



工业和信息化部“十二五”规划教材

应用统计学数据处理 与综合教程

主编 宫琳 阎艳 ●

APPLIED STATISTICS
DATA PROCESSING AND
COMPREHENSIVE EXPERIMENT



北京理工大学出版社

BEIJING INSTITUTE OF TECHNOLOGY PRESS



清华大学“985工程”项目

应用统计学数据处理 与综合教程

主编 曹 明 副主编

APPLIED STATISTICS
DATA PROCESSING AND
COMPREHENSIVE EXPERIMENT

清华大学出版社
Tsinghua University Press



工业和信息化部“十二五”规划教材

应用统计学数据处理 与综合教程

主 编 宫 琳 阎 艳 ●

副主编 薛 庆

APPLIED STATISTICS
DATA PROCESSING AND
COMPREHENSIVE EXPERIMENT



北京理工大学出版社

BEIJING INSTITUTE OF TECHNOLOGY PRESS

图书在版编目 (CIP) 数据

应用统计学数据处理与综合教程/宫琳, 阎艳主编. —北京: 北京理工大学出版社, 2014. 9

ISBN 978 - 7 - 5640 - 9177 - 4

I. ①应… II. ①宫… ②阎… III. ①应用统计学 - 数据处理 - 高等学校 - 教材
IV. ①C8

中国版本图书馆 CIP 数据核字 (2014) 第 090860 号

出版发行 / 北京理工大学出版社有限责任公司

社 址 / 北京市海淀区中关村南大街 5 号

邮 编 / 100081

电 话 / (010) 68914775 (总编室)
82562903 (教材售后服务热线)
68948351 (其他图书服务热线)

网 址 / <http://www.bitpress.com.cn>

经 销 / 全国各地新华书店

印 刷 / 保定市中华美凯印刷有限公司

开 本 / 787 毫米 × 1092 毫米 1/16

印 张 / 9.75

字 数 / 222 千字

版 次 / 2014 年 9 月第 1 版 2014 年 9 月第 1 次印刷

定 价 / 29.00 元

责任编辑 / 施胜娟

文案编辑 / 施胜娟

责任校对 / 周瑞红

责任印制 / 王美丽

本书是高等院校机械工程类专业应用统计学、数据处理综合实验等相关课程的配套教材或计算机实验教材。本书注重应用统计相关知识点、统计分析软件、实际应用案例相结合，主要包括应用统计学概述、相关软件介绍、基本实验指导、综合实验指导等四个部分。涉及应用统计学的统计数据收集及描述、随机变量与概率分布、参数估计及假设检验、方差分析与正交试验设计、相关与回归分析、时序分析和统计决策等知识。通过设计 11 个基本模块和 6 个综合模块，系统学习上述知识，并使用 Excel、SPSS 两种软件进行实验验证。内容涉及机械制造、工业工程、物流、质量、人因等领域，实际案例来源于多种行业。本书旨在引导学生正确掌握应用统计学知识，借助应用统计学软件对工业行业的数据和实际问题进行分析，提高自主分析和解决问题的能力，进而在工业实践中，提高实际应用能力。

全书图文并茂，通俗易懂，既可以作为高等院校机械工程大类本科生、研究生的应用统计学、数据处理综合实验等相关课程的教材或配套教材，也可以作为工业行业中相关管理人员的培训教材。

本书在策划、撰写过程中，得到了工业和信息化部、北京高等学校青年英才计划项目、北京理工大学出版社的支持，使本书得以如期出版，在此作者特表谢忱。

编者

目 录

CONTENTS

第一章 应用统计学概述	001
1.1 统计与统计学	001
1.2 统计学分类	002
1.3 应用统计学概述	003
第二章 相关软件介绍	006
2.1 Excel 功能介绍	006
2.1.1 Excel 简介	006
2.1.2 分析工具库的安装及使用	006
2.2 SPSS 功能介绍	008
2.2.1 SPSS 软件简介	008
2.2.2 SPSS 窗口简介	009
2.2.3 SPSS 数据文件建立	012
2.2.4 SPSS 数据分析过程	014
第三章 基本实验指导	015
3.1 统计数据收集及描述（一）	015
3.1.1 实验目的	015
3.1.2 考查知识点	015
3.1.3 实验工具	016
3.1.4 实验内容与要求	016
3.1.5 实验过程指导	017
3.1.6 思考与分析	027
3.2 统计数据收集及描述（二）	027
3.2.1 实验目的	027
3.2.2 考查知识点	028

3.2.3	实验工具	031
3.2.4	实验内容 with 要求	031
3.2.5	实验过程指导	032
3.2.6	思考与分析	037
3.3	随机变量与概率分布	038
3.3.1	实验目的	038
3.3.2	考查知识点	038
3.3.3	实验工具	041
3.3.4	实验内容 with 要求	041
3.3.5	实验过程指导	041
3.3.6	思考与分析	045
3.4	参数估计及假设检验 (一)	045
3.4.1	实验目的	045
3.4.2	考查知识点	045
3.4.3	实验工具	048
3.4.4	实验内容 with 要求	048
3.4.5	实验过程指导	048
3.4.6	思考与分析	050
3.5	参数估计及假设检验 (二)	051
3.5.1	实验目的	051
3.5.2	考查知识点	051
3.5.3	实验工具	053
3.5.4	实验内容 with 要求	053
3.5.5	实验过程指导	053
3.5.6	思考与分析	057
3.6	参数估计及假设检验 (三)	057
3.6.1	实验目的	057
3.6.2	考查知识点	057
3.6.3	实验工具	058
3.6.4	实验内容 with 要求	058
3.6.5	实验过程指导	058
3.6.6	思考与分析	060
3.7	方差分析与正交试验设计	060
3.7.1	实验目的	060
3.7.2	考查知识点	060
3.7.3	实验工具	062
3.7.4	实验内容 with 要求	062
3.7.5	实验过程指导	064
3.7.6	思考与分析	070

3.8	相关与回归分析（一）	070
3.8.1	实验目的	070
3.8.2	考查知识点	070
3.8.3	实验工具	072
3.8.4	实验内容及要求	072
3.8.5	实验过程指导	072
3.8.6	思考与分析	075
3.9	相关与回归分析（二）	076
3.9.1	实验目的	076
3.9.2	考查知识点	076
3.9.3	实验工具	079
3.9.4	实验内容及要求	079
3.9.5	实验过程指导	080
3.9.6	思考与分析	084
3.10	时序分析	084
3.10.1	实验目的	084
3.10.2	考查知识点	084
3.10.3	实验工具	084
3.10.4	实验内容及要求	084
3.10.5	实验过程指导	085
3.10.6	思考与分析	087
3.11	统计决策	087
3.11.1	实验目的	087
3.11.2	考查知识点	088
3.11.3	实验工具	089
3.11.4	实验内容及要求	089
3.11.5	实验过程指导	089
3.11.6	思考与分析	090
第四章	综合实验指导	091
4.1	区域物流竞争力评价分析	091
4.1.1	背景介绍	091
4.1.2	知识点	091
4.1.3	实验内容及要求	091
4.1.4	实验指导	093
4.2	企业财务数据分析	099
4.2.1	背景介绍	099
4.2.2	知识点	099
4.2.3	实验内容及要求	099

4.2.4	实验指导	100
4.3	人因工程分析	107
4.3.1	背景介绍	107
4.3.2	知识点	107
4.3.3	实验内容及要求	107
4.3.4	实验指导	111
4.4	货运量预测	121
4.4.1	背景介绍	121
4.4.2	知识点	121
4.4.3	实验内容及要求	121
4.4.4	实验指导	121
4.5	企业综合评价	126
4.5.1	背景介绍	126
4.5.2	知识点	126
4.5.3	实验内容及要求	126
4.5.4	实验指导	129
4.6	调查问卷分析	134
4.6.1	背景介绍	134
4.6.2	知识点	135
4.6.3	实验内容及要求	135
4.6.4	实验指导	137
参考文献		145

第一章

应用统计学概述

1.1 统计与统计学

从上古社会的古人类利用石子、绳结等进行计数,到奴隶社会的各种人口、土地、财富的记录,人类进行统计实践活动已有四五千年的历史。但是,真正将统计实践上升到理论层面,并进行概括和总结,使之成为“统计学”,距今却只有 300 多年的历史。从统计学发展的历史来看,对统计学产生过较大影响的主要有以下几个学派:

1. 政治算术学派

该学派的主要代表人物是威廉·配第(1623—1687)和约翰·格朗特(1620—1674)。威廉·配第的代表作是 1690 年出版的《政治算术》,书中对英国、法国、荷兰三国的经济实力进行了分析比较,他主张通过数学、重量等度量方法,对社会结构和政治事项进行解剖分析,以反映国家实力,同时这种方法也为统计学的创立奠定了方法论基础。约翰·格朗特于 1662 年著成《对于死亡表的自然观察与政治观察》一书,书中收集、整理了 1603 年以来的死亡表,研究并发现了人口与社会现象中重要的数量规律性,对伦敦市人口的出生率和死亡率进行分类计算,并利用找到的数理关系,进行推算和预测。

2. 国势学派

该学派的创始人是德国人海尔曼·康令(1606—1681)。后来又陆续出现了一些该学派的代表人物,使该学派持续了将近 200 年。该学派的高特弗里德·阿亨瓦尔(1719—1772)于 1749 年出版的《近代欧洲各国国势学论》中首先使用了“统计学”(德文 Statistik)一词,并定义为研究一国或多数国家显著事项之学。国势学派把统计学理解为国家重要事项的记述,如人口、领土、政治结构等,都用文字阐述,而不用数字计量,向国家统治者提供一些有关各国情况的知识。但是,由于缺乏数字内容,多用文字表述,因此它虽然有统计学之名,但并无统计学之实。

著名的德国经济学家和统计学家克尼斯(1821—1897)于 1850 年发表了论文《独立科学的统计学》,提出将“国家论”作为“国势学”的科学命名,而将“统计学”作为“政治算术”的科学命名,从此统计学才名副其实。

3. 数理统计学派

该学派的创始人是阿道夫·凯特勒(1796—1874),他把法国古典概率论引入统计学,从而使统计的方法有了重大的突破和发展,他运用统计学既研究自然现象,又研究社会规律,大大地扩展了统计学原有的社会学的研究领域,丰富了统计学的研究内容。其最有影响

的著作为 1869 年出版的《社会物理学》。

数理统计学的名称来源于德国的韦特斯坦在 1867 年发表的一篇文章——《关于数理统计学及其在政治经济学和保险学中的应用》。后来高尔顿 (1822—1911)、皮尔逊 (1857—1936)、戈赛特 (1876—1937) 和费希尔 (1890—1962) 等人分别提出并发展了回归和相关、假设检验、卡方分布和 t 分布等理论, 使数理统计学逐渐成为一门独立的科学, 也使其在现代统计理论和实践中的影响越来越大。

在学术观点上, 数理统计学派认为统计学就是数理统计学, 是通用于研究自然现象和社会现象的方法论体系, 是现代应用数学的一个重要分支。

4. 社会统计学派

社会统计学派的创始人是梅尔 (1841—1925), 其著有《统计学与社会学》和《社会生活中的规律性》。另一个代表人物是恩格尔 (1821—1896), 其著有《关于统计学是独立科学或方法问题之我见》和《比利时工人的家庭生活费》等。

社会统计学派着重对社会经济领域的统计方法及其应用进行研究, 继承和发展了“政治算术学派”的理论体系, 又继承和发扬了凯特勒强调的研究自然现象和社会规律的传统。该学派主张统计学是一门独立的实质性社会科学, 而认为数理统计学是一门应用数学。

1.2 统计学分类

统计学是一门搜集、整理和分析统计数据, 并从数据中得到结论的方法科学, 其目的是研究如何正确认识客观事物数理方面的一门科学。统计学基本理论并不依附于任何实质性科学的方法论, 但是, 它又广泛应用于社会科学、自然科学等诸多学科。统计学与统计应用联系紧密, 就像理论与实践的关系。

一般来说, 统计学可以有多种分类方式。

1. 描述统计学和推断统计学

统计学按照侧重点不同, 可以分为描述统计学和推断统计学。

描述统计学 (Descriptive Statistics) 研究如何取得反映客观现象的数据, 并通过图表形式对所搜集的数据进行加工整理和显示, 通过综合、概括与分析得出反映客观现象的规律性数量特征。其主要包括统计数据的收集方法、数据的加工处理方法、数据的显示方法、数据分布特征的概括与分析方法等。

推断统计学 (Inferential Statistics) 研究如何根据样本数据去推断总体数量特征, 在对搜集的样本数据描述的基础上, 对统计总体的未知数量特征做出以概率形式表述的推断。其主要包括参数估计、假设检验、估计、回归等。

2. 理论统计学和应用统计学

统计学按照理论与实践应用的关系, 可以分为理论统计学和应用统计学。

理论统计学 (Theoretical Statistics), 一般又称为数理统计学, 主要是指统计学的一般化、概括化的原理。它主要从纯理论的角度, 以概率论为基础, 对统计方法加以推导论证, 以归纳方法研究统计学的一般理论和统计方法的数学理论。理论统计学是统计方法的理论基础, 属于通用方法论的范畴。

应用统计学 (Applied Statistics) 是研究如何应用统计学一般原理和方法解决实际问题的学科。由于其从所研究的领域或专门问题出发, 通过数据分析来解决实际问题, 因此与不同领域的实质性学科有着非常密切的联系。由于各专业的不同特点, 应用统计学需要有关的专业实质性学科的理论作为指导。目前, 统计学的应用几乎扩展到了所有的科学研究领域。

1.3 应用统计学概述

一般来说, 应用统计学需要结合所研究学科, 采用统计方法进行数据分析和处理, 以解决实际问题。应用统计学应能够提供信息处理、分析咨询、监督控制等基本职能。

应用统计学有很多方法, 具体分析如下。

1. 抽样方法

抽样方法的主要构成内容如图 1-3-1 所示。



图 1-3-1 抽样方法

抽样是获取数据的基本方法。为了确保分析结果的准确性和可靠性, 样本的抽取需要专门的理论方法来支撑。抽样一般分为随机抽样和非随机抽样。建立在概率论与数理统计基础上的样本获取方法属于随机抽样。这类方法可以保证最终推断的结果具备确定的把握程度和准确程度。

针对构成内容比较单一、规模也不大的总体, 可采用简单随机抽样。简单随机抽样具体又分为重复抽样与不重复抽样两种方式。对于那些规模比较大、构成内容比较复杂的总体, 应当结合统计研究的目的以及相关的限制条件选用分层抽样、整群抽样和系统抽样等较为复杂一些的抽样方法。

2. 描述方法

对于需要进行描述分析的总体或者样本, 当总体或者样本数据比较大时, 需要对其进行整理和加工, 以获得总体参数。对样本数据的整理和加工, 以及接下来的统计量的计算过程, 也是对样本数据的概括和描述过程。统计描述方法分为两类: 图表描述与统计量描述。具体内容如图 1-3-2 所示。

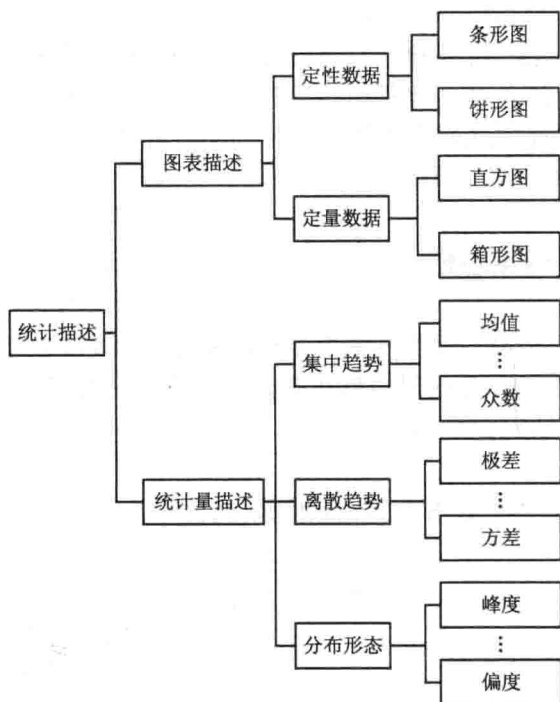


图 1-3-2 描述方法的基本构成内容

对于总体或者样本中所包含的众多个体，对给定变量的各个取值上的分布状况可通过频数分布表来描述。在此基础上制作的频数分布图可以更为直观地显示分布情况。统计中所涉及的变量数据一般分为四种类型：定类数据、定序数据、定距数据和定比数据。定类数据和定序数据，统称为定性数据，也称为品质型数据；定距数据和定比数据，统称为定量数据，也称为数值型数据。适用于定性数据的频数分布图形主要有条形图和饼形图；适用于定量数据的频数分布图形主要有直方图和箱形图。此外，还有茎叶图、点线图。

另外，用于描述的统计量主要有三种类型：描述集中趋势的统计量，具体包括均值、众数、中位数等；描述离散趋势的统计量，具体包括极差、方差、标准差等；描述分布形态的统计量，具体包括峰度、偏度等。此外，分位数，特别是上、下四分位数，也常被用于描述数据的集中趋势和离散趋势。

3. 推断方法

统计推断是指根据带随机性的观测数据（样本）以及问题的条件和假定（模型），而对未知事物做出的、以概率形式表述的推断。由于统计推断是由部分（样本）推断整体（总体），因此根据样本对总体所做的推断，不可能是完全精确的和可靠的，其结论要以概率的形式表达。

统计推断是现代统计学的基本方法，在统计研究中得到了极为广泛的应用，它既可以用于对总体参数的估计，也可以用作对总体某些分布特征的假设检验。针对单个变量或两个变量之间的关系进行推断的方法，属于基础统计方法，其主要构成内容如图 1-3-3 所示。

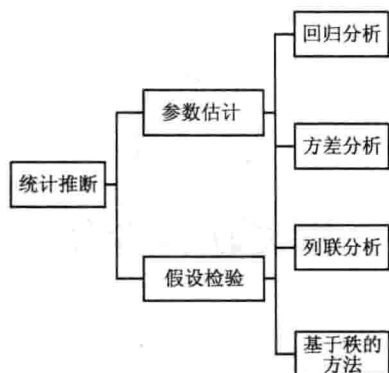


图 1-3-3 推断方法的基本内容

参数估计与假设检验贯穿于统计推断过程的始终。就两变量之间的关系进行统计推断时，由于两个变量所属类型不同，又具体分为：回归分析、方差分析、列联分析、基于秩的方法等。

如果是时间序列样本数据，对应的又有时间序列分析的各种方法；如果是就三个或三个以上变量的关系进行统计推断，则有各种多元统计分析方法，主要的多元统计分析方法有因子分析和聚类分析。

第二章

相关软件介绍

2.1 Excel 功能介绍

2.1.1 Excel 简介

Microsoft Excel 是微软公司 Microsoft Office 桌面办公套件的其中之一。它是用来组织、计算和分析数据的通用型电子表格软件。Microsoft Excel 有多种版本，随着版本的提升，界面和功能有了很多增加和改变。本书采用的是 Microsoft Office 2010 版本的 Excel。

Excel 的功能极其强大，而且由于与 Windows 操作系统同出一家，所以其普及面广，功能界面方便，易于操作。除了可以处理一般的表格编排等基本任务，在数据处理和统计分析方面，Excel 还提供了以下功能：

(1) 公式与函数：Excel 提供了丰富的数学、统计函数，以及自定义公式功能。对于需要结合多种函数才能完成分析的统计任务，可以综合运用 Excel 的公式和函数来完成。

(2) 图表功能：Excel 具有强大的图表功能，除了可以直接绘制柱形图、折线图、饼图等基本图形外，还可以通过分析工具绘制各种统计图形，如直方图、拟合图等。

(3) VBA 编程：在 Excel 中可以使用 VBA，通过 Visual Basic 编辑器，可以对数据进行统计分析处理，还可以按照需求制定各种输出格式。

(4) 分析工具库：Excel 专门提供了数据统计分析工具，该工具集成了常用的统计分析方法，包括描述分析、方差分析、回归、F 检验、t 检验和 z 检验等。需要对数据进行分析时，只需要调用相关功能，并明确相应的输入数据和参数后，Excel 就会输出相应的统计分析结果，部分可以生成图表更加直观。相较之其他统计软件，Excel 在统计方面的功能和分析能力略显一般，但是，对于初学统计的人以及功能要求相对简单的任务，Excel 的功能已经足以完成。

(5) 规划求解：Excel 还提供了用于求解线性、非线性规划的规划求解工具，可以简称为 Excel Solver。其也可以解决部分回归问题，但是本书中不做具体介绍。

2.1.2 分析工具库的安装及使用

1. 分析工具的安装

如果要浏览已有的分析工具，可以单击【数据】选项卡【分析】组中的【数据分析】按钮。

如果【数据分析】没有出现在选项卡上，则必须先加载该分析工具库，具体如下：

(1) 单击【文件】→【选项】，弹出“Excel 选项”对话框，再单击对话框的【加载项】，弹出如图 2-1-1 所示对话框。确保左下角“管理”下拉列表中是“Excel 加载项”（一般默认即是）。



图 2-1-1 Excel 选项对话框

(2) 单击【转到】按钮，弹出“加载宏”对话框，如图 2-1-2 所示。在对话框中，选中“分析工具库”，最后单击【确定】按钮即可。如果没有“分析工具库”，则单击【浏览】进行查找。如果出现一条消息，指出您的计算机上当前没有安装分析工具库，请单击【是】进行安装。

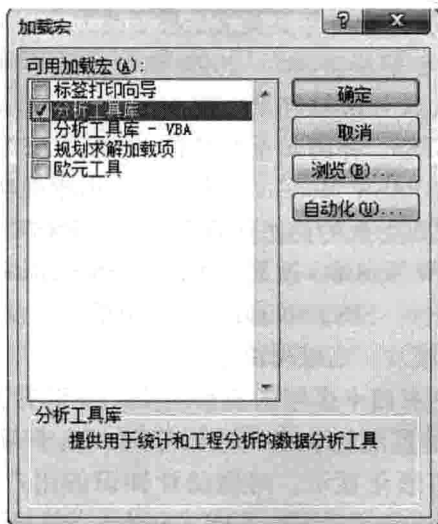


图 2-1-2 加载宏对话框

2. 分析工具的使用

(1) 打开相应的 Excel 文件。

(2) 单击【数据】选项卡【分析】组中的【数据分析】按钮，在弹出的“数据分析”

对话框中双击相应的数据“分析工具”选项，如“直方图”、“抽样”等，如图 2-1-3 所示。

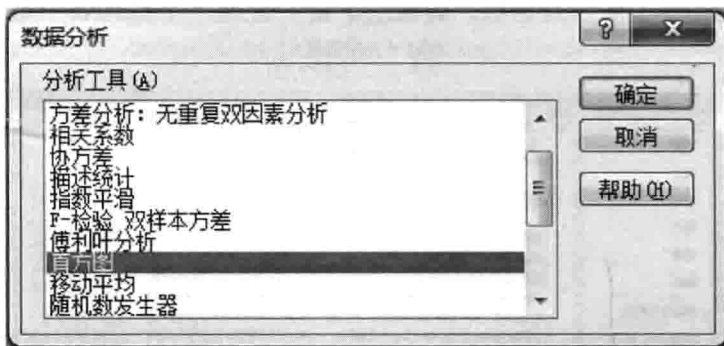


图 2-1-3 数据分析对话框

(3) 在打开的各种分析工具的对话框中，按各种分析工具的要求填写相应的信息，最后按【确定】按钮。

2.2 SPSS 功能介绍

2.2.1 SPSS 软件简介

SPSS 是世界上最早的统计分析软件，由美国斯坦福大学的三位研究生于 20 世纪 60 年代末开发，同时成立了 SPSS 公司，并于 1975 年在芝加哥组建了 SPSS 总部。SPSS 原意为 Statistical Package for the Social Sciences，即“社会科学统计软件包”，后因 SPSS 的产品服务领域的扩大和服务深度的增加，2000 年其英文全称更改为 Statistical Product and Service Solutions，意为“统计产品与服务解决方案”。2009 年 4 月，SPSS 公司宣布重新包装旗下的 SPSS 产品线，定位为预测统计分析软件 PASW (Predictive Analytics Software)。同年 7 月，IBM 花费 12 亿美元收购了 SPSS，并将其重新打包，包括 PASW Statistics（主要为统计分析功能，源于 SPSS Statistics）、PASW Modeler（主要为数据挖掘功能，源于 Clementine）等功能模块。但是，可能是考虑到受众的熟悉程度等原因，PASW Statistics 的名字并未使用太久。2010 年后，IBM 将 PASW Statistics 改回了 IBM SPSS Statistics，简称还是 SPSS Statistics。目前，SPSS 共有七大系列软件，SPSS Statistics 是其中最为大众所熟知的软件，最新版本为 SPSS Statistics 22，包括基本模块、高级模块等。

迄今为止，SPSS 软件已有四十多年的成长历史，是世界上应用最广泛的专业统计软件。SPSS 最突出的特点就是直观的表格式数据管理、基于 Windows 的友好的操作界面、直观的输出结果和美观的图形化显示，略懂统计知识的用户极易上手，分析结果全面。但是，从目前统计软件行业的角度来看，SAS (Statistical Analysis System) 和 R 的实力也非常强大。SAS 相对来说最为昂贵，功能模块丰富。R 提供的基本模块不多，但是由于其为免费软件，受到了越来越多人的追捧，而且众多的拥趸者开发了多种多样的功能包，使得 R 在社会学、生物学等领域的使用率越来越高。当然，直观的表格式数据管理和基于 Windows 的程序菜单仍然是 SPSS 的最大特点，相较之 SAS 和 R 的编程界面，SPSS 对于