



第三讲 因子分析

主要内容

- 1. 因子分析的原理
 - 因子分析的模型
 - 因子分析的步骤
- 2. 求解初始因子
 - 主成分分析法
 - 公因子分析法
 - 因子求解方法对结果的影响
- 3. 因子的解释与应用
 - 正交旋转
 - 斜交旋转
 - 旋转方法的选择
 - 解释因子
 - 因子值及其应用
- 4. 实用举例

西安交通大学管理学院 2008秋 胡平

第三讲 因子分析

3.4 因子分析的实际应用

- 1. 应用的领域
- 2. 几种常用统计软件包的应用举例

西安交通大学管理学院 2008秋 胡平

第三讲 因子分析

3.4.1 应用的领域

- 市场调研和营销
- 组织与人力资源管理
- 经济与金融行为分析
- 工程管理实验
- 信息管理(数据挖掘)
- 社会科学(人口, 心理, 女性)

西安交通大学管理学院 2008秋 胡平

3.4.1 应用的领域

市场调研和营销

- 1. 消费者使用习惯和态度研究 (U&A)

在U&A研究中, 对消费者对产品的态度探查往往需要使用因子分析, 探查影响消费者产品态度的基本因子, 并在此基础上, 利用各因子进行聚类分析对消费群进行细分, 从而达到市场细分的目的。在这里, 基本因子在消费群的细分中具有基础性的作用。

例如, 选取有关啤酒饮用的20个态度语句, 采用5级量表法 (1-非常不满意, 5-非常满意) 对消费者进行了产品态度调查, 因子分析的结果给出了影响产品态度的6个态度因子: 1) 男子气概; 2) 品牌转换; 3) 有吸引的额外利益的啤酒; 4) 独自在家饮; 5) 社会认可; 6) 品牌试验者/社交饮用者。

再如, 因子分析的结果表明, 新车的购买者主要关注购车因子包括经济性、方便性、性能、舒适度和豪华性五个方面, 由此可以产生五个细分市场。

西安交通大学管理学院 2008秋 胡平

3.4.1 应用的领域

- 2、生活形态研究 (life-style)

对生活方式 (life-style) 的研究中不可避免的要用到因子分析技术。目前生活方式的研究主要使用功能句子测量技术, 这些句子数量众多而且具有较强的相关性, 无论从数据本身还是从实际意义出发都适合于因子分析。
- 3、品牌形象和特性研究 (brand-image&attributes)

产品调研中, 经常需要品牌形象和产品属性。在这一方面, 因子分析同样可以发挥重要的作用。例如。顾客是如何评价银行的呢? 对15个有关银行特性的变量的因子分析结果表面, 传统服务 (包括贷款利率、社会声望等)、方便性 (包括服务速度、ATM机的位置等)、可见度 (包括亲友推荐、社会活动的参与度等) 以及能力 (包括雇员能力、辅助服务可得性等) 是评价银行时所使用的四个评价因子。
- 4、满意度研究 (CSR)

满意度研究中同样可能需要使用因子分析对顾客态度进行探查, 以寻求影响顾客满意度评价的基础因子。
- 5、营销渠道研究
- 6、客户关系管理研究

西安交通大学管理学院 2008秋 胡平

市场细分常用方法： 因子分析（Factor Analysis）

- 1) 因子分析是根据其相关程度将多个实测变量(Variables)转换成相互之间互不关联或关联性较低的综合指标(Dimension)的统计方法
- 2) 其主要目的是将多个实测变量所测试到的方方面面的杂乱的信息转化成更加有意义的、易于解释的、有实践运用意义的指标
- 3) 因子分析常用于消费者行为研究等各个方面，在市场营销策划中有着广泛的运用，如用于市场细分等
- 4) 基本上所有的统计软件都有因子分析这个功能，如 SAS、SPSS、SYSTAT、MINITAB等

西安交通大学管理学院 2008秋 胡平

市场细分案例：网民网上购物行为分析

传统购物习惯

针对您在日常生活中的购物习惯（非网上购物），请就下列叙述是否符合您实际购物情况进行打分（1-完全不符合，2-不符合，3-无所谓或难判断，4-符合，5-完全符合）

Name	Width	Label	Values
q6_1	11	0 上网地点	1, 单位 1...99
q6_2	11	1 上网时间	1, 单位 1...99
q6_3	11	2 折扣等促销活动是吸引我购买的主要原因	1, 完全不符合99
q6_4	11	3 对于非常高档商品，我总是等到降价时购买	1, 完全不符合99
q6_5	11	4 有吸引力的促销活动常让我购买一些平时不喜欢的商品	1, 完全不符合99
q6_6	11	5 我在选购商品时最注重商品的品质	1, 完全不符合99
q6_7	11	6 去商店购物是一种享受	1, 完全不符合99
q6_8	11	7 去商店购物时我总是先浏览一下网上的商品	1, 完全不符合99
q6_9	11	8 我愿意为名牌商品多花些钱	1, 完全不符合99
q6_10	11	9 我喜欢在空闲时与家人或朋友一起上街购物	1, 完全不符合99
q6_11	11	10 逛商店、购物对于我而言是一种休闲娱乐活动	1, 完全不符合99
q6_12	11	11 我喜欢在网上浏览商品	1, 完全不符合99
q6_13	11	12 我去购物时总是先浏览一下网上的商品	1, 完全不符合99
q6_14	11	13 我喜欢先在多家商店选择、比较，而后决定是否购买	1, 完全不符合99
q6_15	11	14 我常在网上浏览商品	1, 完全不符合99
q6_16	11	15 我喜欢在网上浏览商品	1, 完全不符合99
q6_17	11	16 我对新技术、新产品很感兴趣	1, 完全不符合99
q6_18	11	17 我常在网上浏览商品	1, 完全不符合99
q6_19	11	18 我常在网上浏览商品	1, 完全不符合99
q6_20	11	19 我常在网上浏览商品	1, 完全不符合99
q6_21	11	20 我常在网上浏览商品	1, 完全不符合99
q6_22	11	21 我常在网上浏览商品	1, 完全不符合99
q6_23	11	22 我常在网上浏览商品	1, 完全不符合99
q6_24	11	23 我常在网上浏览商品	1, 完全不符合99
q6_25	11	24 我常在网上浏览商品	1, 完全不符合99

市场细分案例：网民网上购物行为分析

	q6_1	q6_2	q6_3	q6_4	q6_5	q6_6	q6_7	q6_8	q6_9
1	4	5	4	4	4	1	5	4	4
2	3	3	3	3	4	3	4	3	3
3	3	4	1	4	3	4	3	4	2
4	4	3	3	4	5	3	4	5	3
5	3	3	5	2	3	2	5	5	3
6	4	3	4	4	4	2	4	3	4
7	3	3	2	4	3	4	4	3	3
8	3	4	2	5	5	2	4	3	2
9	3	2	1	2	3	4	5	4	3
10	5	3	5	3	3	5	3	2	2
11	4	5	5	5	3	5	4	3	3
12	3	5	4	3	1	2	2	2	2
13	4	4	2	4	3	4	4	4	3
14	5	4	4	3	1	5	3	1	1
15	4	4	2	4	1	3	3	4	4
16	3	5	4	2	2	4	3	2	4
17	4	5	2	4	4	2	3	5	6
18	3	2	2	4	4	2	4	4	2
19	4	1	5	4	3	2	3	3	3
20	3	2	5	5	5	3	5	4	3
21	1	1	1	1	1	1	1	1	1
22	4	4	3	4	4	4	4	4	3
23	3	2	3	4	3	4	4	3	3

西安交通大学管理学院 2008秋 胡平

市场细分案例：网民网上购物行为分析

对网上购物者的日常购物行为进行的因子分析表明，他们在日常网下购物中追求的主要利益可以归纳为：

- 寻求购物乐趣 (Pleasure)
- 关注商品的品质 (Brand)
- 喜爱新技术含量高的产品 (New-Tech.)
- 冲动购物 (Impulsive Buy)
- 便宜 (Price-Sensitivity)
- 购物的便利 (Convenience)

西安交通大学管理学院 2008秋 胡平

市场细分案例：网民网上购物行为分析

	1	2	3	4	5	6
折扣等促销活动是吸引我购买的主要原因	5.524E-02	5.527E-02	2.831E-02	-.126	.761	-4.06E-02
对于非常高档商品，我总是等到降价时购买	-.3.78E-02	3.369E-04	-1.30E-02	-.177	.724	5.317E-02
有吸引力的促销活动常让我购买一些平时不喜欢的商品	.171	7.540E-02	-.5.09E-03	.552	.494	2.309E-02
我在选购商品时最注重商品的品质	3.053E-02	.768	.146	-.4.60E-02	.145	2.073E-02
去商店购物是一种享受	.750	.191	-.8.69E-03	2.949E-02	2.829E-02	-.104
去商店购物时我总是先浏览一下网上的商品	-.3.99E-02	-.6.54E-02	3.210E-02	-.8.87E-02	4.537E-02	.695
我愿意为名牌商品多花些钱	.115	.829	8.401E-02	3.129E-02	-1.60E-02	7.318E-02
我喜欢在空闲时与家人或朋友一起上街购物	.805	9.133E-02	2.902E-02	3.537E-02	3.878E-02	-.5.76E-02
逛商店、购物对于我而言是一种休闲娱乐活动	.843	5.228E-02	2.348E-02	7.180E-02	4.242E-02	-.170
我喜欢购买名牌商品	.181	.823	.103	6.503E-03	-2.59E-02	-1.58E-02
我去购物时总是先浏览一下网上的商品	1.190E-02	.111	.118	-.581	.107	.430
我喜欢先在多家商店选择、比较，而后决定是否购买	.215	1.364E-02	.106	-.598	.408	2.645E-02
我常在网上浏览商品	-.7.78E-02	9.609E-02	1.065E-02	.194	-.117	.699

西安交通大学管理学院 2008秋 胡平

3.4.1 应用的领域

- 组织与人力资源管理
例如：影响雇员流动的因素分析；组织公平度的测量；工作岗位分析和设计中员工的业绩评价；项目管理与评级分析；
- 经济与金融行为分析
例如：股市业绩的评估，影响投资行为的因素分析；城乡居民消费支出分析；楼盘开发因素分析；城市竞争力比较研究；投资环境的分析
- 工程和管理实验
例如：人因工程设计中计算机的使用对员工疲劳的影响分析，计算机对员工心理压力的研究；影响企业家决策行为的因素分析；

3.4.1 应用的领域

- 信息管理(数据挖掘)
例如:影响知识产权保护的因素分析;
供应链管理中供应商评价因素分析;
电子商务数据仓库中数据挖掘分析;
- 社会科学(人口, 心理, 女性, ...)
例如:老年人的生活满意程度, 生活自理能力;
社会性别与生殖健康评价;育龄妇女生育健康
需求评估;
影响农村城镇化的因素分析;
城市流动人口迁移因素分析

3.4.2 几种常用统计软件包应用举例

- 案例1. 新兴股市的多因素模型
SPSS→Analysis Factor
- 案例2. 35家上市公司评价
Stata的命令句法
Factor
-rotate
-score
-greigen
- 案例3. 1~7月的平均气温
洛杉矶12个地区统计的五个社会经
济指标
- SAS/STAT的FACTOR过程

案例1：新兴股市的多因素模型①

- **模型的建立** 影响股票价格的因素很多, 从长期观点来看, 普遍为经济学家承认的宏观经济因素有国民生产总值 (GNP) 或国内生产总值 (GDP)、通货膨胀率、汇率、利率和失业率, 许多国家的公司 (如日本的大和公司等) 运用这 5 个指标 (依次记为 X1、X 2、X 3、X 4、X 5) 的线性模型作为股票收益率的预测模型, 其具体形式为:
$$Y = \alpha_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \varepsilon$$

其中 $\varepsilon \sim N(0, 2\sigma)$, Y 为预期收益率, 若用 P_t 表示 t 时刻的股票价格, 那么 $Y_{1+t} = (P_{1+t} - P_t) / P_t$ 就是 t+1 时刻的收益率。对股票价格的影响除了上述 5 个宏观因素外, 还有一些微观因素, 诸如上市公司的数量、发行量和交易量。

①何基报、茆诗松:《影响新兴股市的多因素模型及与中国股市的比较》,《统计与信息论坛》,1997年第3期。

案例1：新兴股市的多因素模型

综合这二类因素, 考虑如下股票价格模型:

$$P = \alpha_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \varepsilon$$

其中 $\varepsilon \sim N(0, 2\sigma)$, P 代表股票价格指数, X 1、X 2、X 3 分别为上市公司数量、发行量、交易量, X4、X5、X6、X7 分别为GNP (或国内生产总值GDP)、通胀率、对美元的汇率、利率。前 3 个是微观因素, 后 4 个是宏观因素。

案例1：新兴股市的多因素模型

- **模型应用**
利用模型来对一些主要的新兴股市发展情况进行分析。在运用中X4 表示GDP, X5、X7 分别是累积通货膨胀率、累积利率, 其计算如下:
- 设1984 年、1985 年的通货膨胀率 (或一年期存款利率) 分别为 $1\ a$ 、 $2\ a$, 那么以1983 年为基期, 1984 年、1985 年的 x_5 (或 x_7) 分别为 $1+a_1$ 、 $(1+a_1)$ 、 $(1+a_2)$, 对以后的年份以此类推得 x_5 (或 x_7)。选了14 个具有代表性的新兴股市, 分别是:
拉美: 智利 哥伦比亚 墨西哥 委内瑞拉
亚洲: 印度 韩国 马来西亚 巴基斯坦 菲律宾
中国台湾 泰国
欧洲: 希腊 葡萄牙
非洲: 尼日利亚

案例1：新兴股市的多因素模型

- 在估计参数时, 以年为单位, 诸 x_{ij} 取年平均值, 对每一个新兴股市都取1984~1993 年共10 年的数据, 按年顺序排号为: $(P_j, x_{j1}, x_{j2}, \dots, x_{j7})$ $j = 1, \dots, 10$, 其中 P_j 是1983 + j 年的股票指数的年平均值。为消除量纲的影响, 对每个新兴股市的数据进行如下变换:

$$P_j' = \frac{P_j}{P_1}, X_j' = \frac{X_j}{X_n} \quad j=1, \dots, 7, j=1, \dots, 10 \quad (13-134)$$

案例1：新兴股市的多因素模型

- 对每一新兴股市，用经过（13-134）式变换后的数据去拟合模型（13-133），其计算结果见表13-82第Ⅰ栏。表13-82中第Ⅱ栏、第Ⅲ栏分别列出了用经过（13-134）式变换后的微观因素数据、宏观因素数据拟合每个新兴股市股价线性模型时的一些计算结果。表13-83列出了每个新兴股市的股价指数P与t个变量的相关系数。从表13-82可以看出，模型（13-133）基本上适合于这14个新兴股市。
- 从股价指数与7个变量的复相关系数平方 R^2 来看，除菲律宾为0.7709，台湾为0.8426外，都大于0.90。
- 从模型检验统计量F值来看，除了巴基斯坦、菲律宾、台湾、葡萄牙分别为5.589、0.961、1.530、2.767外，其余都大于临界值 $F_{1.0}(7, 2)=9.35$ 。

西安交通大学管理学院 2008秋 胡平

案例1：新兴股市的多因素模型

- 利用表13-82、表13-83的结果分析新兴股市的发展情况。
1. 新兴股市既受宏观因素影响，又受微观因素影响。从表13-82中第Ⅱ、第Ⅲ栏中可以看出，有7个新兴股市的

$R^2_{\text{宏观}} > 0.9$ ，而 $R^2_{\text{微观}} > 0.9$ 只有6个，另外 $R^2_{\text{宏观}} = 0.4348 > R^2_{\text{微观}} = 0.2542$ ，

- 因此，总的说来，微观因素对新兴股市的影响程度比宏观因素似乎要强。

$R^2_{\text{宏观}} = 0.9777$ ， $R^2_{\text{微观}} = 0.4348$ $R^2_{\text{宏观}} = 0.9919$ ， $R^2_{\text{微观}} = 0.2542$ ，

- 这又表明和微观因素相比，宏观因素对各新兴股市的影响程度参差不齐，相差悬殊，微观因素对各新兴股市的影响程度差别相对较小，对新兴股市的影响更具有普遍性。

西安交通大学管理学院 2008秋 胡平

案例1：新兴股市的多因素模型

2. 股市的成熟性比较。

股票市场的发展是经济发展的“晴雨表”，因此，好的股票市场的运行应和宏观经济运行相一致。一般认为，如果股市运行和国民经济的运行有比较稳定的相关关系，那么该股市基本上是成熟的。如何反映这种稳定的相关关系？反映宏观经济运行好坏的指标有许多，而最能综合反映国民经济运行的指标是GNP或GDP，如用股价指数P作为衡量股市运行的指标，用GNP或GDP作为衡量国民经济运行的指标，

西安交通大学管理学院 2008秋 胡平

案例1：新兴股市的多因素模型

- 那么就可以用P与 x_4 （即GDP）的相关系数 ρ 度量这种“稳定的相关关系”。取 $\alpha=0.05$ ，相关系数检验临界值 $\rho_0=0.6319$ ，这样就可把新兴股市分为二类：

一类趋于成熟（ $\rho \geq \rho_0$ ），它们为智利、印度、墨西哥、尼日利亚、泰国、菲律宾和委内瑞拉。

另一类为不成熟股市（ $|\rho| < \rho_0$ ），它们为哥伦比亚、希腊、韩国、马来西亚、巴基斯坦、台湾和葡萄牙。

从股价指数P与GDP的相关系数可以看出，新兴股市的运行与国民经济的运行表现出参差不齐。两极分化现象严重。成熟度高的股市（如智利、印度、墨西哥和尼日利亚）与GDP的相关系数为 $\rho > 0.96$ ，股市的总体表现和国民经济的运行比较一致。表现比较差的股市（如韩国）与GDP的相关系数为 $\rho < -0.60$ ，股市的运行和经济的发展不能很好地接轨。

西安交通大学管理学院 2008秋 胡平

案例1：新兴股市的多因素模型

国家	P与T个变量复相关 R^2	检验统计量F值	P与微观因素(R1)2	检验F值	P与宏观因素(R2)2	检验F值
和地区				F 05.0 = 4.76		F 05.0 (4, 5) = 5.19
		F 1.0 (7, 2) = 9.35		F 1.0 (3, 6) = 3.29		F 1.0 (4, 5) = 3.52
智利	0.9981	148.1	0.9195	22.84	0.9734	45.72
哥伦比亚	0.9765	11.87	0.4384	1.561	0.9259	15.63
希腊	0.989	25.68	0.9777	87.78	0.857	7.49
印度	0.9928	39.46	0.9226	23.85	0.9919	153.1
韩国	0.9884	24.26	0.8373	10.29	0.9304	16.71
马来西亚	0.9736	10.53	0.9597	47.6	0.8722	8.53
墨西哥	0.9958	68.25	0.9089	19.94	0.9898	112.7
尼日利亚	0.9922	36.18	0.9739	74.62	0.9898	121.1
巴基斯坦	0.9514	5.589	0.7888	7.471	0.5701	1.658
菲律宾	0.7709	0.961	0.6708	4.075	0.7469	3.688
泰国	0.9727	10.2	0.9352	28.87	0.8728	8.577
中国台湾	0.8426	1.53	0.6824	4.296	0.7239	3.278
委内瑞拉	0.9743	10.85	0.8541	11.7	0.8396	6.544

西安交通大学管理学院 2008秋 胡平

案例1：新兴股市的多因素模型

国家和地区	P与上市公司数量 X_1	P与股票发行量 X_2	P与股票交易量 X_3	P与GDP X_4	P与通胀率 X_5	P与汇率 X_6	P与利率 X_7	P与股市规模（市价总值/GDP）
智利	0.7773	0.796	0.3949	0.9796	0.9702	0.9168	0.979	0.9821
哥伦比亚	-0.1891	-0.1399	0.2072	0.1866	0.2329	0.1759	-0.0519	-0.5223
希腊	0.92	0.9871	0.4402	0.9048	0.8334	0.8672	0.731	0.9273
印度	0.9605	0.035	0.9691	0.9743	0.9953	0.9775	0.9929	0.8972
韩国	0.7557	0.6947	-0.2051	0.6406	-0.743	0.7156	0.9478	0.7802
马来西亚	0.9111	0.9366	-0.6031	0.8865	0.1937	0.8942	0.9753	0.9263
墨西哥	0.4507	0.909	-0.1069	0.9684	0.9151	0.7612	0.9244	0.972
尼日利亚	0.9851	0.6215	0.698	0.9908	0.9676	0.9821	0.9942	0.7388
巴基斯坦	-0.2713	-0.1564	-0.1862	0.3002	0.321	0.4004	0.3207	0.3335
菲律宾	0.8059	0.6748	0.506	0.825	0.8318	0.7597	0.8425	0.9917
中国台湾	0.5313	0.6059	0.6387	0.5772	0.4582	-0.77	0.5515	0.9886
泰国	0.898	0.914	0.7019	0.907	0.8938	-0.1161	0.9029	0.9857
委内瑞拉	0.4627	0.6588	0.9132	0.8471	0.8408	0.834	0.7698	0.9507
葡萄牙	0.8165	0.6104	0.4032	0.4615	0.4898	-0.1468	0.4977	0.9696

西安交通大学管理学院 2008秋 胡平

案例1：新兴股市的多因素模型

- 3. 股票价格指数与宏观因素诸指标的关系。
- 在股价指数 P 与通胀率的相关系数中，除哥伦比亚、巴基斯坦、台湾、葡萄牙都小于 0.5 外，其余10个股市中有 9 个股市的相关系数（韩国为0.6406）都大于0.8，这表明这些股市股价的上涨很大程度上受物价的上涨的影响。股价指数 P 与汇率的相关系数中，韩国、台湾为最小，分别为-0.7430、-0.7700，表明了这两个亚洲股市受美元的负影响比较大。在其余的12个股市中有 7 个股市的相关系数大于0.75，剩下 5 个股市的相关系数都小于0.45，这说明了新兴股市受汇率的影响大小不一，呈现出两极分化趋势。由于 X 7 是累积利率，呈递增趋势，而股价指数在股市发展正常时一般也有增长趋势，因此二者呈现正相关。

西安交通大学管理学院 2008秋 胡平

案例1：新兴股市的多因素模型

- 在成熟的股市，长期说来，二者的相关性较强，但从表13-83中可以看出，哥伦比亚、巴基斯坦、台湾、葡萄牙这些股市中，二者的相关系数不大，说明了二者的增长幅度和趋势相差很大，这几个股市的利率波动不大，因此累积利率 x 7 稳定增长，这就说明了这几个股市股价暴涨暴跌现象严重，并没有一种稳定的增长趋势，而且短期投机者比较，使得股价大起大落。
- 4. 股价指数与微观因素诸指标的关系。
- 在股价指数 P 与上市公司数量 X 1 的相关系数中，有 9 个是大于0.77 的，在 P 与股票发行量 X 2 的相关系数中，也有 9 个是大于0.65 的，在 P 与交易量 X 3 的相关系数中也有 一半是大于0.6319 的，这说明了虽然新兴股市的规模在不断扩大，但对股价的稳定所起的作用不大，在表13-83中的股价指数 P 与股市规模的相关系数中，有10个大于0.95，这更进一步地说明了新兴股市的股市规模的扩大并没有缓解股市的供需紧张局面，投资大众的热情很高，股市发展规模亟需进一步扩大。

西安交通大学管理学院 2008秋 胡平

案例1：新兴股市的多因素模型

国家和地区	常数项	回归方程系数							F 值
		X1	X2	X3	X4	X5	X6	X7	
智利	-10.45	0	0	0	6.966	0	0	0	190.07
哥伦比亚	回归方程不显著								
希腊	0.8785	0	0	0.5132	0	0	0	0	304.6
印度	-6.434	0	0	0	0	0	7.596	0	838.79
韩国	-1.808	3.406	0	0	0	0	0	0	12.44
马来西亚	-4.574	0	0	0.1216	0	5.283	0	0	83.08
墨西哥	-100.81	0	93.478	0	16.88	0	-15.15	0	310.81
尼日利亚	-3.472	0	0	0	0	0	0	3.914	679.83
巴基斯坦	回归方程不显著								
菲律宾	-14.8	0	0	0	0	0	0	13.82	19.57
泰国	-1.196	0	3.048	-0.288	0	0	0	0	50.24
中国台湾	18.076	0	0	0	0	0	-17.472	0	11.65
委内瑞拉	0.224	0	0	2.899	0	0	1	0	40.14
葡萄牙	-2.864	3.028	-1.676	0	2.9128	0	0	0	11.85

西安交通大学管理学院 2008秋 胡平

案例1：新兴股市的多因素模型

- 5. 各个股市受宏观因素和微观因素影响的具体情况不同，找出影响每个股市的重要因素，对其以后的发展无疑是有帮助的。
- 用逐步回归方法找出影响各新兴股市的重要因素，表13-84列出了有关的计算结果（其中引入变量时作检验的 F 统计量 F 1 和剔除时的统计量 F 2 均为3.0）。除哥伦比亚、巴基斯坦不能挑出重要因素外，其他12 个股市挑出的重要因素中，X 3 有 4 个，X 2 、X 4 、X 6 各有 3 个，X 1 和 X 7 各有 2 个，X 5 有 1 个。没有一个因素超过 7 个，说明了虽然同是新兴股市，但不同的股市受不同的重要因素的影响不同。在这 7 个因素中，没有一个共同的影响这些新兴股市的重要因素，各新兴股市的具体情况相差很大，表现出明显的参差不齐，不象发达股市（如美国、英国）无一例外地受到重要因素 GNP、汇率、通胀率、利率的影响。另外，从表13-84中可以看出，大部分的新兴股市只能挑选出一个重要因素，说明了新兴股市在发展过程中没有注意到股市的均衡性，往往只受一个重要因素的左右，这样的股市容易出现大起大落现象。

西安交通大学管理学院 2008秋 胡平

案例1：新兴股市的多因素模型

用主成份分析这 7 个因素，找出几个综合指标，从而在分析时可用较少的变量来代替上述 7 个变量，给分析问题带来很大的方便，将 7 个指标作为自变量即 $X = (x_1, x_2, \dots, x_7)$ ，对每一新兴股市取1984~1993 年共10 年的年平均数据进行（13-134）式变换组成 $X(140, 7)$ 的样本矩阵，以相关阵出发来求主成份。由于样本矩阵 $X(140, 7)$ 数据太多，在这里只列出它们的相关阵 $X(7, 7)$ 如下：

西安交通大学管理学院 2008秋 胡平

案例1：新兴股市的多因素模型

	1	2	3	4	5	6	7
1	1	0.522	0.801	-0.038	-0.063	-0.092	0.01
2	0.522	1	0.692	0.09	0.059	0.024	0.55
3	0.801	0.692	1	0.04	-0.015	-0.085	0.282
4	-0.38	0.09	0.04	1	0.908	0.65	0.314
5	-0.063	0.059	-0.015	0.908	1	0.727	0.245
6	-0.092	0.024	-0.085	0.65	0.727	1	0.315
7	0.01	0.55	0.282	0.314	0.245	0.293	1

- 其特征值为： 2.750、2.442、0.946、0.387、0.254、0.145、0.077，特征值贡献率为： 39.3%、34.9%、13.5%、5.5%、3.6%、2.1%、1.1%，由于前3 个特征值累积贡献率达87.7%，故可取 3 个主成份，得到因子载荷阵如下：

西安交通大学管理学院 2008秋 胡平

案例1：新兴股市的多因素模型

因子（主成份）载荷阵A（7，3）

	1	2	3
1	-0.038	0.954	-0.078
2	0.002	0.644	0.656
3	-0.026	0.916	0.249
4	0.93	0.03	0.112
5	0.962	0	0.049
6	0.842	-0.096	0.13
7	0.217	0.045	0.939

3个主成份对各变量的方差贡献分别为：91.7%、84.5%、90.2%、87.9%、92.8%、73.4%和93.1%。以上3个主成份提取的信息量占87.7%，其中，在第一主成份对以上7个变量的因子载荷中，比较大的有第4、第5、第6分量，分别是0.930、0.962、0.842，因此第一主成份主要反映了GDP、通胀率、汇率3个变量，我们称之为宏观因子。

西安交通大学管理学院 2008秋 胡平

西安交通大学管理学院 2008秋 胡平

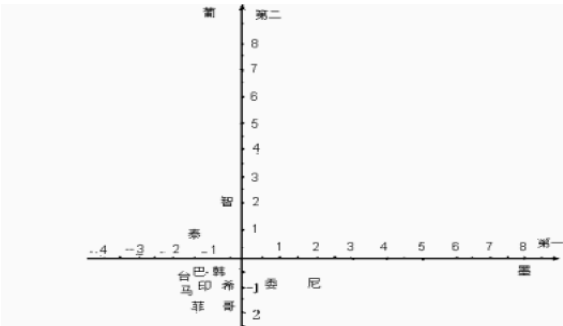
案例1：新兴股市的多因素模型

- 在第二主成份中，以葡萄牙为最大，其值为4.33，而第一主成份仅为-0.372，表明了葡葡股市受微观因素的影响比受宏观因素的影响大得多，而且股市和宏观经济运行联系不大，菲律宾和哥伦比亚的第二主成份最小，说明了相对于宏观因素而言，微观因素对股市影响不大，股市主要受宏观因素影响。亚洲几个股市的第一主成份比较接近，因此，大体上可分成如附图所示的6组，从图中可看出两点：1. 以上7个因素对新兴股市的影响很不均匀，呈两极分化状态，有的只受微观因素的影响（如葡萄牙），有的只是宏观因素起作用（如哥伦比亚），体现了新兴股市发展的不平衡性，这种现象令人担忧。2. 亚洲的几个股市比较靠近，呈板块结构。

西安交通大学管理学院 2008秋 胡平

西安交通大学管理学院 2008秋 胡平

案例1：新兴股市的多因素模型



第一、第二主成份点图

西安交通大学管理学院 2008秋 胡平

西安交通大学管理学院 2008秋 胡平

案例1：新兴股市的多因素模型

- 在第二主成份对以上7个变量的因子载荷中，比较大的有第1、第2、第3分量，分别是0.954、0.644、0.916，因此第二主成份主要反映了上市公司数量、发行量、交易量三个变量，我们称之为微观因子。
- 在第三主成份对7个变量的因子载荷中，比较大的有第2、第7分量，分别是0.656、0.939，因此第三个主成份主要反映了发行量和利率这两个变量，但更主要是反映了利率变量，我们称之为利率因子。通过以上分析我们知道，在众多的影响因素中，宏观因子、微观因子、利率因子是影响新兴股市的主要的共同因素。
- 由于每个新兴股市有10个样本，共组成140个样本，计算经过(3)式变换后的140个数据的前二个主成份的值，对每一新兴股市，分别取属于该股市的10个数据的前二个主成份的均值，列于表13-85，并将它作成附图。从表13-85可以看到，在第一主成份中，墨西哥的数值为最大，而其第二主成份的值仅为-0.141，表明了宏观因素对墨西哥股市的影响比微观因素要大，以马来西亚和台湾的第一主成份的值为最小，均比第二主成份的值小，说明了这两个股市中，宏观因素对股市的总的影

西安交通大学管理学院 2008秋 胡平

案例1：新兴股市的多因素模型

国家和地区	第一主成份（均值）	第二主成分（均值）
智利	-0.129	0.767
哥伦比亚	-0.171	-0.682
希腊	-0.214	-0.86
印度	-0.524	-0.247
韩国	-0.509	-0.09
马来西亚	-0.739	-0.285
尼日利亚	0.865	-0.573
墨西哥	3.97	-0.141
巴基斯坦	-0.605	-0.143
菲律宾	-0.569	-0.633
中国台湾	-0.656	-0.156
泰国	-0.615	0.8
委内瑞拉	0.266	-0.554
葡萄牙	-0.372	4.33

西安交通大学管理学院 2008秋 胡平

案例1：新兴股市的多因素模型

- 我国股市与其他股市比较
- 我国股市起步晚，数据较少，还不能用模型来说明，只能从几个侧面来比较。
- 与发达股市比较。在表13-86中，加拿大、美国、英国股市的股价指数与GNP的相关系数都很大，这些股市与国民经济运行非常一致，我国股市1992~1995年的上证指数与GDP的相关系数为 $\rho = -0.3467$ ，呈负相关，说明了我国的股市发展还没有和宏观经济接轨。这几个发达股市股价指数与GNP、物价指数的复相关系数分别为0.998、1.000、1.000，从而可看出，发达股市与宏观经济是息息相关的。我国股市的上证指数与物价指数、GDP的复相关系数只有0.385，这表明我国股市的发展游离于宏观经济之外。从以上比较可以看出，我国股市与发达股市的差距还很大。

西安交通大学管理学院 2008秋 胡平

案例1：新兴股市的多因素模型

2. 与新兴股市比较。

我国股市的股价指数与G D P的相关系数为 $\rho = -0.3467$ ，与其他新兴股市相比，只比马来西亚的 $\rho = -0.6801$ 大，按 $\rho 0 = 0.6319$ 归类，应归入差的一类，说明了我国股市的总体表现和表现好的新兴股市是有一段差距的。另外，大部分的新兴股市既受宏观因素的影响，又受微观因素的影响，而我国股市受二者的影响不大。一方面表现在国民经济持经济持续稳定增长，但股票价格这几年却大起大落，股市波动频繁。另一方面，股票价格的波动与政策、扩容、消息有密切关系，一有风吹草动，马上就是草木皆兵，股市动荡不安，市场人士反映这几年我国股市是政策市、消息市。

西安交通大学管理学院 2008秋 胡平

案例1：新兴股市的多因素模型

	1986年	1987年	1988年	1989年	股票指数与 GNP相关系数	股票与物价 指数相关系数	复相 关系数R
加拿大多伦多 股票指数	105.7	108.9	116.9	136.9	0.9298	0.9525	0.998
伦敦金融时报 股票指数	116.1	121.4	128.6	169.4	0.8932	0.9536	1
美国道·琼斯 工业平均指数	122.6	125.4	140.2	178	0.9226	0.9453	1
上证指数 1990年=100	1992年 668.52	1993年 1013.4	1994年 674.1	1995年 660.8	与GNP相关系数 -0.3467	-0.384	0.385

资料来源：《证券市场导报》，《国际金融统计》。1995年上证指数为当年1-10月指数

西安交通大学管理学院 2008秋 胡平

案例1：新兴股市的多因素模型

- 通过以上分析可知，和发达股市及其他新兴股市相比，我国股市与国民经济的运行还没有很好地接轨。虽然我国股市的规模不断扩大，但由于发展较晚，使得股市规模相对于国民生产总值来说，显得太小（1994年流通股股票市值只占GNP的不足5%）。这样，股市在国民经济运行中所起的作用也小。因此，我国股市在以后的发展中要进一步扩大股市规模，同时要进行了规范化建设，使市场朝有效化方向发展。唯有这样，才能吸引更多的投资者，更好地保证股市规模进一步扩大，从而使股市与宏观经济运行更好地接轨。

西安交通大学管理学院 2008秋 胡平



西安交通大学管理学院 2008秋 胡平

案例2. 35家上市公司的评价

- 1999年财政部、国家经贸委、人事部和国家计委联合发布了《国有资本金绩效评价规则》。其中竞争性工商企业的评价指标体系包括8大基本指标：净资产收益率、总资产报酬率、总资产周转率、流动资产周转率、资产负债率、已获利息倍数、销售增长率和资本积累率。
- 借助这一指标体系对2000年35家（电力、煤气及水的生产和供应、房地产、信息技术等）中国上市公司的年报数据分析企业的运营情况。

西安交通大学管理学院 2008秋 胡平

案例2. 35家上市公司的评价

行业	公司简称	净资产收益率(%)	总资产报酬率(%)	资产负债率(%)	总资产周转率	流动资产周转率	已获利息倍数	销售增长率(%)	资本积累率(%)
电力、煤气及水的生产和供应	深能源A	16.85	12.35	42.32	0.37	1.78	7.18	45.73	54.54
	深南电A	22	15.3	46.51	0.76	1.77	15.67	48.11	19.41
	原水股份	8.85	8.88	36.2	0.13	0.41	8.53	-11.49	2.44
	大连热电	9.03	7.41	46.89	0.28	0.79	6.86	16.23	-1.52
	龙电股份	12.07	8.7	16.81	0.28	0.68	29.75	4.11	63.06
	华银电力	6.85	6.12	41.93	0.24	0.65	4.38	11.2	3.8
	长春经开	9.85	10.5	31.23	0.34	0.4	17.13	18.05	7.18
	兴业房产	1.07	1.52	66.91	0.21	0.24	1.53	-31.93	1.08
	ST中福	5.17	6.62	65.48	0.16	0.21	1.33	-19.91	23.74
	倍特高新	0.72	2.76	65.39	0.3	0.42	1.24	8.4	0.7
房地产业	三木集团	5.99	4.53	65.17	0.74	0.88	4.14	75.36	0.87
	寰岛实业	0.42	0.2	24.03	0.02	0.03	-8.18	-71.33	0.42
	中关村	9.32	4.48	67.76	0.32	0.37	16.42	-29.42	4.09
	中兴通讯	18.78	11.09	69.15	0.93	1.08	4.79	80.8	23.27
	长城电脑	14.94	9.48	45.53	1.14	1.85	9.51	34.47	35.93
信息技术业	永鼎光缆	9.4	8.67	32.75	0.79	1.25	13.49	41.75	6.33
	博图高科	14.57	7.96	65.86	0.76	0.94	3.95	54.45	15.71
	海星科技	4.06	3.35	36.49	0.48	0.6	4.64	-16.28	1.69
	方正科技	27.48	16.69	57.13	2.51	2.87	7.4	63.27	32.02
	复华实业	5.58	4.1	44.24	0.28	0.41	3.77	12.92	2.3

案例2. 35家上市公司的评价

主成分分析用SPSS操作算例演示

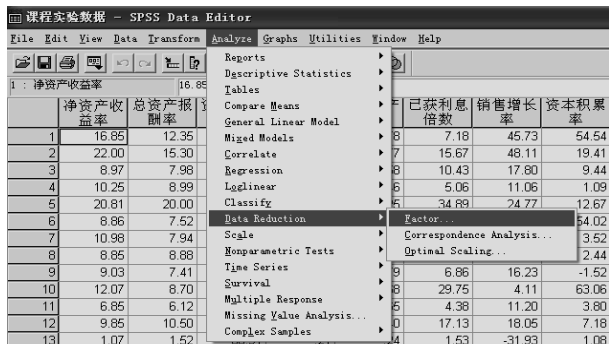
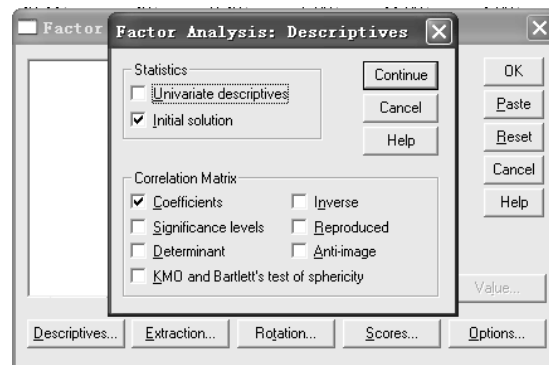


Figure 2-1: SPSS Data Editor showing the menu path: Analyze > Data Reduction > Factor...

- 打开spss数据窗口，在Analysis窗口选择Data Reduction，点击Factor。

西安交通大学管理学院 2008秋 胡平

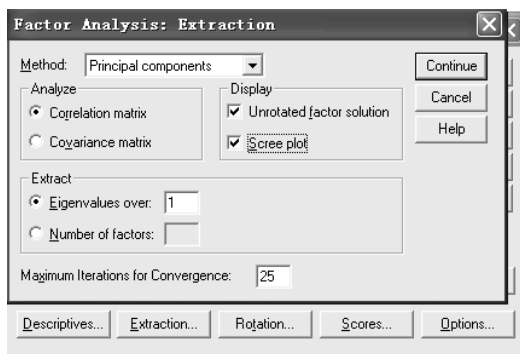
案例2. 35家上市公司的评价



- 点击进入Factor Analysis对话框下面的Descriptives...按钮进入Descriptives对话框，在下面的Correlation Matrix框架中选中Coefficients以输出原始变量的相关矩阵。

西安交通大学管理学院 2008秋 胡平

案例2. 35家上市公司的评价



- 进入Extraction对话框，在Display框架中选取Scree plot以输出碎石图帮助确定应保留主成分个数。

西安交通大学管理学院 2008秋 胡平

案例2. 35家上市公司的评价

Correlation Matrix

	净资产收益率	总资产报酬率	资产负债率	总资产周转率	流动资产周转率	已获利息倍数	销售增长率	资本积累率
净资产收益率	1.000	.891	-.054	.688	.721	.518	.651	.485
总资产报酬率	.891	1.000	-.158	.572	.708	.665	.528	.405
资产负债率	.054	-.158	1.000	.143	-.043	-.407	.161	-.281
总资产周转率	.688	.572	.143	1.000	.782	.142	.547	.342
流动资产周转率	.721	.708	-.043	.782	1.000	.272	.452	.389
已获利息倍数	.518	.665	-.407	.142	.272	1.000	.228	.458
销售增长率	.651	.528	.161	.547	.452	.228	1.000	.402
资本积累率	.485	.405	-.281	.342	.389	.458	.402	1.000

- 除了资产负债率和已获利息倍数与其他指标的相关性较低外，其余指标间的相关性均较强(>0.3),可进行主成分分析。

西安交通大学管理学院 2008秋 胡平

案例2. 35家上市公司的评价

Communalities

	Initial	Extraction
净资产收益率	1.000	.883
总资产报酬率	1.000	.830
资产负债率	1.000	.735
总资产周转率	1.000	.764
流动资产周转率	1.000	.715
已获利息倍数	1.000	.749
销售增长率	1.000	.584
资本积累率	1.000	.501

Extraction Method: Principal Component Analysis.

西安交通大学管理学院 2008秋 胡平

案例2. 35家上市公司的评价

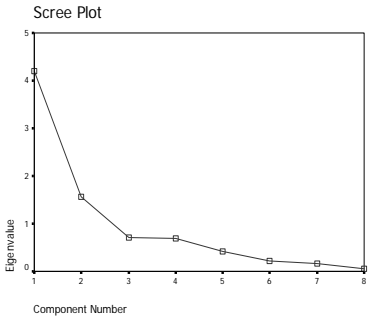
Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loading		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.202	52.531	52.531	4.202	52.531	52.531
2	1.560	19.500	72.030	1.560	19.500	72.030
3	.703	8.793	80.824			
4	.683	8.538	89.362			
5	.416	5.198	94.561			
6	.219	2.739	97.299			
7	.158	1.978	99.277			
8	.058	.723	100.000			

Extraction Method: Principal Component Analysis.

西安交通大学管理学院 2008秋 胡平

案例2. 35家上市公司的评价



西安交通大学管理学院 2008秋 胡平

- 从总方差分解表和碎石图可以看到保留2个或4个主成分 是合适的。
- 若以特征根大于1 为标准，则应保留 2个主成分；
- 若以提取的原始变量总方差大于85 % 为标准，则应保留 4个主成分。

案例2. 35家上市公司的评价

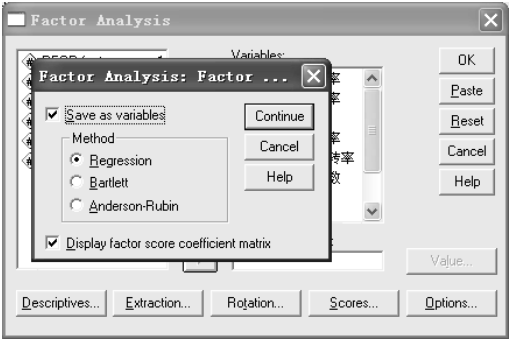
	Component	
	1	2
净资产收益率	.934	.105
总资产报酬率	.903	-.124
资产负债率	-.100	.851
总资产周转率	.771	.411
流动资产周转率	.825	.188
已获利息倍数	.594	-.630
销售增长率	.702	.302
资本积累率	.620	-.341

Extraction Method: Principal Component Analysis.
a. 2 components extracted.

- 给出了标准化的原始变量用主成分线性表示的近似表达

西安交通大学管理学院 2008秋 胡平

案例2. 35家上市公司的评价



- 如果在因子分析窗口中进入Factor Scores对话框，选中Save as variables复选框以输出各公司的主成分得分并存储为变量，选中Display Factor ScoreCoefficient Matrix复选框输出主成分得分矩阵。然后OK之后除了得到上面的输出的结果外，还可以得到下面两张结果表。

案例2. 35家上市公司的评价

	Component	
	1	2
净资产收益率 (%)	.222	.068
总资产报酬率 (%)	.215	-.080
资产负债率 (%)	-.024	.546
总资产周转率	.184	.263
流动资产周转率	.196	.121
已获利息倍数	.141	-.404
销售增长率 (%)	.167	.194
资本积累率 (%)	.148	-.219

Extraction Method: Principal Component Analysis.
Component Scores.

Component	1	2
1	1.000	.000
2	.000	1.000

Extraction Method: Principal Component Analysis.
Component Scores.

西安交通大学管理学院 2008秋 胡平

案例2. 35家上市公司的评价

- 按四个因子提取。
- 第一步：先计算因子载荷矩阵与特殊度。实际上这一部分工作已在上面的主成分分析中完成，按照主成分法提取公因子。输出结果如下所示：

	Communalities	
	Initial	Extraction
净资产收益率	1.000	.919
总资产报酬率	1.000	.943
资产负债率	1.000	.878
总资产周转率	1.000	.881
流动资产周转率	1.000	.911
已获利息倍数	1.000	.908
销售增长率	1.000	.799
资本积累率	1.000	.910

Extraction Method: Principal Component Analysis.

西安交通大学管理学院 2008秋 胡平

案例2. 35家上市公司的评价

Component	Total Variance Explained						
	Initial Eigenvalues			Extraction Sums of Squared Loadings			
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	
1	4.202	52.531	52.531	4.202	52.531	52.531	
2	1.560	19.500	72.030	1.560	19.500	72.030	
3	.703	8.793	80.824	.703	8.793	80.824	
4	.683	8.538	89.362	.683	8.538	89.362	
5	.416	5.198	94.561				
6	.219	2.739	97.299				
7	.158	1.978	99.277				
8	.058	.723	100.000				

Extraction Method: Principal Component Analysis.

西安交通大学管理学院 2008秋 胡平

案例2. 35家上市公司的评价

Component Matrix^a

	Component			
	1	2	3	4
净资产收益率	.934	.105	.002	.189
总资产报酬率	.903	-.124	-.174	.288
资产负债率	-.100	.851	.224	.306
总资产周转率	.771	.411	-.210	-.270
流动资产周转率	.825	.188	-.374	-.237
已获利息倍数	.594	-.630	.053	.395
销售增长率	.702	.302	.461	.042
资本积累率	.620	-.341	.473	-.430

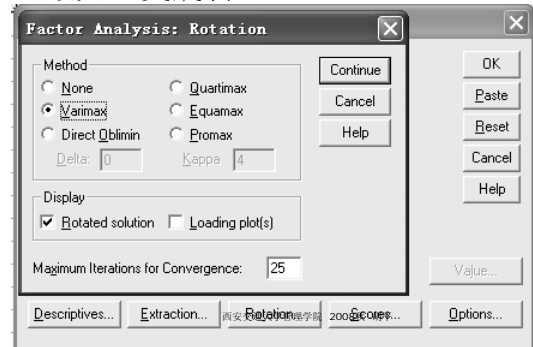
Extraction Method: Principal Component Analysis.

a. 4 components extracted.

西安交通大学管理学院 2008秋 胡平

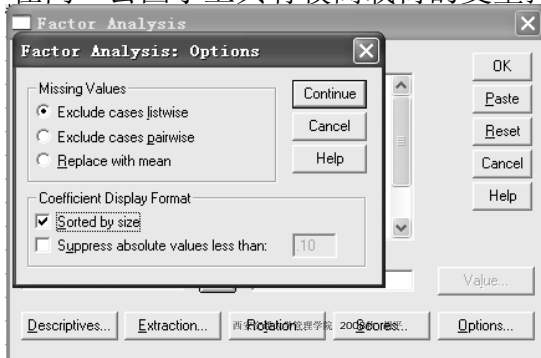
案例2. 35家上市公司的评价

- 第二步：因子旋转。在Factor Analysis对话框中点击Rotation...按钮进入Rotation对话框，选中Varimax进行方差最大正交旋转。



案例2. 35家上市公司的评价

- 在Options对话框中的Coefficient Display Format框架中选中Sorted by Size以使输出的载荷矩阵中各列按载荷系数大小排列，使在同一公因子上具有较高载荷的变量排在一



案例2. 35家上市公司的评价

- 点击OK后可得到如下输出结果：

Rotated Component Matrix^a

	Component			
	1	2	3	4
流动资产周转率	.919	.209	.137	-.066
总资产周转率	.895	.065	.224	.161
净资产收益率	.640	.615	.299	.203
已获利息倍数	.028	.881	.212	-.294
总资产报酬率	.586	.761	.140	-.008
资本积累率	.211	.195	.872	-.260
销售增长率	.369	.306	.592	.467
资产负债率	.044	-.193	-.131	.906

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 6 iterations.

西安交通大学管理学院 2008秋 胡平

案例2. 35家上市公司的评价

Component Transformation Matrix

Component	1	2	3	4
1	.708	.562	.425	.037
2	.381	-.421	-.149	.810
3	-.499	.018	.775	.387
4	-.322	.712	-.443	.440

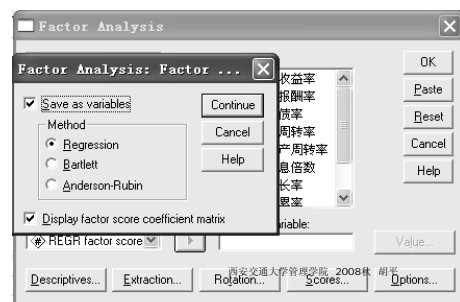
Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

西安交通大学管理学院 2008秋 胡平

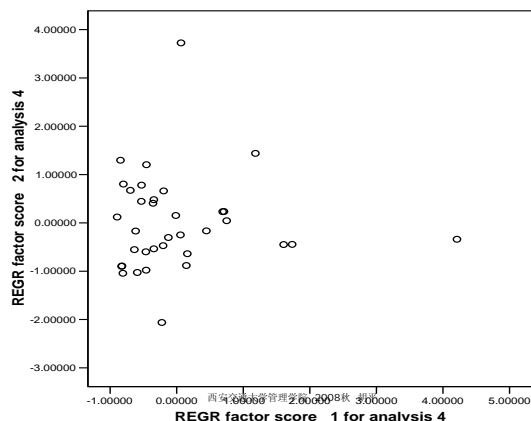
案例2. 35家上市公司的评价

- 第三步：因子得分。对原始八个指标提取公因子后，就可以通过分析少数几个公因子来对各上市公司进行比较研究了。在Factor Analysis对话框中点击Scores...按钮进入Factor Scores对话框，选中save as variables复选框输出因子得分并保存为变量，OK运行。可以看到在数据窗口中多了四个变量fac1_1, fac2_1, fac3_1, fac4_1,这几个变量的值是各公司相应公因子的得分。



案例2. 35家上市公司的评价

- 对第一和第二个公因子做散点图可得到如下输出结果：



因子分析应用的注意事项

■ 应用条件

(1) 变量是计量的，能用线性相关系数（Pearson积叉相关系数）表示。

(2) 总体的同质性

西安交通大学管理学院 2008秋 胡平

■ 样本量

没有估计公式。至少要保证样本相关系数稳定可靠。

■ 因子数目

一般认为，累积贡献要达到80%以上。但要注意Heywood现象。

西安交通大学管理学院 2008秋 胡平

案例2. 35家上市公司评价——Stata操作

■ 变量解释：

Js—流动资产周转率，zb—总资产报酬率，zf—资产负债率，zz—总资产周转率，lb—已获利率倍数，xz—销售增长率，zj—净资产收益率，

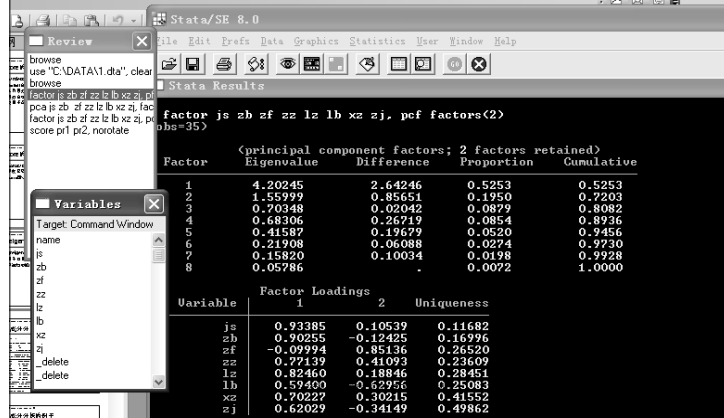
■ 1. 相关性分析

由相关性分析可知，除了资产负债率与已获利息倍数两项指标的相关性较低外，其他指标的相关性均较强（大于0.3），可以尝试主成分分析。

西安交通大学管理学院 2008秋 胡平

案例2. 35家上市公司评价——Stata操作

■ 2. 首先进行factor=2的因子分析



西安交通大学管理学院 2008秋 胡平

案例2. 35家上市公司评价——Stata操作

■ 标准化

$$zj \approx 0.93385 \times \text{prin1} + 0.10539 \times \text{prin2}$$

- **备注：如果我们并不选定factor=2, stata程序则自动提取公因子，根据本题中变量的数量以及变量之间的相互关系，最终，实际提取的公因子仍为2个。**

西安交通大学管理学院 2008秋 胡平

案例2. 35家上市公司评价——Stata操作

- 3. 采用score 命令输出主成分得分系数矩阵

core prin1 pr2, norotate			
Variable	Scoring Coefficients		
	1	2	
js	0.22221	0.06756	
zb	0.21477	-0.07965	
zf	-0.02378	0.54575	
zz	0.18356	0.26342	
lz	0.19622	0.12081	
lb	0.14135	-0.40357	
xz	0.16711	0.19369	
zj	0.14760	-0.21890	

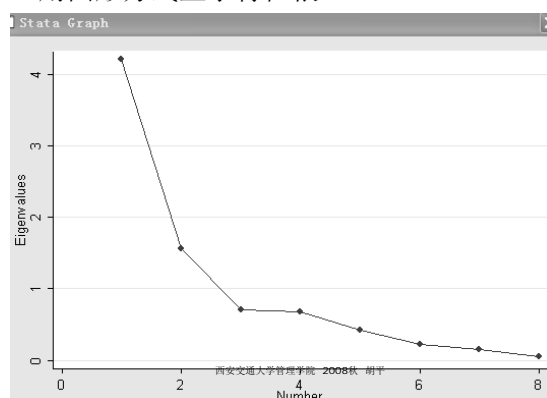
案例2. 35家上市公司评价——Stata操作

- 标准化prin1=标准化js×0.2221+标准化zb×0.21477-标准化zf×0.02378+标准化zz×0.18356+标准化lz×0.19622+标准化lb×0.14135+标准化xz×0.16711+标准化zj×0.14760

西安交通大学管理学院 2008秋 胡平

案例2. 35家上市公司评价——Stata操作

- 4. 用图形方式显示特征根



案例2. 35家上市公司评价——Stata操作

- 正交旋转(varimax)的结果

rotate, varimax factors(2) horst			
(varimax rotation)			
Variable	Rotated Factor Loadings		Uniqueness
	1	2	
js	0.27776	0.89779	0.11682
zb	0.47561	0.77707	0.16996
zf	-0.82003	0.24970	0.26520
zz	-0.06727	0.87143	0.23609
lz	0.15787	0.83100	0.28451
lb	0.81486	0.29184	0.25083
xz	0.00469	0.76450	0.41552
zj	0.56149	0.43140	0.49862

案例2. 35家上市公司评价——Stata操作

- 正交旋转后因子得分

. score prin1 prin2		
(based on rotated factors)		
Variable	Scoring Coefficients	
	1	2
js	0.02718	0.23066
zb	0.15906	0.16483
zf	-0.50952	0.19697
zz	-0.16775	0.27375
lz	-0.03203	0.22819
lb	0.42638	-0.03227
xz	-0.11046	0.23073
zj	0.25971	0.04748

案例2. 35家上市公司评价——Stata操作

- 斜交旋转（promax）结果：

rotate, promax(3.0) factors(2) horst			
(promax rotation)			
Rotated Factor Loadings			
Variable	1	2	Uniqueness
js	0.14503	0.89089	0.11682
zb	0.36668	0.74253	0.16996
zf	-0.87964	0.35678	0.26520
zz	-0.20462	0.90566	0.23609
lz	0.03250	0.83667	0.28451
lb	0.79005	0.20158	0.25083
xz	-0.11420	0.78679	0.41552
zj	0.50854	0.37609	0.49862

案例2. 35家上市公司评价——Stata操作

- 斜交旋转后因子得分

. score pr1 pr2			
(based on rotated factors)			
Scoring Coefficients			
Variable	1	2	
js	0.05401	0.23213	
zb	0.17727	0.18684	
zf	-0.48294	0.11825	
zz	-0.13453	0.24547	
lz	-0.00508	0.22080	
lb	0.41967	0.03210	
xz	-0.08267	0.21154	
zj	0.26349	0.08593	

案例3. 一月和七月平均气温的主成份分析
—— SAS操作

- 用PRINCOMP过程计算主成份分析
- SAS的PRINTCOMP过程有如下功能：
- 完成主成份分析。
- 主成份的个数可以由用户自己确定，主成份的名字可以由用户自己规定，主成份得分是否标准化可自己规定。
- 输入数据集可以是原始数据集、相关阵、协方差阵或叉积阵。输入为原始数据时，用户还可以规定从协方差阵出发还是从相关阵出发进行分析。由协方差阵出发时方差大的变量在分析中起到更大的作用。
- 计算结果有：简单统计量，相关阵或协方差阵，从大到小排序的特征值和相应特征向量，每个主成份解释的方差比例，累计比例等。可生成两个输出数据集：一个包含原始数据及主成份得分，另一个包含有关统计类型为TYPE=CORR或COV。
- 可揭示变量间的共线关系。若某特征值特别接近0说明变量线性相关，这时用这些变量作回归自变量可能得到错误的结果。

西安交通大学管理学院 2008秋 胡平

案例3. 一月和七月平均气温的主成份分析
—— SAS操作

- PRINCOMP主要使用PROC PRINCOMP语句与VAR语句。PROC PRINCOMP语句用来规定输入输出和一些运行选项，包括：
- DATA=输入数据集，可以是原始数据集，也可以是TYPE=CORR, COV的数据集
- OUT=输出包含原始数据和主成份得分的数据集
- OUTSTAT=统计量输出数据集
- COV 要求从协方差阵出发计算主成份。缺省为从相关阵出发计算。
- N=要计算的主成份个数。缺省时全算。
- NOINT 要求在模型中不使用截距项。这时统计量输出数据集类型为TYPE=UCORR或UCOV。
- STD 要求在OUT=的数据集中把主成份得分标准化为单位方差。不规定时方差为相应特征值。
- 用VAR语句指定原始变量。必须为数值型（区间变量）。

西安交通大学管理学院 2008秋 胡平

案例3. 一月和七月平均气温的主成份分析
—— SAS操作

- 例1. 一月和七月平均气温的主成份分析
- 在数据集TEMPERAT中存放有美国一些城市一月和七月的平均气温。我们希望对这两个气温进行主成份分析，希望用一个统一的温度来作为总的可比的温度，所以进行主成份分析。程序如下：
- /*EXAMPLE 1*/
- DATA TEMPERAT;
- INPUT CITY \$1-15 JANUARY JULY;
- CARDS;
- MOBILE 51.2 81.6
- PHOENIX 51.2 91.2
- LITTLE ROCK 39.5 81.4
- SACRAMENTO 45.1 75.2
- DENVER 29.9 73.0
- HARTFORD 24.8 72.7
- WILMINGTON 32.0 75.8
- WASHINGTON DC 35.6 78.7
- JACKSONVILLE 54.6 81.0
- MIAMI 67.2 82.3
- ATLANTA 42.4 78.0
- BOISE 29.0 74.5
- CHICAGO 22.9 71.9
- PEORIA 23.8 75.1
- INDIANAPOLIS 27.9 75.0
- DES MOINES 19.4 75.1
- WICHITA 31.3 80.7
- LOUISVILLE 33.3 76.9
- NEW ORLEANS 52.9 81.9
- PORTLAND, MAINE 21.5 68.0
- BALTIMORE 33.4 76.6
- BOSTON 29.2 73.3
- DETROIT 25.5 73.3
- SAULT STE MARIE 14.2 63.8
- DULUTH 8.5 65.6
- MINNEAPOLIS 12.2 71.9
- JACKSON 47.1 81.7
- KANSAS CITY 27.8 78.8

案例3. 1997至1998年地区统计的50个社会经济指标的主成份分析

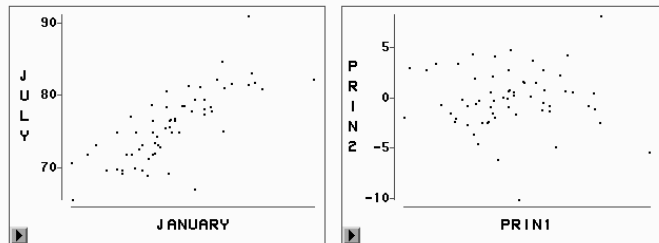
—— SAS操作

- ST LOUIS 31.5 78.6
- GREAT FALLS 20.5 69.3
- OMAHA 22.6 77.2
- RENO 31.9 69.3
- CONCORD 20.6 69.7
- ATLANTIC CITY 32.7 75.1
- ALBUQUERQUE 35.2 78.7
- ALBANY 21.5 72.0
- BUFFALO 23.7 70.1
- NEW YORK 32.2 76.6
- CHARLOTTE 42.1 78.5
- RALEIGH 40.5 77.5
- BISMARCK 8.2 70.8
- CINCINNATI 31.1 75.6
- CLEVELAND 26.9 71.4
- COLUMBUS 28.4 73.6
- OKLAHOMA CITY 36.8 81.5
- PORTLAND, OREG 38.1 67.1
- PHILADELPHIA 32.3 76.8
- PITTSBURGH 28.1 71.9
- PROVIDENCE 28.4 72.1
- COLUMBIA 45.4 81.2
- SIOUX FALLS 14.2 73.3
- MEMPHIS 40.5 79.6
- NASHVILLE 38.3 79.6
- DALLAS 44.8 84.8
- EL PASO 43.6 82.3
- HOUSTON 52.1 83.3
- SALT LAKE CITY 28.0 76.7
- BURLINGTON 16.8 69.8
- NORFOLK 40.5 78.3
- RICHMOND 37.5 77.9
- SPOKANE 25.4 69.7
- CHARLESTON, WV 34.5 75.0
- MILWAUKEE 19.4 69.9
- CHEYENNE 26.6 69.1
- ;
- PROC PRINCOMP COV
- OUT=PRIN;
- VAR JULY JANUARY;
- RUN;

西安交通大学管理学院 2008秋 胡平

案例3. 一月和七月平均气温的主成份分析

—— SAS操作



西安交通大学管理学院 2008秋 胡平

案例3. 一月和七月平均气温的主成份分析

—— SAS操作

- 在 INSIGHT 中打开 WORK.PRIN，分别绘制 JULY 对 JANUARY、PRIN2 对 PRIN1 的散点图（图 1）。从图可以看出主成份为原始变量的一个正交旋转。输出如下：

Principal Component Analysis			
62 Observations			
2 Variables			
Simple Statistics			
	JULY	JANUARY	
Mean	75.92096774	32.55483871	
StD	4.88061193	11.59197967	

西安交通大学管理学院 2008秋 胡平

案例3. 一月和七月平均气温的主成份分析

—— SAS操作

Covariance Matrix				
	JULY	23.8203728	43.4319461	
	JANUARY	43.4319461	134.3739926	
Total Variance = 158.19436542				
Eigenvalues of the Covariance Matrix				
	Eigenvalue	Difference	Proportion	Cumulative
PRIN1	149.396	140.597	0.944380	0.94438
PRIN2	8.799	.	0.055620	1.00000
Eigenvectors				
	PRIN1	PRIN2		
JULY	0.326866	0.945071		
JANUARY	0.945071	-.326866		

西安交通大学管理学院 2008秋 胡平

案例3. 一月和七月平均气温的主成份分析

—— SAS操作

- 输出中，第一部分为简单统计量（均值和标准差），第二部分为协方差的特征值（注意我们在过程中用了 COV 选项，无此选项用相关阵），从这里可以看到贡献率（Proportion）和累计贡献率（Cumulative），第三部分为特征向量。按本结果的特征向量值及用 COV 选项规定使用协方差阵，我们可以知道两个主成份如此计算：

$PRIN1 = 0.326866 (JULY - 75.92) + .945071 (JANUARY - 32.55)$

$PRIN2 = 0.945071 (JULY - 75.92) + (-.326866) (JANUARY - 32.55)$

- 如果没有用 COV 选项，原始变量还需要除以标准差。由系数可见，第一主成份是两个月份的加权平均，代表了一个地方的气温水平，第二主成份系数一正一负，反应了冬季和夏季的气温差别。

西安交通大学管理学院 2008秋 胡平

案例3. Factor —— SAS操作

- SAS/STAT 的 FACTOR 过程可以进行因子分析、分量分析和因子旋转。对因子模型可以使用正交旋转和斜交旋转，可以用回归法计算得分系数，同时把因子得分的估计存贮在输出数据集中；用 FACTOR 过程计算的所有主要统计量也能存贮在输出数据集中。
- FACTOR 过程用法很简单，主要使用如下语句：
- PROC FACTOR DATA= 数据集 选项；
- VAR 原始变量；
- RUN；

西安交通大学管理学院 2008秋 胡平

案例3. Factor —— SAS操作

- 输出结果包括特征值情况、因子载荷、公因子解释比例，等等。为了计算因子得分，一般在 PROC FACTOR 语句中加一个 SCORE 选项和 “OUTSTAT=输出数据集” 选项，然后用如下的得分过程计算公因子得分：
- PROC SCORE DATA=原始数据集
- SCORE=FACTOR 过程的输出数据集
- OUT=得分输出数据集；
- VAR 用来计算得分的原始变量集合；
- RUN；

西安交通大学管理学院 2008秋 胡平

案例3. 洛杉矶12个地区统计的五个社会经济指标—— SAS操作

数据集SOCECON为洛杉矶12个地区统计的五个社会经济指标：人口总数（POP），教育程度（SCHOOL），就业数（EMPLOY），服务业人数（SERVICES），中等的房价（HOUSE）。用FACTOR过程可以进行主成份分析。下例中的SIMPLE选项要求计算变量的简单统计量，CORR 要求输出相关阵。

案例3. 洛杉矶12个地区统计的五个社会经济指标—— SAS操作

```
DATA SOCECON;
  TITLE '五个经济指标的分析';
  INPUT POP SCHOOL EMPLOY SERVICES HOUSE;
  CARDS;
5700 12.8 2500 270 25000
1000 10.9 600 10 10000
3400 8.8 1000 10 9000
3800 13.6 1700 140 25000
4000 12.8 1600 140 25000
8200 8.3 2600 60 12000
1200 11.4 400 10 16000
9100 11.5 3300 60 14000
9900 12.5 3400 180 18000
9600 13.7 3600 390 25000
9600 9.6 3300 80 12000
9400 11.4 4000 100 13000
```

案例3. 洛杉矶12个地区统计的五个社会经济指标—— SAS操作

```
PROC FACTOR DATA=SOCECON SIMPLE CORR;
  TITLE2 '主成份分析';
RUN;
```

结果给出了五个变量的简单统计量，相关阵，和相关阵的特征值、累计贡献：

	1	2	3	4	5
Eigenvalue	2.8733	1.7967	0.2148	0.0999	0.0153
Difference	1.0767	1.5818	0.1149	0.0847	
Proportion	0.5747	0.3593	0.0430	0.0200	0.0031
Cumulative	0.5747	0.9340	0.9770	0.9969	1.0000

2 factors will be retained by the MINEIGEN criterion.

案例3. 洛杉矶12个地区统计的五个社会经济指标—— SAS操作

前两个主成份解释了93.4%的方差，按照缺省的选择因子个数的准则MINEIGEN，取大于1的特征值，所以取两个因子。因子模式阵（factor pattern，或称因子载荷阵）为最重要的结果之一：

	Factor Pattern	
	FACTOR1	FACTOR2
POP	0.58096	0.80642
SCHOOL	0.76704	-0.54476
EMPLOY	0.67243	0.72605
SERVICES	0.93239	-0.10431
HOUSE	0.79116	-0.55818

案例3. 洛杉矶12个地区统计的五个社会经济指标—— SAS操作

它们是用公因子预报原始变量的回归系数。第一主成份（因子）在所有五个变量上都有正的载荷，可见这个因子反应了城市规模的影响。第二主成份在人口、就业上有大的正载荷，在教育程度和住房价格上有大的负载荷，则第二个因子较大的城市人口多但是教育程度和住房价格低。结果还给出了公因子解释能力的估计：

Final Communality Estimates: Total = 4.669974

	POP	SCHOOL	EMPLOY	SERVICES	HOUSE
	0.987826	0.885106	0.979306	0.880236	0.937500

案例3. 洛杉矶12个地区统计的五个社会经济指标—— SAS操作

这里给出了公因子对每一个原始变量的解释能力的量度，这是用原始变量对公因子的复相关系数平方（取0到1间值）来计算的。Communality Estimate是这些复相关系数平方的总和。因为每一个复相关系数平方都比较大，所以可认为两个公因子可以很好地解释原始变量中的信息。但是得到的因子解释不够清楚，于是考虑用其它的因子分析方法。

案例3. 洛杉矶12个地区统计的五个 社会经济指标—— SAS操作

- 我们来进行主因子分析。用FACTOR过程作主因子分析与作主成份分析的不同只是增加一个PRIORS=选项，可以用PRIORS=SMC或者MAX、ONE等。例如：
- PROC FACTOR DATA=SOCECON
priors=smc;
- TITLE2 '主因子分析';
- RUN;

西安交通大学管理学院 2008秋 胡平

案例3. 洛杉矶12个地区统计的五个社会经济指标—— SAS操作

- 主因子法计算简约了的相关阵的特征值（相当于个的估计），所以其特征值可能为负值。选取因子个数的缺省准则是PROPORTION=1，即累计特征值达到特征值总和的100%。这样取了两个因子。结果与主成份分析相似。为了得到好的因子解释，我们在上面的PROC FACTOR语句中再加上一个ROTATE=PROMAX旋转选项，这样将在得到主因子分析后先产生方差最大正交预旋转（VARIMAX）然后进行斜交旋转，并加了一个REORDER选项使输出时把原始变量受相同因子影响的放在一起：
 - PROC FACTOR DATA=SOCECON PRIORS=SMC
ROTATE=PROMAX REORDER;
 - TITLE2 '主因子分析及PROMAX斜交旋转'；
 - RUN;
- 西安交通大学管理学院 2008秋 胡平

西安交通大学管理学院 2008秋 胡平

案例3. 洛杉矶12个地区统计的五个 社会经济指标—— SAS操作

- 在初始的主因子结果之后是方差最大预旋转的结果（只显示了旋转阵和旋转后的因子载荷）；

Orthogonal Transformation Matrix			
	1	2	
1	0.78895	0.61446	
2	-0.61446	0.78895	

Rotated Factor Pattern			
		FACTOR1	FACTOR2
■	HOUSE	0.94072	-0.00004
■	SCHOOL	0.90419	0.00055
■	SERVICES	0.79085	0.41509
■	POP	0.02255	0.98874
■	EMPLOY	0.12625	0.97499

西安交通大学管理学院 2008 秋 模拟

案例3. 洛杉矶12个地区统计的五个社会经济指标—— SAS操作

- 可见第一因子反映了房价、教育水平、服务业人数，这些应该与发达程度有关。第二因子反映了人口和就业情况，与城市规模有关。这样得到的因子已经比较好用。我们再看斜交旋转的结果，这里只给出了旋转后的因子模式阵：

Rotated Factor Pattern (Std Reg Coefs)			
		FACTOR1	FACTOR2
	HOUSE	0.95558	-0.09792
	SCHOOL	0.91842	-0.09352
	SERVICES	0.76053	0.33932
	POP	-0.07908	1.00192
	EMPLOY	0.04799	0.97509

西安交通大学管理学院 2008秋 胡平

案例3. 洛杉矶12个地区统计的五个社会经济指标—— SAS操作

- 从结果看得到的因子比正交旋转没有改进。因为斜交旋转后的公因子是相关的，所以结果中还给出了公因子的相关阵，参考结构 (Reference Structure, 为每个原始变量与公因子扣除其它公因子影响的偏相关)，因子结构 (Factor Structure, 为原始变量与公因子间的相关系数)。

- Prerotation Method: Varimax

	Orthogonal Transformation Matrix	
	1	2
1	0.78895	0.61446
2	-0.61446	0.78895

	Rotated Factor Pattern	
	FACTOR1	FACTOR2
POP	0.02255	0.98874
SCHOOL	0.90419	0.00055
EMPLOY	0.14625	0.97499
SERVICES	0.79085	0.41509
HOUSE	0.94072	-0.00004

西安交通大学管理学院 2008秋 胡平

案例3. 洛杉矶12个地区统计的五个社会经济指标—— SAS操作

- 为了产生因子得分，需要在FACTOR过程中使用SCORE选项和OUTSTAT=选项输出得分系数数据集并调用SCORE过程。比如，为了计算方差最大正交旋转的主因子得分，可以用如下程序：
- PROC FACTOR DATA=SOCECON PRIORS=SMC ROTATE=VARIMAX REORDER SCORE OUTSTAT=OUTF;
- TITLE2 '主因子分析及VARIMAX正交旋转';
- RUN;
- PROC SCORE DATA=SOCECON SCORE=OUTF OUT=OUTS;
- TITLE2 'VARIMAX正交旋转后的主因子得分';
- RUN;

西安交通大学管理学院 2008秋 组别

小结

因子分析是数据缩减的一种多元分析方法，它是基于信息损失最小化而提出的一种非常有效的方法。它把众多的指标综合成几个为数较少的指标，这些指标即因子指标。因子的特点是：第一，因子变量的数量远远少于原始变量的个数；第二，因子变量并非原始变量的简单取舍，而是一种新的综合；第三，因子变量之间没有线性关系；第四，因子变量具有明确解释性，可以最大限度地发挥专业分析的作用。因子分析就是以最少的信息损失，将众多的原始变量浓缩成为少数几个因子变量，使得变量具有更高的可解释性的一种数据缩减方法，是多元分析的主干技术之一。

附加部分

西安交通大学管理学院 2008秋 胡平

1、主成分分析

- 主成分分析的基本思想是通过降维过程，将多个相互关联的数值指标转化为少数几个互不相关的综合指标的统计方法，即用较少的指标来代替和综合反映原来较多的信息，这些综合后的指标就是原来多指标的主要成分。下面通过实例说明用SPSS进行主成分分析的步骤。
- 用主成分分析法减少变量个数
- 例1.1：某研究单位测得20名肝病患者的4项肝功能指标：转氨酶、肝大指数、硫酸锌浊度、甲胎球蛋白，根据数据试做主成分分析。

西安交通大学管理学院 2008秋 胡平

- 如图，打开主成分分析对话框：



- 分别单击[Descriptives]、[Extraction]和[Scores]按钮，得到如下对话框，进行各选项的设置，如图：



西安交通大学管理学院 2008秋 胡平

- 设置好后，单击[Continue]按钮，然后单击[OK]，得到分析结果，以下对结果进行解析。

Descriptive Statistics				
	Mean	Std. Deviation	N	Missing
X1	138.00	88.888	20	
X2	2.33	1.055	20	
X3	15.00	7.420	20	
X4	35.50	21.879	20	

Correlation Matrix					
	X1	X2	X3	X4	
Correlation	1.000	.695	-.148	.025	
	X2	1.000	-.148	.071	
	X3	.219	1.000	.071	
	X4	.025	.135	1.000	
Sig. (1-tailed)	X1	.000	.176	.459	
	X2	.000	.267	.285	
	X3	.176	.267	.383	
	X4	.459	.285	.383	

- 上图是所有原始变量的通用统计描述，包括均数、标准差和总例数。
- 下图为各指标间的相关矩阵，包括偏相关系数及其相应P值。

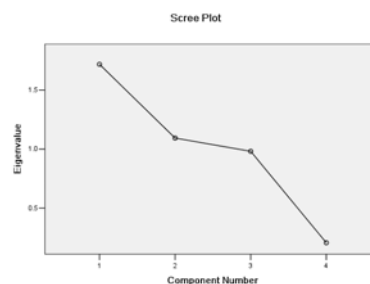
西安交通大学管理学院 2008秋 胡平

Total Variance Explained						
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	1.718	42.956	42.956	1.718	42.956	42.956
2	1.094	27.338	70.295	1.094	27.338	70.295
3	.981	24.534	94.828	.981	24.534	94.828
4	.207	5.172	100.000	.207	5.172	100.000

Extraction Method: Principal Component Analysis.

- 主成分的统计信息，包括特征根由大到小的次序排列，各主成分的贡献率及累计贡献率：第一主成分和第二主成分的特征根分别为1.718和1.094，前两个特征根均大于1，累积贡献率为70.295%。由于第三主成分的特征根接近1，且其贡献率与第二主成分相近，故本例宜取前三个主成分，此时累积贡献率达94.828%。

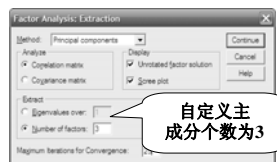
西安交通大学管理学院 2008秋 胡平



- 碎石图，结合特征根曲线的拐点及特征根值，该图从另一个侧面说明取前三个主成分为宜。

西安交通大学管理学院 2008秋 胡平

- 因为主成分个数确定为3，则再回到Factor Analysis: Extraction对话框，在“Number of factors”中选中3，得到该因子负荷矩阵。



	Component		
	1	2	3
X1	.918	.089	-.238
X2	.904	-.297	.058
X3	.115	.945	-.268
X4	.213	.319	.922

Extraction Method: Principal Component Analysis.
a. 3 components extracted.

- 由右图的因子负荷矩阵，可见第一主成分主要包含原变量x1、x2的信息，类似地，第二主成分主要包含原变量x3的信息，第三主成分主要包含原变量x4的信息。

西安交通大学管理学院 2008秋 胡平

- 用主成分分析法解决自变量的多重共线性问题
- 进行多重线性回归分析时，经常碰到自变量之间强相关的问题，即多重共线性。可通过主成分回归来求回归系数，主成分既保留了原指标的绝大部分信息，又有主成分之间互不相关的特点。
- 主成分回归的具体步骤是：
 - ◇采用多重回归分析，进行共线性诊断；
 - ◇进行主成分分析确定所需主成分数；
 - ◇进行主成分回归分析

西安交通大学管理学院 2008秋 胡平

- 例1.2：某研究者收集了13名儿童的性别（x1：男=1，女=2）、年龄（x2：月）、身高（x3：cm）、体重（x4：kg）、胸围（x5：cm）和心象面积（y：cm²）数据，试分析心象面积与性别、年龄、身高、体重和胸围之间的关系。

- 首先进行共线性诊断

ID	X1	X2	X3	X4	Y
1	0	3	5	1	1
2	2	2	4	2	2
3	3	3	5	3	3
4	4	4	6	4	4
5	5	5	7	5	5
6	6	6	8	6	6
7	7	7	9	7	7
8	8	8	10	8	8
9	9	9	11	9	9
10	10	10	12	10	10
11	11	11	13	11	11
12	12	12	14	12	12
13	13	13	15	13	13

西安交通大学管理学院 2008秋 胡平

- 得到回归分析主界面：



西安交通大学管理学院 2008秋 胡平

- 单击[OK]，得到分析结果，以下对结果进行解析。

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.970 ^a	.903	.920	6.53044

a. Predictors: (Constant), X5, X1, X2, X3, X4

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	8106.187	5	1221.837	28.545	.000 ^a
	Residual	298.526	7	42.647		
	Total	8406.713	12			

a. Predictors: (Constant), X5, X1, X2, X3, X4
b. Dependent Variable: Y

- 模型总体的假设检验结果。模型总体拟合很好（ $R^2=0.953$ ）。方差分析表显示结果有统计学意义。

西安交通大学管理学院 2008秋 胡平

Coefficients ^a											
Model		Unstandardized Coefficients		Standardized Coefficients		t	Sig.	95% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta				Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	54.582	124.273			.439	.674	-239.275	348.440		
	X1	-7.763	8.066	-.174		-.962	.368	-25.836	11.311	.203	4.929
	X2	.121	.180	.309		.672	.523	-.304	.545	.031	31.893
	X3	.290	.418	.268		.693	.510	-.598	1.278	.024	42.263
	X4	1.121	2.256	.840		.487	.634	-4.212	6.455	.002	428.936
	X5	-.941	2.332	-.524		-.404	.689	-5.455	4.574	.004	253.206

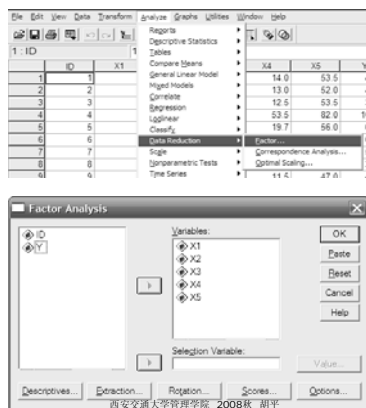
Collinearity Diagnostics ^a									
Model	Dimension	Eigenvalue	Condition Index	Variance Proportions					
				(Constant)	X1	X2	X3	X4	X5
1	1	5.581	1.000	.00	.00	.00	.00	.00	.00
	2	.336	4.084	.00	.01	.01	.00	.00	.00
	3	.071	8.870	.00	.17	.00	.00	.00	.00
	4	.012	22.024	.00	.04	.39	.00	.02	.00
	5	.001	62.706	.01	.02	.59	.66	.00	.01
	6	8.11E-005	262.326	.99	.76	.02	.34	.97	.98

a. Dependent Variable: Y

- 尽管模型总体拟合较好，但参数估计结果显示各偏回归系数均无统计学意义，说明自变量存在共线性。下图显示子变量x4和x5与常数项极度相关。于是我们需要采用主成分回归分析。

西安交通大学管理学院 2008秋 胡平

- 进行主成分分析确定所需主成分数



- 分别单击[Descriptives]、[Extraction]和[Scores]按钮，得到如下对话框，进行各选项的设置，如图：



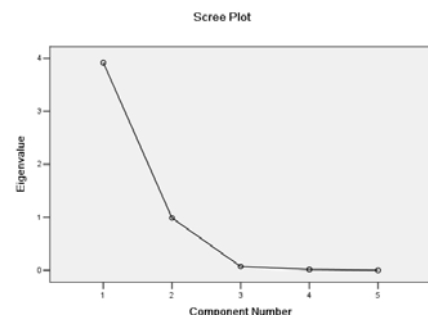
- 单击[OK]，得到分析结果。

Total Variance Explained						
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.918	78.366	78.366	3.918	78.366	78.366
2	.989	19.770	98.136	.989	19.770	98.136
3	.073	1.462	99.598	.073	1.462	99.598
4	.019	.373	99.971	.019	.373	99.971
5	.001	.029	100.000	.001	.029	100.000

Extraction Method: Principal Component Analysis.

- 由图可见，第一主成分的特征根为3.918，它解释了总变异的78.366%；第二主成分的特征根为0.989，接近1，他解释了总变异的19.77%。前两个特征根的累积贡献率为98.136%，即前两个主成分包含了原有5个指标的98.136%的信息，所以本例可以取前两个主成分代替原有的5个指标变量。

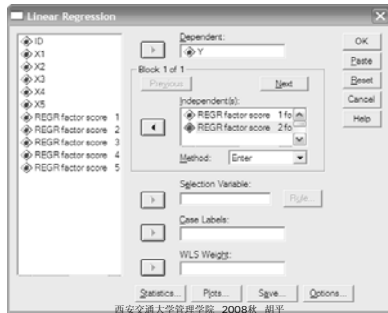
西安交通大学管理学院 2008秋 胡平



- 碎石图，显示前两个主成分的特征值接近1及以上，进一步说明取前两个主成分。

西安交通大学管理学院 2008秋 胡平

- 主成分回归分析
- 根据上面的主成分分析，得到各主成分的数值，下面采用前两个主成分来代替原有的5个变量进行主成分回归分析。



西安交通大学管理学院 2008秋 胡平

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.971 ^a	.943	.932	6.04470

a. Predictors: (Constant), REGR factor score 2 for analysis 1, REGR factor score 1 for analysis 1

Model	Sum of Squares	df	Mean Square	F	Sig.
1 Regression	5041.329	2	2520.665	82.871	.000 ^a
Residual	365.384	10	36.538		
Total	5406.713	12			

a. Predictors: (Constant), REGR factor score 2 for analysis 1, REGR factor score 1 for analysis 1

b. Dependent Variable: Y

- 上图为主成分回归分析的模型拟合情况，结果显示模型拟合较好（ $R^2=0.943$ ，方差分析 $P=0.000$ ）。

西安交通大学管理学院 2008秋 胡平

2、因子分析

- 因子分析是一种寻找隐藏在可测变量中，不能或不易直接观察到，但却影响或支配可测变量的潜在因子，并估计潜在因子对可测变量的影响程度及潜在因子之间关联性的多元统计分析方法。因子分析的主要任务是找出共性因子变量，估计因子模型，计算共性因子变量的取值和对共性因子变量做出合理的解释。
- 例2：为评价医院的医疗工作质量，某研究者收集了近三年的门诊人次、出院人数、病床利用率等9个指标，试用因子分析方法探讨其综合评价体系。

西安交通大学管理学院 2008秋 胡平

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error				Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	58.939	1.676		35.168	.000	55.204	62.675	1.000	1.000
	REGR factor score 1 for analysis 1	22.149	1.745	.959	12.693	.000	18.261	26.037	1.000	1.000
	REGR factor score 2 for analysis 1	-3.589	1.745	-.155	-2.057	.067	-7.477	.299	1.000	1.000

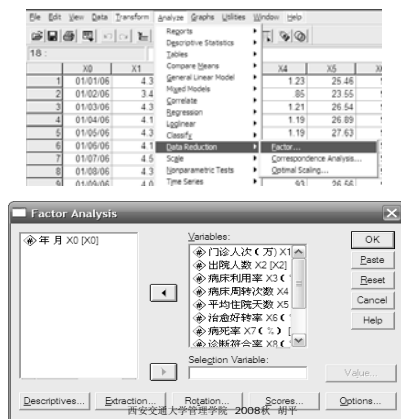
a. Dependent Variable: Y

- 主成分回归分析的参数估计及其假设检验结果，显示 β_0 、 β_1 、 β_2 均有统计学意义，其线性回归方程为：

$$Y=58.939+22.149z_1-3.589z_2$$

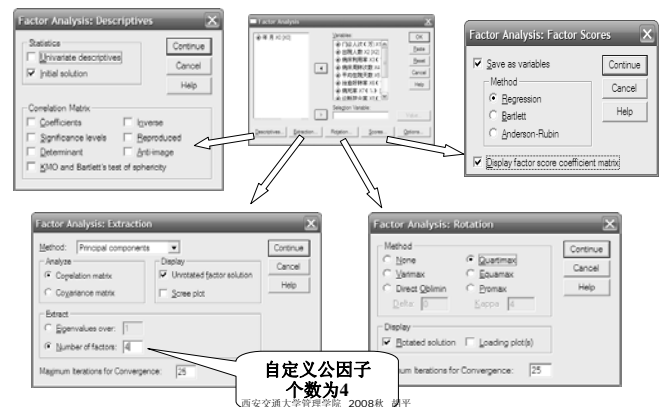
西安交通大学管理学院 2008秋 胡平

- 因子分析过程的操作提示：



西安交通大学管理学院 2008秋 胡平

- 分别单击[Descriptives]、[Extraction]、[Rotation]和[Scores]按钮，得到如下对话框，进行各选项的设置，如图：



西安交通大学管理学院 2008秋 胡平

- 单击[OK], 得到分析结果, 以下对结果进行解析。

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	2.822	31.357	31.357	2.822	31.357	31.357	2.648	29.434	29.434
2	1.992	22.138	53.495	1.992	22.138	53.495	1.847	20.527	49.931
3	1.448	16.091	69.585	1.448	16.091	69.585	1.471	16.340	66.271
4	.784	8.709	78.294	.784	8.709	78.294	1.082	12.023	78.294
5	.668	7.424	85.718						
6	.537	5.995	91.683						
7	.454	5.047	96.730						
8	.175	1.942	98.672						
9	.119	1.328	100.000						

Extraction Method: Principal Component Analysis.

- 主成分信息, 图中显示前3个主成分的特征值大于1, 但它们的累积贡献率仅为69.585%, 故将4个公因子加入, 此时累积贡献率达78.294%, 即约78.3%的总方差可以由4个潜在因子解释。

西安交通大学管理学院 2008秋 胡平

Communalities		
	Initial	Extraction
1 门诊人次 (万) X1	1.000	.880
出院人数 X2	1.000	.873
病床利用率 X3 (%)	1.000	.873
病床周转次数 X4	1.000	.917
平均住院天数 X5	1.000	.767
治愈好转率 X6 (%)	1.000	.796
病死率 X7 (%)	1.000	.683
诊断符合率 X8 (%)	1.000	.573
抢救成功率 X9 (%)	1.000	.684

Extraction Method: Principal Component Analysis.

- 公因子方差比: 结果显示, 每个指标变量的共性方差均在0.5以上, 且大多数接近或超过0.7, 说明4个公因子能够较好地反映原各指标变量的大部分信息。

西安交通大学管理学院 2008秋 胡平

Component Matrix ^a				
	Component			
	1	2	3	4
1 门诊人次 (万) X1	-.260	.769	.009	.469
出院人数 X2	.764	.133	.090	.513
病床利用率 X3 (%)	.239	.778	-.085	-.452
病床周转次数 X4	.684	.666	-.070	-.024
平均住院天数 X5	-.724	.119	.441	.185
治愈好转率 X6 (%)	.039	-.070	.889	-.021
病死率 X7 (%)	-.406	-.183	-.663	.230
诊断符合率 X8 (%)	-.637	.397	.039	-.090
抢救成功率 X9 (%)	.740	-.362	.057	.034

Extraction Method: Principal Component Analysis.
a. 4 components extracted.

- 旋转前的因子载荷阵: 根据0.5原则, 因子1在多数原始指标上有较大载荷; 因子2在x1、x3和x4指标上有较大载荷; 因子3在x6和x7指标上有较大载荷; 因子4在x2指标上有较大载荷。

西安交通大学管理学院 2008秋 胡平

Rotated Component Matrix ^a				
	Component			
	1	2	3	4
1 门诊人次 (万) X1	.512	.285	-.068	.729
出院人数 X2	-.666	.151	.146	.621
病床利用率 X3 (%)	.135	.924	.030	-.010
病床周转次数 X4	-.346	.809	.057	.374
平均住院天数 X5	.736	-.334	.302	.150
治愈好转率 X6 (%)	.044	-.155	.877	.023
病死率 X7 (%)	.199	-.298	-.744	.019
诊断符合率 X8 (%)	.749	.102	-.038	.020
抢救成功率 X9 (%)	-.811	-.022	.156	-.030

Extraction Method: Principal Component Analysis.
Rotation Method: Quartimax with Kaiser Normalization.
a. Rotation converged in 5 iterations.

- 旋转后的因子载荷阵: 通过4次方最大旋转后, 得到了9个指标在4个因子上的新的因子载荷。结果显示, 因子1支配门诊人次、出院人数、平均住院天数、诊断符合率和抢救成功率; 因子2支配病床利用率和病床周转次数; 因子3支配治愈好转率和病人死亡率; 因子4支配门诊人次和出院人数。
- 故可认为: 因子1为综合因子; 因子2为病床利用因子; 因子3为医疗水平因子; 因子4为就诊病人数量因子。

- 通过旋转后计算出的每条记录的4个因子得分作为新变量自动存储到原始数据文件中, 如图所示。

	X1	X2	X3	X4	X5	X6	X7	X8	FACTOR_1	FACTOR_2	FACTOR_3	FACTOR_4
1	399	99.98	1.23	26.46	93.33	3.16	97.51	91.68	-.0825	.7292	-.0681	.7292
2	6102.06	3.45	271	89.28	85	27.55	94.31	2.44	97.94	73.33	.98847	.93818
3	6102.06	4.36	380	103.87	1.21	26.54	92.53	4.62	94.48	76.70	.88845	1.52952
4	6104.06	4.18	377	99.48	1.18	26.89	93.86	2.32	98.41	63.16	1.17940	1.44321
5	6105.06	4.32	376	102.81	1.18	27.63	93.16	1.99	98.71	80.60	.88789	1.73664
6	6106.06	4.13	369	97.55	1.16	27.34	96.63	4.38	99.03	63.16	1.14953	.93933
7	6107.06	4.57	381	91.68	1.14	24.89	96.60	2.73	99.69	73.53	.74863	1.14823
8	6108.06	4.31	389	62.18	.92	31.74	91.67	3.65	99.48	61.11	2.61687	1.13148
9	6109.06	4.06	426	83.27	93	26.56	93.61	3.08	98.48	70.73	.64881	.91840
10	6110.06	4.43	488	52.29	86	24.26	91.52	4.21	98.76	79.57	.36729	.22876
11	6111.06	4.15	498	96.43	1.03	28.76	93.63	3.60	98.16	86.48	1.0281	.77987
12	6112.06	4.18	514	52.89	1.07	26.31	93.24	4.22	100.00	78.95	.37677	.19439
13	6201.06	4.13	498	90.96	.97	26.96	93.66	4.67	98.77	60.53	2.74877	1.18816
14	6202.06	3.53	344	79.86	.88	31.87	94.77	3.38	100.00	61.97	1.06572	1.52799
15	6203.06	4.16	588	98.98	1.01	29.43	95.75	2.77	98.72	42.86	1.02103	1.88644
16	6204.06	4.17	545	92.18	1.06	28.92	94.88	3.14	98.41	82.36	1.7642	.76648
17	6205.06	4.16	587	95.10	1.01	29.82	94.41	2.80	99.36	68.61	.81431	.68864
18	6206.06	4.06	540	93.17	1.07	27.69	93.47	2.77	99.80	70.21	1.68857	.76438
19	6207.06	5.06	552	84.38	1.10	27.56	95.15	3.18	98.63	69.23	.88187	.48038
20	6208.06	4.03	453	72.89	.90	26.63	91.92	4.52	98.05	68.42	.93108	1.53910
21	6209.06	4.16	529	88.53	1.03	27.45	91.52	3.84	98.18	88.42	.22887	.14191
22	6210.06	3.94	510	91.91	1.02	25.44	94.88	2.65	99.38	73.91	.99824	.20609
23	6211.06	4.12	552	89.14	1.10	26.76	92.86	4.67	98.52	68.61	1.74855	1.14342
24	6212.06	4.42	597	90.18	1.18	26.94	93.53	3.76	99.28	73.61	.24776	.57521
25	6201.06	3.95	437	78.87	.87	23.65	94.46	4.03	98.22	87.18	1.36408	1.28784
26	6202.06	3.94	477	87.34	.90	26.78	91.78	4.67	94.28	87.18	1.9386	1.93829
27	6203.06	4.14	638	88.57	1.27	26.53	95.16	1.67	94.50	91.67	1.48312	1.93872
28	6204.06	3.87	583	99.82	1.10	27.66	93.43	3.16	94.49	89.97	1.76340	.97985
29	6205.06	4.08	552	90.19	1.10	27.53	96.36	3.47	97.88	87.14	-.19628	.48073
30	6206.06	4.14	587	91.10	1.09	23.06	91.66	2.47	97.72	87.18	-.88838	.58251
31	6207.06	4.04	574	81.36	1.14	26.60	93.74	1.41	98.26	93.02	-.88144	1.44237
32	6208.06	3.93	510	78.87	1.02	23.88	93.62	3.09	95.48	88.37	1.33964	1.67714

西安交通大学管理学院 2008秋 胡平

西安交通大学管理学院 2008秋 胡平

3、主成分分析与因子分析的联系及区别

- (1)两者都是在多个原始变量中通过它们之间的内部相关性来获得新的变量, 达到既减少分析指标个数, 又能概括原始指标主要信息的目的。
- (2)提取公因子的方法主要有主成分法和公因子法, 主成分法主要从解释变量的变异角度, 尽量使变量的方差能被主成分解释; 公因子法主要是从解释变量的相关性角度, 尽量使变量的相关程度能被公因子解释。

- (3)因子分析提取的公因子比主成分分析提取的更具有可解释性。因子分析校正观察变量的度量误差，且进行因子旋转，使潜在因子的实际意义更明显。
- (4) 两者分析的实质及重点不同。主成分分析主要是综合原始变量的信息，而因子分析重在解释原始变量之间的关系。
- (5)两者的SPSS操作都是通过“Analyze->Data Reduction->Factor...”过程实现，但因子分析除了使用[Descriptives]、[Extraction]和[Scores] 对话框外还要使用[Rotation]对话框进行因子旋转。