因子分析可以看作是主成分分析的一个扩充，因子分析在数学建模中使用的没有主成分分析那么多。关于因子分析和主成分分析的区别可以看一下司守奎老师的“因子分析”那个章节。一开始就有介绍区别。

因子分析

1）、主成分分析法：

例5 研究纽约股票市场上五种股票的周回升率。这里，周回升率＝（本星期五市场收盘价－上星期五市场收盘价）/上星期五市场收盘价。从1975年1月到1976年12月，对这五种股票作了100组独立观测。因为随着一般经济状况的变化，股票有集聚的趋势，因此，不同股票周末回升率是彼此相关的。

相关系数矩阵：



对和，因子分析主成分解见下表。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 变量 | 一个因子 | | 两个因子 | | |
| 因子载荷估计 | 特殊方差 | 因子载荷估计 | | 特殊方差 |
|  |  |
| 1 | 0.7836 | 0.3860 | 0.7836 | -0.2162 | 0.3393 |
| 2 | 0.7726 | 0.4031 | 0.7726 | -0.4581 | 0.1932 |
| 3 | 0.7947 | 0.3685 | 0.7947 | -0.2343 | 0.3136 |
| 4 | 0.7123 | 0.4926 | 0.7123 | 0.4729 | 0.2690 |
| 5 | 0.7119 | 0.4931 | 0.7119 | 0.5235 | 0.2191 |
| 累积贡献 | 0.571342 |  | 0.571342 | 0.733175 |  |

对，残差矩阵为



第一个因子代表了一般经济条件，称为市场因子，所有股票在这个因子上的载

荷都比较大，且大致相等，第二个因子是化学股和石油股的一个对照，两者分别有比较大的负、正载荷。可见使不同的工业部门的股票产生差异，通常称之为工业因子。归纳起来，我们有如下结论：股票回升率由一般经济条件、工业部门活动和各公司本身特殊活动三部分决定。

MATLAB源代码：

clc,clear

r=[1.000 0.577 0.509 0.387 0.462

0.577 1.000 0.599 0.389 0.322

0.509 0.599 1.000 0.436 0.426

0.387 0.389 0.436 1.000 0.523

0.462 0.322 0.426 0.523 1.000];

%下面利用相关系数矩阵求主成分解，val的列为r的特征向量，即主成分的系数

[vec,val,con]=pcacov(r);%val为r的特征值，con为各个主成分的贡献率

f1=repmat(sign(sum(vec)),size(vec,1),1); %构造与vec同维数的元素为±1的矩阵

vec=vec.\*f1; %修改特征向量的正负号，每个特征向量乘以所有分量和的符号函数值

f2=repmat(sqrt(val)',size(vec,1),1);

a=vec.\*f2 %构造全部因子的载荷矩阵

a1=a(:,1) %提出一个因子的载荷矩阵

tcha1=diag(r-a1\*a1') %计算一个因子的特殊方差

a2=a(:,[1,2]) %提出两个因子的载荷矩阵

tcha2=diag(r-a2\*a2') %计算两个因子的特殊方差

ccha2=r-a2\*a2'-diag(tcha2) %求两个因子时的残差矩阵

gong=cumsum(con) %求累积贡献率

该MATLAB源代码运行结果中的a为载荷矩阵。主因子方法是对主成分方法的修正。

2）、主因子分析法：

例 我国上市公司赢利能力与资本结构的实证分析

已知上市公司的数据见表1。

表1 上市公司数据

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 公司 | 销售净利率 | 资产净利率 | 净资产收益率 | 销售毛利率 | 资产负利率 |
| 歌华有线 | 43.31 | 7.39 | 8.73 | 54.89 | 15.35 |
| 五粮液 | 17.11 | 12.13 | 17.29 | 44.25 | 29.69 |
| 用友软件 | 21.11 | 6.03 | 7 | 89.37 | 13.82 |
| 太太药业 | 29.55 | 8.62 | 10.13 | 73 | 14.88 |
| 浙江阳光 | 11 | 8.41 | 11.83 | 25.22 | 25.49 |
| 烟台万华 | 17.63 | 13.86 | 15.41 | 36.44 | 10.03 |
| 方正科技 | 2.73 | 4.22 | 17.16 | 9.96 | 74.12 |
| 红河光明 | 29.11 | 5.44 | 6.09 | 56.26 | 9.85 |
| 贵州茅台 | 20.29 | 9.48 | 12.97 | 82.23 | 26.73 |
| 中铁二局 | 3.99 | 4.64 | 9.35 | 13.04 | 50.19 |
| 红星发展 | 22.65 | 11.13 | 14.3 | 50.51 | 21.59 |
| 伊利股份 | 4.43 | 7.3 | 14.36 | 29.04 | 44.74 |
| 青岛海尔 | 5.4 | 8.9 | 12.53 | 65.5 | 23.27 |
| 湖北宜化 | 7.06 | 2.79 | 5.24 | 19.79 | 40.68 |
| 雅戈尔 | 19.82 | 10.53 | 18.55 | 42.04 | 37.19 |
| 福建南纸 | 7.26 | 2.99 | 6.99 | 22.72 | 56.58 |

MATLAB源代码：

clc,clear

load data.txt; %把原始数据保存在纯文本文件data.txt中

n=size(data,1);

x=data(:,1:4); y=data(:,5); %分别提出自变量x和因变量y的值

——————————————————————————————————

如果不需要检验，则不需要把y列入原始数据中，把矩阵x的大小改变一下，以及下文中的m，m为原始数据中变量的个数。

——————————————————————————————————

m=4;%m为变量的个数

x=zscore(x); %数据标准化

r=cov(x); %求标准化数据的协方差阵，即求相关系数矩阵

[vec,val,con]=pcacov(r); %进行主成分分析的相关计算

c=cumsum(con);

i=1;

while ((c(i)<90)&(con(i+1)>10))

i=i+1;

end

num=i;

f1=repmat(sign(sum(vec)),size(vec,1),1);

vec=vec.\*f1; %特征向量正负号转换

f2=repmat(sqrt(val)',size(vec,1),1);

a=vec.\*f2; %求初等载荷矩阵

am=a(:,1:num); %提出num个主因子的载荷矩阵

[b,t]=rotatefactors(am,'method', 'varimax'); %旋转变换,b为旋转后的载荷阵

bt=[b,a(:,num+1:end)]; %旋转后全部因子的载荷矩阵

contr=sum(bt.^2); %计算因子贡献

rate=contr(1:num)/sum(contr); %计算因子贡献率

fprintf('综合因子得分公式：F=');

for i=1:num

fprintf('+%f\*F%d',rate(i),i);

end

fprintf('\n');

coef=inv(r)\*b; %计算得分函数的系数

coef=coef';

for i=1:num

fprintf('各个因子得分函数为F%d=',i);

for j=1:m

fprintf('+(%f)\*x\_%d',coef(i,j),j);

end

fprintf('\n');

end

%如果仅仅因子分析，程序到此为止

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

score=x\*coef';%计算各个因子的得分

weight=rate/sum(rate); %计算得分的权重

Tscore=score\*weight'; %对各因子的得分进行加权求和，即求各企业综合得分

[STscore,ind]=sort(Tscore,'descend'); %对企业进行排序

display=[score(ind,:)';STscore';ind']; %显示排序结果

fprintf('排序结果如下：');

for i=1:num

fprintf('第%d行为F%d得分,',i,i);

end

fprintf('第%d行为综合因子得分，第%d为原序列\n',num+1,num+2);

disp(display);

[ccoef,p]=corrcoef([Tscore,y]); %计算F与资产负债的相关系数

[d,dt,e,et,stats]=regress(Tscore,[ones(n,1),y]);%计算F与资产负债的方程

fprintf('因子分析法的回归方程为：F=%f+(%f\*y)',d(1),d(2));

if (stats(3)<0.05)%判断是否通过显著性检验的结果

fprintf('\n在显著性水平0.05的情况下，通过了假设检验。\n');

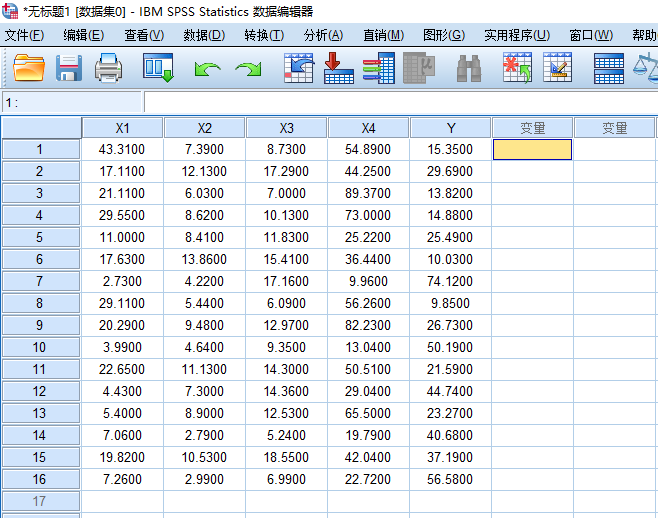
else

fprintf('\n在显著性水平0.05的情况下，通不过假设检验。\n');

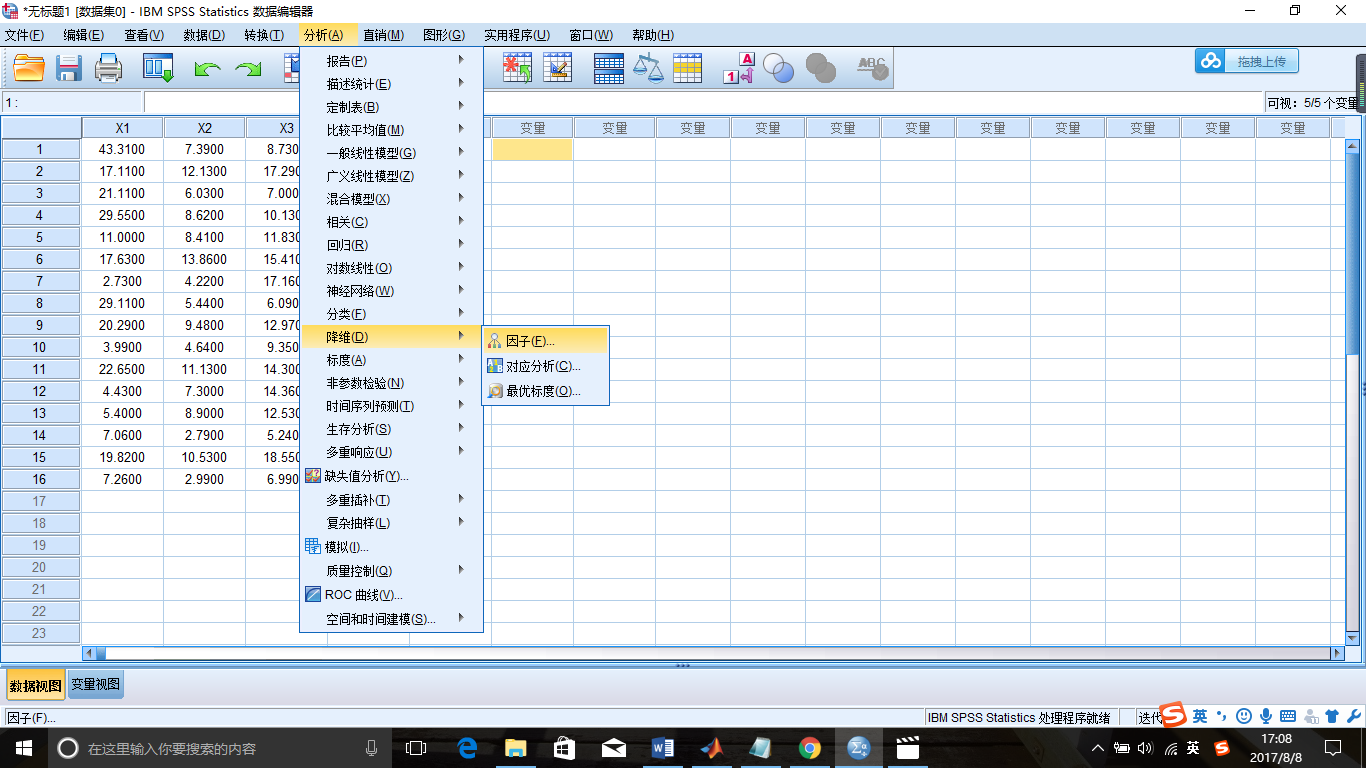
end该MATLAB源代码的displsy为最终排序结果。

本题用SPSS求解（因子分析中的主成分分析法）：

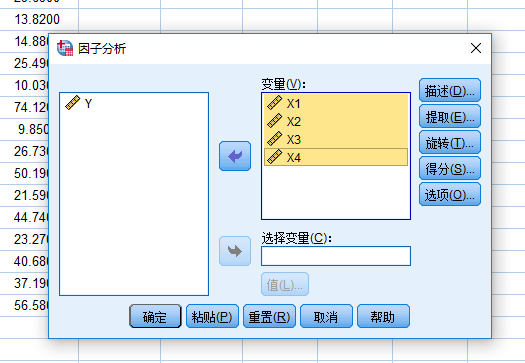
1.把原始数据输入SPSS中，如图：



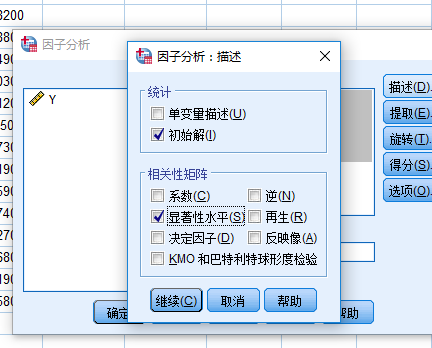
2.依次点击“分析”→“降维”→“因子”，如图：



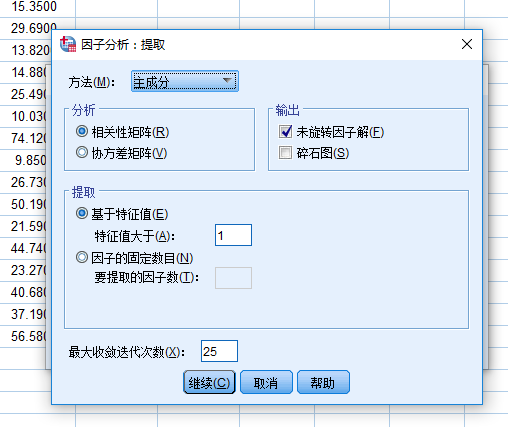
3.在“因子分析”对话框中，把X1、X2、X3、X4移入“变量”中，如图：



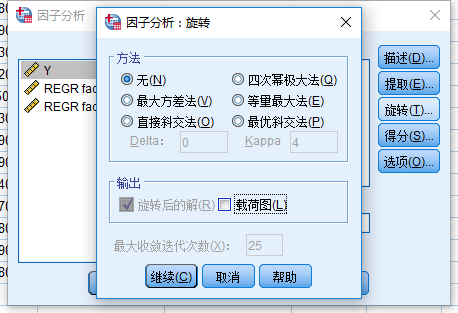
4.单击“描述”，出现“因子分析：描述”对话框；在“统计”栏中，勾选“初始解”；在“相关性矩阵”栏中，勾选“显著性水平”，如图，单击“继续”。



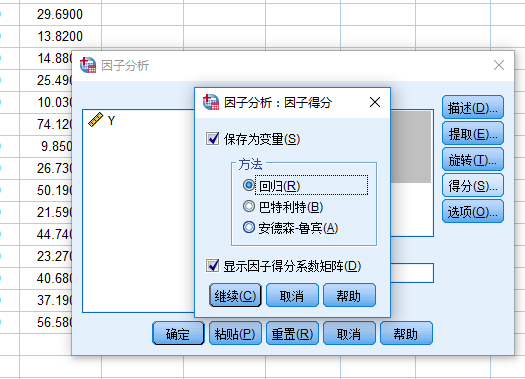
5.单击“提取”，出现“因子分析：提取”对话框，在“方法”中，选择“主成分”；在“分析”中，选择“相关性矩阵”；在“输出”中，勾选“未旋转因子解”；在“提取”中，选择“基于特征值”，并设定特征值大于“1”，如图，单击“继续”。



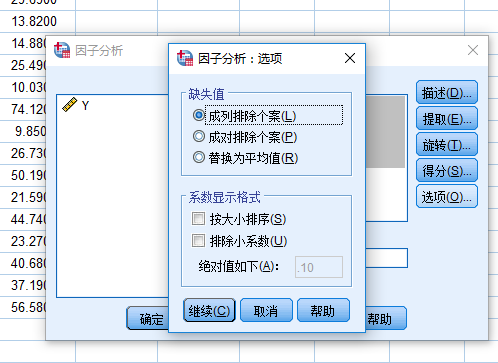
6.单击“选择”，在“因子分析：旋转”对话框中的选项默认即可，如图，单击“继续”。



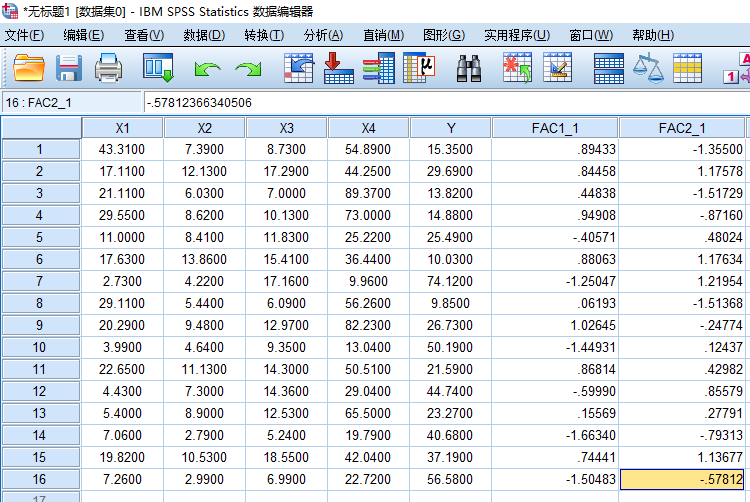
7.单击“得分”，在“因子分析：因子得分”对话框中，勾选“保存为变量”；在“方法”中，选择“回归”；并勾选“显示因子得分系数矩阵”，如图，单击“继续”。



8.单击“选项”，在“因子分析：选项”对话框中的选项默认即可，如图，单击“继续”。



9.在“因子分析”对话框中，单击“确定”，如图。



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **相关性矩阵** | | | | | |
|  | | X1 | X2 | X3 | X4 |
| 显著性 （单尾） | X1 |  | .114 | .263 | .006 |
| X2 | .114 |  | .002 | .096 |
| X3 | .263 | .002 |  | .304 |
| X4 | .006 | .096 | .304 |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **总方差解释** | | | | | | |
| 成分 | 初始特征值 | | | 提取载荷平方和 | | |
| 总计 | 方差百分比 | 累积 % | 总计 | 方差百分比 | 累积 % |
| 1 | 1.897 | 47.429 | 47.429 | 1.897 | 47.429 | 47.429 |
| 2 | 1.550 | 38.740 | 86.169 | 1.550 | 38.740 | 86.169 |
| 3 | .393 | 9.826 | 95.995 |  |  |  |
| 4 | .160 | 4.005 | 100.000 |  |  |  |
| 提取方法：主成分分析法。 | | | | | | |

综合因子分析得分公式为：



|  |  |  |
| --- | --- | --- |
| **成分矩阵a** | | |
|  | 成分 | |
| 1 | 2 |
| X1 | .731 | -.513 |
| X2 | .818 | .503 |
| X3 | .359 | .897 |
| X4 | .752 | -.477 |
| 提取方法：主成分分析法。 | | |
| a. 提取了 2 个成分。 | | |

|  |  |  |
| --- | --- | --- |
| **成分得分系数矩阵** | | |
|  | 成分 | |
| 1 | 2 |
| X1 | .385 | -.331 |
| X2 | .431 | .325 |
| X3 | .189 | .579 |
| X4 | .396 | -.308 |
| 提取方法：主成分分析法。  组件得分。 | | |

各个因子得分函数：





|  |  |  |
| --- | --- | --- |
| **成分得分协方差矩阵** | | |
| 成分 | 1 | 2 |
| 1 | 1.000 | .000 |
| 2 | .000 | 1.000 |
| 提取方法：主成分分析法。  组件得分。 | | |