06. 使用SAS函数

**（一）创建并定义新变量**

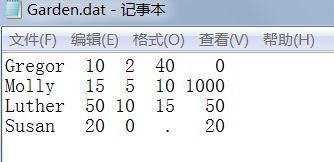
可以在数据步中，利用赋值语句创建或重新定义新变量（列），语法：

变量名=表达式;

注：（1）表达式可以是数值或字符常量、数学表达式、函数等；

（2）加减乘除幂：+ - \* / \*\*，也可以使用括号。

**例1** 下面数据（C:\MyRawData\Garden.dat）是番茄、南瓜、豌豆、葡萄的重量：



代码：

**data** homegarden;

infile 'c:\MyRawData\Garden.dat';

input Name $ **1**-**7** Tomato Zucchini Peas Grapes;

Zone = **14**;

Type = 'home';

Zucchini = Zucchini \* **10**;

Total = Tomato + Zucchini + Peas + Grapes;

PerTom = (Tomato / Total) \* **100**;

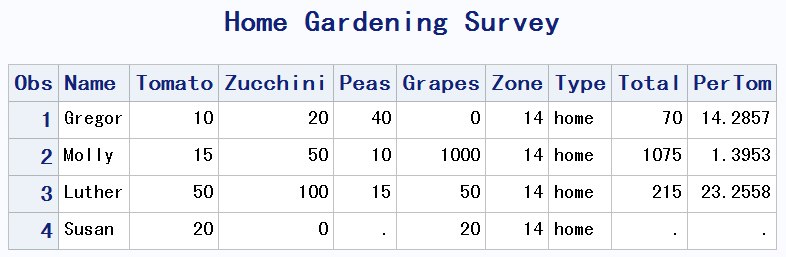
**run**;

**proc** **print** data = homegarden;

title 'Home Gardening Survey';

**run**;

运行结果：



程序说明：

（1）5-8行使用了5个赋值语句，第6行改写旧变量，运行结果将输出所有变量；

（2）若观测值中有缺失值，相应的新变量也出现缺失值。

**（二）使用SAS函数**

**一、SAS函数概述**

SAS自带数百个函数，主要包括以下函数类：

字符/字符串函数

日期/时间函数

距离/州/邮编代码函数

数学函数

概率/描述统计/随机数函数

金融函数

宏函数

变量信息函数

函数基本调用形式（可以嵌套调用）：

函数名(参数1,参数2,…);

**例2**南瓜雕刻比赛的数据（C:\MyRawData\pumpkin.dat）包含了参赛者的名字、年龄、雕刻的南瓜类型、报名日期、五位裁判给出的分数：



代码：

**data** contest;

infile 'c:\MyRawData\Pumpkin.dat';

input Name $16. Age **3.** +**1** Type $1. +**1** Date MMDDYY10.

(Scr1 Scr2 Scr3 Scr4 Scr5) (**4.1**);

AvgScore = mean(Scr1, Scr2, Scr3, Scr4, Scr5);

DayEntered = day(Date);

Type = upcase(Type); /\* 转化为大写 \*/

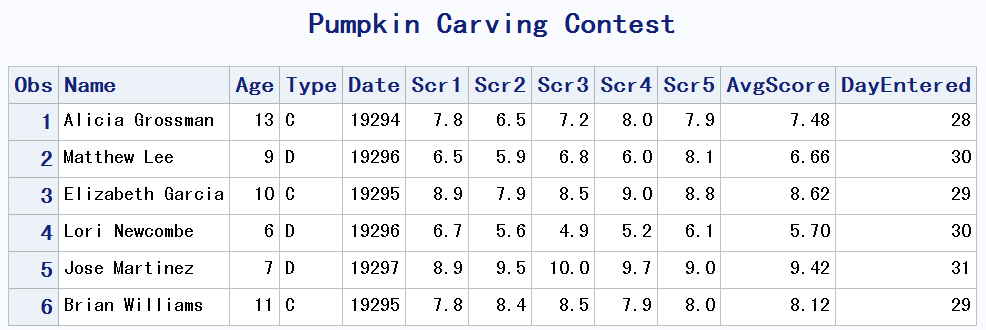
**run**;

**proc** **print** data = contest;

title 'Pumpkin Carving Contest';

**run**;

运行结果：



程序说明：

1. 均值函数mean( )返回非缺失值参数的平均值；
2. day( )函数返回日期值距离1961年1月1日的天数。

**二、常用的数学/统计/金融函数**

|  |  |
| --- | --- |
| int(x) | 取整（舍去小数） |
| round(x) | 取整（四舍五入） |
| abs(x) | 取绝对值 |
| sqrt(x) | 求平方根 |
| exp(x) | e的x次幂 |
| log(x) | 取自然对数 |
| log10(x) | 取以10为底的对数 |
| sin(x) | 求正弦值 |
| max(x1,x2,…) | 返回最大值（忽略缺失值） |
| min(x1,x2,…) | 返回最小值（忽略缺失值） |
| mean(x1,x2,…) | 返回平均值（忽略缺失值） |
| sum(x1,x2,…) | 返回求和值（忽略缺失值） |
| n(x1,x2,…) | 返回非缺失值的个数 |
| nmiss(x1,x2,…) | 返回缺失值的个数 |
| LAGn(x) | 求给定变量x滞后为n的值 |
| DIFn(x) | 求给定变量x的第n阶差 |
| std(x) | 求标准差 |
| probnorm(x) | 求标准正态分布函数 |
| compound( ) | 计算复利 |
| irr( ) | 计算内部盈利率 |
| npv( ) | 计算净现值 |
| saving( ) | 计算定期储蓄的本金和利息 |

**三、常用的日期/时间函数**

|  |  |
| --- | --- |
| datejul(day) | 将日期数值day转化为Julian日期值 |
| today() | 返回今天的日期数值 |
| mdy(mm,dd,yy) | 将月mm、日dd、年yy转化为日期数值 |
| day(date) | 返回日期数值的日 |
| month(date) | 返回日期数值的月（1-12） |
| year(date) | 返回日期数值的年 |
| qtr(date) | 返回日期数值的季（1-4） |
| weekday(date) | 返回日期数值是周几（1=星期日） |
| time(t) | 返回时间数值t对应的时间 |
| yrdif(date1,date2, ‘AGE’) | 计算两个日期的“年龄差”  示例：a=MDY(4,13,2000); b=MDY(8,13,2012);  YRDIF(a,b,'AGE'); 返回 12.3342 |

注：（1）日期数值 = 当前日期距离1960年1月1日的天数；

（2）Julian日期：5位或7位，前2或4位是年，后3位为该年的第多少天，例如1960年2月1日是60032或1960032.

**四、常用的字符串函数**

|  |  |
| --- | --- |
| anyalnum(str, n) | 返回str从n位置开始，第一次出现字母或数字的位置 |
| 示例：a='123 E St, #2 '; anyalnum (a); （默认）返回 1； anyalnum (a,10); 返回12 |
| anyalpha(str, n) | 返回str从n位置开始，第一次出现字母的位置 |
| anydigit(str, n) | 返回str从n位置开始，第一次出现数字的位置 |
| anyspace(str, n) | 返回str从n位置开始，第一次出现空格的位置 |
| cat(str1,str2,…) | 拼接字符串str1,str2,…，开头和末尾加上空格 |
| cats(str1,str2,…) | 拼接字符串str1,str2,…，忽略空格 |
| catx(‘sign’,str1,str2,…) | 拼接字符串（忽略空格），字符串间加分隔符sign |
| compress(str, ‘sign’) | 压缩字符串（去掉符号sign，默认去掉空格） |
| index(str, ‘sign’) | 返回字符或子串sign首次在字符串str中出现的位置 |
| left(str) | 左对齐字符串str  示例：a=' my cat'; left(a); 返回 'my cat ' |
| lenth(str) | 返回str的长度（忽略末尾空格，缺失值长度是1） |
| propcase(str) | 将字符串str第一个字母大写，其余小写 |
| substr(str,pos,n) | 从str第pos位置开始提取长度为n或直至结束的子串 |
| translate(str, ‘to’, ‘from’) | 将str中的字符‘from’换成‘to’ |
| tranwrd(str, ‘from’,‘to’) | 将str中的字符串‘from’替换为‘to’  示例：a='Main Street'; tranwrd(a,'Street','St');  返回 'Main St' |
| trim(str) | 去掉str末尾的空格 |
| upcase(str) | 将str所有字母转化为大写 |