

# 提高种猪生产成绩的研究

## 摘要

在现代化畜牧养殖场中，往往已实现了半自动化，自动化设备可以反馈一些状态数据，而且通过传感器可以收集到一定的结构化数据，同时，需人工判断与记录的数据也会很好地按既定格式保存在服务器中，以上我们称之为养殖大数据。本文通过讨论影响种猪生产成绩的大部分因素得出能够获得理想的种猪生产成绩的最优解或最优解的趋势。同时通过对全国多个省份各个指标的讨论，得出可以在2018年4月建设种猪场以获得最大利润的省份区域。以下程序大部分使用MATLAB工具实现，少部分使用SPSS工具实现，在使用SPSS工具时我们会适当的地方给与提示。

首先，由于可用以评价种猪的生产成绩的影响指标因素过多，通过聚类分析法将单个指标因素按照关联度和相似度分为互不影响的五大类。在应用聚类分析的过程中，考虑到猪每天温度、湿度和喂养的饲料量的不同，在用 $\chi^2$ 拟合优度检验发现占总猪数为87.4%的猪在一个周期里（即配种和怀孕生产所用时间119天）其温度的变化在显著性水平 $\alpha = 0.05$ 下服从正态分布，占总猪数为100%的猪在一个周期的不同湿度里，其湿度的变化在显著性水平 $\alpha = 0.05$ 下服从正态分布，所以，我们完全可以用每只种猪在一个周期里温度和湿度的平均值来表示种猪温度和湿度这个变量，这样的处理方法是有很大的统计意义的。考虑到健康仔猪数和健康仔猪数占生成猪总数的比例对评价种猪的生产成绩所占比例不同，因此我们用层次分析法来进行对这两个变量的决策。最后得出两者占总生产成绩的比例。最后为了能够建立关系分析模型表征变量a，b对结果c的影响程度，我们运用灰色关联度分析的方法来算出各个指标对种猪的生产成绩的影响，并得出前三个影响程度较大的因素分别为一个周期内种猪样本的平均温度，平均湿度和种猪的肥瘦程度。

其次由于绝对值排名前3的显著因子分别为一个周期的平均温度，一个周期的平均湿度和种猪的肥瘦情况，根据这三个显著因子的实际数据，由于所给样本多达15000个，根据统计学分析在这些样本中我们有很大的概率可以获得能使种猪生产成绩最优化的最优解或最优解的趋势。利用主层次分析法，我们可以计算得到每个种猪样本的得分，其中得分最高的种猪样本所处的环境即为影响种猪生产成绩的最优解。经计算我们得知，最优解种猪样本的平均温度，平均湿度和种猪的肥瘦程度分别为25.24，82.58，20，所以当以25.24度的温度，以82.58度的平均湿度喂养种猪时，在一个周期结束后，种猪的体重程度为20，生出的健康小猪数为15只，健康小猪占的比例为100%，此时种猪的生产成绩最好。

最后,由于中国的生猪生产布局变动受到经济因素、市场因素、自然资源等条件的影响,我们按照经济学中的资源禀赋理论和综合比较优势理论来确定建设猪场的最佳位置,据此我们将建设范围锁定在湖南省,四川省,河南省,贵州省这些省份中,最后通过利用有季节特点的时间序列模型来预测出两年后湖南省,四川省,河南省,贵州省这些省份的生猪价格和玉米价格,通过将养猪的成本简化为养猪饲料玉米的成本,用生猪的价格减去玉米的价格即可大概获得养殖场每增加一斤生猪肉可获得的利润,将这几个利润进行比较,得知在四川省养殖种猪能够获得最大利润。这个结果也验证了四川被称为中国养猪第一大省这一说法,所以 2018 年 4 月应在四川省这一省份区域建设种猪场方可获得最大利润。

**关键词:** 正态分布 聚类分析法 层次分析法 灰色关联度 主成分分析法 资源禀赋理论 综合比较优势 时间序列

## 一、问题的提出

### 1.1 问题的背景

我国养猪业生产逐步向工厂化、集约化发展已是大势所趋,与之相适应的现代化管理体系建设显得十分重要,而高效的生产数据统计方式对于一个规模养殖场管理体系的建立来说,尤为重要。一个猪场的核心竞争力,除了优良的品种,更重要是准确的数据统计与分析。数据化管理涉及到猪场的温度和湿度,耗料数据、药品管理、猪群健康管理、母猪受孕率等多项指标,只有通过对各项指标的精细准确的分析,才能找到生产管理中有哪些问题制约了猪场效益,并快速查询到这些问题,比如是猪群健康问题,还是配种问题,亦或用药问题,或者成本控制问题。

### 1.2 问题的重述

在现代化畜牧养殖场中,往往已实现了半自动化,自动化设备可以反馈一些状态数据,而且通过传感器可以收集到一定的结构化数据,同时,需人工判断与记录的数据也会很好地按既定格式保存在服务器中,以上我们称之为养殖大数据。

现有某企业在全国各地建有多个种猪场(饲养和管理母猪,用于生产仔猪的养殖场),也因此收集到大量的养殖大数据。数据主要分为三类:

a.传感器采集的饲养方案数据。如温度、湿度、饲喂饲料量等。未来随着物联网技术的发展,将有更多的数据可通过传感器收集,如体温、运动图像等。

b.人工采集并记录的种猪数据。如品种、胎次、疾病情况、肥瘦程度等。

c.种猪的生产成绩。主要是生仔总数和健康仔猪数。其中以健康仔猪数为最重要的指标，但健康仔猪数占生仔总数的比例也需考虑提升。

如附件所示。现需要：

(1) 对于 a.饲养方案数据和 b.种猪数据，是影响 c.生产成绩的因素，但是，这只是一个经验判断，试建立关系分析模型，以参数形式表征变量 a, b 对结果 c 的影响程度（或正或负）。注意，变量间并非一定相对独立，可对变量进行分组，使分组与分组之间相对独立，再表征分组对结果的影响程度。

(2) 对 (1) 中影响程度较大的因素记为显著因子（绝对值排名前 3），试建立显著因子对 c.生产成绩造成影响的数学模型，并求出最优解或最优解的趋势。

(3) 该企业对种猪场的建设速度为 180 天，猪场前期审批和建成后引种需 180 天，配种和怀孕生产周期为 17 周，养户饲养生猪从进猪到售卖约为 180 天。也就是说，做出建设种猪场的决定到养育成生猪在市场售卖，约有两年的时间滞后（以两年时间计算）。现请根据全国人口分布，及生猪价格的分布和变动趋势，决策 2018 年 4 月应在哪些区域建设种猪场。

## 二、问题的分析

### 2.1 问题一

题目中虽然有比较丰富的数据支持，但是由于用以评价种猪的生产成绩的影响指标因素过多，我们决定应用聚类分析法，层次分析法和灰色关联预测法来对各项指标进行决策。

首先我们利用聚类分析法将单个指标因素按照关联度和相似度分为互不影响的五大类：第一类包括猪的日平均温度，日平均湿度，日平均喂食量和猪的肥瘦情况；第二类包括猪怀孕期间患蓝耳病，患圆环病，患仔猪黄白痢病和患支原体肺炎情况；第三类包括猪的品种；第四类包括猪的胎次；第五类包括猪怀孕期间患病毒性腹泻病情况；在应用聚类分析的过程中，考虑到猪每天温度、湿度和喂养的饲料量的不同，在用 $\chi^2$ 拟合优度检验发现占总猪数为 87.4%的猪在一个周期的不同温度里，其温度的变化在显著性水平 $\alpha = 0.05$ 下服从正态分布，占总猪数为 100%的猪在一个周期的不同湿度里，其湿度的变化在显著性水平 $\alpha = 0.05$ 下服从正态分布，因此，我们完全可以用每只种猪在一个周期里温度和湿度的平均值来表示种猪温度和湿度这个变量，从而使数据更方便处理。

其次考虑到健康仔猪数和健康仔猪数占生成猪总数的比例对评价种猪的生产成绩所占比例不同，因此我们用层次分析法来进行对这两个变量的决策。在构造出评价矩阵之后，判断其一致性，最后得到健康仔猪数和健康仔猪数占生成猪总数的比例对评价种猪的生产成绩所占比例。

最后，为了能够建立关系分析模型，以参数形式表征变量  $a$ 、 $b$  对结果  $c$  的影响程度，在这里我们运用灰色关联度分析的方法来算出各个指标对种猪的生产成绩的影响，并得出前三个影响程度较大的因素用于第二问的研究。

## 2.2 问题二

从第一题的结果可知，绝对值排名前 3 的显著因子分别为一个周期的平均温度，一个周期的平均湿度和种猪的肥瘦情况，根据这三个显著因子的实际数据，由于所给样本多达 15000 个，所以根据统计分析可得，在这么大的数据样本里，我们可以通过主成分分析法计算每个种猪样本的得分，得分最高的种猪样本所处的环境即为影响种猪生产成绩的最优解。

## 2.3 问题三

中国的生猪生产布局变动受到经济因素、市场因素、自然资源等条件的影响。按资源禀赋理论，拥有丰富的饲料资源和较好的环境优势的区域，具有生猪生产的比较优势。因此，基于上述分析，通过资源禀赋系数法和综合比较优势的测算测定中国生猪各生产区域生产的比较优势状况，并以此来大概确定 2018 年 4 月应在哪些区域建设种猪场的省份。

经过寻找资料分析了解，影响生猪价格的主要因数有以下 3 个方面。

1. 生猪的生产周期。生猪不是工业产品，其需要一个自然的生长周期，从母猪怀孕到仔猪生产再到生猪长成后投放市场，一般需要 1 年半到 2 年。在这样的客观规律下，养殖户在看到市场上猪肉价格上涨时，会先增加母猪养殖数量，以便生出更多猪仔，养大后才能提高出栏量。同样的道理，当看到市场价格开始走低时，养殖户只能被动接受，而无法立即削减存栏量，要等到猪出栏。

2. 生猪饲料成本。现在养猪基本靠饲料，我国生猪饲料成本在养猪的总成本中占 70% 以上，它是影响生猪养殖规模与生猪供应的一个重要因素。分析生猪价格波动影响因素少不了饲料成本，由于生猪饲料的原材料主要是玉米等农作物，关注生猪饲料成本主要还是关注这些用来制作生猪饲料的玉米等原材料价格变动情况。

通过以上分析影响生猪价格的因素可知，为了简化模型，我们假设生猪的价格主要受生猪饲料成本，也即是玉米这种原材料价格变动情况。根据查找的资料，生猪每长一斤肉大约要消费 2.8 斤的玉米，因此我们粗略地用生猪一斤肉的价格减去每长一斤肉大约消费的玉米价

格等于养猪场企业生猪每一斤肉的纯利润，通过比较利润的大小我们找出应在 2018 年 4 月建设种猪场的省份，由于我们只能知道最近月份每斤玉米的价格和生猪一斤肉的价格而不能确定未来每斤玉米的价格和生猪一斤肉的价格，并且观察数据可知，每斤玉米的价格和一斤生猪肉的价格符合具有季节特点的时间序列模型，所以我们可以利用有季节特点的时间序列模型来预测出两年后湖南省，四川省，河南省，贵州省这些省份的生猪价格和玉米价格，通过将养猪的成本简化为养猪饲料玉米的成本，用生猪的价格减去玉米的价格即可大概获得养殖场每增加一斤生猪肉可获得的利润，通过比较可以得知在哪个省份饲养种猪会产生最大的生产成绩。

### 三、模型假设

- (1) 假设相同的温度湿度喂食量对每只猪有相同的影响；
- (2) 假设生猪的养殖成本主要来源是购买猪粮玉米的费用；
- (3) 假设数据中大部分猪的情况能代表所有的猪的实际情况；
- (4) 假设大部分人的消费习惯和政治环境不会发生改变；
- (5) 假设主要影响的条件在温度、湿度、喂食量、喂养量、品种、胎次、蓝耳病、圆环病、仔猪黄白痢、病毒性腹泻、支原体肺炎中被包含；

### 四、符号说明

符号	说明
$P_i$	表示概率
$A_i$	表示不同区间
$f_i$	表示落在区间的样本个数
$r$	表示相关系数
$\omega$	权重向量
$\varepsilon$	关联系数
$b_j$	累计贡献率
$d$	两类变量之间的距离
$EF$	资源禀赋系数

五、模型的建立和求解

5.1 问题一

5.1.1 正态分布模型的证明和求解

因为我们有多天的猪场温度和湿度数据，不方便直接用聚类分析法来对 a，b 变量直接进行分类，因为多天的温度和湿度数据呈现规律性变化，我们先证明多天的温度和湿度变化在显著性水平 $\alpha = 0.05$ 下是否服从正态分布，如果满足我们就用每只小猪在一个周期里温度和湿度的平均值来表示温度和湿度这两个变量。

首先，我们采用矩估计方法估计这些参数的取值，先从所给的数据算出样本均值和标准差，我们先用附件中的第一只猪所在猪场温度随天数的变化情况来看是否满足正态分布，通过计算可得第一只猪所在猪场温度在 119 天里样本的平均值和标准差如下：

$$\bar{x}= 24.1, s= 2.0$$

在显著性水平 $\alpha = 0.05$ 下，检验假设：

$H_0$ : 第一只猪在 119 天里每天的温度变化服从正态分布  $N(24.1, 2.0^2)$ 。

样本观察值的最大值为 28.8，最小值为 20，将区间 $[20, 28.8]$ ，分成互不相交的 9 个区间： $A_1, A_2, \cdots A_9$ ，如下表所示。以  $f_i (i=1, 2, 3, \cdots 9)$  记样本观察值落在  $A_i$  中的个数，以  $A_i$  记事件  $\{ X \in A_i \}$ 。若  $H_0$  为真， $X$  服从正态分布  $N(24.1, 2.0^2)$ ，可以计算

$$P_i = P(A_i) = P\{ a_i < X \leq a_{i+1} \}, i=1, 2, 3 \cdots 9。$$

计算结果列于下表：（如表 1 所示）（代码见附录 1）

表 1  $\chi^2$  检验数据表

$p_i$	$n p_i$	$f_i^2 / (n p_i)$
0.0543	0.0543	7.5861
0.0659	0.0659	10.3301
0.1084	0.1084	25.1168
0.1486	0.1486	5.6560
0.1697	0.1697	19.8124
0.1614	0.1614	11.7142

0.1279	0.1279	16.8153
0.0845	0.0845	22.3804
0.0794	0.0794	8.5760
		$\Sigma=127.98$

令 $\chi^2=8.89$ ，由 $\alpha=0.05$ ， $k=9$ ， $r=2$  知

$$\chi^2(k-r-1)=\chi^2_{0.1}(4)=10.64>8.89=\chi^2$$

故在显著性水平 $\alpha=0.05$  下，接受假设  $H_0$ ，认为这些数据是来自正态分布总体。

同理，对于其他的小猪所在猪场温度和湿度随天数的变化情况，我们用同样的方法检测后发现，在这 15000 个样本中，有 87.40%的样本小猪所在猪场温度随天数的变化情况服从正态分布，所有的样本小猪所在猪场湿度随天数的变化情况全部服从正态分布，故我们在比较大的概率下，完全可以用每个样本的均值和方差来表示这组样本的温度随天数的变化情况，每个样本温度和湿度的均值和方差位于附件中。

### 5.1.2 聚类分析法在影响因素归类中的应用<sup>[1]</sup>

R 型聚类法可以研究变量之间的相似关系，按照变量之间的相互关系把各个变量聚合成若干类，从而可以方便地找出影响体系的主要因素。

1) 首先我们用变量相似性度量。在对变量进行聚类分析时，第一步就是确定变量的相似性度量，本文中,采取的相似性度量为相关系数，具体方法如下：

记变量 $x_j$  的取值  $(x_{1j}, x_{2j}, \dots, x_{219j})^T \in R_n (j=1, 2, \dots, 11)$  。则可以用两变量  $x_j$  与  $x_k$  的样本相关系数作为它们的相似性度量， 为

$$r_{jk} = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)}{[\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 \sum_{i=1}^n (x_{ik} - \bar{x}_k)^2]^{\frac{1}{2}}}$$

2) 其次我们用变量聚类法将以上影响因素分类。在本文中， 采取最长距离法解决变量聚类问题， 具体过程如下：

在最长距离法中， 定义两类变量的距离为

$$R(G_1, G_2) = \max \{d_{jk}\},$$

其中，

$d_{jk} = 1 - r_{jk}$  或  $d_{jk}^2 = 1 - r_{jk}^2$ ， 此时,  $R(G_1, G_2)$  与两类中相似性最小的两变量间的相似性度量值有关。我们将各影响因素之间的关联系数矩阵作为输入参数， 经过聚类分析将相关

程度比较大的影响因素作为输出。利用 SPSS 工具进行聚类分析得到的聚类结果如下图 1 所示：

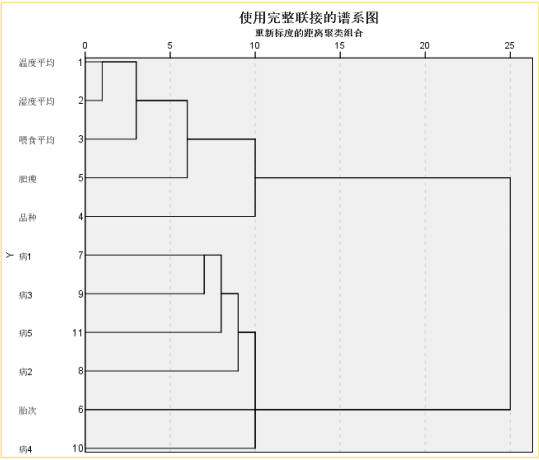


图 1

利用聚类分析法将单个指标因素按照关联度和相似度分为互不影响的五大类：第一类包括猪的日平均温度，日平均湿度，日平均喂食量和猪的肥瘦情况；第二类包括猪怀孕期间患蓝耳病，患圆环病，患仔猪黄白痢病和患支原体肺炎情况；第三类包括猪的品种；第四类包括猪的胎次；第五类包括猪怀孕期间患病毒性腹泻病情况；

5.1.3 层次分析法构建评价体系<sup>[1]</sup>

1) 建立层次结构模型。将决策问题分解为两个层次，最上层为目标层 M，即种猪的生产成绩；最下层为准则层，即健康仔猪数和健康仔猪数占生仔总数的比例两个指标(如图 2 所示)；

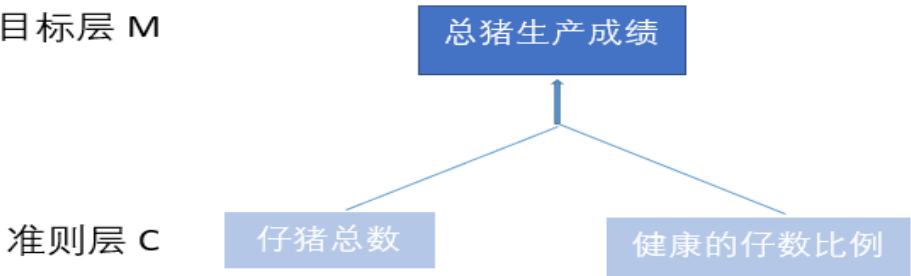


图 2

2) 模型求解

①构造判断矩阵 M-C：将基准层 C 中两个元素 C1，C2 两两比较，得成对比较矩阵。(如表 2 所示)

M	C1	C2
---	----	----



C1	1	5
C2	0.2	1

表 2 比较矩阵

求解 M-C 的特征值，易解得 $\lambda_{\max}=2.0$ ，且权重向量 $\omega_i=(0.8333,0.1667)$

由公式  $CI=\frac{\lambda_{\max}-n}{n-1}$ ，于是根据 $CR=\frac{CI}{RI}$ ，计算得到  $CR=0.0$ ，通过了一致性检验。（如表 3 所示）（代码见附录 2）

n	2	3	4	5	6	7	8	9	10
RI	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.49

表 3 n 与 RI 的关系

5.1.3 灰色关联度计算影响程度

1) 灰色关联度的介绍与建立

客观世界的很多实际问题，其内部的结构、参数以及特征并未被人们全部了解，人们不可能象研究白箱问题那样将其内部机理研究清楚，只能依据某种思维逻辑与推断来构造模型。对这类部分信息已知而部分信息未知的系统，我们称之为灰色系统。而大千世界里的客观事物往往现象复杂，因素繁多。我们经常要对系统进行因素分析，这些因素中哪些对系统来讲是主要的，哪些是次要的，哪些需要发展，哪些需要抑制，哪些是潜在的，哪些是明显的。一般来讲，这些都是我们极为关心的问题。事实上，因素间关联性如何、关联程度如何量化等问题是系统分析的关键。

对于两个系统之间的因素，其随时间或不同对象而变化的关联性大小的量度，称为关联度。在系统发展过程中，若两个因素变化的趋势具有一致性，即同步变化程度较高，即可谓二者关联程度较高；反之，则较低。因此，灰色关联分析方法，是根据因素之间发展趋势的相似或相异程度，亦即“灰色关联度”，作为衡量因素间关联程度的一种方法。

下面计算应用灰色关联度计算时的量化分析四步曲：

1. 确定反映系统行为特征的参考数列和影响系统行为的比较数列：反映系统行为特征的数据序列，称为参考数列。（Y）设参考数列（又称母序列）为  $Y=\{Y(k) \mid k=1,2,\cdots,n\}$ ；影响系统行为的因素组成的数据序列，称比较数列。（X）比较数列（又称子序列） $X_i=\{X_i(k) \mid k=1,2,\cdots,n\}$ ， $i=1,2,\cdots,m$ 。

2. 标准化（无量纲化）：由于系统中各因素列中的数据可能因量纲不同，不便于比较或在比较时难以得到正确的结论。因此为了保证结果的可靠性，在进行灰色关联度分析时，一般都要进行数据的无量纲化处理。以参照数列（取最大数的数列）为基准点，将各数据标准化成介于 0 至 1 之间的数据最佳。

3. 根据公式采用适当的值，产生对应差数列表，内容包括：与参考数列值差（绝对值）、最大差、最小差、 $\xi$  (Zeta) 为分辨系数， $0 < \xi < 1$ ，可设  $\xi = 0.5$  (采取数字最终务必使关联系数计算： $\xi_i(k)$  小于 1 为原则)。

4. 计算关联系数。 $X_0(k)$  与  $X_i(k)$  的关联系数

$$\varepsilon_i(k) = \frac{\min_s \min_t |x_0(k) - x_i(k)| + \rho \max_s \max_t |x_0(k) - x_i(k)|}{|x_0(k) - x_i(k)| + \rho \max_s \max_t |x_0(k) - x_i(k)|}$$

为比较数列  $x_i$  对参考数列  $x_0$  在第  $k$  个指标上的关联系数，其中  $\rho \in [0, 1]$  为分辨系数。其中，称  $\min_s \min_t |x_0(t) - x_s(t)|$ 、 $\max_s \max_t |x_0(t) - x_s(t)|$  分为两级最小差及两级最大差。 $\rho$  越小，分辨力越大，一般  $\rho$  的取值区间为  $(0, 1)$ ，具体取值可视情况而定。当  $\rho \leq 0.5463$  时，分辨力最好，通常取  $\rho = 0.5$ 。 $\xi_i(k)$  为比较数列  $x_i$  的第  $k$  个元素与参考数列  $x_0$  的第  $k$  个元素之间的关联系数。最后求各系数的平均值即是  $X_i$  与  $X_0$  的关联度  $r_i$ 。

5. 计算关联度：因为关联系数是比较数列与参考数列在各个时刻（即曲线中的各点）的关联程度值，所以它的数不止一个，而信息过于分散不便于进行整体性比较。因此有必要将各个时刻（即曲线中的各点）的关联系数集中为一个值，即求其平均值，作为比较数列与参考数列间关联程度的数量表示，关联度  $r_i$  公式如下：

$$r_i = \sum_{k=1}^n u_i \xi_i(k)$$

2) 灰色关联度的预求解

关联度的大小直接反映系统中的各因素对目标值的影响程度，根据表格中的数据，运用灰色关联分析法进行因素分析的步骤如下：

### 第一步，确定分析数列

由于给的数据样本多达上万个，所以在这里我们只是列出来数据中前 10 的小猪数据，其他种猪计算方式和以下计算方式一样。下面计算各指标对种猪生出健康种猪数的影响。

$X_0(k) = \{0.01, 0.5, 1.5, 10.8, 13, 16.3, 18, 19.3, 14.8, 10.3, 8, 1\}$

$X_1(k) = \{24.1597, 24.3387, 23.9042, 24.2857, 23.0798, 22.6824, 24.8462, 25.2092, 25.4824, 24.9345\}$

$$X_2(k) = \{83.4017, 83.0773, 83.4849, 78.4134, 79.0916, 77.8319, 82.5387, 80.2193, 81.8319, 82.5168\}$$

$$X_3(k) = \{2.9706, 2.7412, 2.9647, 2.9706, 2.3118, 2.7235, 2.3176, 2.9706, 2.9765, 2.9882\}$$

$$X_4(k) = \{1, 2, 1, 1, 1, 2, 1, 1, 2, 1\}$$

$$X_5(k) = \{12, 15, 12, 12, 17, 15, 17, 12, 12, 12\}$$

$$X_6(k) = \{2, 1, 3, 2, 2, 3, 1, 7, 4, 2\}$$

$$X_7(k) = \{0, 0, 0, 0, 0, 0, 0, 0, 0, 0\}$$

$$X_8(k) = \{0, 0, 0, 0, 1, 0, 0, 0, 1, 0\}$$

$$X_9(k) = \{1, 0, 0, 0, 0, 0, 0, 0, 0, 0\}$$

$$X_{10}(k) = \{0, 0, 0, 0, 0, 0, 0, 1, 0, 0\}$$

$$X_{11}(k) = \{0, 0, 0, 1, 0, 0, 1, 0, 0, 0\}$$

## 第二步：计算关联系数

$$\min \min |x_0(k) - x_i(k)| = 0$$

$$\max \max |x_0(k) - x_i(k)| = 1.9091$$

$\rho$ , 称为分辨系数。 $\rho$  越小, 分辨力越大, 一般  $\rho$  的取值区间为 (0, 1), 具体取值可视情况而定。

通常取  $\rho = 0.5$ 。

$\xi_i(k)$  继续比较数列  $x_i$  的第  $k$  个元素与参考数列  $x_0$  的第  $k$  个元素之间的关联系数。

$$\begin{aligned} \varepsilon_i(k) &= \frac{\min \min |x_0(k) - x_i(k)| + \rho \max \max |x_0(k) - x_i(k)|}{|x_0(k) - x_i(k)| + \rho \max \max |x_0(k) - x_i(k)|} \\ &= \frac{0 + 0.5 \times 1.9091}{|x_0(k) - x_i(k)| + 0.5 \times 1.9091} \\ &= \frac{0.95455}{|x_0(k) - x_i(k)| + 0.95455} \end{aligned}$$

将相应  $x_0(k)$  与  $x_i(k)$  的数值代入式  $\varepsilon_i(k) = \frac{\Delta_{\min} + \Delta_{\max}}{\Delta + \rho \Delta_{\max}}$  中, 得

## 第三步：算出关联度

由公式  $R_i = \frac{1}{n} \sum_{k=1}^{10} \varepsilon_i(k)$  分别计算出日平均气温 ( $X_1$ )、日平均湿度 ( $X_2$ )、日平均喂食量 ( $X_3$ )、

品种 ( $X_4$ )、肥瘦 ( $X_5$ )、胎次 ( $X_6$ )、疾病 1 ( $X_7$ )、疾病二 ( $X_8$ )、疾病三 ( $X_9$ )、疾病四 ( $X_{10}$ )、

疾病五 ( $X_{11}$ ) 关于种猪生出健康种猪数 ( $X_0$ ) 的关联度  $R_1, R_2, R_3, R_4, R_5, R_6, R_7, R_8, R_9, R_{10}, R_{11}$ 。

$$R_{01}=(R_1\ R_2\ R_3\ R_4\ R_5\ R_6\ R_7\ R_8\ R_9\ R_{10}\ R_{11})=(0.7605\ 0.8091\ 0.4386\ 0.3567\ 0.8359\ 0.4588\ 0.3415\ 0.3419\ 0.3408\ 0.3424\ 0.3418);$$

**第四步：比较关联度大小得出结论**

由  $R_5 > R_1 > R_2 > R_6 > R_3 > R_4 > R_{10} > R_8 > R_{11} > R_7 > R_9$  说明种猪生出健康种猪数（X0）与猪的肥瘦情况（X5）、日平均气温（X1）、日平均湿度（X2）最密切，而胎次（X6）、日平均喂食量（X3）、品种（X4）、疾病四（X10）、疾病二（X8）、疾病五（X11）、疾病 1（X7）和疾病三（X9）的密切程度依次较小。

以上结果说明影响种猪生出健康种猪数的主要影响因子是猪的肥瘦情况。（代码见附录 3）  
 由于各指标对种猪生出健康种猪数占总小猪数的比例的影响的计算方式和上面计算方式一样，在这里我们便不一一列举出求解过程，通过利用灰色关联度来表示各指标对种猪生出健康种猪数占总小猪数的比例的影响，我们计算出各指标的关联度大小如下：

$$R_{02}=(R_1\ R_2\ R_3\ R_4\ R_5\ R_6\ R_7\ R_8\ R_9\ R_{10}\ R_{11})=(0.6106\ 0.7612\ 0.8455\ 0.9205\ 0.6001\ 0.8020\ 0.8705\ 0.8720\ 0.8669\ 0.8741\ 0.8718);$$

在前面，我们已经计算出健康仔猪数和健康仔猪数占生存总数的比例占评价种猪的生产成绩的比例，即为：

$$P1: P2=0.8333: 0.1667$$

故各指标对种猪的生产成绩的影响的计算公式为：

$$E=(E1\ E2\ E3\ E4\ E5\ E6\ E7\ E8\ E9\ E10\ E11)=(R_{01}',R_{02}')*P$$

其中 P=（P1，p2），计算结果如下**表 4** 所示：

E1	E2	E3	E4
0.7355	0.8011	0.5064	0.4507
E5	E6	E7	E8
0.7966	0.5160	0.4297	0.4303
E9	E10	E11	
0.4285	0.4310	0.4302	

**表 4**

根据表格中的数据可知，

$$E2 > E5 > E1 > E6 > E3 > E4 > E10 > E8 > E11 > E7 > E9$$

从影响结果的排序可知，对种猪的生产成绩影响最大的指标分别为：日平均湿度（X2）、猪的肥瘦情况（X5）、日平均气温（X1）、胎次（X6）、日平均喂食量（X3）、品种（X4）、疾病四（X10）、疾病二（X8）、疾病五（X11）、疾病 1（X7）和疾病三（X9）。

## 5.2 问题二

### 1) 主成分分析法的介绍

在一般的数据分析过程中，我们希望用较少的变量去解释原来资料中的大部分变量，将我们手中许多相关性很高的变量转化成彼此相互独立或不相关的变量。通常是选出比原始变量个数少，能解释大部分资料中变量的几个新变量，即所谓主成分，并用以解释资料的综合性指标。由此可见，主成分分析实际上是一种降维方法。运用主成分分析法进行数据分析的过程如下：

1. 将原始数据按行排列组成矩阵 X
2. 对 X 进行[数据标准化](#)，使其均值变为零
3. 求 X 的协方差矩阵 C
4. 将特征向量按特征值由大到小排列，取前 k 个按行组成矩阵 P
5. 通过计算  $Y = PX$ ，得到降维后数据 Y
6. 用下式计算每个特征根的贡献率  $V_i; V_i = x_i / (x_1 + x_2 + \dots)$

根据特征根及其特征向量解释主成分[物理意义](#)。

### 2) 主成分分析法的应用<sup>[2]</sup>

#### 主成分分析法的步骤

(1) 对原始数据进行标准化处理。假设进行主成分分析的指标变量有 m 个，分别为  $x_1, x_2, \dots, x_m$  共有 n 个评价对象，第 i 个评价对象的第 j 个指标的取值为  $a_{ij}$ 。将各指标值  $a_{ij}$  转换成标准化指标值  $\bar{a}_{ij}$  有

$$\bar{a}_{ij} = \frac{a_{ij} - \mu_j}{s_j}, \quad i=1, 2, \dots, n; \quad j=1, 2, \dots, m;$$

式中：

$$\mu_j = \frac{1}{n} \sum_{i=1}^n a_{ij}; s_j = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (a_{ij} - \mu_j)^2}, \quad j = 1, 2, \dots, m,$$

即  $\mu_j, s_j$  为第 j 个指标的样本均值和样本标准差。

对应地，称

$$\bar{x}_j = \frac{x_j - \mu_j}{s_j}, \quad j=1, 2, \dots, m;$$

为标准化指标变量。

(2) 计算相关系数矩阵 R。相关系数矩阵  $R=(r_{ij})_{m \times n}$  有

$$r_{ij}=\frac{\sum_{k=1}^n \bar{a}_{ki} \bar{a}_{kj}}{n-1}, i, j=1, 2, \cdots, m$$

式中： $r_{ii}=1$ ， $r_{ij}=r_{ji}$ ， $r_{ij}$  为第  $i$  个指标与第  $j$  个指标的相关系数。  
经过计算，我们得到的相关系数矩阵为：（如表 5 所示）

R	R1	R2	R3
R1	1	0.334	0.0076
R2	0.334	1	0.0030
R3	0.0076	0.0030	1

表 5

(3) 计算特征值和特征向量。计算相关系数矩阵 R 的特征值  $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$  及对应的特征向量  $u_1, u_2, \cdots, u_n$ ，其中  $u_j=[u_{1j}, u_{2j}, \cdots, u_{nj}]^T$ ，由特征向量组成  $m$  个新的指标变量：

$$\begin{aligned} y_1 &= u_{11} \bar{x}_1 + u_{21} \bar{x}_2 + \cdots + u_{m1} \bar{x}_m, \\ y_2 &= u_{12} \bar{x}_1 + u_{22} \bar{x}_2 + \cdots + u_{m2} \bar{x}_m, \\ &\vdots \\ y_m &= u_{1m} \bar{x}_1 + u_{2m} \bar{x}_2 + \cdots + u_{mm} \bar{x}_m, \end{aligned}$$

式中， $y_1$  为第 1 主成分， $y_2$  为第二主成分， $y_3$  为第 3 主成分；经过带入数据后计算得：

$$\begin{aligned} y_1 &= 0.7070 \bar{x}_1 + 0.7069 \bar{x}_2 + 0.0225 x_m, \\ y_2 &= -0.0090 \bar{x}_1 - 0.0227 x_2 + 0.997 x_m, \\ y_3 &= -0.7072 x_1 + 0.7070 x_2 + 0.0097 x_m, \end{aligned}$$

(4) 选择  $p$  ( $p \leq m$ ) 个主成分，计算综合评价值。

Step1: 计算特征值  $\lambda_j$  ( $j=1, 2, \cdots, m$ ) 的信息贡献率和累积贡献率。称

$$b_j = \frac{\lambda_k}{\sum_{k=1}^m \lambda_k}, j=1, 2, \cdots, m$$

为主成分  $y_j$  的信息贡献率，同时，有

$$\alpha_p = \frac{\sum_{k=1}^p \lambda_k}{\sum_{k=1}^m \lambda_k}$$

为主成分  $y_1, y_2, y_3$  的累积贡献率。当  $\alpha_p$  接近于 1（一般取  $\alpha_p=0.85, 0.90, 0.95$ ）时，则选择前  $p$  个指标变量  $y_1, y_2, y_3$  作为  $p$  个主成分，代替原来  $m$  个指标，经过计算我们可以得到三个成分的累加贡献率如下：（如表 6 所示）（代码见附录 4）

序号	特征根	贡献率	累加贡献率
1	1.335	44.50	45.50
2	0.999	33.32	77.82
3	0.665	22.17	100

表 6

所以，分别以三个主成分的贡献值为权重，构建主成分综合评价模型，即：

$$Z=0.445y_1+0.3332y_2+0.2217y_3$$

将三个主成分值分别代入得：

当一个周期中平均温度为 25.24 度，平均湿度为 82.58 度，把种猪养到肥瘦程度为 20 的时候，种猪的生产成绩达到最大，此时种猪生出的小猪数为 15 只，生出的小猪中健康猪占的比例为 100%，因此当一个周期中平均温度为 25.24 度，平均湿度为 82.58 度，把种猪养到

肥瘦程度为 20 时，养殖种猪可获得最优生产成绩，即最优解。

5.3 问题三

5.3.1 运用资源禀赋系数法<sup>[3]</sup>

资源禀赋系数是用来反映某个国家或地区某一资源相对丰富程度的指标。

$$EF = \frac{P_i/P_{wi}}{Y_i/Y_{wi}} \tag{1}$$

$P_i$  代表某一地区拥有的猪肉产量； $P_{wi}$  代表全国拥有的猪肉产量； $Y_i$  代表该地区生产总值； $Y_{wi}$  代表全国的国民生产总值。如果  $0 < EF < 1$ ， 则说明该地区生猪缺乏或不具有资源禀赋比较优势；如果  $1 < EF < 2$ ， 则说明该地区生猪拥有一定的资源禀赋比较优势；如果  $EF > 2$ ， 则说明该地区生猪拥有较强的资源禀赋比较优势。如表 7 所示。

表 3 中国生猪资源禀赋具有比较优势的省份（自治区）

年份 1<EF<2 的省份（自治区） EF>2 的省份（自治区）

1990	河北省、安徽省、福建省、湖北省、广西、云南省、甘肃省、江西省	湖南省、四川省、贵州省
1995	河北省、安徽省、湖北省、广西、云南省、甘肃省、河南省、吉林省	湖南省、四川省、贵州省、江西省
2000	河北省、内蒙古、安徽省、江西省、河南省、湖北省、海南省、重庆市、吉林省	湖南省、广西、四川省、贵州省、云南省
2005	河北省、吉林省、安徽省、江西省、河南省、湖北省、广西、海南省、重庆市、甘肃省	湖南省、四川省、贵州省、云南省
2010	河北省、吉林省、安徽省、江西省、湖北省、广西、海南省、山东省	湖南省、四川省、贵州省、云南省、河南省
2015	河北省、吉林省、河南省、江西省、山东省、湖北省、广西	湖南省、四川省、河南省、贵州省

注：表 7 数据根据《中国统计年鉴》《中国畜牧业年鉴》计算所得；内蒙古自治区简称为内蒙古

由表 3 可见：1990—2015 年， 湖南省、四川省和贵州省等地的资源禀赋系数大于 2。分析可知， 由于生猪的生长受到温度的影响， 相对于四季分明的北方而言， 南方四季如春的特点更适宜生猪的养殖；其次， 生猪的养殖需要大量的水， 而西南区域如四川省、湖南省等气候温润、水资源丰富， 使得其具有较强的资源禀赋比较优势。河南省、

河北省、山东省和吉林省等地区的资源禀赋系数大于 1，即具有一定的资源禀赋比较优势。由于生猪的养殖过程非常消耗粮食，因此粮食主产区具有相对的优势。玉米是最主要的饲料原料之一，河南省等省份（自治区）的玉米产量排名较靠前，可供充分的饲料原料，低成本优势使得生猪生产由沿海、西南等水稻产区转向中东部地区。

5.3.2 运用综合比较优势指数法

运用综合比较优势法对我国各地区生猪产业发展进行测算，考虑数据的可得性，以生猪的产量比作为参数计算。

$$AAI = \frac{A_i/A_{it}}{A/A_t} \tag{2}$$

式中： $A_i$  为某一地区猪肉的产量； $A_{it}$  为该地区所有肉类总产量； $A$  为全国猪肉总产量； $A_t$  为全国肉类总产量。 $AAI>1$  表明该地区生猪生产具有优势，其值越大，优势越强； $AAI<1$  表明该地区生猪生产不具有优势； $AAI=1$  则处于临界状态。依据不同年份  $AAI$  从大到小的顺序，列出 1995—2015 年主要年份中国在生猪生产上具有综合比较优势的地区，如表 8 所示。

表 4 中国生猪资源禀赋具有综合比较优势的省份（自治区）

年份  $AAI>1$  的省份（自治区）

年份	省份
1995 年	湖南省、贵州省、云南省、重庆市、江西省、湖北省、四川省、浙江省、福建省、广西、陕西省、山西省
2000 年	重庆市、湖南省、贵州省、云南省、江西省、湖北省、福建省、浙江省、四川省、广西、陕西省、山西省
2005 年	湖南省、云南省、福建省、贵州省、重庆市、四川省、湖北省、浙江省、江西省、广西、陕西省、广东省、陕西省、海南省
2010 年	湖南省、贵州省、福建省、陕西省、重庆市、江西省、湖北省、云南省、浙江省、四川省、山西省、河南省、安徽省
2015 年	湖南省、贵州省、浙江省、陕西省、湖北省、云南省、江西省、重庆市、四川省、山西省、河南省、广东省

注：表 8 数据根据《中国统计年鉴》《中国畜牧业年鉴》计算所得。

由表 4 分析发现，湖南省、贵州省在生猪的生产上具有较强的综合比较优势。河南省、广东省、陕西省、山西省等地综合比较优势上升较快，跻身于具有综合比较优势的省



份行列中。广西、福建省、海南省等地曾经是具有综合比较优势的省份（自治区）。通过利用资源禀赋系数法和综合比较优势指数法我们可得知，无论是猪肉资源相对比较丰富还是具有较强的综合比较优势，湖南省，四川省，河南省，贵州省都有比其他省份更强的优势，所以 2018 年 4 月应在哪些区域建设种猪场的决策问题，我们将区域缩小到在湖南省，四川省，河南省，贵州省这些省份中选择。

### 5.3.3 利用具有季节性特点的时间序列模型预测价格

#### 1) 季节性时间序列模型的介绍

在不少实际问题中，时间序列有很明显的周期规律性，例如气温、雨量、用电量等。由季节性因素或其它周期因素引起的周期性变化的时间序列，我们称为季节性时间序列，相应的模型为季节性模型。这里提到的季节，可以是自然季节，也可以是某种产品的销售季节等。显然，在现实的经济活动中，表现为季节性的时间序列是非常多的。比如，空调、取暖设备、季节性服装的生产与销售所产生的数据等。对于季节性时间序列的预测，要从数学上完全拟合其变化曲线是非常困难的。但预测的目的是为了找到时间序列的变化趋势，尽可能地做到精确。

从收集到的数据可知，每个月一斤玉米的价格和生猪的价格明显随着季节变化，因此，我们可以用具有季节性特点的时间序列来预测在 24 个月（即为两年时间）后各个省份每斤玉米的价格和生猪的价格，其中，我们以一个月为一个单位，一年分为 12 个单位，此单位的意思和季度的意思相同，但是为了区别人们定义的季度，我们用单位来表示月份。

#### 2) 季节性时间序列模型的求解

季节系数法的具体计算步骤如下：

1、收集  $m$  年的每年各季度或者各月份（每年  $n$  个季度）的时间序列原本数据  $x_{ij}$ 。其中， $i$  表示年份的序号（ $i=1,2,\dots,m$ ）， $j$  表示季度或者月份的序号（ $j=1,2,\dots,n$ ）

2、计算每年所有的季度或所有月份的算术平均值  $\bar{x}$ ，即：

$$\bar{x} = \frac{1}{k} \sum_{i=1}^m \sum_{j=1}^n x_{ij}, \quad k=mn$$

3、计算同季度或同月份数据的算术平均值  $\bar{x}_j = \sum_{i=1}^m x_{ij}$ ， $j=1,2,\dots,n$

4、计算季节系数或月份系数  $\beta_j = \bar{x}_j / \bar{x}$ 。其中  $j=1,2,\dots,n$  为季度或者月份的序号。

5、预测计算。当时间序列是按季度列出，先求出也测年份（下一年）的年加权平均：

$$Y_{m+1} = \frac{w_1 y_1 + w_2 y_2 + \dots + w_m y_m}{w_1 + w_2 + \dots + w_m}$$

式中， $y_i = \sum_{j=1}^n x_{ij}$  为 i 年份的年合计数； $w_i$  为 i 年份权数。按自然数列取值。再计算预测年份的季度平均值  $\bar{y}_{m+1}$ ： $\bar{y}_{m+1} = y_{m+1} / 4$ 。最后。预测年份第 i 季度的预测值为：

$$\bar{y}_{m+1,j} = \bar{y}_{m+1} \cdot \beta_j$$

### 3) 利用季节性时间序列模型对生猪价格的预测

由于生猪价格符合季节性时间序列模型的特点，所以我们可用这个模型来求解 24 个月 后各个省份的生猪的价格，在这里我们以湖南省的生猪价格预测为例，其他省份生猪价 格的预测方法和以下方法一致。

计算过程如下：

1. 收集 2016 年五月-2018 年五月各月份的时间序列原本数据  $x_{ij}$  [4]。其中，i 表示年份的序号 ( $i=1,2,...,m$ ) ,j 表示月份的序号 ( $j=1,2,...,n$ )，则：

$X_{2016-j} = (21.8 \ 20.54 \ 18.74 \ 18.84 \ 18.64 \ 16.87 \ 16.82 \ 17.32 \ 18.25 \ 17.43 \ 16.4$   
15.65)

$X_{2017-j} = (13.87 \ 13.37 \ 13.63 \ 14.57 \ 14.7 \ 14.3 \ 14.37 \ 15.11 \ 15.29 \ 13.45 \ 11.13$   
10.51)

2、计算每年所有月份的算术平均值  $\bar{x}$ ，即：

$$\bar{x}_{2016-2017} = \frac{1}{k} \sum_{i=1}^m \sum_{j=1}^n x_{ij} = 18.10$$

$$\bar{x}_{2016-2017} = \frac{1}{k} \sum_{i=1}^m \sum_{j=1}^n x_{ij} = 13.69$$

3、计算同季度或同月份数据的算术平均值  $\bar{x}_j = \sum_{i=1}^m x_{ij} = (17.84, 16.96, 16.18, 16.71, 16.67,$   
15.58, 15.6, 16.22, 16.77, 15.44, 13.77, 13.08)  $j=1,2,...,n$

4、计算季节系数或月份系数  $\beta_j = \bar{x}_j / \bar{x}$ 。其中  $j=1,2, ...,n$  为季度或者月份的序号。得到的 季节系数如下表：

$\beta_j = (1.1217, 1.0664, 1.0179, 1.0506, 1.0484, 0.9802, 0.9808, 1.0198, 1.0547, 0.9711$   
0.8657, 0.8226)

5、预测计算。当时间序列是按季度列出，先求出也测年份（下一年）的年加权平均：

$$Y_{m+1} = \frac{w_1 y_1 + w_2 y_2 + \cdots + w_m y_m}{w_1 + w_2 + \cdots + w_m} = 181.9667$$

式中， $y_i = \sum_{j=1}^n x_{ij}$  为 i 年份的年合计数； $w_i$  为 i 年份权数。按自然数列取值。再计算预测

年份的季度平均值  $\bar{y}_{m+1}$ :  $\bar{y}_{m+1} = y_{m+1} / 4$ 。最后。预测年份第  $i$  季度的预测值为: (代码见附录 5)

$$\bar{y}_{m+1,j} = \bar{y}_{m+1} \cdot \beta_j = (17.0093 \quad 16.1700 \quad 15.4357 \quad 15.9316 \quad 15.8982 \quad 14.8635 \\ 14.8730 \quad 15.4643 \quad 15.9936 \quad 14.7252 \quad 13.1277 \quad 12.4744)$$

所以预测 2019 年 4 月分生猪的价格为 12.4777 元/千克, 由于我们要算出两年后生猪的价格, 故让 2018 年 5 月到 2019 年 4 月份的数据假设作为已知条件, 利用 2017 年 5 月到 2019 年 4 月份生猪每个月的价格数据作为时间序列, 重复以上操作可以得到 2019 年 5 月到 2020 年 4 月份的生猪价格, 每个月的平均价格如下:

$$y_{m+1,j} = (15.7022 \quad 15.0212 \quad 14.7800 \quad 15.5102 \quad 15.5593 \quad 14.8297 \quad 14.8702 \quad 15.5471 \\ 15.9078 \quad 14.3272 \quad 12.3351 \quad 11.6877)$$

故当时时间为 2020 年 4 月份时, 湖南省的平均生猪价格为

$$S1=11.6877 \text{ 元/千克}$$

重复利用季节性时间序列模型, 我们可以得到四川省, 河南省, 贵州省这些省份在 2020 年 4 月份时 (即两年后) 的平均生猪价格分别为:

$$S2=11.8605$$

$$S3=11.1742$$

$$S4=11.9918$$

同样的道理, 由于玉米的价格明显随着季节变化, 因此, 我们也可以用具有季节性特点的时间序列来预测在 24 个月 (即为两年时间) 后各个省份每斤玉米的价格, 思路和上面求解生猪的价格一样, 在这里我们不加以重复, 根据具有季节性特点的时间序列模型, 我们可以算出湖南省, 四川省, 河南省, 贵州省这些省份在 2020 年四月份的玉米价格, 分别为: (每个省份 2016 年 5 月到 2018 年 4 月的价格如附录 )

$$C1=2111.3 \text{ 元/吨}$$

$$C2=2098.7 \text{ 元/吨}$$

$$C3=1855.5 \text{ 元/吨}$$

$$C4=2184.1 \text{ 元/吨}$$

成本方面虽然可以选取玉米价格 ( $X_1$ ) 和仔猪价格 ( $X_2$ ) 2 个变量。但由于饲料价格占生猪养殖成本的 80% 以上, 玉米又是饲料的主要组成部分, 故选择玉米价格表示养殖成本。所以当生猪的体重没增加 1 千克, 养殖户可以获得的利润等于;

$$P=S-C$$

因此通过此计算方法得当生猪的体重每增加 1 千克，位于湖南省，四川省，河南省，贵州省养殖户可以获得的利润分别等于：

$$P1=5.7761 \text{ 元}$$

$$P2=5.9841 \text{ 元}$$

$$P3=5.9788 \text{ 元}$$

$$P4=5.8763 \text{ 元}$$

由于  $P2 > P1 > P4 > P3$ ，所以我们可得知，2018 年 4 月在四川省区域内建设种猪场，在两年后我们将获得最大利润，这个结果同时也印证了四川是我国的养猪大省这一说法，很多养殖户由于多种因素大部分会选择去四川省养殖种猪。

## 六、模型的评价

### 6.1 模型的优点

- 1、灰色关联度：灰关联分析是按发展趋势做分析，对样本量的多少没有过多的要求，也不需要典型的分布规律，而且计算量比较小，其结果与定性分析结果会比较吻合。
- 2、主成分分析：可消除评价指标之间的相关影响。在对原指标变量进行变换后形成了彼此相互独立的主成分，而且实践证明指标之间相关程度越高，主成分分析效果越好。在综合评价函数中，各主成分的权数为其贡献率，它反映了该主成分包含原始数据的信息量占全部信息量的比重，这样确定权数是客观的、合理的，它克服了某些评价方法中认为确定权数的缺陷。
- 3、层次分析法：把研究对象作为一个系统，按照分解、比较判断、综合的思维方式进行决策，成为继机理分析、统计分析之后发展起来的系统分析的重要工具。系统的思想在于不割断各个因素对结果的影响，而层次分析法中每一层的权重设置最后都会直接或间接影响到结果，而且在每个层次中的每个因素对结果的影响程度都是量化的，非常清晰、明确。这种方法尤其可用于对无结构特性的系统评价以及多目标、多准则、多时期等的系统评价。

### 6.2 模型的缺点

- 1、关于灰色关联度模型：灰关联分析模型不是函数模型，是序关系模型，其技术内涵为：获取序列间的差异信息，建立差异信息空间；建立和计算差异信息比较测度；建立因子间的序关系。灰色关联空间涉及到灰关联因子空间、灰关联差异信息空间等。灰色关联因子空间是灰关联分析的基础，其是由具备“可比性”、“可接近性”、“极性一致性”的序列构成，灰关联差异信

息空间则是灰关联分析的依据。

2、关于主成分分析模型：主成分的解释其含义一般多少带有点模糊性,不像原始变量的含义那么清楚、确切,这是变量降维过程中不得不付出的代价。因此,提取的主成分个数  $m$  通常应明显小于原始变量个数  $p$  (除非  $p$  本身较小),否则维数降低的“利”可能抵不过主成分含义不如原始变量清楚的“弊”。

3、层次分析法：不能为决策提供新方案层次分析法的作用是从备选方案中选择较优者。这个作用正好说明了层次分析法只能从原有方案中进行选取,而不能为决策者提供解决问题的新方案。这样,我们在应用层次分析法的时候,可能就会有这样一个情况,就是我们自身的创造能力不够,造成了我们尽管在我们想出来的众多方案里选了一个最好的出来,但其效果仍然不够人家企业所做出来的效果好。而对于大部分决策者来说,如果一种分析工具能替我分析出在我已知的方案里的最优者,然后指出已知方案的不足,又或者甚至再提出改进方案的话,这种分析工具才是比较完美的。但显然,层次分析法还没能做到这点。

## 七、参考文献

[1] 小区开放对道路通行的影响, 2016 年数模国赛优秀论文

[2] 姜启源, 谢金星, 叶俊, 《数学模型 (第三版)》高等教育出版社, 2003. 8.

第 236 页至第 237 页

[3] 谭莹, 张俊艳, 何勤英. 中国生猪生产布局演进历程及区域优化研究[A]广州: 华南农业大学, 2017

[4] 猪 e 网 <http://bj.zhue.com.cn/zoushi.php>

[5] 中国统计年鉴 <http://www.stats.gov.cn/tjsj/nds/j/>

## 八、附录

附录 1

```
clc;clear;
fid=fopen('zhuchangwendu.txt','r');
good1=0;
good2=0;
good3=0;
good4=0;
good5=0;
sum1=262;
for i=1:sum1
```

```

A=fscanf(fid,'%f',119);
alpha=0.05;
[mu,sigma]=normfit(A);
p1=normcdf(A,mu,sigma);
[H1,s1]=kstest(A,[A,p1],alpha)
n=length(A);
if H1==0
    good1=good1+1;
    %disp('该数据源服从正态分布。')
else
    %disp('该数据源不服从正态分布。')
end
phat=gamfit(A,alpha);
p2=gamcdf(A,phat(1),phat(2));
[H2,s2]=kstest(A,[A,p2],alpha)
if H2==0
    good2=good2+1;
    %disp('该数据源服从 $\gamma$ 分布。')
else
    %disp('该数据源不服从 $\gamma$ 分布。')
end
lamda=poissfit(A,alpha);
p3=poisscdf(A,lamda);
[H3,s3]=kstest(A,[A,p3],alpha)
if H3==0
    good3=good3+1;
    %disp('该数据源服从泊松分布。')
else
    %disp('该数据源不服从泊松分布。')
end
mu=expfit(A,alpha);
p4=expcdf(A,mu);
[H4,s4]=kstest(A,[A,p4],alpha)
if H4==0
    good4=good4+1;
    %disp('该数据源服从指数分布。')
else
    %disp('该数据源不服从指数分布。')
end
[phat,pci]=raylfit(A,alpha);
p5=raylcdf(A,phat);
[H5,s5]=kstest(A,[A,p5],alpha)
if H5==0
    good5=good5+1;

```

```

        %disp('该数据源服从 rayleigh 分布。')
    else
        %disp('该数据源不服从 rayleigh 分布。')
    end
end
end
sprintf('服从正态分布的数据的比例 %.5f', good1/sum1)
sprintf('服从  $\gamma$  分布的数据的比例 %.5f', good2/sum1)
sprintf('服从泊松分布的数据的比例 %.5f', good3/sum1)
sprintf('服从指数分布的数据的比例 %.5f', good4/sum1)
sprintf('服从 rayleigh 分布的数据的比例 %.5f', good5/sum1)

```

附录 2

```

clc, clear
P=[1.0000 5.0000;
    0.2 1.0000];
[L, G]=eig(P);
w=L(:,1)/sum(L(:,1))
max=max(eig(P))
n=size(P);
n=size(P);
CI=(max-2)/(2-1);

```

附录 3

```

clear;clc;
load yuanshishuju.txt
[m,n]=size(yuanshishuju);
yangben=yuanshishuju(:,1:2);
yangben=yangben';
fangzhen=yuanshishuju(:,3:n);
fangzhen=fangzhen';%待判数据
[rows,cols]=size(fangzhen);
p=0.5;%分辨系数
[m,n]=size(yangben);
R=[];
for irow=1:rows
yy=fangzhen(irow,:);

data=[yy;yangben];
data_gyh1=mean(yangben)
for i=1:m+1
for j=1:n
data_gyh(i,j)=data(i,j)/data_gyh1(j);
end
end
for i=2:m+1
for j=1:n

```

```

Dij(i-1,j)=abs(data_gyh(1,j)-data_gyh(i,j)); end
end
Dijmax=max(max(Dij))
Dijmin=min(min(Dij))
for i=1:m
for j=1:n
Lij(i,j)=(Dijmin+p*Dijmax)/(Dij(i,j)+p*Dijmax);
end
end
LijRowSum=sum(Lij');
for i=1:m
Rij(i)=LijRowSum(i)/n;
end
R=[R;Rij];
end
R

```

#### 附录 4

```

clc,clear
load zcffxf.txt %把原始数据保存在纯文本文件 gj.txt 中
gj=zscore(zcffxf); %数据标准化
r=corrcoef(gj); %计算相关系数矩阵-606-
%下面利用相关系数矩阵进行主成分分析， x 的列为 r 的特征向量，即主成分的系数
[x,y,z]=pcacov(r) %y 为 r 的特征值， z 为各个主成分的贡献率
contr=cumsum(z)
f=repmat(sign(sum(x)),size(x,1),1); %构造与 x 同维数的元素为±1 的矩阵
x=x.*f; %修改特征向量的正负号，每个特征向量乘以所有分量和的符号函数值
num=3; %num 为选取的主成分的个数
df=gj*x(:,1:num); %计算各个主成分的得分
tf=df*z(1:num)/100; %计算综合得分
[stf,ind]=sort(tf,'descend'); %把得分按照从高到低的次序排列
stf=stf', ind=ind'

```

#### 附录 5

```

clc,clear
a= [21.8    20.54    18.74    18.84    18.64    16.87    16.82    17.32    18.25
    17.43    16.4     15.65
13.87    13.37    13.63    14.57    14.7     14.3     14.37    15.11    15.29    13.45
    11.13    10.51
];%计算湖南在 2020 年 4 月份的生猪价格
[m,n] = size(a);
a_mean = mean(mean(a));
aj_mean = mean(a);
bj = aj_mean/a_mean
w = 1:m;

```



```
yhat = w*sum(a,2)/sum(w)
yjmean = yhat/n
yjhat = yjmean*bj
a(3,:)=yjhat;
a=a(2:3,:)
[m,n] = size(a);
a_mean = mean(mean(a));
aj_mean = mean(a);
bj = aj_mean/a_mean
w = 1:m;
yhat = w*sum(a,2)/sum(w)
yjmean = yhat/n
yjhat = yjmean*bj
format
```