模型的建立

主成分分析是一种降维算法，它能将多个指标转换为少数几个主成分，这些主成分是原始变量的线性组合，且彼此之间互不相关，其能反映出原始数据的大部分信息。计算步骤如下：

记样本矩阵为x



将数据进行标准化得到矩阵X，计算X的协方差矩阵R



其中，



计算矩阵R的特征值，特征向量

计算主成分贡献率以及累计贡献率，

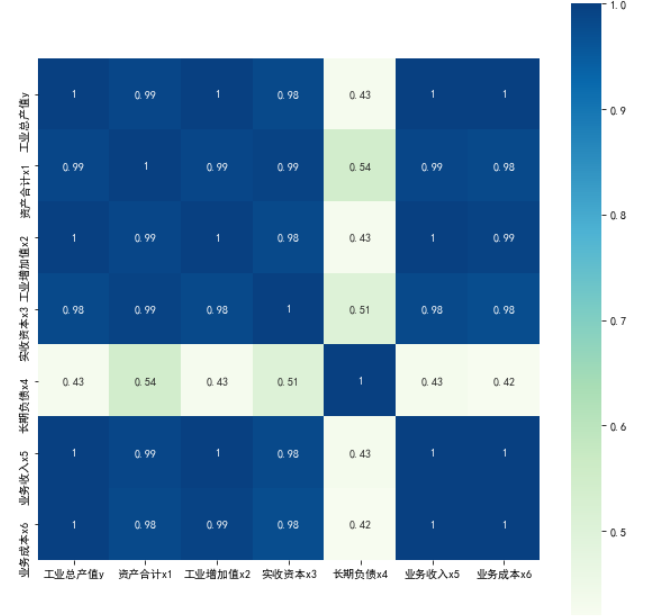


选取贡献率较高的(一般为前两个)主成分，则第i个主成分：。

对于某个主成分而言，指标前面的系数越大，表明该指标对该主成分的影响越大。

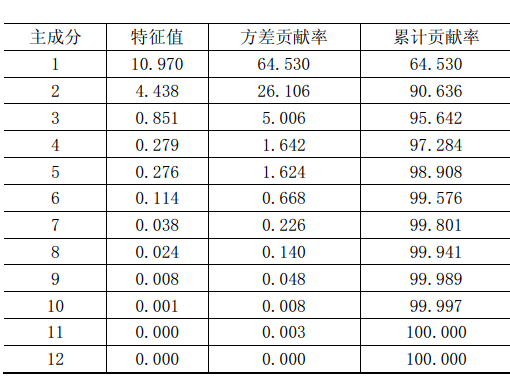
模型的求解

关系系数矩阵

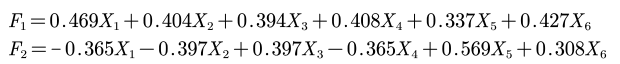


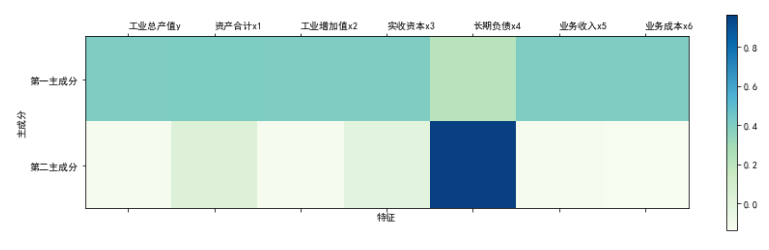
可以看出，许多变量之间具有很强的相关性，因此数据适合利用主成分分析进行降维。

进行主成分降维：



从图表中可以看到，前两个主成分的累计贡献率已经达到了90%，已经代表了原始数据大部分的信息，从而可以选取前两个主成分来代表原始数据。其中





(颜色深度表示贡献率)

一通解释