


# 因子分析在python上的实现

原创 阳望 于 2020-03-24 14:19:09 发布 4494 收藏 77

分类专栏：机器学习

版权

 机器学习 专栏收录该内容

0 订阅 11 篇文章 订阅专栏

**因子分析**（Factor Analysis）是指研究从变量群中提取共性因子的统计技术，这里的共性因子指的是不同变量之间内在的隐藏因子。例如，一个学生的数学、物理、化学成绩都很好，那么潜在的共性因子可能是智力水平高。因此，因子分析的过程其实是寻找共性因子和个性因子并得到最优解释的过程。

因子分析有三个核心问题：一是检验是否适合因子分析，二是如何构造因子变量，三是如何对因子变量进行命名解释。

检验数据是否适合因子分析的方法有kmo检验和Bartlett's球状检验。

KMO值越接近于1，意味着变量间的相关性越强，原有变量越适合作**因子分析**；

Bartlett's球状检验是一种数学术语。用于检验相关阵中各变量间的相关性，是否为单位阵，即检验各个变量是否各自独立。因子分析前，首先进行KMO检验和巴特利球体检验。在因子分析中，若拒绝原假设（p值小于等于0.01），则说明可以做因子分析，若不拒绝原假设，则说明这些变量可能独立提供一些信息，不适合做因子分析。

检测类别	值的范围	因子分析适合情况
KMO值	大于0.9	非常适合
	0.8~0.9	很适合
	0.7~0.8	适合
	0.6~0.7	不太适合
	0.5~0.6	勉强适合
	小于0.5	不适合
BartlettP值	小于或者等于0.01	适合

## 案例

数据集：学生期末考试成绩

姓名	语文	数学	英语	政治	历史	物理	化学
小明	82	83.1	87	73.8	92.4	80.1	76.5
小李	77	90.0	82	69.3	88.9	87.0	82.0
小张	57	84.4	62	51.3	74.9	81.4	77.5
小王	53	84.0	58	47.7	72.1	81.0	77.2
小陈	83	89.7	88	74.7	93.1	86.7	81.7
小吴	81	73.7	86	72.9	91.7	70.7	69.0
小孙	74	70.1	79	66.6	86.8	67.1	66.1
小周	61	70.3	66	54.9	77.7	67.3	66.3
小郑	73	73.2	78	65.7	86.1	70.2	68.6

小钟	80	70.1	85	72	91	67.1	66.0
----	----	------	----	----	----	------	------

需求：提取七个科目的共性因子，对10位同学的综合能力进行量化。

python代码

```
from factor_analyzer import FactorAnalyzer
import pandas as pd
import seaborn as sns
from matplotlib import pyplot as plt
import numpy as np
from scipy.stats import bartlett
#参数设置
data=pd.read_excel('期末考试.xlsx')
data=data.set_index('姓名')
n_factors=2#因子数量

#用检验是否进行
corr=list(data.corr().to_numpy())
bartlett(*corr)

#开始计算
fa = FactorAnalyzer(n_factors=n_factors,method='principal',rotation="varimax")
fa.fit(data)
communalities= fa.get_communalities()#共性因子方差
loadings=fa.loadings_#成分矩阵，可以看出特征的归属因子

#画图
plt.figure()
ax = sns.heatmap(loadings, annot=True, cmap="BuPu")
plt.title('Factor Analysis')

factor_variance = fa.get_factor_variance()#贡献率
fa_score = fa.transform(data)#因子得分

#综合得分
complex_score=np.zeros([fa_score.shape[0],])
for i in range(n_factors):
    complex_score+=fa_score[:,i]*factor_variance[1][i]#综合得分
```