

SPSS医学统计

速学速用

主编 — 于小林
— 刘淑贞
— 张翠香

 人民军医出版社
PEOPLE'S MILITARY MEDICAL PRESS

SPSS YIXUE TONGJJI SUXUE SUYONG

SPSS 医学统计速学速用

SPSS YIXUE TONGJI SUXUE SUYONG

主 编 于小林 刘淑贞 张翠香

副主编 周长政 杜国威 王志海 李海鹰
程晓峰

编 委 (以姓氏笔画为序)

于小林	王志海	王素凤	卢 琳
刘淑贞	孙 丽	孙荣丽	李海鹰
李绪彤	杜国威	张翠香	张鹤君
陈 洁	周长政	赵 杰	赵清叶
荆 航	程晓峰	樊金焱	



人 民 军 医 出 版 社

People's Military Medical Press

北 京

图书在版编目(CIP)数据

SPSS 医学统计速学速用/于小林,刘淑贞,张翠香主编. —北京:人民军医出版社,2007.8
ISBN 978-7-5091-1171-0

I. S… II. ①于…②刘…③张… III. 医学统计—统计分析—软件包,SPSS IV. R195.1-39

中国版本图书馆 CIP 数据核字(2007)第 123886 号

策划编辑:王显刚 文字编辑:荣 堃 责任审读:余满松

出 版 人:齐学进

出版发行:人民军医出版社

经销:新华书店

通信地址:北京市 100036 信箱 188 分箱

邮编:100036

质量反馈电话:(010)51927270;(010)51927283

邮购电话:(010)51927252

策划编辑电话:(010)51927300-8707

网址:[www. pmmp. com. cn](http://www.pmmp.com.cn)

印刷:北京京海印刷厂 装订:京兰装订有限公司

开本:787mm×1092mm 1/16

印张:19.75 字数:480 千字

版、印次:2007 年 8 月第 1 版第 1 次印刷

印数:0001~4000

定价:42.00 元

版权所有 侵权必究

购买本社图书,凡有缺、倒、脱页者,本社负责调换

内 容 提 要

SPSS 是非统计医学专业人员应用最多的统计工具。本书共分 10 章,从简单实用的目的出发,针对临床医护人员在医疗实践和医学科研中经常遇到的医学统计学问题,以日常医学统计中经常用到的统计功能为目标,以实例分析、统计具体步骤等通俗易懂的形式来讲解 SPSS 统计软件的应用。重点介绍各种常见的医学数理统计学的资料分析、方法选择以及程序具体操作等内容,便于读者了解资料的性质和结构,对所收集的资料进行理解分析,模仿、套用恰当的统计方法及统计程序操作,达到顺利完成整个数理统计过程的目的。本书适用于掌握了最基本的医学统计学理论而疏于计算机动手操作的非统计医学专业人员,可作为医学或生物学专业本科生、研究生的辅助教材、培训教材和继续教育用书,同时也可供广大非统计专业的医学临床、科研和教学人员自学用。

前 言

SPSS 是国际上最流行、最具有权威性的统计分析软件之一,也是非统计专业人员应用最多的统计工具。本书从简单实用的目的出发,针对临床医护人员在医疗实践和医学科研中经常遇到的医学统计学问题,以日常医学统计中经常用到的统计功能为目标,以实例分析、上机操作步骤等通俗易懂的形式来讲解 SPSS 统计软件的应用,重点介绍各种常见的医学数理统计学的资料分析、方法选择以及程序操作等具体处理方法,并在内容上根据临床统计实际应用的频度做一定调整,对目前教科书在列但临床应用较少的项目从略;而教科书中没有但实际应用较多的内容进行补充介绍,如反映治疗效果的非参数检验 Ridit 分析,非线性回归分析常用的曲线估计以及随访资料的生存分析等。本书不同于传统的数理统计学书籍,不将理论知识方面的讲解作为重点,但在实践操作方面做了详细的介绍和描述。

医学领域范围广阔,医学科研实践中经常会遇到高难度、复杂性和专业性特别强的问题,常常令非本分支专业的临床工作者迷惑和费解。为减少和避免因此而导致的对资料分析理解困难,影响统计学程序的实际操作和减少数据录入的工作量,本书未采用尖端技术领域的医学科研实例,而是列举通俗易懂的简单例子,便于读者了解资料的性质和结构,结合自己的专业特点,对所收集的资料进行理解分析,模仿、套用恰当的统计方法及统计程序操作,达到顺利完成整个数理统计过程的目的。用于操作练习的实例设计简单,样本量也不足,有些甚至不符合医学科研设计要求,不必拘泥其数字技术性、科学性、准确性和刻意分析实例的研究设计问题,更勿模仿进行科研设计,仅用于实际操作练习。

本书适用于掌握了最基本的医学统计学理论而疏于计算机动手操作的非统计医学专业人员,可作为医学或生物学专业本科生、研究生的辅助教材、培训教材和继续教育用书,同时也可供广大非统计专业的医学临床、科研和教学人员自学用。

本书引用的参考文献较多,除书中所列出的主要参考文献之外,还参阅不少近年来国内医学统计学文献,在此特向相关作者表示衷心的感谢。

由于我们的水平有限,特别是所有编写人员均为非统计和计算机专业的医务工作者,书中缺点和错误难免,恳望医学统计学界各位老师及读者不吝赐教。

编 者

2007 年 3 月于青岛

目 录

第 1 章 SPSS 医学统计基本操作	(1)
第一节 在收集资料的同时完成数据编码	(1)
第二节 SPSS 数据编辑和变量设置窗口	(3)
第三节 定义与设置变量	(4)
第四节 数据的录入和编辑	(9)
第五节 Excel 文件的读取	(14)
第六节 列联表资料的数据录入	(17)
第七节 2 个数据文件的合并	(19)
第八节 变量间的运算	(22)
第九节 数据的划分和分组	(27)
第十节 数据的排序	(31)
第 2 章 常用统计学描述与资料的检验	(34)
第一节 常用统计数据类型	(34)
第二节 临床资料的常用统计描述	(36)
第三节 偏度的计算	(42)
第四节 资料的频数分布分析	(45)
第五节 定量资料的正态性检验	(49)
第六节 频数资料的正态性检验	(52)
第七节 方差齐性检验	(55)
第八节 原始资料构成比和率的计算	(58)
第 3 章 t 检验	(62)
第一节 单样本 t 检验	(62)
第二节 两独立样本 t 检验	(66)
第三节 配对样本 t 检验	(75)
第 4 章 方差分析	(82)
第一节 单因素方差分析	(82)
第二节 随机区组设计资料的多因素方差分析	(91)
第三节 析因设计资料的方差分析	(101)
第四节 重复测量设计的方差分析	(107)
第五节 协方差分析	(109)
第 5 章 χ^2 检验及其他率比较的统计学方法	(118)

第一节	四格表资料的 χ^2 检验	(118)
第二节	$R \times C$ 表资料的 χ^2 检验	(127)
第三节	原始数据资料的 χ^2 检验	(136)
第四节	配对资料的 χ^2 检验	(140)
第五节	样本率与总体率比较的检验	(143)
第六节	分层危险度估计的 χ^2 检验	(145)
第七节	与总体二项分布比较的检验	(151)
第 6 章	秩和检验和其他非参数检验方法	(154)
第一节	定量资料两独立样本的秩和检验	(154)
第二节	成组设计等级/频数表资料的秩和检验	(159)
第三节	成组设计多个样本率比较的秩和检验	(165)
第四节	定量资料两配对样本的秩和检验	(168)
第五节	有序多分类资料试验组与标准组比较的 Ridit 分析	(171)
第六节	两个对比组比较的 Ridit 分析	(178)
第 7 章	相关与回归	(185)
第一节	相关分析	(185)
第二节	线性回归分析	(196)
第三节	曲线估计与非线性回归分析	(203)
第四节	Logistic 回归分析	(206)
第 8 章	诊断性试验与方法学评价	(220)
第一节	诊断方法的灵敏度、特异性及正确率等指标的计算	(220)
第二节	临床诊断一致性 Kappa 检验	(225)
第 9 章	生存分析	(230)
第一节	寿命表生存分析	(230)
第二节	Kaplan-Meier 生存分析	(234)
第三节	COX 回归分析	(240)
第 10 章	常用统计图的制作	(246)
第一节	饼图	(246)
第二节	条图	(251)
第三节	散点图	(265)
第四节	线图	(269)
第五节	箱图	(272)
第六节	面积图	(276)
第七节	直方图	(280)
附录	SPSS 常用词汇英中对照表	(282)

第 1 章 SPSS 医学统计基本操作

第一节 在收集资料的同时完成数据编码

医学领域的科研离不开资料的收集和积累。科学有效的方法能大大提高工作效率,减少差错。临床资料的收集在整个科研工作中起着举足轻重的作用。资料收集工作完成后,数据的计算机录入工作相当费工费时。这里介绍一种简便可行的方法,即在临床资料的收集过程中同时完成数据的编码工作,并进行计算机的数据录入。所谓的数据编码,就是用 SPSS 软件程序所能接受的代码来代替所收集临床资料的内容,比如用代码 1 代表男性,代码 2 代表女性等。做到边收集,边编码,边录入,提高工作效率。

统计实例 为探讨房性心律失常(ARA)与左心房解剖结构的关系,随机对体表心电图检出 ARA 的患者进行超声心动图左心房直径(LAD)和心电图 P 波时间(PW)分析,同时随机选取无 ARA 者进行对照。资料统计表如下。

编号	性别	年龄	诊断	DCGARA	LAD(ms)	Pmax(ms)	Pmin(ms)	PD(ms)
1	男	46	冠心病	Ⅲ级	39.4	126	74	
2	男	38	健康	0级	30.3	108	68	
3	男	54	高血压	Ⅳ级	40.5	134	66	
4	女	32	心肌炎	Ⅱa级	29.2	89	62	
5	女	47	高血压	V级	38.4	122	71	

为方便操作和理解,仅选了 5 例,观察项目也不多。实际在临床科研工作中,资料收集量非常大。下面简单介绍一下资料收集和编码过程中的几个简便规则。

编号:变量名多采用 ID。一般表示所收病例或记录 case 的先后顺序。由于多数的 SPSS 程序统计运算不需要编号,对 case 的顺序也无特殊要求,因此可以省略。但在数据资料较大的情况下,观察项目的查找、比较、补充等方面,有 ID 还是很方便的。需要注意的是,EXCEL 和 SPSS 数据编辑区均有记录号,但与 ID 不是一回事,当 case 顺序发生变化时,记录号不随之发生改变。因此,在需要明确 case 顺序先后的情况下,建一个 ID 非常重要。

性别:一般变量名取 gender 或 sex。如果在资料的收集过程中就以编码的形式进行,如将 1 作为男性代码,0 或 2 为女性代码,就能给以后的计算机输入提供了很大的方便。这样转化的数据能为所有版本的 SPSS 程序直接调用,无论软件程序是否支持汉化。其他的二项分布类资料,如生存/死亡、阴性/阳性、有/无、是/否等,均可用 1/0 或 1/2 来进行代码标记。有些 SPSS 程序规定最小值为阳性事件,因此在编码时要特别注意,最好一开始就养成习惯,将阳性或异常事件用最低的数字代码表示。编码后的数据资料属性不变。

诊断:变量名取一个熟知的英文名,如 Dise 或 Dig。在 SPSS 变量设置窗口中的“Label”

选项中可加中文标注,达到统计过程和结果读取时直观易懂的目的。如果所诊断的疾病种类有多个,可以用数字作为代码分别表示。比如,1 表示健康,2 表示高血压病,3 表示冠心病……等等。以编码的形式记录数据,不但快捷简便,还为以后的计算机录入提供方便。需要注意的是,编码之后的数据应当始终作为疾病诊断的编码对待,切不可误作定量资料的数值进入统计运算,否则会出现“性别的均数为 0.56”,“病种的均数为 2.73”等,造成统计结果的错误。

DCGARA:意指 Holter 心电监护所记录的 24h 房性心律失常的布氏分级情况。无房性心律失常为 0 级,有房性心律失常者按表现的轻重分为 I、II a、II b、III、IV 和 V 级,连同 0 级共有 7 个等级。具备这种特征的数据资料称为有序资料或等级资料。编码可按 1 个等级用 1 个数字,比如本例用 1~7 或 0~6 表示不同的房性心律失常分级情况。需要注意的是,等级资料编码后,数据属性依然是等级资料。例如,本例第 1 例的 4 绝不代表数值 4,而是代表布氏分级 III 级,千万不可将其混为定量资料的数值进入统计运算,从而导致统计结果的不真实。

LAD(ms):指左心房内径,属于定量资料。按实际测算的数值录入即可。如果设计中有分组的情况,如 $LAD > 33\text{mm}$ 为左心房增大,其中轻度 $34 \sim 40\text{mm}$,中度 $40 \sim 50\text{mm}$,重度 $> 50\text{mm}$ 等,不必另开项目进行录入,在统计程序操作的过程中交给 SPSS 完成即可。只要有原始数据,SPSS 可以根据操作者的要求计算和产生新的变量数组来。

Pmax(ms)、Pmin(ms)和 PD(ms)是体表心电图测得的 3 个 P 波时间阈指标,其中前两项为实测数值,属于定量资料,后者也是定量资料,由前两项计算得来,关系式为: $PD = P_{\max} - P_{\min}$ (单位是 ms)。

本例可以将每例 PD 用手工或计算器计算出来,然后录入到 SPSS 数据区中,但是显然不如让 SPSS 程序进行变量的计算同时产生一个新的变量 PD 那样工作效率高。因此,凡是能用变量间计算产生新变量的情况,不必急于手工计算和录入,交给 SPSS 程序完成,这样既可以节省工时,又避免计算和录入出错。

经过上述编码和整理,统计表可换成下述格式。

ID	gender	age	Dise	DCGARA	LAD(mm)	PW(ms)	Pmin
1	1	46	3	4	39.4	126	74
2	1	38	1	0	30.3	108	68
3	1	54	2	5	40.5	134	66
4	2	32	4	2	29.2	89	62
5	2	47	2	6	38.4	122	71

为了方便数据的收集和录入,可以根据编码的内容制作一个变量数据编码表,在录入的时候参考,以免弄错。例如本例:

变量名	name	0	1	2	3	4	5	6
性别	gender	/	男	女	/	/	/	/
诊断	Dise	/	健康	高血压	冠心病	心肌炎	糖尿病	其他
DCGARA	DCGARA	0 级	I 级	II a 级	II b 级	III 级	IV 级	V 级

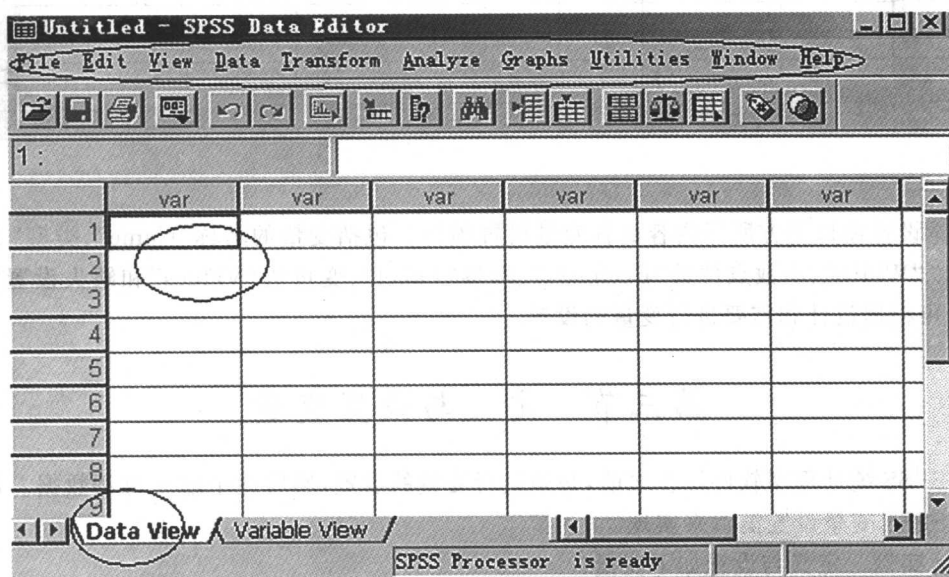
如果有新的资料需要追加,可以直接用上述编码的形式进行记录,这样就实现了在收集资料的同时完成数据编码工作,并且直接录入计算机中,大大提高工作效率。

多人或多个单位同时在不同地点所收集的病例数据文件,可用 SPSS 程序进行合并。

关于如何操作变量间的计算和产生新变量,如何合并 2 个或多个数据文件,将在后面章节中介绍。

第二节 SPSS 数据编辑和变量设置窗口

和大多数 windows 程序一样,SPSS 程序操作的形式是弹出窗口和人机会话。数据编辑和变量设置是 SPSS 的主要和最常用的功能界面。启动 SPSS 后,首先进入 SPSS 主画面,也就是数据编辑窗口。



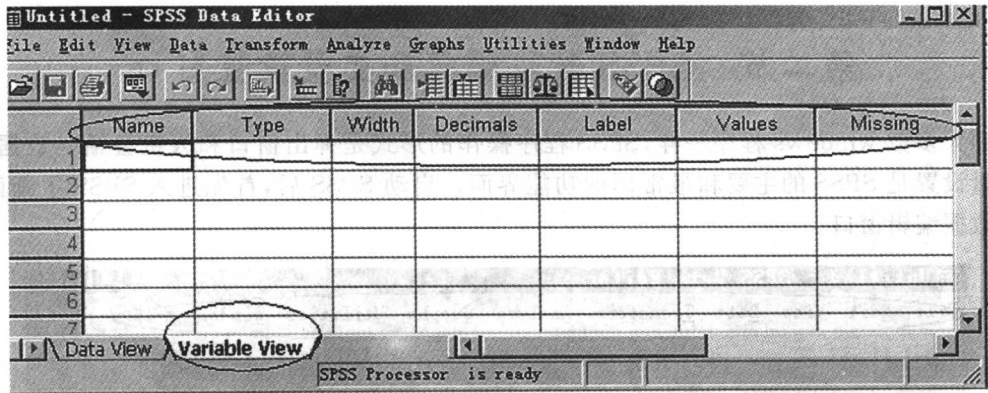
窗口的上列是 SPSS 的各程序菜单栏。相关功能在以后的章节中介绍。

中间标示的区域为数据区。其中,列表示变量,每个变量为 1 列,如果一个数据文件有 4 个变量,在数据区应当有 4 列。行为数据记录。如下图:

File Edit View Data Transform Analyze Graphs Utilities Window						
23 :						
	group	sex	freq	var	var	
1	1	1	56			
2	1	0	78			
3	2	1	64			
4	2	0	39			
5						

该数据文件共有 3 个变量,分别是分组变量“group”、性别变量“sex”和频数变量“freq”。3 个变量各有 4 条记录。

左下方是数据区选择书签“Data View”。其右边是变量设置选择书签“Variable View”。点击相应书签即可进入相应的窗口。



变量设置窗口上方所标内容是各变量设置项目。包括变量的名称“Name”、类型“Type”、字符宽度“Width”、小数点位数“Decimals”、标签“Label”、变量值“Values”和缺失设置“Missing”等,可根据设计和需要进行变量的设置。

第三节 定义与设置变量

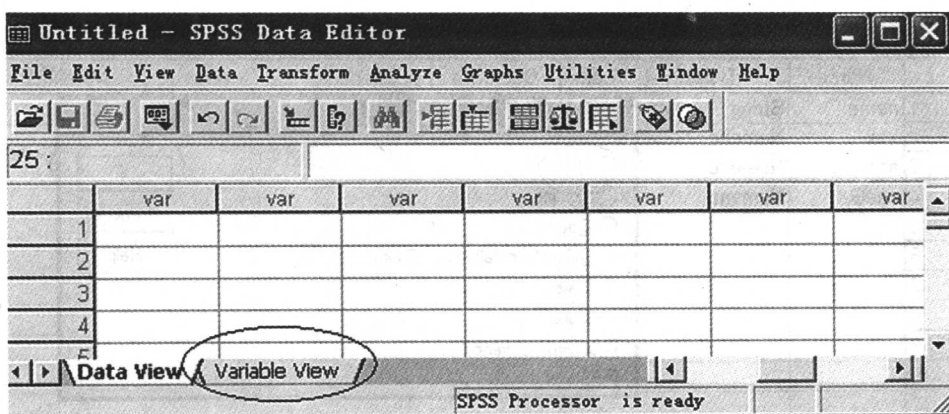
在 SPSS 统计程序软件操作之前,必须进行变量的设置,然后逐个输入每例数据。设置变量通过 SPSS 变量设置窗口来实现。

统计实例 定义与设置如下变量:

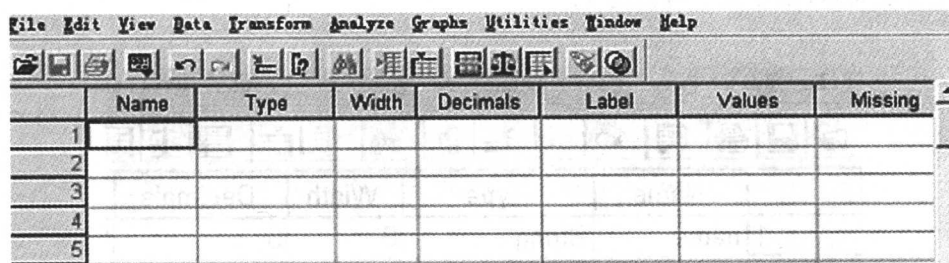
名称	类型	长度	小数点位数	标签	变量值标签	缺失值定义	对齐方式
name	字符	8	0	姓名	/	/	居中
sex	数值	2	0	性别	*	/	居中
age	数值	3	0	年龄	/	/	居中
ageG	数值	2	1	年龄分级	**	/	居中

表中,* 性别变量 sex;1=男,2=女 ** 年龄段分级变量 ageG:1= ≤ 30 岁,2=31~,3=41~,4=51~,5=61~,6= ≥ 70

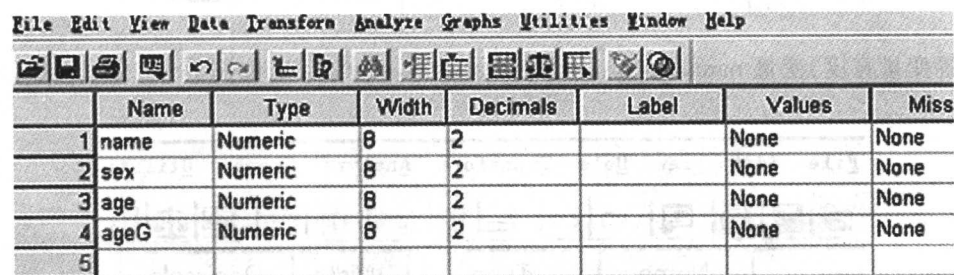
启动 SPSS,进入主画面,在窗口的下方点击变量设置书签“Variable View”。



进入变量设置窗口。

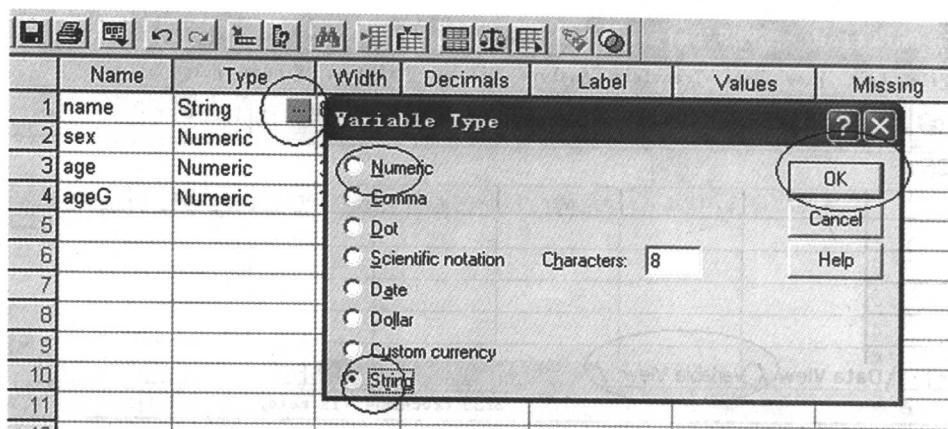


1. 定义变量名 直接在“name”列中键入各项变量名。程序立即给出各变量各项设置的默认值。



2. 设置变量类型 除变量 name 外,其他要定义的各项变量都是数值型,与 SPSS 默认值吻合,因此不必另设。ageG 变量可由 age 变量计算出并替换,暂不作介绍。这里仅对变量 name 类型进行设置。

点击变量 name 的“type”单元格,激活右边的小灰色按钮,点击该按钮出现“Variable Type”对话框。选中字符型变量类型“string”,通过右上方“OK”按钮退回,如图。



3. 设置变量宽度和小数点位数 先设小数点位数再设变量宽度,操作方法如图。
设置变量 ageG 的小数点位数为 1:

	Name	Type	Width	Decimals
1	name	String	8	0
2	sex	Numeric	8	0
3	age	Numeric	8	0
4	ageG	Numeric	8	1
5				

- 设置变量宽度:变量 name 为 8,sex 为 2,age 为 3,ageG 为 2,如图。

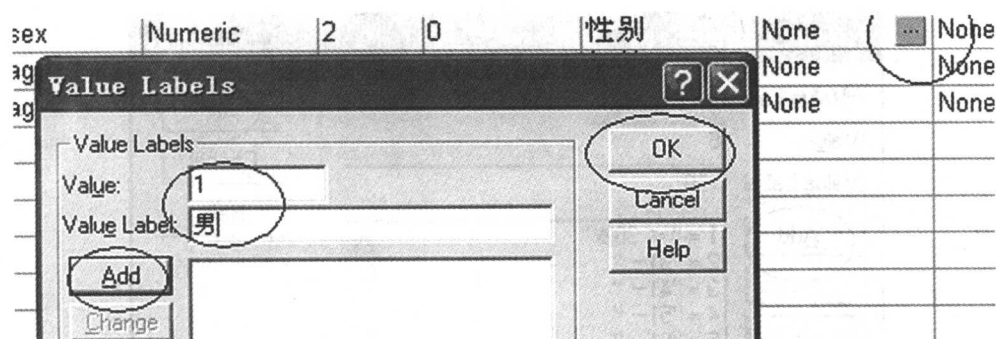
	Name	Type	Width	Decimals
1	name	String	8	0
2	sex	Numeric	2	0
3	age	Numeric	3	0
4	ageG	Numeric	2	1
5				

4. 设置变量标签 直接在相应的单元格内键入标签名称即可,如图。

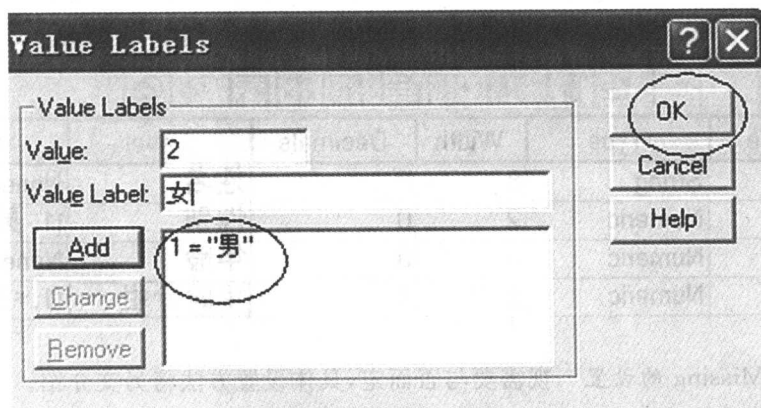
	Name	Type	Width	Decimals	Label
1	name	String	8	0	姓名
2	sex	Numeric	2	0	性别
3	age	Numeric	3	0	年龄
4	ageG	Numeric	2	1	年龄分组

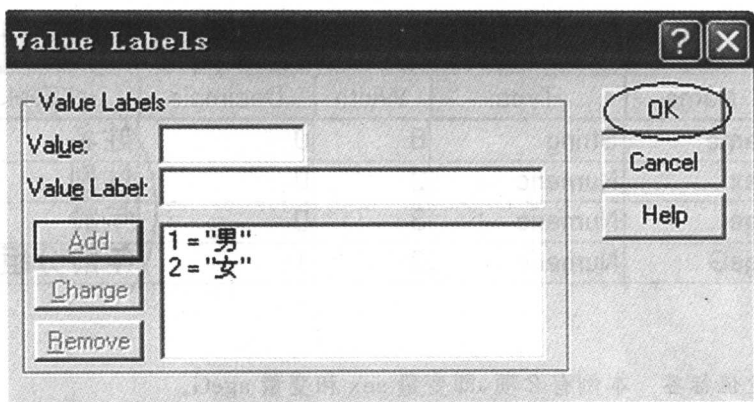
5. 定义变量值标签 本例有2项,即变量 sex 和变量 ageG。

先定义变量 sex。点击变量 sex 对应的 Values 单元格,右边出现小灰色选择框,点击小框弹出“Value Labels”对话框。取值“1”为变量值“男”,通过按钮“add”调入下方的大框内,如图。

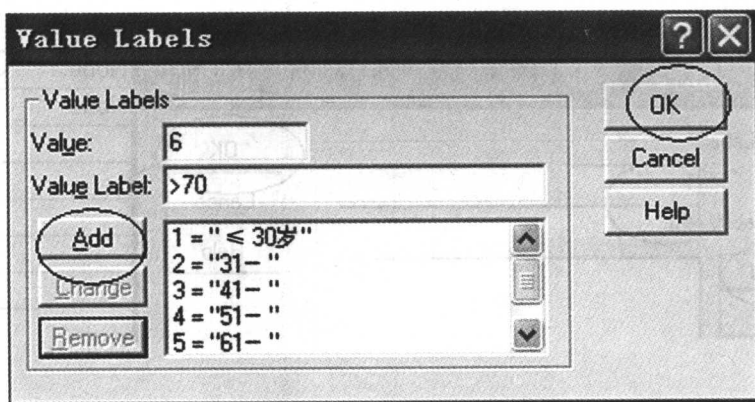


框内出现“1=男”字样。取值“2”为变量值“女”,通过按钮“add”调入下方的大框内,如图。





框内出现 1 = “男”, 2 = “女”字样。Sex 的变量标签定义完毕, 点击按钮“OK”退出对话框。用同样的方法定义年龄分组 ageG 的变量值标签, 如图。



定义完后的窗口界面如图。

Edit View Data Transform Analyze Graphs Utilities Window Help						
	Name	Type	Width	Decimals	Label	Values
1	name	String	8	0	姓名	None
2	sex	Numeric	2	0	性别	{1, 男}...
3	age	Numeric	3	0	年龄	None
4	ageG	Numeric	2	1	年龄分组	{1, ≤30岁}...

6. 缺失值 Missing 的设置 视需要与否而定, 具体设置方法将另文介绍。

7. 设置变量的排列方式 Align 设定各单元格数据居中。

点击“Align”相应单元格右边, 激活一箭头按钮, 点击按钮出现一小下拉菜单, 选择“Cen-

ter”(居中)。

Label	Values	Missing	Columns	Align	Measure
姓名	None	None	8	Center	Nominal
性别	{1, 男}...	None	8	Center	Scale
年龄	None	None	8	Left	Scale
年龄分组	{1, ≤30岁}...	None	8	Right	Scale
				Center	

至此,4个变量均定义和设置完毕。点击窗口下方的“Data View”书签切换到 SPSS 数据编辑窗口。在数据区填入各数据后,均为居中显示,界面如下。

File Edit View Data Transform Analyze Graphs Utilities Wind					
21:					
	name	sex	age	ageG	var
1	zhangsan	1	43	3.0	
2	lisi	2	34	2.0	
3	wangwu	1	26	1.0	
4	zhaoliu	1	48	3.0	
5					

第四节 数据的录入和编辑

在临床科研和统计工作中,当完成了资料的收集工作之后,要进行整理、计算机录入和统计学分析等一系列工作,其中数据的计算机录入工作占有较大的比重,特别是在一些观察项目多、样本收集量大的情况下更是如此。这就需要程序操作者有较广泛的计算机软件知识和娴熟的操作技术。数据的录入与编辑是必不可少的工作项目之一。

统计实例 1 数据的直接录入

将下述统计结果录入 SPSS 数据区中(变量名部分不作为数据录入内容)。

sex	age	HR	SBP	DBP	weight
1	46	76	160	100	62
2	32	68	130	90	56
1	67	72	120	80	60

【统计步骤】 在 SPSS 数据编辑窗口, 点击要录入数据的单元格, 出现黑框表示激活, 可直接将数据录入。

	sex	age	HR	SBP
1				
2				
3				
4				

	sex	age	HR	SBP
1	1	1		
2				
3				

	sex	age	HR	SBP	DBP	weight
1	1	46	76	160	100	62
2	2	32	68	130	90	56
3	1	67	72	120	80	60
4						

统计实例 2 数据的修改

将“HR”的第 2 条记录由“68”改为“88”。

【统计步骤】 点击要修改的记录项, 出现黑框表示被激活, 直接键入要修改的数字“88”。

	sex	age	HR	SBP	DBP
1		46	76	160	100
2		32	68	130	90
3		67	72	120	80

	age	HR	SBP	DBP
46	76	160	100	
32	88	130	90	
67	72	120	80	

统计实例 3 变量名的修改

将变量名“sex”改为“gender”。

【统计步骤】 点击窗口下方的“Variable View”书签, 进入变量设置窗口。