

一、统计计算工具

统计计算工具是我们完成统计分析的必要条件, 算盘、计算器、计算机这些工具为我们完成大量的数据处理立下的汗马功劳。随着计算机软硬件的飞速发展, 对数据的处理与分析基本倾向于用统计分析软件, 目前常用统计分析软件有: SAS、SPSS、STATISTICA、MINITAB 等等, 每个统计分析软件都有各自的组织数据的方式以及分析界面。80 年代以后, 电子表格成为数据组织形式的主流, 著名的电子表格 LOOTS 1-2-3 和 Microsoft Excel 风靡一时。本教材使用 Excel 和 SPSS 作为数据处理的主要工具。

(一) 统计分析软件简介

1. SAS 系统全称为 Statistics Analysis System, 最早由北卡罗来纳大学的两位生物统计学研究生编制, 并于 1976 年成立了 SAS 软件研究所, 正式推出了 SAS 软件。SAS 早期的主要功能是统计分析, 如今 SAS 打出的标牌是“Superior software that gives you The Power to know”(卓越的软件带给你获取知识的力量), 其产品与解决方案除统计分析外, 新增了数据整合、企业智能等, 统计分析功能也在不断增加。SAS 用户遍及金融、医药卫生、生产、运输、通讯、政府和教育科研等领域。在英美等国, 能熟练使用 SAS 进行统计分析是许多公司和科研机构选材的条件之一。在数据处理和统计分析领域, SAS 系统被誉为国际上的标准软件系统。SAS 统计系统是由多个功能模块组合的, 其基本部分是 BASE SAS 模块, 它是 SAS 统计分析系统的核心, 承担着主要的数据管理任务, 并管理用户使用环境, 进行用户语言的处理, 调用其他 SAS 模块和产品。在 BASE SAS 的基础上, 可以增加模块而增强数据分析功能: SAS/STAT(统计分析模块)、SAS/GRAPH(绘图模块)、SAS/QC(质量控制模块)、SAS/ETS(经济计量学和时间序列分析模块)、SAS/OR(运筹学模块)、SAS/IML(交互式矩阵程序设计语言模块)、SAS/FSP(快速数据处理的交互式菜单系统模块)、SAS/AF(交互式全屏幕软件应用系统模块)等等。一般认为使用 SAS 需要编写程序, 比较适合统计专业人员, 而对非统计专业人员则比较困难。事实上 SAS 的很多功能也可以通过菜单操作实现, 如 Insight 就是一个菜单操作的界面。

网络链接: SAS 研究所 http://www.sas.com/

2. SPSS 是软件英文名称的首字母缩写, 原意为 Statistical Package for the Social Sciences, 即“社会科学统计软件包”, 2000 年 SPSS 公司将英文全称更改为 Statistical Product and Service Solutions, 意为“统计产品与服务解决方案”, 标志着 SPSS 的战略方向做出了重大调整。SPSS 公司成立于 1968 年, 现在的客户超过了 25 万, 该公司陆续购并了 SYSTAT 公司、BMDP 软件公司等, 逐渐由原来的单一统计产品开发与销售转向企业、教育科研及政府机构提供全面信息统计决策支持服务, 成为走在了最新流行的“数据仓库”和“数据挖掘”领域前沿的一家综合统计软件公司。SPSS 最突出的特点就是操作界面极为友好, 采用类似 EXCEL 表格的方式输入与管理数据, 数据接口通用, 能方便的从其他数据库中读入数据。2009 年 SPSS 重新包装旗下产品线, 定位为预测统计分析软件(Predictive Analytics Software)。它包括 PASW Statistics 统计分析、PASW Modeler 数据挖掘、Data

Collection family 数据收集和 PASW Collaboration and Deployment Services 企业应用服务四部分。2010 年 SPSS 公司被 IBM 公司并购，各产品家族名称前面统一加上 IBM SPSS 字样。2011 年推出 IBM SPSS Statistics 20。

网络链接： SPSS 公司 <http://www-01.ibm.com/software/analytics/spss/>

3. Statistica 是由美国俄克拉荷马州的 StatSoft 公司研制的大型专业统计图表分析软件包。Statistica 数据统计分析项目有 15 个主命令，130 多个子命令，其功能是：基本统计分析（Basic Statistics）、非参数统计分析（Nonparametrics）、方差分析（General ANOVA）、多元回归分析（Multiple Regression）、非线性估计（Nonlinear Estimation）、时间序列预测（Time Series / Forecasting）、聚类分析（Cluster Analysis）、因子分析（Factor Analysis）、典型分析（Canonical Analysis）、多维尺度分析（Multidimensional Scaling）、路径分析（SEPATH）、可靠性/项目分析（Reliability/Item Analysis）、判别分析（Discriminant Analysis）、对数线性分析（Log-linear-analysis）和生存分析（Survival Analysis）。Statistica 的图形功能很完备，显示输出的图形细腻美观，有 13 个主命令，80 多个子命令，主要包括：快速统计图（Quick Stats Graphs）、二维统计图（Stats 2D Graphs）、三维序列统计图（Stats 3D Sequential Graphs）、三维 XYZ 统计图（Stats 3D XYZ Graphs）和统计矩阵图（Stats Matrix Graphs）等。

网络链接： StatSoft 公司 <http://www.statsoft.com/>

4. MINITAB 是由美国宾夕法尼亚州立大学在 1972 年研制的统计分析软件包，它以无与伦比的易学性、可靠性以及完善的功能而著名于世，是教授统计学，实施六西格玛和其他质量改进项目的理想选择，它具备以下特征：与人们学习和工作方式相适应的逻辑界面；包含完整的数据管理功能、强大的文件导入和导出、数据操作和电子表格格式的数据窗口；配备详尽的文档，例如帮助解释输出结果的 StatGuide 以及 500 多个图文并茂的术语。MINITAB 包括统计分析模块和绘制图形模块，统计分析模块有 13 个主命令：基本统计分析（basic statistics）、回归分析（regression）、方差分析（ANOVA）、实验设计（DOE, Design of Experiments）、控制图（control charts）、质量编制计划工具（Quality Tools）、可靠性/生存分析（reliability/survival）、多变量分析（Multivariate）、时间序列分析（Time series）、统计报表和列联表检验（Tables）、非参数检验（Nonparametrics）、探索性数据分析（EDA, exploratory data analysis）、效能与样本量分析（power and sample size）。MINITAB 与六西格玛解决方案紧密结合，包括通用电气、福特汽车等在内的上千家杰出跨国公司以及 4000 多所学院与大学都在使用该统计分析软件包。

网络链接： Minitab 公司 <http://www.minitab.com/>

5. 马克威分析系统是由上海天律信息技术有限公司开发的中国第一套完全自主知识产权的大型统计分析和数据挖掘系统。马克威分析系统用于从海量信息和数据中寻找规律和知识，通过数据挖掘和统计分析等技术建立概念模型，为决策者提供科学的决策依据。它是一

套集分析、挖掘、预测、决策支持于一体的知识发现工具，适用于企业、政府、科研、教育、军队等单位 and 机构。马克威分析系统在技术上的特点是：将数据挖掘、统计分析、图形展示和智能报表融为一体，为用户提供完整配套的决策支持工具；提供独创的优化算法体系和完备的数据挖掘模型；将可视化数据分析与数据挖掘有机地融合在一起，并将自主开发的嵌入式数据库管理系统同其它关系型数据库实现了无缝连接；它在设计上充分考虑了中国用户的实际情况和使用习惯，将实用性和科学性结合在一起。2003 年，马克威分析系统被国家信息化测评中心选为中国信息化 500 强企业指定数据挖掘和信息分析软件。

网络链接： 上海天律信息技术有限公司 <http://www.tenly.com/>

6. R 是一套完整的数据处理、计算和制图软件系统，功能有数据存储和处理系统、数组运算、完整连贯的统计分析、优秀的统计制图、简便而强大的编程语言。**R** 是 **S** 语言的一种实现，**S** 语言 1980 年前后在统计领域广泛使用，它是由 AT&T 贝尔实验室开发的一种用来进行数据探索、统计分析、作图的解释型语言。**R** 是一个免费的自由软件，有 UNIX、LINUX、MacOS 和 WINDOWS 版本，CRAN（Comprehensive R Archive Network，R 综合典藏网的简称）收藏了 **R** 的执行档下载版、源代码和说明文件，收录了各种用户撰写的软件包。

网络链接： R 网站 <http://www.r-project.org/>

(二) Excel 实现数据处理的主要途径

我们使用 Excel 处理数据主要涉及两方面：一是 Excel 的公式与函数，二是 Excel 的数据分析工具。


1. 公式和函数

公式和函数是 Excel 工作表的核心，公式是连续的一组数据和运算符组成的序列，就象手工或计算器做运算那样工作；函数只要我们输入相应的参数，就会自动地计算出所需要的函数值。Excel 有灵活多变的公式和丰富多彩的函数。

Excel 提供了 435 个内部函数进行数学、财务、统计等计算工作。函数的基本格式为：

=SUM(A1:A5,C1:C3)

“SUM”是函数名称，它通过参数接受数据，参数要写在函数名字后面的括号内，“A1:A5”和“C1:C3”都是区域地址，两值之间需要用“，”分开。每个函数要求自己特定的参数类型，如数值、单元地址、文本或逻辑值等。极少数函数可以不要参数，但也不能省略括号，如 PI()。

对大多数函数，我们很难也没有必要记住它的语法，可以使用“函数向导”完成函数的输入过程。无论是在单元格内直接输入函数，还是在一个公式中包含函数，都有两种方式用以导出“函数向导”：①调用菜单“插入”—“函数”；②单击“常用”工具栏的“粘贴函数  ”按钮。

进入函数向导后弹出对话框，见图 1.1。“函数分类”栏告诉我们 Excel 将内部函数分为了十三类。

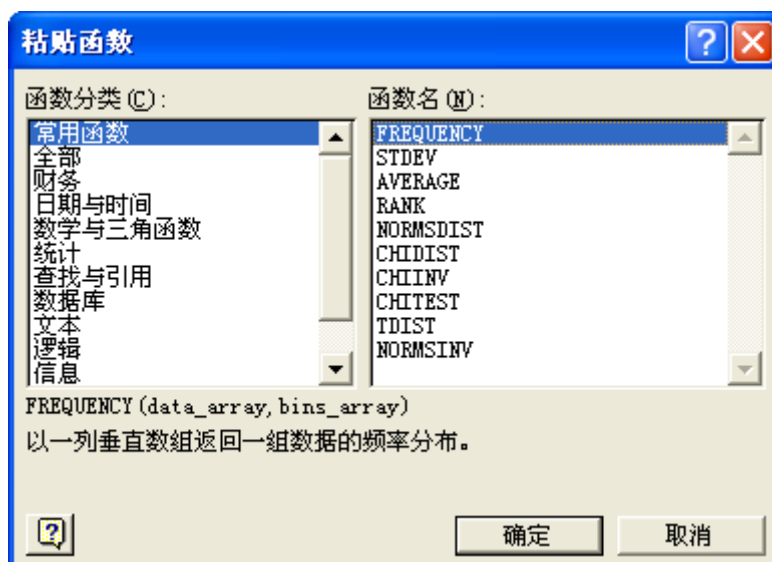


图 1.1 粘贴函数对话框

在“函数分类”中的选择了想使用的函数类别后，右边的“函数名”栏中就列出了该类函数中所有具体的函数。如果我们选中了“常用函数”中的“SUM”函数，就可单击“确定”按钮，进入下一个对话框，见图 1.2。在该框内填入各参数所需的数据（可以是常量、单元格或区域引用、名称等多种形式），每个参数都给予一定的提示。在对话框下方示出当前填入参数的计算结果。

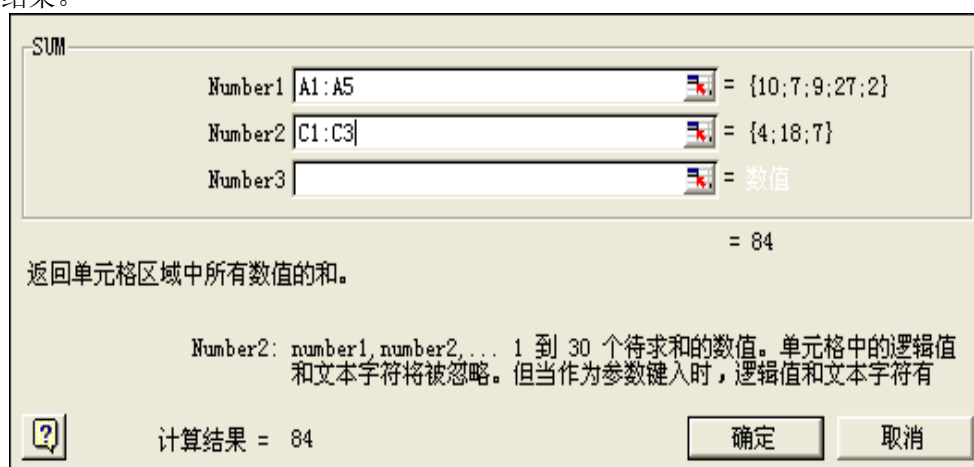


图 1.2 SUM 函数的参数对话框

通过“函数向导”录入函数的最大优点是步步有提示，只要明确想做什么，不必多虑如何做。

2. 数据分析工具

数据分析工具实际上是一个外部宏(程序)模块，它提供了 19 种专门用于数据分析的实用工具。在进行数据分析前，打开“工具”菜单，查看一下此菜单上有没有“数据分析”命令。如果没有，表明数据分析工具库尚未安装到正在运行的 Excel 中。这时需要选择菜单“工具”——“加载宏”，打开“加载宏”对话框，从其中的宏表中选定“分析工具库”宏。安装了数据分析工具库后，每次启动 Excel 时，“分析工具库”宏就被自动加载。调用“数据分析工具”的操作如

下:

(1) 选择菜单“工具”—“数据分析”，打开对话框，见图 1.3。

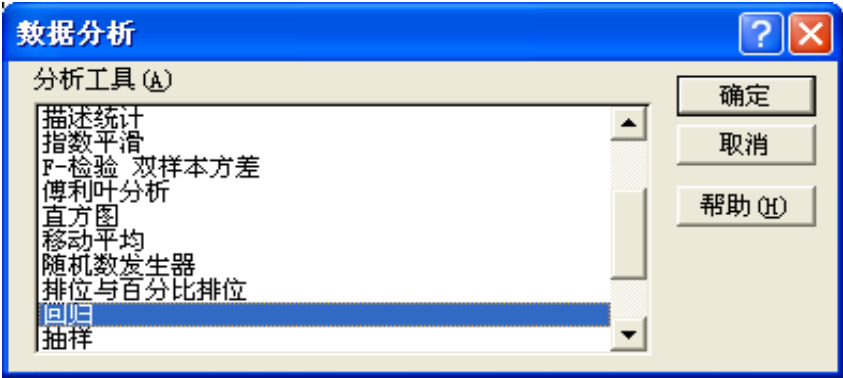


图 1.3 数据分析对话框

(2) 从“数据分析”对话框中分析工具列表中选择一种工具，打开相应的分析工具对话框。如，选择“回归”工具，打开回归分析工具对话框，见图 1.4。



图 1.4 “回归”分析工具对话框

(3) 按分析工具对话框的提示，将数据范围键入对话框，并设置各选项，然后单击“确定”按钮。

(三) SPSS 实现数据处理的主要途径

SPSS 可以实现数据处理的数据编辑与转换、表格与图形的生成和各种统计分析。

1. 数据编辑功能

数据编辑窗口，见图 1.5，在“变量视图”中定义变量的名称、类型、宽度、标签、值、缺失等等；在“数据视图”中根据对变量的定义进行数据输入，利用“数据”菜单，可以进行排序、转置、拆分、加权等操作，对多个数据文件可以对变量或样本进行合并。与 Excel 不同的是，变量名列示在列上，行对应的是变量值，我们常说的样本或 case。

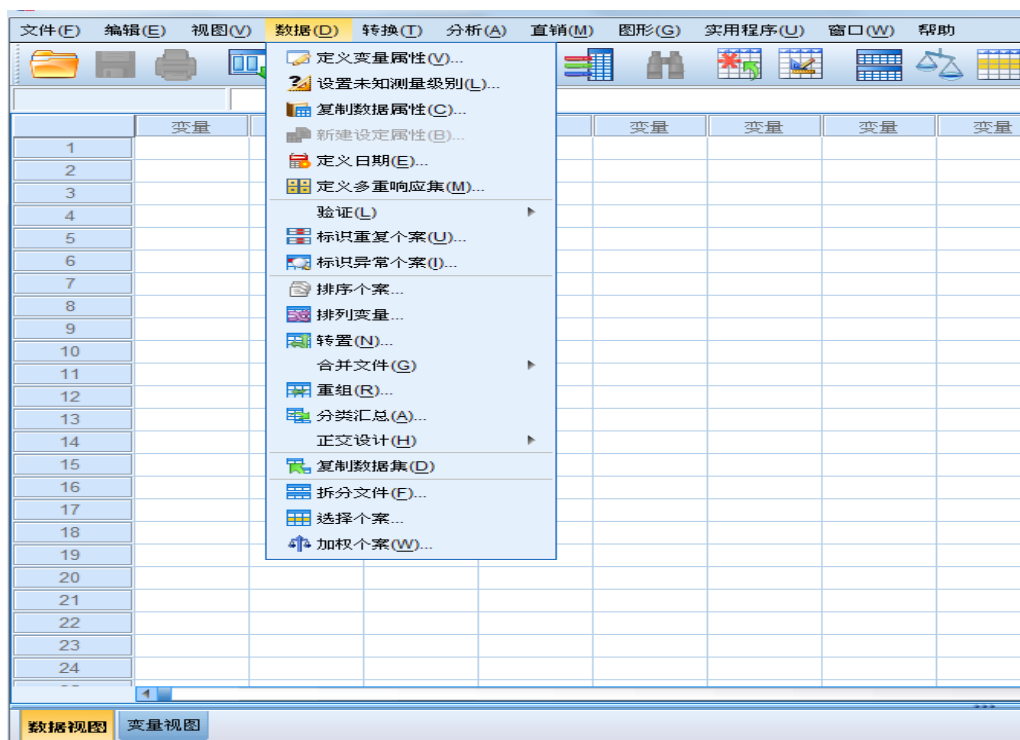


图 1.5 数据编辑窗口的“数据”菜单

2. 图表的生成功能

利用 SPSS 可以生成数十多种基本图和交互图。基本图包括条形图、线图、面积图、饼图、帕累托图、控制图、箱图、散点图、直方图、P-P 概率图、Q-Q 概率图、时间序列图等，有的基本图中又可进一步细分。交互图比基本图更漂亮，包括不同风格的二维、三维图，如条形交互图、饼形交互图。

利用 SPSS 可以生成数十多种不同风格的表格，概括来说有一般表、多响应表和频数表等。

3. 统计分析功能

SPSS 提供的统计分析功能，集中在“分析”菜单中，见图 1.6，共有 23 个菜单，各菜单下有子菜单，如回归下就有 12 个子菜单。这些功能主要包括：数据的描述性分析、参数检验、非参数检验、方差分析、相关分析、回归分析、聚类分析、判别分析、因子分析、时间序列分析、质量控制等。

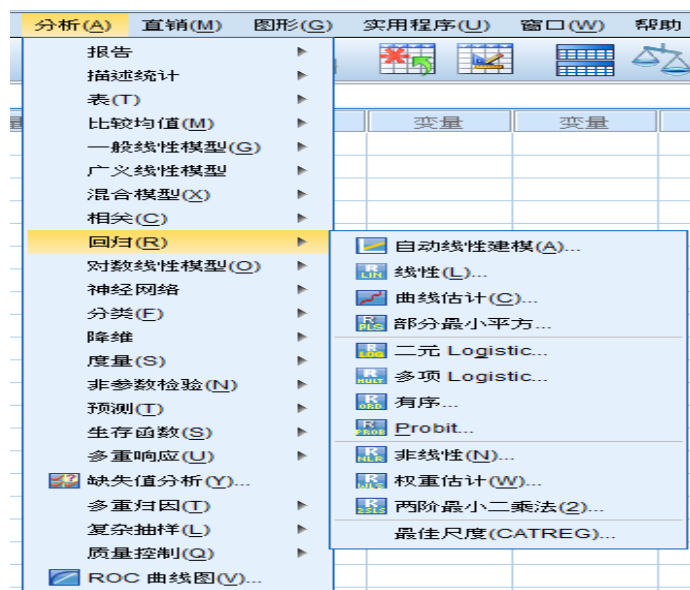


图 1.6 SPSS 的统计分析功能-“分析”菜单

教师：在信息化的时代，计算机的使用非常重要，对于做统计的，如果没有这个工具，那么就肯定是“巧妇难为无锅之炊”了。