国民幸福指数GNH是在20世纪70年代被提出的，直到2006年达到热议阶段。部分学者认为推动国家发展光看GDP是不够的，更具表现力的幸福指数会是更好的指标。

而目前，学术界对农村居民幸福指数统计较少。在中国知网经济与管理科学学科下，……。

此外，现有的农村幸福指数调查研究较为局限，多为区域性调查；且现有文献多关注于幸福感的影响因素，缺少对地区间幸福感差异的原因的探索

而这正是寻求创新的突破口。

因子分析法的基本思想是将观测变量进行分类，将相关性较高，即联系比较紧密的变量分在同一类中，而每个变量就是公共因子。

这是相关性矩阵，这是因子相关性表格。可以看到这些变量在第一因子下相关度最高，归为第一因子……

这是因子分析法主要部分的简单数学解释，可以看到μ是期望，sigma是方差，为了让各变量处于平等地位（即消除量纲），我们将数据标准化，结果为z。这里的F就是（切换到上一页）的因子，就是我在该研究中归类的社会服务，生活水平等，F由成分得分系数与每个省市的相关系数相乘得到。l是因子载荷，即原始变量与因子分析时抽取出共同因子的相关性。

S是因子载荷阵中某列的平方总和，也叫特征值（指ppt），该值越高说明该因子的重要程度越高。综合得分函数以各因子的[方差](https://baike.baidu.com/item/%E6%96%B9%E5%B7%AE)贡献率为权，将各因子线性组合得到。

这是我研究的数据集，来源于EPS平台，选取2015年我国31个省市、自治区有关衡量幸福感的14个指标。

至于为什么不用更方便的SPSS分析而用python分析，从公司层面来说，建模的目的是要卖钱的，也就是要嵌入到公司开发的产品上去，用 Python 写比找SPSS的接口更容易接受。

先是数据清洗。由于是利用EPS平台对所需数据调整后导入，所以不存在错误数据、多余数据与重复数据，故只简化表格与缺失值处理。先将空行，无关行，和缺失值大于1的行删去。

之后利用较为精确的拉格朗日插值对空缺值数据补上并将文件导出。

之后进行标准化处理，KMOtest以及 Bartlett test以确定该数据集适不适合因子分析。KMO值为0.74表现为适合分析。这里有一个小插曲。求完Pearson’relativity后，这个datasetcorr的第ij元素是之前表格第i行与第j行的相关系数，也就是说datasetcorr是对称阵，但是调试时发现用numpy求逆后逆矩阵不是对称阵。

使用这组数据进行测试后发现同一个矩阵用……求逆的结果都不同，甚至正负号 数量级别都不同，但是三个软件对接近整数的数据处理结果都一样，事后发现是精度问题。

而Bartletttest结果表示显著性水平小于0.05，通过了球体检验，故适合作因子分析。

这之后所有的分析我用factoranalyzer库和不用库只从数学原理上都实现了一遍。这里两种方式都表现出到第五个因子时累计率高于85%，故取四个因子。

命名前需先求出成分矩阵便于归类。从图中可以看出，比如这两元素在这个因子下因子载荷最大，故归于该因子。我将这四个因子根据分类情况命名为社会服务因子，生活水平因子，经济实力因子和农村建设因子。可以看到使用库和不使用库除在数据上有细微区别外，与共同因素的对应完全一致。

根据因子得分表，可以进一步计算各省市的因子得分，并推出方差贡献率以求得幸福指数评分模型（指）。

最后，求得因子得分排名。