1. SAS介绍及基础
2. SAS介绍

1. SAS软件是由SAS公司开发的集数据仓库、大规模数据处理、数据挖掘、统计分析、图表制作、网页连接等为一体的计算机软体系统。

SAS是专业的统计分析软件，它对表格数据进行操作和统计分析比用Matlab更方便更专业。

2. SAS，SPSS，EXCEL区别

都能用于处理数据和统计分析，高级程度：

SAS > SPSS > EXCEL

EXCEL——侧重表格（办公），只能处理一些简单的数据分析，公式丰富，一般多用于计算（统计分析结果不全面）；

SPSS——专业统计分析软件，界面菜单式操作，简单易学，缺点是通用性不好（处理同样的数据，也要重复点菜单）；

SAS——更专业统计分析软件，代码编程实现（通用性好），大多用于金融、医药等领域，用于大企业的数据分析，缺点是较难学。

3. 要学好SAS，也离不开《统计学》。

4. SAS应用前景

现在是大数据时代，数据挖掘、数据库维护、做数据分析，在第一时间内获得或者找到最有价值的信息和资源，成为正确决策的依据，这对金融、保险、医药、政府等企业和部门都至关重要。

SAS学习者的就业方向有：金融，银行、保险，证券、投行、临床研究，医药开发，市场调查，政府监管和教育研究部门。

5. 我适合学SAS吗？学习SAS需要懂高等数学和统计学吗？

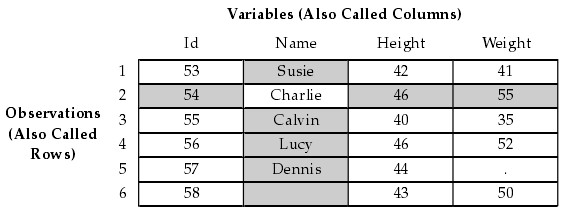
实际情况是，任何专业背景的人，都可以学习并掌握SAS，一经学会，终生受用。不少人听说SAS是个统计分析软件，就自认为需要懂得高深数学和统计学的人才能学习，其实这是一个误区。

SAS在创办起，其宗旨是着重于80%工作量的统计分析前的数据处理，至于统计分析一旦数据就绪，通过相应的分析模块，几乎象傻瓜相机一样，谁都可以操作运行。如果需要的统计方法学上提高或突破，恐怕不是统计分析的日常工作，而是统计方法学的科研了。可见学习SAS不需要具备高等数学和统计学基础，只是需要有一般逻辑思维训练基础即可。

所以，只要遵循正确的学习道路并且获得有效的指导，就可以掌握扎实的SAS编程技能和技巧，再经过一定的经验积累，您也可以成为SAS行家里手，一直有许多学中(西)医、MBA、计算机、信息管理、经济(金融)、机械自动化，甚至英语专业的朋友成功转行SAS的例子。

1. SAS基础
2. SAS语句（可随便跨行、空行，注释可以用 /\*……\*/）
3. 不区分大小写；
4. 以SAS命令关键词开始，以分号结尾；
5. 数据集

（1）即“数据表”，一行称为一个观测值；一列称为一个变量（属性）



（2）数据类型：数值型、字符型

注：“Id”等根据需要既可当作数值型或字符型。

（3）缺失值

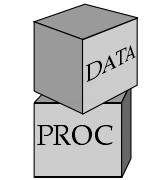
数值型缺失值用 . 表示，字符型缺失值用空格表示；

（4）命名规则：字母开始，由字母、数字、下划线组成。

注1：数据集和变量可以有描述性信息，比如数据集有创建日期、观测值数、变量数等；变量有类型、长度、格式等。

注2：数据集按一行（观测值）一行（观测值）执行。

1. SAS程序两个基本模块



数据步：以DATA开始，负责读入和修改数据，创建数据集；

过程步：以PROC开始，分析处理数据集，呈现结果或报表；

注1：好的习惯是，在它们的结尾都加上RUN，指示该模块的结束，告诉SAS去执行所有之前的程序行。

注2：一个程序可以有多个数据步、过程步。

1. 学生测验成绩：编号、性别、科目1、科目2、家庭作业。

**data** Test;

input Name $**1**-**9** Subject **11**-**12** Gender $ **14** Exam1 **16**-**18** Exam2 **20**-**22** HW\_Grade $ **24**;

datalines; /\* datalines, 数据行开始标志 \*/

Xiao Er 10 M 80 84 A

Zhang San 7 M 85 89 A

Li Si 4 F 90 86 B

Wang Wu 20 M 82 85 B

Zhao Liu 25 F 94 94 A

Liu Qi 14 F 88 84 C

; /\* 分号, 为数据行结束标志 \*/

**run**;

**proc** **print** data=Test;

title '学生测验成绩';

**run**;

**proc** **means** data=Test;

title '学生测验成绩分析';

**run**;

程序说明：Test为数据集名称；input变量名称及其在数据列中的位置，例如1-9表示第1至第9列，变量是字符型需要加上$；datalines表示数据开始，注意数据中空格数目与input中列数的对应。

运行结果：



1. SAS程序执行过程

编完SAS程序，点“提交”按钮，将执行代码，报错和执行过程出现在日志窗口，运行结果出现在结果窗口（输出结果也可以改用html格式输出）。

注：修改语法错误后，最好先清空日志窗口，再再次提交程序。

1. SAS数据集的存取及访问

（1）SAS数据集采用二级名称，即

libref.filename

其中，libref表示路径名，filename表示数据文件名。例如在路径d:\SASWorks下建立数据集data1，可以用如下代码：

libname mylib 'd:\SASWorks';

**data** mylib.data1;

程序说明：libname 定义路径变量mylib用来存储路径'd:\SASWorks'，之后就可以使用该路径变量，在其路径下创建数据集。

（2）SAS数据集的存储分为临时存储和永久存储

不指定存储位置，会默认存入Work目录下，此时为临时存储（退出SAS数据将消失）；其它目录下的存储都是永久存储。

注：二级名称“Work.data1” 等同于一级名称“data1”。