

SAS统计分析系统简介

SAS系统概述

一、什么是 SAS系统

二、 SAS系统的特点

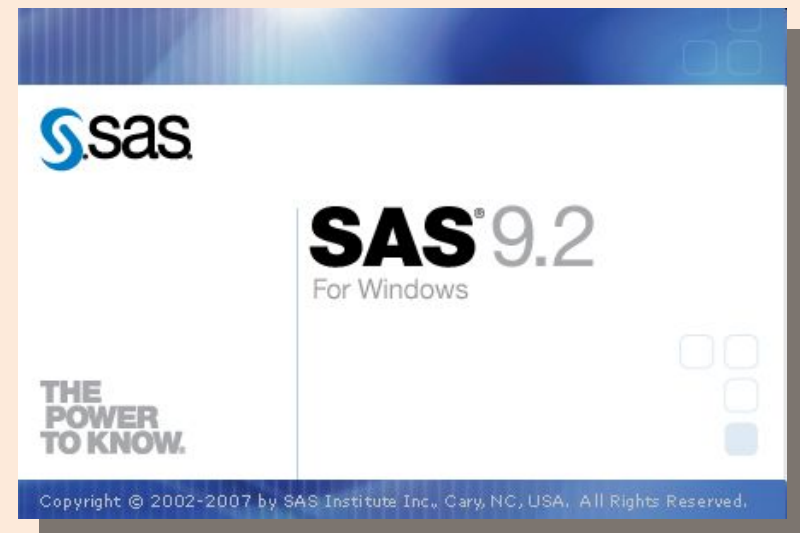
二、 SAS/BASE简介

一、什么是 SAS系统

- SAS系统是模块化的集成软件系统,提供数据访问、管理、分析、呈现。
- SAS最早只是一个数学统计软件，由两位生物统计学研究生所编写制定。
- SAS在1976年由Jim Goodnight及John Sall博士等人成立公司，并且正式推出相关软件。

1. 基本部分（核心部分）： BASE SAS;
2. 统计分析： SAS/STAT;
3. 高级绘图： SAS/GRAPH;
4. 矩阵运算： SAS/IML;
5. 运筹学和线性规划： SAS/OR;
6. 经济预测和时间序列分析： SAS/ETS;
7. 面向任务的菜单驱动界面： SAS/ASSIST

...



二、SAS系统的特点

1. 使用灵活方便、功能齐全
2. SAS语言是编程能力强且简洁易学的非过程语言
3. SAS系统把数据处理与统计分析融为一体
4. 适用性强、应用面广

经过多年的发展，统计分析系统已经遍布全世界，使用的单位遍及金融、医药卫生、生产、运输、通讯、科学研究、政府和教育等领域；在资料处理和统计分析领域，统计分析系统被誉统计软件界的巨无霸。

三、 SAS/BASE简介

可以单独使用，也可以同其它软件产品一起组成一个用户化的SAS系统。

主要功能是：

- 1.数据管理

- 2.基础统计计算功能

- 3.报表生成和图形显示功能

1.数据管理功能：

（1）信息存贮和检索：

多种格式读入数据值，组成SAS数据集。可以用文件操作管理方法把不同数据集的数据组合在一起，供SAS过程分析处理，生成报表。

（2）数据修改与程序设计

（3）文件操作

用于对数据进行编辑、整理、连接、合并及更新的功能，还能同时处理多个输入文件。

2.基础统计计算功能

（1）计算简单的描述性统计量

（2）SAS函数可用于计算概率分布函数、分位数、产生随机数等

3.报表生成和图形显示功能

SAS使用概述

一、SAS的DMS

二、SAS语句概述

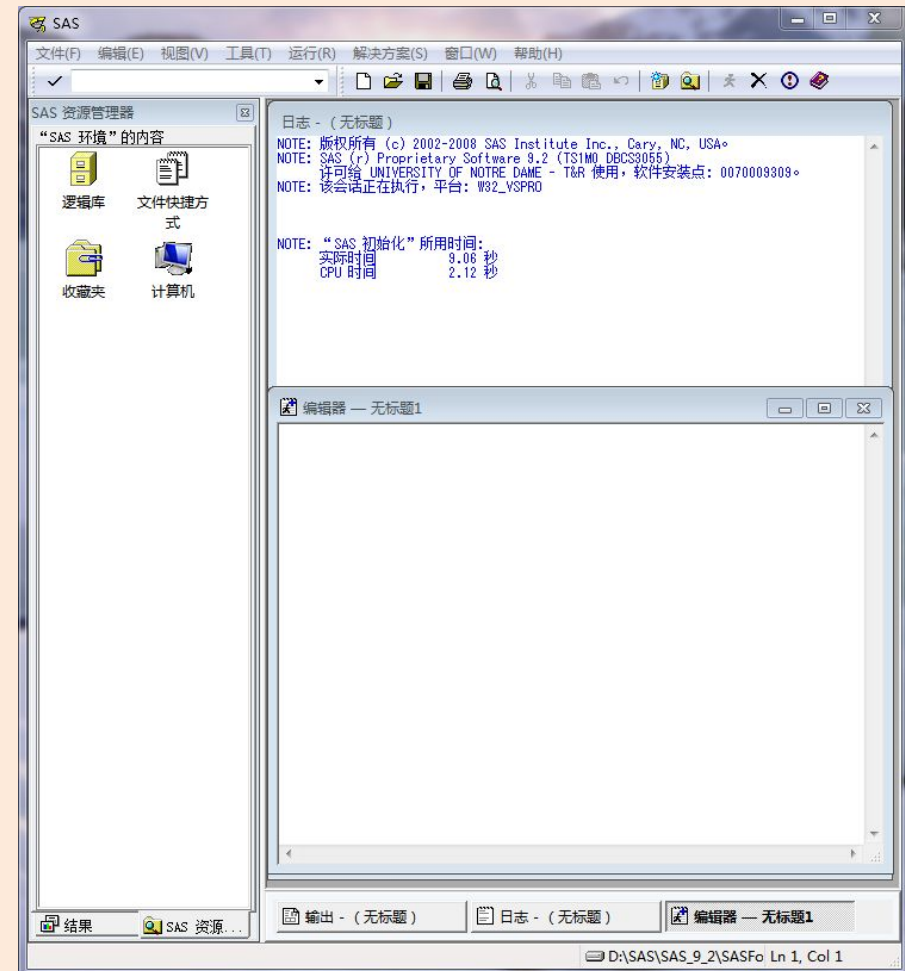
三、SAS程序概述

一、SAS显示管理系统

SAS显示管理系统(Display Manager System, DMS)

用户和SAS系统交互式会话环境；点击SAS图标后，即可进入SAS的显示管理系统DMS在DMS中有四个主要的窗口：

- (1) 编辑窗口：编辑程序和数据文件。
- (2) 日志窗口：记录运行情况，显示ERROR信息。
- (3) 输出窗口：输出运行的结果。
- (4) 图形窗口：输出图形。



二、SAS语句概述

一个SAS语句是由SAS关键词，SAS名字，特殊字符和运算符组成的字符串，并以分号（；）结尾。

SAS关键词: 如data语句，proc语句，infile语句等

SAS名字: 最多由1-8个字符组成，首字符必须是字母或下划线（字符包括：字母、数字或下划线）
如：变量名、SAS数据集名、过程名等

SAS语句 一般形式：

SAS关键词 参数…<item|item|item> 选项;

注：SAS规定，每个语句以“；”作为该语句结束的标志。

二、SAS语句概述

SAS数据类型

- 1.数值，小数、符号
- 2.字符，单括号括起来的1/ 200
字符组成
- 3.日期和时间
- 4.十六进制数值，十六进制字

SAS逻辑运算符

&	AND
	OR
^	NOT

SAS运算符

=	EQ
^=	NE
	GT
<	LT
=	GE
<=	LE
	IN

**

*

/

+

/

其他

<	MIN 最小值
<	MAX
	JOIN 连接字符

三、SAS程序概述

一序列语句组成一个SAS程序，完成一定的功能。

SAS程序中语句分为两种：

1.DATA语句:产生SAS数据集

2.PROC语句:对SAS数据集内数据进行分析处理,输出结果

SAS程序：

DATA语句开始,告诉程序有关的数据集
中间是数据
最后是PROC语句
RUN;

程序示例：

```

data ex;
  input no$ sex$ age h w;
  cards;
  10 f 47 156.3 47.1
  24 m 38 172.4 61.5
  53 m 41 169.2 64.5
  46 f 52 158.2 53.6
  38 f 39 160.1 48
  ;
proc sort;
  by sex;
proc means;
  by sex;
  var h w;
run;

```

临时表work.ex内容

	no	sex	age	h	w
1	10	f	47	156.3	47.1
2	46	f	52	158.2	53.6
3	38	f	39	160.1	48
4	24	m	38	172.4	61.5
5	53	m	41	169.2	64.5

SAS 系统

sex=f

MEANS PROCEDURE

变量	N	均值	标准差	最小值	最大值
h	3	158.2000000	1.9000000	156.3000000	160.1000000
w	3	49.5666667	3.5218366	47.1000000	53.6000000

sex=m

sex=m

MEANS PROCEDURE

变量	N	均值	标准差	最小值	最大值
h	2	170.8000000	2.2627417	169.2000000	172.4000000
w	2	63.0000000	2.1213203	61.5000000	64.5000000

程序输出结果

DATA步入门

一、什么是DATA步

二、创建与合并SAS数据集

三、条件、循环等语句概述

一、什么是DATA步

DATA步是用DATA语句开始的一组或几组SAS语句。其目的：

- 1.检索:把输入数据转化为SAS数据集；
- 2.编辑:检查数据中的错误并修改它们，计算新变量；
- 3.按照用户设计的规格输出报告并存为磁盘文件；
- 4.从已经存在的数据集中通过取子集、合并和更新原数据的方法来产生新的数据集。

二、创建与合并SAS数据集

1.创建：从作业流中输入数据产生SAS数据集

一般形式为：

{ data 语句;
input 语句;
(用于data步的其它SAS语句;)
cards;
数据行
;
}

Ex.1

```
data d1;  
input x1 x2 x3;  
y=x1+x2+x3;  
cards;  
1 2 3  
4 5 6  
;
```

	x1	x2	x3	y
1	1	2	3	6
2	4	5	6	15

注：数据之间以空格隔开；数据行与行之间不用分号。

二、创建与合并SAS数据集

2.创建：从磁盘上读入数据文件产生SAS数据集

一般形式为：

{
data 语句;
infile 语句;
input 语句;
(用于data步的其它SAS语句);
run;
}

1	1	2	3
2	4	5	6

Ex.2
data d2;
infile 'C:\tmp.dat';
input x1 x2 x3;
y=x1+x2+x3;
run;

	x1	x2	x3	y
1	1	2	3	6
2	4	5	6	15

二、创建SAS数据集

3.创建：数据来自EXCEL，可以从多个EXCEL文件导入数据集。

一般形式为：

```
PROC IMPORT OUT=数据集  
DATAFILE=文件路径  
DBMS=EXCEL REPLACE;  
RANGE='Sheet1 %';  
GETNAMES=YES;  
MIXED=NO;  
SCANTEXT=YES;  
USEDATE=YES;  
SCANTIME=YES;  
RUN;
```

	A	B	C	D
1	x1	x2	x3	
2		1	2	3
3		4	5	6

Ex.3

```
PROC IMPORT OUT=d3  
DATAFILE='C:\tmp.xls'  
DBMS=EXCEL REPLACE;  
RANGE='Sheet1 %';  
GETNAMES=YES;  
MIXED=NO;  
SCANTEXT=YES;  
USEDATE=YES;  
SCANTIME=YES;  
RUN;
```

	x1	x2	x3
1	1	2	3
2	4	5	6

SAS9.2不支持xlsx格式，可以转换为CVS格式文件后，使用import导入。

二、创建与合并SAS数据集

4.合并：从一个或几个已经存在的SAS数据集合并成新数据集。

一般形式为：

{ data 语句;
set|merge语句;
(用于data步的其它SAS语句);
run;

Ex.5

```
data d5;  
input x4;  
cards;  
1  
4  
;  
run;
```

```
data d6;  
merge d1 d5;  
run;
```

横向合并两个数据集

Ex.4

```
data d4;  
set d1 d2;  
run;
```

纵向合并两个数据集

	x1	x2	x3	y	x4
1	1	2	3	6	1
2	4	5	6	15	4

	x1	x2	x3	y
1	1	2	3	6
2	4	5	6	15
3	1	2	3	6
4	4	5	6	15

三、条件、循环等语句概述

1.条件语句 if

IF expression then statement1;
[else statement2]

2.循环语句

(1)计数DO循环

do variable = <initial value| TO
<end value| BY <length|;
Statement;
end;

(2)当型循环

do while(循环继续条件);
循环体语句;
end;

(3)直到型循环

do until(循环退出语句);
循环执行体;
end;

(4)where语句
where SAS表达式;

(5)delete语句

三、条件、循环等语句概述

```
data d7;  
input age math chinese@@;  
if age < 16 then delete;  
Sum = math+chinese;  
cards;  
15 76 67  
14 67 89  
17 45 65  
12 60 56  
;  
proc print data = d7;  
run;
```

SAS 系统				2015年
Obs	age	math	chinese	Sum
1	15	76	67	143
2	14	67	89	156
3	12	60	56	116

```
data d8;  
input age math chinese@@;  
do while(age < 16);  
delete;  
end;  
Sum = math+chinese;  
cards;  
15 76 67  
14 67 89  
17 45 65  
12 60 56  
;  
proc print data = d8;  
run;
```

```
data d10;  
set d9;  
where age < 16;  
proc print data = d10;  
run;
```

PROC步入门

一、什么是PROC 步

二、 PROC语句概览

一、什么是PROC 步?

PROC 步是用PROC语句开始的一组或几组SAS语句，是调用并执行一个计算处理过程。

一般形式为：

{ **PROC** SAS过程名 <options>;
 (用于proc步的其它SAS语句;)
run;

注：

1.不同SAS过程<options>也是不同的。

2. Proc步中常出现的SAS控制语句：

- (1) var 变量名;
- (2) by 变量名; (要求排序)
- (3) class 变量名;
- (4) ID 变量名;

Ex:

proc print

data=d1;

run;

SAS 系统				2015年
Obs	x1	x2	x3	y
1	1	2	3	6
2	4	5	6	15
3	5	4	2	8
4	1	2	8	12

Ex:

proc print DOUBLE

data=d1;

run;

在观测值间画
一条白线

SAS 系统				2015:
Obs	x1	x2	x3	y
1	1	2	3	6
2	4	5	6	15
3	5	4	2	8
4	1	2	8	12

二、PROC语句概览

1. proc print;
2. proc plot;
3. proc means; 提供变量的简单描述
4. proc sql;
5. proc univariate 详细的变量分布描述
6. proc anova 方差分析
7. proc reg 多元线性回归与相关
8. proc corr 偏相关系数
9. proc princomp 主成分分析
10. proc factor 因子分析
11. proc varclus 指标聚类,cluster、fastclus
12. proc logistic Logistic回归分析
13. proc ...

谢谢