

全国高等医药院校配套教材

医学课程学习纲要与强化训练

医学统计学学习指导

罗家洪 唐 军 主编



科学出版社
www.sciencep.com

全国高等医药院校配套教材
医学课程学习纲要与强化训练

医学统计学学习指导

罗家洪 唐 军 主编

科学出版社

北 京

全国高等院校医学类教材
医学统计学学习指导与强化训练

内 容 简 介

本书是全国高等医学院校规划教材案例版《医学统计学》配套教材,由原班作者亲自编写,帮助、指导学生或读者学习医学统计学。每一章内容均按照统一模式介绍目的要求、实例分析与电脑操作、习题参考答案、补充习题、补充习题参考答案。该书突破传统模式,采用案例式教学,增加易学易用的国际统计软件包 SPSS 操作,配合案例版《医学统计学》教材教学,结合医学科研实例、案例引导教学,由浅入深、层次分明、针对性强、突出“三基”内容、知识点明确、学生易学易懂。

图书在版编目(CIP)数据

医学统计学学习指导 / 罗家洪,唐军主编. —北京:科学出版社,2007
全国高等医药院校配套教材·医学课程学习纲要与强化训练

ISBN 978-7-03-019516-6

I. 医… II. ①罗…②唐… III. 医学统计-医学院校-教学参考资料
IV. R195.1

中国版本图书馆 CIP 数据核字(2007)第 117704 号

责任编辑:夏 宇 李国红 / 责任校对:李奕莹

责任印制:刘士平 / 封面设计:黄 超

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>

铭浩彩色印装有限公司印刷

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

2007 年 8 月第一次印刷 印张:15

印数:1—5 000 字数:347 000

定价:24.00 元

(如有印装质量问题,我社负责调换(环伟))

《医学统计学学习指导》编委会

主 编 罗家洪 唐 军
副主编 贾 红 祁爱琴 赵若望 董莉萍 石武祥
薛 茜 袁秀琴 程晓萍 孟 琼 毛 勇
学术秘书 毛 勇

编委名单 (按姓氏笔画排列)

马金凤 (新疆医科大学)	马洪林 (辽宁医学院)
王良君 (辽宁医学院)	韦丽琴 (内蒙古科技大学医学院)
毛 勇 (昆明医学院)	石武祥 (大理学院)
石德文 (滨州医学院)	刘军祥 (泸州医学院)
祁爱琴 (滨州医学院)	孙红卫 (滨州医学院)
杜瑞红 (北华大学)	杨庭仕 (大理学院)
李晓梅 (昆明医学院)	何利平 (昆明医学院)
张玉娥 (内蒙古科技大学医学院)	陈冬富 (泸州医学院)
罗家洪 (昆明医学院)	孟 琼 (昆明医学院)
赵若望 (内蒙古科技大学医学院)	赵拥军 (滨州医学院)
胡志宏 (北华大学)	袁秀琴 (南华大学)
贾 红 (泸州医学院)	徐天和 (滨州医学院)
唐 军 (滨州医学院)	曹明芹 (新疆医科大学)
章丽娟 (大理学院)	董莉萍 (北华大学)
喻 箴 (昆明医学院)	程晓萍 (辽宁医学院)
童玲玲 (南华大学)	谢红卫 (南华大学)
薛 茜 (新疆医科大学)	

前 言

《医学统计学学习指导》是中国科学院教材建设专家委员会规划教材、高等医学院校规划教材、案例版《医学统计学》的配套教材。案例版《医学统计学》是在借鉴国外先进教学模式的基础上编写的适合中国国情的全新案例式教材,已由科学出版社于2006年8月出版。本教材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,具有先进性、科学性、启发性、适用性、易学易用等特点。

医学统计学的学习注重理论知识的掌握,更注重在实际工作中的灵活应用。为配合案例版《医学统计学》的学习,帮助广大医学生及医务工作者有效掌握医学统计学的常用统计分析方法及其原理,灵活应用统计知识搞好科研工作,我们以案例版《医学统计学》为基础,由原版作者亲自编写,帮助、指导学生及医务工作者学习医学统计学。每一章内容结构均由目的要求、实例分析与电脑操作、思考练习参考答案、补充思考练习、补充思考练习参考答案共五个模块构成。最后附有四套模拟考试题。

1. 《医学统计学学习指导》各个模块的具体内容

(1) 目的要求:包括了解、熟悉、掌握、重点难点几个方面:①了解:只要知道有这些内容,不必要掌握。②熟悉:需要熟记的内容,知道来龙去脉及其应用。③掌握:不仅需要熟记,还要会计算、会灵活应用。④重点难点:教学或学习要点或关键点以及疑难点。

(2) 实例分析与电脑操作:以教材实例介绍本章统计方法在 SPSS 统计软件包上操作:①分析:该资料属于那一种类型资料,属何种设计方案,应该用何种统计方法处理。②操作:a. 详细介绍在 SPSS 上怎样建立数据库、输入数据、调用哪一个过程、选择哪些方法或统计指标等。b. 对于不能直接调用 SPSS 菜单计算的资料,采用编写小程序方法处理。③结果:列出 SPSS 输出的全部或主要结果。④解释:对 SPSS 输出的结果给予解释,并结合专业做出结论。

(3) 思考练习参考答案:对案例版《医学统计学》教材中的思考练习进行详细解答和评析。

(4) 补充思考练习:补充大量的是非题、选择题和应用分析题。

(5) 补充思考练习参考答案:对补充的是非题、选择题和应用分析题进行解答。

(6) 附录《医学统计学》期末模拟考试题:附有不同类型、不同难易程度的四套《医学统计学》期末模拟考试题。

2. 《医学统计学学习指导》主要特点和创新之处

(1) 目的明确、重点突出:为避免学习中主次不分的情况,每章的第一部分介绍学习的目的要求,明确学习的重点和难点,应该掌握、熟悉和了解的内容。在分清主次的基础上,调动学习者的主动性和积极性,以事半功倍地学好医学统计学。

(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,医学生和医务工作者必须具备应用 SAS、SPSS、PEMS 等统计软件包的能力。为使学

习者掌握 SPSS 的常用统计分析方法,《医学统计学学习指导》以教材为基础,以实例分析和电脑操作的形式,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学习 SPSS 的应用。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习者进行每一步操作(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,并结合专业知识,做出统计结论。

(3) 灵活编程、易学易用:采用 SPSS 编程窗口,编写了常用小程序,处理不能直接调用 SPSS 菜单的数据或资料(如用均数、标准差资料进行 t 检验或 z 检验),以补充 SPSS 菜单功能的不足,并突出了 SPSS 的强大编程功能。学习者只要将程序输入,简单修改数据或部分程序就可计算输出结果。

(4) 分析正误、解惑答疑:全面解答了案例版《医学统计学》的思考练习题和本书内的补充思考练习题。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,让学习者感到好学、易懂、过目不忘。

(5) 题库丰富、适应备考:为适应医学生参加执业医师、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,其中大部分应用分析题都是医学科研热点的浓缩,具有时代性。最后四套《医学统计学》期末模拟考试题,可供不同层次学习者应付各种考试的实战模拟训练,也可供教师及有关人员出题参考。

总之,《医学统计学学习指导》突破传统模式,在内容、框架、体例上有所创新;采用案例式教学,增加易学易用的国际统计软件包 SPSS 操作;配合案例版《医学统计学》教学,结合医学科研实例、案例引导教学,由浅入深、层次分明、针对性强;突出“三基”内容、知识点明确、易学易懂。

《医学统计学学习指导》凝聚了各位编者的大量心血,是各位编者多年来教学经验总结或科研成果的综合反映。在编写和出版过程中,得到了科学出版社和各参编医科院校的大力支持;同时,昆明医学院院长姜润生教授,副院长李燕主任医师,副院长李松教授,教务处杨凌处长、李飞副处长、于建云副处长,公共卫生学院刘革院长、张有福书记、万崇华副院长、吴锡南副院长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。

由于编者的水平和编写经验,难免存在缺点和错误,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中不断总结经验,提出宝贵意见,以便进一步修改完善。

罗家洪

2007年4月于春城昆明

目 录

前言	
第一章 绪论	(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	(9)
四、补充思考练习	(11)
五、补充思考练习参考答案	(12)
第二章 计量资料的统计描述	(14)
一、目的要求	(14)
二、实例分析与电脑操作	(14)
三、思考练习参考答案	(20)
四、补充思考练习	(23)
五、补充思考练习参考答案	(27)
第三章 分类资料的统计描述	(30)
一、目的要求	(30)
二、实例分析与电脑操作	(30)
三、思考练习参考答案	(35)
四、补充思考练习	(39)
五、补充思考练习参考答案	(41)
第四章 统计表与统计图	(44)
一、目的要求	(44)
二、实例分析与电脑操作	(44)
三、思考练习参考答案	(56)
四、补充思考练习	(60)
五、补充思考练习参考答案	(61)
第五章 总体均数估计与假设检验	(63)
一、目的要求	(63)
二、实例分析与电脑操作	(63)
三、思考练习参考答案	(76)
四、补充思考练习	(81)
五、补充思考练习参考答案	(83)
第六章 方差分析	(86)
一、目的要求	(86)
二、实例分析与电脑操作	(86)
三、思考练习参考答案	(98)
四、补充思考练习	(102)

五、补充思考练习参考答案·····	(104)
第七章 χ^2 检验 ·····	(107)
一、目的要求·····	(107)
二、实例分析与电脑操作·····	(107)
三、思考练习参考答案·····	(117)
四、补充思考练习·····	(120)
五、补充思考练习参考答案·····	(123)
第八章 秩和检验 ·····	(126)
一、目的要求·····	(126)
二、实例分析与电脑操作·····	(126)
三、思考练习参考答案·····	(139)
四、补充思考练习·····	(141)
五、补充思考练习参考答案·····	(144)
第九章 直线相关与回归 ·····	(146)
一、目的要求·····	(146)
二、实例分析与电脑操作·····	(146)
三、思考练习参考答案·····	(155)
四、补充思考练习·····	(159)
五、补充思考练习参考答案·····	(164)
第十章 调查设计 ·····	(166)
一、目的要求·····	(166)
二、实例分析与电脑操作·····	(166)
三、思考练习参考答案·····	(168)
四、补充思考练习·····	(171)
五、补充思考练习参考答案·····	(174)
第十一章 实验设计 ·····	(176)
一、目的要求·····	(176)
二、实例分析与电脑操作·····	(176)
三、思考练习参考答案·····	(181)
四、补充思考练习·····	(185)
五、补充思考练习参考答案·····	(188)
第十二章 剂量反应 ·····	(190)
一、目的要求·····	(190)
二、实例分析与电脑操作·····	(190)
三、思考练习参考答案·····	(198)
四、补充思考练习·····	(201)
五、补充思考练习参考答案·····	(204)
附录一 《医学统计学》期末模拟考试题(一) ·····	(205)
附录二 《医学统计学》期末模拟考试题(二) ·····	(210)
附录三 《医学统计学》期末模拟考试题(三) ·····	(214)
附录四 《医学统计学》期末模拟考试题(四) ·····	(221)

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 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紊的规定: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要说明算法。本书的实例分析是以 SPSS for Windows 11.5 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。对于某些章的统计程序计算,如果 SPSS 11.5 版本在某些计算机上不能运行,改用 SPSS 13.0 版本就可以运行。

(二) SPSS 的启动

双击桌面 SPSS 快捷图标或单击开始→程序→SPSS for Windows→SPSS 11.5 for Windows,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 10 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(变量视图)或 Data View(数据视图),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。

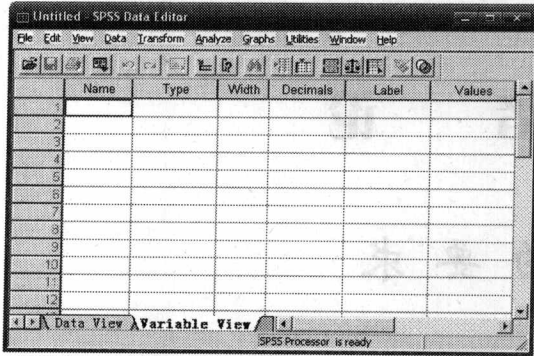


图 1-1 SPSS 的 Variable View 窗口

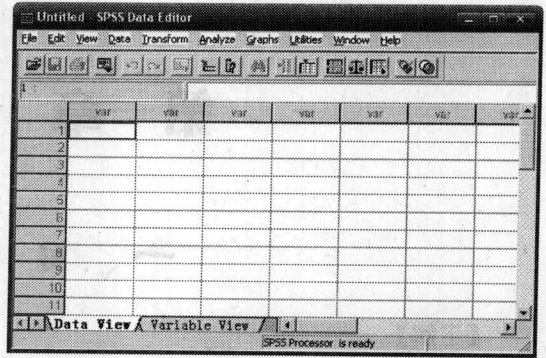




图 1-2 SPSS 的 Data View 窗口

(三) File 菜单(文件操作)

1. **New** 新建文件, 快捷键为 $\text{Ctrl}+\text{N}$ 。包括新建 Data(数据)、Syntax(程序)、Output(输出结果)、Draft Output(草案输出结果)、Script(手稿)。


2. **Open** 打开文件, 快捷按钮为 , 快捷键为 $\text{Ctrl}+\text{O}$ 。包括打开 Data、Syntax、Output、Script、Other(其他类型文件)。


3. **Save** 保存文件, 快捷按钮为 , 快捷键为 $\text{Ctrl}+\text{S}$ 。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“.spo”。

4. **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。


5. **Exit** 退出 SPSS 系统, 快捷键为 $\text{Alt}+\text{F4}$ 。

(四) Edit 菜单(编辑)

1. **Undo/Undo Set Cell Value** 撤消单元输入值, 快捷按钮为 , 快捷键为 $\text{Ctrl}+\text{Z}$ 。

2. **Redo** 恢复单元输入值, 快捷按钮为 , 快捷键为 $\text{Ctrl}+\text{R}$ 。


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。

4. **Find** 查找数据, 快捷按钮为 , 快捷键为 $\text{Ctrl}+\text{F}$ 。

5. **Options** 选项, 选择 SPSS 参数。

(五) View 菜单(视图)

1. **Status Bar** 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. **Fonts** 自定义字体。

4. **Grid Lines** 显示或隐藏表格线。

5. **Value Labels** 变量值标签/变量值显示切换, 系统默认不显示变量值标签, 快捷按钮

为⑨。变量值标签的定义方法参见例 4-1。

6. Variables 变量视图/数据视图切换,快捷键为 Ctrl+T。

(六) Data 菜单(数据操作)

1. Define Dates 定义日期,主要用于时间序列模型。

2. Insert Variable 插入变量,在当前列插入新变量,快捷按钮为①。

3. Insert Cases 插入个案(或记录),在当前行插入新个案(或记录),快捷按钮为②。

4. Go to Case 个案(或记录)定位,到达指定记录号的个案(或记录),快捷按钮为③。
当数据较多时,非常有用。

5. Sort Cases 过程 个案(或记录)排序,按个案(或记录)排序,排序方式有升序(Ascending)和降序(Descending)两种。

6. Transpose 过程 数据行列转置,可以将原来的一条记录转成为一条变量,或将原来的一个变量转成为一条记录。原变量名会自动保存在系统生成的 case-|b| 的字符变量中。

7. Restructure 过程 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 8-8)。

8. Merge Files 过程 合并数据文件。

(1) Add Cases 过程:增加个案(或记录),从外部数据文件中增加个案(或记录)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 过程: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。

9. Aggregate 过程 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。

10. Orthogonal Design 过程 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。

11. Split File 过程 拆分数据文件,用于数据文件的分组处理,快捷按钮为④。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。

12. Select Cases 过程 选择个案(或记录),用于选择需分析的个案(或记录),快捷按钮为⑤。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。

13. Weight Cases 过程 个案(或记录)加权或频数加权,快捷按钮为⑥。在使用频数表格式录入数据时(如 χ^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 7-1)。

(七) Transform 菜单(数据转换)

1. Compute 过程 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. Random Number Seed 过程 随机数字种子,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(参见例 10-1)。在临床试验等情况下,可以人为指定一个种子,结果就可重现(参见例 11-1)。

3. Count 过程 计数,用于计数每个个案(或记录)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。

4. Recode 过程 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。

(1) Into Same Variables 过程:在相同变量中重新编码。

(2) Into Different Variables 过程:在不同变量中重新编码(参见例 2-1)。

5. Categorize Variables 过程 分类变量,用于将连续型变量按要求转化为分类变量,根据需要在 Number of categories(分类数)框中输入数字,比如输入 2 时,原连续型变量就可转化为二分类的新变量。

6. Rank Cases 过程 个案(或记录)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(参见例 8-6)。

7. Automatic Recode 过程 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。

8. Create Time Series 过程 建立时间系列,用于创建时间序列变量。

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

(八) Analyze 菜单(统计分析)

1. Reports 菜单(统计报表)

(1) 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)、Grouped Median(分组中位数)、Std. Error of Mean(标准误)等。

(2) Case Summaries 过程:个案(或记录)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)、Grouped Median(分组中位数)、Standard Deviation(标准差)、Std. 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。

(3) Report Summaries in Rows 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。

(4) 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:探索性分析,进行数据分布状况的探索性分析,例如正态性检验。

(4) Crosstabs 过程:列联表分析,进行列联表资料的分析,用于分类资料/等级资料的统计描述及各种假设检验,例如 χ^2 检验、McNemar 检验等(参见第七章)。

(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。

3. 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。

(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)。

4. General Linear Model 菜单(广义或一般线性模型)

(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。

5. 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 9-1、例 9-2)。

(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。

6. 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 9-3、例 9-4)。

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

(3) 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 9-5)。

(4) Multinomial Logistic 过程:多分类 Logistic 回归分析,进行无序多分类和有序多分

类资料的 Logistic 回归分析。

(5) Probit 过程: 概率单位法, 用于分析剂量反应关系(参见例 12-1、例 12-2)。

7. Loglinear 菜单(对数线性模型分析)

(1) General 过程: 广义对数线性模型分析, 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。

8. 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, 称为 K-均值聚类法(快速聚类法、逐步聚类法)。

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

(4) Discriminant 过程: 判别分析, 进行判别分析。

9. Data Reduction 菜单(数据简化或降维分析)

(1) Factor 过程: 因子分析, 进行因子分析/主成分分析。

(2) Correspondence Analysis 过程: 对应分析, 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。

10. Scale 菜单(尺度分析)

(1) Reliability Analysis 过程: 可靠性分析或信度分析, 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 Scaling 过程: 多维尺度分析, 进行多维尺度分析, 反映多个研究事物间的不相似程度。

(3) Multidimensional Scaling(PROXSCAL)过程: 多维邻近尺度分析, 进行多维邻近尺度分析, 反映多个研究事物间的相似或不相似程度。

11. Nonparametric Tests 菜单(非参数检验)

(1) Chi-Square 过程: 单样本 χ^2 检验, 用于检验二项/多项分类变量的分布, 即检验分类数据样本所在总体分布(各类别所占比例)是否与已知总体分布相同即拟合优度检验。

(2) Binomial 过程: 二项式检验, 用于检验二项分类变量的分布是否服从指定概率参数 P 的二项分布。

(3) Run 过程: 游程检验, 用于检验样本序列的随机性。

(4) 1-Sample K-S 过程: 单样本 K-S 检验(1-Sample Kolmogorov-Smirnov Test, 即 1-Sample K-S), 用于检验样本是否服从各种常用分布, 如正态分布、均匀分布、指数分布。

(5) 2 Independent Samples 过程: 两独立样本秩和检验, 进行完全随机设计两样本资料的秩和检验(参见例 8-3、例 8-4)。

(6) K Independent Samples 过程: 多个独立样本秩和检验, 进行完全随机设计多个样本资料的秩和检验(参见例 8-5)。

(7) 2 Related Samples 过程: 两个相关样本秩和检验(配对秩和检验), 进行配对设计资料的秩和检验、单样本资料的秩和检验(参见例 8-1、例 8-2)。

(8) K Related Samples 过程:多个相关样本秩和检验(随机区组秩和检验),进行随机区组设计资料的秩和检验(参见例 8-7)。

12. Survival 菜单(生存分析)

(1) Life Tables 过程:寿命表方法,用于分析分组生存资料,求出不同组段的生存率。当样本量较大时(如 $n > 50$),可以把资料按不同时间段分成几段,观察不同时间点的生存率。

(2) Kaplan-Meier 过程:Kaplan-Meier 法(Kaplan-Meier, 1958),用于样本含量较小时,不能给出特定时间点的生存率。

(3) Cox Regression 过程:Cox 回归分析,用于拟合 Cox 比例风险模型,这是生存分析中最重要的一个分析方法。

(4) Cox w/Time-Dep Cov 过程:含时间一依赖协变量的 Cox 回归分析(Cox 时效协变量回归分析),当所研究的危险因素取值随时间不断变化,或危险因素强度随时间不断变化时,不符合 Cox 模型的适用条件,此时需要对模型加以修正,就必须用这个过程进行分析。

(九) Graphs 菜单(统计图形制作与编辑)

1. Gallery 画廊窗口,将统计图制作过程做了简单介绍,相当于一个自学向导。

2. Interactive 菜单 交互绘图,用于绘制各种交互式统计图。

3. Map 菜单 地图,用于绘制统计地图。

4. 其他菜单项 可绘制多种统计图:Bar(条图)、Line(线图)、Area(面积图)、Pie(圆图)、High-Low(高-低图)、Pareto(Pareto 图)、Control(控制图)、Boxplot(箱式图)、Error Bar(误差图)、Scatter(散点图)、Histogram(直方图)、P-P(P-P 图)、Q-Q(Q-Q 图)、Sequence(序列图)、ROC Curve(ROC 曲线)、Autocorrelations(自相关图)、Cross-Correlations(互相关图)、Spectral(频谱图)(绘图方法参见第四章)。

(十) SPSS 的数据输入与保存(数据准备)

【例 1-1】 抽样调查某地 110 名健康成年男子的红细胞数($\times 10^{12}/L$),结果如表 1-1,试建立 SPSS 数据库并输入数据。(教材 P25 应用分析题第 1 题)

表 1-1 某地 110 名健康成年男子红细胞数($\times 10^{12}/L$)

5.25	4.60	5.20	5.00	5.62	4.80	5.53	4.71	5.40	5.25
5.10	4.40	4.80	5.56	5.20	4.10	5.10	4.62	5.60	5.86
4.30	4.52	5.52	4.32	4.10	4.40	4.42	4.80	5.50	5.16
4.49	4.99	5.21	5.10	5.00	4.85	5.00	4.29	4.36	4.60
4.22	5.30	4.26	4.78	4.50	4.92	4.52	5.45	5.10	5.06
4.90	4.75	4.63	4.72	4.10	5.12	4.72	5.28	4.95	4.88
4.81	4.50	5.20	4.64	4.25	5.14	5.42	5.11	4.55	4.60
5.40	5.01	5.70	4.89	4.07	5.20	4.65	4.96	5.13	5.04
5.10	5.30	5.05	5.32	4.82	5.19	4.86	5.11	5.16	4.85
5.60	3.99	5.20	5.64	4.72	5.18	5.13	5.01	5.15	5.90
4.84	4.78	5.30	5.50	4.89	5.60	5.45	5.36	5.15	4.65

【分析】 该资料为同一个观察指标(红细胞数)的测量数值,应当输入到同一个变量中,本例的变量为红细胞数;同一观察对象的数据应当独占一行,为一个记录,本例应该有110个记录。

【操作】

1. 建立数据库

(1) 定义变量:激活 SPSS 的数据编辑窗口,单击窗口左下角的 Variable View(变量视图),切换到 SPSS 的 Variable View 窗口。变量视图的每一行代表对一个变量的定义,每一列代表定义该变量时用到的某种属性,包括 Name(变量名)、Type(变量类型)、Width(宽度)、Decimals(小数点位数)、Label(变量名标签)、Values(变量值标签)、Missing(缺失值)等。

在第 1 行第 1 列中输入“红细胞数”,敲击 Enter 键(回车)或用鼠标点中,就可以依次对 Type、Width、Decimals、Label、Values 等进行定义,本例均为系统默认,即 Type 为 Numeric(数值型),宽度为 8,小数点位数为 2。这样就完成了定义变量的过程,如图 1-3 所示。

(2) 保存文件:选择菜单 File→Save 或 Save as,弹出 Save Data As(数据存为…)对话框。值得注意的是,系统默认的存盘目录是 SPSS 系统的根目录,为保证文件的安全性,应将文件储存到 C 盘以外的其他本地磁盘(如 D 盘、E 盘等)或可移动磁盘(如移动硬盘、U 盘等)。选择好存盘目录后,在“文件名”框中输入“例 1-1”,保存类型为默认的 SPSS(*.sav)文件,单击“保存”,该文件就以“例 1-1.sav”的文件名保存了(默认后缀名“.sav”自动加上),如图 1-4 所示。

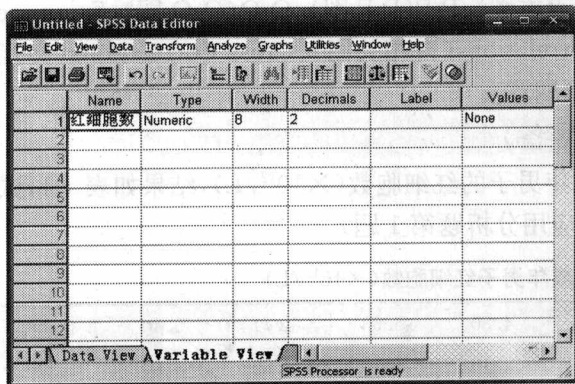


图 1-3 SPSS 的 Variable View 窗口

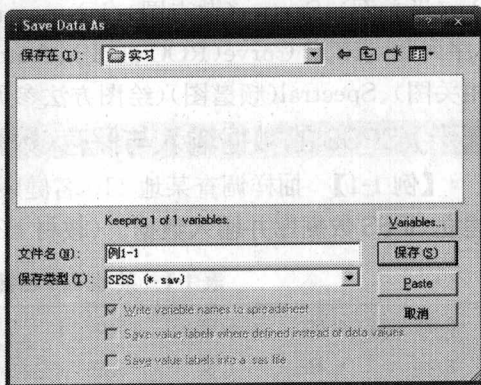


图 1-4 Save Data As 对话框

2. 输入数据 点击数据编辑窗口左下角的 Data View(数据视图),切换到 SPSS 的 Data View 窗口,可见第 1 列的名称为深色显示,即刚定义的变量“红细胞数”,其余各列名称仍为灰色的“var”,表示尚未使用。同样,各行的标号仍为灰色的“1、2、3…”,表明该数据集中没有记录。

从第 1 行第 1 列开始输入数据,输入第 1 个数据 5.25,如图 1-5 所示,敲击 Enter 键下移一行,继续输入下一条记录。依此方法,将 110 个数据全部输入,记录号一直到 110 都是深色显示,其后均为灰色显示,表示全部记录数为 110,如图 1-6 所示。

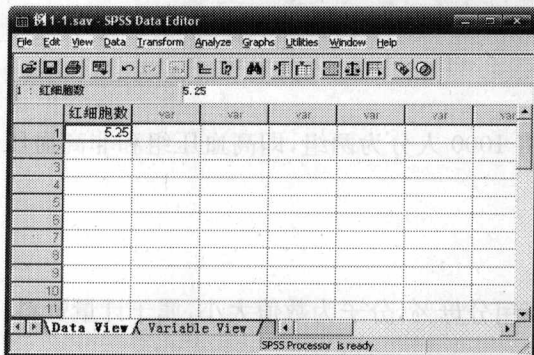


图 1-5 SPSS 的 Data View 窗口

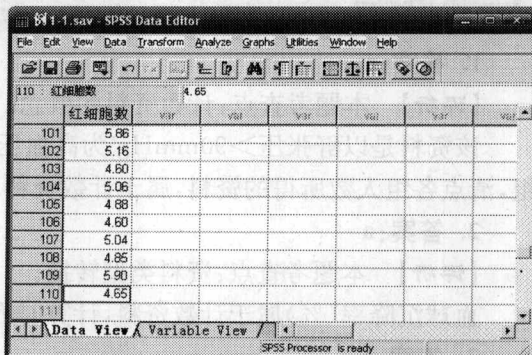


图 1-6 数据输入完毕

三、思考练习参考答案

(一) 名词解释

1. 【解答】 参见教材 P6【知识点 1-4】第 1 条。
2. 【解答】 参见教材 P5。
3. 【解答】 参见教材 P6【知识点 1-4】第 2 条。
4. 【解答】 参见教材 P6【知识点 1-4】第 4 条。

(二) 是非题

1. 答案：—

【评析】 本题考查点：抽样误差。

抽样误差的根源在于个体变异，在抽样研究中是不可避免的，但其规律可以认识。

2. 答案：+

【评析】 本题考查点：小概率事件的概念。

$P \leq 0.05$ 的随机事件称为小概率事件，小概率事件的原理是在一次实验中是不大可能发生的。

3. 答案：—

【评析】 本题考查点：统计描述与统计推断的区别。

统计描述是对已知的样本(或总体)的分布情况或特征进行分析表述。统计推断是根据已知的样本信息来推断未知的总体。

4. 答案：+

【评析】 本题考查点：普查与抽样研究的区别。

对全部研究对象都进行调查或测定，称为普查。此时没有抽样误差，但有非抽样误差。

5. 答案：+

【评析】 本题考查点：分类资料的概念。

分类资料是把观察单位按某种属性(性质)或类别进行分组，清点各组观察单位数所得资料。各观察单位是定性的，一般无度量衡单位。各属性之间互不相容。