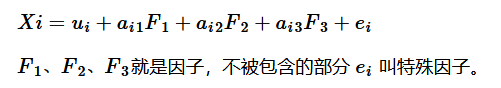
# 起源

因子分析最早由英国心理学家C.Spearman发表了第一篇有关因子分析的文章《对智力测验得分进行统计分析》，从中提出的：他发现学生的英语、法语和古典语成绩非常有**相关性**，他认为这三门课程背后有一个**共同的因素驱动**，最后将这个因素定义为“语言能力”。由此解开了因子分析的序幕。

# 基本思想

因子分析通过研究众多变量间的内部依赖关系，探求观测数据中的基本结构，并用少数几个假想变量（因子）来表示基本的数据结构。原始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。

如在企业品牌形象研究中，消费者可通过24个指标构成的评价体系，对商场的方方面面进行优劣评价，但是消费者会主要关心3个放米娜，即商店的环境、商店的服务和商品的价格。因子分析可以通过24个变量，找出反应商店环境、商店服务水平和商品价格的3个潜在因子，对商店进行综合评价。



# 因子分析特点

* 因子变量的数量 << 原有的指标变量数量。
* 因子变量不是对原有变量的取舍，而是根据原有变量的信息进行重新组合，能够反映原有变量大部分的信息。
* 因子变量不存在线性相关关系。
* 因子变量具有命名解释性，即该变量是对某些原始变量信息的综合与反应。

# 算法用途

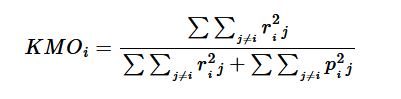
* 降维，减少分析变量的数量
* 分类，将内部具有关联变量/样本划分一类

# 分析步骤

* a. 选择分析变量，并进行标准化处理
* b. 计算变量间的相关系数矩阵、相关系数矩阵的特征值、特征向量
* c. 充分性检验：KMO和巴特莱特球度检验，验证变量是否适合做因子分析。

KMO(Kaiser-Meyer-Olkin)检验

KMO检验统计量是用于比较变量间简单相关系数矩阵和偏相关系数的指标。数学定义为：



是否适合因子分析：KMO在0.9以上非常适合；0.8表示适合，0.7表示一般，0.6表示不太适合；0.5以下表示极不适合。

巴特莱特球度检验

该检验以原有变量的相关系数矩阵为出发点，其零假设

是：相关系数矩阵为单位矩阵，即相关系数矩阵主对角元素均为1，非主对角元素均为0（即原始变量之间无相关关系）。

依据检验统计量服从卡方分布，如果统计量卡方值较大且对应的sig值小于给定的显著性水平

时，零假设不成立。即说明相关系数矩阵不太可能是单位矩阵，变量之间存在相关关系，适合做因子分析。

* d. 提取公因子：只取方差>1或特征值>1（方差小于1的因子其贡献可能很小。）；因子的累积方差贡献率达到80%。
* e. 因子旋转：提取因子的实际意义更容易解释

正交旋转和斜交旋转，其中主要使用正交旋转的方差最大旋转法。

* f. 计算因子得分

# **因子分析应用场景**

因子分析方法主要用于三种场景，分别是：

* 信息浓缩：将多个分析项浓缩成几个关键概括性指标。比如将多个问卷题浓缩成几个指标。如果偏重信息浓缩且关注指标与分析项对应关系，使用因子分析更为适合。
* 权重计算：利用方差解释率值计算各概括性指标的权重。在信息浓缩的基础上，可进一步计算每个主成分/因子的权重，构建指标权重体系。
* 综合竞争力：利用成分得分和方差解释率这两项指标，计算得到综合得分，用于综合竞争力对比（综合得分值越高意味着竞争力越强）。此类应用常见于经济、管理类研究，比如上市公司的竞争实力对比。

参考: <https://www.cnblogs.com/wxyz94/p/16153555.html>

<https://blog.csdn.net/ryo007gnnu/article/details/120587418>

[https://blog.csdn.net/weixin\_42058609/article/details/120764932?utm\_medium=distribute.pc\_relevant.none-task-blog-2~default~baidujs\_baidulandingword~default-0-120764932-blog-120587418.235^v38^pc\_relevant\_default\_base3&spm=1001.2101.3001.4242.1&utm\_relevant\_index=3](https://blog.csdn.net/weixin_42058609/article/details/120764932?utm_medium=distribute.pc_relevant.none-task-blog-2~default~baidujs_baidulandingword~default-0-120764932-blog-120587418.235%5ev38%5epc_relevant_default_base3&spm=1001.2101.3001.4242.1&utm_relevant_index=3)