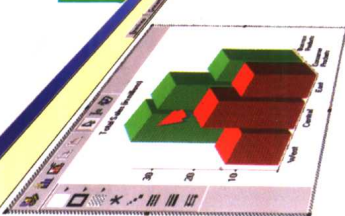
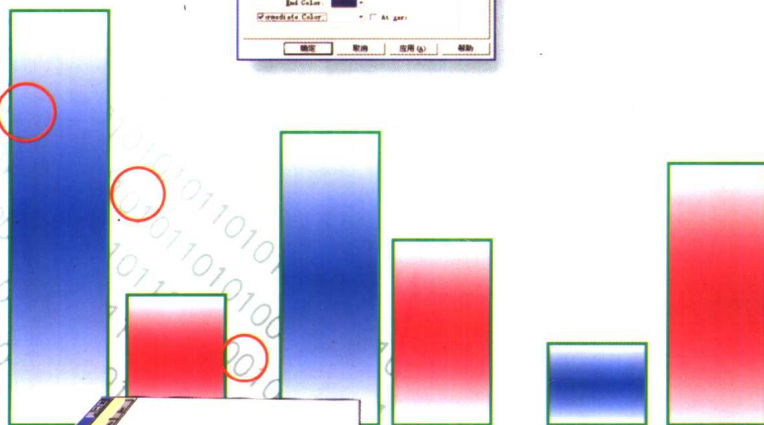
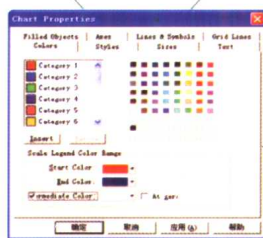


SPSS10.0 统计软件

主编 陈平雁 黄浙明

SPSS 10.0
TONGJI RUANJIAN
GAOJI YINGYONG
JIAOCHENG

高级应用教程



人民军医出版社

PEOPLE'S MILITARY MEDICAL PUBLISHER

SPSS10.0 统计软件高级应用教程

SPSS10.0 TONGJI RUANJIAN
GAOJI YINGYONG JIAOCHENG

主 编 陈平雁 黄浙明

副主编 易 东 安胜利 欧春泉 魏朝晖 邹尔新

编 者 (以姓氏拼音为序)

安胜利 蔡湛宇 陈平雁 何礼明 胡 静

黄浙明 莫一心 欧春泉 魏朝晖 易 东

张 超 张彦琦 邹尔新



人 民 军 医 出 版 社

People's Military Medical Press

北 京

图书在版编目(CIP)数据

SPSS10.0 统计软件高级应用教程/陈平雁,黄浙明主编. —北京:人民军医出版社,2004. 1
ISBN 7-80157-986-0

I. S… II. ①陈…②黄… III. 统计分析-软件包, SPSS10.0-教材 IV. C819

中国版本图书馆 CIP 数据核字(2003)第 059207 号

主 编:陈平雁 黄浙明

出 版 人:齐学进

策划编辑:张怡泓

版式设计:赫英华

封面设计:吴朝洪

出版发行:人民军医出版社

地址:北京市复兴路 22 号甲 3 号,邮编:100842,电话:(010)66882586、66882585、51927258

传真:(010)68222916,网址:www.pmmp.com.cn

印 刷:北京天宇星印刷厂

装 订:桃园装订厂

版 次:2004 年 1 月第 1 版,2004 年 1 月第 1 次印刷

开 本:787mm×1092mm 1/16

印 张:19.75

字 数:480 千字

印 数:0001~4500

定 价:33.00 元

(凡属质量问题请与本社联系,电话:(010)51927289、51927290)

内 容 提 要

本书为《SPSS10.0 统计软件应用教程》的姊妹篇，侧重介绍高级统计方法以及高级作图的 SPSS 过程。主要内容有：自定义统计表，多元方差分析，联合分析，对应分析，信度分析，交互统计图形，统计地图，确切检验方法，时间序列分析，缺失值估计，回归模型，对数线性模型。

本书以非统计专业人员为主要对象，围绕医学或生物学科科研数据，偏重实际应用，较少涉及统计理论背景知识，对操作步骤以及统计分析结果有较详尽的解释，通俗易懂，便于自学。

本书可作为医学或生物研究生和本科生的教材，亦可用作培训和继续教育的教材，同时也可供广大非统计专业的科研和教学人员自学用。对统计专业人员，本书亦有其重要参考价值。

责任编辑 张怡泓

前 言

本书是《SPSS10.0 统计软件应用教程》一书的姊妹篇。本书虽然冠名为“高级”，但仍继承了《SPSS10.0 统计软件应用教程》的风格，即以非统计专业人员为主要对象，侧重应用背景、操作过程和结果的详尽解释，所选资料力求为科研实际数据。

本书共有 12 章。第 1 章介绍的自定义统计表，在数据汇总和报告方面较之 **Frequencies** 和 **Descriptives** 过程具有很大的灵活性和很强的功能。第 2 章是关于多元方差分析的，用于多组均向量的比较。第 3 章的联合分析方法在市场调查领域的应用并不鲜见，但在卫生管理领域的应用还是比较新的方法。本章所列举的应用实例具有参考价值。第 4 章的对应分析是一种主要用于定性二维或多维列联表数据分析的类似于主成分分析的变量降维分析方法。第 5 章的信度分析除了介绍过程外，还介绍了信度分析的报告内容和参考方式，量表评价中可以借鉴。第 6 章介绍了交互统计图形，充分体现了 SPSS 强大的图形绘制功能，是撰写研究报告和制作电子媒体的有用工具。第 7 章是关于统计地图的，为一般统计软件所不具有的功能。第 8 章介绍了 SPSS 提供的确切检验方法，许多过程中都嵌入了这种方法，在此章中给以概括。第 9 章是关于时间序列分析的，包括指数平滑，自回归和 ARIMA 模型等内容。第 10 章是缺失值估计。除了直线回归和曲线拟合内容在上本书中已做介绍外，第 11 章较全面地介绍了各种回归分析方法，包括多分类 Logistic 回归分析，等级回归，概率单位回归，加权回归，非线性回归，二阶段最小二乘回归，最优标度回归等。第 12 章关于对数线性模型的内容，在结果的解释上着墨较重，因为这是对数线性模型应用的难点。

本书配有大量实例，实例尽可能贴近医学科研实际。每一实例均有相对应的数据文件，与该书配套的数据文件集可从第一军医大学网站下载，下载链接为：<http://www.fimmu.edu.cn/spss10advdata>。下载后，用 WinZip 进行解压缩即可。

本书根据编者应用 SPSS10.0 的体会以及教学和数据处理的经验撰写而成。但由于我们的水平有限，不足之处在所难免，恳请读者提出宝贵意见，便于日后完善。

陈平雁 黄浙明

2003 年 5 月，于广州，香港

目 录

1 自定义统计表 (Custom Tables)	1
1.1 基本统计表 (Basic Tables)	1
1.1.1 基本统计表一般定义	2
1.1.2 基本统计表的统计量输出	5
1.2 多项应答统计表 (Multiple Response Tables)	7
1.2.1 多项应答统计表的一般过程	7
1.2.2 示例	8
1.3 综合统计表 (General Tables)	9
1.3.1 综合统计表的一般过程	9
1.3.2 示例	10
1.4 频数统计表 (Tables of Frequencies)	11
2 多元方差分析 (Multivariate Analysis of Variance MANOVA)	13
2.1 多元方差分析的一般过程	13
2.2 实例	19
3 联合分析 (Conjoint Analysis)	27
3.1 联合分析的一般过程	27
3.2 实例	37
4 对应分析 (Correspondence Analysis)	43
4.1 对应分析的一般过程	43
4.2 实例	47
5 信度分析 (Reliability Analysis)	56
5.1 信度分析的一般过程	56
5.2 信度评价实例	59
6 交互统计图形 (Interactive)	69
6.1 交互图形的绘制	69
6.1.1 条图(Bar)	69
6.1.2 点图、线图、带图和垂线图 (Dot, Line, Ribbon Chart and Drop-Line)	78
6.1.3 面积图	80
6.1.4 圆图	83
6.1.5 箱丝图 (Boxplot, Box-and-Whisker Plot)	87
6.1.6 误差条图(Error Bar)	89

6.1.7 直方图(Histogram)	90
6.1.8 散点图.....	93
6.2 交互图形的编辑.....	97
6.2.1 选择所要编辑的对象.....	97
6.2.2 一般编辑 (Edit)	98
6.2.3 显示 (VIEW)	114
6.2.4 插入图形元素 (INSERT)	115
6.2.5 格式编辑 (Format)	116
7 统计地图 (Maps)	121
7.1 地理集文件的建立.....	121
7.1.1 地理集文件管理器 (Geoset Manager)	121
7.1.2 地理词典管理器 (Geodictionary Manager)	125
7.2 统计地图绘制.....	127
7.2.1 数据结合 (Data Binding)	128
7.2.2 数值域统计地图 (Rang of Values Maps)	129
7.2.3 刻度符号统计地图 (Graduated symbol maps)	132
7.2.4 点密度统计地图 (Dot Density Maps)	134
7.2.5 个值统计地图 (Individual Values Maps)	136
7.2.6 条图统计地图 (Bar Chart Maps)	138
7.2.7 圆图统计地图 (Pie Chart Maps)	140
7.2.8 多主题统计地图 (Multiple Theme Maps)	142
7.3 统计地图编辑.....	144
7.3.1 编辑主题.....	145
7.3.2 编辑图层.....	151
7.4 统计地图度量单位的定义.....	152
8 确切检验方法 (Exact Test)	153
8.1 确切法.....	153
8.2 Monte Carlo 法.....	154
8.3 应用确切检验的时机.....	156
8.4 如何获得确切检验统计量.....	158
9 时间序列分析 (Time Series)	161
9.1 概述.....	161
9.1.1 时间序列分析的意义和策略.....	161
9.1.2 SPSS 时间序列分析的基本方法.....	162
9.1.3 时间序列数据的一般处理.....	162
9.2 指数平滑法 (Exponential Smoothing)	164
9.3 自回归分析 (Autoregression)	172
9.4 ARIMA 分析(ARIMA)	179

	3
9.4.1 非季节性 ARIMA 分析.....	180
9.4.2 季节性 ARIMA 分析.....	190
10 缺失值估计(Replace Missing Values)	199
10.1 缺失值替代过程 (Replacing Missing Values...)	199
10.2 缺失值分析 (Missing Value Analysis...)	200
11 回归模型(Regression)	213
11.1 多分类 Logistic 回归分析 (Multinomial Logistic Regression)	213
11.2 等级回归 (Ordinal Regression)	221
11.3 概率单位回归 (Probit...)	232
11.4 加权回归(Weighted Least-squares Regression).....	238
11.5 非线性回归 (Nonlinear Regression)	241
11.6 二阶段最小二乘回归 (Two stage least-squares regression)	248
11.7 最优标度回归 (Optimal scaling regression)	252
12 对数线性模型 (Loglinear)	259
12.1 对数线性模型的选择 (Model Selection...)	259
12.2 广义对数线性模型分析 (General...)	265
12.2.1 完全表 (Complete Table) 的 Genlog.....	265
12.2.2 不完全表 (Incomplete Table) 的 Genlog.....	274
12.2.3 生存参数模型 (Survival Parametric Model) 的 Genlog.....	278
12.2.4 列联表标准化 (Table Standardization) 的 Genlog.....	281
12.2.5 Poisson 对数线性回归分析 (Poisson Loglinear Regression)	284
12.3 Logit 对数线性模型分析 (Logit...)	289
12.3.1 一个二分类因变量的 Logit 分析.....	289
12.3.2 两个二分类因变量的 Logit 分析.....	292
12.3.3 一个多分类因变量的 Logit 分析.....	296
12.3.4 一个按等级分类因变量的连续比 Logit 分析.....	301
参考文献	307

1 自定义统计表 (Custom Tables)

在对分类变量较多的数据进行分类汇总，或在处理由多项应答组成的问卷时，SPSS 的自定义统计表 (Custom Tables) 体现了很大的灵活性和方便性。虽然 **Crosstabs** 过程也能进行分类变量的列表，但 Custom Tables 具有更强的列表功能、更大的灵活性和更多的输出格式选择。

自定义统计表菜单下包括以下四种过程，见图 1-1。

- Basic Tables: 基本统计表。
- General Tables: 综合统计表。
- Multiple Response Tables: 多项应答统计表。
- Tables of Frequencies: 频数统计表。

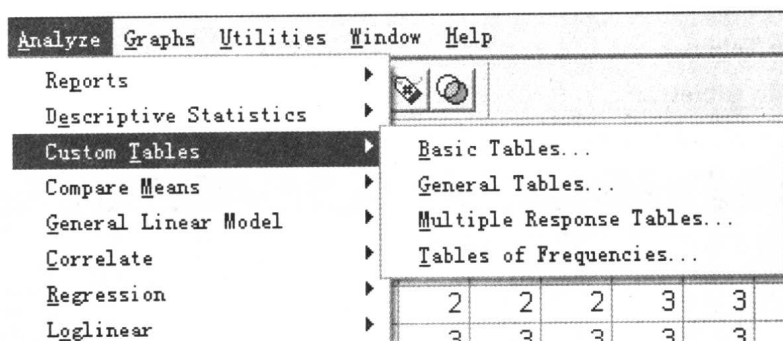


图 1-1 自定义统计表下拉菜单

例 1-1: 在一项老年护理现状的调查研究中，对 255 名护士进行了问卷调查，调查的部分数据见数据文件 “nurse_survey.sav”。文件中，除了年龄、职称、学历和科室等变量外，变量 “c1” ~ “c10” 是有关护士对医院某些方面满意程度的应答，为单选应答，即 10 个方面中，每个方面只能选五级满意程度（很不满意，不满意，一般，比较满意，很满意）中的一级。变量 “prob1” ~ “prob6” 则是护士对老年护理最紧迫问题的应答，为多选应答，即最紧迫的 6 个问题（“prob1” ~ “prob6”）中可以选 1 个，也可以选多个。本章将围绕本数据介绍创建统计表的四种过程。

1.1 基本统计表 (Basic Tables)

基本统计表过程可用以创建三维统计表。所谓三维统计表，是指由行变量 (row/down variable)、列变量 (column/across variable) 和分层变量 (layer/separate variable) 构成的统计表。由于统计表是平面的，实际上每个表只是行变量与列变量组合的统计表，而分层变量则

体现在具有相同行列变量的不同统计表上,如图 1-2,“年龄组”为行变量;“学历”为列变量;“科室”为分层变量。

科室 内科		学历			
		中专	大专	本科	不适用
		Count	Count	Count	Count
年龄组	<=25yr	41	4		37
	26—	10	4	1	10
	36—		2		5

科室 外科		学历			
		中专	大专	本科	不适用
		Count	Count	Count	Count
年龄组	<=25yr	40	3	2	29
	26—	6	5	3	14
	36—	1	1	1	4

图 1-2 三维统计表示例

1.1.1 基本统计表一般定义

从菜单选择

Analyze

Custom Tables

Basic Tables...

弹出 Basic Tables (基本统计表) 主对话框 (图 1-3)。

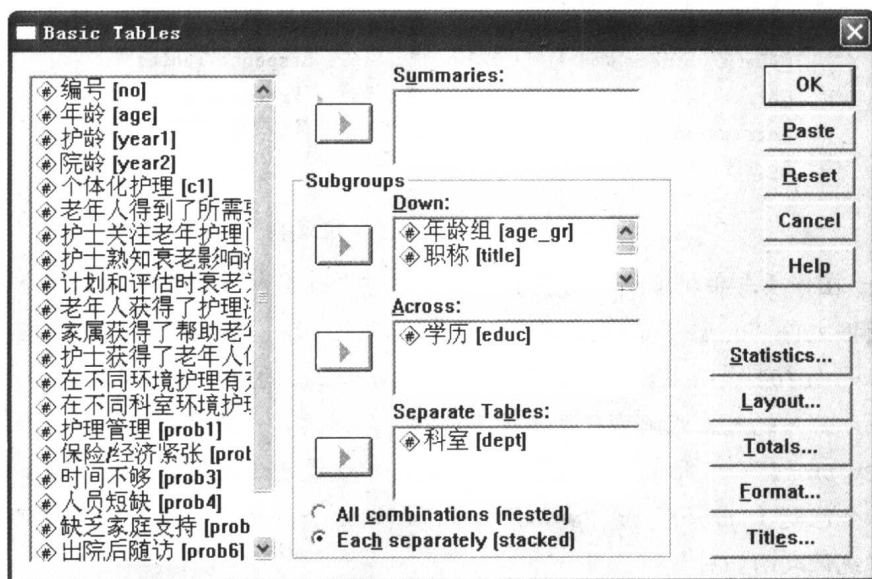


图 1-3 基本统计表主对话框

◆ **Summaries:** 汇总变量,一般为定量变量。选入汇总变量后,击 **Statistics...**按钮,可在统计量对话框中选择汇总统计量,如均数、标准差、标准误等。

◇ **Subgroups** 定义分组变量。

◆ **Down:** 行变量,一般为分类变量。

◆ **Across:** 列变量，一般为分类变量。

◆ **Separate Tables:** 分层变量，一般为分类变量。

○ **All combinations (nested):** 嵌套排列。当行变量框、列变量框或分层变量框内同时选入两个或两个以上变量时，该选项被激活。如果图 1-3 中选该项，意味着行变量框内选入“age_gr”和“title”两个变量，而且变量“title”被嵌套在变量“age_gr”内，结果如图 1-4 所示。

○ **Each separately (stacked):** 平行排列。当行变量框、列变量框或分层变量框内同时选入两个或两个以上变量时，该选项被激活。如果图 1-3 中选该项，意味着行变量框内选入“age_gr”和“title”两个变量，而且变量“age_gr”与变量“title”呈并列关系，结果如图 1-5 所示。

科室 内科				学历			
				中专	大专	本科	不适用
年龄组	<=25yr	职称	护士	41	4		37
			护师				
	26—	职称	护士	10	1		4
			护师		2	1	2
36—	职称	主管护师	主管护师		1		4
			(副)主任护师		2		5

图 1-4 图 1-3 中选中行变量嵌套排列

科室 内科				学历			
				中专	大专	本科	不适用
年龄组	<=25yr			41	4		37
	26—			10	4	1	10
	36—				2		5
职称	护士			51	5		41
	护师				2	1	2
	主管护师				3		9
	(副)主任护师						

图 1-5 图 1-3 中选中行变量平行排列

★ **Statistics...** 统计量

击 **Statistics...**按钮，弹出 Basic Tables: Statistics (基本统计表的统计量选项)对话框(图 1-6)。Statistics 框为统计量备选列表；Cell Statistics 框为选入统计量列表；此外还有输出格式选项，如百分位点、字节宽度、小数位、标签、排列顺序等。

★ **Layout...** 输出方式

击 **Layout...**按钮，弹出 Basic Tables: Layout (基本统计表的输出方式)对话框(图 1-7)。

- ◆ Summary Variable Labels: 汇总变量标签。
 - ☐ Down the left side: 标签在横标目位置。
 - ☐ Across the top: 标签在纵标目位置。
 - ☐ In separate tables: 标签在分层变量位置。
- ◆ Statistics Labels: 统计量标签。
- ◆ Groups in Summary Variable Dimension:
 - ☐ All summary variables under each group: 汇总变量在分组变量下级。
 - ☐ All groups under each summary variable: 汇总变量为最上级。
- ☐ Label groups with value labels only: 用变量值标签给分组变量标记。

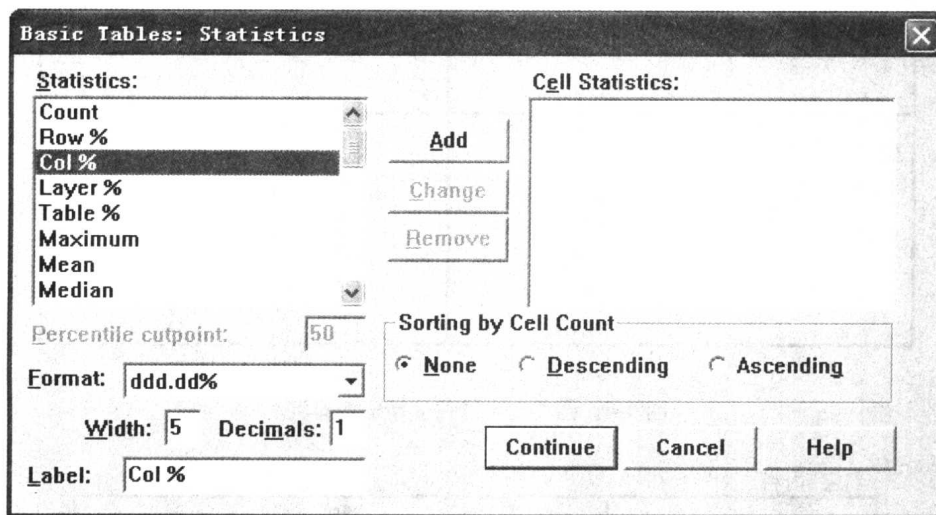


图 1-6 基本统计表的统计量选项对话框

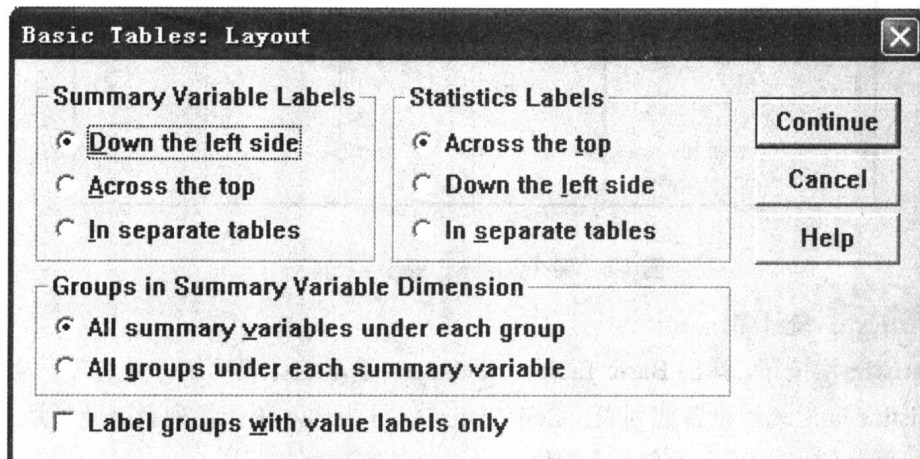


图 1-7 基本统计表的输出方式对话框

★ Totals... 合计项输出

击 **Totals...**按钮, 弹出 Basic Tables: Totals (基本统计表的合计项输出) 对话框 (图 1-8)。

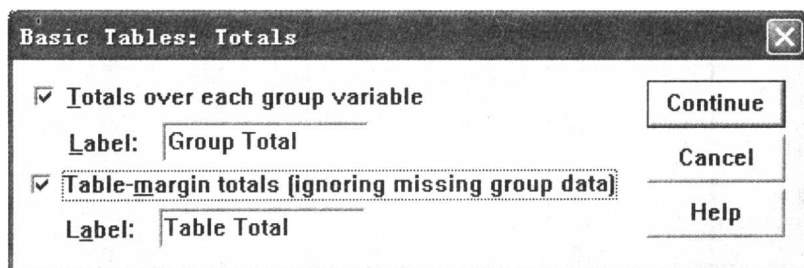


图 1-8 基本统计表的合计项输出对话框

☐ Totals over each group variable: 分组合计。

☐ Table-margin totals (ignoring missing group data): 总合计 (缺失分组数据不列出)。

★ Format... 频数为 0 的输出格式

击 **Format...**按钮, 弹出 Basic Tables: Format (基本统计表的频数 0 输出格式) 对话框 (图 1-9)。频数为 0 时有两种输出方式, 即空白 (Blank) 和数值 0 (Zero), 前者为系统默认。另外还有缺失统计量的输出方式, 系统默认 “.”。

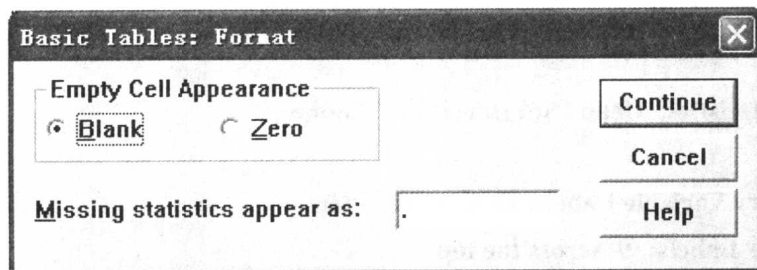


图 1-9 基本统计表的频数 0 输出格式对话框

★ Titles... 标题 (略)

1.1.2 基本统计表的统计量输出

1. 输出百分数与合计项

在图 1-2 选项的基础上, 在 **Statistics** 选项对话框中将统计量 “Count” 和 “Row%” 选入; 在 **Layout** 选项对话框中, Statistics Labels 下选 ☒ Across the top, 以及选 ☒ All summary variables under each group; 在 **Totals** 选项对话框中选 ☒ Totals over each group variable。输出结果见图 1-10。

科室 内科										
	学历								Group Total	
	中专		大专		本科		不适用		Count	Row %
	Count	Row %	Count	Row %	Count	Row %	Count	Row %		
年龄 ≤25yr	41	50.0%	4	4.9%			37	45.1%	82	100.0%
组 26—	10	40.0%	4	16.0%	1	4.0%	10	40.0%	25	100.0%
36—			2	28.6%			5	71.4%	7	100.0%
Group Total	51	44.7%	10	8.8%	1	.9%	52	45.6%	114	100.0%
职称 护士	51	52.6%	5	5.2%			41	42.3%	97	100.0%
护师			2	40.0%	1	20.0%	2	40.0%	5	100.0%
主管护师			3	25.0%			9	75.0%	12	100.0%
(副)主任护师										
Group Total	51	44.7%	10	8.8%	1	.9%	52	45.6%	114	100.0%

图 1-10 依行变量百分数及分组合计的输出结果

2. 汇总变量的统计量输出

操作过程:

Analyze**Custom Tables****Basic Tables**► **Summaries:** *age*► **Down:** *title*► **Across:** *educ***Statistics**► **Cell Statistics:** Mean / Std Deviation / Count**Layout****Summary Variable Labels:** ☉ Across the top**Statistics Labels:** ☉ Across the top**Groups in Summary Variable Dimension:** ☉ All summary variables under each group**Totals**☒ **Totals over each group variable****Label:** Group total

结果见图 1-11。

图 1-11 汇总变量统计量(均数、标准差、样本量)的输出

1.2 多项应答统计表 (Multiple Response Tables)

调查表或量表常会设计多选应答项, 如例 1-1 中, 某一条目陈述为“老年护理目前最紧迫的问题是什么?” 该条目下有 6 个备选项 (“*prob1*” ~ “*prob6*”), 应答时可以选 1 项, 也可以选多项。对这种多选应答条目, 用 1 个变量是无法表示应答结果的, 而需要用相当于备选项项目数的变量来表示, 如本例需要 6 个变量表示。多项应答统计表的好处在于可以将每个多选应答项下的多个变量看作 1 个变量来处理。

1.2.1 多项应答统计表的一般过程

从菜单选择

Analyze

Custom Tables

Multiple Response Tables...

弹出 Multiple Response Tables (多项应答统计表) 主对话框 (图 1-12)。

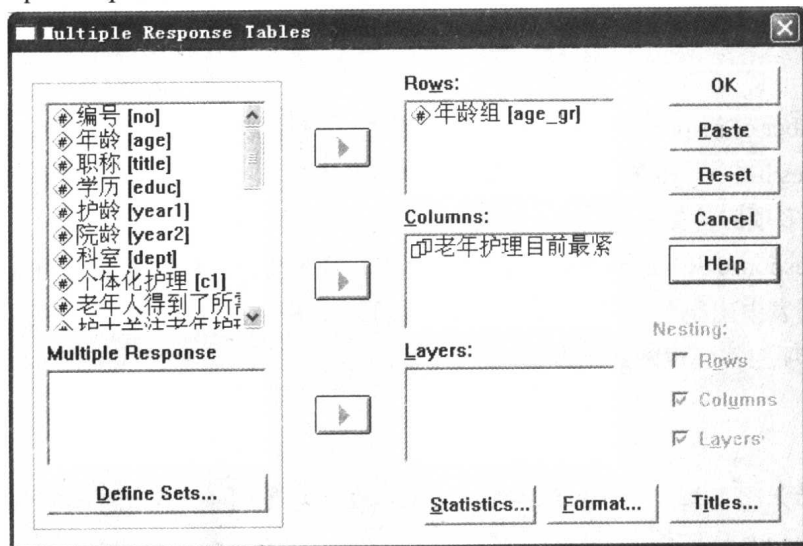


图 1-12 多项应答统计表主对话框

★ Define Sets... 定义多项应答集

击按钮, 弹出 Define Multiple Response Sets (定义多项应答集) 对话框 (图 1-13)。

◇ Variables Are Coded As: 分类变量的选择, 有二分类和多分类两种选择。

○ Dichotomies Counted value: 二分类变量阳性事件赋值, 本例选此项, 如“1”为阳性事件赋值, 即表示该问题被认为是最紧迫的问题。

○ Categories: 多分类变量。

◆ Name: 多项应答集名称, 当多项应答集被定义后, 该名称前加美元符号 “\$”。

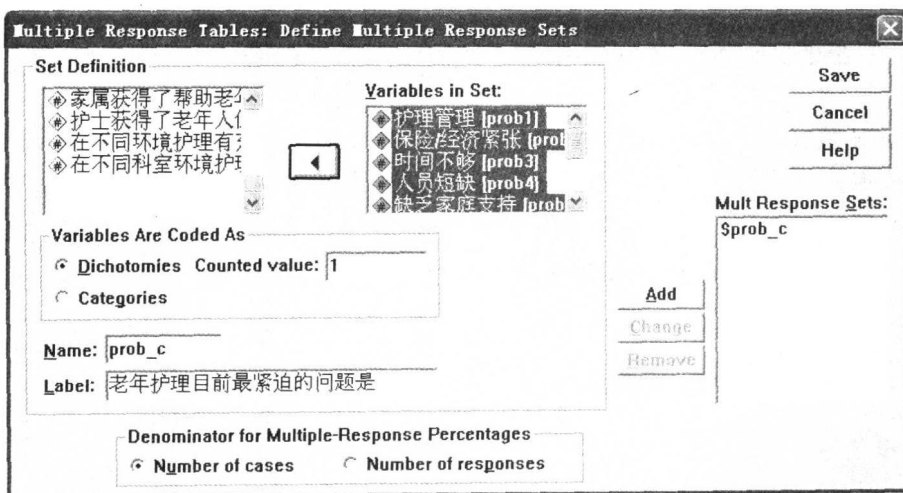


图 1-13 定义多项应答集对话框

◆ Label: 多项应答集标签。

◆ Denominator for Multiple-Response Percentages: 定义计算百分数的分母。

○ Number of cases: 以观察单位数为分母计算单元格的百分数, 分母可以小于阳性应答总数。

○ Number of responses: 以阳性应答总数为分母计算单元格的百分数。

◆ Variables in Set: 多项应答集中所含的变量, 本例 6 个变量 “prob1” ~ “prob6” 被选入, 构成一个多项应答集。

◆ Mult Response Sets: 多项应答集列表。当上述选项完成后, 本框边的 **Add** 键被激活, 击 **Add** 键, 列表中出现新定义的多项应答集名称, 击 **Save** 键后, 定义完毕。当保存该数据文件后, 此多项应答集也随之保存。

1.2.2 示例

图 1-14 是在图 1-12 和图 1-13 选项基础上, 在统计量选项 (Statistics) 对话框中选了 ☒ Column Percentages 后的结果。

			\$PROB_C					
			护理管理	保险/经济紧张	时间不够	人员短缺	缺乏家庭支持	出院后随访
年龄组	<=25yr	Count		100	75	113	95	81
		Column %		68.0	71.4	70.2	72.5	73.6
	26—	Count	12	34	23	35	28	21
		Column %	44.4	23.1	21.9	21.7	21.4	19.1
	36—	Count	15	13	7	13	8	8
		Column %	55.6	8.8	6.7	8.1	6.1	7.3

图 1-14 多项应答频数表

1.3 综合统计表 (General Tables)

虽然上述两种统计表过程在创建绝大多数统计表时有很强的功能和很大的灵活性,但是,当有些统计表的横标目或纵标目既有平行排列变量,又有嵌套变量时;当不同变量用不同的统计量时;当合计项的要求更为复杂时,需要用综合统计表过程满足这些要求。

1.3.1 综合统计表的一般过程

从菜单选择

Analyze

Custom Tables

General Tables...

弹出 **General Tables** (综合统计表) 主对话框 (图 1-15)。

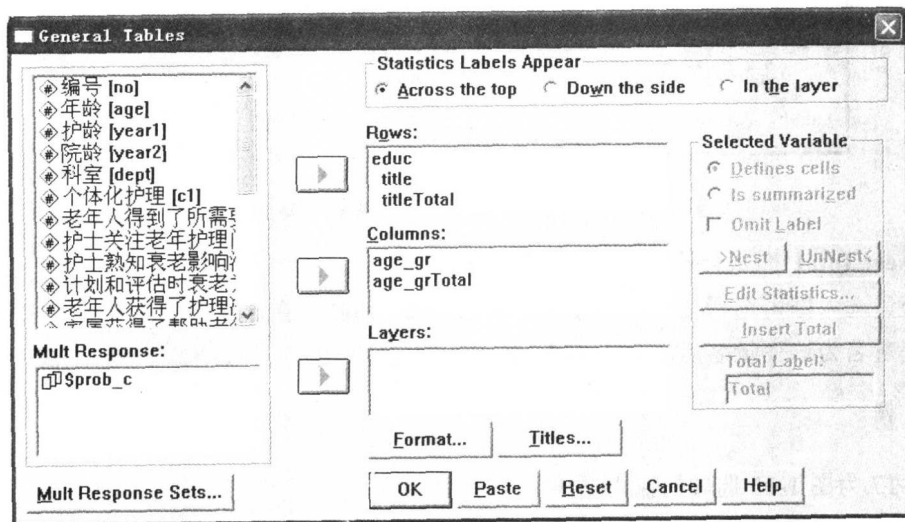


图 1-15 综合统计表主对话框

◇ **Statistics Labels Appear** 统计量标签的位置。

- ⊙ Across the top: 统计量标签在纵标目。
- Down the side: 统计量标签在横标目。
- In the layer: 统计量标签在分层变量位置。

◇ **Selected Variable** 定义变量类型。

- ⊙ Defines cells: 定义为分类变量, 系统默认。
- Is summarized: 定义为汇总变量, 一般为定量变量, 选此项后, 原变量名后自动加“(S)”, 如变量“educ”被定义为汇总变量后, 变量名显示为“educ(S)”。
- Omit Label: 不显示变量标签。

★ **Mult Response Sets...** 定义多项应答集 (见 1.2 节)