

第一讲：SPSS 数据预处理操作

教学目的：能应用 SPSS 软件进行：数据文件的建立、数据录入与修改、变量计算及转换

教学内容：1) SPSS 概述
2) SPSS 数据文件的建立
3) SPSS 数据整理与转换

教学重点：SPSS 数据文件的建立

教学难点：SPSS 数据转换

教学时间：1 学时

SPSS 概述

1.1 SPSS 简介

SPSS 是英文 Statistical Package for the Social Science（社会科学统计软件包）的缩写。SPSS 名为社会科学统计软件包，这是为了强调其在社会科学应用的一面(因为社会科学研究中的许多现象都是随机的，要使用统计学来进行研究)，而实际上广泛应用于经济学、社会学、生物学、教育学、心理学、医学以及体育、工业、农业、林业、商业和金融等各个领域。

SPSS 的基本功能包括数据管理、统计分析、图表分析、输出管理等等。SPSS 统计分析过程包括描述性统计、均值比较、一般线性模型、相关分析、回归分析、对数线性模型、聚类分析、数据简化、生存分析、时间序列分析、多重响应等几大类，每类中又分好几个统计过程，比如回归分析中又分线性回归分析、曲线估计、Logistic 回归、Probit 回归、加权估计、两阶段最小二乘法、非线性回归等多个统计过程，而且每个过程中又允许用户选择不同的方法及参数。SPSS 也有专门的绘图系统，可以根据数据绘制各种图形。

1.2 SPSS 的运行环境

硬件环境：能运行Windows 95/NT/2000 或以上版本的微机。

软件环境：SPSS 能在中英文Windows 9X、Windows NT 4.0、Windows2000 及更高版本的Windows 操作系统上运行。

1.3 SPSS 的进入和退出

进入 Windows 后，逐一选取开始程序 SPSS for Windows—SPSS 11.0 for Windows，单击 SPSS 11.0 for Windows 后，即可进入 SPSS。首先是一个对话框如图 1.1，显示最近打开的数据文件、结果文件和程序命令文件。可选取你要打开的文件，按 OK 按钮打开。如不喜欢这个对话框，可选取对话框下方的 Don't show this dialog in the future 选项则以后不会再显示。

要退出 SPSS，在任一 SPSS 窗口下逐一选取 File——Exit，或在 SPSS Data Editor 窗口单击右上角关闭按钮。若生成了新文件或对以前文件进行了修改，退出前会提醒你对其进行保存。

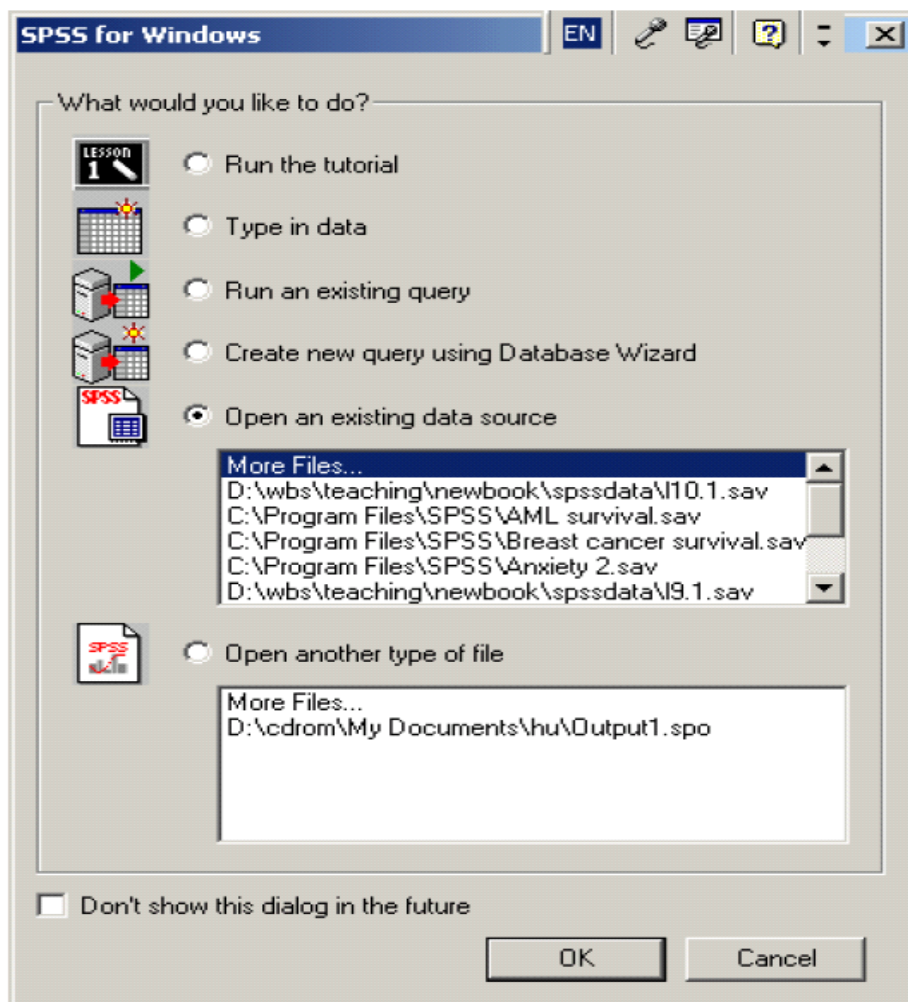


图 1.1 SPSS 进入对话框

1.4 SPSS 窗口及其功能

SPSS 有数据编辑窗口 SPSS Data Editor、结果窗口 SPSS Viewer 和语句窗口 Syntax。这里只介绍数据编辑窗口和结果窗口。

1.4.1 数据编辑窗口

启动 SPSS 后，首先进入数据编辑窗口 SPSS Data Editor，如图 1.2 的 Data View 窗口，可在此窗口下录入数据、修改数据等。

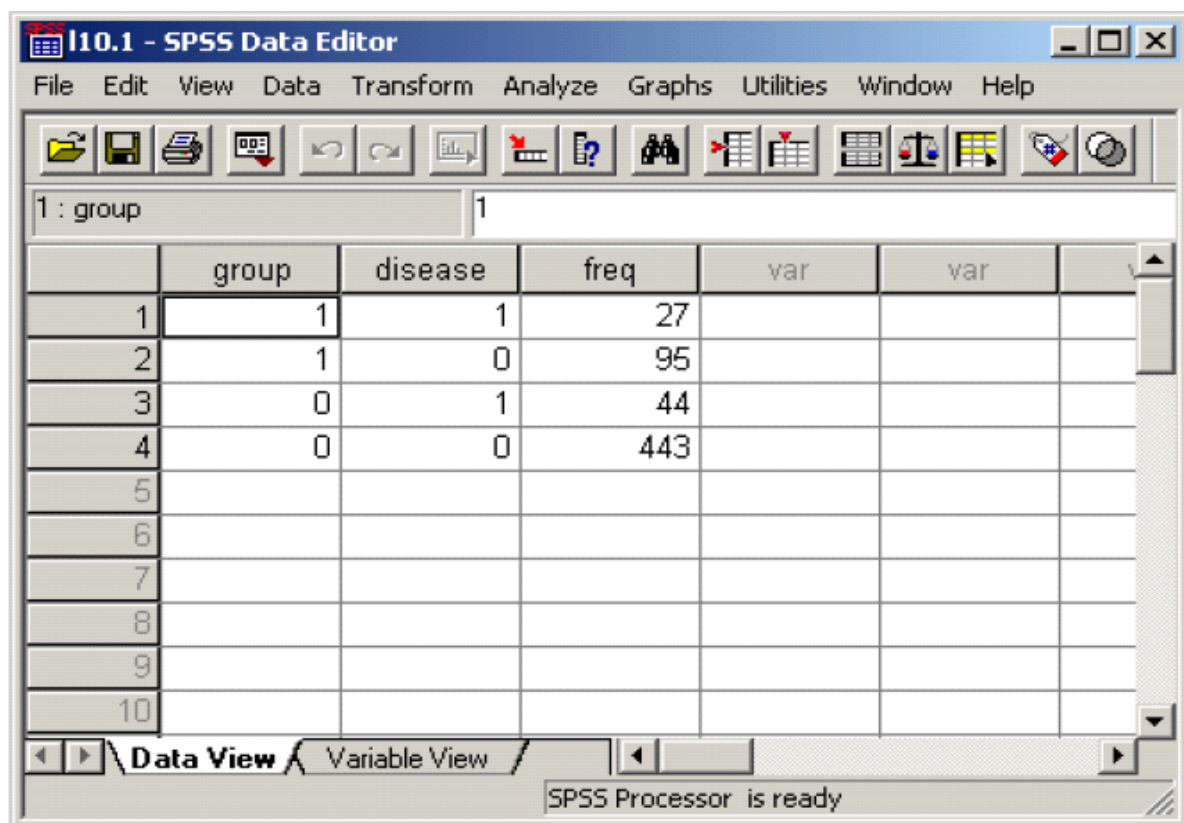


图 1.2 Data View 窗口

点击下方 Variable View 按钮可进入变量定义窗口，如图 1.3。在此窗口下可定义、修改变量或察看变量的设置情况。在下图中可看到变量的名字 (Name)、宽度 (Width)、类型 (Type)、小数位数 (Decimals)、标签 (Label) 等。

SPSS 的几个窗口菜单大致相同其中的命令也基本相同。菜单下的工具按钮均为常用的命令。将鼠标移至按钮附近，则弹出该按钮的解释，以后就不再说明。下面将简介，各命令的具体作用请参考实验指导教材。

● File 菜单中常用的命令为 New (创建新文件)、Open (打开已有文件)、Save (保存)、Save As (另存为)、Close (关闭窗口，SPSS Data Editor 窗口下无此命令，该窗口关闭则退出 SPSS)、Recently used Data (最近打开的数据)、Recently used Files (最近打开的文件)、Exit (退出) 等。

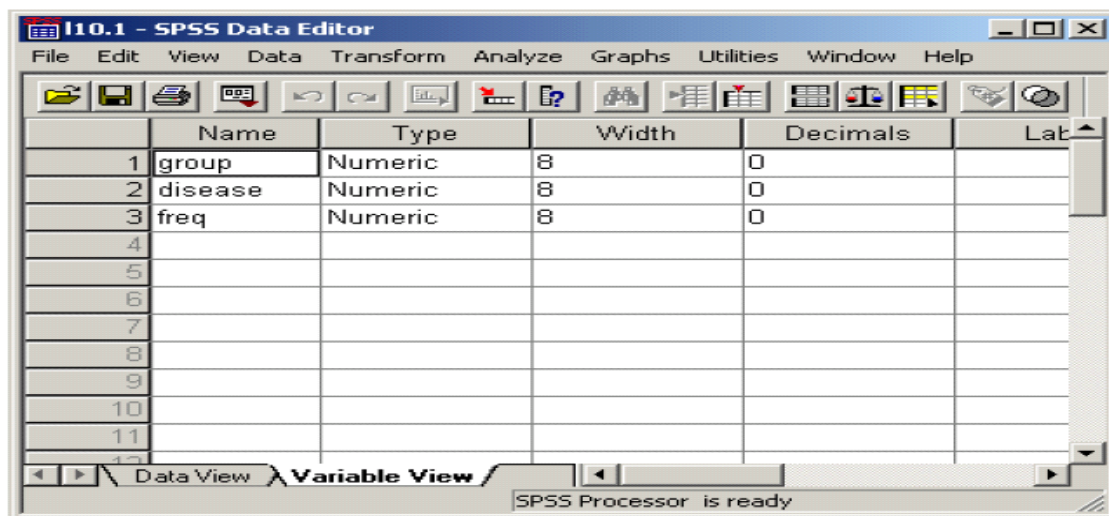


图 1.3 Variable View 窗口

●Edit 菜单下的常用命令有 Undo（撤销上次操作）、Redo（重复上次操作）、Cut（剪切）、Copy（复制）、Paste（粘贴）、Clear（清除）、Find（查找）、Options（选项）、Options（可对 SPSS 的一些设置进行调整）等。

●View 菜单中的命令可对窗口的设置进行调整, 有 Status Bar, Tool Bars, Fonts, Grid lines, Value Labels。

●Data 菜单中是数据进行编辑整理的命令, 包括 Define Dates, Insert Variable, Insert Case, Go to Case, Sort Cases, Transpose, Merge Files, Aggregate 等

●Transform 中命令有: Compute; Random Number Seed; Count; Recode; Categorize Variable; 等

●Utilities 菜单中命令有: Variables; File Info; Define Sets; Use Sets; Run Script; Menu Editor; Analyze; Graph 等

菜单下的工具按钮均为常用的命令。将鼠标移至按钮附近, 则弹出该按钮的解释, 在此不再一一说明。

1.4.2 结果窗口

该窗口中显示 SPSS 处理的输出信息。如处理成功, 则显示处理结果; 如处理失败, 则显示错误信息。

结果窗口可分为左右两个窗口。左窗口为大纲窗口, 其内容即结果的目录; 右窗口为内容窗口, 显示具体结果。如点击大纲窗口的项目, 则内容窗口会显示相应内容, 与资源管理器操作类似, 如图 1.4。

SPSS 结果一般由以下几个部分组成: 题头 (Heading, 即处理的名称)、标题 (Title)、记录 (Notes, 包括数据文件名、缺失值处理、操作所用时间等, 默认选项为隐藏, 只有双击大纲窗口中该项目才显示出来)、结果具体内容 (文本、图形等)。单击题头 (对准题头左边黄色的盒状图标), 则可选中上述的整个处理的结果; 双击题头时, 可使整个结果折叠或打开。单击题头前的 “+” 或 “—” 符号, 也可达到同样效果。

双击大纲窗口中某个项目, 则可使相应内容隐藏或显示, 同时大纲窗口中该项目名称前图标由一本打开的书变为合上的书或相反。

如要改变结果中表格或图形的位置, 可在大纲窗口或结果窗口单击该项目, 将它拖到指定位置即可。在此窗口中还可对结果进行编辑如选择删除复制等。

该窗口菜单同数据编辑窗口类似, 新增加项目如图所示: File; Edit; View 等

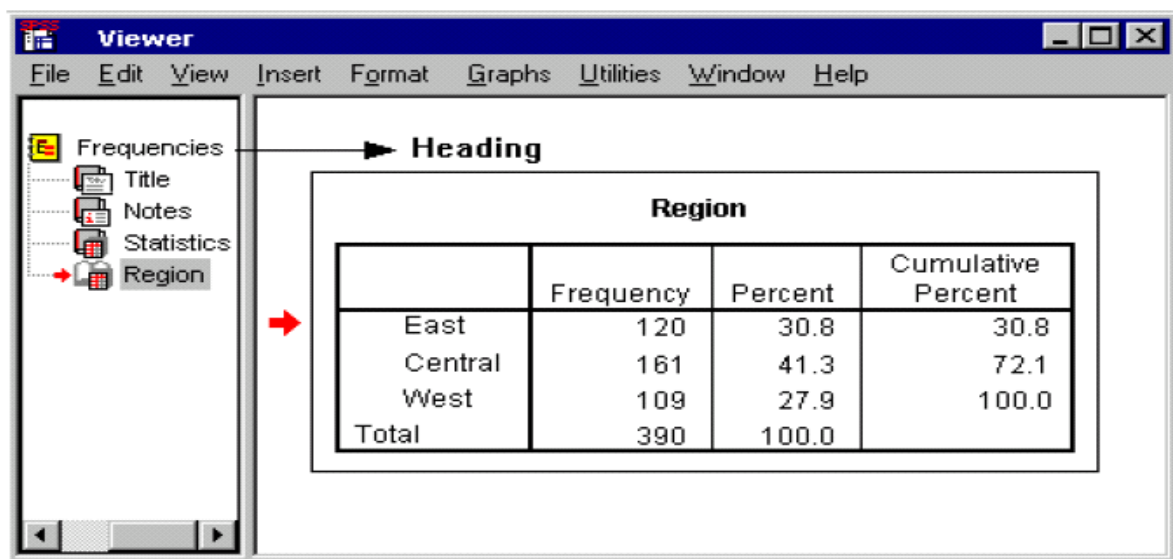


图 1.4 结果窗口 Viewer

SPSS 数据文件的建立

如上一章所述，建立 SPSS 数据文件有两种方法：一是在 SPSS 中直接输入，二是从其他数据文件转入。重点介绍第一种方法，它类似于在 FOXPRO 建立数据库。首先，在 SPSS Variable View 窗口中建立数据文件结构，如定义变量类型、宽度、标签等；然后在 SPSS Data View 窗口输入数据，并保存为“.sav”格式的 SPSS 数据文件。通常情况下，调查问卷的数据文件建立与录入都是采用第一种方法。在问卷数据录入前，还必须对问卷选项进行编码。

2.1 问卷设计及数据编码

2.1.1 问卷

一份调查问卷的结果通常表现为文字型和数字型两大类，其中文字型包括二项选择问题、多项选择问题、排序问题、开放式问题、连线问题等调查结果，而数字型包括百分比、绝对数等。对于数字型，可直接录入数据；而文字型则需要进行事前或事后编码，且不同类型的调查结果有不同的编码表现。此外，无论是数字型还是文字型，都有可能碰到调查数据缺省或不应该有的情况。这同样需要事先编码。

下面将以为了解大学生购书情况进行调查而获得的一份问卷为例，介绍几种常见问卷数据编码。

2.1.2 数据编码

1、单项选择题的编码

由于单项选择题的答案是唯一的，所以对各答案依次编码为 1、2、3…。如问卷中的第 1 题有四个答案，可依次编码“1”代表答案“300 元以下”，“2”代表答案“300—500 元”，“3”代表答案“500—700 元”，“4”代表答案“700 元以上”。于是，被访问者选择的任何答案都可以用 1，或 2，或 3，或 4 来表示。

2、多项选择题的编码

问卷中的第 5 题是多项选择题，共有 6 个答案项。编码时，可将 6 个答案看作 6 个变量，用“1”表示选择了该项答案，用“0”表示没有选择该项答案。若第 5 题答案顺序是从左到右、一行一行的排列，则 012 号问卷结果的编码是“0 0 1 1 1 0”。此外，该题的第 6 个答案项没有唯一的表现，应该视为开放性问题的编码（详细解释见后面）。

3、排序题的编码

对于排序题的编码，其处理方法是：将每个序位当作一个变量，而需要排序的那些选项作为每个变量的取值。如问卷中第 3 题有六个选项要排序，将“购书时考虑的第一因素”作为第一个变量，记为“Q301”；其取值可为“内容”、“价格”、“包装”、“实用性”、“知名作家的作品”、“热门畅销书”中的任何一个，可分别编码为“1”、“2”、“3”、“4”、“5”、“6”。同理，可将“购书时考虑的第二因素”、“购书时考虑的第三因素”等依次作为第二、第三变量等，记为“Q302”、“Q303”等；而每个变量的取值都与 Q301 相同，同样也编码为 1、2、3、4、5、6。于是，012 问卷第 3 题排序结果可转化为 6 个变量的取值依次为：2，4，1，5，6，3。

4、开放式问题的编码

对于开放式问题，如第 5 题的“其他”选项，第 7、8、9 题，有的答案是数字型，可直接录入，但有的答案则需要归类编码。其中第 5 题的“其他”选项和第 7 题的第一个空，由于事先不知道有多少类答案，需事后抽查部分调查问卷结果统计后才能初步确定与编码。这就是“事后编码”，它类似于二项选择问题的编码。至于第 9 题的“性别”、“专业”、“年级”等问题的所有可能答案，在事先都容易确定，可以直接编码录入数据。这就是“事前编码”，如“性别”有两个选项——男或女，可用“1”表示“男”，“0”表示“女”，也类似于二项选择问题的编码。

1. 您的月收入大概有多少?

- ☒ 300 元以下 ☐ 300—500 元
☐ 500—700 元 ☐ 700 元以上

2. 您购书的主要目的是:

- ☒ 学习、考试需要 ☐ 了解时尚、热点
☐ 休闲娱乐 ☐ 个人爱好及收藏

3. 您购书时主要考虑的因素依次是: (请在选项前横线上标出序号)

- 3 内 容 1 价 格
6 包 装 2 实用性
4 知名作家的作品 5 热门畅销书

4. 您较能接受的书的价格为:

- ☒ 20 元以下 ☐ 20—50 元
☐ 50 元以上 ☐ 需要, 则无所谓

5. 您购书的地点一般是: (可多选)

- ☐ 新华书店 ☐ 校内书店
☒ 学校附近书摊 ☒ 菜园坝书刊市场
☒ 商家在校园内的图书展销 ☐ 其他 跳蚤市场

6. 您对商家来校园内进行图书展销的看法是:

- ☐ 既方便又实惠, 很好
☒ 无所谓, 视其内容、价格而定
☐ 没兴趣, 宁愿到正规书店购买

7. 您最近一次购书的时间是 两个月前; 花费为 5.60 元;

8. 您平均每个月的购书支出占月收入的比例为 2%;

9. 您的性别 女, 专业 经济学, 年级 大二, 电话 62650029。

5. 缺失值的编码

问卷中, 若遇到被调查者不回答的一些选项, 则应该当作缺失值。缺失值又称为用户缺失值 (User Missing Value)。如第 8 题“您平均每个月的购书支出占月收入的比例”有缺失值, 可编码为 99%。又如, 第 7 题“花费”有缺失值, 可编码为 9999.99。如果可以选择的项目有 9 项, 但被调查者正好选择了第 9 个选项, 则以“0”表示缺失值。当然, 缺失值也可用研究者自己能够识别的其他数字来表示, 如“0”或“9”为用户缺失值。

用户缺失值与系统缺失值 (System Missing) 的含义不同。系统缺失值主要是指计算机默认的缺失方式, 如果在输入数据时空缺了某些数据或输入了非法的字符, 计算机就把其界定为缺失值, 这时的数据标记为“.”, 而用户界定的缺失值则不会在数据显示时出现“.”。

6. “不适用情况”的编码

当碰到被调查者不适用的选项时, 被调查者不需要对这些题进行回答, 这种情况下, 研究者可以用“8”、“98”等值来赋值表示“不适用”。若答案正好是 8, 可以以 0 替补。

2.2 在 SPSS 数据窗口直接输入数据

进入 SPSS 系统时, 系统就已经生成了一个空数据文件, 即空白的数据管理界面 (见图 1.2)。在输入

数据之前，必须先在 Variable View 窗口下定义变量。通常的做法是：把问卷中的问题定义成变量，把答案项编码定义成取值，并定义变量的类型、数据项宽度、标签、缺失值（或叫缺省值）等。然后，在 Data View 窗口下输入数据，并存盘为 “*.sav” 文件就行了。

2.2.1 定义变量

在 Variable View 窗口下，从列顶端可看到 10 种功能选项，分别是：Name, Type, Width, Decimals, Label, Values, Missing, Columns, Align 和 Measure 等变量属性。见图 2.1。

●Name，即变量名。SPSS 变量命名原则如下：

- 1) SPSS 变量的变量名由不多于 8 个字符组成；
- 2) 首字符是字母，其后可为字母或数字，或除 “?”，“!” 和 “*” 以外的字符。但应该注意不能以下划线 “_” 和圆点 “.” 作为变量名的最后一个字符；
- 3) 变量名不能与 SPSS 保留字相同。SPSS 的保留字有 all, and, by, eq, ge, gt, le, lt, ne, not, or, to, with, crosstabs, thru 等。
- 4) SPSS 中虽不区别大小写字符，但程序中的命令和关键词要用大写字母，表示系统内定；变量名等宜用小写字母，表示可人为指定。
- 5) 可以用中文做变量名，但最好不用，因为涉及一个兼容性问题。

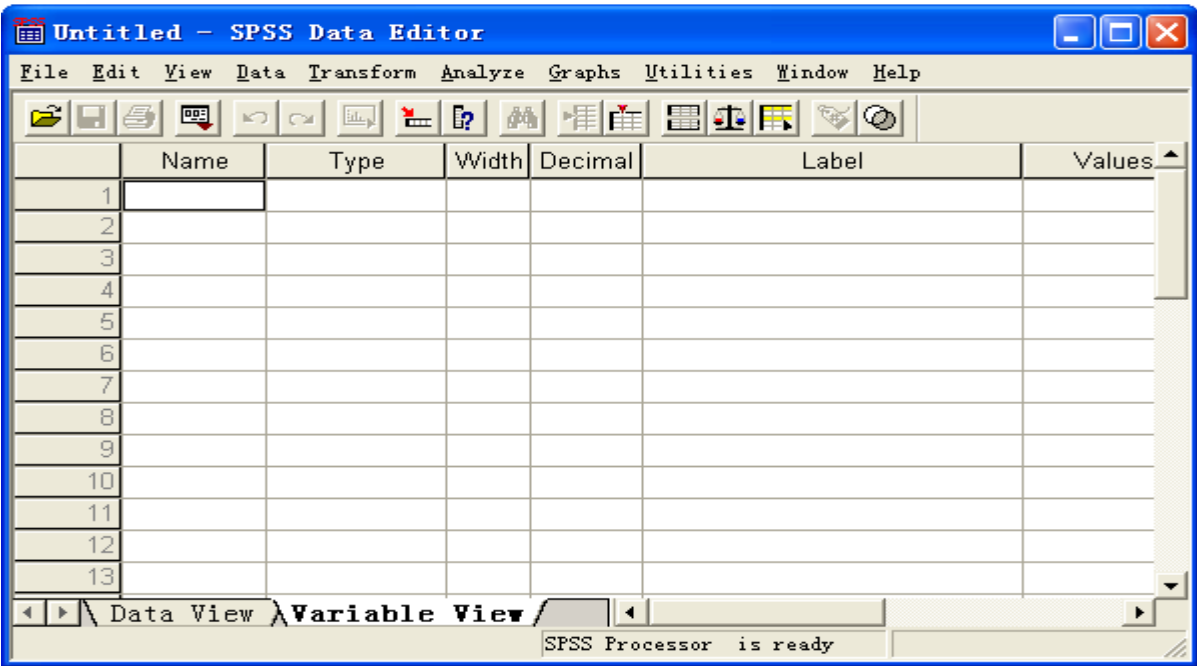


图2.1 定义变量

●Type：变量类型。

有 8 种变量类型可供选择，包括 Numeric（数值型变量），Comma（带逗号的数值型变量），Dot（带圆点的数值型变量），Scientific notation（科学计数法），Date（日期型变量），Dollar（货币型变量），Custom currency（自定义型变量），String（字符型变量）等。其中，常用的有三种 Numeric, String 和 Date；系统默认的变量类型为标准数值型变量（Numeric）。实践中，应尽量采用数值型变量，以方便以后的统计分析。

Width：变量的长度。系统默认长度为 8。

Decimals：小数位数。若为数值型变量，必须指定小数位数。系统默认小数位数为 2。

Label：变量标签，是对变量的进一步说明。变量较多时，需采用变量标签对变量的含义加以解释，

以免混淆。在数据统计分析过程中，会在变量名相对应的位置上显示该变量的标签，有利于分析结果得出结论。

Values: 标签变量值。标签是对变量的可能的取值所附加的进一步说明。无论是问卷结果显示是数字型还是文字型，只要答案是分类的（也称分类变量），如问卷中第 9 题“性别”，编码后都要定义其取值的标签。具体操作：单击行中 Values 格，再单击格中的按钮，即可定义值标签，如图 2.2。

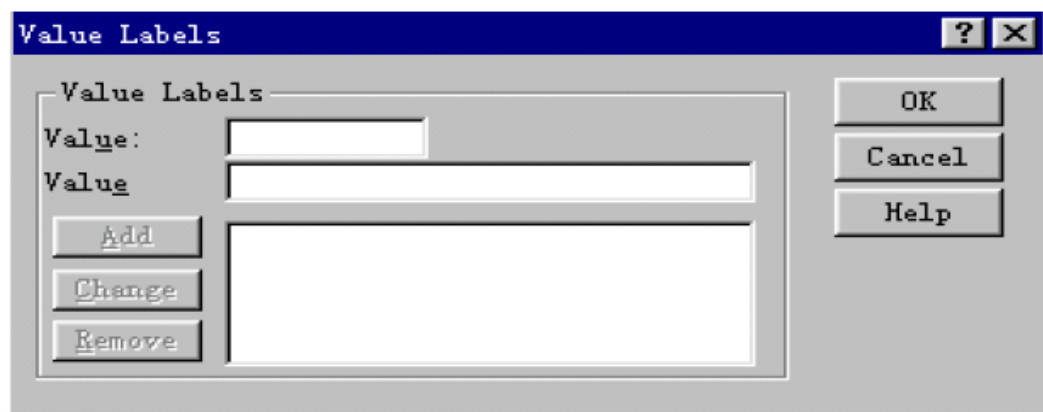


图 2.2 值标签的定义

Missing: 缺失值的定义。可指定缺失值。

Columns: 列宽度。

Align: 对齐方式，分为左对齐、右对齐、中间对齐。

Measure: 变量的度量类型。Scale 为定量变量，Ordinal 为等级变量，Nominal 为定性变量。

以问卷中第 1 题为例，可定义变量名称 Name 为“Q1”，数值型变量 Numeric，宽度 Width 默认为 8，小数位数 Decimals 为 0，变量标签 Label 为“被调查者的月收入”，标签值 Values 为“1=300 元以下”，“2=300—500 元”，“3=500—700 元”，“4=700 元以上”，其他则均选择系统默认。此外，定义问卷中的变量时，必须定义问卷编号变量，便于差错和统计分析。也就是，录入问卷调查结果之前，应先录入问卷编号，如上节问卷编号为“12”。

如果有许多个变量的类型相同，可以先定义一个变量，然后把该变量的定义信息复制给新变量。具体操作为：先定义一个变量，在该变量的行号上单击右键，弹出如图 2.3 (A) 所示的快捷菜单，选择 Copy；然后用鼠标右键选择多行，弹出如图 2.3 (B) 所示的快捷菜单，选择 Paste；再把自动产生的新变量名称（如 Var0001、Var0002、Var0003、……）改为所要的变量名称。

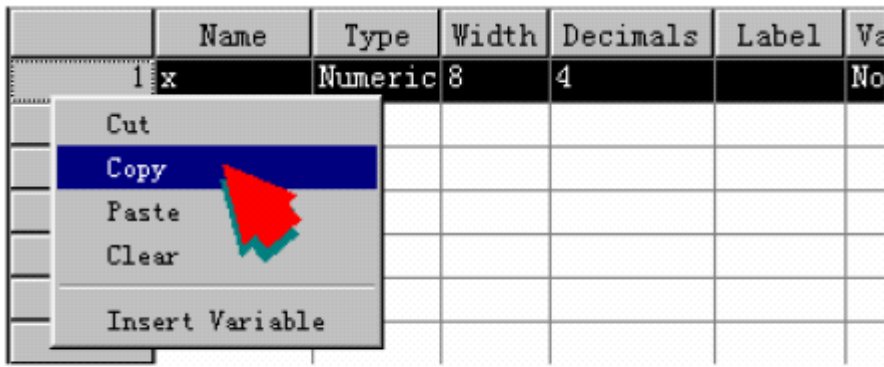
2.2.2 数据的直接输入

定义好了变量就可以开始输入数据。如果样本不大，变量不多，可以直接在 SPSS 中 Data View 窗口输入数据。输入方法就如同在 Excel 中输入一样，每一个变量占一列，每一行代表一个被观测个体的记录或一份问卷（Case），数据就录在行列交叉处。

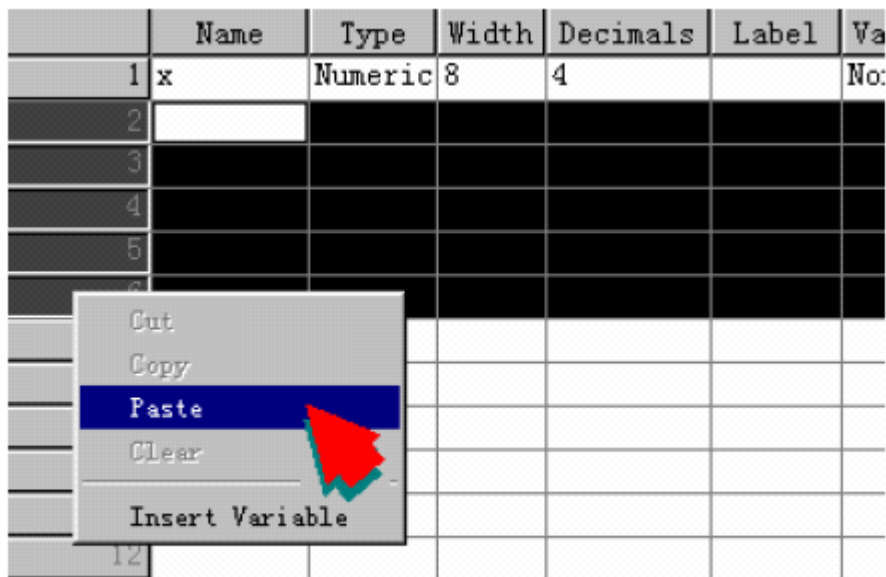
和其他常用统计软件相比，SPSS 数据界面最大的优势就是支持鼠标的拖放操作，以及拷贝粘贴等命令，下面的数据输入技巧就是对这些功能的利用。

1、连续多个相同值的输入

如前面 group 变量有连续多个 1，如果直接输入，可以在第一格内输入 1 并回车，然后回到刚才的单元格并单击右键，选择 copy，最后用拖放方式选中所有应输入 1 的单元格，单击右键并选择 paste，所有选中的单元格就会都被刚才拷贝的 1 填充。



(A) 复制



(B) 粘贴

图2.3 复制变量定义信息

2、将 EXCEL 数据直接引入 SPSS

若已有 Excel 数据文件，而且数据量较少，可直接打开原数据，用拷贝粘贴的方法将数据引入 SPSS：先在 EXCEL 中选中所有的数据（不包括变量名），然后选择拷贝命令；然后切换到 SPSS，最好使行 1 列 1 单元格成为当前单元格，然后执行粘贴命令，数据就会全部转入 SPSS，再修改相应的变量类型、宽度、小数位数等即可。输入完毕，从 File 菜单上选取 Save As 命令将其保存。SPSS 数据文件扩展名为 .SAV。单击保存类型下拉菜单，可选择保存为其他格式的数据文件，如 Excel 文件、dBase 文件等。如图 2.4 所示。

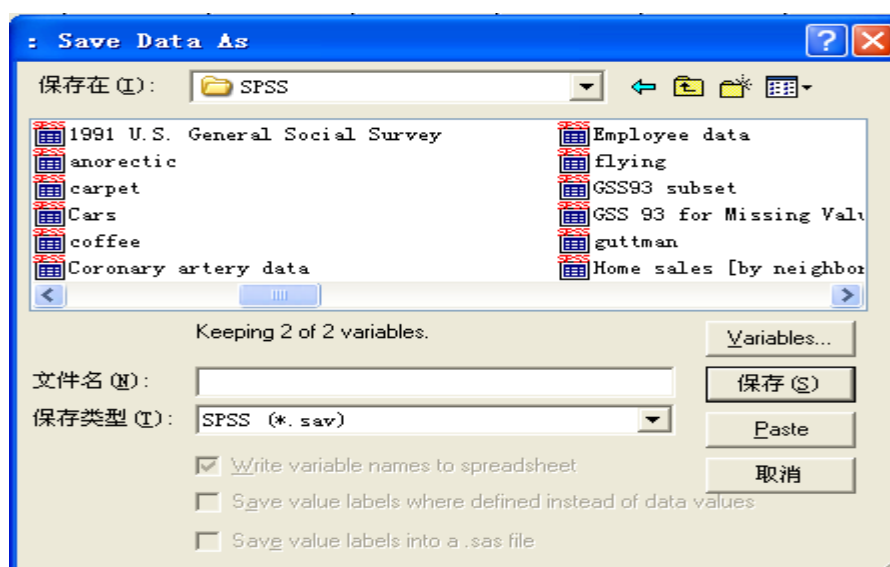


图 2.4 Save As 对话框

2.3 从其他文件转换成 SPSS 数据文件

●数据读入方法

若数据文件已以其他格式存在，如 EXCEL 格式、LOTUS 格式、dBase 格式、TEXT 格式等，可以在 SPSS Data Editor 窗口直接读入。


选择菜单 File→Open→Data 或直接单击快捷工具栏上的“”按钮，系统就会弹出 Open File 对话框，单击“文件类型”列表框，在里面能看到直接打开的数据文件格式，见表 2.2。选择所需的文件类型，然后选中需要打开的文件，SPSS 就会按你的要求打开你要使用的数据文件，并自动转换为数据 SPSS 格式。

表 2.2 SPSS 可读入的数据文件列表

SPSS (*.sav)	SPSS 数据文件（6.0~10.0 版）
SPSS/PC+ (*.sys)	SPSS 4.0 版数据文件
Systat (*.syd)	*.syd 格式的 Systat 数据文件
Systat (*.sys)	*.sys 格式的 Systat 数据文件
SPSS portable (*.por)	SPSS 便携格式的数据文件
EXCEL (*.xls)	EXCEL 数据文件（从 5.0 版~2000 版）
Lotus (*.w*)	Lotus 数据文件
SYLK (*.slk)	SYLK 数据文件
dBase (*.dbf)	dBase 系列数据文件，（从 dBase II~IV）
Text (*.txt)	纯文本格式的数据文件
data (*.dat)	纯文本格式的数据文件

●读入 EXCEL 数据文件的步骤

首先，选取 File→Open→Data，打开已保存的 SPSS 数据文件或其他类型的数据文件；然后单击文件类型，下拉菜单选择 Excel (*.xls)，如图 2.6。如 Excel 文件第一行为变量名，则选中 Read Variable Names from the first row of Data 选项，如图 2.7；并表名，如 sheet2 [A1: D5]。

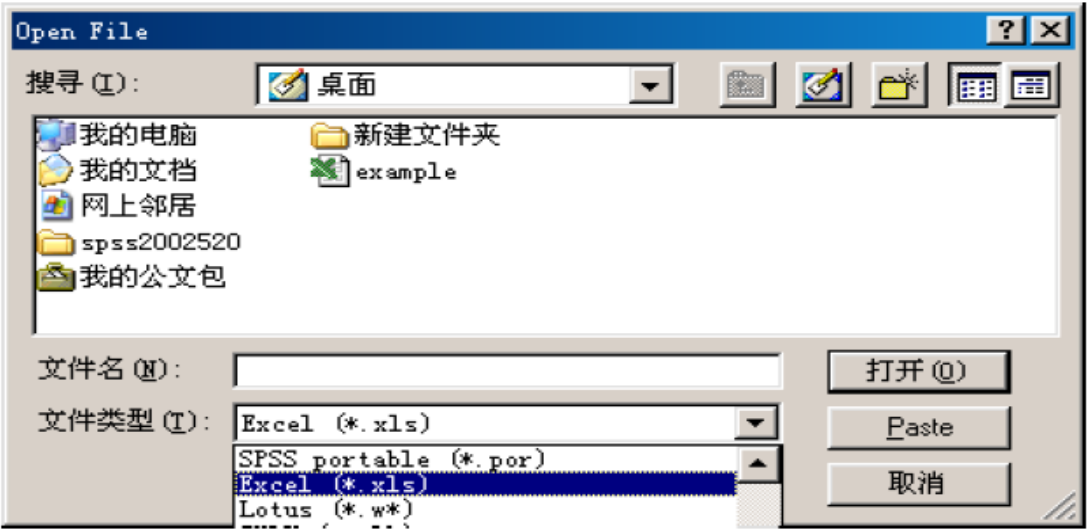


图 2.6 open file 对话框

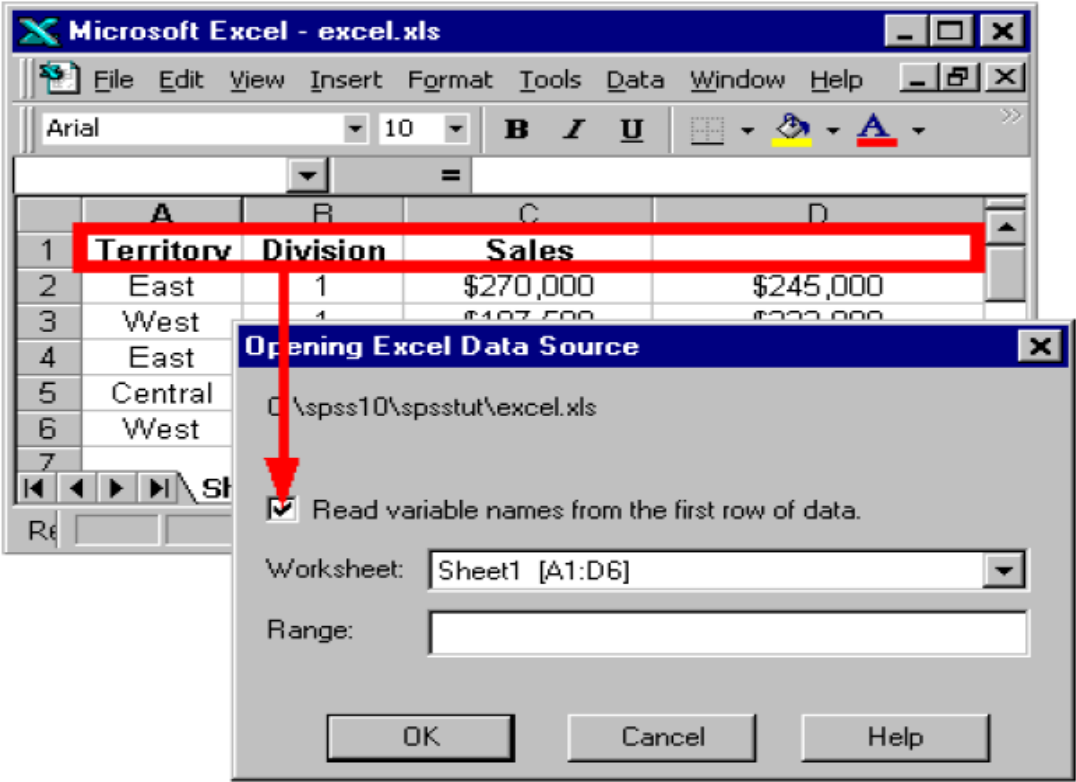


图 2.7 读入 EXCEL 变量名选项

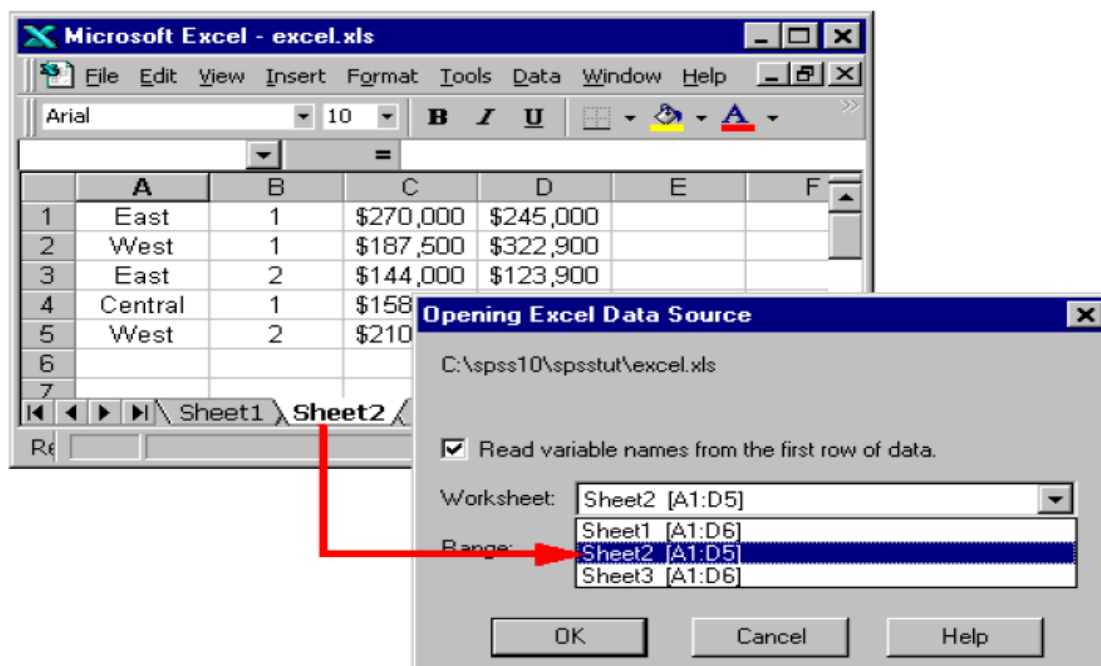


图 2.8 读入 EXCEL 数据表名的选择

SPSS 数据整理与转换

输入数据以后，需要先对数据进行一些整理，如根据数据分析目的，按变量分组、合并、加权、重新定义或计算新变量等，以为最终的统计分析做准备。这些功能集中在 Data 和 Transform 菜单项中，下面将以一个案例来介绍一些常用的功能。

3.1 一个案例

有研究者为了解重庆市南坪地区促销方式特征，进行一次“南坪商圈促销方式及效果调研”活动，所设计的调查问卷如表 3.1，收集了 312 份有效问卷，并建立了 SPSS 数据文件——111.sav。打开该数据文件后，图 3.1a 显示的是该文件的 Data View 窗口，而 b 图是 Variable View 窗口。

表 3.1 南坪商圈促销方式与效果调研调查问卷

1. 你从事的职业是
①销售类 (停止调查) ②非销售类
2. 你最近一个月内是否接受过类似调查:
①是 (停止调查) ②否
3. 你的性别: ①男 ②女
4. 你所属的年龄阶段是:
①25 岁以下 ②25—35 岁 ③35—45 岁 ④45 岁以上
5. 你的月收入属于下列哪个阶段:
①1000 元以下 ②1000—1500 元 ③1500—2000 元 ④2000 元以上
6. 你见过哪些促销方式: (多选)
①特价销售 ②免费品尝 ③买商品赠礼品 ④买满一定金额返代金券
⑤换购 ⑥广告促销 ⑦其他_____
7. 请将第 5 题的促销方式的代号进行排序: (按喜欢程度由强到弱)

8. 你最近一次参加的促销活动的促销方式是什么?
①特价销售 ②免费品尝 ③买商品赠礼品 ④买满一定金额返代金券
⑤换购 ⑥广告促销 ⑦其他_____
9. 你最近一次参加促销活动的消费是_____元?
10. 你对商场的促销活动的态度:
①反感 ②不感兴趣 ③喜欢
11. 遇到促销活动, 你会:
①一次买很多 ②不理睬 ③有需要就买 ④认为物美价廉就买
.....

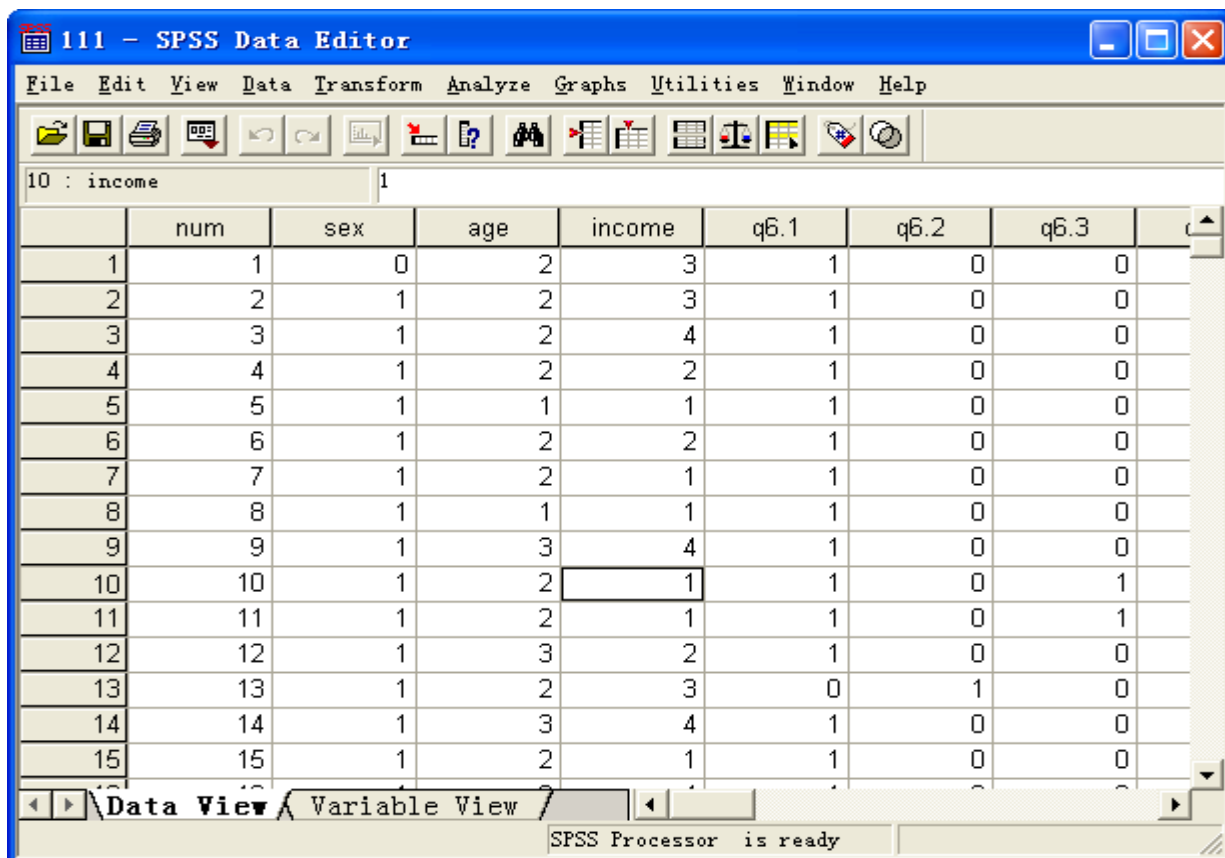


图 3.1a 111.sav 的 Data View 窗口

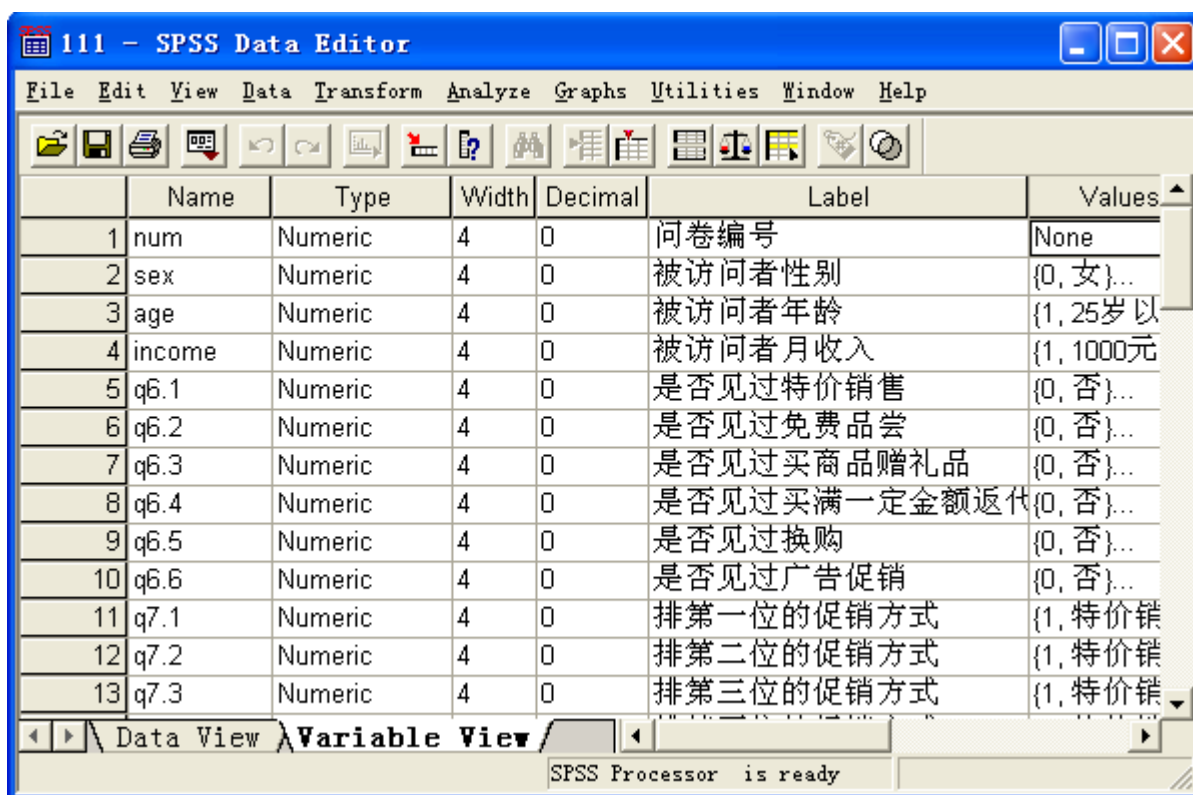


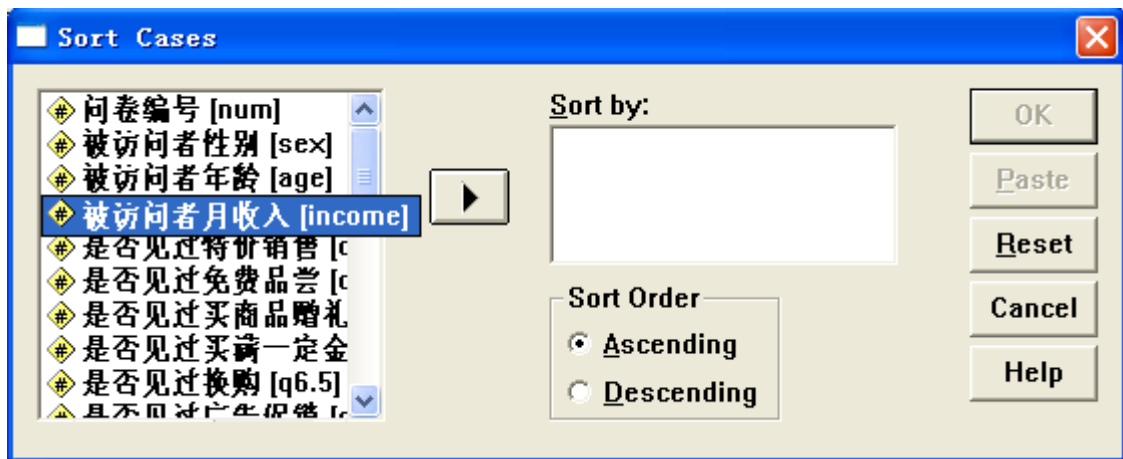
图 3.1b 111.sav 的 Variable View 窗口

3.2 数据整理

●排序 Sort Cases

选择菜单 Data→Sort Cases，系统弹出 Sort Cases 对话框。该对话框并不复杂，左方是可用于排序的变量，中间是一个箭头符号，右边上面是已确定的排序变量（在 Sort by 下的框里），右边下方则是 Sort Order 单选钮，有升序和降序两种选择。具体方法是：先可选择某一变量（如图 3.2a 中的分组变量 [income]），点击中间的箭头使选中变量到 Sort by 框里，然后选择升序 (Ascending) 或降序 (Descending) 单选钮。

若还要按其他变量排序，可继续上面的做法，只是排在前面的变量有排序优先权。如图 3.2a 中的可排序变量 income 和 age，若 Sort by 框中 income 在 age 的前面（见图 3.2b），排序时就会先满足 income 的排序要求，再满足 age 的排序要求。



a 选择 income 按升序排



b 再选择 age 排序

图 3.2 Sort Cases 对话框

●合并数据文件 Merge Files

若案例中收集到的有效问卷超过 1000 份，研究者就必须建立至少两个同变量结构的 SPSS 数据文件，假定除了 111.sav 以外，还有 222.sav；或者，因为变量个数太多，研究者建立了两个各含半数变量的 SPSS 数据文件——111.sav 和 333.sav。现要把上述两种情况下的两个数据文件分别合并起来，就需分别采用

Merge Files 中的 Add Cases 和 Add Variable，见图 3.3a。

1. 从外部数据文件中增加记录到当前数据文件中，称为纵向合并，用 Add Cases 对话框完成，相互合并的数据文件中应该有相同的变量。选择菜单 Data→Merge Files→Add Cases，系统首先弹出打开数据文件对话框。选中需要添加的数据文件并按 OK，系统才弹出 Add Cases 对话框（见图 3.3b），左侧显示的是新、老数据文件中不匹配的变量名（图 3.3b 中显示没有不匹配的变量名），右侧显示的是已经匹配的变量名。可以用 Rename 按钮对不匹配变量改名（先选中）或用鼠标强行匹配（即先按 Ctrl 键选中匹配的两个变量再单击 Pair 钮）。右下方的 Indicate case source as variable 复选框用于定义一个新变量以区分哪些记录是后来添加的。选择停当后单击 OK，该操作就完成了。

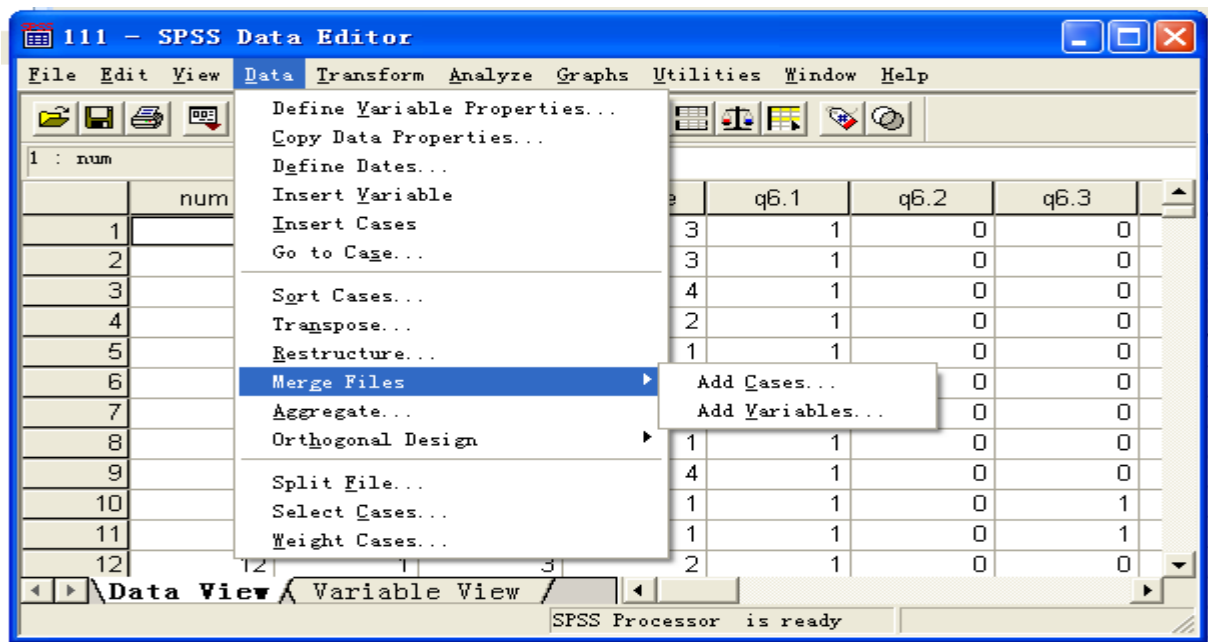


图 3.3a Merge Files 的两种方式

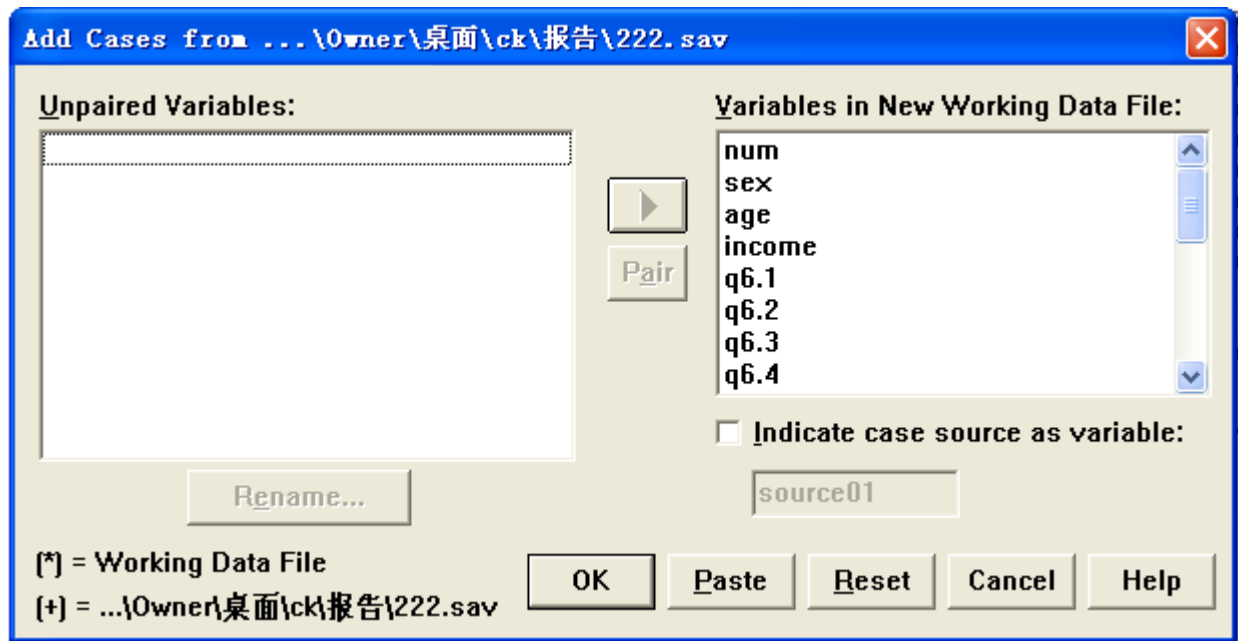


图 3.3b Add Cases 的对话框

2. 从外部数据文件增加变量到当前数据文件，称为横向合并，用 Add Variable 对话框完成，相互合并的数据文件中应包含同样的记录。选择菜单 Data→Merge Files→Add Variable 对话框，系统同样先弹出打开数据文件对话框，单击 OK 后弹出和前面相似的 Add Variable 对话框（见图 3.3c）。左侧是外部数据文件中多出来的变量（图 3.3c 中是 333.sav 多余的变量，与 111.sav 中的变量一致），右侧是合并后新文件包含的所有变量。其他选项，可按需选择即可。

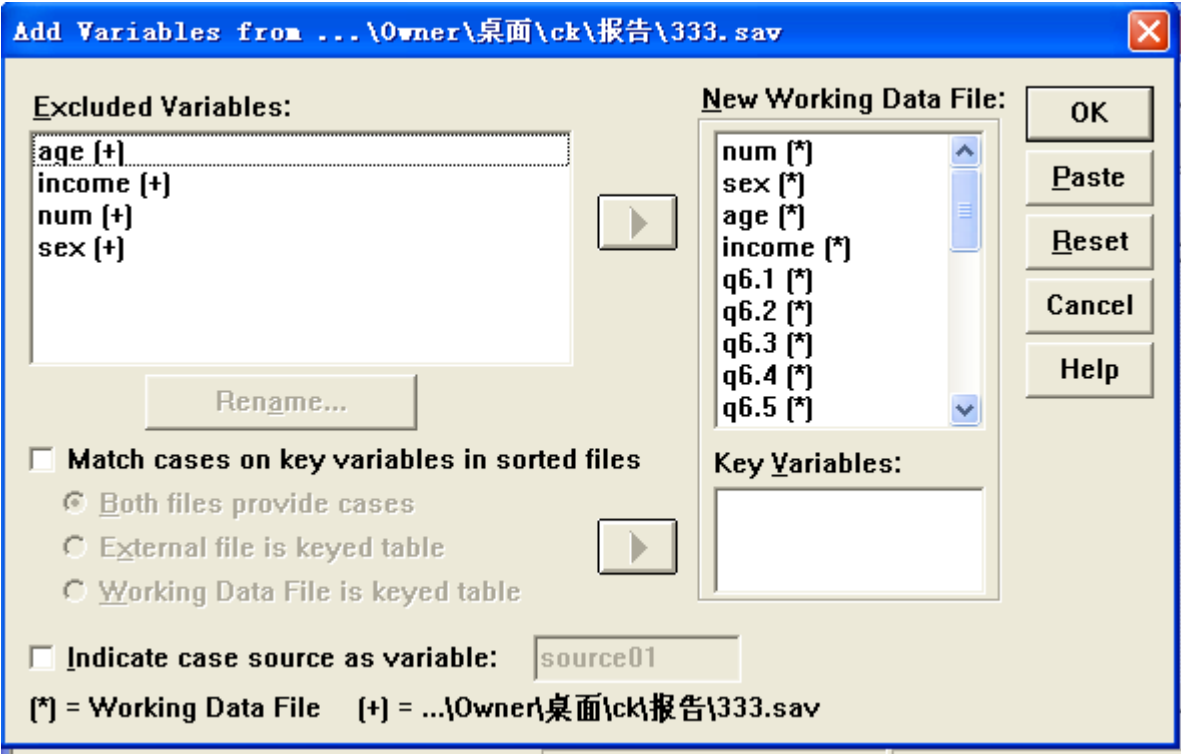
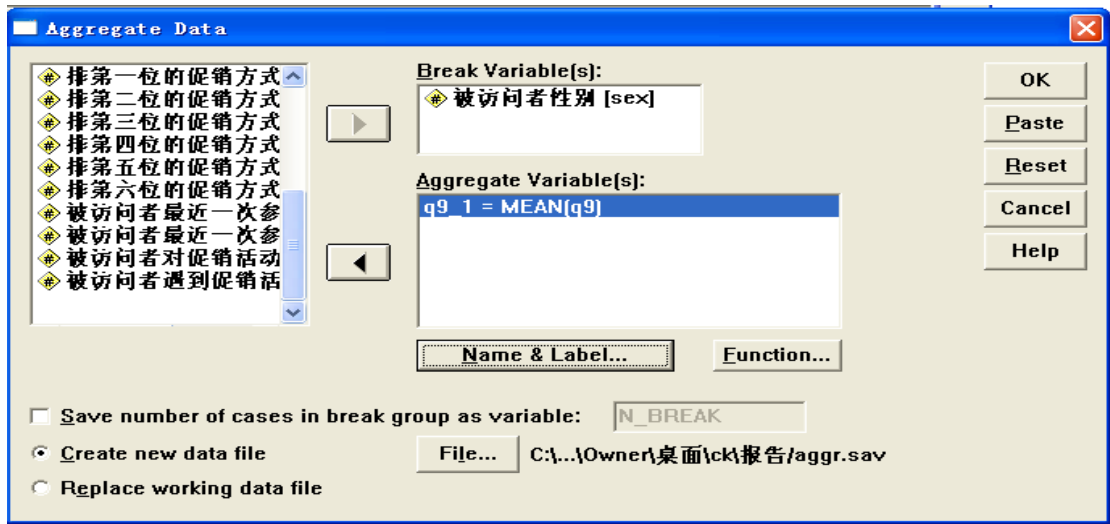


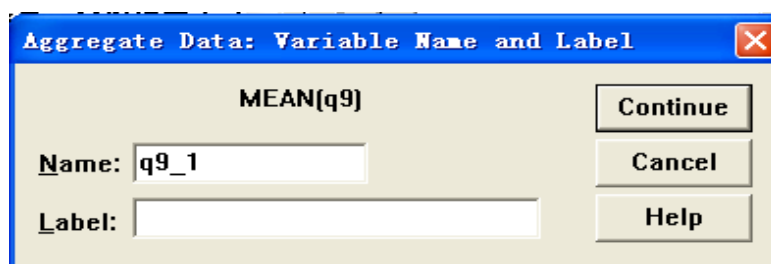
图 3.3c Add Variables 的对话框

●数据分类汇总 Aggregate

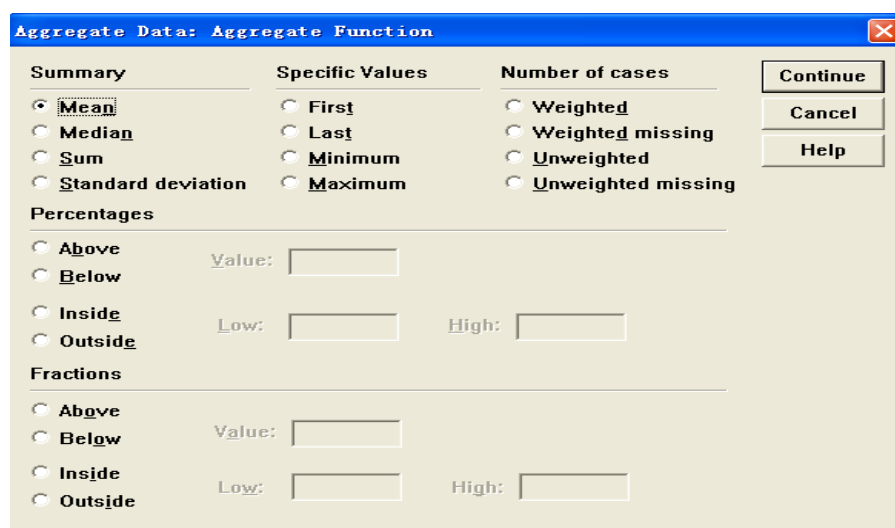
所谓分类汇总就是按指定的分类变量对观测值进行分组，对每组记录的各变量值求指定的描述统计量，结果可以存入新数据文件，也可以替换当前数据文件，见图 3.4。



a 选择 Aggregate Variables 的界面



b 定义新汇总变量的名称和标签的界面



c 定义汇总函数

图 3.4 Aggregate Data 对话框

图 3.4 中各个部分的含义如下：

- Break Variables 框：用于选择分组变量。这里选择了 sex；
- Aggregate Variables 框：用于选择被汇总的变量，可选择 q9（即被访问者最近一次参加促销活动的消费），如图 3.4a；
- Name&Label 钮：用于定义新产生的汇总变量的名称和标签，如图 3.4b；
- Function 钮：用于定义汇总函数，共有三组函数，以最常用的第一组为例，可选的函数有均值、同组的第一个观测值、最后一个观测值、同组记录数、标准差、最小值、和、最大值共 8 个，如图 3.4c；
- Save Number of cases in break group as variable 复选框：用于定义一个新变量以存储同组的记录数；
- Create new data file 单选钮：定义一个新文件以存储汇总的结果，右侧的 File 钮用于具体文件名的定义，默认文件名为 AGGR.sav；
- Replace working data file 单选钮：用汇总的结果替换原来的数据。

●数据文件分组 Split File

若要对不同群体作同类分析，可使用 Split File 将数据文件分割成不同的组来进行处理。选择 Data→Split Files，出现 Split Files 对话框，见图 3.5。图中显示，本次分析是按 age（即被访问者年龄）分组分析。

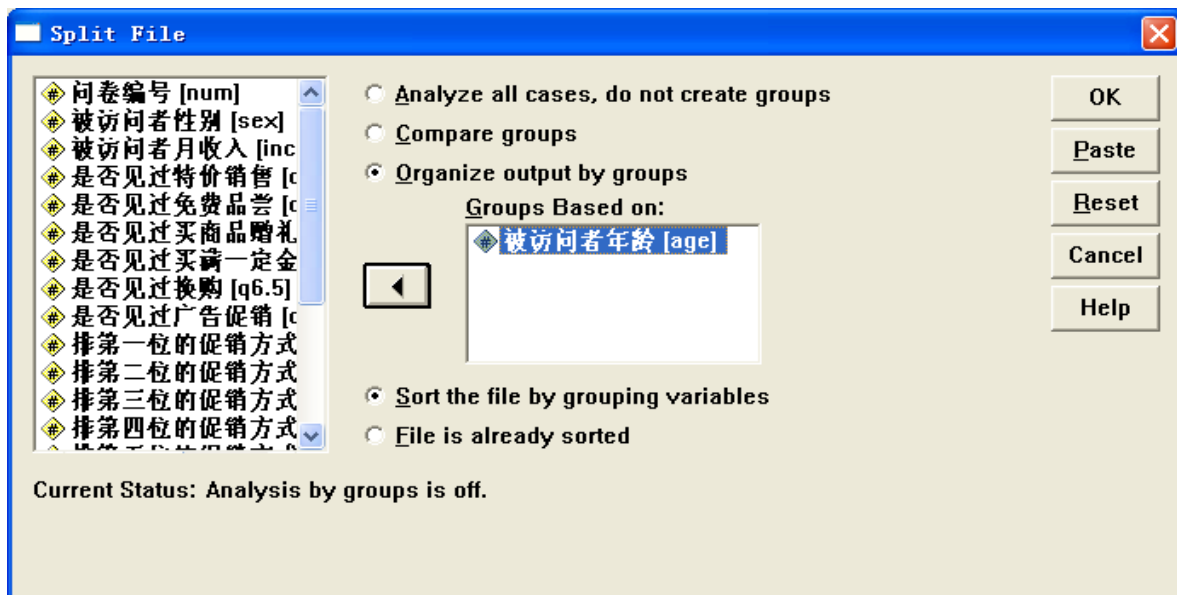


图 3.5 Split File 对话框

图 3.5 种各部分的含义如下：

- Analyze all cases 单选框：和下面的两个单选框为一组，选中本框不拆分文件；
- Compare groups 单选框：按所选变量拆分文件，各组分析结果紧挨在一起便于相互比较；
- Organize output by groups 单选框：按所选变量拆分文件，各组分析结果单独放置；
- Groups based on 框：用于选择拆分数据文件的变量；
- Sort the file by grouping variables 单选框：将数据按所用的拆分变量排序；
- File is already sorted 单选框：数据保持原状，不按所用的拆分变量排序。

●选择分析组 Select Cases

有些时候不需要分析全部的数据，而是按某种要求分析其中的一部分。比如，案例中的研究者只想对年龄（age）在 35—45、月收入（income）在 2000 元以上的女性对促销方式的看法与行为；或者只想对 200 个数据进行分析以了解大概情况，等等。这时需使用 Select Cases 对话框，来简化工作。

选择 Data→Select Cases，弹出对话框如图 3.6a 所示。其中：

- All cases 单选钮：和下面的 4 个单选钮为一组，选中它则分析所有的记录；
- If condition is satisfied 单选钮：只分析满足条件的记录；
- If 按钮：和 If 单选钮一起使用，单击后弹出 If 对话框；
- Random sample of cases 单选钮：从原数据中按某种条件抽样；
- Sample 按钮：和 Random 单选钮一起使用，可以设定按百分比抽取记录，或者精确设定从前若干个记录中抽取多少个记录；
- Based on time or case range 单选钮：基于记录序号来选择记录；
- Range 按钮：和 Based 单选钮一起使用，用于输入记录序号范围；
- Use filter variable 单选钮：使用筛选指示变量来选择记录，必需在下面选入一个筛选指示变量，该变量取值为非 0 的记录将被选中，进入以后的分析；
- Filtered 单选钮：和下面的 Deleted 单选钮为一组，表示未被选中的记录只是被隔离，这些记录的记录号会被加上斜杠以示区别；
- Deleted 单选钮：未被选中的记录将被删除，一般不要使用。

所以，先前的第一种情况应选择 If Condition is Satisfied，单击 If 后弹出 If 对话框，见图 3.6b。图中右上方框里的“sex=0 and age=3 and income=4”就是需要满足的条件，其中“sex=0”、“age=3”、

“income=4”代表女性、年龄 35—45 岁、月收入 2000 元以上；而中间的“and”表明三个条件都必须满足，若只需满足其中任意一项都行，则用“or”连接。此外，条件表达式也可从 Functions 框中选择。当填好条件表达式后，可单击 Continue 回到图 3.6a，再单击 ok 就完成了。

而先前的第二种情况，则应该选择 Based on time or case range，并单击 Range 弹出对话框，见图 3.6c。在 First case 框输入 1，在 End case 框输入 200，单击 Continue 回到图 3.6a，再单击 ok 就完成了。

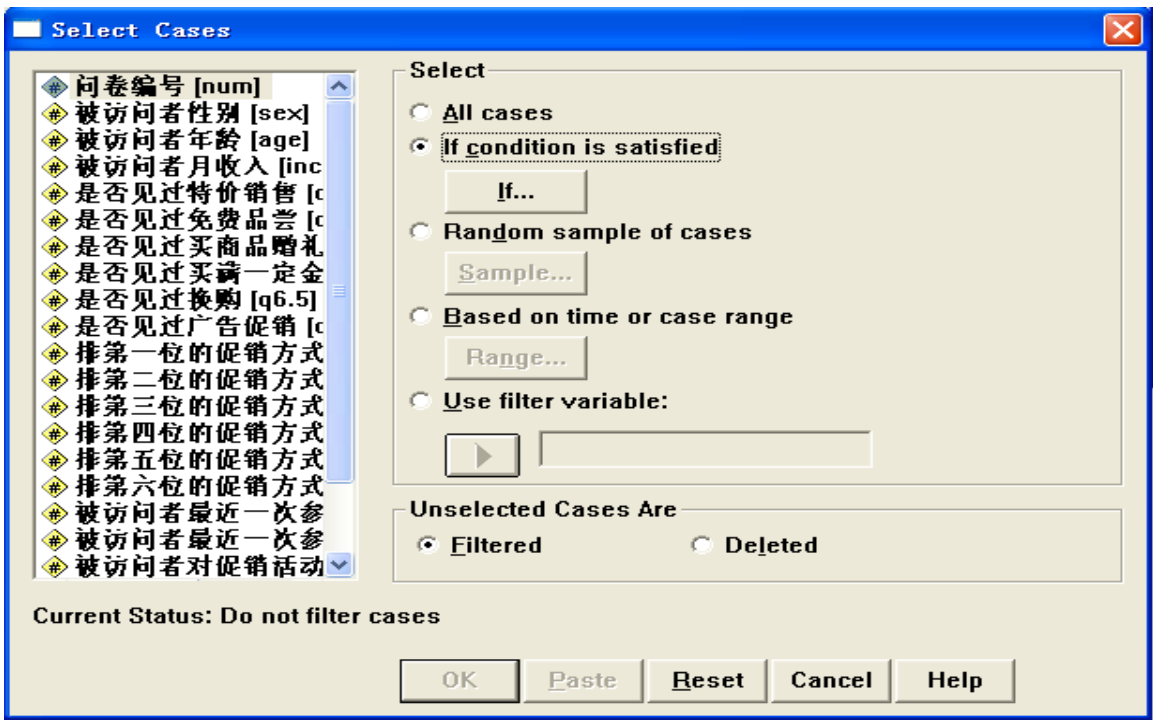


图 3.6a Select Cases 对话框

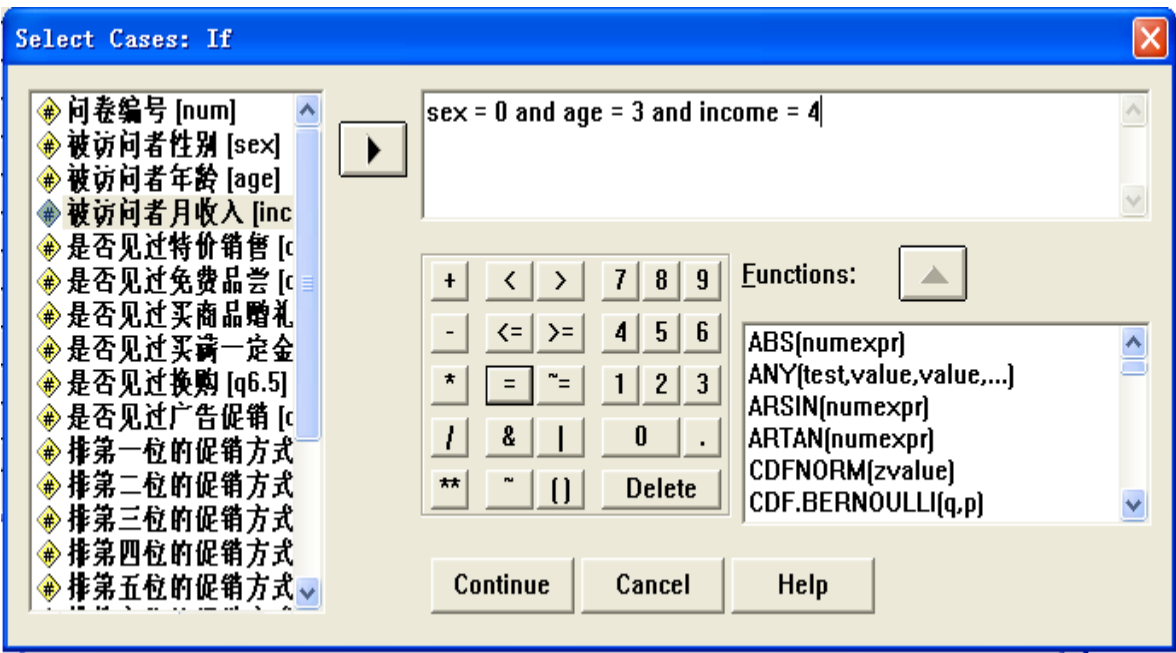


图 3.6b If 对话框

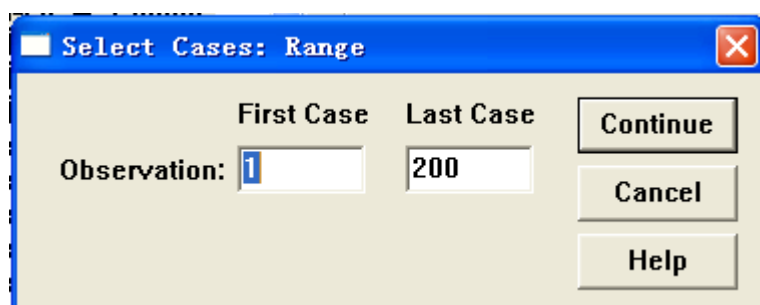


图 3.6c Range 对话框

当对数据集做出筛选后，所做的筛选将在以后的分析中一直有效，直到再次改变选择条件为止。同时在多数情况下，系统会自动产生一个名为 filter_\$ 的筛选指示变量，被选中的记录该变量取值为 1，反之则为 0。

● 加权分析组 Weight Cases

当两组个案 (Cases) 数目相差悬殊（比如男多女少）时，为了尽量反映“平等”的社会现象，有时需要给予适当的权重（比如对弱势群体赋予大的权重，反之亦然），以利于统计分析。在默认情况下，每一行就是一条记录，如果希望在分析过程中利用不同的变量对数据进行加权处理，就需要用到 Weight Cases 对话框，见图 3.7。该对话框的使用极为简单，界面上有两个单选钮，分别是不权重记录和用某变量权重记录，如果选择后者，则需要选中一个权重变量。图中显示以性别 sex 加权。

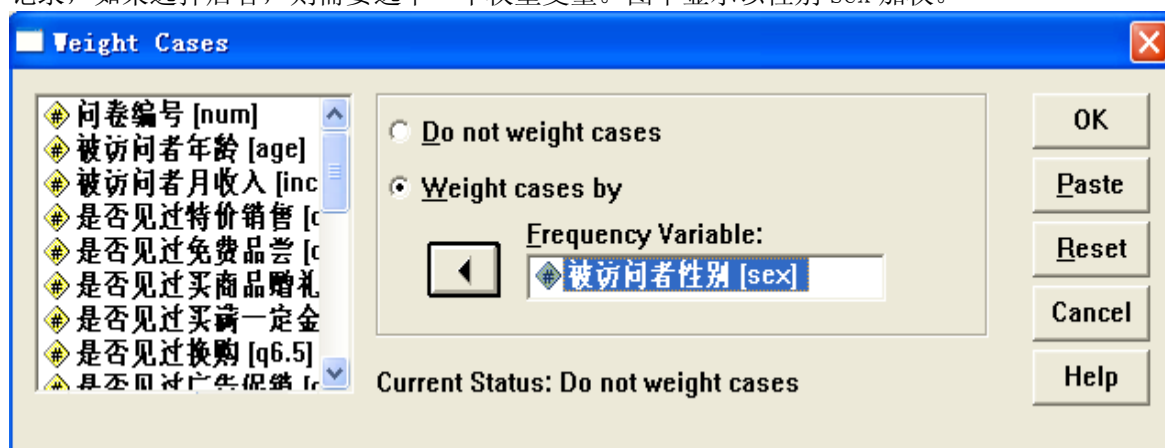


图 3.7 Weight Cases 对话框

● Data 菜单中的其余对话框

- Define dates 对话框：可以自动生成时间变量。
- Insert Variable 命令：在当前列插入新变量。
- Insert cases 命令：在当前行插入新记录。
- Goto cases 对话框：到达指定记录号的记录，该命令在记录数极多时（1000 条以上）非常有用。

3.3 数据转换

从调查得来的原始数据是非常重要的。但是有些原始数据没有归类，比如案例中的消费，不象年龄、月收入等只有几个取值，如果直接打印出频率表，则既浪费纸张又欠直观。为此，需要复制一份原始数据（即备份），再对复制后的原始数据重新编码，比如按消费段编码成几组，然后分组比较和打印显示。同样，对原始数据也可以进行重新计算。例如从消费 q9 计算出一个新变量——绝对消费差异。

类似这种数据变换，可用以下 4 条命令。

- (1) COMPUTE：对原始数据进行四则运算等，进而派生出新的变量。

(2) RECODE: 重新编码数据, 重新安排次序。

(3) COUNT: 创建一个新变量用以计算某些变量共同发生的频次(即计数)。

(4) IF: 如果表达式成立, 则可以有条件地创建一个新变量。

以上 4 条数据变换的命令分别又有 2 种执行方式: 其一是直接采用命令来实现。其二是采用 Transform 过程中的对话框来实现。这里只介绍使用对话框实现的操作。

● 计算 Compute

在案例中, 若研究者想分析每个被访问者最近一次参加促销活动的消费额与经验平均消费额 110 元的绝对差异, 定义变量名为 absq9。可采用如下操作:

点击 Transform→Compute, 进入 Compute Variable 对话框, 如图 3.8。Target variable 为新产生的变量, 可输入新变量名 absq9。该栏下方为一按钮 Type & Label, 单击该按钮后可选择新变量类型和定义变量标签。

Numeric Expression 为表达式, 有两种方法在该栏中写入表达式:

1) 直接写出。如对 SPSS 中函数比较熟悉, 可在此栏中直接填入 abs(q9-110), 即计算消费额 q9 与经验平均消费额 110 元的绝对差异, 然后单击 OK 就行了。

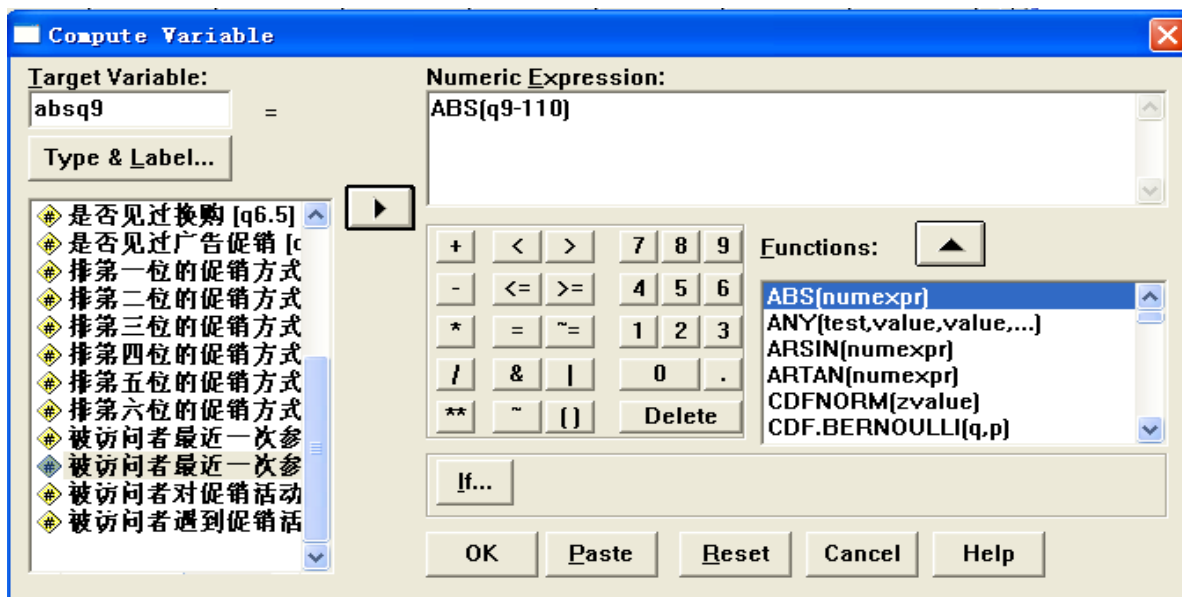


图 3.8 新变量的产生

2) 从对话框选取。Numeric Expression 栏下方为 Functions 栏, 该栏中列出了常用的函数, 可从该栏选取所要的函数。每个函数后的括号中注明了所需变量的类型或其他特殊的变量 (numexpr 表示数值变量, strexpr 表示字符变量)。选中后单击 Functions 上方的箭头, 进入 Numeric Expression 栏。该栏对话框左下方列出了数据库中的变量, 选中所要的变量, 单击其右上方的箭头将其选入框内, 或进入函数后的括号内。如上例, 首先从 Functions 栏选中 ABS(numexpr), 单击上方箭头进入 Numeric Expression, 此时 ABS 后的括号内为一问号。从左下方栏中选中变量 q9, 单击右上方箭头进入 ABS 后的括号内, 并减去 110。此时 Numeric Expression 内表达式为 abs(q9-110), 如图 3.8, 再单击 OK, 即产生新变量 absq9。

如不明白函数意义, 可在该函数名称处单击右键即弹出该函数的解释。此外, 可以单击 If 按钮, 设定计算新变量的 Cases 范围。比如说, 只对女性计算 absq9; 或者只对消费 q9 大于 50 的 Cases 计算 absq9; 等等。

● 计数 Count

计数 Count 是计算变量值中含有相同范围值的次数。如案例中, 研究者要计算第一喜欢“特价销售”

或“购商品赠礼品”促销方式的消费者人数，诸如此类，就要采用 Count。其具体操作步骤如下：

- 1、选择 Transform→Count，弹出 Count 对话框，如图 3.9a；
- 2、在 Target Variable 框中输入新变量名 no1q7，在 Target Label 框中输入新变量的标签；
- 3、从左侧的变量列表中选择 q7.1（即排第一位的促销方式）到 Numeric Variables 框内；
- 4、单击 Define Values 定义 q7.1 的取值为 1 或 3，见图 3.9b。分别在左侧的 Value 框中输入 1、3，并依次 Add（添加）到 Values to Count 框中；
- 5、单击 Continue 回到图 3.9a，再单击 ok 即可。

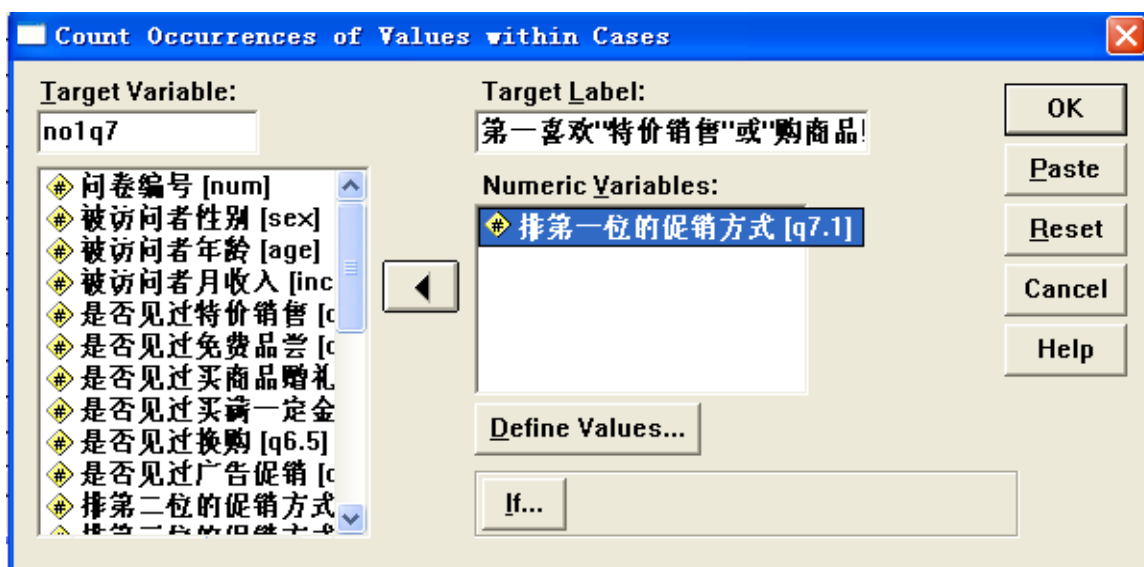


图 3.9a Count 对话框

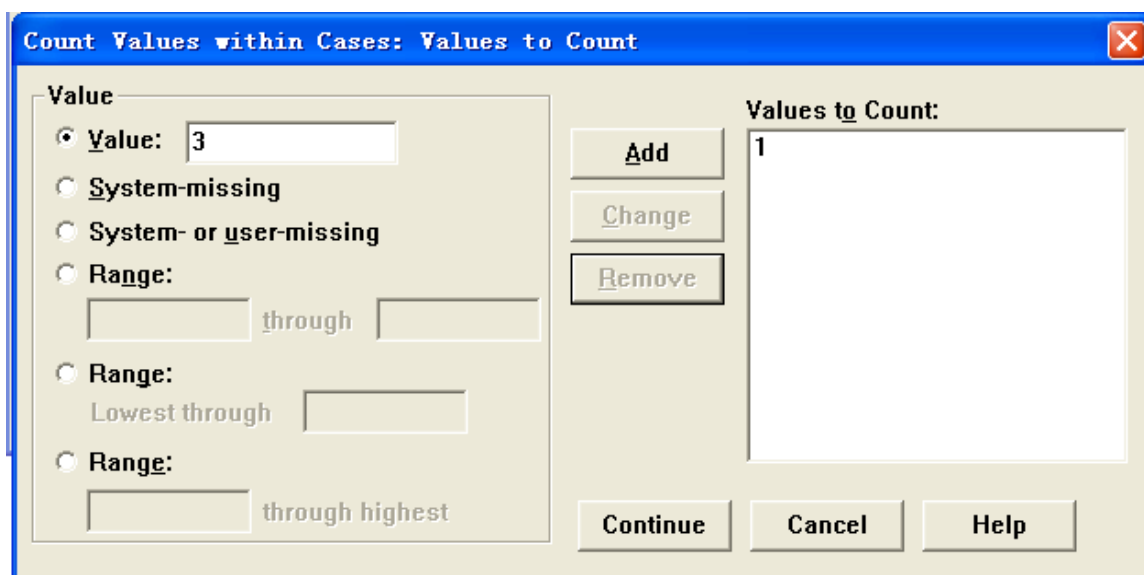


图 3.9b Define Values 对话框

●重新赋值 Recode

若研究者需要按消费 q9 分组分析，可采用 Recode 对 q9 重新赋值。Recode 对话框用于从原变量值按照某种一一对应的关系生成新变量值，可以将新值赋给原变量，也可以生成一个新变量。比如说将 q9 分成 5 组，即“50 元以下”、“50-100 元”、“100-150 元”、“150-250 元”、“250 元以上”，然后分别赋值为 1、

2、3、4、5。其具体操作如下：

1、选择 Transform→Record→Into Different Variables, Recode 对话框如图 3.10a。也可选择 Transform→Record→Into Same Variables, 但这样就替换掉了原始数据, 不利于其他统计分析, 所以常常选择 Into Different Variables。

2、从变量列表中选择 q9 到 Numeric Variable→Output 框中, 显示 q9→?;

3、在 Output Variable 下 Name 框中输入新变量名 groupq9, 并定义 label, 再单击 Change, 将原来的 “q9→?” 变成 “q9→groupq9”;

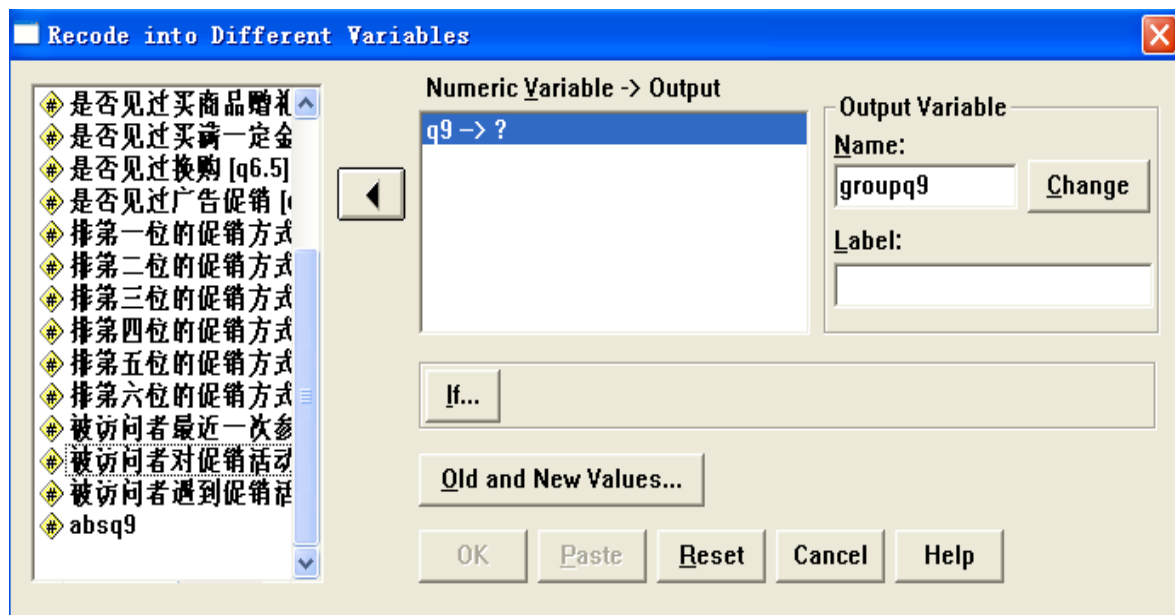


图 3.10a Recode 对话框

4、单击 Old and New Values, 系统弹出变量值定义对话框, 如图 3.10b。左侧是老数据 Old Value, 右侧是新赋值 New Value; 每指定一个或一组老数据后, 就要在 New Value 框中输入一个新值, 并单击 Add, 使老数据被重新赋予新值。其中左侧的 Old Value 中:

- Value: 当原始数据为离散, 且数据范围较小时, 可一个一个地对老数据赋予新值;
- Range + Through: 指定一组老数据, 在第一个框里输入下限, 第二个框里输入上限;
- Range + Lowest Through: 在右边的筐里输入一个上限, 可指定小于该上限的老数据组;
- Range + Through Highest: 在左边的筐里输入一个下限, 可指定大于该下限的老数据组;
- All other values: 指定所有剩余的、还没有被赋值的老数据为一个组。

前面要求重新赋值的 q9, 第一组 “50 以下” 应选择 Range + Lowest through, 在其右侧框中键入 50; 再在右上方的 Value 右侧框中键入对应的新变量值 1, 并单击下方的 Add 键, 此时 Old→New 框中就会加入 “Lowest thru 50→1”。按照类似的方法依次加入另四组数据转换, 最终 Old→New 框中共有 Lowest thru 50→0、50 thru 100→2、100 thru 150→3、150 thru 250→4、Else→5 等 5 条。

5、赋值完了以后, 单击 Continue 回到图 3.10a, 再单击 OK, 系统就会按要求生成新变量 groupq9。

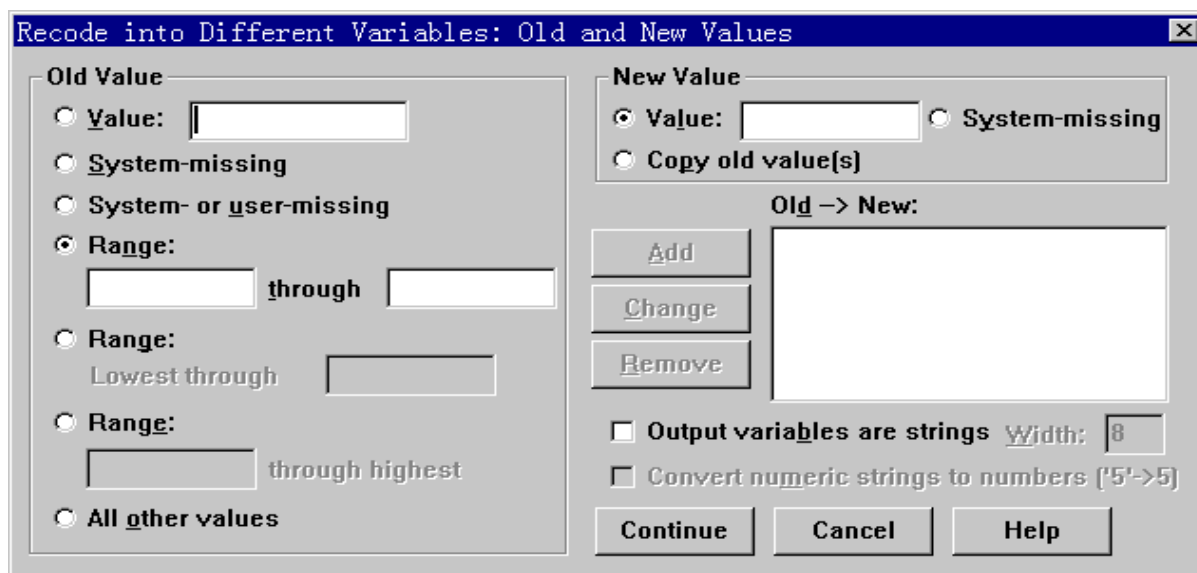


图 3.10b Recode 赋值对话框

●条件转换 If

案例中，若研究者要按 sex（性别）和 q10（被访问者对促销活动的态度）对被访问者分组，将“反感”的“男性”编入第 1 组，“反感”的“女性”编入第 2 组，“不感兴趣”的“男性”编入第 3 组，“不感兴趣”的“女性”编入第 4 组，“喜欢”的“男性”编入第 5 组，“喜欢”的“女性”编入第 6 组，或与此类似的有条件的变换，则可采用 IF 来转换。

If 并出现在 Transform 的下拉菜单中，而是在各种转换过程中出现，如前三种转换过程中都能看到 If 的身影。在 Computer 过程中，对 If 有过简单的解释，其他过程中 If 的作用也是类似的，这里就不再一一赘述了。

附：实验项目 1：SPSS 数据文件建立与数据预处理操作

实验项目	能应用 SPSS 软件进行：数据文件的建立、数据录入与修改、变量计算及转换
实验日期	
实验环境	SPSS for WINDOWS
实验内容	1、熟悉 SPSS 窗口菜单功能 2、根据给定的问卷，建立 SPSS 数据文件 3、录入数据 4、数据预处理操作
实验步骤	根据实验自己认真填写.
实验结论 (或实验体会)	1. 写出求解问题的主要结果。 2. 谈谈实验体会。
实验批改	