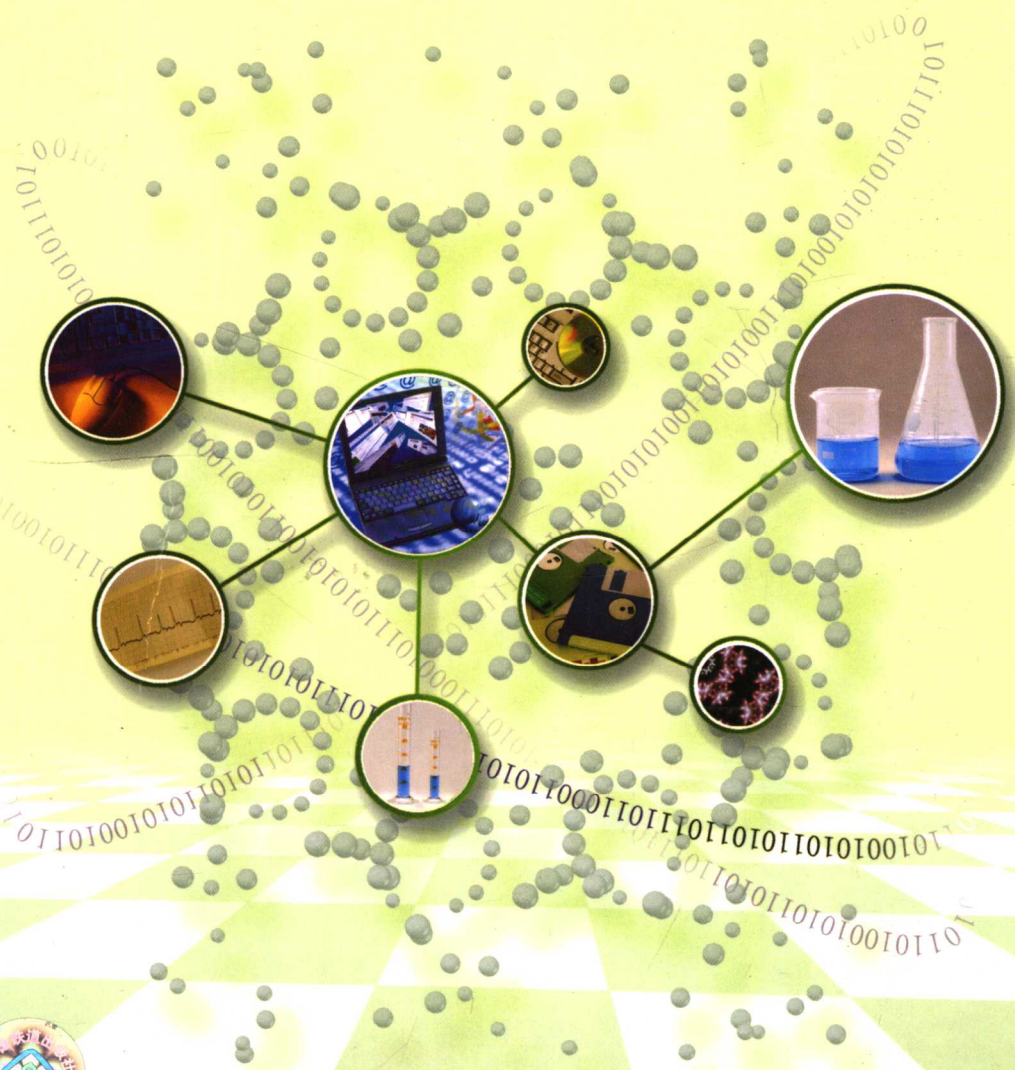




全国高等医药类院校计算机规划教材

医学统计实用技术教程

刘尚辉 张筠莉 主编



中国铁道出版社
CHINA RAILWAY PUBLISHING HOUSE

内 容 简 介

本书从使用计算机处理医药学数据的角度,运用典型案例详细地讲述了 SPSS 与 Excel 在医药学统计中的应用。

我们的写作对象是非统计专业的医药院校的大学生、研究生以及科研人员,写作目的是使不懂医学统计的学生与科研人员通过对本书的学习,能够独立进行各种常用的医学统计分析工作。全书分上、下两篇,上篇介绍 SPSS 在医药学统计中的应用,下篇介绍 Excel 在医药学统计中的应用,共计 16 章内容。

通过对本书的学习,使读者首先能够针对自己的数据资料来选择适当的统计方法;然后运用 SPSS 或 Excel 将数据资料建立起数据文件,或将不同类型的数据文件转换成所需的资料形式;接下来操作统计软件输出统计分析结果;最后学会分析结果并得出正确的结论。

应用是学习的动力,简明是对教材的基本要求。我们力求使本教材简明扼要,从应用实例出发,每例都配有问题分析、操作实现及结果解释,以达到使读者在较短的时间内就能学会使用常用统计方法的目的。

图书在版编目(CIP)数据

医学统计实用技术教程/刘尚辉,张筠莉主编. —北京:
中国铁道出版社,2007.1
全国高等医药类院校计算机规划教材
ISBN 978-7-113-07621-4

I. 医… II. ①刘…②张… III. ①医学统计—统计分析—
软件包, SPSS—医学院校—教材②电子表格系统, Excel—
应用—医学统计—医学院校—教材 IV. R195.1-39

中国版本图书馆 CIP 数据核字(2007)第 008051 号

书 名: 医学统计实用技术教程

作 者: 刘尚辉 张筠莉 等

出版发行: 中国铁道出版社(100054,北京市宣武区右安门西街8号)

策划编辑: 严晓舟 秦绪好

责任编辑: 苏 茜 赵 轩 徐盼欣

封面设计: 薛 为

封面制作: 白 雪

责任校对: 王 欣

印 刷: 三河市国英印务有限公司

开 本: 787×1092 1/16 印张: 14.5 字数: 330 千

版 本: 2007 年 3 月第 1 版 2007 年 3 月第 1 次印刷

印 数: 1~5 000 册

书 号: ISBN 978-7-113-07621-4/TP·2231

定 价: 20.00 元

版权所有 侵权必究

本书封面贴有中国铁道出版社激光防伪标签,无标签者不得销售

凡购买铁道版的图书,如有缺页、倒页、脱页者,请与本社计算机图书批销部调换。

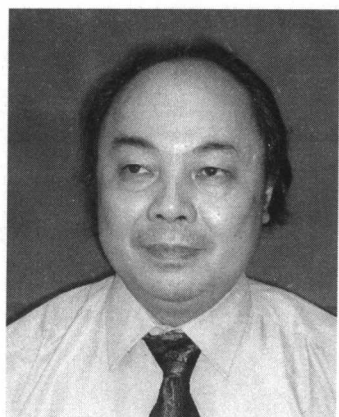
编审委员会主任、主审简介



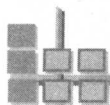
王世伟, 1950 年 12 月生, 现任中国医科大学计算机中心主任, 教授, 硕士研究生导师, 理工学部副主任。现担任中华医院管理学会信息管理专业委员会委员, 全国高等院校计算机基础教育研究会医学专业委员会副主任委员, 辽宁省高等院校计算机基础教育研究会副理事长, 辽宁省卫生信息化建设专家组首席专家, 辽宁省医学影像技术学会常务理事等学术职务。20 年来一直从事计算机基础教育与科研工作。在国家级核心期刊发表论文 30 余篇, 主持国家、省级科研课题 3 项, 其中“构建医学特色的《大学计算机基础》课程体系”系辽宁省“十一五”规划课题, 在此基础上主编出版了《医学信息系统教程》、《现代医学影像技术》、《网站的规划与建设》等 20 余册全国高等医药院校计算机规划系列教材。



周怡, 1954 年 4 月生, 广东药学院医药信息工程学院院长, 医药信息学教授, 东南大学本科及硕士研究生毕业。主要研究方向有: 计算机科学与网络技术在医药中的应用、医药信息整合与数据挖掘、智能化医药信息系统工程等。教育部高等学校(本科)计算机基础教学指导委员会医药类计算机基础课程教学指导分委员会委员(2006~2010)。全国高等院校计算机基础教育研究会医学专业委员会副主任委员(2005~2008)。中国卫生信息学会卫生信息技术应用专业委员会委员(2004~2007)。广东省“南粤教书育人优秀教师”。



邹赛德, 中国医药信息学会常务理事、中华医院管理协会信息专业委员会委员、广东医院管理学会信息管理专业委员会主任委员、广东医药信息学会理事长、中国电子学会医药信息学分会常务委员、广州市卫生局的咨询小组成员; 是《医药信息技术基础》、《计算机应用基础》、《医学计算机应用基础》、《医学计算机应用基础实验教程》四本本科规划教材的主编,《医学计算机实用教程》研究生规划教材的副主编和《医院管理学信息管理分册》专著的编委。在全国医药院校计算机教学中具有表率 and 影响的作用, 同时在广东医学信息学领域内居学术领导地位。在国内同行中具有较高的威信和影响力。



全国高等医药类院校计算机规划教材

编审委员会

- | | | |
|------|-----------|---------|
| 主 审: | 邹赛德 | 中山大学 |
| 主 任: | 王世伟 | 中国医科大学 |
| | 周 怡 | 广东药学院 |
| 副主任: | 刘建平 | 辽宁中医药大学 |
| | 董鸿晔 | 沈阳药科大学 |
| | 王延红 | 沈阳医学院 |
| | 李祥生 | 山西医科大学 |
| 委 员: | (按姓氏拼音排序) | |
| | 高 昱 | 锦州医学院 |
| | 韩 滨 | 大连医科大学 |
| | 刘 燕 | 中山大学 |
| | 刘尚辉 | 中国医科大学 |
| | 肖 锋 | 大连医科大学 |
| | 晏峻峰 | 湖南中医学院 |
| | 张筠莉 | 锦州医学院 |

林姝收财财真竹外刺类茂国碧高国全

本

书

编

委

会

学大山中 彭赛胜 : 审 主
主 编: 刘尚辉 张筠莉 王 王 主

副主编: 刘红波 于 净 周 王 主 编

编 委: (按姓氏笔画排序) 董 董

于 净 包 红 刘丹萍 刘红波
刘尚辉 张筠莉 李 新 姚 林
高 轶 (音 书 力 致 若) 员 委

呈 高
寒 韩
燕 岐
群 尚 岐
群 肖
郭 郭 晏
陈 陈 洪

序

随着 21 世纪科教兴国战略的实施及信息化社会进程的加速,形成了以信息化带动医药卫生事业现代化的整体发展趋势,并深刻地影响与改变着传统的医药科学,使今天的医学工作者和医药院校的师生们同样面临着 IT 知识更新的机遇和挑战。

我国卫生信息化建设同时面临着两个十分紧迫又必须解决的重要问题:对高素质医学人才不断快速增长的需求;现行医药院校的计算机与信息技术基础教学体系中的 IT 知识结构不够全面以及专业领域的 IT 操作技能不适用。

本着“社会的进步靠科学,科学的进步靠人才,人才的培养靠教育,教育的发展靠理念”的精神,必须努力构建具有医学特色的“大学计算机基础与应用”课程体系,紧密结合本学科信息化建设与应用的方向,用科学发展观来培养能进行自主学习,且适应未来社会发展及医药信息化建设需求的合格医学人才。为了推进全国高等医学院校计算机基础课程体系的教学改革,作好教材建设先行的工作,我们编写了“全国高等医药类院校计算机规划教材”丛书。

“全国高等医药类院校计算机规划教材”丛书,先期包括计算机与信息技术基础类、程序设计基础类及医学 IT 实用技术基础类,共三大类 12 本教材。涵盖了全国高等医药院校本、专科各专业的计算机与信息技术应用基础课程的教学用书。教材内容覆盖面和知识点的取舍本着全面系统、科学合理、结合专业、注重实用、知识宽泛、关注发展的六项原则,比较完整地构建了具有医学特色的计算机与信息技术基础课程教材体系。其中:

- 计算机与信息技术基础类:包括《医学计算机与信息技术基础教程》、《医学计算机与信息技术基础实验指导》、《医学多媒体实用技术教程》、《医学网络实用技术教程》;
- 程序设计基础类:包括《Visual FoxPro 程序设计教程》、《Visual FoxPro 程序设计实验指导与习题集》、《Visual Basic 程序设计教程》、《Visual Basic 程序设计实验指导与习题集》;
- 医学 IT 实用技术基础类:包括《医学统计实用技术教程》、《医学信息系统教程》、《医学影像实用技术教程》、《医学数据挖掘与网格计算教程》。

“全国高等医药类院校计算机规划教材”丛书的编写与出版,得到了国内许多著名医药院校的鼎力支持与合作,教材的编者包括国内医学院校知名教育专家、教育部医药类计算机基础课程教学指导分委员会委员,大量从事一线教学工作、具备丰富教学经验的教师。他们

视角独特，洞见非凡，匠心独运地将目前开展得如火如荼的 IT 与医疗卫生信息化建设与实践以这套全国高等医药类院校计算机规划教材丛书的形式表现了出来。中国铁道出版社对该系列教材进行了远见卓识的精心策划、科学论证、倾力帮助、编辑发行等大量认真而卓有成效的工作。此间，还有全国许多从事计算机基础教学方面的领导、专家、老师和同学们给了我们很大的支持和帮助，在此一并表示衷心的感谢。

在此，我们愿把这套规划教材丛书奉献给全国医药院校的师生们，为构建我国具有医学特色的计算机与信息技术应用基础课程体系，共同走出一条新路，在深化教学改革及先行教材建设中做出贡献。

王世伟 周怡

2006 年 7 月

前言

医学院校的师生以及医疗卫生系统中的科研工作者经常会面临这样的问题：大量的科研数据需要有效的处理，对处理后的结果进行正确的统计分析，从而达到预期的目的。解决这些问题的关键就是要掌握一种得心应手的统计分析工具，利用它来对自己的科研数据进行恰当的处理和分析，以得到正确的结论。本书就是力求提供给读者这样一些实用的，既简单又易学的统计分析软件，从而能够进行各种常用的医学统计分析，这对读者的学习与科研都是非常需要的。

本书从计算机处理医药学数据的角度，运用典型案例详细地讲述了 SPSS 的统计操作方法与使用技巧，又考虑到目前很多读者对 Excel 较为熟悉，因此对 Excel 中的统计函数与统计分析工具也做了介绍，以满足不同读者的应用需求。本书面对的主要对象是非统计专业的医药院校的大学生、研究生以及科研人员，写作目的是使不懂统计的学生与科研人员通过对本书的学习，能够独立进行各种常用的统计分析处理。本书分上、下两篇，上篇介绍了 SPSS 在医药学统计中的应用，下篇介绍了 Excel 在医药学统计中的应用，共计 16 章内容。

在上篇的 13 章中，我们本着可读性好、实用性强、简单明了、举例典型的写作特点来介绍 SPSS 这一统计软件包。主要内容有：①数据文件的建立，包括原始数据的录入与编辑方法，以及当数据是来自其他类型的文件时，如何将其转换为 SPSS 类型的数据文件等；②数据资料的简单描述性分析；③计量、计数资料的差异比较；④资料的相关性与相关程度的分析方法；⑤生存资料分析方法；⑥常用统计图形的制作方法；等等。在每一部分内容的介绍中，我们从非统计专业用户的角度出发，用一个典型的实例引导用户逐步完成全部操作。使读者知道为什么选择这种统计方法；如何去建立数据文件；怎样操作统计软件；如何看结果和正确的解释结论，等等。这些内容是进行医学统计最需要解决的问题，也是学习 SPSS 的关键与核心问题。

下篇介绍了 Excel 在医药学统计中的应用。大多数读者对 Excel 并不陌生，但对 Excel 中的统计函数与统计工具的使用了解较少。借助读者已经具备的知识条件，无需再学习与安装其他的统计软件，就可以很快地处理自己手中的数据，达到统计分析的目的。因此，在下篇中我们对 Excel 中的统计函数与统计工具做了较全面且详细的介绍，并对 Excel 中的图表功能也一并做了介绍。

应用是学习的动力，简明是对教材的基本要求。我们力图使本教材简明扼要，从应用实例出发，每例都配有问题分析、操作实现及结果解释，以达到使读者在短时间内学会常用统计分析方法的目的。

全书共 16 章，第 1 章、第 2 章、第 11 章和第 13 章由刘尚辉编写，第 3 章和第 4 章由姚琳编写，第 5 章和第 6 章由张筠莉编写，第 7 章由包红编写，第 8 章和第 12 章由刘红波编写，第 9 章由李新编写，第 10 章由刘丹萍编写，第 14 章、第 15 章和第 16 章由于净编写。在此谨向在本书编写过程中给予帮助的同志们表示衷心的感谢。

SPSS 大而全，在某种意义上没有必要对其全部了解。本书只介绍在医学中常用的统计内容与方法，以达到实用的写作目的。对于未涉及的内容，读者可查阅 SPSS 原版使用手册。

由于编者水平有限，书中难免有不妥之处，敬请批评指正。

编 者

目 录

第 1 章 常用医药学统计软件概述.....	1
1.1 SPSS 系统介绍	1
1.2 SAS 系统介绍	4

上篇 SPSS 在医药学统计中的应用

第 2 章 SPSS for Windows 概述.....	7
2.1 SPSS for Windows 的主要分析功能介绍	7
2.2 SPSS for Windows 的运行环境介绍	8
2.3 SPSS for Windows 的启动与退出	8
2.4 SPSS for Windows 的主窗口介绍	9
习题	13
第 3 章 数据文件的建立	14
3.1 数据的录入	14
3.2 变量、观察值的插入与删除	18
3.3 数据文件的打开与保存	20
3.4 SPSS 中基本文件类型	21
3.5 其他类型的数据文件的读入	21
习题	26
第 4 章 分析前的准备工作	28
4.1 数据查找	28
4.2 观察值(行)的排序与选择	29
4.3 变量值的排秩	30
4.4 数据文件的行列转置	33
4.5 观察值(行)的加权	34
4.6 数据文件的拆分	34
4.7 数据文件的合并	36
4.7.1 纵向合并	36
4.7.2 横向合并	39
4.8 数据的分类汇总	43
习题	46
第 5 章 数据转换	47
5.1 变量计算	47
5.2 随机数种子	50

5.3 特定变量值计数	51
5.4 重新赋值	53
5.5 连续变量离散化	55
习题	56
第6章 数据资料的描述性分析	57
6.1 频数分析过程	57
6.1.1 频数分析步骤	57
6.1.2 频数分析实例	59
6.2 描述统计量过程	62
6.2.1 描述统计量过程	62
6.2.2 描述统计量过程实例	63
6.3 数据探索过程	65
6.3.1 数据探索过程	65
6.3.2 数据探索实例	67
6.4 用 MEANS 过程进行常用描述性统计量的计算	72
6.4.1 均值比较过程	72
6.4.2 均值比较过程实例及分析	73
习题	76
第7章 两组计量资料的均数差异比较	77
7.1 t 检验概述	77
7.2 单样本 t 检验	77
7.3 两独立样本 t 检验	79
7.4 配对设计样本 t 检验	84
习题	88
第8章 多组计量资料的均数差异比较	90
8.1 方差分析概述	90
8.2 单因素方差分析及两两比较	90
8.3 双因素方差分析及两两比较	94
8.4 析因设计的方差分析	98
习题	101
第9章 计数资料样本间差异的比较	103
9.1 χ^2 检验概述	103
9.2 $R \times C$ 表资料 χ^2 检验	104
9.3 一般四格表 χ^2 检验	105
9.3.1 Pearson χ^2 检验	106
9.3.2 校正 χ^2 检验	107
9.3.3 Fisher 精确概率检验	108
9.4 配对 (四格表) χ^2 检验	109

9.5 配对 $R \times R$ 表一致性检验	110
习题	111
第 10 章 偏态分布计量资料或等级资料样本间差异的比较	113
10.1 秩和检验概述	113
10.2 配对设计资料的秩和检验	113
10.3 随机设计两个样本的秩和检验	115
10.3.1 原始资料的分析	116
10.3.2 等级资料的分析	117
10.4 随机设计多个样本的秩和检验	118
10.4.1 原始数据的分析	119
10.4.2 等级资料的分析	120
10.5 随机区组设计多个样本的秩和检验	122
习题	123
第 11 章 变量间关系的分析方法	125
11.1 相关分析概述	125
11.2 双变量相关分析	125
11.2.1 皮尔逊 (Pearson) 相关分析	125
11.2.2 斯皮尔曼 (Spearman) 等级相关分析	128
11.3 回归分析概述	129
11.4 直线回归	130
11.5 多元线性回归	133
11.6 多元逐步回归分析	137
11.7 logistic 回归	138
习题	142
第 12 章 生存资料的分析方法	144
12.1 生存资料分析概述	144
12.2 Kaplan-Meier 方法	145
12.3 寿命表方法	147
12.4 Log-rank 检验	150
习题	152
第 13 章 统计图形的制作	154
13.1 条形图	154
13.2 饼图	161
13.3 线图	163
13.4 直方图	167
13.5 散点图	167
13.6 帕累托图	168
习题	170

下篇 Excel 在医药学统计中的应用

第 14 章 Excel 中数据处理与图形制作	171
14.1 Excel 中数据的处理	171
14.1.1 Excel 中数据的输入与格式化	171
14.1.2 Excel 中数据的条件格式与有效性检验	172
14.1.3 Excel 中数据的排序、汇总与筛选	173
14.2 Excel 中图形的制作与编辑	175
14.2.1 统计图表的建立	175
14.2.2 图表的选项	176
14.2.3 柱形图	177
14.2.4 条形图	178
14.2.5 折线图	178
14.2.6 饼图	178
14.2.7 XY 散点图	179
14.2.8 面积图	179
14.2.9 应用举例	180
习题	182
第 15 章 Excel 中常用统计函数的应用	185
15.1 数据资料的描述类函数	185
15.2 数据资料的差异比较类函数	188
15.3 数据资料的相关性类函数	189
习题	190
第 16 章 Excel 中统计工具的应用	192
16.1 描述统计	192
16.2 t 检验	194
16.3 方差分析	198
16.4 相关与回归分析	204
16.5 F 检验	210
16.6 z 检验	211
习题	212
模拟试卷	215
参考文献	217

第1章 常用医药学统计软件概述

在信息时代的今天,人们每天都会通过各种方式获取大量的信息,要在这些信息中获取有科学价值的结果,必须对信息进行存储、整理、分析和处理。由于对数据的分析大都是基于基本的统计原理,并利用计算机来实现的,国外的一些公司多年来研制了许多统计软件包,如 SPSS、SAS、R、Stata、S-PLUS、NCSS、Egret、MiniTab、Statistica 和 Excel 等,在国内利用 SPSS、SAS 和 Excel 进行数据统计分析的用户较多。国内的学者们经过多年的努力也研制出了许多用于数据处理与分析的软件,如 DAS (Drug And Statistics) 药物与统计软件、Qstat (Quick Statistical Analysis Software) 快速数据统计分析程序、简明统计软件 (CS2000)、PEMS (Package for Encyclopaedia of Medical Statistics) 《中国医学百科全书·医学统计学》统计软件包、SPLM (Statistical Program for Linear Modeling) 统计软件、多元统计分析绘图系统 PPCH3.2、DPS (Data Processing System) 数据处理系统和 NDST 新药统计程序等,它们都是各具特色的有效信息处理工具,为许多用户接受与使用。

上述每一种数据处理工具都是按照统计原理与科学的算法研制出来的,使用时可以根据个人情况选择其中一种或几种。在医药学上常用的统计方法很多,它们各自适用于不同类型的数据。因此,在对数据资料进行统计处理时,必须根据数据的具体情况科学正确的选用。这是因为,数理统计是根据一定的数据分布推导出一定的统计方法,它仅适用于一定的数据。例如, t 检验及方差分析是由正态分布数据以及各组方差相同推导出来的,因而要求数据为正态分布(至少接近正态分布)且方差齐性。如果数据不满足这一要求而用此方法来处理,将得出不正确的结论。再例如,分类数据与计量数据的统计方法也是不同的,等等。应用错误的统计方法会使整个精心进行的研究得出错误的结论,如何根据数据的情况来正确地选择统计方法进行数据分析是非常重要的。因此,统计软件包的选择、统计方法的选择、软件的操作以及对结果的正确解释是信息处理的重要环节与组成部分。

在各具特色的众多统计软件中,本书选择两种分上下篇加以介绍,上篇介绍 SPSS 统计软件包的统计分析方法;下篇介绍 Excel 中的常用统计函数、统计工具与图表的制作方法。由于微软的办公自动化软件得到广泛使用,目前利用 Excel 进行数据统计分析的用户也很多。Excel 采用窗口式的工作界面,操作简单,不需要编程,可以利用其特有的统计函数与统计工具来实现各种统计分析功能,并可以利用 Excel 的图表功能将各种统计分析结果制作成可视化的图表。因此,Excel 也是一种有效的统计分析工具,可以将其运用于日常的工作、学习及科研中。

下面我们对在医药学上较常用的 SPSS 与 SAS 统计软件包进行简要介绍。

1.1 SPSS 系统介绍

1. SPSS 系统概况

SPSS 是软件英文名称 Statistical Package for the Social Sciences 的首字母缩写,即“社会科学统计软件包”。随着 SPSS 产品服务领域的扩大和服务深度的增加,SPSS 公司已于 2000

年正式将英文全称更改为 Statistical Product and Service Solutions, 意为“统计产品与服务解决方案”。这标志着 SPSS 这一软件的应用领域正向着更宽更远的方向发展。

SPSS 是世界上最早的统计分析软件之一, 由美国斯坦福大学的三位研究生于 20 世纪 60 年代末研制, 同时成立了 SPSS 公司, 并于 1975 年在芝加哥组建了 SPSS 总部。1984 年 SPSS 总部首先推出了世界上第一个统计分析软件微机版本 SPSS/PC+, 开创了 SPSS 微机系列产品的开发方向, 极大地扩充了它的应用范围, 并使其很快应用于自然科学、技术科学、社会科学等各个领域。SPSS 软件应用于通讯、医疗、银行、证券、保险、制造、商业、市场研究、科研教育等多个领域和行业, 是世界上应用最广泛的专业统计软件之一。

1994 至 1998 年间, SPSS 公司陆续购并了 SYSTAT 公司、BMDP 软件公司、Quantime 公司、ISL 公司等, 并将各公司的主打产品收纳在 SPSS 旗下, 使得 SPSS 公司由原来的单一统计产品的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, 成为在目前较为流行的“数据仓库”和“数据挖掘”领域前沿的一家综合统计软件公司。

2. SPSS 系统组成

SPSS 系统由多个模块构成, 其中 SPSS Base 为基本模块, 其余模块为 Advanced Models、Regression Models、Tables、Trends、Categories、Conjoint、Exact Tests、Missing Value Analysis 和 Maps, 分别用于完成某一方面的统计分析功能, 它们均需挂接在 SPSS Base 上运行。

3. SPSS 系统的操作方式

SPSS 最突出的特点就是操作界面较为友好, 输出结果清晰明了。它将几乎所有的功能都以统一、规范的界面展现出来, 使用 Windows 的窗口方式展示各种管理和分析数据方法, 使用对话框展示出各种功能选择项。用户只要掌握一定的 Windows 操作技能, 并且知道一些统计分析原理, 就可以使用该软件为特定的科研工作服务。SPSS 采用类似 Excel 表格的方式输入与管理数据, 数据接口较为通用, 能方便地从其他数据库中读入数据 (如 dBase、Excel、Lotus 等)。其统计过程包括了常用的统计方法, 如描述性统计、平均值比较、相关分析、回归分析、聚类分析、数据简化、生存分析、多重响应等, 每类中又下含同类多种统计过程, 比如回归分析中又分线性回归分析、非线性回归分析、曲线估计等多个统计过程, 而且每个过程中允许用户选择不同的方法及参数进行统计分析。因此, 除可以实现常规的各种统计外, 还可用来做一些不常用的分析处理, 完全可以满足专业与非统计专业研究人员的工作需要。存储时是专用的 SPO 格式, 可以转存为 HTML 格式和文本格式。对于熟悉老版本编程运行方式的用户, SPSS 还特别设计了语法生成窗口, 用户只需在菜单中选好各个选项, 然后单击“粘贴”按钮就可以自动生成标准的 SPSS 程序。极大地方便了中、高级用户。

4. SPSS 系统的功能特点

自 1985 年 SPSS INC 公司推出 SPSS/PC V1.0 以来, SPSS 的版本不断更新, 软件功能不断完善, 操作越来越简便, 与其他软件的接口也越来越多。现在的 SPSS for Windows 具有以下几种功能:

(1) SPSS 的数据编辑功能

在 SPSS 的数据编辑器窗口中, 不仅可以对打开的数据文件进行增加、删除、复制、剪

切和粘贴等常规操作,还可以对数据文件中的数据进行排序、转置、拆分、聚合、加权等操作,对多个数据文件可以根据变量或观察值进行合并。

(2) 表格的生成和编辑

利用 SPSS 可以生成数十种风格的表格,利用专门的编辑窗口或直接在查看器中可以编辑所生成的表格。在 SPSS 的高版本中,统计结果多被归纳为表格和(或)图形的形式。

(3) 图形的生成和编辑

利用 SPSS 可以生成数十种基本图和交互图。其中基本图包括条形图、线图、面积图、饼图、高低图、帕累托图、控制图、箱图、误差条图、散点图、直方图、P-P 概率图、Q-Q 概率图、序列图和时间序列图等,有的基本图又可进一步细分。交互图比基本图更漂亮,可有不同风格的二维、三维图。交互图包括条形交互图、点形交互图、线形交互图、带形交互图、饼形交互图、箱形交互图、误差条形交互图、直方交互图和散点交互图等。图形生成以后,可以对其进行编辑。

(4) 与其他软件的联接

SPSS 能打开 Excel、dBase、FoxBase、Lotus 1-2-3、Access、文本编辑器等生成的数据文件。SPSS 生成的图形可以保存为多种图形格式。现在的 SPSS 软件支持 OLE 技术和 ActiveX 技术,生成的表格和交互图对象可以与其他同样支持该技术的软件进行自动嵌入与链接。SPSS 还有内置的 VBA 客户语言,可以通过 Visual Basic 编程来控制 SPSS。

(5) SPSS 的统计功能

SPSS 的统计功能是 SPSS 的核心部分,利用该软件,几乎可以完成所有的数理统计任务。具体来说,SPSS 的基本统计功能包括:

样本数据的描述和预处理、假设检验(包括参数检验、非参数检验及其他检验)、方差分析(包括一般的方差分析和多元方差分析)、列联表、相关分析、回归分析、对数线性分析、聚类分析、判别分析、因子分析、对应分析、时间序列分析、生存分析、可靠性分析等。

5. 利用 SPSS 进行统计处理的基本过程

SPSS 功能强大,操作起来非常简单,这一特点从下面的流程图便可看出。SPSS 进行统计处理的基本过程如图 1-1 所示。

其基本步骤如下:

(1) 数据的录入

将数据以电子表格的方式输入到 SPSS 中,也可以从其他可转换的数据文件中读入数据。数据录入的工作分两个步骤:一是定义变量,

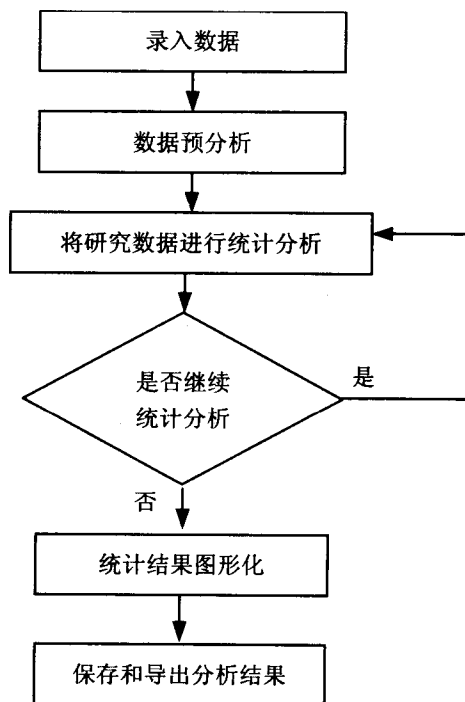


图 1-1 SPSS 统计处理流程图

二是输入变量值。

(2) 数据的预分析

在原始数据录入完成后,要对数据进行必要的预分析,如数据分组、排序、分布图、平均数、标准差的描述等,以掌握数据的基本特点和基本情况,保证后续工作的有效性,为采用正确的统计检验方法提供依据。

(3) 统计分析

按研究的要求和数据的情况确定统计分析方法,然后对数据进行统计分析。

(4) 统计结果图形化

在统计过程进行完成后,SPSS 会自动生成一系列数据表,其中包含了统计处理产生的整套数据。为了能更形象地呈现数据,需要利用 SPSS 提供的图形生成工具将所得数据图形化。SPSS 提供了许多图形来进行数据的图形化处理,使用时可根据数据的特点和研究的需求来进行选择。

(5) 保存和导出分析结果

数据结果生成之后,可将它以 SPSS 自带的数据格式进行存储。同时也可利用 SPSS 的输出功能以常见的数据格式进行输出,以供其他软件系统使用。

1.2 SAS 系统介绍

1. SAS 系统概况

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,进行 SAS 的维护、开发、销售和教育工作。1985 年以前,SAS 只能运行在大型机上,1985 年开始推出能运行在 IBM PC/XT 及其兼容机上的 SAS 版本。SAS 6.04 版是最后一个 DOS 版本。以后都是运行在 Windows 操作系统下的软件。SAS 软件被广泛用于金融、医药卫生、政府、教育、科研、生产、运输、通信和商业等各种从事数据管理与数据分析处理的领域中,是用于决策支持的太型集成信息系统。但该软件系统最早的功能限于统计分析,至今,统计分析功能也仍是它的重要组成部分和核心功能。它使用简单,能用极简单的命令去做你想做的各种数据整理和分析工作。

2. SAS 系统组成

SAS 系统是一个组合软件系统,它由多个功能模块组合而成,其基本部分是 BASE SAS 模块。BASE SAS 模块是 SAS 系统的核心,承担着主要的数据管理任务,并管理用户的使用环境,进行用户语言的处理,调用其他 SAS 模块和产品。当运行 SAS 系统时,首先必须启动 BASE SAS 模块,它本身不仅具有数据管理、程序设计及描述统计计算功能,还有与其他产品或模块共同构成一个完整的系统的功能。各模块的安装及更新都可通过其安装程序非常方便地进行。SAS 系统具有灵活的功能扩展接口和强大的功能模块,在 BASE SAS 的基础上,还可以增加如下不同的模块来增加不同的功能: 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。