

社会科学统计软件包

SPSS/PC⁺

- 李燕琛 程士寅 编著
- 科学技术文献出版社

社会科学统计软件包

SPSS/PC⁺

李燕琛 程士寅 编著

科学技术文献出版社

(京)新登字130号

内 容 简 介

本书系统地介绍了目前国际上社会科学研究中广泛使用的 SPSS/PC⁺软件包,并对该软件系统的构成、语言、工作方式及数据管理功能等进行了详细说明。为了便于阅读在叙述软件包统计功能的同时,还对有关统计方法作了必要的、通俗易懂的介绍。

主要读者对象是社会科学工作者,也可作为高等院校社会学、心理学、人口学、计划统计学以及经济类有关学科的本科生、研究生的教材或参考资料。

社会科学统计软件包 SPSS/PC⁺

李燕琛 程士寅 编著

科学技术文献出版社出版

(北京复兴路15号 邮政编码 100038)

中国科学技术情报研究所印刷厂印刷

新华书店北京发行所发行 各地新华书店经售

*

787×1092毫米 16开本 11.125印张 283千字

1991年10月第1版 1991年10月第1次印刷

印数: 1—2000册

科技新书目: 253—158

ISBN 7-5023-1499-7/TP·83

定 价: 6.90元

前 言

随着对社会现象考察、分析的精确化、科学化,不少数学方法特别是数理统计方法已经不可避免地渗透到社会科学领域的研究中来。电子计算机一般都配有社会科学统计软件包。正确地使用这些软件包来解决有关实际问题,已经成为衡量研究工作水平先进性的一个重要标志。正因为如此,我们认为比较系统地介绍一下社会科学统计软件包的情况是很有必要的。

SPSS 统计软件包是国际上广泛流行的著名统计软件包,是美国 SPSS 公司的软件产品,能够在几十种型号计算机上运行。SPSS 是英文 Statistical Package for the Social Science (社会科学统计软件包)的缩写,它适用于社会学、人口学、心理学、医学、经济管理学、统计学等学科的研究,是各级计划、统计、管理部门实现科学管理、决策的有力工具。从 SPSS 软件包提供的统计方法及实际使用情况来看,SPSS 软件包还能用来解决相当多的自然科学和工程技术问题。

鉴于 IBM PC 机及其兼容机是我国最普及的微机,本书只介绍运行于 IBM PC 机的 SPSS 统计软件包。对于 SPSS 用户来说,读懂了这本小册子,也就不难掌握其它版本的 SPSS 软件包。

本书前九章介绍 SPSS/PC 软件包,一至三章对数据和文件的管理,如何控制运行等有一个简单的说明;四至八章介绍统计命令,着重说明每一个统计命令对应的统计功能,以及命令的使用方法和上机步骤。对常见的数理统计方法我们将不作解释,而对于一般中文统计书上不常见的统计方法,则给以简洁说明;第九章介绍对样品的查询、显示和输出的命令;第十章介绍 SPSS/PC 的新版本 SPSS/PC+,着重讲述它较之 SPSS/PC 扩充的功能。

限于笔者的经验水平,本书难免存在错误和不妥之处,敬请读者批评指正。

作者

1990.12

目 录

第一章 SPSS/PC 软件包概况	(1)
第一节 SPSS/PC 系统的构成	(1)
第二节 SPSS/PC 命令语言	(1)
第三节 SPSS/PC 的工作方式	(3)
第四节 SPSS/PC 的文件	(4)
第二章 SPSS/PC 的数据和文件管理系统	(6)
第一节 SPSS/PC 的数据定义与转换命令	(6)
第二节 SPSS/PC 的文件操作与管理命令	(12)
第三章 SPSS/PC 的运行操作与控制系统	(15)
第一节 系统的设置命令—SET 命令	(15)
第二节 执行命令—INCLUDE 命令	(19)
第三节 PC 命令	(19)
第四章 数据的一般统计处理方法	(20)
第一节 计算单变量的统计量—DESCRIPTIVES 命令	(20)
第二节 单变量的频率表与直方图—FREQUENCIES 命令	(23)
第三节 多变量的列联表与统计量—CROSSTABS 命令	(27)
第五章 对比组的统计分析	(36)
第一节 单因子的方差分析—ONEWAY 命令	(36)
第二节 计算多因子各试验条件下的统计量—MEANS 命令	(40)
第三节 多因子的方差分析与协方差分析—ANOVA 命令	(42)
第四节 两个独立样本与样本对的 t 检验—T-TEST 命令	(48)
第六章 相关分析与回归分析	(51)
第一节 相关分析—CORRELATIONS 命令	(51)
第二节 回归分析—REGRESSION 命令	(53)
第七章 非参数检验—NPAR TESTS 命令	(65)
第八章 多元分析	(78)
第一节 聚类分析—CLUSTER 命令	(78)
第二节 因子分析—FACTOR 命令	(89)
第九章 数据显示过程	(103)
第一节 活动文件的信息显示—LIST、WRITE、DISPLAY 命令	(103)
第二节 绘制二维图形—PLOT 命令	(105)
第三节 输出报表的控制—REPORT 命令	(110)
第四节 上机操作技巧点滴	(116)
第十章 SPSS/PC+增加的命令与功能	(119)
第一节 判别分析—DISCRIMINANT 命令	(119)
第二节 对文件的操作管理命令	(135)
一、对当前活动文件的聚合—AGGREGATE 命令	(135)

二、合并系统内的文件—JOIN 命令	(138)
三、修改活动文件的变量—MODIFY VARS 命令	(141)
四、对活动文件的样品操作—SORT CASE 命令	(143)
五、对 SPSS / PC+系统的管理—SPSS MANAGER 命令	(144)
六、显示系统文件的信息—SYSFILE INFO 命令	(146)
七、外部数据文件与活动文件的转换—TRANSLATE 命令	(146)
第三节 辅助运行命令	(149)
一、命令文件的注释	(149)
二、在 SPSS / PC+系统中运行其它软件系统的程序或命令 —EXECUTE 命令	(149)
三、定义输出文件的标题与子标题 —TITLE 命令与 SUBTITLE 命令	(150)
第四节 REVIEW 编辑器	(150)
一、REVIEW 的进入、退出与工作方式切换	(151)
二、菜单使用初步	(153)
附录: SPSS / PC+ V2.0 的命令一览表	(155)
参考文献	(172)

第一章 SPSS / PC 软件包概况

在应用数理统计方法的过程中,最大的困难是两个:第一,要处理的数据多、结构复杂、计算量大;第二,对于待解决的问题往往要先进行探索性分析,即通过不同统计方法反复试验比较,寻找有效的综合性的处理手段。统计软件包正是为克服这两个困难而产生、发展起来的。统计软件包 (Statistical Package) 不同于针对某一特殊统计方法而编制的专项程序,它包含了覆盖面大、使用频率高的各种统计方法的命令;统计软件包也不同于由若干统计程序组成的统计库 (Statistical Library),统计库必须依附于编写其程序的语言系统才能工作,而统计软件包则自成系统,它对数据和文件有较强的管理功能,运行软件包中的程序不需要依赖于别的系统而是独立地进行。

第一节 SPSS / PC 系统的构成

SPSS / PC 和其它统计软件包一样,由三部分构成:统计命令,数据、文件管理系统,和运行操作控制系统。这三部分对应着三类程序(命令)集合,它们之间有紧密联系,但作用各不相同。

一般说来,统计软件的主要功能是由它的统计命令来体现。但是统计命令处理分析的数据必须是 SPSS 系统内部的活动文件和系统文件,这就需要用 SPSS 的数据、文件管理命令把外部数据转化成活动文件或系统文件,并且对它们进行操作、管理;而无论是进行数据、文件的管理还是执行统计命令,都必须通过操作控制系统来完成。统计软件包的数据、文件管理系统和操作控制系统都是针对统计数据的特点,为方便各种统计方法的实施而设计的。

SPSS / PC 与用户之间通过语言沟通。SPSS 有自己的语言,它不同于计算机的一般高级语言(过程语言)。SPSS 语言是命令语言,即非过程语言。它的一个语句就是一条命令,每条命令可调用软件包中相应的程序模块。使用 SPSS 语言可编写成 SPSS 的源程序,SPSS 源程序又叫命令文件。命令文件可通过 SPSS 的执行命令提交系统运行。

第二节 SPSS / PC 命令语言

SPSS / PC 命令语言中有 40 多条命令,它们分成三类,即统计命令、数据与文件的管理命令、辅助和运行命令,分别对应着软件包的三个组成部分。当处理一个具体的统计问题时,我们可以选择合适的命令按 SPSS / PC 的语法规则和一定的次序编写成程序。程序开始部分一定是数据定义部分,然后方可调用各种统计命令。当然,在做某一统计分析之前,有时要按问题要求先进行一系列数据变换。

一、命令语言的语法规则

命令文件由一条条命令组成。每条命令由命令关键字和命令说明部分组成,它们之间要

用空格分隔开。命令说明部分是该命令相应的参数，它们可以是变量名、文件名、子命令名、运算符、关键字、变量值、分隔符、标号等。在编写每条命令时要遵守以下语法规则：

1) 一条命令可以从任意列开始写，以符号‘/’作为结束符。但程序中的数据行必须从第一列开始。

2) 一条命令在一行中写不下时，可以续行，续行内容可以从任意列开始写。

3) 一条命令中若有多个子命令，子命令之间须用斜杠‘/’作为分隔符。

4) 命令中的关键字可以简写，但至少取前三个字符。

例句：

```
LIST VARIABLES=AGE INCOME /CASES=FROM 10 TO 60  
/FORMAT=SINGLE NUMBERED.
```

这是一条显示样品信息的命令，LIST 是显示命令关键字，VARIABLES、CASES 和 FORMAT 是子命令关键字，AGE 和 INCOME 是变量名，FROM、TO、SINGLE 和 NUMBERED 是关键字。

二、命令中参数的约定

1) 文件名、变量名不能超过 8 个字符，首字符必须是字母，其后可跟数字或字母，但不能有空格，建立源程序时变量不能拆开在两行写。给变量或文件取名不能与系统关键字相重。取名要尽量便于记忆，最好结合实际意义。特别要注意，文件名需要用引号括起来。

2) 定义变量表时可以使用关键字 TO，以使书写简便。例如定义变量 A1, A2, A3, ..., A10 可以写成 A1 TO A10。又例如用一系列变量名定义了一个变量表：AGE, SEX, NAME, EDUCATN, SCORE1 TO SCORE5，当统计命令使用它时可写成 AGE TO SCORE5。

3) 命令说明部分的关键字一般指保留关键字，它们各有特定的意义，例如 TO、BY、WITH、EQ、LT、AND、OR、NOT、LOWEST、THRU、HIGHEST 等。

4) 变量值可以是数值型或字符型，数值型值可以是整数或小数，字符型值必须用引号括起来。

5) 标号是指对变量或变量值所作的说明，也可以说标号是变量或变量值的扩充信息，它可以是短语，也可以是句子。标号虽然可以跨行书写，但它的长度有限制，这一点将在相应的命令中详讲。例如对变量 EDUCATN 加标号‘EDUCATION OF HEAD OF HOUSEHOLD’，说明变量 EDUCATN 是指家庭户主的受教育程度。

6) 分隔符有普通的与专用的两种。两个元素或两个变量之间一般用空格或逗号分隔，空格与逗号是普通分隔符；函数符号与自变量用圆括号分隔，子命令关键字与参数用等号分隔，子命令之间用斜杠分隔等，这些分隔符都属于专用分隔符。

7) 算术运算符是+、-、*、/、**等，分别对应加、减、乘、除、乘方等运算。

第三节 SPSS / PC 的工作方式

运行 SPSS / PC 软件包有两种方式：对话方式和批作业方式。在讲这两种方式之前，先简单介绍一下如何进入 SPSS / PC 环境。

一、SPSS / PC 系统的进入与退出

IBM PC 机或其兼容机的用户，必须首先把存有 SPSS / PC 软件包的 10 块软盘上的信息全部拷贝到硬盘 SPSS 子目录中。以后每次上机时只需用西文 DOS 启动，设置工作盘为 C 盘，然后按以下几步键入信息：

1) 在 DOS 提示符 C> 下，键入 CD \ SPSS (回车) (进入 SPSS 子目录)；

2) 仍在提示符 C> 下，键入 NOKEY (回车)；

3) 当再出现提示符 C> 时，键入 SPSSPC (回车)，随后出现提示符 SPSS / PC:，表示已进入 SPSS / PC 系统，此时可以使用 SPSS / PC 的一切命令。

不管是在何种工作方式下，当要结束 SPSS / PC 软件包的运行时，在 SPSS / PC: 提示符下键入 STOP 命令，即可退出 SPSS / PC 系统返回 DOS。

二、对话方式

在 SPSS / PC: 提示符下，键入一条 SPSS / PC 命令，系统就执行这条命令。当执行这条命令成功时，输出执行结果，以及命令执行过程中占用计算机资源的情况，如 CPU 时间、存贮空间等，并提示该命令的使用限制，然后又出现提示符 SPSS / PC:，再键入一条命令……如此反复进行；如果执行命令失败，则显示出错信息，给出 ERROR NO(错误号)及错误性质，然后出现提示符 SPSS / PC:，等待重新输入。如输入的命令在一行中写不下，可以续行，换行时按回车键，换行后的提示符将变为 :，冒号前空出 7 个字符的位置。

当要注销键入的当前行时，在未按回车键之前，按 ESC 键。

当要注销当前命令时，按 Ctrl+Break 键。

DOS 下的编辑功能键 F1~F3 在 SPSS / PC: 下仍然有效。

三、批作业方式

对话方式只适用于处理简单的统计分析问题，因而一般常采用批作业工作方式。使用批作业工作方式时，首先要根据具体统计问题编写 SPSS 的命令文件(源程序)，命令文件用 EDLIN 等编辑软件建立，并且存贮在 C 盘 SPSS 子目录中，然后进入 SPSS / PC 系统，在 SPSS / PC: 提示符下，键入执行命令 INCLUDE '文件名' (回车)，这里的文件名是指用户指定的命令文件名。这个命令文件立即运行，与此同时输出相应的结果。

第四节 SPSS / PC 的文件

SPSS / PC 软件包涉及的文件主要有 8 种类型:

文件定义	扩展名	文件定义	扩展名
命令文件	.INC	列表文件	.LIS
数据文件	.DAT	携带文件	.POR
活动文件	.ACT	帮助文件	.HLP
系统文件	.SYS	记录文件	.LOG

1. 命令文件(Include file)

命令文件就是在 EDLIN 等编辑软件中用 SPSS / PC 语言编写的源程序。用 INCLUDE 命令可以把命令文件提交给 SPSS / PC 系统执行, SPSS / PC 系统将把命令文件执行的过程写入记录文件。

2. 数据文件(Data file)

数据文件一般是指包含原始数据的外部文件, 如果是 ASC II 码文件, 其数据按固定格式或自由格式组织。DATA LIST 命令是定义读数据文件的命令。SPSS 统计过程中产生的数据矩阵文件也是一种数据文件, 统计命令中的 WRITE 子命令将这种数据文件写入由 SET RESULTS 命令定义的盘文件中。

3. 活动文件(Active file)

活动文件是直接提供给统计命令使用的文件, 是属于 SPSS / PC 系统的内部文件, 它包含数据和字典信息, 其组织结构是: 文件—样品—变量。而外部数据文件的结构是文件—记录—字段, 不能直接供统计命令使用, 必须通过 DATA LIST 命令转换成活动文件。在生成活动文件的同时还产生 3 个系统变量, 它们是 \$ DATE、\$ CASENUM 和 \$ WEIGHT。系统变量名以 \$ 为首字符, \$ DATE 表示读入样品数据的日期; \$ CASENUM 表示样品的序号; \$ WEIGHT 表示样品的权重, 其初始值为 1, 在使用 WEIGHT 命令之后, 其值将改变。注意不能把系统变量名定义为用户变量名, 一般也不能在命令中对系统变量进行操作。

4. 系统文件(System file)

把活动文件中的数据及描述数据的字典信息存入磁盘, 则生成系统文件。系统文件通过 SAVE 命令生成, 文件名可由用户定义, 隐含文件名为 SPSS.SYS。用 GET 命令可取出系统文件, 以生成活动文件供统计命令使用。

5. 携带文件(Portable file)

携带文件也是由活动文件存盘生成, 包含活动文件的数据和字典信息。它用于不同版本 SPSS 之间的通讯。EXPORT 和 INPORT 分别是存、取携带文件的命令。

6. 列表文件(List file)

列表文件是 SPSS / PC 作业运行的结果, 可以将其输出至显示器、磁盘和打印机。输出方式由环境设置命令 SET 选择, 隐含在显示器上输出, 写盘时隐含文件名为 SPSS.LIS。

7. 记录文件(Log file)

记录文件是 SPSS / PC 作业运行过程的记录。它包括执行过的 SPSS 命令、执行过程中的出错信息、输出结果及计算机资源占用情况。

8. 帮助文件(Help file)

帮助文件是 SPSS / PC 软件包提供给用户使用的，它告诉用户如何使用 SPSS / PC 软件包，它可以列出各种统计命令的格式，以及它们所使用的关键字。在 SPSS / PC: 提示符下键入 HELP.(回车)，便显示出如何获得进一步帮助的提示信息。再键入 HELP TOPICS.(回车)则可得到帮助文件的内容提要，据此提要选择地键入帮助命令就可得到需要的帮助信息。

必须强调指出，SPSS / PC 软件包的统计命令只能处理活动文件。活动文件与其它文件的关系可用图 1 表示。

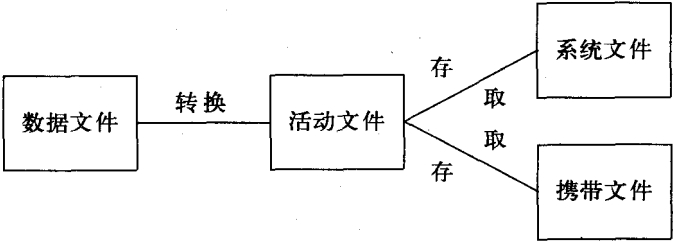


图 1

第二章 SPSS / PC 的数据和文件管理系统

通过第一章的介绍, 我们已经知道 SPSS / PC 软件包的统计命令只能直接处理活动文件, 这是因为活动文件中的数据是按统计数据的特点来组织的。因此, 在作统计分析之前要先把原始数据文件转换成活动文件, 这个转换是通过数据定义命令来完成的。另外, 为了统计分析的方便, 还经常需要对活动文件中的变量、样品进行操作, 对文件进行管理, 这些工作均由 SPSS / PC 软件包中相应的一系列命令来完成, 这样的一系列命令组成了 SPSS / PC 的数据、文件管理系统。

第一节 SPSS / PC 的数据定义与转换命令

一、数据定义命令

1. DATA LIST 命令

命令的功能: 定义一个活动文件。当遇到其后第一条统计命令时, 把原始数据按一定的格式读入到这个活动文件中。

命令的格式: 定义活动文件根据数据来源的不同而有不同的格式, 一种情况是活动文件的数据需从数据文件中读入, 另一种情况是把原始数据置于命令文件之中, 下面通过一个示意性的简单例子来说明:

例: 在一次问卷中, 获得 10 个人年龄、性别、收入的数据, 把这 10 个样品的原始数据转变成 SPSS / PC 的活动文件。

第一种方法(从原始数据文件中读入数据):

1) 用 EDLIN 行编辑软件建立一个包含 10 行原始数据的数据文件, 取名 RAW.DAT, 存在 C 盘 SPSS 子目录中(在 C>提示符下键入 EDLIN \SPSS\RAW.DAT (回车), 然后输入数据)。注意每个人的数据占一行并按固定的格式输入, 例如变量 AGE (年龄) 的数据占第 1、2 列, 变量 SEX (性别) 的数据占第 4 列, 两变量之间空一列, 变量 INCOME (收入) 占第 7~10 列 (前 3 列是整数位, 后一列是小数位), 它与 SEX 变量之间空两列。数据文件的内容是:

```
65 1 2500
52 1 1750
51 1 1460
51 1 1720
53 1 1760
51 1 1460
46 2 1250
50 2 1030
57 2 1770
```

49 2 1250

其中第 4 列对非数值型变量 SEX 进行了编码, 1 表示男, 2 表示女。当然, 变量 SEX 也可以用字母表示, 例如, M 表示男, F 表示女, 此时 SEX 是字符型变量。

2) 进入 SPSS/PC 系统, 在 SPSS/PC: 提示符下输入一条数据定义命令:

DATA LIST FILE = 'RAW.DAT' / AGE 1-2 SEX 4 INCOME 7-10(1).(回车)

当这条命令执行完, 例题要求的活动文件就定义好了。

这里 DATA LIST 是命令关键字; 命令的说明部分自 FILE 子命令开始, 由它指定一个数据文件的名称, 此例的文件名为 RAW.DAT; 斜杠后给出一个样品数据的格式, 系统将按照这个格式把外部数据文件 RAW.DAT 中的数据读入活动文件。这里规定的格式必须与数据文件的组织格式是一致的, 即一个样品对应一行数据, 每行数据的第 1、2 列赋给变量 AGE, 第 4 列赋给变量 SEX, 第 7~10 列赋给实数型变量 INCOME, 第 10 列是小数位上的数据。若取 SEX 为字符型变量, 则命令中的 'SEX 4' 应改为 'SEX 4(A)', 字母 A 是字符型变量的标志。

由此, 不难写出 DATA LIST 命令的一般格式:

DATA LIST FILE = '文件名' / 变量名 数据读入格式.....

第二种方法(原始数据置于命令文件中):

1) 编写一个命令文件, 取名为 EXAMP.INC, 存在 C 盘 SPSS 子目录中。命令文件的内容如下:

DATA LIST / AGE 1-2 SEX 4 INCOME 6-9(1).

BEGIN

65 1 2500

52 1 1750

51 1 1460

51 1 1720

53 1 1760

51 1 1460

46 2 1250

50 2 1030

57 2 1770

49 2 1250

END DATA.

这里原始数据放置在命令文件中, 并且用 BEGIN DATA 与 END DATA 把这些数据行括起来。与第一种方法下的 DATA LIST 命令格式比较, 省略了 FILE 子命令而其它内容不改变。用这个命令定义的活动文件生成时, 也是按斜杠后给定的格式把数据行的数据读入活动文件中。

2) 进入 SPSS/PC 系统, 在 SPSS/PC: 提示符下输入执行命令:

INCLUDE 'EXAMP.INC'。(回车)

即可定义好活动文件, INCLUDE 是执行命令的关键字, 其说明部分指定一个命令文件的名称。显然这个命令的功能就是向系统提交待执行的命令文件。

通过以上这个例子基本上可以了解 DATA LIST 命令的功能、格式和使用,

DATA LIST 命令还可选择 FREE 和 MATRIX 两个关键字，相应地有下面另外两种格式：

DATA LIST FILE='文件名' FREE / 变量表.

DATA LIST FILE='文件名' MATRIX / 变量表.

可以看出这两种格式与例中第一种方法使用的格式类似，只是在 FILE 子命令后用了关键字 FREE 或 MATRIX，并且斜杠后的变量格式表有所不同，这里只列出变量名，未规定它对应数据行的哪些列。

FREE 表示以自由方式读取数据文件。在 FREE 方式下，所读数据文件不是按每个样品数据占一行组织的，而是所有的数据在文件隐含宽度内依次排列，仅在数据之间用空格分隔。仍以上面的例子来说明，假如变量 SEX 定义为字符型变量，M 表示男性，F 表示女性，数据文件的内容是：

65 M 250.0 52 M 175.0 51 M 146.0.....50 F 103.0.....49 F 125.0

其中数值型数据以空格作为结束标记，它可以为小数；读字符型数据以命令中指定的宽度为依据。这时，数据定义命令应写为：

DATA LIST FILE='RAW.DAT' FREE / AGE SEX(A1) INCOME.

例中变量 SEX 后面的(A1)表示读宽度为 1 的字符型数据，字符 A 后的数字表示字符型数据的宽度，若 A 后不写数字即(A)，则表示读宽度为 8 的字符型数据，这是隐含约定。

关键字 MATRIX 表示读入一个矩阵。这个矩阵的意义由配合使用的统计命令所选择的 READ 子命令中的关键字表示出来，例如作回归分析时直接读入变量的相关系数矩阵，命令是：

DATA LIST FILE='CORRMAT' MATRIX / V1 TO V20.

REGRESSION READ CORRELATION / VARIABLES=V1 TO V20

/DEPENDENT=V1 /METHOD=STEPWISE.

必须注意，在运行 SPSS 作业过程中只允许存在一个活动文件。如果命令文件里不止有一个 DATA LIST 命令，那么当执行到后面的 DATA LIST 命令时，前面的活动文件将被该命令所定义的新的活动文件取代。

2. VARIABLE LABELS 命令

命令的功能：给活动文件中的所有变量或部分变量定义标号。标号是一短语，用来注释变量的意义。定义标号的长度有限制，不能超过 40 个字符，但一个标号可以折在两行写。由 VAR LABELS 命令定义的标号不仅作为字典信息进入活动文件，而且可在输出文件中使用，以提高输出结果的可读性。

命令的格式：

VARIABLE LABELS 变量名 '标号'.

例句：

VAR LABELS AGE 'AGE OF THE RESPONDENT' SEX 'SEX OF THE
RESPONDENT' INCOME 'MONTHLY INCOME'.

注明变量 AGE 和 SEX 分别表示被调查者的年龄和性别，变量 INCOME 是表示月收入。

3. MISSING VALUES 命令

命令的功能：给样品中的缺失数据定义标记。所谓缺失数据是指某些样品的某些变量的数据没有得到，这些样品就带上了缺失数据。那么在原始数据中如何表示这些缺失数据呢？

不妨把它们定义成这些变量最可能取到的值或不可能取到的值，这取决于选用何种方式处理缺失数据。例如变量 AGE、INCOME 的缺失值可定义为-1，变量 SEX 的缺失值可定义为 'N'。或者干脆把这些缺失数据空起来，由 SPSS/PC 系统把它们定义为系统缺失值，SPSS/PC 系统有一整套处理缺失值的办法。这样，即使是对于带缺失值的样品，也往往能从中挖掘出有用的信息。

命令的格式：

MISSING VALUES 变量名(缺失值).

例句：

MISSING VALUES AGE INCOME (-1) / SEX ('N').

定义 AGE、INCOME 的缺失值为-1，SEX 缺失值为 'N'。

4. VALUE LABELS 命令

命令的功能：给活动文件中变量的取值(包括缺失值)定义标号。这种标号对变量值提供描述性的解释。它们与变量标号一样不仅作为字典信息进入活动文件，而且可在输出文件中使用，提高输出文件的可读性。定义的值标号不能超过 20 个字符。

命令的格式：

VALUE LABELS 变量名 变量值 '标号'.

例句：

VALUE LABELS SEX 'M' 'MALE' 'F' 'FEMALE' 'N' 'NO ASCERTAINED'.

变量 SEX 的值 'M' 表示 'MALE' (男性)，'F' 表示 'FEMALE' (女性)，'N' 表示 'NO ASCERTAINED' (数据缺失)。

5. FORMATS 命令

命令的功能：用这个命令可以定义变量的输出格式。如果命令文件中没有使用这个命令，则活动文件中定义的变量将按 DATA LIST 命令规定的格式输出；活动文件中通过数据转换命令产生的新变量将按隐含格式 F8.2 输出，即输出 8 位，其中有 2 位小数。

命令的格式：

FORMATS 变量名 (格式符).

其中格式符可以使用 FW.d、DOLLARW.d 和 COMMAW.d。其中 W 为数的总位数，d 为小数位数，F 表示实数，DOLLAR 表示在输出的数值前加货币号 \$，COMMA 表示在输出的数值中加分节号。

例句：

FORMATS AGE (F 2.0) INCOME (DOLLAR 6.2).

二、数据转换命令

这类命令用于对活动文件中的变量进行操作，或是修改老变量，或是生成新变量。

1. COMPUTE 命令

命令的功能：做统计分析时，用 COMPUTE 命令可在活动文件中引入作为老变量算术函数的新变量，或者对活动文件中的老变量通过算术运算进行修改。

命令的格式：

COMPUTE 新变量 = 算术表达式.

例句：

COMPUTE AGIN=AGE*INCOME.

COMPUTE INCOME=INCOME+8.

前者使活动文件中已存在的变量 AGE 与 INCOME 相乘得到新变量 AGIN, 后者对已存在的变量 INCOME 进行修改, 使它在原数据上加 8。

算术表达式中能够使用的操作符号和函数有:

+	加法运算	EXP	自然指数函数
-	减法运算	LG10	常用对数函数
*	乘法运算	LN	自然对数函数
/	除法运算	RND	四舍五入函数
**	乘方运算	NORMAL	正态分布的伪随机数函数
ABS	绝对值函数	UNIFORM	均匀分布的伪随机数函数
ATAN	反正切函数	MOD10	取 10 的模
SIN	正弦函数	SYSMIS	对系统缺失值作逻辑检查
COS	余弦函数	MISSING	对用户定义的缺失值作逻辑检查
SQRT	平方根函数	VALUE	把缺失值作为有效值处理
TRUNC	截尾函数	YRMODA	年、月、日函数
LAG	样品的延迟函数		

以上操作符和绝大部分函数都是我们所熟悉的, 这里不作介绍。此外, 只介绍函数 SYSMIS、MISSING 和 VALUE, 它们是用来处理带缺失数据的变量; NORMAL 和 UNIFORM 分别是产生正态分布和均匀分布伪随机数的函数, 如 NORMAL(n)产生服从均值为 0、标准差为 n 的正态分布的随机数, UNIFORM(n)产生在 0~n 区间服从均匀分布的随机数; 函数 LAG(变量名, n)的值是前 n 个样品的指定变量的值。

例句: $Y=X-5$

当 X 取值为用户定义的缺失值或系统缺失值时, 由上式计算 Y 就会失败, 因此必须先用 SYSMIS 与 MISSING 函数检查 X 是否是缺失值, 如果 X 是缺失值, 则用 VALUE(X) 函数处理。在讲述 IF 命令之后, 可以对这个例句进行修改。

2. IF 命令

命令的功能: 类似于 COMPUTE 命令, 但变量的转换必须在 IF 命令指定的逻辑条件下才能进行。

命令的格式:

IF (逻辑表达式) 新变量名=算术表达式.

逻辑表达式中能够使用的关系符和逻辑操作符有:

GT	大于符	NE	不等于符	AND	逻辑与运算符
LT	小于符	LE	小于等于符	OR	逻辑或运算符
EQ	等于符	GE	大于等于符	NOT	逻辑非运算符

例句:

IF (SYSMIS(X) OR MISSING(X)) $Y=VALUE(X)-5$.

这是上例句的修改语句。表示当 X 为系统的缺失值或用户定义的缺失值时, 把 X 的缺失值当成有效值处理, 将其减去 5 后作为新变量 Y 的值。

3. RECODE 命令

命令的功能：对活动文件中的变量进行重编码。作统计分析前，往往需要对活动文件中的变量进行重编码。例如作变量 AGE 与 INCOME 的列联表，考察老、中、青三种人分别在高收入、中等收入、低收入三个档次中所占的比例。首先要对变量 AGE 与 INCOME 重编码，比如规定变量 AGE 取值在 35 以下为青年用 1 表示，36~55 为中年用 2 表示，55 以上为老年用 3 表示；变量 INCOME 取值在 100 以下为低收入用 1 表示，100~200 为中等收入用 2 表示，200 以上为高收入用 3 表示。这样变量 AGE 与 INCOME 各划分为 3 个类目，可以作出 3×3 的列联表。活动文件中经过重编码的变量值是用新值代替了老值。重编码的变量可以是各种类型的变量。

命令的格式：

RECODE 变量名(重编码表).

对于数值型变量，重编码前的一些数值可以选用以下关键字表示：

LOWEST	表示最低值	THRU	用来表示某一取值范围
HIGHEST	表示最高值	MISSING	表示用户定义的缺失值
SYSMIS	表示系统缺失值	ELSE	表示其它尚未重编码的值

重编码后的数值只可选用 SYSMIS 关键字。

对于字符型变量，以上关键字均不能使用，并要注意字符变量重编码前后的字符串长度应该相同，字符变量值必须用单引号括起来。

例句：

```
RECODE AGE(LOWEST THRU 35=1)
      (36 THRU 55=2) (56 THRU HIGHEST=3)
      /INCOME(LO THRU 100=1) (100 THRU 200=2) (200 THRU H1=3).
RECODE SEX('M'='1') ('F'='2') ('N'='3').
```

注意重编码变量取值为实数时，为避免漏掉数值，必须使重编码各档中上档的下界值与下档的上界值相重合。比如例句中的 INCOME 变量，取值在 100~200 的重编码为 2，上界值为 200。其上一档即取值大于 200 的重编码为 3，下界值定为 200，若定为 201，则 200.5 就会被漏掉。

如果要在活动文件中保留变量重编码前的值，可通过 COMPUTE 命令引入新变量，对新变量重编码而保留老变量值。

例句：

```
COMPUTE AGE1=AGE.
COMPUTE INCOME1=INCOME.
```

引入新变量 AGE1、INCOME1，可对其重编码而保留变量 AGE、INCOME 的值。

4. COUNT 命令

命令的功能：记下样品中指定的标准变量取值在值表中的数目。

命令的格式：

COUNT 计数变量=标准变量表(值表).

例句：

```
COUNT A1=AGE INCOME (3).
```

表示指定活动文件中经重编码后的变量 AGE、INCOME 为标准变量，当一个样品变量