



因子分析



《美国数学建模竞赛》
完整课程请长按下方二维码





因子分析

因子分析模型是主成分分析的推广。它也是利用降维的思想，由研究原始变量相关矩阵内部的依赖关系出发，把一些具有错综复杂关系的变量归结为少数几个综合因子的一种多变量统计分析方法。

因子分析的思想始于1904年Charles Spearman对学生考试成绩的研究。近年来，随着电子计算机的高速发展，人们将因子分析的理论成功地应用于心理学、医学、气象、地质、经济学等各个领域，也使得因子分析的理论和方法更加丰富。



因子分析的基本思想是根据相关性大小把原始变量分组，使得同组内的变量之间相关性较高，而不同组的变量间的相关性则较低。每组变量代表一个基本结构，并用一个不可观测的综合变量表示，这个基本结构就称为**公共因子**。对于所研究的某一具体问题，原始变量就可以分解成两部分之和的形式，一部分是少数几个不可测的所谓公共因子的线性函数，另一部分是与公共因子无关的**特殊因子**。



$$X = AF + \varepsilon$$

称 F_1, F_2, \dots, F_m 为公共因子, 是不可观测的变量,
 $A = (a_{ij})_{p \times m}$ 称为因子载荷阵, a_{ij} 表示第 i 个变量
在第 j 个因子上的载荷 (loading). ε_i 是特殊因子,
是不能被前 m 个公共因子包含的部分。并且满足
 $cov(F, \varepsilon) = 0$, F, ε 不相关。



因子分析的解决的三个基本问题：

- 1.因子载荷阵 A 的估计；
- 2.当因子难以得到合理的解释时，对因子载荷阵作正交变换,即因子旋转。目的：对因子的实际意义做出合理的解释；
- 3.给出每个变量（或样品）关于 m 个公共因子的得分，通常表示为原始变量的线性组合，即因子得分函数。目的：对公共因子作出估计。



因子分析一般步骤:

- 1) 类似主成分分析, 计算 \bar{x}_k 及 $s_k, k, j = 1, 2, \dots, m$, 建立基本方程组;
- 2) 用主成分分析法确定因子载荷阵A;
- 3) 方差极大正交旋转, 对变量系数极值化(尽量趋于0或1);
- 4) 得到因子得分函数, 计算样本因子得分。



例：已知12个地区的5个经济指标：人口总数、学校总数、就业人口、服务业总数、房子个数。依据已知的5个经济指标，对12个地区的综合经济实力进行分析。

分析：由于指标个数较多，不便于分析排序。因此，考虑先做因子分析找出指标的共同因子，再计算因子得分，通过分析因子得分来评价该地区的经济指标。



```
data ex842;  
input objects$ pop school employ services house@ @;  
cards;  
/*数据省略*/  
;  
proc factor data=ex842 /*factor表示调用因子分析模块*/  
method=principal rotate=varimax /*rotate表示因子旋转*/  
percent=0.8 /*要求累计贡献率大于0.8*/  
score outstat=ex1; /*计算因子得分*/  
var pop school employ services house; run;  
proc score data=ex842 score=ex1 out=ex2;  
var pop school employ services house; run;  
proc print data=ex1; proc print data=ex2;  
run;
```



Eigenvalues of the Correlation Matrix: Total = 5 Average = 1

	Eigenvalue	Difference	Proportion	Cumulative
1	2.87331359	1.07665350	0.5747	0.5747
2	1.79666009	1.58182321	0.3593	0.9340
3	0.21483689	0.11490283	0.0430	0.9770
4	0.09993405	0.08467868	0.0200	0.9969
5	0.01525537		0.0031	1.0000

2 factors will be retained by the PROPORTION criterion.

结果表明，5个因子对应的特征值，特征值表示因子贡献率。通常确定因子个数时，要求因子累计贡献率大于80%。结果表明应选取2个因子，记为F1,F2 贡献率分别为57.47%、35.93%。



确定因子载荷阵系数，得到初始的特征向量：

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

由于对应实际问题，公共因子的实际意义不好解释。因此考虑将指标的系数极值化，即让系数趋于1或0，趋于1说明公共因子与该指标密切相关，否则趋于0时说明相关程度很低。由此，要做因子旋转实现系数的极值化。



因子旋转程序运行结果：

The FACTOR Procedure		
Rotation Method: Varimax		
Orthogonal Transformation Matrix		
	1	2
1	0.82069	0.57137
2	-0.57137	0.82069
Rotated Factor Pattern		
	Factor1	Factor2
pop	0.01602	0.99377
school	0.94076	-0.00882
employ	0.13702	0.98007
services	0.82481	0.44714
house	0.96823	-0.00605

Orthogonal Transformation Matrix 表示因子旋转阵， Rotated Factor Pattern 为旋转后得到的因子载荷。Standardized Scoring Coefficients为因子得分系数结果。



Standardized Scoring Coefficients

	Factor1	Factor2
pop	-0.09052	0.48389
school	0.39233	-0.09631
employ	-0.03883	0.46536
services	0.29949	0.13776
house	0.40349	-0.09764

于是得到两个因子为:

$$F_1 = -0.09052pop + 0.39233school - 0.03883employ + 0.29949services + 0.40349house$$

$$F_2 = 0.48389pop - 0.09631school + 0.46536employ + 0.13776services - 0.09764house$$



根据得到的因子得分函数计算各个样本的因子得分

3. 因子得分:

The SAS System							08:50 Monday, October 5, 2009 148	
Obs	objects	pop	school	employ	services	house	Factor1	Factor2
1	A	5700	12.8	2500	270	25000	1.20297	-0.03080
2	B	1000	10.9	600	10	10000	-0.85918	-1.38351
3	C	3400	8.8	1000	10	9000	-1.25937	-0.76740
4	D	3800	13.6	1700	140	25000	1.11491	-0.79697
5	E	4000	12.8	1600	140	25000	0.93710	-0.76320
6	F	8200	8.3	2600	60	12000	-1.22514	0.54857
7	G	1200	11.4	400	10	16000	-0.16819	-1.54932
8	H	9100	11.5	3300	60	14000	-0.44127	0.73444
9	I	9900	12.5	3400	180	18000	0.32032	0.91306
10	J	9600	13.7	3600	390	25000	1.57628	1.02554
11	K	9600	9.6	3300	80	12000	-0.94628	0.96184
12	L	9400	11.4	4000	100	13000	-0.45215	1.10776

以上为12个样本的5个指标值，以及2个公共指标的得分，因子F1中J地区的得分最高，说明该地区的服务教育服务设施较好。因子F2中L地区的得分最高，说明该地区的就业情况较好。



因子分析与主成分分析的区别与联系

因子分析、主成分分析都是重要的降维方法（数据简化技术），因子分析可以看作是主成分分析的推广和发展。

主成分分析不能作为一个模型来描述，它只能作为一般的变量变换，主成分是可观测的原始变量的线性组合；因子分析需要构造因子模型，公共因子是潜在的不可观测的变量，一般不能表示为原始变量的线性组合。

因子分析是用潜在的不可观测的变量和随机影响变量的线性组合来表示原始变量，即通过这样的分解来分析原始变量的协方差结构（相依关系）。