1 **PrincipalComponentAnalysis<-function(dataset, yname = NULL, xname = NULL,**

**scale=TRUE, threshhold = 0.85, plotstr = NULL, compname = NULL, screepname = NULL)**

功能：

主成分分析。通过正交变换将一组可能存在相关性的变量转换为一组线性不相关的变量。适合数值型数据。

输入：

**dataset:**。两列及以上，均为数值型向量。字符串型数字向量(‘1’,’2’,’3’)会被自动转为数值型（1，2，3）; 字符串型非数字向量（“男”，“女”）会被自动转为（1，2），但是强烈建议事先转化字符串型非数字向量。

**yname**：因变量名称，类型为字符串，如果缺省，则仅仅对**xname**指定的变量进行主成分分析；如果不缺省，则在利用主成分对**yname**指定变量进行多元线性回归。

**xname**：自变量名称，类型为字符串向量，如果**yname**为空，则缺省为所有变量，否则需要指定。

**scale:** 是否标准化dataset数据（减去均值，除以标准差）。

**threshhold**：主成分回归累积方差阈值，类型为数值型，一般取值范围为

**[0.8,1]**。如果**yname**不为空，则选择累积方差小于**threshhold**的

主成分参与回归；如果**yname**为空，则不要求该参数。

**plotstr**：图片输出文件夹目录，类型为字符串，缺省不输出。

**compname** ：前两大主成分散点图的输出名称，类型为字符串，缺省不输出。

**screepname** ：碎石图的输出名称，类型为字符串，缺省不输出。

输出：

**PCAResultRowName**：主成分分析结果矩阵的行名，类型为字符串向量。

**PCAResultColName**：主成分分析结果矩阵的列名，类型为字符串向量。

**PCAResult**：主成分分析结果矩阵，包括主成分标准差(SD)，主成分方差(Deviation)， 主成分方差贡献率(ProportionalVar)，主成分方差累积贡献率(CumulativeVar)。

**compname:** 前两大主成分散点图。

**screepname** ：碎石图。

当yname不为空，额外输出主成分回归结果，和LinearRegression结果一致，

但不输出LinearRegression中的图。

四．附录

1. **Formula**

**formulastring** 是指R中用字符串表示的回归公式，如（“y ~ x1 + x2 + x1:x2”）。~左边的是因变量，右边的X1， x2 是自变量，其中x1:x2表示两者的交互效应，作为第三个自变量出现。自变量之间用+连接。