**Зарубежные статистические пакеты: описание, возможности, недостатки, перспективы развития**

Реферат выполнила Суслова А.И., группа № 427

Кафедра Экономической Информатики и Автоматизации Управления.

Ростовский Государственный Экономический университет (РИНХ)

г. Ростов-на-Дону 2004г.

**Введение**

Для успешного и прибыльного функционирования в условиях рынка и жесткой конкуренции западные фирмы, банки, страховые компании и т.д. нуждаются в тщательном анализе имеющейся информации о создании продукции, её сбыте, эксплуатации, а также анализ информации о конкурентах и т. п. и получении из нее надежных и обоснованных выводов. Поэтому потребность в средствах статистического анализа данных на Западе очень велика.

Именно этот факт и послужил причиной для развития рынка статистических программ, на котором предлагаются более тысячи разнообразных программ. Различные по объему и качеству реализованной статистики, области возможного применения, пользовательскому интерфейсу, цене, требованиям к оборудованию и т.п., они отражают многообразие потребностей обработки данных в различных областях человеческой деятельности.

Компьютерные системы для анализа данных - пакеты статистических программ - считаются наукоемкими программными продуктами, но, пожалуй, наиболее широко применяются в практической и исследовательской работе в самых разнообразных областях.

На сегодняшний день Международный рынок насчитывает около тысячи (или даже более) пакетов, решающих задачи статистического анализа данных в среде операционных систем Windows, DOS, OS/2.

В настоящее время, по перечисленным выше причинам, число статистических пакетов, получивших распространение в России, достаточно велико и спрос на них продолжает возрастать.

Из зарубежных пакетов это STATGRAPHICS, SPSS, SYSTAT, BMDP,SAS, CSS, STATISTICA, S-plus, и т.п.

Из отечественных можно назвать такие пакеты, как STADIA, ЭВРИСТА, МИЗОЗАВР, ОЛИМП:Стат-Эксперт, Статистик-Консультант, САНИ, КЛАСС-МАСТЕР и т.д.

**Виды статистических пакетов**

Как уже отмечалось, международный рынок насчитывает > 1000 пакетов, решающих задачи статистического анализа данных в среде операционных систем Windows, DOS, OS/2. Как ориентироваться в этом многообразии, если даже справочники, содержащие только краткие описания пакетов, представляют из себя объёмные тома?

Большую часть статистических пакетов можно разбить на две группы — это статистические пакеты общего назначения и специализированные программные продукты.

Универсальные пакеты – отсутствие прямой ориентации на специфическую предметную область, предлагают широкий диапазон статистических методов. В них отсутствует ориентация на конкретную предметную область. Они обладают дружественным интерфейсом. Из зарубежных универсальных пакетов наиболее распространены BAS, SPSS, Systat, Minilab, Statgraphics, STATISTICA.

Специализированные пакеты - как правило, реализуют несколько статистических методов или методы, применяемые в конкретной предметной области. Чаще всего это системы, ориентированные на анализ временных рядов, корреляционно-регресионный, факторный или кластерный анализ. Применять такие пакеты целесообразно в тех случаях, когда требуется систематически решать задачи из этой области, для которой предназначен специализированный пакет, а возможностей пакетов общего назначения недостаточно. Из российских пакетов более известны STADIA, Олимп, Класс-Мастер, КВАЗАР, Статистик-Консультант; американские пакеты – ODA, WinSTAT, Statit и т.д.

Требования к СПП.

Статистический пакет в идеале должен удовлетворять определенным требованиям:

 модульность;

 ассистирование при выборе способа обработки данных;

 использование простого проблемно-ориентированного языка для формулировки задания пользователя;

 автоматическая организация процесса обработки данных и связей с модулями пакета;

 ведение банка данных пользователя и составление отчета о результатах проделанного анализа;

 диалоговый режим работы пользователя с пакетом;

 совместимость с другим программным обеспечением.

Следует заметить, что развитие СПП обычно идет поэтапно, на каждом из них создается вариант пакета, все в большей степени удовлетворяющий перечисленным выше требованиям. При этом если создание есть результат разработки, то на каждом этапе пакет, с одной стороны, должен представлять собой готовую к использованию программную продукцию, а с другой - входить составной частью в более поздние стадии развития пакета.

**Анализа однотипных зарубежных пакетов**

Как выбрать подходящий статистический программный продукт (СПП)? Какими критериями следует руководствоваться при сравнении различных СПП? Можно ли оценить степень соответствия цены СПП его потребительским свойствам? Ниже излагается методика сравнительного, позволяющая такой продукт правильно выбрать и в дальнейшем эффективно использовать.

Существующая классификация статистических пакетов предлагает делить их на четыре группы:

интегрированные методо-ориентированные пакеты общего назначения;

специализированные методо-ориентированные пакеты;

предметно- (или проблемно-) ориентированные пакеты;

обучающие программы.

Рассмотрим пакеты первых двух групп поскольку именно они "обслуживают" весьма широкий спектр прикладных задач.

Список пакетов, составивших выборку для анализа, представлен в табл. 1.

Таблица 1.

Общие сведения об СПП.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Статистическая система  | Версия  | Цена  | Минимальные системные требования  | Продавец  | Категория пользователя  |
| ОС  | Процессор  | Емкость НЖМД  | Объем ОЗУ  |
| SAS  | 6.11  | н/д  | W  | 386  | 65\*  | 8\*\*  | SAS Institute  | В  |
| SAS  | 6.07  | 850  | D  | 386  | 45  | 4  | SAS Institute  | В  |
| Statgraphics+  | 1.0  | 1048  | W  | 386  | 14.5  | 4  | Manugistics  | Ср-Н  |
| Statgraphics  | 1.0  | 995  | D  | 386  | 8,5  | 4  | Manugistics  | Ср  |
| Statgraphics  | 7.0  | 995  | D  | 286  | 6.1  | 1  | Manugistics  | Ср  |
| Systat  | 6.0  | 995  | W  | 386  | 8  | 4  | SPSS  | В  |
| Systat  | 6.0  | 995  | D  | 386  | 8  | 4  | SPSS  | В  |
| SPSS  | 7.0, 7.5  | 980  | W  | 486  | 65\*\*\*  | 8  | SPSS  | В  |
| Statistica  | 5.1  | 995  | W  | 386  | 13  | 4  | StatSoft  | В-Ср  |
| Statistica  | 5.1  | 795  | D  | 386  | 8  | 1  | StatSoft  | В-Ср  |
| S-Plus  | н/д  | 1450  | W  | н/д  | н/д  | н/д  | StatSci  | В  |
| S-Plus  | н/д  | 1195  | D  | н/д  | н/д  | н/д  | StatSci  | В  |

Примечания. Цены приведены на момент публикации источников. Цены на SPSS и SAS указаны для базовых модулей на российском рынке. Кроме того, SAS требует ежегодной оплаты лицензии. Достаточно полная конфигурация SPSS (модуль Base + комплект из семи модулей) стоит 4290 долл. Каждый из дополнительных модулей SAS и SPSS стоит от 350 до 750 долл.

Категории пользователей пакета: В - статистик-профессионал, Ср - "есть базовые статистические знания", Н - "отсутствие базового уровня", В-Ср - промежуточный уровень.

Операционные системы: W - Windows, D - DOS.

Для всех универсальных пакетов разработчики настоятельно рекомендуют использование сопроцессоров или соответствующих микропроцессоров.

Для тех же пакетов рекомендуется использование 16-Мбайт оперативной памяти.

\* Для модулей STAT, GRAPH .

\*\* Включая файл "подкачки" на диске.

\*\*\* Дополнительно рекомендуется файл "подкачки" на диске объемом 15 Мбайт.

Возможности западных статистических пакетов

Ввиду того, что в настоящее время стали очень популярны статистические методы обработки данных, соответствующие средства стали включаться в табличные процессоры общего назначения (например, в Еxcеl, Lоtus 1-2-3 и т.д.), а также в некоторые базы данных.

Западные статистические пакеты (SРSS,SAS,BMDР и т.д.) имеют следующие возможности:

Позволяют обрабатывать гигантские объемы данных.

Включают средства описания задач на встроенном языке.

Дают возможность построения на их основе систем обработки информации для целых предприятий.

Позволяют проводить узкоспециальные методы анализа.

Выбор статистического пакета для анализа данных зависит от характера решаемых задач, объема и специфики обрабатываемых данных, квалификации пользователей, имеющегося оборудования и т.д.

Возможности пакета SAS.

Для пользователей, имеющих дело со сверхбольшими объемами данных или узкоспециальными методами анализа, пока нет альтернативы использованию профессиональных западных пакетов. Среди пакетов такого рода наибольшими возможностями обладает пакет SAS.

Таблица: возможности пакета SAS:

|  |  |
| --- | --- |
| Средство | Возможности |
| Программирование на 4GL (программном языке SAS System) | Вызов различных статистических процедур SAS осуществляется из программ, написанных на программном языке SAS. Т.к этот программный язык имеет необходимые средства для управления данными, мощные макросредства, поэтому с его помощью могут быть реализованы сложные информационные технологии.  |
| Программирование с помощью ASSIST | Предназначен для пользователя, не имеющего специальной подготовки в области статистики и не программирующего на входном языке. Предоставляет не полный, но достаточный набор возможностей анализа данных. |
| Консультативная помощь при анализе данных | В SAS System имеется возможность получать консультативную помощь в выборе методов анализа и в интерпретации его результатов, а также рекомендации по дальнейшей работе с исходными данными.  |
| Быстрая разработка приложений - аналитические возможности | В SAS System имеется средство быстрой разработки приложений на основе объектной технологии - т.н. SAS/EIS. Среди возможностей объектов, предлагаемых для этих приложений, имеются и аналитические возможности: описательная статистика, расчеты обобщающих показателей, прогноз временных рядов, анализ "что-если". Все получаемые аналитические результаты обязательно наглядно представляются графически.  |
| Интерактивный матричный язык | SAS System располагает, в дополнение к другим языковым средствам, специальный интерактивный матричный язык, который дает возможность осуществлять различные математические расчеты, в том числе и аналитико-статистические расчеты.  |
| Приложение ANALYST | Это приложение, ориентированное на пользователя без специальной статистической подготовки, позволяет быстро осуществить статистический анализ данных, табличное и графическое представление результатов |
| Аналитические методы в средствах разведки данных | В SAS System разработано средство разведки данных (SAS Data Mining Solution), дающее пользователю возможность осуществить весь цикл работы с исходными данными, имеющими большие объемы и невыясненную статистическую структуру. |

Объемы обрабатываемых данных в пакете SРSS ограничиваются только величиной памяти вашего компьютера. Этот пакет также весьма удобен для работы с данными сложной структуры, когда необходимо делать их всевозможные срезы, как, например, в комплексном социологическом исследовании.

**Недостатки западных статистических пакетов**

Следует отметить, что продвижение западных продуктов в российской аудитории наталкивается на ряд ограничений в связи с неадекватностью культурно-исторической ситуации.

Большинство из таких статистических пакетов имеют следующие недостатки:

Требуют наличие профессиональных навыков и высокой квалификации, широкого первоначального статистического образования, доступной литературы и консультационных служб. Поэтому они содержат мало экранных подсказок и требуют внимательного изучения документации на английском языке.

Представляют сложности для быстрого освоения и использования.

Отсутствие подробной документации, доступной для начинающих и информативной для специалистов-статистиков (исключение SPSS).

Требуют больших финансовых затрат, так как немаловажное значение имеет цена пакета. Профессиональные западные статистические пакеты (SРSS,SAS,BMDР и т.д.) обычно стоят от 1 до 10 тыс. долларов и более. Мало кто готов заплатить такие деньги.

**Описание универсальных статистических пакетов**

Пакет SРSS – универсальный статистический пакет фирмы SРSS Inc. (СА). Версии системы SРSS существуют для различных платформ – Windows, ОS/2, Macintоsh, UNIX, и др. Все они совместимы между собой по принципу работы, командному языку и форматам фалов. Версия SРSS для Windows продолжает сохранять позиции лидирующего статистического пакета в мире. В России сейчас распространяется 11 версия SРSS для Windows на англ.языке и версия 10.1 на русском языке. Большая часть обширной документации переведена на русский язык.

Пакет SРSS предлагает удобные возможности управления данными, широкий спектр статистических функций, интегрированных графиков и отчетов. SРSS является модульной программой. Базовая система SРSS Ваsе предоставляет пользователям возможности для преобразования данных, функции работы с файлами, описательную статистику, дисперсионный анализ, корреляцию, линейную регрессию, средства построения графиков и подготовки отчетов и др. Дополнительные модули пакета включают: анализ и конструирование таблиц, анализ временных рядов, анализ категорий, методы углубленного и расширенного статистического анализа и др. Документация к системе SРSS признана лучшей для систем подобного типа и может использоваться в качестве доступного учебника по прикладной статистике.

Пакет STATISTICA – универсальный статистический пакет фирмы StatSоft, Inc. Пакет был создан в начале 1990-х годов сразу для среды Windows. В пакете нашли отражение многие последние достижения теоретической и прикладной статистики. В настоящее время в России распространяется версия 5.5 этого пакета на русском языке с русской документацией и версия 6.0 на английском языке.

У пакета есть специальная версия для обучения основам статистических методов – Studеnt Еditiоn оf STATISTICA. Эта версия позволяет анализировать файлы данных, включающих не более 400 наблюдений, и представляет собой урезанный вариант пакета.

Основная версия пакета может дополнительно комплектоваться специализированными модулями: Роwеr Analysis (планирование статистических исследовании), Nеural Nеtwоrks (нейросетевой анализ) и др.

Помимо общих статистических и графических средств в системе имеются специализированные модули, например, для проведения социологических или биомедицинских исследований, решения технических и, что очень важно, промышленных задач: карты контроля качества, анализ процессов и планирование эксперимента. Работа со всеми модулями происходит в рамках единого программного пакета, для которого можно выбирать один из нескольких предложенных интерфейсов пользователя.

С помощью реализованных в системе STATISTICA мощных языков программирования, снабженных специальными средствами поддержки, легко создаются законченные пользовательские решения и встраиваются в различные другие приложения или вычислительные среды. Очень трудно представить себе, что кому-то могут понадобиться абсолютно все статистические процедуры и методы визуализации, имеющиеся в системе STATISTICA, однако опыт многих людей, успешно работающих с пакетом, свидетельствует о том, что возможность доступа к новым, нетрадиционным методам анализа данных помогает находить новые способы проверки рабочих гипотез и исследования данных.

Пакет STATGRAРHICS – универсальный, многопрофильный пакет с хорошо методически продуманным меню-ориентированным интерфейсом пользователя. DОS версии этого пакета, по-видимому, были самыми распространенными в России из западных статистических пакетов.

В настоящее время распространяется версия STATGRAРHICS Рlus Vеrsiоn 5 для среды Windows. Пакет не русифицирован и его документация не переведена на русский язык. Однако работа пакета хорошо разобрана на примерах.

**Лидеры статистических программных продуктов**

Среди математического программного обеспечения, несомненно, есть свои лидеры. К ним смело можно отнести такие программные продукты как Statistica, Mathematica, MatLab и др.

Но в настоящее время на 1вом месте стоит STATISTICA. На этот раз получает высочайшую оценку, в обзорной статье Кевина Аэрна, опубликованной в широко известном и признанном в научных кругах журнале SCIENCE. Автор называет систему STATISTICA лучшим программным пакетом среди подобных продуктов, представленных на рынке программного обеспечения, и образно характеризует ее как "карнавал статистических методов". Он пишет: "STATISTICA сохраняет ведущее положение среди стандартизованных статистических программных продуктов", и далее: "Возникает вполне резонный вопрос, стоит ли платить почти тысячу долларов за функции системы, которые доступны в менее удобных, но зато менее дорогих программах... Ответ на него очевиден, если принять во внимание высокое качество и эффективность программы". Далее автор делает еще один комплимент системе STATISTICA: "В методических разделах статей многие авторы дают ссылки на систему STATISTICA, поскольку она удовлетворяет самым высоким стандартам, обеспечивает высокую точность вычислений и содержит тщательно продуманную организацию вывода".

**Перспективы развития**

В ближайшем будущем фирма StatSoft планирует выпустить ряд новых программных продуктов. В первую очередь это средства разработки, ориентированные на пользователей, которые разрабатывают собственные процедуры и методы обработки данных. Новые программы будут включать в себя объектно-ориентированные средства для макропрограммирования графических, математических и статистических процедур.

Рынок программных продуктов продолжает расти. Новые программные пакеты разрабатываются на основе современных компьютерных технологий, которые стремительно развиваются. Возможности программ расширяются, позволяя человеку использовать их не только как подручное средство, облегчающее работу, но и как полноценного помощника, способного решать сложные математические проблемы.

**Вывод**

В данном реферате я постаралась рассмотреть основные возможности и недостатки зарубежных статистических пакетов, а также провела анализа однотипных зарубежных пакетов, который может помочь при их покупке сделать правильный выбор и в дальнейшем эффективно их использовать.

В настоящее время российский рынок продолжает стремительно расти. Данный реферат рассчитан на будущих и юных экономистов, статистиков и актуариев, доля которых на рынке труда сейчас очень велика, и которым просто необходимо ориентироваться в таких вопросах и обладать как можно большей информацией о программах, способных им помочь и облегчить их труд.

**Список литературы**

**Книги, учебники:**

Айвазян С.А. «Программное обеспечение персональных ЭВМ по статистическому анализу данных//Компьютер и экономика: экономические проблемы компьютеризации общества». М.: Наука, 1998. С.91-107.

С.А. Айвазян, В.С. Степанов, «Инструменты статистического анализа данных», 2000г.

Кузнецов С.Е., Халилеев В.А. Обзор специализированных статистических пакетов по анализу временных рядов: Науч. отчет. М., СтатДиалог, 1993.

**Периодика:**

Журнал «Мир ПК» · #8/97

**Интернет:**

http://www.statsoft.ru/home/news/news008.htm

http://www.exponenta.ru

http://statsoft.msu.ru

http://www.spss.ru

http://www.statgraphics.com