



普通高等教育“十二五”规划教材

SPSS(PASW)17.0 在医学统计中的 应用 (第五版)

马斌荣 编著



科学出版社

普通高等教育“十二五”规划教材

SPSS(PASW)17.0 在医学统计中的应用

(第五版)

马斌荣 编著

科学出版社

北 京

内 容 简 介

本书是《SPSS(PASW)17.0 在医学统计中的应用》(第四版)的升级版。全书从医学科研中的实际问题出发,深入浅出地介绍了如何正确使用 SPSS 统计软件包,期望给读者提供一个简单、明了、正确的进行医学数据处理的方法。

和前几版相比,除了 SPSS 17.0 带有的新功能外,本书还包含了医学科研统计中常用的统计方法,并增加了生存分析、ROC 分析、Cox 回归分析模型及量表的信度效度分析。本书同时附赠例题的原始 SPSS 数据电子文件,以供主讲教师教学使用。

医学科研统计的主要问题有:如何建立数据文件;如何选择统计方法;如何正确使用统计软件包;如何恰当地解释统计分析结果。实际上,这四者是密切联系、不可分割的,本书把它们有机地整合在一起。

本书适用于医学院校师生、医疗卫生系统的科研工作者,以及相关领域的研究生、参加继续教育的医务工作者。

图书在版编目(CIP)数据

SPSS(PASW)17.0 在医学统计中的应用/马斌荣编著.—5 版.—北京:科学出版社,2014

普通高等教育“十二五”规划教材

ISBN 978-7-03-041839-5

I. ①S… II. ①马… III. ①医学统计-统计分析-软件包-高等学校-教材

IV. ①R195.1-39

中国版本图书馆 CIP 数据核字(2014)第 207883 号

责任编辑:刘 畅 / 责任校对:张小霞
责任印制:霍 兵 / 封面设计:迷底书装

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>

大厂书文印刷有限公司印刷

科学出版社发行 各地新华书店经销

*

1998 年 9 月第 一 版 开本:720×1000 1/16

2015 年 1 月第 五 版 印张:19

2015 年 1 月第十七次印刷 字数:427 000

定价:39.80 元

(如有印装质量问题,我社负责调换)

第五版前言

SPSS 是世界上通用的统计软件包之一, 2009 下半年改名为 PASW, 故本书第四版定名为《SPSS (PASW) 17.0 在医学统计中的应用》。本书是第四版的升级版。

《SPSS (PASW) 17.0 在医学统计中的应用》(第四版) 自 2010 年 6 月出版发行以来, 随着医学科研的发展, 与时俱进, 不断增加了许多新的内容, 本次已是第五版。本书已被全国很多医学院校选为医学硕士生和博士生教材。有些医学科研机构和临床医院也使用本书作为提高医学科研能力的培训教材。据学员们反映, 本书可读性好、实用性强、深入浅出、简单明了。

随着 SPSS 软件的升级和医学科研的发展, 本书新版作了如下的几点修改:

1. 在第一章和第二章, 在介绍建立数据文件的同时, 重点讲解了“数据清理”, 以保证后续统计分析的科学性。

2. 在第十一章增加了 ROC 分析。它是对不同疾病不断发现新的诊断试验方法、早期诊断和预警模型的效度分析的一个有效方法。

3. 增加了第十二章生存分析。它是对生存时间进行分析的统计学方法的总称。它是对治疗肿瘤及其他慢性疾病的、有助于延长带瘤生存期的新药及新的治疗方法进行评价的一个有效方法。第五版还增加了 Cox 回归分析模型。

4. 第五版增加了第十四章量表的信度效度分析, 为调查研究的质量评价提供方法。

本书也保留了前几版的一贯特点:

1. 本书不是按 SPSS 使用手册的纲目进行罗列讲解; 而是以医学科研统计分析要解决的实际问题为纲目, 引导读者如何调用 SPSS 中的功能来进行数据处理。举例如下。

(1) 医学科研中, 在建立数据文件后, 进行统计分析之前, 必须对数据进行清理。但是, 做哪些清理? 如何清理? SPSS 没有交代, 本书结合医学实际, 引导读者如何使用 SPSS 中的一些功能进行数据清理。

(2) 医学统计描述, 首先要确定观测变量的正态性, 然后根据正态性和非正态性选用不同的统计量, 因此, 本书把正态性检验放在统计描述中整合在一起来讲解。

(3) 在计数资料的 χ^2 检验中, 医学科研工作者经常具有频数表资料而非原始资料, 本书把“Data”中的“Weight Cases”与 χ^2 检验整合在一起介绍。

(4) 在多组的行 X 列 χ^2 检验中, 经常要做行 X 列分割。如何做行 X 列分割? 本书把 χ^2 检验与“Missing Value”整合处理。

(5) 在协方差分析中, 首先要确定协变量, 本书介绍了在做协方差分析之前如何确定协变量, 并且在操作过程中如何正确使用“Split file”。

(6) 医学科研中, 计量资料的两组或多组均数的统计学检验中, 首先要确定观测变量的正态性和方差齐性, 本书把相关内容整合在一起进行统计分析。

如此等等, 贯穿全书, 因此, 本书是 SPSS 与医学科研统计方法的整合。

2. 为了使读者能正确使用统计方法, 对大样本或小样本, 要校正或不要校正, 什么条件下校正, 如何校正都作了适当的补充和叙述, 尤其在四格表 χ^2 统计分析中, 补充了较完善的例题。

应该指出 SPSS 是世界上通用的统计软件包之一。它无须编写程序, 完全可以在 Windows 下通过“菜单”、“对话框”, 使用鼠标来操作, 为读者提供极大方便, 深受用户欢迎。

在医学科研中存在大量统计问题, 主要有: 如何建立数据文件; 如何选择统计方法; 如何正确使用统计软件包; 如何恰当地解释统计分析结果。实际上, 这四者是密切联系、不可分割的, 本书把它们有机地整合在一起。

作者积 40 余年的医学中统计分析的经验, 从医学科研中实际问题出发, 深入浅出地介绍如何正确使用 SPSS 统计软件包, 期望给读者提供一个简单、明了、正确地进行医学数据处理的方法。

SPSS 内容很多, 涉及面也很宽, 本书只介绍在医学中常用的统计内容和方法, 有很强的实用价值。本书中未提及的内容, 读者可查阅 SPSS 的原版使用手册。

我的博士研究生陈卉副教授参加了本书第十一章、第十二章及第十五章的编写和修改工作, 对她的认真细致的工作深表感谢。

首都医科大学医学专业的博士生、硕士生在学习本教材及做论文课题的过程中, 对本教材提出了许多建设性建议, 相当一部分在本次修改中采纳了, 在此向他们表示感谢!

由于编者水平有限, 书中难免有不妥之处, 望读者批评指正, 以利提高。

马斌荣

2014 年 9 月

于首都医科大学

目 录

第五版前言

第一章 数据文件的建立	1
第一节 数据编码	3
第二节 定义变量和构建数据文件框架结构	5
第三节 数据录入	9
第四节 数据文件的存储与读入	11
第五节 根据已存在的变量建立新变量	15
第二章 清理数据与基本统计分析	20
第一节 对不合理数据的检查与清理	20
第二节 相关变量之间的逻辑检查与清理	23
第三节 统计描述	25
第三章 t 检验	32
第一节 单个样本 t 检验	32
第二节 配对样本 t 检验	35
第三节 两组独立样本的 t 检验	39
第四章 方差分析	45
第一节 单因素方差分析	45
第二节 方差分析中均数的两两比较 (multiple comparison in ANOV)	49
第三节 随机区组设计的方差分析	52
第四节 多因素方差分析	59
第五节 析因分析	62
第五章 χ^2 检验	68
第一节 行 $\times$ 列 χ^2 检验	68
第二节 四格表 χ^2 检验	77
第三节 配对 χ^2 检验	85
第六章 秩和检验	90
第一节 配对比较的秩和检验	90
第二节 两独立样本比较的秩和检验	92
第三节 有序变量的两独立样本比较的秩和检验	95
第四节 多个独立样本比较的秩和检验	98
第七章 相关与回归	103
第一节 一元线性相关与回归	103
第二节 多元相关分析	117

第三节 多元线性回归	120
第四节 多元逐步回归	127
第五节 曲线回归	129
第八章 Logistic 回归	141
第一节 Logistic 回归的基本概念	141
第二节 Logistic 回归的统计分析	143
第九章 协方差分析	156
第一节 完全随机设计的协方差分析	156
第二节 随机配伍组设计的协方差分析	163
第三节 协方差分析中的组间的两两比较	169
第十章 聚类分析	172
第一节 对样品聚类	172
第二节 对指标聚类	179
第三节 K-Means 聚类	184
第十一章 诊断试验的评价与 ROC 曲线分析	194
第一节 诊断试验的评价	194
第二节 诊断试验的一致性检验——Kappa	200
第三节 诊断试验评价的 ROC 曲线分析	207
第十二章 生存分析	220
第一节 生存分析的基本概念	220
第二节 寿命表法	221
第三节 Kaplan-Meier 法	226
第四节 Cox 比例风险回归模型	234
第十三章 主成分分析	238
第一节 主成分分析的基本概念	238
第二节 主成分分析的操作步骤	240
第十四章 量表的信度效度分析	247
第一节 量表的信度分析	247
第二节 量表的效度分析	256
第十五章 统计图形	263
第一节 条形图	263
第二节 饼图	276
第三节 线图	280
第四节 直方图	286
第五节 散点图	287
第十六章 统计表	290
第一节 一般的二维统计表	290
第二节 嵌套式二维统计表	294

第一章 数据文件的建立

数据文件的建立是指把科研工作过程中采集的各种信息(information)、数据(data)以某种方式存入计算机的磁盘中,建立可随时存取、修改、统计分析的数据文件的全过程。一般来说,它包括数据编码、建立数据文件的框架结构、数据录入、选定数据文件名及保存数据等几个步骤。

例 1.1 北京儿童医院李龙教授进行了“围手术期输血与先天性巨结肠术后感染关系的探讨”的研究课题,共观察了 86 个病例,每个病例采集的数据为性别、年龄(月龄)、红血球压积、手术方式、疾病部位、手术持续时间(min)、手术中失血量(mL)、手术中输血次数、手术中输血量(mL/kg 体重)及感染与否等。

其原始数据如表 1.1 所示。

表 1.1 围手术期输血对先天性巨结肠术后感染关系的探讨的数据

病例号	性别	月龄	红血球压积	手术方式	疾病部位	手术持续时间/min	手术中失血量/mL	手术中输血次数	手术中输血量/(mL/kg 体重)	感染与否
1	男	11	56.4	环型	乙状结肠	200	40	1	10.0	无感染
2	男	4	32.5	Z 型	结肠	215	40	1	15.2	无感染
3	男	10	37.8	Z 型	直肠	190	40	1	13.5	无感染
4	男	22	37.9	吻合型	结肠	250	40	2	30.0	感染
5	女	7	47.8	环型	乙状结肠	145	40	1	16.7	无感染
6	女	6	47.4	吻合型	直肠	205	60	2	18.3	无感染
7	男	45	54.7	吻合型	直肠	210	40	3	21.4	无感染
8	男	1	98.3	吻合型	直肠	270	20	3	30.3	感染
9	男	1	47.0	吻合型	直肠	180	40	3	31.3	无感染
10	男	4	31.6	吻合型	乙状结肠	180	40	1	20.0	无感染
11	男	15	49.7	Z 型	直肠	190	40	1	10.0	无感染
12	男	5	31.8	吻合型	乙状结肠	170	40	1	25.0	无感染
13	男	1	52.3	环型	直肠	135	30	4	35.7	无感染
14	男	9	46.6	Z 型	直肠	245	40	1	12.5	感染
15	男	1	76.4	吻合型	乙状结肠	200	20	3	32.3	感染
16	男	144	48.1	吻合型	乙状结肠	325	40	1	7.4	感染
17	男	11	80.8	吻合型	乙状结肠	280	100	2	18.8	感染
18	男	2	56.1	吻合型	直肠	225	20	2	22.7	无感染

续表

病例号	性别	月龄	红血球 压积	手术 方式	疾病 部位	手术持 续时间 /min	手术中 失血量 /mL	手术中 输血 次数	手术中输血 量/(mL/kg 体重)	感染 与否
19	男	17	41.2	吻合型	直肠	225	40	1	16.7	感染
20	男	60	41.9	吻合型	结肠	270	40	1	10.5	无感染
21	男	2	52.2	吻合型	直肠	165	30	4	40.3	无感染
22	男	78	53.7	Z型	直肠	275	40	2	10.5	无感染
23	男	5	33.8	吻合型	乙状结肠	140	40	1	16.7	无感染
24	女	4	58.7	环型	乙状结肠	110	40	4	58.3	感染
25	男	16	43.3	吻合型	乙状结肠	165	40	1	9.1	无感染
26	男	3	45.2	环型	乙状结肠	130	10	3	32.6	感染
27	男	38	48.5	吻合型	乙状结肠	175	40	2	14.8	感染
28	男	1	57.1	环型	直肠	140	20	2	29.4	无感染
29	男	8	50.2	吻合型	结肠	225	20	3	28.9	感染
30	男	120	55.6	吻合型	乙状结肠	230	40	1	7.2	无感染
31	男	15	49.7	Z型	直肠	190	30	1	11.1	无感染
32	男	13	54.5	环型	直肠	155	40	2	13.6	无感染
33	男	4	42.2	环型	直肠	130	10	2	14.7	无感染
34	男	29	61.2	环型	直肠	120	40	2	30.0	感染
35	男	3	34.6	吻合型	结肠	175	40	2	20.8	无感染
36	男	24	43.0	环型	直肠	170	40	1	10.0	无感染
37	男	3	45.5	吻合型	乙状结肠	165	40	2	30.0	无感染
38	男	65	72.9	环型	直肠	240	40	1	11.1	无感染
39	男	26	48.1	Z型	结肠	180	100	1	11.1	感染
40	男	6	46.9	吻合型	直肠	195	30	1	16.1	感染
41	男	2	59.1	吻合型	直肠	230	50	3	39.1	无感染
42	男	10	45.8	Z型	乙状结肠	205	40	1	11.8	无感染
43	女	54	46.3	Z型	乙状结肠	270	40	1	7.4	感染
44	男	5	41.4	吻合型	乙状结肠	210	40	2	13.3	无感染
45	男	14	36.7	吻合型	直肠	160	40	1	10.0	感染
46	男	16	49.7	吻合型	直肠	250	100	2	11.5	无感染
47	男	10	80.8	吻合型	乙状结肠	280	100	2	18.9	感染
48	男	2	54.7	吻合型	乙状结肠	210	50	3	46.9	感染
49	男	26	82.4	吻合型	乙状结肠	230	50	2	20.0	无感染
50	男	25	39.8	吻合型	结肠	210	40	1	23.5	感染
51	女	3	34.7	吻合型	直肠	225	50	4	34.1	无感染
52	男	84	46.3	Z型	乙状结肠	215	40	3	55.6	无感染
53	男	46	54.9	吻合型	直肠	270	50	1	11.8	无感染
54	男	5	45.7	吻合型	乙状结肠	175	40	1	14.3	无感染

续表

病例号	性别	月龄	红血球 压积	手术 方式	疾病 部位	手术持 续时间 /min	手术中 失血量 /mL	手术中 输血 次数	手术中输血 量/(mL/kg 体重)	感染 与否
55	男	3	48.4	吻合型	乙状结肠	165	40	2	30.1	无感染
56	女	7	35.9	吻合型	结肠	175	20	2	20.0	感染
57	男	1	86.0	吻合型	结肠	260	20	4	61.5	无感染
58	男	9	46.4	Z型	直肠	205	30	1	5.9	无感染
59	男	8	50.2	吻合型	结肠	270	100	3	26.4	感染
60	女	1	80.2	吻合型	乙状结肠	195	30	3	33.3	无感染
61	男	4	34.8	吻合型	乙状结肠	180	40	0	0	无感染
62	女	15	35.8	吻合型	直肠	210	50	0	0	无感染
63	男	19	34.6	吻合型	结肠	200	50	0	0	无感染
64	女	45	47.7	吻合型	乙状结肠	210	40	0	0	无感染
65	男	13	74.5	Z型	乙状结肠	150	40	0	0	无感染
66	男	96	42.7	吻合型	直肠	225	80	0	0	无感染
67	男	60	58.5	Z型	结肠	190	100	0	0	无感染
68	女	96	45.9	吻合型	乙状结肠	235	150	0	0	无感染
69	男	48	42.6	吻合型	乙状结肠	180	40	0	0	无感染
70	男	10	51.5	吻合型	直肠	250	100	0	0	无感染
71	女	16	34.9	吻合型	直肠	190	40	0	0	无感染
72	男	60	45.8	Z型	乙状结肠	190	100	0	0	无感染
73	女	72	36.0	Z型	直肠	175	40	0	0	无感染
74	男	6	35.6	吻合型	直肠	230	40	0	0	无感染
75	男	3	48.2	Z型	乙状结肠	190	50	0	0	无感染
76	男	20	40.5	吻合型	结肠	140	40	0	0	无感染
77	男	36	34.2	吻合型	直肠	235	40	0	0	无感染
78	男	52	35.7	吻合型	直肠	150	20	0	0	无感染
79	男	6	48.0	环型	乙状结肠	140	40	0	0	感染
80	男	39	73.2	环型	直肠	165	50	0	0	无感染
81	男	16	47.7	环型	直肠	125	20	0	0	无感染
82	女	44	68.5	Z型	直肠	175	50	0	0	无感染
83	男	13	63.7	环型	乙状结肠	190	40	0	0	无感染
84	男	120	51.6	环型	直肠	135	50	0	0	无感染
85	男	11	52.6	环型	直肠	140	50	0	0	无感染
86	男	65	41.9	Z型	直肠	215	40	0	0	无感染

第一节 数据编码

在建立数据文件之前首先要进行数据编码(data code)。它包括如下三个层次。

一、定义数据项的变量名(variable name)

在例 1.1 中，涉及病例号、性别、月龄、红血球压积、手术方式、疾病部位、手术持续时间、手术中失血量、手术中输血次数、手术中输血量及感染与否等 11 个数据项，应该对每一个数据项确立一个变量名。本例中的变量名如下：

num	病例号
sex	性别
age	月龄
hct	红血球压积
pt	手术方式
da	疾病部位
time	手术持续时间
lb	手术中失血量
bn	手术中输血次数
bc	手术中输血量
infect	感染与否

为数据项取变量名时，应该遵循如下原则：

- (1) 变量名可以是多个字符，可以用汉语拼音或英文缩写，并且可以是汉字。
- (2) 变量名不能使用 SPSS 的保留字。SPSS 的保留字有 ALL、AND、OR、NOT、EQ、GE、GT、LE、LT、NE、TO、WITH 及一些常用的函数符号等。
- (3) 系统中不区分变量名中的大小写字母。例如，ABCD 与 abcd 被认为是同一变量。

二、定义数据项的变量标签(variable label)

定义变量标签是对该变量名所表示的数据项内涵的进一步说明，而变量标签不受字符位数的限制，可以较详细地表达该变量的内涵。变量标签可以用英文，也可以用中文。在统计分析过程的输出中，会在变量名对应的位置显示该变量的英文或中文的标签。这将有助于理解输出结果。

变量标签是一个可选择(optional)的属性，可以定义，也可以不定义。例 1.1 的变量标签如表 1.2 所示。

表 1.2 例 1.1 的变量标签

变量名	变量标签	变量名	变量标签
num	病人的病例号	time	手术时间
sex	病人的性别	lb	手术中失血量
age	病人的月龄	bn	手术中输血次数
hct	红血球压积	bc	手术中输血量
pt	手术方式	infect	感染与否
da	疾病部位		

三、定义变量取值的变量值标签(value label)

每个变量名对应一个数据项,每个变量取不同的值(简称为变量值),表示数据项中的不同信息。有时为了更好地理解统计分析过程中的输出结果,要进一步给变量不同的取值赋以不同的标签,以便对不同的信息作出解释。在输出结果的相应位置上会出现该标签,使读者一目了然。

并不是所有变量值都要定义标签,一般来说,只有离散变量才给变量值定义标签。例 1.1 的变量值标签如表 1.3 所示。

表 1.3 例 1.1 的变量值标签

变量名	变量值	变量标签
sex	1	男
	2	女
pt	1	环型
	2	Z 型
	3	吻合型
da	1	结肠
	2	乙状结肠
	3	直肠
infect	0	无感染
	1	感染

第二节 定义变量和构建数据文件框架结构

定义变量包括定义变量名、变量类型、变量长度(小数点位数)、变量标签、变量值标签、缺失值和变量显示格式(宽度、对齐方式)等。

定义变量的步骤如下。

1. 启动 SPSS Statistics 17.0, 进入定义变量的对话框

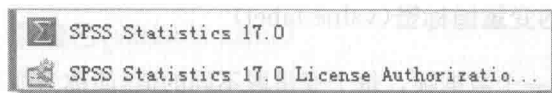
(1) 首先进入 Windows 2003/XP, 在“开始”菜单中找到 SPSS Statistics 17.0 的图标, 如图 1.1(a)或(b)所示。

(2) 单击图标中的“SPSS Statistics 17.0”或“PASW Statistics 17”, 即可启动 SPSS Statistics 17.0。

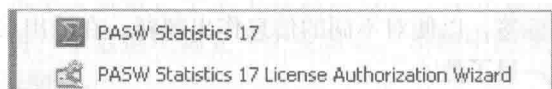
一旦启动, 将在屏幕上显示 SPSS Statistics 17.0 主界面(变量定义对话框), 如图 1.2 所示。

在图 1.2 显示的 SPSS Statistics 17.0 主界面的第一行出现了数据编辑(data editor)栏目, 它由 11 个菜单项目组成的主菜单。其内容如下:

① File 文件操作;



(a)



(b)

图 1.1 SPSS Statistics 17.0 图标

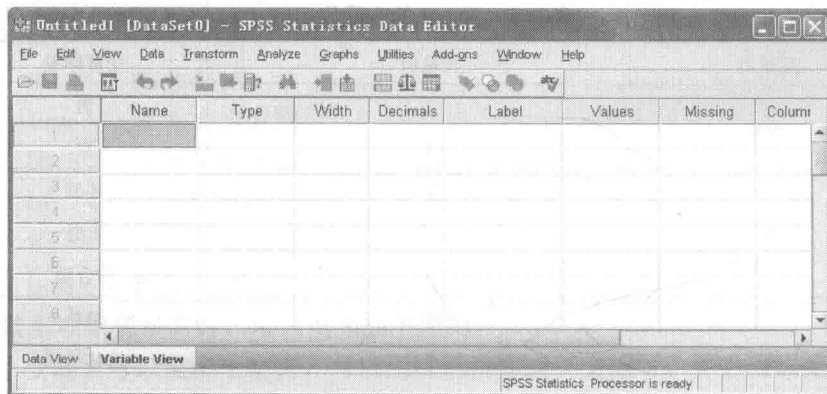


图 1.2 SPSS Statistics 17.0 主界面(变量定义对话框)

- ② Edit 文本编辑；
- ③ View 视图；
- ④ Data 数据文件的建立和编辑；
- ⑤ Transform 数据转换；
- ⑥ Analyze 统计分析；
- ⑦ Graphs 统计图表的建立和编辑；
- ⑧ Utilities 实用的一些工具；
- ⑨ Add-ons 辅助信息；
- ⑩ Window 窗口信息与控制；
- ⑪ Help 帮助。

每个菜单都包括一系列功能，用鼠标单击可出现下拉式菜单，供读者进一步选择和操作。这些功能将在后面章节中陆续介绍。

从图 1.2 左下角可见“Variable View”，表示当前它是定义变量的视图。

此时，屏幕出现与变量有关的字段如下：

Name	变量名
Type	变量类型
Width	变量宽度

Decimals	小数点
Label	变量标签
Values	变量值标签
Missing	缺失值
Columns	变量显示宽度
Align	变量对齐方式(左、右或中心对齐)

Measure 度量类型(定量变量、等级变量、定性变量)

2. 定义变量

以例 1.1 中的变量 num、sex、age、het、pt、da、time、lb、bn、bc、infect 为例,定义变量。

首先对病例号“num”进行操作。

(1) 在图 1.2 中的“Name”下键入“num”。

(2) 点击“Type”,在其下出现一个矩形框,再点击该矩形框右侧的图标,弹出定义变量类型的对话框,如图 1.3 所示。

在图 1.3 的左边出现 8 种可供选择的变量类型,如表 1.4 所示。

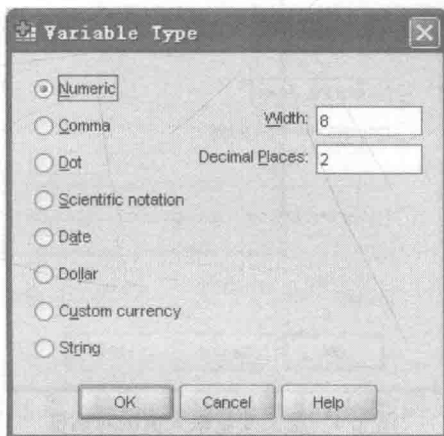


图 1.3 定义变量类型的对话框

表 1.4 8 种可供选择的变量类型

Numeric	数值型
Comma	带逗点的数值型
Dot	逗点作小数点的数值型
Scientific notation	科学记数法
Date	日期型
Dollar	带有美元符号的数值型
Custom currency	用户自定义型
String	字符串型

读者可以用光标选择其中一个类型,并点击“OK”。

如果选了“Numeric”,则在图 1.3 中“Numeric”右侧的“Width”、“Decimal Places”中,应调整相应的数值型变量的宽度(长度)及小数位数。其默认宽度为 8,默认小数位数为 2。应该注意,数值型的宽度应该大于该变量的数据的“整数位数+小数位数+1”。

如果选了“String”,则在图 1.3 中部的“Width”中键入字符串的长度。

本例的“num”,可选变量类型为“Numeric”,其“Width”为 4,“Decimal

Places”为0。

在图 1.3 的下方有三个按钮,“OK”表示确认,“Cancel”表示取消,“Help”表示获取帮助信息。此处点击“OK”表示确认,并回到图 1.2。

(3) 在图 1.2 的“Label”中应录入相应的变量标签,对于“num”应录入“病人的病例号”;对于变量“num”没有变量值标签,故不用“Values”。



图 1.4 定义变量缺失值的对话框

(4) 点击图 1.2 中的“Missing”,在其下出现一个矩形框,再点击该矩形框中右侧的图标,弹出定义变量“Missing Values”(缺失值)的对话框,如图 1.4 所示。

在 SPSS 中缺失值有两类,即系统缺失值和用户缺失值。系统缺失值不需定义,系统自动生成;若无用户缺失值,则可选图 1.4 中的“No missing values”。若有用户缺失值,并且变量的观察值是离散值,则可选图 1.4 中的“Discrete missing values”,并输入相应的缺失值。

如果变量的缺失值既有某个范围,也有离散值,则可选图 1.4 中的“Range plus one optional discrete missing value”。

本例不需定义缺失值,故可不操作此部分内容。

在图 1.4 的下方点击“OK”,回到图 1.2。

(5) 在图 1.2 的“Columns”中,调整变量显示宽度的数值,本例可选“5”。

(6) 在图 1.2 的“Align”中,可选“Right”(右对齐)。

(7) 在图 1.2 的“Measure”中,可选“Scale”(定量变量)。

至此,变量“num”已定义完毕。

其次对性别“sex”进行操作。

在图 1.2 的“Name”下录入“sex”,同时可选变量类型为“Numeric”,其“Width”为4,“Decimal Places”为0;“Label”中应录入相应的变量标签,对于“sex”录入“病人的性别”;对于变量“sex”应有变量值标签,故点击“Values”,在其下出现一个矩形框,再点击该矩形框中右侧的图标,可弹出定义变量值标签“Value Labels”(变量值标签)的对话框,如图 1.5 所示。

在图 1.5 中,读者用键盘键入变量值标签:先键入变量值,再键入该变量值对应的标签,然后再点击“Add”;重复该过程,直到所有的变量值及其标签都键完为止。在本例中,先键入变量值“1”,再键入该变量值的标签“男”,点击“Add”;键入变量值“2”,再键入该变量值的标签“女”,点击“Add”。

在图 1.5 的下方点击“OK”,回到图 1.2。

在图 1.2 的“Columns”中,调整变量显示宽度的数值,本例可选“4”。

在图 1.2 的“Align”中,可选“Right”(右对齐)。

在图 1.2 的“Measure”中,可选“Nominal”(定性变量)。



图 1.5 定义变量值标签的对话框

读者应该模仿此例，对变量“age”、“hct”、“pt”、“da”、“time”、“lb”、“bn”、“bc”及“infect”作出相应的变量名及其属性的定义。最后，得到了例 1.1 的数据文件的变量的框架结构，如图 1.6 所示。

	Name	Type	Width	Decimals	Label	Values	Missing	Columns
1	num	Numeric	3	0	病人的病例号	None	None	4
2	sex	Numeric	4	0	病人的性别	{1, 男}...	None	4
3	age	Numeric	5	0	病人的月龄	None	None	6
4	hct	Numeric	5	1	红血球压积	None	None	6
5	pt	Numeric	4	0	手术方式	{1, 环型}...	None	5
6	da	Numeric	4	0	疾病部位	{1, 结肠}...	None	4
7	time	Numeric	6	0	手术时间	None	None	6
8	lb	Numeric	6	0	手术中失血量	None	None	5
9	bn	Numeric	3	0	手术中输血次数	None	None	4
10	bc	Numeric	6	1	手术中输血量	None	None	6
11	infect	Numeric	3	0	感染与否	{0, 无感染}...	None	4
12								
13								
14								
15								

图 1.6 例 1.1 的数据文件的变量的框架结构

第三节 数据录入

在图 1.2 的下部，点击“Data View”，可得到数据文件的二维表格，如图 1.7 所示。

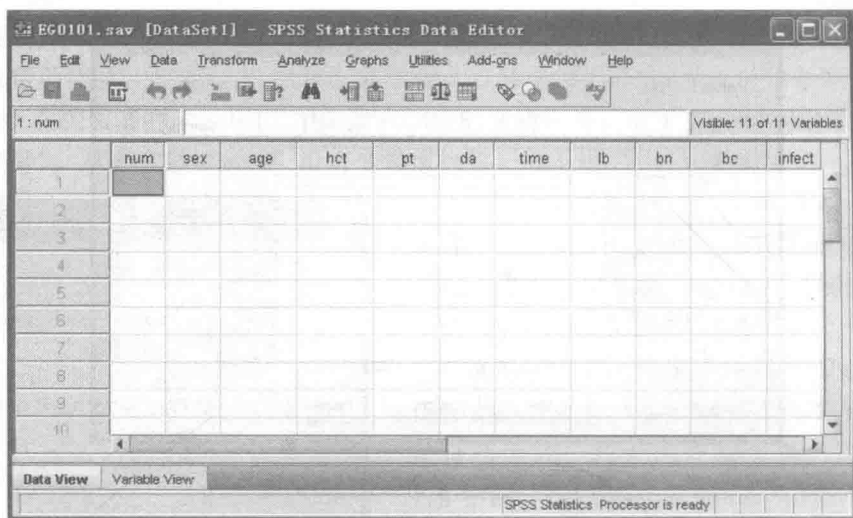


图 1.7 例 1.1 的数据文件的二维表格

图 1.7 中表格的顶部标有已定义的变量名“num”、“sex”、“age”、“hct”、“pt”、“da”、“time”、“lb”、“bn”、“bc”、“infect”；表格的左侧有观察量 1, 2, 3, 4, … 的序号。一个变量名和一个观察量序号就对应了二维表格中的一个单元格，反之亦然。

一、按变量输入数据

把光标移到要输入值的变量名对应的一列的顶部，即序号为 1 的单元格，并单击之，使该单元格为当前操作的单元格，输入该变量的第一个值，回车；此时，序号为 1 的单元格接受该数值，并且当前操作单元格下移到序号为 2 的单元格，输入第二个值，回车；如此下去，直到把该变量的数值输完为止。

在本例中，把光标移到变量名“num”列的顶部，对应观察量序号为 1 的单元格，并单击之，使之成为当前操作单元格。键入 1，回车；2，回车；3，回车；4，回车；5，回车；6，回车；……86，回车。把 86 个病人的病例号全部输入完毕。

再把光标移到变量名“sex”列的顶部，对应观察量序号为 1 的单元格，并单击之，使之成为当前操作单元格。键入 1，回车；1，回车；1，回车；1，回车；2，回车；2，回车；……1，回车。把 86 个病人的性别全部输入完毕。

然后，把光标移到变量名为“age”列的顶部，对应序号为 1 的单元格，并单击之，使之成为当前操作单元格，键入 11，回车；4，回车；10，回车；22，回车；7，回车；……65，回车。

同样可对变量“hct”、“pt”、“da”、“time”、“lb”、“bn”、“bc”、“infect”的数据进行录入。