



中国科学院教材建设专家委员会规划教材  
全国高等医药院校规划教材

案例版™

供医学硕士研究生和预防医学、公共管理、卫生检验、卫生监督、  
营养等专业本科生使用

# 卫生统计学 计算机操作教程

第2版

罗家洪 郭秀花 主编



中国科学院教材建设专家委员会规划教材

全国高等医药院校规划教材

供医学硕士研究生和预防医学、公共管理、卫生检验、卫生监督、  
营养等专业本科生使用

案例版™

# 卫生统计学

# 计算机操作教程

第2版

主编 罗家洪 郭秀花

科学出版社

北京

· 版权所有 侵权必究 ·

举报电话: 010-64030229; 010-64034315; 13501151303(打假办)

## 内 容 简 介

本书是《卫生统计学》(第2版,罗家洪、郭秀花主编)的配套教材,全书24章,以易学易用的国际统计软件SPSS软件包为基础,每章有“目的要求”;在“实例分析与电脑操作”中,除操作步骤外,还对操作和结果进行“解释”,帮助学生进一步理解。本书除与《卫生统计学》匹配的思考练习参考答案外,增加了有一定难度的“补充思考练习”,并附有“补充思考练习参考答案”。

本书是3年制医学硕士研究生和5年制预防医学专业本科生教材,也可供公共管理、卫生检验、卫生监督、营养等专业本科生使用,同时适用于医学研究生毕业后执业医师考试和博士研究生入学考试,也可作为在职医疗卫生人员继续教育培训教材,还可以作为在职医疗卫生人员科研参考书。

### 图书在版编目(CIP)数据

卫生统计学计算机操作教程:案例版 / 罗家洪, 郭秀花主编. —2版.  
—北京: 科学出版社, 2015.12  
中国科学院教材建设专家委员会规划教材·全国高等医药院校规划教材  
ISBN 978-7-03-044933-7

I. ①卫… II. ①罗… ②郭… III. ①计算机应用-卫生统计-医学院校-教材 IV. ①R195.1-39

中国版本图书馆CIP数据核字(2015)第128487号

责任编辑: 朱 华 / 责任校对: 胡小洁 彭 涛  
责任印制: 赵 博 / 封面设计: 陈 敬

版权所有, 违者必究。未经本社许可, 数字图书馆不得使用

科学出版社 出版

北京东黄城根北街16号

邮政编码: 100717

<http://www.sciencep.com>

三河市骏杰印刷有限公司 印刷

科学出版社发行 各地新华书店经销

\*

2007年8月第 一 版 开本: 787×1092 1/16  
2015年12月第 二 版 印张: 39  
2015年12月第四次印刷 字数: 949 000

定价: 69.00 元

(如有印装质量问题, 我社负责调换)

# 《卫生统计学计算机操作教程》

## 编委会

### 第2版

主 编 罗家洪 郭秀花  
副主编 李 健 姚应水 贾 红 刘启贵 赵若望  
董莉萍 曹明芹 程晓萍 许传志 李秀央  
谢红卫 尹家祥 孟 琼 罗艳侠  
学术秘书 毛 勇 陈 莹

编 委 (按姓氏笔画排序)

王 山(浙江大学)	罗艳侠(首都医科大学)
王 莹(辽宁医学院)	罗家洪(昆明医科大学)
王 倩(新疆医科大学)	金岳龙(皖南医学院)
韦丽琴(内蒙古科技大学包头医学院)	孟 琼(昆明医科大学)
毛 勇(昆明医科大学)	赵若望(内蒙古科技大学包头医学院)
尹家祥(大理大学)	郝金奇(内蒙古科技大学包头医学院)
孔丽娅(浙江大学)	胡志宏(北华大学)
叶运莉(四川医科大学)	侯瑞丽(内蒙古科技大学包头医学院)
田小兵(川北医学院)	姚应水(皖南医学院)
尔西丁·买买提(新疆医科大学)	姚雪梅(新疆医科大学)
朱 陶(川北医学院)	贺连平(皖南医学院)
刘军祥(四川医科大学)	贾 红(四川医科大学)
刘启贵(大连医科大学)	高 琦(首都医科大学)
齐亚莉(北华大学)	高晓凤(川北医学院)
许传志(昆明医科大学)	郭 晋(首都医科大学)
孙艳春(昆明医科大学)	郭秀花(首都医科大学)
杜瑞红(北华大学)	曹明芹(新疆医科大学)
巫秀美(大理大学)	常 巍(昆明医科大学)
李 岩(北华大学)	章丽娟(大理大学)
李 健(川北医学院)	董莉萍(北华大学)
李秀央(浙江大学)	喻 箴(昆明医科大学)
何 电(首都医科大学)	程晓萍(辽宁医学院)
宋 湛(川北医学院)	童玲玲(南华大学)
张俊辉(四川医科大学)	谢红卫(南华大学)
陈 莹(昆明医科大学)	詹志鹏(辽宁医学院)
陈 润(四川医科大学)	

# 前 言

本教材是根据教育部《2003~2007年教育振兴行动计划》的精神,本着与时俱进、改革与创新医学生培养模式、教学方法的宗旨,在借鉴国外先进教学模式——案例式教学模式的基础上,在供医学研究生和预防医学、卫生管理、卫生检验、卫生监督、营养等专业本科生使用的《卫生统计学》(案例版,第2版)(罗家洪、郭秀花主编)的基础上编写的《卫生统计学计算机操作教程》。本教程可与《卫生统计学》(案例版,第2版)配套使用,也可以与其他教材配套使用,还可以独立供本科生、研究生等学习 SPSS 统计软件包使用,本教程以 SPSS17.0 撰写计算机操作过程。

本教程是常年从事医学统计学和卫生统计学教学工作的各位主编、副主编及编委的经验总结,也是医学科研统计方法的综合反映,具有目的明确、重点突出、操作明细、解释详尽等主要特点。

《卫生统计学计算机操作教程》主要突出了以下几个方面的特点:

**1. 目的明确、重点突出** 为避免学习中主次不分的情况,每章的第一部分就向学生介绍本章的目的要求,让学生知道学习的重点和难点,应该掌握、熟悉和了解的内容。在分清主次的基础上,提高学习的主动性和积极性,事半功倍地学好卫生统计学。

**2. 操作明细、解释详尽** 为适应信息时代的发展,准确高效地分析处理各种医学科研资料,医学生有必要具备应用 SAS、SPSS、PEMS 等统计软件包的能力。为使医学生掌握 SPSS 的常用统计分析方法,本教程以教材为基础,以实例分析和计算机操作的形式,按统一模式构建例题(分析→操作→结果→解释),由浅入深地指导学生应用 SPSS。具体模式:首先,分析资料的研究目的、类型及其设计方案,据此决定应该采用何种统计方法进行分析,调用何种 SPSS 过程实现;然后,以图文并茂的形式指导学生进行每一步操作(建立数据库→输入数据→统计分析),并以统计术语的形式,翻译每个新出现的英语单词或短语;最后,根据 SPSS 的输出结果,逐行逐字地解释了每个结果的意义,并做出统计结论。

**3. 分析正误、解惑答疑** 为方便学生学习,本教材全面解答了《卫生统计学》(案例版,第2版)的思考练习题和本书内的补充思考练习题,均按解答、评析的形式,具体分析每道习题错误的原因或正确的道理,特别是针对反例式应用分析题,既详尽分析了错误原因,又给出正确做法。这样的解题方式,相信会让学生感到好学、易懂、过目不忘。

**4. 题库丰富、适应备考** 为适应医学生参加执业医师、博士研究生入学等综合性考试的备考需求,我们在教材思考练习的基础上,根据教学大纲的要求,增加了大量的补充思考练习题,并附多套模拟考卷,其中大部分应用分析题都是最近一年医学科研热点的浓缩,相信会有所帮助。

本教程的编写融入了主编、各位副主编和编委的大量心血,饱含着所有参编院校对莘莘学子的期望和爱心。在本教程编写和出版过程中,得到了科学出版社和各参编医科院校的大力支持;同时,昆明医科大学校长姜润生教授,副校长李松教授,副校长李燕主任医师,教务处章宗籍处长,公共卫生学院殷建忠院长、王贵义书记等也给予了大力支持并提出了宝贵意见,我谨代表全体编委一并鸣谢。

本教程是全新案例版教材,限于我们的水平,可能有不少的缺点和错误,热忱欢迎广大师生和同行批评指正,并希望各医学院校在使用过程不断总结经验,提出宝贵意见,以便进一步修改完善。

罗家洪

2015年11月于春城昆明

# 目 录

## 前言

<b>第一章 绪论</b> .....	1
一、目的要求.....	1
二、实例分析与电脑操作.....	1
三、思考练习参考答案.....	11
四、补充思考练习.....	15
五、补充思考练习参考答案.....	17
<b>第二章 计量资料的统计描述</b> .....	19
一、目的要求.....	19
二、实例分析与电脑操作.....	19
三、思考练习参考答案.....	27
四、补充思考练习.....	33
五、补充思考练习参考答案.....	35
<b>第三章 分类资料的统计描述</b> .....	39
一、目的要求.....	39
二、实例分析与电脑操作.....	39
三、思考练习参考答案.....	44
四、补充思考练习.....	50
五、补充思考练习参考答案.....	53
<b>第四章 统计表与统计图</b> .....	55
一、目的要求.....	55
二、实例分析与电脑操作.....	55
三、思考练习参考答案.....	83
四、补充思考练习.....	89
五、补充思考练习参考答案.....	91
<b>第五章 总体均数估计与假设检验</b> .....	95
一、目的要求.....	95
二、实例分析与电脑操作.....	95
三、思考练习参考答案.....	108
四、补充思考练习.....	112
五、补充思考练习参考答案.....	114
<b>第六章 方差分析</b> .....	116
一、目的要求.....	116
二、实例分析与电脑操作.....	116
三、思考练习参考答案.....	141

四、补充思考练习 .....	147
五、补充思考练习参考答案 .....	150
<b>第七章 二项分布与 Poisson 分布 .....</b>	<b>153</b>
一、目的要求 .....	153
二、实例分析与电脑操作 .....	153
三、思考练习题参考答案 .....	162
四、补充思考练习 .....	167
五、补充思考练习参考答案 .....	169
<b>第八章 <math>\chi^2</math> 检验 .....</b>	<b>172</b>
一、目的要求 .....	172
二、实例分析与电脑操作 .....	172
三、思考练习参考答案 .....	186
四、补充思考练习 .....	193
五、补充思考练习参考答案 .....	196
<b>第九章 秩和检验 .....</b>	<b>199</b>
一、目的要求 .....	199
二、实例分析与电脑操作 .....	199
三、思考练习参考答案 .....	213
四、补充思考练习 .....	221
五、补充思考练习参考答案 .....	223
<b>第十章 直线相关与回归 .....</b>	<b>225</b>
一、目的要求 .....	225
二、实例分析与电脑操作 .....	225
三、思考练习参考答案 .....	230
四、补充思考练习 .....	235
五、补充思考练习参考答案 .....	237
<b>第十一章 调查设计 .....</b>	<b>239</b>
一、目的要求 .....	239
二、实例分析与电脑操作 .....	239
三、思考练习参考答案 .....	241
四、补充思考练习 .....	247
五、补充思考练习参考答案 .....	250
<b>第十二章 实验设计 .....</b>	<b>251</b>
一、目的要求 .....	251
二、实例分析与电脑操作 .....	251
三、思考练习参考答案 .....	256
四、补充思考练习 .....	259
五、补充思考练习参考答案 .....	260
<b>第十三章 剂量反应 .....</b>	<b>262</b>
一、目的要求 .....	262



二、实例分析与电脑操作	262
三、思考练习与参考答案	274
四、补充思考练习	280
五、补充思考练习参考答案	283
<b>第十四章 Ridit 分析</b>	284
一、目的要求	284
二、实例分析与电脑操作	284
三、思考练习与参考答案	293
四、补充思考练习	300
五、补充思考练习参考答案	302
<b>第十五章 寿命表</b>	304
一、目的要求	304
二、实例分析与电脑操作	304
三、思考练习参考答案	312
四、补充思考练习	318
五、补充思考练习参考答案	321
<b>第十六章 随访资料生存分析</b>	326
一、目的要求	326
二、实例分析与电脑操作	326
三、思考练习参考答案	339
四、补充思考练习	348
五、补充思考练习参考答案	351
<b>第十七章 病例对照研究设计与分析</b>	353
一、目的要求	353
二、实例分析与电脑操作	353
三、思考练习参考答案	362
四、补充思考练习	367
五、补充思考练习参考答案	369
<b>第十八章 队列研究</b>	370
一、目的要求	370
二、实例分析与电脑操作	370
三、思考练习参考答案	373
四、补充思考练习	378
五、补充思考练习参考答案	382
<b>第十九章 多重线性回归与相关</b>	386
一、目的要求	386
二、实例分析与电脑操作	386
三、思考练习参考答案	392
四、补充思考练习	399
五、补充思考练习参考答案	404

第二十章 Logistic 回归	405
一、目的要求	405
二、实例分析与电脑操作	405
三、思考练习参考答案	431
四、补充思考练习	446
五、补充思考练习参考答案	449
第二十一章 聚类分析	451
一、目的要求	451
二、实例分析与电脑操作	451
三、思考练习参考答案	469
四、补充思考练习	474
五、补充思考练习参考答案	477
第二十二章 判别分析	485
一、目的要求	485
二、实例分析与电脑操作	485
三、思考练习参考答案	503
四、补充思考练习	513
五、补充思考练习参考答案	515
第二十三章 主成分分析和因子分析	517
一、目的要求	517
二、实例分析与电脑操作	517
三、思考练习参考答案	536
四、补充思考练习	547
五、补充思考练习参考答案	553
第二十四章 诊断和筛检试验的设计与分析	557
一、目的要求	557
二、实例分析与电脑操作	557
三、思考练习参考答案	565
四、补充思考练习	572
五、补充思考练习参考答案	574
附录 模拟考卷	577
附录一 2015 级硕士研究生《卫生统计学》模拟试卷(A 卷)	577
附录二 2015 级在职研究生课程进修班《卫生统计学》模拟试卷(A 卷)	585
附录三 2011 级预防医学专业《卫生统计学》模拟试卷(A 卷)	593
附录四 2013 级卫生检验与检疫、食品卫生与营养学《卫生统计学》模拟试卷(A 卷)	600
附录五 2013 级公管专业《卫生统计学》模拟试卷(A 卷)	607

# 第一章 绪 论

## 一、目的要求

### 【了解】

医学统计学的概念及主要内容。

### 【熟悉】

统计工作的基本步骤。

### 【掌握】

1. 统计资料的类型及其相应的分析方法。
2. 统计学的几个基本概念。

### 【重点难点】

1. 重点是统计资料的类型和统计学的基本概念。
2. 难点是正确区别统计资料的类型。

## 二、实例分析与电脑操作


### (一) SPSS 的概述

SPSS (Statistical Package for Social Sciences, 社会科学统计软件包) 是美国 SPSS 公司开发的大型统计软件包, 适用于社会科学、医学、经济学、心理学等各个领域。在国际学术界有条不成文的规定: 凡是用 SAS 和 SPSS 统计分析的结果, 在国际学术交流中可以不必说明算法。本书的实例分析是以 SPSS Statistics 17.0 版本为基础, 结合教材的例题和思考练习题, 具体介绍 SPSS 的应用。

### (二) SPSS 的启动

双击桌面 SPSS 快捷图标或单击开始 → 程序 → SPSS Statistics → SPSS Statistics, 即可启动(激活)SPSS 的数据编辑窗口; SPSS 数据编辑窗口的第一行是主菜单, 共包含 11 个菜单项, 第二行是数据编辑快捷工具栏。单击窗口左下角的 Variable View (变量视窗) 或 Data View (数据视窗), 可以在变量视图窗口(图 1-1)和数据视图窗口(图 1-2)之间互相切换。下面介绍主要菜单。

### (三) File 菜单(文件操作)

1. **New** 新建文件, 快捷键为 Ctrl+N。包括新建 Data (数据)、Syntax (程序)、Output (输出结果)、Draft Output (草案输出结果)、Script (脚本)。
2. **Open** 打开文件, 快捷按钮为 , 快捷键为 Ctrl+O。包括打开 Data (数据)、Syntax (程序)、Output (输出结果)、Script (脚本)、Other (其他类型文件)。
3. **Open Database** 打开数据库, 包括 New Query (新建查新)、Edit Query (编辑查新) 和 Run Query (运行查新)。

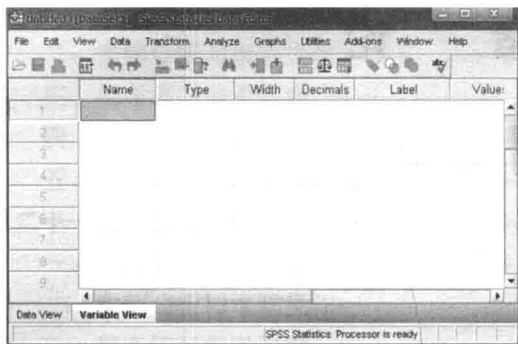


图 1-1 SPSS 的 Variable View 窗口

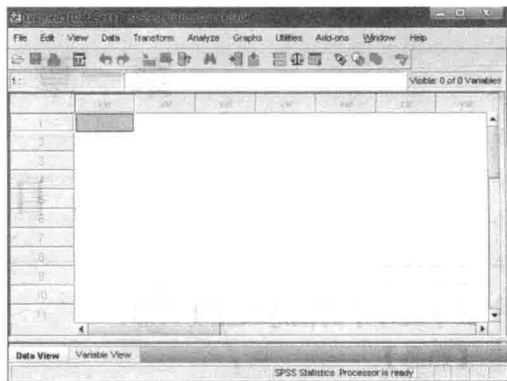



图 1-2 SPSS 的 Data View 窗口

4. **Read Text Data** 打开文本数据。

5. **Close** 关闭, 快捷键 Ctrl+F4。

6. **Save** 保存文件, 快捷按钮为 , 快捷键为 Ctrl+S。保存文件时, 文件的格式不同, 其后缀名也不同。SPSS 数据文件的默认后缀名为“.sav”; 程序文件的默认后缀名为“.SPS”; 输出结果的默认后缀名为“.spv”。

7. **Save as** 另存为…。若要保留原数据文件, 可用 **Save as** 将修改过的数据以新文件名保存。可用 **Save as** 将 SPSS 数据文件转换为 Excel、SAS、Stata 等数据文件。

8. **Save All Data** 保存所有数据。

9. **Export To DataBase** 导出到数据库, 将 SPSS 数据文件导出为 Excel、SAS、Stata 等数据文件, 或用 **Export...** 将 SPSS 结果转换为 Word 文件等。

10. **Mark File Read Only** 将文件标记只读。

11. **Rename Dataset** 重新命名数据集。

12. **Display Data File Information** 显示数据文件信息, 包括 **Working File**(工作文件)和 **External File**(外部文件)。

13. **Cache Data** 缓存数据。

14. **Stop Processor** 停止处理程序, 快捷键 Ctrl+Period。

15. **Switch Server** 开关服务器。

16. **Predictive Enterprise Repository** 预测企业存储库, 包括 **Connect** (链接)、**Store From SPSS Statistic** (从 SPSS Statistics 存储)、**Publish to Web** (发布到互联网)、**Add a File** (添加一个文件)、**Retrieve to SPSS Statistics** (检索到 SPSS 统计软件包)、**Download a File** (下载一个文件)。

17. **Print Preview** 打印预览。


18. **Print** 打印, 快捷键 Ctrl+P。


19. **Recently Used Data** 最近使用过的数据。

20. **Recently Used File** 最近使用过的文件。

21. **Exit** 退出 SPSS 系统, 快捷键为 Alt+F4。

#### (四) Edit 菜单(编辑)


1. **Undo** 撤消单元输入值, 快捷按钮为 , 快捷键为 Ctrl+Z。

2. **Redo** 恢复单元输入值, 快捷按钮为, 快捷键为 Ctrl+Y。

3. **Cut/Copy/Paste/Clear** 剪切/复制/粘贴/清除数据或变量, 此四项既可在菜单中选择, 也可以选中单元格或变量后, 单击鼠标右键调出。前三项可依次用快捷键 Ctrl+X、Ctrl+C、Ctrl+V。

4. **Insert Variables** 插入变量。

5. **Insert Cases** 插入个案。

6. **Find** 查找数据, 快捷按钮为, 快捷键为 Ctrl+F。

7. **Find Next** 查找下一个, 快捷键为 F3。

8. **Replace** 替换, 快捷键为 Ctrl+H。

9. **Go to Case** 转向个案。


10. **Go to Case** 转向变量。

11. **Go to Imputation** 转向归因。

12. **Options** 选项, 选择 SPSS 参数。通过 **Options** (选项) 语言选择调整, 实现中英文 SPSS 统计软件包转换。

### (五) View 菜单(视图)


1. **Status Bar** 状态栏, 显示或隐藏状态栏。

2. **Toolbars** 工具栏定义, 系统默认为 **Data Editor** (数据编辑) 工具栏。小技巧: 工具栏中的为 **Dialog Recall** (重复调用对话框), 单击之, 所弹出的下拉列表中就依次列出了最近几次使用的一些过程名(包括非统计分析过程), 直接从中选择需要的过程, 就可重复已做过的分析。

3. **Menu Editor** 菜单编辑器, 可自定义菜单。

4. **Fonts** 自定义字体。

5. **Grid Lines** 显示或隐藏表格线。

6. **Value Labels** 变量值标签/变量值显示与隐藏切换, 系统默认不显示变量值标签, 快捷按钮为。变量值标签的定义方法参见例 4-1。

7. **Mark Imputed Data** 标识归因数据。

8. **Customize Variable View** 自定义变量视图。

9. **Variables/Data** 变量视窗/数据视窗切换, 快捷键为 Ctrl+T。

### (六) Data 菜单(数据操作)

1. **Define Variable Properties** 定义变量属性。

2. **Copy Data Properties** 复制变量属性。

3. **New Custom Attribute** 新建设定属性。

4. **Define Dates** 定义日期, 主要用于时间序列模型。

5. **Define Multiple Response Set** 定义多重响应集。

6. **Validation** 验证, 包括 **Load Predefined Rules** (加载预先定义规则)、**Define Rules** (定义规则) 和 **Validate Data** (验证数据)。


7. **Identify Duplicate Cases** 标识重复个案。


8. **Identify Unusual Cases** 标识异常个案。


9. **Sort Cases** 个案排序, 按个案排序, 排序方式有升序(**Ascending**)和降序(**Descending**)

两种。

**10. Sort Variables** 变量排序, 按变量排序, 排序方式有升序(Ascending)和降序(Descending)两种。

**11. Insert Variable** 插入变量, 在当前列插入新变量, 快捷按钮为 。在 V17.0 中, 仅有快捷按钮, 菜单不再显示。

**12. Insert Cases** 插入个案, 在当前行插入新个案, 快捷按钮为 。在 V17.0 中, 仅有快捷按钮, 菜单不再显示。

**13. Go to Case** 个案定位, 到达指定记录号的个案, 快捷按钮为 。当数据较多时, 非常用。在 V17.0 中, 仅有快捷按钮, 菜单不再显示。

**14. Transpose** 数据行列转置, 可以将原来的一条记录转成为一个变量, 或将原来的一个变量转成为一个记录。原变量名会自动保存在系统生成的 `case_b` 的字符变量中。

**15. Restructure** 数据重排, 例如, 进行随机区组设计资料的秩和检验时, 处理组各为一个变量, 但是, 当对其进行两两比较时(秩变换分析方法), 变量定义需按随机区组设计方差分析的形式, 这时就可以调用 **Restructure** 过程完成这一任务(参见例 9-8)。

**16. Merge Files** 合并数据文件。


(1) **Add Cases** 增加个案(或记录), 从外部数据文件中增加个案到当前数据文件中, 称为纵向合并。注意: 相互合并的数据文件中应该有相同的变量;


(2) **Add Variables** 增加变量, 从外部数据文件中增加变量到当前数据中, 称为横向合并。注意: 横向合并时默认按照相同个案(或记录)数进行合并, 否则会丢弃一部分记录。


**17. Aggregate** 数据分类汇总, 分类汇总是按指定的分类变量(选入 **Break Variables** 框)对观察值(选入 **Aggregate Variables** 框)进行分组, 在 **Function** 子对话框中定义需描述的统计量。

**18. Orthogonal Design** 正交设计, 用于自动生成正交设计表格。分为 **Generate**(生成)、**Display**(显示)两个过程。

**19. Copy Dataset** 复制数据集。

**20. Split File** 拆分数据文件, 用于数据文件的分组处理, 快捷按钮为 。选择某分组变量(如性别、职业、实验分组等)对数据文件进行分组后, 就可以对数据文件进行分组统计分析。

**21. Select Cases** 选择个案, 用于选择需分析的个案, 快捷按钮为 。当不需要分析某变量的全部数据时, 可调用该过程进行选择。

**22. Weight Cases** 个案加权或频数加权, 快捷按钮为 。在使用频数表格式录入数据时(如  $\chi^2$  检验), 相同取值的观察(如处理、疗效)或组段只录入一次, 另加一个频数变量用于记录该数值共出现了多少次, 分析时需要用 **Weight Cases** 过程将频数加权即将频数变量的数据乘以组段。(参见例 2-2、例 8-1)

## (七) Transform 菜单(数据转换)

**1. Compute Variable** 计算变量, 用于对变量进行计算。主要特点:

(1) 目标变量可以是新变量, 也可以是已有的变量。

(2) **Compute** 过程中赋给变量的值可以是一个常数, 也可以是从已有变量值或系统函数计算而得的值, 系统函数可以从 **Function** 框中选择。

(3) 操作记录既可以是所有记录, 也可以设定逻辑条件。**Compute** 过程可以直接调用菜单

进行(参见例 5-4),也可以编写程序进行(参见例 3-1)。

**2. Count Values with Cases** 计数,用于计数每个个案在多个变量中相同数值的发生次数,或某个值或某些值在某个变量取值中出现的次数,并生成一个新变量。

**3. Shift Values** 转换值

**4. Recode Into Same Variables** 在相同变量中重新编码, **Recode** 过程即重新编码(或重新赋值),用于将原变量值按照某种一一对应的关系生成新变量值。

**5. Recode Into Different Variables** 在不同变量中重新编码。(参见例 2-1)

**6. Automatic Recode** 自动重新编码,自动按原变量大小生成新变量,功能与 Rank Cases 过程类似。

**7. Visual Binning** 可视离散化,根据现存的连续变量进行分组,并产生一个新的分组变量。可用于将连续变量创建一个分类变量,也可以将多个有序分类合并为少数的分类变量。

**8. Optimal Binning** 最优离散化,将一个或多个连续变量的值分布到“块”中进行离散化。根据块进行分析。

**9. Rank Cases** 个案排秩,用于变量的秩变换。可根据某变量的大小进行编秩,并将秩次结果存入新变量。

**10. Date and Time Wizard** 日期和时间向导。

**11. Create Time Series** 创建时间序列,用于创建时间序列变量。

**12. Replace Missing Value** 替换缺失值,用于时间序列模型数据的预处理。

**13. Random Number Generator** 随机数字生成器,用于设定伪随机种子数,默认情况下随机种子随着时间在不停改变,这样所计算出的随机数值无法重复。在临床试验等情况中,可以人为指定一个种子,结果就可重现。

**14. Run Pending Transform** 运行挂起的转换。

## (八) Analyze 菜单(统计分析)

**1. Reports 菜单(统计报表)**

(1) **Codebook** 过程:代码本,显示活动数据集中所有或指定变量和多重响应集的变量信息。

(2) **OLAP Cubes (Online Analytical Processing Cubes)** 过程:在线分层分析,用于对分组变量的各组之间或不同变量之间进行统计,可计算 **Sum**(总和)、**Number of case**(个案例数)、**Mean**(均数)、**Median**(中位数)、**分组中位数 (Grouped Median)**、**Standard Error of mean**(标准误)等。

(3) **Case Summaries** 过程:个案汇总,用于计算 **Number of case**(例数)、**Mean**(均数)、**Median**(中位数)、**Harmonic mean**(调和均数)、**Geometric mean**(几何均数)、**分组中位数 (Grouped Median)**、**Standard deviation**(标准差)、**Standard Error of mean**(标准误)等。注:几何均数和调和均数的计算常调用 **Case Summaries** 过程实现。

(4) **Report summaries in row** 过程:按行报表汇总,用于按行形式表达变量或指标的统计量输出报告。可计算 **Sum of values**(总和)、**Mean of values**(均数)、**Standard deviation**(标准差)、**Percentage above value**(高于某值的百分数)、**Percentage below value**(低于某值的百分数)、**Percentage inside low ...high ...**(介于 **Low** 与 **High** 之间的百分比)等。

(5) **Report summaries in columns** 过程:按列报表汇总,用于按列形式表达变量或指标的统计量输出报告。可计算按行报表汇总的指标。

## 2. Descriptive Statistics 菜单(描述性统计分析)

(1) **Frequencies** 过程: 频数分布分析, 用于生成详细的频数表, 并可按要求计算描述统计量, 生成常用的条图、圆图、直方图等(参见例 2-1)。

(2) **Descriptive** 过程: 描述性分析, 进行一般性的统计描述, 适用于服从正态分布的计量资料。

(3) **Explore** 过程: 探索行分析, 进行数据分布状况的探索性分析, 可进行分层统计描述, 例如正态性检验, 不同性别的统计描述比较。

(4) **Crosstabs** 过程: 行列表(或列联表, 或交叉表)分析, 进行行列表资料的分析, 用于分类资料/等级资料的统计描述及各种假设检验, 例如  $\chi^2$  检验、McNemar 检验等(参见第 8 章)。

(5) **Ratio** 过程: 比率统计分析, 用于对两个连续型变量计算相对比指标。

(6) **P-P Plots** 过程: 绘制 P-P 图。

(7) **Q-Q Plots** 过程: 绘制 Q-Q 图。

3. **Tables**(统计表格)菜单 包括 **Custom Tables**(设定表格)、**Multiple Response Set**(定义多重响应集)过程。

4. **RFM Analysis** 菜单 包括 **Transaction Data**(交易资料)和 **Customer Data**(顾客资料)过程。

## 5. Compare Means 菜单(均数间的比较)

(1) **Means** 过程: 平均数分析, 用于对样本进行统计描述, 即检验前的预分析。可计算各个指标的平均值, 或根据一个或多个分类指标计算相关指标的统计量, 并可进行单因素方差分析等。

(2) **One-Sample T Test** 过程: 单样本  $t$  检验, 进行单样本  $t$  检验。(参见例 5-1)

(3) **Independent-Samples T Test** 过程: 独立样本  $t$  检验, 进行完全随机设计两样本均数比较的  $t$  检验。(参见例 5-3)

(4) **Paired-Samples T Test** 过程: 配对  $t$  检验, 进行配对  $t$  检验。(参见例 5-2)

(5) **One-Way ANOVA** 过程: 单因素方差分析, 进行完全随机设计的方差分析。(参见例 6-1)

## 6. General Linear Model 菜单(一般线性模型)

(1) **Univariate** 过程: 单变量方差分析, 当应变量为一个时, 进行随机区组设计的方差分析。(参见例 6-2)

(2) **Multivariate** 过程: 多变量或多元方差分析, 当应变量为多个时, 进行多元方差分析。

(3) **Repeated Measures** 过程: 重复测量方差分析, 进行重复测量资料的方差分析。

(4) **Variance Components** 过程: 方差成分分析, 对层次数据拟合方差成分模型分析。

7. **Generalized Linear Models** 菜单(广义线性模型) 包括 **Generalized Linear Models**(广义线性模型)过程和 **Generalized Estimating Equations**(广义估计方程)过程。

## 8. Mixed Models(混合模型)菜单 Linear(线性混合模型)。

## 9. Correlate 菜单(相关分析)

(1) **Bivariate** 过程: 双变量相关分析, 进行两个/多个变量间的参数/非参数相关分析。如果是多个变量, 则给出两两相关的分析结果。(参见例 10-1、例 10-2)

(2) **Partial** 过程: 偏相关分析, 进行偏相关分析。如果需要进行分析的两个变量的取值均受到其他变量的影响, 就可以利用偏相关分析对其他变量进行控制, 输出控制其他变量影响后的相关系数。

(3) **Distances** 过程: 距离相关分析, 可对同一变量内部各观察单位的数值或各个不同变量间进行相似性或不相似性(距离)分析。



## 10. Regression 菜单(回归分析)

(1) **Linear** 过程: 线性回归分析, 用于拟合线性回归模型, 包括简单线性回归(1 个自变量)和多重线性回归(多个自变量)。(参见例 10-3、例 10-4)

(2) **Curve Estimation** 过程: 曲线参数估计法, 用于拟合常用曲线。

(3) **Partial Least Squares Regression** 过程: 部分(偏)最小平方回归。

(4) **Binary Logistic** 过程: 二分类 **Logistic** 回归分析, 进行二分类资料的 **Logistic** 回归分析。(参见例 19-1)

(5) **Multinomial Logistic** 过程: 多分类 **Logistic** 回归分析, 进行无序多分类资料的 **Logistic** 回归分析。

(6) **Ordial** 过程: 有序分类回归, 进行有序分类资料的 **Logistic** 回归分析。

(7) **Probit** 过程: 概率单位法, 用于分析剂量反应关系。(参见例 13-1、例 13-2)

(8) **Nonlinear** 过程: 非线性回归, 对非线性关系资料进行回归分析。

(9) **Weight Estimation** 过程: 权重估计法, 对观测值精确度不同的资料进行回归分析。

(10) **2-Stage Least Squares** 过程: 二阶段最小二乘回归。

(11) **Optimal Scalling (CATREG)**: 分类回归分析, 对分类资料进行回归分析。

## 11. Loglinear 菜单(对数线性模型分析)

(1) **General** 过程: 广义线性模型分析, 进行一般对数线性模型分析, 主要用于证实性研究。

(2) **Logit** 过程: Logit 对数线性模型分析, 当应变量为两分类时, 可以用 Logit 过程提供的对数线性模型来分析。

(3) **Model Selection** 过程: 模型选择对数线性分析, 用于拟合分层对数线性模型。

**12. Neural Networks (神经网络) 菜单** 包括 **Multilayer Perceptron** (多层感知器) 和 **Radial Basis Function** (径向基函) 过程。

## 13. Classify 菜单(分类分析)

(1) **TwoStep Cluster** 过程: 二阶段聚类分析, 可进行两步聚类分析。

(2) **K-Means Cluster** 过程: 快速聚类分析, 对记录进行快速聚类, 又称为 *K*-均值聚类分析(快速聚类分析、逐步聚类分析、*Q* 型聚类分析)。

(3) **Hierarchical Cluster** 过程: 系统聚类分析, 对数据进行系统聚类或分层聚类分析。

(4) **Tree** (决策树) 过程。

(5) **Discriminant** 过程: 判别分析, 对数据进行判别分析。

(6) **Nearest Neighbor Analysis** 过程: 最近邻元素分析。

## 14. Dimension Reduction 菜单(降维分析)

(1) **Factor** 过程: 因子分析, 对数据进行因子分析/主成分分析。

(2) **Correspondence Analysis** 过程: 对应分析, 对数据进行简单对应分析。

(3) **Optimal Scaling** 过程: 最优尺度分析, 可进行同质性分析、最优尺度分析。

## 15. Scale 菜单(信度分析)

(1) **Reliability Analysis** 过程: 信度分析或可靠性分析, 进行内在的信度分析, 用于评价问卷的稳定性和可靠性。

(2) **Multidimensional Unfolding (PREFSCAL)** 过程: 多维展开分析, 对数据进行多维展开分析, 尝试查找允许您直观地检查两组对象之间关系的公共定量尺度。

(3) **Multidimensional Scaling (PROXSCAL)** 过程: 多维邻近尺度分析, 对数据进行多维邻