流行病学与循证医学研究中SAS软件实用教程

主编　闫永平　肖　丹

第四军医大学出版社

流行病学与循证医学研究中SAS软件实用教程

主　　编　闫永平　肖　丹

副 主 编　门　可　吉兆华

编　　委　（按汉语拼音排序）

吉兆华　李　端　李婵娟　龙　泳

门　可　邵中军　苏海霞　王　波

王安辉　肖　丹　闫永平　张　磊

张景霞　张维璐　张玉海

学术秘书　姜　艳

第四军医大学出版社·西安

图书在版编目（CIP）数据

流行病学与循证医学研究中SAS软件实用教程/闫永平，肖丹主编. —西安：第四军医大学出版社，2014.4

ISBN 978-7-5662-0476-9

Ⅰ.①流… Ⅱ.①闫…②肖… Ⅲ.①流行病学-统计分析-应用软件-教材②临床医学-统计分析-应用软件-教材 Ⅳ.①R18-39②R4-39

中国版本图书馆CIP数据核字（2014）第055596号

liuxingbingxue yu xunzhengyixue yanjiuzhong SAS ruanjian shiyong jiaocheng

流行病学与循证医学研究中SAS软件实用教程

出版人：富　明　　责任编辑：张永利　郑　爱

出版发行：第四军医大学出版社

地址：西安市长乐西路17号　邮编：710032

电话：029-84776765　　传真：029-84776764

网址：http://press.fmmu.edu.cn

制版：绝色设计

印刷：陕西奇彩印务有限责任公司

版次：2014年4月第1版　2014年4月第1次印刷

开本：787×1092　1/16　　印张：14.5　　字数：335千字

书号：ISBN 978-7-5662-0476-9/R·1323

定价：38.00元

前　言

流行病学是一门运用描述性、分析性和实验性研究等方法从群体角度探索疾病流行规律和评价干预措施效果的应用性学科，它作为临床医学和预防医学研究的方法学被应用于相关的各个学科。循证医学强调严格地测量、观察和评价研究结果而形成最佳证据，从而提高研究质量和医疗实践水平。群体研究过程中资料整理和数据处理是结果分析的重要一环，SAS(statistical analysis system)作为当今国际上最著名的数据分析软件之一，在数据处理和统计分析领域被誉为国际上的标准软件系统。因此掌握该软件在流行病学与循证医学研究中的应用无疑具有非常重要的实用价值。

本书在总结我们应用SAS软件进行各类流行病学和循证医学资料整理和统计处理中的大量实例和经验的基础上，吸取该领域国内外新近的先进理论和进展，编制了适用于我国临床医学和预防医学等专业的研究生和工作人员使用的常用SAS软件操作教程。

全书分为十五个章节。第一章是SAS软件的简介；第二章介绍了原始数据的加工与整理方法；第三章至第五章是SAS软件在病因研究中的应用和编程方法，包括现况研究、病例对照研究和队列研究；第六章介绍了应用SAS软件进行传染病暴发调查中潜伏期、再生数、传播危险因素等的计算方法；第七章至第十章是SAS软件在临床各类研究中的应用和编程方法，包括筛检与诊断试验、临床试验、疾病预后和Meta分析；第十一章至第十五章是SAS软件在资料处理中的应用技巧，包括缺失值的处理、常用抽样方法的实现、批量数据的读入和双录入数据的处理、大规模流行病学调查与统计量的计算、信度与效度分析。各章均以实例为基础，结合流行病学与循证医学的基本理论，突出SARS软件的应用和编程方法，希望达到举一反三的目的。上述相关章节所涉及的原始资料、数据库和SAS程序可以通过第四军医大学流行病学学科专业网站下载(http://lxbx.fmmu.edu.cn/nr.jsp?urltype=news.NewsContentUrl&wbnewsid=142923&wbtreeid=1910)。

由于此类图书甚少，而且多为实践操作程序，尽管我们反复推演，但限于水平，书中缺点及错误之处可能未完全避免，敬请各位读者批评指正。

闫永平

2013年11月20日于西安

目　录

第一章 SAS 软件简介

本章内容

1. SAS 软件的功能与结构。
2. SAS 数据库与 SAS 数据集。
3. SAS 过程与 SAS 程序。
4. SAS 程序的基本运行环境和工作机制。

SAS(statistical analysis system)是当今国际上最著名的数据分析软件之一,由美国北卡罗来纳大学(North Carolina State University)的 SAS 软件有限公司研制成功,并于 1976 年正式推出。SAS 软件可用于数据管理、统计分析、运筹决策等工作,在数据处理和统计分析领域,该软件被誉为国际上的标准软件系统。目前,SAS 软件的用户遍及全球金融、医药卫生、生产、运输、通讯、科学研究、政府和教育等领域的 131 个国家和地区的约 55000个企业、政府和大学,2011 年《财富》全球 500 强企业的前 100 名中,有 90 家企业使用 SAS 解决方案。

第一节 SAS 软件的功能与结构

一、SAS 软件的功能

SAS 软件将数据管理和数据分析融为一体,可实现以数据为中心的数据交换、管理、分析和呈现等功能。

(一)数据交换

SAS 软件可读入多种格式的数据文件,并将源数据转换成 SAS 数据集,供 SAS 过程分析处理。也可将 SAS 数据集中的数据转换成其他格式的数据文件,供其他软件处理。

(二)数据管理

数据的加工和处理可在两个阶段进行,分别是读入外部源数据文件时和外部源数据转换为 SAS 数据集后。对于某些较为复杂的数据加工和处理任务,可先在数据读入过程中进行初步加工,然后再对 SAS 数据集进行进一步加工。

SAS 软件提供了完备的 SAS 语句和函数,可用于数据的加工和处理。例如,赋值语句可用于建立新的变量,控制语句 DO/END、IF-THEN/ELSE 等可用于选择符合条件的数据构成新的数据集,信息语句 DROP、KEEP 等可用于指定在新数据集中删除或保留的原数据集中的变量,操作语句 INPUT、SET、MERGE 等可用于数据的录入、数据集的合并、拼

接等，日期函数 TODAY、TIME、INTCK 等用于计算当前的日期和时间、两个时间点的时间间隔等，字符运算函数 SUBSTR、INDEX、SCAN 等用于提取或置换字符值、在字符串中寻找一个字符、由字符串中返回一个特定位置的词等，转换函数 INPUT 和 PUT 用于数值与字符之间的转换。

（三）数据分析

通过调用 SAS 软件的一些基本过程和 SAS 函数，可进行多种统计分析，包括：

1. 计算描述性统计量，如均数、标准差、标准误、总和、平方和、极差、相关系数、峰度系数、偏度系数等 40 多项。

2. 计算概率分布函数、分位数、产生随机数。

3. 对数据进行标准化、编秩并计算其统计量。

4. 产生并分析列联表。

5. 进行方差分析、相关与回归分析、线性模型拟合、属性数据分析、多变量数据的判别和聚类分析、非参数统计分析、生存分析、时间序列分析、实用预测、质量控制、运筹学统计分析等。

6. 绘制二维与三维的基本统计图，如条图、直方图、圆图、散点图、等差和等比线图、曲线拟合图、时间序列图等。

（四）数据呈现

SAS 可以输出和呈现多种类型的数据，包括：

1. 打印输出数据集中的数据和统计分析结果。

2. 将数据分析过程中产生的中间数据存储为一个新的数据集，以便进一步分析和处理。

3. 合并多个数据分析过程中产生的中间数据，存储为一个新的数据集，以便数据输出或进一步分析和处理。

SAS 分析结果的呈现形式也很多样化，主要包括列表报告、汇总报告以及用户自定义报表等。

二、SAS 软件的结构

SAS 是一个组合软件系统，由 50 多个功能模块组合而成，各个模块之间既相互独立又相互补充。其中，BASE SAS 模块是 SAS 系统的基础和核心，它提供了 SAS 系统的基本运行环境——显示管理系统（display manager system），即运行 SAS 必须先启动 BASE SAS 模块；SAS/STAT、SAS/IML、SAS/INSIGHT、SAS/OR、和 SAS/QC 模块是 SAS 数据分析的核心，各模块侧重于不同的分析方向；SAS/AF、SAS/EIS 和 SAS/GRAPH 模块是 SAS 开发和呈现的工具，为用户提供便捷的面向对象的开发工具；SAS/ACCESS、SAS/CONNECT、SAS/SHARE 和 SAS/WA 模块主要用于 SAS 对分布处理模式的支持及其数据仓库设计。

（一）BASE SAS 模块

SAS 系统的核心模块，承担主要的数据管理任务，并管理交互应用环境，进行用户语言处理，调用其它 SAS 模块和产品。该模块提供丰富的数据管理功能，支持用标准 SQL

语言对数据进行操作，能够制作各种复杂的统计报表，可进行基本的描述性统计分析、相关分析、正态分布检验等。

SAS 系统具有灵活的功能扩展接口和强大的功能模块，在 BASE SAS 的基础上，还可以通过增加不同的模块而增加不同的功能，组成一个用户化的 SAS 系统，以满足不同用户的多样化需求。

（二）SAS/STAT 模块

统计分析模块，是国际上统计分析领域的标准软件，包括方差分析、回归分析、定性数据分析、多变量分析、判别分析、聚类分析、生存分析、心理测验分析和非参数统计分析等 60 多个过程。其中，针对不同的模型或不同特点的数据，SAS/STAT 模块提供了十多个过程用于回归分析，包括正交回归、响应面回归、Logistic 回归等。该模块为多种实验设计提供了方差分析和协方差分析的工具，为一般线性模型和广义线性模型的分析和处理提供了专用过程，为主成分分析、典型相关分析、判别分析和因子分析提供了许多专用过程。

（三）SAS/IML 模块

交互式矩阵程序设计语言模块，该模块提供了一套完整的面向矩阵的交互式矩阵编程语言（interactive matrix language，IML），帮助用户研究新算法或解决 SAS 中没有现成算法的问题。SAS/IML 模块中的基本数据元素是数据矩阵，数据可以是数值型的，也可以是字符型的。用这种语言可以很方便地用较少的语句描述复杂的计算过程，进而在控制语句的帮助下实现许多复杂的算法。

（四）SAS/INSIGHT 模块

交互数据分析模块，是一个进行交互式数据探索和分析的可视化工具。将统计方法和交互式图形显示融合在一起，为用户提供一种全新的使用统计分析方法的环境。该模块具有很强的图像表现能力，可同时打开多个窗口对数据和图像进行比较、探索和分析。

（五）SAS/OR 模块

运筹学模块，可提供全面的运筹学方法，是一种强有力的决策支持工具。SAS/OR 模块包含通用的线性规划、混合整数规划和非线性规划的求解，也为专门的规划问题提供更为直接的解决办法，如网络流问题、运输问题等，还包含用于项目管理、时间安排和资源分配等问题的一系列方法。

（六）SAS/QC 模块

质量管理模块，可提供一套全屏幕菜单系统，引导用户进行标准的统计过程控制和试验设计。该模块可进行多种控制图的制作和分析，如 Pareto 图（排列图）可用于发现需要优先考虑的因素，Ishikawa 图（鱼骨图）可使用户直观地进行因果分析。

（七）SAS/AF 模块

交互式全屏幕软件应用系统模块，是一个应用开发工具。用户可将包含众多功能的 SAS 软件作为方法库，利用 SAS/AF 的屏幕设计能力以及 SQL 语言的处理能力来快速开发各种功能强大的应用系统。该模块也采用了先进的 OOP（面向对象编程）技术，使用户可以方便快速地开发各类具有图形用户界面的应用系统。

(八)SAS/EIS 模块

决策支持表现模块,是一个快速应用开发工具。采用面向对象的编程模式,以图表等直观的方式将关键性或总结性信息呈现给用户。

(九)SAS/GRAPH 模块

绘图模块,能够绘制多种图形,如直方图、圆图、线图、条图、散点图、星形图、三维曲面图、等高线图和地理图等。该模块还提供全屏幕图形编辑器和丰富的中西文矢量图形字体,用户可以对图形进行任意修改、拼接以及绘制文字及图形元素等。

(十)SAS/ACCESS 模块

数据库接口模块,利用该模块可对多种不同格式的外部源数据进行查询、访问和分析。SAS/ACCESS 模块提供了双向的数据接口,既可将数据读入 SAS,也可在 SAS 中更新外部数据或将 SAS 数据加载到外部数据库中。对于偶尔使用的数据,可不将外部源数据真正读入 SAS 数据库,而只需在 SAS 中建立描述文件,将此文件当作 SAS 数据集使用;对一些经常使用的外部数据,可利用 SAS/ACCESS 模块将数据真正提取进入 SAS 数据库,存储为 SAS 数据集。

SAS/ACCESS 模块支持的数据库主要有:IMS-DL/I, SQL/DS, DB2, ADABAS, Rdb, ORACLE, SYBASE, INGRES, INFORMIX, DBF/DIF, EXCEL, ODBC 等。

(十一)SAS/CONNECT 模块

远程连接模块,通过该模块可使各平台的 SAS 系统建立内在联系,实现分布处理,从而有效地利用各平台的数据和资源。SAS/CONNECT 模块既提供远程计算服务,又提供远端数据服务,支持 MVS、CMS、VSE、OpenVMS、UNIX、OS/2、Windows、DOS、AOS/VS、PRIMOS 等常用系统之间的几乎各种互连方式,以及 TCP/IP、APPC、DECNet、NETBIOS、TELNET、ASYNC、HLLAPI、3270 等多种通讯方法。

(十二)SAS/SHARE 模块

数据库的并发性控制模块,使两个或更多的客户端能同时操作同一个 SAS 文件。

(十三)SAS/WA 模块

数据仓库管理工具,它在 SAS 软件基础上提供了一个建立数据仓库的管理层,包括:定义数据仓库和主题,数据转换和汇总,汇总数据的更新,Metadata 的建立、管理和查询,Data Marts 和 Info Marts 的实现。

(十四)SAS/ASSIST 模块

面向任务的菜单驱动模块,集成了 SAS 系统其他模块的各种功能,提供了一个面向任务的菜单驱动用户界面。利用 SAS/ASSIST 模块,用户不需编程,仅靠鼠标操作就可以方便地使用 SAS 系统的其它产品。它自动生成的 SAS 程序既可以辅助有经验的用户快速编写 SAS 程序,又可以帮助用户学习 SAS 语言。

(十五)SAS/ETS 模块

计量经济学和时间序列分析模块,包含全面的时间序列时域分析和谱域分析,如实用预测(逐步自回归、指数平滑、Winters 方法)、序列相关校正回归、分布滞后回归、ARIMA模型、状态空间方法、谱分析和互谱分析等,还提供许多处理时间序列数据的实用

程序,如时间频率转换和插值、季节调整等。

(十六)SAS/FSP 模块

快速数据处理的交互式菜单系统模块,SAS/FSP 模块提供对 SAS 数据集的屏幕浏览和编辑功能,能快速打开 SAS 数据集,具有全屏幕数据录入、编辑和查询以及数据文件创建等功能。

(十七)SAS/Enterprise Miner 模块

数据挖掘模块,基于"SEMMA"理念,为用户提供抽样工具、数据重组、神经元网络、数据回归、结果显示等新过程。

(十八)SAS/GIS 模块

空间数据分析模块,主要用于空间数据的分析和展示。该模块将空间数据分层存储,每一层可以是某些地理元素(如坐标、海拔等),也可以与用户自定义的属性数据(如人口、发病数等)相关联。用户可交互式地缩小或放大地图,设定各层次是否显示,并利用各种交互式工具进行数据分析和展示。

第二节 有关 SAS 的基本概念

一、SAS 数据库与 SAS 数据集

(一)SAS 数据库

SAS 数据库是存放 SAS 数据集的地方,在 Windows 操作系统中相当于硬盘上的某个文件夹。除了 SAS 数据集以外,SAS 数据库还可以存放其他类型的 SAS 文件。

1. *库标记* 为了使用SAS数据库,需要为每一个数据库指定一个库标记来识别。库标记又称为库逻辑或库关联名,是 SAS 文件的物理位置在 SAS 系统中的一个逻辑标识。来自不同文件夹的文件可以被分别指定为不同的库标记,也可以指定为同一个库标记;同一个文件夹也可以被指定为不同的库标记。

2. *数据库的类型* SAS 数据库分为永久型和临时型。临时型数据库的库标记是 WORK,在 SAS 启动后自动生成,退出 SAS 后,该文件夹及其内所有文件将被删除。永久型数据库的库标记由用户自行定义,退出 SAS 后,该文件夹及其内所有文件不会被删除。启动 SAS 后,系统会自动生成 3 个永久型数据库,其库标记分别是 MAPS、SASHELP 和 SASUSER。

(二)SAS 数据集

SAS 数据集是存放数据及其属性的地方,相当于硬盘上的某个文件。SAS 数据集是关系型结构,通常分为描述部分和数据部分。描述部分用于存放数据属性信息,如变量名称、类型和长度等;数据部分用于存放数据值。SAS 数据值在数据集中以交叉表的形式存储,表格中的列称为变量(variable),表格中的行称为观测(observation)。

1. *变量类型* SAS 数据集中的变量可分为数值型和字符型。

数值型变量只允许变量值为数字,以浮点的形式存储,长度为 3 ~ 8 个字节,默认长

度为8个字节。SAS过程可以对这些数字进行统计运算，系统默认数值型变量小数点后保留2位有效值。缺失值以圆点“.”表示。

字符型变量允许变量值为中、英文字母，各种符号和数字，以ASCII码的形式存储，长度为1~32767个字符。此时数字被当作字符处理，不能进行统计计算。缺失值以空格表示。

2. *数据集的命名* 每个SAS数据集都有一个两级文件名，第一级是库标记，第二级是文件名，两者之间用“.”分隔。如名为WORK.DATA的数据集表示数据集DATA存储在临时数据库WORK中，当用户彻底退出SAS系统后，此数据集就自动消失，临时数据库的库标记WORK可以省略；名为AA.DATA的数据集表示数据集DATA存储在永久数据库AA中，当用户彻底退出SAS系统后，此数据集不会消失。

二、SAS过程与SAS程序

（一）SAS过程

SAS过程（SAS PROCEDURE）是模块化的子程序，是SAS研制者为解决特定问题而编写的、经过编译的SAS程序，以“.DLL”为扩展名，存储在SAS软件的各子目录下，供用户调用。

（二）SAS程序

SAS程序（SAS PROGRAM）是SAS用户运用SAS语言编写的一段程序。其目的是将用户的实验数据与变量名称联系在一起，并告诉SAS系统调用特定的SAS过程来完成某项任务。称为SAS引导程序，简称SAS程序。

一个简单的SAS程序包括一个SAS数据步和一个SAS过程步。数据步的作用是建立SAS数据集，以DATA语句开始，RUN语句结束，中间可包含众多执行特定功能的数据加工语句和控制语句。过程步的作用是激活SAS过程，执行特定的数据处理与分析功能和其它功能，以PROC语句开始，RUN语句结束，中间可包含各类控制语句。如果数据步或过程步后还有其它语句，则RUN语句可以省略。

第三节 SAS程序的运行

一、SAS的基本运行环境

SAS系统最常用的窗口有三个，分别是：编辑窗口（Editor）、日志窗口（Log）和输出窗口（Output）。

启动SAS后，一般直接进入编辑窗口，在此窗口可输入数据和编写SAS程序。程序编写完成后，单击工具栏中的小人像图标，SAS系统就开始执行操作，完成指定的任务。如果程序正确，计算结果将显示在输出窗口，运行过程中的各种信息显示在日志窗口；如果程序有错，则系统停止运行，并将错误信息显示在日志窗口。

编辑窗口可根据语句的不同类型，自动以四种颜色显示，分别是：深蓝、浅蓝、黄底黑

字和白底黑字。如果程序语句有误,则系统自动以红色显示错误语句。

如果运行错误语句,则日志窗口会出现错误提示。其中,绿色字体起到警告(warning)作用,提示存在小错误,SAS 系统会自动纠正;红色字体提示程序出现错误(error),系统停止运算,需根据提示修改程序。

二、SAS 程序工作机制

1. 一段 SAS 程序由一个或多个数据步和(或)过程步组成,它们将被同时提交运行。如果只需运行其中的某一部分程序,可先选中需要运行的程序,再单击工具栏中的小人像图标。

2. SAS 的程序步依次按顺序执行,位置靠前者先执行。

3. 一段程序可包含多个数据步,可创建多个数据集。

4. 同一数据集可依次被多个过程步多次使用。

5. 一个过程步可以立即使用位于其前面的程序所新建的数据集。

(肖 丹 李 端)

第二章 原始数据的加工与整理

本章内容

1. 利用编程法和菜单操作从外部文件读取数据(LIBNAME 语句、IMPORT 过程、CONTENTS 过程)。

2. 利用菜单操作创建 SAS 数据集。

3. 对数据集中的变量和观测进行筛选、拆分、拼接和更新(DATA 语句、SET 语句、IF-THEN 语句、KEEP 和 DROP 语句、KEEP 和 DROP 选项、OUTPUT 语句、SELECT 语句、MERGE 语句、赋值语句、MEANS 过程)。

4. 日期型数据的加工与整理(TRANSLATE 函数、INPUTN 函数、YEAR 函数、MONTH 函数、DAY 函数、INTCK 函数、PRINT 过程)。

5. 利用编程法和菜单操作将 SAS 数据写入外部文件(EXPORT 过程)。

原始数据中包含丰富的信息,有些信息可以直接利用,有些信息则需通过各种加工和整理才能为我们所用。本章将通过实例演示,详细介绍 SAS 系统中原始数据的加工与整理方法。

分析数据截取自“离退休干部健康状况调查”,数据文件名称是 data_chap2. xls。EXCEL格式的部分原始数据见图 2 -1。数据库中各指标及其取值的意义如下:

A	B	C	D	E	F	G
birthdate	armydate	health	height	weight	bp1	bp2
20.06.1930	01.01.1945	3	173	70	140	82
01.12.1930	01.08.1945	2	188	88	138	78
10.11.1927	01.06.1945	3	162	65	125	75
01.11.1927	01.01.1945	3	165	53	145	80
06.11.1926	01.04.1938	4	172	70	135	85
18.04.1929	01.05.1949	3	173	66	130	80
18.01.1931	01.05.1949	3	170	66	130	75
10.08.1928	01.05.1944	3	170	65	132	75
24.03.1927	01.07.1941	3	169	75	135	78
09.10.1923	01.05.1945	3	159	68	160	100
05.06.1923	01.06.1940	3	167	83	150	90

图 2 -1 EXCEL 格式的部分原始数据

1. birthdate: 出生日期,数据格式为“日.月.年”;
2. armydate: 入伍日期,数据格式为“日.月.年”;
3. health: 健康状况自我评分,0 ~5 分别代表健康状况由差到好;
4. height: 身高,厘米;
5. weight: 体重,千克;
6. bp1: 收缩压,千帕;

7. bp2：舒张压，千帕。

第一节 创建 SAS 数据集

一、从外部文件读取数据

在 SAS 系统中，对数据的管理、分析和呈现都是面向 SAS 数据集的，但是实际的研究数据往往是以各种不同的格式存储在不同的数据库中。因此，首先要从外部数据文件中读取数据，并将其转化为 SAS 数据集，才能在 SAS 系统中进行数据的加工、整理和分析。

（一）编程法

1. 程序 2－1

第一步

```
libname zl "E:\SAS\data"; /*创建逻辑库 zl*/
run;
```

第二步

```
proc import datafile="E:\SAS\data\data_chap2.xls"
  out=zl.health DBMS=EXCEL replace;
  sheet="sheet1";
  getnames=yes;
run;
```

第三步

```
proc contents data=zl.health;
run;
```

2. 程序说明　程序 2－1 第一步创建 SAS 逻辑库 zl，用于存储 SAS 数据，并指定逻辑库的物理存储位置是“E:\SAS\data”。第二步从 EXCEL 格式的外部数据文件 data_chap2. xls 中读入数据，并转换成 SAS 数据集 zl. health。

SAS 数据集分为描述部分和数据部分（详见第一章），程序 2－1 的第三步调用 CONTENTS过程，查看数据集 zl. health 的描述部分。

3. 结果说明　图 2－2 展示了 SAS 数据集 zl. health 的部分变量和观测。

	birthdate	armydate	health	height	weight
1	01.02.1924	01.02.1945	0	160	63
2	01.10.1929	01.03.1949	0	175	72
3	01.02.1931	01.02.1945	1	177	74
4	01.11.1926	01.09.1945	1	159	52
5	01.12.1930	01.08.1945	2	188	88
6	01.06.1931	01.06.1947	2	164	83
7	14.11.1922	01.12.1940	2	174	80
8	01.02.1925	01.07.1945	2	166	70
9	23.08.1929	01.10.1946	2	172	74
10	01.11.1923	01.10.1939	2	173	66
11	16.11.1927	01.06.1940	2	171	77
12	23.05.1923	03.01.1939	2	176	76

图 2－2　SAS 数据集 zl. health 的部分变量和观测

CONTENTS 过程的运行结果输出在 Output 窗口(图 2-3),主要展示 SAS 数据集 zl. health 的描述部分和变量属性,包括数据集的基本信息、引擎/主机相关的信息以及按字母排序的变量和属性列表。

CONTENTS PROCEDURE

数据集名	ZL. HEALTH	观测数	111
成员类型	DATA	变量数	7
引擎	V9	索引数	0
创建时间	2011 年 04 月 24 日 星期日 下午 03 时 42 分 43 秒	观测长度	80
上次修改时间	2011 年 04 月 24 日 星期日 下午 03 时 42 分 43 秒	删除的观测数	0
保护		已压缩	NO
数据集类型		已排序	NO
标签			
数据表示法	WINDOWS_32		
编码			

引擎/主机相关的信息

数据集页面大小	8192
数据集页数	2
首数据页	1
每页最大观测数	101
首数据页的观测数	80
数据集修复数	0
文件名	E:\SAS\data\health. sas7bdat
创建版本	9. 0202M2
创建主机	XP_PRO

按字母排序的变量和属性列表

#	变量	类型	长度	输出格式	输入格式	标签
2	armydate	字符	20	$20.	$20.	armydate
1	birthdate	字符	20	$20.	$20.	birthdate
6	bp1	数值	8			bp1
7	bp2	数值	8			bp2
3	health	数值	8			health
4	height	数值	8			height
5	weight	数值	8			weight

图 2-3 CONTENTS 过程的运行结果

4. 程序一般格式

(1)LIBNAME 语句　用于定义逻辑库,并指定该逻辑库的存储位置,一般格式是:

```
LIBNAME 逻辑库名“物理存储位置”;
```

注意程序中物理存储位置的引号不能遗漏,且必须是英文格式。逻辑库名(即库标记)可根据使用者的习惯自行命名。

(2)IMPORT 过程　用于从外部文件读取数据并创建 SAS 数据集,一般格式是:

```
PROC IMPORT DATAFILE = “文件地址及全名” | TABLE = “表名”
  OUT = <逻辑库名. >生成数据集名 <DBMS = 标识名 > <REPLACE >;
  <其它语句 >;
RUN;
```

注意在 SAS 环境中录入或读入数据时,变量名必须是英文格式。

PRPLACE 选项表示在生成文件时,自动替换已经存在的同名文件。

程序 2 - 1 的第二步中,语句“getnames = yes”表示从第一行读入变量名,在 EXCEL 格式的数据中常用。这是系统默认的一种读入变量名称的方式,因此,该语句也可以省略。

不同格式的外部文件使用不同的 DBMS 标识名,常见的数据格式及其 DBMS 标识名见表 2 - 1。

表 2 - 1　不同格式外部文件使用的 DBMS 标识名

DBMS 标识名	数据格式	后缀
ACCESS	Microsoft Access 数据库	. MDB
DBF	dBASE 文件	. DBF
WK1	Lotus 1 表	. WK1
WK3	Lotus 3 表	. WK3
WK4	Lotus 4 表	. WK4
EXCEL	EXCEL V4 或 V5 表	. XLS
EXCEL4	EXCEL V4 表	. XLS
EXCEL5	EXCEL V5 表	. XLS
EXCEL2000	EXCEL 97 或 2000 表	. XLS
DLM	固定分隔符的文本文件(缺省以空格为分隔符)	. *
CSV	逗号为分隔符的文本文件	. CSV
TAB	表格符为分隔符的文本文件	. TXT

(3)CONTENTS 过程　用于显示指定 SAS 数据集的各种属性信息,以及数据集中的全部变量及其属性。变量信息列表将按照字母顺序排列,变量属性信息包括变量类型、长度、标签以及格式等。CONTENTS 过程的一般格式是: