

21 世纪高等院校教材

SAS 软件实用教程

第2版

主编 张 瑛 雷毅雄



科学出版社

21 世纪高等院校教材

SAS 软件实用教程

第 2 版

主 编	张 瑛	雷毅雄	
副主编	张丕德	郜艳晖	李丽霞
编 委	(按姓氏笔画排序)		
	叶小华	李丽霞	李燕芬
	邹宗峰	张丕德	张 敏
	张 瑛	周舒冬	郜艳晖
	徐 英	雷毅雄	

科 学 出 版 社

北 京

· 版权所有 侵权必究 ·

举报电话:010-64030229;010-64034315;13501151303(打假办)

内 容 简 介

本书以简明、实用的写作手法,介绍 SAS 软件统计分析的相关基础知识和应用。全书分为两部分共十章,第一部分由第一章和第二章组成,介绍 SAS 系统的操作环境,SAS 语言的语句、函数、程序结构,SAS 数据集的建立与修改;第二部分由第三章至第十章组成,通过“例题-题意分析-程序-结果-结论”的分析主线,介绍 SAS 常用程序在描述性统计和推断性统计中的应用。

本书既可以作为高等院校本科生和研究生《卫生统计学》理论教材的配套的 SAS 实验教材,又可以作为医学工作者学习统计软件的参考书和工具。

图书在版编目(CIP)数据

SAS 软件实用教程 / 张瑛, 雷毅雄主编. —2 版. —北京: 科学出版社, 2014. 6

21 世纪高等院校教材

ISBN 978-7-03-041274-4

I. S… II. ①张… ②雷… III. 统计分析-应用软件-高等学校-教材
IV. C812

中国版本图书馆 CIP 数据核字(2014)第 130734 号

责任编辑:周万灏 杨鹏远 王 颖 / 责任校对:刘亚琦

责任印制:肖 兴 / 封面设计:范璧合

版权所有,违者必究。未经本社许可,数字图书馆不得使用

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>

安泰印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2009 年 4 月第 一 版 开本:787×1092 1/16

2014 年 6 月第 二 版 印张:8

2014 年 6 月第四次印刷 字数:180 000

定价:23.80 元

(如有印装质量问题,我社负责调换)

序

SAS 软件是国际上最为权威的标准数据管理和统计分析的系统,广泛地应用于金融、医药卫生、政府、生产、通讯、社会心理和教育科研等各个领域,被誉为统计软件界的巨无霸。一般的统计教材主要传授理论知识,SAS 软件手册侧重于对程序语句和过程步详细而专业的阐述,这两种书籍对于初学者或不擅长于编程的人来说,要快速掌握 SAS 软件解决数据分析问题有一定的难度。本书摒弃传统手册式介绍软件的写作方法,着意从实用性出发,以《卫生统计学》教材所涉及统计分析方法为范围,沿着“例题—题意分析—程序—结果—结论”的分析主线,详细介绍各种研究目的下各类型数据统计分析的 SAS 程序设计思路与技巧,并结合例题的专业背景对程序运行的结果进行详尽的解释。本书将统计理论和 SAS 实际应用两者有效结合,可使初学者快速利用 SAS 软件进行统计分析。本书在每个例题分析环节里,以专业背景下的研究目的为出发点,运用统计学理论知识分析例题数据的研究设计特点和数据类型具备的统计方法应有条件,指出解决例题问题适宜选用的统计分析方法和适宜应用的 SAS 过程步,并在解释环节中,依据统计学理论知识,选择适合于该例题程序运行的结果来做出统计推断。在“题意分析—程序—结果”的全过程,强化统计逻辑思维观的建立和 SAS 程序的有效运用,这是一本兼顾统计学知识和 SAS 软件应用的好教材。

我校卫生统计学教研室自 20 世纪 90 年代中期在卫生统计学实验课上开展 SAS 软件教学,多年来积累了丰富的 SAS 软件实验课教学经验,本书的前身即为教研室老师自编的 SAS 实验讲义。在自编讲义修改基础上,本书第 1 版作为《卫生统计学》SAS 软件应用的配套教材已于 2009 年出版,经过 5 年的实践,以内容全面、实例丰富、浅显易懂、实操性强的特点获得读者用户的赞誉。现在,第 2 版已正式出版,增加了实际数据分析中经常用到的数据集的标签、格式和卡方检验中原始数据的分析等内容,使第 2 版的实用性较第 1 版有更进一步的提高。

我鼎力推荐此书给大学生、研究生和卫生从业人员,因为该书较其他 SAS 入门图书更实用易懂,涉及的统计方法也较为全面,能满足医药卫生科研人员日常科研工作的基本统计分析需求。

陈思东

2014 年 4 月

第2版前言

为了更好地配合《卫生统计学》的实验教学工作,我们在校内自编讲义基础上于2009年出版了与《卫生统计学》理论教材配套的《SAS 软件实用教程》,经过5年使用得到的反馈,获得读者们较好的评价,但也发现存在一些不足,为使其与最新软件有更好的结合度并更具实用性,我们决定对第1版进行修订。

本书仍分为两大板块共十章。第一板块由第一、二章组成,主要介绍 SAS 软件的语句、函数、程序结构和 SAS 数据集的建立与修改。第二板块由第三章至第十章组成,介绍 SAS 软件在统计学上的应用,其中第三章为 SAS 常用程序在描述性统计中的应用,第四章至第十章为 SAS 常用程序在推断性统计中的应用。

与第1版相比,第2版修订的内容主要有以下几个方面:①教材程序均以较新的 SAS 9.3 版进行运算;②第二章 SAS 数据集建立的方法中,增加了条件语句、标签语句和输出格式语句;③第六章卡方检验补充了原始数据的统计分析,考虑到关联性分析方法种类繁多,本章仅介绍检验分类变量间的独立性问题,为了避免概念的混淆,把第1版中的关联性分析的名称改为独立性检验;④为保持前后格式和陈述风格的统一,对某些章节的表述进行了适当的调整。

本书保留第1版由浅入深、简明易懂和注重实用性的特点。各章均由例题、题意分析、程序与程序解释、结果与结果解释四部分组成,以帮助初学者通过理解题意、解题思路与方法、程序运用、数据处理结果的研判与推论的过程建立严谨的统计逻辑思维,提高正确运用 SAS 程序分析数据的能力。

书中例题既有来自不同版本的《卫生统计学》和《医学统计学》教材,也有来自编写组教师的科研数据,在书中不一一注明出处,都列入参考文献中,在此向原作者致谢!

本书在修订过程中,得到全体编委的大力支持和通力合作,使本书得以顺利完成修订工作,在此衷心感谢编委们付出的辛勤劳动。感谢李燕芬和陈艳丽老师为书中程序运行验算、文字校对和截图工作付出的努力。

尽管我们力求做得更好,但难免有错漏,敬请读者、前辈和同行批评指正。

张 瑛 雷毅雄

2014 年 3 月

目 录

第一章 SAS 软件概述	(1)
第一节 SAS 简介	(1)
第二节 SAS 窗口工作环境	(2)
第三节 SAS 语言的语句和程序	(4)
第二章 建立 SAS 数据集	(10)
第一节 SAS 数据集概述	(10)
第二节 SAS 数据集的建立	(12)
第三节 SAS 数据集的修改	(15)
第三章 常用统计描述过程	(21)
第一节 定量资料的统计描述	(21)
第二节 定性资料的统计描述	(28)
第三节 统计图制作	(29)
第四章 t 检验	(32)
第一节 单样本资料的 t 检验	(32)
第二节 配对设计资料的 t 检验	(34)
第三节 两独立样本资料的 t 检验	(36)
第五章 方差分析	(39)
第一节 完全随机设计资料的方差分析	(39)
第二节 随机区组设计资料的方差分析	(44)
第三节 析因设计资料的方差分析	(47)
第四节 重复测量资料的方差分析	(50)
第六章 χ^2 检验	(54)
第一节 完全随机设计资料两组频率分布的 χ^2 检验	(54)
第二节 完全随机设计资料多组频率分布的 χ^2 检验	(60)
第三节 配对设计资料的 χ^2 检验	(62)
第四节 分类变量的独立性检验	(63)
第七章 基于秩次的非参数统计	(69)
第一节 单样本资料的符号秩和检验	(69)
第二节 配对设计资料的符号秩和检验	(70)
第三节 独立样本资料的秩和检验	(72)
第四节 随机区组设计资料的秩和检验	(80)

第八章 线性相关与回归	(83)
第一节 线性相关	(83)
第二节 秩相关	(85)
第三节 简单线性回归	(86)
第四节 多重线性回归与相关	(90)
第九章 Logistic 回归	(97)
第一节 非条件 Logistic 回归模型	(97)
第二节 条件 Logistic 回归模型	(103)
第十章 生存分析	(107)
第一节 生存率估计与非参数检验	(107)
第二节 COX 模型	(113)
参考文献	(122)

第一章 SAS 软件概述

第一节 SAS 简介

一、SAS 的创立和发展

SAS (Statistical Analysis System, SAS) 软件是 20 世纪 60 ~ 70 年代美国北卡罗来纳州立大学研制并发展起来的。自 1976 年成立 SAS 软件研究所 (SAS Institute Inc.) 以来,经过多年的发展和完善,SAS 软件已由最初的统计分析软件,成为一个管理和分析数据、编写报告、商业智能的大型集成应用软件系统,具有完备的数据访问、管理和分析、呈现、应用开发及决策支持等功能。目前全球有 120 个国家的 50 000 多家客户都在采用 SAS 解决方案,行业遍及金融、医药卫生、生产、运输、通讯、政府和教育科研等领域。2010 年《财富》全球 500 强企业前 100 名企业中,有 92 家是 SAS 客户。尽管现在 SAS 已不再表示任何含义的首字母缩写,但其数据处理和统计分析仍是它的重要组成部分和核心功能。SAS 软件被誉为国际上标准统计分析软件,目前处于世界领先水平,是全球范围使用最为广泛的统计软件。

二、SAS 系统的组成部分

SAS 系统是一个模块化的组合软件系统,随着版本不断升级,模块逐渐增多,功能也越来越强大,如 2011 年推出的 SAS 9.3 版本中约有 200 多个模块,各个模块之间既相互独立又相互交融补充。本书使用最多的是 Base SAS 模块和 SAS/STAT 模块,书中所有程序均在 SAS 9.3 版本中实现。

Base SAS 模块是 SAS 系统的基础。它既可以单独使用,也可以与其他模块组成一个用户化的 SAS 系统,但是其他模块必须与之结合起来才能使用。Base SAS 主要具有数据及用户使用环境的管理、SAS 语言程序的处理、基本的数据分析和报告等功能,以及调用其他 SAS 模块和产品的作用。

SAS/STAT 提供了当今流行的主要统计分析方法,是国际上统计分析领域的标准权威软件。它具有回归分析、方差分析、属性数据分析、多变量分析、聚类分析、判别分析、非参数分析、生存分析和心理测量分析等统计功能。该软件时常更新,反映统计方法的新进展。

SAS 软件中其他常用的模块还有:

SAS/GRAPH 是完成多种绘图功能的图形软件包,可以绘制包括二维和三维的曲线图、直方图、圆饼图、区块图、星形图、地理图以及各种映像图。这些图形非常形象、直观地表达各变量之间的关系及数据的分布状态,对解决各种实际问题起着重要的辅助作用。

SAS/IML 主要用于矩阵运算,它提供功能强大的面向矩阵运算的编程语言。它处理的基本数据元素是一个矩阵,用户可直接用矩阵代数的记号来组成 IML 的程序语句,用于研

究新算法或作为解决 SAS 系统中没有现成方法的工具。

SAS/ETS 主要用于经济预测和时间序列分析。它提供了经济分析、时间序列分析、时间序列预测、建立计量经济和财务模型、周期性调整、金融分析和报表、经济和金融数据集接口以及时间序列资料管理等过程。此外, SAS/ETS 也包括了一个进行交互式时间序列预测的菜单驱动系统。

SAS/INSIGHT 是可视化的数据探索工具,是进行数据挖掘的有力工具。它采用交互式数据分析,制作各种统计图形以及用于方差分析、线性拟合、Logistic 回归和 Poisson 回归等。

SAS/ASSIST 模块为 SAS 系统提供了面向任务的、菜单驱动用户的友好界面。它可免去用户学习 SAS 语言的负担,同时 SAS/ASSIST 生成的 SAS 程序既可帮助用户学习 SAS 语言,又可辅助有经验的用户快速编写 SAS 程序。

SAS/ACCESS 是对目前许多流行数据库的接口组成的接口集,这种接口是透明和动态的。

SAS/OR 主要用于运筹学和线性规划。是运筹学和工程管理的专用软件,该软件不仅包含通用的线性规划、混合整数规划和非线性规划的方法,还包含解决项目管理、时间安排和资源分配等问题的一整套方法。

SAS/QC 提供了根据产品观测数据进行产品质量管理的各种分析工具。它是质量管理的专用软件,可提供完整的实验设计和质量管理的菜单系统,用于实验设计、质量管理和过程控制。

SAS/FSP 提供全屏幕交互式数据输入、编辑、查询功能,可设计数据输入屏幕,并具有书信撰写方面的功能。

SAS/AF 是一个应用灵活的开发工具,利用 SAS/AF 的屏幕设计能力及 SCL 语言的处理能力可快速开发各种功能强大的应用系统,通过 SAS/AF 的 OOP(面向对象编程)技术,可建立用户化菜单系统,连接各种应用程序。

详细的各类模块列表和介绍可参见 <http://support.sas.com/software/index.html>。

第二节 SAS 窗口工作环境

一、启动 SAS

双击桌面上的快捷方式图标,或从开始→程序菜单中找到 SAS→SAS 9.3(英语),点击启动(图 1-1)。

二、SAS 窗口环境

启动 SAS 后,可看到图 1-2 的界面,其内有五个主要的 SAS 窗口,分别是 **Editor** 窗口、**Log** 窗口、**Output** 窗口、**Explorer** 窗口和 **Results** 窗口。这些窗口可以帮助我们完成很多最基本的 SAS 任务。点击窗口条上相应的按钮可将某窗口移至前台,成为当前活动窗口。

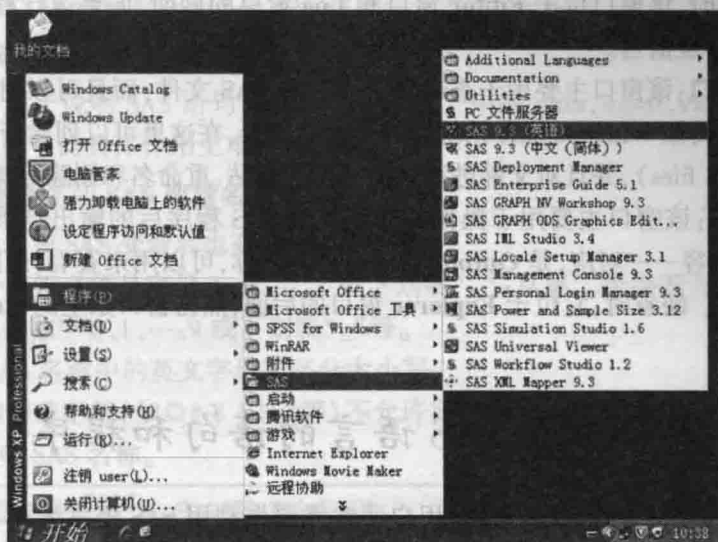


图 1-1 启动 SAS

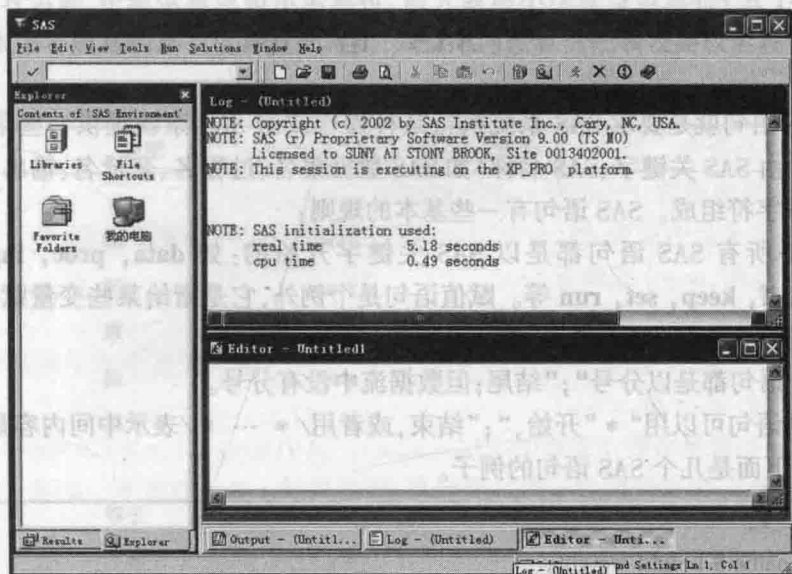


图 1-2 SAS 的窗口工作界面

下面我们分别介绍五个窗口的主要功能。

Editor 窗口:主要用于打开 SAS 程序文件(SAS 程序文件扩展名为 *.sas)、编辑和修改 SAS 程序、提交全部或部分 SAS 程序。根据程序中编码的性质可以显示不同的颜色,并且对 SAS 语言进行语法检查。在 SAS 中可同时打开多个 **Editor** 窗口进行操作。

Log 窗口:该窗口显示有关 SAS 会话和提交 SAS 程序的信息,包括程序的运行时间和出错信息等(log 文件的扩展名为 *.log)。

Output 窗口:该窗口主要用于显示提交 SAS 程序后的运行结果(output 文件的扩展名为 *.lst)。缺省时,该窗口位于 **Editor** 窗口和 **Log** 窗口的后面,如果运行程序有结果输出时,该窗口自动移至前台。

Explorer 窗口:该窗口主要用于查看和管理所有 SAS 文件,而且可以对非 SAS 文件创建快捷方式。它类似 Windows 操作系统中的资源管理器,在这里可以创建新的库(Libraries)和 SAS 文件(SAS files),并且对文件进行移动、复制、粘贴、重命名和删除等操作。

Results 窗口:该窗口主要用于操作和管理提交 SAS 程序后的输出结果。它的内容与 **Output** 窗口的内容一一对应,是 **Output** 窗口内容的名称,可以用来查看、删除、保存和打印部分或全部结果。缺省时,它位于 **Explorer** 窗口的后面,点击窗口条上的 **Results** 按钮可将它移至前台。

第三节 SAS 语言的语句和程序

为运用 SAS 系统管理和分析数据,用户通常需要先利用 SAS 语言编写一系列指令,称为 SAS 程序。通过 SAS 程序,用户可以定义数据以及指定对数据做哪些统计分析。学习用 SAS 语言编写 SAS 程序,可以帮助应用者更灵活地应用 SAS。

一、SAS 语句

一个 SAS 语句就是要求 SAS 系统执行某种操作或给 SAS 系统提供一些信息的命令。SAS 语句通常由 SAS 关键字、SAS 名称(如 SAS 数据集名、过程名、变量名、输出格式名等)、运算符及特殊字符组成。SAS 语句有一些基本的规则:

(1) 几乎所有 SAS 语句都是以 SAS 关键字开始的:如 **data**, **proc**, **input**, **cards**, **model**, **class**, **if**, **keep**, **set**, **run** 等。赋值语句是个例外,它是指给某些变量赋值或把某些变量的值赋给其他变量,如 $x=2$; $y=3$; $z=x+y$ 等。

(2) SAS 语句都是以分号“;”结尾;但数据流中没有分号。

(3) 注释语句可以用“*”开始,“;”结束,或者用/* ... */表示中间内容是注释语句。

例 1.1 下面是几个 SAS 语句的例子。

```
data sas1_1;
input id name $ height weight;
bmi=weight/height * * 2 ;      /* 把 weight 除以 height 平方的值赋给变量 bmi */
cards;                          /* 数据流开始 */
1 Judy 1.56 45
2 Lucy 1.67 53
;                                /* 数据流结束 */
proc print data=sas1_1;
run;
```

其中 **data**, **input**, **cards**, **proc** 和 **run** 等是 SAS 关键字; **sas1_1** 是数据集名; **id**, **name**,

height, weight 和 bmi 是变量名, name 是个用 \$ 定义的字符变量, 其余是数值变量; “=”、“/”和“**”属于 SAS 运算符; print 是 SAS 过程名。

1. SAS 关键字

SAS 关键字用于说明 SAS 语句的类型。如例 1.1 中的 data、input、cards、proc 和 run 语句等。SAS 关键字的具体作用见本节第二点。

2. SAS 数据集名和 SAS 变量名

SAS 数据集名和 SAS 变量名也有一些基本规则:

- (1) 第一个字符必须是字母 A、B、...、Z 或下划线“_”; 从第二个字符开始, 可以为字母 A、B、...、Z, 阿拉伯数字 0、1、...、9 或下划线“_”等。
- (2) 所有 SAS 名称中的英文字母不区分大小写。
- (3) 空格和特殊字符(如 ©#¥ ¥%\$ 等)不允许在 SAS 名称中使用。

例 1.2 一些 SAS 名称。

```
xt5_1;_num3;year2008;_n_;age;sex;name
```

3. SAS 运算符

SAS 运算符是用于比较运算、算术运算或逻辑运算等的符号。常用的 SAS 运算符包括: 算术运算、比较运算、逻辑运算或布尔运算符, 最大或最小连接等运算符(表 1-1)。SAS 表达式的运算次序与通常的算术运算规则相同, 如括号内运算优先和较高级运算符优先等。

表 1-1 SAS 运算符

运算符	说明	例子
算术运算符		
+	加	x+y
-	减	x-y
*	乘	x*y
/	除	x/y
**	乘方	x**y
比较运算符		
=	等于	x=y
^=	不等于	x^=y
>	大于	a>b
>=	大于等于	a>=b
<	小于	a<b
<=	小于等于	a<=b
逻辑运算符		
And/&	逻辑与	x>2 and y>3
Or/	逻辑或	x>2 or y>3

续表

运算符	说明	例子
Not/^	逻辑非	
其他		
<>	最大	3<>5:结果为 5
><	最小	3><5:结果为 3
	连接	A="my name is",B="SAS",C=A B,那么 C="my name is SAS"

4. SAS 函数

SAS 函数是一个独立的子程序,它对 0 个或多个自变量进行计算后返回一个值,每个函数都有一个关键字名,为了调用一个函数,需要先写出它的函数名,再用括号将 0 个或多个自变量括起来,跟在函数名后面,表示这个函数对这些自变量执行某种运算。函数一般形式为:函数名(自变量,自变量,...)。

SAS 函数有多种,这里介绍部分常用函数。

(1) SAS 常用概率密度函数。

1) 标准正态分布函数:PROBNORM(x)。

该函数计算服从标准正态分布的随机变量 u 小于 x 的概率。即 $p(u < x)$ 。如 $y = \text{probnorm}(-1.96)$,结果为 0.025。

2) t 分布概率函数:PROBT(x, df, nc)。

计算自由度为 df ,非中心参数为 nc 的 t 分布随机变量小于 x 的概率,当 $nc=0$ 或缺省时,该分布为中心分布。如 $y = \text{probt}(0.95, 100)$,结果为 0.8278。

3) F 分布概率函数:PROBF($x, df1, df2, nc$)。

计算服从分子自由度为 $df1$,分母自由度为 $df2$ 的 F 分布的随机变量小于 x 的概率,当分布为中心分布时, $nc=0$ 或缺省。

4) χ^2 分布概率函数:PROBCHI(χ, df, nc)。

计算服从自由度为 df ,非中心参数为 nc 的 χ^2 分布的随机变量小于 x 的概率。如 $nc=0$ 或缺省,即为中心 χ^2 分布。

5) 二项分布概率函数:PROBBNML(p, n, m) 其中 $0 \leq p \leq 1, n \geq 1, 0 \leq m \leq n$ 。

计算概率为 p ,样本例数为 n 的二项分布,随机变量 $x \leq m$ 的概率。如求 $p(x=k)$ 的值,可计算 $\text{probbnml}(p, n, k) - \text{probbnml}(p, n, k-1)$ 。

6) 泊松分布概率函数:POISSON(m, n) 其中 $m \geq 0, n \geq 0$ 。

计算参数为 m 的泊松分布的随机变量 $x \leq n$ 的概率。如计算 $P(x=k)$ 的值,可用 $\text{Poisson}(m, k) - \text{Poisson}(m, k-1)$ 。

(2) SAS 常用分位数函数。

1) 正态分布分位数函数:PROBIT(p) ($0 \leq p \leq 1$)。

计算标准正态分布的分位数,是概率函数 PROBNORM 的逆函数。如 $\text{probit}(0.025)$,结果为 -1.96。

2) t 分布的分位数函数:TINV(p, df, nc)。

计算自由度为 df ,非中心参数为 nc 的 t 分布的 p 分位数,如 nc 缺省或 $nc=0$,则计算中

心 t 分布的 p 分位数。

3) F 分布的分位数函数: $\text{FINV}(p, df1, df2, nc)$ 。

计算分子自由度为 $df1$, 分母自由度为 $df2$, 非中心参数为 nc 的 F 分布的 p 分位数。如 nc 缺省或 $nc=0$, 则计算中心 F 分布的 p 分位数。

4) χ^2 分布的分位数函数: $\text{CINV}(p, df, nc)$ 。

计算自由度为 df , 非中心参数为 nc 的 χ^2 分布的 p 分位数, 如 nc 缺省或 $nc=0$, 则计算中心 χ^2 分布的 p 分位数。

(3) 其他 SAS 常用函数见表 1-2。

表 1-2 其他 SAS 常用函数

函 数	说 明
算术函数	
$\text{ABS}(X)$	取 X 的绝对值。
$\text{SQRT}(X)$	计算 X 的平方根。
$\text{MAX}(X_1, \dots, X_n)$	求 $X_1, \dots, X_n$ 中的最大值。
$\text{MIN}(X_1, \dots, X_n)$	求 $X_1, \dots, X_n$ 中的最小值。
$\text{MOD}(X, Y)$	求 X/Y 的余数。 如 $\text{MOD}(10, 3) = 1; \text{MOD}(6, 2) = 0$ 。
$\text{SIGN}(X)$	当 $X < 0$ 时其值为 -1 ; 当 $X > 0$ 时其值为 1 ; 当 $X = 0$ 时其值为 0 。 如 $\text{SIGN}(3.5) = 1; \text{SIGN}(-5.4) = -1; \text{SIGN}(0) = 0$ 。
$\text{EXP}(X)$	计算 e 的 X 次幂。 $\text{EXP}(X) = e^X$ 。
$\text{LOG}(X)$	对自变量 X 求以 e 为底的自然对数。
$\text{LOG2}(X)$	对自变量 X 求以 2 为底的对数。
$\text{LOG10}(X)$	对自变量 X 求以 10 为底的对数。
截取函数	
$\text{CEIL}(X)$	取 $\geq X$ 的最小整数。 如 $\text{CEIL}(5.7) = 6; \text{CEIL}(-2.3) = -2$ 。
$\text{FLOOR}(X)$	取 $\leq X$ 的最大整数。 如 $\text{FLOOR}(6.9) = 6; \text{FLOOR}(-7.2) = -8$ 。
$\text{INT}(X)$	取 X 的整数部分。 如 $\text{INT}(5.6) = 5; \text{INT}(-3.7) = -3$ 。
$\text{ROUND}(X, n)$	X 按 n 指定的精度取舍入值。 如 $\text{ROUND}(73.58, 0.1) = 73.6$ 。
随机数函数	
$\text{UNIFORM}(\text{seed})$ 或 $\text{RANUNI}(\text{seed})$	产生服从均匀分布 $\text{UNI}(0, 1)$ 的随机数。
$\text{NORMAL}(\text{seed})$ 或 $\text{RANNOR}(\text{seed})$	产生服从标准正态分布 $N(0, 1)$ 的随机数。经如下变换: $M + s * \text{NORMAL}(\text{seed})$, 可得到服从正态分布 $N(M, s^2)$ 的随机数。
$\text{RANEXP}(\text{seed})$	产生一个参数 $\lambda = 1$ 的指数分布的随机数。 如果 $Y = \text{RANEXP}(\text{seed}) / \lambda$, 产生参数为 λ 的指数分布随机数。
$\text{RANBIN}(\text{seed}, n, p)$	产生服从均值 np , 方差为 $np(1-p)$ 的二项分布的随机数。
$\text{RANPOI}(\text{seed}, \lambda)$	产生服从均数为 λ 的泊松分布的随机数。

注: 表中所列统计符号、函数等均为 SAS 程序自动列出, 故为真实反映软件的运行情况, 我们未做任何改动。全书余同。

二、SAS 程 序

将一系列 SAS 语句按逻辑顺序排列起来,构成 SAS 程序。通常 SAS 程序包含数据步和过程步两部分。数据步以 data 语句开头,以 run 语句结束,其主要作用是建立 SAS 数据集。过程步以 proc 开头,以 run 语句结束,其主要作用是激活 SAS 过程对数据进行处理和分析。通常,在数据处理过程中,可有多个数据步和多个过程步混合使用。后一个 data 或 proc 语句起到前一步中 run 语句的作用,故两步中间的 run 语句常省略,但最后一步的最后必须有 run 语句,否则最后一步将不能运行。Endsas 语句可以终止交互式的 SAS 工作。

SAS 语句书写格式自由,可在各行的任意位置开始输入程序,一条 SAS 语句可以连续写在几行中,也可以一行写几个语句,每个语句的最后一定要用“;”号结束。

1. 数据步

数据步的主要功能有:

- (1) 创建新的数据集。
- (2) 读入并且修改数据。

用户在编写数据步程序时最好养成两个好的习惯:

- (1) 尽可能少地使用数据步(当然大多数情况下至少需要一个数据步)。
- (2) 将主数据集程序代码保存在一个单独的程序中,在另外的程序中进行分析。用恰当的容易理解的名字命名或用注释语句等。一个好程序应该做到清晰易读。

2. 过程步

过程步的主要功能有:

- (1) 使用存在的数据集进行一些特定的分析。
- (2) 产生结果或报告。

一个典型的 SAS 程序由数据步(DATA 步)开始,创建一个 SAS 数据集,然后接一系列的过程步,对数据进行分析。

例 1.3 SAS 程序的例子。

```
data sas1_3;                                /* 创建名为 sas1_3 的 SAS 数据集 */
input id name $ height weight;
bmi=weight/height**2;                      /* 在 cards 语句之前创建分析用的所有变量 */
cards;                                     /* cards 表示开始输入数据 */
1 Judy 1.56 45                             /* 数据流中每个数据值之间最少有 1 个空格 */
2 Lucy 1.67 53
;                                           /* ;表示数据输入结束 */
run;                                       /* 数据步结束 */
proc print data=sas1_3;                   /* 打印输出 sas1_3 数据集到 Output 窗口 */
run;                                       /* 过程步结束 */
```

SAS 程序编辑完成后,可选择以下任一方式提交 SAS 程序:

- (1) 点击程序编辑窗口工具栏上的提交图标 。

(2) 在 run 下拉菜单中选择 submit。

(3) 使用 F3 键。

(4) 在菜单条的命令框中输入 submit。

注意:用户可以用光标选定部分程序进行提交。

SAS 程序提交后,SAS 会在 Log 窗口写入一些信息(一定要去读这些信息!),这些信息是非常有用的,可以帮助你调试 SAS 程序。Log 窗口中通常包括三类信息,分别用蓝色、绿色和红色表示。

NOTE:用蓝色表示,主要是 SAS 程序运行的一般情况,一些有用的提示等。

WARNING:用绿色表示,不算错误,但 SAS 会通知你程序可能有一些问题。SAS 处理过程不会终止,SAS 仍然会创建数据集。

ERROR:用红色表示,反映程序代码中有错误,SAS 将不能处理数据步,并终止对数据的处理和分析。如果你正在运行数据步来代替一个旧的数据集,那么旧数据集将不会被代替。

当运行一个 SAS 程序后,如果有输出结果,SAS 会将它写在 Output 窗口中,同时将标题加到 Results 窗口中。用户可以通过点击工具栏上的图标,来清除窗口中内容。

(邹艳晖)

(1) 在 SAS 程序中使用 libname 语句,如 libname libref libref-path; 创建数据集。

例 2.2 创建和使用 SAS 数据集的例子。

【程序】

```
libname j 'd:\jinh'; /* 创建名为 j 的 SAS 数据集 */
data j.wz; /* 创建数据集 j.wz */
input id name $ weight;
cards;
1 july 120 999
2 july 107 23
3 july 107 23
4 july 128 65
run;
```

第二章 建立 SAS 数据集

第一节 SAS 数据集概述

一、SAS 数据集基本格式

SAS 具有创建和处理 SAS 数据集的功能。SAS 数据集的扩展名为 *.sas7bdat。SAS 数据集含有如观测数、变量名、文件更新的时间、数据的长度和格式等描述数据集的信息。我们常说的 SAS 数据集多指形如图 2-1 的数据表。

id	name	height	weight
1	Judy	156	45
2	Lucy	167	53
3	Susan	165	55
4	Tony	178	62

图 2-1 SAS 数据集(数据表)的观测和变量

在图 2-1 中的矩形数据表中,每一行数据值称为一个观测或称为一条记录,每一列数据值称为一个变量,所以在 SAS 数据集中,每一个观测由各个变量的数据值组成。

SAS 数据集集中的变量有两种类型:数值型变量和字符型变量。数值型变量的取值只能是数值,前面可以直接加正号(+)或负号(-)表示正负值,加小数点(.)表示小数,加 E 表示科学计数法。字符型变量的取值可以是字母、特殊符号或数字。

在 SAS 中所有程序处理缺失值的形式都是相同的。如果观测中的某一个或多个变量值为缺失值,SAS 处理时会将该观测全部删去。因此,如果你的数据集中含有缺失值,应该经常查看 Log 窗口的信息,并在 Output 窗口打印数据列表,查看观测的数目。在 SAS 语句中,可指令数据流中“999”所代表的缺失值在 Output 窗口中以“.”表示。

例 2.1 缺失值的例子。

【程序】

```
data sas2_1;  
input id name $ height weight;  
if name='999' then name=.; /* 字符型变量值缺失 */  
if weight=999 then weight=.; /* 数值型变量值缺失 */  
cards;  
1 Judy 156 999  
2 Lucy 167 53  
3 999 165 55  
4 tony 178 62  
;
```