

苏州大学数学科学学院

统计计算与 SAS 软件包实验报告

姓名： 熊雄 学号： 1907402030

实验 3 SAS 基础

实验目的：熟悉 SAS 软件

实验内容：

1. 给定三种类型的数据文件分别为：data1.txt, data2.csv , data3.xls
2. 按文件exam12.xls的要求完成实验
3. set 语句后选项的意义（自己给出相应的程序）

结果与分析：

一、 将 data1.txt 转化为 SAS 数据集

输入以下代码：

```
filename data1 'E:\sas class\实验三\data1.txt';
data ex1;
    infile data1 firstobs=2 dlm='09'x ; /*从第二行开始读起，分隔符为Tab*/
    length name $15.;
    input
        name$
        sex$
        age;
run;
proc print data = ex1;
run;
```

提交后输出如下：

SAS 系统

Obs	name	sex	age
1	阿尔弗雷德	男	14
2	爱丽丝	女	13
3	芭芭拉	女	13
4	菲利普	男	16
5	亨利	男	14
6	简	女	12
7	杰弗瑞	男	13

二、 将 data2.csv 转化为 SAS 数据集

输入以下代码：

```
filename data2 'E:\sas class\实验三\data2.csv';
data ex2;
    infile data2 firstobs=2 dlm =','; /*从第二行开始读起, 分隔符为,* /
    length name $15.;
    input
        name$
        height
        weight;
run;
proc print data = ex2;
run;
```

提交后输出如下：

SAS 系统

Obs	name	height	weight
1	阿尔弗雷德	69.0	112.5
2	爱丽丝	56.5	84.0
3	芭芭拉	65.3	98.0
4	菲利普	72.0	150.0
5	亨利	63.5	102.5
6	简	59.8	84.5
7	杰弗瑞	62.5	84.0

三、 将 data3.xls 转化为 SAS 数据集

输入以下代码：

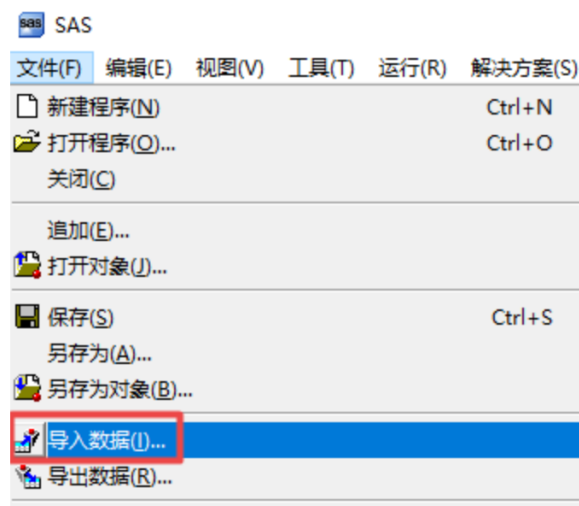
```
proc import datafile = 'E:\sas class\实验三\data3.xls' ;  
    DBMS=xls  
    OUT = ex3 replace;  
    GETNAMES = YES ; /*第一行作为数据集的列变量名*/  
    MIXED = YES; /*读取字符和数值混合的数据表时，将所有数据转化为字符*/  
  
run;  
proc print data = ex3;  
run;
```

提交后得到的结果如下：

SAS 系统

Obs	Name	gender	Age	Height	Weight
1	XI	7	11	57.5	85
2	©	7	15	66.5	112
3	女y	s	15	62.5	112.5
4	~	7	12	59	99.5
5	y漂	7	12	57.3	83
6	1	s	14	64.3	90

可以发现,excel 里面的中文在这里导入会变成乱码,在网上查了用 filename 然后 encoding 的方法,但好像该方法只能使用在 proc import+dml 引擎上,用在.xls 的直接 SAS 重启。故在此更换使用菜单栏里的导入方式:



由此可以得到临时数据集 ex3:

	Name	gender	Age	Height	Weight
1	托马斯	男	11	57.5	85
2	威廉	男	15	66.5	112
3	雅妮特	女	15	62.5	112.5
4	约翰	男	12	59	99.5
5	詹姆斯	男	12	57.3	83
6	茱迪	女	14	64.3	90

四、 用 DATA 步编程实现上述数据集的合并

得到三个临时数据集后，先利用 merge 语句合并数据集 Ex1 与 Ex2，再利用 keep 语句，继续合并 Ex3。输入以下代码：

```
data Ex1AndEx2;  
merge Ex1 Ex2; /*合并ex1与ex2*/  
run;  
data result;  
length name $15.;  
set Ex1AndEx2 Ex3(rename = (gender = sex)); /*重新命名ex3中的变量gender*/  
run;  
proc print;  
run;
```

得到数据集 result:

	Name	gender	Age	Height	Weight
1	阿尔弗	男	14	69	112.5
2	爱丽丝	女	13	56.5	84
3	芭芭拉	女	13	65.3	98
4	菲利普	男	16	72	150
5	亨利	男	14	63.5	102.5
6	简	女	12	59.8	84.5
7	杰弗瑞	男	13	62.5	84
8	托马斯	男	11	57.5	85
9	威廉	男	15	66.5	112
10	雅妮特	女	15	62.5	112.5
11	约翰	男	12	59	99.5
12	詹姆斯	男	12	57.3	83
13	茱迪	女	14	64.3	90

【注：该数据集变量 Name 尝试了很多方法也无法修改变量长度，目前也没找到相关的资料】

五、 从 result 中挑选第 1, 5, 8 观测

利用 set 语句，紧接着输入以下代码：

```
data result1;  
do n=1,5,8;  
set result point=n;  
output;  
end;  
stop;  
run;
```

挑选后的数据集 result1 如下：

SAS 系统

Obs	name	sex	age	height	weight
1	阿尔弗	男	14	69.0	112.5
2	亨利	男	14	63.5	102.5
3	托马斯	男	11	57.5	85.0

六、 输出 result 中最后一条观测

紧接着输入以下代码：

```
data ResultLast;  
set result nobs=last point=last;  
output;  
stop;  
run;
```

得到最后一条观测：

SAS 系统

Obs	name	sex	age	height	weight
1	茱迪	女	14	64.3	90

七、统计 result 中“男”性人数

输入以下代码，即新建数据集 man，装载着 result 中 sex 为“男”的观测：

```
data man;  
set result;  
if sex='男' then output man;  
run;  
proc print;  
run;
```

输出得到：

SAS 系统

Obs	name	sex	age	height	weight
1	阿尔弗	男	14	69.0	112.5
2	菲利普	男	16	72.0	150.0
3	亨利	男	14	63.5	102.5
4	杰弗瑞	男	13	62.5	84.0
5	托马斯	男	11	57.5	85.0
6	威廉	男	15	66.5	112.0
7	约翰	男	12	59.0	99.5
8	詹姆斯	男	12	57.3	83.0