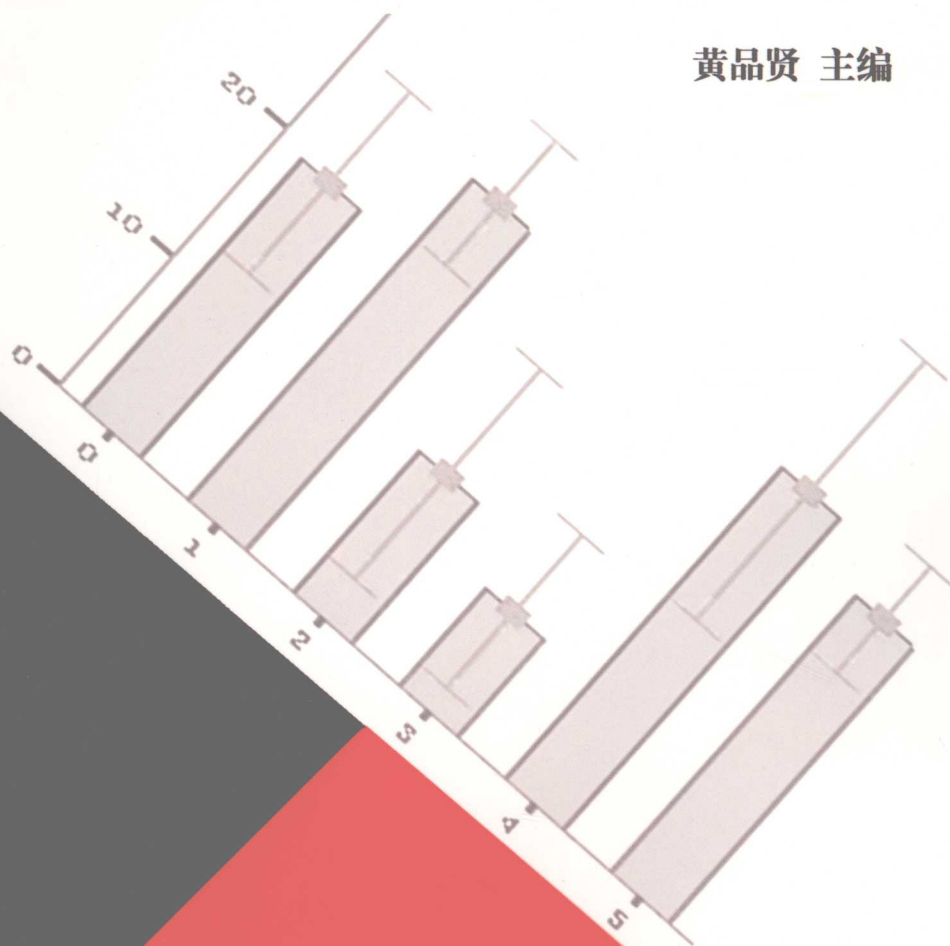


中医统计学实习指导 及SPSS15.0的应用

黄品贤 主编



科学出版社
www.sciencepress.com

中医统计学实习指导

张其成 主编

第1版 第1次印刷

中国中医药出版社

中医统计学实习指导 及 SPSS15.0 的应用

黄品贤 主 编

科 学 出 版 社

北 京

内 容 简 介

目前,越来越多的中医药科研实践需要用到高级统计分析方法,而要实现这样的统计并得到准确的结果,只能借助于统计软件来完成。SPSS 统计软件是现阶段国际上最流行的权威性统计分析软件之一,也是非统计专业人员应用最多的统计软件。中医药院校本科生和研究生的《医学统计学》教材虽品种繁多,但缺乏能提高解决统计工作中实际问题能力的实习指导。因此,编写《中医统计学实习指导及 SPSS15.0 的应用》,目的是帮助学生更好地理解 and 掌握医学统计学方法,提高学生解决统计工作中实际问题的能力。

本书的特点是将统计学理论和方法与 SPSS15.0 操作衔接,注重教材的基础性、先进性和实用性,尤其注重常用的研究设计、多元方差分析、多元线性回归分析、生存分析及流行病学信度和效度分析等。

本书适应面广,可供中医药院校各专业的本科生、七年制学生使用,也可供硕士研究生、博士研究生和其他研究人员参考和使用。

图书在版编目(CIP)数据

中医统计学实习指导及 SPSS15.0 的应用/黄品贤主编.
北京:科学出版社,2009
全国高等中医院校选用教材
ISBN 978-7-03-024830-5

I. 中… II. 黄… III. 中国医药学-医学统计-统计分析-软件包,SPSS 15.0-中医学院-教材 IV. R2-32

中国版本图书馆 CIP 数据核字(2009)第 102888 号

责任编辑:潘志坚 钱 鑫/ 责任校对:刘珊珊
责任印制:刘 学 / 封面设计:殷 靓

科 学 出 版 社 出 版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

南京展望文化发展有限公司制版

江苏省句容市排印厂印刷

科学出版社出版 各地新华书店经销

*

2009 年 7 月第 一 版 开本: B5(720×1 000)

2009 年 7 月第一次印刷 印张: 17 3/4

印数: 1—3 200 字数: 348 000

定价: 21.00 元

前 言

医学统计学(Medical Statistics)是运用统计学的原理和方法,研究医学科研中有关数据的搜集、整理、分析和推断的一门的科学。近年来,随着计算机技术、统计软件与现代信息学的发展,使医学统计学得到了突飞猛进的发展,并在中医药研究中发挥了显著的作用。目前,许多国际性研究项目、国家科技支撑计划项目、新药临床试验等均要求有医学统计学人员参加,越来越多的中医药科研与实践需要用到高级统计分析以及建立模型,而要实现这样的统计并得到准确的结果,只能借助于统计软件来完成。SPSS 统计软件是目前国际上最流行的权威性统计分析软件之一,其特点是易于掌握、操作方便、功能优异,也是非统计专业人员应用最多的统计软件。目前中医药院校本科生和研究生的《医学统计学》教材繁多,但缺乏结合中医药研究实例的实习指导。因此,我们编写《中医统计学实习指导及 SPSS15.0 的应用》,目的是帮助学生更好地理解医学统计学的基本概念、基本理论,掌握各种统计方法的使用条件、统计分析思路、正确解释统计结果,最终提高学生实际问题能力。

本书结合中医药研究,将统计学理论和方法与 SPSS15.0 操作衔接,在内容上重点围绕学生在学习和应用医学统计学时经常遇到的重点、难点,中医药研究与实践中的应用问题,同时注重研究设计、多元线性回归分析、生存分析、聚类分析与判别分析、主成分分析与因子分析、信度和效度分析等的编写。题型有是非题、最佳选择题、简答题、计算分析题。

本书适应面广,可供中医药院校各专业的本科生、七年制学生使用,也可供硕士研究生、博士研究生和中医药科研人员学习参考。

限于我们的经验、水平和能力,加之编写时间仓促,书中难免有错漏之处,恳请各位专家、广大师生批评指正,以便日后修订完善。

为方便大家运用,与本书配套的 SPSS 数据文件集可以登陆“www.sciencep.sh.cn”下载,文件名为“中医统计学实习指导及 SPSS15.0 的应用—数据文件”。

黄品贤

2009 年 3 月于上海

hpx8388@hotmail.com

目 录

前言

SPSS15.0 概述	1
第一节 数据编辑窗口的基本操作	1
第二节 数据文件的整理	7
第三节 结果窗口基本操作	17
实习一 中医统计学基础	20
习题	20
参考答案	22
实习二 统计表和统计图	23
习题	23
参考答案	28
实习三 数值变量资料的统计描述	44
习题	44
参考答案	48
实习四 数值变量资料的统计推断——t 检验	56
习题	56
参考答案	64
实习五 数值变量资料的统计推断——方差分析	85
习题	85
参考答案	94
实习六 分类变量资料的统计描述和统计推断	120
习题	120
参考答案	126
实习七 非参数检验	141
习题	141
参考答案	145
实习八 相关回归分析	158
习题	158

参考答案	162
实习九 实验设计	172
习题	172
参考答案	176
实习十 多元线性回归分析	181
习题	181
参考答案	184
实习十一 Logistic 回归分析	193
习题	193
参考答案	197
实习十二 生存分析	202
习题	202
参考答案	207
实习十三 聚类分析和判别分析	216
习题	216
参考答案	219
实习十四 主成分分析和因子分析	233
习题	233
参考答案	235
实习十五 信度与效度的分析	241
习题	241
参考答案	243
实习十六 异常值的处理	252
习题	252
参考答案	254
实习十七 Meta 分析	257
习题	257
参考答案	259
综合试题一	261
综合试题二	269
参考书目	275

SPSS15.0 概述

提要

1. 介绍 SPSS15.0 for Windows 数据编辑窗口的基本操作,包括:操作入门,File 菜单简介。
2. 数据文件的整理:对变量进行变换,对数据文件进行分组、合并、加权及筛选等。
3. 统计分析结果的导出。

第一节 数据编辑窗口的基本操作

一、数据编辑窗口操作入门

(一) 数据编辑窗口

当打开 SPSS 程序后,展现在面前的界面就是数据编辑窗口(图 1-1):

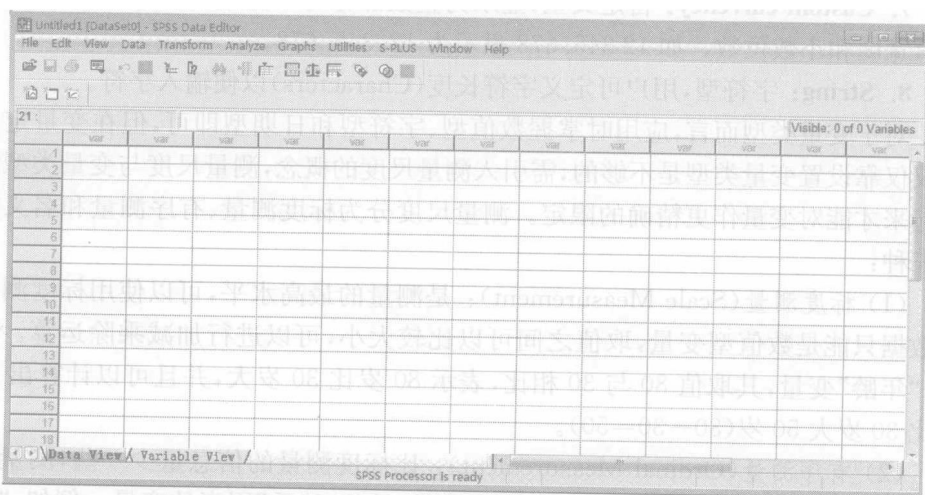


图 1-1 数据编辑窗口

这是一个典型的 Windows 软件界面,有菜单栏和工具栏。工具栏下方是数据栏,数据栏下方是数据编辑窗口的主界面,该界面和 Excel 的界面非常相似,由若干个行和列组成,每行对应一条记录,每列对应一个变量。当没有输入任何数据

时,所有的行和列的标号都如图所示是灰色的。

(二) SPSS 中变量类型与测量尺度

SPSS 中变量共有两种基本类型,即数值型和字符型,具体分为 8 个类型:

1. **Numeric:** 数值型,可同时定义数值的宽度(Width),即整数部分、小数点和小数部分的位数,默认为 8 位;定义小数位数(Decimal Places),默认为 2 位。

2. **Comma:** 加逗号的数值型,即整数部分每 3 位数加一逗号,其余定义方式同数值型。

3. **Dot:** 3 位加点数数值型,无论数值大小,均以整数形式显示,每 3 位加一小数点(但不是小数点),可定义小数位数,但都显示 0,且小数点用逗号表示。如 1.234 5 显示为 12.345,00(实际是 $12345E-4$)。

4. **Scientific notation:** 科学记数型,同时定义数值宽度和小数位数,在数据管理窗口中以指数形式显示。如定义数值宽度为 9,小数位数为 2,则 345.678 显示为 $3.46E+02$ 。

5. **Date:** 日期型,用户可从系统提供的日期显示形式中选择自己需要的。如选择 mm/dd/yy 形式,则 1995 年 6 月 25 日显示为 06/25/95。

6. **Dollar:** 货币型,用户可从系统提供的货币显示形式中选择自己需要的,并定义数值宽度和小数位数,显示形式为数值前有 \$。

7. **Custom currency:** 自定义型,显示为整数部分每 3 位加一逗号,用户可定义数值宽度和小数位数。如 12 345.678 显示为 12,345.678。

8. **String:** 字符型,用户可定义字符长度(Characters)以便输入字符。

对于变量类型而言,应用时掌握数值型、字符型和日期型即可,但在变量定义时仅仅靠设置变量类型是不够的,需引入测量尺度的概念,测量尺度与变量类型结合起来才能对变量作更精确的限定。测量尺度分为标度测量、有序测量和名义测量三种:

(1) 标度测量(Scale Measurement): 是测量的最高水平,可以使用标度测量的数据只能是数值型变量,取值之间可以比较大小,可以进行加减乘除运算。例如,“年龄”变量,其取值 80 与 30 相比,表示 80 岁比 30 岁大,并且可以计算出 80 岁比 30 岁大 50 岁($80-30=50$)。

(2) 有序测量(Ordinal Measurement): 比标度测量的信息量少一些,测量数值不代表数量的大小,而是代表有序分类情况,是基于“质”因素的变量。例如:“疗效”变量分为治愈、好转、无效、恶化四个等级,变量取值依次代表疗效越来越差。

(3) 名义测量(Nominal Measurement): 是三种测量尺度中测量精确度最低、最粗略的基于“质”因素的变量,其变量取值仅代表观测对象的分类或属性,无法用变量的取值比较大小,如“性别”变量、“职业”变量等。

(三) 定义变量

建立 SPSS 数据文件,多数情况下需要定义新变量,只需单击数据编辑窗口左下方的 **Variable View** 就可以切换到变量定义界面开始定义新变量了。如 anxiety.sav 的变量定义(图 1-2):

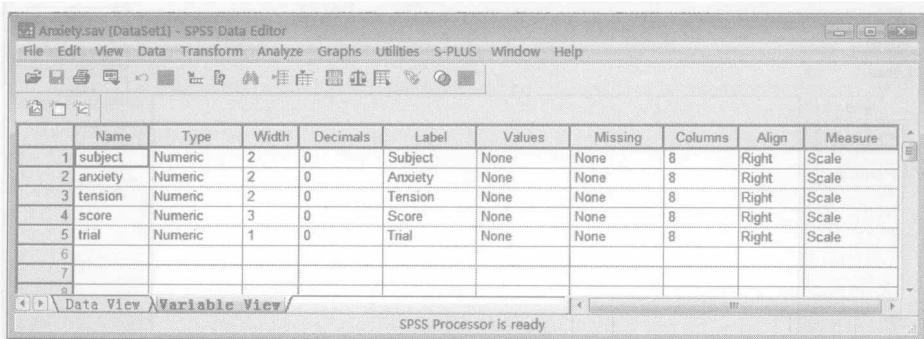


图 1-2 数据编辑窗口变量定义界面

以变量 subject 为例:变量名为 subject,类型为 Numeric,宽度为 2 位,小数位数 0 位,变量名标签为 Subject,变量值标签未定义……

下面具体介绍变量视图中各个栏目的意义和用法:

1. Name 栏: 用于设定变量名,SPSS 中变量名长度应在 64 位以内,推荐尽量使用英文变量名,因为变量名采用中文有时会有潜在冲突;并且,当将数据转换为其他格式时,中文变量名可能会受影响。例如,强大的 SAS 软件较低的版本就不认中文变量名。

2. Type 栏: 选择该框时右侧出现  的按钮,单击按钮会弹出变量类型对话框(图 1-3),可设置变量类型,并更改变量运算宽度等格式,一般按默认数值型即可。

3. Width 栏: 设置变量运算宽度,如数值型默认为 8 位,一般不需更改。

4. Decimals 栏: 设置小数位,默认为 2 位,即数值型变量默认情况下为 5 位整数、1 位小数点位和 2 位小数。

5. Label 栏: 用于定义变量名标签,是对变量名的进一步描述,该标签会在结果中输出以方便阅读。如果设置了英文变量名感觉阅读不是很方便,可在此设置中文变量名标签,变量标签可长达 256 个字符,需要时可用变量标签对变量名的含

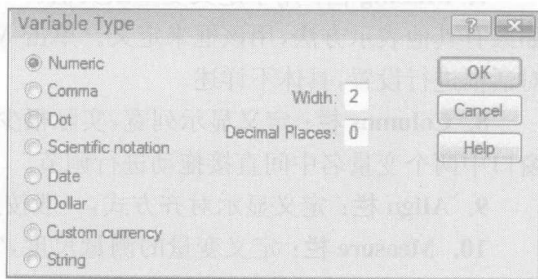


图 1-3 变量类型设置对话框

义加以解释。

6. Values 栏: 用于定义变量值标签,以 anxiety 为例,单击 Value 框右侧的按钮,会弹出变量值标签对话框(图 1-4)。

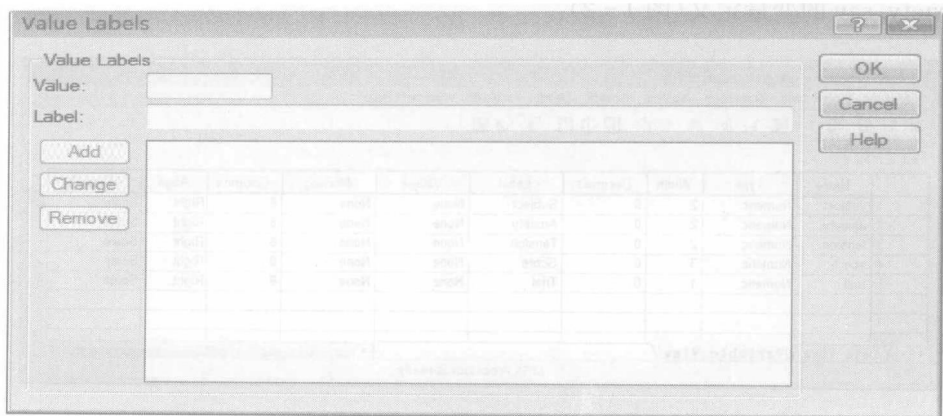


图 1-4 变量值标签对话框

上部的两个文本框分别为变量值输入框和变量值标签输入框,分别在其中输入“1”和“有”,此时下方的 Add 按钮变黑,单击它,该变量值标签就会被加入下方的标签框内。同样方法定义变量值“2”为“无”,最后按 OK,变量值标签设置完成,分析结果中会有相应的标签出现。在数据编辑窗口中切换回 Data View 界面,然后选择菜单 View,点 Value Labels,便可以看到变量设定值标签后的效果。

7. Missing 栏: 用于定义变量缺失值。一般按默认设置,即缺失值用“.”表示,如果有其他表示方法,用该框来定义。单击 Missing 框右侧的省略号,弹出缺失值对话框进行设置,具体不详述。

8. Columns 栏: 定义显示列宽,实际很少用到这里的设置,多通过鼠标在数据窗口中两个变量名中间直接拖动进行调节。

9. Align 栏: 定义显示对齐方式,一般按照默认的右对齐即可。

10. Measure 栏: 定义变量的测量尺度,它可以对变量进行精确的定义,特别是在绘制交互式统计图等方法非常有用。

以上介绍中以 Name、Label 和 Measure 栏较为常用,其他了解即可。定义了变量后就可以输入数据了,或者有时已经输入的数据往往有错误,这就需要对数据进行修改,Windows 界面下直接输入、修改非常方便,在此不再赘述。

二、File 菜单简介

除了统计分析功能外,统计软件的数据管理能力也非常重要,在 SPSS 中数据

文件的管理功能基本上集中在 File 菜单上,该菜单的组织结构和常用软件如 Word 等极为相似,本部分只介绍 SPSS 中常用的几个菜单项。

(一) 新建数据文件

当进入 SPSS 系统时,系统就已经生成了一个空的数据文件(图 1-1)。用户只要按照需要定义变量,输入数据存盘即可。

(二) 打开其他格式的数据文件

其他格式的数据文件可以直接在 SPSS 中打开,SPSS 提供了以下几种方法来打开其他格式的数据文件:

1. 直接打开: 在 SPSS 中可以直接读入很多常用格式的数据文件,其中包括 Excel 各个版本的数据文件。选择菜单 **File** → **Open** → **Data** 或直接点击快捷工具栏上的  按钮,弹出 **Open File** 对话框,单击“文件类型”列表框,SPSS 里能直接打开的数据文件格式有:

SPSS(*.sav)	SPSS 数据文件
SPSS/PC(*.sys)	SPSS4.0 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
Lotus(*.w*)	Lotus 数据文件
SYLK(*.slk)	SYLK 数据文件
dBase(*.dbf)	dBase 系列数据文件
SAS Long File Name(*.sas7bdat)	SAS7 版本以上长文件名类型数据文件
SAS Short File Name(*.sd7)	SAS7 版本以上短文件名类型数据文件
SAS v6 for Windows(*.sd2)	SAS6 版(for Windows)数据文件
SAS v6 for UNIX(*.ssd01)	SAS6 版(for UNIX)数据文件
SAS Transport(*.xpt)	SAS 编写格式的数据文件
Text(*.txt)	纯文本格式的数据文件
data(*.dat)	纯文本格式的数据文件

选择所需的文件类型,然后选中需要打开的文件,SPSS 就会按照要求打开相应的数据文件,并转换到 SPSS 数据文件中。

2. 数据库查询方式打开: 对于不能在 SPSS 中直接打开的数据文件,SPSS 还提供了一种使用范围更广、较为专业的数据接口,即数据库查询。例如使用数据库查询方式打开 Access 2003 的数据文件 demo.mdb。

选择菜单 **File** → **Open Database** → **New Query**, 系统会弹出数据库向导的第一个窗口,列出所使用电脑上已安装的所有数据源格式(图 1-5):

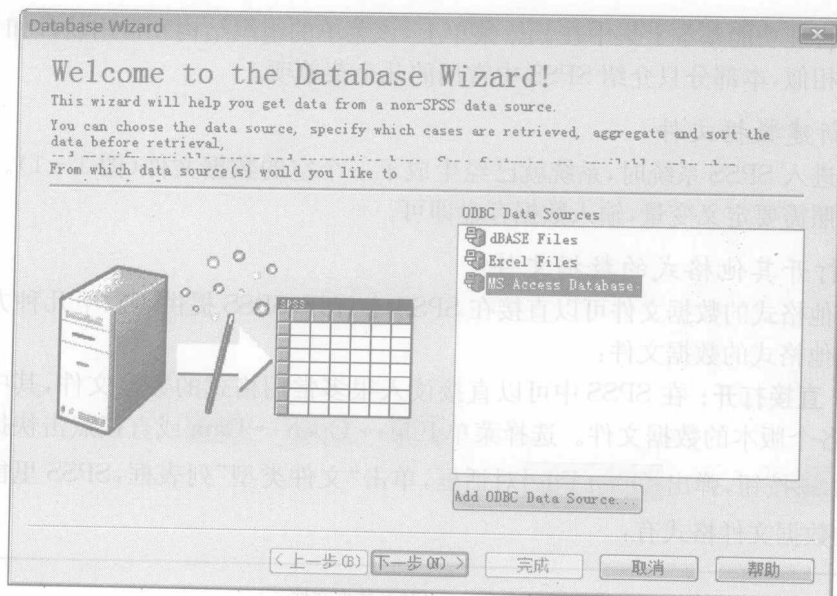


图 1-5 数据库向导初始对话框中的数据源列表

选中 MS Access Database 然后单击 **下一步** 按钮,在弹出的文件搜索对话框中找到 demo. mdb 文件并确定,出现向导窗口(图 1-6),采用拖放的操作将所需变量引入右侧框中,一般不需复杂设置,直接点击完成即可。

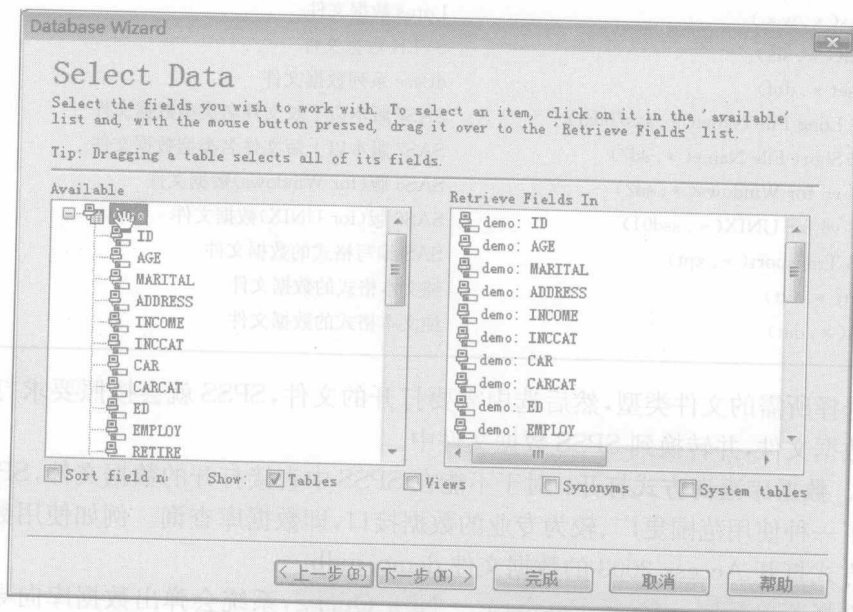


图 1-6 打开数据库向导变量选择窗口

目前 SPSS 直接打开功能越来越完善,数据库查询接口主要用于和一些流行的数据库进行数据交换,如 SQL Server、DB2、Oracle 等。

(三) 保存数据文件

对数据进行修改后,就需要对数据进行保存。选择 File → Save,如果数据文件曾经存储过,则系统会自动按原文件名保存数据;否则会弹出和选择 Save As 菜单时相同的 Save Data As 对话框,选择合适的类型保存即可。

需要注意的是,有时在分析过程中生成的一些临时变量,如果不希望全部保存下来,那么可以使用 Save Data As 对话框中的 Variables 按钮来指定需要保存的变量。图 1-7 是在保存文件 demo.sav 时 Variables 子对话框的内容,发现在每个变量的最左侧都有一个复选框,表明它们都会被保存在文件中。如果不想保存哪个变量,则单击相应变量的复选框去除选择。

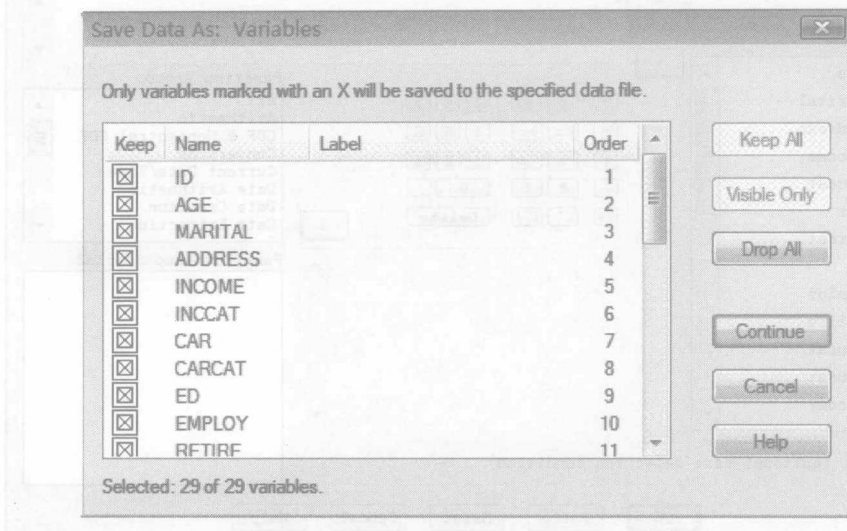


图 1-7 Save Data As: Variables 子对话框内容

第二节 数据文件的整理

一、Transform 菜单

Transform 菜单主要集中了一些对变量进行变换的过程,例如对原变量进行四则运算、对数据重新编码、产生秩次变量等,这些过程在统计分析的预处理中起着重要作用,尤其是 Compute 过程,将重点介绍。

(一) Compute 过程

Compute 过程是 Transform 菜单中最常用的过程,用于给变量赋值,可以产生新的目标变量,也可以对已存在的变量进行赋值,可以对所有记录进行操作,也可以设定逻辑条件,只对满足条件的记录加以赋值。

例如:在数据集 demo. sav 中建立新的变量 age2,令其值当变量 $\text{age} \leq 30$ 时为 1,当 $30 < \text{age} \leq 45$ 时为 2,其余取值为 3,试图通过赋值过程实现将原连续型变量 age 转变成三分类的分类变量。

选择菜单 Transform $\rightarrow$ Compute Variable,弹出 Compute Variable 对话框(图 1-8):

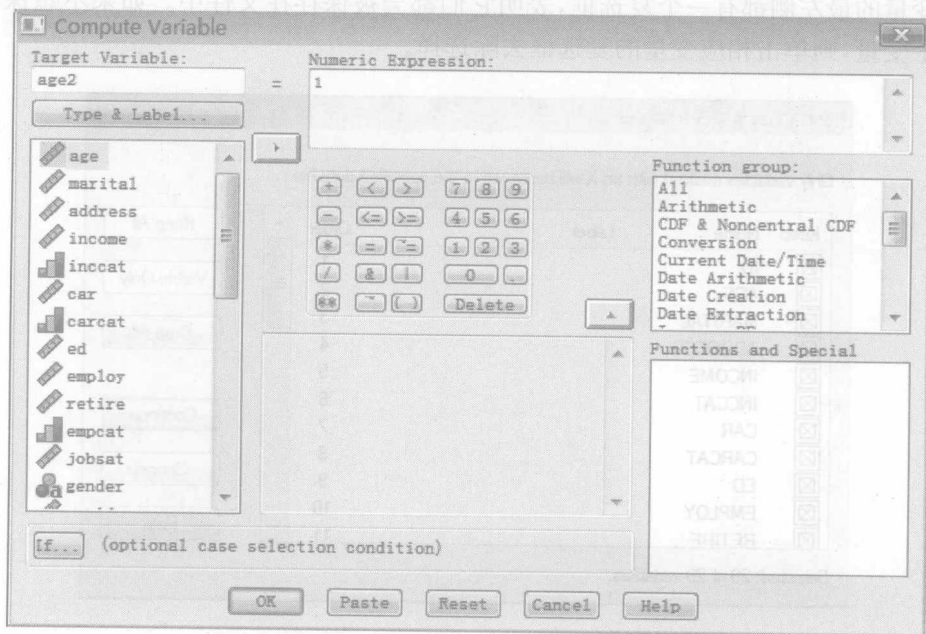


图 1-8 Compute Variable 对话框

现对此对话框进行详细讲解:

1) 左上角 Target Variable 框,在其中输入需要计算的变量名,在此键入 age2 后,Type & Label 按钮就会变黑,点击这个按钮可以对“age2”变量进行详细的定义,多数情况下不需设置。

2) 左下方为候选变量列表,列出了原数据集中所有的变量,用于将计算所需变量直接引入 Numeric Expression 框。

3) 右上方的 Numeric Expression 框,用于给目标变量具体赋值,先输入“1”。

4) 对话框中部的软键盘,可以用鼠标按键输入数字和符号,软键盘右侧为函

数窗口,在赋值过程中可以找到需要的函数调用之。

5) 对话框下部的 **if** 按钮,用于设定进行赋值的逻辑条件。点击此按钮,出现对话框(图 1-9)。

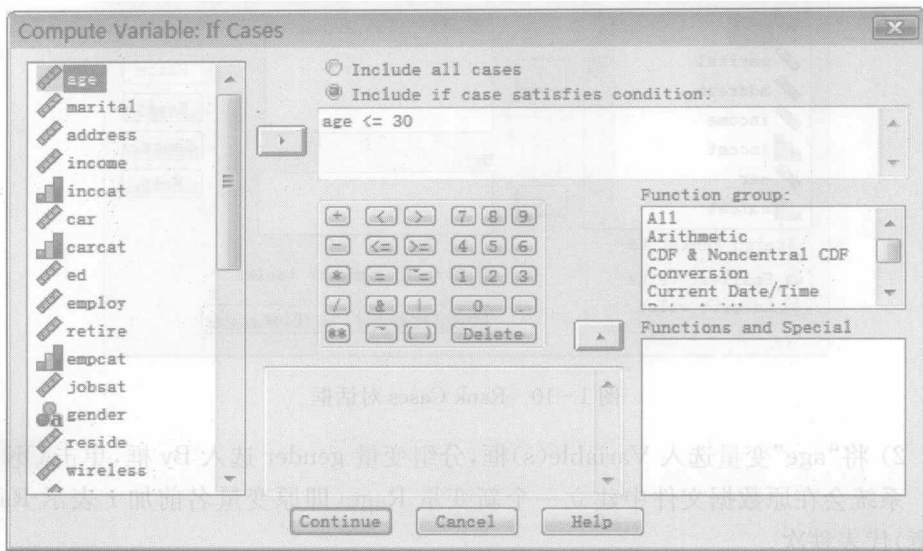


图 1-9 记录选择对话框

此例需要根据 age 变量的取值来对 age2 变量赋值,因此选中第二个选项“Include if case satisfies condition”,左侧选中“age”变量,然后单击“**+**”,age 变量被引入了右侧的变量框,用鼠标点击软键盘或通过电脑键盘直接敲击完成表达式“age <= 30”。现在 **Continue** 按钮变黑,点击它,系统回到 **Compute Variable** 对话框,然后单击 **OK** 按钮。通过以上步骤,将 age 取值 ≤ 30 岁的观察对象 age2 变量赋值为 1,依此方法可完成 age2 变量的所有赋值。需要注意的是,接下来当对 age2 赋值 2 或 3 时,系统会弹出一个确认对话框,确认对 age2 变量的替换即可。

软键盘中几个符号解释如下:

~ =	&		**	~
不等于	逻辑符号 AND	逻辑符号 OR	乘方,相当于函数 EXP()	逻辑符号 NOT

(二) Rank Cases 过程

在许多时候,资料不满足参数检验的条件,则不得不使用非参数检验方法,简单的非参数检验可以直接用特定对话框来完成,但复杂的非参数方法则需要先计算秩次再进行相应的分析,这时就需用到 Rank Cases 过程。Rank Cases 可对变量取值进行排次序(秩次),并将秩次结果存储到一个新变量中去。

例如：在 demo. sav 数据中，按 gender 变量分两组计算 age 变量值的秩次。

1) 选择 Transform → Rank Cases，弹出 Rank Cases 对话框(图 1-10)：

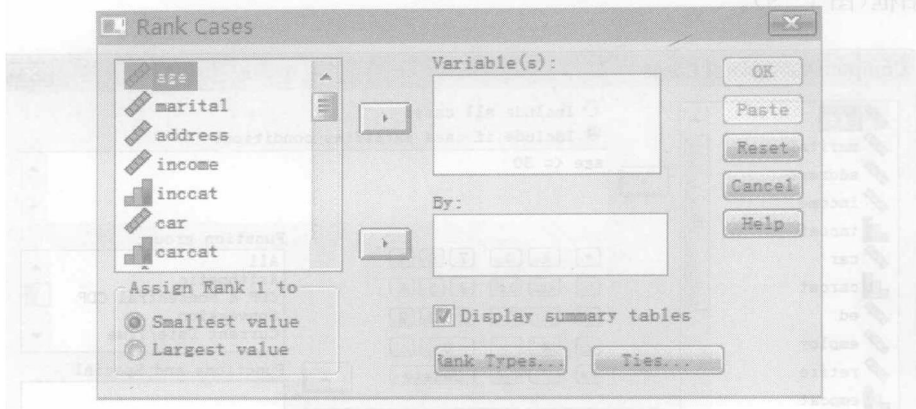


图 1-10 Rank Cases 对话框

2) 将“age”变量选入 Variable(s)框，分组变量 gender 选入 By 框，单击 OK 即可。系统会在原数据文件中建立一个新变量 Rage(即原变量名前加 r 表示 Rank 之意)代表秩次。

二、Data 菜单

Transform 菜单下的过程多是对变量进行操作，但实际应用中，经常还需对整个数据文件进行加工整理，比如对数据进行分组、合并、加权及筛选等操作，这些功能主要集中在 Data 菜单中。

(一) Restructure 过程

在进行统计分析时，往往需要根据要求改变数据的排列格式，如 SPSS 重复测量模型中要求同一对象的多次测量并排在一行上，如果原来数据不是按要求的格式输入，就需要用到这里的 Restructure 过程。

例如：SPSS 自带数据文件 anxiety. sav，记录了 12 名精神病患者在接受治疗后四个时间点的精神状态评分。在 anxiety. sav 中每一次测量成为一条记录，用变量 subject 和 trial 区分是哪位病人的第几次测量，anxiety 和 tension 作为协变量在相同病人的记录中重复出现，这种格式称为长型格式；而若用 SPSS 重复测量模型进行分析需对原格式进行转换，要求同一对象的多次测量并排在一行上，每位观察对象为一条记录，四次测量分别用 trial1~trial4 这四个变量来分别记录，不需单独设置变量 trial 区分测量次数，subject、anxiety 和 tension 也不再重复出现，这种格式被称为宽型格式，转换后数据存为 anxiety2. sav。