

一、数学函数

ABS(x) 求 x 的绝对值。

MAX(x1, x2, ..., xn) 求所有自变量中的最大一个。

MIN(x1, x2, ..., xn) 求所有自变量中的最小一个。

MOD(x, y) 求 x 除以 y 的余数。

SQRT(x) 求 x 的平方根。

ROUND(x, eps) 求 x 按照 eps 指定的精度四舍五入后的结果，比如
ROUND(5654.5654, 0.01) 结果为 5654.57, ROUND(5654.5654, 10) 结果为 5650。

CEIL(x) 求大于等于 x 的最小整数。当 x 为整数时就是 x 本身，否则为 x 右边最近的整数。

FLOOR(x) 求小于等于 x 的最大整数。当 x 为整数时就是 x 本身，否则为 x 左边最近的整数。

INT(x) 求 x 扔掉小数部分后的结果。

FUZZ(x) 当 x 与其四舍五入整数值相差小于 1E-12 时取四舍五入。

LOG(x) 求 x 的自然对数。

LOG10(x) 求 x 的常用对数。

EXP(x) 指数函数。

SIN(x), COS(x), TAN(x) 求 x 的正弦、余弦、正切函数。

ARSIN(y) 计算函数 $y=\sin(x)$ 在区间的反函数, y 取 [-1, 1] 间值。

ARCOS(y) 计算函数 $y=\cos(x)$ 在的反函数, y 取 [-1, 1] 间值。

ATAN(y) 计算函数 $y=\tan(x)$ 在的反函数, y 取间值。

SINH(x), COSH(x), TANH(x) 双曲正弦、余弦、正切

ERF(x) 误差函数

GAMMA(x) 完全函数

此外还有符号函数 SIGN, 函数一阶导数函数 DIGAMMA, 二阶导数函数 TRIGAMMA , 误差函数余函数 ERFC, 函数自然对数 LGAMMA, ORDINAL 函数, AIRY 函数, DAIRY 函数, Bessel 函数 JBESSEL, 修正的 Bessel 函数 IBESSEL, 等等。

二、数组函数

数组函数计算数组的维数、上下界, 有利于写出可移植的程序。数组函数包括:

DIM(x) 求数组 x 第一维的元素的个数(注意当下界为 1 时元素个数与上界相同, 否则元素个数不一定与上界相同)。

DIM k(x) 求数组 x 第 k 维的元素的个数。

LBOUND(x) 求数组 x 第一维的下界。

HBOUND(x) 求数组 x 第一维的上界。

LBOUND k(x) 求数组 x 第 k 维的下界。

HBOUND k(x) 求数组 x 第 k 维的上界。

三、字符函数

较重要的字符函数有:

TRIM(s) 返回去掉字符串 s 的尾随空格的结果。

UPCASE(s) 把字符串 s 中所有小写字母转换为大写字母后的结果。

LOWCASE(s) 把字符串 s 中所有大写字母转换为小写字母后的结果。

INDEX(s, s1) 查找 s1 在 s 中出现的位置。找不到时返回 0。

RANK(s) 字符 s 的 ASCII 码值。

BYTE(n) 第 n 个 ASCII 码值的对应字符。

REPEAT(s, n) 字符表达式 s 重复 n 次。

SUBSTR(s, p, n) 从字符串 s 中的第 p 个字符开始抽取 n 个字符长的子串

TRANWRD(s, s1, s2) 从字符串 s 中把所有字符串 s1 替换成字符串 s2 后的结果。

其它字符函数还有 COLLATE, COMPRESS, INDEXC, LEFT, LENGTH, REVERSE, RIGHT, SCAN, TRANSLATE, VERIFY, COMPBL, DEQUOTE, INDEXW, QUOTE, SOUNDEX, TRIMN, INDEXW。

四、日期和时间函数

常用日期和时间函数有：

MDY(m, d, yr) 生成 yr 年 m 月 d 日的 SAS 日期值

YEAR(date) 由 SAS 日期值 date 得到年

MONTH(date) 由 SAS 日期值 date 得到月

DAY(date) 由 SAS 日期值 date 得到日

WEEKDAY(date) 由 SAS 日期值 date 得到星期几

QTR(date) 由 SAS 日期值 date 得到季度值

HMS(h, m, s) 由小时 h、分钟 m、秒 s 生成 SAS 时间值

DHMS(d, h, m, s) 由 SAS 日期值 d、小时 h、分钟 m、秒 s 生成 SAS 日期时间值

DATEPART(dt) 求 SAS 日期时间值 dt 的日期部分

INTNX(interval, from, n) 计算从 from 开始经过 n 个 in 间隔后的 SAS 日期。其中 interval 可以取 'YEAR'、'QTR'、'MONTH'、'WEEK'、'DAY' 等。比如，INTNX('MONTH', '16Dec1997' d, 3) 结果为 1998 年 3 月 1 日。注意它总是返回一个周期的开始值。

INTCK(interval, from, to) 计算从日期 from 到日期 to 中间经过的 interval 间隔的个数，其中 interval 取 'MONTH' 等。比如，INTCK('YEAR', '31Dec1996' d, '1Jan1998' d) 计算 1996 年 12 月 31 日到 1998 年 1 月 1 日经过的年间隔的个数，结果得 2，尽管这两个日期之间实际只隔 1 年。

其它日期和时间函数还有 DATE、TODAY、DATETIME、DATEJUL、JULDATE、 HOUR、MINUTE、SECOND 、TIME、TIMEPART 等。详见《SAS 系统—Base SAS 软件使用手册》、《SAS 系统—SAS/ETS 软件使用手册》。

五、分布密度函数、分布函数

作为一个统计计算语言，SAS 提供了多种概率分布的有关函数。分布密度、概率、累积分布函数等可以通过几种统一的格式调用，格式为

分布函数值 = CDF(' 分布', x <, 参数表>);

密度值 = PDF(' 分布', x <, 参数表>);

概率值 = PMF(' 分布', x <, 参数表>);

对数密度值 = LOGPDF(' 分布', x <, 参数表>);

对数概率值 = LOGPMF(' 分布', x <, 参数表>);

CDF 计算由' 分布' 指定的分布的分布函数， PDF 计算分布密度函数值，PMF 计算离散分布的分布概率，LOGPDF 为 PDF 的自然对数，LOGPMF 为 PMF 的自然对数。函数在自变量 x 处计算，<, 参数表>表示可选的参数表。

分布类型取值可以为：BERNOULLI, BETA, BINOMIAL, CAUCHY, CHISQUARED, EXPONENTIAL, F, GAMMA, GEOMETRIC, HYPERGEOMETRIC, LAPLACE, LOGISTIC, LOGNORMAL, NEGBINOMIAL, NORMAL 或 GAUSSIAN, PARETO, POISSON, T, UNIFORM, WALD 或 IGAUSS, and WEIBULL。可以只写前四个字母。

例如，PDF(' NORMAL', 1.96) 计算标准正态分布在 1.96 处的密度值 (0.05844)，CDF(' NORMAL', 1.96) 计算标准正态分布在 1.96 处的分布函数值 (0.975)。PMF 对连续型分布即 PDF。

除了用上述统一的格式调用外，SAS 还单独提供了常用的分布的密度、分布函数。

PROBNORM(x) 标准正态分布函数

PROBT(x, df<, nc>) 自由度为 df 的 t 分布函数。可选参数 nc 为非中心参数。

PROBCHI(x, df<, nc>) 自由度为 df 的卡方分布函数。可选参数 nc 为非中心参数。

PROBF(x, ndf, ddf<, nc>) F(ndf, ddf) 分布的分布函数。可选参数 nc 为非中心参数。

PROBBNML(p, n, m) 设随机变量 Y 服从二项分布 B(n, p)，此函数计算 P(Y = m)。

POISSON((lambda, n) 参数为 lambda 的 Poisson 分布 Y = n 的概率。

PROBNEGB(p, n, m) 参数为 (n, p) 的负二项分布 Y = m 的概率。

PROBHYP(N, K, n, x<, r>) 超几何分布的分布函数。设 N 个产品中有 K 个不合格品，抽取 n 个样品，其中不合格品数小于等于 x 的概率为此函数值。可选参数 r 是不匀率，缺省为 1，r 代表抽到不合格品的概率是抽到合格品概率的多少倍。

PROBBETA(x, a, b) 参数为 (a, b) 的 Beta 分布的分布函数。

PROBGAM(x, a) 参数为 a 的 Gamma 分布的分布函数。

PROBMC 计算多组均值的多重比较检验的概率值和临界值。

PROBBNRM(x, y, r) 标准二元正态分布的分布函数，r 为相关系数。

六、分位数函数

分位数函数是概率分布函数的反函数。其自变量在 0 到 1 之间取值。分位数函数计算的是分布的左侧分位数。SAS 提供了六种常见连续型分布的分位数函数。

PROBIT(p) 标准正态分布左侧 p 分位数。结果在 -5 到 5 之间。

TINV(p, df <, nc>) 自由度为 df 的 t 分布的左侧 p 分位数。可选参数 nc 为非中心参数。

CINV(p, df<, nc>) 自由度为 df 的卡方分布的左侧 p 分位数。可选参数 nc 为非中心参数。

FINV(p, ndf, ddf, <nc>) F(ndf, ddf) 分布的左侧 p 分位数。可选参数 nc 为非中心参数。

GAMINV(p, a) 参数为 a 的伽马分布的左侧 p 分位数。

BETAINV(p, a, b) 参数为 (a, b) 的贝塔分布的左侧 p 分位数。

七、随机数函数

SAS 可以用来进行随机模拟。它提供了常见分布的伪随机数生成函数。

1. 均匀分布随机数

有两个均匀分布随机数函数：UNIFORM(seed)，seed 必须是常数，为 0，或 5 位、6 位、7 位的奇数。RANUNI(seed)，seed 为小于 $2^{31}-1$ 的任意常数。在同一个数据步中对同一个随机数函数的多次调用将得到不同的结果，但不同数据步中从同一种子出发将得到相同的随机数序列。随机数种子如果取 0 或者负数则种子采用系统日期时间。

2. 正态分布随机数

有两种，NORMAL(seed)，seed 为 0，或 5 位、6 位、7 位的奇数。RANNOR(seed)，seed 为任意数值常数。

3. 指数分布随机数

RANEXP(seed)，seed 为任意数值，产生参数为 1 的指数分布的随机数。参数为 lambda 的指数分布可以用 RANEXP(seed)/lambda 得到。

另外若 $Y = \alpha - \beta * \text{LOG}(\text{RANEXP}(\text{seed}))$ ，则 Y 为位置参数为 alpha，尺度参数为 beta 的极值分布。若 $Y = \text{FLOOR}(-\text{RANEXP}(\text{seed}) / \text{LOG}(p))$ ，那么 Y 是具有参数 p 的几何分布变量。

4. 伽马分布随机数

RANGAM(seed, alpha), seed 为任意数值常数, $\alpha > 0$, 得到参数为 alpha 的伽马分布。设 $X = \text{RANGAM}(\text{seed}, \alpha)$, 则 $Y = \beta * X$ 是形状参数为 alpha, 尺度参数为 beta 的 GAMMA 分布随机数。如果 alpha 是整数, 则 $Y = 2 * X$ 是自由度为 $2 * \alpha$ 的卡方分布随机数。

如果 alpha 是正整数, 则 $Y = \beta * X$ 是 Erlang 分布随机数, 为 alpha 个独立的均值为 beta 的指数分布变量的和。

如果 $Y_1 = \text{RANGAM}(\text{seed}, \alpha)$, $Y_2 = \text{RANGAM}(\text{seed}, \beta)$, 在 $Y = Y_1 / (Y_1 + Y_2)$ 是参数为 (alpha, beta) 的贝塔分布随机数。

5. 三角分布随机数

RANTRI(seed, h), seed 为任意数值常数, $0 < h < 1$ 。此分布在 0 到 1 取值, 密度在 0 到 h 之间为 $2x/h$, 在 h 到 1 之间为 $2(1-x)/(1-h)$ 。

6. 柯西分布随机数

RANCAU(seed), seed 为任意数值常数。产生位置参数为 0, 尺度参数为 1 的标准柯西分布随机数。 $Y = \alpha + \beta * \text{RANCAU}(\text{seed})$ 为位置参数为 alpha, 尺度参数为 beta 的一般柯西分布随机数。

7. 二项分布随机数

RANBIN(seed, n, p) 产生参数为 (n, p) 的二项分布随机数, seed 为任意数值。

8. 泊松分布随机数

RANPOI(seed, lambda) 产生参数为 $\lambda > 0$ 的泊松分布随机数, seed 为任意数值。

9. 一般离散分布随机数

RANTBL(seed, p_1, \dots, p_n) 生成取 1, 2, \dots , n 的概率分别为 p_1, \dots, p_n 的离散分布随机数。

八、样本统计函数

样本统计函数把输入的自变量作为一组样本，计算样本统计量。其调用格式为“函数名（自变量 1，自变量 2，…，自变量 n）”或者“函数名（OF 变量名列表）”。比如 SUM 是求和函数，如果要求 x1, x2, x3 的和，可以用 SUM(x1, x2, x3)，也可以用 SUM(OF x1-x3)。这些样本统计函数只对自变量中的非缺失值进行计算，比如求平均时把缺失值不计入内。

各样本统计函数为：

MEAN 均值

MAX 最大值

MIN 最小值

N 非缺失数据的个数

NMISS 缺失数值的个数。

SUM 求和

VAR 方差

STD 标准差

STDERR 均值估计的标准误差，用 STD/SQRT(N) 计算。

CV 变异系数

RANGE 极差

CSS 离差平方和

USS 平方和

SKEWNESS 偏度

KURTOSIS 峰度

注意：数据集的存储一般是每行为一个个体的观测值，每列是个体的一个属性(变量)，所以统计一般应该对列进行，而不是象这里对行进行，把各变量作为一个样本的各个观测处理。这里提供的函数主要用于进行一些自编程序的计算。

九、SAS/IML 矩阵功能简介

SAS/IML 是 SAS 提供的一个可以进行矩阵运算编程的工具，详细使用请参见有关资料或系统帮助（Help | Extended Help | SAS System Help: Main Menu | Help for SAS Products | SAS/IML）。它可以用来进行交互的矩阵运算，也可以编好一个程序再一起运行。程序可以使用分支、循环、模块化子程序等控制结构。数据步中的函数大都能在 SAS/IML 中使用，SAS/IML 也提供了一些特有的函数。SAS/IML 的一个方便之处是它可以直接读取 SAS 的数据集并把结果写成 SAS 数据集，它也有存取外部文件的功能。

要交互运行 SAS/IML，只要在程序窗口输入

```
proc iml;

    reset print;
```

提交此程序，就可以进入交互的 SAS/IML 运算状态。退出用 **QUIT** 语句。SAS/IML 中可以使用标量、行向量、列变量和矩阵，可以使用字符型数据。变量取名规则遵循 SAS 语言的统一规定，变量可以存储标量、向量和矩阵。

赋值用等号。例如：

```
sc = 15.25;vh = {1 2};

vh1=5:9;

vv = {3, 4};

mat1 = {1 2 3,
        4 5 6};

mat2 = {"Li" "Ming",
        "Zhang" "Chong"};
```

我们注意赋值语句是一个 SAS 语句，它以分号结尾。上面定义了标量 sc，行向量 vh 和 vh1，列向量 vv，两行两列的矩阵 mat1，字符型矩阵 mat2。由上可见，写矩阵常量时，行的元素之间以空格分隔，行之间以逗号分割。可以用“开始：结尾”的写法生成一个等差数列行向量。

矩阵之间可以用<、=等符号进行元素两两间的比较。要得到一个标量的结果，可以用 **ALL()** 函数表示自变量的各元素均为真（非零），用 **ANY()** 函数表示自变量的元素中至少一个为真。可以用&、|、^ 连接两个逻辑型矩阵（元素间的与、或、非）。

用||表示矩阵左右连接，用//表示矩阵上下连接。x`表示 x 的转置。

SAS/IML 中可以进行通常的矩阵加减乘（+、-、*）运算，也可以进行对应元素之间的乘除（#、/）运算。矩阵加减必须两个矩阵大小完全相等，或者其中一个为标量。矩阵乘法要求第一矩阵的列数等于第二矩阵的行数，或其中一个为标量。矩阵元素之间的乘除运算是对应元素进行乘除，当两个矩阵大小完全相同时可以进行运算，其中有一个为标量时可以进行运算，另外，如果其中有一个是行（列）向量而其长度与另一个矩阵的列数（行数）相同也可以进行运算。矩阵逆要用 **INV()** 函数运算。

为了读入一个数据集，先打开数据集，用如

```
USE sasuser.c9501;
```

然后用 **READ ALL VAR{ 变量列表}** 的格式读入数据集的各变量值，例如：

```
READ ALL VAR {name math chinese};
```

```
print name math chinese;
```

这时三个变量都可以作为列向量来使用。