

特别说明

此资料来自豆丁网(<http://www.docin.com/>)

您现在所看到的文档是使用**下载器**所生成的文档

此文档的原件位于

<http://www.docin.com/p-19965047.html>

感谢您的支持

抱米花

<http://blog.sina.com.cn/lotusbaob>

SPSS 中主成分分析的基本操作

Xiaowenzi22 与 pinksss 共同制作

阐述主成分分析法的原理

主成分分析是设法将原来众多具有一定相关性（比如 P 个指标），重新组合成一组新的互相无关的综合指标来代替原来的指标。通常数学上的处理就是将原来 P 个指标作线性组合，作为新的综合指标。最经典的做法就是用 F_1 （选取的第一个线性组合，即第一个综合指标）的方差来表达，即 $\text{Var}(F_1)$ 越大，表示 F_1 包含的信息越多。因此在所有的线性组合中选取的 F_1 应该是方差最大的，故称 F_1 为第一主成分。如果第一主成分不足以代表原来 P 个指标的信息，再考虑选取 F_2 即选第二个线性组合，为了有效地反映原来信息， F_1 已有的信息就不需要再出现在 F_2 中，用数学语言表达就是要求 $\text{Cov}(F_1, F_2)=0$ ，则称 F_2 为第二主成分，依此类推可以构造出第三、第四，……，第 P 个主成分。

主成分模型：

$$\begin{cases} F_1 = a_{11}X_{11} + a_{21}X_{21} + \dots + a_{p1}X_p \\ F_2 = a_{12}X_{12} + a_{22}X_{22} + \dots + a_{p2}X_p \\ \dots \\ F_p = a_{1m}X_{11} + a_{2m}X_{22} + \dots + a_{pm}X_p \end{cases}$$

其中 $a_{1i}, a_{2i}, \dots, a_{pi}$ ($i=1, \dots, m$) 为 X 的协方差阵 Σ 的特征值对应的特征向量， X_1, X_2, \dots, X_p 是原始变量经过标准化处理后的值（因为在实际应用中，往往存在指标的量纲不同，所以在计算之前先消除量纲的影响，而将原始数据标准化）。

$A = (a_{ij})_{p \times m} = (\lambda_1, \lambda_2, \dots, \lambda_m)$ ， $R_{ii} = \lambda_i$ ，R 为相关系数矩阵， $\lambda_1, \lambda_2, \dots, \lambda_p$ 是相应的特征值和单位特征向量， $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$

上述方程组要求：

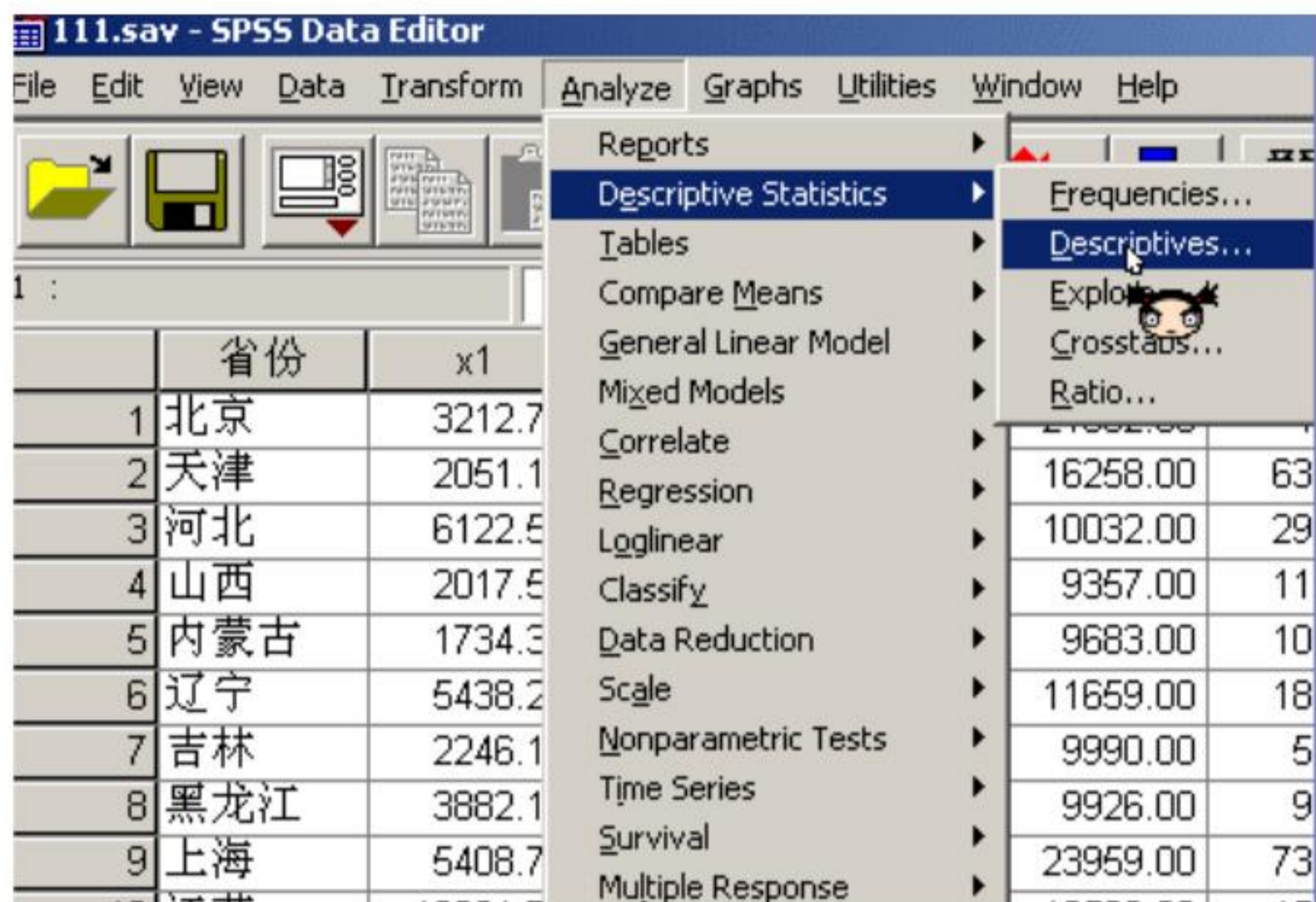
1、 $a_{1i}^2 + a_{2i}^2 + \dots + a_{pi}^2 = 1$ ($i=1, \dots, m$)

2、 $AA^T = I_m$ ($A = (a_{ij})_{p \times m} = (\lambda_1, \lambda_2, \dots, \lambda_m)$ ，A 为正交矩阵)

3、 $\text{Cov}(F_i, F_j) = \lambda_{ij}$ ， $\lambda_{ij} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$

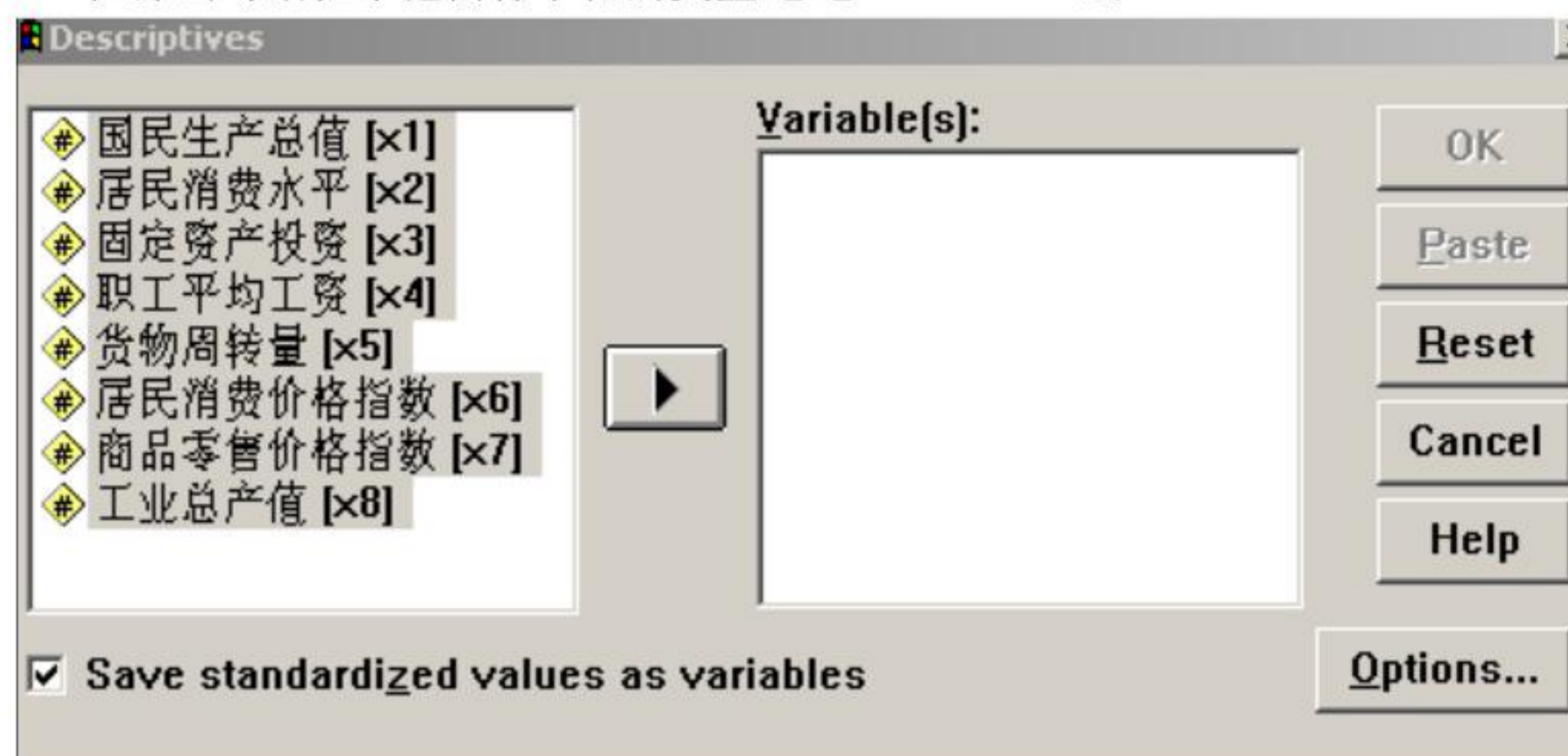
操作步骤：

一、数据标准化



1、

2、在弹出对话框中把需标准化的变量选进 Variable 去



并在下面的提示前打钩



3、然后点“OK”

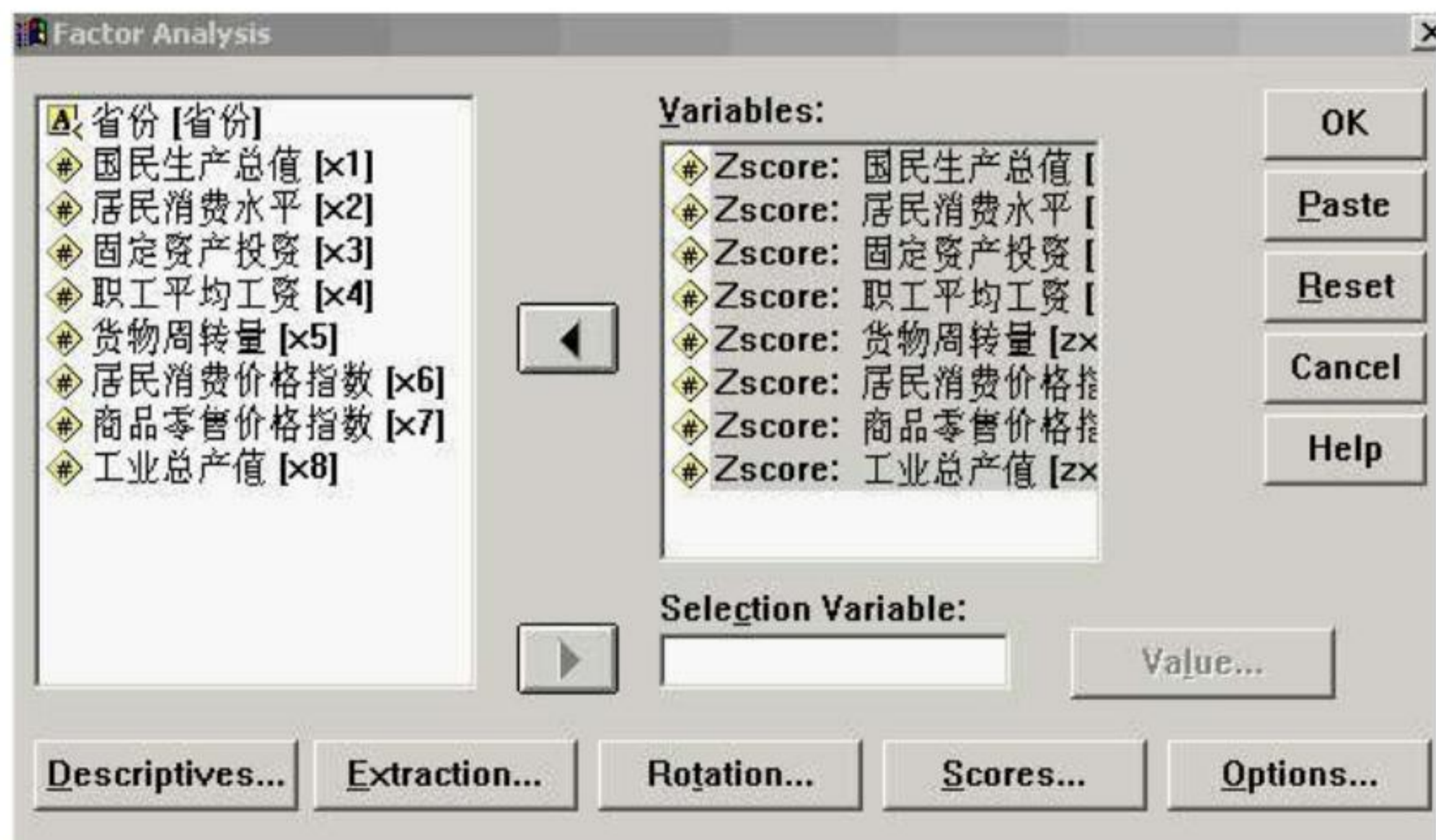
4、数据编辑窗内将出现结果

二、主成分分析基本操作

1、

The screenshot shows the SPSS 'Analyze' menu. The 'Data Reduction' option is selected, and its submenu is open, showing 'Factor...' as the first option. Other options in the 'Data Reduction' submenu include 'Correspondence Analysis...', 'Optimal Scaling...', 'Cluster...', 'Multidimensional Scaling...', 'Two-Way Tables...', 'Path Analysis...', 'Latent Class Analysis...', 'Latent Markov Models...', 'Latent Transition Analysis...', 'Latent Class Growth Analysis...', 'Latent Class Cluster Analysis...', 'Latent Class Regression Analysis...', 'Latent Class Structural Equation Modeling...', 'Latent Class Growth Modeling...', 'Latent Class Cluster Modeling...', 'Latent Class Regression Modeling...', and 'Latent Class Structural Equation Modeling Modeling...'.

2、选择后弹出下面的对话框

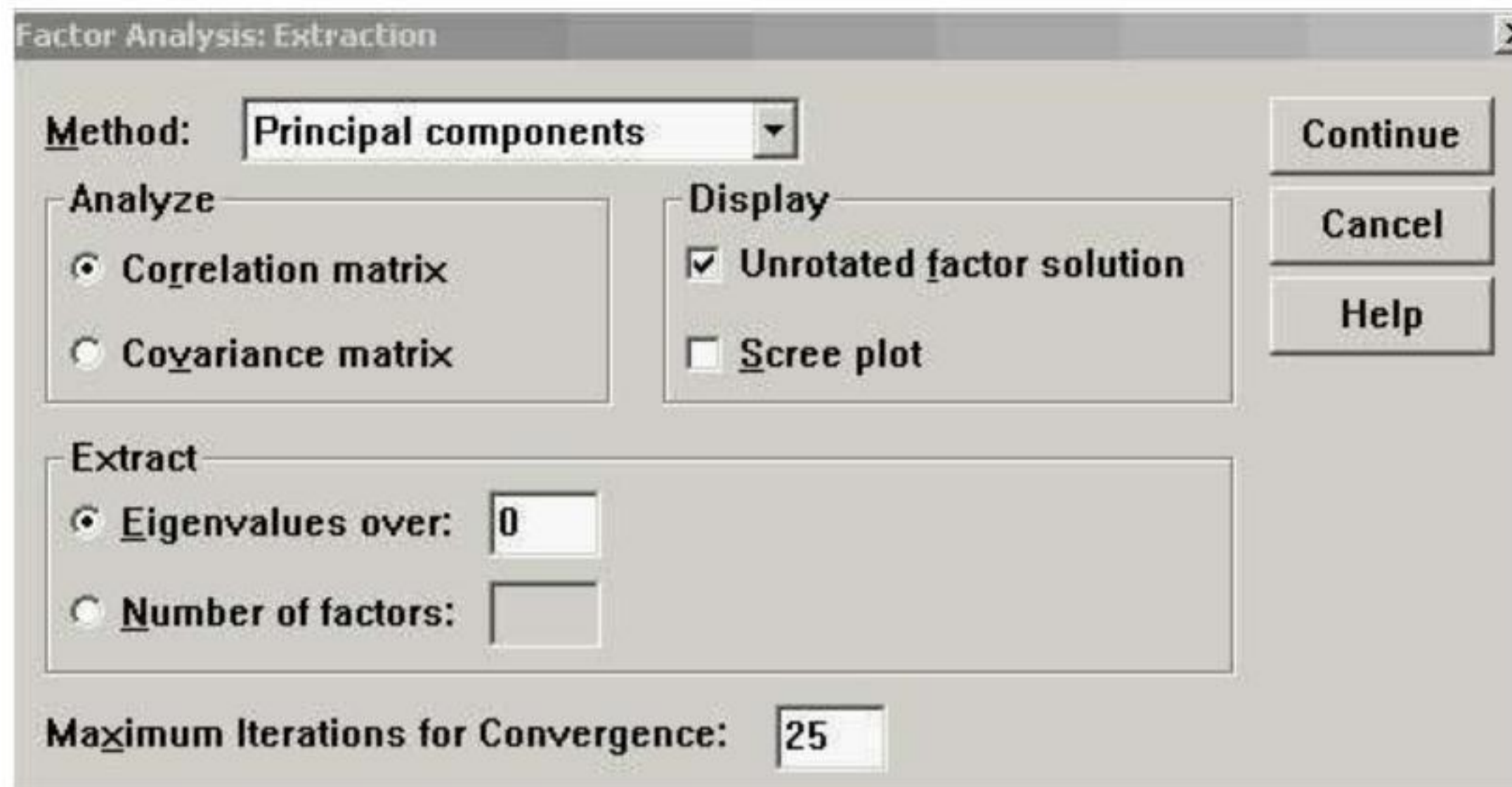


3、把标准化后的数据都选进 Variables 去

4、点击

Extraction...

5、弹出现下面的对话框



The image shows the 'Factor Analysis: Extraction' dialog box in SPSS. The 'Method' dropdown is set to 'Principal components'. The 'Analyze' section has 'Correlation matrix' selected. The 'Display' section has 'Unrotated factor solution' checked and 'Scree plot' unchecked. The 'Extract' section has 'Eigenvalues over:' selected with a value of '0'. The 'Maximum iterations for convergence' is set to '25'. Buttons for 'Continue', 'Cancel', and 'Help' are on the right.

Factor Analysis: Extraction

Method: Principal components

Analyze

- ☒ Correlation matrix
- ☐ Covariance matrix

Display

- ☒ Unrotated factor solution
- ☐ Scree plot

Extract

- ☒ Eigenvalues over: 0
- ☐ Number of factors:

Maximum iterations for convergence: 25

Continue
Cancel
Help

6、在对话框的空白处填 0，记得上面的图中要选中前面的点



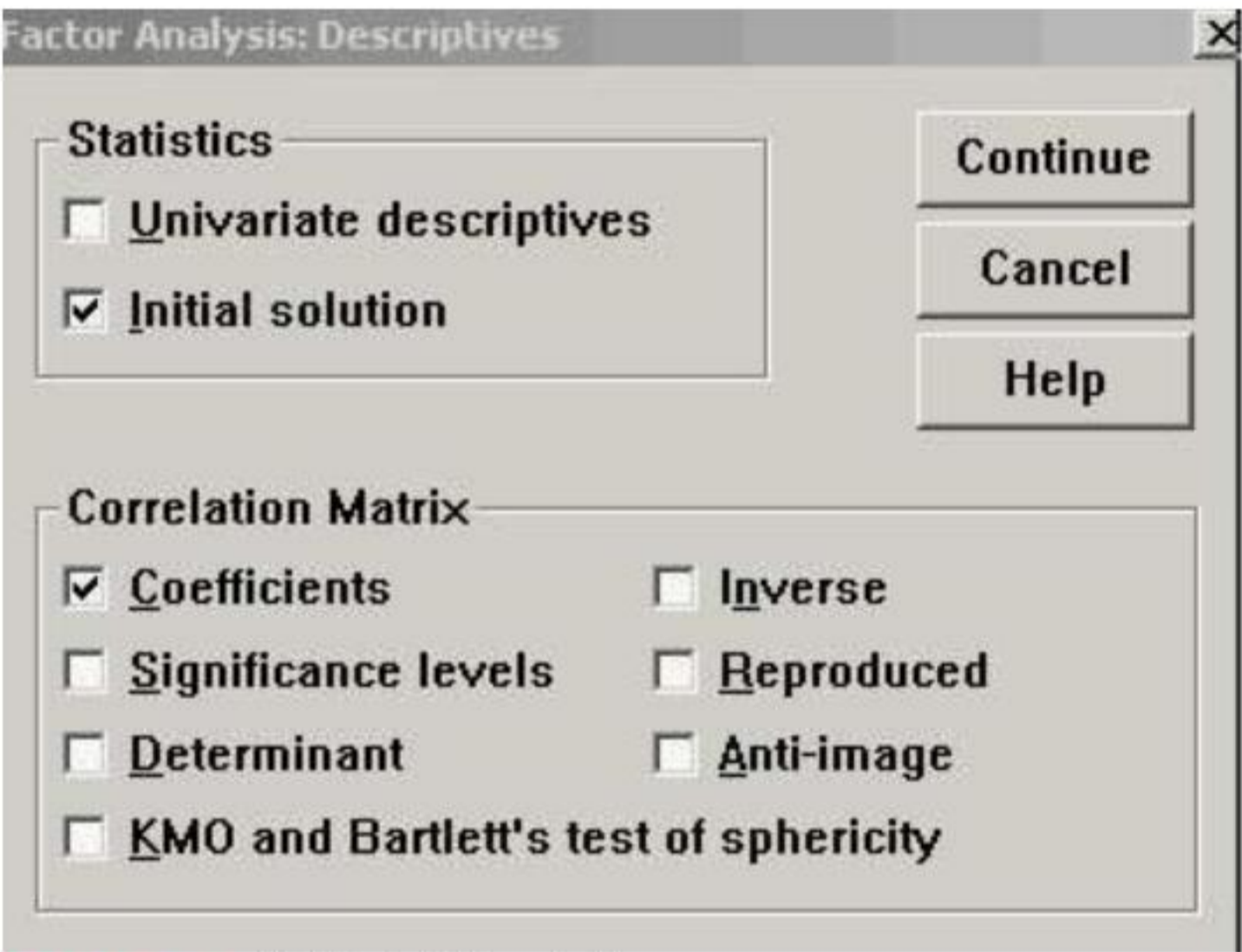
7、点击 continue 钮

8、返回上个对话框

9、如需要得到相关系数矩阵，点击



10、弹出下面的对话框



在 **Coefficients** 前的方框打上钩

11、然后点击 continue 钮

12、返回上个对话框，点击 “OK”

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.849	48.118	48.118	3.849	48.118	48.118
2	1.808	22.594	70.712	1.808	22.594	70.712
3	1.306	16.329	87.042	1.306	16.329	87.042
4	.595	7.443	94.485	.595	7.443	94.485
5	.289	3.608	98.092	.289	3.608	98.092
6	.078	.977	99.069	.078	.977	99.069
7	.057	.718	99.787	.057	.718	99.787
8	.017	.213	100.000			

Extraction Method: Principal Component Analysis.

三、提取特征向量

1、在计算主成分的步骤中将出现因子载荷矩阵，我们可以取得每个主成分的方差，即特征根，它的大小表示了对应主成分能够描述原来所有信息的多少（更多情况下是由方差贡献率来反映）。一般来讲，为了达到降维的目的，我们只提取前几个主成分，由于前 3 个特征值累计贡献率达到 87.042%，根据累计贡献率大于 85%的原则，故选取前三个特征值。所以决定用三个新变量来代替原来的七个变量。但这三个新变量的表达还不能从输出窗口中直接得到，因为“Component Matrix”是指因子载荷矩阵，每一个载荷量表示主成分与对应变量的相关系数。

Component Matrix ^a							
	Component						
	1	2	3	4	5	6	7
1.úÃñÉú²ú×ÜÖµ(x1)	.855	.477	-.025	.049	-.133	-.098	.069
¾ÓÃñÏû·ÑË®Æ½(x2)	.747	-.614	.083	.103	.086	.179	.088
1.¶×Ê³ú¶×Ê(x3)	.916	.352	-.030	.103	-.094	-.007	.089
Öº¼Æ½¾¼û¹×Ê(x4)	.554	-.688	.330	.231	.169	-.169	-.031
»ðÏÖÜ×ªÁ¿(x5)	.627	-.078	.371	-.680	.028	-.009	-.021
¾ÓÃñÏû·Ñ¼Û,ñÖ,Êý(x6)	-.379	-.095	.851	.132	-.325	.027	.000
ÉÏÆ·ÁãÊ¼¼Û,ñÖ,Êý(x7)	-.285	.682	.569	.086	.346	.024	.046
¹×Öµ×Ü²úÖµ(x8)	.893	.355	.063	.179	.001	.081	-.183

Extraction Method: Principal Component Analysis.

a. 7 components extracted.

2、将前三个因子载荷矩阵输入（可用复制粘贴的方法）到数据编辑窗口（为变量 B1、B2、B3），然后利用“Transform→compute”，在对话框中输入“A1=B1/SQR(3.849)”[注：第二主成分 SQR 后的括号中填 1.808，第三主成分 SQR 后的括号中填 1.306]，即可得到特征向量 A1。同理，可得到 A2、A3。然后就可以得出主成分表达式。

四、主成分排名

将特征向量与标准化后的数据相乘，就可以得到各个主成分得分 Z_1 、 Z_2 、 Z_3 ，若需求综合评价函数，还需在 Transform→compute 输入综合评价函数， Z_1 、 Z_2 、 Z_3 前的系数是主成分的方差贡献率。

参考文献

[1] 张文彤主编《SPSS11 统计分析教程（高级篇）》[M]，北京希望电子出版社，2002 年 6 月。

[2] 王芳 《主成分分析与因子分析的异同比较及应用》，《统计教育》，2003 年第 5 期。

[3] 于秀林 任雪松，《多元统计分析》，中国统计出版社，1999 年 8 月。