

图解SAS软件统计分析（一）：基本知识和界面操作

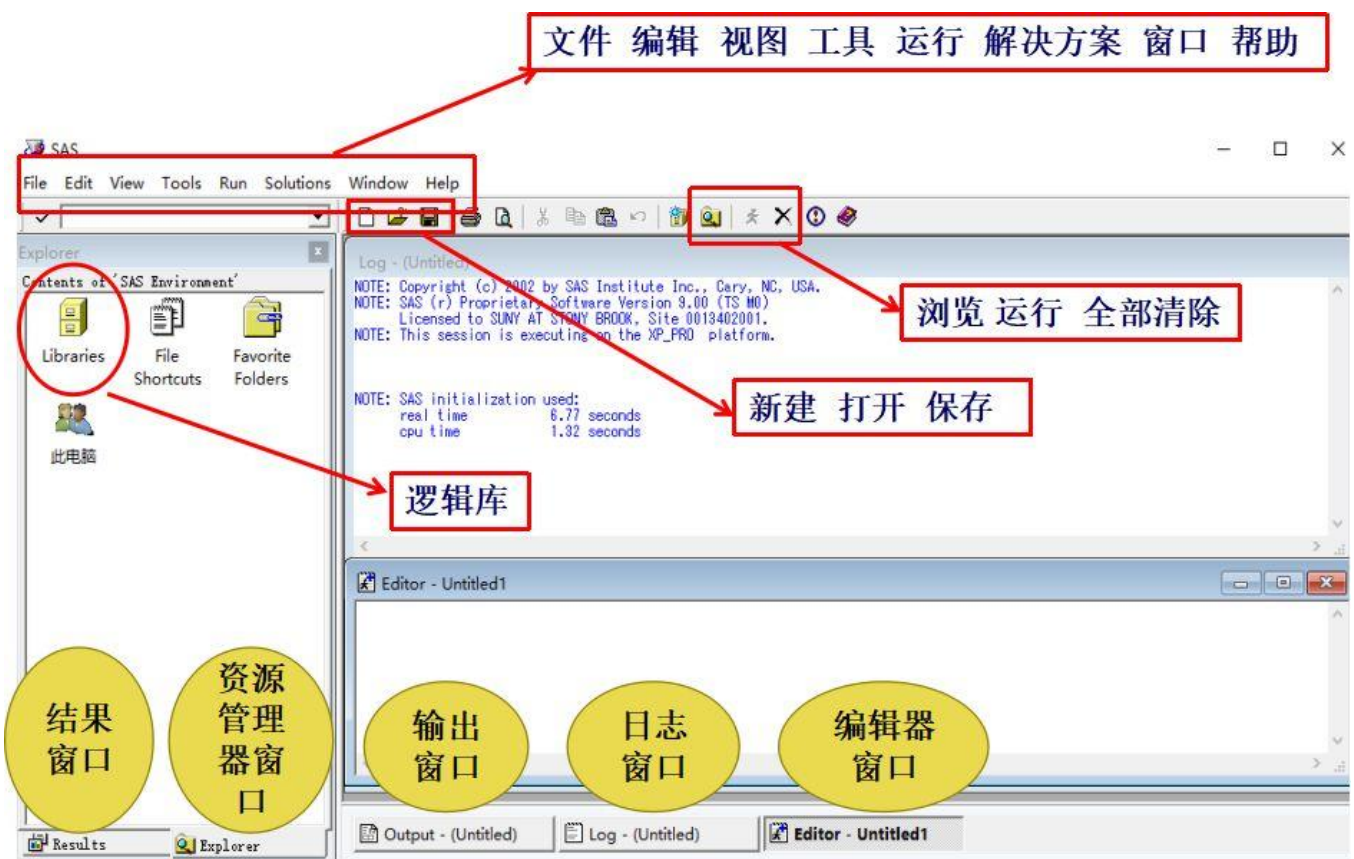
作者：writer 来源：爱科学

本文原地址：<https://www.iikx.com/news/statistics/9794.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

图解SAS软件统计分析(一)：基本知识和界面操作

。本文介绍常用统计分析软件SAS，大多数研究者一听到SAS要编程，就想着会很难，但当你熟悉SAS软件的基本操作界面之后，基本上通过拷贝复制就能得到你想要的结果。接下来详细介绍SAS的基本知识和界面操作。



一、SAS窗口简介

1、编辑器窗口

主要功能是编辑SAS程序语句，并用不同的颜色显示SAS语句，同时进行语句的逻辑检查，如：

出现红色字体，说明语句拼写有误，需要修改;深蓝色字体表示数据步或过程步的开始;浅蓝色表示关键语句;黄底色表示数据流。

2、日志窗口

主要作用是显示运行程序后的有关信息。显示的信息内容包括所建立的数据集名称、建立数据集包括多少个变量和观测、执行了什么过程、执行过程运行了多长时间、语句中有什么错误等等。

3、输出窗口

主要作用是显示程序运行的结果。

注：当程序写好时需要点击运行(小人的图标哟)，另外需要注意无论是数据集的名称还是SAS程序，SAS均不区分大小写。

二、SAS数据集与SAS数据库

SAS数据集存放于SAS数据库中，类似于excel文档存放在一个文件夹中。SAS库的类型可分为永久数据库(关闭SAS后数据集保存)和临时数据库(关闭SAS软件后SAS数据集不保存，但可以保存进建立临时数据集的程序)。

三、SAS运算符及常用函数

1、SAS运算符

+	加	-	减
*	乘	/	除
=	等于	^=	不等于
>	大于	<	小于
>=	大于等于	<=	小于等于
and	与	or	或
not	非		

2、SAS常用函数

max(x)	最大值	min(x)	最小值
mean(x,y,...)	均数	N(x,y,...)	例数
std(x,y,...)	标准差	stderr(x,y,...)	标准误
sum(x,y,...)	总和	var(x,y,...)	方差

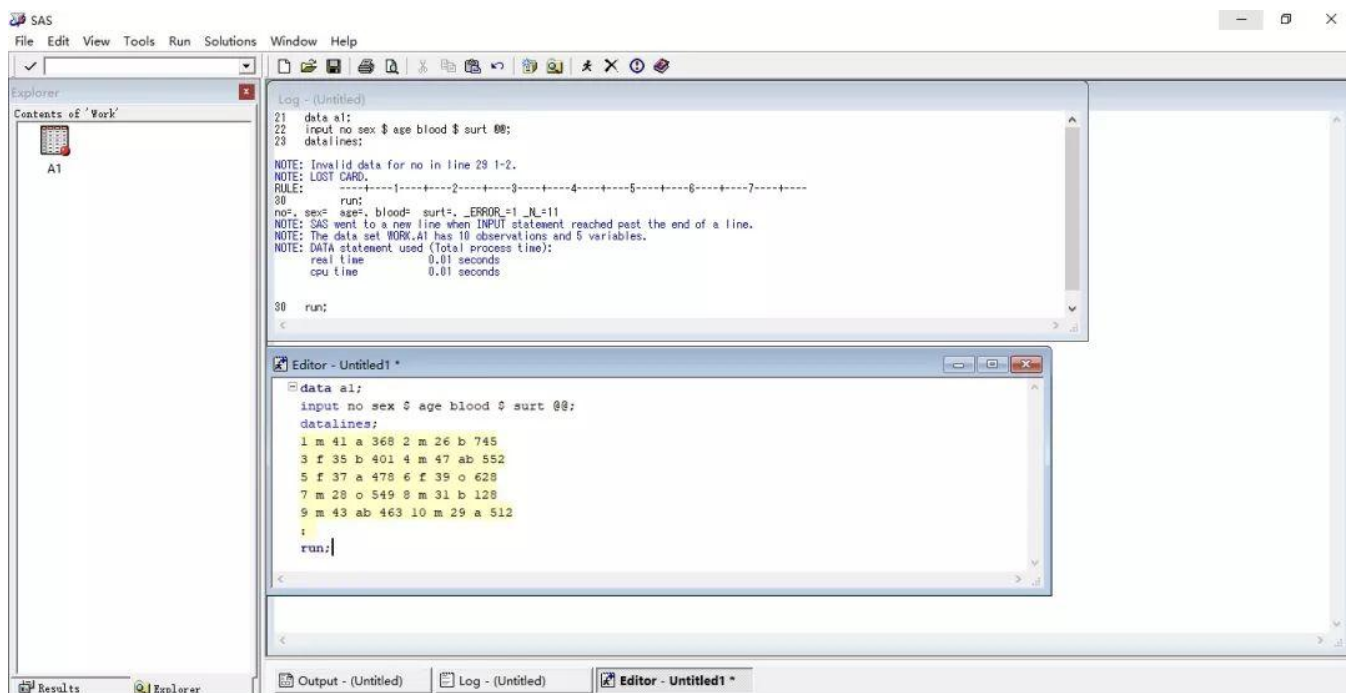
四、SAS程序简介

- 1、SAS程序结构：数据步(data step)+过程步(proc step);以run结束
- 2、SAS程序语法规则：每个语句以“;”作为结束符号;同一行可以有多个语句，一个语句也可以分几行编写。

五、建立数据集

下面就以一个实例，了解建立SAS数据集。

SAS的编辑窗口和日志窗口如下：



建立的SAS数据a1如下：

	no	sex	age	blood	surt
1	1	m	41	a	368
2	2	m	26	b	745
3	3	f	35	b	401
4	4	m	47	ab	552
5	5	f	37	a	478
6	6	f	39	o	628
7	7	m	28	o	549
8	8	m	31	b	128
9	9	m	43	ab	463
10	10	m	29	a	512

六、整理数据

常用的SAS语句：

1、if-then/else

2、删除不需要变量或保留需要变量：drop/keep语句(data步中)

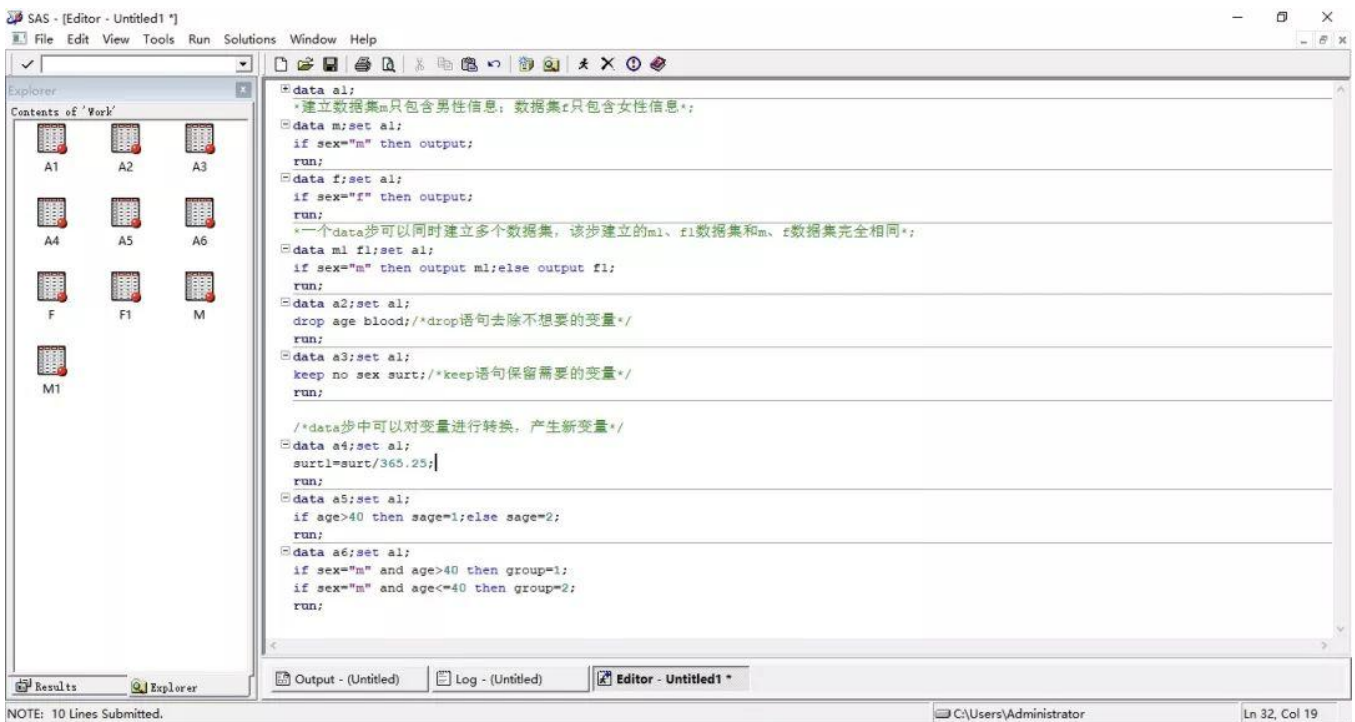
3、产生新变量：直接等号赋值即可，if-then/else产生新变量

4、数据集的排序：procsort过程步

5、数据集的合并：procmerge过程步(横向合并)

6、数据集的连接：set语句(纵向连接)

详细SAS程序如下：



```
data a1;
  *建立数据集m只包含男性信息；数据集f只包含女性信息；
data m;set a1;
  if sex="m" then output;
run;
data f;set a1;
  if sex="f" then output;
run;
/*一个data步可以同时建立多个数据集，该步建立的m1、f1数据集和m、f数据集完全相同*/
data m1 f1;set a1;
  if sex="m" then output m1;else output f1;
run;
data a2;set a1;
  drop age blood; /*drop语句去除不想要的变量*/
run;
data a3;set a1;
  keep no sex surt; /*keep语句保留需要的变量*/
run;

/*data步中可以对变量进行转换，产生新变量*/
data a4;set a1;
  surt1=surt/365.25;
run;
data a5;set a1;
  if age>40 then sage=1;else sage=2;
run;
data a6;set a1;
  if sex="m" and age>40 then group=1;
  if sex="m" and age<=40 then group=2;
run;
```

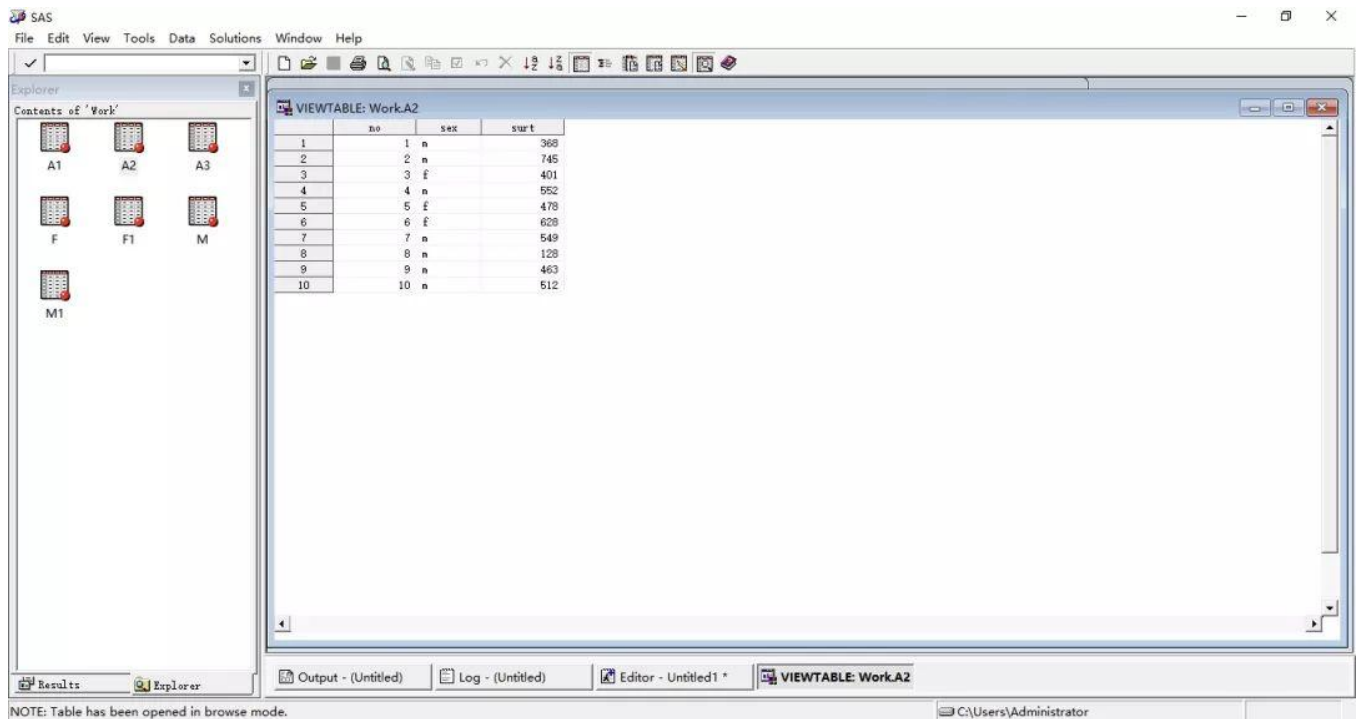
数据集M如下：(只包含男性信息)

	no	sex	age	blood	surt
1	1	n	41	a	368
2	2	n	26	b	745
3	4	n	47	ab	552
4	7	n	28	o	549
5	8	n	31	b	128
6	9	n	43	ab	463
7	10	n	29	a	512

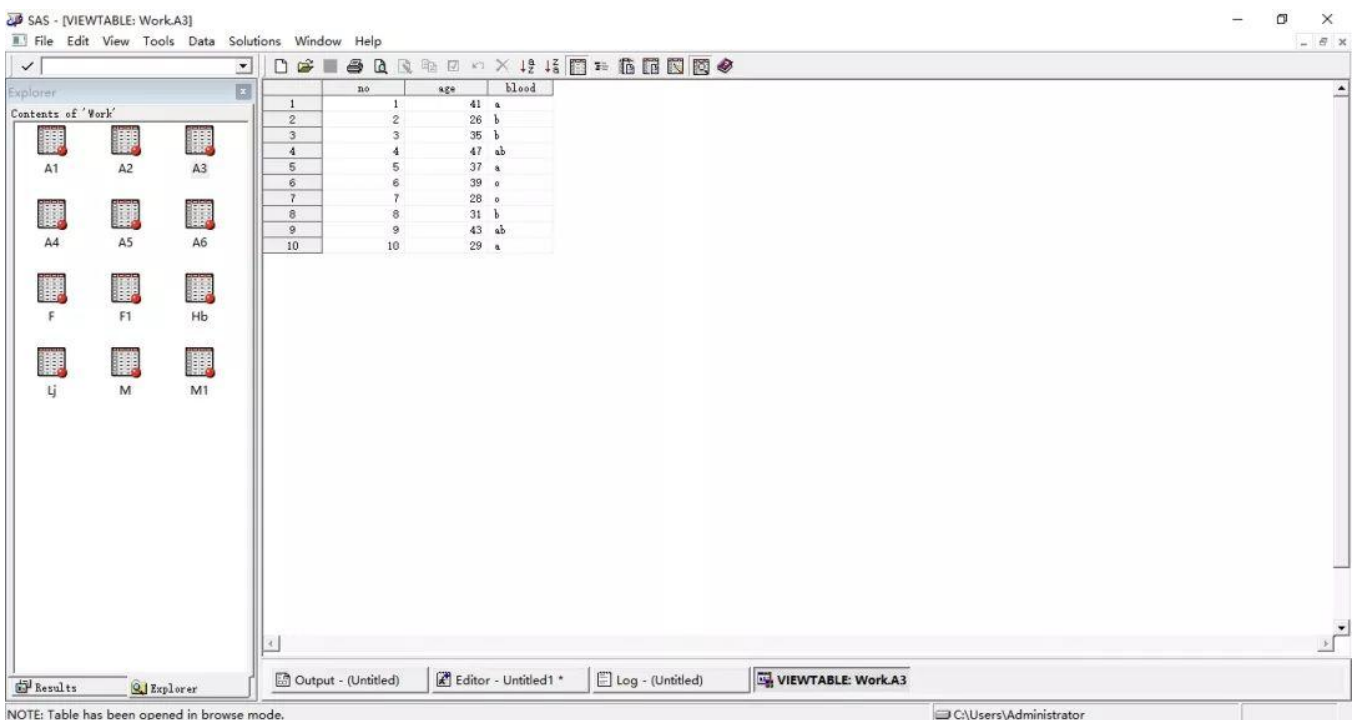
数据集F如下：(只包含女性信息)

	no	sex	age	blood	surt
1	3	f	35	b	401
2	5	f	37	a	478
3	6	f	39	o	620

数据集A2如下：(使用drop语句去除A1数据集中的 age blood变量)



数据集A3如下：(使用keep语句保留A1数据集中的 age blood变量)



A4、A5、A6如下：(产生新变量)

SAS - [VIEWTABLE: Work.A4]

File Edit View Tools Data Solutions Window Help

Explorer

Contents of 'Work'

A1 A2 A3

A4 A5 A6

F F1 M

M1

	no	sex	age	blood	surt	surt1
1	1	n	41	a	368	1.0075290897
2	2	n	26	b	745	2.0396988364
3	3	f	35	b	401	1.0978781656
4	4	n	47	ab	552	1.5112936345
5	5	f	37	a	478	1.3086926762
6	6	f	39	o	628	1.7193702943
7	7	n	28	o	549	1.5030800821
8	8	n	31	b	128	0.3504449008
9	9	n	43	ab	463	1.2676249144
10	10	n	29	a	512	1.401779603

Results Explorer

Output - (Untitled) Log - (Untitled) Editor - Untitled1 * VIEWTABLE: Work.A4

C:\Users\Administrator 音频服务未运行。

/*数据集的排序:按照年龄降序排序(若按升序排列变量名前的ascending可以省略不写)*/

```
proc sort data=a1;by decending age;run;
```

Proc sort之后的A1如下：

SAS - [VIEWTABLE: Work.Lj]

File Edit View Tools Data Solutions Window Help

Explorer

Contents of 'Work'

A1 A2 A3

A4 A5 A6

F F1 Lj

M M1

	no	sex	age	blood	surt
1	1	n	41	a	368
2	2	n	26	b	745
3	4	n	47	ab	552
4	7	n	28	o	549
5	8	n	31	b	128
6	9	n	43	ab	463
7	10	n	29	a	512
8	3	f	35	b	401
9	5	f	37	a	478
10	6	f	39	o	628

Results Explorer

Output - (Untitled) Log - (Untitled) Editor - Untitled1 * VIEWTABLE: Work.Lj

C:\Users\Administrator 2019年2月10日 星期日

NOTE: Table has been opened in browse mode.

/*数据集的连接:m、f数据集众向连接成lj数据集*/

data lj;set m f;run;

lj数据集如下：

	no	sex	age	blood	surt
1	1	m	41	a	368
2	2	n	26	b	745
3	4	n	47	ab	552
4	7	n	28	o	549
5	8	n	31	b	128
6	9	n	43	ab	463
7	10	n	29	a	512
8	3	f	35	b	401
9	5	f	37	a	478
10	6	f	39	o	628

/*数据集的合并:a1、 b1横向合并成hb数据集*/

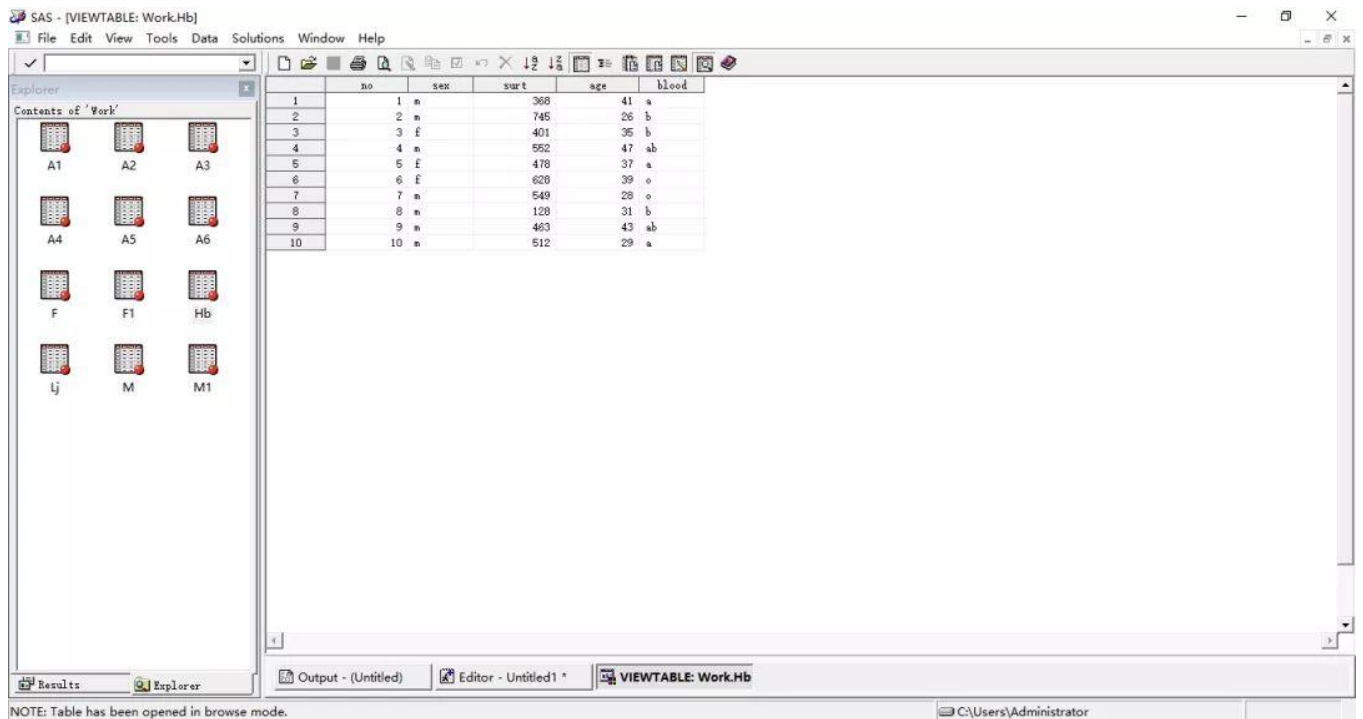
/*需要对原数据集的关联词排序，即对by后的变量进行排序*/

```
proc sort data=a1;by no;run;
```

```
proc sort data=b1;by no;run;
```

```
data hb;merge a1 b1;by no;run;
```

数据集HB如下：

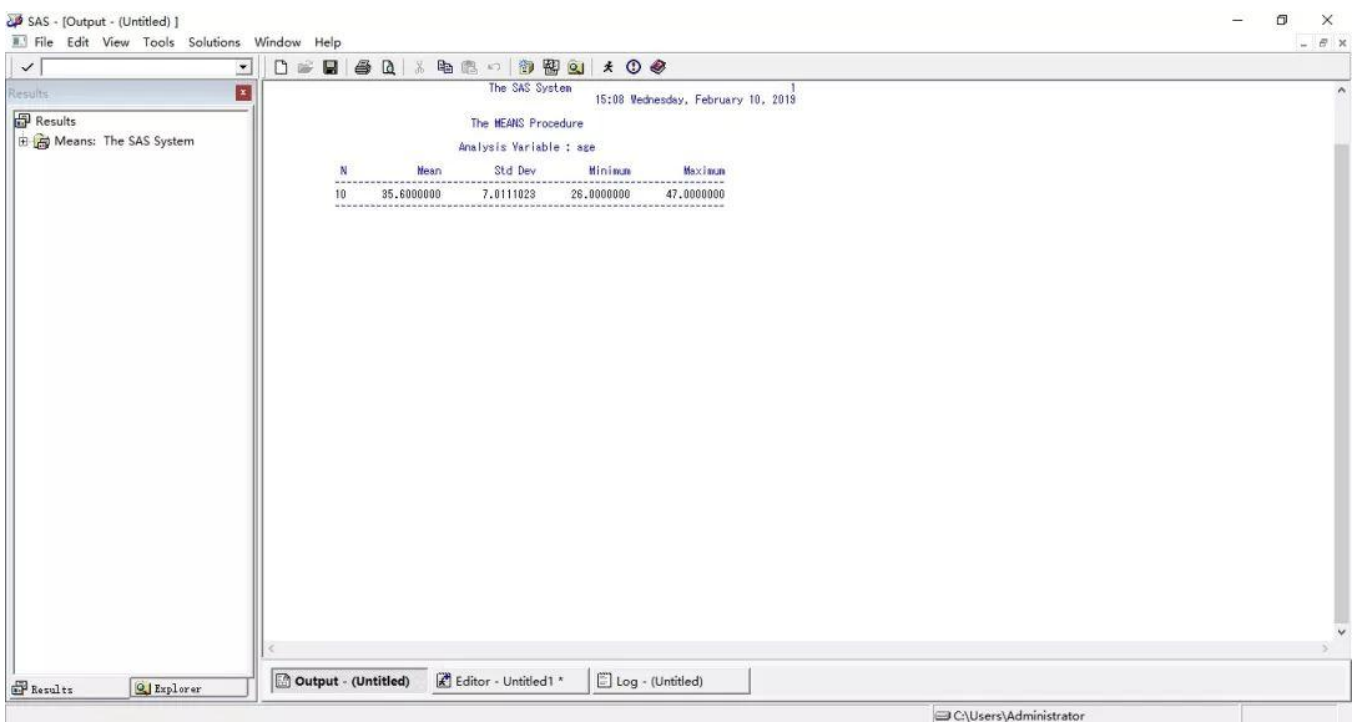


七、描述性统计分析

数值型变量的均值(means 过程)：

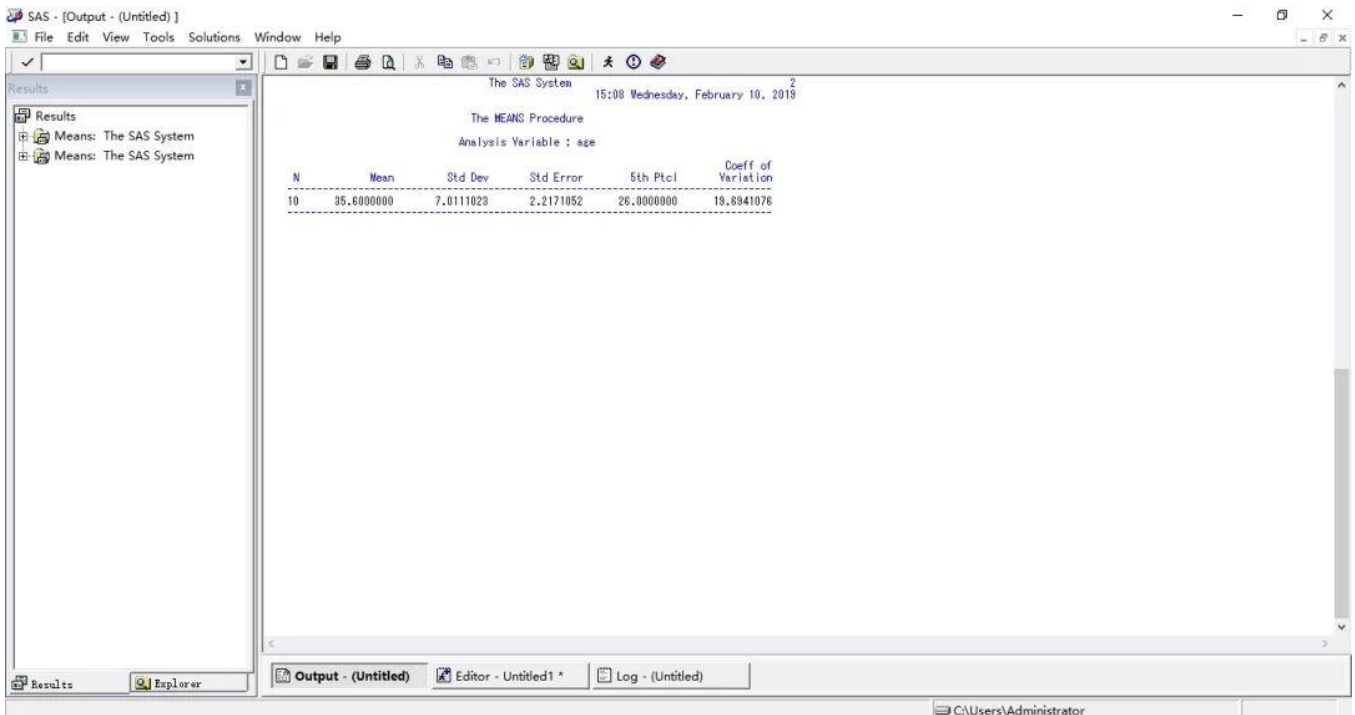
程序：proc means data=a1;var age;run;

结果如下：



可以限制means过程步的输出结果，如下需要输出age的n mean std stderr p5 cv。

程序：proc means data=a1 n mean std stderr p5 cv;var age;run;



The screenshot shows the SAS Output window with the following content:

The SAS System 15:08 Wednesday, February 10, 2015

The MEANS Procedure

Analysis Variable : age

N	Mean	Std Dev	Std Error	5th Pctl	Coeff of Variation
10	95.6000000	7.0111023	2.2171052	26.0000000	19.6941076

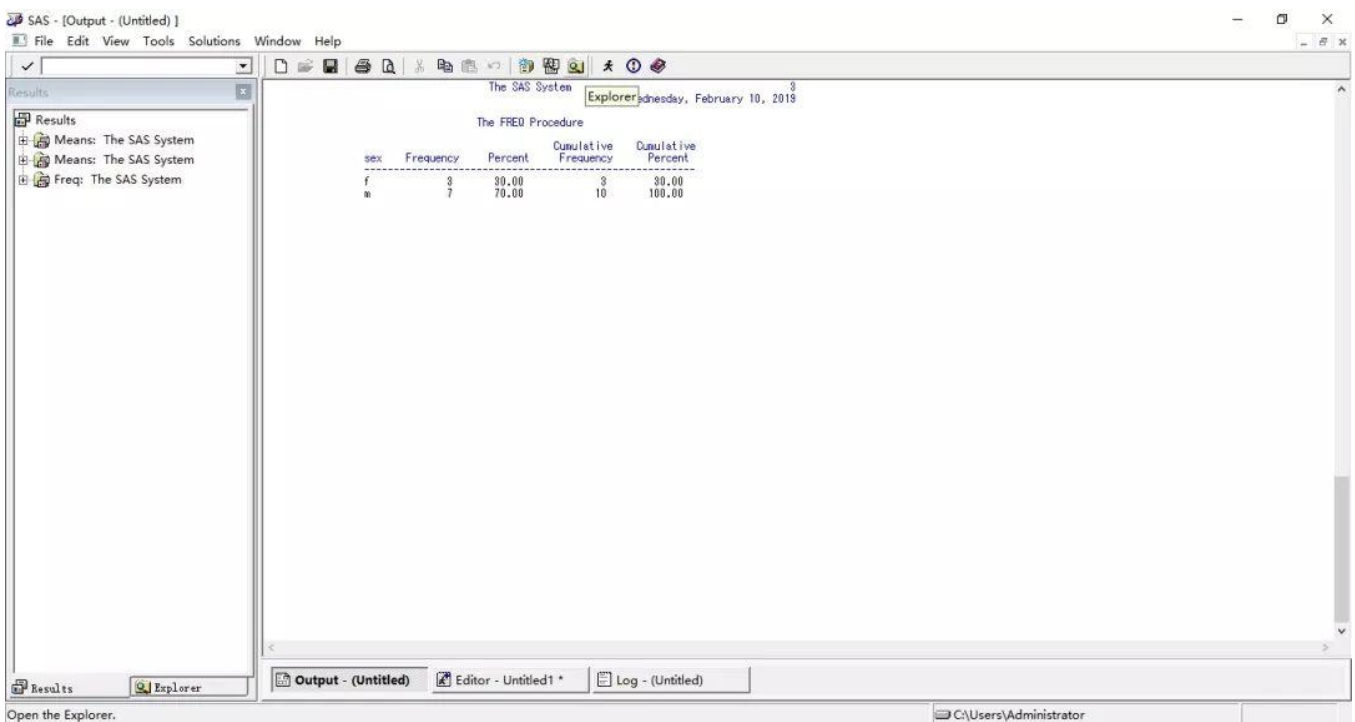
Results Explorer

Output - (Untitled) Editor - Untitled1 * Log - (Untitled)

C:\Users\Administrator

分类变量描述频数：

程序：proc freq data=a1;tables sex;run;



The screenshot shows the SAS Output window with the following content:

The SAS System Explorer Wednesday, February 10, 2015

The FREQ Procedure

sex	Frequency	Percent	Cumulative Frequency	Cumulative Percent
f	9	90.00	9	90.00
m	1	10.00	10	100.00

Results Explorer

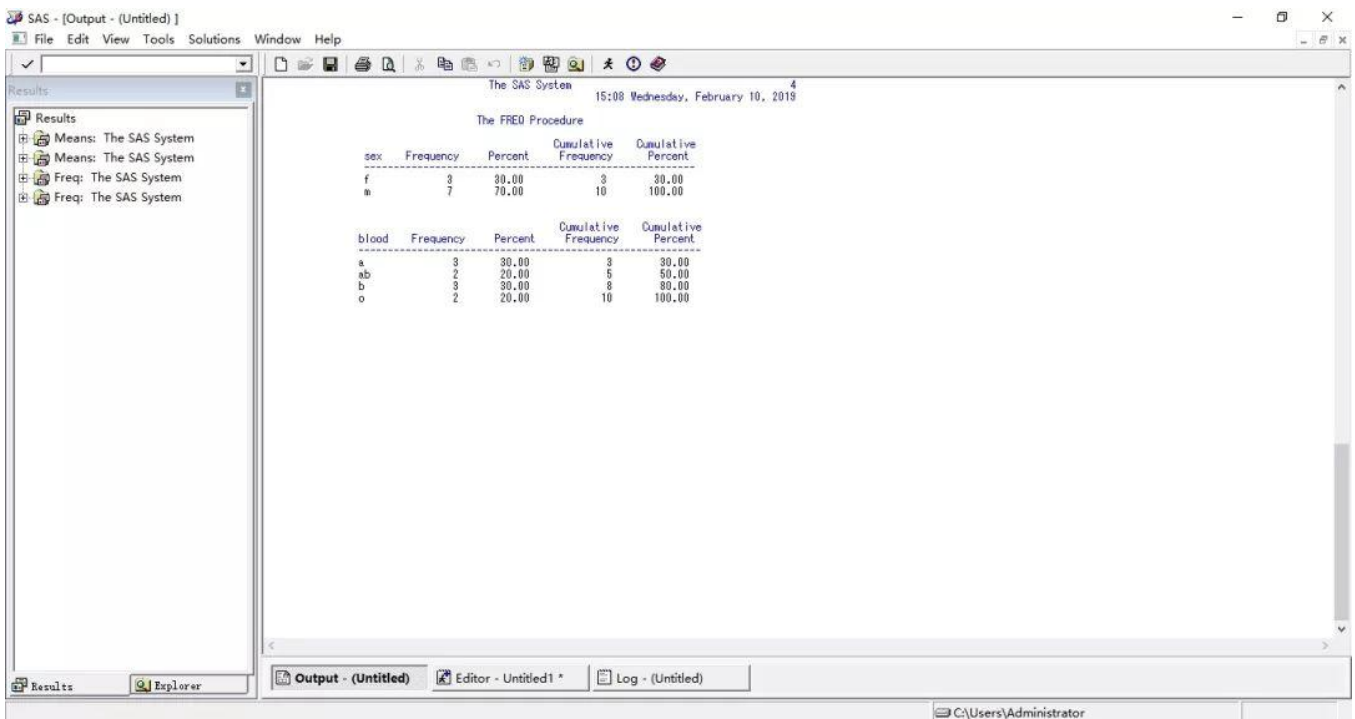
Output - (Untitled) Editor - Untitled1 * Log - (Untitled)

C:\Users\Administrator

Open the Explorer.

可以同时看多个变量的频数，如下同时看性别和血型的频数。

程序：proc freq data=a1;tables sex blood;run;



今天就到这里了，期待下期统计推断的具体案例实际操作，不见不散哟。

参考资料：

《SAS统计软件应用》，贺佳主编，人民卫生出版社，2014年4月

更多 统计方法 请访问 <https://www.iikx.com/news/statistics/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](http://www.iikx.com)转发