



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

供临床、基础、口腔、麻醉、影像、药学、检验、护理、法医等专业使用

案例版™

医学统计学计算机操作教程

第3版

主 编 罗家洪 郭秀花

第3版



科学出版社

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™

供临床、基础、口腔、麻醉、影像、药学、检验、护理、法医等专业使用

医学统计学计算机操作教程

第3版

主 编 罗家洪 郭秀花
副主编 姚应水 贾 红 刘启贵 赵若望 董莉萍 程晓萍
毛 勇 李秀央 谢红卫 孟 琼 刘 芬
学术秘书 何利平 彭林珍
编 者 (按姓氏汉语拼音排序)

常 巍 (昆明医科大学)
陈 莹 (昆明医科大学)
董莉萍 (北华大学)
郭秀花 (首都医科大学)
何利平 (昆明医科大学)
贺连平 (皖南医学院)
胡志宏 (北华大学)
李晓梅 (昆明医科大学)
刘 芬 (首都医科大学)
刘启贵 (大连医科大学)
罗家洪 (昆明医科大学)
毛 勇 (昆明医科大学)
彭林珍 (云南交通职业技术学院)
童玲玲 (南华大学)
王耶盈 (昆明医科大学)
肖媛媛 (昆明医科大学)
闫宇翔 (首都医科大学)
叶运莉 (西南医科大学)
喻 箴 (昆明医科大学)
张 凤 (首都医科大学)
赵若望 (内蒙古科技大学包头医学院)

陈 燕 (皖南医学院)
程晓萍 (锦州医科大学)
杜瑞红 (北华大学)
郝金奇 (内蒙古科技大学包头医学院)
和丽梅 (昆明医科大学)
侯瑞丽 (内蒙古科技大学包头医学院)
贾 红 (西南医科大学)
李秀央 (浙江大学)
刘军祥 (西南医科大学)
刘 艳 (南华大学)
罗 健 (昆明医科大学)
孟 琼 (昆明医科大学)
宋桂荣 (大连医科大学)
王良君 (锦州医科大学)
吴梦吟 (浙江大学)
谢红卫 (南华大学)
姚应水 (皖南医学院)
俞婉琦 (浙江大学)
詹志鹏 (锦州医科大学)
张俊辉 (西南医科大学)

科学出版社

郑重声明

为顺应教育部教学改革潮流和改进现有的教学模式,适应目前高等医学院校的教育现状,提高医学教育质量,培养具有创新精神和创新能力的医学人才,科学出版社在充分调研的基础上,引进国外先进的教学模式,独创案例与教学内容相结合的编写形式,组织编写了国内首套引领医学教育发展趋势的案例版教材。案例教学在医学教育中,是培养高素质、创新型和实用型医学人才的有效途径。

案例版教材版权所有,其内容和引用案例的编写模式受法律保护,一切抄袭、模仿和盗版等侵权行为及不正当竞争行为,将被追究法律责任。

图书在版编目(CIP)数据

医学统计学计算机操作教程 / 罗家洪,郭秀花主编. —3版. —北京:科学出版社,2018.3

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 978-7-03-056280-7

I. ①医… II. ①罗… ②郭… III. ①计算机应用-医学统计-高等学校-教材 IV. ①R195.1-39

中国版本图书馆CIP数据核字(2018)第006623号

责任编辑:朱 华 / 责任校对:郭瑞芝

责任印制:赵 博 / 封面设计:王 融

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>

新科印刷有限公司印刷

科学出版社发行 各地新华书店经销

*

2007年8月第 一 版 开本:850×1168 1/16

2018年3月第 三 版 印张:18 1/2

2018年3月第十六次印刷 字数:612 000

定价:59.80元

(如有印装质量问题,我社负责调换)

前 言

本教材是根据《国家中长期教育改革和发展规划纲要(2010—2020年)》《国家中长期人才发展规划纲要(2010—2020年)》《国家中长期科学和技术发展规划纲要(2006—2020年)》等的精神,在本科案例版《医学统计学》(第2版)和《医学统计学计算机操作教程》(第2版)等的基础上编写的第3版《医学统计学计算机操作教程》,本教程本着与时俱进、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,编写的适合中国国情的全新案例式教程。本教程可与本科案例版《医学统计学》第3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SPSS20.0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,具有先进性、科学性、启发性、实用性等主要特点。

《医学统计学计算机操作教程》主要突出了以下几个方面特点:

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,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、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,昆明医学院院长姜润生教授,副院长李松教授,副院长李燕主任医师,教务处章宗籍处长,公共卫生学院陆林院长、张有福书记、殷建忠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。

本教程是全新案例版《医学统计学计算机操作教程》(第3版),限于我们的水平和缺乏编写经验,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

2017年8月于春城昆明

目 录

第 1 章 绪论	1
一、目的要求	1
二、实例分析与计算机操作	1
三、思考练习参考答案	10
四、补充思考练习	12
五、补充思考练习参考答案	13
第 2 章 计量资料的统计描述	15
一、目的要求	15
二、实例分析与计算机操作	15
三、思考练习参考答案	21
四、补充思考练习	24
五、补充思考练习参考答案	26
第 3 章 分类资料的统计描述	28
一、目的要求	28
二、实例分析与计算机操作	28
三、思考练习参考答案	32
四、补充思考练习	37
五、补充思考练习参考答案	39
第 4 章 统计表与统计图	41
一、目的要求	41
二、实例分析与计算机操作	41
三、思考练习参考答案	63
四、补充思考练习	66
五、补充思考练习参考答案	68
第 5 章 总体均数估计与假设检验	70
一、目的要求	70
二、实例分析与计算机操作	70
三、思考练习参考答案	81
四、补充思考练习	88
五、补充思考练习参考答案	90
第 6 章 方差分析	92
一、目的要求	92
二、实例分析与计算机操作	92
三、思考练习参考答案	102
四、补充思考练习	106
五、补充思考练习参考答案	109
第 7 章 二项分布与 Poisson 分布	110
一、目的要求	110
二、实例分析与计算机操作	110
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	122
第 8 章 χ^2 检验	124

一、目的要求	124
二、实例分析与计算机操作	124
三、思考练习参考答案	132
四、补充思考练习	137
五、补充思考练习参考答案	139
第 9 章 秩和检验	142
一、目的要求	142
二、实例分析与计算机操作	142
三、思考练习参考答案	154
四、补充思考练习	157
五、补充思考练习参考答案	159
第 10 章 直线相关与回归	161
一、目的要求	161
二、实例分析与计算机操作	161
三、思考练习参考答案	167
四、补充思考练习	172
五、补充思考练习参考答案	174
第 11 章 调查设计	177
一、目的要求	177
二、实例分析与计算机操作	177
三、思考练习参考答案	179
四、补充思考练习	181
五、补充思考练习参考答案	183
第 12 章 实验设计	185
一、目的要求	185
二、实例分析与计算机操作	185
三、思考练习参考答案	190
四、补充思考练习	194
五、补充思考练习参考答案	195
第 13 章 剂量反应	197
一、目的要求	197
二、实例分析与计算机操作	197
三、思考练习与参考答案	206
四、补充思考练习	210
五、补充思考练习参考答案	212
第 14 章 多因素分析	213
一、目的要求	213
二、实例分析与计算机操作	213
三、思考练习参考答案	226
四、补充思考练习	230
五、补充思考练习参考答案	231
主要参考文献	233
附录一 《医学统计学》期末模拟考试题(一)	234
附录二 《医学统计学》期末模拟考试题(二)	237
附录三 《医学统计学》期末模拟考试题(三)	240
附录四 《医学统计学》期末模拟考试题(四)	266

第 1 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、实例分析与计算机操作

(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要说明算法。本书的实例分析是以 IBM SPSS Statistics 20.0 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 → IBM 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 12 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。

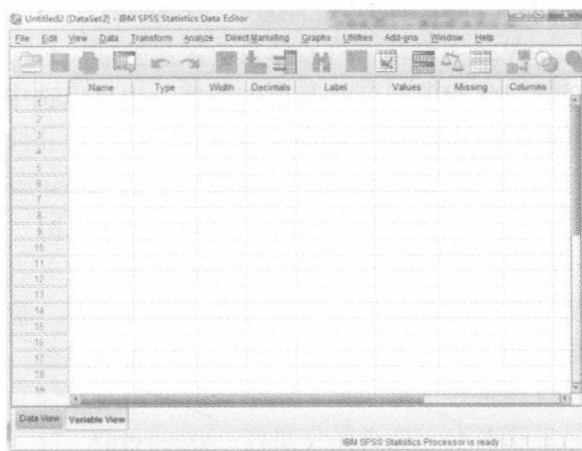


图 1-1 SPSS 的 Variable View 窗口

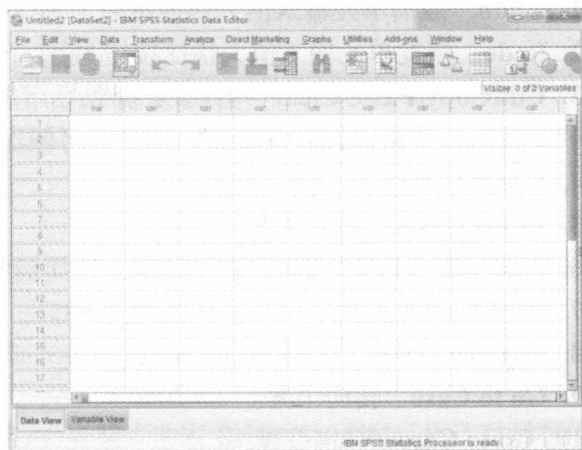



图 1-2 SPSS 的 Data View 窗口




(三) File 菜单 (文件操作)

1. **New** 新建文件, 快捷键为 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)。
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 后, 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 后, 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, 非常有用。在 V17.0 后, 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 χ^2 。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 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)、**Grouped Median**(分组中位数)、**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)、**Grouped Median**(分组中位数)、**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 χ^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) **Ordinal** 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, 对数据进行多维邻近尺度分析, 反映多个研究事物间的相似或不相似程度。
- (4) **Multidimensional Scaling** 过程: 多维尺度分析, 对数据进行多维尺度分析, 反映多个研究事物间的不相似程度。

16. 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 9-3、例 9-4)。
- (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 9-5)。
- (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 9-1、例 9-2)。

(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 9-7)。

17. Forecasting 菜单(预测分析、时间系列分析) 可用于尝试预测序列的未来值。解释过去值的序列的模型还可以预测下几个值是增加还是减少, 以及增减的幅度有多大。

(1) **Create Models** 过程: 建立时间序列模型, 用数据建立时间序列模型。

(2) **Apply Models** 过程: 应用时间序列模型, 从外部文件加载现有的时间序列模型, 并将它们应用于活动数据集。使用此过程, 可以在不重新建立模型的情况下获得其新数据或修订数据可用的序列的预测值。

(3) **Seasonal Decomposition** 过程: 季节分解法, 可将一个序列分解成一个季节性成分、一个组合趋势和循环的成分和一个“误差”成分。

(4) **Spectral Analysis** 过程: 频谱分析, 可用于标识时间序列中的周期行为。它不需要分析一个时间点与下一个时间点之间的变异, 只要按不同频率的周期性成分分析整体序列的变异。平滑序列在低频率具有更强的周期性成分; 而随机变异(“白噪声”)将成分强度分布到所有频率。

(5) **Sequence Chart** 过程: 系列图, 通过系列图说明该时间系列是否具有稳定性。

(6) **Autocorrelations** 过程: 自相关图, 通过自相关图说明该时间系列是否具有平稳性。

(7) **Cross-Correlations** 过程: 交叉相关图, 通过 2 个或以上时间序列的正、负或零延迟的交叉相关系数形成的图形, 用于时间序列中识别具有其他变量的前导指数的变量。

18. 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—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。

19. Multiple Response 菜单(多重响应)

(1) **Define Variable Sets** 过程: 定义多重响应数据集, 将多个基础变量定义为多重响应数据集(多重二分类变量集或多重分类集), 以便进行多重响应分析。

(2) **Frequencies** 过程: 多重响应频率分析, 可产生多重响应数据集的频数表, 包括例数、响应百分比、个案百分比等。

(3) **Crosstab** 过程: 多重响应行列表分析, 可产生多重响应数据集、基础变量或组合变量的行列表, 包括例数、各种百分数等。

20. Missing Value Analysis 菜单 缺失值分析。

21. Multiple Imputation 菜单 **Analyze Patterns** (分析模式) 和 **Impute Missing Data Value** (归因缺失数据值)。

22. Complex Sample 菜单(复杂抽样) 包括 **Select a Sample** 过程(选择抽样), **Prepare for Analysis** 过程(准备分析), **Complex Sample Plan** 过程(以下统计分析的复杂计划): **Frequency** (频率分析)、**Descriptive** (描述性分析)、**Crosstabs** (交叉表分析)、**Ratios** (比率分析)、**General Linear Model** (一般线性模型)、**Logistic Regression** (Logistic 回归)、**Ordinal Regression** (有序回归) 和 **COX Regression**。

23. Quality Control 菜单 质量控制, 绘制质量控制图。

24. ROC Curve 菜单 通过 **ROC Curve** 过程绘制计量资料或分类资料的 **ROC** 曲线, 并计算曲线下的面积, 对诊断价值做出判断。

(九) Graphs 菜单(统计图形制作与编辑)

1. Chart Builder 菜单(图表构建程序) 可绘制 **Bar** (条图)、**Line** (线图)、**Area** (面积图)、**Pie** (圆图)、**Scatter/Dot** (散点图/点图)、**Histogram** (直方图)、**High-Low** (高-低图)、**Boxplot** (箱式图) 和 **Dual**

Axes (双轴图)。

2. Graphboard Template Chooser 菜单 (图形画板模板选择程序) 可绘制 **Bar** (条图)、**Line** (线图)、**Area** (面积图)、**Pie** (圆图)、**Scatter** (散点图)、**Histogram** (直方图) 等。

3. Legacy Dialogs (旧对话框) 传统的图形构建界面, 可绘制多种统计图: **Bar** (条图)、**3-D Bar** (3-D 条形图)、**Line** (线图)、**Area** (面积图)、**Pie** (饼图或圆图)、**High-Low** (高-低图)、**Boxplot** (箱式图)、**Error Bar** (误差图)、**Population Pyramids** (人口金字塔)、**Scatter/Dot** (散点图/点图)、**Histogram** (直方图)。**Interactive** (交互式绘图): 可绘制 **Bar** (条图)、**Dot** (点图)、**Line** (线图)、**Area** (面积图)、**Pie** (饼图或圆图)、**Boxplot** (箱式图)、**Error Bar** (误差图)、**Scatterplot** (散点图) 等。(绘图方法参见第四章)

(十) Utilities 菜单 (实用程序)

Utilities 菜单包括 **Variables** (变量)、**OMS Control Panel** (OMS 控制面板)、**OMS Identifiers** (OMS 标识符)、**Data File Comments** (数据文件注释)、**Define Variable Set** (定义变量集)、**Use Variable Set** (使用变量集)、**Show All Variables** (显示所有变量)、**Spelling** (拼写)、**Run Script** (运行脚本)、**Production Job** (生产工作)、**Custom Dialog Builder** (自定义对话框构建程序)、**Install Custom Dialog** (安装自定义对话框)。

(十一) Add-ons 菜单 (附加软件)

Add-ons 菜单包括 **Applications** (应用) (包括 AMOS 多水平分析等)、**Service** (服务)、**Programmability** (可编程序)、**Statistics Guides** (统计指导)。

(十二) Windows 菜单 (窗口)

Window 菜单包括 **Split** (拆分)、**Minimize All Windows** (将所有窗口最小化)。

(十三) Help 菜单 (窗口)

Help 菜单包括 **Topics** (主题)、**Tutorial** (教程)、**Case Studies** (个案研究)、**Statistics Coach** (统计辅导)、**Command Syntax Reference** (指令语法参考)、**Developer Central** (开发者中心)、**About** (关于)、**Algorithms** (算法)、**SPSS Inc. Home** (SPSS 公司主页)、**Check For Updates** (检查更新)、**Product Registration** (产品注册)。

(十四) SPSS 的数据输入与保存 (数据准备)

【例 1-1】 某市 2016 年 120 名 18 岁女孩身高 (cm) 资料表 1-1, 试建立 SPSS 数据库并输入数据。

表 1-1 某地 2014 年 120 名 18 岁女孩身高 (cm) 资料

169.1	171.5	169.4	167.5	165.9	166.6	169.8	167.9	166.3	167.7
165.6	169.8	167.7	166.8	167.9	158.6	169.8	168.9	164.5	165.3
165.1	164.5	159.9	171.6	169.7	168.3	163.6	158.5	165.6	165.9
161.5	166.3	168.5	167.1	161.0	159.0	167.3	157.2	164.7	165.1
166.1	167.5	166.0	158.5	161.2	167.5	158.2	154.7	159.6	169.7
168.8	165.6	167.2	162.8	163.7	173.5	159.8	168.3	159.4	168.8
169.2	168.9	159.8	168.6	166.5	163.5	169.8	166.8	163.9	169.2
171.2	163.4	164.5	166.5	165.9	168.4	167.2	167.7	169.5	159.9
165.2	169.6	162.3	164.6	165.4	173.6	162.8	163.4	164.6	167.2
165.7	161.3	165.5	166.4	168.9	165.0	164.2	165.3	162.7	168.5
168.5	162.6	158.8	165.8	175.5	166.9	165.8	164.5	167.9	165.8
168.1	169.8	166.2	167.5	165.9	166.6	169.4	167.9	164.5	169.9

【分析】 该资料为同一个观察指标（身高）的测量数值，应当输入到同一个变量中，本例的变量为身高；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120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“身高”，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。

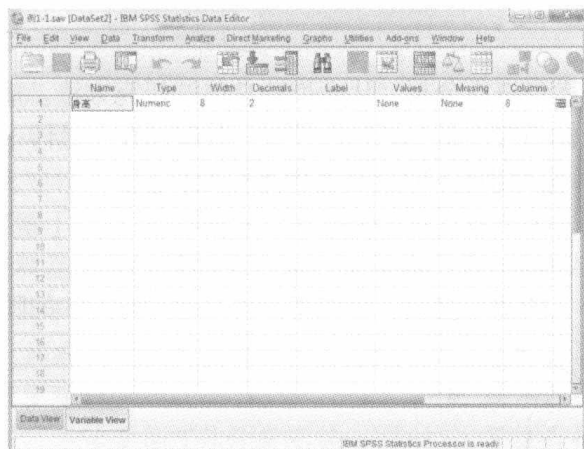


图 1-3 SPSS 的 Variable View 窗口

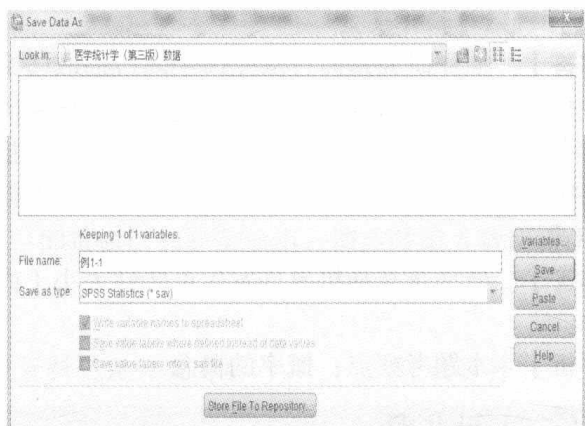


图 1-4 Save Data As 对话框

(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。

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“身高”，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169.1，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120个数据全部输入，记录号一直到120都是深色显示，其后均为灰色显示，表示全部记录数为120，如图1-6所示。

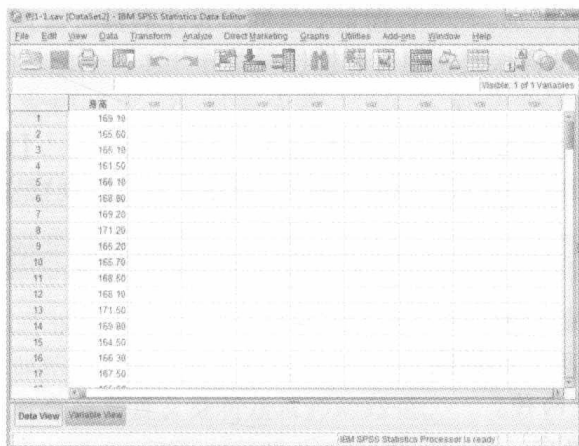


图 1-5 SPSS 的 Data View 窗口

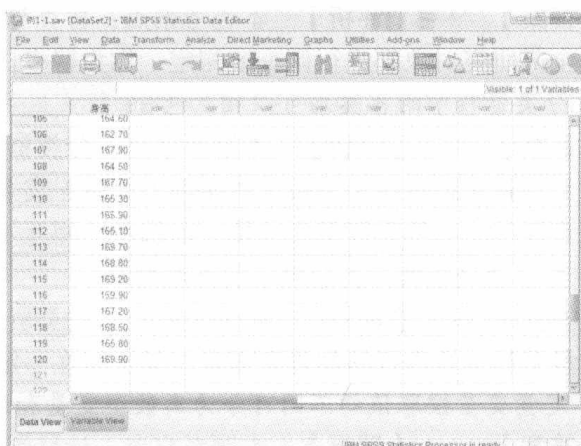


图 1-6 数据输入完毕

三、思考练习参考答案

(一) 名词解释

1. 总体与样本

【解答】 总体是根据研究目的确定的同质研究对象的全体即研究对象的全体。样本是总体中具有代表性的一部分个体。

【评析】 本题考察点：总体与样本的概念。参见教材【知识点 1-4】第 1 条。

2. 参数与统计量

【解答】 总体的指标称为参数，一般用希腊字母表示，如总体均数用 μ 表示，总体标准差用 σ 表示。样本指标称为统计量，如样本均数用 \bar{x} 表示，样本标准差用 s 表示。

【评析】 本题考察点：参数与统计量的概念。

3. 抽样研究与抽样误差

【解答】 抽样研究是通过从总体中随机抽取样本，对样本信息进行分析，从而推断总体的研究方法。抽样误差是由随机抽样造成的样本指标与总体指标之间、样本指标与样本指标之间的差异。其根源在于总体中的个体存在变异性。只要是抽样研究，就一定存在抽样误差，不能用样本的指标直接下结论。

【评析】 本题考察点：抽样研究与抽样误差的概念。

4. 概率

【解答】 概率是某随机事件发生可能性大小（或机会大小）的数值度量。小概率事件是指 $P \leq 0.05$ 的随机事件。

【评析】 本题考察点：概率的概念。

(二) 是非题

1. 只要增加样本例数就可以避免抽样误差。

【答案】 - 【评析】 本题考察点：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. 某医院发生的医疗事故属于小概率事件。

【答案】 + 【评析】 本题考察点：小概率事件。

发生概率 $P \leq 0.05$ 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。一般医院发生的医疗事故均小于 0.05。

3. 统计描述就是用样本推断总体的统计过程。

【答案】 - 【评析】 本题考察点：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. 如果对全部研究对象都进行了调查或测定就没有抽样误差。

【答案】 + 【评析】 本题考察点：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. 分类资料中的各类别可以相互包含。

【答案】 - 【评析】 本题考察点：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。各观察变量值是定性的，一般无度量衡单位。各属性之间互不相容。

6. 医学领域中的三类资料不能互相转换。

【答案】 - 【评析】 本题考察点：三类资料的概念及其应用。

计量资料、计数资料和等级资料三类资料在一定条件下可以互相转换。

7. 定量变量按取值的不同可分为离散型变量和连续型变量两种。

【答案】 + 【评析】 本题考察点：计量资料资料的概念。

定量变量按取值的不同可分为离散型变量（discrete variable）和连续型变量（continuous variable）两种，前

者取值范围是有限个值或者一个数列构成的，常取 0 和正整数值，如现有子女数，儿童的龋齿数，胎次等。连续型变量则可以取实数轴上的任何数值，如身高、体重、血红蛋白等。

8. 科研结果的好坏取决于研究设计的好坏，研究设计是统计工作的基础和关键，决定着整个统计工作的成败。

【答案】 + 【评析】 本题考察点：科研设计的意义。

科研结果的好坏取决于研究设计的好坏，一定的设计决定了一定的数据分析方法，不同设计下获得的资料常用不同的方法来分析。

9. 没有较好的统计学知识，就不可能进行较好的科学研究，更不可能写出一篇高质量的科研论文。 ()

【答案】 + 【评析】 本题考察点：统计学的意义。

10. 用 SAS 和 SPSS 统计分析的结果，在国际学术交流中可以不说明算法。 ()

【答案】 + 【评析】 本题考察点：国际统计软件包的应用。

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。

(三) 选择题

1. 若成年男子以血红蛋白 $<125\text{g/L}$ 为贫血，调查某地 1000 人中有多少个贫血患者，这是_____。

- a. 计量资料 b. 还不能决定是计量资料还是计数资料 c. 计数资料
d. 既可作计量也可作计数资料 e. 等级资料

【答案】 c 【评析】 本题考察点：计数资料的概念。

该资料是以血红蛋白 $<125\text{g/L}$ 为贫血，将 1000 人分为两组，即贫血组和非贫血组，清点各组人数所得的资料，属于计数资料。

2. 一批病人的淋巴细胞转换率 (%) 是_____。

- a. 计量资料 b. 还不能决定是计量资料还是计数资料 c. 计数资料
d. 既可作计量也可作计数资料 e. 等级资料

【答案】 a 【评析】 本题考察点：资料类型转换。

淋巴细胞转换率 (%) 属于计数资料，计算时因不用分母%，分子为数值大小，属于计量资料。

3. 统计一批糖尿病患者的住院天数是_____。

- a. 计量资料 b. 还不能决定是计量资料还是计数资料 c. 计数资料
d. 既可作计量也可作计数资料 e. 等级资料

【答案】 a 【评析】 本题考察点：计量资料的概念。

住院天数 (天) 有度量衡单位和数值大小，属于计量资料。

4. 测量某病患者的抗体滴度 ($1:2, 1:4, 1:8, \dots$)，是_____。

【答案】 a 【评析】 本题考察点：计量资料的概念。

抗体滴度为比例或分数，计算时用其倒数，有具体的数值大小，属于计量资料。

5. 调查某医院医生的工作状况，医生一天内上班的时间是_____。

- a. 变量 b. 总体 c. 个体 d. 变量值 e. 统计指标

【答案】 b 【评析】 本题考察点：计量资料的概念。

时间是变量，具体上班时间是变量值。

6. 治疗结果分为有效和无效的资料，严格说来属于_____。

- a. 等级资料 b. 计数资料 c. 计量资料 d. 等级或计量均可 e. 计数或计量均可

【答案】 e 【评析】 本题考察点：等级资料的概念。

治疗结果分为有效和无效的资料，一般属于计数资料，严格说来有效和无效有程度差异，属于等级资料。

(四) 简答题

1. 某医师根据自己 20 年来收集的胆结石病例进行分析，认为：胆结石的发病和居住地有关，某些地区特别容易发生胆结石。女性发生胆结石的机会比男性大。从治疗效果看，保守治疗的效果不如手术治疗的效果好。请从统计学的角度，分析该医师的结论。

【解答】 (1) 该医师将患者按不同居住地、不同性别、不同疗法分组，分别清点各组的患者例数，属于二分类的计数资料；其设计方案为完全随机设计。由于该医师所收集的胆结石病例资料为样本资料，不可避免