

所属类别	2020 年“华数杯”全国大学生数学建模竞赛	参赛编号
本科组		CM11305

脱贫帮扶绩效评价

摘要

党的十八大以来，国家把扶贫开发工作纳入“四个全面”战略布局，作为实现第一个百年奋斗目标的重点工作，摆在更加突出的位置。因此，能否正确评价扶贫开发工作的绩效，对其后续工作的开展尤为重要。

本文依据题中所给数据解决了以下问题：

针对问题一：为了验证是否存在不同年份间对应各项评分有直接关联的规律。首先，在 6 个不同类型帮扶单位属性的基础上，通过 Pearson 相关性分析，分别建立指标相关系数矩阵（见图 1.1），得到显著性水平均小于 0.05，且系数均处在 0.5~0.7 之间，即任意两个指标间存在较为紧密的关系，验证了上述规律的存在。其次，以该矩阵为基础，得到不同类型帮扶单位的相关系数增长率（见表 1.2），建立被帮扶村庄成长稳定性模型，求得 gse （见表 1.3），即各指标间关系渐趋于线性相关，且为正相关。

针对问题二：首先，由表 1.2，得到帮扶绩效显著性得分（见表 2.1），并乘以对应扶贫单位类型规模比重，得到扶贫单位类型绩效综合得分，建立扶贫单位类型绩效评价模型。其次，计算每个村庄每个指标 2015-2020 年的增值后，使用熵值法求得村庄综合得分。并用每个帮扶单位类型的平均村庄综合得分，表示帮扶单位绩效综合得分，建立帮扶单位绩效评价模型。

针对问题三：为了找到单项评价排名前五的帮扶单位。本文根据问题一中的表，求出稳定性影响因子，并结合问题二中的 DPS，建立了特色指标绩效评价模型，分别求出了各项指标前五名的帮扶单位（见表 3.3）。

针对问题四：为了找到哪些因素对荣誉称号的获得有影响。本文首先自定义了相似度，并在相同帮扶单位中，找到与这 10 个村庄相似度最高的被帮扶村庄，后联立分式方程，建立基于相似度匹配的指标预测模型，求得缺失的数据。其次，以熵值法为基础，选取绩效综合得分，平均特性，单位类型分数，村庄基础，村庄现状为指标建立“脱贫先进”综合评价模型，得到综合得分前 10000 的村庄，以及各指标的权重 $w=0.1343, 0.0079, 0.4553, 0.1227, 0.2799$ 。所以一级、二级称号的得分标准为 0.545127738, 0.14658427。所以 10 个村庄中 39257 和 25149 号村庄可以评上“脱贫先进二级村庄”，无一人能评上“脱贫先进一级村庄”称号。

针对问题五：本文根据前四问探究所得的结果，给国家扶贫办写了一封信，向相关部门介绍了本文的研究成果，以助力国家制定更加合理的脱贫帮扶政策，并在信中阐述了相关观点及建议。

关键词： Pearson 系数 熵值法 相似度匹配及预测模型 粘合稳定性影响因子

一、 问题重述

1.1 背景

为了更好的激励各帮扶单位提高扶贫效率。五年前，国家启动了脱贫帮扶绩效评价机制。某科研团队接受任务后，对全国 32165 个需要帮扶的贫困村进行了初步的贫困调查。从居民收入(记为 SR)、产业发展(记为 CY)、居住环境(记为 HJ)、文化教育(记为 WJ)、基础设施(记为 SS)等五个评价指标给出了评分。以此为依据，将被帮扶的村庄划分为 160 个集合，每个集合指定帮扶单位（标记为 0-159）进行帮扶。这 160 个帮扶单位按照单位属性（如国企还是民营企业等）标记为 0-5 等 6 个类型。

2020 年，研究团队再次进行了调研，得到了被帮扶的这些村庄居民收入、产业发展、居住环境、文化教育、基础设施等五个方面的评分数据以及总分数据。为了便于比较和研究，所有数据都进行了标准化处理（标准化后的数值越大表示评分越高）。

1.2 需要解决的问题

问题一

本问题中五年前的评分与2020年对应的各项评分是否有着直接的关联，如是否存在五年前的居民收入不错，现在的居民收入也会是不错的规律。请分析各个评价指标的对应关系。

问题二

请运用附件的数据，阐明什么类型的帮扶单位，哪些帮扶单位在脱贫帮扶上面有较高的绩效？请给不同类型的帮扶单位绩效排序，给出脱贫帮扶绩效前十名的帮扶单位编号。

问题三

哪些帮扶单位分别在居民收入、产业发展、居住环境、文化教育、基础设施等评价指标上帮扶业绩明显？请列出各单项评价指标前五名的帮扶单位编号。

问题四

全国计划给予 10000 个村庄“脱贫先进村庄”称号。请问，哪些因素对获得这个荣誉称号有着非常重要的影响？数据表中最后有 10 个村庄的 2020 年的评价分数被删除，请你判断他们能评上“脱贫先进村庄”称号吗？如果称号分为一级和二级（一、二级称号比例为 1:3），这 10 个村庄中谁能评上“脱贫先进一级村庄”称号？

问题五

依据你的研究成果，向国家扶贫办写一封 500 字左右的信，阐述你的观点和建议。

二、 问题分析

2.1 问题一的分析

此问题要求我们分析五年前的各项评分是否与 2020 年对应的各项评分有直接关联，以及各个评价指标的对应关系。考虑到扶贫单位对各村庄的发展有较大影响，故可以转化为分析该村庄成长稳定性变化。若该村庄成长稳定性增加则判定为五年前的各项评分与 2020 年对应的各项评分有直接关联。

故本文需要建立一个被帮扶村庄成长稳定性分析模型，可以对不同指标间进行 Pearson 相关性分析。再利用其相关系数增长率，得到一个被帮扶村庄成长稳定性得分 gse ，来衡量该村庄未来发展的稳定性，得分大于 0 说明该村庄发展越趋于稳定，逐渐脱贫，即各指标间关系渐趋于线性相关，且为正相关。

2.2 问题二的分析

此问题要求我们分析什么类型的帮扶单位，哪些帮扶单位在脱贫帮扶上面有较高的绩效，并给出排序。考虑到仅用 2020 年各村庄评分高低无法有效体现一个帮扶单位的扶贫工作是否有较高绩效，故本问可继续沿用第一问的成长稳定性分析模型中的成长稳定性增长率，对每类的所有相关系数进行相加得到总分。根据总分排序可以得到每类帮扶单位的绩效排名。

对于每个帮扶单位的绩效排序我们需要先计算出每个帮扶单位的绩效。即需要求出该帮扶单位下每个贫困村庄的绩效，再取平均值得到该帮扶单位的总体绩效。可以利用熵值法对 5 项评价指标确定客观权重，计算出此帮扶单位下每个村庄五年间每项指标的差值，再用加权求和得出每个村庄的平均绩效值。

2.3 问题三的分析

此问题要求我们分析哪些帮扶单位分别在五种评价指标上帮扶业绩明显，即可以转化为求出每个贫困村庄的特色指标得分，考虑到帮扶数量对帮扶单位的帮扶业绩有较大影响，所以我们可以对其求和得到该帮扶单位的特色指标得分。

为此我们需要探究某一特色指标与其他指标间的成长稳定性增长率，考虑到不同种类帮扶单位对于村庄扶贫工作的侧重点不同，所以继续沿用问题一按照帮扶单位种类的分类评价指标值，以及成长稳定性分析模型中的成长稳定性增长率。可以找到某一指标与所有其他指标间的成长稳定性增长率平均值较高，则可以认定为该指标为这一类型的特色指标。结合问题二得到的每个村庄五年间每项指标的差值，进行一个加权求和可以得到该村庄的特色指标得分。

2.4 问题四的分析

此问题要求我们分析哪些因素对“脱贫先进村庄”有着非常重要的影响，以及预测未知的 10 个村庄 2020 年的评价分数，并判断他们中哪些能获得“脱贫先进村庄”一级称号。考虑到贫困村庄的脱贫工作与帮扶单位有着很大的关联，因此我们可以找到与未知村庄同一帮扶单位的其他村庄，并通过计算各项 2015 年指标间的差值绝对值之和，来匹配到同一帮扶单位下与未知村庄 2015 年各项指标最符合的已知村庄。再通过比值法预测未知村庄 2020 年各项指标值。

考虑到帮扶单位、特色指标以及 2015 和 2020 年总分多方面因素对评选“脱贫先进村庄”的影响，我们可以以村庄特色指标得分、帮扶单位绩效、帮扶单位类型绩效、2015

年总分、2020 年总分这 5 个因素来确定最终该村庄用以评选“脱贫先进村庄”的荣誉值，再对这 5 个因素使用熵值法确定客观权重，得到这些因素对“脱贫先进村庄”评选的影响程度，并且根据加权计算出该村庄的荣誉值，按荣誉值排名确定出一级、二级荣誉分别所需要的最低荣誉值，来判断最后十个未知村庄中谁能评上“脱贫先进一级村庄”。

2.5 问题五的分析

本问要求根据前四问的研究结果，向国家扶贫办写一封信。本质上就是对前四问问题的一个总结性概括，因此本问决定利用前四问分析中出现的问题，以及解决的方法思路，向国家扶贫办介绍本问的研究结果和对未来扶贫工作的一些建议。

三、模型假设

1. 假设所有扶贫单位的扶贫措施都是真实有效，落实到位的，不存在人为干扰。
原因：人为阻断扶贫资金或物资将影响扶贫措施落实到位
2. 假设扶贫过程中，被帮扶地区无自然灾害，公共安全事故发生。
原因：例如疫情，地震，洪水等非人为因素可能阻碍扶贫绩效增长或对其造成实质性的打击。
3. 假设所有扶贫单位对其负责村庄的实际帮扶时间相差不大，不存在两个月以上的差距。
原因：帮扶时间的长短可能会直接影响到各指标的绩效，需确保时间统一。
4. 假设帮扶过程中，被帮扶村庄都积极配合帮扶单位工作。
原因：当地村民或村干部的不配合可能会放缓扶贫进度而直接影响扶贫绩效。

四、符号说明

符号	符号说明
r_{ij}	第i类帮扶单位类型第j对指标间的相关系数
$r = (r_{ij})_{m \times n}$	评价指标相关系数矩阵
S_{ij}	第i类帮扶单位类型第j对指标间相关系数5年间的增长率
gse_i	被帮扶村庄成长稳定性增长率熵值，即第i类帮扶单位类型所有指标间相关系数增长率均值
$Aess_i$	帮扶绩效显著性得分，即第i类帮扶单位类型所有指标间相关系数增长率之和
$x = (x_{ij})_{p \times q}$	已知村庄评价指标数据矩阵
Dps_{ij}	直接绩效分值，即第i个村庄第j项指标五年间差值
Dps_w_j	直接绩效分值第j项指标的熵权
V_Pscore_i	村庄绩效综合得分，即第i个村庄的所有直接绩效分（ Dps_{ij} ）加权（ w_j ）后总分
Pau_Pscore	扶贫单位绩效综合得分，即该扶贫单位所扶贫村庄的绩效综合得分（ V_Pscore_i ）的平均值
$Paut_Pscore$	扶贫单位类型绩效综合得分，即该扶贫单位类型帮扶绩效显著性得分（ $Aess_i$ ）和所帮扶村庄占比的乘积
$num1$	表示该类型扶贫单位总共帮扶的村庄个数。
Asf_{ik}	粘合稳定性影响因子，即类型i类扶贫单位下第k个指标与其他指标之间的成长稳定性增长率平均值
Vci_Sscore_{hk}	村庄特色指标显著性得分，即粘合稳定性影响因子与直接绩效分值的乘积，中h表示当前帮扶单位所帮扶的村庄序号，k为指标序号
$Pauci_Sscore_{dk}$	扶贫单位特色指标显著性得分，即每个单位对应村庄特色指标显著性得分相加之和，d为扶贫单位序号
$y = (y_{ij})_{t \times q}$	未知村庄评价指标数据矩阵
$Vsdv_i$	村庄相似度差异值，即已知村庄与第i个未知村庄间每项评价指标的绝对差值之和
Vci_Ssmean_h	村庄特色指标显著性得分均值，每个村庄的所有特色指标显著性得分均值
$Vhvh_h$	村庄荣誉值

注：部分说明见文中

五、模型的建立与求解

5.1 问题一模型的建立与求解

5.1.1 数据的处理

考虑到不同类型的帮扶单位对不同村庄的帮扶绩效不同，直接进行相关性分析，可能与真实值会有较大的偏差。所以，本文按照题干中的 6 类帮扶单位类型，分别进行 Pearson 相关性分析。

5.1.2 模型建立

1. Pearson 相关性分析：

用于度量两个变量 X 和 Y 之间的相关程度（线性相关），其值介于-1 与 1 之间。两个变量之间的皮尔逊相关系数定义为两个变量之间的协方差和标准差的商：

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} \quad (1)$$

上式定义了总体相关系数，常用希腊小写字母 ρ 作为代表符号。估算样本的协方差和标准差，可得到样本相关系数。

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (2)$$

2. 模型一：建立被帮扶村庄成长稳定性模型

对 6 个不同类型的指标分别进行 Pearson 相关性分析，得到 6 个相关系数矩阵，以类型 0 为例（见图 1）

Correlations										
	2015 SR	2015 CY	2015 HJ	2015 WJ	2015 SS	2020 SR	2020 CY	2020 HJ	2020 WJ	2020 SS
2015 SR	1	.545**	.592**	.566**	.571**	.554**	.528**	.618**	.465**	.490**
Pearson Correlation		.000	.000	.000	.000	.000	.000	.000	.000	.000
Sig. (2-tailed)										
N	2483	2483	2483	2483	2483	2483	2483	2483	2483	2483
2015 CY		1	.656**	.734**	.707**	.530**	.716**	.673**	.619**	.620**
Pearson Correlation			.000	.000	.000	.000	.000	.000	.000	.000
Sig. (2-tailed)										
N		2483	2483	2483	2483	2483	2483	2483	2483	2483
2015 HJ			1	.666**	.654**	.573**	.617**	.803**	.541**	.565**
Pearson Correlation				.000	.000	.000	.000	.000	.000	.000
Sig. (2-tailed)										
N			2483	2483	2483	2483	2483	2483	2483	2483
2015 WJ				1	.783**	.555**	.709**	.688**	.665**	.660**
Pearson Correlation					.000	.000	.000	.000	.000	.000
Sig. (2-tailed)										
N				2483	2483	2483	2483	2483	2483	2483
2015 SS					1	.553**	.692**	.679**	.626**	.643**
Pearson Correlation						.000	.000	.000	.000	.000
Sig. (2-tailed)										
N					2483	2483	2483	2483	2483	2483
2020 SR						1	.603**	.674**	.559**	.588**
Pearson Correlation							.000	.000	.000	.000
Sig. (2-tailed)										
N						2483	2483	2483	2483	2483
2020 CY							1	.753**	.767**	.768**
Pearson Correlation								.000	.000	.000
Sig. (2-tailed)										
N							2483	2483	2483	2483
2020 HJ								1	.710**	.742**
Pearson Correlation									.000	.000
Sig. (2-tailed)										
N								2483	2483	2483
2020 WJ									1	.773**
Pearson Correlation										.000
Sig. (2-tailed)										
N									2483	2483
2020 SS										1
Pearson Correlation										.000
Sig. (2-tailed)										
N										2483

** . Correlation is significant at the 0.01 level (2-tailed).

图 1 类型 0 的 Pearson 相关性分析结果

由于每个矩阵都包含 2015 年和 2020 年指标的相关性系数，因此每个矩阵都可以转化为 10 个不重复的相关系数增长率，以类型 0 为例（见图 2）

表 1.1 类型 0 的相关系数增长率

相关系数	2015年	2020年	增长率 (s _n)
r ₀₁ (SR_CY)	0.545	0.603	0.106
r ₀₂ (SR_HJ)	0.592	0.674	0.139
r ₀₃ (SR_WJ)	0.566	0.559	-0.012
r ₀₄ (SR_SS)	0.571	0.588	0.03
r ₀₅ (CY_HJ)	0.656	0.753	0.148
r ₀₆ (CY_WJ)	0.734	0.767	0.045
r ₀₇ (CY_SS)	0.707	0.768	0.086
r ₀₈ (HJ_WJ)	0.666	0.71	0.066
r ₀₉ (HJ_SS)	0.654	0.742	0.135
r ₀₁₀ (WJ_SS)	0.783	0.773	-0.013

而有 6 类帮扶单位类型，每一类都有 10 个相关性系数增长率，所以将 6 个不同类型的增长率放到一个矩阵中（见图 2）

表 1.2 不同类型帮扶单位的相关系数增长率

S _{ij}	类型5	类型4	类型3	类型2	类型1	类型0
S _{i1}	0.617	0.38	0.599	0.341	0.706	0.106
S _{i2}	0.295	0.806	0.291	0.144	0.371	0.139
S _{i3}	0.43	2.228	0.403	0.247	0.586	-0.012
S _{i4}	0.442	4.561	0.365	0.274	0.635	0.03
S _{i5}	0.464	2	0.425	0.351	0.499	0.148
S _{i6}	0.281	1.028	0.285	0.193	0.469	0.045
S _{i7}	0.302	6.581	0.261	0.211	0.49	0.086
S _{i8}	0.631	9.012	0.396	0.236	0.395	0.066
S _{i9}	0.365	4.818	0.271	0.219	0.419	0.135
S _{i10}	0.092	0.045	0.108	0.085	0.269	-0.013

我们得到了一系列相关系数的增长率。当增长率大于 0，说明该类贫困村庄中，这两个指标的相关性在逐渐变大，即这两个指标渐趋于线性相关，且为正相关；反之，则为负线性相关。也可以理解为该贫困村这两项指标的稳定性逐渐提升，以及对应产业正以一个较快的速度发展。

设被帮扶单位成长稳定性得分

$$gse_i = \frac{1}{n} \times \sum_{j=1}^m gse_{ij}, \quad i = 1, 2, 3, 4, 5, 6 \quad (3)$$

得到

表 1.3 帮扶单位成长稳定性得分 (gse)

相关系数增长率均值 (gse)
0.458
0.341
0.647
1.051
0.648
0.384
1.322
1.789
1.038
0.098

结论:

1. 通过 Pearson 相关性分析得到的 Pearson 系数均处在区间[0.5, 0.7], 且显著性水平均小于 0.05, 所以可以说 2020 年各项指标与 2015 年存在直接关联且较为紧密。
2. 通过被帮扶村庄成长稳定性模型计算得出, *gse*均大于 0, 因此可以说各个指标间存在着渐趋于正线性相关的关系。

5.2 问题二

5.2.1 模型建立

模型二: 扶贫单位类型绩效评价模型

在问题一的基础上, 利用相关系数增长率矩阵表建立扶贫单位类型绩效评价模型。求出该类型相关系数增长率之和, 定义为帮扶绩效显著性得分 (Significance score of assistance effect), 用来衡量该类型扶贫单位 5 年来的扶贫工作为贫困村庄带来的发展程度。符号用 Aess 来表示:

$$Aess_i = \sum_{j=1}^n S_{ij} \quad (4)$$

模型三: 扶贫单位绩效评价模型

Step1 计算直接绩效分值

对于每个帮扶单位的绩效排序我们需要先计算出每个帮扶单位的绩效, 本文将问题转化为求出该帮扶单位下每个贫困村庄的绩效。首先设该扶贫单位下有 p 个被扶贫村庄, q 项评价指标, 形成原始值表数据矩阵 $x = (x_{ij})_{p \times q}$, ($j \leq 5$) i 代表村庄, j 表示评价指标。通过评价指标值计算出此帮扶单位下每个村庄五年间每项指标的差值, 定义为直接绩效分值 (Direct performance score), 符号用 Dps 来表示:

$$Dps_{ij} = x_{ij2020} - x_{ij2015} \quad (5)$$

Step2 熵值法确定客观权重

考虑到每个指标对整体绩效的影响力度不同, 我们利用熵值法确定个评价指标的客

观权重，再进行加权计算得到最终绩效得分。熵值法是一种客观赋权法。熵 (Entropy) 是利用概率论来确定信息不确定性的一个量度。对于某项指标 x ，指标值 x_{ij} 的差距越大，则该指标在综合评价中所起的作用越大；如果某项指标的指标值全部相等，则该指标在综合评价中几乎不起作用。给定一系列的方案和属性，熵值法能够确定出各属性的客观权重值。

根据公式 (9) 对指标进行标准化及非负化处理。

$$p_{ij} = (x_{ij} - \bar{x}_j) / Sd_j \quad (6)$$

其中 \bar{x}_j 为第 j 项指标的均值， Sd_j 为第 j 项指标的标准差。

再计算第 j 项指标的熵值 e_j 和信息熵冗余度 d_j

$$\begin{aligned} e_j &= -k \sum_{i=1}^n p_{ij} \ln(p_{ij}) \\ d_j &= 1 - e_j \end{aligned} \quad (7)$$

式中常数 $k = \frac{1}{\ln(n)}$ ，满足 $e_j \geq 0$ ， d 是第 j 项指标的信息效用值。

$$Dps_w_j = \frac{d_j}{\sum_{i=1}^n d_j} = \frac{1 - e_j}{n - \sum_{i=1}^n e_j} \quad (8)$$

Dps_w_j 代表第 j 项指标的熵权。

Step3 计算扶贫单位绩效综合得分

故本问可继续沿用第一问的成长稳定性分析模型中的成长稳定性增长率，并在此基础上利用熵值法对 5 项评价指标确定客观权重，进行加权计算出该村庄绩效综合得分 (Comprehensive village performance score)，符号用 V_Pscore 表示：

$$V_Pscore_i = \sum_{j=1}^q (Dps_{ij} \times Dps_w_j) \quad (9)$$

按每类帮扶单位分类对村庄绩效综合得分求平均，则可以得到每个扶贫单位绩效综合得分 (Comprehensive performance score of poverty alleviation unit)，以此衡量该扶贫单位 5 年来所有扶贫工作的总体绩效，符号用 Pau_Pscore 表示：

$$Pau_Pscore = \frac{\sum_{i=1}^p Vps_i}{p} \quad (10)$$

按照扶贫单位绩效综合得分对各扶贫单位进行绩效排序，可以得到脱贫帮扶绩效前十名的帮扶单位编号。

5.2.2 模型求解

模型二：扶贫单位类型绩效评价模型

得到表 2.1 帮扶绩效显著性得分

表 2.1 相关系数增长率矩阵表及其帮扶绩效显著性得分

Sij	类型5	类型4	类型3	类型2	类型1	类型0
S ₁₁	0.617	0.38	0.599	0.341	0.706	0.106
S ₁₂	0.295	0.806	0.291	0.144	0.371	0.139
S ₁₃	0.43	2.228	0.403	0.247	0.586	-0.012
S ₁₄	0.442	4.561	0.365	0.274	0.635	0.03
S ₁₅	0.464	2	0.425	0.351	0.499	0.148
S ₁₆	0.281	1.028	0.285	0.193	0.469	0.045
S ₁₇	0.302	6.581	0.261	0.211	0.49	0.086
S ₁₈	0.631	9.012	0.396	0.236	0.395	0.066
S ₁₉	0.365	4.818	0.271	0.219	0.419	0.135
S ₁₁₀	0.092	0.045	0.108	0.085	0.269	-0.013
帮扶绩效显著性得分 (A _{ess})	3.919	31.459	3.404	2.301	4.839	0.73
帮扶绩效显著性得分排名	3	1	4	5	2	6

由上表可以得到类型 4 的帮扶单位绩效是最高的，且远超其他类型，类型 0 的帮扶单位绩效是最低的，且远低于其他类型。

模型三：扶贫单位绩效评价模型

Step1 熵值法确定客观权重

通过 MATLAB 根据式 (9) (10) (11) 对已有数据进行分析得到权重比值，如表 2.2:

表 2.2 评价指标权重比值

	居民收入 (SR)	产业发展 (CY)	居住环境 (HJ)	文化教育 (WJ)	基础设施 (SS)
权重	0.1354	0.1704	0.2169	0.261	0.2162

Step2 计算扶贫单位绩效综合得分

根据式 (8) (12) (13) 得到每个扶贫单位的绩效综合得分，如表 2.3:

表 2.3 扶贫单位的绩效综合得分排序

排名	帮扶单位(0-159)	绩效综合平均得分	帮扶单位类型 (0-5)
1	113	0.702450648	3
2	149	0.700651573	5
3	158	0.660086765	5
4	153	0.579445738	5
5	115	0.488880787	5
6	131	0.438128884	3
7	155	0.382185235	5
8	54	0.362884041	2
9	132	0.343087957	5
10	45	0.339698075	2
.....		
159	20	-0.72814344	2
160	100	-0.779989071	5

由上表可以得到脱贫帮扶绩效前十名的帮扶单位，编号为：113、149、158、153、115、131、155、54、132、45。

5.2.3 模型二优化

由模型三扶贫单位类型绩效评价模型所得出的帮扶绩效显著性得分结果来分析，可以明显看出各类得分差异较大。经过查看原始数据后发现，帮扶绩效显著性得分最高的类型4帮扶单位总共被帮扶的村庄只有29个，只占总共被帮扶村庄的0.09%。基于此现象我们提出问题，若只考虑帮扶绩效显著性得分是否忽略了被帮扶村庄数量对扶贫单位绩效的影响。

基于此想法，我们进一步优化模型，将该类型单位所帮扶村庄所占总村庄的百分比计算出来，再结合原模型得到的帮扶绩效显著性得分计算出最终的扶贫单位类型绩效综合得分（Comprehensive performance score of poverty alleviation unit type），以此衡量该类型扶贫单位5年来的扶贫工作为贫困村庄带来的发展程度。符号用 Paut_Pscore 来表示：

$$Paut_Pscore = \frac{num1}{32165} * Pau_Pscore \quad (11)$$

num1 表示该类型扶贫单位总共帮扶的村庄个数。

最终计算得出扶贫单位类型绩效综合得分，如表 2.4：

表 2.4 扶贫单位类型绩效综合得分排序

帮扶单位类型	5	4	3	2	1	0
帮扶绩效显著性得分 (Aess)	3.919	31.459	3.404	2.301	4.839	0.73
帮扶村庄数量 (num)	3741	29	11280	2133	12499	2483
帮扶村庄所占百分比	11.63%	0.09%	35.07%	6.63%	38.86%	7.72%
扶贫单位类型绩效综合得分	0.4558	0.0284	1.1938	0.1526	1.8804	0.0564
扶贫单位类型绩效综合得分排名	3	6	2	4	1	5

由上表可以得到类型1的帮扶单位绩效是最高的，类型4的帮扶单位绩效是最低的，

相比于之前的排名，此次模型优化消除了被帮扶村庄的数量对帮扶单位绩效的影响，得分结果更加可靠，符合实际。

5.3 问题三

5.3.1 模型的建立

模型四：基于粘合稳定性影响因子的特色指标绩效评价模型

Step1：计算粘合稳定性影响因子

沿用问题一得到的相关系数增长率（见表 1.3），对每一个指标都建立一个相关系数增长率矩阵，可以得到这一指标与其他指标之间的成长稳定性增长率平均值，若值越大，代表该单位以这一特色指标进行扶贫的帮扶业绩更明显。定义该值为粘合稳定性影响因子（Adhesive stability factor），符号用 Asf_{ik} 表示，考虑到不同种类帮扶单位对于村庄扶贫工作的侧重点不同，所以继续沿用问题一按照帮扶单位种类的分类评价指标值。i 为 0-5 类扶贫单位，k 为指标序号：

$$\begin{aligned}
 Asf_{i1} &= S_{ij_{SR_CY}} + S_{ij_{SR_HJ}} + S_{ij_{SR_WJ}} + S_{ij_{SR_SS}} \\
 Asf_{i2} &= S_{ij_{CY_SR}} + S_{ij_{CY_HJ}} + S_{ij_{CY_WJ}} + S_{ij_{CY_SS}} \\
 Asf_{i3} &= S_{ij_{HJ_SR}} + S_{ij_{HJ_CY}} + S_{ij_{HJ_WJ}} + S_{ij_{HJ_SS}} \\
 Asf_{i4} &= S_{ij_{WJ_SR}} + S_{ij_{WJ_CY}} + S_{ij_{WJ_HJ}} + S_{ij_{WJ_SS}} \\
 Asf_{i5} &= S_{ij_{SS_SR}} + S_{ij_{SS_CY}} + S_{ij_{SS_HJ}} + S_{ij_{SS_WJ}}
 \end{aligned} \tag{12}$$

其中，j 表示每两个评价指标之间的相关关系。 Asf_{i1} 表示第 i 类扶贫单位的居民收入（SR）粘合稳定性影响因子，其他 Asf_{i2} 、 Asf_{i3} 、 Asf_{i4} 、 Asf_{i5} 含义依次为其他四种评价指标的粘合稳定性影响因子。

Step2：计算村庄特色指标显著性得分：

结合使用问题二得出的直接绩效分值，计算出村庄特色指标显著性得分（Significance score of village characteristic index），以此衡量该村庄 5 年扶贫工作中得到较大提升的指标得分，符号用 Vci_Score_{hk} 表示，其中 h 表示当前帮扶单位所帮扶的村庄序号，k 为指标序号：

$$Vci_Score_{hk} = Asf_{ik} * Dps_{hk} \tag{13}$$

表示第 h 个村庄第 k 个指标的特色指标显著性得分。

Step3：计算扶贫单位特色指标显著性得分：

考虑到帮扶数量对帮扶单位的帮扶业绩有较大影响，因此计算出每个单位对应村庄特色指标显著性得分相加之和，以此衡量该扶贫单位 5 年扶贫工作中得到较大提升的指标得分，定义为扶贫单位特色指标显著性得分（Significance score of characteristic

indicators of poverty alleviation units), 符号用 Pauci_Sscore_{dk} 表示, 其中 d 为扶贫单位序号:

$$\text{Pauci_Sscore}_{dk} = \sum_{h=1}^{\text{num2}} \text{Vci_Sscore}_{hk} \quad (14)$$

其中, num2 为该扶贫单位总共帮扶的村庄个数。 Pauci_Sscore_{dk} 表示第 d 个扶贫单位第 k 个特色指标显著性得分。对 5 个评价指标的 Pauci_Sscore 进行排序, 得到每个指标最高的前 5 个帮扶单位的编号。由此可以知道哪些帮扶单位分别在居民收入、产业发展、居住环境、文化教育、基础设施等评价指标上帮扶业绩明显。

5.3.2 模型的求解

Step1: 计算粘合稳定性影响因子

由问题一得出的表 1.3 相关系数增长率矩阵可以变化得到该指标和其他指标相关系数矩阵表, 根据式 (15) 计算出这一指标与其他指标之间的成长稳定性增长率平均值, 即粘合稳定性影响因子, 见表 3.1 (以居民收入 SR 指标为例):

表 3.1 SR 指标成长稳定性增长率矩阵及其粘合稳定性影响因子

单位类型	5	4	3	2	1	0
S_{SR_CY}	0.617	0.38	0.599	0.341	0.706	0.106
S_{SR_HJ}	0.295	0.806	0.291	0.144	0.371	0.139
S_{SR_WJ}	0.43	2.228	0.403	0.247	0.586	-0.012
S_{SR_SS}	0.442	4.561	0.365	0.274	0.635	0.03
Asf_{SR}	0.446	1.994	0.415	0.252	0.575	0.066

其他 4 个指标的粘合稳定性影响因子及其粘合稳定性影响因子见附录表 1

Step2: 计算扶贫单位特色指标显著性得分:

由问题二中模型四得到的直接绩效分值 Dps 联合式 (13) (14) 进一步得到帮扶单位的特色指标显著性得分, 见表 3.2:

表 3.2 部分帮扶单位特色指标显著性得分

帮扶单位(0-159)	SR特性	CY特性	HJ特性	WJ特性	SS特性
115	88.5938	44.7004	31.8644	53.0816	34.8038
130	69.8509	-4.6631	23.3573	47.65	16.3205
111	56.0372	-8.1206	-3.395	-54.2762	-66.6339
...
47	-83.6102	-65.0422	-22.7138	-63.3166	-66.6534

对每项特色指标显著性进行排序得到表 3.3:

表 3.3 各指标显著性得分前五帮扶单位编号

名次	SR显著性得分 单位编号	CY显著性得分 单位编号	HJ显著性得分 单位编号	WJ显著性得分 单位编号	SS显著性得分 单位编号
1	115	115	46	115	115
2	130	116	89	116	156
3	111	113	138	130	116
4	134	146	75	156	131
5	146	114	129	113	133

从该表中可以得出单项评价指标前五名的帮扶单位编号。

5.4 问题四

5.4.1 模型的建立

模型五：基于村庄相似度匹配的指标预测模型

Step1: 计算村庄相似度差异值

预测村庄 2020 年评价指标值的一个基本原则是找到已知村庄中与未知村庄 2015 年扶贫前基本情况最相似的，以差异度最小村庄 2020 年评价指标来预测未知村庄 2020 年评价指标。考虑到贫困村庄的脱贫工作与帮扶单位的扶贫工作有着很大的关联，因此本文对差异度最小的村庄搜索范围缩小到与未知村庄同一帮扶单位的其他村庄。

继续沿用问题二中的已知村庄评价指标数据矩阵 $x = (x_{ij})_{p \times q}$ ，（ $j \leq 5$ ），设有 t 个未知村庄，则可形成未知村庄评价指标数据矩阵 $y = (y_{ij})_{t \times q}$ ，（ $i \leq 10, j \leq 5$ ），其中 i 代表村庄， j 表示评价指标。本文定义了一个村庄相似度差异值（Village similarity difference value），即已知村庄与未知村庄间每项评价指标的绝对差值之和，符号用 $Vsdv_i$ 表示：

$$Vsdv_i = \sum_{j=1}^q |x_{ij} - y_{ij}| \quad (15)$$

Step2: 比值法求未知村庄评价指标预测值

匹配到每个未知村庄最相似的已知村庄后，利用已知村庄的 2015 年与 2020 年评价指标值和未知村庄的 2015 年评价指标值通过比值法求出未知村庄 2020 年评价指标预测值 y_{ij2020} ：

$$\frac{x_{ij2015}}{x_{ij2020}} = \frac{y_{ij2015}}{y_{ij2020}} \quad (16)$$

其中， x_{ij2015} 为已知村庄的 2015 年评价指标值， x_{ij2020} 为已知村庄的 2020 年评价指标值， y_{ij2015} 为未知村庄的 2015 年评价指标值。

模型六：多因素脱贫先进村庄综合评价模型

Step1：熵值法确定多因素权重

考虑到帮扶单位、特色指标以及 2015 和 2020 年总分多方面因素对评选“脱贫先进村庄”的影响，本文使用问题二和问题三得到的：村庄特色指标显著性得分、扶贫单位类型绩效综合评分、扶贫单位绩效综合平均得分结合原数据的 2015 年、2020 年总分共 5 个因素来衡量该村庄是否可以评选“脱贫先进村庄”称号，其中村庄各项特色指标显著性得分转化为村庄特色指标显著性得分均值，符号用 Vci_Ssmean_h 表示，h 表示为村庄序号：

$$Vci_Ssmean_h = \frac{Vci_Score_{hk}}{5} \quad (17)$$

利用熵值法确定多因素权重 $\{w_1, w_2, w_3, w_4, w_5\}$ ，以此衡量这些因素对“脱贫先进村庄”称号评选的影响程度。

Step2：计算村庄荣誉值

根据村庄特色指标显著性得分均值 (Vci_Ssmean_h)、扶贫单位类型绩效综合评分 (Pau_Pscore)、扶贫单位绩效综合平均得分 ($Paut_Pscore$)、2015 年总分 ($Score_{2015}$)、2020 年总分 ($Score_{2020}$) 以上五种因素加权求和得到村庄荣誉值 (Village Honor value)，以此衡量该村庄在“脱贫先进村庄”称号评选中的排序，用符号 Vhv_h 表示：

$$Vhv_h = w_1 * Vci_Ssmean_h + w_2 * Pau_Pscore + w_3 * Paut_Pscore + w_4 * Score_{2015} + w_5 * Score_{2020} \quad (18)$$

依据题意全国计划给予 10000 个村庄“脱贫先进村庄”称号，且称号分为一级和二级，一、二级称号比例为 1:3。所以本文给前 2500 个村庄授予一级称号，给 2501~10000 授予二级称号。根据式 (21) 对村庄荣誉值进行排序确定出一级、二级荣誉分别所需要的最低荣誉值，以此判断最后十个未知村庄中谁能评上“脱贫先进一级村庄”。

5.4.2 模型的求解

模型五：基于村庄相似度匹配的指标预测模型

基于村庄相似度匹配的指标预测模型，根据式 (17)、(18) 找到未知村庄最匹配的已知村庄（如表 4.1）

表 4.1

待匹配村庄编号	帮扶单位	帮扶单位类型	匹配结果
39257	116	1	39135
25149	89	1	25273
12722	7	2	12705
12916	10	2	13018
21570	47	1	21484
22096	48	1	22024
47883	138	1	47866
34208	78	1	33840
34276	78	1	34192
52436	151	3	52462

利用比值法对其进行计算后得到表 4.2

表 4.2 预测值

村庄编号	2020 SR	2020 CY	2020 HJ	2020 WJ	2020 SS
39257	1.2305	0.3800	0.0200	0.6891	-0.1462
25149	0.5421	0.4742	0.9452	-0.0072	-0.3309
12722	-0.9016	-0.0269	0.2698	-0.3649	0.8261
12916	-8.5986	-0.0758	0.1126	-0.1043	-0.0143
21570	-0.0796	0.1925	0.3549	-0.3740	-0.1601
22096	-0.1707	0.6725	-0.0400	0.0177	-0.0655
47883	0.3380	0.1191	-0.0292	-0.1590	-0.0236
34208	0.2851	0.0094	0.1509	0.8902	0.2512
34276	0.2559	0.3686	0.1511	-0.2578	-0.6392
52436	-0.2397	-0.0988	-0.6455	-0.8371	-0.7265

模型六：多因素脱贫先进村庄综合评价模型

Step1：熵值法确定多因素权重

根据村庄特色指标显著性得分均值 (Vci_Ssmean_n)、扶贫单位类型绩效综合评分 (Pau_Pscore)、扶贫单位绩效综合平均得分 ($Paut_Pscore$)、2015 年总分 ($Score_{2015}$)、2020 年总分 ($Score_{2020}$) 以上五种因素按熵值法确定多因素权重比值。

表 4.3 多因素权重比值

	村庄特色指标显著性得分均值 (Vci_Ssmean_n)	扶贫单位类型绩效综合 评分 (Pau_Pscore)	扶贫单位绩效综合得分 ($Paut_Pscore$)	2015年总分 ($Score_{2015}$)	2020年总分 ($Score_{2020}$)
权重	0.0079	0.4553	0.1343	0.1227	0.2799

根据式 (21) 得到村庄荣誉值如下表

表 4.4

村庄编号	绩效综合得分	平均特性	单位类型分数	村庄基础	村庄现状	荣誉值
39257	-0.4896	-0.2334	0.1517	1.0448	0.4861	0.2657
25149	-0.9377	-0.4414	0.0307	1.4405	0.4154	0.1776
12722	-0.9885	-0.2433	-0.1143	1.1845	-0.0222	-0.0476
12916	-2.2299	-0.4543	-0.2761	1.2310	-0.0190	-0.2830
21570	-0.9179	-0.3973	-0.4316	1.0448	0.0161	-0.1902
22096	-0.4672	-0.1906	-0.4180	0.6375	-0.1601	-0.2212
47883	-0.5563	-0.2421	-0.1671	0.6840	-0.0135	-0.0726
34208	-0.0875	-0.0436	-0.2171	0.5676	-0.0174	-0.0461
34276	-0.9290	-0.4118	-0.2171	1.0332	-0.0301	-0.1085
52436	-0.6392	-0.1375	-0.0511	0.0556	-0.1250	-0.1383

根据上表得出，一级荣誉最低值为 0.545127738、一级荣誉最低值为 0.14658427。对比 10 个未知村庄所计算出的村庄荣誉值，如下表

表 4.5 荣获称号的“先进”村

村庄编号	荣誉值	评定结果
39257	0.2657	脱贫先进二级村庄
25149	0.1776	脱贫先进二级村庄
12722	-0.0476	/
12916	-0.2830	/
21570	-0.1902	/
22096	-0.2212	/
47883	-0.0726	/
34208	-0.0461	/
34276	-0.1085	/
52436	-0.1383	/

发现 10 个未知村庄中编号为 39257 和 25149 的两个村庄最终可以被评选为“脱贫先进村庄”二级称号。

5.5 问题五：推荐信

尊敬的中国扶贫办领导：

我们是 xx 大学数学建模团队，在此全面建设小康的关键时期，为凝心聚力，确保如期打赢脱贫攻坚战，我们对研究团队所给的数据进行了深入研究，并结合国内扶贫实际情况及对应问题做出了相应问题。

首先，我们建立了基于 Person 相关性分析的被帮扶单位成长稳定性模型，得到 $gse > 0$ 的结论，即居民收入，产业发展等各指标的关系逐渐趋于线性正相关，这意味着国家战略落实后，符合预期规律。被帮扶村庄在帮扶单位的帮助下，各项指标皆趋于稳定，人民生活逐渐富足，因贫困而造成的成长不确定性在不断减小。

第二，因单一的数据无法体现扶贫单位的实际落实效果，我们基于帮扶绩效显著性，单位规模比重，直接绩效分值等指标，结合熵值法建立了扶贫单位及其类型的绩效评价模型，最终评价出了扶贫单位类型绩效排名（135204），以 113，149，158 为代表的十个突出扶贫单位。

第三，扶真贫，真扶贫，扶贫先扶智，扶志，或是授之以渔，或是提供资源，或是大力发展教育，不同扶贫单位各有其长处短处。取长补短，相互学习，共同进步才能促进国家扶贫事业发展，因此，为找到在某一指标上具有特色，富有经验的扶贫单位，我们建立了基于粘合稳定性影响因子的特色指标绩效评价模型，根据每一单位对应指标的影响因子及其特色指标显著性得分，最终得到了各单项评价指标前五名的帮扶单位编号。

第四，为进一步完善“脱贫先进村庄”称号评选的合理性，我们基于熵值法建立了多因素脱贫先进村庄综合评价模型，找到了扶贫单位，村庄绩效综合得分等主要影响因素，最终结合各因素得出村庄的“荣誉值”，根据荣誉值进行评选。

以上就是我们团队的研究结果，希望我们的研究能对扶贫开发工作有所帮助，也期望在伟大中国共产党的领导下，早日打赢脱贫攻坚战！

六、 参考文献

- [1]黄承伟.中国扶贫开发道路研究:评述与展望[J].中国农业大学学报(社会科学版). 2016(05)
- [2]葛安佳.安徽省大别山连片特困地区扶贫绩效评价与模式研究[D].安徽财经大学,2018.
- [3]陈升,潘虹,陆静.精准扶贫绩效及其影响因素:基于东中西部的案例研究[J].中国行政管理.2016(09)
- [4]孙晗霖,王志章,刘芮伶.民族地区财政扶贫绩效及其影响因素研究——基于渝东南地区的实证研究
- [5].西南大学学报(自然科学版),2020,42(06):65-78.
- [6]刘帅,王晓江.海南旅游精准扶贫绩效影响因素的研究[J].财经界(学术版),2018(22):8-9.
- [7]郭显光.熵值法及其在综合评价中的应用[J].财贸研究,1994(6):56-60.
- [8]姜启源.数学实验与数学建模[J].数学的实践与认识,2001(5):613-617.

七、 附录

表 1 所有指标成长稳定性增长率矩阵及其粘合稳定性影响因子

	5	4	3	2	1	0
S_{SR_CY}	0.617	0.380	0.599	0.341	0.706	0.106
S_{SR_HJ}	0.295	0.806	0.291	0.144	0.371	0.139
S_{SR_WJ}	0.430	2.228	0.403	0.247	0.586	-0.012
S_{SR_SS}	0.442	4.561	0.365	0.274	0.635	0.030
Asf_{SR}	0.446	1.994	0.415	0.252	0.575	0.066
S_{CY_SR}	0.617	0.380	0.599	0.341	0.706	0.106
S_{CY_HJ}	0.464	2.000	0.425	0.351	0.499	0.148
S_{CY_WJ}	0.281	1.028	0.285	0.193	0.469	0.045
S_{CY_SS}	0.302	6.581	0.261	0.211	0.490	0.086
Asf_{CY}	0.416	2.497	0.393	0.274	0.541	0.096
S_{HJ_SR}	0.806	0.291	0.144	0.371	0.139	0.806
S_{HJ_CY}	0.464	2.000	0.425	0.351	0.499	0.148
S_{CY_WJ}	0.631	9.012	0.396	0.236	0.395	0.066
S_{CY_SS}	0.365	4.818	0.271	0.219	0.419	0.135
Asf_{CY}	0.439	4.159	0.346	0.238	0.421	0.122
S_{HJ_SR}	0.295	0.806	0.291	0.144	0.371	0.139
S_{HJ_CY}	0.464	2.000	0.425	0.351	0.499	0.148
S_{HJ_WJ}	0.631	9.012	0.396	0.236	0.395	0.066
S_{HJ_SS}	0.365	4.818	0.271	0.219	0.419	0.135
Asf_{HJ}	0.439	4.159	0.346	0.238	0.421	0.122
S_{WJ_SR}	0.430	2.228	0.403	0.247	0.586	-0.012
S_{WJ_CY}	0.281	1.028	0.285	0.193	0.469	0.045
S_{WJ_HJ}	0.631	9.012	0.396	0.236	0.395	0.066
S_{WJ_SS}	0.092	0.045	0.108	0.085	0.269	-0.013
Asf_{WJ}	0.359	3.078	0.298	0.190	0.430	0.022
S_{SS_SR}	0.442	4.561	0.365	0.274	0.635	0.030
S_{SS_CY}	0.302	6.581	0.261	0.211	0.490	0.086
S_{SS_HJ}	0.365	4.818	0.271	0.219	0.419	0.135
S_{SS_WJ}	0.092	0.045	0.108	0.085	0.269	-0.013
Asf_{SS}	0.300	4.001	0.251	0.197	0.453	0.060

熵值法:

```
function [s,w]=shang(x)
[n,m]=size(x); % n=23 个国家, m=5 个指标
%% 数据的归一化处理
[X,ps]=mapminmax(x');
ps.ymin=0.002; % 归一化后的最小值
ps.ymax=0.996; % 归一化后的最大值
ps.yrange=ps.ymax-ps.ymin; % 归一化后的极差,若不调整该值,则逆运算会出错
X=mapminmax(x',ps);
% mapminmax('reverse',xx,ps); % 反归一化,回到原数据
X=X'; % X 为归一化后的数据
%% 计算第 j 个指标下,第 i 个记录占该指标的比重 p(i, j)
for i=1:n
    for j=1:m
        p(i, j)=X(i, j)/sum(X(:, j));
    end
end
%% 计算第 j 个指标的熵值 e(j)
k=1/log(n);
for j=1:m
    e(j)=-k*sum(p(:, j).*log(p(:, j)));
end
d=ones(1,m)-e; % 计算信息熵冗余度
w=d./sum(d); % 求权值 w
s=w*p'; % 求综合得分[\code]
s = s';
```

相似度匹配及预测:

```
clc;
clear;
format short;
a = importdata('disiti89.txt');
[m,n] = size(a);
minsum = 9999;
for i = 1:m - 1
    sum = 0;
    for j = 1:n
        sum = sum + abs(a(i, j) - a(m, j));
    end
    if(sum < minsum)
```

```

        minsum = sum;
        mini = i;
    end
end
for j = n - 5:n
    a(m, j) = a(mini, j)/(a(mini, j - 6)/a(m, j - 6));
end
for i = 1:n
    ans(1, i) = a(mini, i);
    ans(2, i) = a(m, i);
end

```