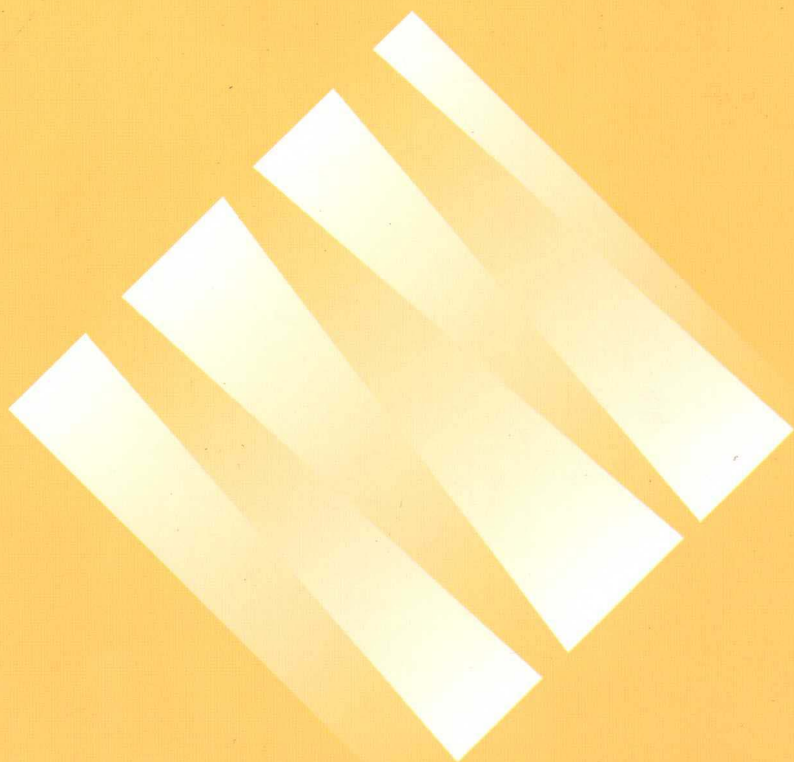


统计分析及其SAS实现

Statistical Analysis with SAS

张晓冉 编著



清华大学出版社

统计分析及其SAS实现

Statistical Analysis with SAS

张晓冉 编著

清华大学出版社
北京

内 容 简 介

本书介绍了常用统计分析方法及其在 SAS 软件中的实现. 全书共 11 章, 内容包括: 统计分析的 SAS 基础, 描述统计, 假设检验, 线性回归分析与回归诊断, 方差分析, 分类数据的列联表分析, 主成分分析, 因子分析, 典型相关分析, 判别分析和聚类分析. 每章末附有一定数量的习题, 用于补充正文或提供统计模型. 书中在讲述各种统计方法时, 力争使统计学原理、实例和 SAS 程序及输出有机结合, 以增强读者对统计学中问题和原理的理解.

本书可作为统计学、概率论与数理统计和经济管理等专业高年级本科生统计分析软件课程的教材使用, 也可供相关专业研究生、青年教师和统计分析工作者参考.

版权所有, 侵权必究. 侵权举报电话: 010-62782989 13701121933

图书在版编目(CIP)数据

统计分析及其 SAS 实现/张晓冉编著. --北京: 清华大学出版社, 2011.5
ISBN 978-7-302-25231-3

I. ①统… II. ①张… III. ①统计分析—应用软件, SAS IV. ①C812

中国版本图书馆 CIP 数据核字(2011)第 061868 号

责任编辑: 陈 明

责任校对: 赵丽敏

责任印制: 王秀菊

出版发行: 清华大学出版社

地 址: 北京清华大学学研大厦 A 座

<http://www.tup.com.cn>

邮 编: 100084

社 总 机: 010-62770175

邮 购: 010-62786544

投稿与读者服务: 010-62776969, c-service@tup.tsinghua.edu.cn

质 量 反 馈: 010-62772015, zhiliang@tup.tsinghua.edu.cn

印 装 者: 北京市清华园胶印厂

经 销: 全国新华书店

开 本: 185×230

印 张: 16.75

字 数: 365 千字

版 次: 2011 年 5 月第 1 版

印 次: 2011 年 5 月第 1 次印刷

印 数: 1~3000

定 价: 28.00 元

产品编号: 039632-01

前 言

随着社会进步和经济发展,越来越多的社会和经济问题需要对数据进行定量的统计分析,以提供给决策者最优决策方案.为满足社会经济发展对高校教学的需要,国内各大院校相继为统计学专业和经济管理等专业高年级大学生开设了统计分析软件课程,以培养学生分析和解决实际统计问题的能力,增强其社会竞争力.

统计分析软件课程主要讲授统计分析方法如何在统计分析软件中实现.值得注意的是,要想真正具有处理实际统计问题的能力,不仅要会用统计软件,更重要的是要理解、掌握各种统计分析方法.也就是说,统计分析软件只是统计计算的工具,统计分析离不开统计工作者.统计分析软件的使用也有助于统计工作者对统计方法的认识和理解.

众所周知,理论与实践相结合会使学习效率最高,教学效果也最好,这是因为在解决实际问题的过程中对理论的认识最深刻.在讲授统计方法原理时,如果能将统计方法的背景和原理通过实例介绍清楚,再随之讲授该方法在软件中怎样实现,这样会更易于为学生所理解和接受,这也是本书追求的目标.本书在介绍各种统计分析方法时,都是先介绍方法的背景思想和原理,然后结合实例介绍方法在 SAS 软件中的实现.介绍统计分析方法时,希望用尽可能少的数学推导,尽量把统计原理讲清楚.在讲述统计方法的 SAS 实现时,先是解释程序中重要语句和选项,然后解释软件输出中的重要内容,最后给出结论.全书力争做到理论与应用自然的结合,且力争建立起理论方法与软件应用之间的桥梁,建立起软件输出与理论公式间的联系,使之更易于读者理解、接受和应用.

本书在注重统计知识应用与实现的同时,也注重统计理论的讲述.内容的选取具有相对独立性.本书对多数与应用有关的例题和习题,直接给出了解题过程或解决方法,以使读者可以对照所学知识,更方便地解决实际问题中遇到的问题.对与理论有关的中等难度习题,本书也提供了证明方法或思路,以帮助读者更好地理解书中讲授的原理和方法.

全面讲授本书所需要的学时可如下配备(括号内为上机学时):第 1 章统计分析的 SAS 基础 6(6)学时,第 2 章描述统计 4(2)学时,第 3 章假设检验 10(4)学时,第 4 章线性回归分析与回归诊断 8(4)学时,第 5 章方差分析 8(4)学时,第 6 章分类数据的列联表分析 6(3)学时,第 7 章主成分分析 4(2)学时,第 8 章因子分析 6(2)学时,第 9 章典型相关分析 6(2)学时,第 10 章判别分析 8(4)学时,第 11 章聚类分析 6(3)学时,共讲授 72 学时左右(上机 36 学时).

授课教师可以根据不同要求适当调整学时和授课内容. 如果对学生统计学理论要求不是很高, 授课时可以略去部分理论与计算细节, 而只讲授统计分析原理与上机实现部分. 例如, 本书中的部分理论推导放到了每章的补充与习题一节, 可供教师选讲. 再如, 可略去 3.2.1* 节似然比检验统计量的推导, 5.1.4* 节可以只讲授多重比较的重复 t 检验和 Bonferroni t 检验. 再有, 5.1.3* 节、5.2.2* 节、5.4* 节、6.4* 节、8.2.3* 节和 11.2* 节都可作为可选内容. 在讲授本书的同时, 结合上机实践效果更佳. 如果条件允许, 应适当增加上机学时.

由于编者水平有限, 书中难免有错误和不恰当之处, 敬请专家和读者批评指正.

编 者

2010.10.5

目 录

第 1 章 统计分析的 SAS 基础	1
1.1 SAS 概述	1
1.1.1 显示管理方式	2
1.1.2 系统配置	3
1.2 SAS 编程基础	4
1.2.1 数据步基本语句	5
1.2.2 input 语句	6
1.2.3 数据集与数据库	9
1.3 几个常用过程	10
1.3.1 print 过程	10
1.3.2 sort 过程	11
1.3.3 gplot 过程	12
1.3.4 transpose 过程(转置数据集)	14
1.3.5 standard 过程(标准化数据)	15
1.4 常用结构语句	16
1.4.1 选择结构	16
1.4.2 循环结构	19
1.5 SAS 数据集管理	22
1.5.1 编辑数据集	22
1.5.2 数据集的横向合并	23
1.5.3 与其他系统数据集间的转换	26
1.6 SAS 的常用统计函数	27
1.6.1 随机数生成函数	28
1.6.2 累积分布函数和概率密度函数	31
1.6.3 分位数函数	32
1.7 补充与习题	33

第 2 章 描述统计	38
2.1 描述性统计量	39
2.1.1 描述总体分布位置的统计量	39
2.1.2 描述总体分散程度的统计量	42
2.1.3 描述总体分布形状的统计量	43
2.2 图形描述总体	44
2.2.1 条形图和直方图	45
2.2.2 箱线图(Box-and-Whisker)	47
2.3 补充与习题	49
第 3 章 假设检验	51
3.1 假设检验与区间估计	51
3.1.1 参数的假设检验	51
3.1.2 参数的区间估计	53
3.2 正态总体均值和方差的检验	54
3.2.1* 正态总体均值检验	55
3.2.2 正态总体方差检验	57
3.3 两总体均值比较	61
3.3.1 独立组均值比较	61
3.3.2 成对组均值比较	66
3.4 正态性检验	70
3.4.1 χ^2 检验法	71
3.4.2 经验分布函数(ECDF)检验法	74
3.4.3 Shapiro-Wilk 检验	74
3.5 补充与习题	76
第 4 章 线性回归分析与回归诊断	78
4.1 统计模型	78
4.2 模型参数的最小二乘估计和因变量的预测	79
4.2.1 模型参数的最小二乘估计及其性质	79
4.2.2 因变量的预测	80
4.3 回归分析中的假设检验	81
4.3.1 回归方程的显著性检验	81
4.3.2 单个回归系数的显著性检验	82
4.3.3 回归系数的一般线性假设检验	83

4.3.4	应用例子	85
4.4	回归诊断	91
4.4.1	残差分析	92
4.4.2	影响分析	94
4.4.3	多重共线性	97
4.5	多重共线性处理方法	101
4.5.1	筛选回归变量	101
4.5.2	岭回归	107
4.6	补充与习题	110
第 5 章	方差分析	114
5.1	单因子方差分析	114
5.1.1	模型及其显著性检验	114
5.1.2	方差齐性检验	117
5.1.3*	可估函数与对比	119
5.1.4*	多重比较	121
5.2	无交互效应的两因子方差分析	127
5.2.1	模型及因子效应的显著性检验	127
5.2.2*	因子显著性检验统计量分布的推导	129
5.2.3	多重比较与同时置信区间	130
5.3	有交互效应的两因子方差分析	133
5.3.1	模型及因子效应的显著性检验	134
5.4*	一般线性模型(glm)过程	138
5.5	补充与习题	139
第 6 章	分类数据的列联表分析	143
6.1	用 freq 过程生成列联表	143
6.2	单向表分析	148
6.2.1	二项分布参数的检验与区间估计	148
6.2.2	多项分布参数检验	148
6.3	双向表分析	151
6.3.1	独立性的 χ^2 检验	151
6.3.2	小样本列联表独立性的 Fisher 精确检验	152
6.4*	2×2 表分析	155
6.4.1	风险和风险差	155

6.4.2	相对风险和优比比率	156
6.5	补充与习题	160
第 7 章	主成分分析	163
7.1	总体主成分	163
7.1.1	主成分定义及求法	163
7.1.2	主成分性质	164
7.1.3	标准化变量的主成分	166
7.2	样本主成分与主成分得分	168
7.2.1	样本主成分	168
7.2.2	主成分得分	169
7.3	补充与习题	171
第 8 章	因子分析	174
8.1	正交因子模型及模型中各量的统计意义	174
8.1.1	正交因子模型	174
8.1.2	正交因子模型的性质	175
8.1.3	因子载荷阵中各量的统计意义	176
8.2	模型参数估计	177
8.2.1	模型参数的主成分估计	177
8.2.2	模型参数的主因子与迭代主因子估计	178
8.2.3*	模型参数的极大似然估计	183
8.3	因子旋转与因子得分	183
8.3.1	因子旋转	183
8.3.2	因子得分	184
8.4	补充与习题	188
第 9 章	典型相关分析	190
9.1	总体典型相关变量	190
9.1.1	第一对典型相关变量的求法	190
9.1.2	一般典型相关变量求法	193
9.1.3	典型相关变量的性质	194
9.2	典型相关变量系数的估计和典型变量得分	195
9.3	典型相关系数检验	197
9.3.1	第一个典型相关系数的检验	197

9.3.2	一般典型相关系数的检验	201
9.4	补充与习题	203
第 10 章 判别分析		206
10.1	多总体距离判别	206
10.1.1	马氏距离	206
10.1.2	马氏距离判别	207
10.2	正态总体判别分析中的检验问题	209
10.2.1	多元方差分析——MANOVA	209
10.2.2	多元正态两总体均值比较	210
10.2.3	多元正态总体协差阵的齐性检验	210
10.3	Bayes 判别分析	211
10.3.1	先验概率的确定	212
10.3.2	后验概率与广义平方距离	212
10.3.3	Bayes 判别准则	213
10.4	非参数判别	217
10.4.1	密度函数核估计	218
10.4.2	k -最近邻估计	219
10.5	逐步判别与 stepdisc 过程	221
10.5.1	逐步判别的理论基础	222
10.5.2	逐步判别的计算	223
10.6	补充与习题	228
第 11 章 聚类分析		231
11.1	系统聚类法	231
11.1.1	系统聚类法分类	231
11.1.2	系统聚类法的性质	235
11.1.3	类个数的确定	236
11.2	动态聚类法与 fastclus 过程	244
11.2.1	初始凝聚点的选取	245
11.2.2	初始分类与修改分类	246
11.3	补充与习题	250
附录 A 辅助知识		252
A.1	矩阵知识	252

A. 1. 1	分块矩阵的行列式与逆	252
A. 1. 2	投影阵	253
A. 2	回归问题	254
A. 3	随机向量的投影与相关系数	255
参考文献		258

第 1 章

统计分析的 SAS 基础

1.1 SAS 概述

SAS(Statistical Analysis System)由美国 SAS 软件研究所研制,是一套大型集成式应用软件,具有完备的数据存取、管理和分析功能.在数据处理和统计分析领域,SAS 被誉为国际权威软件,广泛应用于科研、教育、行政管理、生产和金融等不同领域.本书目的是介绍常用统计分析方法,用到的工具是 SAS 8.2 和 SAS 9 版本.

SAS 是由多个功能模块组合而成的集成软件,其核心是 SAS/BASE 模块,即运行 SAS 首先要启动 SAS/BASE 模块. SAS/BASE 模块可单独存在,也可以与其他模块共同构成一个完整的系统.在 SAS/BASE 基础上,可以增加不同模块以达到增加不同功能的目的.可增加的常用模块有 SAS/STAT(统计分析),SAS/GRAPH(绘图),SAS/ETS(经济计量学与时间序列),SAS/ACCESS(访问外部数据库),SAS/IML(交互式矩阵语言)等.

本书涉及的 SAS 内容包括 SAS/BASE, SAS/STAT 和 SAS/GRAPH. SAS/BASE 有自己完整的语言结构,并且提供了各种内置函数(如随机数生成函数和分位数函数),使用户可以方便地实现各种统计计算要求. SAS/STAT 提供了多个统计分析过程,每个过程都含有多个选项,使用户轻松控制统计计算方法和程序输出内容. SAS/GRAPH 是一个绘图模块,可以绘制多种数据图.

注 1.1 SAS 支持多个操作系统.拿到 SAS 软件后,首先要看其授权(license)文件,确认其 license 属哪个操作系统,是 PC 版还是 Server 版.如果 license 是 Windows PC 版,则 SAS 只能在 Windows 专业版操作系统上使用,而不能在 Windows 服务器版上使用.另外还需注意,计算机的系统时间必须在 SAS 的授权期内,否则 SAS 不能运行.

注 1.2 如果安装 SAS 软件时没用到所有光盘,那是因为没有完全安装 SAS.建议完全安装 SAS,因为完全安装后只要有可行 license,即可使用相应的模块功能.

在 Windows 环境下,常用的 SAS 启动方法有两种:

- (1) 单击任务栏“开始→程序→The SAS System→The SAS System for Windows”.
- (2) 建立 SAS 程序的桌面快捷启动方式,通过单击桌面快捷方式启动 SAS,方便快捷.

1.1.1 显示管理方式

显示管理方式是运行 SAS 程序最常用的方式,它通过窗口、菜单和工具栏来运行、调试 SAS 程序,实现人机对话.显示管理方式有三个基本窗口(Editor, Log, Output)和一些特殊功能窗口(Explorer, Results, Help, Options, Keys, Libname 等),下面予以简单介绍.

(1) Editor 编程窗口,提供编写 SAS 程序的文本编辑器,但对汉字兼容性较差.建议不要在 SAS 程序中写汉字,如果非要写汉字,应在记事本或 Word 文档中将程序写好,然后复制到 Editor 窗口.对有汉字的 SAS 程序行,不要在 Editor 窗口中编辑.

(2) Log 日志窗口,用于记录数据步和过程步的执行情况、所花机时以及程序错误信息等.

(3) Output 输出窗口,在该窗口显示 SAS 程序运行的输出结果.

(4) Explorer 资源管理器窗口,默认 Explorer 窗口中包括 SAS 数据库、SAS 文件关联库、喜好文件夹、我的电脑.用户可在其中建立新的 SAS 数据库和 SAS 文件关联(参见 1.2.3 节).

(5) Result 结果窗口,其中条目是程序输出的超链接,以方便用户查找 Output 窗口中的输出.

(6) Help 帮助窗口,提供 SAS 的各种帮助信息,窗口快捷键为 F1. SAS 软件提供了丰富的在线帮助,尤其是 SAS 9 及其后继版本.要学好 SAS,需充分利用 SAS 的在线帮助.

(7) Options 设置窗口,用于查看和改变 SAS 的系统设置.

(8) Keys 快捷键设置窗口,用于建立、修改各种功能快捷键.在不同 SAS 版本中,一些命令的快捷键可能不同,但命令快捷键可以通过 Keys 窗口来设置,以满足不同用户需求.此窗口的快捷键一般为 F9.

(9) Libname 数据库窗口,查看已存在的 SAS 数据库关联.单击右键选择 New 选项,可建立新的 SAS 数据库关联.在 SAS 8.2 及其后继版本中,此功能可在 Explorer 窗口实现.

图 1.1 是 SAS 启动后经过调整的窗口.如果对已调好的界面满意,例如图 1.1,希望下次运行 SAS 时还出现同样的界面,则需存储当前设置.在 SAS 8.2 或 SAS 9 中,完成此任务方法如下:通过菜单按顺序选择 Tools→Options→Preferences→General,得到个人喜好的 General 标签,在其中选中 Save settings on exit.在上述 General 标签中,还有一个可选项 Confirm exit,如果不希望每次退出 SAS 都进行确认,则取消该选项.

命令框提供快速输入 SAS 命令的方法,如图 1.2 所示.图 1.2 中命令框的右侧是 SAS 的工具栏,为 SAS 的常用命令和功能提供快捷方式.用户可以通过菜单 Tools→Customize 进入自定义对话框,然后在 Customize 标签中调整工具栏上图标排列顺序.

下面是几个可在命令框中使用的简单命令.

(1) nums Editor 为活动窗口状态下,在命令框中输入 nums,确认(回车或单击其左侧对勾)后会在 Editor 窗口出现行号. nums 为开关命令,即再次执行 nums 后行号消失.

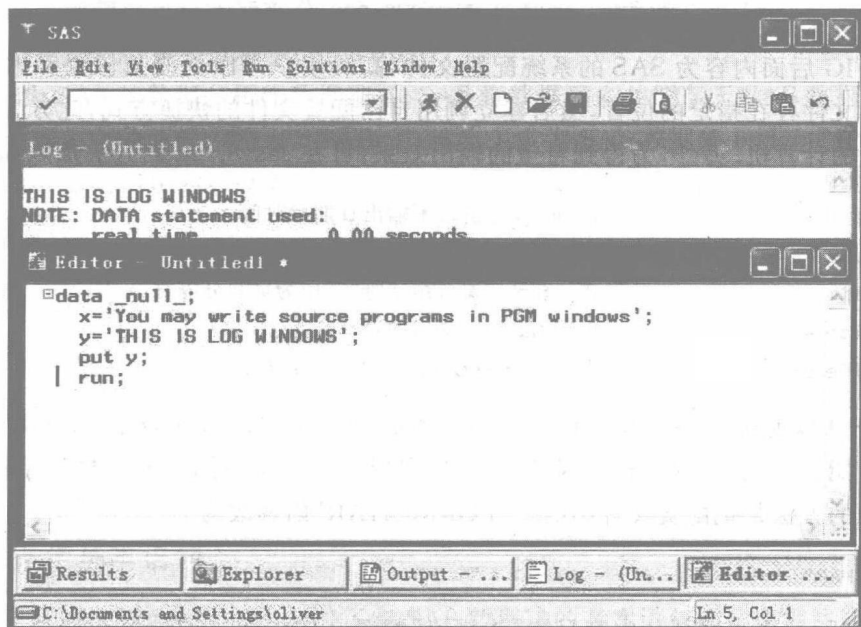


图 1.1 调整后的 SAS 启动界面



图 1.2 Editor 窗口工具栏

(2) x 进入 DOS 状态,此时输入 Exit,返回 SAS. x 命令后可直接加 DOS 命令,如“x dir”.

(3) clear 清除活动窗口内容,快捷键 Ctrl+E. 注意,统计分析的调试过程中会得到很多输出和日志,为清晰起见,通常要把无用的输出和日志清空. 此时很少用 clear 命令,而是在 Log 或 Output 窗口为活动状态下,单击工具栏上的 New 图标来达到清空目的.

(4) end 保留该活动窗口修改内容并退出该窗口,快捷键 F3.

(5) file “filename” 将窗口中内容存储到指定文件,相当于保存某文件.

(6) include “filename” 读入指定文件,相当于打开某文件.

(7) 可在命令框中输入窗口或工作站名称,确认后相应窗口或工作站出现. 例如,keys 得到快捷键设置窗,miner 得到数据挖掘工作站,analyst 得到程序员应用工作站.

1.1.2 系统配置

高级用户根据实际需要,可以配置自己的 SAS 系统文件. 找到 SAS 程序快捷方式启动项,右击打开属性对话框,在快捷方式的目标栏中会看到系统配置文件的位置,如

```
D:\SAS\v8\sas.exe - CONFIG d:\sas\v8\SASV8.CFG
```

其中-CONFIG 后面内容为 SAS 的系统配置文件,高级用户可以更改其位置和内容。例如,可以将配置文件存在指定的位置,然后建立调用自己配置文件的快捷方式,此方法尤其适用于有保护卡的计算机。另外,为得到希望的输出效果,可在 sasv8.cfg 的开头加入如下内容:

```
-nodate           /* 在 Output 窗口不输出日期和时间 */
-nocenter        /* 使 Output 窗口输出内容左对齐 */
-formdlm_       /* 在 Output 窗口用字符“_”作为页间分隔符 */
-pagesize 40     /* 设置 Output 窗口的页行数为 40 */
-linesize 88     /* 设置 Log 和 Output 窗的行宽为 88 */
```

对 SAS 9 以前版本,在 Output 窗口输出图形(如 plot 过程)或表格(如 tabulate, freq 过程)时,可能会出现一些乱码,原因是 SAS 的配置文件中构造图形或表格的字符有问题。解决该问题的方法是把配置文件中的“-FORMCHAR”内容改为

```
-FORMCHAR"|-----|+|-----+=|-\<>*"
```

注意,实际上更改这部分内容只需“注释掉”配置文件中的当前内容和取消上述内容的注释。

有三种方法提交 Editor 窗口中的程序:单击工具栏中行走的小人;单击 Run→Submit (快捷键为 F8);在命令框内输入 submit 命令。

1.2 SAS 编程基础

下面通过实例学习 SAS 编程。

例 1.1 某健身俱乐部有若干男性和女性学员,现有这些人的身体脂肪含量百分比数据(见程序)。要求建立 SAS 数据集,并在 Output 窗口打印输出数据集。

```
data bodyfat;                               /* 建立名为 bodyfat 的数据集 */
  input gender $ fatpct @@;                 /* 输入两个变量,其中 gender 为字符变量 */
  cards;                                   /* 下面为要输入的数据 */
m 13.3 f 22.0 m 19.0 f 26.0 m 20.0 f 16.0 m 8.0 f 12.0 m 30.4
m 18.0 f 21.7 m 22.0 f 23.2 m 20.0 f 21.0 m 31.0 f 28.0 m 20.2
m 21.0 f 30.0 m 12.0 f 23.0 m 16.0 m 12.0 m 24.0 f 40.5 m 15.3
;                                           /* 空语句,告诉 SAS 数据到此结束 */
proc print data= bodyfat;                 /* 打印过程,在 Output 窗口打印数据集 */
  title '男女体脂数据';                   /* 在输出结果的每页开头给出标题 */
  footnote "men and women's data of fattiness"; /* 给出脚注 */
run;                                       /* 执行以上语句 */
```

总体上,上面程序可分为两部分:第一部分准备数据,称为数据步(Data Step);第二部分处理(或分析)数据,称为过程步(Proc Step)。

数据步为 SAS 的数据分析作数据准备。SAS 将数据读入并储存于 SAS 数据集中,因为只有 SAS 数据集中的数据才能为 SAS 所用。输出 1.1 为 SAS 数据集 bodyfat 在 Output 窗口打印输出的一部分,由此可见 SAS 数据集的矩阵形状:矩阵的行为观测,列为变量,Obs 列为观测序号。

输出 1.1 男女体脂数据的部分输出

Obs	gender	fatpct
1	m	13.3
2	f	22.0
3	m	19.0

1.2.1 数据步基本语句

结合例 1.1 和下面的例 1.2,可以了解 SAS 数据步的基本语句和几个需注意的地方。

例 1.2 下面程序中给出了一个班学生基本情况的部分数据,要求建立名为 student 的 SAS 数据集,然后在 Output 窗口输出数据集 student。

```
data student;
  input name $ gender $ height weight born : yymmdd10. @@;
  label born='birthday'; format born yymmdd10.;
  cards;
Ray   male   175   68.2   82-2-6      Oliver male   .       74.5   81-12-10
Rose  female 170   62.7   1982/2/25   Branda female 162   52.5   83.1.7
;
proc print data=student; run;
```

1. data 语句 给出要建立的 SAS 数据集的名字。

2. input 语句 告诉 SAS 输入数据集中的变量信息,包括变量名、变量类型(数值型、字符型、日期型、时间型)和变量输入格式。input 语句内容非常丰富,1.2.2 节还将进一步讨论,下面是使用 input 语句最应注意的几点。

(1) SAS 的变量名和数据集名都由字母、数字和“-”(下划线)构成,且必须以字母或“-”开头。

(2) 若数据中有多于一个观测的数据行,则应在 input 语句尾加“@@”(行固定标识符),目的是让 SAS 读完一个观测后读取数据的指针仍停留在当前行。事实上,数据步是个循环过程,默认情况下一个数据步读取一个观测,然后读取数据的指针转向下一个数据行,以进行下一个数据步的读取。若数据中存在多个观测的数据行,为把数据读全,需使用“@

@”符,使一个数据步执行完后读取数据的指针停留在当前位置,而不是转向下一行。

(3) input 语句中字符型变量名后必须加“\$”符。对日期型和时间型变量,也必须在 input 语句中指定变量的输入格式和格式修饰符,如例 1.2 中 born 的输入格式为 yymmdd10.,格式修饰符为“:”。

3. label 语句 标记变量语句。用在数据步中,为变量附加标记。

4. format 语句 指定变量输出格式。如例 1.2 中指定 born 的输出格式为 yymmdd10. 与之类似的还有 informat 语句,用来指定变量输入格式,可替代 input 语句中变量的输入格式。

5. cards 语句 告诉 SAS 下面为数据行。当数据行中有分号或其他特殊符号时,需把 cards 改为 datalines。

6. 缺失值 若数据中有缺失值,需在数据行相应位置加“.”(默认的缺失值标志)。

7. 空语句 数据结束后一个分号独占的一行称为空语句,其作用是结束数据行。若紧接着的语句是以“;”结尾的语句,则空语句可省。

注 1.3 空语句必须单独占一行,即“;”前面不能有数据。

8. 程序注释 为便于分析者调试程序,用户常常要在程序中加些注释,程序运行时并不解析这些注释。SAS 语言中有以下两种程序注释方法:

(1) “* 注释内容;” 注释星号后单个语句。注意,该方法只能注释掉星号与该行第一个分号之间内容。若该行无分号,则系统报错。

(2) “/* 注释内容 */” 注释掉程序中“/*”和“*/”之间内容。该方法在调试大型程序时常用,但要注意不能嵌套使用。

9. title 语句 为每个输出页定义标题。用户最多可定义 10 个标题,如

```
title "身体脂肪含量"; title2 "参加体型锻炼的男女";
```

如果想取消所有 title 定义的内容,只需运行空 title 语句“title;”。

10. footnote 语句 为每个输出页定义脚注,用法与 title 相同。

11. run 语句 执行该语句前面的所有语句,用在每个数据步和过程步的结尾。若程序中有错误,用该语句有助于发现程序中错误的位置。

1.2.2 input 语句

input 语句有着丰富的用法,是 SAS 数据步非常重要的内容。下面结合实例讲述 input 语句的使用方法。

例 1.2 中已经遇到了输入格式修饰符“:”。input 语句有三个输入格式修饰符“:”,“&”和“~”,其作用如下。

(1) “:”——指明变量的输入值是从后面的非空格列中读起,直到指针遇到数据分隔符(通常数据分隔符是空格),或读满指定的变量长度,或到达数据行尾而结束。