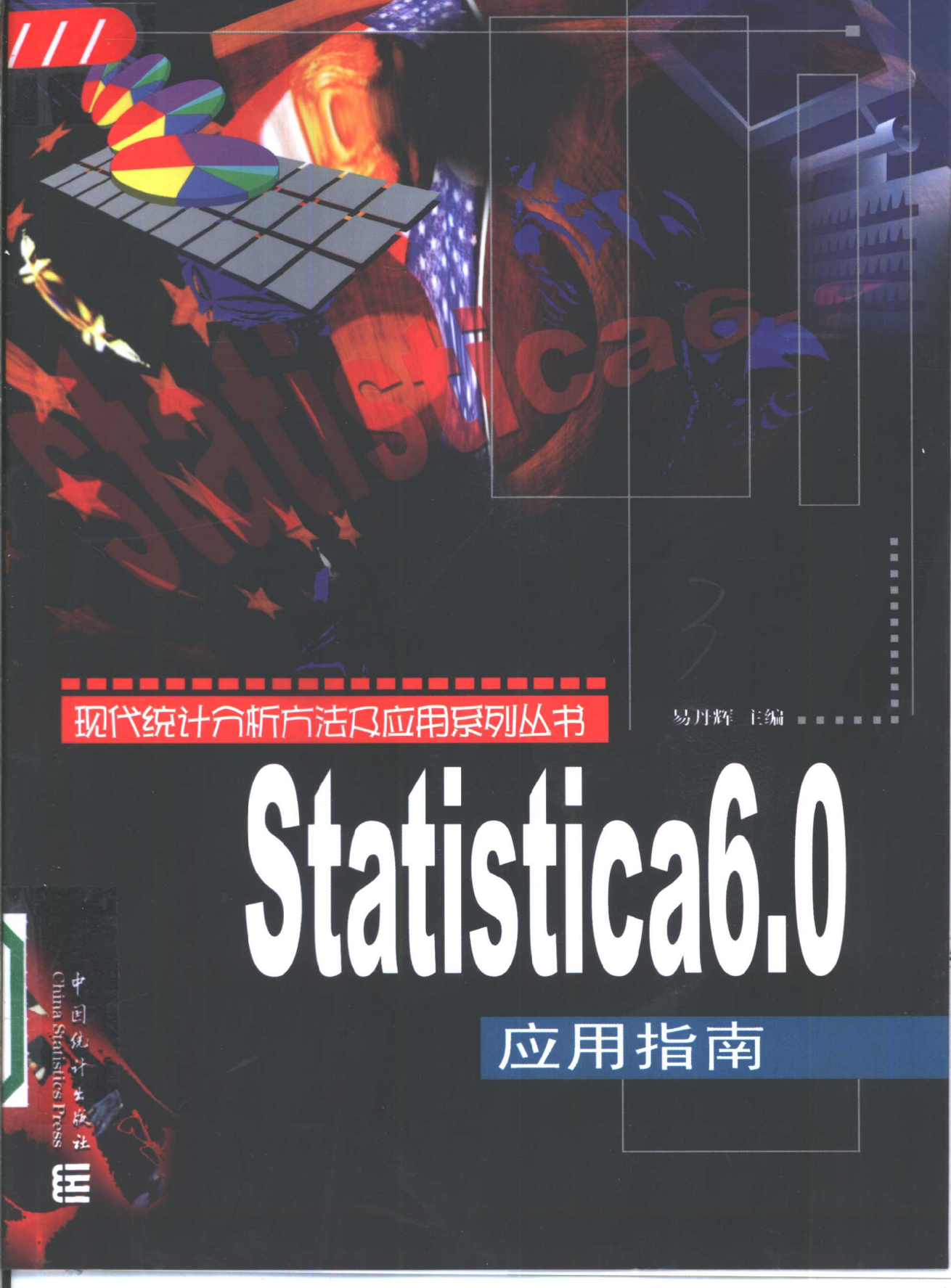




现代统计分析方法及应用系列丛书

易丹辉 主编

Statistica6.0

应用指南

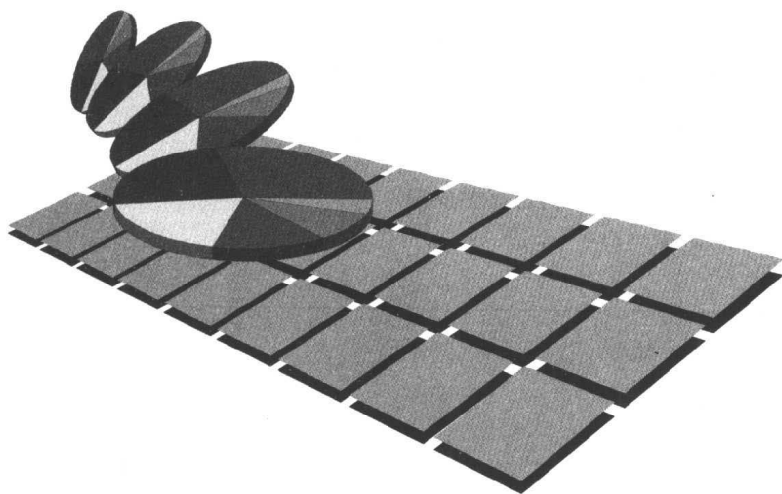
中国统计出版社
China Statistics Press



现代统计分析方法及应用系列丛书

Statistica6.0 应用指南

易丹辉主编



中国统计出版社
China Statistics Press



(京)新登字 041 号

图书在版编目(CIP)数据

Statistica 6.0 应用指南 / 易丹辉编著.

—北京: 中国统计出版社, 2002.10

(统计分析方法与应用丛书)

ISBN 7-5037-3935-5

I.S... II. 易... III. 统计分析—应用软件,

Statistica 6.0—指南 IV. C819-62

中国版本图书馆 CIP 数据核字 (2002) 第 082380 号

责任编辑 / 吕 军 徐 颖

设 计 / 刘国宁 张建民

出版发行 / 中国统计出版社

通信地址 / 北京市西城区月坛南街75号 邮政编码 / 100826

办公地址 / 北京市丰台区西三环南路甲6号

电 话 / (010)63459084、63266600-22500 (发行部)

印 刷 / 北京市顺义兴华印刷厂

经 销 / 新华书店

开 本 / 787 × 1092mm 1/18

字 数 / 380 千字

印 张 / 20

印 数 / 1-3000 册

版 别 / 2002 年 10 月第 1 版

版 次 / 2002 年 10 月第 1 次印刷

书 号 / ISBN 7-5037-3935-5/C · 2003

定 价 / 39.00 元

中国统计版图书, 版权所有, 侵权必究。

中国统计版图书, 如有印装错误, 本社发行部负责调换。

前 言

StatSoft 公司, 1984 年成立于美国俄克拉荷马州的土桑 (Tulsa, Oklahoma)。成立初期, 公司推出的产品 Psychostat-2 和 Psychostat-3 以易于操作、富有弹性的特点, 获得学术界和大众的喜爱; 后来经过不断研发和改良, 1988 年推出 STATISTICA 0 版, 同时于德国汉堡成立第一家 StatSoft 国外分公司—StatSoft Europe, 并将所有的技术投入于 STATISTICA 系列的研发工作。1999 年, STATISTICA 5.5 版以高质量的统计图表及最佳功能操控组合问世, 目前已成为市场上拥有超强统计功能的统计软件之一。目前 STATISTICA 6.0 版已经开始发售, 并以强大的统计功能、简易的操作方式和丰富多彩的图形编辑功能取得不俗业绩。目前, StatSoft 在全球已成立了 19 家国外或地区分公司, 包括英国、法国、德国、意大利、波兰、葡萄牙、比利时、南非、捷克共和国、西班牙、斯堪地那维亚半岛 (挪威、瑞典)、俄国、日本、韩国、澳洲、巴西、以色列、东南亚国协 (新加坡、马来西亚、泰国、菲律宾与印尼) 以及台湾。各国或地区分公司负责推广销售 STATISTICA, 提供完善的咨询服务和训练课程。StatSoft 大中华地区 STATISTICA 产品使用者的技术支持及服务工作由台湾分公司 (即台湾统软顾问有限公司) 负责, 同时 StatSoft 还在深圳成立了办事处, 更好地服务于大陆地区的 STATISTICA 产品使用者。

1.1 本书读者剪影

本书读者群是那些手头有一些数字资料, 希望利用统计学知识从中挖掘出一些有用信息的人。这些人懂一些计算机基本操作和统计学

的基础知识，并且有 STATISTICA 软件，但对统计学以及 STATISTICA 的具体应用又不太清楚。

一般常用的统计方法其实并不太困难，当然也并不很容易。本书每一章节首先会简要回顾涉及到的统计学知识，主要内容则侧重于讲述如何用 STATISTICA 来实现这些统计功能。读者不要期望读了本书后就能「恢复求学时的功力」，因为不管怎么说，统计学还是一门专业功课。

1.2 统计软件的作用

统计软件扮演的角色，其实与计算器、汽车、高速公路等一样，它只是工具或者手段，它代替你做「大部分」的苦工，并且极少犯错，犯错也多半是使用者自己的错。软件的使用者只要注意几件事：

- (1) 输入的资料是否有误；
- (2) 所用的方法是否正确；
- (3) 解读计算机计算的结果。

在这里，人和计算机有明确的分工。计算机会做的部分留给软件去执行；而我们因为计算机做了大量且不会算错的计算，便可以将心力更有效地用在专业思考上。计算机并未将原来的问题变得浅显，它只是让使用者腾出功夫来想更深的问题。

STATISTICA 是一个标准统计软件，与其他的统计软件相比，STATISTICA 最大优点在于其菜单式的指令，操作相当方便，同时 STATISTICA 的图也画得比较细腻，因为这是针对视窗系统发展的软件。别的统计软件并非不用 Windows；但还不象 STATISTICA 这样几乎每一个指令都用鼠标点选。统计软件是给在公司业务或一般研究中可能碰到数据分析问题的人员使用的，真正大的问题还需要另行改写。这里强调的是本质，本文认为统计软件的作用在于帮助人们解决分析数据的需求：

需求→工具→满足需求

统计软件全套想法是：

Data → Software → (useful) Information

这与软件使用者的专业知识有很大关系。软件只是帮助处理机械式的运算，又快、又精确，画一个图只要一眨眼的时间，而且要画几个图，就可以画几个图。但是除非使用者能够读懂结果，了解其意义，否则很难将它翻译成有用信息。同样的「加减乘除」，在店员手上或者在会计师手上算出的结果就颇有层面深浅上的不同。软件不能取代功力，它的功用是替你省下大量做「加减乘除」的时间，让你腾出空来想更深一层的问题。

1.3 STATISTICA产品系列

目前，几乎全世界各地的主要大学、公司及政府机构都使用 StatSoft 的产品。STATISTICA 于 1993 年推出视窗版本，是 StatSoft 的主要产品系列，该版本获得了使用者及评论家的赞赏与好评。至目前为止，STATISTICA 在每一个独立的比较评论中，都获得了最高的评价。并且在所有全新 STATISTICA 系列产品中支持图表转换 HTML 档案功能，使用者可通过网页浏览器浏览。

STATISTICA 产品包括下列产品系列：Quick STATISTICA、STATISTICA、STATISTICA QC Charts、STATISTICA Process Analysis、STATISTICA DOE、STATISTICA Neural Networks、STATISTICA Power Analysis、STATISTICA Data Mining、STATISTICA Student Version 等，不同产品包含不同的功能，有些产品如 Quick STATISTICA、STATISTICA 等可以独立发售、安装与使用，有些产品如 STATISTICA Power Analysis 等则需要配合其他产品才能安装使用，使用者可根据自己的需要选择适当模块。

1.4 本书的使用

本书内容以 STATISTICA 6.0 版为例讲解，6.0 版与前几个版本比

较，在结构方面有较大改变，因此其他版本使用者参考时要加以注意。本书前两章讲述了STATISTICA的基本概念与基本操作，学习这两章后读者可以掌握使用该软件进行统计分析的基本方法以及对软件状态进行个性化设置的技巧，STATISTICA软件的老用户可以略过这一部分。以后八章主要介绍各种统计分析方法的实现。最后一章，简介STATISTICA较高级的应用，即STATISTICA的BASIC语言。因此对STATISTICA软件不很熟悉的用户，可以首先学习前两章，掌握软件基本使用方法后，再有针对性地阅读后面各章，在做具体工作时翻阅相应的章节。

1.5 致谢

本书的写作得到了国家统计局统计教育中心徐加林等老师的鼓励与帮助。本书共分十章，其中第三章、第四章主要由王静完成，第七章主要由黄旭安完成，第八章主要由景昕完成，第九章由王燕完成。全书的规划、整理以及其它各章节的写作由谢邦昌、易丹辉和戴稳胜共同完成。在此谨对国家统计局统计教育中心徐加林等老师以及王静、黄旭安、景昕、王燕表示感谢，并感谢统计出版社责编的辛勤工作。

本书是大陆地区STATISTICA的第一本教材，由于时间紧张等因素，书中的不当之处在所难免。我们热忱欢迎读者提出批评与建议，不止限于对本书的斧正，也欢迎其他的意见建议。来函请发往：wensdai@rucstat.com或wensdai@163.com，作者将不胜感激。

作者

2002年9月

中国人民大学统计学系数据挖掘中心

目 录

第一章 STATISTICA 数据文件及其管理

- 第一节 STATISTICA 简介.....(1)
- 第二节 STATISTICA 数据文件及其编辑.....(4)
- 第三节 STATISTICA 统计分析前：数据文件的预处理.....(13)

第二章 STATISTICA 工 具 栏

- 第一节 数据电子表格工具栏.....(33)
- 第二节 STATISTICA 的图表、结果工具及结果图表的定制.....(39)

第三章 基本统计分析

- 第一节 基本统计分析概述.....(84)
- 第二节 描述统计量分析 (Descriptive Statistics)(84)
- 第三节 相关分析 (Correlation Matrices)(90)
- 第四节 Frequency Tables 频数分析.....(94)
- 第五节 列联分析 (Tables and Banners)(101)
- 第六节 Multiple Response Tables 多应答分析.....(108)

第四章 均值比较与检验

- 第一节 均值比较的概念.....(116)
- 第二节 单一样本 t 检验.....(116)
- 第三节 独立样本 t 检验.....(121)
- 第四节 相依样本 t 检验.....(128)

第五章 方差分析简介

- 第一节 方差分析的基本概念.....(132)
- 第二节 BREAKDOWN AND ONE-WAY ANOVA.....(133)
- 第三节 重复测量单因子方差分析.....(137)

第六章 回归分析

第一节 一元回归	(146)
第二节 多元回归	(152)

第七章 多元统计分析

第一节 聚类分析	(163)
第二节 因子分析	(173)
第三节 主成分和分类分析	(183)
第四节 典型相关分析	(190)
第五节 可靠度和项目分析	(198)
第六节 分类树	(207)
第七节 对应分析	(220)
第八节 多维标度分析	(235)
第九节 判别分析	(243)
第十节 广义判别分析	(252)

第八章 非参数检验

第一节 单样本非参数检验	(263)
第二节 两样本的非参数检验	(267)
第三节 多个样本的非参数检验	(277)

第九章 生存分析

第一节 引言	(287)
第二节 非参数方法	(289)
第三节 参数方法	(303)
第四节 假设检验	(312)
第五节 半参数方法	(330)

第十章 STATISTICA BASIC 语言简介

第一节 什么是 STATISTICA BASIC 语言?	(346)
第二节 SVB 示例	(349)

1

STATISTICA数据 文件及其管理

第一节 STATISTICA 简介

§ 1.1.1 STATISTICA 的模块组织

STATISTICA 是一个优秀的标准统计软件，其简单的菜单式操作、细腻的图形化结果呈现等优良特性为它赢得了世界范围内大量的用户。

STATISTICA 采用模块化设计，在 5.5 及以前版本中，STATISTICA 通过一个“模块呼叫器”调用不同模块。在 6.0 版本中，STATISTICA 将不同模块均置于“Statistics”菜单下，每个模块包含一连串该模块可应用到的统计分析程序。通过 **Statistics** 下拉菜单的选取，使用者可以根据目前数据的内容与需求，选取其中一个或一个以上的模块进行数据的统计分析任务。

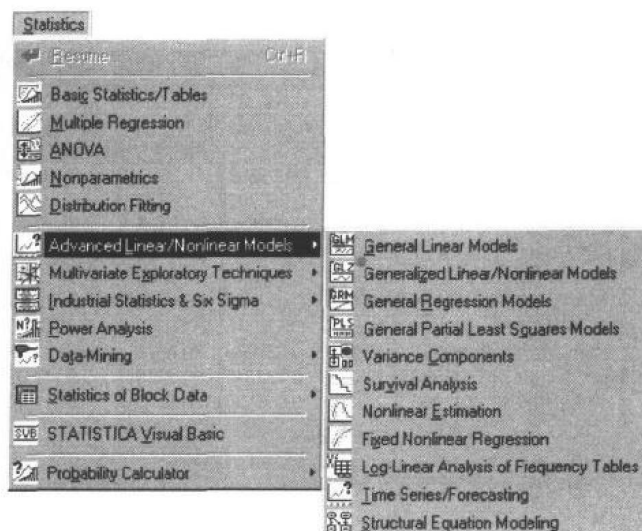


图 1.1.1 统计模块菜单图

使用者可以同时开启几个 STATISTICA 模块，每一个模块可以用于执行相同或不同的分析目的。STATISTICA 功能更强的是，使用者可以在同一个 STATISTICA 应用环境中，执行多个统计分析工作。这些工作任务可以相同也可以不同（比如执行三次多元回归与两次变异数分析），并且每一个分析模块都可以利用相同或不同的输入数据文件。

§ 1.1.2 STATISTICA 输出结果的组织

STATISTICA 对统计分析的结果提供三种不同的输出模式，以三种不同方式查看这些输出内容。这三种模式分别为工作簿（workbooks）、报告（reports）与独立查看窗口（stand-alone windows）。使用者可以通过过 **Tools** 下拉窗体中的 **Options** 选项进入，并通过 **Output Manager** 按钮修改输出内容的呈现方式。具体内容详见“结果表格的订制”一节。下面简单介绍一下结果输出三种模式。

工作簿（WORKBOOKS）

工作簿是 STATISTICA 隐含的输出模式，每一个输出的结果都会以两个标签展示并存储在工作簿中，其中一个标签位于工作簿左侧，形如一般资源管理器中的文件夹，这些输出文件被组织成阶层式的目录，成为一个树状结构组织图，每一份文件为单一分析任务执行后的产出结果，并储存在根目录下的个别子目录中。另一个标签位于右下，形如卡片，点击每张卡片，就会展示一个统

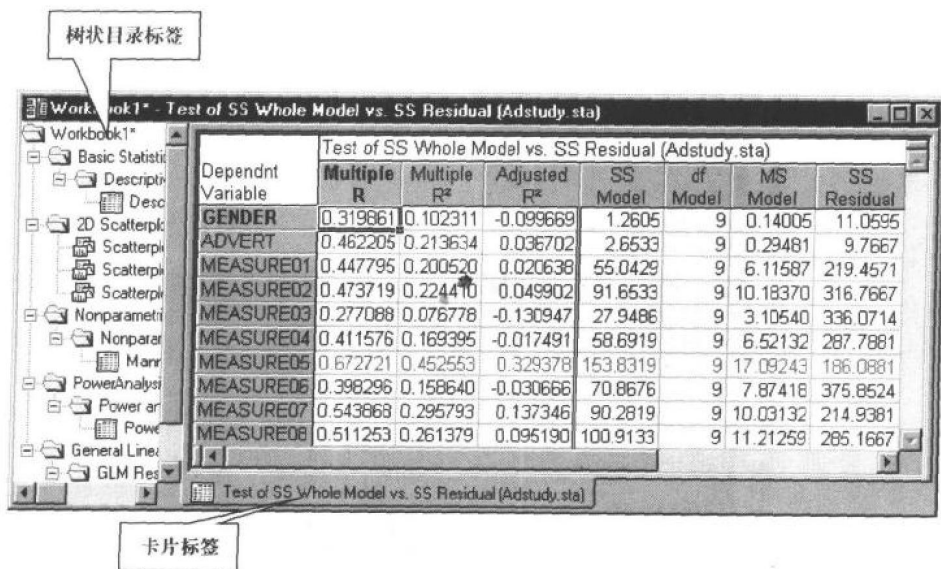


图 1.1.2 结果工作簿图

计分析的结果。

举例来说, 你可以利用复制文件或移动文件的方式, 将工作簿中的部分文件萃取至报告窗口或者独立的窗口中。你也可以将整个树状结构中的特定树干 (branches), 加以复制、移动并重新组织一个新的特定目的之组织结构图。另外, 结果工作簿还有一个特性, 它可以将整个数据表格视为结果工作簿中的一部分, 输出到结果中。

报告 (REPORTS)

STATISTICA 报告提供一个比较传统的输出内容管理方式, 每一个输出文件以文字编辑器一样的文件格式呈现在窗口中。这种文件格式以及它的应用环境都是以文字编辑器作业环境为基础的, 其优点在于使用者可以利用窗口的滚动轴, 随时对特定的表格或图表文件进行编辑, 比如插入注记 (note) 或说明。

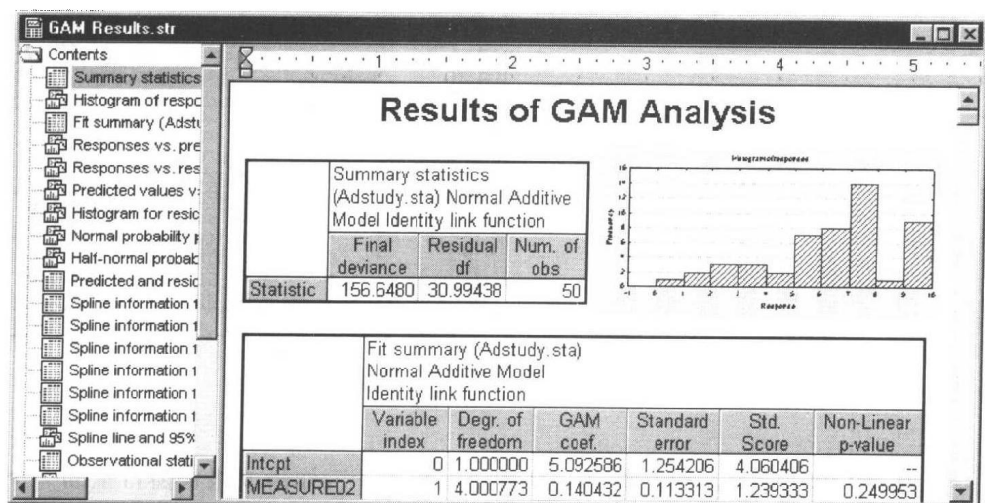


图 1.1.3 报告编辑窗口

报告输出模式包括输出文件的补充信息, 提供一个分析过程中众多选项条件详细纪录。每个报告输出相当于工作簿里的一个文件夹, 内容为一个分析过程所有的结果。

独立窗口 (STAND-ALONE WINDOWS)

STATISTICA 输出文件也可以直接显示为一连串的单窗口, 就象 5.5 以前版本展示的那样。使用者可以轻易地自定义排列这些 STATISTICA 应用环境的作业接口; 对于查看新的输出结果而言, 以这种输出方式建立参考文件 (reference documents) 相当有用。

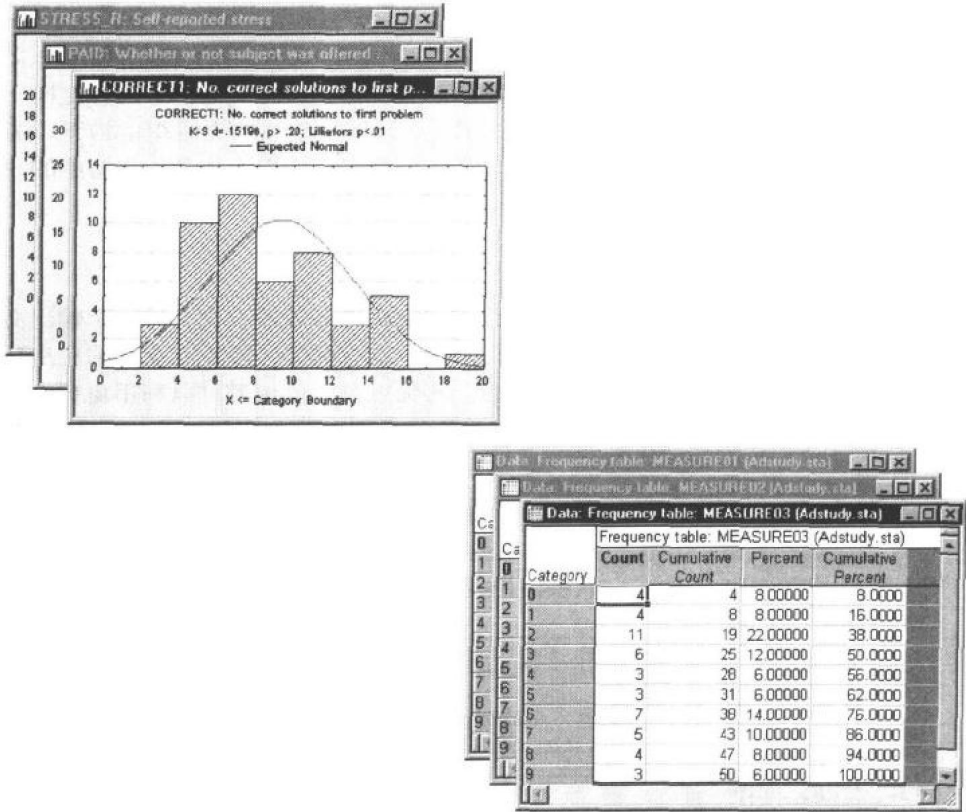


图 1.1.4 独立输出窗口模式

如图 1.1.4 所示，独立输出窗口实际上相当于工作簿或报告窗口里的一个单独文件，是一次分析过程中选择某一选项条件下的分析结果。

第二节 STATISTICA 数据文件及其编辑

STATISTICA 运行时，屏幕上显示数据文件编辑窗口，称为 spreadsheet。如图 1.2.1 所示，用户可以在此窗口上建立与编辑数据文件，下面简要介绍 STATISTICA 数据文件编辑窗口的组成及文件的编辑。

§ 1.2.1 数据文件编辑窗口 SPREADSHEET 的组成

标题栏

在数据文件编辑窗口的最上端是标题栏，标题栏中显示目前正在编辑或展示、处理的数据文件的名称、该数据文件的变量和记录数目。如下图所示，该

August 10th	1 GENDER	2 ADVERT	3 MEASURE01	4 MEASURE02	5 MEASURE03	6 MEASURE04	7 MEASURE05	8 MEASURE06	9 MEASURE07
R. Rafuse	MALE	PEPSI	9	1	6	8	1	2	
T. Laker	MALE	COKE	6	7	1	8	0	0	
E. Bizot	FEMALE	COKE	9	8	2	9	8	8	
K. French	MALE	PEPSI	7	9	0	5	9	9	
E. Van Landuyt	MALE	PEPSI	7	1	6	2	8	9	
K. Harrell	FEMALE	COKE	6	0	0	8	3	1	
W. Noren	FEMALE	COKE	7	4	3	2	5	7	
W. Willden	MALE	PEPSI	9	9	2	6	6	8	
S. Kohut	FEMALE	PEPSI	7	8	2	3	6	9	
B. Madden	MALE	PEPSI	6	6	2	8	3	6	
M. Bowling	FEMALE	PEPSI	4	6	6	5	6	8	
J. Wilcoxson	MALE	COKE	7	3	3	7	0	6	
J. Landrum	MALE	PEPSI	6	2	3	1	8	1	
M. Taylor	MALE	COKE	7	2	4	8	1	2	
N.S. Madden	FEMALE	PEPSI	6	2	7	5	7	2	
K. Ridgway	FEMALE	PEPSI	3	2	5	4	4	4	
L. Cunha	MALE	COKE	2	9	9	3	1	4	
F. Wind	FEMALE	PEPSI	1	0	7	5	2	4	

图 1.2.1 数据编辑窗口

数据文件名称为 *Adstudy.sta*，包含 25 个变量与 50 笔记录。

信息栏

标题栏下方是一个白色区域字段，这个字段用以对数据文件的说明进行编辑，使用者可以在这个信息栏内标明该数据文件的内容说明。编辑时只需要将鼠标光标移到这个字段上，双击左键即可。

附加信息框

数据表额外的信息，如果未在信息栏内标明，可以在 spreadsheet 左上方的 **Info Box** 中输入。同样，使用者也可利用鼠标左键双击该字段进行编辑，在上图示例中，**Info Box** 的内容为 *August 10th*。

变量与记录

STATISTICA 数据文件依据变量与记录来组织，本书中，将一个观察值称为一个记录，而每个记录所涉及的指标称为变量。Spreadsheet 的每一列表示一个变量，而每一行表示一个观测值，即一个记录。

变量的名称在数据文件上方的灰色单元格中由左到右标示，隐含名称为 *Var1*, *Var2*, ……。记录名称则标示在文件左侧的灰色格子中（由上至下）。在上图的例子中，第一个变量的名称为 *GENDER*，第一笔记录的名称为 *R. Rafuse*。

§ 1.2.2 STATISTICA 数据文件的建立与管理

数据文件的建立

如前所述，STATISTICA 数据通过变量和记录来组织。建立数据文件的步

骤，可以通过菜单项：file → New，或点击快捷键□进行。作上述操作后，会出现如图 1.2.2 所示的对话框，显示有四个选项卡，分别是 Macro(SVB) Program、Report、Workbook、Spreadsheet。

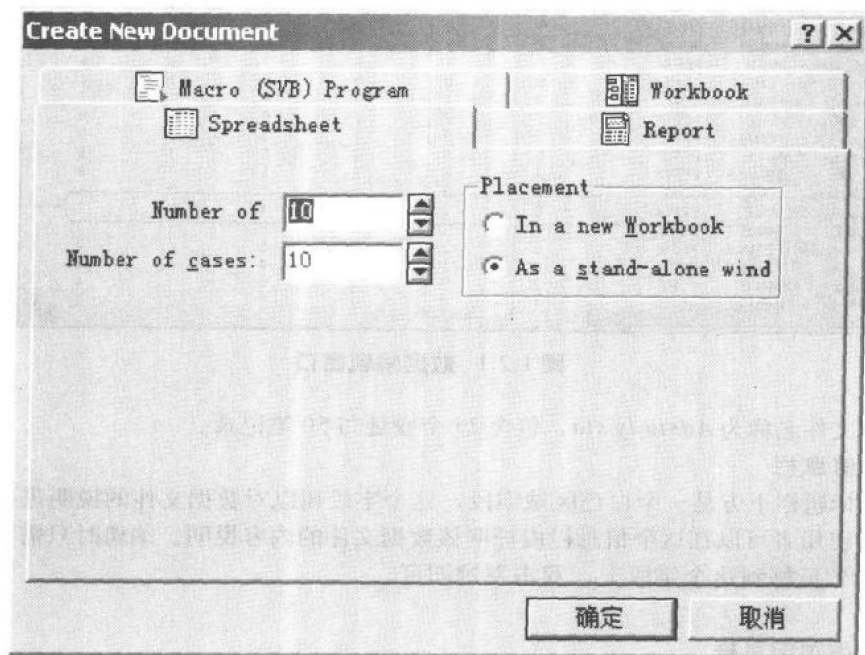


图 1.2.2 新建文件对话框

其中 Workbook、Report 上文已作简介，分别是一种输出文件的格式。Macro (SVB)Program 是新建一个 STATISTICA BASIC 程序，该部分内容将于最后一章讲解。

我们要建立的是一个数据文件，因此选择 spreadsheet。在本选项卡上有三项内容由使用者决定，其中 Number of Var 和 Number of Case 是让使用者定义该数据文件的变量个数和记录个数。实际上，STATISTICA 的 spreadsheet 的形式与操作与常见电子表格 Excel 的外形和操作方法相差无几，其中最大的差别就在这里：Excel 表格可以无限制地任意添加列和行的内容，而 spreadsheet 表格中，你只能按已定义的变量个数和记录个数进行编辑，如果有新增变量或记录的情况，只能先通过变量、记录编辑器增加空的变量、记录后才能进行，具体增加办法见后述。

新建 spreadsheet 选项卡右侧还有一个选择，即 Placement: 1, In a new workbook; 2, As a stand-alone window。前者表示将新建的数据表格作为一个新工作簿的部分；后者表示将新建数据表格作为一单独窗口。

变量与记录的操作

新建数据表格后，需要对变量与记录进行操作以定义合适的变量名称、类型、格式以及定义记录名称等。

变量与记录可以通过几种方式进行修改，例如新增 (add) 变量、移动 (move) 记录、重新计算 (recalculate) 变量等等。大多数的变量与记录修改选项，可通过  和  工具栏下按钮执行。

变量 (VARIABLES) 格式及其管理

查看与编辑变量格式

每一个变量都有一组特定的属性或规格，查看与编辑变量的格式有两种方式。

第一种方式，选择某一变量，并从 **Vars** 工具栏按钮中选取 **Specs...**，打开 **Variable** 规格对话框。或者，选择某个变量，双击该变量标题栏，也可以打开变量规格对话框。如图 1.2.3 所示。

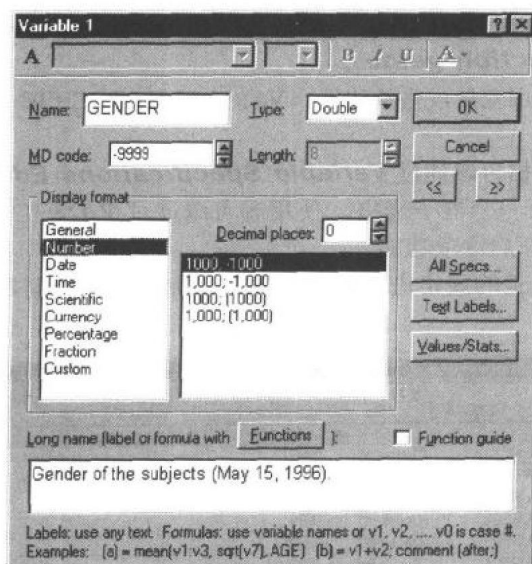


图 1.2.3 变量规格对话框

该对话框中各名称及按钮的含义分别如下：

Name 文本框为该变量的变量名称，在电子表格每一列的表头显示。

Type 文本框为该变量的数据类型，有四种分别为：**Double**，是 STATISTICA 中针对数值储存的预设类型，每一笔数值可以有一个相对应的文字值卷标予以注记；**Integer**，用于数据为数字类型，使用者不可以在这种类型的变量中输入包含小数点在内的数值，且每个数字值可以有相对应的文字值卷标同时注记；

Byte, 为最长限制为 255 位的整数 (integers) 数据类型, 使用者不可以在这种类型的变量中输入包含小数点在内的数值, 且每个数字值可以有相对应的文字值卷标同时注记, 此种数据类型的优点, 在于它是一种最经济的变量数据值储存方式; **Text** 用于文字型数据。

MD code 文本框用于指定缺失值的编码工作, 或者也可以用于指定某一特定值, 将该值排除在本次分析工作之外不参加所作分析。

Length 框只在使用者选取变量的数据属性为 **Text** 时才可以获得, 用于定制文本最大长度。

Display format 选择框下, 提供使用者选取变量在数据文件中显示的格式。当选定某一特定类型格式后, 右侧的窗体将展示并提供相关的详细格式设立条件。

Decimal places 文本框只有在变量格式设定为 **Number**、**Scientific**、**Currency** 或 **Percentage** 时才可用, 用于设定电子表格中变量数据的小数点后的位数。

Long name 框提供使用者对变量作标记, 以描述说明该变量, 并在统计结果输出打印时同时打印出来, 另外也可以用于制定电子表格中变量间的公式计算 (可参考 **Function guide**)。

第二种方式, 用变量规格编辑器 (VARIABLE SPECIFICATIONS EDITOR) 展示与编辑。

所有变量的规格可以通过 **Variable Specifications Editor** 加以预览, 这个编辑器看起来象一个电子表格, 使用者可以通过 **Vars** 工具列窗体中的 **All Specs** 选项选取开启这项功能。

同时对几个变量的规格进行定制或预览, 尤其当需要复制或粘贴其它变量

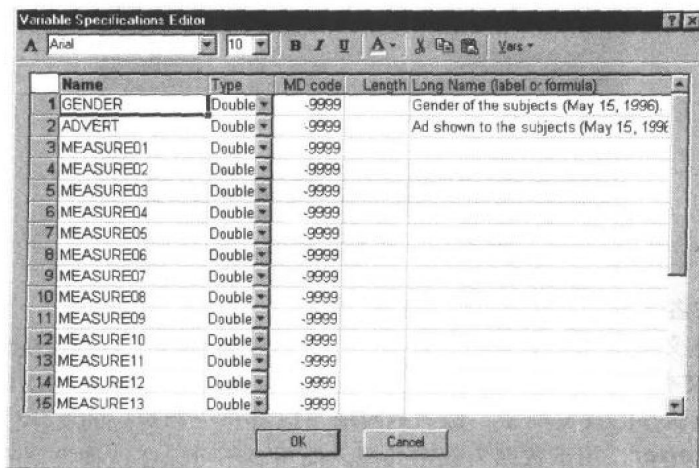


图 1.2.4 变量规格编辑器