第 47 章　综合评价与多属性决策*

唐 启 义[a,b]
（a. 浙江大学，b.杭州睿丰信息技术有限公司）

综合评价指数在各个领域广泛应用，尤其是在人文、环保、科技、多元化、可持续发展等新兴战略领域，其应用更是无处不在。在综合评价指数构建过程中，如何合理第给多因子赋权是整个过程的关键阶段。各个因子在构建综合评价指数时的权重是准确表达各个评价因子的作用、反映评价主体价值取向的关键要素，更是决定多因子综合评价结果科学性、合理性、公允性的核心。因此，赋权方法是多因子综合评价研究的重点、难点内容。Greco 等(2019)对综合评价指数构建过程中的加权、聚集和稳健性问题进行了系统的回顾和分析。

综合评价指数应用广泛，但如何对构建综合评价指数的各个因子进行加权处理，或者说，**计算各个因子的加权系数是综合评价中的核心问题，也是令人头疼的问题**（郭亚军，2007，第 31 页）。在现有求解权重系数的方法中，没有一个是大家公认的标准模型。

针对同一评价对象集，如果采用多种逻辑上可行的综合评价方法， 那么可能得到的评价分档、排序结果也存在差异。选择不同的评价方法可能得到不同的评价结果。这是目前综合评价领域中不可回避又亟待解决的一个难题。

实际上目前没有哪一种加权体系是无可非议的，每种方法都有其优点和缺点，并没有明确的胜过其他方法的最终方案。如欧盟 OECD/EU-JRC(2008)编制的《综合评价指数编制手册》中介绍了多种综合评价指数计算的加权方法，并对这些方法的优缺点进行了分析。由于综合评价过程中加权处理的方法不能统一，这就导致有人(Sharpe, A., 2004；Stiglitz,2009)对多因子综合评价权重的合理性、公平性提出了质疑，并**认为在其建构综合指数过程中主观性的存在是主要缺陷。**

在 OECD/EU-JRC(2008)编制的《综合评价指数编制手册》中，作者用较大篇幅介绍了 Melyn(1994)和 Cherchye 等(2007;2008)参照数据包络分析(DEA)模型提出的质疑补偿内生赋权法(Benefit of Doubt, BoD)作为综合评价分析时权重处理工具。但同时也对 BoD 方法依旧存在的较为严重的理论问题进行了讨论，如原始 BoD 模型权重计算结果可能存在部分分量为 0，甚至还可能出现某个因子权重非 0 而其他因子权重均为 0 的极端情况，即 0-1 权问题；其次，BoD 模型存在多重解问题；再由于原始 BoD 模型每个评价对象权向量都不相同，且权重系数之和可能不等于 1，导致采用 BoD 计算的综合评价指数理论上不具备可比性。

在国内，随着我国市场经济发展，现有的综合评价方法越来越难以为决策者提供准确的测度新形势下我国各地区经济、社会发展的真实情况。同时，对于因排名靠后而受

*基于最大熵-最小二乘优化算法的综合评价与多属性决策新模型,是浙江大学唐启义教授于 2023 年 12 月首先提出。该算法用于综合评价与多属性决策新技术。杭州睿丰信息技术有限公司实现了该算法的计算机计算过程，作为该公司统计软件产品“DPS 数据处理系统”的一组新增功能提供给用户应用。本文亦系相应的 DPS 数据处理系统的增补教材章节。文献引用格式为: Tang, Q.-Y. and Zhang, C.-X. (2013), Data Processing System (DPS) software with experimental design, statistical analysis and data mining developed for use in entomological research. Insect Science. 20(2): 254-260.

到影响的评价对象，如果我们没有客观的、标准的综合评价体系，受评对象就有理由对评价结果的公平性、合理性提出质疑，这会导致综合评价的权威性降低。

因此，对综合评价分析来说，一方面随着可供利用信息的急剧增长，综合评价指数的应用越来越广泛，如 Bandura(2011)根据一些经济、政治、社会或环境因子确定了 400 多个官方综合评价指数，对国家进行排名或评估；在联合国开发计划署的一份补充报告中，Yang(2014)记录了 100 多个人类进步的综合评价指数。虽然与现有应用的实际数量相比，这些清单还远远不够详尽，但它们让我们很好地了解综合评价指数的普及程度，对数据解释和整合的需求越来越大。但另一方面由于没有一个构建综合分析指数时确定各个因子的权重的“金标准”，即使人们从选择理论框架、花费人力财力收集数据、力求综合评价过程实施得完美无缺，但因无标准的综合分析指数模型可用，评价过程仍可能存在某些缺陷。尽管存在上述问题，但综合指数已被一些国际组织广泛用于衡量经济、环境和社会现象。因此，他们在发展过程中仍提供了一个极为相关的工具(OECD, 2008)。

如前所述，目前综合评价或多属性决策尚处于“**盲人摸象**”阶段。本研究将统计学与信息学有机结合来解决综合评价或多属性决策中各个因子权重估计，即如何给参与构建综合评价的因子一个科学、客观、公平、合理的权重系数。目的是对构建综合评价指数的方法论方面有所贡献。

作者认为，综合评价指数和各个评价因子的关系类似物理学中功(work)与能(energy)的关系。综合评价指数是各个评价因子在综合评价中的表现，各个评价因子是综合评价指数所具有的能力的来源。综合评价指数和各个评价因子之间存在有定量的转换关系。

目前的现状是，我们已经建立了综合评价指数和各个评价因子之间定量转换关系的函数（方程），但尚未找到求解这个定量转换函数的计算方法。到目前为止，所有的探索综合评价指数的数学方法都不是试图定量地求解这个转换函数，而是从其它方面探索综合评价指数和各个评价因子之间的相关性。以致这些偏离综合评价指数和各个评价因子之间定量的转换函数的非定量求解方法都是徒劳的。

为解决这一问题，作者提出了基于最大信息熵（Jaynes 1957；Amos G. and John H.，2023）–最小二乘残差的非线性拟合方法、和加权回归分析类似、定量地拟合综合评价指数和各个评价因子之间的关系，将权重系数估计和指标融合方法作为一个整体，从而求解综合评价指数中各个因子权重，从根本上解决综合评价指数建模过程中存在的问题。

为解决这一问题，作为建立综合分析指数的基础，提高加权处理过程的透明度和合理性，作者提出了基于最大信息熵（Jaynes 1957；Amos G. and John H.，2023）-最小二乘残差拟合的综合评价指数模型因子权重估计方法，以从方法论方面解决目前综合评价指数中存在的问题。

47.1　最大熵-最小残差综合评价指数模型

47.1.1　综合评价指数模型

假定有 n 个样本，p 个因子，每个因子指标的观测值为 x_{ij} (i=1，2，3，…，n；j=1，

2，3，…，p）。构造多因子综合评价指数，从统计模型的观点来看，是需要估计出各个因子的权重系数 w_j（j=1，2，3，…，p），以计算综合评价指数 y_i（i=1，2，3，…，n）。构建综合评价指数统计模型 y_i 的计算公式为：

$$y_i = x_{i1}w_1 + x_{i2}w_2 + x_{i3}w_3 + \cdots + x_{ip}w_p \tag{47.1}$$

式中 $w_1 + w_2 + w_3 + \cdots + w_p = 1$ 且 $w_j \geq 0$.

综合评价指数（y_i）从理论上来说，可以看成是由多个理论组分 C_{ij} 组成，C_{ij} 可定义为综合评价指数模型中第 i 个样本，第 j 个因子指标的综合评价指数理论值。

根据权重系数 w_j 可将综合评价指数理论值 y_i 分解为各个因子指标对应于综合评价指数理论值 y_i 的组成成分 C_{ij}。第 i 个综合评价指数（y_i）的第 j 个组分 C_{ij} 的值等于综合评价指数（y_i）乘以相应的权重 w_j，即：

$$C_{ij}=y_i \cdot w_j \tag{47.2}$$

也就是说，综合评价指数(y_i)我们也可以用理论组分之和的形式来表示：

$$y_i = C_{i1} + C_{i2} + C_{i3} + \cdots + C_{ip} \tag{47.3}$$

因前面的式(47.1)和式(47.3)式左边 y_i 相等，有理由认为等式右边的 $x_{ij}w_j$ 和 C_{ij} 一一对应的。如果我们将 $x_{ij}w_j$ 作为综合评价指数的观察组分值，C_{ij} 作为综合评价指数的理论组分值，那么综合评价指数模型中，观察组分值和理论组分值越接近，综合评价指数越能代表各个因子，这时的综合评价指数也越有代表性。

我们令相应的观察组分值和理论组分值之差为残差，$\sigma_{ij}=x_{ij} \cdot w_j - C_{ij}$，则 σ_{ij} 可视为第 i 个样本，第 j 个因子指标的拟合误差（式 47.4）。在回归模型中，参数估计值是通过最小化残差（σ）来确定的

$$\begin{aligned}\sigma &= \sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{p}\left(C_{ij} - x_{ij}w_j\right)^2}{(n-1)\cdot p}} \\ &= \sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{p}\left(y_i w_j - x_{ij}w_j\right)^2}{(n-1)\cdot p}} \\ &= \sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{p}\left[\left(x_{ij} - y_i\right)w_j\right]^2}{(n-1)\cdot p}}\end{aligned} \tag{47.4}$$

式（47.4）可以认为第 i 个样本第 j 个因子的拟合残差 σ_{ij}。它是第 i 个样本第 j 个因子观察值和第 i 个综合指数（理论值 y_i）之差，乘以第 j 个权重系数 w_j，即以 w_j 进行加权后的残差。某一因子权重系数越大，对残差的影响就越大，该因子越是“重要”。**这一**

做法实际上就是加权回归分析过程。这也和多因子综合分析中，因子权重系数越大、该因子越重要是一致的。

因此在综合评价指数统计模型中，为求解各个因子的权重系数，应该使得第 i 个样本，第 j 个因子指标的理论组分值 C_{ij} 和相应的观察组分值(x_{ij}·w_j)之间的残差尽可能地小。假设拟合误差项σ_{ij} 在各个因子之间是独立的，因为理想情况下，每个单独因子的组分可认为独立于其他因子来衡量综合评价指数特征的一个特定方面。综合评价模型中，为求解各个因子的权重系数，应该使得第 i 个样本，第 j 个因子指标的理论组分值 C_{ij} 和相应的观察组分值(x_{ij}·w_j)之间的残差尽可能地小。类似于线性回归模型中离差平方和的分解，综合评价模型的 Root Mean Square Error。式(47.4)中(n–1)·p 为自由度。式(47.4)可作为多因子综合评价模型极小化误差函数。即使得残差 σ 最小以求出权重系数 w_j。

式(47.4)中(n–1)·p 为自由度。式(47.4)可作为多因子综合评价指数目标函数模型。即使得残差 σ 最小以求出权重系数 w_j。

不过，在上述综合评价指数公式(47.1)中，如果第 j 个因子指标的权重系数为 1，其它因子的权重系数为零，这时综合评价指数 y_i 即为其中某一因子的观察值 x_{ij}，拟合残差 y_i–x_{ij} 为零，达到最小。显然，直接根据式(47.4)对残差 σ 进行优化得到的结果必是某个因子的权重系数为 1，其他因子权重系数为零，综合评价指数等于某个因子值。尽管这时的残差 σ 等于零，但它失去了我们定义的综合评价指数的意义！

47.1.2　最大熵-最小残差的综合评价指数模型

熵（Entropy）是信息论中的一个重要概念，是度量数据集中信息总量的度量单位。Jaynes(1957)的最大熵原理（MaxEnt）提供了一种从不完全信息中进行推理的系统和最佳方法。对于科学中的问题，拥有的信息越多，我们能够构建的模型和理论就越好。然而，对于复杂系统，如多因子综合评价体系的建立，我们永远不会有足够的信息来明确地预测各个因子的权重的概率分布。对于这样的系统，理论和模型的构建是一个欠定的推理问题。然而，如果我们把可用的信息指定为约束条件，并直接建立在最大熵原理之上，我们就有可能确保在所有与我们所拥有的信息相一致的可能模型中，所选择的模型是偏差最小的。

建立综合评价模型，因各个指标量纲不一致，因此一般的做法是先对数据进行最小-最大规格化，或其它规格化处理，使得处理后各个因子的转换值在 0-1 之间。这我们可以将其视为 0-1 之间的比率值。作者认为比率值是某个因子的在评价范围内分布的概率。

Jaynes（1982 年）指出，只要将 Shannon(1948)提出的熵视为信息的有效度量，并参考随机事件的概率分布，那么最大熵推理就是在给定有限信息（以该分布的矩形式）的情况下估计未知概率分布的正确方法。此外，Jaynes（1982 年）指出，直观地说，熵更高的分布具有更大的多样性；它们可以以更多种方式在自然界中实现，因此更有可能被观察到。因此可见，最大熵推理从给定信息中给出了最佳的概率估计，而不需要假设除约束集中包含的知识之外的任何进一步知识。根据最大熵推理，任何其他形式都将基于非最优推理，要么使用比可用信息更少的信息，要么表达不合理的偏见。

在多因子综合评价指数研究中，我们将应用最大熵原理来确定综合评价指数中各因

子权重系数。如果我们将各因子权重系数视为各因子在综合评价指数中重要性的某种概率分布，那么概率分布的最大熵形式可通过最大化 Shannon 信息熵得到。综合评价指数中各个因子权重系数 w_j 的 Shannon 信息熵公式为

$$E=-\sum_{j=1}^{p} w_j \ln(w_j) \tag{47.5}$$

从 Shannon 信息熵的公式(47.5)来看，等权重时，信息熵最大；某因子权重系数为 1，其他因子权重系数为零时，信息熵最小；这和残差大小的趋势是一致的。这似乎和我们优化模型时，希望利用最大熵原理，以达到信息熵尽可能大、残差尽可能小来优化综合评价指数模型相矛盾。

从 Shannon 信息熵的公式来看，等权重时，信息熵最大；某因子权重系数为 1，其他因子权重系数为零时，信息熵最小；这和残差大小的趋势是一致的。这似乎和我们优化模型时，希望利用最大熵原理，以达到信息熵尽可能大、残差尽可能小来优化综合评价指数模型相矛盾。

将信息熵用于综合评价指数建模，直观的想法是类似 LASSO 回归等压缩回归方法，建立如下形式的综合评价模型的优化函数

$$\min f(\sigma,E)=\sigma-\lambda E$$

当式中的λ取不同的值，会有不同的优化结果，当λ给值“适当”时，会有我们希望的模型。但这样做最大的问题是λ怎样取值？同时优化的目标函数是一个杂合体，因为残差和信息熵的量纲是不同的，这样以和/差的形式组合在一起作为优化的目标函数本身欠妥，因为综合评价模型实质是要优化式（47.4）所定义的残差（σ）。

这里自然可以想到，以和/差形式对拟合残差和信息上进行组合不能达到我们预期结果，那么以乘/除形式进行组合，即

$$\min f(\sigma,E)=\sigma/E$$

这种形式实际上只是将残差 σ 放大或缩小若干倍，是一种线性关系，和根据式(47.4)定义的残差进行优化结果相同，并不能得到所期望的统计模型。

鉴于前面两种形式均不适合构建综合评价模型，一种自然的想法是指数函数形式来建立综合评价误差函数模型：以残差（σ）为底，信息熵(E)为指数（47.6）。

$$\min f(\sigma,E)=\text{minimize}(\sigma^{E}) \quad , \tag{47.6}$$

以式 47.6 为误差函数，因建立综合评价或多属性决策模型，须预先进行最小-最大（或其它）规格化处理。处理后各个因子取值在 0-1 之间，这时拟合得到的残差标准差（σ）值必然是小于 1。当残差小于 1 时，其指数(信息熵)越大，误差函数 47.6 越小。根据这一特点，作者提出了一个基于最大熵-最小残差的建立综合评价模型(ME-MR)，其极小化误差函数 $f(\sigma,E)$ 的公式为：

式(47.6)中 σ 表示残差，E 表示各因子权重的 Shannon 信息熵。由于该模型中指数函数的底，即残差（σ）总是小于 1 的，这时指数（信息熵 E）越大，其函数值越小。因

此式 47.6 误差函数的这一特点可以理解为在信息熵的空间内，极小化残差的过程。

如果原始数据不进行转换，这时可以将各个原始数据观察值除以所有观察值中（绝对值）的最大值，将所有观察值的取值转换到 0—1 之间，成为比率值，再进行优化建模。

综上所述，综合评价指数分析模型可以表达为：

$$\begin{cases} f(\sigma, E) = \min \sigma^E \\ \text{s.t} \sum_{j=1}^{p} w_j = 1, \quad w_j \geq 0 \end{cases} \tag{47.7}$$

式中，x_{ij}是第i个样本第j个变量因子观察值，C_{ij}是第i个样本第j个变量因子的理论组分，w_j为第j个因子的权重系数。σ为拟合残差，E为参数估计值的信息熵。式中，s.t. 代表约束条件，这是一个以w_j，$j = 1, 2, \ldots, p$为优化变量的非线性回归方程模型。

带约束条件非线性回归方程模型式(47.7)的优化，可采用多种带约束条件的非线性优化方法进行求解。由于不同的优化算法收敛速度有快有慢，进行模型权重参数Bootstrap抽样估计需要花费时间较长，因此选用运行稳定、速度较快的算法值得探索。作者初步选用的是序列无约束极小化方法(Sequential Unconstrained Minimization Technique，SUMT)进行综合评价指数模型的优化求解，可以用较少的时间得到全局最优的结果。作者已将SUMT优化算法收录在作者主持开发的统计分析计算机软件——DPS数据处理系统之中(Tang and Zhang, 2013)。在以下的叙述中，我们将作者提出的综合分析指数模型得到的指数称之为MEMR指数，相应的权重系数称之为MEMR权重。

最大熵-最小二乘建模，从通俗意义上来讲，就是在尽可能全面地兼顾各个因子（信息）前提下，运用最小二乘技术提取各个因子所具有的共同趋势来建立综合评价模型或进行多属性决策。这与人们开展综合评价或多属性决策研究的动机是高度一致的。

47.1.4　综合评价模型统计检验

综合评价指数计算，最简单的是等权重(Equal weighting,EW)模型，即所有变量都被赋予相同的权重，如有 p 个变量，所有的指标的权重系数均等于 $1/p$。显然，实际综合评价和多属性决策过程中，各个评价指标在综合评价和多属性决策过程中的重要性是不尽相同的。这里我们将等权重模型作为综合评价模型的参照（零模型），将等权重模型作为最大熵-最小二乘（MEMR）模型的一个子模型，以嵌套模型的方式进行统计检验。这意味着我们可以使用统计检验方法来比较 MEMR 综合指数模型模型和等权重模型，即检验 MEMR 综合指数模型是否可以简化为等权重模型（零假设）。其统计检验可以用 F 检验。 各个模型的残差平方和如表 47-1。

表 47-1　多因子综合指数模型方差分析

变异来源	残差平方和	自由度	均方
EW 模型	$RSS(w_{EW})$	n-1	$RSS(w_{EW})/(n-1)$
MEMR 模型	$RSS(w_{MEMR})$	$n-p$	$RSS(w_{MEMR})/(n-p)$
EW 模型- MEMR 模型	$RSS(w_{EW})- RSS(w_{MEMR})$	$p-1$	$(RSS(w_{EW})- RSS(w_{MEMR})/(p-1))$

F 检验统计量是以残差平方和项表示等权重模型和最大熵-最小二乘综合指数模型之间距离的度量。较大的 F 统计值表明，这两个模型相差很远，意味着等权重模型因各个指标的权重系数肯不完全相等，不能很好地对数据进行描述，因此，更一般的最大熵-最小二乘综合指数模型是更合适的模型。因此需要采用 MEMR 模型来进行综合评价和多属性决策。

F 检验统计量是是一个比值，分子是比较的两个模型之间的方差项之差除以两个模型之间的参数之差，分母是 MEMR 模型（全模型）均方误差，统计检验公式如下：

$$F=\frac{\dfrac{RSS(\hat{w}_{EW})-RSS(\hat{w}_{MEMR})}{n-1-(p-1)}}{\dfrac{RSS(\hat{w}_{MEMR})}{p-1}} \tag{47.8}$$

F 检验统计量与线性模型中的定义相同，但与线性模型不同的是，F 分布仅在近似上成立（随着样本量的增加而变得更加接近）。根据 F 值及其自由度，我们不难得到显著性概率 p 值。

从 F 统计检验中我们可以得出，如果统计检验不显著（例如，采用 0.05 的显著性水平），则我们可以接受等权重模型作为综合评价模型，不过，一般说来，MEMR 模型总能或多或少地改善综合评价或多属性决策的效果；如果显著性概率 p 值小于 0.05，那么等权重模型作为综合评价模型可能不适用，而应采用 MEMR 模型进行综合评价或多属性决策。这是因为综合评价指数模型中的权重系数至少有一个不等，优化过程有意义，优化产生的权重系数与等权重系数相比，显著地增加了综合指数与观察值的贴近程度。另外，我们根据 EW 模型残差平方和和 MEMR 模型拟合残差平方和的差值大小，较 EW 模型残差平方和下降的百分比定义了模型优化指数 R

$$R=\frac{\text{EW模型残差平方和}-\text{MEMR模型残差平方和}}{\text{EW模型残差平方和}}\times 100\% \tag{47.9}$$

模型优化指数的含义是，经过优化计算，和原假设，即各个因子取相等权重相比，综合分析指数模型在多大程度上得到了优化，即残差平方和下降的百分比，或可理解为拟合程度提高的程度。

47.1.5 权重系数的置信区间

传统的参数推断主要依赖中心极限定理，因为它规定在大样本条件下，抽样分布都是服从正态分布。但对于某些抽样分布未知或难以计算的统计量，Bootstrap 法就十分有用了。

由于 MEMR 模型是一个非线性模型，因此很难准确地估计各个权重系数的标准误差。我们这里应用 bootstrap 方法进行估计。针对多因子最大熵最小二乘综合评价指数模型的 Bootstrap 法估计，每个 Bootstrap 样本中可计算出一组权重系数。如果我们抽取 1000

个 Bootstrap 样本，可得到 1000 组权重系数。

a. **权重系数 95%置信区间计算**

根据这 1000 组权重系数，计算出它们的第 2.5 百分位数和第 97.5 百分位数，得到各个权重系数 95%置信区间。如果权重系数 95%置信区间不包含 1/*p*, 则可以认为该权重系数和简单的平均权重值比较有统计学意义，这时多因子综合评价不适宜采用简单的平均权重来计算综合评价指数。

b. **权重系数标准误及其变异系数**

根据 Bootstrap 抽样的若干个样本，每个样本容量和初始抽取样本容量相同，每个样本计算出样本统计量的均值和方差，如均值为 $\overline{x}_1, \overline{x}_2, \overline{x}_3, \cdots, \overline{x}_n$，这时得到均值的抽样分布为 $\hat{\mu} = \sum_{i=1}^{n} \overline{x}_i / n$。然后安装公式 $\hat{\sigma}^2 = \sum_{i=1}^{n} (\overline{x}_i - \hat{\mu})^2 / (n-1)$。每个变量的变异系数为 $CV = \sigma / \hat{\mu}$。

DPS 系统计算各个因子的变异系数的平均值，用于考察综合评价模型中各个因子权重系数的波动幅度，以及权重系数在综合评价过程中的稳定程度。根据变异系数大小，从几个模型中选取变异系数较小的模型作为应用参考。

47.2　综合评价研究设计、实施与数据预处理

47.2.1 综合评价研究的设计与实施

欧盟 OECD/EU-JRC(2008)编制的《综合评价指数编制手册》中介绍了综合评价研究的理论框架，并指出：综合评价不仅要要有科学的计算方法，更要求有一个完善的理论框架，这是构建综合指数、多属性决策的起点。该框架应明确界定要测量的现象及其子成分，选择反映其相对重要性和整体复合维度的个别指标体系。这一过程应以需要测量的内容为基础，而不是以可用的指标为基础。因此研究人员应清楚地了解综合指标所衡量的是什么，应怎样将各个子群体和底层指标联系起来，并与综合指标的结构之间有明确的联系。

其次，在众多的因子中如何确定子群体。多维概念可以分成几个子组。这些子组不需要统计上彼此独立，现有的联系应该在理论上或经验上尽可能地描述。这样的嵌套结构，可提高用户对综合指标背后驱动力的理解。也可能更容易确定不同因素之间的相对权重。研究人员在这一步骤进行过程中应尽可能要利益攸关方参与进来，以便考虑到多种观点，并增加概念框架和指标集的稳健性。且调查的各个因子指标应尽可能精确地测量或表述(投入和产出度量)。

综合指标的优劣在很大程度上取决于底层变量的质量。理想情况下，应该根据变量的相关性、分析的合理性、及时性、可及性等因素来选择变量。保证综合指标基本数据集质量。虽然指标的选择必须以综合指标的理论框架为指导，但数据选择过程可能相当主观，因为可能没有单一的确定指标集。缺乏相关数据也可能限制开发人员构建完善的

复合指标的能力。由于缺乏国际上可比较的定量(硬)数据，综合指标通常包括来自调查或政策审查的定性(软)数据。

在无法获得所需数据或跨国可比性有限的情况下，可以使用代理措施。例如，可能无法获得有关使用计算机的雇员人数的数据。相反，可以使用可以访问计算机的员工数量作为代理。与软数据的情况一样，在使用代理指标时必须谨慎。在数据允许的范围内，应通过相关性和敏感性分析来检验代理指标的准确性。构建者还应密切关注所讨论的指标是否依赖于 GDP 或其他与规模相关的因素。为了在小国和大国之间进行客观比较，需要通过适当的规模度量来缩放变量，例如人口、收入、贸易额和人口稠密的土地面积等。最后，所选变量的类型——投入、产出或过程指标——必须与预期的复合指标的定义相匹配。

综合指标的质量和准确性应与数据收集和指标制定方面的改进同步发展。目前在一系列政策领域构建国家绩效综合指标的趋势可能进一步推动改进数据收集、确定新的数据来源和加强统计数据的国际可比性。另一方面，我们并不认同利用现有数据就一定足够的观点。糟糕的数据会在“垃圾输入、垃圾输出”的逻辑下产生糟糕的结果。然而，从实用的角度来看，在构造复合时需要做出妥协。我们认为至关重要的是这些妥协的透明度。

数据的缺失往往会妨碍的综合评价工作的进行。数据缺失可能的情形是:

完全随机缺失。缺失值不依赖于感兴趣的变量或数据集中任何其他观察到的变量。例如，如果(i)不报告收入的人与报告收入的人平均收入相同，则可变收入中缺失的值将属于完全随机缺失;并且，如果(ii)对于没有报告收入的人和报告了收入的人来说，数据集中的其他每个变量平均而言都必须相同。

随机缺失。缺失值不依赖于感兴趣的变量，而是以数据集中的其他变量为条件。例如，如果收入数据缺失的概率取决于婚姻状况，那么收入中缺失的值将是随机，但在婚姻状况的每个类别中，收入缺失的概率与收入的价值无关。

非随机丢失。取决于值本身。如高收入家庭不太可能报告他们的收入。不过非随机缺失没有统计检验，通常也没有判断数据是随机缺失还是系统缺失，而大多数估算缺失值的方法都需要随机缺失的假定。

处理缺失数据的一般方法是删除那一行记录，即简单地从分析中省略缺失的记录。然而，这种方法忽略了完整和不完整样本之间可能的系统差异，并且只有在删除的记录是原始样本的随机子样本时才产生无偏估计。此外，考虑到使用的信息较少，标准误差通常会在减少的样本中更大。DPS 可在缺失数据较少时，采用随机森林回归法予以插补。

47.2.2 数据转换

采集得来的数据可能是各种各样的性态。统计分析一般要求数据呈正态分布。当数据分布呈现非正态时，我们可以将原始数据通过某种函数转换，使偏态资料正态化，从而满足统计检验或其他统计分析方法对资料正态分布的要求，对数据资料进行正态转换常用方法有：

1、取对数， $X' = \ln(X)$

对数变换常用于：1）使服从对数正态分布的数据正态化。如环境中某些污染物的分布，人体中某些微量元素的分布等，可用对数正态分布改善其正态性。2）使数据达到方差齐性，特别是各样本的标准差与均数成比例或变异系数 CV 接近于一个常数时。此外，国家间国民收入差异很大，稍等几百美元，多的十几万美元，数量上相差了几个数量级，如果不取对数转换一下，得分根本没有可比性。

2、平方根变换 $X'=\sqrt{X}$

即将原始数据 X 的平方根作为新的分布数据。平方根变换常用于：1）使服从 Poission 分布的计数资料或轻度偏态资料正态化，可用平方根变换使其正态化。2）当各样本的方差与均数呈正相关时，可使资料达到方差齐性。

3、倒数变换 $X'=1/X$

即将原始数据 X 的倒数作为新的分析数据。这种转换常用于资料两端波动较大的资料，可使极端值的影响减小。

47.2.3 数据规格化

数据规格化的目的是将参与综合评价或多属性决策的多个因子，按某种规则对其变化、使得各个因子取值范围相同，达到各个因子之间在数量上具有可比性。

在多指标评价体系中，由于不同变量样本常常具有不同的单位和不同大小的数量值。如第一个变量的单位是 kg，第二个变量的单位是 cm，第三个变量是百分数。在计算绝对距离时将出现将三个变量样本值绝对值之差相加的情况，这时三个绝对值的单位不一致，如 3kg+5cm+35%，这 3 个数直接相加得到的和没有什么实际专业意义。不同样本数量值差异相差较大时，会使在计算出的关系系数中，不同变量所占的比重大不相同。例如如果第一个变量的数值在 2%到 4%之间，而第二个变量的数值范围都在 1000 与 5000 之间，第二个变量将决定关系系数；如果变量量纲相同、数量等级接近，变量样本间差异显著，对计算结果也会产生较大影响。为了消除量纲影响和数值大小的影响，故须将各个因子指标进行标准化处理。

在进行综合评价时，DPS 提供了 3 种数据规格化方法，即 Min-Max 规格化，标准化-正态概率和标准化-线性转换。

（1）*a*-*b* 线性(Min-Max 规格)化

综合评价指数计算过程中，许多综合评价指数组成因子是用不同的单位来衡量的。数值越高，表现越好，大多数因子被认为是“好”。例外情况是“长期失业率”和“因慢性疾病而妨碍日常活动”，这两个因子都被认为是“不好的”，因为数值越高，表现越差。为了将所有因子放在一个共同的基础上，我们使用重新缩放技术将数据归一化。根据一个因子被认为是“好”还是“坏”，*a*-*b* 线性(规格)化分别应用(10*a*)和(10*b*)进行数据转换

$$x'_{ij}=\frac{x_{ij}-a}{b-a}\quad(47.10a)\quad 和\quad x'_{ij}=\frac{b-x_{ij}}{b-a}\quad(47.10b)$$

如果$a=x_{\min}, b=x_{\max}$，那么公式(47.10*a*)和(47.10*b*)分别为：

$$x'_{ij}=\frac{x_{ij}-x_{\min j}}{x_{\max j}-x_{\min j}} \quad (47.11a) \quad 和 \quad x'_{ij}=\frac{x_{\max j}-x_{ij}}{x_{\max j}-x_{\min j}} \quad (47.11b)$$

即最大、最小规格化转换。转换后所有归一化因子 x'_{ij} 都具有相同的范围(0,1)，其归一化值为 0 和 1 分别对应于因子 1 上最差和最强的表现。这是综合评价中常用的数据预处理方式。在 DPS 里面，**各个因子最大值、最小值可以在用户界面中进行调整**，如果将最小值调整为零，那么这时的归一化结果相当于每个指标除以最大值。如果将最小值调整为零，最大值调整为均值，那么这时的归一化结果相当于每个指标除以平均值，即均值化。

如果$a=0, b=x_{\max}$，那么公式(47.10*a*)和(47.10*b*)分别为：

$$x'_{ij}=\frac{x_{ij}}{x_{\max j}} \quad (47.12a) \quad 和 \quad x'_{ij}=\frac{x_{\max j}-x_{ij}}{x_{\max j}} \quad (47.12b)$$

即为各个变量除以它的最大值。

如果$a=0, b=\bar{x}$，那么公式(47.10*a*)相当于各变量除以各自的平均值，即均值化转换。

这里的转换参数 *a* 和 *b* 的取值，可根据相应的专业背景定义有专业意义的值，如在联合国开发计划署人类发展指数研究中，首先根据根据四个基础数据，算出三个指数：寿命指数、教育指数、收入指数。三个指数都试图归一化成 0 到 1 之间的一个数值，数值越大越好。教育指数又是由两个次级的教育指数算术平均得来的。最后，HDI 指数是寿命指数、教育指数、收入指数三个数值的几何平均，即三个指数相乘再开三次方，也是一个 0 到 1 的数值。

其中寿命指数，令 *a*=20，*b*=85。寿命指数 = (预期人均寿命- *a*) / (*b*-*a*)。这里 *b*=85 是有意义的，相当于设置了一个最高的值，没有国家与地区高于这个高限。2014 年预期人均寿最长的是香港 84 岁和日本 83.7 岁。*a*=20 是一个最低值，如果人均寿命只有 20，得分就是 0。这个值不是随便设计的，历史研究表明，如果一个族群的人均寿命低于 20 这个生殖年龄，族群就会消失。当然没有国家这么低，再差的国家多少也能得一些分。

再如教育指数计算，*a*=0，*b*=15。平均学校教育年数指数 MYSI =(MYS-*a*)/(*b*-*a*)。这个归一化处理更简单，平均教育年数高达 15 的，指数值是 1.0。联合国认为没有国家的教育是完美的。最差的也不会是 0，多少会有些分数。再如预期学校教育年数指数 EYSI = EYS/18。这里归一化处理和 MYSI 一样，只不过上限变成了 18。对于有些国家超过了 18 年的国家，就技术处理成 18 年，指数值为 1.0。

（2）标准化（z-score 标准化）-正态概率转换或线性转换

标准化是统计学里面最常用的方法。经过标准化处理之后的各个指标，其均值为 0，

标准差为 1。这样各个指标之间的数据就具有了可比性。转换公式为：

$$x'_{ij}=\frac{x_{ij}-\overline{x}_j}{\sigma_j} \tag{47.13}$$

式中，$\overline{x}_j$ 为原始数据均值，σ_j 为原始数据的标准差，

$$\sigma_j=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_{ij}-\overline{x}_j\right)^2}$$

这是目前数据统计分析时用得最多的数据标准化方式。如果是低优数据，标准化转换公式则为

$$x'_{ij}=\frac{\overline{x}_j-x_{ij}}{\sigma_j}。$$

标准差分数可以回答这样一个问题，即“给定数据距离其均值多少个标准差”，在均值之上的数据会得到一个正的标准化分数，反之会得到一个负的标准化分数。这种转换方式因有负值存在，且不在 0-1 范围之内，因此一般将标准化值进一步做正态概率转换：

$$z'_{ij}=\frac{1}{\sqrt{2\pi}}e^{-\frac{\left(x'_{ij}\right)^2}{2}} \tag{47.14}$$

或线性转换：

$$z'_{ij}=(x'_{ij}+3)/6 \tag{47.15}$$

标准化转换不受最小值、最大值设置的影响，因此如果要进行标准化转换，那就无需设置各个因子的最小值(a)、最大值(b)。Min-Max 规格化转换是两端对齐：即各个因子的最小值、最大值都调整到接近 0 和 1，使得各个因子的取值在 0-1 之间。但标准化转换是向平均值(中心)处看齐，各个因子标准化转换后两端的取值可能不一样，即各因子的最小值、最大值可能不等，因为它不是对齐到 0-1 两端。

47.2.3 Kendall 协同系数检验

对 n 个样本 m 个指标进行综合评价，在研究各个因子权重，建立定量综合分析模型之前，我们可以检验一下这 m 个指标的趋势是否一致。如果一致，那么综合分析模型的建立更有意义。因此在应用综合评价模型时先检查一下这 m 个指标的趋势是否一致是有参考意义的，尽管这里的检验没有考虑到各个因子的权重大小。

Kendall 协同系数检验原假设 H_0 为“这些评估指标之间是不相关的或者是随机的”而备择假设 H_1 为：“它们之间呈正相关的或者是多少一致的”。Kendall 和 Slith(1939)提出的协同系数(coefficient of concordance)定义为

$$W=\frac{12S}{m^2(n^3-1)}$$

这里 S 是个体的总秩与平均秩的偏差的平方和，每个因子(共 m 个)对于所有参加排序的个体有一个从 1 到 n 的排列(秩)，而每个个体有 m 个打分(秩)，记 R_i 为第 i 个个体的秩的和(i □ l，2，…，n)，则

$$S=\sum_{i=1}^{n}\left(R_i-\frac{m(n+1)}{2}\right)^2$$

因此 Kendall 协同系数 W 还可以写成下面的形式

$$W=\frac{12\sum_i^n R_i^2-3m^2n(n+1)^2}{m^2n(n^2-1)}$$

上式右边的表达式计算起来较方便，W 的取值范围是 0~1。当 n 大时，可以利用大样本性质：在零假设下，对固定的 m，当 $n\to\infty$时

$$m(n-1)W=\frac{12S}{mn(n+1)}\to\chi^2_{n-1}$$

W 的值大(显著)，意味着各个因子趋势无明显不同，这样所产生的评估结果更有道理的；而如果 W 不显著，意味着各个因子的趋势各不相同，这时进行综合评价，有足够的理由相信这些因子能够产生一个共同的评估结果。

47.3 DPS 系统下最大熵-最小二乘建模

47.3.1 最大熵-最小二乘建模用户界面

DPS 系统下建立最大熵最小二乘综合评价及多属性决策分析，其数据格式是一行一个样本，一列一个变量（因子）。数据上面一行可以将各个因子的名称放入，这样输出结果更为直观。例如对我国 2007 年各地区农村居民家庭平均每人生活消费支出水平进行综合指数分析，将原始数据按一行一个地区、一列一个指标方式编辑如图 47-1 所示形式。

然后执行“专业统计”—“综合评价与多属性决策”—“最大熵-最小二乘模型”功能，这时系统出现供用户对各个指标特性进行设置的对话框(图 47-2 左)。

这里我们应用相加型方式汇集各个因子，在数据汇总方式选择框中选择“相加型”。因不需要进行数据转换，故在数据转换下面的选择框中选择“不转换”。数据标准化下面的多选框中，因这里是相加型模型，一般选择 Min-Max 规格化。

在选择框下面，系统给出了每个指标是否是低优指标、以及各个指标的平均值、最小值、最大值；以及是否强行限制到 0-1 之间。

	A	B	C	D	E	F	G	H	I
1	地区	食品	衣着	居住	家电	交通和通讯	文娱	医疗保健	其他
2	北　京	2132.51	513.44	1023.21	340.15	778.52	870.12	629.56	111.75
3	天　津	1367.75	286.33	674.81	126.74	400.11	312.07	306.19	64.30
4	河　北	1025.72	185.68	627.98	140.45	318.19	243.30	188.06	57.40
5	山　西	1033.68	260.88	392.78	120.86	268.75	370.97	170.85	63.81
6	内蒙古	1280.05	228.40	473.98	117.64	375.58	423.75	281.46	75.29
7	辽　宁	1334.18	281.19	513.11	142.07	361.77	362.78	265.01	108.05
8	吉　林	1240.93	227.96	399.11	120.95	337.46	339.77	311.37	87.89
9	黑龙江	1077.34	254.01	691.02	104.99	335.28	312.32	272.49	69.98
10	上　海	3259.48	475.51	2097.21	451.40	883.71	857.47	571.06	249.04
11	江　苏	1968.88	251.29	752.73	228.51	543.97	642.52	263.85	134.41
12	浙　江	2430.60	405.32	1498.50	338.80	782.98	750.69	452.44	142.26
13	安　徽	1192.57	166.31	479.46	144.23	258.29	283.17	177.04	52.98
14	福　建	1870.32	235.61	660.55	184.21	465.40	356.26	174.12	107
15	江　西	1492.02	147.71	474.49	121.54	277.15	252.78	167.71	61.08
16	山　东	1369.20	224.18	682.13	195.99	422.36	424.89	230.84	71.98
17	河　南	1017.43	189.71	615.62	136.37	269.46	212.36	173.19	62.26
18	湖　北	1479.04	168.64	434.91	166.25	281.12	284.13	178.77	97.13
19	湖　南	1675.16	161.79	508.33	152.60	278.78	293.89	219.95	86.88
20	广　东	2087.58	162.33	763.01	163.85	443.24	254.94	199.31	128.06
21	广　西	1378.78	86.90	554.14	112.24	245.97	172.45	149.01	47.98
22	海　南	1430.31	86.26	305.90	93.26	248.08	223.98	95.55	73.23
23	重　庆	1376	136.34	263.73	138.34	208.69	195.97	168.57	39.06
24	四　川	1435.52	156.65	366.45	142.64	241.49	177.19	174.75	52.56
25	贵　州	998.39	99.44	329.64	70.93	154.52	147.31	79.31	34.16
26	云　南	1226.69	112.52	586.07	107.15	216.67	181.73	167.92	38.43
27	西　藏	1079.83	245	418.83	133.26	156.57	65.39	50	68.74
28	陕　西	941.81	161.08	512.40	106.80	254.74	304.54	222.51	55.71
29	甘　肃	944.14	112.20	295.23	91.40	186.17	208.90	149.82	29.36
30	青　海	1069.04	191.80	359.74	122.17	292.10	135.13	229.28	47.23
31	宁　夏	1019.35	184.26	450.55	109.27	265.76	192	239.40	68.17
32	新　疆	939.03	218.18	445.02	91.45	234.70	166.27	210.69	45.25
33									

图 47-1　各地区农村居民家庭平均每人生活消费支出 (2007 年)

综合评价指数计算过程中，许多综合评价指数组成因子是用不同的单位来衡量的。数值越高，表现越好，大多数因子被认为是“好”。例外情况是“长期失业率”和“因慢性疾病而妨碍日常活动”，这两个因子都被认为是“不好的”，因为数值越高，表现越差。如前所述，在计算权重系数之前，这里可根据每个指标的特性，设置该指标是否是“低优指标”。

各个指标均值、最小值和最大值均显示在这里，供用户参考分析各个指标的趋势。大小及变化情况的参考。各个指标的最小值、最大值一栏，用户可以根据需要进行修改。在最小-最大规格化转换的处理时，系统将根据修改后的最小值、最大值进行数据预处理

计算。如将最小值全部改成零，这时相当于各个指标进行除以该指标最大值，即为除以最大值的转换；最小值改成零，最大值改成各个指标的均值，即为均值化转换。如果是考试成绩，满分是 120 分，60 分作为及格线，那么最大值可设为 120，最小值可设为 60。总之，根据各个指标的专业背景意义来设置最大值和最小值，可使得转换后的数值，既在各个指标之间具有可比性，又体现了各个因子的专业意义。

图 47-2 综合指数模型分析用户界面

如需要进行各个权重系数置信区间的估计，可以在下面的 Bootstrap 抽样次数输入框中输入抽样次数，缺省情形下是“0”，表示不进行 Bootstrap 抽样估计。注意，进行 Bootstrap 抽样时需要耗费的时间会较长，特别是因子指标比较多的时候。

原始数据经过我们设置的各个指标进行预处理之后，用最大熵-最小残差模型公式(7)对预处理之后的数据，经 SUMT 算法优化计算，给出如下结果。

相加型综合评价指数模型

数据 Min-Max 规格化处理

因子	是否低优	均值	标准差	最小值(a)	最大值(b)	是否 0-1 限制
食品	否	1424.9461	514.4770	939.0300	3259.4800	否
衣着	否	213.4490	101.1272	86.2600	513.4400	否
居住	否	601.6335	367.0977	263.7300	2097.2100	否
家电	否	155.3713	82.1620	70.9300	451.4000	否
交通和通讯	否	347.9865	179.6311	154.5200	883.7100	否
文娱	否	323.1948	200.4646	65.3900	870.1200	否

医疗保健	否	231.2929	124.3624	50	629.5600	否
其他	否	78.4332	43.2870	29.3600	249.0400	否

熵值=2.980593

残差=0.011633

目标函数对数值=-13.2752

模型拟合方差分析表

模型残差	平方和	自由度	均方
等权重模型	0.0366	247	0.0001
MEMR 模型	0.0014	240	0.0000
模型残差差值	0.0352	7	0.0050

F 统计量=4.3388　　显著性 p 值=0.0001

优化指数=11.2334%

权重系数 Bootstrap 抽样 1000 次估计,时间 0 分 5 秒。

因子	权重系数	抽样值均值	标准差	中位数	95%置信区间	
食品	0.1171	0.1158	0.0071	0.1155	0.0958	0.1392
衣着	0.1039	0.1048	0.0102	0.1049	0.0729	0.1369
居住	0.1186	0.1181	0.0088	0.1177	0.0879	0.1514
家电	0.1580	0.1572	0.0101	0.1574	0.1263	0.1884
交通和通讯	0.1612	0.1611	0.0082	0.1615	0.1350	0.1826
文娱	0.1096	0.1097	0.0088	0.1092	0.0793	0.1454
医疗保健	0.1072	0.1073	0.0075	0.1075	0.0842	0.1345
其他	0.1245	0.1260	0.0146	0.1248	0.0843	0.1654
熵值	2.9806	2.9766	0.0062	2.9769	2.9465	2.9946
			正态近似置信区间		2.9645	2.9888

样本名次及综合指数

地区	综合指数	排序	地区	综合指数	排序	地区	综合指数
上　海	0.9782	12	湖　南	0.2279	22	四　川	0.1493
北　京	0.7264	13	黑龙江	0.2273	23	青　海	0.1429
浙　江	0.7146	14	湖　北	0.2179	24	新　疆	0.1225
江　苏	0.4529	15	山　西	0.1851	25	广　西	0.1219
福　建	0.3313	16	河　北	0.1828	26	重　庆	0.1194
广　东	0.3205	17	江　西	0.1699	27	云　南	0.1133
山　东	0.3	18	安　徽	0.1661	28	海　南	0.1124
辽　宁	0.2837	19	河　南	0.1659	29	西　藏	0.1044
天　津	0.2751	20	宁　夏	0.1547	30	甘　肃	0.0621
内蒙古	0.2513	21	陕　西	0.1509	31	贵　州	0.0298
吉　林	0.2385						

47.3.2 多因子综合指数模型结果解读

系统首先输出最大熵-最小二乘非线性含条件约束模型的信息熵 E 等于 2.9806。拟合残差标准差(σ=0.01163)以及模型目标函数自然对数值为-13.2753。

模型输出结果的第二部分是模型的方差分析结果。检验的原假设为 H_0：因子 j 权重

系数等于 1/p；备择假设 H_1：因子中各个因子 j 权重系数不是都等于 1/p。这部分结果解读要点是从统计学角度，检验各个指标的权重系数的大小是否有差异。如果没有差异，那我们可以采用等权重的方式进行综合指数计算。如果这里的显著性概率 p 值小于 0.05，那我们就应该采用模型给出的各个指标的权重系数来计算综合评价指数。本例中显著性检验 p=0.0002，远小于 0.05，因此这里的数据不宜采用等权法估计综合评价指数，而应采用模型估计值进行综合指数的计算。

输出结果第三部分为各个指标的权重系数，是综合评价分析模型的核心和重点。它既是各个因子对形成综合指数的贡献大小，又是综合评价指数中各个因子的权重重要性的体现，是各个因子和综合指数相互之间，从集中趋势、离散程度及多指标之间的相互关系诸方面在最大熵条件下、拟合误差达到最小的结果。

这里的第一列是各个指标权重系数的估计值。它反映了各个因子在综合评价指数中的重要性，权重系数越大，该指标在综合评价中越是重要。本例中，根据权重系数来看，交通和通讯、家庭设备及服务两个指标对综合指数（综合评价）的影响较大；其次是其他商品及服务、居住和食品；文教娱乐用品及服务、医疗保健等其他指标的影响较小。

“抽样均值”和“标准差”两栏是根据 Bootstrap 抽样（这里是 1000 次）结果计算出的估计值，这里的标准差即为权重系数在正态假设下的标准误。根据标准误可以估计其 95%的置信区间。

右边 3 列，即“中位数”和“95%置信区间”是根据 Bootstrap 抽样的百分位数估计。一般来说 95%置信区间值大于 1/p 时，该权重系数显著偏大；小于 1/p 时，该权重系数显著偏小。如这里的家电和交通通讯两个因子，权重系数 95%置信区间分别为0.1317 ~0.1828 和 0.1269~0.1820,大于 1/p（0.125），故可认为这两个指标在统计学意义下，显著偏大。

Bootstrap 抽样估计，一方面是用于统计检验；另一方面是表达了各个权重系数的灵敏度。Bootstrap 抽样的百分位数 95%置信区间值的范围越小，说明该因子的权重系数在评估综合指数或多属性决策时较为稳健；反之权重系数在估计综合指数和多属性决策时更应慎重。通过 Bootstra 模拟抽样我们了估计每个因子权重系数的标准误。为考查估计精度，我们计算了各个因子标准误的均值，以及各个因子权重系数的均值（等于 1/p）。然后计算权重系数。变异系数通常用于比较不同数据集的离散程度。当数据集的变异系数较大时，表示数据的波动较大，离散程度较高；反之，当数据集的变异系数较小时，表示数据的波动较小，离散程度较低。因此，通过比较不同数据集的变异系数，我们可以判断它们的稳定性和一致性。并对加性模型、乘积模型以及概率加法模型，应用变异系数进行判断、选取变异系数较小的综合评价模型。

我们将各种数据规格化方法和各种模型对应起来进行建模，通过 Bootstrap 自助抽样我们得到相应的变异系数列于表 47-2。从表 47-2 可以可出，将最小最大规格化数据采用乘积模型进行分析不是很好的搭配，因为他们变异系数显著偏高。

注意，如果没有进行 Bootstrap 抽样，变异系数将无法计算。

表 47-2　不同数据规格化方法与模型因子权重变异系数

模型	最小最大规格化	标准化-正态概率	标准化-线性转换

加法模型	7.58	7.70	6.75
乘积模型	44.53	11.77	10.11
概率加法模型	7.80	6.07	5.59
Logit 模型	87.45	15.57	76.35
Probit 模型	23.83	9.78	30.08
重对数模型	46.75	12.97	65.52

界面的第二个表页是综合指数输出图示。图中以综合指数为横坐标，纵坐标是综合指数和各个因子的规格化转换值，图形以泡泡图方式显示（图 47-3）。该图表达的含义有 4 个方面：一是各个评价对象的排序及相似性，距离越近越相似；二是综合指数直线的变化趋势；三是各个因子规格化值围绕综合指数直线分布情形，越是集中，排序、或评价效果越好；四是各个因子权重系数，在综合指数图中每个因子以一个颜色显示，每个因子显示的泡泡大小是根据该因子的权重系数大小确定，权重系数越大泡泡越大，反之亦然，这可识别不同因子分布趋势和综合指数直线的一致性。

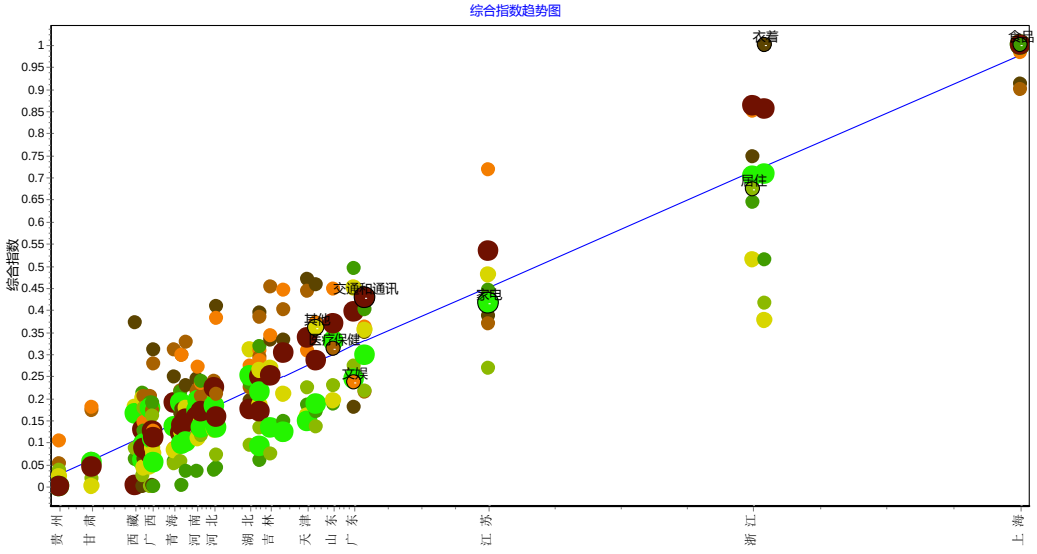


图 47-3　综合指数及各因子规格化值泡泡图

从图 47-3 中可以看出，交通和通讯、家电两个因子权重系数较大，在图中所显示的泡泡也较大。这两个指标的散点图趋势和综合指数蓝色直线的趋势的吻合程度最高。衣着、医疗与保健、文娱这三个因子权重系数较小，在图中显示的泡泡亦较小,且这三个因子规格化值偏离综合指数蓝色直线程度较大。一般来说，权重系数越大、泡泡越大的因子，和蓝色的综合指数直线的距离越近、趋势越一致。这也直观地说明了模型权重估计的合理性。从该图我们还可以看出各评价对象按综合指数排序的状况，分析综合指数曲线与各个因子指标点的离散程度。此外，各因子离散程度越小，综合指数代表性越好。

最后我们可根据系统给出了各个样本最大熵-最小残差模型综合评价指数数值结果，将评价指数从大到小进行排序，可得到各个样本排序结果。本例估计得到各地综合评价指数从大到小为：上海为 0.9782，北京 0.7264，浙江 0.7146，江苏 0.4529，福建 0.3313，广东 0.3205，山东0.3000，辽宁 0.2837，天津 0.2751，内蒙古 0.2513，吉林 0.2385，湖南 0.2279，黑龙江 0.2273，湖北 0.2179，山西 0.1851，河北 0.1828，江西 0.1699，安徽 0.1661，河南 0.1659，宁夏 0.1547，陕西 0.1509，四川 0.1493，青海 0.1429，新疆 0.1225，广西 0.1219，重庆0.1194，云南 0.1133，海南 0.1124，西藏 0.1044，甘肃 0.0621 及贵州为 0.0298。根据排序结果进行专业意义上的解读和评价对象的位次比较。如本例中，上海、北京和浙江排名为前三，消费水平最高；其次是江苏、福建、广东和山东，消费水平次之，等等…。

47.4 综合评价模型和目前常用模型比较

例 1　技术成就指数

这里仍以前面技术成就**指数**- TAI 例子来分析。各种方法在计算权重系数之前，均按(47.10*a*)式对原始数据进行了归一化处理。原始数据归一化处理之后，我们采用等权法、主成分分析法(PCA，Dunteman G. H. 1989)、因子分析法(FA，Harman H H . 1967)、熵权法（Ent.）、标准离差法(SD，Irik Mukhametzyanov, 2021)、Criteria importance though intercrieria correlation method 法(CRITIC，Diakoulaki, D., Mavrotas, G., & Papayannakis, L. 1995），以及本文提出的基于最大熵-最小残差的综合评价**指数**赋权法计算出各个因子的权重系数如图 47-4。

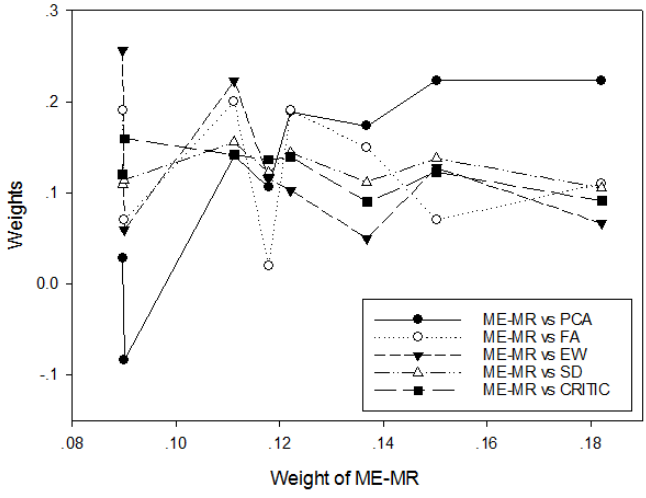


图 47-4　技术成就**指数**数据几种赋权法权重系数

将 MEMR 权重、第一主成分、因子分析、熵值权重、变异系数权重、以及标准离差权进行相关分析发现，仅 MEMR 权重和第一主成分之间高度正相关。主成分分析是

业界公认的数据降维方法。例子数据结果显示，MEMR 权重和主成分分析的第一主成分的因子载荷系数的大小趋势一致，相关系数为 0.8263。这表明 DPS 权重在本质上也是一种数据降维方法。

MEMR 权重和第一主成分这两种权重系数和其它的赋权法达到的权重系数多为负相关。这表明这几种权重尽管解释了评价因子的变化情况，但并没有起到数据降维、综合各因子为综合**指数**的作用。在欧盟 OECD/EU-JRC(2008)在编制的《综合评价**指数**编制手册》中，熵值权重、变异系数和 CRITIC 这几种赋权方法未提及。

分别将 MEMR 法、熵权法、变异系数法和标准离差法估计得到的综合指数和各因子权重及规格化值关系做成直线-泡泡组合图。泡泡图直径统一定义为 $0.025\times w_i/w_{max}$。这里 w_i 是第 i 个因子的权重系数，w_{max} 是所有权重系数中的最大值。各种估计方法得到的组合图列于图 47-5。图 47-5 的左上方是 MEMR 法综合指数和各个因子权重及规格化值关系图，右上方是熵权法、左下方是变异系数法、右下方是标准离差法。

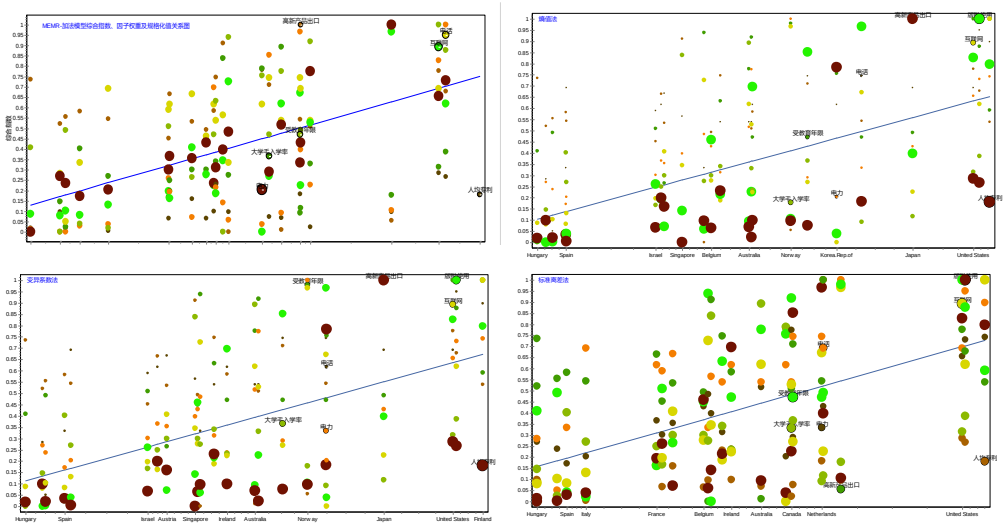


图 47-5　不同方法综合指数估计值和各因子权重及规格化值关系图

从图 47-5 可以看出，只有 MEMR 法建立的模型，其权重较大的因子靠近综合指数直线的两侧分布而权重系数较小的因子分布在离综合指数直线较远的位置。因子权重大小的分布和综合指数大小趋势一致。但熵权法、变异系数法和标准离差法则没有类似的趋势。这说明熵权法、变异系数法和标准离差法得到的因子权重似乎不能反映由此估计的综合指数的某种趋势。

例 2　虚拟例子（正交表）

这里我们以一个 2 水平 1 因素+3 水平 1 因素+4 水平 1 因素+8 水平 1 因素，共 4 个因素 24 个样本的虚拟多因子数据集(表 47-3)。

根据表 47-3 的模拟数据，我们看比较几种主要对因子权重系数赋权方法的计算结果。因为这时各个因子相互正交，不存在因子之间的相关性；同时，各个因子重要程度也可定性地直观地辨别。

表 47-3 中 4 个变量数据直观分析可知：一是该表是正交表，各因子之间相互独立、整齐可比、没有随机误差；二是各个因子水平不同，水平数较多的因子所包含的信息量要大，赋权时可能有较大的权重系数。在计算权重系数之前，我们按(47.13*a*)式对原始数据进行了归一化处理。

表 47-3　24 处理混合水平正交表

处理号	x1	x2	x3	x4
1	1	1	1	1
2	1	1	2	4
3	2	1	3	3
4	2	1	4	2
5	2	1	1	8
6	2	1	2	5
7	1	1	3	6
8	1	1	4	7
9	1	2	1	2
10	1	2	2	3
11	2	2	3	4
12	2	2	4	1
13	2	2	1	7
14	2	2	2	6
15	1	2	3	5
16	1	2	4	8
17	1	3	1	2
18	1	3	2	3
19	2	3	3	4
20	2	3	4	1
21	2	3	1	7
22	2	3	2	6
23	1	3	3	5
24	1	3	4	8

然后我们采用主成分分析法、熵权法、标准离差法及 criteria importance though intercrieria correlation method(CRITIC)法及本文提出的基于最大熵-最小残差的综合评价指数赋权法计算各个因子的权重系数。由于因各个因子间相关系数为零，CRITIC 法(Diakoulaki 等,1995)计算权重和标准离差法权重结果相同。几种主要赋权方法计算得到的权重系数结果分析如表 47-4。

如表 47-4 所示，主成分分析法，加权处理只是为了矫正两个或多个相关因子之间的重叠信息，而不是衡量相关指标理论重要性。如果因子之间没有相关性，那么就不能用这种方法来估计权重，本例即不适合主成分分析估计。因为从分析结果可以看出，作为权重系数的规格化特征向量是主对角线上元素都为 1，其余元素全为 0 的 4 阶单位矩阵。

表 47-4 正交表虚拟样本数据几种赋权法权重系数

因子	水平数	MEMR	熵值	标准离差	CRITIC	PCA1
x_1	2	0.2040	0.3850	0.3109	0.3109	1
x_2	3	0.2469	0.2567	0.2538	0.2538	0
x_3	4	0.2642	0.2082	0.2317	0.2317	0
x_4	8	0.2849	0.1502	0.2035	0.2035	0

熵权法得到的权重系数，当各个因子的水平数分别为 2，3，4 和 8 时，权重系数值分别为 0.3850，0.2567，0.2082 和 0.1502。呈现随着因子水平数越大，权重系数越小的趋势。这有悖于我们对数据所含信息量大小的直观分析。即，如果我们将各个因子的各个水平视为等级变量，应该是等级多的变量比等级较少的变量通常具有更大的信息量，能够提供更多的分析和推断可能。这里 x_4 有 8 个等级水平，接近数量指标，包含信息量理应较大，即在综合评价指数中含有较大的信息量，具有较大的权重系数。但这里的权重系数却不到两水平因子(x_1)的一半。这说明熵权法作为计算权重系数的方法值得商榷。

标准离差法和 CRITIC 法，这两种方法权重系数相同：当各个因子的水平数分别为 2，3，4 和 8 时，权重系数值分别为 0.3109，0.2538，0.2317 和 0.2035。这个结果和熵权法类似，即呈现随着因子水平数越大，权重系数越小的趋势。这有悖于我们对数据所含信息量大小的直观分析。因此和熵权法一样，标准离差法和 CRITIC 法作为计算权重系数的方法值亦得商榷。

采用本文提出的直接拟合综合评价指数模型的计算方法，计算结果显示，当各个因子的水平数分别为 2，3，4 和 8 时，权重系数值分别为 0.2040，0.2469，0.2642 和 0.2849。即呈现随着因子水平数增加，权重系数增大。这和我们的直观判断，即等级较多的变量比等级较少的变量通常具有更大的信息量的认知一致。

另外，以正交表作为例子，如果对任意正交表中各个因子进行减去平均值、除以标准差的标准化转换，那么这时各个因子的均值为零、标准差为 1。这时根据模型(4)计算出的各个因子的权重系数相同。这也从另一个方面说明正交表具有均匀分散、整齐可比的特点。

例 3 虚拟例子（企业综合发展状况评价（苏为华等，2021））

一家大型投资公式为提高投资效率，拟对 15 家企业从 10 个指标进行评估。评估结果经 min-max 规格化转换后列表于表 47-5。

根据表 47-5 数据，苏为华(2021)采用多种 BoD 及其改进方法构造权重并计算综合指数。相应的 BoD 方法及其结果如表 47-6。

非补偿型 BoD 方法，由 Cherchye L 等(2007)进行了系统的介绍。BoD 方法存在较为严重的 0-1 权重问题，实践中往往需要对 BoD 的权重计算结果中的 0 权重按照一定规则进行补偿，而未对权重进行补偿的 BoD 模型通常称为非补偿型 BoD。企业综合发展状况各指标采用非补偿型 BoD 结果列于表 47-6。从结果来看，无法用于各个企业的名次排序。

Cherchye L （2007）在介绍了非补偿性 BoD 方法之后，提出在非补偿型 BoD 方法

约束中加入一个刚性的分值补偿约束，以解决过柔性问题，并列举了 5 种权重补偿方法。这里我们以绝对补偿为例，将约束值设置为 0.85，这时计算得到的企业综合发展状况各指标采用绝对补偿型 BoD 排名指数如下（表 47-6）。从结果来看，无法完成各个企业的名次排序。

表 47-5　15 家企业发展的综合状况规格化数据

企业	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10
1	1.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
2	0.9	0.2	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
3	0.8	0.9	0.7	0.1	0.2	0.3	0.4	0.5	0.6	0.7
4	0.7	0.8	0.9	0.3	0.1	0.2	0.3	0.4	0.5	0.6
5	0.6	0.7	0.8	0.9	1.0	0.1	0.2	0.3	0.4	0.5
6	0.5	0.6	0.7	0.8	0.9	1.0	0.1	0.2	0.3	0.4
7	0.4	0.5	0.6	0.7	0.8	0.9	1.0	0.1	0.2	0.3
8	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	0.1	0.2
9	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.6	0.1
10	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
11	0.8	0.3	0.2	0.5	0.6	0.5	0.8	0.6	0.4	0.7
12	0.9	1.0	0.2	0.5	0.6	0.5	0.8	0.6	0.4	0.7
13	0.8	0.2	1.0	0.6	0.6	0.5	0.8	0.6	0.4	0.7
14	0.8	0.2	0.6	1.0	0.6	0.5	0.8	0.6	0.4	0.7
15	0.8	0.2	0.6	0.2	0.6	0.5	0.8	0.6	1.0	0.7

表 47-6　15 家企业综合评价指数

企业	Non-Compen-Satory BoD	Compensatory BoD	Generalized BoD	two sets of weights BoD	ABC 法	SEW	MEMR
1	1.0	0.9548	1.0	0.5	1.00	0.55	0.5529
2	1.0	0.9724	1.0	0.5	0.95	0.47	0.4692
3	1.0	0.9985	1.0	1.0	0.90	0.52	0.5007
4	1.0	0.9985	1.0	0.5057	0.85	0.48	0.4604
5	1.0	1	1.0	0.5	0.80	0.55	0.5478
6	1.0	0.9985	1.0	0.5	0.75	0.55	0.5573
7	1.0	No solution	1.0	0.6264	0.70	0.55	0.5575
8	1.0	1	1.0	0.5	0.65	0.55	0.5634
9	1.0	0.9868	1.0	0.5	0.60	0.51	0.5223
10	1.0	No solution	1.0	0.5	0.55	0.55	0.5621
11	1.0	0.385	0.9938	0.5	0.80	0.54	0.5449
12	1.0	No solution	1.0	0.5	0.90	0.62	0.6111
13	1.0	1	1.0	0.5	0.80	0.62	0.6228
14	1.0	1	1.0	0.5	0.63	0.62	0.6268
15	1.0	1	1.0	0.5	0.80	0.60	0.6020

广义 BoD。这是 Rogge N（2017）将 BoD 中所有计算符号放宽为一般累加算子“⊕”

和一般累乘算子“⊗”，并采用几何加权汇总模式的数学规划方法。企业综合发展状况各指标采用几何平均-广义 BoD 计算的结果列于表 47-6。从计算结果来看，亦无法用于各个企业的名次排序。

为了增强 BoD 的区分度，Zhou 等（2007）提出了从正反两个方面计算“好指数”和“坏指数”的思想的 two sets of weights 方法，为原始的 BoD 分析结果增加了更多的信息，以期获得区分度更高的双向综合指数。企业综合发展状况例子分别计算好指数和坏指数之后，以调节参数λ=0.5 计算综合分值，这时各指标综合分值列于表 47-6。从计算结果来看，亦无法用于各个企业的名次排序。

如果了解了评价体系中各个指标相对重要性时，这时可采用 Wan(2007)提出的一种类似 BoD 的多准则 ABC 分析法，其模型形式与 BoD 大体相同。这里，我们假定企业综合发展状况各指标的重要性从左到右依次递减，这时采用 ABC 分析法可以获得 15 家企业有序加权后的综合分值（表 47-6）。

最后，我们采用等权法（SEW）和 MEMR 计算各个企业的综合指数，将各个企业综合指数列于表 47-6 的最右边两列。

从表 47-6 可以看出各种 BoD 方法均不适合该数据集的综合评价。对 ABC 法、SEW 法和 MEMR 给出的结果进行相关分析得知，ABC 法和 SEW 法、MEMR 法的相关系数分别为-0.1694和-0.2934，均为负相关，MEMR 法和 SEW 法的相关系数为 0.9801，高度正相关。这表明表明 ABC 法不能反映各个企业综合指数的趋势，通过 MEMR 得到的综合指数反映了各个企业的综合趋势，可以较好地完成各个企业间的优先排序。

47.5 概率加法模型

47.5.1 概率加法模型

概率加法原理是概率论中的一个基本概念，用于计算多个事件的联合概率。基于概率加法规则的多属性决策概率加法模型，可以表达在多个指标同时呈现时，响应（即决策）发生的概率如何增加。根据这个模型，在多个指标情况下进行多属性的决策是单个指标呈现的效应的概率之和。

概率加法模型应用概率加法公式，计算两个或多个事件概率之和。具体来说，设 A 和 B 是任意二事件，则概率加法公式为：$P(A \cup B) = P(A) + P(B) - P(A \cap B)$。其中，$P(A \cup B)$表示事件 A 或事件 B 发生的概率，$P(A)$和 $P(B)$分别表示事件 A 和事件 B 发生的概率，$P(A \cap B)$表示事件 A 和事件 B 同时发生的概率。

简单来说，概率加法模型表明，当有多个独立指标或信息源可用时，综合指数或多属性决策的总体概率会随着每个独立指标对决策过程的贡献而增加。这个模型通常用于解释人类如何整合和处理来自不同感官模式的信息，以做出决策或判断。因此概率加法模型是一种完整的多属性决策方法。

和加型、乘积型综合评价模型相比，概率加法模型更充分地考虑系统诸因子之间

的相互关系。多指标综合评价中，各个因子之间的相互关系对综合评价结果的影响一直是人们关注的话题，但也是在综合评价过程中没有得到很好地解决的一个方面。应用概率加法模型我们可以相当完美地解决了这个问题。概率加法模型完全、充分地将各个指标之间的互作效应在综合评价过程中加以考虑，强调系统多因子互作关系对评价结果的影响时可考虑使用概率加法多因子综合评价模型。

当综合评价过程是概率加法模型形式时，那么含两个因子权重系数按概率加法得到的评价指数(y_i^{pr})为：

$$y_i^{pr} = x_{ij}w_j + w_{ik}w_k - (w_{ij}w_j)(x_{ik}w_k) \tag{47.16}$$

当因子个数为 3 时，有：

$$\begin{aligned} y_i^{pr} &= x_{ij}w_j + w_{ik}w_k + w_{il}w_l \\ &-(w_{ij}w_j)(x_{ik}w_k) - (w_{ij}w_j)(x_{il}w_l) - (w_{ik}w_k)(x_{il}w_l) \\ &+(w_{ij}w_j)(x_{ik}w_k)(w_{il}w_l) \end{aligned} \tag{47.17}$$

当评价指标有 p 个时，按概率加法规则，得到的评价指数如式（47.18）所示

$$\begin{aligned} y_i^{pr} &= \sum_{j=1}^{p} x_{ij}w_j - \sum_{1\le j\le k\le p}^{p} (x_{ij}w_j)(x_{ik}w_k) \\ &+ \sum_{1\le j\le k\le l\le p}^{p} (x_{ij}w_j)(x_{ik}w_k)(x_{il}w_l) + \cdots + (-1)^{p-1}(x_{i1}w_1)(x_{i2}w_2)\cdots(x_{ip}w_p) \end{aligned} \tag{47.18}$$

式中 $w_j \ge 0$， $w_1 + w_2 + w_3 + \cdots + w_p = 1$。

在概率加法模型中，当约束各因子的权重系数之和为 1 时，因加法规则考虑了因子间交互作用，估计所得到的模型理论值小于加法模型的理论值，如两因子(式 47.16)，和加法模型理论值相比，应用概率加法得到的评价指数减少了一个乘积项的值。因此在概率加法公式 47.18 中，我们通过引入一个类似回归方程中的常数项来矫正因交互项而导致概率加法模型理论值偏小的问题（式 47.19）。常数项的加入可以提高拟合精度，即使得残差平方和更小一些，但不影响综合指数即多属性决策结果的排序。

$$\begin{aligned} y_i &= w_0 \times y_i^{pr} \\ &= w_0 \times \Big(\sum_{j=1}^{p} x_{ij}w_j - \sum_{1\le j\le k\le p}^{p} (x_{ij}w_j)(x_{ik}w_k) \\ &+ \sum_{1\le j\le k\le l\le p}^{p} (x_{ij}w_j)(x_{ik}w_k)(x_{il}w_l) + \cdots + (-1)^{p-1}(x_{i1}w_1)(x_{i2}w_2)\cdots(x_{ip}w_p)\Big) \end{aligned} \tag{47.19}$$

从式（47.19）可以看出，概率加法模型将所有的因子间的互作效应都考虑进去了，且是一个极为复杂，随着因子数量的增加、计算工作量急剧增大的统计模型。

和前面加法模型类似，可将综合评价指数（y_i）看成是由多个理论组分 C_{ij} 组成，第 i 个综合评价指数(y_i)的第 j 个组分 C_{ij} 的值等于综合评价指数(y_i)乘以相应的权重 w_j，即：

$$C_{ij}=y_i \cdot w_j$$

也就是说，综合评价指数(y_i)我们也可以用理论组分之和与其常系数项的乘积的形式来表示：

$$y_i = C_{i1} + C_{i2} + C_{i3} + \cdots + C_{ip}$$

我们令相应的观察组分值和理论组分值之差为残差，$\sigma_{ij}=x_{ij} \cdot w_j - C_{ij}$，则 σ_{ij} 可视为第 i 个样本，第 j 个因子指标的拟合误差(残差)。假设拟合误差项 σ_{ij} 在各个因子之间是独立的，因为理想情况下，每个单独因子的组分可认为独立于其他因子来衡量综合评价指数特征的一个特定方面。

综合评价指数统计模型中，为求解各个因子的权重系数，应该使得第 i 个样本，第 j 个因子指标的理论组分值 C_{ij} 和相应的观察组分值($x_{ij} \cdot w_j$)之间的残差尽可能地小。类似于线性模型中离差平方和的分解，综合评价指数模型的均方根误差，即残差 σ 为

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{p}\left(C_{ij} - x_{ij}w_j\right)^2}{(n-1)\cdot p}} \tag{47.20}$$

式(47.20)中 $(n-1)\cdot p$ 为自由度。式(47.20)可作为多因子综合评价指数目标函数模型。即使得残差 σ 最小以求出权重系数 w_j。

在多因子综合评价指数研究中，我们将各因子权重系数视为各因子在综合评价指数中重要性的某种概率分布，那么概率分布的最大熵形式可通过最大化 Shannon 信息熵得到。综合评价指数中各个因子权重系数 w_j 的 Shannon 信息熵公式为

$$E = -\sum_{j=1}^{p} w_j \ln(w_j) \tag{47.21}$$

仍和前面的加法模型一样，用于估计综合评价指数中各个因子的权重系数，综合评价指数(MEMR)模型误差函数 $f(\sigma, E)$ 的形式定义为

$$f(\sigma, E) = \min \sigma^E \tag{47.22}$$

式(47.22)中 σ 表示残差，E 表示各因子权重的 Shannon 信息熵。由于该模型中指数函数的底，即残差(σ)总是小于 1 的，这时指数（信息熵 E）越大，其函数值越小。因此指数方程的这一特点可以用于实现综合评价指数模型中最大熵-最小残差的优化目标。

如果原始数据不进行转换，这时可以将各个原始数据观察值除以所有观察值中（绝对值）的最大值，将所有观察值的取值转换到 0—1 之间，成为比率值，再进行优化建模。

这时在最大熵条件下使得残差σ最小以求出权重系数w_j。极大熵-极小残差为目标函数的乘积型综合评价指数(MEMR)模型的误差函数公式如下：

$$\begin{cases} \min \sigma^E \\ \text{s.t} \sum_{j=1}^{p} w_j = 1.0, \quad w_{j\geq} 0 (j = 0,1,..., p) \end{cases} \tag{47.23}$$

非线性回归方程模型式(47.23)的优化，亦可采用序列无约束极小化(SUMT)方法进行综合评价指数模型的优化求解。在 DPS 数据处理系统之中，概率加法模型的统计计算的用户界面和加法模型等公用一个用户界面。

47.5.2 概率加法模型应用示例

这里以 Tzeng G. H., Huang J. J.(2011)采用 VIKOR 法、TOPSIS 法对不同类型动力汽车评价、研究例子数据为例介绍概率加法模型的应用。该例研究的汽车动力类型包括：

①常规柴油机：柴油发动机是所有现有内燃机中效率最高的，是目前动力源的主要竞争者，将其引入备选方案以便与新燃料模式进行比较。

②压缩天然气(CNG)：压缩天然气已经在全球范围内商业化，技术成熟，排放二氧化碳量少，具有高辛烷值。

③液化气(LPG)：亦具排放二氧化碳量少的特点。

④燃料电池(氢)：可以将氢和氧转化为车辆的动力，由于能量来自氢和氧之间的化学反应，因此不会产生有害物质，只排放以空气形式存在的纯净水。

⑤甲醇：发动机可以在任何气体与甲醇的混合率下平稳运行，甲醇将作为替代燃料，有助于减少黑烟和氧化亚氮的排放。但甲醇的热能低于汽油，其连续行驶能力不如传统车辆。此外，燃烧甲醇产生的乙醛化合物会形成强酸。研究人员应该对这种燃料模式给予更多的关注。

⑥机会充电电动：动力源是负载电池和公共汽车停车空闲时趁机会快速充电的结合。每当公共汽车从车站出发时，它的电池就会充满电，提供足够的电力使其移动到下一个公共汽车站。

⑦直接充电：动力主要来自装载的电池。当电池电量不足时，车辆必须返回工厂充电。开发一种合适的电池对这种模式的车辆至关重要。如果能在电池中储存更多的电量，该车的巡航距离将会增加。

⑧可更换电池：可更换电池的电动巴士的目标是实现快速充电和更长的巡航距离。对公交车进行了改装，以创造更多的车载电池空间，并调整了车载电池的数量，以满足不同路线的需求。快速交换设施必须准备好进行快速电池交换，以保持车辆的机动性。

⑨汽油发动机混合动力：以电动机为主要动力源，配以小型汽油发动机。当电力中断时，汽油发动机可以接管并继续旅行。在行驶过程中产生的动能将转化为电能，以增加车辆的巡航距离。

⑩柴油混合动力：以电动机和小型柴油发动机为主要动力源。当电力中断时，柴油发动机可以接管并继续行程，而行驶过程中产生的动能将转化为电能，以增加车辆的巡

航距离。

⑪带 CNG 发动机的混合动力电动:以电动机和小型 CNG 发动机为主要动力来源。当电力故障时，CNG 发动机接管并提供动力，将产生的动能转换为电能，以允许连续行驶。

⑫带 LPG 发动机的混合动力：以电动机和小型 LPG 发动机为主要动力源。当电力故障时，液化石油气发动机立即接管并提供动力，将产生的动能转换为电能，以允许持续行驶。

综合评价准则：从不同方面对各种燃料模式进行评价。这里考虑了四个方面的评价标准:社会、经济、技术和交通。为了对备选方案进行评价，建立了 11 个评价标准，并采用专家评分方法，给出各个指标的分值，这些指标有:

①能源供应:该标准基于每年可供应的能源量，能源供应的可靠性，能源储存的可靠性以及能源供应的成本。②能源效率:该标准表示燃料能源的效率。③空气污染:该标准是指一种燃料模式对空气污染的影响程度，因为不同燃料模式的车辆对空气的影响是不同的。④噪声污染:本标准是指车辆运行过程中产生的噪声。⑤产业关系:传统的汽车工业是机车工业，与其他工业生产有着错综复杂的关系;每一种替代品与其他工业生产的关系被作为标准。⑥实施成本:该标准是指生产和实施替代车辆的成本。⑦维护成本:替代车辆的维护成本是标准。⑧车辆性能:该标准代表巡航距离、爬坡和平均速度。⑨道路设施:该标准是指替代车辆运行所需的道路特征(如路面和坡度)。⑩交通速度:该标准是指在一定交通条件下，不同车辆的平均速度的比较。如果交通流的速度高于车辆的速度，车辆将不适合在某些路线上运行。⑪舒适性:这一标准是指舒适度这一特定问题，是指用户倾向于关注车辆的配件(空调、自动门等)。针对这些条目，项目组聘请多位专家进行评分，多位专家评分结果列于图 47-6 所示形式。

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		能源供应	能源效率	空气污染	噪音污染	产业关系	实施成本	维护成本	车辆能力	道路设施	通勤速度	舒适度	
2	常规柴油机	0.82	0.59	0.18	0.42	0.58	0.36	0.49	0.79	0.81	0.82	0.56	
3	压缩天然气(CNG)	0.77	0.70	0.73	0.55	0.55	0.52	0.53	0.73	0.78	0.66	0.67	
4	液化气(LPG)	0.79	0.70	0.73	0.55	0.55	0.52	0.53	0.73	0.78	0.66	0.67	
5	燃料电池(氢)	0.36	0.63	0.86	0.58	0.51	0.59	0.74	0.56	0.63	0.53	0.70	
6	甲醇	0.40	0.54	0.69	0.58	0.51	0.52	0.68	0.52	0.63	0.60	0.70	
7	**电动汽车**	**0.69**	**0.76**	**0.89**	**0.60**	**0.72**	**0.80**	**0.72**	**0.54**	**0.35**	**0.79**	**0.73**	
8	**直接充电**	**0.77**	**0.79**	**0.89**	**0.59**	**0.73**	**0.80**	**0.72**	**0.47**	**0.44**	**0.87**	**0.75**	
9	**可更换电池**	**0.77**	**0.79**	**0.89**	**0.59**	**0.73**	**0.80**	**0.72**	**0.51**	**0.48**	**0.87**	**0.75**	
10	汽油混合动力	0.77	0.63	0.63	0.52	0.66	0.63	0.65	0.67	0.70	0.80	0.74	
11	柴油混合动力	0.77	0.63	0.51	0.58	0.66	0.63	0.65	0.67	0.70	0.80	0.74	
12	电动 CNG	0.77	0.73	0.80	0.48	0.63	0.66	0.65	0.67	0.71	0.62	0.78	
13	电动 LPG	0.77	0.73	0.80	0.48	0.63	0.66	0.65	0.67	0.71	0.62	0.78	
14													

图 47-6　专家对各评价指标评价均值

然后执行“专业统计”—“综合评价与多属性决策”—“最大熵-最小二乘模型”功能，这时系统出现供用户对各个指标特性进行设置的对话框(图 47-7 左)。

这里各个评价指标已是标准化的数值，无需进行转换。这时转换方式可选 Min-Max

规格化，并将各个指标的最小值都设置成 0，最大值设置成 1。

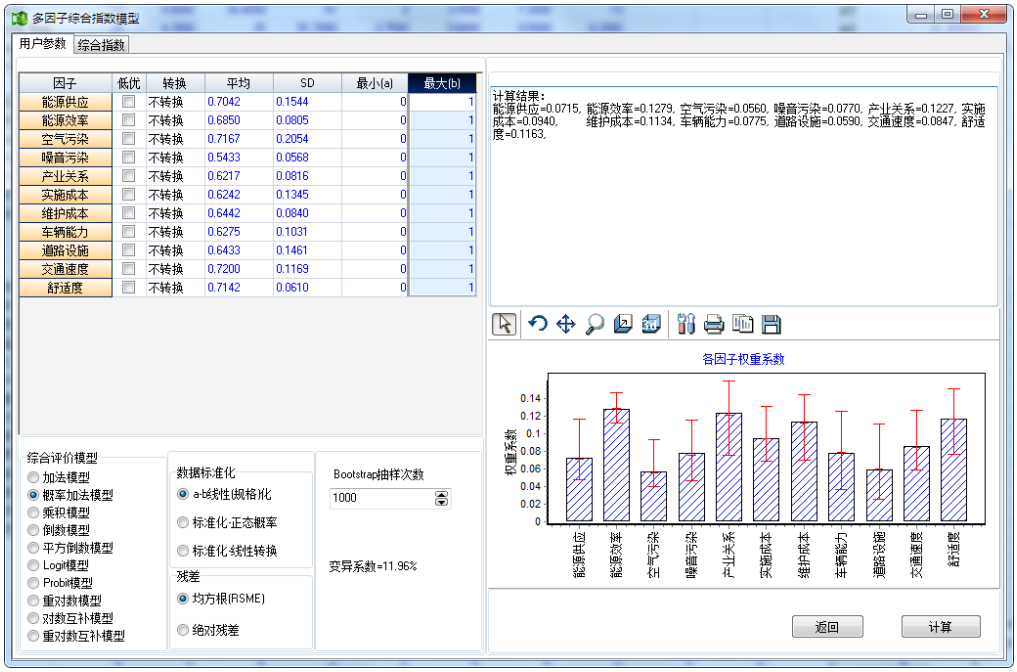


图 47-7 概率加法综合指数模型分析用户界面

这里我们作为例子，可以考虑各个评价指标之间的互作，选用概率加法综合评价方式汇集各个因子，在数据汇总方式选择框中选择“概率加法”。为估计各个权重系数的置信区间，将 Bootstrap 模拟抽样次数设置为 1000。模型残差指定为均方根误差。

点击“计算”按钮，系统即执行优化计算。因这里需要进行 1000 次的 Bootstrap 模拟抽样，即拟合 1000 个概率加法的非线性模型优化，因此要花费一些时间，用户须稍等一会，系统才能完成计算。完成计算之后，系统输出和加法型模型输出格式完全相同的计算结果如下。

概率加法型综合评价与多属性决策
Min-Max 规格化处理

因子	是否低优	数据转换	均值	标准差	最小值(a)	最大值(b)
能源供应	否	不转换	0.7042	0.1544	0	1
能源效率	否	不转换	0.6850	0.0805	0	1
空气污染	否	不转换	0.7167	0.2054	0	1
噪音污染	否	不转换	0.5433	0.0568	0	1
产业关系	否	不转换	0.6217	0.0816	0	1
实施成本	否	不转换	0.6242	0.1345	0	1
维护成本	否	不转换	0.6442	0.0840	0	1
车辆能力	否	不转换	0.6275	0.1031	0	1
道路设施	否	不转换	0.6433	0.1461	0	1
交通速度	否	不转换	0.7200	0.1169	0	1
舒适度	否	不转换	0.7142	0.0610	0	1

熵值=3.406961
残差=0.009566
目标函数对数值=-15.8407

模型拟合方差分析表

模型残差	平方和	自由度	均方
等权重模型	0.0153	131	0.0001
MEMR 模型	0.0111	121	0.0001
模型残差差值	0.0042	10	0.0004

F 统计量=4.6346　　显著性 p 值=0.0001
优化指数=27.6948%

权重系数 Bootstrap 抽样 1000 次估计,时间 0 分 18 秒。

因子	权重系数	抽样均值	标准差	中位数	95%置信区间	
常数项	1.3379	1.3397	0.0122	1.3394	1.3005	1.3901
能源供应	0.0715	0.0723	0.0093	0.0711	0.0482	0.1237
能源效率	0.1279	0.1287	0.0048	0.1287	0.1157	0.1487
空气污染	0.0560	0.0571	0.0084	0.0556	0.0389	0.0896
噪音污染	0.0770	0.0751	0.0106	0.0748	0.0414	0.1124
产业关系	0.1227	0.1209	0.0101	0.1205	0.0900	0.1566
实施成本	0.0940	0.0943	0.0102	0.0931	0.0696	0.1261
维护成本	0.1134	0.1114	0.0095	0.1118	0.0819	0.1433
车辆能力	0.0775	0.0778	0.0135	0.0771	0.0344	0.1278
道路设施	0.0590	0.0601	0.0135	0.0586	0.0276	0.1178
交通速度	0.0847	0.0856	0.0090	0.0843	0.0641	0.1155
舒适度	0.1163	0.1167	0.0113	0.1170	0.0764	0.1497
变异系数	11.85					
熵值	3.4070	3.3990	0.0197	3.4021	3.3023	3.4388
			正态近似		3.3604	3.4376

样本转换值及综合指数

样本	能源供应	能源效率	空气污染	噪音污染	产业关系	实施成本	维护成本	车辆能力	道路设施	交通速度	舒适度	综合指数
常规柴油机	0.82	0.59	0.18	0.42	0.58	0.36	0.49	0.79	0.81	0.82	0.56	0.6028
压缩天然气(CNG)	0.77	0.70	0.73	0.55	0.55	0.52	0.53	0.73	0.78	0.66	0.67	0.6474
液化气(LPG)	0.79	0.70	0.73	0.55	0.55	0.52	0.53	0.73	0.78	0.66	0.67	0.6484
燃料电池(氢)	0.36	0.63	0.86	0.58	0.51	0.59	0.74	0.56	0.63	0.53	0.70	0.6236
甲醇	0.40	0.54	0.69	0.58	0.51	0.52	0.68	0.52	0.63	0.60	0.70	0.6013
电动汽车	0.69	0.76	0.89	0.60	0.72	0.80	0.72	0.54	0.35	0.79	0.73	0.6923
直接充电	0.77	0.79	0.89	0.59	0.73	0.80	0.72	0.47	0.44	0.87	0.75	0.7054
可更换电池	0.77	0.79	0.89	0.59	0.73	0.80	0.72	0.51	0.48	0.87	0.75	0.7090
汽油混合动力	0.77	0.63	0.63	0.52	0.66	0.63	0.65	0.67	0.70	0.80	0.74	0.6703
柴油混合动力	0.77	0.63	0.51	0.58	0.66	0.63	0.65	0.67	0.70	0.80	0.74	0.6689
电动 CNG	0.77	0.73	0.80	0.48	0.63	0.66	0.65	0.67	0.71	0.62	0.78	0.6764
电动 LPG	0.77	0.73	0.80	0.48	0.63	0.66	0.65	0.67	0.71	0.62	0.78	0.6764

输出结果的第一部分是模型的方差分析，检验的原假设为各因子权重系数相等；备择假设是至少有一个因子权重系数不相等。若原假设成立，那我们可用等权重方式进行综合指数计算。这里方差分析显著性概率 p 值小于 0.01，表明不宜采用等权法估计综合

评价指数，而应采用综合评价模型进行综合指数的计算更合理一些。

各个指标权重系数的输出结果，既显示了各因子对形成综合指数的贡献大小，又是综合评价指数中各个因子的权重重要性的体现。和加法模型相比，概率加法模型更是体现了各个因子和综合指数相互之间的交互作用。

从权重系数大小来看，能源效率(0.1279)和产业关系(0.1227)两个指标对综合评价的影响较大；其次是舒适度(0.1163)和维护成本(0.1134)。这些因子是开发不同能源类型汽车需优先考虑的。

Bootstrap 抽样百分位数估计，一般来说 95%置信区间值大于各个权重系数的平均值(这里为 1/12=0.08333)时，该权重系数显著偏大；小于它时，该权重系数显著偏小。如这里的能源效率因子权重系数 95%置信区间为0.1159 ~0.1441, 大于权重系数均值（0.0833），故可认为该指标在统计学意义下，显著偏大。其次是产业关系，亦可认为在统计学意义下，显著偏大。

概率加法模型估计得到的综合评价指数，和 Tzeng G. H., Huang J. J.(2011)的原文中采用 VIKOR 法、TOPSIS 法计算得到的综合指数比较，其名次排序结果如表 47-7。

表 47-7 不同综合评价方法估计综合指数及其排序

动力类型	概率加法模型		VIKOR		TOPSIS(1)		TOPSIS(2)	
	秩次	综合指数	秩次	综合指数	秩次	综合指数	秩次	综合指数
可更换电池	1	0.7090	2	0.1720	1	0.9450	1	0.9750
直接充电	2	0.7054	4	0.2530	3	0.9310	2	0.9670
电动汽车	3	0.6923	3	0.2240	2	0.9330	3	0.9640
电动 CNG	4	0.6764	8	0.5100	5	0.7000	4	0.8890
电动 LPG	5	0.6764	9	0.5100	6	0.7000	5	0.8890
汽油混合动力	6	0.6703	1	0.1680	4	0.7490	9	0.7560
柴油混合动力	7	0.6689	10	0.8060	12	0.3010	12	0.0970
液化气(LPG)	8	0.6484	6	0.4790	11	0.3450	8	0.8300
压缩天然气(CNG)	9	0.6474	7	0.4800	10	0.3990	7	0.8300
燃料电池(氢)	10	0.6236	12	0.9250	8	0.5630	6	0.8650
常规柴油机	11	0.6028	5	0.2810	7	0.7000	11	0.4880
甲醇	12	0.6013	11	0.8520	9	0.5270	10	0.6980

从表 47-7 可以看出，加法概率模型和 TOPSIS（2）法的结果，排名前 5 的名次顺序相同，动力类型依次为，可更换电池、直接充电、电动汽车、电动 CNG 和电动 LPG；而与 VIKOR 法估计出来的排名顺序差异较大。

在概率加法模型中，我们充分地考虑的各个因子交互作用对综合指数的影响。因此这样建立起来的综合评价模型，以及综合指数理论值包含了各个因子的交互作用。为探讨各个因子互作项对综合指数的影响，我们可以选中 DPS 电子表格中各个因子转换值以及综合指数，点击按钮“  ”计算相关系数。然后选定输出结果的相关系数矩阵，点击通径分析按钮“  ”，系统根据相关系数矩阵进行通径分析，生成通径分析图（图 47-8）。

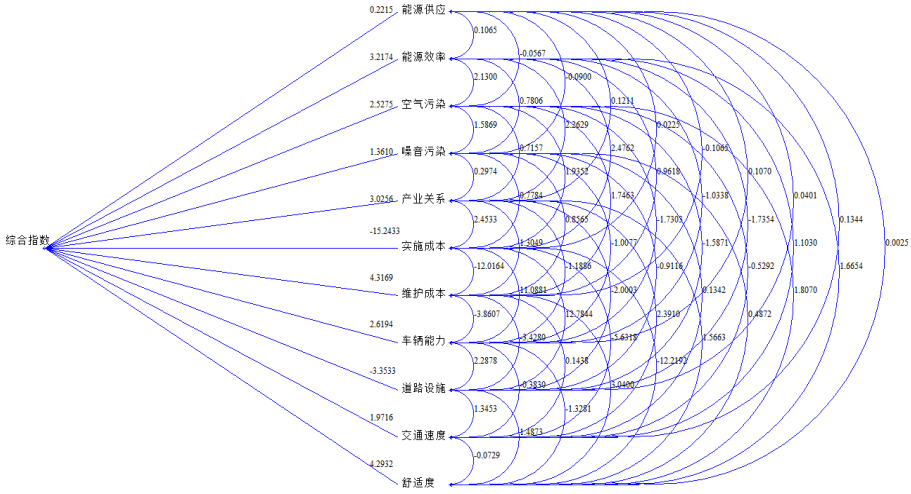


图 47-8　各因子指标和综合指数的通径分析图

图 47-8 直观地显示了各个因子指标和综合指数的关系，便于我们识别各个因子。及其交互作用对综合指数的影响。

47.6 乘积型综合评价模型

47.6.1 乘积型综合指数模型

乘积型综合评价指数模型理论上也是一种常用的多指标综合评价方法。该模型将各指标之间的关系通过乘积的方式结合起来，以得出最终的综合评价结果。

和加型综合指数模型相比，乘积型多因子综合评价指数模型更考虑系统诸因子中少数因子的短板效应。短板效应意味着即使其他大部分因子表现优异，但系统的整体（指数）表现也会受到最薄弱的因子的制约，即排名名称高低因某个因子而一票否决。因此强调系统多因子平衡发展时可考虑使用乘积型多因子综合评价指数模型。

当综合评价指数模型是乘积型形式时，那么式(47.1)加法型模型表达方式的相应乘积型形式为：

$$y_i = x_{i1}^{w_1} \cdot x_{i2}^{w_2} \cdot x_{i3}^{w_3} \cdot \cdots \cdot x_{ip}^{w_p} = \prod_{j=1}^{p} x_{ij}^{w_j} \text{ 或 } y_i = e^{x_{i1}w_1 + x_{i2}w_2 + \cdots + x_{ip}w_p} \tag{47.24}$$

式中 $w_1 + w_2 + w_3 + \cdots + w_p = 1$。

为求解权重系数，我们可以对式（47.24）进行取对数操作，转换成加法型综合评价指数模型形式（47.25）：

$$\ln(y_i)=\ln(x_{i1})\cdot w_1+\ln(x_{i2})\cdot w_2+\ln(x_{i3})\cdot w_3+\cdots+\ln(x_{ip})\cdot w_p \tag{47.25}$$

记 $y_i'=\ln(y_i), x_{ij}'=\ln(x_{ij})$，式(47.25)可以写成

$$y_i'=x_{i1}'\cdot w_1+x_{i2}'\cdot w_2+x_{i3}'\cdot w_3+\cdots+x_{ip}'\cdot w_p$$

同理，对数转换后的理论组分也可用理论组分之和的形式来表示：

$$y_i'=C_{i1}'+C_{i2}'+C_{i3}'+\cdots+C_{ip}'$$

式中，$C_{ij}'=y_i'\ w_j$，这时乘积型模型的残差 σ' 为：

$$\begin{aligned}\sigma'&=\sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{p}\left(C_{ij}'-x_{ij}'w_j\right)^2}{(n-1)\cdot p}}\\&=\sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{p}\left(y_i'w_j-x_{ij}'w_j\right)^2}{(n-1)\cdot p}}\\&=\sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{p}\left[\left(y_i'-x_{ij}'\right)w_j\right]^2}{(n-1)\cdot p}}\end{aligned} \tag{47.26}$$

显然，经对数转换后的各个因子的取值不再在 0-1 区间，数值大小不等，均值和方差的大小因样本不同而异。拟合残差理论上不一定处在 0-1 区间。因此不能应用式(47.7)，采用最大熵-最小二乘方法进行数值优化。

为将拟合模型的方差控制在 0-1 区间，一个简单的方法是除以样本经对数转换后的标准差（记为σ_T）

$$\sigma_T=\sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{p}\left(\ln x_{ij}-\bar{X}\right)^2}{n\cdot p-1}},\quad \bar{X}=\frac{\sum_{i=1}^{n}\sum_{j=1}^{p}\ln x_{ij}}{n\cdot p}$$

将拟合过程中的残差都除以样本经对数转换后的标准差σ_T

$$\sigma_{\mathrm{M}}=\frac{\sigma'}{\sigma_T} \tag{47.27}$$

这样亦相当于对样本的离散程度进行标准正态分布转换，即将样本方差转换到 1。这时，我们可应用式(47.7)进行乘积型综合评价模型的最大熵-最小二乘估计，这时综合评价模型的极小化误差函数为：

$$\min f(\sigma_M,E)=\text{minimize}\quad \sigma_M^E \tag{47.28}$$

乘积型综合评价模型表达为：

$$\begin{cases} \min \quad \sigma_M^E \\ \text{s.t} \sum_{j=1}^{p} w_j = 1, \quad w_j \geq 0 \end{cases} \tag{47.29}$$

和加法型模型(式47.7)一样，带约束条件非线性回归方程模型式(47.29)的优化，亦可采用序列无约束极小化(SUMT)方法进行综合评价指数模型的优化求解。作者将乘积型综合评价指数模型的统计分析收录在作者主持开发的DPS数据处理系统之中。

乘积型多因子综合评价指数模型目前应用较少，主要原因可能是因为系统中各个因子的权重系数难以客观估计。联合国开发计划署每年均出版了含有人类发展指数的人类发展报告[8,20]。人类发展指数采用了Health index，Education index和Income index这3个指标，以等权重的方式、几何平均计算方法计算，这种方法即为等权重的乘积型的多因子综合评价。

47.6.2 乘积型综合指数模型应用示例

这里仍以我国 2007 年各地区农村居民家庭平均每人生活消费支出水平进行综合指数分析，将原始数据按一行一个地区、一列一个指标方式编辑如图 47-1 所示形式。

然后执行“专业统计”—“综合评价与多属性决策”—“最大熵-最小二乘模型”功能，这时系统出现供用户对各个指标特性进行设置的对话框(图 47-9 左)。

图 47-9　乘积型综合指数模型分析用户界面

这里我们应用乘积型方式汇集各个因子，在数据汇总方式选择框中选择“乘积型”即可。因不需要进行数据转换，故在数据转换下面的选择框中选择“不转换”。数据标准化下面的多选框中，因这里是乘积型模型，一般选择标准化正态概率来实施规格化。

在选择框下面，系统给出了每个指标是否是低优指标、以及各个指标的平均值、最小值、最大值。

综合评价指数计算过程中，许多综合评价指数组成因子是用不同的单位来衡量的。数值越高，表现越好。但是也有例外情形，如感染率。在计算权重系数之前，这里可根据每个指标的特性，可设置某些指标为“低优指标”。

如需要进行各个权重系数置信区间的估计，可以在下面的 Bootstrap 抽样次数输入框中输入抽样次数，缺省情形下是“0”，表示不进行 Bootstrap 抽样估计。注意，进行 Bootstrap 抽样时需要耗费的时间会较长，特别是因子指标比较多的时候。

原始数据安装我们设置的方式进行预处理，然后用最大熵-最小残差模型公式(47.29)对预处理之后的数据，经 SUMT 算法优化计算，给出如下结果。

乘积型综合评价指数模型

标准化-正态概率处理

因子	是否低优	数据转换	均值	标准差	最小值(a)	最大值(b)
食品	否	不转换	1424.9461	514.4770	939.0300	3259.4800
衣着	否	不转换	213.4490	101.1272	86.2600	513.4400
居住	否	不转换	601.6335	367.0977	263.7300	2097.2100
家电	否	不转换	155.3713	82.1620	70.9300	451.4000
交通和通讯	否	不转换	347.9865	179.6311	154.5200	883.7100
文娱	否	不转换	323.1948	200.4646	65.3900	870.1200
医疗保健	否	不转换	231.2929	124.3624	50	629.5600
其他	否	不转换	78.4332	43.2870	29.3600	249.0400

熵值=2.947230

残差=0.034385

目标函数对数值=-8.1373

模型拟合方差分析表

模型残差	平方和	自由度	均方
等权重模型	0.3567	247	0.0014
MEMR 模型	0.2838	240	0.0012
模型残差差值	0.0729	7	0.0104

F 统计量=8.8121　　显著性 p 值=0.0001

优化指数=20.4467%

权重系数 Bootstrap 抽样 1000 次估计,时间 0 分 6 秒。

因子	权重系数	抽样均值	标准差	中位数	95%置信区间	
食品	0.0945	0.0937	0.0139	0.0924	0.0557	0.1441
衣着	0.0817	0.0835	0.0122	0.0818	0.0564	0.1484
居住	0.1321	0.1332	0.0170	0.1329	0.0843	0.1847
家电	0.1428	0.1425	0.0165	0.1422	0.0818	0.2192
交通和通讯	0.1995	0.1975	0.0125	0.1983	0.1372	0.2285

文娱	0.1274	0.1270	0.0157	0.1263	0.0793	0.1928
医疗保健	0.0958	0.0960	0.0125	0.0954	0.0598	0.1510
其他	0.1263	0.1266	0.0111	0.1264	0.0870	0.1677
变异系数	11.77					
熵值	2.9472	2.9399	0.0134	2.9407	2.8797	2.9710
			正态近似置信区间		2.9137	2.9662

样本转换值及综合指数

样本	综合指数	样本	综合指数	样本	综合指数
北　京	0.9403	安　徽	0.3473	四　川	0.3092
天　津	0.5171	福　建	0.6181	贵　州	0.1603
河　北	0.3788	江　西	0.3523	云　南	0.2638
山　西	0.3608	山　东	0.5787	西　藏	0.2225
内蒙古	0.4823	河　南	0.3490	陕　西	0.3240
辽　宁	0.5372	湖　北	0.4258	甘　肃	0.1956
吉　林	0.4553	湖　南	0.4435	青　海	0.2996
黑龙江	0.4414	广　东	0.5767	宁　夏	0.3273
上　海	0.9985	广　西	0.2766	新　疆	0.2693
江　苏	0.7920	海　南	0.2578		
浙　江	0.9761	重　庆	0.2594		

系统首先输出最大熵-最小二乘非线性含条件约束模型的信息熵 E 等于 2.9472。拟合残差标准差(σ=0.034385)以及模型目标函数自然对数值为-8.1373。和加法模型比较，信息熵较小，说明各个权重系数之间差异比加法模型大。但要注意，模型拟合残差和加法模型没有可比性。然后输出模型的方差分析结果。

检验的原假设是各个因子权重系数等于 1/p。这部分结果解读要点是从统计学角度，检验各个指标的权重系数的大小是否有差异。如果没有差异，那我们可采用等权重的方式进行综合指数计算。如果这里的显著性概率 p 值小于 0.05，那我们就应该采用模型给出的各个指标的权重系数来计算综合评价指数。本例中显著性检验 p<0.0001，小于 0.05，因此这里的数据不宜采用等权法估计综合评价指数，而应采用模型估计值进行综合指数的计算。从模型的优化指数可以看出，乘积型模型的优化程度要高，各个因子的权重系数差异要大，这和信息熵的大小也可以看出。

输出结果第二部分为各个指标的权重系数，是综合评价分析模型的核心和重点。它既是各个因子对形成综合指数的贡献大小，又是综合评价指数中各个因子的权重重要性的体现，是各个因子和综合指数相互之间，从集中趋势、离散程度及多指标之间的相互关系诸方面在最大熵条件下、拟合误差达到最小的结果。

这里的第一列是各个指标权重系数的估计值。它反映了各个因子在综合评价指数中的重要性，权重系数越大，该指标在综合评价中越是重要。本例中，根据权重系数来看，交通和通讯、家庭设备及服务两个指标对综合指数（综合评价）的影响较大；其次是其他商品及服务的影响较小。

“抽样均值”和“标准差”两栏是根据 Bootstrap 抽样（这里是 1000 次）结果计算出的估计值，这里的标准差即为权重系数在正态假设下的标准误。根据标准误可以估计其 95%

的置信区间。权重系数 95%置信区间均包含 1/p（0.125），故可认为权重系数在统计学意义下不显著。

最后系统给出了各样本最大熵-最小残差模型综合评价指数。将评价指数从大到小进行排序，可得到各个样本排序结果，根据排序结果进行专业意义上的解读和评价对象的位次比较。如本例中，上海、浙江和北京排名为前三，消费水平最高；其次是江苏、福建、广东、山东等，消费水平次之。这和应用相加型模型结果某些地区的排名有些差异。

47.7 倒数型综合评价指数模型

47.7.1 倒数型综合指数模型

倒数型综合评价指数模型，系将各指标之间的关系通过和的倒数方式和综合指数结合起来，以得出最终的综合评价结果。

当综合评价指数模型是倒数型形式时，那么式(47.1)加法型模型表达方式的相应倒数型形式为：

$$y_i = 1\Big/\left(x_{i1}w_1 + x_{i2}w_2 + x_{i3}w_3 + \cdots + x_{ip}w_p\right) \tag{47.30}$$

式中 $w_j \geq 0, w_1 + w_2 + w_3 + \cdots + w_p = 1$。

为求解权重系数，须对式（47.30）中两边取倒数转换。转换成加法型的线性模型：

$$1/y_i = x_{i1}w_1 + x_{i2}w_2 + x_{i3}w_3 + \cdots + x_{ip}w_p$$

记 $y_i' = 1/y_i$，因这里 xij 是隐性因变量，同样进行取倒数转换，$x_i' = 1/x_i$，这时转换后的公式为

$$y_i' = x_{i1}' \cdot w_1 + x_{i2}' \cdot w_2 + x_{i3}' \cdot w_3 + \cdots + x_{ip}' \cdot w_p$$

同理，对数转换后的理论组分也可用理论组分之和的形式来表示：

$$y_i' = C_{i1}' + C_{i2}' + C_{i3}' + \cdots + C_{ip}'$$

式中，$C_{ij}' = y_i'\ w_j$，这时乘积型模型的残差 σ' 为：

$$\sigma' = \sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{p}\left(C_{ij}' - x_{ij}'w_j\right)^2}{(n-1)\cdot p}} = \sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{p}\left(y_i'w_j - x_{ij}'w_j\right)^2}{(n-1)\cdot p}} = \sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{p}\left[\left(y_i' - x_{ij}'\right)w_j\right]^2}{(n-1)\cdot p}}$$

显然，经对倒数转换后的各个因子的取值不再在 0-1 区间，数值大小不等，均值和方差的大小因样本不同而异。拟合残差理论上不一定处在0-1区间。因此不能应用式(47.7)，采用最大熵-最小二乘方法进行数值优化。

为将拟合模型的方差控制在 0-1 区间，一个简单的方法是和乘积型综合评价模型一样，除以样本经倒数转换后的标准差（记为σ_T）

$$\sigma_T=\sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{p}\left(\frac{1}{x_{ij}}-\bar{X}\right)^2}{n\cdot p-1}},\quad \bar{X}=\frac{\sum_{i=1}^{n}\sum_{j=1}^{p}\frac{1}{x_{ij}}}{n\cdot p}$$

将拟合过程中的残差都除以样本经倒数转换后的标准差σ_T

$$\sigma_{\mathrm{R}}=\sigma'/\sigma_T \tag{47.31}$$

这样亦相当于对样本的离散程度进行标准正态分布转换，即将样本方差转换到 1。这时，我们可应用式(47.32)进行倒数型综合评价模型的最大熵-最小二乘估计，这时综合评价模型的极小化误差函数为：

$$\min f(\sigma_M,E)=\text{minimize}\quad \sigma_M^E \tag{47.32}$$

倒数型综合评价模型表达为：

$$\begin{cases}\min\quad \sigma_M^E\\ \text{s.t}\sum_{j=1}^{p}w_j=1,\quad w_j\geq 0\end{cases} \tag{47.33}$$

和加法型模型(式47.7)一样，带约束条件非线性回归方程模型式(47.33)的优化，亦可采用序列无约束极小化(SUMT)方法进行综合评价指数模型的优化求解。作者将乘积型综合评价指数模型的统计分析收录在作者主持开发的DPS数据处理系统之中。

47.7.2 倒数平方型综合指数模型

对倒数型综合评价指数模型进行扩展，采用倒数的平方形式将各指标和综合评价指数结合起来，以得出最终的综合评价结果，形式为：

$$y_i=\frac{1}{(x_{i1}w_1+x_{i2}w_2+x_{i3}w_3+\cdots+x_{ip}w_p)^2} \tag{47.34}$$

式中 $w_j\geq 0, w_1+w_2+w_3+\cdots+w_p=1$。

为求解权重系数，我们须对式（47.34）中两边取倒数、然后取平方根转换。转换成加法型的线性模型：这时的转换方式为

$$1/\sqrt{y_i} = x_{i1}w_1 + x_{i2}w_2 + x_{i3}w_3 + \cdots + x_{ip}w_p \tag{47.35}$$

和倒数型多因子综合评价指数模型一样推导可得到应用最大熵最小二乘求解的综合指数模型。

47.7.3 倒数型综合指数模型应用示例

这里仍以我国 2007 年各地区农村居民家庭平均每人生活消费支出水平进行综合指数分析，将原始数据按一行一个地区、一列一个指标方式编辑如图 47-1 所示形式。

然后执行“专业统计”—“综合评价与多属性决策”—“最大熵-最小二乘模型”功能，这时系统出现供用户对各个指标特性进行设置的对话框(图 47-9 左)。

这里我们在综合评价模型里面选“倒数型模型”，数据格式化检验选“标准化-正态概率”，因线性转换有的值很小、取倒数后的值会很大，形成极端异常值。其他选项用默认值。经最大熵-最小二乘算法优化，给出各个指标权重如，食品=0.1112、衣着=0.0669、居住=0.1445、家电=0.1625、交通和通讯=0.2078、文娱=0.1082、医疗保健=0.0634 及其他=0.1355。这一结果和前面的乘积型模型中各个因子的权重系数相近。

最后系统给出了各样本最大熵-最小残差模型综合评价指数。将评价指数从大到小进行排序，可得到各个样本排序结果，根据排序结果进行专业意义上的解读和评价对象的位次比较。如本例中，上海、浙江和北京排名为前三，消费水平最高；其次是江苏、福建、广东、山东等，消费水平次之。这和应用乘积模型的排名相同。

47.8 Logistic 和 Probit 等其它综合评价模型

前面加法模型、概率加法模型都是对 0-1 规格化数据直接进行建模。鉴于 0-1 区间数据更多地是表达比率，若参与评价的因子是二项分布，即各个因子值都是由“好”、“坏”，“通过”、“不通过”等属于二项分布变量构成的时候，理论上对这些比率值的建模，须采用类似广义线性模型中对于比率值进行线性化处理，建立线性模型。

因子为比率值时，Logistic 模型和 Probit 模型是用的最多的两个表达二项分布的统计模型。这里在介绍这两个模型后，我们简单地介绍一下其他 3 个表达比率值的统计模型：重对数变换、对数互补模型和重对数互补模型。

47.8.1 Logistic 型综合指数模型

当综合评价指数模型是 Logistic 型形式时，那么综合评价模型表达方式的为

$$y_i = \frac{1}{1+e^{-\left(x_{i1}w_1 + x_{i2}w_2 + \cdots + x_{ip}w_p\right)}} \tag{47.36}$$

式中 $w_1 + w_2 + w_3 + \cdots + w_p = 1$。

Logistic 回归模型因变量可以有多种不同的表达形式，为求解权重系数，可将等式表达为概率的函数与自变量之间的线性表达式

$$\ln\left[y_i/(1-y_i)\right]=\sum w_i x_i \tag{47.37}$$

因为我们这里等号右边的变量 x_{ij} 实际上也是隐性因变量，因此在综合评价指数模型中也需做 Logit 转换，转换后的 Logit 综合评价指数模型形式（47.38）：

$$\ln(\frac{1}{1-y_i})=\ln(\frac{1}{1-x_{i1}})\cdot w_1+\ln(\frac{1}{1-x_{i2}})\cdot w_2+\cdots+\ln(\frac{1}{1-x_{i3}})\cdot w_p \tag{47.38}$$

记 $y_i'=\ln(\frac{1}{1-y_i}),x_{ij}'=\ln(\frac{1}{1-x_{ij}})$，式(47.38)可以写成

$$y_i'=x_{i1}'\cdot w_1+x_{i2}'\cdot w_2+x_{i3}'\cdot w_3+\cdots+x_{ip}'\cdot w_p$$

同理，Logit 转换后的理论组分也可用理论组分之和的形式来表示：

$$y_i'=C_{i1}'+C_{i2}'+C_{i3}'+\cdots+C_{ip}'$$

式中，$C_{ij}'=y_i'\ w_j$，这时 Logit 模型的残差 σ' 为：

$$\sigma'=\sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{p}\left(C_{ij}'-x_{ij}'w_j\right)^2}{(n-1)\cdot p}}=\sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{p}\left[\left(y_i'-x_{ij}'\right)w_j\right]^2}{(n-1)\cdot p}} \tag{47.39}$$

和乘积型综合评价分析模型一样，经 Logit 转换后的各个因子的取值不再在 0-1 区间，且标准 Logistic 分布的方差是$\pi^2/3$，比标准正态分布方差的 1 大。为将拟合模型的方差控制在 0-1 区间，一个简单的方法是除以样本经对数转换后的标准差（记为σ_T）

$$\sigma_T=\sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{p}\left(\ln\frac{1}{1-x_{ij}}-\bar{X}\right)^2}{n\cdot p-1}},\quad \bar{X}=\frac{\sum_{i=1}^{n}\sum_{j=1}^{p}\ln\frac{1}{1-x_{ij}}}{n\cdot p}$$

将拟合过程中的残差都除以样本经对数转换后的标准差σ_T

$$\sigma_{\text{Logit}}=\sigma'/\sigma_T \tag{47.40}$$

这样亦相当于对样本的离散程度进行标准正态分布转换，即将样本方差转换到 1。这时，我们可应用式(47.7)进行乘积型综合评价模型的最大熵-最小二乘估计。这时综合评价模型的极小化误差函数为：

$$\min f(\sigma_{Logit},E)=\text{minimize}\quad \sigma_{Logit}^{E} \tag{47.41}$$

Logit 型综合评价模型表达为：

$$\begin{cases} \min \quad \sigma_{Logit}^{E} \\ \text{s.t} \sum_{j=1}^{p} w_j = 1, \quad w_j \geq 0 \end{cases} \tag{47.42}$$

47.8.2 Probit 型综合指数模型

当综合评价指数模型是 Probit 模型形式时，那么综合评价模型表达方式的为

$$\begin{aligned} P(y_i = 1) &= g(x_{i1}w_1 + x_{i2}w_2 + \cdots x_{ip}w_p) \\ &= \int_{-\infty}^{x_{i1}w_1 + x_{i2}w_2 + \cdots x_{ip}w_p} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \mathrm{d}u \end{aligned} \tag{47.43}$$

式中 $w_1 + w_2 + w_3 + \cdots + w_p = 1$。

Probit 回归模型线性化转换形式为 $\Phi^{-1}(u)$。因为我们这里等号右边的变量 x_{ij} 实际上也是隐性因变量，因此在综合评价指数模型中也需做逆正态 $\Phi^{-1}(u)$ 转换，转换后的 Probit 综合评价指数模型形式（47.44）：

$$\Phi^{-1}(y_i) = \Phi^{-1}(x_{i1}) \cdot w_1 + \Phi^{-1}(x_{i2}) \cdot w_2 + \cdots + \Phi^{-1}(x_{ip}) \cdot w_p \tag{47.44}$$

记 $y_i' = \Phi^{-1}(y_i), x_{ij}' = \Phi^{-1}(x_{ij})$，式(47.30)可以写成

$$y_i' = x_{i1}' \cdot w_1 + x_{i2}' \cdot w_2 + x_{i3}' \cdot w_3 + \cdots + x_{ip}' \cdot w_p$$

同理，Probit 转换后的理论组分也可用理论组分之和的形式来表示：

$$y_i' = C_{i1}' + C_{i2}' + C_{i3}' + \cdots + C_{ip}'$$

式中，$C_{ij}' = y_i' \ w_j$，这时 Probit 模型的残差 σ' 为：

$$\sigma' = \sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{p}\left(C_{ij}' - x_{ij}' w_j\right)^2}{(n-1)\cdot p}} = \sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{p}\left[\left(y_i' - x_{ij}'\right)w_j\right]^2}{(n-1)\cdot p}} \tag{47.45}$$

和乘积型综合评价分析模型一样，经 Probit-转换后的各个因子的取值不再在 0-1 区间，且标准 Logistic 分布的方差是$\pi^2/3$，比标准正态分布方差的 1 大。为将拟合模型的方差控制在 0-1 区间，一个简单的方法是除以样本经对数转换后的标准差（记为σ_T）

$$\sigma_T = \sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{p}\left(\Phi^{-1}(x_{ij}) - \bar{X}\right)^2}{n \cdot p - 1}}, \quad \bar{X} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{p}\ln \Phi^{-1}(x_{ij})}{n \cdot p}$$

将拟合过程中的残差都除以样本经对数转换后的标准差σ_T

$$\sigma_{\text{Probit}} = \sigma' / \sigma_T \tag{47.46}$$

这样亦相当于对样本的离散程度进行标准正态分布转换，即将样本方差转换到 1。这时，我们可应用式(47.7)进行乘积型综合评价模型的最大熵-最小二乘估计。这时综合评价模型的极小化误差函数为：

$$\min f(\sigma_{Probit}, E) = \text{minimize} \quad \sigma_{Probit}^{E} \tag{47.47}$$

Probit 型综合评价模型表达为：

$$\begin{cases} \min \quad \sigma_{Probit}^{E} \\ \text{s.t} \sum_{j=1}^{p} w_j = 1, \quad w_j \geq 0 \end{cases} \tag{47.48}$$

47.8.3 其它综合评价分析模型

DPS 提供的其他综合评价分析模型有：

①重对数模型

$$y_i = \exp(-\exp(-(x_{i1}w_1 + x_{i2}w_2 + \cdots + x_{ip}w_p))) \tag{47.49}$$

②对数互补模型

$$y_i = 1 - \exp(x_{i1}w_1 + x_{i2}w_2 + \cdots + x_{ip}w_p) \tag{47.50}$$

③重对数互补模型

$$y_i = 1 - \exp(-\exp(x_{i1}w_1 + x_{i2}w_2 + \cdots + x_{ip}w_p)) \tag{47.51}$$

式中 $w_1 + w_2 + w_3 + \cdots + w_p = 1$。

对于重对数模型，线性化转换公式为 $y_i' = -\ln(-\ln(y_i))$，$x_{ij}' = -\ln(-\ln(x_{ij}))$；对数互补模型 $y_i' = -\ln(1 - y_i)$，$x_{ij}' = \ln(1 - x_{ij})$；重对数互补模型 $y_i' = \ln(-\ln(1 - y_i))$，$x_{ij}' = \ln(-\ln(1 - x_{ij}))$。然后计算各因子转换后的标准差$\sigma_T$。以标准差$\sigma_T$规范化拟合残差 σ'。

$$\sigma' = \sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{p}\left(C_{ij}' - x_{ij}'w_j\right)^2}{(n-1)\cdot p}} = \sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{p}\left[\left(y_i' - x_{ij}'\right)w_j\right]^2}{(n-1)\cdot p}} \tag{47.52}$$

式中 $C_{ij}' = y_i'\ w_j$，$y_i' = C_{i1}' + C_{i2}' + C_{i3}' + \cdots + C_{ip}'$。拟合过程中残差都除以样本经对数转换后的标准差$\sigma_T$

$$\sigma_a = \sigma'/\sigma_T \tag{47.53}$$

这样亦相当于对样本的离散程度进行标准正态分布转换，即将样本方差转换到 1。这时，我们可应用式(47.7)进行乘积型综合评价模型的最大熵-最小二乘估计。这时综合评价模型的极小化误差函数为

$$\min f(\sigma_a, E) = \text{minimize} \quad \sigma_a^E$$

含约束条件的非线性综合评价模型表达式为：

$$\begin{cases} \min \quad \sigma_a^E \\ \text{s.t} \sum_{j=1}^{p} w_j = 1, \quad w_j \geq 0 \end{cases} \tag{47.54}$$

47.5.3 操作示例

这里以Logistic模型为例，带约束条件非线性回归方程模型式(47.42)的优化，亦可采用序列无约束极小化(SUMT)方法进行综合评价指数模型的优化求解。作者这里以烟叶感官质量评价为例，应用Logistic综合评价指数模型进行统计分析。

	A	B	C	D	E	F	G	H
1	样品号	香气质	香气量	杂气	刺激	余味	劲头	浓度
2	1	8	6	8	8	8	4	6
3	2	6	6	6	6	6	4	4
4	3	6	6	6	6	6	4	4
5	4	6	6	6	4	6	4	4
6	5	8	6	8	8	8	4	6
7	6	6	6	6	6	6	4	6
8	7	6	6	6	6	6	4	6
9	8	6	6	4	6	6	4	6
10	9	8	8	8	6	8	6	6
11	10	8	8	6	6	6	6	6
12	11	8	8	8	6	6	6	6
13	12	8	8	6	6	6	6	6
14	13	8	8	8	6	6	8	8
15	14	8	8	6	6	6	6	6
16	15	6	6	6	6	6	6	6
17	16	6	6	6	6	6	6	6
18	17	6	6	6	6	6	6	6
19	18	6	8	8	6	6	8	8
20	19	6	8	6	6	6	8	8
21	20	6	6	6	6	6	8	8
22	21	6	6	6	6	6	8	8
23	22	6	6	4	6	4	8	8
24	23	4	6	4	6	4	8	8
25	24	4	6	4	4	4	8	8
26	25	6	6	4	4	4	4	6
27	26	6	6	6	8	6	4	6
28	27	6	6	6	6	6	6	6
29								

图 47-10 烟叶感官质量指标 Logit 综合评价数据格式

这里以胡建军等（2001）关于烟叶感官质量研究中 7 个因子，即香气质、香气量、

杂气、刺激、余味、劲头和浓度，由专家以 1~10 级评分数据进行综合指数分析，将原始数据按一行一个烟叶样本、一列一个感官质量指标方式编辑如图 47-10 所示形式。然后执行“专业统计”—“综合评价与多属性决策”—“最大熵-最小二乘模型”功能，这时系统出现供用户对各个指标特性进行设置的对话框(图 47-11 左)。

图 47-11　Logit 综合评价模型用户界面

这里我们应用乘积型方式汇集各个因子，在数据汇总模型选择框中选择“Logit 型”即可。因不需要进行数据转换，故在数据转换下面的选择框中选择“不转换”。数据标准化下面的多选框中，这里可选择默认值，即 *a-b* 线性规格化。这里的各个因子评分是按 1~10,分 10 个等级进行的，所以这里的最小值取 1，最大值取 10。

如需要进行各个权重系数置信区间的估计，可以在下面的 Bootstrap 抽样次数输入框中输入抽样次数，缺省情形下是“0”，表示不进行 Bootstrap 抽样估计。一般这里取 1000。设置完成后，点击”计算”，系统即执行 SUMT 算法优化计算，给出如下结果。

因子	是否低优	数据转换	均值	标准差	最小值(a)	最大值(b)
香气质	否	不转换	6.4444	1.1547	1	10
香气量	否	不转换	6.5926	0.9306	1	10
杂气	否	不转换	6.0741	1.2987	1	10
刺激	否	不转换	6	0.9608	1	10
余味	否	不转换	5.9259	1.0350	1	10
劲头	否	不转换	5.8519	1.6572	1	10
浓度	否	不转换	6.3704	1.2449	1	10

模型：Logit 模型

数据规格化方法：a-b 线性(规格)化

熵值=2.716438

残差=0.061468

目标函数对数值=-6.1770

模型拟合方差分析表

模型残差	平方和	自由度	均方
等权重模型	0.9566	188	0.0051
MEMR 模型	0.6877	182	0.0038
模型残差差值	0.2689	6	0.0448

F 统计量=11.8631　　显著性 p 值=0.0001

优化指数=28.1141%

权重系数 Bootstrap 抽样 1000 次估计,时间 0 分 7 秒。

因子	权重系数	抽样均值	标准差	中位数	95%置信区间	
香气质	0.1751	0.1749	0.0232	0.1754	0.0821	0.2650
香气量	0.1443	0.1402	0.0245	0.1403	0.0712	0.2138
杂气	0.1843	0.1860	0.0220	0.1849	0.0986	0.2842
刺激	0.1637	0.1646	0.0266	0.1629	0.0952	0.2616
余味	0.1902	0.1935	0.0233	0.1921	0.1256	0.2842
劲头	0.0613	0.0603	0.0131	0.0598	0.0230	0.1075
浓度	0.0810	0.0804	0.0178	0.0797	0.0317	0.1629
变异系数	16.3711					
熵值	2.7165	2.6953	0.0524	2.7049	2.4262	2.7910
			正态近似置信区间		2.5925	2.7981

样本转换值及综合指数

样本	香气质	香气量	杂气	刺激	余味	劲头	浓度	综合指数
S1	0.7778	0.5556	0.7778	0.7778	0.7778	0.3333	0.5556	0.7112
S2	0.5556	0.5556	0.5556	0.5556	0.5556	0.3333	0.3333	0.5232
S3	0.5556	0.5556	0.5556	0.5556	0.5556	0.3333	0.3333	0.5232
S4	0.5556	0.5556	0.5556	0.3333	0.5556	0.3333	0.3333	0.4857
…	…	…	…	…	…	…	…	…
S27	0.5556	0.5556	0.5556	0.5556	0.5556	0.5556	0.5556	0.5556

系统首先输出最大熵-最小二乘非线性含条件约束模型的信息熵 E 等于 2.71644。拟合残差标准差(σ=0.061468)以及模型目标函数自然对数值为-6.1770。

然后输出模型的方差分析结果。检验的原假设是各个因子权重系数等于 $1/p$。这里的显著性检验 p=0.0001，小于 0.05，因此表明这里的数据不宜采用等权法估计综合评价指数，而应采用 Logit 模型进行综合指数的计算。从模型的优化指数可以看出，乘积型模型的优化程度要高，各个因子的权重系数差异要大，这和信息熵的大小也可以看出。

输出结果第二部分为各个指标的权重系数，是综合评价分析模型的核心和重点。它既是各个因子对形成综合指数的贡献大小，又是综合评价指数中各个因子的权重重要性的体现，是各个因子和综合指数相互之间，从集中趋势、离散程度及多指标之间的相互关系诸方面在最大熵条件下、拟合误差达到最小的结果。

这里的第一列是各个指标权重系数的估计值。它反映了各个因子在综合评价指数中的重要性，权重系数越大，该指标在综合评价中越是重要。本例中，根据权重系数来看，杂气、余味和香气质这 3 个指标对综合指数（综合评价）的影响较大；其次是刺激和香气量，其他因子的影响较小。

“抽样均值”和“标准差”两栏是根据 Bootstrap 抽样（这里是 1000 次）结果计算出的估计值，这里的标准差即为权重系数在正态假设下的标准误。根据标准误可以估计其 95%的置信区间。权重系数 95%置信区间均包含 1/p（0.125），故可认为权重系数在统计学意义下不显著。

最后系统给出了各样本最大熵-最小残差 Logit 模型综合评价指数。将评价指数从大到小进行排序，我们就可得到各个样本排序结果，这里不再赘述。

47.5.4 几个模型结果比较

对烟草评价数据，我们用几个模型分别进行综合评价，数据规格化均选用标准化-正态概率法。得到的权重系数估计见表47-8。

表47-8 各种综合分析模型的权重系数

因子	Logistic	Probit	重对数模型	对数互补模型	重对数互补模型
香气质	0.1856	0.1852	0.1920	0.1928	0.1744
香气量	0.1386	0.1337	0.1257	0.1221	0.1574
杂气	0.2134	0.2098	0.2180	0.2214	0.2019
刺激	0.1268	0.1345	0.1242	0.1149	0.1293
余味	0.1709	0.1773	0.1677	0.1573	0.1700
劲头	0.0836	0.0800	0.0858	0.0958	0.0874
浓度	0.0811	0.0795	0.0867	0.0957	0.0797

从表47-8可以看出，各个模型的权重系数大小趋势一致：权重系数最大的因子是杂气、接下来的是香气质和余味。权重系数最小的两个因子是浓度、劲头。不同模型之间的权重系数不大。

对各个综合评价模型计算出来的综合指数进行因子（R型）聚类分析，得到聚类图如47-12。

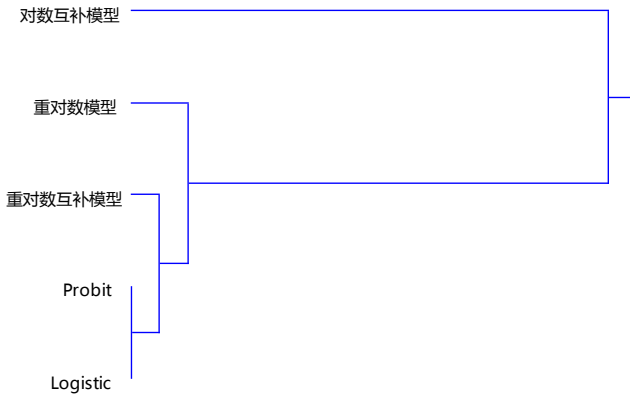


图 47-12 不同模型之间综合指数的 R 型聚类图

从图 47-12 可以看出，Logistic 模型和 Probit 模型得到的综合评价指数结果最接近。秩次排名，仅有一处不一致。其次是重对数互补模型、重对数模型，其综合评价指数跟 Logistic 模型和 Probit 模型的相似程度也较高。只有对数互补模型的综合评价指数较其他模型的综合评价指数差异大一些。各个综合评价模型综合指数排名结果如表 47-9。

从表 47-9 可以看出，各个模型的名次排名大体上是一致的。只是对数互补模型排名和其他综合评价模型的差异较大，排名第一的样本和其它模型不一样。

表 47-9　不同综合评价模型的综合指数及其名称排序结果

样本	Logistic	名次	Probit	名次	重对数	名次	对数互补	名次	重对数互补	名次
S9	0.8780	1	0.8647	1	0.8840	1	0.8878	3	0.8409	1
S1	0.8574	2	0.8408	2	0.8808	2	0.8902	1	0.7552	3
S5	0.8574	3	0.8408	3	0.8808	3	0.8902	2	0.7552	4
S13	0.8512	4	0.8373	4	0.8592	4	0.8697	4	0.8253	2
S11	0.7939	5	0.7761	5	0.8061	5	0.8190	5	0.7524	5
S18	0.7681	6	0.7462	6	0.7853	6	0.8088	6	0.7252	6
S10	0.6844	7	0.6685	7	0.6986	7	0.7166	7	0.6497	7
S12	0.6844	8	0.6685	8	0.6986	8	0.7166	8	0.6497	8
S14	0.6844	9	0.6685	9	0.6986	9	0.7166	9	0.6497	9
S19	0.6509	10	0.6328	10	0.6689	10	0.7006	10	0.6211	10
S20	0.5276	11	0.5208	11	0.5569	11	0.5975	11	0.4969	11
S21	0.5276	12	0.5208	12	0.5569	12	0.5975	12	0.4969	12
S26	0.5120	13	0.5045	13	0.5555	13	0.5931	13	0.4481	13
S15	0.4293	14	0.4316	14	0.4355	14	0.4409	16	0.4228	14
S16	0.4293	15	0.4316	15	0.4355	15	0.4409	17	0.4228	15
S17	0.4293	16	0.4316	16	0.4355	16	0.4409	18	0.4228	16
S27	0.4293	17	0.4316	17	0.4355	17	0.4409	19	0.4228	17
S6	0.3883	18	0.3941	18	0.3987	18	0.4064	20	0.3776	18
S7	0.3883	19	0.3941	19	0.3987	19	0.4064	21	0.3776	19
S2	0.3314	20	0.3459	20	0.3569	20	0.3800	23	0.3153	20
S3	0.3314	21	0.3459	21	0.3569	21	0.3800	24	0.3153	21
S8	0.2609	22	0.2770	22	0.2898	23	0.3232	26	0.2517	22
S22	0.2532	23	0.2697	23	0.3511	22	0.4861	14	0.2176	23
S4	0.2307	24	0.2494	24	0.2780	24	0.3301	25	0.2116	24
S23	0.1520	25	0.1751	25	0.2572	25	0.4434	15	0.1307	25
S25	0.1044	26	0.1123	27	0.1296	27	0.1811	27	0.1008	26
S24	0.0979	27	0.1123	26	0.1851	26	0.3986	22	0.0841	27

47.7　绝对残差综合评价指数模型

均方误差和绝对误差是在统计学和机器学习领域中常用的两种评估指标，用于衡量预测值与真实值之间的差异。均方误差（Mean Squared Error，MSE）是预测值与真实值之间差异的平方的平均值。均方误差越小表示模型的预测结果与真实值之间的差异越小。亦即预测值和实际观测的因变量数值的差异，数值大，意味着预测的值越不准确（因为

预测值距离观测到的数值很远）。

绝对误差（Mean Absolute Error，MAE）是预测值与真实值之间差异绝对值的平均值。绝对误差直接衡量了预测值与真实值之间的平均差异，不考虑方向。与均方误差相比，绝对误差更加鲁棒，因为它不受极端值（离群值）的影响，但相对来说，其梯度在零点不连续。

在实际应用中，可以根据具体情况选择使用均方误差或绝对误差作为评估指标，以衡量模型的预测准确性。如果我们将误差函数的定义从均方差改为绝对误差，这时综合评价指数模型中式(47.4)残差的定义为式(47-55)：

$$\sigma=\sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{p}\left|C_{ij}-x_{ij}w_j\right|}{(n-1)\cdot p}} \tag{47-55}$$

基于最大熵-极小化残差优化模型和式（47.7）相同，亦可采用多种带约束条件的非线性优化方法进行求解，并进行模型权重参数Bootstrap抽样估计。

表 47-10　不同误差选项综合指数及其排序

样本	均方误差	绝对差	Rank
常规柴油机	0.6028	0.6106	11
压缩天然气(CNG)	0.6474	0.6588	9
液化气(LPG)	0.6484	0.6600	8
燃料电池(氢)	0.6236	0.6315	10
甲醇	0.6013	0.6083	12
电动汽车	0.6923	0.7022	3
直接充电	0.7054	0.7163	2
可更换电池	0.7090	0.7200	1
汽油混合动力	0.6703	0.6801	6
柴油混合动力	0.6689	0.6773	7
电动 CNG	0.6764	0.6869	4
电动 LPG	0.6764	0.6869	5

在操作界面 47.2 中，将误差函数改选为“绝对差”，这时系统按绝对差进行优化计算，得到结果：能源供应=0.1535,能源效率=0.1985,空气污染=0.1286,噪音污染= 0.1244,产业关系=0.1742,实施成本=0.1536,维护成本=0.1731,车辆能力=0.1465,道路设施=0.1383,交通速度=0.1544,舒适度=0.1960。这个结果和采用均方误差相比（表 47-10），因子间权重系数大小的波动要小一些。不过在本例中，绝对残差模型计算得到的综合指数和均方残差模型的综合指数，各个动力类型的分值排序完全相同（表 47-10）。

47.8 因子加权综合评价模型

在进行多指标综合评价的过程中，如果决策者对某个评价对象有某种偏好，或对某些因子的有了很大程度上的了解，并且由相关专家给出了大体的权重估计。这时可以在专家对各个评价对象、或各个因子给出的经验权重基础上进行综合评价，实现综合评价过程主客观的统一，这时的综合评价指数模型可以由加权综合评价模型来实现。

47.8.1 因子加权的综合分析指数模型

因子加权处理是专家主观权重和客观估计权重系数相互融合、相互作用的过程。综合评价指数生成过程中，主观和客观不是割裂的，而是相互联系、相互作用的。在认识

事物和解释现象时，需要综合考虑主观因素和客观的赋权方法，并将二者有机地结合起来，以获取更全面、准确的综合评价指数。这样，综合指数既可体现专家知识在综合评价指数形成过程中的作用，在主观权重的基础上客观估计权重系数、使得综合评价过程更科学。这种主客观融合，是主客观之间的互动与统一，更能对综合评价指数有更深入、全面的认知。

这里的因子加权综合指数计算，反映的主观因素对客观评价的影响，和以往根据主观因素进行综合评价、或多属性决策本质上是不同的。以往的主观定权没有融合客观数据的信息，主客观之间没有互动与统一，仅是硬性地将主观权重和客观的因子数量进行组合而已。

如果令 ω_j（j=1，2，…，p）为专家评定的第 j 个因子重要性值的权重得分，我们可以在公式 47.7 中的残差函数公式中，对对每个因子的残差进行加权处理，加权后的残差项 σ_ω 为

$$\sigma_\omega=\sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{p}\frac{\left(y_i w_j-x_{ij}w_j\right)^2}{\omega_j^2}}{(n-1)\cdot p}}=\sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{p}\left(\frac{(y_i-x_{ij})w_j}{\omega_j}\right)^2}{(n-1)\cdot p}} \tag{47.56}$$

这里的权重得分 ω_j 是各个因子重要性值进行均值化处理后的转换值。这时多因子综合评价指数模型为

$$\begin{cases}\min\sigma_\omega^E\\ \text{s.t}\sum_{j=1}^{p}w_j=1,\quad w_j\geq 0\end{cases} \tag{47.57}$$

47.8.2 因子加权综合分析指数模型例子

某公司为了进行设备采购，初选了 4 个签章目标。对于设备采购的评价标准，选择了产值(万元)，性能，成本(万元)，利润和排污量 5 个评价指标。相关专家对各个评价因子给出的重要性权重分别为，产值=0.28，性能=0.18，成本=0.18，利润=0.18 和排污量=0.18。权重因子以一行的方式存放到电子表格（如图 47-13）。注意，各指标重要性权重数据区域是在选中各个因子评价矩阵后，按下 Ctrl 键的同时，用鼠标选中。

	A	B	C	D	E	F	J
1		产值(万元)	性能	成本(万元)(低优)	利润	排污量(低优)	
2	A1	6800	5230	4300	0.7500	0.1500	
3	A2	7280	5450	3580	0.6000	0.1200	
4	A3	8800	8870	7690	0.6200	0.1400	
5	A4	8450	7950	4000	0.6700	0.2200	
6							
7	指标重要性权重	0.28	0.18	0.18	0.18	0.18	
8							

图 47-13 因子加权综合指数模型分析数据格式

然后执行“专业统计”—“综合评价与多属性决策”—“最大熵-最小二乘模型”功能，这时系统出现供用户对各个指标特性进行设置的对话框(图 47-14 左)。

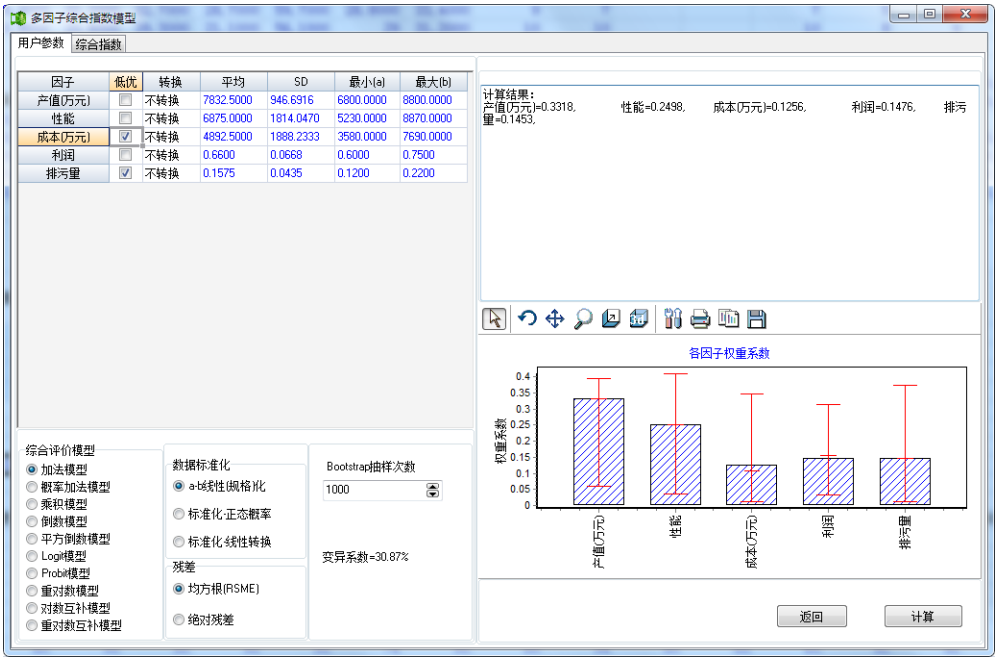


图 47-14　因子加权综合指数模型分析用户界面

这里的评价指标需要标准化处理，我们设置各个因子最小值为零。转换方式选 Min-Max 规格化，即对各个指标按式 47.15*a* 和 47.15*b*（低优指标）进行转换。

作为例子，选用加法型综合评价模型。为估计各个权重系数的置信区间，将 Bootstrap 模拟抽样次数设置为 1000。模型残差指定为均方根误差。

点击”计算”按钮，系统即执行优化计算。因这里需要进行 1000 次的 Bootstrap 模拟抽样，即拟合 1000 个概率加法的非线性模型优化，因此要花费一些时间，用户须稍等一会，系统才能完成计算。完成计算之后，系统输出和加法型模型输出格式完全相同的计算结果如下。

相加型综合评价指数模型
数据 Min-Max 规格化处理

因子	是否低优	均值	标准差	最小值(a)	最大值(b)	是否 0-1 限制
产值(万元)	否	7832.5000	946.6916	6800	8800	是
性能	否	6875	1814.0470	5230	8870	是
成本(万元)(低优)	是	4892.5000	1888.2333	3580	7690	是
利润	否	0.6600	0.0668	0.6000	0.7500	是
排污量(低优)	是	0.1575	0.0435	0.1200	0.2200	是

熵值=2.215634
残差=0.085979
目标函数对数值=-5.4364
对各个因子进行加权处理。

模型拟合方差分析表

模型残差	平方和	自由度	均方
等权重模型	0.1349	19	0.0071
MEMR 模型	0.1062	15	0.0071
模型残差差值	0.0287	4	0.0072

F 统计量=1.0141　　显著性 p 值=0.4435

优化指数=21.2869%

权重系数 Bootstrap 抽样 1000 次估计,时间 0 分 3 秒。

因子	权重系数	抽样均值	标准差	中位数	95%置信区间	
产值(万元)	0.3318	0.3163	0.0624	0.3318	0.0593	0.3955
性能	0.2498	0.2451	0.0597	0.2506	0.0359	0.4093
成本(万元)(低优)	0.1256	0.1306	0.0564	0.1070	0.0133	0.3471
利润	0.1476	0.1514	0.0342	0.1543	0.0320	0.3143
排污量(低优)	0.1453	0.1566	0.0721	0.1453	0.0133	0.3742
熵值	2.2156	2.1692	0.1005	2.1785	1.7178	2.3126
			正态近似置信区间		1.9723	2.3219

样本转换值及综合指数

样本	产值(万元)	性能	成本(万元)(低优)	利润	排污量(低优)	综合指数
A1	0	0	0.8248	1.0000	0.7000	0.3529
A2	0.2400	0.0604	1.0000	0	1.0000	0.3656
A3	1.0000	1.0000	0	0.1333	0.8000	0.7174
A4	0.8250	0.7473	0.8978	0.4667	0	0.6420

系统首先输出最大熵-最小二乘非线性含条件约束模型的信息熵 E 等于 2.215634。拟合残差标准差(σ=0.085979)以及模型目标函数自然对数值为-5.4364。

模型输出结果的第二部分是模型的方差分析结果。检验的原假设为 H_0：因子 j 权重系数等于 $1/p$；备择假设 H_1：因子中各个因子 j 权重系数不是都等于 $1/p$。这部分结果解读要点是从统计学角度，检验各个指标的权重系数的大小是否有差异。如果没有差异，那我们可以采用等权重的方式进行综合指数计算。如果这里的显著性概率 p 值小于 0.05，那我们就应该采用模型给出的各个指标的权重系数来计算综合评价指数。本例中方差分析 F 值等于 1.0141，显著性检验 p=0.4435，远大于 0.05。因此这里的数据和采用等权法估计综合评价指数差异不显著。从权重系数 Bootstrap 抽样 1000 次估计也可以看出，所有的权重系数 95%置信区间都包含简单均值法的权重系数。

各个指标权重系数的输出结果，既显示了各因子对形成综合指数的贡献大小，又是综合评价指数中各个因子的权重重要性的体现。和不加权的加法模型相比，产值项因有较大权重，权重系数要大很多。

47.9 评价对象加权综合评价模型

47.9.1 对评价对象进行加权的综合分析指数模型

在进行多指标综合评价的过程中，如果决策者对方案有偏好信息，决策者在建立综合评价模型时，充分利用这些偏好信息，调整不同评价对象对模型训练或评估的影响。

通过对评价对象进行加权处理，可以更准确地反映不同评价对象在数据集中的重要性，以达到更好的模型性能。根据评价对象进行加权是一种有效的方法，可以帮助模型更好地适应数据分布，提高模型的泛化能力和性能。

如果令 ω_i（i=1，2，…，n）为为第 i 个样本权重，我们可以在公式 47.7 中的残差函数公式中，对每个样本的残差进行加权处理，加权后的残差函数 σ_ω 为

$$\sigma_\omega = \sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{p}\left(\frac{y_i w_j - x_{ij} w_j}{\omega_i}\right)^2}{(n-1)\cdot p}} = \sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{p}\frac{\left(y_i w_j - x_{ij} w_j\right)^2}{\omega_i^2}}{(n-1)\cdot p}} \tag{47.58}$$

这里的权重系数 ω_i 是对各个样本的重要性权数进行均值化处理后的转换值。这时多因子综合评价指数模型为

$$\begin{aligned} &Q_{\min} = \sigma_\omega^E \\ &\text{s.t}\sum_{j=1}^{p} w_j = 1, \quad w_j \geq 0 \end{aligned} \tag{47.59}$$

47.9.2 对评价对象进行加权例子

决策者为了把德才兼备人才徐泽水(2004)在进行多指标综合评价的过程中，如果决策者对方案有偏好信息，决策者在建立综合评价模型时，充分利用这些偏好信息，调整不同评价对象对模型训练或评估的影响。相关专家对各个评价对象的偏好信息以一列的方式存放到电子表格中，并在各个评价因子数据选中之后，按下 Ctrl 键的同时，用鼠标选中（图 47-15）。

	A	B	C	D	E	F	G	H	I	J
1		政策水平	工作作风	业务水平	口才	写作	健康状况		主观偏好	
2	A1	0.9500	0.9000	0.9300	0.8500	0.9100	0.9500		0.82	
3	A2	0.9000	0.8800	0.8500	0.9200	0.9300	0.9100		0.85	
4	A3	0.9200	0.9500	0.9600	0.8400	0.8700	0.9400		0.90	
5	A4	0.8900	0.9300	0.8800	0.9400	0.9200	0.9000		0.75	
6	A5	0.9300	0.9100	0.9000	0.8900	0.9200	0.9500		0.95	
33										

图 47-15　评价对象加权综合指数模型分析用户界面

然后执行“专业统计”→“综合评价与多属性决策” →“最大熵-最小二乘模型”功能，这时系统出现供用户对各个指标特性进行设置的对话框(图 47-16 左)。

这里的评价指标已是规格化处理的数据，这时我们可设置各个因子最小值为零、最大值为 1，转换方式选 Min-Max 规格化，即相当于转换后原来的数值不变。

作为例子，选用加法型综合评价模型。为估计各个权重系数的置信区间，将 Bootstrap 模拟抽样次数设置为 1000。模型残差指定为均方根误差。

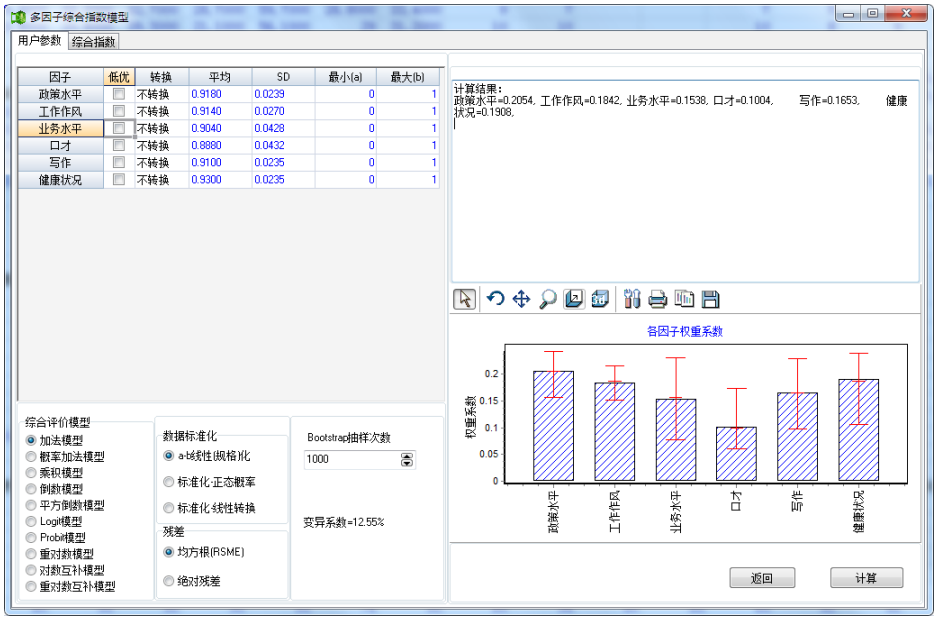


图 47-16　评价对象加权综合指数模型分析用户界面

点击”计算”按钮，系统即执行优化计算。因这里需要进行 1000 次的 Bootstrap 模拟抽样，即拟合 1000 个加法模型优化，因此要花费一些时间，用户须稍等一会，系统才能完成计算。完成计算之后，系统输出和加法型模型输出格式完全相同的计算结果如下。

相加型综合评价指数模型
数据 Min-Max 规格化处理

因子	是否低优	均值	标准差	最小值(a)	最大值(b)	是否 0-1 限制
政策水平	否	0.9180	0.0239	0	1	是
工作作风	否	0.9140	0.0270	0	1	是
业务水平	否	0.9040	0.0428	0	1	是
口才	否	0.8880	0.0432	0	1	是
写作	否	0.9100	0.0235	0	1	是
健康状况	否	0.9300	0.0235	0	1	是

熵值=2.552286
残差=0.004918
目标函数对数值=-13.5649
对各个评价对象加权处理。

模型拟合方差分析表

模型残差	平方和	自由度	均方
等权重模型	0.0008	29	0.0000
MEMR 模型	0.0006	24	0.0000
模型残差差值	0.0002	5	0.0000

F 统计量=1.6422　显著性 p 值=0.1790
优化指数=25.4914%

权重系数 Bootstrap 抽样 1000 次估计,时间 0 分 3 秒。

因子	权重系数	抽样均值	标准差	中位数	95%置信区间	
政策水平	0.2054	0.2058	0.0199	0.2077	0.1576	0.2430
工作作风	0.1842	0.1865	0.0076	0.1866	0.1602	0.2251
业务水平	0.1538	0.1542	0.0221	0.1547	0.0834	0.2145
口才	0.1004	0.1010	0.0195	0.0972	0.0510	0.1710
写作	0.1653	0.1654	0.0276	0.1631	0.0819	0.2298
健康状况	0.1908	0.1871	0.0214	0.1895	0.0890	0.2472
熵值	2.5523	2.5413	0.0185	2.5462	2.4183	2.5675
			正态近似置信区间		2.5051	2.5775

样本转换值及综合指数

样本	政策水平	工作作风	业务水平	口才	写作	健康状况	综合指数
A1	0.9500	0.9000	0.9300	0.8500	0.9100	0.9500	0.9211
A2	0.9000	0.8800	0.8500	0.9200	0.9300	0.9100	0.8975
A3	0.9200	0.9500	0.9600	0.8400	0.8700	0.9400	0.9192
A4	0.8900	0.9300	0.8800	0.9400	0.9200	0.9000	0.9077
A5	0.9300	0.9100	0.9000	0.8900	0.9200	0.9500	0.9199

系统首先输出最大熵-最小二乘非线性含条件约束模型的信息熵 E 等于 2.5523。拟合残差标准差(σ=0.004918)以及模型目标函数自然对数值为-13.5649。

模型输出结果的第二部分是模型的方差分析结果。本例中方差分析 F 值等于 1.6422，显著性检验 p=0.1790，大于 0.05。因此这里的数据和采用等权法估计综合评价指数差异不显著。从权重系数 Bootstrap 抽样 1000 次估计也可以看出，所有的权重系数 95%置信区间都包含简单均值法的权重系数。

输出结果第三部分为各个指标的权重系数，是综合评价分析模型的核心和重点。它既是各个因子对形成综合指数的贡献大小，又是综合评价指数中各个因子的权重重要性的体现，这里可以看出，因子“政策水平”的权重系数最大，其次是健康状态和工作作风。这是在最大熵条件下、拟合误差达到最小的结果。

最后综合评价值，按大小排序是 A1>A5>A3>A4>A2。如果不进行加权处理，同样应用加法模型评价，最后综合评价值，按大小排序是 A5>A3>A1>A4>A2。看来专家意见对评价结果影响也是很大的。

47.10 分亚组综合评价体系的集成

47.10.1 分亚组综合评价模型

分亚组综合评价是一类包含如果子组的多因子综合体系。一个系统分成若干亚组，可以更细致地了解不同子群体的情况，避免因为整体数据的平均效果而掩盖了各个子群体的特点。这种方法在调查、研判或决策过程中能够提供更全面和准确的信息，有助于做出更具针对性的决策。

本章提出的综合分析模型亦可用于多个子组的组合评价，集成多个子系统。这些子

系统既包括同一评价范围指标体系、或不同层次、不同评价方法的“集成”，甚至可将不同评价模型混合而成为新的评价体系的综合，如不同形式的模糊隶属函数，灰色白化函数等作为多因子指标进行组合。不管不同评价方法的评价值取值方向与数量级别、区间长度是否一致，只要通过数据标准化处理，变换为有相同区间的评价值即可进行综合评价模型中实现这种“集成”。

不同子系统的集成，各子集权重如何分配，尽管在以往认为是十分棘手的问题，但在这里已经得到了很好的解决。因为多指标综合评价过程中最基本的要素仍然是权：单项指标价值、子集的权。只要权数科学、合理，子集集成问题就迎刃而解。

针对每个亚组，进行独立的评价和分析。这可能涉及到对每个亚组应用特定的评价方法或模型，以了解各个亚组的特点、优劣势等。

根据各个亚组的评价结果，再应用综合评价模型将各个亚组的评价结果综合起来，得出总体评价。最后，根据总体评价结果对各个亚组的表现进行比较和解释，以便更好地理解整体情况。分亚组综合评价模型可描述为：

若有 N 个待评价的对象，每个对象有 G 个子组，每个子组有 $G(g)$个因子指标。

$$y_i=\sum_{g=1}^{G}\eta_g\left(\sum_{k=1}^{G(g)}x_{igk}w_{gk}\right) \tag{47.60}$$

式中 x_{igk} 为第 i 个评价对象，第 g 个子组的第 k 个因子，w_{gk} 是第 g 个子组的第 k 个因子的权重系数；η_g 是第 g 个子组的权重系数。

进行综合评价和多属性决策时，我们可以先求出每个子组的权重系数，即 g=1，2，…，G 时的权重系数 w_{gk}，然后计算每个子组的综合评价值 C_{ig}。再根据每个子组的综合评价值 C_{ig} 估计各个子组的权重系数 η_g。最后计算各个评价对象的综合评价指数。

综合指数的集成过程较为复杂，没有统一的表达形式，可根据实际问题确定计算模式，可表示为各个指标的相加或相乘，或概率加法模型。这可应用式（47.7）和式（47.15）进行。两个层次的组合评价，可在 DPS 里面直接完成。

47.10.2 分亚组综合评价模型界面与操作

例 1 这里以某医院历年 4 个方面、11 项指标的工作质量进行评价。分析时，将数据编辑成如图 47-17 格式。

	A	B	C	D	E	F	G	H	I	J	K	L	
	指标类型	医疗质量			工作质量			动态指标		病床使用			
	指标名称	治疗率	病死率	成功率	感染率	愈合率	死亡率	门诊量	出院人数	使用率	周转率	住院日	
	1994 年	94.70	2.30	88.10	5.20	97.60	0.23	606.0	15236	89.8	30.6	10.6	
	1995 年	95.30	1.70	89.50	4.50	98.80	0.19	727.0	15201	86.6	30.1	10.3	
	1996 年	95.00	1.60	87.80	4.10	98.00	0.15	776.0	16940	90.1	33.5	9.7	
	1997 年	95.40	1.30	90.80	3.60	97.10	0.18	778.0	17101	82.8	32.3	9.1	
	1998 年	95.10	1.30	88.40	3.60	98.30	0.17	814.0	17783	82.1	33.2	8.1	

图 47-16 某医院 1994~1998 年 11 项指标分组数据编辑格式

按图 47-17 方式编辑并选中数据后，执行“专业统计”下面的“综合评价及多属性决策”→“多子组最大熵-最小二乘模型”功能，这时系统出现供用户对各个指标特性进行设置的对话框(图 47-18)。

多子组多因子综合评价或多属性决策，因子转换、子组内聚合方式、以及子组间的聚合方式都有可能不同，因此用户界面设计须考虑各种组合情况，由用户自己设计各人的组合模式。故多子组的用户界面稍微复杂一点。

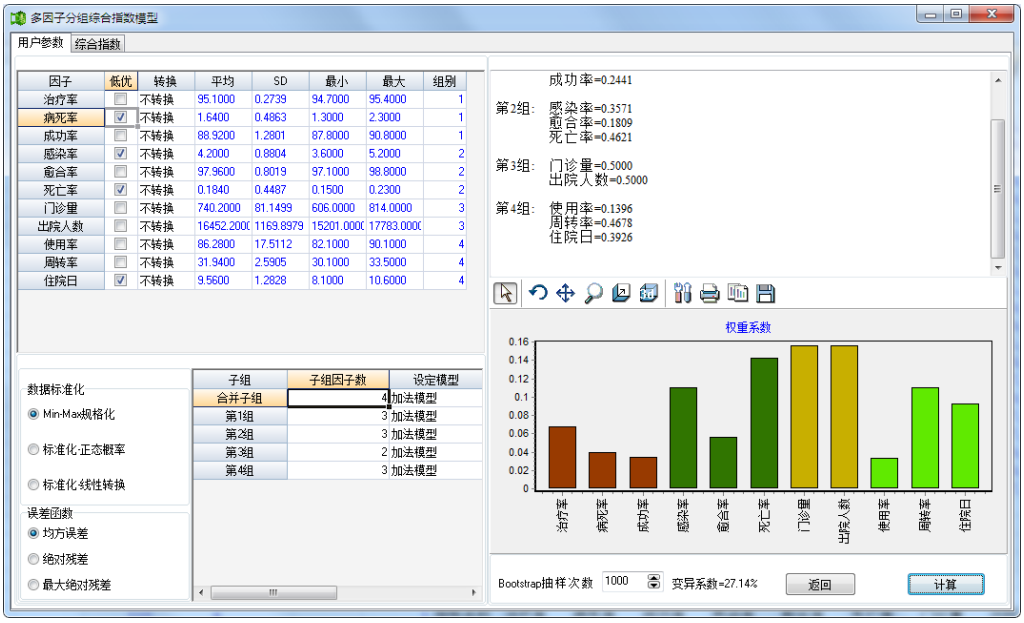


图 47-18 多子组数据综合指数模型用户界面

在分组数据综合指数模型用户界面左边，和一般的综合指数模型相比，左上角各个因子基本情况栏的最右边多了一列，在这里是由用户输入该因子属于第几组的顺序数字号（必须输入自然数）。因子组编号是从 1 开始，因此输入自然以表示各个子组的次序。即用数码 1，2，3，4……所表示。

如果把某个因子的序号（自然数）删去，则在综合分析建模时就不考虑这个因子。直分销其他有自然数的因子。

本例中，可将 11 个因子分成医疗质量、工作质量、动态指标和病床使用 4 个亚组。我们就可以将每个因子后面输入对应 4 个亚组的数字序号，用 1、2、3 和 4 表示 4 个亚组。即填写到最右边那一列中（如图 47-18）。当然在因子表中，每个子组的各个因子的位置不必是邻接的，可以在任意的行。

用户界面左下方用户可以选择数据格式化方法，和各个子组采用什么样的模型来建模（默认是加法模型），且可以是每个子组的模型各不相同。几个子组的合并方式也可以根据用户的需要来设定。

注意，这里的例子中，病死率、院内感染率、新生儿死亡率、平均住院日为低优指标，需在前面进行勾选，以让系统识别这几个是数值越低越好的 “低优指标”。

其它设置，如数据标准化（格式化）方式、是否进行 Bootstrap 模拟抽样，都有默认值，用户可以根据需要进行设置。

参数设置完成之后，点击”计算”按钮，这时系统给出分析结果如下：

分析结果的第一部分是综合每个子组的评价分析，这是根据每个亚组的综合指数，基于最大熵-最小残差综合分析指数模型计算得到的。结果中，首先是汇集各个子组的方差分析，以从统计学角度，检验各子组权重系数大小是否有差异。如果没有差异，那我们可以采用等权重的方式来综合个子组的综合指数。这里的 F 值为 2.7246，显著性概率 p 值为 0.0787，大于 0.05，理论上各个子组的权重系数可采用等权法估计综合评价指数。

因子	是否低优	均值	标准差	最小值(a)	最大值(b)	组别
治疗率	否	95.1000	0.2739	94.7000	95.4000	1
病死率	是	1.6400	0.4863	1.3000	2.3000	1
成功率	否	88.9200	1.2801	87.8000	90.8000	1
感染率	是	4.2000	0.8804	3.6000	5.2000	2
愈合率	否	97.9600	0.8019	97.1000	98.8000	2
死亡率	是	0.1840	0.4487	0.1500	0.2300	2
门诊量	否	740.2000	81.1499	606	814	3
出院人数	否	16452.2000	1169.8979	15201	17783	3
使用率	否	86.2800	17.5112	82.1000	90.1000	4
周转率	否	31.9400	2.5905	30.1000	33.5000	4
住院日	是	9.5600	1.2828	8.1000	10.6000	4

系统各子组加法模型

熵值=1.940178

残差=0.037099

目标函数对数值=-6.3913

方差分析表

变异来源	平方和	自由度	均方	F 值	p 值
模型	0.0113	3	0.0038	2.7247	0.0787
残差	0.0220	16	0.0014		

优化指数=33.8134%

权重系数 Bootstrap 抽样 1000 次估计:

因子	权重系数	抽样均值	标准差	中位数	95%置信区间	
第 1 组	0.1425	0.1382	0.0337	0.1425	0.0491	0.2510
第 2 组	0.3093	0.3205	0.0286	0.3135	0.2369	0.3973
第 3 组	0.3121	0.3007	0.0402	0.3091	0.1666	0.3961
第 4 组	0.2361	0.2406	0.0377	0.2354	0.1355	0.3871

样本	第 1 组	第 2 组	第 3 组	第 4 组	综合指数
1994 年	0.0245	0.0532	0.0068	0.2032	0.0700
1995 年	0.7140	0.5681	0.2909	0.1257	0.3979
1996 年	0.4001	0.8033	0.7454	0.7487	0.7149
1997 年	0.9999	0.6459	0.7814	0.5505	0.7161
1998 年	0.6010	0.8313	1.0000	0.8191	0.8483

第 1 组因子加法模型

熵值=1.521121

残差=0.062300

目标函数对数值=-4.2223
方差分析表

变异来源	平方和	自由度	均方	F 值	p 值
模型	0.0231	2	0.0115	2.9714	0.0895
残差	0.0466	12	0.0039		

优化指数=33.1208%

权重系数 Bootstrap 抽样 1000 次估计:

因子	权重系数	抽样均值	标准差	中位数	95%置信区间	
治疗率	0.4753	0.4563	0.0866	0.4811	0.0446	0.5448
病死率	0.2806	0.2947	0.0559	0.2806	0.2104	0.5115
成功率	0.2441	0.2489	0.0527	0.2432	0.0000	0.4438

样本	治疗率	病死率(低优)	成功率	综合指数
1994 年	0.0001	0	0.1000	0.0245
1995 年	0.8570	0.6000	0.5667	0.7140
1996 年	0.4286	0.7000	0.0000	0.4001
1997 年	0.9999	1.0000	1.0000	0.9999
1998 年	0.5714	1.0000	0.2000	0.6010

第 2 组因子加法模型
熵值=1.491387
残差=0.070614
目标函数对数值=-3.9530
方差分析表

变异来源	平方和	自由度	均方	F 值	p 值
模型	0.0427	2	0.0214	4.2848	0.0394
残差	0.0598	12	0.0050		

优化指数=41.6614%

权重系数 Bootstrap 抽样 1000 次估计:

因子	权重系数	抽样均值	标准差	中位数	95%置信区间	
感染率	0.3571	0.3649	0.0439	0.3617	0.0664	0.4848
愈合率	0.1809	0.1888	0.0511	0.1809	0.0186	0.4266
死亡率	0.4621	0.4464	0.0677	0.4621	0.2070	0.5543

样本	感染率(低优)	愈合率	死亡率(低优)	综合指数
1994 年	0	0.2941	0	0.0532
1995 年	0.4375	0.9999	0.5000	0.5681
1996 年	0.6875	0.5294	1.0000	0.8033
1997 年	1.0000	0.0001	0.6250	0.6459
1998 年	1.0000	0.7059	0.7500	0.8313

第 3 组因子加法模型
熵值=1.000000
残差=0.075783
目标函数对数值=-2.5799
方差分析表

变异来源	平方和	自由度	均方	F 值	p 值
模型	0.0000	1	0.0000	0.0000	1.0000
残差	0.0459	8	0.0057		

优化指数=0.0000%

权重系数 Bootstrap 抽样 1000 次估计:

因子	权重系数	抽样均值	标准差	中位数	95%置信区间	
门诊量	0.5000	0.5000	0.0000	0.5000	0.5000	0.5000
出院人数	0.5000	0.5000	0.0000	0.5000	0.5000	0.5000

样本	门诊量	出院人数	综合指数
1994 年	0	0.0136	0.0068
1995 年	0.5817	0	0.2909
1996 年	0.8173	0.6735	0.7454
1997 年	0.8269	0.7359	0.7814
1998 年	1.0000	1.0000	1.0000

第 4 组因子加法模型

熵值=1.438902

残差=0.086801

目标函数对数值=-3.5169

方差分析表

变异来源	平方和	自由度	均方	F 值	p 值
模型	0.1087	2	0.0543	7.2128	0.0088
残差	0.0904	12	0.0075		

优化指数=54.5894%

权重系数 Bootstrap 抽样 1000 次估计:

因子	权重系数	抽样均值	标准差	中位数	95%置信区间	
使用率	0.1396	0.1326	0.0812	0.1366	0.0185	0.5080
周转率	0.4678	0.4723	0.0302	0.4708	0.3846	0.5373
住院日	0.3926	0.3951	0.0625	0.3942	0.0000	0.5205

样本	使用率	周转率	住院日(低优)	综合指数
1994 年	0.9625	0.1471	0	0.2032
1995 年	0.5625	0	0.1200	0.1257
1996 年	1.0000	1.0000	0.3600	0.7487
1997 年	0.0875	0.6471	0.6000	0.5505
1998 年	0.0000	0.9118	1.0000	0.8191

组别	组权重	因子	权重系数	总比例
第 1 组	0.1425	治疗率	0.4753	0.0677
第 1 组	0.1425	病死率	0.2806	0.0400
第 1 组	0.1425	成功率	0.2441	0.0348
第 2 组	0.3093	感染率	0.3571	0.1104
第 2 组	0.3093	愈合率	0.1809	0.0559
第 2 组	0.3093	死亡率	0.4621	0.1429
第 3 组	0.3121	门诊量	0.5000	0.1560
第 3 组	0.3121	出院人数	0.5000	0.1561
第 4 组	0.2361	使用率	0.1396	0.0330
第 4 组	0.2361	周转率	0.4678	0.1105
第 4 组	0.2361	住院日	0.3926	0.0927

综合模型中各个指标权重系数，即医疗质量、工作质量、动态指标和病床使用这 4 个指标，其权重系数分别为 0.1425、0.3093、0.3121 和 0.2361。权重系数 Bootstrap 检验，其置信区间都包含 0.25(等权权重)，没有显著大于等权权重的指标，这和方差分析结果一致。

各个亚组中以动态指标的权重系数最大，工作质量和病床使用率次之，医疗质量所占权重较小。然后给出了各年的综合指数，1994~1998 年综合指数分别为 0.0700,0.3979,0.7149,0.7161 和 0.8483；各年指数呈上升趋势，可以认为自 1994 年以来，该医院工作状况逐步进入佳境。

分析结果的第二部分是每个子组的综合评价指数建模结果。以第一子组为例，该组的治疗率、成功率、病死率这 3 个因子指标的权重系数分别为 0.4753、0.2806 和 0.2441。显然成功率在第一子组里面的权重所占的比例最大，其次才是成功率和病死率。1994~1998 年综合指数分别为 0.0245、0.7140、0.4001、0.9999 和 0.6010；各年指数总体来讲呈上升趋势。

47.10.3 人类发展指数研究

联合国开发计划署人类发展指数研究中，首先根据四个基础数据，预期寿命、预期教育年数、在校平均年数和收入指数，将其中的预期教育年数和在校平均年数计算均值，作物教育指数。然后将寿命指数、教育指数、收入指数归一化成 0 到 1 之间的一个数值，数值越大越好。最后，人类发展指数(HDI)是寿命指数、教育指数、收入指数三个数值的几何平均，即三个指数相乘再开三次方，也是一个 0 到 1 的数值。

研究中，对寿命指数计算，令 a=20，b=85。寿命指数= (预期人均寿命- a) / (b-a)。这里 b=85 是有意义的，相当于设置了一个最高的值，没有国家与地区高于这个高限。2022 年预期人均寿最长的香港 84.3 岁。a=20 是一个最低值，如果人均寿命只有 20，得分就是 0。这个值不是随便设计的，历史研究表明，如果一个族群的人均寿命低于 20 这个生殖年龄，族群就会消失。当然没有国家这么低，再差的国家多少也能得一些分。

教育指数计算中，平均学校教育年数指数 MYSI =(MYS-a)/(b-a)，这里 a=0，b=15。这个归一化处理更简单，平均教育年数高达 15 的，指数值是 1.0。2022 年平均教育年数高的德国,为 14.3。联合国认为没有国家的教育是完美的。最差的也不会是 0，多少会有些分数。因此最低值取 0。

预期学校教育年数指数定义为 EYSI = EYS/18。归一化处理和 MYSI 一样，最低值为 0，最高值为 18。2022 年预期教育年数最高的是澳大利亚的 21.1，超过了 18 年，就技术处理成 18 年，指数值为 1.0。最低的是厄立特里亚的 4.1，尼日尔的 5.4。中国预期教育年数是 13.1，EYSI 指数值是 0.728。

教育指数= (MYSI + EYSI) /2，即两个学校教育指数作算术平均就是最终的教育指数。

收入指数。各国的收入指数差距拉得比较大，可能有几十上百倍的差距，比国民收入与寿命相比，差距很大。所以不能拿绝对值差异来算收入指数，不然一堆国家的指数值都是接近于 0。故对收入数值先取自然对数，相当于用对数坐标显示收入差距，就能看出一些层次了。在规格化时，高限取 75000，而不是更高的值，如果有国家收入比这

还高就当是 75000。如 2022 年人均国民收入最高的是 146673 美元，也只取 75000（再高也没意义了）。最低的是南苏丹，581 美元，也多少有点得分。

最终的 HDI 指数。联合国开发计划署定义 HDI 指数等于寿命指数*教育指数*收入指数，然后开三次方作几何平均。用几何平均，而不用算术平均，这可能是联合国想搞“一票否决”。因 2010 年前是算术平均。如果用算术平均，有两个数值过得去，能把另一个低分给补一下。用几何平均算法，得一个低分，另两个高分作用也不大，最终还是低分。

这里我们应用 2022 年联合国开发计划署定义计算 HDI 指数的原始数据（https://hdr.undp.org/sites/default/files/2023-24_HDR/HDR23-24_Statistical_Annex_HDI_Table.xlsx），对寿命指数、教育指数和收入指数应用我们提出的乘积型综合知识模型，计算各个因子的权重系数。

首先将全球 193 个国家的预期寿命、预期教育年数、在校平均年数和收入指数这 4 个因子数据读入到 DPS 电子工作表中。然后用鼠标选中待分析的数据。再在 DPS 数据处理系统菜单方式下，执行“专业统计”下面的“综合评价及多属性决策”→“多子组最大熵-最小二乘模型”功能，这时系统出现供用户对各个指标特性进行设置的对话框(图 47-19)。

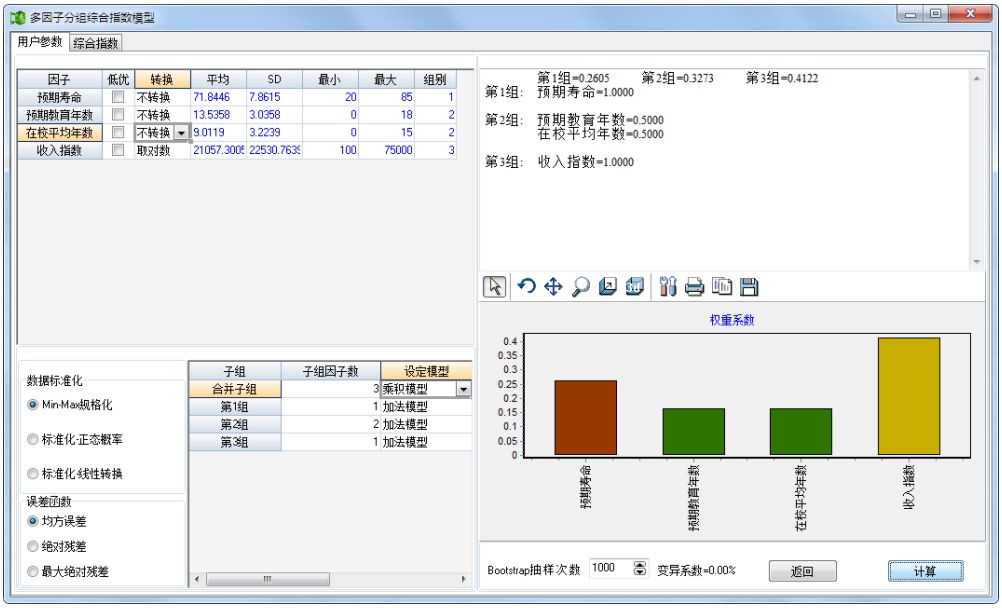


图 47-19　人类发展指数多子组综合指数模型用户界面

设置时，因输入指数需要对原始数据进行取对数转换，因子在数据转换一栏中将收入指数后面的“不转换”改选成“取对数”。

并按图示，将各个指标的最小值、最大值、通过直接输入修改成前面讨论过的几个阈值。

在各个因子基本情况列表最右边是设置各个因子分组的序号。这里第 1 个因子为第一组，所以输入“1”，第 2 个、第 3 个因子是第二组，所以在他们的右边标记上组号“2”，第 4 个因子为第 3 组，组号的值改为 3。注意这里的第一组、第三组都只有 1 个因子。

在数据规格化选择框中，选择默认的“Min-Max 规格化”，误差函数为“均方误差”。在设定模型选择框中，子组合并过程设定为乘积模型，以和联合国开发计划署的综合指数计算算法原则上一致。

最后，将 Bootstrap 随机抽样次数设为 1000。点击”计算”后，系统采用式(47.16)乘积型综合评价指数模型、并进行了 1000 个样本的 Bootstrap 随机抽样，构建乘积型人类发展指数模型。优化计算，得到寿命指数、教育指数和收入指数这 3 因子的综合评价指数模型计算结果如下：

Min-Max 规格化处理

因子	是否低优	数据转换	均值	标准差	最小值(a)	最大值(b)
预期寿命	否	不转换	71.8446	7.8615	20	85
预期教育年数	否	不转换	13.5358	3.0358	0	18
在校平均年数	否	不转换	9.0119	3.2239	0	15
收入指数	否	取对数	21057.3005	22530.7639	100	75000

数据标准化：Min-Max 规格化
误差函数：均方差
系统各子组乘积模型
熵值=1.448703
残差=0.035126
目标函数对数值=-2.4195
方差分析表

变异来源	平方和	自由度	均方	F 值	p 值
模型	0.1923	2	0.0962	57.0068	0.0000
残差	0.9718	576	0.0017		

优化指数=16.5234%

权重系数 Bootstrap 抽样 1000 次估计:

因子	权重系数	抽样均值	标准差	中位数	95%置信区间	
第 1 组	0.2605	0.2606	0.0100	0.2611	0.2288	0.2913
第 2 组	0.3273	0.3270	0.0122	0.3266	0.2843	0.3653
第 3 组	0.4122	0.4124	0.0084	0.4127	0.3864	0.4354
变异系数	3.1916					

样本	第 1 组	第 2 组	第 3 组	综合指数
Switzerland	0.9892	0.9244	0.9883	0.9672
Norway	0.9754	0.9367	0.9878	0.9676
Iceland	0.9662	0.9600	0.9523	0.9584
Hong Kong, China	0.9892	0.9044	0.9724	0.9539
Singapore	0.9862	0.8661	1.0000	0.9506
…	…	…	…	…
South Sudan	0.5477	0.3456	0.2920	0.3635
Somalia	0.5554	0.2744	0.3583	0.3681

组别	组权重	因子	权重系数	总比例
第 1 组	0.2605	预期寿命	1	0.2605
第 2 组	0.3273	预期教育年数	0.5000	0.1637
第 2 组	0.3273	在校平均年数	0.5000	0.1637
第 3 组	0.4122	收入指数	1	0.4122

输出结果的第一部分是几个子组的合并评价模型。系统各子组乘积模型熵值=1.448703，残差=0.041074，目标函数对数值=-4.9800。并从统计学角度，检验各个指标的权重系数的大小是否有差异。如果没有差异，那我们可以采用等权重的方式进行综合指数计算。这里方差分析 F 值为 57.0068，显著性概率 p 值小于 0.0001。可以认为这里的数据不宜采用等权法估计综合评价指数，而应采用综合指数模型进行综合指数的计算。

模型检验之后，系统输出了各个子组的权重系数。如果进行了 Bootstrap 模拟抽样，在权重系数后面输出 Bootstrap 模拟抽样估计的抽样均值、标准差、中位数和95% 置 信区间。从这里结果看出，3 个子组的权重系数分别为 0.2605、0.3273 和 0.4122。显然和各因子平均的权重系数(0.3333)相比，第一子组（期望寿命）显著偏小，而第三子组（国民收入）的权重系数明显偏大。因此不适合应用几何平均方法来进行综合评价。

表 47-11 Logistic 模型分组综合指数和因子分析样本得分

样本	第 1 组	第 2 组	FY1	FY2
S1	0.7511	0.4415	1.6521	-0.8893
S2	0.5556	0.3333	-0.1192	-1.1470
S3	0.5556	0.3333	-0.1192	-1.1470
S4	0.5146	0.3333	-0.2402	-1.1645
S5	0.7511	0.4415	1.6521	-0.8893
S6	0.5556	0.4415	-0.1376	-1.1483
S7	0.5556	0.4415	-0.1376	-1.1483
S8	0.5040	0.4415	-0.9176	-1.2617
S9	0.7441	0.5556	1.7131	0.3382
S10	0.6422	0.5556	0.2773	0.1304
S11	0.6938	0.5556	1.0572	0.2438
S12	0.6422	0.5556	0.2773	0.1304
S13	0.6938	0.7778	1.0877	1.4678
S14	0.6422	0.5556	0.2773	0.1304
S15	0.5556	0.5556	-0.0888	0.0769
S16	0.5556	0.5556	-0.0888	0.0769
S17	0.5556	0.5556	-0.0888	0.0769
S18	0.6466	0.7778	0.8547	1.4341
S19	0.5918	0.7778	0.0748	1.3207
S20	0.5556	0.7778	-0.0583	1.3009
S21	0.5556	0.7778	-0.0583	1.3009
S22	0.4486	0.7778	-1.4941	1.0931
S23	0.4022	0.7778	-1.7270	1.0594
S24	0.3632	0.7778	-1.8480	1.0419
S25	0.4083	0.4415	-1.6945	-1.3735
S26	0.6007	0.4415	-0.0166	-1.1308
S27	0.5556	0.5556	-0.0888	0.0769

权重系数输出之后，系统给出了 193 个国家（地区）3 个子组规格化值，及其根据模型优化得到的权重系数，应用乘积模型计算得到的综合指数。根据综合指数大小可完成对 193 个国家（地区）的名次排序。这样的结果应该更为客观、合理和公平。

输出结果第二部分为各个子组综合评价模型优化结果。因这里的子组 1 和子组 3 分别仅含 1 个因子，因此无需模型计算，直接将权重系数置为 1.0 即可。模型的综合指数值即为该组的因子指标值。

子组 2 仅含两个因子。当应用规格化数据、加法模型估计权重系数时，两个因子的权重系数都是 0.5.这和标准化后的数据进行主成分分析，第一主成分在两个变量上的规格化特征向量的绝对值都是 0.7071（0.5 的平方根值）。模型的综合指数值即为该组的两个因子指标值的算术平均值。

47.10.4 综合评价因子分亚组研究

在采用 Logistic 模型对胡建军等（2001）关于烟叶感官质量 7 个因子综合评价研究结果发现，7 个因子中，香气质、香气量、杂气、刺激和余味这 5 个指标均有较大的权重系数，说明这 5 个因子在评价过程中趋势大体一致，代表了各个样品的综合趋势。而另外两个因子，劲头和浓度的权重较小，这两个因子的趋势可能好前面 5 个因子的共同趋势不是很一致。因此，另一种想法是将前面 5 个因子作为一个亚组，后面两个因子作为一个亚组。

将香气质、香气量、杂气、刺激和余味这 5 个因子采用 Logistic 模型计算综合指数的结果列于表 47-11（第 1 组），劲头和浓度这两个因子采用 Logistic 模型计算综合指数的结果列于表 47-11（第 2 组）。并对烟叶感官质量 7 个因子采用极大似然估计方法进行因子分析，得到两因子的样本得分，分别列于表 47-11 中的 YF1、YF2 两列。

综合评价研究结果发现，7 个因子中，香气质、香气量、杂气、刺激和余味这 5 个指标均有较大的权重系数，说明这 5 个因子在评价过程中趋势大体一致，代表了各个样品的综合趋势。而另外两个因子，劲头和浓度

对表 47-11 中两种方法得到的样本综合知识（得分）进行相关分析，相关分析结果列表于表 47-12。

表 47-12 两种建模方法样本得分相关系数

变量	第 2 组	FY1	FY2
第 1 组	-0.1613	**0.9817**	-0.0193
第 2 组		-0.1648	**0.9613**
FY1			-0.0106

从相关分析结果来看，第一组综合指数和因子分析第一主因子的样本得分的相关系数为 0.9817；第二组综合指数和因子分析第二主因子的样本得分的相关系数为 0.9613。均显示出了极高的相关程度。由此可见，应用最大熵最小二乘技术建立的模型所给出的综合评价指数客观地反映了各个因子的综合效应。和因子分析等线性变换的多元统计分析方法相比，可望获得的综合指数更能客观地反映各个因子综合效应。

47.11　面板数据多因子综合模型

传统的多准则综合评价方法只适用于一个“截面”的多指标数据分析。面对多个截面，如不同地点、不同时间、或不同专家群体决策获取的面板数据时几乎失去了评价的能力。因为受限于传统综合评价权重系数估计方法的局限，不仅不是很合理、客观，同样也无法对面板数据进行精准的多因子综合评价分析、估计时间权重、专家权重或地点权重等表达界面间变化趋势的参数。

多因子综合评价及多属性决策研究，因受因子权重系数难以估计、不能建立多因子和综合指数之间定量模型的影响，更难开展多因子面板数据综合评价的研究。因此面板数据的综合评价和多属性决策几乎无相关科研工作者涉足，目前仍是一片空白。

但是在自然、社会科学研究中，面板数据是常用的一种数据形式，特别是将面板数据各个截面数据、同时考虑各个截面（如按时间的顺序，截面间具有某种趋势）的变化趋势、将各个截面整合起来进行分析，提取多指标面板数据的重要信息，即各个因子的权重系数和各个界面的权重系数，是一个目前亟需解决的难题。

应用最大熵最小二乘综合评价模型，深入开展面板数据的综合评价及多属性决策研究目前尚为空白。这方面研究应该具有广阔的前景。特别是多个专家在多属性决策方面的研究、探讨。

作者提出的最大熵-最小二乘多因子综合评价模型，其实质是一类特殊的，即不含显式的因变量的回归方程模型。因此，应用这类特殊的回归模型，对多因子面板数据，亦可建立相应综合评价模型。只要我们给出适合面板数据分析的恰当的数据模型，应用最大熵最小二乘，不难对面板数据综合评价模型进行优化。通过模型优化，不仅可以估计

面板中各个因子的权重系数，亦可估计各个横截面间的变化系数（斜率系数）。允许不同界面在共同效应上拥有不同的权重（载荷）系数。通过一个多因子综合评价模型，来计算多因子面板数据的综合指数、或进行多属性决策。

47.11.1 面板数据综合分析指数模型

面板数据的多因子综合评价指数统计模型，在前面的统计模型中增加了面板因子，这种情形下，我们可以在拟合残差项中增加一个面板因子ω。

若有 t 个时段，每个时段有 n 个 DUMs，p 个因子指标。在每个面板中每个因子的权重系数相同，即每个因子所占比例相同。不同面板之间，权重系数相差比例系数是 τ_t。这时面板数据多因子指数评价模型为

$$\begin{aligned}
y_{1i} &= x_{1i1}\tau_1 w_1 + x_{1i2}\tau_1 w_2 + x_{1i3}\tau_1 w_3 + \cdots + x_{1ip}\tau_1 w_p \\
y_{2i} &= x_{2i1}\tau_2 w_1 + x_{2i2}\tau_2 w_2 + x_{2i3}\tau_2 w_3 + \cdots + x_{2ip}\tau_2 w_p \\
&\qquad\qquad (47.61) \\
y_{\mathrm{ti}} &= x_{ti1}\tau_t w_1 + x_{ti2}\tau_t w_2 + x_{ti3}\tau_t w_3 + \cdots + x_{tip}\tau_t w_p
\end{aligned}$$

均方误差残差等于

$$\sigma = \sqrt{\frac{\sum_{h=1}^{t}\sum_{i=1}^{n}\sum_{j=1}^{p}\left(y_{hi}\tau_h w_j - x_{ij}\tau_h w_j\right)^2}{(n-1)\cdot p - t}} \qquad (47.62)$$

由此，整个面板数据综合评价指数模型的形式为：

$$\begin{cases} \min \sigma^E \\ \text{s.t} \begin{cases} \sum_{j=1}^{p} w_j = 1, \quad w_j \geq 0 \\ \sum_{h=1}^{t} \tau_h = t, \quad \tau_h \geq 0 \end{cases} \end{cases} \qquad (47.63)$$

47.11.2 面板数据综合分析指数例子

例1 医疗质量评估 面板数据综合指数模型建立，这里以湖南省某医院信息科获得该院各临床科室年年的相关资料(王一任，2012,综合评价方法若干问题研究及其医学应用(博士论文))，限于篇幅， 从中选取个科室从工作量、治疗质量、诊断质量、效率质量等方面进行动态综合评价。评价指标，出院人数(人); 入出院诊断符合率(%); 治疗有效率(%); 平均床位使用率(%); 病床周转次数(次)；出院者平均住院日(日)。面板数据综合分析指数数据格式如图47-20。

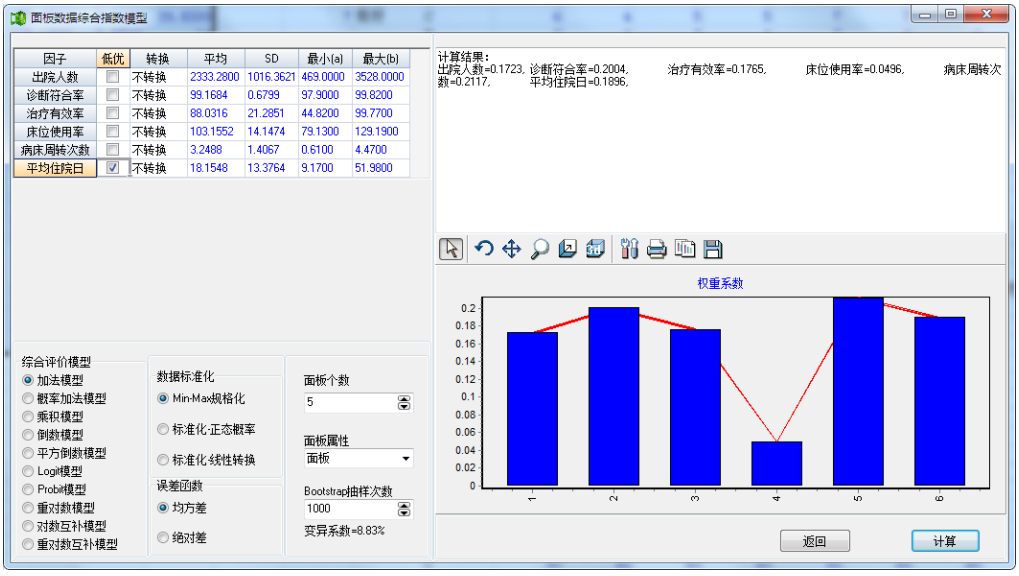
	A	B	C	D	E	F	G	H
1	年份	科室	出院人数	诊断符合率	治疗有效率	床位使用率	病床周转次数	平均住院日
2	2007	A	2282	99.22	98.22	83.65	2.84	14.36
3	2007	B	469	97.90	47.20	87.27	0.61	39.83
4	2007	C	2412	99.12	98.01	105.34	3.09	14.47
5	2007	D	3000	99.53	98.01	87.45	3.71	10.51
6	2007	E	2579	99.60	99.02	98.20	4.22	10.06
7	2008	A	1800	99.38	98.27	94.53	3.78	16.65
8	2008	B	530	97.90	47.20	87.27	0.61	39.83
9	2008	C	2856	99.78	98.82	129.19	3.97	14
10	2008	D	3072	99.54	98.04	102.08	3.88	9.64
11	2008	E	3048	98.98	97.89	79.13	4.32	9.69
12	2009	A	2400	99.34	98.26	93.17	4.30	13.99
13	2009	B	538	97.90	47.20	107.01	0.64	51.98
14	2009	C	2529	99.82	98.45	120.88	3.10	13.08
15	2009	D	3433	99.53	98.08	104.03	4.28	9.45
16	2009	E	3081	99.67	99.77	117.86	4.47	9.19
17	2010	A	2108	99.45	98.98	89.92	3.92	13.98
18	2010	B	540	97.91	44.82	101.08	0.68	42.21
19	2010	C	2532	99.71	98.93	120.08	3.25	13.37
20	2010	D	3436	99.55	98.11	107.29	4.35	9.19
21	2010	E	3184	99.76	98.12	99.91	4.21	9.67
22	2011	A	2736	99.12	98.01	96.06	4.17	12.50
23	2011	B	521	97.90	45.29	107.29	0.68	44.73
24	2011	C	2520	99.69	98.82	124.37	3.30	13.10
25	2011	D	3528	99.55	98.10	107.43	4.38	9.17
26	2011	E	3198	99.36	99.17	128.39	4.46	9.22
27								

图47-20　面板数据综合评价数据格式

然后执行“专业统计”—“综合评价及多属性决策”—“面板数据最大熵-最小二乘模型”功能，这时系统出现供用户对各个指标特性进行设置的对话框(图 47-21 左)。

这里的用户界面（图 47-2）相比，基本相同。只是增加了一个面板设置对话框。这里模型可以选相加型或乘积型；数据转换、数据标准化，以及误差函数的设置也相同。和前面的图这里可根据每个指标的特性，设置指标是否是“低优指标”、数据是否进行转换、或标准化。

在面板属性选择框中，面板、时期、地区、专家。这里是各个年份的面板数据，因此这里选“时期”。如果进行Bootstrap抽样，则在文本输入框中输入抽样次数（一般1000次以上）。



47-21 面板数据综合评价分析用户界面

然后点击”计算”按钮，应用多因子综合指数模型进行统计分析，得到简要结果，各个指标的权重系数直方图如图 47-21 右所示。计算完成输出结果如下:

相加型综合评价指数模型
数据 Min-Max 规格化处理

因子	是否低优	均值	标准差	最小值(a)	最大值(b)	是否 0-1 限制
出院人数	否	2333.2800	1016.3621	469	3528	否
诊断符合率	否	99.1684	0.6799	97.9000	99.8200	否
治疗有效率	否	88.0316	21.2851	44.8200	99.7700	否
床位使用率	否	103.1552	14.1474	79.1300	129.1900	否
病床周转次	否	3.2488	1.4067	0.6100	4.4700	否
平均住院日	是	18.1548	13.3764	9.1700	51.9800	否

熵值=2.487323
残差=0.021624
目标函数对数值=-9.5362

方差分析表

变异来源	平方和	自由度	均方	F 值	p 值
等权模型	0.1550	148	0.0010		
MEMR 模型	0.0902	9	0.0100	21.4829	0.0000
面板	0.0003	4	0.0001	0.1591	0.9586
因子	0.0899	5	0.0180	38.5419	0.0000
残差	0.0648	139	0.0005		

变量指标	权重系数
出院人数	0.1723

诊断符合率	0.2004
治疗有效率	0.1765
床位使用率	0.0496
病床周转次	0.2117
平均住院日	0.1896

各时期调整的权重系数

时期	时期 1	时期 2	时期 3	时期 4	时期 5
时期权重	0.9962	0.9936	1.0020	1.0043	0.9953
出院人数	0.1716	0.1712	0.1726	0.1730	0.1715
诊断符合率	0.1996	0.1991	0.2008	0.2012	0.1994
治疗有效率	0.1759	0.1754	0.1769	0.1773	0.1757
床位使用率	0.0494	0.0492	0.0497	0.0498	0.0493
病床周转次	0.2109	0.2104	0.2121	0.2126	0.2107
平均住院日	0.1888	0.1883	0.1899	0.1904	0.1887

不同时期综合指数及其均值

DUMs	时期 1	时期 2	时期 3	时期 4	时期 5	均值	指数模型值
A	0.7021	0.7419	0.8168	0.7919	0.8089	0.7723	0.7898
B	0.0693	0.0725	0.0409	0.0742	0.0680	0.0650	0.0637
C	0.7329	0.9003	0.8405	0.8388	0.8363	0.8297	0.8357
D	0.8422	0.8726	0.9242	0.9369	0.9356	0.9023	0.9186
E	0.8695	0.8139	0.9498	0.9275	0.9256	0.8973	0.9156

时期	时期 1 系数	时期 2 系数	时期 3 系数	时期 4 系数	时期 5 系数
指数模型权重	0.0866	0.1037	0.2434	0.2843	0.2820

Bootstrap 抽样 1000 次权重系数置信区间估计

因子	模型值	抽样值	标准差	中位数	95%置信区间	
出院人数	0.1723	0.1717	0.0187	0.1719	0.1014	0.2422
诊断符合率	0.2004	0.2003	0.0161	0.2006	0.1498	0.2417
治疗有效率	0.1765	0.1771	0.0102	0.1768	0.1499	0.2143
床位使用率	0.0496	0.0487	0.0077	0.0480	0.0284	0.0842
病床周转次数	0.2117	0.2143	0.0096	0.2139	0.1855	0.2527
平均住院日	0.1896	0.1880	0.0145	0.1884	0.1369	0.2329
时期 1	0.9954	0.9945	0.0101	0.9953	0.9523	1.0221
时期 2	0.9929	0.9959	0.0066	0.9962	0.9678	1.0135
时期 3	1.0026	1.0020	0.0077	1.0024	0.9696	1.0212
时期 4	1.0054	1.0036	0.0079	1.0043	0.9632	1.0229
时期 5	0.9950	0.9953	0.0072	0.9954	0.9533	1.0175
变异系数	7.6899					

系统首先输出最大熵-最小二乘非线性综合评价模型的信息熵 E 等于 2.487323、拟合残差标准差(σ=0.021624)、以及模型目标函数自然对数值为-9.5262。

模型输出结果的第二部分是模型的方差分析结果。本例中 MEMR 模型方差分析 F 值等于 21.4829，显著性检验 p<0.0001。因此这里的数据和采用等权法估计综合评价指

数差异极显著。其中面板间模型方差分析 F 值等于 0.1591，显著性检验 p=0.9586，无统计学意义。因此面板间的没有差异；因子间权重模型方差分析 F 值等于 38.5419，显著性检验 p<0.0001，极显著。说明各个因子的权重系数差别较大，不宜采用等权法估计综合评价指数。

输出结果第三部分为各个指标的权重系数，包括各个因子权重系数即各个截面的斜率系数。从各个因子的权重来看，病床周转次数的权重最大，其次是诊断符合率和平均住院日数。对于面板间，即这里的各个时期变化趋势的系数来看，各个年份的斜率系数均包含 1，因此不能说各个时期间的权重系数有某种趋势。王一任在原文章中对时间面板的权重估计是咨询专家的意见来确定的。

最后综合评价值，从不同科室的综合评价值来看，科室 D 和科室 E 得分最高，科室 B 得分最低。不同时期得分来看，有逐步上升的趋势，尤其是科室 D、科室 E 和科室 A。

例 2 多属性决策 考虑航天装备的评估问题，首先制定 8 个评估指标（属性），导弹预警能力(x_1)，成像侦察能力(x_2)，通信保障能力(x_3)，电子侦察能力(x_4)，卫星测绘能力(x_5)，导航定位能力(x_6)，海洋监测能力（x_7）和气象预报能力(x_8)。现由四位决策者 D_k（k=1,2,3,4）对每个型号装备的各项指标进行评估打分（得分从 0 到 100）。4 个专家的评估得分如图 47-22 所示。试确定最佳航天装备。

	A	B	C	D	E	F	G	H		I
1	专家	型号	导弹预警	成像侦察	通信保障	电子侦察	卫星测绘	导航定位	海洋监测	气象预报
2	D1	A	0.85	0.90	0.95	0.60	0.70	0.80	0.90	0.85
3	D1	B	0.95	0.80	0.60	0.70	0.90	0.85	0.80	0.70
4	D1	C	0.65	0.75	0.95	0.65	0.90	0.95	0.70	0.85
5	D1	D	0.75	0.75	0.50	0.65	0.95	0.75	0.85	0.80
6	D2	A	0.60	0.75	0.90	0.65	0.70	0.95	0.70	0.75
7	D2	B	0.85	0.60	0.60	0.65	0.90	0.75	0.95	0.70
8	D2	C	0.60	0.65	0.75	0.80	0.90	0.95	0.90	0.80
9	D2	D	0.65	0.60	0.60	0.70	0.90	0.85	0.70	0.65
10	D3	A	0.60	0.75	0.85	0.60	0.85	0.80	0.60	0.75
11	D3	B	0.80	0.75	0.60	0.90	0.85	0.65	0.85	0.80
12	D3	C	0.95	0.80	0.85	0.85	0.90	0.90	0.85	0.85
13	D3	D	0.60	0.65	0.50	0.60	0.95	0.80	0.65	0.70
14	D4	A	0.80	0.80	0.85	0.65	0.80	0.90	0.70	0.80
15	D4	B	0.85	0.70	0.70	0.80	0.95	0.70	0.85	0.85
16	D4	C	0.90	0.85	0.80	0.80	0.95	0.85	0.80	0.90
17	D4	D	0.65	0.70	0.60	0.65	0.90	0.85	0.70	0.75
18										

图47-22　面板数据综合评价数据格式

按图 47.22 方式编辑、选中数据之后，执行“专业统计”—“综合评价及多属性决策”—“面板数据最大熵-最小二乘模型”功能，这时系统出现供用户对各个指标特性进行设置的对话框(图 47-23 左)。

在这里选择模型相加型或乘积型；数据转换、数据标准化，以及误差函数，以及各个指标特性的设置（图 47-23 左）。主要这里是 100 分制得分，因此各指标统一设置最小值为 0，最大值为 100，进行数据的规格化处理。

47-23 面板数据综合评价分析用户界面

在面板属性选择框中，这里是各个专家的评价数据，因此这里选“专家”。如果进行 Bootstrap 抽样，则在文本输入框中输入抽样次数（一般 1000 次以上）。然后点击”计算”按钮，应用多因子综合指数模型进行统计分析，得到简要结果，各个指标的权重系数直方图如图 47-23 右所示。计算完成输出结果如下:

因子	是否低优	数据转换	均值	标准差	最小值(a)	最大值(b)
导弹预警	否	不转换	75.3125	13.0982	0	100
成像侦察	否	不转换	73.7500	8.4656	0	100
通信保障	否	不转换	72.5000	15.5991	0	100
电子侦察	否	不转换	70.3125	9.5688	0	100
卫星测绘	否	不转换	87.5000	7.9582	0	100
导航定位	否	不转换	83.1250	8.9209	0	100
海洋监测	否	不转换	78.1250	10.3078	0	100
气象预报	否	不转换	78.1250	7.0415	0	100

模型：加法模型

数据规格化：Min-Max 规格化

熵值=2.961847

残差=0.011941

目标函数对数值=-13.1144

方差分析表

变异来源	平方和	自由度	均方	F 值	p 值
等权模型	0.0213	126	0.0002		
MEMR 模型	0.0048	10	0.0005	3.3462	0.0007
面板	0.0003	3	0.0001	0.7051	0.5509
因子	0.0045	7	0.0006	4.4781	0.0002
残差	0.0165	116	0.0001		

各专家调整的权重系数

变量指标	权重系数	专家 1	专家 2	专家 3	专家 4
导弹预警	0.1237	0.1230	0.1213	0.1233	0.1239
成像侦察	0.1551	0.1542	0.1521	0.1546	0.1554
通信保障	0.0927	0.0921	0.0909	0.0924	0.0928
电子侦察	0.1196	0.1189	0.1172	0.1191	0.1198
卫星测绘	0.0889	0.0884	0.0872	0.0886	0.0891
导航定位	0.1066	0.1059	0.1045	0.1062	0.1067
海洋监测	0.1327	0.1320	0.1301	0.1323	0.1330
气象预报	0.1807	0.1797	0.1772	0.1801	0.1810

样本转换值及其综合指数

样本	导弹预警	成像侦察	通信保障	电子侦察	卫星测绘	导航定位	海洋监测	气象预报	综合指数
A	0.85	0.90	0.95	0.60	0.70	0.80	0.90	0.85	0.8202
B	0.95	0.80	0.60	0.70	0.90	0.85	0.80	0.70	0.7796
C	0.65	0.75	0.95	0.65	0.90	0.95	0.70	0.85	0.7856
D	0.75	0.75	0.50	0.65	0.95	0.75	0.85	0.80	0.7505
A	0.60	0.75	0.90	0.65	0.70	0.95	0.70	0.75	0.7291
B	0.85	0.60	0.60	0.65	0.90	0.75	0.95	0.70	0.7295
C	0.60	0.65	0.75	0.80	0.90	0.95	0.90	0.80	0.7701
D	0.65	0.60	0.60	0.70	0.90	0.85	0.70	0.65	0.6802
A	0.60	0.75	0.85	0.60	0.85	0.80	0.60	0.75	0.7146
B	0.80	0.75	0.60	0.90	0.85	0.65	0.85	0.80	0.7780
C	0.95	0.80	0.85	0.85	0.90	0.90	0.85	0.85	0.8614
D	0.60	0.65	0.50	0.60	0.95	0.80	0.65	0.70	0.6733
A	0.80	0.80	0.85	0.65	0.80	0.90	0.70	0.80	0.7854
B	0.85	0.70	0.70	0.80	0.95	0.70	0.85	0.85	0.8011
C	0.90	0.85	0.80	0.80	0.95	0.85	0.80	0.90	0.8583
D	0.65	0.70	0.60	0.65	0.90	0.85	0.70	0.75	0.7226

不同专家综合指数及其均值

DUMs	专家 1	专家 2	专家 3	专家 4	均值	指数模型值
A	0.8202	0.7291	0.7146	0.7854	0.7623	0.7626
B	0.7796	0.7295	0.7780	0.8011	0.7721	0.7736
C	0.7856	0.7701	0.8614	0.8583	0.8189	0.8215
D	0.7505	0.6802	0.6733	0.7226	0.7066	0.7068

专家	专家 1 系数	专家 2 系数	专家 3 系数	专家 4 系数
指数模型权重	0.2338	0.2333	0.2513	0.2816

Bootstrap 抽样 1000 次权重系数置信区间估计

因子	模型值	抽样值	标准差	中位数	95%置信区间	
导弹预警	0.1237	0.1232	0.0116	0.1227	0.0936	0.1682
成像侦察	0.1551	0.1559	0.0093	0.1556	0.1285	0.1876
通信保障	0.0927	0.0914	0.0086	0.0917	0.0644	0.1241
电子侦察	0.1196	0.1204	0.0137	0.1205	0.0783	0.1678
卫星测绘	0.0889	0.0879	0.0106	0.0872	0.0559	0.1275
导航定位	0.1066	0.1049	0.0113	0.1045	0.0744	0.1385
海洋监测	0.1327	0.1350	0.0144	0.1338	0.1005	0.1832
气象预报	0.1807	0.1811	0.0055	0.1810	0.1629	0.1994

专家 1	0.9941	0.9943	0.0055	0.9941	0.9731	1.0130
专家 2	0.9804	0.9806	0.0034	0.9804	0.9711	0.9941
专家 3	0.9965	0.9963	0.0036	0.9964	0.9847	1.0090
专家 4	1.0017	1.0014	0.0023	1.0014	0.9931	1.0079
变异系数	9.0952					

系统首先输出最大熵-最小二乘非线性含条件约束模型的信息熵 E 等于 2.961847。拟合残差标准差(σ=0.011941)以及模型目标函数自然对数值为-13.1144。

模型输出结果的第二部分是模型的方差分析结果。本例中 MEMR 模型方差分析 F 值等于 3.3462，显著性检验 p=0.0007(p<0.01)。因此这里的数据和采用等权法估计综合评价有极显著差异。其中面板间模型方差分析 F 值等于 0.7051，显著性检验 p=0.5509，无统计学意义。因此面板间的差异不大，即各位专家的评分的影响不显著；因子间权重模型方差分析 F 值等于 4.4781，显著性检验 p=0.0.0002（p<0.01），有极显著的统计学意义。说明各个因子的权重系数差别大，不宜采用简单的等权重方法进行评价。

输出结果第三部分为各个指标的权重系数，从各个因子权重系数来看，各个评价因子的权重系数从大到小依次是气象预报(0.1807)、成像侦察(0.1551)、海洋监测(0.1327) 、导弹预警(0.1237)、电子侦察(0.1196)、导航定位(0.1066)、通信保障(0.0927)和卫星测绘(0.0889)。

最后综合评价值，从不同方案综合评价值来看，方案 C 得分最高，其次是方案 B。再次的是型号 A，型号 D 得分最低。

47.12　广义熵-最小二乘模型

47.11.1 广义熵最小二乘模型

最大熵-最小二乘模型从根本上解决了综合评价及多属性决策建模及其权重系数估计的问题。前面章节，我们通过较多的例子展示了其模型在综合评价及多属性决策领域的合理性、可行性。

线性模型文献中常常介绍了自动模型构建算法，如向前选择、向后消除、所有子集回归以及各种组合等。其目的是希望自动生成基于一些协变量 x_1、x_2、…、x_p 来预测响应 y 的“良好”线性模型。“良好”通常以预测准确性来定义，但简洁性也是一个重要的标准：为了深入了解自变量和因变量的关系，大家更倾向于选择更为简约的模型。多因子综合评价分析也是一样，有时我们可能需要从众多的因子指标里面选取一些重要的因子，应用这些“重要”的因子来建立综合评价模型或进行多属性决策。

如果我们能将基于信息熵条件的最小二乘一般化，建立一个基于广义熵的最小二乘模型，那对我们考察综合评价或多属性决策体系各个因子权重系数的变化趋势，便于识别系统中的关键因子、或能突出某些“重要”因子的权重。即类似 LASSO 回归技术，从多个评价因子中挑选“主要”因子，建立较简单的模型。

作为最大熵最小二乘建模工具的一般化，我们这里提出了一种新的基于信息熵的最小二乘估计模型，广义熵最小残差(Generalized Entropy-Minimum reduls, GEMR)模型。

GEMR 在满足权重系数之和等于 1 的条件下，最小化残差平方和。广义熵最小二乘是不局限在最大信息熵的条件下进行最小二乘估计，而是在由研究工作者设定的信息熵的条件下进行最小二乘估计，产生的有一部分权重系数较大，另一些权重系数较小，或接近为零的结果。以方便研究工作者根据专业背景对模型阐释，从而给出了可解释的模型。即产生类似于子集选择的可解释模型。

在最大熵最小二乘模型（式 47.6）中的表达式 σ^E 中，σ为拟合残差，E 为信息熵值。为拓展模型，我们可以在模型的信息熵(E)项中增加一个指数项(λ)，λ取值在 0-1 之间，将式 47.6 改进为

$$\sigma=\sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{p}\left((y_i-x_{ij})w_j\right)^2}{(n-1)\cdot p}}$$

$$E=-\sum_{j=1}^{p}w_j\ln(w_j) \tag{47.64}$$

$$Q_{\min}=\sigma^{E^{\lambda}},\quad 0\le\lambda\le 1$$

$$\text{s.t}\sum_{j=1}^{p}w_j=1,\quad w_j\ge 0$$

在式 47.64 中，λ大于等于 0，且小于等于 1，其实质是调整信息熵的空间大小。如果λ趋近于 0，$E^{\lambda\to 0}=E^0=1$，即熵值项趋近于 1，优化模型退化为不受信息熵约束的，仅含σ的最小二乘模型。这时模型优化后必然是某个因子权重系数为 1，其它系数为零，残差σ=0。

当λ逐渐增大时，权重系数值逐步分摊到各个因子上面。λ=1 即为最大熵-最小二乘优化估计。如果λ趋于无穷大时，各因子权重系数趋于相等，即趋于简单均值模型。

因此随着λ值变化，各个因子权重系数亦表现变化。这里仍以前面技术成就指数各指标例子为例，式 47.64 中的λ的变化范围从 1.0 到 0.0，得到各个因子权重系数估计值变化趋势如图 47-24。

图 47.24 演示了在不同的信息熵的空间，进行最小二乘优化得到的多因子综合评价指数模型的各个因子的权重系数。因此式 47.64 充分地反映了多因子综合评价指数模型中权重系数、残差、熵值之间的关系，是一个基于广义熵的最小二乘多因子指标综合评价模型。

从图 47.24 可以看出，如果λ趋近于 0，系统内有一个指标的权重系数的值接近为 1，而其它因子权重系数接近与零。图中随着λ值的逐步增大，各个权重系数值逐步趋近于系数相等，p 个因子指标时，权重值趋近于 1/p。这和我们模型的假设的结论完全一致。

同时，最大熵最小二乘综合评价指数模型（式 47.7）是当λ=1 时的特例。从某种意义上来讲，是信息熵最大和拟合误差最小之间达到的一种动态平衡；或者说是在保留了尽可能多的因子指标信息的同时，体现了各个因子指标权重系数的特异性的辩证的统一。这正是多因子综合指数需要表达的涵义。

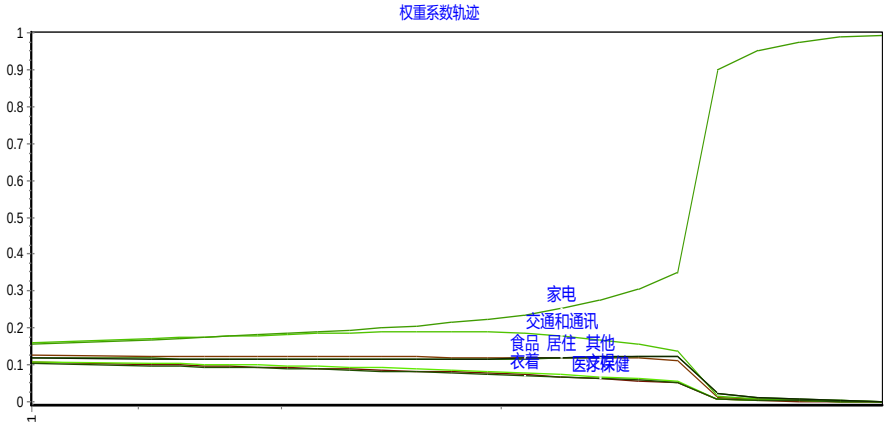


图 47.24 各因子权重系数变化趋势图

47.11.2 广义熵最小二乘模型

应用广义熵最小二乘模型(GEMR)模型进行综合评价及多属性决策分析，其数据格式、用户界面和最大熵最小二乘(MEMR)相同：一行一个样本，一列一个变量（因子）。在数据上面一行可以将各个因子的名称放入，这样输出结果更为直观。例如对我国 2007 年各地区农村居民家庭平均每人生活消费支出水平进行综合指数分析，将原始数据按一行一个地区、一列一个指标方式编辑如图 47-1 所示形式。

然后执行“专业统计”→“综合评价与多属性决策” →“广义熵最小二乘模型”功能，这时系统出现供用户对各个指标特性进行设置的对话框（图 47-25）。

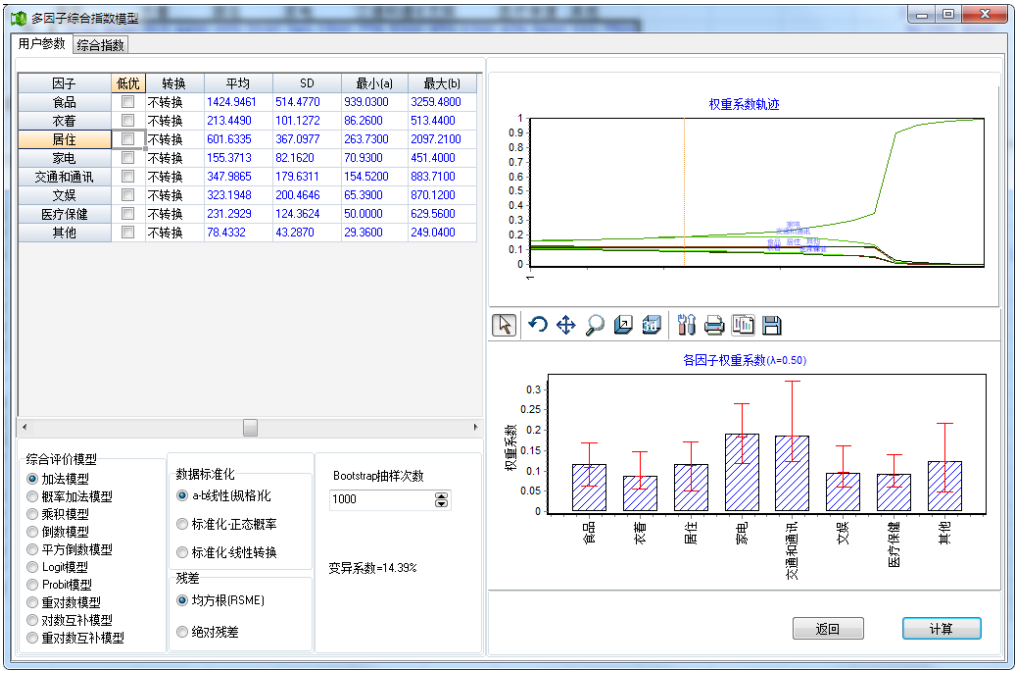


图 47-25　广义熵最小二乘用户界面

该界面看起来和图 47-2 类似。是的，最大熵最小二乘里面的用户设置，这里都有。因此我们可以用广义熵最小二乘模型来建立最大熵最小二乘模型。

但广义熵最小二乘界面左下方还有一个调整信息熵指数λ设定值的横向工具条。用鼠标左右移动工具条上面的按钮，即可随意地在 0-1 之间调整信息熵指数λ设定值。设定值确定之后，按下右下方的”计算”按钮，系统就会在当前λ值设定值进行广义熵最小二乘建模。

用户界面其他设置和前面的最大熵最小二乘建模分析里的设置相同。图 47-25 中，我们是将λ设置成 0.5 时进行优化建模。当λ设置成 0.5 时建模结果和最大熵最小二乘结果比较见表 47-11。

从表 47-11 可以看出，随着λ设定值变小，拟合结果的信息熵越来越小，但优化指数越来越高。各个因子的系数大小的相差愈来愈大，即因子权重越来越往一些“重要”的因子上聚集，即某些因子在综合评价或多属性决策过程中的重要性越来越大。通过 GEMR 模型，我们可以认识到哪些因子在综合评价或多属性决策中发挥的作用大小，以及随着信息熵的变化，因子权重系数的变化。

表 47-11　不同模型权重系数估计比较

因子	MEMR	GEMR (λ=0.50)	MEMR (λ=0.25)
食品	0.0985	0.1150	0.1220
衣着	0.0869	0.0874	0.0617
居住	0.1324	0.1155	0.1206
家电	0.1423	0.1895	0.2815
交通和通讯	0.1845	0.1849	0.1656
文娱	0.1281	0.0952	0.0668
医疗保健	0.0997	0.0905	0.0618
其他	0.1276	0.1221	0.1202
熵值	2.9625	2.9378	2.8067
优化指数(%)	17.90	18.73	29.04

从图 47-25 右上方权重系数轨迹，我们可以看出，随着λ设定值的变化，各因子权重系数的变化：当λ设定值增大时各因子权重值趋于最大熵时的各个因子权重系数等于 1/p。当λ设定值接近于 0 时，只有一个因子的权重系数为 1，其余均为 0。

GEMR 的结果输出，除给出最大熵最小二乘建模所有的结果外，还输出了权重系数随λ设定值的变化而变化的轨迹，由轨迹可以看出信息熵的变化对权重系数大小的影响。以及不同λ值改变时，综合指数的统计结果。这些结果可供用户进行比较分析。

47.11.3 多子组及面板数据广义熵最小二乘模型

根据广义熵最小二乘模型原理，和前面最大熵最小二乘多子组、面板数据最大熵最小二乘建模过程相同，多子组和面板数据一样可以建立广义熵最小二乘模型。

多子组和面板数据广义熵最小二乘的操作和相应的最大熵最小二乘模型建模过程类似，只是在用户界面左下边增加了设定信息熵指数λ值的工具条。

拖动按钮选择信息熵指数λ值后，点击”计算”按钮，系统即可给出当前用户需要的分析结果。

47.13　关于综合评价指数模型的讨论

多因子综合评价是现代社会中广泛应用的测度手段，多因子综合评价结果更是人类科学、理性决策的重要辅助工具。在多因子综合评价过程中，如何对各个因子指标赋权，在此之前一直是尚未解决的技术难题。公平、合理地计算因子权重，可以为其理性、公平决策提供科学、公允的理论依据；反之难以为决策主体提供客观的评价结果，并有可能因指数不合理而误导问题或引发争议。

作者基于最大熵-最小残差原理提出的在信息熵极大化、模型拟合残差极小化条件下的综合评价指数模型，实现了直接应用多因子权重和综合指数的关系估计的模型建立。这样在基于理论组分和观测值组分的最小二乘拟合模型，理论组分模型值和实际组分观察值之间贴近程度最高。通过数值优化技术得到的综合评价指数在最大程度上代表了各个因子的集中趋势和离散程度，因此最具有代表性。

信息熵尽可能大的条件下拟合残差尽可能小的优化过程，是保留尽可能多的信息前提下，提取多因子、多属性的共同特征。这从某种意义上来说，是采纳尽可能多的建议的基础上集中体现某种趋势，尽可能地综合各个因子所表达的共同点（求同存异）的过程。这符合综合评价和多属性决策的基本原则。

本文提出的综合指数模型，和目前一下常用方法相比，有如下特点：

这里提出的指数模型和等权法相比，总要比等权法的因子权重更好地反映各个因子指标的影响，更贴近综合评价指数所表达的实际意义。该模型在最次情况下，也具有和当前广泛应用的等权法一样的评价效果。

一般情形下，本模型产生的因子权重比主成分法得到的权重因子更能反映各个因子指标的权重。主成分分析法主要考虑的是各个因子之间的相关性，因子之间相关性小的情形下，主成分分析结果不是很理想；同时主成分赋权一直没有一个确定的，需要选取多少个主成分标准，主成分个数不同、综合评分结果也可能不一样。这时评价对象及其利益相关方则会对赋权的“合理性”有所质疑。

模拟例子结果显示，熵值法、标准离差法、CRITIC 权重等，主要是根据多因子的离散程度等特性进行权重系数的估计、没有考虑各个因子集中趋势对权重因子的影响，估计的权重系数，并不是基于尽可能考虑所有的多因子信息且提炼多因子集中趋势的综合评价模型计算出来的。这些赋权方法，没有考虑到各个因子、及其权重和估计出来综合指数之间的定量关系，更多地是从数据外表特征的某一方面，为了解释某个现象而对数据进行展示。这就像把数据作为花瓶一样、按某些需求向人们展示，并不是深刻地揭示数据所包含的信息，应用模型进行数值优化，提炼出的权重系数。其共同特征是，这些赋权技术仅考虑了参与评价的各个因子变异程度，忽略了各个因子的集中趋势，并不普适于所有综合评价过程。什么时候可用、什么时候不可用，亦没有一个取舍标准。应用 BoD 的进行综合评价时这一问题尤为显著。因此，目前文献介绍的所有赋权方法，不能适用于所有定量因子综合评价场合，失去了作为综合分析指数评价的普适性工具的条件。

这里提出的最大熵-最小二乘综合评价分析建模技术，得到的综合评价指数不仅反映

了各个因子集中趋势、而且反映了各个因子的离散程度。在多元统计分析领域，最大熵-最小二乘模型采用非线性优化技术，使得高维空间数据比主成分分享更有效地映射到一维空间。在各种情形下我们都可应用最大熵-最小二乘模型进行多元数据的降维分析。

本文提出的综合评价指数模型，是一个结合信息学和统计学的全新的研究领域，其理论研究有待进一步深入。作者认为该本文提出的综合评价指数模型在下述一些方面有待深入研究：

(1).最大熵-最小二乘综合评价指数模型参数估计算法及统计推断：包括优化现有的非线性数值方法和开发新的数值优化算法，提高数值优化过程的速度和稳健性；最大熵-最小二乘模型参数的推断准确性和置信区间的计算等。以及如何进行模型的假设检验，以及如何诊断模型拟合的好坏和残差的合理性的进一步研究。

(2).含属性变量的综合评价指数模型权重系数估计问题。

(3).综合评价指数模型选择与比较：如加法型模型和乘法型模型，如何选择最适合数据的模型类型，以及如何在不同模型之间进行比较和评估。

(4).本文提出的综合评价指数模型，亦可应用于多属性决策中权重估计。在多属性决策中，需要研究如何评估应用综合评价指数模型估计出来的各个因子权重在预测未来数据或新观测值时的前瞻性和准确性。

(5).综合评价指数模型应用的扩展，例如面板数据，时间序列数据等的综合评价指数模型的建模问题；不同领域,如医学、生态学、金融等不同领域的具体应用；考虑专家对每个因子给出权重的情形下模型的建立；以及如何根据具体领域的需求对综合评价指数模型进行定制和改进。

参 考 文 献

郭亚军，2007，综合评价理论、方法及应用[M]，北京：科学出版社

胡建军,马明,李耀光,俞长育, 2001, 烟叶主要化学指标与其感官质量的灰色关联分析[J], 烟草科技,1:3-7

宋逢明，陈涛涛,1999,高科技投资项目评价指标体系的研究[J],中国软科学，1:90-93

苏为华等，2021 年，综合评价基本理论与前沿问题研究[M]，北京：科学出版社

徐泽水，2004，不确定多属性决策方法及应用[M]，北京：清华大学出版社

Amos Golan and John Harte，2023. Information theory: A foundation for complexity science[J]. Proceedings of the National Academy of Sciences (PNAS), 119 (33) e2119089119. https://www. pnas.org/doi/10.1073/pnas.2119089119

Bandura, R. 2011. Composite indicators and rankings: Inventory 2011. Technical report, Office of Development Studies, United Nations Development Programme (UNDP), New York.

Cherchye L , Moesen W, Rogge N, et al.,2007. An Introduction to 'Benefit of the Doubt' Composite Indicators[J]. Social Indicators Research, 82(1):111-145.

Cherchye L, Moesen W, Rogge N, et al.,2008. Creating composite indicators with DEA and robustness analysis: the case of the Technology Achievement Index[J]. Journal of the Operational Research Society, 59(2): 239-251. DOI：10.2307/30132778

Diakoulaki, D., Mavrotas, G., & Papayannakis, L. (1995). Determining objective weights in multiple criteria problems: The critic method[J]. Computers & Operations Research, 22(7), 763-770. DOI：10.1016/0305- 0548(94)00059-H https://doi.org/10.1016/0305-0483(92)90021-X

Dunteman G. H. 1989. Principal Components Analysis. Newbury Park: Sage Publications.

Efron B. Bootstrap methods: another look at the jackknife[M]//Breakthroughs in statistics: Methodology and distribution. New York, NY: Springer New York, 1992: 569-593.

Fiacco A. V. and G. P. McCormick, 1964. Computational Algorithm for the Sequential Unconstrained Minimization Technique for Nonlinear Programming[J],Management Science, 10(4): 601-617. https://www.jstor.org/stable/2627508

Greco S., Ishizaka A., Tasiou M., Torrisi G.,2019. On the methodological framework of composite indices: A review of the issues of weighting, aggregation, and robustness[J], Soc Indic Res, 141 (1): 61-94. doi:10.1007/S11205-017-1832-9

Irik Mukhametzyanov, 2021. Specific character of objective methods for determining weights of criteria in MCDM problems: Entropy, CRITIC, SD[J].Decision Making Applications in Management and Engineering 30(2):2620-0104. DOI: 10.31181/dmame210402076i

Irik Z. Mukhametzyanov, 2023. Normalization of Multidimensional Data for Multi-Criteria Decision Making Problems-Inversion, Displacement, Asymmetry. New York, NY: Springer New York.

Jaynes E. T., 1957. Information Theory and Statistical Mechanics, The Physical Review, Vol. 106, no. 4, pp. 620 - 630. https:// doi.org/10.48550/arXiv.1105.5662 .

Jaynes, E. T., 1982. On the Rationale of Maximum-Entropy Methods[J], in Proc. IEEE, 70: 939-952. DOI：10.1109/PROC.1982.12425

Melyn W., Moesen W.,1994. Towards a synthetic indicator of macroeconomic performance: unequal weighting when limited information is available[J]. Public economics research papers, 1-24. https://lirias.kuleuven.be/handle/123456789/103555

OECD & Joint research centre，2008. Handbook on constructing composite indicators: methodology and user guide. Paris: OECD. https://www.istat.it/en/files/2014/06/Handbook-on-Constructing-Composite-Indicators.pdf

Sandra Huber,Martin Josef GeigerAdiel Teixeira de Almeida,2019. Multiple Criteria Decision Making and Aiding Cases on Models and Methods with Computer Implementations: Cases on Models and Methods with Computer Implementations

Shannon C. E.,1948. A mathematical theory of communication[J]. Bell Syst. Tech. J. 27, 379–423. http://dx.doi.org/10.1002/j.1538-7305.1948.tb01338.x

Sharpe, A., 2004. Literature review of frameworks for macro-indicators[M]. Ottawa: Centre for the Study of Living Standards.

Stiglitz, J. E., A. Sen, and J. Fitoussi，2009. Report by the commission on the measurement of economic performance and social progress. Technical report, www.stiglitz-sen-fitoussi.fr.

Tang, Q.-Y. and Zhang, C.-X. (2013), Data Processing System (DPS) software with experimental design, statistical analysis and data mining developed for use in entomological research. Insect Science. 20(2): 254-260. http://onlinelibrary.wiley.com/doi/10.1111/j.1744-7917.2012.01519.x/pdf

Tzeng G. H., Huang J. J., 2011. Multiple Attribute Decision Making: Methods and Applications. CRC Press, Taylor & Francis Group: USA,

United Nations, 2023. Human Development Report 2021/2022. United Kingdom: Oxford University Press. http://www.undp.org

Yang, L., 2014. An inventory of composite measures of human progress, Technical report, United Nations Development Programme Human Development Report Office. http://www.undp.org

A.T. de Almeida et al., 2015. Multicriteria and Multiobjective Models for Risk, Reliability and Maintenance Decision Analysis[M], International Series in Operations Research & Management Science 231, DOI 10.1007/978-3-319-17969-8_1dels Maintenance