

Программные средства статистического анализа

Артюхов А.А., Жаднов В.В., Чекуров С.И.
jadnov@mitme.ru

Внедрение систем менеджмента качества (СМК) на российских предприятиях-производителях РЭС и сертификация этих систем на соответствие требованиям стандартов ISO серии 9000 обуславливают все более широкое применение статистических методов, которые являются одним из элементов СМК. Каждое предприятие должно определить потребности в статистических методах, применяемых при разработке, управлении и проверке возможностей технологических процессов и характеристик надежности РЭС.

Очевидно, что автоматизация управления качеством требует применения программных средств (ПС), реализующих статистические методы. На сегодняшний день эти ПС широко представлены на рынке программного обеспечения (ПО). Чтобы ориентироваться в этом секторе рынка ПО приведем краткую характеристику наиболее распространенных ПС.

Универсальные статистические пакеты общего назначения (SAS, SPSS, SYSTAT, STATISTICA, MINITAB, STATGRAPHICS, STATISTICA, STATVIEW)

SAS [<http://www.sas.com>, SAS Institute, Inc.]. SAS обладает превосходными возможностями по набору статистических алгоритмов. Кроме того, SAS предоставляет пользователю возможность подключения его оригинальных алгоритмов. SAS/IDS - это интеграция весьма разнообразных возможностей доступа к данным и управления ими, средств анализа данных, способов представления информации и генерации отчетов. Пакет имеет модульную структуру и легко может быть сконфигурирован под специфические особенности его пользователя. Основными достоинствами SAS являются мощное интеллектуальное ядро, поддержка архитектуры «клиент-сервер», возможность доступа и интеграции данных из любых источников и наличие объектно-ориентированной технологии быстрой разработки приложений. В практическом плане, SAS сообщает пользователю, какие переменные определены, какого они типа, какие переменные являются активными, какой тип модели Вы бы хотели использовать для анализа и специальные требования на используемый метод вычислений. Формирование отчетов нетрудно, если пользователь владеет соответствующими командами пакета. Главные недостатки пакета - громоздкость, большие трудности в освоении, высокие требования к статистической квалификации пользователя, жесткие требования к аппаратной части ПЭВМ, большой размер, занимаемый на винчестере.

SPSS [<http://www.SPSS.com>, SPSS, Inc.]. SPSS (рис. 1) пакет предназначен в первую очередь для статистиков-профессионалов, так как имеет достаточно мощный аппарат статистического анализа, вполне соизмеримый по мощности с SAS. SPSS версии 8.0 для Windows 2000/XP

отличается разнообразными возможностями по управлению данными и манипулированию полученными результатами, по работе с электронными таблицами. SPSS предоставляет достаточно удобную графику (более 50 типов диаграмм), а также развитые средства подготовки отчетов. Эта версия отличается повышенной производительностью и скоростью вычислений, а также расширенным функциональным наполнением. Аналитические параметры отображаются на экране в виде простых и ясных меню и диалоговых окон. С введением новой панели инструментов требуется лишь один щелчок мышки для запуска часто используемых процедур. Для снижения трудоемкости освоения работы с SPSS в эту версию введено новое обучающее средство «Навигатор». SPSS обладает весьма полным набором статистических (всего их более 60-ти) и графических процедур, а также процедур создания отчетов. Интерфейс прост и удобен. Пакет отличается высокой точностью вычислений.

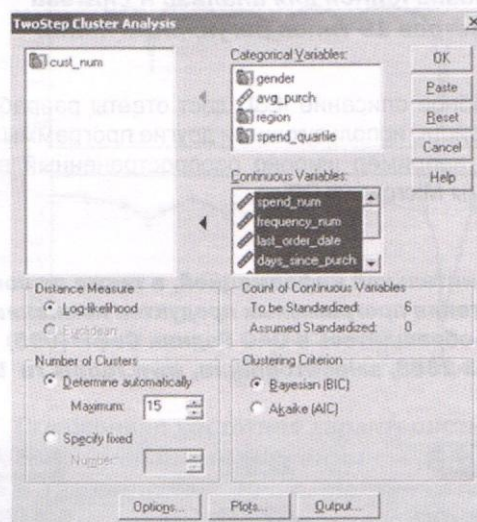


Рис.1

SYSTAT [<http://www.spss.com>, SPSS, Inc.]. Инструкции по пошаговой работе, как правило, полезны и точно ведут к цели, а указатель просто превосходен и очень полезен для пользователя, начинающего работу с SYSTAT. Пакет использует табулированное окно для ввода данных и их редактирования. Предоставляемые возможности по преобразованиям данных легки в использовании. Однако результаты таких преобразований являются «статистическими»: однажды вычисленные, данные уже не преобразуются, если Вы вдруг впоследствии внесли изменения в исходные данные. Функциональные имена в окне преобразований являются кодами, например, XDF и XCF обозначают плотность и кумулятивную функцию распределения случайной величины хи-квадрат. Пропущенные символьные значения кодируются в виде пробелов; пропуски в числовых признаках кодируются как самое отрицательное число и появляются в редакторе как периодические вещественные числа. Пакет обладает прекрасными возможностями отображения на экране исходных данных и полученных результатов разведочного анализа, имея в своем распоряжении около 30-ти различных способов графического отображения: гистограммы, ящики с «усами», стебли с листьями, иконки, 2-D и 3-D диаграммы рассеяния и т.д. Кроме того, имеются матрицы диаграмм рассеяния, графики функций и географических карт. В рекламных материалах подчеркивается, что эта версия пакета доминирует среди аналогов

по разнообразию типов графиков. Кроме того, пакет позволяет порождать и изображать сложные поверхности, что полезно для визуализации сложных функций. SYSTAT обладает высокой точностью, используя много превосходных алгоритмов.

Пакет дает удобные возможности работы во всех областях статистики, но особенно он силен в области дисперсионного анализа и планирования экспериментов. В последней версии появилось множество дополнительных процедур для дискриминантного анализа, матричной алгебры, логлинейных моделей, планирования экспериментов, структурного анализа и карт контроля качества. Также были добавлены робастные алгоритмы, дающие точные и корректные результаты при почти вырожденных данных. Кроме того, эта версия предоставляет пользователю наиболее широкие возможности анализа общей линейной статистической модели. Несколько важных статистических методов решительно не являются дружественными к пользователю, например по непарному t-критерию и простому однофакторному дисперсионному анализу.

STATISTICA [<http://www.statsoft.ru>, StatSoft, Inc.]. Пакет не стоит использовать пользователю-новичку в статистике, т.к. предполагает владение статистической терминологией. Наличие достаточно широкого спектра функциональных алгоритмов делает пакет STATISTICA (Рис. 2.) версии 5.5 достаточно привлекательным для статистиков-профессионалов. Однако относительно распространена точка зрения, в соответствии с которой удобство работы с этим пакетом оценивается весьма невысоко. В частности, STATISTICA по своей структуре как бы состоит из нескольких связанных между собой модулей. Эти модули взаимодействуют друг с другом, имея одинаковый формат файлов. Так, если Вам нужен раздел линейной регрессии, то Вы должны покинуть оболочку главного модуля СПП и загрузить модуль линейной регрессии.

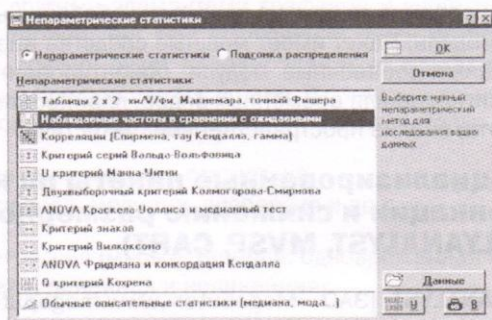


Рис.2

В плане функционального наполнения пакет, например, по сравнению с программой STSC+, отличается большим разнообразием, включая в себя и те разделы анализа, которые STSC+ содержит лишь в дополнительных модулях. Например, он включает в себя ряд непараметрических методов анализа, методы многомерного анализа: дискриминантного, факторного кластерного логлинейного и др.

Если Вы забыли задать ту или иную переменную или параметр метода, то пакет тут же заметит это, и сделает запрос к пользователю с необходимой подсказкой. STATISTICA имеет возможность работы в пакетном режиме, используя свой командный язык SCL. Можно использовать и наборы команд, объединяемые в последовательности или макро-

сы. Используемые в пакете макрокоманды предоставляют более широкие возможности, чем стандартные Windows-макросы. Например, с помощью макрокоманд удобно готовить презентации. Кроме того, пакет дает средства составления отчетов. Наиболее сильной стороной пакета является графика и средства редактирования графических материалов. Представлены сотни типов графиков: типа 2-D или 3-D (имеются даже графики типа 4-D), матрицы и пиктограммы. Есть возможность разработать свой дизайн графика и добавить его в меню.

К сожалению, в пакете отсутствуют методы планирования экспериментов и графика по методам контроля качества. В целом пакет STATISTICA-5.5 по мощности уступает перечисленным выше пакетам.

MINITAB [Minitab Inc., <http://www.minitab.com>]. MINITAB продуман по разделу описательной (дескриптивной) статистики, хорошо сконструирован и управляется с помощью очень удобного меню, или, по желанию пользователя, через команды, составляя которые помогают диалоговые окна пакета. Так, в версии 10.0 были добавлены алгоритмы факторного и кластерного анализа. Кроме того, эта версия позволяет получать множество хороших и сложных многоцветных графиков. Часто используемые команды можно запускать и по их первой букве. Общее число команд превышает 200. Можно составлять и специальные макросы для выполнения последовательностей команд. Таким образом, пакет ориентирован на неспециалистов в области математической статистики, в частности, на студентов университетов. Пользователь MINITAB может легко и быстро решать практически все типовые задачи, в основном из области одномерного анализа и анализа временных рядов. Фирмой Minitab Inc. хорошо налажена поддержка пользователей и обмен опытом через группу пользователей MINITAB. Кроме того, фирма распространяет минируководство по пакету для тех студентов, которые, не имея его хотят больше узнать о пакете.

STATGRAPHICS [<http://www.manugistics.com>, Manugistics, Inc.]. Пакет предназначен, в основном, для тех пользователей, которые уже имеют определенный опыт в статистике. Особенно это касается модуля работы с многомерными методами. Для корректного их использования пользователь должен иметь базовые знания по статистике и знать допущения и ограничения тех или иных статистических критериев и многомерных методов. STATGRAPHICS хорошо известен российскому пользователю (иногда он обозначается STSC). Пакет был изначально разработан и нацелен, в первую очередь, на графические возможности компьютерной статистики. Однако постоянное его совершенствование в плане функциональных алгоритмов и способов управления данными существенно усилило его базовую мощность. Важнейшим «плюсом» пакета считается удачное соединение математического аппарата обработки данных с современной интерактивной графикой. Другие, менее существенные его достоинства - это широкие возможности взаимодействия с электронными таблицами и СУБД, а также с разнообразной периферией. В плане классификации и смежных с нею задач, пакет содержит такие разделы, как дескриптивная статистика, разведочный анализ, многомерный анализ. Относительными недостатками STSC являются нечеткости в его справочной системе и отображение результатов расчетов с точностью до 4-5 значащих цифр.

STATVIEW [Abacus Conceptr, <http://www.abacus.com>.] Фирма Abacus Conceptr представляет свой продукт как интегрированную систему для анализа данных и презентаций, ориентированную на исследователей и аналитиков. По их мнению, STATVIEW версии 5.0 является хорошим, и возможно даже превосходным средством для решения этих задач, особенно для данных малого-среднего размера (меньше 1000 наблюдений). В STATVIEW имеются два основных типа документов для данных и для их анализа. Первый тип имеет вид электронных таблиц. Документы для анализа могут комбинировать таблицы, текст, графику, с тем, чтобы можно было создать отчеты на основе документов первого типа. STATVIEW использует интерфейс типа электронных таблиц для ввода данных и манипуляций над ними. Имеющиеся преобразования данных на основе заданных Вами формул, основываются на более чем 150-ти встроенных арифметических, алгебраических, тригонометрических, статистических и других функциях. Преобразования могут быть статистическими или динамическими. В последнем случае изменение входных данных автоматически отслеживается в преобразованных данных. Поддерживается обработка данных с пропущенными значениями. Имеются возможности извлечения подвыборок, сортировки или разбиения данных на подмножества.

STATVIEW предлагает дескриптивную статистику, частотные распределения, вычисления процентных точек, t-критерии, доверительное оценивание, анализ корреляций и ковариаций, регрессию (простую, множественную, полиномиальную, пошаговую), дисперсионный анализ, анализ таблиц сопряженности, непараметрические методы и разведочный факторный анализ.

Возможность доступа к любому виду анализа обеспечивается графическим пользовательским интерфейсом. Всего имеется четыре вида базовых компонент для окон: стандартные «маковские» падающие вниз меню, иерархическая палетка анализа, палетка переменных и окно анализа результатов.

По своим графическим возможностям, STATVIEW является одним из лучших пакетов для графического представления данных.

Все рассмотренные выше пакеты позволяют импортировать или экспортировать данные формата ASCII, dBASE, Lotus 1-2-3. Кроме того, они способны компоновать данные для анализа из различных файлов, осуществлять выборку подмножеств данных, их ранжирование или сортировку по тем или иным критериям. И, наконец, всегда имеется возможность отредактировать данные в среде пакета и добавить описательный текст к анализируемому набору данных (файлам).

Специализированные пакеты, включающие мощную статистическую компоненту (IMSL, S-PLUS)

IMSL [Visual Numerics, <http://www.vni.com>]. Пакет IMSL по сути является библиотекой процедур и ориентирован прежде всего на исследователей и на разработчиков прикладных программных продуктов. IMSL можно использовать в виде библиотеки исходных текстов на C++. Библиотеки пакета IMSL можно использовать отдельно или же с подключением графической библиотеки Exponent Graphics или другой объектно-ориентированной библиотеки на языке

C++, называемой Object Suite. Последняя библиотека составлена из математического и графического модуля. Первый, кроме математических функций, поддерживает также разработку объектно-ориентированного интерфейса с пользователем.

IMSL состоит из двух больших разделов: Math/Library и Stat/Library, кроме которых к нему также можно подключить графику. Последняя представлена либо мощной библиотекой графических функций Exponent Graphics-2.1 или же графическим модулем из библиотеки Object Suite.

IMSL содержит исчерпывающе полный набор математических и статистических функций, и, в частности, по набору первых даже конкурирует со специализированной вычислительной библиотекой NAG.

S-PLUS [AT&T Bell Labs, <http://www.mathsoft.com>]. S-PLUS представляет собой интерактивную среду, которая включает в себя полноценный графический анализ данных и S-язык программирования, который является расширяемым и удобным для использования. Среда S-PLUS создана аналогично библиотеке IMSL. S-язык разработан в AT&T Bell Labs и является единственным объектно-ориентированным языком, специально предназначенным для анализа данных. Именно поэтому получаемые результаты имеют неограниченную свободу при проведении исследований, анализа и моделирования данных в науке и технике.

S-PLUS может быть очень полезной для статистика-аналитика, умеющего составлять свои программы на основе объектно-ориентированной технологии. S-PLUS имеет в своем составе более 1650 функций, включая регрессию и дисперсионный анализ, многомерные методы, временные ряды, анализ сигналов и др. Имеются и современные робастные (устойчивые) методы. В плане классификации имеются современные непараметрические методы: древообразные модели классификации, модели целенаправленного проецирования данных на плоскость, обобщенные аддитивные модели. Пакет имеет мощные средства визуализации, а также добавочные модули, ориентированные под анализ сигналов или временных рядов, планирование эксперимента, анализ пространственной статистики.

Специализированные пакеты по классификации и снижению размерности (POLYANALYST, MVSP, CART)

POLYANALYST [ЗАО «Megaputer», www.megaputer.com]. Этот пакет, работающий под управлением OS/2 Warp фирмы IBM, умеет автоматически обнаруживать знания, скрытые в базах данных. Пакет работает с данными произвольного типа: вещественными или целыми, логическими или качественными. Он позволяет проводить поиск многофакторных зависимостей в данных в виде аналитических формул, структурных правил и алгоритмов, а также делать вывод по обучающим примерам правил для классификации новых данных.

Выводимые пакетом законы имеют разнообразную форму и включают в себя алгебраические операторы, функции, циклы, условные конструкции. POLYANALYST относится к довольно новому направлению искусственного интеллекта, называемому «символическими методами извлечения знаний из баз данных» (data mining & knowledge discovery) или «ДМ-технологией».

CART [SciTech Software, Inc., www.scitechsoft.com]. Пакет предназначен для автоматического разведочного анализа данных и работает на многих платформах. Пакет строит классификационные и регрессионные деревья на основе оригинальной программы, предназначенной для анализа данных на основе метода древовидных структур. CART автоматически строит и минимизирует такие деревья, которые похожи на дендрограммы. В справочной системе приведены примеры работы с пакетом, подробно разобранные от начала до конца.

Пакеты этой группы позволяют существенно экономить место на винчестере и другие ресурсы компьютера. Они весьма удобны, если прикладная задача пользователя попадает под их специализацию, так как специализированные пакеты предоставляют определенный раздел прикладной статистики с максимально возможной полнотой, включают новейшие и мощные методы, редко встречающиеся в универсальных пакетах, и обеспечивают предлагаемому методу наилучшие характеристики по скорости или точности вычислений и другим ресурсам.

Пакеты для решения задач, смежных с классификацией (BMDP, TURBO SPRING-STAT-WIN, STATISTIX, SIGMASTAT, STATXACT-3, MS-EXCEL 2003)

BMDP [<http://www.SPSS.com>, SPSS, Inc]. Этот пакет, называемый в каталоге BMDP New System Pro, и обозначаемый ниже для краткости BMDP, специально был разработан для Windows фирмой BMDP, которая была хорошо известна специалистам, работающим на «больших машинах». В последние годы эта фирма является дочерней компанией корпорации SPSS. Важным плюсом пакета в управлении данными является связь объектов. Это обеспечивает большие удобства при установлении соответствия между точками или другими элементами графиков и ячейками электронных таблиц. Так, при исключении точек-«выбросов» из данных, график будет перерисован и результат будет немедленно предъявлен на экране. В функциональном плане реализованы следующие статистические методы:

- вся описательная статистика (среднее, медиана, дисперсия и др.);
- регрессия (множественная, одновременная и пошаговая, линейная и нелинейная);
- корреляционный и стандартный дисперсионный анализ (только без пропусков в ячейках).

BMDP управляется только через меню и не имеет командного языка. Через окно Transform к имеющимся преобразованиям можно добавлять и свои.

TURBO SPRING-STAT-WIN [SciTech Software, Inc., <http://www.scitechsoft.com>]. Интерактивный, меню-управляемый, модульный пакет с широким набором статистических процедур, с 2-D и 3-D графикой и встроенной экспертной системой.

STATISTIX [SciTech Software, Inc., <http://www.scitechsoft.com>]. Пакет содержит довольно широкий спектр средств статистического анализа: описательную статистику, t-критерии, дисперсионный, регрессионный и компоне-

нтный анализ, проверку гипотез, временные ряды, графику, анализ таблиц сопряженности, логистическую регрессию и др.

SIGMASTAT [Jandel Scientific, www.jandel.com]. Этот пакет является показательным примером целенаправленного ПС, который в своем классе методов является, скорее всего, самым лучшим. Преимущества SIGMASTAT проявляются в интерфейсе пользователя, а также в его Консультанте (Advisor), который особенно полезен для начинающего пользователя. По форме Консультант представляет собой диалоговое окно, которое через серию вопросов побуждает пользователя активно мыслить и действовать, что должно приводить пользователя к нужным статистическим методам. В случае же, если пользователь уверенно себя чувствует в окружающем его меню и в процедурах, то он может выбрать нужные ему методы непосредственно из меню.

Пакет явно не претендует на то, чтобы считаться универсальным. Но в то же время его методы достаточно разнообразны. Они включают в себя, кроме традиционной описательной статистики, задачи регрессионного и дисперсионного анализа, а также ряд непараметрических тестов. Пакет имеет много встроенных преобразований данных и небольшой подраздел программирования, в котором пользователь может задать оригинальные преобразования.

Сильной стороной пакета является присущая ему комбинация легкости использования, усиливаемая для новичка превосходными алгоритмами отбора часто используемых статистических методов, с подобной документацией.

STATXACT [Cytel Statistical Software, Inc., www.cytel.com]. Это самый полный пакет статистического анализа, ориентированный, прежде всего, на анализ малых выборок на основе точных (не асимптотических) непараметрических статистических методов. Пакет имеет специальные версии для стыковки с универсальными пакетами SYSTAT и SPSS.

Статистические экспертные системы (STATISTICAL NAVIGATOR PRO, STAREX)

STATISTICAL NAVIGATOR PRO [SciTech Software, Inc., <http://www.scitechsoft.com>]. Этот пакет помогает неспециалисту в математической статистике провести квалифицированный анализ имеющихся у него данных. Он консультирует пользователя по использованию более чем 200 различных методов, включая многомерный анализ и классификацию. STATISTICAL NAVIGATOR PRO задает ряд вопросов относительно цели исследования данных и их характера, сопровождая каждый из вопросов серией подсказок. Пакет фокусирует внимание пользователя на ключевых моментах, подчеркивая, что необходимо принять во внимание для принятия верного решения.

Специализированные программные пакеты по анализу результатов испытаний на надежность (RELEX WEIBULL 7.7, WEIBULL6++)

RELEX WEIBULL 7.7 [Relex Software Inc., <http://www.relex-software.com>]. RELEX WEIBULL (Рис. 3.) - модуль статистического анализа, анализируемые виды распределений включают нормальное, Вейбулла, логнор-

мальное, равномерное, экспоненциальное, Гумбеля, Рэ-ля, биномиальное и др. Представление и анализ данных для выбранных классов параметрических распределений проводится с использованием метода «вероятностной бумаги». На «вероятностной бумаге» анализируемое распределение представляется прямой линией, что обеспечивает наглядность и позволяет естественным образом применять все методы регрессионного анализа.

В частности, проверки адекватности модели и значимости коэффициентов регрессии (Фишеровский анализ). Для оценок параметров распределений предлагается большой набор методов, например, методы Хазена (Hazen), Бенарда (Benard) и их модификации, биномиальное оценивание, метод средних, метод максимального правдоподобия и его модификация и др. Необходимо отметить, что квалифицированное применение модуля RELEX WEIBULL требует от пользователей весьма высокой подготовки в области математической статистики в приложении к задачам теории надежности.

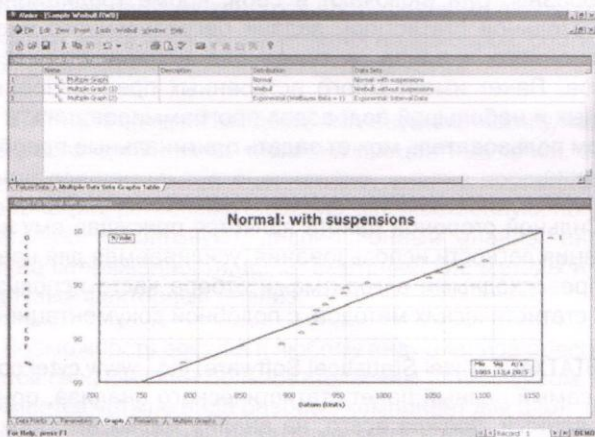


Рис.3

WEIBULL6++ [ReliaSoft Inc., <http://www.weibull.com>]. Пакет (Рис. 4) статистического анализа, анализируемые виды распределений включают нормальное, Вейбулла, логнормальное, экспоненциальное. Не требует специальных знаний по математической статистике. Исходные данные - наработки до отказа, которые вносятся в таблицу и после выбора распределения рассчитываются его параметры.

Приведенные выше ПС являются наиболее известными. Вместе с тем, существуют и другие специализированные и универсальные пакеты, характеристики ряда которых приведены ниже.

STATISTICAL SOFTWARE [NCSS Statistical Software, <http://www.ncss.com>]. Пакет Statistical Software, называемый для краткости NCSS, легок в использовании и управляется через меню. В NCSS имеется и отдельный модуль Углубленной статистики. Базовый модуль включает управление данными, описательную статистику, t-критерии, кросстабуляцию, дисперсионный, регрессионный, факторный, кластерный и дискриминантный анализ, повторяющиеся измерения и непараметрические критерии. В модуле Advanced Statistics содержатся методы канонической корреляции, логистической и нелинейной регрессии, поверхности отклика, дополнительные кластерные алгоритмы,

графики на вероятностной бумаге, компонентный анализ и многомерный дисперсионный анализ.

ODA [SciTech Software, Inc., <http://www.scitechsoft.com>]. Пакет обеспечивает наиболее адекватную модель для любой заданной конфигурации исходных данных. Он включает ряд стандартных статистических процедур. Применительно к классификации это кластерный и дискриминантный анализ, включая методы типа jack-knife (методы «складного ножа», т.е. «скользящего экзамена»), логистическая регрессия, нелинейные модели и деревья классификации. Также имеется очень много методов, смежных с классификационными.

SOLO [SciTech Software, Inc., <http://www.scitechsoft.com>]. Пакет для углубленного статистического анализа, содержащий широкий класс статистических методов. Из классификационных и близких к ним методов содержит модель логистической регрессии, многомерное шкалирование и анализ соответствий, дискриминантный и кластерный анализ. Методы для углубленного анализа данных реализованы в виде дополнительного модуля. Пакет разработан для пользователя, который довольно редко обращается к статистическому анализу данных. В нем используется графика с высокой разрешающей способностью.

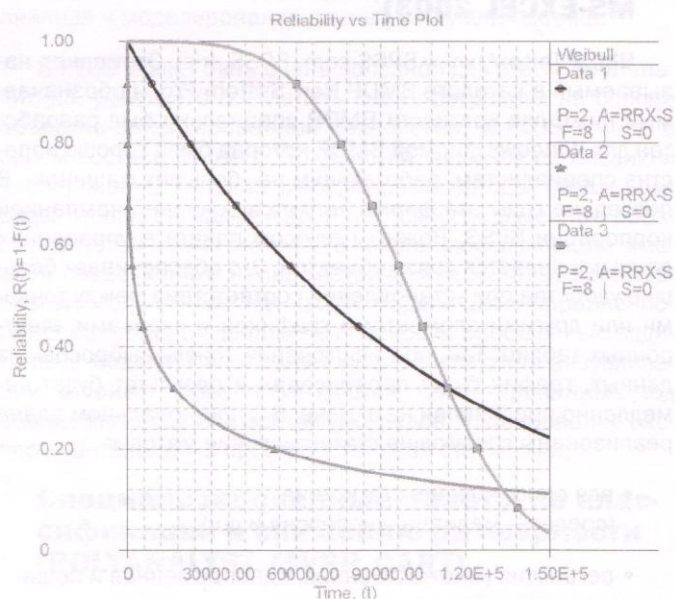


Рис.4

STATLAB PRO [SciTech Software, Inc., <http://www.scitechsoft.com>]. Пакет полезен для графического анализа данных (используется более 60 типов графиков). Он имеет модульную структуру и адаптирован ко многим платформам. Содержит методы описательной статистики, регрессию, корреляцию, анализ таблиц сопряженности, развитый аппарат проверки гипотез, метод главных компонент, анализ соответствий, факторный и дискриминантный анализ. В пакете нет ограничений на число переменных и наблюдений, он импортирует файлы формата ASCII, dBASE и Excel. В новой версии используется графический интерфейс с пользователем, кластерный анализ, непараметрические критерии проверки гипотез и множественная регрессия. Для обмена данными с другими Windows-приложениями используется стандарт ODBC.

WINSTAT [SciTech Software, Inc., <http://www.scitechsoft.com>]. Пакет предоставляет полные возможности работы под Windows с падающими меню, диалоговыми окнами, поддержкой мыши и средствами вырезания и вставки. Включает описательную статистику, t-критерии, корреляционный, дисперсионный, факторный, кластерный и дискриминантный анализ, временные ряды и графику.

MULTIVARIANCE 7 [SciTech Software, Inc., <http://www.scitechsoft.com>]. Это работающий с хорошей скоростью методо-ориентированный пакет для регрессионного, дисперсионного и ковариационного анализа, включающая многомерные методы. В частности, он содержит методы дискриминантного и компонентного анализа, анализа канонических корреляций и анализа данных о продолжительности жизни.

JMP STATISTICAL [<http://www.sas.com>, SAS Institute, Inc.]. JMP версии 5.1 для Windows является инструментарием для «статистического открытия». Пакет предназначен для исследователей, аналитиков, статистиков. Основная ориентация пакета - статистическая графика, поэтому он может служить хорошим инструментарием для проведения разведочного статистического анализа данных. Пакет обеспечивает богатую, интегрированную, графическую среду для различных методов прикладного статистического анализа, включая кластерный анализ и планирование эксперимента. В пакете хорошо развит интерактивный графический анализ, его результаты динамически соединяются с исходными данными. В частности, при выделении мышкой точки на одном из графиков, она автоматически выделяется на всех других.

В JMP 5.1 принят скорее подход с вычислением и последующим представлением пользователю полного набора имеющихся средств анализа и графики по выбранной задаче, чем подход с ведением его по всей технологической цепочке последовательности шагов. JMP привлекает исключительно широким диапазоном методов многомерной линейной модели, включая пробит- и логит-анализ. Функциональные возможности в меню «анализ» включают в себя: распределения и подгонка зависимостей; построение и анализ линейной модели; построение и анализ нелинейной модели (включая пробит- и логит-регрессию); анализ матрицы рассеяния, главные компоненты, факторный анализ; анализ данных типа времени жизни; планирование эксперимента.

DATA DESK [Data Description, Inc.]. Этот пакет был изначально разработан как программное средство для Макинтошей и предназначен, в основном, для проведения разведочного анализа данных, а также для обучения этому разделу компьютерной статистики. Пакет использует тип интерфейса, близкий к «Маковскому» значительно в большей степени, чем другие статистические пакеты. Он построен так, чтобы разведочный анализ был удобным и легким в управлении. Пакет имеет хорошую скорость работы и весьма компактен.

Средства разведочного анализа, такие как «вычищение части данных» (brushing) и «разрезание на слои» (slicing), а также связывание через «горячие клавиши» (hot links) для одних и тех же данных всех относящихся к ним

графиков и изображений, - все это выделяет данный пакет в области интерактивного разведочного анализа данных. В дополнение к этому, так называемые «гипервзгляды» (hyperviews) позволяют пользователю активизировать список дальнейших действий, которые могут производиться либо через иконку в окне результатов, либо указанием почти на любой элемент на экране. Например, «гипервзгляды» дают возможность в рамках линейной регрессии, построить регрессию сразу после выдачи диаграммы рассеяния или показать график остатков, относящихся к уравнению регрессии.

Любые операции с преобразованиями данных применяются к признакам на основе синтаксиса, похожего на тот, что используется в языках программирования. Полученный результат преобразований хранится в автоматически создаваемой переменной, так что данное преобразование может быть сделано всего один раз. В меню предлагается широкий набор стандартных преобразований. Оригинальным здесь является применение динамического преобразования признака типа Бокса-Кокса, в котором экспонента управляется через операцию «скольжения». Это легко делается с помощью мышки. Если Вы сделали некоторое преобразование, то на «автомате» все активные окна, в которые входит данный признак, будут тут же пересчитаны, но уже для преобразованных данных.

Сильной чертой пакета является его графические возможности, относящиеся именно к разведочному анализу данных. Особенно выделяется здесь пакет в плане использования цвета и символов, техники drag-and-drop, «горячего сцепления» (hot links) данных из различных окон. Однако единственными многомерными графиками являются матрицы с диаграммами рассеяния и график с возможностью вращения осей координат. Пропущен при разработке и ряд других традиционных графиков (лица Чернова и др.).

В последнюю версию 5.1.1. включены «шаблоны», которые дают возможность через технику «тащи и роняй» перемещать новые наборы данных к ранее записанным результатам анализа и далее использовать нужные Вам модули. Статистическая подсистема включает в себя в этой версии нелинейную и логистическую регрессию, а также «оконтуривание» диаграмм рассеяния. В документации на пакет описана как методология, так и статистика и способы управления пакетом. Однако, оперативная документация, построенная по принципу гипертекста, трудна в использовании.

Естественно, что приведенные выше пакеты не исчерпывают все многообразие ПС статистического анализа. Более подробную информацию об этих и других пакетах можно найти на сайтах фирм-производителей, материалы которых были использованы в данной статье.

За дополнительной информацией, а также по вопросам приобретения программных продуктов обращайтесь в ОАО Родник Софт: (095) 113-7001, 113-2688, sales@rodnik.ru, www.rodnik.ru ■