

社会统计学及SPSS软件应用

STATISTICS WITH SPSS

Instructor: 王荣欣

Email: rxwang@qq.com

周一3-4节、单周周四3-4节, 3A106-2

2020年11月23日

CONTENTS

1 因子分析

- 因子分析例子
- 统计软件操作

Example 2.1

下面有一个量表，表中的0代表我国根本不存在这方面的问题，10代表这方面的问题在我国非常严重，请您在量表中选择—个数字表示您的态度

- | | |
|------|------|
| • 环保 | • 言论 |
| • 教育 | • 腐败 |
| • 国防 | • 出版 |
| • 稳定 | • 犯罪 |
| • 民主 | • 酗酒 |
| • 耕地 | |

Component Matrix^a

	Component		
	1	2	3
国防 c1c	.697		
教育 c1b	.615		
民主 c1e	.717		
犯罪 c1j	.598	.540	
环保 c1a	.569		.545
稳定 c1d	.726		
耕地 c1f	.544		
腐败 c1h	.543		
言论 c1g	.654		
酗酒 c1k	.541	.516	
出版 c1i	.654		

Extraction Method: Principal Component Analysis.

a. 3 components extracted.

Rotated Component Matrix^a

	Component		
	1	2	3
国防 c1c	.749		
教育 c1b	.536		.516
民主 c1e	.711		
犯罪 c1j		.759	
环保 c1a			.734
稳定 c1d	.731		
耕地 c1f			.616
腐败 c1h			.624
言论 c1g	.716		
酗酒 c1k		.775	
出版 c1i	.591	.546	

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 15 iterations.

- 1 第一个因子与国防、教育、政治民主等相关，是全局政治。
- 2 第二个因子与犯罪、酗酒问题有关，是社会状况。
- 3 第三个因子与环保、耕地、腐败问题相关，是具体的治理。

Example 2.2

英超进球谁最强（2012-2013赛季）

变量说明 			
变量类型	变量名	详细说明	备注
因变量	下赛季进球数	计数变量 单位：个	—
球员基本情况	球员姓名、年龄、所属球队、球衣号码、场上位置	场上位置是定性变量 有3个水平（前锋、中场、后卫）	年龄和场上位置 用于建模
球员出场情况	出场次数、首发次数、出场时间（分钟）	3者正向线性相关性较高	首发次数 用于建模
球员射门表现	射门次数、射正次数、进球数（头球、左脚、右脚、任意球、点球、乌龙球）	射门成功率=射正次数/射门次数， 是评价射手的重要指标	—
球员场上表现	助攻、传球、过人、抢断、赢得点球机会、拦截、解围（头球、后场）、头球争顶成功	单位：次	—
球员犯规情况	越位、犯规、红牌、黄牌、	单位：次	—



- 为了评价球员在球场上的整体表现，从数据集选取了22个技术指标进行因子分析。
- 这22个指标之间的相关性如图所示。

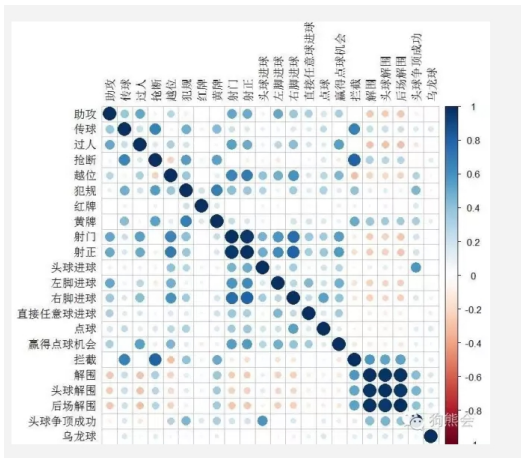


Figure 2.1: 相关系数矩阵

	进攻因子	防守中场因子	防守后卫因子	共性方差
助攻	0.57	0.23	-0.36	0.50
过人	0.57	0.21	-0.36	0.50
越位	0.78	-0.22	0.17	0.69
射门	0.93	0.07	0.01	0.87
射正	0.95	-0.02	0.04	0.90
头球进球	0.52	-0.13	0.53	0.56
左脚进球	0.64	-0.04	-0.15	0.44
右脚进球	0.80	0.04	0.02	0.64
直接任意球进球	0.42	0.18	-0.13	0.23
点球	0.46	0.22	0.00	0.26
赢得点球机会	0.66	-0.05	0.03	0.44
传球	0.18	0.83	0.00	0.72
抢断	-0.08	0.91	0.00	0.83
拦截	-0.21	0.86	0.21	0.84
黄牌	0.21	0.64	0.34	0.58
犯规	0.44	0.57	0.27	0.59
解围	-0.33	0.38	0.78	0.87
头球解围	-0.30	0.35	0.80	0.86
后场解围	-0.35	0.36	0.77	0.84
头球争顶成功	0.26	0.05	0.79	0.69
红牌	0.02	0.12	0.14	0.03
乌龙球	0.02	0.21	0.14	0.06

- 因子分析的结果表明，保留三个公共因子，累计方差贡献率可以达到58.82%。
- 根据主成分法估计的因子载荷矩阵，三个因子分别为进攻因子、防守中场因子和防守后卫因子。

Example 2.3

许多美国人认为出于医学或治疗原因的堕胎是可以接受的，但他们反对出于个人偏好或个人利益的堕胎。用1984年美国GSS数据检验。

对堕胎的态度

Please tell me whether or not you think it should be possible for a pregnant woman to obtain a legal abortion...(Yes=1, No=2)

- 1. If there is a strong chance of serious defeat in the baby?
(如果胎儿很可能存在严重的缺陷?) abdefec
- 2. If she is married and does not want any more children?
(如果她已婚, 但不想要更多的孩子?) abnomor

- 3. If her health is seriously endangered by the pregnancy?
(如果她自身的健康因怀孕受到严重危害?) abhlth
- 4. If the family has a very low income and cannot afford any more children?
(如果家庭收入很低, 不能抚养更多的孩子?) abpoor

- 5. If she became pregnant as a result of rape?
(如果她是被强奸而怀孕的?) abrape
- 6. If she is not married and does not want to marry to the man?
(如果她未婚, 且不想同那个男人结婚?) absingl
- 7. If she wants it for any reason.
(如果她无论如何都想堕胎?) abany

- 1 corr abdefec abnomor abhlth abpoor abrape absingl abany
(1) 或者 *corr abdefec-abany*
- 2 factor abdefec abnomor abhlth abpoor abrape absingl abany, ipf
(1) 或者 *factor abdefec-abany, ipf*
- 3 The Eigenvalue is the amount of the total variance across all items being analyzed that is explained by each factor.

```
. factor abdefec-abany [aw=hhwt],ipf means
(sum of wgt is 1.4730e+03)
(obs=1473)
```

Factor analysis/correlation	Number of obs	=	1473
Method: iterated principal factors	Retained factors	=	6
Rotation: (unrotated)	Number of params	=	21

Factor	Eigenvalue	Difference	Proportion	Cumulative
Factor1	3.86834	2.86508	0.7658	0.7658
Factor2	1.00326	0.93092	0.1986	0.9644
Factor3	0.07233	0.00896	0.0143	0.9787
Factor4	0.06337	0.03378	0.0125	0.9912
Factor5	0.02959	0.01457	0.0059	0.9971
Factor6	0.01502	0.01526	0.0030	1.0000
Factor7	-0.00024	.	-0.0000	1.0000

LR test: independent vs. saturated: $\chi^2(21) = 6563.70$ Prob> $\chi^2 = 0.0000$

Figure 3.1: 主因子法

- 1 The cumulation is possible because the factors are orthogonal, that is, uncorrelated. Here we see that more than 96 per cent of the variance across the seven items is explained by the first two factors.
- 2 The Factor Loadings are the correlations between the observed variables and the underlying hypothetical factors, six in this case.
- 3 The Uniqueness is the proportion of variance in each item not explained by the common factors.

- 1 There is a strong first factor: all of the abortion acceptance items share a substantial amount of common variance (the variance shared between the underlying factor and the observed variable is given by the square of the loading).
- 2 But there is also a second factor that divides the items into two groups, those with positive and negative loadings. This is a typical outcome, and is somewhat awkward to interpret.

- 1 Factor matrices are rotated to improve the interpretation of the loadings.
- 2 A varimax rotation maximizes the variance of each factor, and tends to give the clearest separation of factors among the various orthogonal rotations.

因子载荷

Factor loadings (pattern matrix) and unique variances

Variable	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Uniqueness
abdefec	0.6220	0.5330	0.0500	-0.0882	-0.0853	0.0047	0.3114
abnomor	0.8641	-0.2628	0.0887	-0.1620	0.0327	0.0058	0.1491
abhlth	0.4825	0.5416	0.0408	0.0895	0.0519	0.0686	0.4567
abpoor	0.8308	-0.1834	-0.1747	-0.0303	0.0540	0.0365	0.2404
abrape	0.6124	0.4120	-0.0733	0.0146	0.0427	-0.0911	0.4394
absingl	0.8688	-0.2490	-0.0635	0.0848	-0.1129	0.0055	0.1592
abany	0.8257	-0.3022	0.1427	0.1142	0.0326	-0.0246	0.1918

初步结论

- 1 The items loading strongly on Factor 1 are those concerned with the acceptability of abortion for personal preference reasons, while those loading strongly on Factor 2 are those concerned with the acceptability of abortion for therapeutic reasons.
- 2 X_2 、 X_4 、 X_6 、 X_7 (abnomor、abpoor、absingl和abany) 主要负荷于因子1, 在因子2上的负荷很弱。
- 3 X_1 、 X_3 、 X_5 (abdefec、abhlth和abrape) 则相反。

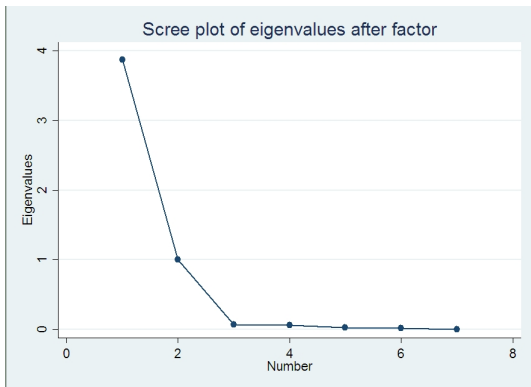


Figure 3.2: 碎石图 (命令 greigen)

因子旋转

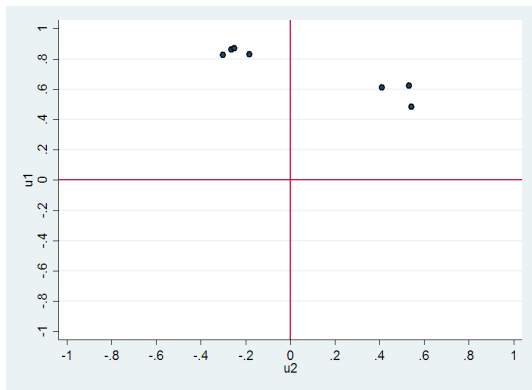


Figure 3.3: 未旋转前

因子旋转

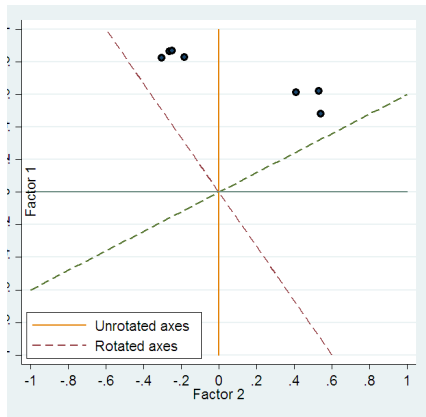


Figure 3.4: 旋转后

rotate, varimax

Rotated factor loadings (pattern matrix) and unique variances

Variable	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
abdefec	0.2736	0.7771	-0.0110	-0.0292	-0.0871	-0.0365
abncomr	0.8802	0.2082	0.0014	-0.1703	-0.0580	-0.0195
abhlth	0.1490	0.7082	0.0088	0.0307	0.1351	-0.0155
abpoor	0.8051	0.2471	0.2229	0.0143	0.0004	0.0230
abrape	0.3209	0.6569	0.0926	0.0219	0.0036	0.1304
absingl	0.8763	0.2167	0.0404	0.1510	-0.0382	-0.0030
abany	0.8704	0.1527	-0.1163	0.0161	0.1053	0.0489

Figure 3.5: 旋转后的因子载荷

因子的命名

- 1 因子1上载荷较强的四个题项：个人偏好型堕胎的态度
- 2 因子2上载荷较强的三个题项：治疗型堕胎的态度

参考文献

- 1 李连江，2017，《戏说统计：文科生的量化方法》，北京：中国政法大学出版社。
- 2 金在温、查尔斯.米勒，2012，《因子分析：统计方法与应用问题》，上海：格致出版社、上海人民出版社。
- 3 唐启明（Donald Treiman），2012，《量化数据分析——通过社会研究检验想法》.北京：社会科学文献出版社。