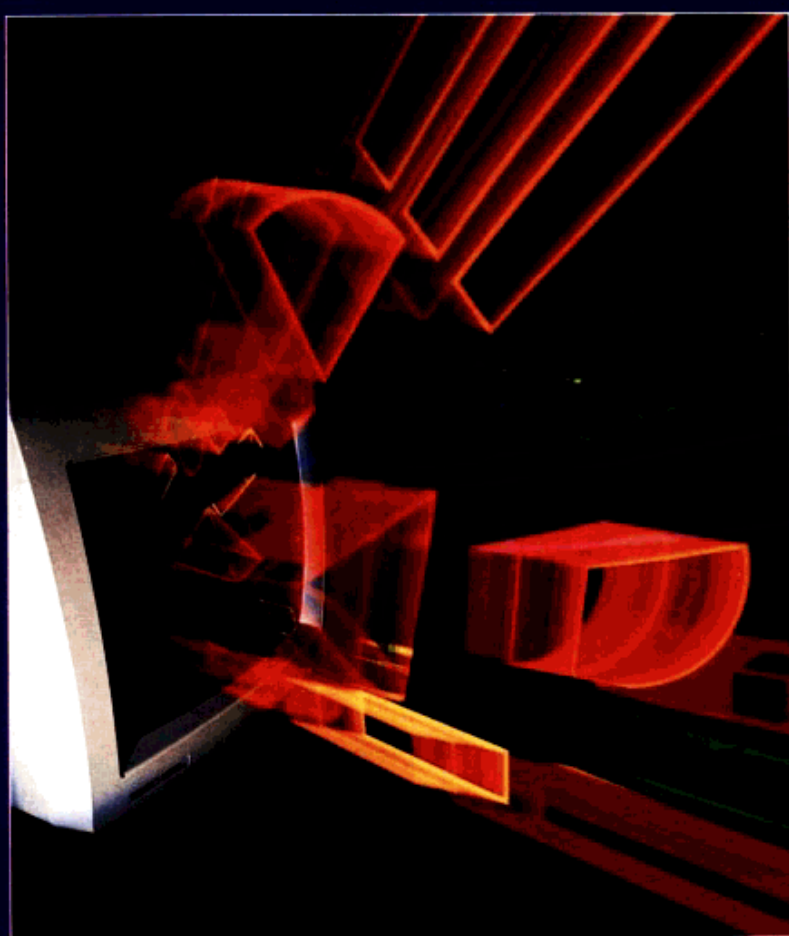


全国统计系统计算机应用技术证书考试 (NIT)

SAS基础

王吉利 主编



 中国统计出版社
China Statistics Press

前 言

本书是根据国家统计局统计教育中心和教育部考试中心联合下发《关于在统计系统开展全国计算机应用技术证书考试项目的通知》有关规定编写的,是统计系统全国计算机应用技术证书考试《SAS 基础》模块的指定教材。

统计系统全国计算机应用技术证书考试(NIT)项目由国家统计局统计教育中心与教育部考试中心教育测量中心共同组织、管理。全国计算机应用技术证书考试(NIT)是教育部考试中心主办的计算机应用技能的培训与考试系统,它采用了系统化的设计、模块化的结构、个性化的教学、规范化的考试和国际化的标准,为用人单位提供了一个客观、统一、规范的标准,适合各种行业人员岗位培训的需要。《SAS 基础》模块是学习统计分析软件中有关的数据采集、数据管理、数据显示和统计分析应用的模块。它适用于需要运用统计分析软件进行基本统计分析和数据处理的各级政府统计人员、企事业管理人员、信息分析人员和数据处理人员。

为了适应 NIT 这种新型的考试模式,本书改变了传统教材的写法,采用任务驱动,引导学生具体任务的操作中掌握知识,充分体现了 NIT 的教学思想。除作为统计系统 NIT 的教学及培训教材外,本书还可作为 SAS 基础知识的自学和参考用书。

本书由十二个单元构成,每个单元又包含若干个任务。书后附有本模块的培训与考试大纲和作业设计等内容。

本书由国家统计局统计教育中心组织编写,由国家统计局统计教育中心主任王吉利同志主编,教育部考试中心教育测量中心主任潘阳同志、国家统计局统计教育中心高级工程师乐宜同志任副主编。

本书的第 1、2、3、4、5、6、7、9、12 单元由国家统计局统计教育中心吴小武同志编写,第 8、10、11 单元由国家统计局统计教育中心徐加林同志编写。在编写过程中得到了上海宏能软件公司总经理张少华博士、中国人民大学周山英教授和清华大学徐士良教授、国家统计局粮农中心李文海同志和农调队于新华同志的指导和帮助,在此表示诚挚的谢意。

由于编者知识及水平有限,本书存在不少缺点和不足,希望在不断积累经验的基础上,根据情况的发展及时修改,请广大读者指正。

编者
2001 年 1 月

目 录

| | |
|---------------------------------|----|
| 第一单元 SAS 概述 | 1 |
| 一、SAS 系统主体的构成 | 1 |
| 1. SAS 数据库部分 | 1 |
| 2. SAS 分析核心 | 1 |
| 3. SAS 开发工具及呈现工具 | 1 |
| 4. SAS 对分布处理模式的支持及数据仓库的设计 | 1 |
| 二、SAS 各模块的主要功能 | 2 |
| 第二单元 SAS 基础 | 6 |
| 一、SAS 数据集 | 6 |
| 二、SAS 数据库 | 7 |
| 1. SAS 数据库 | 7 |
| 2. SAS 文件的两级名 | 7 |
| 三、SAS 变量 | 8 |
| 1. 类型 | 8 |
| 2. 命名规则 | 8 |
| 3. 输入、输出格式 | 8 |
| 4. 标签 | 10 |
| 四、SAS 日期、时间及日期时间值 | 10 |
| 1. 存贮方式 | 11 |
| 2. 输入格式的描述 | 11 |
| 3. 输出格式的描述 | 11 |
| 五、SAS 表达式 | 13 |
| 1. SAS 常数 | 13 |
| 2. SAS 算符 | 13 |
| 第三单元 SAS 校园学习版概述 | 16 |
| 任务一 SAS 校园学习版简介 | 16 |
| 一、功能及特色 | 16 |
| 二、运行环境要求 | 16 |
| 三、光盘包括的内容 | 17 |
| 任务二 SAS 校园学习版的安装和运行 | 20 |

| | |
|--------------------|-----------|
| 一、安装步骤 | 20 |
| 二、运行 | 20 |
| 任务三 定义数据库(LIBRARY) | 22 |
| 任务四 系统设置 | 23 |
| 任务五 SAS 显示管理系统 | 24 |
| 任务六 更新密码 | 25 |
| 练习题 | 26 |
| 第四单元 数据采集 | 27 |
| 任务一 文本文件快速转换 | 27 |
| 一、定义外部数据源 | 27 |
| 二、数据预览 | 29 |
| 三、修改数据属性 | 30 |
| 任务二 文本文件转换 | 31 |
| 一、定义外部数据源 | 32 |
| 二、生成变量 | 33 |
| 三、修改已生成的变量 | 35 |
| 四、保存数据文件 | 36 |
| 任务三 创建、查看数据集 | 36 |
| 一、菜单功能简介 | 36 |
| 二、查看数据集 | 38 |
| 三、创建数据集 | 38 |
| 练习题 | 40 |
| 第五单元 数据管理 | 41 |
| 任务一 合并 SAS 数据表 | 41 |
| 一、方式一 | 42 |
| 二、方式二 | 43 |
| 三、方式三 | 44 |
| 任务二 SAS 数据表转置 | 44 |
| 任务三 SQL 查询 | 46 |
| 练习题 | 52 |
| 第六单元 图形报表 | 53 |
| 任务一 SAS 报表制作 | 53 |
| 任务二 SAS 图形应用 | 56 |
| 任务三 SAS 图形高级应用 | 59 |
| 练习题 | 62 |

| | |
|---|-----|
| 第七单元 统计分析应用 | 63 |
| 任务一 统计分析的基本概念 | 63 |
| 一、数据的收集 | 63 |
| 二、总体、样本和统计量 | 63 |
| 三、描述统计和推断统计 | 64 |
| 四、统计分析的过程 | 64 |
| 任务二 分析员应用 | 65 |
| 任务三 交互数据分析 | 68 |
| 练习题 | 72 |
| | |
| 第八单元 描述统计 | 73 |
| 任务一 频数统计 | 73 |
| 任务二 计算常用的描述性统计量 | 79 |
| 一、常用的描述性统计量 | 79 |
| 二、常用描述性统计量的计算 | 84 |
| 任务三 数据分析图形化 | 94 |
| 一、生成柱状图(Bar Chart)和直方图(Histogram) | 94 |
| 二、盒型图(Box and Whisker Plots) | 104 |
| 三、用“分析员应用”绘制柱状图、直方图、盒型图和累积分布图 | 108 |
| 练习题 | 112 |
| | |
| 第九单元 置信估计与假设检验 | 114 |
| 任务一 置信估计 | 114 |
| 一、概述 | 114 |
| 二、用“交互数据分析”做均值的置信区间 | 115 |
| 三、用“分析员应用”做置信区间 | 117 |
| 任务二 总体参数检验 | 117 |
| 一、概述 | 117 |
| 二、用“交互数据分析”做均值检验 | 118 |
| 三、用“分析员应用”做参数检验和置信区间 | 119 |
| 任务三 总体参数比较 | 128 |
| 一、概述 | 128 |
| 二、用“交互数据分析”做总体参数比较 | 128 |
| 三、用“分析员应用”做总体参数比较 | 128 |
| 练习题 | 137 |
| | |
| 第十单元 方差分析 | 139 |
| 任务一 单因素方差分析 | 139 |
| 一、方差分析的基本概念 | 139 |

| | |
|---|-----|
| 二、在“交互数据分析”环境作单因素方差分析 | 145 |
| 三、在“分析员应用”环境作单因素方差分析 | 151 |
| 任务二 均值估计与比较 | 155 |
| 一、概述 | 155 |
| 二、在“分析员应用”中进行均值比较 | 156 |
| 练习题 | 159 |
| | |
| 第十一单元 回归分析 | 160 |
| 任务一 绘制散点图考察变量间的关系 | 160 |
| 一、散点图的概念 | 160 |
| 二、在 SAS 系统中生成散点图 | 161 |
| 任务二 计算相关系数 | 169 |
| 一、相关系数的基本概念 | 169 |
| 二、相关系数的计算 | 169 |
| 任务三 制作散点图矩阵和置信椭圆 | 174 |
| 一、制作散点图矩阵 | 174 |
| 二、生成散点图矩阵和置信椭圆 | 176 |
| 任务四 拟合一元线性回归方程 | 177 |
| 一、简单线性回归的概念 | 177 |
| 二、在 SAS 系统中拟合线性回归 | 180 |
| 任务五 利用回归方程进行预测 | 188 |
| 一、产生预测值 | 188 |
| 二、回归与预测值的置信区间 | 189 |
| 练习题 | 190 |
| | |
| 第十二单元 SAS V6.12 介绍 | 193 |
| 任务一 启动 SAS V6.12 版 | 193 |
| 任务二 SAS 校园学习版部分功能的实现 | 194 |
| | |
| 全国计算机应用技术证书考试(NIT)培训与考试大纲(SAS 基础) | 196 |
| | |
| 附件一 全国计算机应用技术证书考试(NIT) | |
| 学员评估记录表 SAS 基础模块 | 200 |
| 附件二 全国计算机应用技术证书考试(NIT) | |
| 作业设计考核表 SAS 基础模块 | 202 |
| 附件三 SAS 基础模块作业设计 | 203 |

SAS 是用于数据分析和决策支持的大型集成式、模块化系统。由于该系统在 60 年代末及 70 年代以统计分析及线性数学模型为主并且在当时以此闻名于世，所以该系统早期的全称为“STATISTICAL ANALYSIS SYSTEM”（统计分析系统）。这个全称的英文缩写“SAS”便很自然地成为开发此系统的公司的名称“SAS INSTITUTE INC.”（SAS 研究所）及其产品的总商标。

SAS 系统经过二十多年的发展，以其卓越的数据处理能力为在线数据分析、数据仓库、数据挖掘和决策支持提供了全面解决方案。SAS 系统是一个由三十多个专用模块组成的大型集成式软件包。其主要功能包括：客户机/服务器计算、数据访问、数据存储及管理、应用开发、图形处理、数据分析、报告编制、质量控制、项目管理、计算机性能评估、运筹学方法、计量经济学与预测等等。在数据处理和统计分析领域，SAS 系统被誉为国际标准软件。

一、SAS 系统的主体构成

尽管 SAS 显得很庞大，但它基本上由四大部分构成。

1. SAS 数据库部分

主要由 BASE SAS 模块提供常见的商业型数据库管理功能；BASE SAS 是 SAS 的基础，所有其他模块都必须与之结合起来使用。

2. SAS 分析核心

这一部分是 SAS 的灵魂，也是 SAS 与其他各种软件系统的本质区别，换言之，SAS 分析与决策支持功能是严肃的、权威的，这一点无论是从商业应用，还是从科学研究方面来讲都是可以确认的。属于这一部分的主要模块有：SAS/STAT, SAS/CPE, SAS/ETS, SAS/IML, SAS/INSIGHT, SAS/LAB, SAS/OR, SAS/PH-CLINICAL, SAS/QC 与 SAS/SPECTRAVIEW。

3. SAS 开发工具及呈现工具

主要由 SAS/AF, SAS/EIS 与 SAS/GRAPH 模块提供便捷的面向对象的开发工具，该工具支持客户/服务器的应用开发。

4. SAS 对分布处理模式的支持及其数据仓库设计

属于这一部分的主要模块有：SAS/ACCESS, SAS/CONNECT, SAS/SHARE 与 SAS/WA。

二、SAS 各模块的主要功能

BASE SAS 是 SAS 的核心,其它各模块均在 BASE SAS 提供的环境中运行。用户可根据应用需要选择相应的模块与 BASE SAS 一起构成一个用户化的 SAS 系统。具体来讲,BASE SAS 负责数据管理、交互应用环境管理、进行用户语言处理、调用其他 SAS 产品等。BASE SAS 除对 SAS 的数据库提供丰富的数据管理功能外,还支持标准的 SQL 语言对数据进行操作。BASE SAS 能够制作从简单列表到比较复杂的统计报表和用户自定义式样的复杂报表。BASE SAS 可以进行基本的描述性统计、变量间相关系数的计算、正态分布检验等。同时它还支持长数据名,并具有强化了了的 WEB 功能。

SAS/ACCESS 提供了与目前许多流行数据库软件的接口,从而可以实现对众多不同格式的数据进行查询、访问和分析。利用 SAS/ACCESS,可建立外部异构数据库的一个统一的公共数据界面。SAS/ACCESS 提供的与外部数据库的接口是透明和动态的。用户不必将数据真正读入 SAS 数据库,而只需在 SAS 中建立对外部数据的描述(即 VIEW)文件,便可将此文件当作真正存储着数据的 SAS 数据集一样使用。对一些反复经常使用的外部数据,可以利用 SAS/ACCESS 将数据真正提取进入 SAS 数据库。SAS/ACCESS 提供的接口是双向的,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。目前,SAS/ACCESS 支持的数据库主要有:IMS/DL1、SQL/DS、DB2、ADABAS、Rdb、ORACLE、Sybase、INGRES、Informix、DBF/DIF、ODBC 等。

SAS/AF 是一个应用开发工具。用户使用 SAS/AF 可将包含众多现成功能的 SAS 软件作为方法库,利用 SAS/AF 的屏幕设计能力以及 SCL 语言的处理能力来快速开发各种应用系统。SAS/AF 采用了先进的面向对象编程(OOP—Oriental Object Program)技术,使用户可方便快速地开发各类具有图形用户界面(GUI—Graphic User Interface)的应用系统。

SAS/ASSIST 提供面向任务的菜单驱动界面,借助它可以通过菜单系统来使用 SAS 的其他产品。它自动生成的 SAS 程序既可辅助有经验的用户快速编写 SAS 程序,又可帮助用户学习 SAS 语言。

SAS/CALC 为功能完善的多维电子表格软件,它具有财务分析、数值建模、数据整合及管理的能力。初学者和有经验的人员都能通过一个标准表格界面管理和分析电子表格中的数据。程序员和应用开发人员可以使用强有力的编程窗口,利用电子表编程语言,建立完整的应用。

SAS/CFO Vision 用于财务整合和报告,内部包含了会计知识,为日常财务工作提供了现成的程序,并提供了访问所有主要数据源的接口。它主要用于:访问财务和非财务的有关信息,整合财务数据,通过一个财务信息仓库来管理业务结构,通过财务报告和分析帮助理解财务的结果,并在组织内交流关键的业务结果信息。

SAS/CONNECT 在目前标准的网络环境下,通过 SAS/CONNECT 可使各平台的 SAS 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机器资源。SAS/CONNECT 既提供远程计算服务,也提供远端数据服务。SAS/CONNECT 支持多种分布处理模式,包括流

行的 Client/Server 方式。目前, SAS/CONNECT 支持 MVS、CMS、VSE、OpenVMS、UNIX、OS/2、Windows、Windows95、Windows NT、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、ASYN、HLLAPI、3270 等多种通讯协议。

SAS/CPE 是计算机及网络系统运行状况分析评估软件。

SAS/EIS 是决策支持表现工具, 也是一个快速应用开发工具。SAS/EIS 完全采用新兴的面向对象的编程模式 (OOP)。EIS 是以生动直观的图表方式将关键性或总结性信息呈现给使用者的非常易用的系统。

SAS/EM 是一个用于数据挖掘的产品。它基于“SEMMA”这一理念, 为用户提供了从抽样工具、数据重组、神经网络、数据回归到结果显示的许多新的过程, 使 SAS 的分析功能得到了全新的发挥。

SAS/ENGLISH 为用户提供利用自然英语获取信息的能力。

SAS/ETS 提供丰富的计量经济学和时间序列分析方法, 是研究复杂系统和进行预测的有力工具。它提供方便的模型设定手段、多样的参数估计方法。

SAS/FSP 是一个用来进行数据处理的交互式菜单系统, 它具有全屏幕数据录入、编辑和查询等功能, 同时也是一个开发工具。

SAS/GEO 可把各种地理信息数据和不同的应用整合在一个桌面系统中。它可以完成: 从 SAS 访问、修改、建立常用的地理信息数据; 实现不同格式地理信息数据间的转换; 对数据进行分析与显示。

SAS/GIS 集地理信息系统功能与空间数据的显示分析于一体。它提供层次化的地理信息, 每一层可以是某些地理元素, 也可与用户定义的主题 (例如人口、产值等) 相关联。用户可交互式地缩小或放大地图, 设定各层次显示与否, 并利用各种交互式工具进行数据显示与分析。

SAS/GRAPH 是强有力的图形软件包, 可将数据及其包含的深层信息以多种图形生动地呈现出来, 如直方图、饼形图、星形图、散点相关图、曲线图、三维曲面图、等高线图以及地图等等。SAS/GRAPH 提供一个全屏幕编辑器, 提供多种设备驱动程序, 支持非常广泛的图形输出设备以及标准的图形交换文件。

SAS/IML 提供功能强大的面向矩阵运算的编程语言, 帮助用户研究新算法或解决 SAS 中没有现成算法的专门问题。SAS/IML 中的基本数据元素是矩阵。它包含大量的数学运算符、函数和例程序, 用户用很少的语句便可描述很复杂的计算过程。

SAS/INSIGHT 是一个可视化的数据探索工具。它将统计方法与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。

SAS/IntrNet 为 SAS Web 应用提供数据服务和计算服务, 包括: htmSQL (为一 UNIX Web 服务器的 CGI 程序), 使得可以通过 Web 浏览器动态查询 SAS 数据或外部的关系型数据库; SAS ODBC Driver, 使得能通过支持 ODBC 的 Windows Web 服务器来访问 SAS 数据; SAS Driver for JDBC, 使得可以通过 Java applet 来查询 SAS 数据; SAS/IntrNet Application

Dispatcher, 使得可以通过 Web 浏览器动态地递交 SAS 程序到 SAS 应用服务器执行, 并将结果返回浏览器。

SAS/ITSV (IT Service Vision) 是企业的全面 IT 服务的性能评估和管理的软件, 这些 IT 服务包括计算机系统、网络系统、WEB 服务器和电话系统等。ITSV 将不同来源的数据进行整理和组织, 以占有最少磁盘空间同时又能反映足够信息的方式, 存放于性能数据仓库中, 用 GUI 或批处理的方式产生组织内任意层面的报告。系统程序员及网络工程师能借此识别、研究并解决有关问题, 效益及业务分析人员能借此制定资源管理的总体策略, CIO 和数据中心经理能借此定期地得到所需的汇总和分析报告。

SAS/LAB 是菜单驱动、面向任务的解释、引导式数据分析软件。

SAS/MDDB Server 是 SAS 的多维数据库产品, 主要用于在线分析处理 (OLAP), 可将数据仓库或其它数据源来的数据以立体阵列的方式存储, 以便于用多维数据浏览器等工具快速和方便地访问数据。

SAS/NVISION 是用于高标准影像制品的设计和制作软件。

SAS/OR 提供全面的运筹学方法, 是一种强有力的决策支持工具。它辅助人们实现对人力、时间以及其他各种资源的最佳利用。它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、分配问题等。它还包含用于项目管理、时间安排和资源分配等问题的一整套方法。

SAS/QC 为全面质量管理提供了一系列工具。它也提供一套全屏菜单系统, 引导用户进行标准的统计过程控制以及试验设计。SAS/QC 提供了多种不同类型控制图的制作与分析。PARETO 图 (排列图) 可用于发现需优先考虑的因素; ISHIKAWA 图 (鱼骨图) 可用于直观地进行因果分析。

SAS/SHARE 是 SAS 中进行数据库的并发性控制的专门软件。它对同时有多个用户操作同一文件的系统提供一个可靠、高效的多用户数据访问环境。

SAS/SPDS (Scalable Performance Data Server) 是一高性能的多用户数据服务器, 主要针对数据仓库应用中的数据检索来设计。该服务器充分利用了 SMP 结构的特性以提高 I/O 的效率。

SAS/SPECTRAVIEW 是进行多维数据观察、分析、研究的一个交互式、立体、可视化工具。

SAS/STAT 覆盖了所有的实用数理统计分析方法, 是国际统计分析领域的标准软件。SAS/STAT 提供了十多个过程, 可进行各种不同模型或不同特点数据的回归分析, 如: 正交回归、响应面回归、LOGISTIC 回归、非线性回归等, 且具有多种模型选择方法。可处理的数据有实型数据、有序数据和属性数据, 并能产生各种有用的统计量和诊断信息。在方差分析方面, SAS/STAT 为多种试验设计模型提供了方差分析工具。更一般的, 它还有处理一般线性模型和广义线性模型的专用过程。在多变量统计分析方面, SAS/STAT 为主成分分析、典型相关分析、判别分析, 对应分析和因子分析等提供了许多专用过程。SAS/STAT 还包含多种聚类准则的聚类分析方法。

SAS/TOOLKIT 是 SAS 功能扩充工具。用户可使用各种高级语言编写 SAS 过程、函数和输入输出格式等，然后通过该软件转化为 SAS 的组成部分。

SAS/WA (WAREHOUSE ADMINISTRATOR) 是建立数据仓库的集成工具，它在其他 SAS 软件的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, Info marts 的实现。

为了适应市场的需要，加大对用户的支持，SAS 还根据行业的需要，将模块整合成系统，为不同的行业提供适合行业的专用解决方案。目前已形成的有专用于医药行业的 **SAS/PH-CLINICAL**，专用于企业财务管理的 **SAS/CFO**，专用于制作各种企业报表和报告的 **SAS/ER** (ENTERPRISE REPORTER) 等。

第二单元

SAS 基础

用户在使用 SAS 系统之前必须了解一些 SAS 的基础知识，特别是 SAS 系统如何对数据进行管理，这一单元我们就针对 SAS 校园学习版来介绍有关这方面的内容，这部分内容也适用于 SAS V6.12 版本。

一、SAS 数据集

在 SAS 中，只有 SAS 数据集才能被 SAS 过程（PROCEDURE）使用。SAS 数据集是关系型的，它通常分为两部分：描述部分和数据部分。描述部分包含了一些关于数据属性的信息；数据部分包括数据值。SAS 的数据值被安排在一个矩阵式的表状结构中，如图 2.1 所示。表的列称之为变量（VARIABLE），变量类似于其他文件类型的域或字段（FIELD）；表的行称之为观测（OBSERVATION），观测相当于记录（RECORD）。

| | NAME | SEX | AGE | GRADE |
|---|------|-----|-----|-------|
| 1 | 孙成钢 | 男 | 19 | 605 |
| 2 | 刘晓菲 | 女 | 18 | 598 |
| 3 | 唐能通 | 男 | 20 | 587 |
| 4 | 王楠 | 女 | 19 | 591 |
| 5 | 叶滔 | 男 | 19 | 602 |
| 6 | 李媛媛 | 女 | 18 | 595 |

图 2.1 SAS 数据集的矩阵式结构

在 SAS 中共有两种类型的数据集：SAS 数据文件（SAS DATA FILE）和 SAS 数据视窗（SAS DATA VIEW）。

SAS 数据文件不仅包括描述部分，而且包括数据部分，其数据部分包含着所有数据的实体，它们一条观测紧接着另一条观测排列在文件的内部。SAS 数据视窗只有描述部分，没有数据部分，它实际上并不包含数据实体，只是包含了与其他数据文件或者其他软件数据的映射关系，它的描述部分包含了足够的信息以便使 SAS 的所有过程可以访问到实际上并不包含在 SAS 数据视窗内部的数值。

二、SAS 数据库

1. SAS 数据库

SAS 数据集存储在被称为 **SAS 数据库 (LIBRARY)** 的文件集中。SAS 数据集是最常用的一种 SAS 文件类型，但 SAS 数据库可以包含其他类型的 SAS 文件。这些数据库使得 SAS 能够在 SAS 程序中指明并查阅文件。

SAS 数据库在不同的操作系统下有不同的表示和命名规定，它们大致等同于操作系统的文件管理方式。以 WINDOWS 操作系统为例，SAS 数据库的组织方式是以目录为基础的。

无论使用哪一种操作系统都需要为每个 SAS 数据库指定一个库标记来识别，库标记仅仅是 SAS 文件的物理位置在 SAS 中的一个统一的逻辑标识，用这个方法可以将不同操作系统的特性隐藏起来。对用户而言，只要访问不同的库标记就可以访问到存储在不同操作系统中任何位置的 SAS 文件。

在编程状态下，库标记的指定可使用“LIBNAME”语句，例如：在 WINDOWS 操作系统中，要指定目录“C:\EXERCISE”为库标记“EXERCISE”，可在 PROGRAM EDITOR（程序编辑窗）中提交下面的语句：

```
PROGRAM EDITOR - (Untitled)
LIBNAME EXERCISE 'C:\EXERCISE';
```

SAS 数据库分为永久库和临时库两种。临时库只有一个，名为 WORK，它在每次启动 SAS 后自动生成，结束 SAS 后库中的数据被自动删除。用户可以使用“LIBNAME”语句指定永久库的库标记，永久库中的所有文件将被保留，但库标记仍然是临时的。每次启动 SAS 时都自动指定两个库标记：SASUSER 和 WORK。库标记 WORK 被 SAS 定义为缺省的数据库，在 SAS 程序中引用 WORK 库中的数据集可以省略库标记。

在 SAS 校园学习版中，库标记的指定和使用请参考第三单元中的有关内容。

2. SAS 文件的两级名

每个 SAS 文件都有一个两级名，第一级是库标记，第二级是文件名，中间用“.”隔开，其一般形式为：*libref.SAS-filename*，其中：*libref* 是库标记，*SAS-filename* 是文件名。在程序中通过指定两级名来识别 SAS 文件。例如：下面的语句中引用了 EXERCISE 库中的名为 CLASS 的数据集。

```
PROC PRINT DATA=EXERCISE.CLASS;
```

在 SAS 校园学习版中，当通过菜单驱动来打开一个 SAS 数据集时，通常情况下系统会弹出一个窗口，让用户进行选择，首先应先选择一个 SAS 数据库，再从库中的数据文件中选择一个 SAS 数据集，这种操作方法也体现了如何在菜单驱动中通过 SAS 文件的两级名来识别 SAS 文件。

三、SAS 变量

描述给定特征的数值集合构成变量 (VARIABLE)。在 SAS 数据集中每条观测包含每个变量的一个值。在图 2.1 中每一列数值就是一个变量。例如：第一列构成变量 NAME，它包含所有成员的名字；第二列构成变量 SEX，包含了他们的性别值，等等。

1. 类型

SAS 变量共有两种类型：字符型变量 (CHARACTER) 和数值型变量 (NUMERIC)。

- 字符型变量：以 ASCII 码存放，最大长度不超过 200 个字符，数据值可以是字符、字母、特殊字符以及数字。

- 数值型变量：以浮点数存放，最大长度为 8 个字节，数据值只能是数值。

SAS 数据集的矩阵式结构要求每个观测的每个变量值都必须存在，因此，如果某个数据值缺失，系统将自动补上一个缺失值。对于数值型变量，这个值显示为一个点“.”，而对于字符型变量，这个值显示为空格。

2. 命名规则

SAS 变量命名规则有：变量名长度为 1-8 个字符；始于一个字母或下划线“_”；紧接数字、字母或下划线。

注意：不能用中文字符来命名一个 SAS 变量。

3. 输入、输出格式

SAS 数据集的数据值的内部存放格式并不一定与该数据值的输入和输出格式一致，如：我们日常使用的数值“1,000,000”，通常不能认为它是一个数值，因为它包含逗号，但我们可以指示 SAS，不管逗号，把它读为“1000000”；另外输出位数较多的数值时，加上逗号更利于阅读，我们也可以指示 SAS 将数值“1000000”输出为“1,000,000”或“\$1,000,000”，这些操作的实现都要利用 SAS 的两个重要功能：输入格式 (INFORMAT) 和输出格式 (FORMAT)。输入格式指示 SAS 如何读入数据；而输出格式指示 SAS 如何输出数据。它们的一般形式如下。

`<$>informat-name<w>.<d>`

`<$>format-name<w>.<d>`

\$ 表示变量是字符型的

informat-name 输入格式名

format-name 输出格式名

w 输入或输出的宽度

· 必须的分隔符

d 小数部分的长度

注意：所有输入（或输出）格式名都必须包含一个点“.”作为名字的一部分。

用一个输出格式的例子“DOLLAR10.2”说明如下：①格式中没有“\$”符号，表示该变量是数值型变量。②格式名为“DOLLAR”。③*w*的值为10，它规定了输出这个值最多用10列。④*d*的值为2，它规定对这个值取两位小数。⑤小数点本身占一列。⑥剩下的7列(10-2-1)是这个值中所有其他字符，包括其余的数值、“\$”号和“,”号，如果是负数，还有一个负号。

SAS 提供的几种常用的输入、输出格式描述如下。

(1) 输入格式

①字符型变量的输入格式描述

| 输入格式名 | 格式描述 | 宽度范围 |
|-------------------|---------------|-------|
| <i>\$w</i> | 读入标准的字符数据 | 1-200 |
| <i>\$CHARw</i> | 保留开头和结尾的空格 | 1-200 |
| <i>\$HEXw</i> | 转换十六进制数据为字符数据 | 1-200 |
| <i>\$VARYINGw</i> | 读入可变长度的字符值 | 1-200 |

②数值型变量的输入格式描述

| 输入格式名 | 格式描述 | 宽度范围 |
|-----------------|---|------|
| <i>w.d</i> | 读入标准的数值数据 | 1-32 |
| <i>COMMAw.d</i> | 移走数值中被嵌入的字符（如：逗号、空格、\$、%、破折号和圆括号），当从带有圆括号的数值中移走圆括号时，在这个值域的开头加上负号。 | 1-32 |
| <i>Ew</i> | 读入用科学计数法E表示的数值 | 7-32 |

(2) 输出格式

①字符型变量的输出格式描述

| 输出格式名 | 格式描述 | 宽度范围 |
|-------------|----------------|-------|
| \$w. | 输出标准的字符数据 | 1-200 |
| \$CHARw. | 输出字符数据时保留开头的空格 | 1-200 |
| \$HEXw. | 转换字符数据为十六进制数据 | 1-200 |
| \$VARYINGw. | 输出字符值时长度可变化 | 1-200 |

②数值型变量的输出格式描述

| 输出格式名 | 格式描述 | 宽度范围 |
|-----------|---|------|
| w.d | 输出标准的数值数据 | 1-32 |
| BESTw. | SAS 选择最好的表示法 | 1-32 |
| COMMAw.d | 输出含有逗号和十进制点等的数值， 如果规定 <i>d</i> 值，它必须是 0 或 2。 | 2-32 |
| DOLLARw.d | 输出含有逗号和十进制点等的数值，并在数字前加上美元符号“\$”，如果规定 <i>d</i> 值，它必须是 0 或 2。 | 2-32 |
| Ew. | 输出用科学计数法 E 表示的数值 | 7-32 |
| PERCENTw. | 输出数值为百分数 | 4-32 |
| ROMANw. | 输出数值为罗马数字 | 2-32 |

4. 标签

变量的标签是一个代替变量名的描述性标识。它至多可用 40 个字符，这些字符串可以在一些确定的 SAS 过程中替代变量名被打印出来。

在 SAS 校园学习版中，变量标签的规定请参考第四单元中的有关内容。

四、SAS 日期、时间及日期时间值

日期、时间以及日期时间值在 SAS 内部以数值型变量存储，为了让日期、时间以及日期时间值同一般数值一起操作，用户必须了解日期、时间以及日期时间值在 SAS 中的存储方式、输入格式和输出格式。

1. 存贮方式

SAS 存贮日期值为 1960 年 1 月 1 日和这个日期之间的天（日）数。

SAS 存贮时间值为从午夜开始到此刻的秒数。

SAS 存贮日期时间值为从 1960 年 1 月 1 日午夜到这个日期时间之间的秒数。

2. 输入格式的描述

SAS 提供的几种常用的输入格式描述如下。

| 输入格式名 | 格式描述 | 可读取的数值形式 | 宽度范围 |
|-----------------------------------|-------------------------------------|--|-------|
| <i>DATE</i> _w | 读日、月、年形式的日期值 (ddmmyy) | 1-oct-1999;1 oct 1999; 01 oct 99 | 7-32 |
| <i>DATETIME</i> _{w.d} | 读日期时间值 (ddmmyy hh:mm:ss.ss) | 25oct99:10:54:10.2; 25oct1999/10:54:10 | 13-40 |
| <i>DDMMYY</i> _w | 读日、月、年形式的日期值 (ddmmyy) | 251099;25/10/99; 25-10-99;25 10 99 | 6-32 |
| <i>JULIAN</i> _w | 读 Julian 日期（天文计算中常用的 计时法）(yyddd) | 99298（表示 99 年第 298 天即 1999 年 10 月 25 日） | 5-32 |
| <i>MMDDYY</i> _w | 读月、日、年形式的日期值 (mmddy) | 102599;10/25/99; 10-25-99;10 25 99 | 6-32 |
| <i>MONYY</i> _w | 读月和年份数据值 (mmyy) | oct99（1999 年 10 月） | 5-32 |
| <i>TIME</i> _{w.d} | 读小时、分、秒形式的时间值 (hh:mm:ss.ss) | 14:34:35.30 | 5-32 |
| <i>YYMMDD</i> _w | 读年、月、日形式的日期值 (yymmdd) | 991025,99/10/25 99-10-25,99 10 25 | 6-32 |
| <i>YYQ</i> _w | 读这一年的季节（1、2、3 和 4） 值 | 99Q4（相应于 1999 年 10 月 1 日-12 月 31 日） | 4-32 |

3. 输出格式的描述

SAS 提供的几种常用的输出格式描述如下。