

因子分析实例

- 对我国各省市自治区的农业生产情况作因子分析。
- 从农业生产条件、生产结果和效益出发，选取六项指标分别为：

X_1 : 乡村劳动力人口（万人）；

X_2 : 人均经营耕地面积（亩）；

X_3 : 户均生产性固定资产原值（元）；

X_4 : 家庭基本纯收入（元）；

X_5 : 人均农业总产值（千元/人）；

X_6 : 增加值占总产值比重（%）。

原始资料数据如下页表：

序号	地 区	X_1	X_2	X_3	X_4	X_5	X_6
1	北 京	66.9	0.93	2972.41	3290.73	2.525	49.7
2	天 津	80.2	1.64	4803.54	2871.62	1.774	49.6
3	河 北	1621.8	2.03	4803.54	2871.81	0.8004	54
4	山 西	635.4	2.76	2257.66	1499.14	0.555	56.2
5	内 蒙 古	514.1	10.17	5834.94	1550.15	0.9051	66.4
6	辽 宁	605.1	2.96	3108.86	2059.35	1.4752	53.1
7	吉 林	534.2	4.73	4767.51	1940.46	1.1154	63.1
8	黑龙江	494.8	8.24	5573.02	2075.42	1.6283	57.8
9	上 海	66	1.02	1660.03	4571.81	3.0448	35.6
10	江 苏	1530.2	1.26	2826.86	2868.33	1.1921	50.6
11	浙 江	1123.1	0.94	5494.23	3289.07	0.8565	63.3
12	安 徽	1953.6	1.44	3573.62	1508.24	0.5756	59.2
13	福 建	775.8	0.82	2410.05	2295.19	1.1496	62.8
14	江 西	1103.2	1.3	2310.98	1804.93	0.6649	59.9
15	山 东	2475.1	1.44	3109.11	1989.53	0.8809	55
16	河 南	2815.8	1.5	3782.26	1508.36	0.5823	58.5
17	湖 北	1296.5	1.6	2291.6	1754.13	0.8799	62.8
18	湖 南	2089.3	1.42	2348.72	1719.18	0.587	64.7
19	广 东	1439.8	0.88	3249.61	2928.24	1.096	59.7
20	广 西	1579.9	1.43	3090.17	1590.9	0.5694	64.5
21	海 南	165.9	1.35	4454.77	1575.49	0.3535	65.2
22	四 川	3903.7	1.08	2870.45	1340.61	0.4443	64.1
23	贵 州	1376.6	1.18	2282.27	1206.25	0.2892	65.4
24	云 南	1642.2	2.42	4025.06	1096.73	0.3456	64.2
25	西 藏	88.6	2.51	11559.83	1257.71	0.4349	70.4
26	陕 西	1046.1	2.6	2228.55	1091.96	0.4383	59.7
27	甘 肃	672	5.86	2879.36	1037.12	0.4883	57.2
28	青 海	137.1	2.62	6725.11	1133.06	0.4096	70.3
29	宁 夏	139.1	4.01	5607.97	1346.89	0.4973	62.5
30	新 疆	288.5	3.96	7438.13	1161.71	1.4939	57.8

- 第一步：将原始数据标准化

即令

$$x_{ij}^* = \frac{x_{ij} - \bar{x}_j}{s_j}, \quad 1 \leq i \leq 30, \quad 1 \leq j \leq 6.$$

其中，

$$\bar{x}_j = \sum_{i=1}^{30} x_{ij} / 30, \quad s_j^2 = \sum_{i=1}^{30} (x_{ij} - \bar{x}_j)^2 / 29, \quad 1 \leq j \leq 6.$$

- 第二步：建立指标间的相关系数阵**R**

即对 $1 \leq i, j \leq 30$, 计算

$$r_{ij} = \frac{\sum_{k=1}^{30} (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{\sqrt{\sum_{k=1}^{30} (x_{ki} - \bar{x}_i)^2} \sqrt{\sum_{k=1}^{30} (x_{kj} - \bar{x}_j)^2}},$$

得

$$\mathbf{R} = \begin{bmatrix} 1 & -0.3325 & -0.3710 & -0.2026 & -0.3955 & 0.1413 \\ -0.3325 & 1 & 0.3492 & -0.2980 & -0.0014 & 0.1654 \\ -0.3710 & 0.3492 & 1 & -0.2481 & -0.1308 & 0.4044 \\ -0.2026 & -0.2980 & -0.2481 & 1 & 0.8145 & -0.7112 \\ -0.3955 & -0.0014 & -0.1308 & 0.8145 & 1 & -0.7967 \\ 0.1413 & 0.1654 & 0.4044 & -0.7112 & -0.7967 & 1 \end{bmatrix}$$

- 第三步：求**R**的特征根和特征向量

序 号	特征根	贡献率（%）	累计贡献率（%）
1	2.7765	46.2756	46.2756
2	1.7409	29.0160	75.2917
3	0.7116	11.8612	87.1529
4	0.4334	7.2248	94.3778
5	0.2369	3.9484	98.3263
6	0.1004	1.6736	100

$$\text{第 } i \text{ 个变量的贡献率} = \frac{\lambda_i}{\sum_{j=1}^6 \lambda_j} \times \%, \quad 1 \leq i \leq 6.$$

$$\text{前 } i \text{ 个变量的累计贡献率} = \frac{\sum_{j=1}^i \lambda_j}{\sum_{j=1}^6 \lambda_j} \times \%, \quad 1 \leq i \leq 6.$$

- 由于前三个特征根累计贡献率已达87.15%，所以取前三个特征根所对应的特征向量。
如下：

α_1	α_2	α_3
0.1460	-0.6242	-0.1854
0.1631	0.5270	0.7547
0.2421	0.5272	0.5369
-0.5463	0.0153	0.2325
-0.5455	0.2317	-0.0422
0.5453	0.0225	0.2276

- 第四步：计算因子载荷矩阵

令 $a_i = \sqrt{\lambda_i} \alpha_i$, $1 \leq i \leq 3$. 则因子载荷矩阵 $A = (a_1, a_2, a_3)$.

A 表示如下

因子 指标	a_1	a_2	a_3	h_i^2
X_1	0.2433	-0.8236	-0.1564	0.7621
X_2	0.2718	0.6954	0.6366	0.9629
X_3	0.4035	0.6957	0.4529	0.8520
X_4	-0.9103	0.0202	0.1961	0.8675
X_5	-0.9089	0.3057	-0.0356	0.9210
X_6	0.9086	0.0296	0.1920	0.8634

其中 h_i^2 是公共因子对变量 x_i 的贡献, $1 \leq i \leq 6$.

- 第五步：因子**Varimax**旋转

旋转后的因子载荷矩阵**A**如下

因子 指标	$a_1 (F_1)$	$a_2 (F_2)$	$a_3 (F_3)$
X_1	-0.3793	-0.7252	-0.3036
X_2	-0.1046	0.2178	0.9510
X_3	-0.2957	0.8698	0.0890
X_4	0.8862	0.0265	-0.2852
X_5	0.9499	0.1206	0.0645
X_6	-0.8976	0.2402	-0.0009

- 第六步：将指标按高载荷分类，并结合专业知识给出各因子的命名（解释）
- 在第一因子中， X_4 、 X_5 、 X_6 三项指标有较大的载荷，这些都从产出效益方面描述农业情况的，所以称为产出及效益因子。
- 在第二个因子中， X_1 、 X_3 有较大的载荷，这主要是人们对农业的生产工具、人力等的投入，所以称为人为投入条件因子。
- 在第三个因子中， X_2 有较大的载荷，这主要从自然条件方面刻画农业的生产条件状况，所以称为自然条件因子。

	高载荷指标	命 名
因子一	X_4 ：家庭基本纯收入 X_5 ：人均产值 X_6 ：增加值占总产值比重	产出及效益因子
因子二	X_1 ：乡村劳动力人口 X_3 ：户均生产性固定资产 原值	人为投入条件因子
因子三	X_2 ：人均经营耕地面积	自然条件因子

- 第七步：因子得分

计算因子得分矩阵

$$\mathbf{B} = \mathbf{A}'\mathbf{R}^{-1}.$$

B为

$$\mathbf{B} = \begin{bmatrix} -0.1592 & 0.0271 & -0.1063 & 0.3083 & 0.3562 & -0.3336 \\ -0.4986 & -0.1906 & 0.7097 & 0.1330 & 0.0738 & 0.2086 \\ -0.0771 & 0.9657 & -0.2667 & -0.2659 & 0.0904 & -0.1590 \end{bmatrix}$$

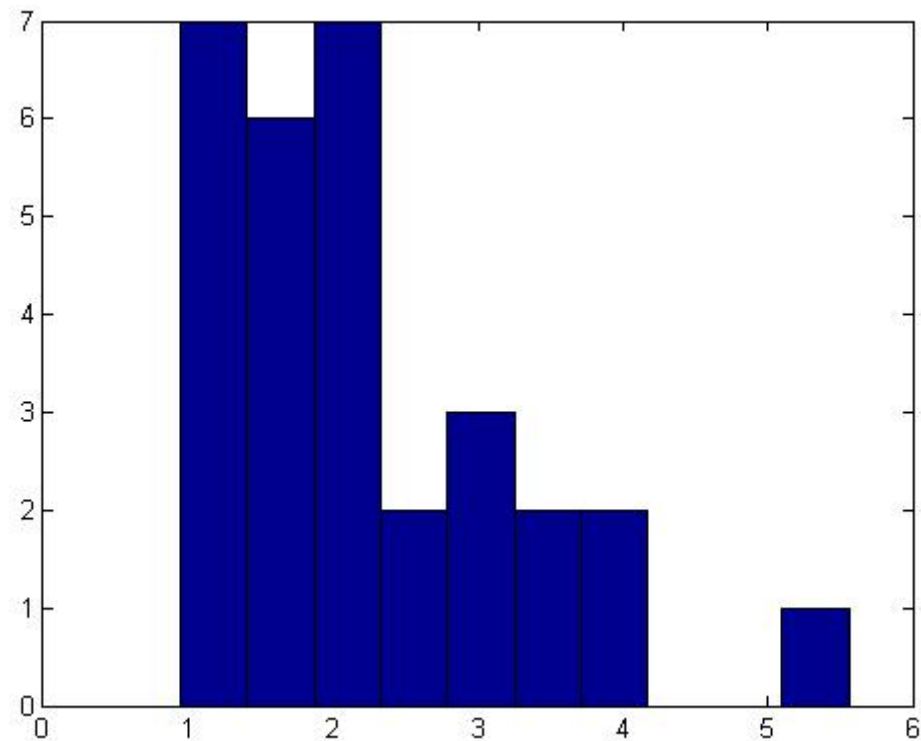
计算因子得分估计

$$\hat{\mathbf{F}} = \mathbf{B}\mathbf{X}^*.$$

计算三个公共因子绝对得分之和

行政区	得分总和	行政区	得分总和	行政区	得分总和
北京	2.9342	浙江	2.5237	海南	1.9911
天津	2.5903	安徽	1.5356	四川	3.6627
河北	0.9471	福建	0.9435	贵州	1.9248
山西	1.1655	江西	1.0140	云南	1.3281
内蒙古	3.9199	山东	1.5117	西藏	5.5579
辽宁	1.4929	河南	2.1293	陕西	1.7389
吉林	1.3497	湖北	1.0150	甘肃	2.9330
黑龙江	3.4426	湖南	2.0175	青海	2.9342
上海	3.9771	广东	1.4801	宁夏	1.9522
江苏	1.9923	广西	1.5219	新疆	2.0541

直方图



可以根据得分值将样本聚类