**主成分**

主成分分析也称作主分量分析，是霍特林（Hotelling）在1933年首先提出。主成分分析是利用降维的思想，在损失较少信息的前提下把多个指标转化为较少的综合指标。转化生成的综合指标即称为主成分，其中每个主成分都是原始变量的线性组合，且各个主成分互不相关。stata对主成分分析的主要内容包括：主成分估计、主成分分析的恰当性（包括负偏协方差矩阵和负偏相关系数矩阵、KMO（Kaiser-Meyer-Olkin）抽样充分性、复相关系数、共同度等指标测度）、主成分的旋转、预测、各种检验、碎石图、得分图、载荷图等。

**文本, 信件

描述已自动生成**

**表格

描述已自动生成**

其中，a称为得分，b称为载荷。主成分分析主要的分析方法是对相关系数矩阵（或协方差矩阵）进行特征值分析。

stata中可以通过负偏相关系数矩阵、负相关系数平方和KMO值对主成分分析的恰当性进行分析。负偏相关系数矩阵即变量之间两两偏相关系数的负数。非对角线元素则为负的偏相关系数。如果变量之间存在较强的共性，则偏相关系数比较低。因此，如果矩阵中偏相关系数较高的个数比较多，说明某一些变量与另外一些变量的相关性比较低，主成分模型可能不适用。这时，主成分分析不能得到很好的数据转化效果。

Kaiser-Meyer-Olkin抽样充分性测度也是用于测量变量之间相关关系的强弱的重要指标，是通过比较两个变量的相关系数与偏相关系数得到的。KMO介于0与1之间。KMO越高，表明变量的共性越强。如果偏相关系数相对于相关系数比较高，则KMO比较低，主成分分析不能起到很好的数据转化效果。根据Kaiser（1974），一般的判断标准如下：0.00-0.49  /不能接受(unacceptable)/;0.50-0.59  /非常差（miserable）/;0.60-0.69  /勉强接受（mediocre）/;0.70-0.79  /可以接受（middling）/;0.80-0.89  /比较好（meritorious）/;0.90-1.00  /非常好（marvelous）/。

SMC即一个变量与其他所有变量的负相关系数的平方，也就是负回归方程的可决系数。SMC比较高表明变量的线性关系越强，共性越强，主成分分析就越合适。

成分载荷、KMO、SMC等指标等可以通过extat命令进行分析。

多元方差分析是方差分析在多元中的扩展，即模型含有多个响应变量。本章介绍多元（协）方差分析以及霍特林（Hotelling）均值向量T检验。

**正文**

**一、熵值法/熵权法含义**

熵值法与熵权法是一个含义，是一种客观赋权法，是指根据各项指标观测值所提供的信息的大小来确定指标权重。在信息论中，熵是对不确定性信息的一种度量。从统计学角度来看（同信息学有本质区别，也不同于热力熵）：数据离散程度越大，则信息量越大，熵值也就越小，应当赋予更大的权重；数据离散程度越小，则信息量越小，熵值也越大，应当赋予更小的权重。

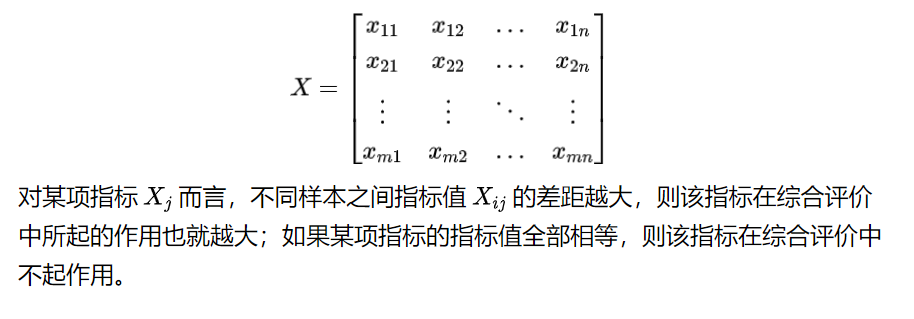
**二、熵值法处理方法**

**（1）数据形式及预处理**

**1.1 数据形式**

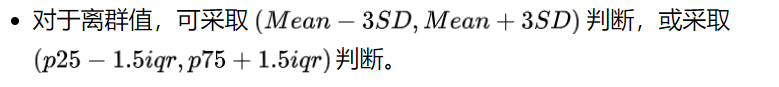
根据指标的特性，我们可以用熵值来判断某个指标的离散程度：指标熵值越小，离散程度越大，该指标对综合评价的影响（即权重）也越大。

假设现在有这样一段截面数据，m 个样本与 n 个评价指标，形成原始数据矩阵如下：



**1.2 预处理**

* 对于空值需要采取一定的手段填充，若缺失比例过大，则直接删除变量。



### （2）正式处理

#### 2.1 数据标准化

对于不同类型的指标，需要按照不同的方式分别进行标准化。

正向指标：

文本, 信件

描述已自动生成

#### 2.2 计算指标比重

计算第 j 个指标下第 i 个样本占该指标的比重：

文本

描述已自动生成

#### 2.3 计算熵值

计算第 j 个指标的熵值：

文本

描述已自动生成

文本

描述已自动生成