聚类分析

```
/*系统聚类法，method 有 single-最短距离法，complete-最长距离法，median-中间距离法，centroid-重心法，average-类平均法，ward-离差平方和法(Ward 法)*/
proc cluster data=yourdata method=ward outtree=outtree standard;
id region;run;
/*画树形图*/
proc tree data=outtree horizontal out=result n=5;run;
/*快速聚类法，先标准化*/
proc standard data=tmp1.exe3_1 m=0 std=1 out=stout;run;
/*选择凝聚点种子，分几类就几个种子*/
data seed;
set stout;
if _n_=10 then output;if _n_=20 then output;if _n_=30 then output;run;
proc fastclus maxclusters=4 data=stout seed=seed mean=stat out=output;run;
```

判别分析

```
# 距离判别，listerr 是回判，pool=no 表示认为协差阵不等，yes 是相等
proc discrim data=tmp1.exe4_1ts listerr testdata=tmp1.exe4_1ds out=out testout=testout
outstat=outstat pool=no;
class x5;var x1-x4;run;
# 贝叶斯判别，多了 priors
proc discrim data=tmp1.exe4_1ts listerr testdata=tmp1.exe4_1ds out=out testout=testout
outstat=outstat pool=no;
class x5;var x1-x4;priors '1'=0.2 '2'=0.8;run;
# 费歇判别，画前两个典型变量的散点图
proc candisc data=tmp1.exe4_2 out=out;
class x6;var x1-x5;run;
proc plot data=out;
plot can2*can1=x6;run;
```

主成分分析

```
/*主成分分析*/
proc princomp data=yourdata out=out5_1;run;
/*取前三个主成分，并绘制两个主成分的得分图*/
proc princomp data=yourdata n=3 out=out5_1 plot=score(ncomp=2) plot=pattern(ncomp=2);
id industry;run;
/*根据第一主成分降序排序*/
proc sort data=out5_1;by descending prin1;run;
/*打印样品名与第一主成分得分*/
proc print;var state prin1;run;
/*主成分回归*/
proc reg data=yourdata outest=out;
model y=x1-x3/pcomit=1,2;run;quit;
/*打印结果*/
proc print data=out;run;
/*岭回归*/
proc reg data=yourdata outest=out;
model y=x1-x3/ridge=0 to 2 by 0.1;run;
/*向前逐步回归*/
proc reg data=yourdata outest=out;
model y=x1-x3/selection=stepwise slentry=0.1 slstay=0.15;run;
```

```
/*向后逐步回归*/
proc reg data=yourdata outest=out1;
model y=x1-x3/selection=backward slstay=0.15;
run;quit;
proc print data=out;
run;
```

因子分析

```
/*因子分析，选定 5 个因子，r=v 是方差最大，r=q 是四次方最大*/
proc factor data=yourdata method=prin r=v n=5 out=a1 outstat=stat1 reorder;
run;
/*计算因子得分，15 是指标数*/
data a2;
set a1;
f=(5.6327*factor1+2.7072*factor2+2.2692*factor3+1.3137*factor4+1.0431*factor5)/15;
keep f factor1 factor2 factor3 factor4 factor5;
run;
/*按综合因子得分降序排列*/
proc sort data=a2 out=a3;by descending f;run;
/*画散点图*/
proc plot data=a1;
plot factor2*factor1$region='*' href=0 vref=0;run;
```

对应分析

```
/*用于列联表形式的对应分析*/
proc corresp data=yourdata out=a2 rp cp short;var x1-x4;id region;run;
/*用于尚未变成列联表的原始数据的对应分析*/
proc corresp data=yourdata out=a4 rp cp short;tables row, column;run;
```

典型相关分析

```
proc cancorr data=yourdata out=out1 outstat=outstat1 all vdep / vreg;
with y1-y3;var x1-x3;run;
```