列车选取说明

# 一、需求

根据列车的运行时间、车组数、途径城市、车厢数等指标选取列车。使选取尽可能少的列车覆盖41个目标城市。

# 二、选取标准：

覆盖41个目标城市

列车运行时间长

硬座车厢多

# 三、数据整理

根据列车选取要求及标准将列车指标整理为：运行时间、车组数、途径城市数量、硬座数量、经过目标城市数量。

# 四、数据分析思路：

**Step1：**从指标属性上将指标分为三类，第一类评价列车运行情况，第二类评价列车的硬座情况，第三类为经过目标城市情况。

**Step2：**通过对第一类、第二类指标计算出列车的综合评分，再通过综合得分\*经过目标城市数量得到列车的最终得分。最终得分表示该列车无论是运行时间、车组数、途径城市数量、硬座数量以及经过目标城市数量各指标的表现均比较好。

**Step3：**通过将列车按经过的目标站点进行分类，并按降序排列，最终选出通过该目标站点的最优列车。

# 五、分析方法：

1、指标分类

采用因子降维（详情见附件一）思想，将指标进行分类：

|  |  |
| --- | --- |
| **类别** | **指标** |
| 第一类  列车运行情况 | 运行时间 |
| 车组数 |
| 途径城市数量 |
| 第二类  列车硬座情况 | 硬座数量 |
| 第三类  经过目标城市情况 | 经过目标城市数量 |

2、列车综合得分计算

根据第一类、第二类的因子贡献率，计算出各因子的权重（详情见附件一）。列车综合得分计算公式：

通过列车综合得分高低可以初步分辨出列车的优劣情况。得分越高表示该列车第一类和第二类的各项指标均表现优良。得分越低则相反。

3、列车最终得分计算

由于列车综合得分的高低能初步分辨列车的优劣，但考虑到选取的列车要覆盖大量的目标城市，因此得到最终得分的计算方式：

该最终得分表示出该列车不仅第一类、第二类的各项指标表现优良外同时表示经过的目标站点数量较多。

4、列车选取

通过目标城市将列车进行分类，并将列车的最终得分进行降序排列，得分最高者即为经过该目标城市最优的列车。

# 附件一：

**列车指标降维处理及各类别权重确定（详细）**

# 一、列车指标降维处理

（一）因子分析检验：

利用SPSS软件的因子分析功能，数据进行处理得到：

| **表一 KMO 和 Bartlett 的检验** | | |
| --- | --- | --- |
| 取样足够度的 Kaiser-Meyer-Olkin 度量。 | | .606 |
| Bartlett 的球形度检验 | 近似卡方 | 154.697 |
| df | 6 |
| Sig. | .000 |

可以从上表中知道，可以对列车的各类指标进行降维处理，并且具有统计学意义。

（二）因子选取

| **表二 解释的总方差** | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 成份 | 初始特征值 | | | 提取平方和载入 | | | 旋转平方和载入 | | |
| 合计 | 方差的 % | 累积 % | 合计 | 方差的 % | 累积 % | 合计 | 方差的 % | 累积 % |
| 1 | 2.209 | 53.969 | 53.969 | 2.159 | 53.969 | 53.969 | 2.159 | 53.941 | 53.941 |
| 2 | 1.165 | 26.943 | 80.912 | 1.078 | 26.943 | 80.912 | 1.078 | 30.400 | 84.342 |
| 3 | .571 | 14.282 | 95.194 |  |  |  |  |  |  |
| 4 | .192 | 4.806 | 100.000 |  |  |  |  |  |  |
| 提取方法：主成份分析。 | | | | | | | | | |

从上表中可以看到，第一类、第二类指标对总体的累计方差贡献率达到84.342%，因此可以选取第一类指标、第二类指标进行分析。

（三）因子分类：

| **旋转成份矩阵a** | | |
| --- | --- | --- |
| 指标 | 成份 | |
| 1 | 2 |
| 运行时间 | .926 |  |
| 车组数 | .825 | -.243 |
| 硬座数 |  | .968 |
| 经过站点数 | .838 | .352 |
| 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。 | | |

从旋转成分矩阵中可以了解到：

1、第一类中包含3个指标，即运行时长、车组数、经过站点数，这些指标和列车的运行情况有关，因此将第一类命名为列车运行情况；

2、第二类中包含1个指标，即硬座数，该指标表示了列车的组成情况，因此将第二类命名为列车的硬座情况。

（四）因子得分计算

| **表四 成份得分系数矩阵** | | |
| --- | --- | --- |
| 指标 | 成份 | |
| 1 | 2 |
| 运行时间 | .415 | -.027 |
| 车组数 | .380 | -.253 |
| 硬座数 | -.041 | .868 |
| 经过站点数 | .361 | .281 |
| 提取方法 :主成份。  旋转法 :具有 Kaiser 标准化的正交旋转法。  构成得分。 | | |

第一类得分：

Y1=0.415\*运行时间+0.380\*车组数+0.041\*硬座数+0.361\*经过站点数

第二类得分：

Y2=-0.027\*运行时间+（-0.253）\*车组数+0.868\*硬座数+0.281\*经过站点数

# 二、第一类、第二类指标的权重确定

根据各类别的总体方差贡献率，计算出权重：

|  |  |  |
| --- | --- | --- |
| **类别** | **总体方差贡献率** | **权重** |
| 第一类 | 53.94% | 0.64 |
| 第二类 | 30.40% | 0.36 |
| 合计 | 84.34% | - |