# 第2章

퉈

# 数据的建立与管理

# 

- 掌握 SPSS 数据文件变量的设置。
- 掌握 SPSS 数据文件的数据录入。
- 掌握数据文件的合并及拆分的 SPSS 操作。
- 掌握数据排序、选择个案及计算变量的 SPSS 操作。

如何建立数据文件是使用统计软件的第一步,因为数据是所有统计研究的基础,没有数据,统计分析也就无从谈起,因此,介绍 SPSS 数据的建立是本章的首要任务。

在实际的统计研究中,我们所建立的原始数据往往不能直接应用于最终的统计分析。 这不仅是因为数据库可能有工作人员录入错误或原始问卷记录错误等情况,使得数据库中 包含不正确的数据;还因为针对同一个研究目的,往往要从不同的侧面对数据进行研究, 采用多种统计分析方法进行分析,而不同的统计方法对数据文件结构的要求也不尽相同, 这就需要对数据文件的结构进行重新调整或转换,以便适合于相应的统计方法使用。以上 这些工作被统称为数据管理,数据管理是统计分析工作中非常重要的一个环节,直接关系 到数据分析的结果,是统计分析工作中不可缺少的一个关键步骤。在 SPSS 中,数据文件的 管理功能基本上都集中在"数据"和"转换"两个菜单中,前者的功能主要是实现文件级 别的数据管理,如个案排序、选择个案、文件合并和拆分等;而后者主要实现数据变量级 别的数据管理,如计算新变量、变量取值重编码等,主要与变量数值的转换有关。本章将 主要介绍这两个菜单的相应功能。

# 2.1 数据的建立

在 SPSS 中建立数据文件大致有两种情况:一种是将原始数据直接录入 SPSS;另一种 是利用 SPSS 读取其他数据格式的资料。数据录入就是把每个个案(公司、被调查者等)的每 个指标(变量)录入到软件中。在录入数据时,大致可归纳为三个步骤:定义变量名,即给每 个指标起个名字;指定每个变量的各种属性,即对每个指标的一些统计特性作出指定;录 入数据,即把每个个案的各指标值录入为电子格式。因此,我们有必要先了解变量的各种 属性。

# 2.1.1 变量的属性

任何一个变量都有相应的变量名与之对应,但为了进一步满足统计分析的需要,除了 变量名外,往往还要对每一个变量进一步定义许多附加的变量属性,如变量类型、变量宽 度和小数位等。如图 2-1 所示,在变量视图中 SPSS 为每个变量指定了 11 种变量属性。



图 2-1 变量视图

### 第2章 数据的建立与管理

### 1. 名称

该单元格主要的目的是定义变量名称, SPSS 中变量名定义应符合以下要求。

(1) 在一个数据文件中变量名必须是唯一的,不能重名。

(2) 变量名不区分大小写;变量名长度不能超过 64 个字符(32 个汉字)。

(3) 首字符必须是字母、汉字或特殊符号@,但不能是空格或数字;其后的字符可为 字母、数字、中文及特殊符号".""\$""@",但不能为"?""!""\*"等字符。变 量名的首尾都不能是".""。"或"-",以免引起误会。

(4) 一些逻辑词语不能作为变量名, 如 all、and、or、by、to、with、not 等。

如果用户不指定变量名, SPSS 软件会以"VAR"开头来命名变量,后面跟五位数字,如 VAR00001、VAR00019 等。

### 2. 变量类型的设置

SPSS 中变量有三种基本类型:数值型、字符串型和日期型。根据不同的显示方式,数 值型又被细分成了六种,所以 SPSS 中的变量类型共有八种。在变量视图中选择【类型】单 元格时,右侧会出现 按钮,单击 按钮会打开【变量类型】对话框,如图 2-2 所示。左 侧为具体的变量储存类型,右侧用于进一步定义变量宽度和小数位。

· 变量类型 · · · · · · · · · · · · · · · · · · ·
⑥ 数值(N)
····································
◎ 点(D) 小数位(P): 2
◎科学计数法(S)
○ 日期(A) ○ 关二(1)
◎ 2,000/ ◎ 设定货币(U)
◎ 字符串(民)
确定取消常期

图 2-2 【变量类型】对话框

1) 数值型

在三种基本变量类型中,数值型是 SPSS 最常用的变量类型。数值型的数据是由 0~9 的阿拉伯数字和其他特殊符号,如美元符号、逗号或圆点组成。数值型数据根据内容和显 示方式的不同,可以分为标准数值型、每三位用逗号分隔的数值型、每三位用圆点分隔的 圆点数值型、科学计数型、显示带美元符号的美元数值型和自定义货币型等六种不同的表 示方法。其中,最为常用的只有标准数值型,作为初学者,其他几种使用频率较低,如有 兴趣可以自行查阅软件中的帮助信息即可了解详细内容,在此不过多赘述。

2) 字符串型

字符串也是 SPSS 中较为常用的数据类型,变量值是一串字符,字符串变量中的大小写 是被区分的,但字符串变量不能参与算数运算,只能在频率与交叉表分析中显示。

数值型变量可以直接转换为字符型变量,不过字符串型变量转换为数值型时,数字数

据不会丢失,但非数字数据则会丢失。例如,"部门"变量录入数据时的数据为"后勤" 等字符串,若将其变量类型由字符串改为数值型,则"后勤"数据会消失。但若"部门" 的数据为数字,例如,用"1"代表"后勤",尽管此时"部门"的变量类型为字符串,但 将其改为数值型时数据还会保留。

3) 日期型

它可以用来表示日期或时间。日期型数据的显示格式有很多,SPSS 在对话框右侧会以 列表的方式列出各种显示格式以供用户选择。

### 3. 变量宽度的设置

"宽度"是指数据视图中数据所占的列宽,一般使用系统默认的设置,默认宽度为 8 个字符宽度,用户也可以根据需要调整。在电脑中的编辑中每个汉字占 2 个字符,每个字 母和数字占 1 个字符。

在此需要注意的是,用户设定好字符型变量宽度后,所录入的数据长度将自动被限定 在宽度之内。字符型数据的长度不能超过变量设定的宽度。如果字符串的长度超过变量宽 度,超过部分将被系统截掉。例如,某单元格的数据为 abcdef,但当设置该变量宽度为 4 时,后面的 ef 将自动被截掉。

### 4. 小数位的设置

"小数"用于设置变量数值的小数位数,数值型变量默认为 2 个小数位,字符型变量 SPSS 自动设置为 0。小数点的设置只影响显示的位数,而不影响实际数值。例如 0.3456, 在小数位为 2 个时将显示为 0.35(自动四舍五入),但其数值大小依然为 0.3456 而非改为 0.35。 当变量小数位为 2 个时,输出结果的均值默认为 4 个小数位,标准差默认为 5 个小数位; 当变量小数位为 0 时,则输出结果的均值为 2 个小数位,标准差为 3 个小数位。

录入数据时,系统会默认将数值型变量小数位设置为2,如果录入的数据多为整数,为 了数据视图的视觉效果更简洁,建议手动将整数变量设置小数位为0。

### 5. 变量标签与值标签的设置

1) 变量标签的设置

变量标签是对变量名含义进行注释说明的标记,目的是使人更清楚明确地了解该变量的含义。有时一个变量的全称太长,不适合直接作为变量名,此时就用简略词语给变量起名,然后在变量标签中附注完整的名称或具体含义,如图 2-3 中的 "q1"变量所示。

文件(E)	编辑(	<u>E</u> ) 视图	⊘	数据( <u>D</u> )	转换 <mark>(</mark> ]	) 分析(A	) 直销( <u>M</u> )	图形( <u>G</u> )	实用程序 <mark>(U</mark> )	窗口( <u>W</u> ) 帮助	
					2		<b>*</b>	<b>81</b>			
		名称		类型		宽度	小数		标签		值
1	q	1		数值(N)	8	}	2	您认为公	司目前采用的	是什么薪酬制度	无
2	q	2		数值(N)	8	}	2				无
3											

图 2-3 变量标签的设置

当设置变量标签时,在各种统计分析操作的变量列表以及输出结果中,该变量就会以

变量标签出现而不是以原变量名出现,如果变量标签比较长,在命令窗口常常只能见到标 签,见不到变量名,如图 2-4 所示,这给使用者带来了不便。如果不想让变量标签代替原变 量名出现,则可以选择【编辑】→【选项】命令,在【选项】命令中的【常规】选项卡的 【变量列表】选项组中选择【显示名称】,如图 2-5 所示。这时如果我们再打开命令分析对 话框,变量的标签就不再显示了,如图 2-6 所示。这样设置后我们做统计分析时在选择变量 时就十分清楚明了了。



图 2-4 显示变量标签的【频率】对话框

选项	×
常规 查看器 数据 货币 输出标签 图表 枢轴表 文件位置	脚本 多重归因 语法编辑器
交里列表 ◎显示标签(L) ◎显示名称(L)	輸出 一表格中较小的数值没有科学记数法(Q)

图 2-5 变量名显示的设置

111 频率(F)		×
<mark>∳ q1</mark> ∳ q2	变 <u>里()</u> : ❤	统计里( <u>S</u> ) 图表( <u>C</u> ) 格式( <u>F</u> ) <u>B</u> ootstrap(B)
✓ 显示频率表格(D) 确定	お助 の 重 置 の 取 消	帮助

图 2-6 未显示变量标签的【频率】对话框

### 2) 值标签的设置

由于 SPSS 只能对数值型数据进行算术统计分析,因此在 SPSS 中录入的内容以数值为

主。但数字本身是没有具体意义的,只有在特定的研究项目中才有特定的意义,因而我们 需要对变量数据的各种取值的含义进行注释说明,即设置值标签。例如,性别的数据中有 "1"和"2"两种取值,具体它们分别代表哪种性别,则需要在值标签中说明,如图 2-7 所示。

文件 <mark>(E</mark> )	编辑	(E)	视图(⊻)	数据( <u>D</u> )	转换(I	) 分析(	y) 直销	( <u>M</u> )	图形( <u>G</u> )	实用程序	( <u>U)</u> 7	笥口( <u>₩</u> )
		e		]	1			3	- M	×,	4	
		-	名称	类型	민	宽度	小婆	釹	标	Ϋ́.		值
1		性别		数值 <mark>(N)</mark>	8	3	2				无	
2			1 值标签	E(V)							23	S
3											-	
4			「值标签	(V)								
5			值 <mark>(U</mark> ):	2						拼罩	≣(S)	
6			标签(L	.): <del>'</del>								
7						1						
8			9	Stn(A)	1.00 = " <del>⊊</del>	3						
9				MUL								
10				亘改(C)								
11			88	ll除(R)								
12												
13							The sale	Tent				
14						・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・	取消	帮助				
45		C										

图 2-7 值标签的设置

当变量数据的含义非常直接明确时,可以不设置值标签,如年级。除此之外,读者还 可以只对部分取值设置值标签,而不一定对所有的值设置。需要注意的是,值标签一般是 针对离散变量(定类变量和定序变量)设置,连续变量(定距变量和定比变量)不需要设置,因 为连续变量的数值可以反映数值大小,有具体的意义。离散变量、连续变量等变量类型的 相关知识点我们将在"3.1 变量类型"中做详细介绍。

3) 变量属性及值标签的批量设置

如果我们需要将数据中很多变量的属性和值标签设置为相同,可以采用以下两种方法处理。

(1) 复制粘贴数据整体属性法。可以通过选择【数据】→【复制数据属性】命令完成, 过程并不复杂,感兴趣的读者可以自己尝试操作。

(2) 复制粘贴数据单个属性法。直接单击要复制的变量的某个属性或值标签单元格, 选择复制选中目标变量对应的属性单元格,然后粘贴到新变量中即可,这和一般的复制粘 贴过程是一样的,读者可以自己完成,在此不再赘述。

6. 缺失值的设置

缺失值是指某个样本缺少特定变量的数据信息,它将不被纳入各种统计分析中。SPSS 中的缺失值有系统缺失值和用户缺失值两大类。

1) 系统缺失值

当变量中某个样本没有提供信息或者提供的是非法格式的信息时,系统自动将其设置为缺失值。在 SPSS 中,对于数据型变量数据,系统缺失值默认用"."表示,而字符串型 变量就是空字符串。

### 第2章 数据的建立与管理

2) 用户缺失值

用户缺失值是指用户根据特定目的设置的、自己能够识别的数值。例如,不符合题目 要求的答案,不适合某项统计分析条件的数值、录入错误的数据等。一般用特殊的数字表 示,如"99""98"等。设置用户缺失值可以保留最原始信息,同时又避免错误数据被纳 入统计分析而造成结果偏误。在变量视图中,单击【缺失】下面的单元格出现 按钮,单 击 步按钮弹出【缺失值】对话框,有三种方式可供定义用户缺失值,如图 2-8 所示。

·····································
<ul> <li>没有缺失值(N)</li> </ul>
○ 离散缺失值( <u>D</u> )
◎范围加上一个可选离散缺失值(民)
低(L): 高(H):
离散值(S):
确定 取消 帮助

图 2-8 【缺失值】对话框

(1) 没有缺失值:默认为没有用户缺失值,只有系统缺失值。

(2) 离散缺失值:缺失值是 1~3 个不连续的数值。

(3) 范围加上一个可选离散缺失值:缺失值是一个区间范围,且还可以设置某个零散的缺失值。

需要注意的是,如果数据中有用户缺失值,那就一定要在变量属性中设定,要不就将 所有用户缺失值都设定为系统缺失值,即删除为空。

### 7. 变量列宽、对齐、度量标准的设置

(1) 列宽:数据区域中变量所在列的宽度。设置时宽窄要适度,以变量名不换行为佳。

(2) 对齐:字符型变量自动左对齐,数值型变量自动右对齐。建议统一用居中对齐。

(3) 度量标准:字符型、分类变量可以设置为"名义",等级顺序变量设置为"序号", 连续变量设置为"度量",也可以采用系统的默认设置。

### 8. 角色的设置

该属性是源自于数据挖掘方法体系中要求某些对话框支持用于预先选择分析变量的预 定义角色。当打开其中一个对话框时,满足角色要求的变量将自动显示在目标列表中。由 于此类对话框在现有的 SPSS 中很少,因此一般用户可以直接忽略这一属性。

# 2.1.2 数据的直接录入

将非电子化的原始问卷资料录入到 SPSS 软件中,需要注意数据录入的以下基本原则: ①每个个案要设置标记 ID,以便核对数据信息及作为其他数据处理的关键变量;②变量信 息要简单独立,一个属性就是一个变量,避免一个变量多重属性(如"农村男生""女研究

生");③统计指标(变量)在列,样本在行,一个变量一列,一个样本一行;④录入的数据 为原始数据而不是汇总数据;⑤数据应先对变量进行分类编码(数字化)之后再录入。问卷中 不同的问题类型录入的方式有所不同,接下来以案例 2-1 "员工薪酬调查问卷"为例,介绍 问卷中常见的几种题型的录入方法。



### 员工薪酬调查问卷

填写说明:

1. 本问卷的目的在于了解填写者对公司员工薪酬问题的建议和看法。

2. 所有填写均为匿名填写,我们承诺对填写者的个人信息保密。

(注:薪酬概念涵盖员工在公司工作所获的各项收益,包括工资、奖金、津贴、保险以 及各项福利等)请在您选择的答案序号上画"√"。

### 一、个人信息

S1. 性别:

51.	17711.	
	A. 男	B. 女
\$2	在脸.	

32.	平时:			
S3.	您所在的部门:			
	A. 业务部门	B. 管理职能部门		
S4.	您的学历:			
	A. 大专以下	B. 大专	C. 本科及以上	
S5.	您在公司的工作年	尾:		
	A.1 年以下	B. 1~2 年	C. 2~3年	D.3年以上
S6.	您在本职位的累计	工作年限:		
	A.2年以下	B. 2~5年	C. 6~10年	D. 10 年以上
二、	薪酬调查			
q1.	您认为公司目前采	用的是什么薪酬制	度?	
	A. 岗位(职务)工资	欠	B. 业绩工资	
	C. 技能工资		D. 以岗位工资为	基础的绩效工资

E. 其他(请说明)
q2. 您认为现行的薪酬制度
A. 非常合理
B. 合理
C. 一般
D. 不合理
E. 非常不合理
q3. 从您进入公司以来,薪酬制度是否做过调整?
A. 是
B. 否

- q4. (如果上题回答"是")最近一次薪酬制度调整是哪年? \_\_\_\_\_年
- q5. 您认为调整后的薪酬制度与以前的相比:

第2章 数据的建立与管理

A. 改进了很多 B. 有-	一些改进 C. 一般	D. 不合理	E. 非常不合理
q6. 您认为您的薪酬所得在同	行业中		
A. 高很多 B. 偏高	哥 C. 一般	D. 偏低	E. 低很多
q7. 您认为您的薪酬所得在本:	地区属于		
A. 很高的 B. 偏离	高 C. 一般	D. 偏低	E. 很低的
q8. 您认为您的薪酬所得在本	职业属于		
A. 很高的 B. 偏语	哥 C. 一般	D. 偏低	E. 很低的
q9. 您认为您的薪酬所得同您	的工作难度和责任是否	对等?	
A. 很对等 B. 较对 B. 较对	寸等 C. 一般	D. 较不对等	E. 很不对等
q10. 您认为您的薪酬所得同位	计出的努力相比是否对等	<b>等</b> ?	
A. 很对等 B. 较对 B. 较对	寸等 C. 一般	D. 较不对等	E. 很不对等
q11. 在现行的薪酬制度下,公	公司多长时间给您调整-	一次工资?	
A. 半年以下 B.1年	C.2年	D.3年	E. 没有明确
q12. 您认为薪酬变动的依据是	是什么?(最多选三项)		
A. 岗位或职务变动 B.	个人业绩大小 C	. 定期升降	
D. 公司利润大小 E.	工作年限 F.	其他	
q13. 目前薪酬调整的趋势是			
A. 只升不降 B. 升降	备结合 C. 只降不知	升 D. 没有变	乏化
q14. 您认为公司目前的工资等	级设置与工资级差(每个	工资等级之间的	差距)是否合理?
1			· · · · · · · · · · · · · · · · · · ·
A. 工资等级与工资级差让	<b>设置比较合理</b>		
<ul><li>A. 工资等级与工资级差</li><li>B. 工资等级设置太多(少)</li></ul>	设置比较合理 )		
<ul> <li>A. 工资等级与工资级差</li> <li>B. 工资等级设置太多(少)</li> <li>C. 工资级差太大(小)</li> </ul>	设置比较合理 )		
<ul> <li>A. 工资等级与工资级差试</li> <li>B. 工资等级设置太多(少)</li> <li>C. 工资级差太大(小)</li> <li>D. 工资等级与工资级差试</li> </ul>	设置比较合理 ) 设置都不合适		
<ul> <li>A. 工资等级与工资级差试</li> <li>B. 工资等级设置太多(少)</li> <li>C. 工资级差太大(小)</li> <li>D. 工资等级与工资级差试</li> <li>q15. 目前薪酬中奖金与业绩的</li> </ul>	设置比较合理 ) 设置都不合适 的挂钩程度为		
<ul> <li>A. 工资等级与工资级差试</li> <li>B. 工资等级设置太多(少)</li> <li>C. 工资级差太大(小)</li> <li>D. 工资等级与工资级差试</li> <li>q15. 目前薪酬中奖金与业绩的</li> <li>A. 20%及以下 B. 30%</li> </ul>	设置比较合理 ) 设置都不合适 ) 力挂钩程度为 C.40%	D. 50%	E. 60%以上
<ul> <li>A. 工资等级与工资级差试</li> <li>B. 工资等级设置太多(少)</li> <li>C. 工资级差太大(小)</li> <li>D. 工资等级与工资级差试</li> <li>q15. 目前薪酬中奖金与业绩的</li> <li>A. 20%及以下 B. 30%</li> <li>q16. 您认为奖金与业绩的挂钩</li> </ul>	设置比较合理 设置都不合适 的挂钩程度为 C. 40% 的程度应在	D. 50%	E. 60%以上
<ul> <li>A. 工资等级与工资级差试</li> <li>B. 工资等级设置太多(少)</li> <li>C. 工资级差太大(小)</li> <li>D. 工资等级与工资级差试</li> <li>q15. 目前薪酬中奖金与业绩的</li> <li>A. 20%及以下 B. 30%</li> <li>q16. 您认为奖金与业绩的挂钩</li> <li>A. 20%及以下 B. 30%</li> </ul>	<ul> <li>设置比较合理</li> <li>)</li> <li>设置都不合适</li> <li>的挂钩程度为</li> <li>C.40%</li> <li>动程度应在</li> <li>30%</li> </ul>	D. 50% . 40%	E. 60%以上
<ul> <li>A. 工资等级与工资级差试</li> <li>B. 工资等级设置太多(少)</li> <li>C. 工资级差太大(小)</li> <li>D. 工资等级与工资级差试</li> <li>q15. 目前薪酬中奖金与业绩的</li> <li>A. 20%及以下 B. 30%</li> <li>q16. 您认为奖金与业绩的挂钩</li> <li>A. 20%及以下 B.</li> <li>D. 50% E.</li> </ul>	<ul> <li>2</li> <li>2</li> <li>2</li> <li>2</li> <li>2</li> <li>3</li> <li>3</li> <li>4</li> <li>5</li> <li>4</li> <li>5</li> <li>5</li> <li>5</li> <li>5</li> <li>5</li> <li>6</li> <li>5</li> <li>5</li></ul>	D. 50% . 40% 不浮动	E. 60%以上
<ul> <li>A. 工资等级与工资级差试</li> <li>B. 工资等级设置太多(少)</li> <li>C. 工资级差太大(小)</li> <li>D. 工资等级与工资级差试</li> <li>q15. 目前薪酬中奖金与业绩的</li> <li>A. 20%及以下 B. 30%</li> <li>q16. 您认为奖金与业绩的挂钩</li> <li>A. 20%及以下 B.</li> <li>q17. 您认为奖金的浮动对自己</li> </ul>	<ul> <li>20置比较合理</li> <li>20置都不合适</li> <li>20括钩程度为</li> <li>C.40%</li> <li>20程度应在</li> <li>30%</li> <li>C</li> <li>60%及以上</li> <li>F.</li> <li>20的工作产生</li> </ul>	D. 50% . 40% 不浮动	E. 60%以上
<ul> <li>A. 工资等级与工资级差试</li> <li>B. 工资等级设置太多(少)</li> <li>C. 工资级差太大(小)</li> <li>D. 工资等级与工资级差试</li> <li>q15. 目前薪酬中奖金与业绩的</li> <li>A. 20%及以下 B. 30%</li> <li>q16. 您认为奖金与业绩的挂载</li> <li>A. 20%及以下 B.</li> <li>D. 50% E.</li> <li>q17. 您认为奖金的浮动对自己</li> <li>A. 积极帮助 B.</li> </ul>	<ul> <li>役置比较合理</li> <li>役置都不合适</li> <li>勿挂钩程度为</li> <li>C.40%</li> <li>勾程度应在</li> <li>30%</li> <li>C</li> <li>60%及以上</li> <li>F.</li> <li>己的工作产生</li> <li>一些帮助</li> <li>C</li> </ul>	D. 50% . 40% 不浮动 . 没影响	E. 60%以上
<ul> <li>A. 工资等级与工资级差试</li> <li>B. 工资等级设置太多(少)</li> <li>C. 工资级差太大(小)</li> <li>D. 工资等级与工资级差试</li> <li>q15. 目前薪酬中奖金与业绩的</li> <li>A. 20%及以下 B. 30%</li> <li>q16. 您认为奖金与业绩的挂转</li> <li>A. 20%及以下 B.</li> <li>q17. 您认为奖金的浮动对自己</li> <li>A. 积极帮助 B.</li> <li>D. 一些困扰 E.</li> </ul>	<ul> <li>役置比较合理</li> <li>分置都不合适</li> <li>为挂钩程度为</li> <li>C.40%</li> <li>动程度应在</li> <li>30%</li> <li>C</li> <li>60%及以上</li> <li>F.</li> <li>合い工作产生</li> <li>一些帮助</li> <li>C</li> <li>很大困扰</li> </ul>	D. 50% . 40% 不浮动 . 没影响	E. 60%以上
<ul> <li>A. 工资等级与工资级差试</li> <li>B. 工资等级设置太多(少)</li> <li>C. 工资级差太大(小)</li> <li>D. 工资等级与工资级差试</li> <li>q15. 目前薪酬中奖金与业绩的</li> <li>A. 20%及以下 B. 30%</li> <li>q16. 您认为奖金与业绩的挂载</li> <li>A. 20%及以下 B.</li> <li>D. 50% E.</li> <li>q17. 您认为奖金的浮动对自己</li> <li>A. 积极帮助 B.</li> <li>D. 一些困扰 E.</li> <li>q18. 您认为目前的薪酬体系中</li> </ul>	<ul> <li>役置比较合理</li> <li>役置都不合适</li> <li>勿挂钩程度为</li> <li>不 40%</li> <li>勾程度应在</li> <li>30%</li> <li>C</li> <li>60%及以上</li> <li>F.</li> <li>2的工作产生</li> <li>一些帮助</li> <li>て</li> <li>很大困扰</li> <li>基本工资与奖金的比例</li> </ul>	D. 50% . 40% 不浮动 . 没影响 刘相比	E. 60%以上
<ul> <li>A. 工资等级与工资级差试</li> <li>B. 工资等级设置太多(少)</li> <li>C. 工资级差太大(小)</li> <li>D. 工资等级与工资级差试</li> <li>q15. 目前薪酬中奖金与业绩的</li> <li>A. 20%及以下 B. 30%</li> <li>q16. 您认为奖金与业绩的挂转</li> <li>A. 20%及以下 B.</li> <li>q17. 您认为奖金的浮动对自己</li> <li>A. 积极帮助 B.</li> <li>D. 一些困扰 E.</li> <li>q18. 您认为目前的薪酬体系中</li> <li>A. 比较合理 B.</li> </ul>	<ul> <li>役置比较合理</li> <li>分置都不合适</li> <li>为挂钩程度为</li> <li>C.40%</li> <li>勾程度应在</li> <li>30%</li> <li>C</li> <li>60%及以上</li> <li>F.</li> <li>合い工作产生</li> <li>一些帮助</li> <li>化</li> <li>很大困扰</li> <li>基本工资与奖金的比例</li> <li>奖金比例偏大</li> <li>C</li> </ul>	<ul> <li>D. 50%</li> <li>. 40%</li> <li>不浮动</li> <li>. 没影响</li> <li>· 没影响</li> <li>· 利相比</li> <li>. 奖金比例偏小</li> </ul>	E. 60%以上
<ul> <li>A. 工资等级与工资级差试</li> <li>B. 工资等级设置太多(少)</li> <li>C. 工资级差太大(小)</li> <li>D. 工资等级与工资级差试</li> <li>q15. 目前薪酬中奖金与业绩的</li> <li>A. 20%及以下 B. 30%</li> <li>q16. 您认为奖金与业绩的挂载</li> <li>A. 20%及以下 B.</li> <li>D. 50% E.</li> <li>q17. 您认为奖金的浮动对自己</li> <li>A. 积极帮助 B.</li> <li>D. 一些困扰 E.</li> <li>q18. 您认为目前的薪酬体系中</li> <li>A. 比较合理 B.</li> <li>q19. 您清楚公司提供的福利和</li> </ul>	登置北较合理                  受置都不合适                  か程度应在                  30%                  402                  30%                  60%及以上                 60%及以上                 7.40%                 20%                 40世界助                 7.10%                 20%                 7.10%                 7.10%                 7.10%                 7.10%                 7.10%                7.10%                 7.10%                 7.10%                 7.10%                 7.10%                 7.10%                 7.10%                 7.10%                 7.10%                 7.10%                 7.10%                 7.10%                 7.10%	D. 50% . 40% 不浮动 . 没影响 利相比 . 奖金比例偏小	E. 60%以上
<ul> <li>A. 工资等级与工资级差试</li> <li>B. 工资等级设置太多(少)</li> <li>C. 工资级差太大(小)</li> <li>D. 工资等级与工资级差试</li> <li>q15. 目前薪酬中奖金与业绩的</li> <li>A. 20%及以下 B. 30%</li> <li>q16. 您认为奖金与业绩的挂钩</li> <li>A. 20%及以下 B.</li> <li>A. 20%及以下 B.</li> <li>q17. 您认为奖金的浮动对自己</li> <li>A. 积极帮助 B.</li> <li>D. 一些困扰 E.</li> <li>q18. 您认为目前的薪酬体系中</li> <li>A. 比较合理 B.</li> <li>q19. 您清楚公司提供的福利和</li> <li>A. 清楚 B.</li> </ul>	2       出較合理         2       2         2       2         3       C         4       4         4       4         4       4         5       C         4       4         5       C         6       0%         6       0%         2       0%         6       0%         2       0%         2       0%         2       0%         2       0%         2       0%         2       0%         2       0%         2       0%         2       0%         3       0%         2       0%         3       0%         2       0%         3       0%         3       0%         3       0%         3       0%         3       0%         3       0%         3       0%         3       0%         3       0%         3       0%         3       0% <td>D. 50% . 40% 不浮动 . 没影响 列相比 . 奖金比例偏小 . 不清楚</td> <td>E. 60%以上</td>	D. 50% . 40% 不浮动 . 没影响 列相比 . 奖金比例偏小 . 不清楚	E. 60%以上
<ul> <li>A. 工资等级与工资级差试</li> <li>B. 工资等级设置太多(少)</li> <li>C. 工资级差太大(小)</li> <li>D. 工资等级与工资级差试</li> <li>q15. 目前薪酬中奖金与业绩的</li> <li>A. 20%及以下 B. 30%</li> <li>q16. 您认为奖金与业绩的挂载</li> <li>A. 20%及以下 B.</li> <li>D. 50% E.</li> <li>q17. 您认为奖金的浮动对自己</li> <li>A. 积极帮助 B.</li> <li>D. 一些困扰 E.</li> <li>q18. 您认为目前的薪酬体系中</li> <li>A. 比较合理 B.</li> <li>q19. 您清楚公司提供的福利和</li> <li>A. 清楚 B.</li> <li>q20. 除工资、奖金外,公司损</li> </ul>	登置比较合理                  受置都不合适                  か建度应在                  30%                  74度应在                  30%                  60%及以上                 74                74                70%                 75                76                76                76                76                76                70%                 76                76                76                76                76                76                76                76                70%                70%                70%                70%                70%                70%                70%                70%                70%                70%                70% <td< td=""><td>D. 50% . 40% 不浮动 . 没影响 列相比 . 奖金比例偏小 . 不清楚 . 不清楚 . 不清楚</td><td>E. 60%以上</td></td<>	D. 50% . 40% 不浮动 . 没影响 列相比 . 奖金比例偏小 . 不清楚 . 不清楚 . 不清楚	E. 60%以上
<ul> <li>A. 工资等级与工资级差试</li> <li>B. 工资等级设置太多(少)</li> <li>C. 工资级差太大(小)</li> <li>D. 工资等级与工资级差试</li> <li>q15. 目前薪酬中奖金与业绩的</li> <li>A. 20%及以下 B. 30%</li> <li>q16. 您认为奖金与业绩的挂垫</li> <li>A. 20%及以下 B. 30%</li> <li>q16. 您认为奖金与业绩的挂垫</li> <li>A. 20%及以下 B.</li> <li>D. 50% E.</li> <li>q17. 您认为奖金的浮动对自己</li> <li>A. 积极帮助 B.</li> <li>D. 一些困扰 E.</li> <li>q18. 您认为目前的薪酬体系中</li> <li>A. 比较合理 B.</li> <li>q19. 您清楚公司提供的福利和</li> <li>A. 清楚 B.</li> <li>q20. 除工资、奖金外,公司提</li> <li>A. 住房补贴 B.</li> </ul>	2       出較合理         2       2         2       2         3       C         4       4         4       4         4       4         5       C         4       4         4       4         5       C         6       0%         6       0%         2       0%         6       0%         2       0%         0%       C         6       0%         0%       C         1       C         1       C         1       C         1       C         1       C         1       C         1       C         1       C         1       C         1       C         1       C         1       C         1       C         1       C         1       C         1       C         1       C         1       C         1       C	D. 50% . 40% 不浮动 . 没影响 列相比 . 奖金比例偏小 . 不清楚 §式有(可选择多巧 . 带薪休假	E. 60%以上

F. 财产或人身保险 G. 其他(请说明)
q21. 您希望公司增加哪些福利?(请说明)\_\_\_\_\_
q22. 您认为在现行的薪酬制度下,不同层级之间员工薪酬水平的差距\_\_\_\_\_
q23. 您对公司薪酬制度的改革和完善有什么意见和建议?(请说明)\_\_\_\_\_\_

### 1. 开放题录入

开放题即没有固定答案选项,由被访者自己填写答案,在变量视图中的名称单元格输入变量名称,并且设置其他变量属性,开放题的变量属性根据问题回答答案是数值或者文字来选择变量类型,在"员工薪酬调查问卷"中的S2题目为"年龄",为数值型变量,则在变量类型中选择"数值"。在此需要注意的是,开放题中由于变量的输入内容具体的实际意义,且内容丰富,因此我们就不需要对每个问题的答案进行 "值"标签设置。通常情况下,开放题的数据主要作用是研究人员对于了解研究对象的补充资料,无法进行算术分析。

### 2. 单选题录入

单选题的录入方法与开放题相似,不同的是,在单选题中需要将选项进行编码后定义 变量的值标签。例如,"员工薪酬调查问卷"中S1题目为"性别",有两个选项,将A选 项"男"编码为"1",B选项"女"编码为"2",在值标签中输入选项编码的标签含义。 具体操作步骤如下:在变量视图中,单击性别变量【值】标签单元格右端。按钮,在弹出 的【值标签】对话框中输入各个取值以及其标签,每输入完成一个值标签单击【添加】按 钮,依次添加,最后单击【确定】按钮即可,如图2-9所示。通过这种方式,既可以减少数 据录入的工作量,而且可以方便后面的数据分析工作。

文件(E)	编辑	i(E)	视图(⊻)	数据( <u>D</u> )	转换①	分析(A)	直销( <u>M</u> )	图形(G)	实用程序	(U) 窗口(W)
<b>a</b>	H	e			1		▙	tt	X	
			名称	类型	ų.	宽度	小数	标	签	值
1		性别		数值(N)	8		2			无
2		ſ	11 值标签	5(V)						23
3										
4			值标签	E(V)						
5			值(U):	2					拼写	5(S)
6			标签(	L): 女		_			ı —	
7									1	
8			l s	₹tn(A)	.00 = '男'					
9				4670H (C)						
10				E改(C)						
11			Bi	削除( <u>R</u> )						
12										
13										
14						确定 1	以消 帮助			
47		U								

图 2-9 值标签的编辑

### 3. 多选题录入

多选题,又被称为多重响应,是在社会调查和市场调研中极为常见的一种数据记录类型。一般情况下,对于问卷中的一个单选题一个被访者只有一个答案,对应的变量只有一个取值。而多选题,例如,"员工薪酬调查问卷"中"薪酬调查"部分的 q12 题和 q20 题,均为多选题,被访者可以选择两个或者更多选项。这样一来,多选题中每道题都可能有一

个以上的答案,多选题就不能用一个变量来直接编码,否则无法进行分析,这时需要使用 几个变量来进行记录。

在统计软件中多选题的录入方法通常有两种:多重二分法和多重分类法。

1) 多重二分法

所谓多重二分法,是指在编码时,对应每个选项都要定义一个变量,有几个选项就有 几个变量,这些变量分别代表其中一个选项的选择结果,一般分为"选中"选项与"未选" 选项两类。在 SPSS 中,多选题的录入程序与单选题相同,均是先在变量视图中进行变量设 置,然后直接录入数据。但多选题的不同之处是变量的定义方式不同,在数据录入完毕, 分析数据之前,还需要定义多选题变量集,利用变量集进行多重响应分析,而不能直接使 用变量分析。

在定义变量时,每个选项对应一个变量,例如"员工薪酬调查问卷"中 q20 题,对应 所选择的七种选项,分别设定了 q20.1、q20.2、q20.3、q20.4、q20.5、q20.6、q20.7 这七个 变量,均以 0 表示未选中,1 表示选中,如图 2-10 所示。可见第 1 个个案除工资、奖金外, 公司还提供给他的其他福利形式有住房补贴、交通补贴、带薪休假、股票期权、医疗补贴 及其他补贴,但是没有提供财产或人身保险。而第 4 个个案只有住房补贴、交通补贴和医 疗补贴。

文件(E)	编辑	(E) 视图	(V)	救据( <u>□</u> ) 转	讀( <u>T</u> )	分析( <u>A</u> )	直销( <u>M</u> )	图形(	G) 实用程	ş(U) 窗口(V	⊻) 帮助	
2					2				<b>88 </b>		- <b>4</b> 2	4
1 : ID			1								可见: 1	7
	[	q20.1		q20.2	q	20.3	q20.4		q20.5	q20.6	q20.7	
1			1		1	1		1	1	(	)	1
2			1		1	1		1	1	(	)	1
3			1		1	1		1	1	(	)	1
4			1		1	0		0	1	(	)	0
5			1		1	0		0	1	(	)	0

图 2-10 多重二分法的数据录入

在多重二分法中,无论多选题拆分成多少个变量,其变量值标签的定义应该一致,否则将会出现混乱。多重二分法适用于未限定选择答案数量的多选题。

2) 多重分类法

多重二分类法实际上是多选题的标准格式,但这种数据格式有时会给数据录入带来麻烦,比如"员工薪酬调查问卷"中 q12题,每个被访者被限制回答最多三个选项,但总选项数量有六个,显然,如果使用多重二分法录入,则有一半的数据需要录入为"未选中",徒增了许多数据录入的工作。对于这类限定选择选项数量的多选题,则较多使用多重分类法进行数据录入。

多重分类法与多重二分法一样,也是利用多个变量对一个多选题的答案进行定义,应 该用多少个变量来定义,由被访者实际可能给出的答案数量而定。这些变量采用一套值标 签,并且每个变量都是多分类的,每个变量代表被访者的一次选择。

以 q12 题为例,由于限定最多回答三个选项,因此只需要设定 q12.1、q12.2、q12.3 三 个变量即可。如图 2-11 所示,个案 1 选择了"个人业绩大小""定期升降""公司利润大 小"三个选项;个案 2 只选择了"个人业绩大小""公司利润大小"两个选项,q12.3 为缺 失值,这种数据缺失现象在多重分类法中其实是一种正常现象。



图 2-11 多重分类法的数据录入

### 4. 定义变量集

多选题录入完毕后 SPSS 只会默认它们是若干分散独立的变量,并不会把它们识别成一 道多选题,只有将其设定为多选题变量集(也称为多重响应集),SPSS 才会对其进行正确的 识别,从而将多选题的全部变量当成一整道题目来判断。但需要注意的是,统计分析的逻 辑是利用样本去估计分析总体,只有当样本值是唯一时对总体的估计分析才能准确,而多 选题的被访者的回答存在多种可能,变量的取值并不是唯一的。因此,我们认为多选题的 分析只适合进行简单的频率分析,而不适合进行更高级的统计分析,事实上,SPSS 到目前 为止也没有提供任何直接分析多选题数据的统计方法和功能模块。

SPSS 中提供了多种方式处理多选题,如图 2-12(a)所示,在【分析】菜单中的【多重响应】子菜单项的【定义变量集】模块,以及如图 2-12(b)所示,在【数据】菜单中的【定义 多重响应集】子菜单项,都可以用来设定多选题变量集。所不同的是,【多重响应】菜单 项的【定义变量集】定义的多选题变量集不能在 SPSS 数据文件中保存,关闭数据文件后相 应的信息就会丢失,如果再次使用,则必须重新定义变量集;而【数据】菜单中的【定义 多重响应集】模块可以保存所定义的信息。这两个过程的操作基本相同,现在以【分析】 菜单设定为例介绍如何定义多选题。



图 2-12 定义多重响应集

**步骤 1:** 打开本章数据"员工薪资调查",依次选择【分析】→【多重响应】→【定义 变量集】命令,如图 2-12(a)所示。

步骤 2: 单击【定义变量集】进入其对话框,将表示同一多选题的变量一起选入右侧的 【集合中的变量】框中。在【将变量编码为】选项组中选中变量编码方式,多重二分法方 式的题目选中【二分法】,需要在其右侧的【计数值】文本框中填入数字"1"。多重分类 法方式的题目选中【类别】,需要设定变量的取值范围,在该范围内的记录值将纳入分析。 将新定义的变量名称和标签填上,该例在【名称】文本框中填上"Q20",并在【标签】文 本框中填上"其他福利形式",如图 2-13 所示。然后将【集合中的变量】框中的变量添加 到右侧的【多响应集】框中,单击【关闭】按钮即可。

步骤3: 重新依次选择【分析】→【多重响应】命令时就会发现,原来呈现灰色的【频率】和【交叉表】命令现在已经被激活,如图2-14所示。这时我们便可以对定义的多重响应集进行分析了,不过用【分析】菜单所定义的多重响应集只能做频率分析和交叉表分析,对于其他分析使用的集,可以使用【数据】菜单上的定义多重响应集,有兴趣的读者可以自己查阅相关数据学习,限于篇幅,这里就不再详细列举过程了。本书在第3章时会介绍频率分析和交叉表分析,完成第3章学习后读者可以再回来尝试完成多重响应集的分析。



图 2-13 【定义多重响应集】对话框



图 2-14 多重响应集

# 2.2 数据的打开与保存

# 2.2.1 外部数据的打开

SPSS 软件在数据文件兼容性方面做得非常出色,除了可以打开 SPSS 格式文件数据(.sav) 以外,还可以直接读入许多常用格式的数据文件,包括 Excel、dBase、SAS、Stata 和 txt 格 式等,本书只介绍最常见的 Excel 和 txt 文件的读取。

### 1. 读取 Excel 文件

在读入数据前,首先要打开 Excel 数据,观察数据的基本结构是否与 SPSS 数据视图一 致,是否一行表示一个个案、一列表示一个变量。如果与 SPSS 数据视图不一致,需要在 Excel 工作表中进行数据处理,转置单元格行与列。然后关闭 Excel 工作表,再进行接下来 的读入数据的操作,依次选择【文件】→【打开】→【数据】菜单命令调出打开数据对话框。 因为系统会默认打开".sav"文件,所以需要在【文件类型】下拉列表框中选择 Excel(.\*xls,\*xlsx,\*xlsm)文件,这时 Excel 文件会显示在数据框中,如图 2-15 所示。选择要 打开的 Excel 文件,单击【打开】按钮,弹出【打开 Excel 数据源】对话框,如图 2-16 所 示。【从第一行数据读取变量名】选项用于确定 Excel 数据文件的第一行是否应被识别为变 量名称。在【工作表】下拉列表框中选择 Excel 数据文件的一个工作表(如果存在多个工作 表的话)。在【范围】文本框中指定被读取数据在 Excel 工作表中的位置,用单元格的起(所 要选择的 Excel 数据区域左上角单元格名,如 A1)止(所要选择的 Excel 数据区域右下角单元 格名称,如 F6)位置来表示,中间用":"隔开。例如,A1:F6表示选择宽度为A1-A6、长度 为F1-F6 的方块区域数据。设置完毕后,单击【确定】按钮数据就会被顺利读入 SPSS 中。 如果要读入整个 Excel 文档,则不需要设置"范围"。

1开数据	×
查找范围(L): 🔰 SPSS 🔹 💽 🔯 🔝 🔛	
@ 文件1xls @ 文件2xls	
	打开
文件类型(T): Excel (*.xis, *.xisx, *.xism) 🔽	粘贴(巴)
根据观察值,最小化字符串宽度 从存编库检索文件(B)	取消 帮助

图 2-15 选择 Excel 文件

#### 2. 读取 txt 文件

SPSS 可以通过两种菜单操作方式读取文本数据:一种是选择【文件】→【打开文本数据】菜单项,如图 2-17 所示;另一种方式与打开 Excel 文件的方式一样,选择【文件】→【打开】→ 【数据】菜单项,两种途径是一样的,系统会弹出打开数据对 话框,只是第一种方式的文本类型自动跳到了 Text(.txt),后者 需要在"文件类型"下拉列表框中进行选择。

文件(E)	编辑( <u>E</u> )	视图(⊻)	数据( <u>D</u> )	转			
新建	( <u>N</u> )			۲			
打开	-( <u>O</u> )			۲			
打开数据库(B)							
₩ 一 一 一 一 一 一 一 一 一 一							

图 2-17 打开文本数据

X

文本数据的读取与 Excel 数据一样,首先打开该数据,观察这数据的基本结构,例如, 变量间是固定宽度的,还是用某种分隔符区分的,第一行是否为变量名等。然后关掉这个 文本文件,再进行 SPSS 读入数据操作。以导入本章"001 数据"文本数据为例,在【打开 文件】对话框中【文件类型】下拉列表中选择"文本格式(\*.txt,\*.dat)",然后选中相应的 文本,单击右侧的【打开】按钮后会弹出"文本导入向导"对话框,如图 2-18(a)所示,从

C:Users\hpiDesktop\SPSS\数据1.xls ✓ 从第一行数据读取变量名。 工作表: Sheet1 [A1:X1602] ▼ 范围: 字符串列的最大宽度: 32767 通定 取消 帮助

图 2-16 【打开 Excel 数据源】对话框

11 打开 Excel 数据源

.

对话框标题中可以看到该导入导向共分六步,具体如下。

**步骤 1:** 系统首先会询问有无预定义格式,如图 2-18(a)所示,如果将要打开的文本数 据有预定义格式,则在此处选择相应的预定义格式文件,在下方为按预定格式读入的数据 文件的预览效果。若没有预定格式,保持默认的选择【否】并直接单击【下一步】按钮即可。

**步骤 2:** 在如图 2-18(b)所示的对话框中设定变量排列方式和变量名称,变量的排列方 式有两种选择,一种变量间是采用某些符号进行分隔的,在【变量是如何排列的?】选项 组中选择【分隔】;另一种变量间采用的是固定顶宽度来分隔变量,选择【固定宽度】, 然后在下方的【文本文件】选项组中调整标尺上的分隔线位置来设定变量的固定宽度。如 果文件中有变量名称,则需要将【变量名称是否包括在文件的顶部?】选项组中选择【是】, 单击【下一步】按钮。



图 2-18 文本导入向导的第 1、2 步

**步骤 3:** 在如图 2-19(a)所示的对话框中确定数据开始行每个个案所占的行数、希望导入的个案数量,一般前两者的默认设定就是最常见的情况,第三个功能则可以用于个案进行随机抽样。

**步骤 4:** 对变量分隔符以及文本限定符进行设定,如图 2-19(b)所示,根据相应选项的 设定情况,下方会动态显示出数据的预览情况。这里选择的导入文本文件中变量之间采取 的是逗号分隔变量,因此在【变量之间有哪些分隔符?】选项组中选中【逗号】,下方的 数据预览窗口会显示出正确的数据读入情况。右侧的【文本限定符是什么?】选项组提供 了"无""单引号""双引号"和"自定义"四种选择。如果数据中的字符串变量使用了 限定符进行分隔,则需要在此处进行设定。

步骤 5: 在如图 2-20(a)所示的对话框中对各个变量做进一步的属性设定,包括更改变量名和更改数据格式类型,在下方的【数据预览】选项组中选择某一列需要更改的变量即可进行操作,如果这里不需要进行更改,可以直接单击【下一步】按钮。

步骤 6: 在如图 2-20(b)所示的对话框中确定是否希望重复利用本次操作的选择,可以

考虑将这次的文件设定保存为预定义格式文件,或者将本次操作粘贴为 SPSS 语句。如果直接单击【完成】按钮,则向导结束,随后就可以看到 SPSS 成功读入该文本数据。

III 文本导入向导 - 第 3 步 ( 共 6 步 ) 分隔	田 文本导入向导 - 第 4 步 (共 6 步)分隔     □
第一个教报个案从哪个行号开始?(C) 2 2 2 2 2 2 3 3 3 2 5 2 5 2 5 2 5 2 5 2 5 2 5 2 5 2 5 2 5 2 5 5 5 5 5 5 5 5 5 5	交量/尋問書整分陽符?       文本限定符是什么?         創業符()       空格(5)         逆 運号       分号(5)         其他(3):       公式号点(2)         其他(3):       女号(5)
<ul> <li>○ 前① 1000 个个套。</li> <li>○ 个案的随机百分比(近似值)(巴): 10 ● %</li> <li>● 数据预选</li> <li>0.10.20.30.40.50.50.00.</li> </ul>	- - - 数据特別 - - - - - - - - - -
1     b.2,4     2       2     b,5,1     v       3     b,6,1     v       <     L-歩優)     下一歩 >(1)       完成     取消     現約	5 9 8 (<上一步(1)) 下一步 > (1)) 完成 取消 税助

(a)

(b)

图 2-19 文本导入向导的第 3、4 步

文本导入向导 - 第5步(共6步)     文本导入向导 - 第5步(共6步)	文本导入向导-第6步(共6步)
在數据粉壳中边探的变里规范 变量全称(V): 原始名称: 原验 编号 數据格式(D):	
教査	
4.1932.04	
8487975   編号 科目 成绩 1 2 4 2 5 4 3 6 1 4 9 10 5 9 9	$\begin{array}{c ccccccccccccccccccccccccccccccccccc$
(×上−歩@) 下−歩 ×(M) 完成 取消 帮助	< 上一歩(色) 下一歩 >(2) 完成 取消 若助
(a)	(b)

图 2-20 文本导入向导的第 5、6 步

# 2.2.2 | SPSS 数据的保存

SPSS 数据录入并编辑整理完成以后应及时保存,以防数 据丢失, SPSS 的数据文件默认保存格式为".sav",如图 2-21 所示。保存数据文件可以通过【文件】→【保存】或者【文 件】→【另存为】命令来执行。当然,SPSS 也可以将数据另 存为其他格式的文件,例如常用的 Excel 文件,只要在【另存 为】对话框中选择你要存储的格式即可。



图 2-21 数据的保存

# 2.3 数据的管理

# 2.3.1 数据检验

数据录入 SPSS 之后,需要先检查核对数据是否存在录入错误,有的话需要及时修正,以保证在使用数据时得到正确的分析结果。

### 1. 是否存在空行/空列

首先需要核对数据录入时是否存在空行或空列,这些空行或空列并不是数据缺失,而 是由于在数据录入时操作的疏忽所导致的,这会影响到后继的数据分析结果,因此,我们 必须将这些空行或空列查找出来,并删去。检查的方法十分简单,可以单击选中某一列的 变量名,右击后在弹出的菜单栏中选择【升序排序】,如果存在空行,空行将自动呈现在 最前面,如图 2-22 所示。



图 2-22 查找空行

### 2. 变量数值是否超出特定范围

在数据录入过程中有时会出现录入的数值与问卷中的变量值范围不一致的情况。这种 不一致有两种可能性:一种是数据在录入时出现的操作失误,另一种则可能是在调查过程 中被访者的误答。前一种情况我们需要找到原始问卷对录入数据进行修改,后一种情况则 只能将该被访者填答的变量标记为缺失值。检查的方法也可以通过上文提到的对变量进行 升序排列的方式来查找超出特定范围的值。

### 3. 变量数据是否存在重复样本

在进行大量数据录入时,当数据录入工作中断或多人分别录入时,经常会出现重复录入的情况,从而产生重复样本数据。重复样本的检查可以选择【数据】→【标识重复个案】 命令完成,如图2-23所示,现以案例演示其基本过程。



请将本章数据"标识重复个案.sav"中的重复个案找出来。 **案例分析:**标识重复个案最重要的是确定筛选重复个案的 变量,变量越具有区分性越好。例如,身份证号就是一个好的 筛选变量,因为每个个案只有一个号;而性别就不是一个好的 筛选变量,因为个案在该变量取值相同的概率太大。

步骤 1: 打开本章数据"标识重复个案.saw",选择【数据】 →【标识重复个案】命令,如图 2-23 所示。

步骤2:单击【标识重复个案】进入其主对话框,选择"查 重"的依据,将作为筛选重复样本标准的变量置入【定义匹配 个案的依据】框中。在此需要注意的是,除非确认某筛选变量 每个个案的取值是唯一的,否则建议尽可能多选择几个变量作



图 2-23 标识重复个案

为筛选依据,以防误判。这里把"ID"和"s1"作为"查重"的依据,如图2-24所示。

步骤 3: 标识重复个案会生成新的变量,我们需要对这个变量做基本设置。基本个案指示符是指对于重复个案,可以指定其中一个为主个案,其余为多余的"重复"个案。可以将第一个个案或最后一个个案设定为主个案,主个案标识变量取值为1,重复个案标识为0。 这里选择系统默认状态,即【每组中的最后一个个案为基本个案】,如图 2-24 所示。



图 2-24 【标识重复的个案】对话框

步骤 4: 单击【确定】按钮后,数据视窗的左侧将生成新的变量"最后一个基本个案",如图 2-25 所示。我们可以看到,第一个个案的变量值为"0",第二个个案为"1",这就意味着第一个个案和第二个个案是重复的,其余的数据依次类推。最后,在结果输出窗口中还会给出本次操作的信息汇总,如表 2-1 所示,可见一共有三个重复的个案,占总数据的 18.8%。重复个案通常需要删除,可以对"最后一个基本个案"升序排序,然后删除前面取 值为 0 的所有个案即可。

文件(E)	编辑	(E) 视图	( <u>V</u> ) 数据( <u>D</u> )	转换(T) 分析	·(A) 直销(M)	图形( <u>G</u> ) 实用	月程序( <u>U</u> )	窗口( <u>W</u> ) 帮助
	H		🛄 🗠	<b>~</b>	1 📥 🗐	#1	*5 🖬	
1:最后-	-个基2	本个案	0					
	[	q20.2	q20.3	q20.4	q20.5	q20.6	q20.7	最后一个基本个案
1		1	1	1	1	0	1	0
2		1	1	1	1	0	1	1
3		1	1	1	1	0	1	0
4		1	1	1	1	0	1	1
5		1	0	0	1	0	0	0

图 2-25 标识重复个案生成的新变量视图

表 2-1 重复个案输出窗口结果

		频率	百分比	有效百分比	累积百分比
	重复个案	3	18.8	18.8	18.8
有效	主个案	13	81.3	81.3	100.0
	合计	16	100.0	100.0	

### 4. 变量数值之间是否符合特定逻辑

在研究中的变量数值之间可能存在趋同关系、互斥关系和函数关系等逻辑关系。SPSS 提供了数据验证模块帮助用户进行数据的核查,用户可以通过自行定义数据验证的规则对 数据进行检查。例如,性别 s1 只有 1、2 两种取值码;年龄 s2 取值应当在 18~60 岁之间; q3 选择"2"时 q4 应当为缺失等,否则该题被视为无效题。前两种验证方式叫作单变量验 证方式,后一种验证方式叫作交叉变量验证方式,即涉及多个变量的交互关系。连续单变 量验证方式是将变量值设定在某个范围,而离散单变量验证方式是将变量值设定出某些特 殊的取值。这里简单介绍连续单变量的验证不过程,其他形式的验证,有兴趣的读者可以 参考相关书籍尝试。



用【验证数据】命令找出本章数据"员工薪资调查.sav"中年龄不在 30~50 岁的被试。 案例分析:验证数据最重要的是明确验证规则,该例的验证规则很简单,即年龄在 30~50 岁,超出该范围的可能是不符合条件的数据。

步骤 1: 打开本章数据"员工薪资调查.sav",依次选择【数据】→【验证】→【验

证数据】命令,如图 2-26 所示。



图 2-26 验证数据

步骤 2: 单击【验证数据】进入其主对话框,在【变量】选项卡中将需要验证的变量放入【分析变量】框中,这里选择 "s2"。个案标识变量是指如果筛选出不合格的数据,用什么的方式将其标识出来,一般需要选择一些能够唯一确定个案的变量,如学号、工号、编号等。这里选择 "ID 号",如图 2-27 所示。



图 2-27 【验证数据】对话框

步骤 3: 【基本检查】选项卡的选项选择默认,主要对【单变量规则】选项卡的内容进 行设定,单击【单变量规则】选项卡进入其界面,如图 2-28 所示,单击其右下角的【定义 规则】按钮进入【验证数据:定义验证规则】对话框。这里需要对验证规则进行设定,先 命名该规则为"年龄规则",填入其最大值"50"和最小值"30",如图 2-29 所示,单击 【继续】按钮回到上一层界面并勾选刚才所设定的"年龄规则",表示将应用到目标变量 "s1"上。如果需要将筛选出来的个案保存到数据上,可以单击【保存】选项卡进入其界面, 选中【保存用来记录所有验证违规的指示变量】复选框,如图 2-30 所示。最后单击【确定】 按钮,提交系统分析,输出结果如图 2-31 和表 2-2 所示。

验证数据								
变量 基本检查	单变重规则 交叉	安重规则 输	出保存					
夏向某个变量应用共	·····································	重,然后选中·	一个或多个共	砚	_			_
析变量列表根据器	动掘的扫描显示非缺	失值的分布。	则列表显示	同以应用	创所选	变量的所有规	则。	
分析变量( <u>A</u> ):				Inni		规则(U):		
安里	分布	載小値	載大值	RUN		应用	名称	
s2		2	2 5	5 0				
見子(の): 新有変角	<ul> <li>□扫描</li> </ul>	的小贺粉:13					<b>+</b>	:∿±ami
显示(D): 所有変重	▼ 已扫描	的个案数: 13					淀	义规则
显示(D): 所有变量 变量分布	▼ 已扫描	的个案数: 13		10 ku				:义规则
显示(D): 所有変重 変重分布 ☑ 限制扫描的个	▼ 已扫描 案数(L) 个案数(C)	的个案数: 13 : 5000	重新扫描(§	) 限制把	描的	▶案数不会≪BF	定 ?~景响验证的个案数里	:义规则 。
显示①: 所有变量 变量分布 ▼限制扫描的个部	! ▼ 已扫描 案数(L) 个案数(C)	的个案数: 13 : 5000	重新扫描(3	3) 限制拍	描的~	▶ 案数不会 <bf< td=""><td>定 2&gt;景响验证的个案数里</td><td>:义规则 。</td></bf<>	定 2>景响验证的个案数里	:义规则 。

图 2-28 【单变量规则】选项卡

<ul> <li>註 验证数据: 定义验证规则</li> <li>单支量规则</li> </ul>	×
規则(民): 名称 类型 年齡規则 数字	規则定义 名称(M): 年齢規则 推式(E): 「mm/dd/yyy マ
	<ul> <li>有效值():</li> <li>在范围内 ▼</li> <li>最小值():</li> <li>20</li> <li>指定最小值():</li> <li>最大值(2):</li> <li>50</li> <li>第首值将视力在(2)范围内。</li> <li>最大值(3):</li> <li>50</li> <li>第方在范围内使用非标注的值(4)</li> <li>由于长字符串定里没有值标卷,对于这种变量您应该经常选中此选项。</li> <li>ダ 允许在范围内使用非整数值(2)</li> <li>ダ 允许使用用户缺失值(1)</li> <li>マ 允许使用用户缺失值(1)</li> <li>マ 允许使用用户缺失值(1)</li> <li>マ かけ、使用系统缺失值(2)</li> <li>マ かけ、使用の(3)</li> </ul>
*##[0] E9](0) #9*(0)	

图 2-29 【验证数据: 定义验证规则】对话框

变量 基本检查 单变里规则	J 交叉变量规则 輸出	保存	
摘要变量( <u>S</u> ):			
描述		保存	名称
空个案指示器			EmptyCase
双ID组			DuplicateIDGroup
ID 指示器不完整			IncompleteID
验证违规数(总数)			ValidationRuleViolations
替换现有的摘要变量(E)			
🔽 保存用来记录所有验证违规	的指示变量①		

图 2-30 保存违规指示变量

步骤 4: 从图 2-31 中,我们可以看出,数据视窗的最右侧生成了一个新的变量"年龄规则\_s2",其取值为"0"和"1",取值为"0"的表示符合验证条件,取值为"1"的表示不符合验证条件。而表 2-2 也给出了验证结果的详细信息,标识了其个案 ID 号,我们看到第一个个案号为"3"(序号), ID 号为"2",经检验会发现其正是图 2-31 上的第三个个案。

	q20.3	q20.4	q20.5	q20.6	q20.7	年龄规则_s2
1	1	1	1	0	1	0
2	1	1	1	0	1	0
3	1	1	1	0	1	1
4	0	0	1	0	0	0
5	0	0	1	0	0	0

#### 图 2-31 验证数据结果

بن ج	确认违反规则	标识符
系例	单变量 <sup>a</sup>	ID
3	年龄规则 (1)	2
6	年龄规则 (1)	3
9	年龄规则 (1)	7
10	年龄规则 (1)	13
12	年龄规则 (1)	8

表 2-2 验证数据结果报告

a. 违反规则的变量数遵循每个规则。

# 2.3.2 数据的合并

在进行 SPSS 数据分析时,常常遇到这样的情况,即欲分析的数据被分别存储在几个不同的文件中,此时我们需要将这些文件合并成一个总文件才能进行后续的统计分析。针对不同的数据构成情况, SPSS 提供了两种数据文件的合并方式:一种是纵向的合并个案,另一种是横向的合并变量。

### 1. 合并个案

合并个案是将若干个数据集中的数据进行纵向拼接组成一个新的数据集,合并后的数 据集的个案数是原来几个数据集个案数的总和,这一方法也被称为添加个案。添加个案的 特征是,个案被分散在不同的数据文件中,但这些数据文件的变量构成基本相同。需要注 意的是,添加个案并不是只能添加个案,实际上在添加个案的过程中,有些变量也因为是 新的而被添加进去。



将本章数据"合并数据 1.sav"和"合并数据 2.sav"合并。

**案例分析:**观察两份数据的基本结构,发现两份数据的大部分变量是相同的,只是"合并数据1"(见图 2-32(a)),比"合并数据2",(见图 2-32(b))多了一个"年龄"变量,另外从"编号"看其个案数,可以看出两份数据的个案是不同的,对于这样的数据采用"添加个案"进行合并较为妥当。

编号	性别	年齡	民族	职位	工资		编号	性别	民族	职务	工资
1	2	39	1	3	5700	ĺ	6	2	1	3	5000
2	2	37	2	3	6200	Ì	7	2	2	3	6000
3	1	40	1	3	6200		8	1	1	3	6200
4	2	35	1	2	5700		9	2	1	2	5700
5	1	41	1	1	6900		10	1	1	1	7010

(a) 合并数据 1

(b) 合并数据 2

图 2-32 数据比较

步骤 1: 首先打开两份数据文件,以其中任何一份 数据作为源数据进行合并,这里选择"合并数据1"作为 源文件。在"合并数据1"上,依次选择【数据】→【合 并文件】→【添加个案】命令,如图2-33所示。

步骤 2:单击【添加个案】进入到合并数据向导框,如图 2-34 所示,上面提供了已经在桌面打开的数据,如果不想合并已打开的数据,可以重新选择文件。这里选择"合并数据 2",单击【继续】按钮后进入添加个案对话框,如图 2-35(a)所示。在【非成对变量】框中显示的变量是两个数据集中没有成对的变量,这些变量名后面都附加了"\*"或"+"号,"\*"表示该变量名是当前活动数据集中有的变量,"+"表示该变量名是外部待合并数据文件中的变量,从图 2-35(a)中我们可以看出,"年龄""职位"和"职务"这三个变量是没有配对成功的,前两个变量是



图 2-33 添加个案

原来数据的变量,后一个变量是新添加进来的变量。【新的活动数据集中的变量】框中显 示的是将要合并的新数据的变量,它们都是两个待合并的数据中共有的变量名。如果希望 对数据集中的变量名重新命名,可以单击【重命名】按钮重新设置变量名,这里不做改变。

步骤 3: 对于没有能匹配成功的变量,我们需要进一步分析变量的关系,例如,"职位"和"职务"两个变量,经过分析发现是同一个变量,所以需要对其进行强制配对,可以通过 Ctrl 键选中两者,然后单击【对】按钮把两者配对进右侧的【新的活动数据集中的变量】 框。而对于"年龄"这个变量,并没有和它重复且不同名的变量,所以直接单击向右箭头

进入【新的活动数据集中的变量】框便可。如果希望在合并后的数据文件中看出个案的来 源,可以选中【将个案源表示为变量】复选框,此时合并后的数据文件中将自动出现名为 "源 01"的变量,取值为 0 或 1。"0"表示记录来自当前活动的数据集,"1"表示被合 并的外部数据集,这里也选中该选项。所有设置完成后如图 2-35(b)所示,最后单击【确定】 按钮,提交系统分析,可以看到新的数据集已经合成,如图 2-36 所示。



图 2-34 合并数据向导框



图 2-35 添加个案对话框

步骤 4: 从图 2-36 中我们可以看出,"年龄"变量下有部分缺失值,那是因为新增加的数据没有这个变量所致。新数据增加了一个变量,即"源 01",其有"0"和"1"两种取值,"0"取值是指这些个案属于源文件的,"1"取值是指新增加的个案。

	编号	性别	年齡	民族	职位	工资	源01
1	1	2	39	1	3	5700	0
2	2	2	37	2	3	6200	0
3	3	1	40	1	3	6200	0
4	4	2	35	1	2	5700	0
5	5	1	41	1	1	6900	0
6	6	2		1	3	5000	1
7	7	2		2	3	6000	1
8	8	1		1	3	6200	1
9	9	2		1	2	5700	1
10	10	1		1	1	7010	1

图 2-36 合并后的数据

### 2. 合并变量

合并变量是指将若干个数据文件中的变量与己有的数据变量进行合并,即在某个数据 中增加变量(添加列),这一方法也被称为添加变量。添加变量的特征是,数据文件中的个案 基本相同,但是每个数据文件的变量基本不同。需要注意的是,添加变量并不是只能添加 变量,实际上在添加变量的过程中,有些个案也因为是新个案而被添加进去。



将本章数据"合并数据 3.sav"和"合并数据 4.sav"合并。

**案例分析**:观察两份数据的基本结构,发现两份数据的大部分个案是相同的,只是"合并数据3"(见图 2-37(a)),比"合并数据4"(见图 2-37(b))多了一个编号为"6"的个案;观察变量,我们可以看出,两份数据的变量部分相同,但是也有很多是不同的,对于这样的数据采用"添加变量"进行合并较为妥当。

编号	性别	民族	年级	Q6	Q7	Q8	Q9	Q10	绑	号	性别	民族	年级	Q1	Q2	Q3	Q4	Q5
1	2	1	4	2	3	5	2	4		1	2	1	4	4	4	2	5	3
2	2	1	4	2	2	4	3	4	]	2	2	1	4	4	2	1	2	2
3	1	1	4	1	1	1	3	4		3	1	1	4	3	2	3	2	4
4	1	1	4	5	5	1	1	3		4	1	1	4	1	5	1	3	5
5	2	2	4	3	3	5	1	5		5	2	2	4	4	4	1	1	5
									]	6	2	2	4	3	4	1	2	3
4i	(a) 合并数据 3					41			(b)	合并	数打	居 4						

图 2-37 数据比较

步骤 1: 首先打开两份数据文件,以其中任何一份数据作为源数据进行合并,这里选择 "合并数据3"作为源文件。在"合并数据3"上,依次选择【数据】→【合并文件】→【添 加变量】命令,如图2-38所示。

步骤 2: 单击【添加变量】进入到合并数据向导框,如图 2-39 所示。选中"合并数据 4", 单击【继续】按钮后进入添加变量对话框,如图 2-40(a)所示。在【已排除的变量】框中显 示的变量是两个数据集中重复的变量,这些变量的变量名后面都附加了"+"号,从图 2-40(a) 中我们可以看出,"年级""性别""民族"和"编号"是两份数据重复的变量。【新的 活动数据集】框中显示的是合并后的新数据的变量名,该列表框中的变量名后都附加有"\*" 或"+"号,"\*"表示该变量名是当前活动数据集中的变量,"+"表示该变量名是外部待 合并数据文件中的变量。在默认情况下,如果变量名没有在两个数据集中同时出现,则 SPSS 会自动将其列入新数据文件的变量列表中。

步骤 3: 如果两个待合并的数据文件中的记录数据排列的顺序是按照记录编号横向一一 对应的(即个案完全一样),则可以直接单击【确定】按钮完成合并工作,否则必须按照"关 键变量"将两份数据进行匹配,实际上如果数据比较庞大,去检查数据是不是一一对应是 不太方便的,所以一般都是按照匹配关键变量进行操作,这个步骤是合并变量最关键的步骤。被匹配的关键变量名必然因为重名出现在"已排除的变量"框中,由上面的分析可知,

这里有"年级""性别""民族"和"编号"四个变量名是重复的,先选择最优的匹配变量"编号",因为它的取值是唯一的,而其他变量取值都不是唯一的。把重复变量放进【关键变量】框前需要先选中【按照排序文件中的关键变量匹配个案】复选框。但是如果仅仅以"编号"匹配,新的个案的其他重复变量的值是缺失的,所以还需要添加"年级""性别""民族"三个变量到【关键变量】框中,如图 2-40(b)所示,最后单击【确定】按钮,提交系统分析,系统此时会提醒关键变量是否已经按升序排好序,如果未排序,需要关闭命令先对数据进行排序。因为这里"编号"变量已经排好序,所以单击【确定】按钮就可以看到新的数据集已经合成,如图 2-41 所示。





图 2-38 添加变量

图 2-39 合并数据向导框

已排給的变量(E):	新的活动数据集(N):	已排除的变量(E):	新的活动数据集(N):
年級(+) 性勞)(+) 民族>(+) 編号(+) 重命名(A)	編号(*) 性別>(*) 民話>(*) 年級(*) G6(*) G7(*) G8(*) G9(*) G9(*) G10(*) ▼	重命名(5)	O8(*)         C           O7(*)         O8(*)           O8(*)         O8(*)           O8(*)         O8(*)           O2(*)         O3(*)           O2(*)         O3(*)           O4(*)         V
<ul> <li>□ 按照排序文件中的关键交里匹配个案</li> <li>◎ 两个文件都提供个案(B)</li> <li>◎ 非活动数据集为基于关键字的表(C)</li> <li>◎ 活动数据集为基于关键字的表(A)</li> </ul>	关键交重①:	<ul> <li>☑ 按照排序文件中的关键变里匹配</li> <li>◎ 两个文件都提供个案(B)</li> <li>◎ 非活动数据集为基于关键字的表(</li> <li>◎ 活动数据集为基于关键字的表(</li> </ul>	☆ 关键变型(V): 编号 民族* 年級
<ul> <li>□ 将个案源表示为变量①: 源01</li> <li>(*)=活动數据集</li> <li>(*)=数据集2</li> </ul>		<ul> <li>一 将个案源表示为变量(): 原01</li> <li>(*)=活动数据集</li> <li>(*)=数据集2</li> </ul>	
确定 粘贴(P) 重型	B(R) 取消 帮助	确定 粘贴 の	重置(R) 取消 帮助

(a) 添加关键变量前

(b) 添加关键变量后

图 2-40 添加变量对话框

编号	性别	民族	年级	Q6	Q7	Q8	Q9	Q10	Q1	Q2	Q3	Q4	Q5
1	2	1	4	2	3	5	2	4	4	4	2	5	3
2	2	1	4	2	2	4	3	4	4	2	1	2	2
3	1	1	4	1	1	1	3	4	3	2	3	2	4
4	1	1	4	5	5	1	1	3	1	5	1	3	5
5	2	2	4	3	3	5	1	5	4	4	1	1	5
6	2	2	4						3	4	1	2	3

图 2-41 合并后的数据

步骤 4: 从图 2-41 中我们可以看出,新数据集的变量除了两者重复的四个变量,还增加了 Q1~Q5 这五个新的变量。个案数上也由原来的五个增加到了六个。第六个个案中的缺失值是因为该个案在第一份数据(即"合并数据 3")上没有取值。

# 2.3.3 数据的排序

SPSS 数据编辑窗口的记录前后次序在默认情况下是由录入时的先后顺序决定的,在实际工作中,有时希望按照某种顺序来观察一批数据。例如,在"员工薪酬调查"数据中,将数据按照 ID 顺序来进行排列,以便随时检索和浏览。下面简单介绍 SPSS 提供的三种数据排序方式。

### 1. 单变量排序

单变量排序在 SPSS 中操作最为简单,在要排序的列变量名处右击,弹出快捷菜单,选择后两项"升序排序"或"降序排序"即可。

#### 2. 多变量单向排序

多变量单项排序与单变量排序操作步骤类似,同时选中要排序的各个变量后在变量名 处右击,弹出快捷菜单选择"升序排序"或"降序排序"即可。这种个案排列的原理是, 先按第一个变量排序,当第一个变量取值相同时再对相同取值的个案按第二个变量做同向 排序。

### 3. 多变量混合排序

多变量混合排序是指根据多个变量各自不同的排 序方式对个案进行排序,其中有的是升序,有的是降序 排序,这种情况需要使用菜单中的"排序个案"进行操 作。选择【数据】→【排序个案】命令后,如图 2-42 所示,在【排序依据】框中选入排序依据的各个变量, 然后分别单独设置各个变量的排序方式,设置为升序的 变量后有"(A)"标识,设置为降序的变量后有"(D)"



### 图 2-42 【排序个案】对话框

标识。单击【确定】按钮提交系统分析后,系统的结果输出窗口不会输出排序的结果,通 过查看数据视图可以发现个案顺序发生了改变。

# 2.3.4 选择个案

在实际统计分析中,有时并不需要对所有的个案进行统计分析,而只要求对某些特定 的个案进行分析,此时就需要先选出这部分个案才能进行后续分析。例如,只分析男性员 工的数据,或者只分析业务部门员工的数据。从样本中选择部分个案,这可以利用【选择 个案】菜单来操作。



打开本章数据"员工薪酬调查.sav",筛选出业务部门的男性员工。 **案例分析:**这里筛选的条件有两个,一个为业务部门,一个为男 性。当然,筛选的条件不仅可以是一个、两个,还可以是任意多个。 多个条件的合并,需要用字符"&"将条件进行链接。

步骤 1: 打开本章数据"员工薪酬调查.sav", 依次选择【数据】→【选择个案】命令, 如图 2-43 所示。

步骤 2: 单击【选择个案】进入其主对话框,如图 2-44 所示。【选 择个案】对话框由【选择】选项组和【输出】选项组组成,系统提供 了五种选择个案的方式: 第一,"全部个案",表示全部个案都纳入 分析,不进行筛选,这是默认设置; 第二,"如果条件满足",表示 按指定条件进行筛选个案,这是初学者使用最多的方式; 第三,"随 机个案样本",表示从原始数据中按照某种条件随机抽样,使用下方 的【样本】进行具体设定,可以按百分比抽取个案,或者精确设定从 前若干个个案中抽取多少个个案; 第四, "基于时间或个案全距", 表示基于时间或个案序号来选择相应的个案,使用下方的【范围】按 钮设定个案序号范围; 第五,"使用筛选器变量",此时需要在其下 方选择一个筛选指示变量,该变量取值非0的个案将被选中,进行之 后的分析。





图 2-44 【选择个案】对话框

步骤 3: 选择"如果条件满足"方式,单击其下方的【如果】按钮将会打开【选择个案: If】对话框,用于定义筛选条件的数学表达式,如图 2-45 所示。将左侧待筛选的变量选入 右侧顶部空文本编辑框中,利用其下方的小键盘编辑变量的筛选条件,小键盘提供了最基 本的算数运算方法。如果个案的筛选需要进行更复杂的函数运算,小键盘右侧的【函数组】 列表框还提供了更丰富的运算函数,用户可以在【函数组】列表框中单击一个函数选入上 方的文本编辑框,然后在函数公式中插入变量。这里条件有两个,一为业务部门,一为男 性。双击变量 s1 进入右侧的运算框,然后再编写等式,即"s1=1",同理,把"s3=1"在 框中编辑好,因为是两个条件,需要用"&"连接,所以数学表达式最终为"s1=1& s3=1"。 条件设置好后,单击【继续】按钮回到上一层对话框。



图 2-45 【选择个案: If】对话框

步骤 4: 选择个案的输出方式。在图 2-44 所示的对话框中,【输出】选项组提供了三种方式处理选择结果:第一,"过滤掉未选定的个案"。未选定的个案将不包括在分析中, 但仍然保留在数据文件中,使用该选项后会在数据文件中生产命名为"filter\_\$"的变量, 对于选定的个案该变量的值为"1",未选中的个案该变量值为"0",在数据视图中未被 选中的个案号会以"/"加以标记。第二,"将选定个案复制到新数据集"。将选定的个案 复制到新数据集时,原始数据集不会受到影响,只是另外生成了一个只包含被筛选出的个 案的新数据文件。第三,"删除未选定个案"。直接从数据文件中删除未选定个案。需要 注意的是,一旦选择此项操作,原有未被选定的个案数据将从原始数据文件中删除,此外, 由于此项操作不能后退撤销,因此我们要谨慎操作,以免数据丢失。如果不小心选择此项 操作但还没保存文件,那可以退出文件不保存任何修改,这样才能恢复原来的完整数据。 这里选择系统默认设置,即选择【过滤掉未选定的个案】,最后单击【确定】按钮,提交 系统分析,输出结果如图 2-46 所示。

	ID	s1	s2	s3	q3	q4
1	1	1	33	1	2	2014
2	11	1	34	1	2	2014
	2	2	28	1	1	2015
4	10	2	40	2	1	2013
5	12	2	30	2	1	2014
6	3	1	55	2	1	2011
7	4	1	41	1	1	2013
8	5	1	43	1	1	2011

#### 图 2-46 选择生效后的数据界面

步骤 5: 从图 2-46 我们可以看出,有些个案号上面画有一条斜线,表示这些个案不符 合我们分析的要求,是被过滤掉的部分,系统暂时做好标识,当退出数据文件再打开的时 候,这些斜线就会消失。接下来分析的任何命令都是针对已选中的个案进行的,如果想要 重新选择全部个案,则在【选择】选项组中选中【全部个案】即可对所有个案进行分析。

# 2.3.5 计算变量

在数据统计分析的过程中,我们经常需要对数据变量进行各种运算然后得到新的变量, 如数据的求和、函数运算等。在 SPSS 中可以通过选择【转换】→【计算变量】命令来产 生这样的新变量。



打开本章数据"员工薪酬调查.sav",计算第20题各个选项的得分之和。

**案例分析:**利用【计算变量】命令对原始数据进行必要的四则运算是数据整理的常见 工作,初学者需要掌握。【计算变量】命令还可以与【选择个案】命令结合使用。

步骤 1: 打开本章数据"员工薪酬调查.sav",选择【转换】→【计算变量】命令,如 图 2-47 所示。



#### 图 2-47 计算变量

步骤 2: 单击【计算变量】进入其对话框,如图 2-48 所示。在左侧【目标变量】文本 框中输入欲生成的新变量的变量名 "Q20 分数"。单击【类型与标签】按钮,在弹出的对 话框中可以对新变量的类型和标签进行设置,这里不做设置。在【数字表达式】框中输入 新变量的数学表达式,这里输入"q20.1 + q20.2 + q20.3 + q20.4 + q20.5 + q20.6 + q20.7", 如图 2-48 所示。需要注意的是,尽量利用【计算变量】对话框中的小键盘编辑数学表达式, 如果读者要用外置键盘编辑,则要确保在英文状态下编辑数学表达式。如果仅仅对满足特 定条件的个案进行计算,则可以单击【如果】按钮,进入【选择个案】对话框,选择个案 的操作请参考"选择个案"的操作步骤,这里不再重复介绍。最后,单击【确定】按钮, 提交系统分析,则可以看到在数据文件中新生成了一个变量"Q20分数",如图 2-49 所示。

计算变量		×
計算变量 目标变量(D): 020分数 文型与标签(L) ✔ ID ✔ 1D ✔ 10 ✔ 10 ✔ 10 ✔ 10 ✔ 10 ✔ 10 ✔ 10 ✔ 10 ✔ 10 ✔ 10 ✔ 10 ✔ 10 ✔ 12 ♥ 122 ♥ 122 ♥ 122 ♥ 122 ♥ 122 ♥ 122 ♥ 122 ♥ 122 ♥ 122 ♥ 122 ♥ 123 ♥ 120	数字表达式(E):     q20.1 + q20.2 + q20.3 + q20.4 + q20.5 + q20.6 + q20.7]     (	
		Ф J

图 2-48 【计算变量】对话框

q20.4	q20.5	q20.6	q20.7	Q20分数
1	1	0	1	6.00
1	1	0	1	6.00
1	1	0	1	6.00
0	1	0	0	3.00
0	1	0	0	3.00

图 2-49 计算变量结果

# 2.3.6 变量值的重新编码

当我们需要将连续变量转化为等级变量,或者对变量取值进行重新修改或合并时,通 过变量值的重新编码就可以实现。SPSS 提供了两种变量数值重新编码的方式:一种是对原 始变量值直接进行重新编码并替换原数值的"重新编码为相同变量"方式;另一种是根据 原始变量的取值生成新变量来记录重新编码结果的"重新编码为不同变量"方式。

### 1. 重新编码为相同变量



打开本章数据"员工薪酬调查.sav"中变量 q3, "是"被编码为1, "否"被编码为2, 现在请把"是"重新编码为2, "否"重新编码为1。

**案例分析:**修改原变量的取值或合并部分取值为某个取值时,可采用重新编码完成 设置。

步骤 1: 打开本章数据"员工薪酬调查.sav",选择【转换】→【重新编码为相同变量】 命令,如图 2-50 所示。



图 2-50 重新编码为相同变量

步骤 2: 单击【重新编码为相同变量】进入其对话框。将"q3"变量选入【数字变量】 框中,如图 2-51 所示。

步骤 3: 单击【旧值和新值】进入其对话框,在左侧【旧值】选项组中的【值】文本框 中输入"1",在右侧【新值】选项组中的【值】文本框中输入"2",单击【添加】按钮, 表示将"旧值 1"改为"新值 2";依次类推,就可以将"旧值 2"改为"新值 1",如图 2-52 所示。全部旧值和新值转换关系建立之后,单击【继续】按钮回到主对话框,最后单 击【确定】按钮,提交系统分析,这时系统就将原始数据变量 q3 中的取值进行了转化,即 把原来的数据"1"变成了"2",把数据"2"变成了"1"。



图 2-51 【重新编码到相同的变量中】对话框

重新编码成相同变量:旧值和新值	X
<ul> <li>         ・田値和新値         <ul> <li>             日値和新値             </li> <li>             値(少):</li></ul></li></ul>	新信 ◎ 值(V): 1 ◎ 系统缺失(Y) IE -> 新(D): 1-> 2 添加(A) 副於(C) 副第(R)
继续	取消帮助

图 2-52 【重新编码成相同变量:旧值和新值】对话框

此外,当需要将分类或定序变量的某些取值水平合并为若干少数水平时,也可以采用 这种重新编码的方法。例如,将大学四个年级的大一和大二合并为低年级,大三和大四合 并为高年级,这个过程实际上就是将原数据的"1"和"2"重新编码为"1"(表示低年级), "3"和"4"重新编码为"2"(表示高年级),具体的操作方法与上述方法相同。如果需要 生成新的年级分组变量,则可以采用下面介绍的"重新编码为不同变量"的方法。

2. 重新编码为不同变量



打开本章数据"员工薪酬调查.sav",对 q4 划分年份组,其中 2011—2014 年为组 1, 2015 年为组 2,缺失值为组 0。

**案例分析:**在 SPSS 中,将连续变量转换成离散变量,按照某种一一对应的关系生成变量值,可以将新值赋给原变量,也可以生成一个新变量,通过"重新编码为不同变量"这一命令可完成这一任务。

步骤 1: 打开本章数据"员工薪酬调查.sav",选择【转换】→【重新编码为不同变量】 命令,如图 2-53 所示。



图 2-53 重新编码为不同变量

步骤 2: 单击【重新编码为不同变量】进入其对话框。将 q4 变量选入【数字变量→输 出变量】框中,在【输出变量】选项组中的【名称】文本框输入新变量名"T4",单击【更 改】按钮,原来的"q4->?"就会变成"q4->T4",即新老变量名间建立了对应关系,如 图 2-54 所示。



图 2-54 【重新编码为其他变量】对话框

步骤 3: 单击【旧值与新值】按钮进入其对话框,如图 2-55 所示。对话框左侧的【旧 值】选项组为原有变量的取值,右侧的【新值】选项组为新变量的赋值设定。两边设定完 毕后单击【添加】按钮,新旧变量间的对应编码规则就会被加入到右下方的规则列表框中。 这里在【旧值】选项组的【范围】框中设定 2011—2014 年为"新值 1",在【值】文本框 中设定 2015 年为"新值 2",在【系统缺失】处设定系统确实值为"新值 0",最后添加 结果如图 2-55 所示。上述重新编码过程既可以将连续变量转换成数值型或字符型离散变量, 也可以将数值型字符变量转换成数值型变量,只需选中图 2-55 右下角的【将数值字符串移 动为数值】复选框即可。单击【继续】按钮回到主对话框。

旧值	新值
◎ 值(⊻):	◎ 值(_):
	◎ 系统缺失(Y)
◎ 系统缺失( <u>S</u> )	◎ 复制旧值(P)
◎ 系統或用户缺失(U) ◎ 范围:	旧> 新(□):
	2015> 2
	SYSMIS> 0
到(1)	②加(A) 2011 thru 2014> 1
	更改( <u>C</u> )
◎ 范围,从最低到值( <u>G</u> ):	刪除(民)
◎ 范围,从值到最高(E):	
	■ 輸出变量为字符串(B) 宽度(W): 8
◎ 斫有其他值(0)	■ 将教值字符串移动为教值(M)("5'->5)

图 2-55 【重新编码到其他变量:旧值与新值】对话框

步骤 4: 所有操作完成后单击【确定】按钮,提交系统分析,输出结果如图 2-56 所示。 从图 2-56 中我们可以看出,新变量 T4 的取值有"1""2"和"0"三种取值,分别对应 q4 变量的取值范围,即"2011—2014""2015"和"缺失值"。

	s2	s3	q3	q4	T4
1	33	1	2	2014	1
2	34	1	2	2014	1
3	28	1	1	2015	2
4	40	2	1	2013	1
5	30	2	1	2014	1
6	55	2	1	2011	1
7	41	1	1	2013	1
8	43	1	1	2011	1
9	22	1	2		0

图 2-56 旧值与新值转换后的结果

小 结

在数据分析之前,我们需要先建立数据文件,要建立文件需要先设置变量;变量的设置在名称上需要符合多个条件,如不能重名,此外,还需要根据数据的特点设置数据类型、小数位以及设置变量标签和值标签等属性;数据的录入可以采用直接录入的方式,也可以采用导入外部 Excel、txt 等格式数据文件的方式,其中多选题的录入又可以分为多重二分法和多重分类法两种;数据文件建立后,为保证后期数据统计分析结果的准确性,我们还要对空行、超出特定范围的数值、重复个案以及不符合特定逻辑规则的个案进行处理;最后,我们可以对数据进行文件合并、计算变量、选择个案及重新编码等操作,为后期统计分析做准备。



1. 简述 SPSS 变量名的设置应注意的事项。

2. 简述变量标签和值标签的区别。

3. 将本章"合并数据 1.sav"和"合并数据 5.sav"合并。

4. 将本章"合并数据 3.sav"和"合并数据 6.sav"合并。

5. 打开本章数据"员工薪酬调查.sav",筛选出管理部门的女性,并为其建立一个新数据。

6. 表 2-3 是我国的一些经济指标(本章数据"国民经济核算.sav"),请根据以下要求对

该数据进行统计与分析。

(1) 计算出人均国内生产总值,在原数据上生成"人均国内生产总值"变量。

(2) 将三大产业增加值加总,在原数据上生成"三大产业增加值"变量。

(3) 对国内生产总值进行等级划分,在原始数据中生成"规模等级"新变量。其中 300 000 亿元以下为"小规模",编码为"1"; 300 000 亿~500 000 亿元为"中等规模",编码为 "2"; 500 000 亿元以上为"大规模",编码为"3"。

时间	国内生产总值 /亿元	人口数 /亿	第一产业增加值 /亿元	第二产业增加值 /亿元	第三产业增加值 /亿元
2014 年	643 974.00	13.64	58 343.50	277 571.80	308 058.60
2013 年	595 244.40	13.57	55 329.10	261 956.10	277 959.30
2012 年	540 367.40	13.51	50 902.30	244 643.30	244 821.90
2011年	489 300.60	13.44	46 163.10	227 038.80	216 098.60
2010年	413 030.30	13.38	39 362.60	191 629.80	182 038.00
2009年	349 081.40	13.31	34 161.80	160 171.70	154 747.90
2008年	319 515.50	13.25	32 753.20	149 956.60	136 805.80
2007 年	270 232.30	13.18	27 788.00	126 633.60	115 810.70
2006年	219 438.50	13.11	23 317.00	104 361.80	91 759.70
2005 年	187 318.90	13.04	21 806.70	88 084.40	77 427.80

表 2-3 国民经济核算

注:数据来源于国家统计局。