## 214因子分析法（FA）

1、因子分析概述

因子分析(factor analysis)是一种数据简化的技术。它通过研究众多变量之间的内部依赖关系，探求观测数据中的基本结构，并用少数几个假想变量来表示其基本的数据结构。这几个假想变量能够反映原来众多变量的主要信息。原始的变量是可观测的显在变量，而假想变量是不可观测的潜在变量，称为因子。

因子分析特征：1）因子分析与回归分析不同，因子分析中的因子是一个比较抽象的概念，而回归因子有非常明确的实际意义；2）主成分分析与因子分析也有不同，主成分分析仅仅是变量变换，而因子分析需要构造因子模型。主成分分析利用原始变量的线性组合表示新的综合变量，即主成分；3）因子分析：潜在的假想变量和随机影响变量的线性组合表示原始变量。

2、因子分析模型

设为变量，如果表示为





称为公共因子，是不可观测的变量，他们的系数称为因子载荷。 是特殊因子，是不能被前*m*个公共因子包含的部分。并且满足：



即不相关并且互不相关，方差为1，而。

3、因子载荷阵估计方法

要建立实际问题的因子模型，关键是要根据样本数据矩阵估计因子载荷矩阵A,对A的估计方法很多，主要有主成分法、主因子法及最大似然估计法。这里采用较为普遍的主成分方法。

设样本的协差阵的特征值和对应的标准正交化特征向量分别为：



则协差阵可分解为：



当最后*p*-*m*个特征值较小时，协差阵可以近似的分解为



A即为因子协方差阵；当X的协方差阵未知，可以用样本协方差阵S去代替。

4、因子旋转

不管用何种方法确定因子载荷矩阵A，它们都不是唯一的，我们可以由任意一组初始公共因子做线性组合，得到新的一组公共因子，使得新的公共因子彼此之间相互独立，同时也能很好的解释原始变量之间的相关关系。这样的线性组合可以找到无数组，这样就引出了因子旋转。因子旋转的目的是为了找到意义更为明确，实际意义更明显的公因子。因子旋转不改变变量共同度，只改变公因子的方差贡献。

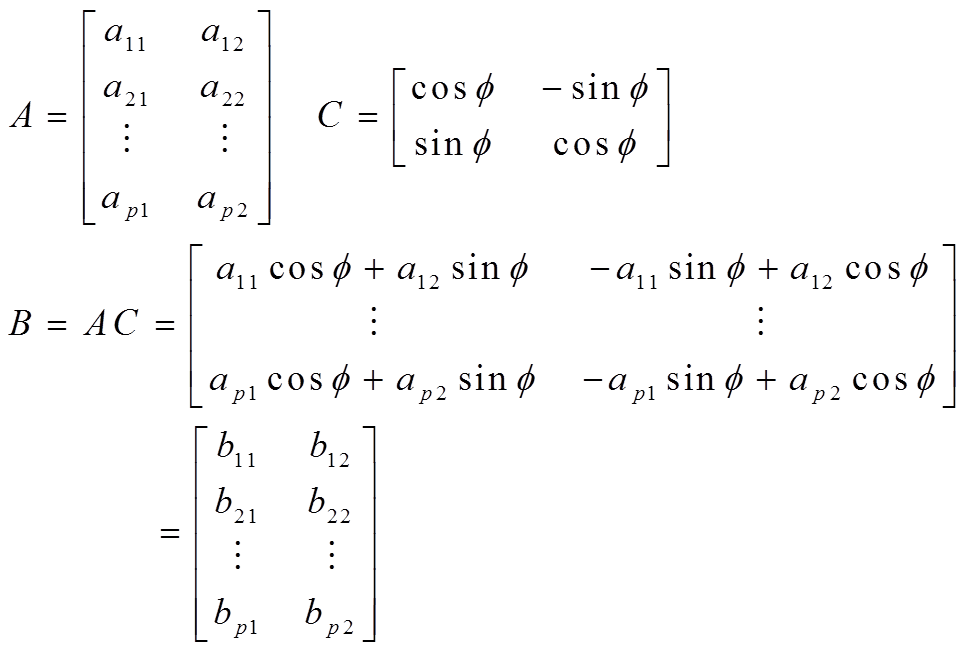
（1）因子旋转分为两种：正交旋转和斜交旋转

特点：1）正交旋转。由因子载荷矩阵A左乘一正交阵而得到，经过旋转后的新的公因子仍然保持彼此独立的性质。正交变化主要包括方差最大旋转法、四次最大正交旋转、平均正交旋转。2）斜交旋转。放弃了因子之间彼此独立这个限制，可达到更简洁的形式，实际意义也更容易解释。

不论是正交旋转还是斜交旋转，都应该在因子旋转后，使每个因子上的载荷尽可能拉开距离，一部分趋近1，一部分趋近0，使各个因子的实际意义能更清楚地表现出来。

（2）接下来接收方差最大化正交旋转：

假设前提：公因子的解释能力能够以其因子载荷平方的方差来度量。先考虑两个因子的平面正交旋转：对A按行计算共同度，考虑到各个变量的共同度之间的差异所造成的不平衡，需对A中的元素进行规格化处理，即每行的元素用每行的共同度除之。规格化后的矩阵，为方便仍记为A，施行方差最大正交旋转（C为正交阵）：



目的：希望所得结果能使载荷矩阵的每一列元素的绝对值尽可能向1和0两极分化，即原始变量中一部分主要与第一因子有关，另一部分主要与第二因子有关，也就是要求（b112,…,bp12），（b122,…,bp22）这两组的方差尽量大。为此，正交旋转的角度必须满足使旋转后得到因子载荷阵的总方差V1+V2=G达最大。即：



经过计算，其旋转角度可按下面公式求得：



如果公共因子多于两个，我们可以逐次对每两个进行上述的旋转，设公共因子数m>2。1）第一轮旋转，每次取两个，全部配对旋转，变换共需进行m(m-1)/2次；2）对第一轮旋转所得结果用上述方法继续进行旋转，得到第二轮旋转结果。每一次旋转后，矩阵各列平方的相对方差之和总会比上一次有所增加；3）当总方差的改变不大时，就可以停止旋转。

5、因子得分函数

因子分析的数学模型是将变量表示为公共因子的线性组合。由于公共因子能反映原始变量的相关关系，用公共因子代表原始变量时，有时更有利于描述研究对象的特征，因而往往需要反过来将公共因子表示成为变量的线性组合，即：



并称上式为因子得分函数。

6、估计因子得分函数的方法

估计因子得分的方法很多，如加权最小二乘方法、回归法等，这里采用回归法估计因子得分函数。回归法是1939年由Thomson提出来的，所以又称为汤姆森回归法。

我们现在仅知道由样本值可得因子载荷阵A,由因子载荷的意义知：









则，我们有如下的方程组：

 *j*=1,2,…,*m*

为原始变量的相关系数矩阵。

其中：

为载荷矩阵的第*j*列；为第*j*个因子得分函数的系数，记为*B*.

于是**就是估计因子得分的计算公式。

在估计出公因子得分后，可以利用因子得分进行进一步的分析，如样本点之间的比较分析，对样本点的聚类分析等，当因子数m较少时，还可以方便地把各样本点在图上表示出来，直观地描述样本的分布情况，从而便于把研究工作引向深入。

7、因子分析的步骤

1）选择分析的变量。用定性分析和定量分析的方法选择变量，因子分析的前提条件是观测变量间有较强的相关性，因为如果变量之间无相关性或相关性较小的话，他们不会有共享因子,所以原始变量间应该有较强的相关性。

2）计算所选原始变量的相关系数矩阵。相关系数矩阵描述了原始变量之间的相关关系。可以帮助判断原始变量之间是否存在相关关系，这对因子分析是非常重要的，因为如果所选变量之间无关系，做因子分析是不恰当的。并且相关系数矩阵是估计因子结构的基础。

3)提取公共因子。这一步要确定因子求解的方法和因子的个数。需要根据研究者的设计方案或有关的经验或知识事先确定。因子个数的确定可以根据因子方差的大小。只取方差大于1(或特征值大于1)的那些因子，因为方差小于1的因子其贡献可能很小；按照因子的累计方差贡献率来确定，一般认为要达到60％才能符合要求。

4）因子旋转。通过坐标变换使每个原始变量在尽可能少的因子之间有密切的关系，这样因子解的实际意义更容易解释,并为每个潜在因子赋予有实际意义的名字。

5）计算因子得分。求出各样本的因子得分，有了因子得分值，则可以在许多分析中使用这些因子，例如以因子的得分做聚类分析的变量，做回归分析中的回归因子。