

## 四 为分析做好准备

学习要求：

- 1、 个案的排序与变量值的排秩
- 2、 合并文件
- 3、 选择个案
- 4、 数据文件的转置
- 5、 变量内容的重新编码

数据录入工作完成以后，就可以进行统计分析了。但由于有些统计分析方法对变量的属性，测量的层次、数据的分布等有些特殊要求。这就需要对数据文件进行整理和再加工。以使变量和数据能够符合统计分析的要求。

### 4.1 个案的排序

#### 4.1.1 个案排序的方法

虽然系统已经按录入的先后将所有的Cases 排队了。但是在进行数据处理时，有时需要按某个变量值重新排序，排序的功能可以通过下述过程来实现：

1. 打开个案排序对话框

执行下述操作：

Data→Sort Cases 打开对话框，如图4-1 所示。

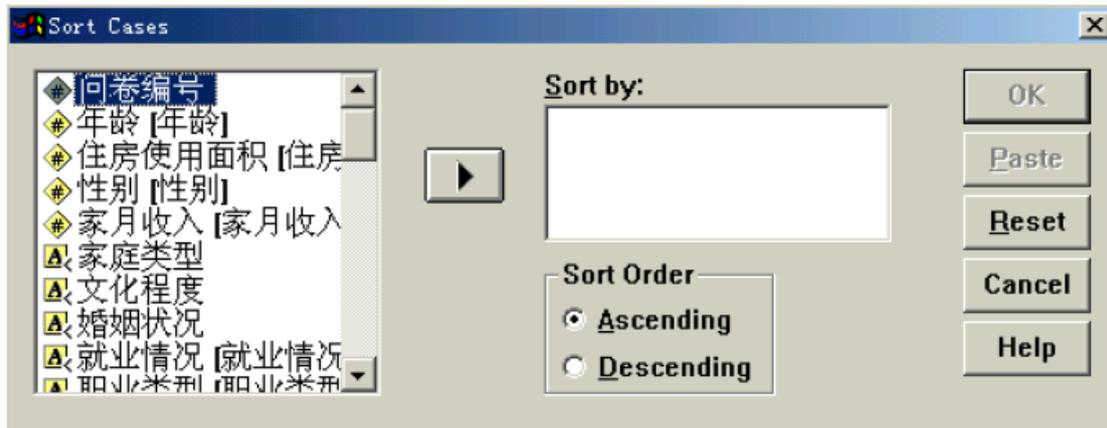


图4-1 cases 排队对话框

2. 选择排队变量

从左侧的源变量窗口中选择一个或多个变量，通过单击中间的箭头按钮，使之进入到Sort by 窗口中，这些变量将作为排队依据的变量。如果选择的是多个变量，系统先按选择的第一个变量排队，第一个变量值相等时，按第二个变量排队，以此类推。

3. 选择排队规则

Sort Order 是排队规则选项栏，栏中包括两个选项。

(1)Ascending 是按照升序的顺序排序。即第一个Case 的变量值是最小的。这是系统默认选项。

(2)Descending 是按照降序的顺序排序。即第一个Case 的变量值是最大的。

上述操作完成以后，单击“OK”按钮，提交运行。可以在数据编辑窗口看到重新排队的数据文件。

#### 4.1.2 个案排序的实验。

实验：“休闲调查”是按“问卷编号”排队的，现将其按“年龄”和“住房面积”两个变量重新排队。

## 4.2 变量值的排秩

有些分析在进行之前要求对变量求秩，秩是指变量排序后的顺序号。在很多非参数检验的方法中，都要对秩进行运算。

### 4.2.1 排秩的方法

求秩的工作可以由下述过程来实现。

#### 1. 打开个案排秩对话框

执行下述操作：Transform→rank cases 打开个案排秩对话框，如图4-2 所示。



图4-2 个案排秩对话框

#### 2. 选择排秩变量。

从左侧源变量窗口中选择一个或多个要排秩的变量进入Variable(s)窗口。

#### 3. 确定排秩的顺序

在个案排秩对话框中源变量窗口的下面有Assign Rank 1 to（把秩值1 分配给）选项栏。这是确定排秩顺序的选项栏。

(1) Smallest value 是将秩值1 分配给最小值。这是按照升序来排秩，也是系统默认选项。

(2) Largest value 是将秩值1 分配给最大值。这是按照降序来排秩

#### 4. 确定秩变量的类型

单击“Rank Types”按钮，打开对话框，如图4-3所示。

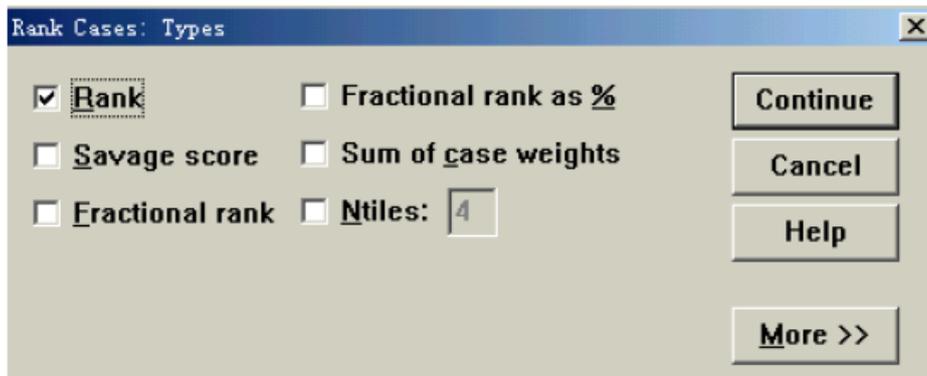


图4-3 秩变量类型对话框

对话框中的Rank 是系统默认选项。即将求得的秩在数据编辑窗口中建立一个新变量来保存。

其余几个选项是对原始的秩变量进行各种计算或变化后形成的新的秩变量的选项。对初学者用处不大，此处不作介绍。上述选项作完后，单击Continue 按钮返回个案排秩对话框。

5. 确定相同值的秩的取值方法。

单击Ties（同值）按钮进入Ties 对话框，打开对话框，如图4-4 所示。

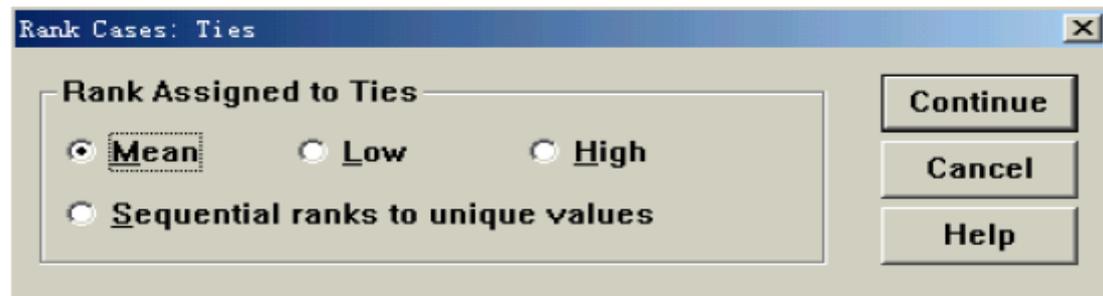


图4-4 相同值的秩的取值方法对话框

在该对话框中Rank Assigned to Ties 是对于相同数值的秩的分配方法选项栏。

- (1) Mean 是指相同值的秩取平均值。这是系统默认状态。
- (2) Low 是相同值的秩取最小值。
- (3) High 是相同值的秩取最大值。
- (4) Sequential ranks to unique values 是相同值的秩次取第一个出现的秩次值。

如果没有特殊要求，对后三个选项可以不作选择，采用系统默认选项即可。

6. 分组排秩的方法

如果想对变量的值进行分组排秩，如分男、女性别不同对住房面积进行分别排秩。则可以选择一个或多个定类变量进入By 下面的窗口中。系统将按By 变量分组排序。

上述操作完成以后，单击“OK”按钮，提交运行。可以在数据编辑窗口看到一个新生成的秩变量。

#### 4.2.2 排秩实验

1. 实验1：在“休闲调查1”数据文件中对“年龄”求秩。
2. 实验2：在“休闲调查1”数据文件中按照男、女分组，对“年龄”求秩。

### 4.3 合并文件

合并文件的功能是将一个外部文件与当前文件合并成一个新的数据文件。合并时既可以增加个案，也可以增加变量。

执行下述操作：data →Merge files（合并文件）拉出三级菜单，如图4-5 所示。

可以看到三级菜单中有两个命令Add Cases（增加个案）和Add Variable（增加变量）。即文件的合并有两种方法。一种是将两个文件中的个案和并在一起，另一种是将两个文件中的变量和并在一起。

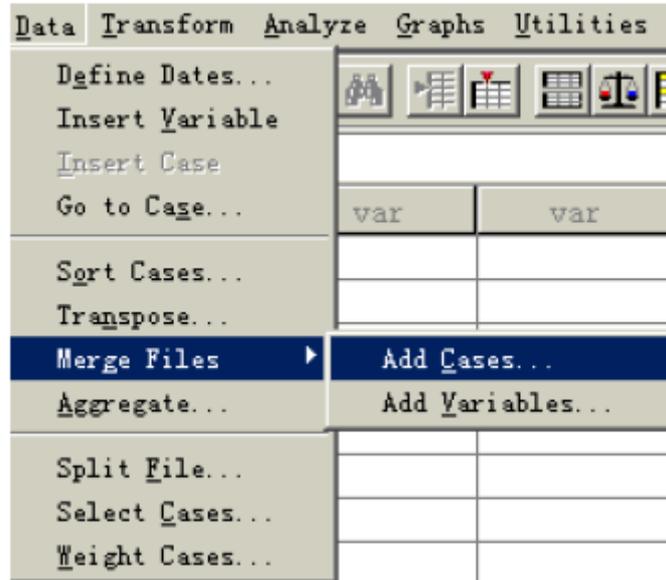


图4-5 Merge files 菜单

#### 4.3.1 增加个案 (Add Cases)

##### 1. 增加个案 (Add Cases) 的合并方法。

增加个案是指将外部数据文件中,与现在的工作变量具有相同变量名的个案追加到现在的工作变量中。也称纵向合并。合并的方法如下:

(1) 打开合并个案对话框

在当前工作文件的数据窗口下执行下述操作:

data→Merge files →Add Cases 打开读取文件 (read File) 对话框,如图4-6 所示。



图4-6 读取文件对话框

这是与第二章,第二节中“打开文件对话框”的内容和使用方法都基本相同的对话框。

(2) 选择要合并的外部文件。从read File 的文件列表窗口中选择要合并的外部文件的文件名,单击“打开”按钮,打开Add cases from 对话框,如图4-7 所示。



图4-7 合并个案对话框

1) Unpaired variable 是不匹配变量窗口。该窗口中的变量是两个文件中变量名不相同的变量或变量名虽相同而变量定义不同的变量。执行合并的命令后，这些变量在新的数据文件中将消失。在该窗口的下面有两项说明：

[\*]=Working Data File: 星号表示工作数据文件。

[+]=后面标出的是外部文件。

2) Variables in new working data File 是新的工作变量窗口。执行合并命令后，在该窗口中的所有变量都在新的数据文件中存在。

3) Indicate case source as variable 是指明Case 来源的选项。如果选择此项，执行和并命令后，系统将生成一个新的数据文件，并在这个数据文件最右侧生成一个表示个案来源的新变量。系统默认的变量名是Source01。也可以输入任意变量名作为表示个案来源的变量。系统默认的该变量的取值为0 和1。数值0 表示该个案来源于原工作数据文件，数值1 表示该个案来源于外部数据文件。

(3) 单击“OK”后，执行合并命令，生成新的数据文件。

## 2. 增加个案的实验

**实验：将“休闲调查1”与“休闲调查2”的Cases 合并。**

### 4.3.2 增加变量 (Add Variable)

增加变量是将外部数据文件中的变量增加到当前的工作文件中。也称横向合并。合并的方法有两种。一种是直接合并，另一种是按关键变量合并。

#### 1. 直接合并

(1) 打开读取文件对话框。在当前工作文件的数据窗口下执行下述操作：

date → Merge files → Add Variable 打开read File (读取文件) 对话框，如图4-8 所示。

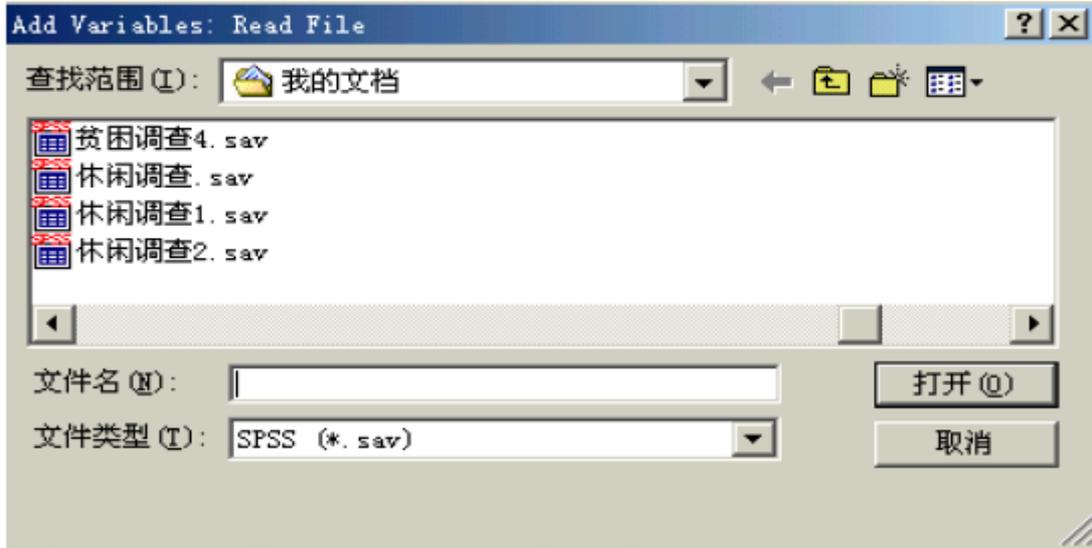


图4-8 读取文件对话框

(2) 选择要合并的外部文件。在read File 对话框中的文件名列表窗口中选择要合并的外部数据文件的文件名。单击“打开”按钮，打开增加变量（ADD Variables from）对话框，如图4-9 所示。

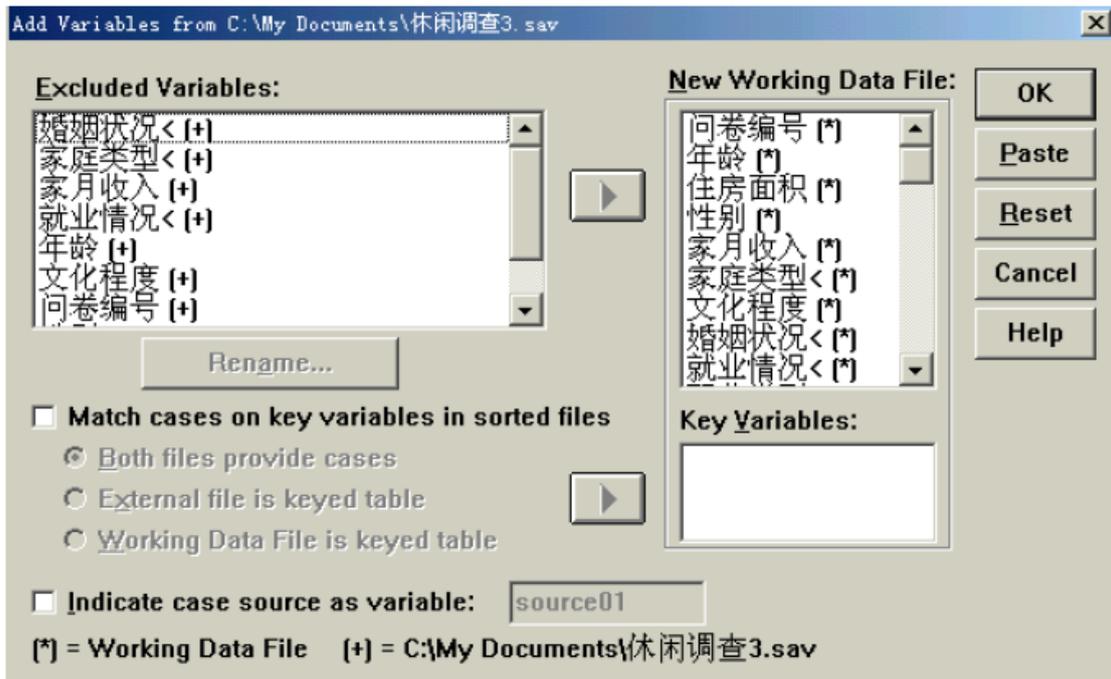


图4-9 增加变量对话框

1) Excluded Variables 是拒绝变量窗口。在将要和并的外部数据文件中，如果存在与当前工作文件的变量名同名的变量，这部分变量被拒绝加到合并后的新的数据文件中。该窗口中所显示的即是这部分变量。

2) New working data File 是新的工作数据文件窗口。该窗口中包括了原工作数据文件中的全部变量及和并进来的外部数据文件中与原工作数据文件不同名的变量。该窗口中的所有变量将存在于新生成的数据文件中。

3) [\*]=Working Data File: 星号表示工作数据文件

4) [+]: 加号表示外部文件。

(3) 单击“OK”，执行合并命令，系统将生成新的数据文件。

## 2. 直接合并的实验

**实验：将“休闲调查1”的变量与“休闲调查3”的变量合并。**

在合并前，“休闲调查1”中共有39个变量。合并后增加了12个变量。

### 3. 按关键变量合并的方法

用上述方法可以对任意两个数据文件进行合并。但如果合并进来的变量的不属于原文件中的个案时，这种合并便毫无意义。为使合并进来的变量确属于原来文件中的个案，可以采用按关键变量合并的方法。按这种方法合并时，要求两个文件至少有一个共同的关键变量。关键变量是具有相同变量名的变量。按关键变量合并的过程如下。

(1) 打开读取文件对话框。在当前工作文件的数据窗口下执行下述操作：data → Merge files → Add Variable 打开read File (读取文件) 对话框，如图4-8所示。

(2) 选择要合并的外部文件。在read File 对话框中的文件名列表窗口中选择要和并的外部数据文件的文件名。单击“打开”按钮，打开增加变量对话框，如图4-9所示。

(3) 确定关键变量。Key Variables 是关键变量窗口。选择在两个文件中具有相同变量名的变量作为关键变量。方法是，从图4-9中的Excluded Variables 窗口中选择一个关键变量，使之进入到Key Variables 窗口中。

(4) 对两个文件按关键变量的升序进行排序。按关键变量合并数据文件时，系统要求对两个数据文件按关键变量的升序来排序。如果没有排序就直接合并，系统会给出提示。

(5) 确定个案的匹配方法。Match cases on key variables in sorted files 是按排序的关键变量匹配个案的选项。选中Match cases on key variables in sorted files 选项后，会激活下面的三个选项。

1) Both files provide case 是由外部文件和当前文件两者提供个案的选项。如选择此项可以激活Indicate case source as variable (生成一个表明个案来源的变量) 选项。选择此项后系统将生成一个变量名为Source01 的新变量。其值为0 表示个案来自当前数据文件，为1 表示个案来自外部数据文件。

2) External file is keyed table 是以外部文件为关键表的选项。选择该选项，则保持当前数据中的个案数不变，外部文件匹配当前数据文件的关键变量值。如匹配成功，外部文件的新变量加入到当前数据文件中。匹配不成功则不加入。

3) Working Data File is keyed table 是以当前数据文件为关键表的选项。保持外部数据文件中的个案数不变，当前数据文件匹配外部文件的关键变量值。如匹配成功，外部文件中的变量加入到当前数据文件的变量中。匹配不成功则舍去当前数据文件中的个案。

(6) 上述选项作完以后，单击OK 按钮提交运行。系统将生成一个变量合并后的新的数据文件。

### 4. 按关键变量合并的实验

**实验：以“问卷编号”为关键变量，将“休闲调查1”的变量与“休闲调查3”的变量并按关键变量合并。**