对2018的数据进行分析和排名，用于解决第三问：

原理运用：主成分分析法

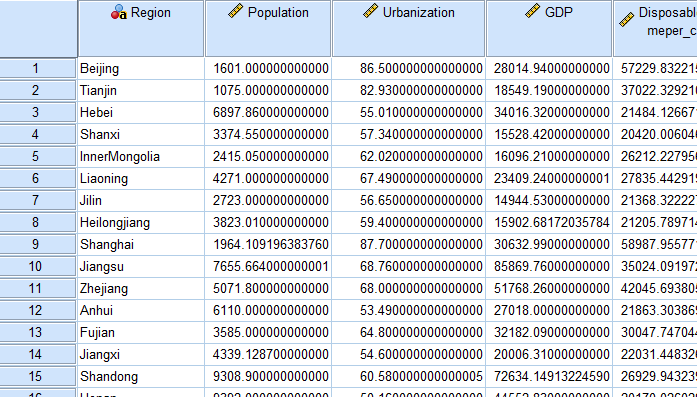
参考文献：

<https://wenku.baidu.com/view/86f92fdc336c1eb91b375d2c.html>（模型建立的原理可以抄这里面的）

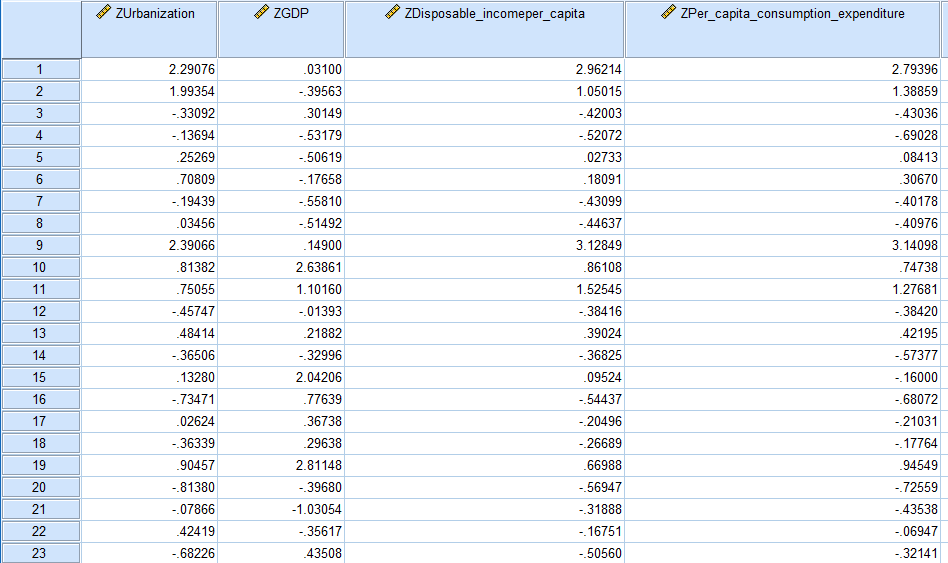
原理上就是对数据降降维，将多个指标转化为少数几个综合指标，作为主成分（具体见上面文献）

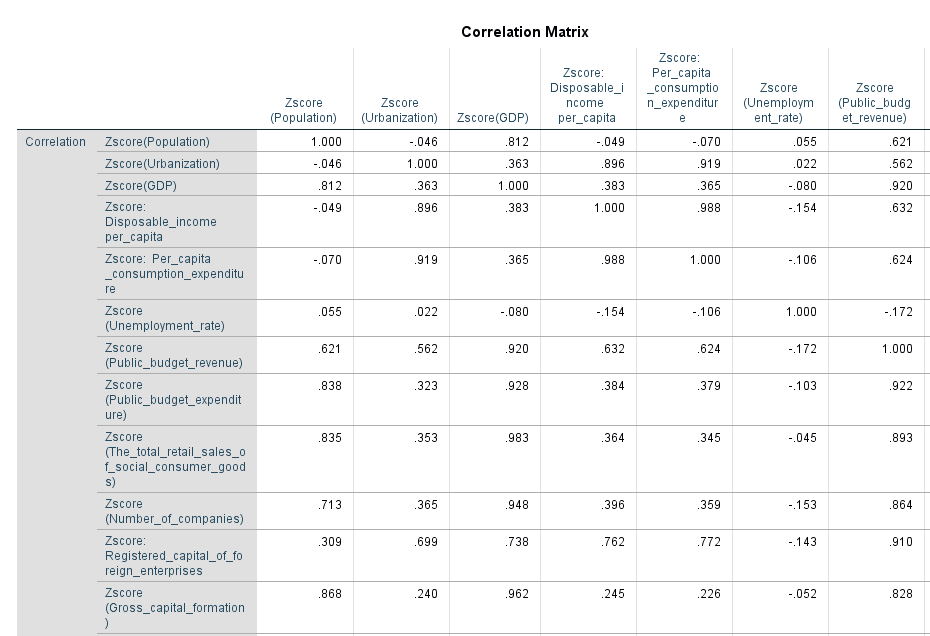
（步骤需要扩充，这里只给出每步大概做了些啥）

1. 先使用SPSS导入数据，见下图

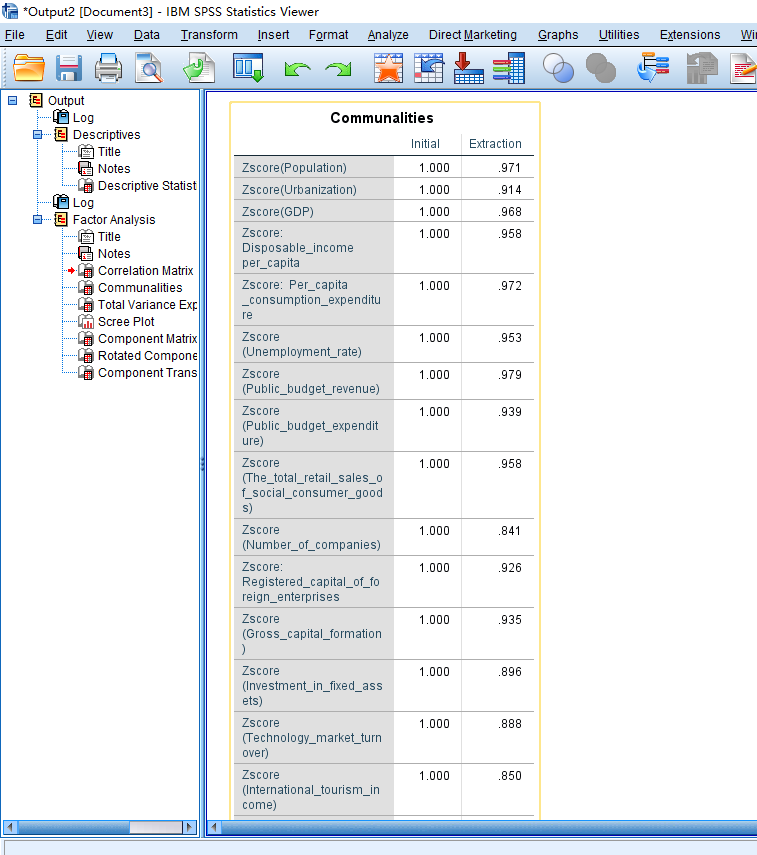


1. 对数据做了标准化处理，见下图1，使用SPSS软件对数据降维，先得到相关系数矩阵，见下图2

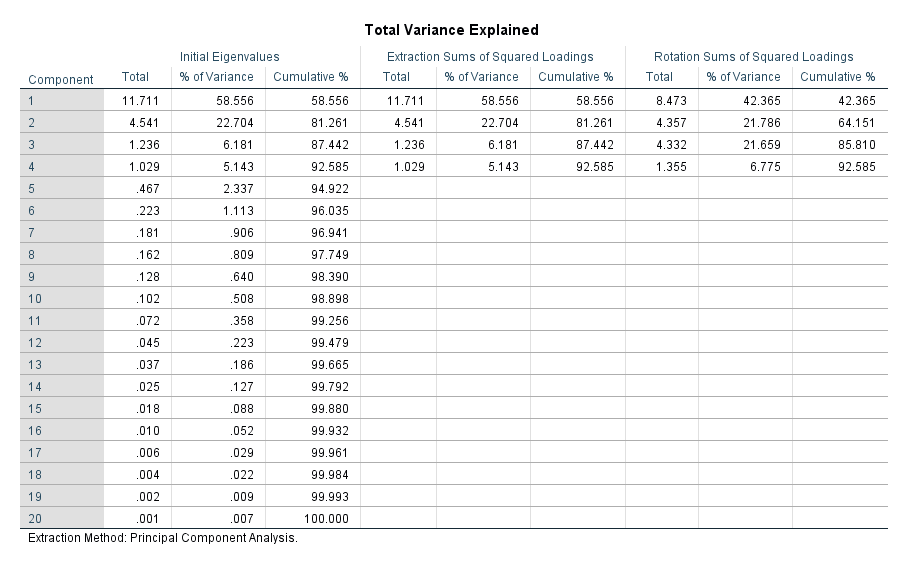




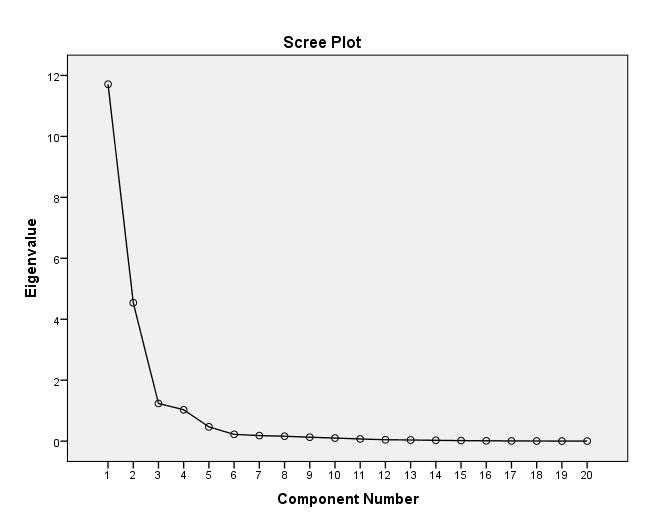
1. 提取主成分，查看信息保存率如下图，大部分数据都在90%以上，因此主成分分析法得到验证，是可行的方法（这一块也可以放到最前面，先分析该方法的可行性）



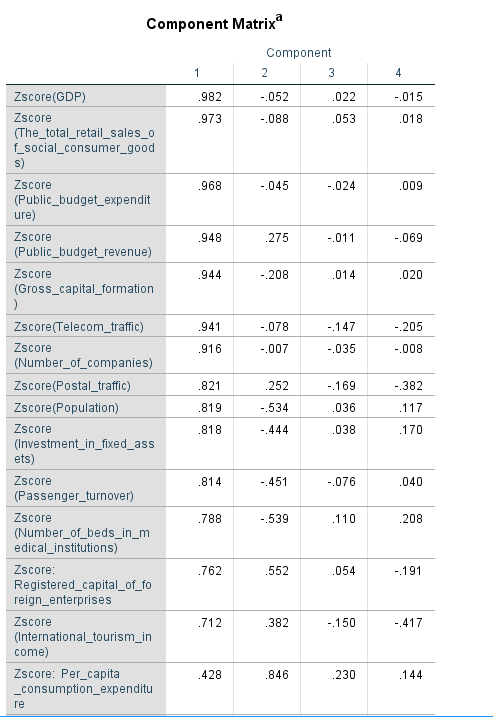
1. 接下来分析主成分和贡献率等，这一块由相关系数矩阵计算得来，筛除掉特征值小于1的（特征值相当于影响非常小了），表中可知，第一、二、三、四主成分总占比超过了92%，因此我们只需要主成分1、2、3、4



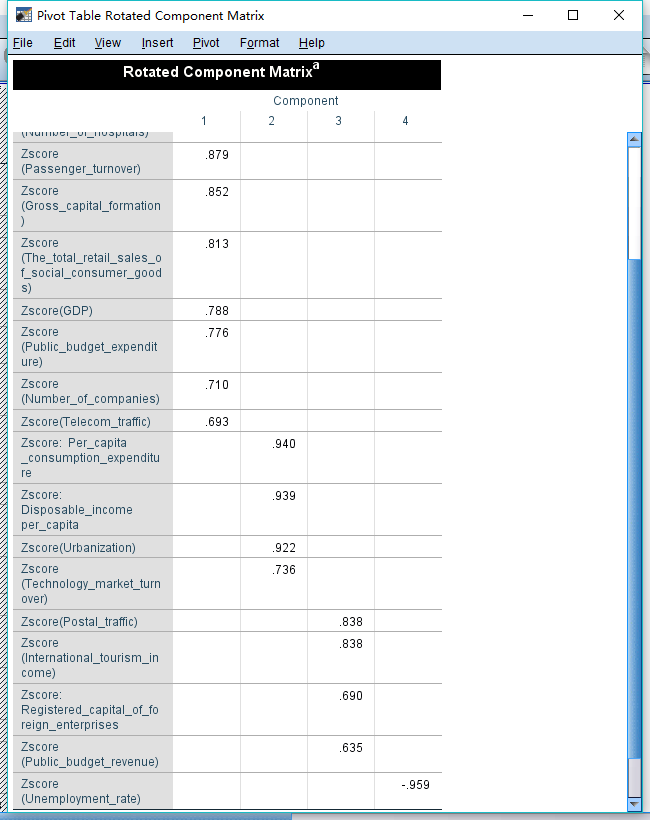
这里的分析还可以加上陡坡图再次论证，每一个主成分是一个点，在下图中，在第五主成分之后的点趋与收敛，因此我们只需要取前四个主成分即可



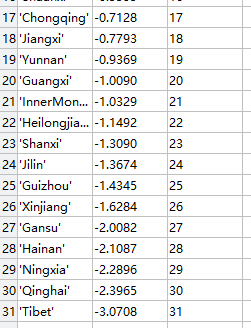
1. 根据特征值算特征向量，再算特征向量，再算在主成分上的载荷，也就是这些所有的变量对主成分的影响。大致如下图



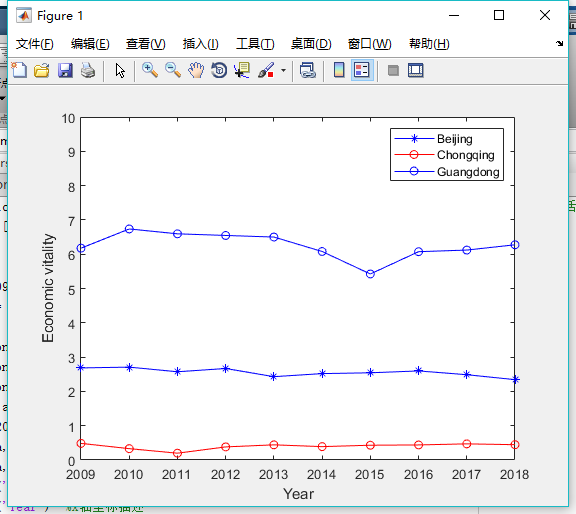
1. 这里看出来对所有的因素都对主成分具有影响，但是这样不好，我们应该对每一个主成分的总因素做出排除，使得主成分之间没有太多的相关性。筛选时将相关系数为0.6一下的剔除掉（0.6以下在person系数中被认为是非强相关），再用Varimax with Kaiser Normalization（凯撒最大方差化进行旋转处理）将每个主成分给剖分开，按照不同的因素，这样每个主成分的影响因子都不同，例如，第一个主成分影响较大的有医院差床位，大学数量，医院数量，人口等，我们可以近似认为第一个主成分与公共设施具有巨大相关性（后面的我先不分析了，因为有可能数据跟最后的数据有出入）



1. 得到了主成分载荷，用文献中公式算各主成分的得分，对每个主成分加权求和，权重是贡献率，可以得到每一年的排名，如下图，列分别代表省份名字，得分以及排名，这一部分用matlab代码实现。



1. 对每一年进行如上分析，我们可以得到每个省份每一年的排名，可以对这个画一个折线图，折现可以在每个区域找几个代表省份做比较，如下图所示



9、这就是整个模型的构建以及城市排名，第三问可以解决掉