

КОМПЬЮТЕРНЫЕ МЕТОДЫ ОБРАБОТКИ ДАННЫХ

Краткий обзор статистических пакетов

Введение

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, в том числе для анализа данных психологических исследований.

В настоящее время число статистических пакетов перевалило за 1000

Примеры зарубежных статистических пакетов: STATGRAPHICS, SPSS, SYSTAT, BMDP, SAS, CSS, STATISTICA, S-plus, R и т.п.

Примеры отечественных статистических пакетов: STADIA, ЭВРИСТА, МИЗОЗАВР, ОЛИМП:Стат-Эксперт, Статистик-Консультант, САНИ, КЛАСС-МАСТЕР и т.д.

Виды статистических пакетов

Универсальные

Специализированные

Универсальные
профессиональные

С открытым кодом
(настраиваемые)

Универсальные пакеты

Примеры:

STADIA, STATGRAPHICS, STATISTICA, S-plus

Универсальность этих пакетов особенно полезна:

- на начальных этапах обработки, когда речь идет о подборе статистической модели или метода анализа данных;
- когда поведение статистических данных выходит за рамки использовавшейся ранее модели;
- в процессе обучения основам статистики.

Основной недостаток таких систем — высокая стоимость

Профессиональные пакеты

Примеры:
SPSS, SAS/IDS, BMDP

Профессиональность этих пакетов заключается в следующем:

- практически полное отсутствие ограничений в проведении расчетов;
- высокая мощность программного продукта;
- широкое распространение;
- наличие профессиональной поддержки продукта;
- достаточное количество обучающих материалов.

Основные недостатки таких систем – очень высокая стоимость и высокие требования к квалификации пользователя.

Специализированные пакеты

Примеры:

EQS, ЭВРИСТА, ОЛИМП: Стат-Эксперт

Профессиональность этих пакетов заключается в следующем:

- пакеты включают в себя отдельные виды или модели анализа;
- созданы для конкретной предметной деятельности или задачи;
- распространены и известны часто только узким специалистам.

Основные недостатки таких систем – узкая специализация, зачастую не в области психологии (кроме EQS).

Пакеты с открытым кодом

Примеры:

R

Особенность этих пакетов заключается в следующем:

- базовый пакет включает в себя отдельные виды или модели анализа;
- существует база дополнений к пакету, существенно расширяющая функциональные возможности программной среды;
- при условии настройки пакета позволяют решать практически любую задачу по статистическому анализу;
- свободно распространяются и модифицируются.

Основные недостатки таких систем – малая популярность в России.

Пакет STADIA (версия 7)

STADIA - 6.0

Файл График=F6 Вычисл=F7 Преобр=F8 Статист=F9 Окона Помощь=F1

Таблица данных: tst.std

	sample	lc1	lc2	f1	f2	PolyrX	PolirY	ExprY
1	1.9	20.1	7.2	20	22.1	1	5	9
2	2.5	23						7
3	3.4	26						5
4	4.1	19						4
5	5.8	16						3
6	6.6	23						5
7	7.2	31						2
8	8	33						5
9	9.199	28						1
10	0.1							
11	2.4	29						
12	3.5	30						
13	4.8							
14	5.6							
15	6.3							
16	7.3							
17	8.1							
18	1.4							
19	3							
20	4.7							

Графики данных

Математические графики

- 1=функциональный
- 2=с отклонениями
- 3=диагр.рассеяния
- 4=распределение
- 5=ящики-с-усами

Трёхмерные графики

- A=диагр.рассеяния
- B=поверхность
- C=сглаживание
- D=картирование

Деловые графики

- 6=столб.диаграмма
- 7=диаграмма-башня
- 8=100%-я башня
- 9=круг.диаграмма
- ☒ столбики ☒ 3D-эффект

Сглаживающие

- E=интерполяция
- G=сглаживание

Шагов= 50 Коэф= 0.5

☒ <Esc>=Отменить

Одна или несколько экспериментальных зависимостей $Y=f(X)$

Dat / Rez /

Пакет STADIA

<http://statsoft.msu.ru>

Пакет STADIA разработан ведущими специалистами Московского государственного университета им. М. В. Ломоносова (главный разработчик - А. П. Кулаичев) совместно с НПО "Информатика и компьютеры". Первая версия пакета была создана в конце 70-х гг.

В настоящее время пакет используется в учебном процессе и научно-практической работе более чем в 150 университетах России, включая 17 университетов медицинского профиля. Среди пользователей пакета: НИИ им. Сербского, НИИ педиатрии РАМН, НИИ дефектологии, поликлиники, больницы, медсанчасти городов РФ. Пакет относительно недорог и функционален.

Пакет STATISTICA (версия 10)



Пакет STATISTICA

<http://www.statsoft.ru>

<http://www.statsoft.com>

Ряд авторов считает, что пакет Statistica является хорошо сбалансированным по соотношению "мощность/удобство". Наличие достаточно широкого спектра функциональных алгоритмов делает его достаточно привлекательным для статистиков-профессионалов.

Средства манипулирования исходными данными в пакете Statistica хорошо развиты. Зачастую для проведения определенного вида анализа требуется несколько щелчков мышью.

Сильной стороной пакета является графика и средства редактирования графических материалов.

Пакет SPSS (версия 19)



Пакет SPSS

<http://www.spss.ru>

<http://spss.com>

Пакет SPSS предназначен в первую очередь для статистиков-профессионалов. Он включает развитый аппарат статистического анализа, соизмеримый по мощности с SAS. Программу SPSS для Windows считают в настоящее время одним из лидеров среди универсальных статистических пакетов.

SPSS имеет удобные графические средства (более 50 типов диаграмм), а также развитые средства подготовки отчетов. Аналитические параметры отображаются на экране в виде простых и понятных меню и диалоговых окон.

Пакет SAS/IDS

The screenshot displays the SAS software environment. The main window shows a SAS script named 'salesreport.sas' with the following code:

```
proc tabulate data=sashelp.prodall;
  where country = 'Canada';
  class country year prodtype product;
  var actual predict;

  keyword all / style = (font_weight=bold font_style=italic);
  table country*prodtype*(product=' ' all*(a=<parent)*(foreground=black
    year*sum=' ' *Z=dollar12.*(actual predict)
    sum='Combined' *Z=dollar12.*(actual predict) ;

run;

options reset=all device=gif hsize=7in vsize =5in vorigin=2.5in
  ftext='Arial' btext=12 pt ctext=black colors=(cream orange steel);

/* create graph */
proc gchart data=temp;
  where _type_='2';
  pieid country / sumvar=actual ffreq_freq_type=sum
  nolegend noborder name='graph';

run; quit;

ods pdf close;
```

The Log window (Log - (Untitled)) shows the following messages:

```
NOTE: There were 3 observations read from the data set WORK.TEMP.
      WHERE _type_='2';
NOTE: PROCEDURE GCHART used (Total process time):
      real time      0.72 seconds
      cpu time       0.19 seconds
```

The Results Viewer window displays a table titled 'Total Product Sales by Country for 1997 & 1998' with the filter 'Country = Canada'. The table shows sales data for 1997, 1998, and Combined years, categorized by Country, Product Type, and Product.

Country	Product Type	1997		1998		Combined	
		Actual Sales	Predicted Sales	Actual Sales	Predicted Sales	Actual Sales	Predicted Sales
Canada	TURNED	\$224,182	\$235,449	\$236,509	\$266,818	\$460,691	\$502,267
	SOFA	\$231,406	\$242,847	\$243,794	\$275,042	\$475,200	\$517,889
	AP	\$437,868	\$478,296	\$486,363	\$516,430	\$924,231	\$994,726
OFFICE	CHAIR	\$202,283	\$230,994	\$255,456	\$263,762	\$457,739	\$494,753
	DESK	\$230,670	\$235,445	\$238,866	\$257,304	\$469,536	\$492,749
	AP	\$432,951	\$466,438	\$484,322	\$521,467	\$917,273	\$987,905
All Products		\$896,551	\$944,712	\$974,625	\$1,057,527	\$1,865,176	\$2,002,259

A pie chart titled 'Total Product Sales by Country for 1997 & 1998' shows the distribution of sales by country:

- U.S.A. \$7419927.70
- Canada \$1865176.30
- Mexico \$1231659.10

Пакет SAS/IDS

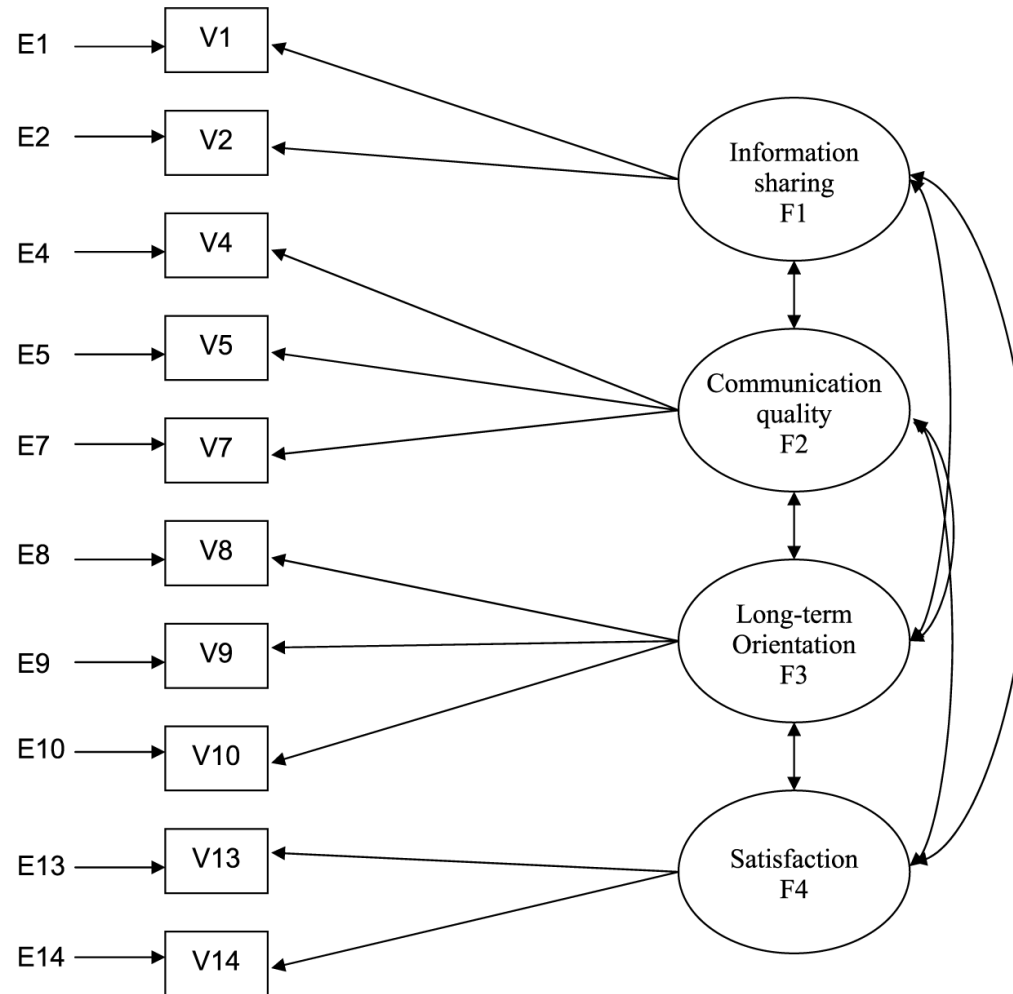
<http://www.sas.com>

Система SAS известна с 1976 г. и способна работать под управлением практически любой операционной системы.

SAS включает свыше 20 различных программных продуктов, объединенных друг с другом Information Delivery System или IDS.

Традиционные пользователи в РФ: предприятия ВПК, крупные бизнесмены (банки, включая Центробанк, биржи, торговые фирмы), некоторые атомные станции, крупнейшие медицинские и геофизические центры, крупные государственные структуры и др. Использование SAS возможно решить практически любые задачи.

Пакет EQS (версия 6.1)



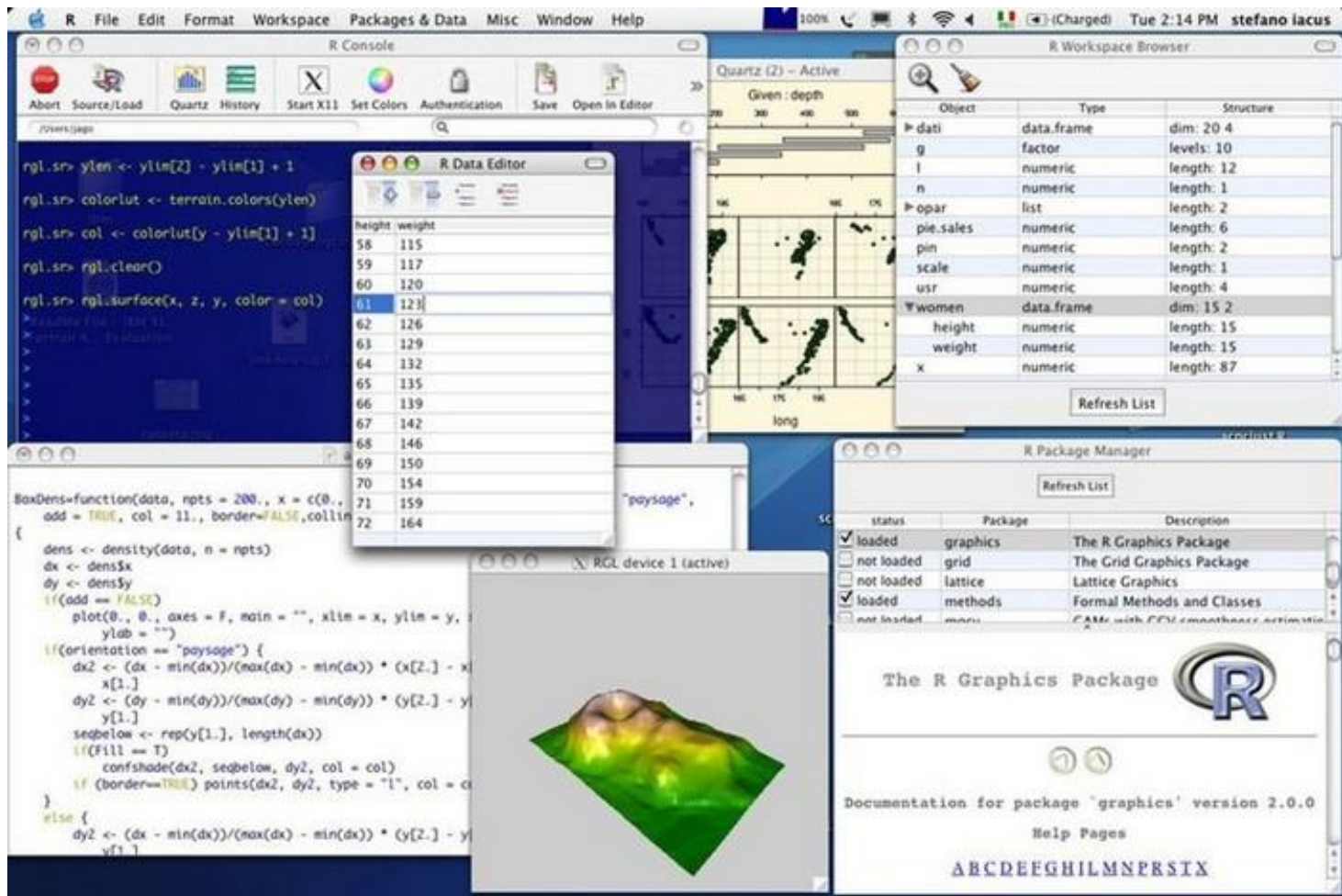
Пакет EQS

<http://www.mvsoft.com/>

Пакет содержит в себе набор уравнений для анализа данных, полученных в психологических исследованиях.

Основной принцип аналитической модели – так называемый Path analysis, предполагающий построение разветвленной сети независимых и зависимых переменных.

Данная модель являет собой более новый и точный объяснительный принцип, каким до недавнего времени являлся факторный анализ.



Пакет R

<http://www.r-project.org/>

Проект R является ответом сообщества свободно распространяемого ПО на создание коммерческих версий статистических пакетов, в частности S-plus.

Базовая версия R включает в себя только основные модели статистического анализа. Кроме того, базовая версия требует от пользователя пакета определенных знаний и навыков в области программного языка R.

В то же время, подключаемые к базовой программе модули позволяют значительно расширить ее функциональность: добавить графический кнопочный интерфейс, модули визуализации и многомерных статистических моделей и др.

Задания

I. Подготовка презентации (итоги)

Презентация (формат ppt, pps и pdf)

II. Отчет по занятиям (по задачам)

1. Собственный интернет-ресурс
2. Работа в MS Excel
3. Работа в SPSS
4. Работа в R (опционально)