

使用 SPSS 软件进行因子分析和聚类分析的方法

◇ 通灵珠宝(中国)有限公司 陈 治

一、方法原理

1. 因子分析 (Factor Analysis)

因子分析是从多个变量指标中选择出少数几个综合变量指标的一种降维的多元统计方法。

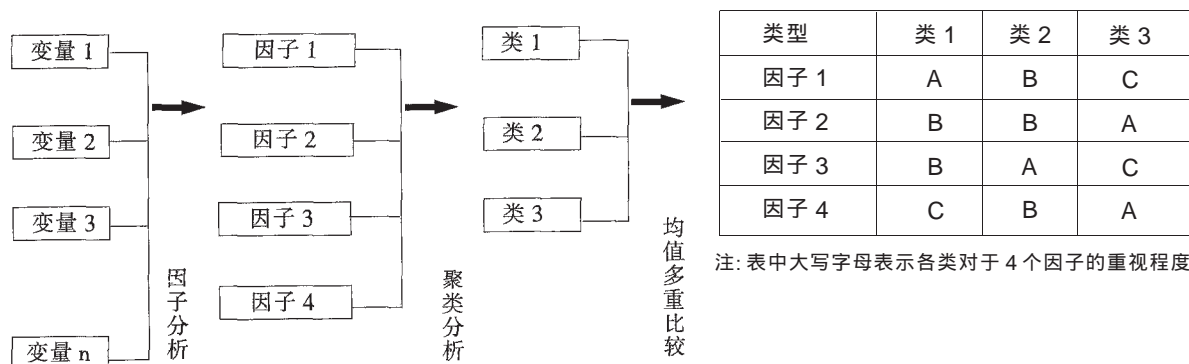
我们在多元分析中处理的是多指标的问题, 观察指标的增加是为了使研究过程趋于完整, 但由于指标太多, 使

得分析的复杂性增加; 同时在实际工作中, 指标间经常具备一定的相关性, 使得观测数据所放映的信息有重叠, 故人们希望用较少的指标代替原来较多的指标, 但依然能放映原有的全部信息, 于是就产生了因子分析方法。

2. 聚类分析 (Cluster Analysis)

聚类分析是根据事物本身特性来研究个体分类的统计方法, 是按照物以类聚的原则来研究的事物分类。

3. 市场细分方法的流程图



二、实证分析

		总人口 (亿)	生产总值 (亿)	消费总额 (亿)	人均年工资 (千)	年度总储蓄 额/万亿	年度财政 总收入/亿
1	启东市	114.38	153.63	50.27	10.88	124.55	10.02
2	江都市	106.69	139.29	43.36	10.64	119.42	11.19
3	丹阳市	80.25	174.75	46.01	13.50	95.81	16.62
4	如皋市	143.87	98.76	37.36	11.29	93.18	9.04
5	兴化市	154.99	103.29	26.00	10.37	78.61	7.55
6	东台市	116.24	135.08	36.02	10.60	85.39	8.20
7	如东县	109.36	102.57	36.80	11.62	83.68	8.87
8	沭阳县	174.54	87.05	21.35	9.15	28.40	3.81
9	邳州市	158.04	92.63	23.79	8.66	40.24	5.70
10	海安县	96.54	93.54	26.44	11.55	101.76	8.51
11	沛 县	119.50	88.60	18.53	8.84	53.51	5.06
12	姜堰市	90.70	86.33	31.51	10.95	76.40	8.62
13	射阳县	104.70	98.15	25.50	9.60	46.48	5.90
14	阜宁县	105.00	73.50	19.70	9.25	40.81	3.85
15	大丰市	73.37	90.60	24.25	9.88	53.33	6.31
16	宝应县	91.99	67.86	20.35	9.75	47.38	4.83
17	建湖县	79.12	81.22	23.26	9.51	46.21	5.82
18	东海县	114.35	58.28	18.24	8.24	28.04	3.08
19	高邮市	83.06	70.51	20.95	10.20	51.59	5.56
20	赣榆县	107.91	61.78	19.29	9.58	27.47	3.02
21	丰 县	109.00	54.20	16.80	8.25	28.85	2.53
22	滨海县	108.00	56.70	14.60	9.39	27.19	3.00
23	盐都县	85.00	90.60	22.00	9.78	12.75	5.01
24	仪征市	59.35	72.40	29.00	14.56	52.35	11.25
25	泗 洪	103.00	58.90	12.30	7.92	22.06	3.36
26	新沂市	95.60	54.20	17.80	8.31	26.15	3.33
27	涟水县	103.00	52.60	14.70	8.03	19.41	2.51
28	灌云县	107.23	40.02	14.51	7.95	19.85	1.97
29	扬中市	27.24	80.11	19.18	13.12	51.22	8.31
30	盱眙县	73.22	58.85	13.99	10.00	16.47	3.08
31	溧水县	40.23	61.26	19.71	13.97	22.23	6.77
32	灌 南	72.71	29.47	9.86	7.95	11.58	2.06
33	响水县	57.00	29.47	8.97	8.94	15.36	2.04
34	金湖县	36.04	31.45	10.40	9.85	17.56	2.73
35	洪泽县	39.00	25.30	11.10	10.25	14.30	2.24

已调查 35 个城市的总人口、生产总值、消费总额、人均年工资、年度总储蓄额、年度财政总收入等数据, 试对上述城市进行分类研究。

1. 因子分析:

- 选用 Analyze Data Reduction Factor
- 引入因子分析的 6 个变量(总人口、生产总值、消费总额、人均年工资、年度总储蓄额、年度财政总收入)
- 提取公因子的方法(Method): 主成分分析法
- 提取(Extract) 可选: 提取特征值大于 1 的因子
- 旋转(Rotation) 的方法: 方差最大正交旋转
- 因子得分(Factor Scores): 作为新变量存入

表 1 方差解释表(Total Variance Explained)

因子	特征值	贡献率	累计贡献率
1	4.08	68%	68%
2	1.43	24%	92%

我们从原始表格中的 5 个变量中提取出 2 个公因子, 用这 2 个公因子就可以解释原始表格中 92% 的信息量

表 2 旋转后的因子负荷矩阵(Rotated Component Matrix)

	因子 1	因子 2
	消费因子	人口因子
总人口		0.914
生产总值	0.944	
消费总额	0.973	
人均年工资	0.630	
年度总储蓄额	0.934	
年度财政总收入	0.945	

根据原始表格中的 5 个变量与新生的 2 个公因子之间的相关程度, 我们为这 2 个公因子取名为消费因子和人口因子

2. 聚类分析:

- 选用 Analyze Classify K- Means Cluster
- 引入聚类分析的 2 个变量(即上面的 2 个公因子)
- 聚类的数目(Number of Clusters): 3 类
- 聚类方法(Method): 仅分类
- 储存新变量(Save New Variables): 聚类成员

表 3 各类数量分布表(Number of Cases in each Cluster)

类别	数量
1	12
2	16
3	7

3. 均值多重比较:

- 选用 Analyze Compare Means One- Way ANOVA
- 将 2 个因子移入因变量, 3 个类移入“Factor”
- 多重比较方法(Multiple Comparisons): 邓肯法 Duncan

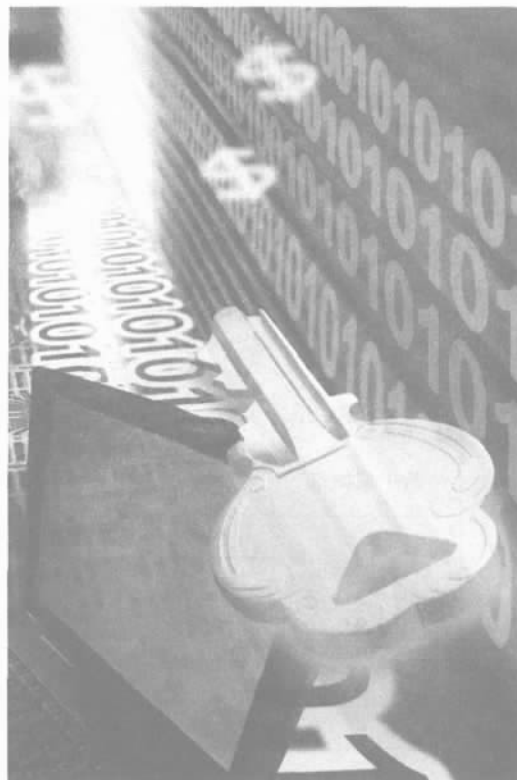




表4 3个类对于因子1的重视程度比较

Cluster Number of Case	Subset for alpha = 0.05	
	1	2
1	-0.5466864	1.4215031
2	-0.2118928	
3		

简化 →

类别	对于因子1的重视程度	
	1	2
1	B	A
2	B	
3		

注:A表示最重视,B表示一般

表5 3个类对于因子2的重视程度比较

Cluster Number of Case	Subset for alpha = 0.05	
	1	2
1	-0.7803180	0.7670201
2	-0.4155008	
3		

简化 →

类别	对于因子2的重视程度	
	1	2
1	B	A
3	B	
2		

注:A表示最重视,B表示一般

4.综合

	第1类	第2类	第3类
	折中型城市	人口型城市	消费型城市
消费因子	B	B	A
人口因子	B	A	B

注:A表示最重视,B表示一般

根据不同类别对于2个公因子的重视程度划分等级,我们将被调查城市划分为3类并分别为其命名:折中型城市、人口(突出)型城市、消费(突出)型城市。



折中型城市	大丰市	宝应县	建湖县	高邮市	盐都县	金湖县
	扬中市	盱眙县	溧水县	灌南	响水县	洪泽县
人口型城市	如皋市	兴化市	东台市	沭阳县	邳州市	沛县
	射阳县	阜宁县	东海县	赣榆县	丰县	滨海县
	泗洪	新沂市	涟水县	灌云县		
消费型城市	启东市	江都市	丹阳市	如东县	海安县	姜堰市
	仪征市					