

1. 概述

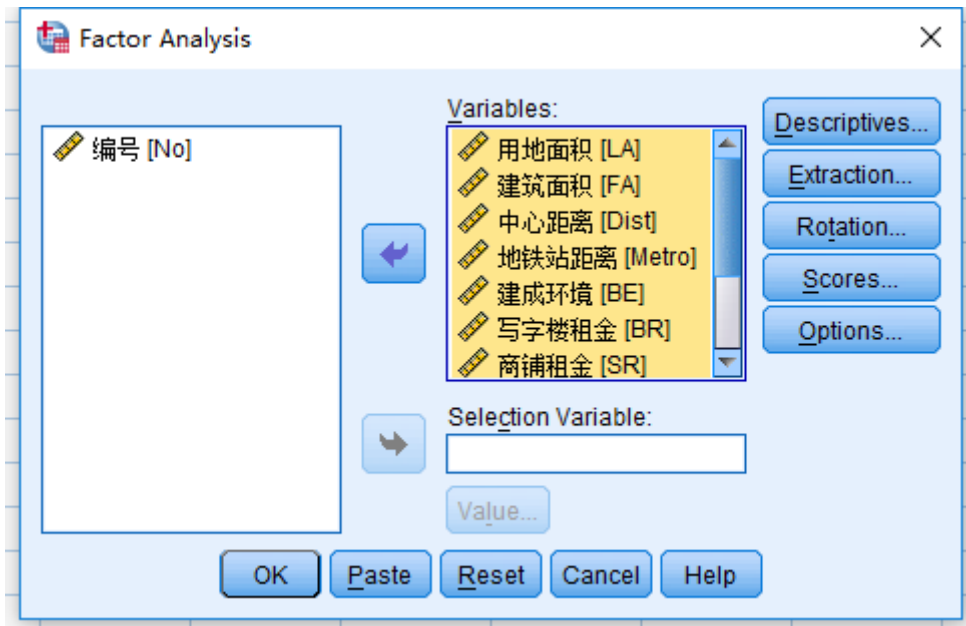
本节以 SPSS 软件演示主成分分析和因子分析的操作过程，使用的数据文件为“pca_fa.sav”，另外，附件中还提供了对应的 csv 数据。在 SPSS 中没有独立的主成分分析功能，而是借助因子分析实现的。因此，在下面的操作演示中将同步介绍主成分分析与因子分析两种方法。

2. 操作过程

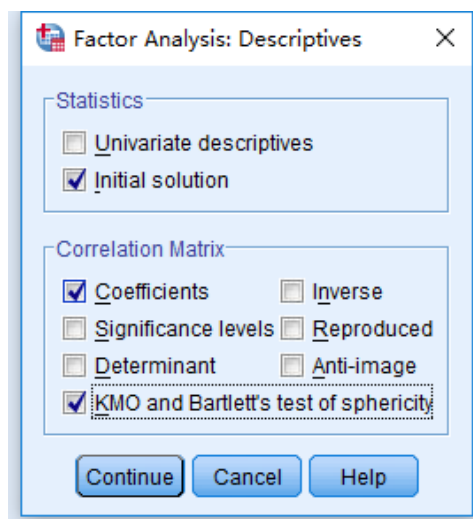
在 SPSS 中打开“pca_fa.sav”，如下所示。

No	LA	FA	Dist	Metro	BE	BR	SR	Loc	AQ	AH	IC
1	.35	.29	.82	556.38	1	3.20	4.95	1	1	12	5.67
2	.49	.46	1.25	473.54	1	3.20	5.68	1	1	16	4.27
3	23.90	11.72	2.72	1290.33	1	2.20	2.73	3	2	15	4.22
4	.23	.14	2.05	1174.59	1	1.80	3.37	3	1	12	4.11
5	.85	1.02	1.17	542.95	2	3.70	5.49	1	3	22	5.97
6	.08	.12	1.10	485.76	1	3.70	5.58	1	1	12	4.41
7	.34	.51	.20	243.80	2	3.50	9.07	1	2	21	4.05
8	1.77	3.34	.46	566.34	2	3.20	6.11	1	3	36	5.02
9	.56	.81	1.08	226.31	1	3.10	6.15	1	1	16	4.00
10	1.22	1.83	.64	540.87	2	2.10	5.29	3	3	32	4.09

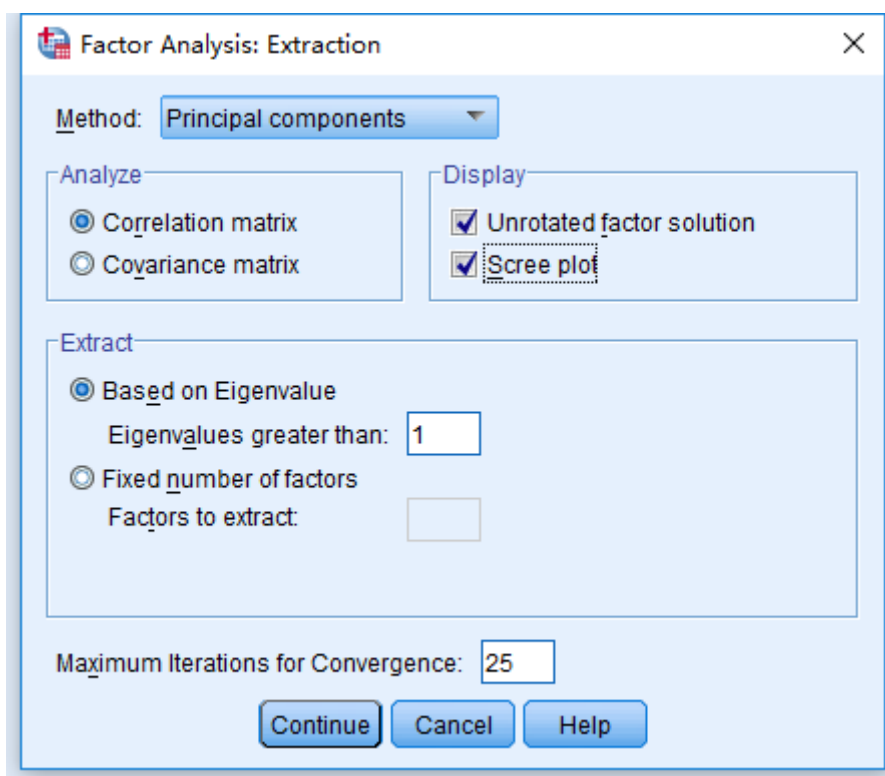
在菜单栏中选择“Analyze – Dimension Reduction – Factor”，进入因子分析对话框。将除了“编号 [No]”以外的所有变量全部选入“Variables”中，如下所示：



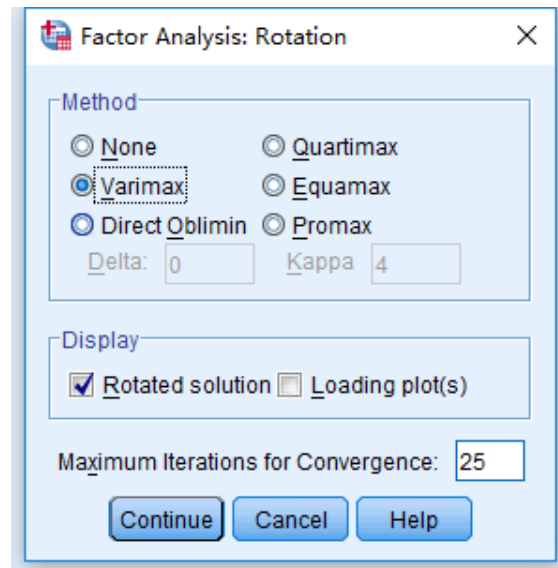
点击“Descriptives”，进入统计量对话框。在“Correlation Matrix”中勾选“Coefficients”，以汇报原始变量之间的相关系数矩阵；同时勾选“KMO and Bartlett's test of sphericity”，以汇报因子分析所需要的适宜性检验——KMO 指标与 Bartlett 球度检验。对于主成分分析而言，该检验并非必需。



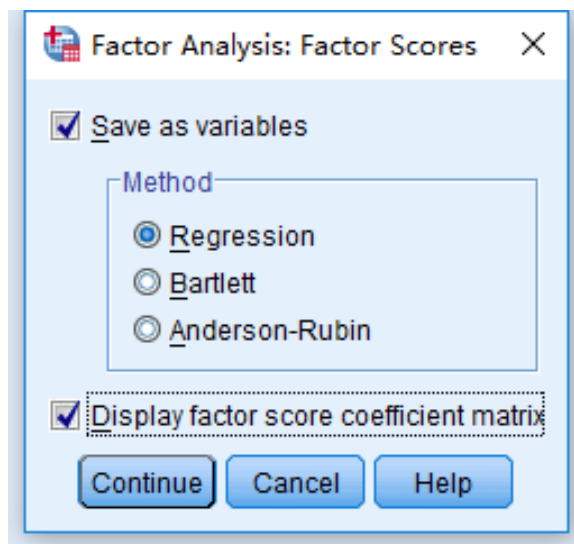
点击“Continue”回到主对话框，点击“Extraction”进入因子提取对话框。确认“Method”下拉菜单中被选中的是“Principal components”，即使用主成分法。这一设定使得我们可以得到主成分分析的结果，同时将使用主成分法作为提取因子的方法。在“Display”中勾选“Scree plot”，以显示碎石图。下方的“Extract”用于设置提取标准，确认“Based on Eigenvalue”被选中，且“Eigenvalues greater than:”的输入框中的数字为 1，即使用“特征值大于 1”的标准提取因子或主成分；如果想提取固定数量的因子或主成分，则可以选中下方的“Fixed number of factors”，并输入希望提取的数量。



点击“Continue”回到主对话框，然后点击“Rotation”进入因子旋转对话框。此时“Method”默认选中“None”，即不进行因子旋转。对于主成分分析而言，这样即可；而对于因子分析而言，一般需要通过因子旋转提高可解释性。这里，我们选中“Varimax”，使用常用的最大方差法进行因子旋转。



点击“Continue”回到主对话框，然后点击“Scores”进入因子得分对话框。请注意，这里设置的是因子得分，而非主成分取值，后者需要自行计算。在因子得分对话框中勾选“Save as variables”，将因子得分保存为变量，“Method”使用默认的“Regression”即可。同时，勾选“Display factor score coefficient matrix”，以显示因子得分矩阵。



点击“Continue”回到主对话框，然后点击“OK”运行分析。SPSS 的结果窗口首先报告了各原始变量之间的相关系数矩阵。

Correlation Matrix

	用地面积	建筑面积	中心距离	地铁站距离	建成环境	写字楼租金	商铺租金	区位	建筑质量	建筑高度	不规则系数	
Correlation	用地面积	1.000	.917	.262	.227	-.105	-.166	-.218	.252	.037	-.040	.007
	建筑面积	.917	1.000	.239	.236	-.044	-.189	-.217	.293	.116	.109	.041
	中心距离	.262	.239	1.000	.753	-.171	-.546	-.715	.583	-.193	-.192	.066
	地铁站距离	.227	.236	.753	1.000	-.102	-.419	-.681	.559	-.111	-.127	.015
	建成环境	-.105	-.044	-.171	-.102	1.000	.065	.070	-.021	.725	.760	-.132
	写字楼租金	-.166	-.189	-.546	-.419	.065	1.000	.361	-.732	.099	-.012	.047
	商铺租金	-.218	-.217	-.715	-.681	.070	.361	1.000	-.608	.115	.076	.003
	区位	.252	.293	.583	.559	-.021	-.732	-.608	1.000	.011	.032	-.031
	建筑质量	.037	.116	-.193	-.111	.725	.099	.115	.011	1.000	.806	-.027
	建筑高度	-.040	.109	-.192	-.127	.760	-.012	.076	.032	.806	1.000	-.090
	不规则系数	.007	.041	.066	.015	-.132	.047	.003	-.031	-.027	-.090	1.000

然后，SPSS 报告了因子分析的适宜性检验结果，即 KMO 指标和 Bartlett 球度检验。

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.661
Bartlett's Test of Sphericity	Approx. Chi-Square	848.081
	df	55
	Sig.	.000

下面报告的是变量共同度，这是因子分析中各原始变量的信息量被公因子所解释的比例。

Communalities

	Initial	Extraction
用地面积	1.000	.959
建筑面积	1.000	.962
中心距离	1.000	.797
地铁站距离	1.000	.704
建成环境	1.000	.818
写字楼租金	1.000	.567
商铺租金	1.000	.683
区位	1.000	.737
建筑质量	1.000	.858
建筑高度	1.000	.876
不规则系数	1.000	.956

Extraction Method: Principal Component Analysis.

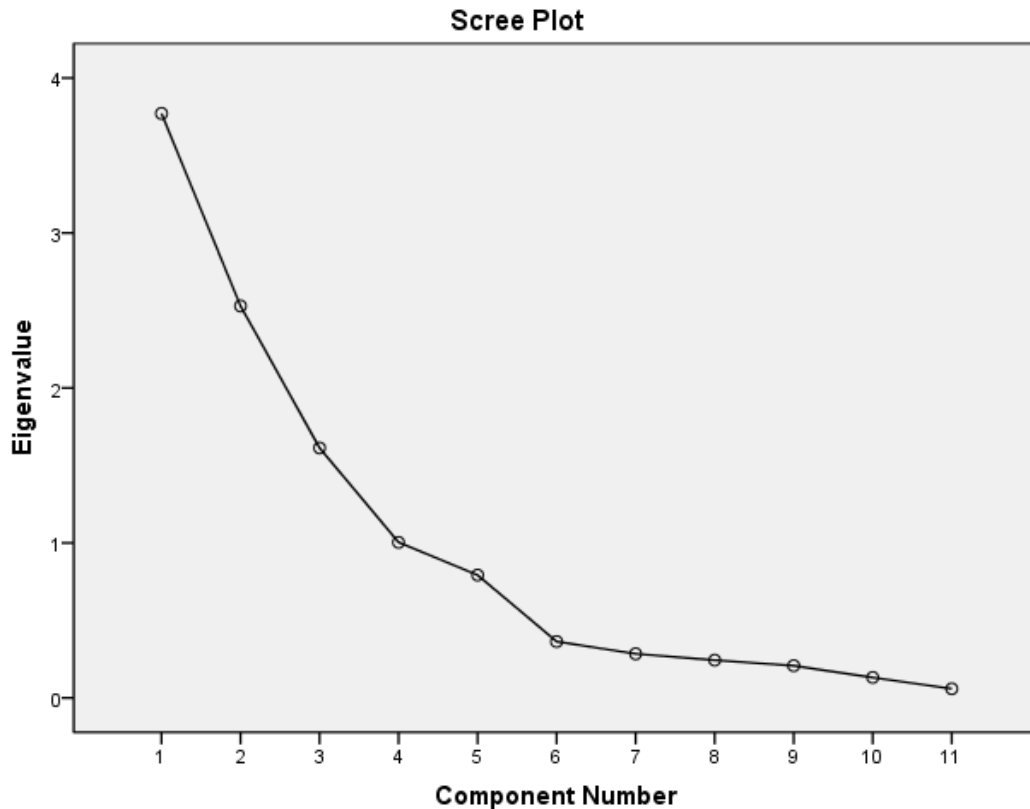
接下来，SPSS 报告了主成分/公因子提取的情况，即每个主成分/公因子对应的特征值、方差贡献率、累计方差贡献率。特别地，对于因子分析而言，包含了因子旋转之前和之后的结果。

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.771	34.283	34.283	3.771	34.283	34.283	3.396	30.870	30.870
2	2.529	22.992	57.274	2.529	22.992	57.274	2.565	23.314	54.184
3	1.613	14.666	71.940	1.613	14.666	71.940	1.932	17.561	71.745
4	1.004	9.127	81.067	1.004	9.127	81.067	1.025	9.321	81.067
5	.793	7.205	88.272						
6	.364	3.305	91.577						
7	.285	2.587	94.164						
8	.243	2.213	96.377						
9	.208	1.890	98.267						
10	.132	1.198	99.465						
11	.059	.535	100.000						

Extraction Method: Principal Component Analysis.

然后，SPSS 以碎石图的形式呈现了各主成分/公因子的特征值（信息量）的变化。



接下来，SPSS 报告了荷载矩阵。对于主成分分析而言，这就可以作为最终结果的主成分荷载矩阵。对于因子分析而言，这是因子旋转前的初始因子荷载矩阵。请注意，对于同样一个矩阵，主成分分析与因子分析具有不同的数学意义。

Component Matrix^a

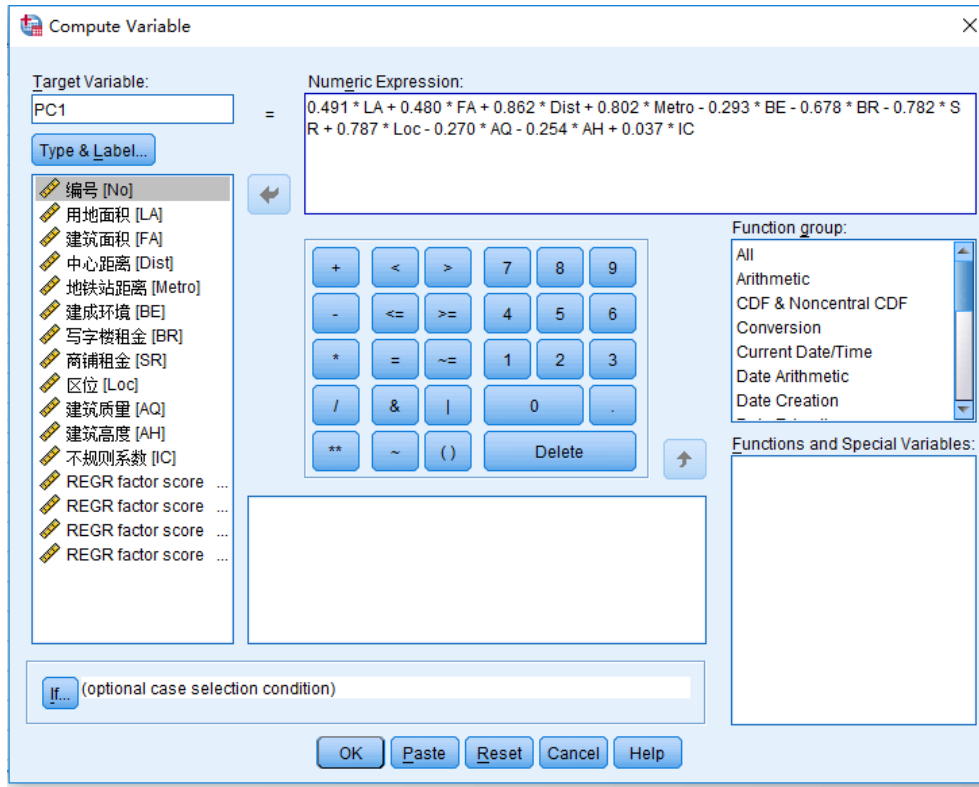
	Component			
	1	2	3	4
用地面积	.491	.223	.813	-.085
建筑面积	.480	.330	.788	-.037
中心距离	.862	.003	-.204	.112
地铁站距离	.802	.071	-.216	.096
建成环境	-.293	.827	-.218	.026
写字楼租金	-.678	-.152	.252	.147
商铺租金	-.782	-.101	.235	-.079
区位	.787	.256	-.219	-.067
建筑质量	-.270	.875	-.018	.137
建筑高度	-.254	.894	-.104	.053
不规则系数	.037	-.139	.128	.959

Extraction Method: Principal Component Analysis.

a. 4 components extracted.

至此，主成分分析的结果就已经全部呈现完毕。如果需要得到每个样本在各个主成分变量上的取值，可以通过“Transform – Compute Variable”，利用荷载矩

阵中的系数计算新变量，计算公式参见课件。例如，如果要计算主成分 1 的取值结果，可以输入如下公式：



下面的结果均为与因子分析相关的结果。在初始因子荷载矩阵之后，SPSS 报告了旋转因子荷载矩阵。在因子分析中，我们一般基于这个矩阵（而非上面的初始因子荷载矩阵）进行解释和因子命名。

Rotated Component Matrix^a

	Component			
	1	2	3	4
用地面积	.156	-.044	.966	-.007
建筑面积	.175	.070	.962	.028
中心距离	.863	-.167	.111	.110
地铁站距离	.825	-.087	.097	.085
建成环境	-.039	.894	-.098	-.081
写字楼租金	-.731	.001	-.061	.170
商铺租金	-.819	.052	-.082	-.063
区位	.837	.075	.148	-.094
建筑质量	-.086	.916	.093	.047
建筑高度	-.037	.934	.033	-.048
不规则系数	-.006	-.054	.016	.976

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 4 iterations.

接下来，SPSS 将报告一个转换矩阵，其意义是表明如何将初始因子荷载矩

阵转换为旋转因子荷载矩阵，我们一般不需要关注。

Component Transformation Matrix

Component	1	2	3	4
1	.903	-.247	.352	.024
2	.167	.950	.246	-.092
3	-.395	-.153	.899	.105
4	.035	.111	-.081	.990

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

然后，SPSS 报告的是因子得分矩阵。请根据课件中的原理介绍，区分主成分荷载、初始因子荷载、旋转因子荷载、因子得分得不同矩阵的区别和联系。

Component Score Coefficient Matrix

	Component			
	1	2	3	4
用地面积	-.070	-.035	.528	-.036
建筑面积	-.058	.014	.519	.006
中心距离	.260	-.024	-.042	.102
地铁站距离	.253	.005	-.047	.083
建成环境	.039	.353	-.071	-.021
写字楼租金	-.229	-.020	.051	.163
商铺租金	-.254	-.018	.055	-.064
区位	.257	.058	-.018	-.085
建筑质量	.002	.363	.039	.100
建筑高度	.026	.368	.001	.011
不规则系数	.002	.039	-.017	.959

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
Component Scores.

在 SPSS 结果窗口的最后，报告的是因子得分的协方差矩阵。由于在计算因子得分时采用的是标准化变量，因此协方差矩阵实质相当于相关系数矩阵。同时，由于采用正交矩阵，所以因子得分两两之间的相关系数均为 0。

Component Score Covariance Matrix

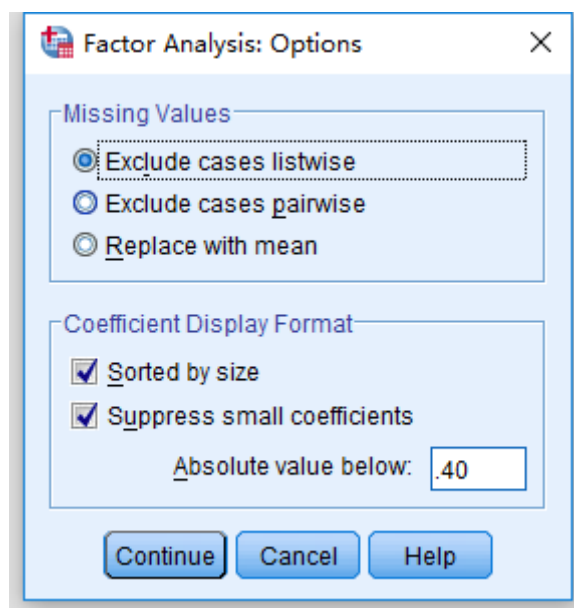
Component	1	2	3	4
1	1.000	.000	.000	.000
2	.000	1.000	.000	.000
3	.000	.000	1.000	.000
4	.000	.000	.000	1.000

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

回到数据视图，我们可以看到 SPSS 在最右侧生成了 4 个新的变量，它们就是 4 个公因子的得分。

Loc	AQ	AH	IC	FAC1_1	FAC2_1	FAC3_1	FAC4_1
1	1	12	5.67	-.44775	-1.09906	-.24469	2.79098
1	1	16	4.27	-.46093	-1.01749	-.08026	-.24464
3	2	15	4.22	.78053	-.68935	9.74152	-.56809
3	1	12	4.11	1.80349	-1.06031	-.63991	-.87186
1	3	22	5.97	-.55084	1.00046	.18846	3.85104
1	1	12	4.41	-.68040	-1.24371	-.27564	.18425
1	2	21	4.05	-1.50213	.30746	.06227	-.84775
1	3	36	5.02	-.76004	1.76410	1.41582	1.53140
1	1	16	4.00	-.72601	-1.03857	.13813	-.96631
3	3	32	4.09	.54998	1.62728	.51938	-.98863

最后介绍一个使输出结果更加清晰易读的操作技巧。重新选择“Analyze – Dimension Reduction – Factor”，进入因子分析对话框，并按照上面的介绍进行设定。在运行分析之前，在主对话框中点击“Options”，进入选项子对话框。在“Coefficient Display Format”中，勾选“Sorted by size”和“Suppress small coefficients”，并在下方的“Absolute value below”中输入0.4，如下图所示。



点击“Continue”回到主对话框，点击“OK”运行分析。这一设定的效果体现在初始因子荷载矩阵和旋转因子荷载矩阵上。下图显示了此时的旋转因子荷载矩阵。可以看到：（1）每个公因子所代表的原始变量已经按荷载绝对值从大到小排序；（2）荷载绝对值低于一定的阈值（这里设置为0.4）时将不显示。这样的结果显然更加清晰，有利于我们快速找到公因子与原始变量的对应关系。

Rotated Component Matrix^a

	Component			
	1	2	3	4
中心距离	.863			
区位	.837			
地铁站距离	.825			
商铺租金	-.819			
写字楼租金	-.731			
建筑高度		.934		
建筑质量		.916		
建成环境		.894		
用地面积			.966	
建筑面积			.962	
不规则系数				.976

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 4 iterations.