

SPSS实现主成分分析与因子分析（二）

作者：王江源 来源：爱科学

本文原地址：<https://www.iikx.com/news/statistics/6061.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

SPSS实现主成分分析与因子分析

(二)。SPSS没有提供单独的主成分分析方法，而是混在因子分析当中，下面通过一个例子来讨论主成分分析与因子分析的实现方法及相关问题。

一、问题提出

男子十项全能比赛包含100米跑、跳远、跳高、撑杆跳、铅球、铁饼、标枪、400米跑、1500米跑、110米跨栏十个项目，总分为各个项目得分之和。为了分析十项全能主要考察哪些方面的能力，以便有针对性的进行训练，研究者收集了134个顶级运动员的十项全能成绩单，将通过因子分析来达到分析目的。

二、分析过程

变量视图：

	名称	类型	宽度	小数	标签
1	百米跑	数值(N)	8	2	100米(秒)
2	跳远	数值(N)	8	2	跳远(米)
3	铅球	数值(N)	8	2	铅球(米)
4	跳高	数值(N)	8	2	跳高(米)
5	四百米跑	数值(N)	8	2	400米(秒)
6	一百一十米栏	数值(N)	8	2	110米栏(秒)
7	铁饼	数值(N)	8	2	铁饼(米)
8	撑杆跳	数值(N)	8	2	撑杆跳(米)
9	标枪	数值(N)	8	2	标枪(米)
10	一千五百米跑	数值(N)	8	2	1500米(秒)
11	总分	数值(N)	8	0	总分

数据视图(部分)：

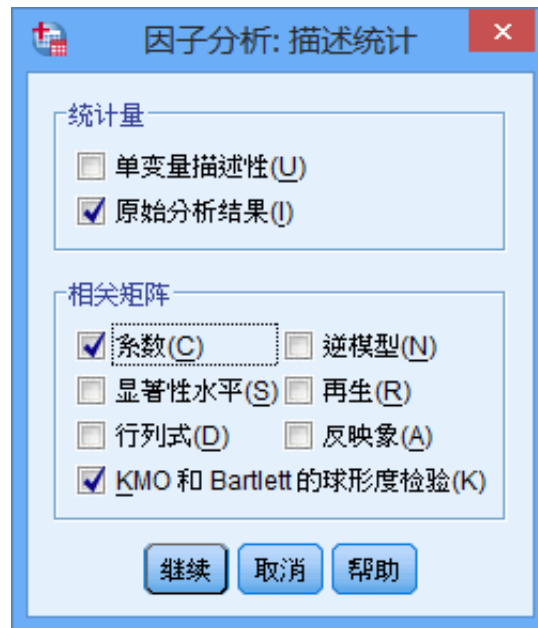
	百米跑	跳远	铅球	跳高	四百米跑	一百一十米栏	铁饼
1	11.25	7.43	15.48	2.27	48.90	15.13	49.28
2	10.87	7.45	14.97	1.97	47.71	14.46	44.36
3	11.18	7.44	14.20	1.97	48.29	14.81	43.66
4	10.62	7.38	15.02	2.03	49.06	14.72	44.80
5	11.02	7.43	12.92	1.97	47.44	14.40	41.20
6	10.83	7.72	13.58	2.12	48.34	14.18	43.06
7	11.18	7.05	14.12	2.06	49.34	14.39	41.68
8	11.05	6.95	15.34	2.00	48.21	14.36	41.32
9	11.15	7.12	14.52	2.03	49.15	14.66	42.36
10	11.23	7.28	15.25	1.97	48.60	14.76	48.02
11	10.94	7.45	15.34	1.97	49.94	14.25	41.86
12	11.18	7.34	14.48	1.94	49.02	15.11	42.76
13	11.02	7.29	12.92	2.06	48.23	14.94	39.54
14	10.99	7.37	13.61	1.97	47.83	14.70	43.88
15	11.03	7.45	14.20	1.97	48.94	15.44	41.66
16	11.09	7.08	14.51	2.03	49.89	14.78	43.20
17	11.46	6.75	16.07	2.00	51.28	16.06	50.66

菜单选择(分析->降维->因子分析)：

打开因子分析的主界面，将十项成绩选入”变量“框中(不要包含总分)，如下：



点击”描述“按钮，打开对话框，选中”系数“和”KMO和Bartlett球形度检验“：



上图相关解释：

”系数“：为变量之间的相关系数阵列，可以直观的分析相关性。

”KMO和Bartlett球形度检验“：用于定量的检验变量之间是否具有相关性。

点击”继续“，回到主界面，点击”抽取“，打开对话框。

”方法“=>”主成分“，”输出“=>”未旋转的因子解“和”碎石图“，”抽取“=>”基于特征值“，其余选择默认。



解释：

因子抽取的方法：选取默认的主成分法即可，其余方法的计算结果可能有所差异。

输出：“未旋转的因子解”极为主成分分析结果。碎石图有助于我们判断因子的重要性(详细介绍见后面)。

抽取：为抽取主成分(因子)的方法，一般是基于特征值大于1，默认即可。

点击“继续”，回到主界面，点击“确定”，进入分析。

输出的主要表格如下：

(1)相关性检验

因子分析要求变量之间有相关性，所以首先要进行相关性检验。首先输出的是变量之间的相关系数矩阵：

相关矩阵

	100米(秒)	跳远(米)	铅球(米)	跳高(米)	400米(秒)	110米栏(秒)	铁饼(米)	撑杆跳(米)	标枪(米)	1500米(秒)
相关 100米(秒)	1.000	-.691	-.420	-.364	.698	.751	-.353	-.627	-.344	.254
跳远(米)	-.691	1.000	.391	.471	-.636	-.654	.375	.632	.446	-.356
铅球(米)	-.420	.391	1.000	.321	-.142	-.489	.856	.643	.703	.202
跳高(米)	-.364	.471	.321	1.000	-.275	-.487	.376	.472	.338	-.132
400米(秒)	.698	-.636	-.142	-.275	1.000	.655	-.154	-.521	-.150	.554
110米栏(秒)	.751	-.654	-.489	-.487	.655	1.000	-.403	-.709	-.350	.155
铁饼(米)	-.353	.375	.856	.376	-.154	-.403	1.000	.620	.618	.288
撑杆跳(米)	-.627	.632	.643	.472	-.521	-.709	.620	1.000	.557	-.070
标枪(米)	-.344	.446	.703	.338	-.150	-.350	.618	.557	1.000	.045
1500米(秒)	.254	-.356	.202	-.132	.554	.155	.288	-.070	.045	1.000

可以直观的看到，变量之间有相关性。但需要检验，接着输出的是相关性检验：

KMO 和 Bartlett 的检验	
取样足够度的 Kaiser-Meyer-Olkin 度量。	.788
Bartlett 的球形度检验 近似卡方	211.586
df	45
Sig.	.000

上图有两个指标：第一个是KMO值，一般大于0.7就说明不了之间有相关性了。第二个是Bartlett球形度检验，P值<0.001。综合两个指标，说明变量之间存在相关性，可以进行因子分析。否则，不能进行因子分析。

(2)提取主成分和公因子

接下来输出主成分结果：

解释的总方差						
成份	初始特征值			提取平方和载入		
	合计	方差的 %	累积 %	合计	方差的 %	累积 %
1	5.024	50.235	50.235	5.024	50.235	50.235
2	2.080	20.799	71.034	2.080	20.799	71.034
3	.735	7.355	78.389			
4	.686	6.857	85.246			
5	.376	3.763	89.009			
6	.302	3.021	92.030			
7	.286	2.855	94.885			
8	.224	2.238	97.123			
9	.205	2.047	99.170			
10	.083	.830	100.000			

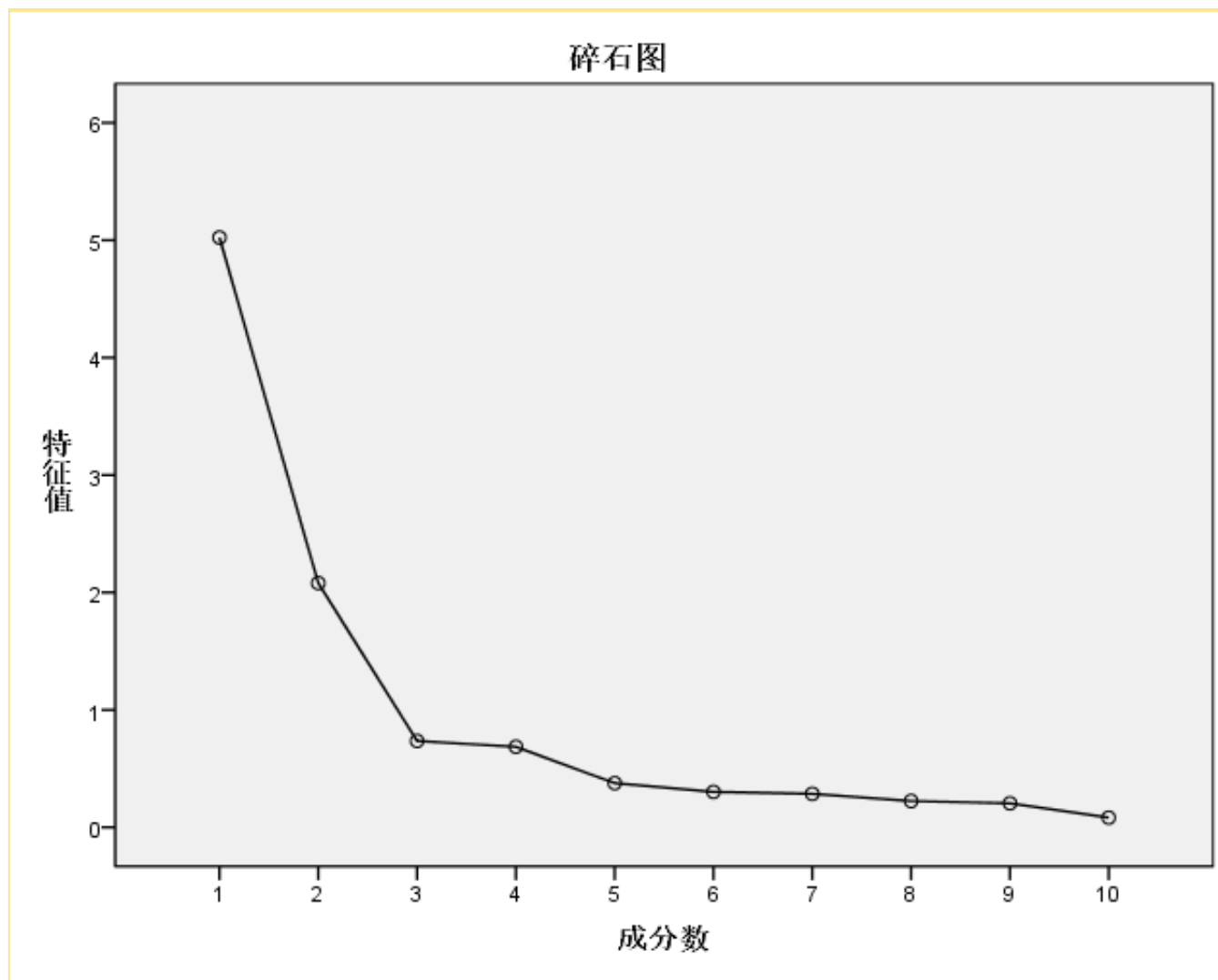
提取方法：主成份分析。

这就是主成分分析的结果，表中第一列为10个成分;第二列为对应的”特征值“，表示所解释的方差的大小;第三列为对应的成分所包含的方差占总方差的百分比;第四列为累计的百分比。一般

来说，选择”特征值“大于1的成分作为主成分，这也是SPSS默认的选择。

在本例中，成分1和2的特征值大于1，他们合计能解释71.034%的方差，还算不错。所以我们可以提取1和2作为主成分，抓住了主要矛盾，其余成分包含的信息较少，故弃去。

下面，输出碎石图，如下：



碎石图来源于地质学的概念。在岩层斜坡下方往往有很多小的碎石，其地质学意义不大。碎石图以特征值为纵轴，成分为横轴。前面陡峭的部分特征值大，包含的信息多，后面平坦的部分特征值小，包含的信息也小。

由图直观的看出，成分1和2包含了大部分信息，从3开始就进入平台了。

接下来，输出提取的成分矩阵：

	成份	
	1	2
100米(秒)	-.804	.294
跳远(米)	.810	-.285
铅球(米)	.726	.569
跳高(米)	.600	-.011
400米(秒)	-.660	.616
110米栏(秒)	-.837	.189
铁饼(米)	.687	.601
撑杆跳(米)	.872	.089
标枪(米)	.657	.430
1500米(秒)	-.187	.787

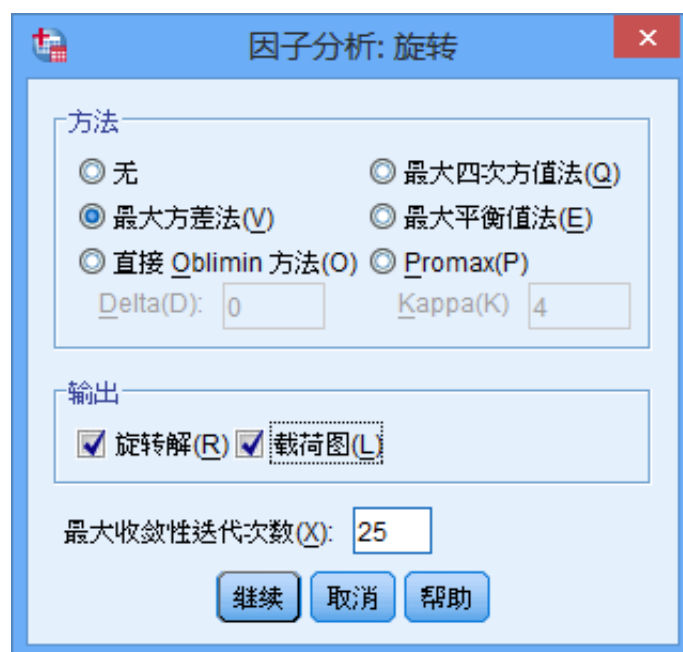
提取方法:主成份。

a. 已提取了 2 个成份。

上表中的数值为公因子与原始变量之间的相关系数，绝对值越大，说明关系越密切。公因子1和9个运动项目都正相关(注意跑步运动运动的计分方式，时间越短，分数越高)，看来只能称为“综合运动”因子了。公因子2与铁饼、铅球正相关，与1500米跑、400米跑负相关，这究竟代表什么意思呢?看来只能成为“不知所云”因子了。

(三)因子旋转

前面提取的两个公因子一个是大而全的“综合因子”，一个不知所云，得到这样的结果，无疑是分析的失败。不过，不要灰心，我们可以通过因子的旋转来获得更好的解释。在主界面中点击“旋转”按钮，打开对话框，“方法”=>“最大方差法”，“输出”=>“旋转解”。



点击“继续”，回到主界面点击“确认”进行分析。输出结果如下：

	成份	
	1	2
100米(秒)	.785	-.341
跳远(米)	-.783	.352
铅球(米)	-.134	.913
跳高(米)	-.442	.406
400米(秒)	.903	-.009
110米栏(秒)	.737	-.440
铁饼(米)	-.084	.909
撑杆跳(米)	-.570	.666
标枪(米)	-.179	.765
1500米(秒)	.678	.441

提取方法:主成份。
旋转法:具有 Kaiser 标准化的正交旋转法。

a. 旋转在 3 次迭代后收敛。

这是选择后的成分矩阵。经过旋转，可以看出：

公因子1得分越高，所有的跑步和跨栏成绩越差，而跳远、撑杆跳等需要助跑类项目的成绩也越差，所以公因子1代表的是奔跑能力的反向指标，可称为“奔跑能力”。

公因子2与铁饼和铅球的正相关性很高，与标枪、撑杆跳等需要上肢力量的项目也正相关，所以该因子可以成为“上肢力量”。

经过旋转，可以看出公因子有了更合理的解释。

(四)结果的保存

在最后，我们还要将公因子储存下来供后续使用。点击“得分”按钮，打开对话框，选中“保存为变量”，方法采用默认的“回归”方法，同时选中“显示因子得分系数矩阵”。



SPSS会自动生成2个新变量，分别为公因子的取值，放在数据的最后。同时会输出一个因子系数表格：

	成份	
	1	2
100米(秒)	-.160	.142
跳远(米)	.161	-.137
铅球(米)	.145	.273
跳高(米)	.119	-.005
400米(秒)	-.131	.296
110米栏(秒)	-.167	.091
铁饼(米)	.137	.289
撑杆跳(米)	.174	.043
标枪(米)	.131	.207
1500米(秒)	-.037	.378

提取方法：主成份。
构成得分。

由上图，我们可以写出公因子的表达式(用F1、F2代表两个公因子，Z1~Z10分别代表原始变量)：

$$F1 = -0.16 \cdot Z1 + 0.161 \cdot Z2 + 0.145 \cdot Z3 + 0.119 \cdot Z4 - 0.131 \cdot Z5 - 0.167 \cdot Z6 + 0.137 \cdot Z7 + 0.174 \cdot Z8 + 0.131 \cdot Z9 - 0.037 \cdot Z10$$

F2同理，略去。

注意，这里的变量Z1~Z10，F1、F2不再是原始变量，而是标准正态变换后的变量。

更多 统计方法 请访问 <https://www.iikx.com/news/statistics/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](http://www.iikx.com)转发