一、因子分析

1 因子分析的基本思想

1.1 因子分析的基本出发点

将原始指标综合成较少的指标，这些指标能够反映原始指标的绝大部分信息（方差），这些综合指标之间没有相关性。

1.2 因子变量的特点

（1）这些综合指标称为因子变量，是原变量的重造；

（2）个数远远少于原变量个数，但可反映原变量的绝大部分方差；

（3）不相关性；

（4）可命名解释性。

2 因子分析的基本步骤

（1）确认待分析的原始变量是否适合作因子分析；

（2）构造因子变量；

（3）利用旋转方法使因子变量具有可解释性；

（4）计算每个样本的因子变量得分。

3 因子分析的数学模型

数学模型（xi为标准化的原始变量；Fi为因子变量；k<p）



也可以矩阵的形式表示为：

X=AF+*ε*

F：因子变量；

A：因子载荷阵；

*a*ij：因子载荷；

*ε*：特殊因子。

4 因子分析的相关概念

（1）因子载荷

在因子变量不相关的条件下，aij就是第*i*个原始变量与第*j*个因子变量的相关系数。aij绝对值越大，则Xi与Fi的关系越强。

（2）变量的共同度(Communality)

也称公共方差。Xi的变量共同度为因子载荷矩阵A中第*i*行元素的平方和。

可见：Xi的共同度反应了全部因子变量对Xi总方差的解释能力。

（3）因子变量Fj的方差贡献

因子变量Fj的方差贡献为因子载荷矩阵A中第j列各元素的平方和

可见：因子变量Fj的方差贡献体现了同一因子Fj对原始所有变量总方差的解释能力，Sj/p表示了第j个因子解释原所有变量总方差的比例。

5 原有变量是否适合作因子分析

计算原有变量的相关系数矩阵，一般小于0.3就不适合作因子分析。

6 确定因子变量--主成份分析

6.1主成份分析法的数学模型



将原有的P个相关变量Xi作线性变换后转成另一组不相关的变量Yi

该方程组要求：

系数uij依照两个原则来确定：

yi与yj(i≠j,i,j=1,2,3,…p)互不相关；

y1是x1,x2,x3,…,xp的一切线性组合（系数满足上述方程组）中方差最大的；y2是与y1不相关的x1,x2,x3,…,xp的一切线性组合中方差次大的；yP是与y1, y2, y3,…yp都不相关的x1,x2,x3,…,xp的一切线性组合中方差最小的；

y1在总方差中所占比例最大，它综合原有变量的能力最强，其余变量在总方差中所占比例依次递减，即：其余变量综合原有变量的能力依次减弱。

6.2主成份分析的基本步骤

（1）将原始数据标准化；

（2）计算变量间简单相关系数矩阵R；

（3）求R的特征值λ1≥λ2≥λ3≥…λp≥0及对应的单位特征向量*μ*1,*μ*2*,μ*3*,…μ*p；

（4）得到：yi=*u*1i*x*1+*u*2i*x*2+…+*u*pi*x*p

6.3确定因子变量—计算因子载荷





7确定因子变量个数

确定k个因子变量

（1）根据特征值λi确定：取特征值大于1的特征根；

（2）根据累计贡献率：一般累计贡献率应在70%以上；







（3）通过观察碎石图的方式确定因子变量的个数。



8因子变量的命名解释

（1）发现：

aij的绝对值可能在某一行的许多列上都有较大的取值，或aij的绝对值可能在某一列的许多行上都有较大的取值。

（2）表明：

某个原有变量xi可能同时与几个因子都有比较大的相关关系，也就是说，某个原有变量xi的信息需要由若干个因子变量来共同解释；同时，虽然一个因子变量可能能够解释许多变量的信息，但它却只能解释某个变量的一少部分信息，不是任何一个变量的典型代表。

（3）结论：因子变量的实际含义不清楚

通过某种手段使：每个变量在尽可能少的因子上又比较高的载荷，即：在理想状态下，让某个变量在某个因子上的载荷趋于1，而在其他因子上的载荷趋于0。这样：一个因子变量就能够成为某个变量的典型代表，它的实际含义也就清楚了。

9计算因子得分

因子得分是因子变量构造的最终体现。

基本思想：是将因子变量表示为原有变量的线性组合，即：通过因子得分函数计算因子得分。

因子得分可看作各变量值的权数总和，权数的大小表示了变量对因子的重要程度。



10因子分析的基本步骤

（1）菜单选项：analyze->Data Reduction->Factor；

（2）选择参与因子分析的变量到Variables框；

（3）Discriptive:分析是否适合做因子分析；

（4）Extraction：选择构造因子变量的方法，默认主成分分析法。

Extract框：指定确定因子个数的标准；

（5）Rotation：择因子载荷矩阵的旋转方法。默认是不进行旋转。一般可以选择Varimax选项采用方差极大法旋转。

（6）Scores：Save as variables：将因子得分存成一个名为FACn\_m的SPSS变量中，其中：n是因子变量的名，以数字序号的形式表示；m表示是第几次作的。

Display factor score coefficient matrix项表示：以矩阵的形式输出因子得分函数。

Method框中提供了估计因子得分的几种方法。