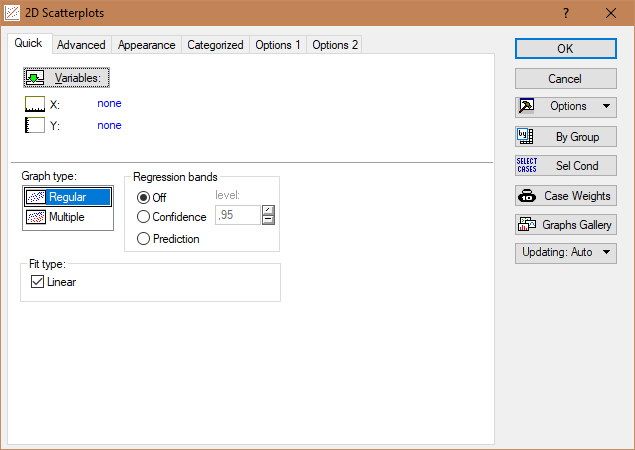


Рисунок 2.1 - Окно задания параметров построения графика

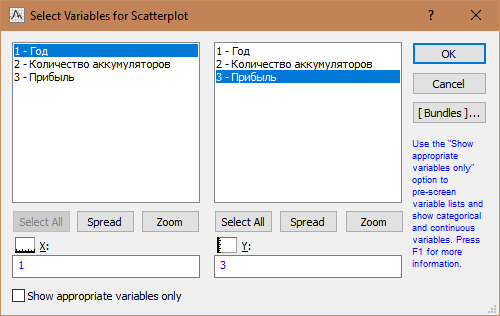


Рисунок 2.2 - Выбор переменной для построения графика

По окончании задания параметров построения графика нажать кнопку «OK». В результате система отобразит графическое распределение исходных данных (рисунок 2.3).

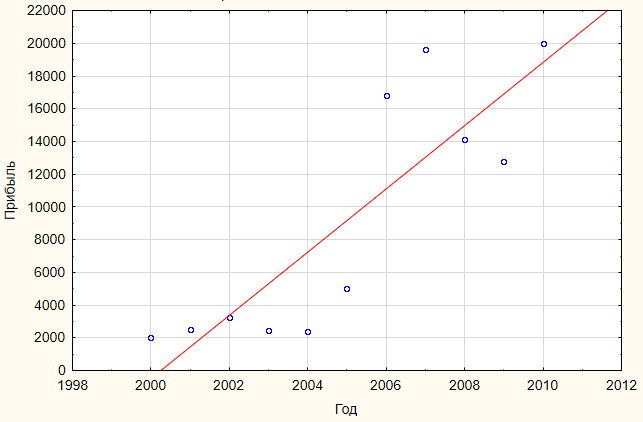


Рисунок 2.3 - Scatterplot

На рисунке 2.3 мы видим график распределения прибыли за каждый год. В 2010 году наглядно видно, что прибыль составила 20000. Это была максимальная прибыль интернет-магазина.

Сделаем еще несколько других графиков.

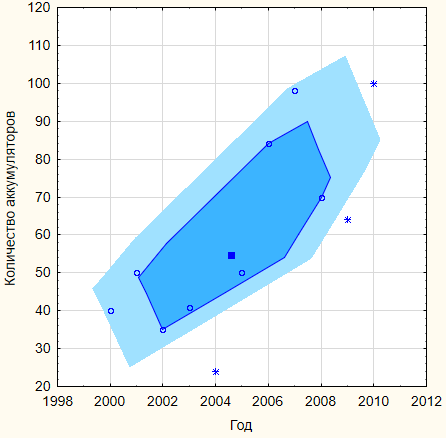


Рисунок 2.4 – Big plot

На рисунке 2.4 мы видим график распределения количества аккумуляторов по годам. Наглядно видно, что в 2002 году было продано минимальное количество аккумуляторов, а уже в 2010 интернет-магазин продал целых 100 аккумуляторов BOSCH.

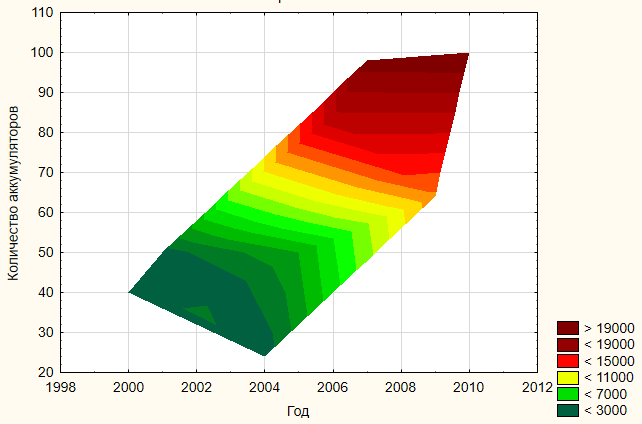


Рисунок 2.5 – 3D Water Plot

На рисунке 2.5 мы видим уже объемный график. Сразу бросается в глаза разноцветная 3D фигура. Цвета обозначают прибыль. Справа внизу мы видим индикацию прибыли по цвету. При этом мы также можем наблюдать количество проданных аккумуляторов за конкретный год.

Интернет-магазин вел более подробный учет продаж каждого товара. В нашем примере в роли товара выступает автомобильный аккумулятор. Аккумулятор имеет различные характеристики (полярность, тип электролита и т.д). В таблице на рисунке 3.1 мы видим аккумуляторы, проданные в 2004 году. Всего аккумуляторов было продано 24.



Рисунок 3.1 – Проданные аккумуляторы в 2004 году

Теперь построим две круговых диаграммы для следующих столбцов: *Полярность*, *Тип электролита*.



Рисунок 3.2 – Pie Chart для полярности

Исходя из диаграммы на рисунке 3.2 мы видим, что аккумуляторы с универсальной полярностью не так были востребованы в 2004 году. Чаще покупали с прямой и обратной полярностью.



Рисунок 3.3 – Pie Chart для типа электролита

Взглянув на эту диаграмму, мы сразу видим, что аккумуляторы с AGM наиболее востребованы среди покупателей интернет-магазина BOSCH.

**Висновок:** В результаті виконаної лабораторної роботи ми ознайомилися з программним пакетом STATISTICA, його перевагами і недоліками. Також навчилися створювати таблиці, заповнювати їх і будувати різноманітні графіки по значенням з цих таблиць.