### 分析流程 数据源： all\_data\_processed4\_pro.csv 算法配置： 算法： 主成分分析(PCA) 变量： 变量：{ age\_at\_event mean，Average Net Result mean，Mark mean，Mark min，current\_match\_ranking mean，Mark var，Average Rank mean，Nat，Years to Olympics，efficiency mean，current\_match\_ranking min }；索引项：{ Olympic Ranking }收起 参数： 主成分个数：{ 1 } 分析结果： 主成分分析通过少数几个主成分尽可能解释所有变量的信息，达到降维的目的：KMO的值为0.668，同时，Bartlett球形检验的结果显示，显著性P值为0.000\*\*\*，水平上呈现显著性，拒绝原假设，各变量间具有相关性，主成分分析有效，程度为不太适合。可以根据特征根（特征根<1）或累计方差解释率（累计方差解释率>90%）来选择最终需要保留的主成分，请看详细结论。

### 分析步骤 1. 首先进行KMO和Bartlett的检验，判断是否可以进行主成分分析。 对于KMO值：0.8上非常合适做主成分分析，0.7-0.8之间一般适合，0.6-0.7之间不太适合，0.5-0.6之间表示差，0.5下表示极不适合，对于 Bartlett的检验，若P小于0.05，拒绝原假设，则说明可以做主成分分析，若不拒绝原假设，则说明这些变量可能独立提供一些信息，不适合做主成分分析。 2. 通过分析方差解释表格和碎石图，确定主成分的数量方差解释表格主要是看主成分对于变量解释的贡献率（可以理解为究竟需要多少主成分才能把变量表达为100%），如果太低（如低于60%）则需要调整主成分数据。 碎石图的作用是根据特征值下降的坡度来确认需要选择的主成分个数，这两者结合可用于确认或调整主成分个数。 3. 通过分析主成分载荷系数与热力图，可以分析到每个主成分中隐变量的重要性，如研究【多金属矿体】中25种有用元素的分布规律，其中各元素视为指标，假设前文确定得到5个主成分，主成分1中，SO、SO2、Na2S、HS、H2S主成分载荷系数较大，因此可将主成分1确定为硫化物成分，以此类推，也可结合具体业务进行各主成分的隐变量分析。 4. 基于主成分载荷图通过将多主成分降维成双主成分或者三主成分，通过象限图的方式呈现主成分的空间分布。如果提取2个主成分时，无法呈现三维载荷主成分散点图，如果提取1个主成分时，无法显示主成分象限图。 5. 通过分析成分矩阵，得出主成分成分公式与权重。 6. 输出主成分分析法综合得分。

### 详细结论

**输出结果1：KMO检验和Bartlett的检验**

|  |  |
| --- | --- |
| KMO检验和Bartlett的检验 | |
| KMO值 | | 0.668 |
| Bartlett球形度检验 | 近似卡方 | 34615.232 |
| df | 55 |
| P | 0.000\*\*\* |
| 注：\*\*\*、\*\*、\*分别代表1%、5%、10%的显著性水平 | |

**图表说明：**

上表展示了KMO检验和Bartlett球形检验的结果，用来分析是否可以进行主成分分析。  
● 若通过KMO检验（KMO>0.6），说明了题项变量之间是存在相关性的，符合主成分分析要求；  
● 若通过Bartlett检验：P<0.05，呈显著性，则可以进行主成分分析。

**智能分析：**

KMO检验的结果显示，KMO的值为0.668，同时，Bartlett球形检验的结果显示，显著性P值为0.000\*\*\*，水平上呈现显著性，拒绝原假设，各变量间具有相关性，主成分分析有效，程度为不太适合。

**输出结果2：方差解释表格**

|  |  |  |  |
| --- | --- | --- | --- |
| 总方差解释 | | | |
| 成分 | 特征根 | | |
| 特征根 | 方差解释率(%) | 累积方差解释率(%) |
| 1 | 3.993 | 36.303 | 36.303 |
| 2 | 1.708 | 15.532 | 51.835 |
| 3 | 1.374 | 12.493 | 64.328 |
| 4 | 0.996 | 9.052 | 73.38 |
| 5 | 0.976 | 8.877 | 82.257 |
| 6 | 0.815 | 7.405 | 89.662 |
| 7 | 0.6 | 5.453 | 95.115 |
| 8 | 0.286 | 2.601 | 97.716 |
| 9 | 0.153 | 1.39 | 99.106 |
| 10 | 0.066 | 0.601 | 99.707 |
| 11 | 0.032 | 0.293 | 100 |

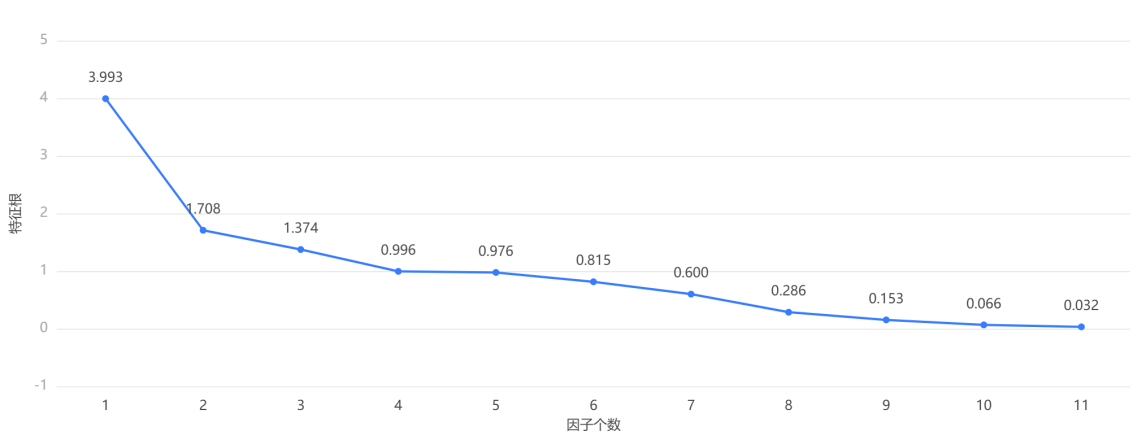
**图表说明：**

上表为总方差解释表格，主要是看主成分对于变量解释的贡献率（可以理解为究竟需要多少主成分才能把变量表达为100%），一般都要表达到90%以上才可以，否则就要调整因子数据。  
● 一般情况下，方差解释率越高，说明该主成分越重要，权重占比也应该越高。

**智能分析：**

方差解释表中，在主成分4时，总方差解释的特征根低于1，变量解释的贡献率达到73.38，以上仅为参考，若特征根小于1临界值过大，也可以集合具体情况具体分析。

**输出结果3：碎石图**



**图表说明：**

● 碎石图是根据各主成分对数据变异的解释程度绘制的图。其作用是根据特征值下降的坡度来确认需要选择的主成分个数，结合方差解释表可用于确认或调整主成分个数。  
● 每一个主成分为一个点，通过“坡度趋于平缓”的位置判断提取主成分的数量。

**输出结果4：因子载荷系数表**

|  |  |  |
| --- | --- | --- |
| 因子载荷系数表 | | |
|  | 因子载荷系数 | 共同度（公因子方差） |
| 主成分1 |
| age\_at\_event mean | -0.062 | 0.004 |
| Average Net Result mean | 0.742 | 0.551 |
| Mark mean | 0.924 | 0.854 |
| Mark min | 0.949 | 0.9 |
| current\_match\_ranking mean | 0.15 | 0.023 |
| Mark var | -0.73 | 0.532 |
| Average Rank mean | 0.884 | 0.781 |
| Nat | -0.354 | 0.125 |
| Years to Olympics | 0.005 | 0.000 |
| efficiency mean | -0.244 | 0.059 |
| current\_match\_ranking min | 0.404 | 0.163 |

**图表说明：**

上表为因子载荷系数表，可以分析到每个主成分中隐变量的重要性。  
● 假设前文确定得到n个因子，因子 i 中a、b、c、d的因子载荷系数较大，因此可将主成分 i 进行总结重命名。

**输出结果5：因子载荷矩阵热力图**



**图表说明：**

上图为载荷矩阵热力图，可以分析到每个主成分中隐变量的重要性。同时可结合具体业务进行各因子的隐变量分析。

**输出结果6：成分矩阵表**

|  |  |
| --- | --- |
| 成分矩阵表 | |
| 名称 | 成分 |
| 成分1 |
| age\_at\_event mean | -0.016 |
| Average Net Result mean | 0.186 |
| Mark mean | 0.231 |
| Mark min | 0.238 |
| current\_match\_ranking mean | 0.038 |
| Mark var | -0.183 |
| Average Rank mean | 0.221 |
| Nat | -0.089 |
| Years to Olympics | 0.001 |
| efficiency mean | -0.061 |
| current\_match\_ranking min | 0.101 |

**图表说明：**

上表为成分矩阵表，意在说明各个成分的所包含的因子得分系数（主成分载荷），用于计算出成分得分，得出因子公式，其计算公式为：线性组合系数\*（方差解释率/累积方差解释率），最后将其归一化即为因子权重得分。  
● 线性组合系数，公式为：因子载荷系数除以对应特征根，即成分矩阵的系数。

**智能分析：**

模型的公式：  
F1=-0.016×age\_at\_event mean+0.186×Average Net Result mean+0.231×Mark mean+0.238×Mark min+0.038×current\_match\_ranking mean-0.183×Mark var+0.221×Average Rank mean-0.089×Nat+0.001×Years to Olympics-0.061×efficiency mean+0.101×current\_match\_ranking min  
由上可以得到：  
F=F1

**输出结果7：因子权重分析**

|  |  |  |  |
| --- | --- | --- | --- |
| 名称 | 方差解释率(%) | 累积方差解释率(%) | 权重(%) |
| 主成分1 | 3.993 | 36.303 | 100% |

**图表说明：**

上表为主成分分析的根据载荷系数等信息所做的主成分权重分析，其计算公式为：方差解释率/旋转后累积方差解释率。

**智能分析：**

主成分分析的权重计算结果显示，主成分1的权重为100%、

**输出结果8：综合得分表**

|  |  |  |  |
| --- | --- | --- | --- |
| 排名 | 行索引 | 综合得分 | 主成分1 |
| 1 | 9 | 2.621 | 2.621 |
| 2 | 9 | 2.516 | 2.516 |
| 3 | 9 | 2.46 | 2.46 |
| 4 | 9 | 2.457 | 2.457 |
| 5 | 9 | 2.425 | 2.425 |
| 6 | 9 | 2.393 | 2.393 |
| 7 | 9 | 2.383 | 2.383 |
| 8 | 9 | 2.358 | 2.358 |
| 9 | 9 | 2.35 | 2.35 |
| 10 | 9 | 2.277 | 2.277 |
| 11 | 9 | 2.26 | 2.26 |
| 12 | 9 | 2.247 | 2.247 |
| 13 | 8 | 2.233 | 2.233 |
| 14 | 9 | 2.214 | 2.214 |
| 15 | 9 | 2.202 | 2.202 |

**图表说明：**

上表格为预览结果，只显示综合排序的前15条数据，全部数据请点击下载按钮导出。  
综合得分根据F值计算得到的综合得分进行降序排序，可得到各个样本的综合得分与排名情况，同时输出因子浓缩后的结果。

### 参考文献 [1] Scientific Platform Serving for Statistics Professional 2021. SPSSPRO. (Version 1.0.11)[Online Application Software]. Retrieved from https://www.spsspro.com. [2] 何晓群.多元统计分析.北京：中国人民大学出版社，2012.

\*主成分分析—因子载荷系数表

主成分分析（PCA）的因子载荷系数表为理解数据中潜在的主成分提供了关键信息。该表揭示了各原始变量与主成分之间的相关性及每个变量在提取主成分时的贡献程度。以下是对给定数据的详细解读：

首先，观察主成分1的因子载荷系数，可以发现“Average Net Result mean”、“Mark mean”、“Mark min”和“Average Rank mean”的载荷系数绝对值均超过0.7，表明这些变量与主成分1之间存在高度的正相关关系。这些变量在提取主成分1时贡献了显著的信息量，因此可以认为主成分1主要反映了这些变量的变化特征。同时，“Nat”和“Years to Olympics”的载荷系数绝对值较小，表明它们与主成分1的相关性较弱，对主成分1的贡献较小。

其次，注意到“Mark var”的载荷系数为负值（-0.73），表示该变量与主成分1之间存在负相关关系。然而，由于其绝对值较大，说明该变量在提取主成分1时同样贡献了显著的信息量，尽管方向相反。

此外，共同度（公因子方差）提供了每个变量被主成分解释的程度的量化指标。例如，“Mark mean”和“Mark min”的共同度分别为0.551和0.9，表明这两个变量的大部分变异可以由主成分1来解释。而“age\_at\_event mean”和“Years to Olympics”的共同度较低，说明这两个变量在主成分1中的代表性较弱。