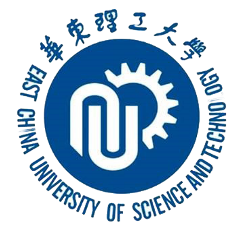
****  ** **

**中国研究生创新实践系列大赛**

**“华为杯”第十七届中国研究生**

**数学建模竞赛**

|  |  |
| --- | --- |
| **学 校** | **长春理工大学** |
| **参赛队号** | **20101860020** |
| **队员姓名** | **1.张耹铭** |
| **2.刘阳** |
| **3.谷晓雁** |

**中国研究生创新实践系列大赛**

**“华为杯”第十七届中国研究生**

**数学建模竞赛**

题 目 （方法+对象+目的）

摘 要：

本文研究了XXX问题。主要创新点在于（创新点1解决了XX问题，取得了XX有益效果），（创新点2解决了XX问题，取得了XX有益效果）以及（创新点3解决了XX问题，取得了XX有益效果）。

问题一：我们（采用了XX方法），主要包括（主要的处理步骤），（获得的有益效果，要有数据支持，关键点加粗）。

问题二：我们（采用了XX方法），主要包括（主要的处理步骤），（获得的有益效果，要有数据支持，关键点加粗）。

问题三：我们（采用了XX方法），主要包括（主要的处理步骤），（获得的有益效果，要有数据支持，关键点加粗）。

最后，对于XXX模型的优缺点进行了评价，本文提出了以下几点（未来的改进措施？）：（1）（2）（3）。

关键字：关键词1；关键词2；关键词3；关键词4；关键词5

1. 问题重述

1.1 问题背景

汽油辛烷值（RON）是车用汽油最重要的品质指标之一，辛烷值越高表示汽油的抗爆性越好，提高辛烷值对汽油动力经济性能有十分重要的意义。辛烷值与汽油中的硫、烯烃等含量有关。为了符合相关的国家标准，必须降低汽油中的硫、烯烃含量，同时也要尽量保持较高的辛烷值。

过去采用数据关联和机理分析的方法对化工过程建模，但由于工艺过程的复杂性、设备的多样性以及操作变量（控制变量）之间具有高度非线性和相互强耦联，而且辛烷值测定具有滞后性，对数据处理有较高的要求。辛烷值和操作变量之间的数学模型分析工作较少，如何控制操作变量、建立损失预测模型对于提高汽油精制的经济效益具有重要意义。

某石化企业的催化裂化汽油精制脱硫装置运行4年，积累了大量的历史数据，其汽油产品的辛烷值损失平均达到了1.37个单位，而同类装置的最小损失值只有0.6个单位。，所以在对辛烷值损失量的建模方面有较大的优化空间。

中国石油化工研究所开发的DSO选择性汽油加氢脱硫技术，不仅具有较高的加氢脱硫活性及脱硫选择性，而且具有很好的原料适应性，可以灵活地控制加氢脱硫反应深度，在达到产品硫含量要求的同时，辛烷值损失较少，能够满足装置长周期运行的要求。DSO技术在催化裂化汽油加氢装置上的应用\_陈晓华.pdf



图1.1 催化裂化汽油精制脱硫装置

1.2 问题重述

题目包含4个附件，附件一提供了某石化企业2017年4月17日至2020年5月26日共325个样本数据以及对应的367个变量，附件二是确定样本数据的方法，附件三提供了285号和313号样本的原始数据，附件四是354个操作变量的信息，要求利用上述数据解决以下问题：

问题一：参考附件一的325个数据样本的预处理结果，根据附件二的样本确定方法对附件三中285号、313号样本原始数据进行整定、筛选，并将数据求取平均值加入到附件一相应的样本号中。

问题二：在建立降低辛烷值损失的模型中，使用先降维后建模的方法，通过降维的方法从367个变量中筛选出具有代表性、独立性的主要变量，并详细说明建模主要变量的筛选过程及合理性。

问题三：采用上述样本和建模的主要变量，通过数据挖掘技术建立辛烷值损失预测模型，并进行模型验证。

问题四：要求在产品硫含量不大于5ug/g的前提下，保证优化过程中原料、待生吸附剂、再生吸附剂的性质保持不变，利用已建立辛烷值损失预测模型获得325个数据样本，求预测的辛烷值不大于附件一中原有辛烷值70%的样本对应主要变量优化后的操作条件。

问题五：针对问题四选取的已优化主要操作变量，结合其每次允许的调整幅度，对133号样本以图形展示其主要操作变量优化调整过程中对应的汽油辛烷值和硫含量的变化轨迹。



2. 模型假设

考虑到实际情况，本文做出如下假设：

1. 模型假设，如果不够可以根据题目条件进行补充。
2. 模型假设，如果不够可以根据题目条件进行补充。
3. 模型假设，如果不够可以根据题目条件进行补充。
4. 模型假设，如果不够可以根据题目条件进行补充。
5. 模型假设，如果不够可以根据题目条件进行补充。
6. 模型假设，如果不够可以根据题目条件进行补充。



3. 符号系统

|  |  |  |
| --- | --- | --- |
| 符号 | 符号说明 | 单位 |
|  | 原始观测变量 | - |
|  | 原始观测变量的公共因子 | - |
|  | 因子载荷矩阵 | - |
|  | 原始观测变量的特殊因子 | - |
|  | 第个原始变量对第个因子的得分 | - |
|  | 回归参数 | - |
|  | 随机扰动 | - |
|  | 残差 | - |
|  | 相对误差 | - |

4. 数据处理

4.1 问题分析

问题一要求按照附件二中的样本确定方法对附件三中285号和313号样本原始数据进行数据预处理，并将获得的数据插入到附件一中。附件一中325个样本数据和附件三的样本原始数据都需要满足附件四中规定的操作变量的取值范围，同时附件一和附件四的数据还存在数据缺失项，采用插值法补全数据缺失项，并根据拉依达准则进行异常值检测。

4.2 数据处理步骤

数据处理内容包括附件一的样本数据和附件三的原始数据，附件三为附件一中两个样本的原始数据，所以我们先对附件三中原始数据进行处理，其处理步骤如下：

1. 最大最小限幅筛选法：根据附件四提供的数据变量操作范围，对附件三的原始数据进行筛选。
2. 插值：在保留的49条原始数据中，有5处数据项存在缺失，使用前后两小时数据平均值进行了插值补全。
3. 均值处理：使用2小时内操作变量的平均值作为辛烷值操作变量数据插入附件一的样本数据中。
4. 拉依达准则（3准则）进行异常值检测。

随后我们对附件一中的样本数据进行处理：

1. 覆盖：将原始数据处理得到的两条样本数据插入到附件一中。
2. 插值：插值补全122项缺失数据，剩下的18个位点仅仅含有部分时间位点数据，无法使用插值补充，直接将此位点删除，保留了336项操作变量。
3. 最大最小限幅筛选法：根据附件四提供的操作变量范围，发现样本号为301、302、303、304的操作变量超出范围，作为无效样本数据删除，最终保留了321条有效样本数据。

4.3 数据处理与结果分析

附件三中名为“非净化风进装置流量”的部分数据在删除前后进行比较，根据附件四可知，非净化风进装置流量的取值范围为0-900，通过最大最小限幅筛选法将原始数据中超出范围的数据删除。图5.1为非净化风进装置流量删除超出范围数据前后的对比，在原始异常数据中，删除了第七次采样超出0-900范围的数据，剩余所有数据都在操作范围内。



表4.1 原始数据删除前后对比

|  |  |
| --- | --- |
| 删除前 | 删除后 |
| 543.3043 | 543.3043 |
| 595.8862 | 595.8862 |
| 648.4681 | 648.4681 |
| 511.6857 | 511.6857 |
| 446.7246 | 446.7246 |
| 463.643 | 463.6430 |
| 910.5219\* |  |
| 784.4914 | 784.4914 |
| 514.8187 | 514.8187 |
| 473.3051 | 473.3051 |
| 463.4757 | 463.4757 |
| 422.0896 | 422.0896 |
| 472.5419 | 472.5419 |
| 866.6923 | 866.6923 |
| 420.2141 | 420.2141 |

注：灰底上标\*为删除值

|  |  |
| --- | --- |
|  |  |
| 图4.1 数据删除前后 | |

（处理了多少数据，有什么有益效果）

根据附件四提供的数据变量操作范围，对附件三的原始数据进行筛选，共计删除了31条样本原始数据，保留了285号样本原始数据40 条，313号样本原始数据9条。在保留的49条原始数据中，有5处数据项存在缺失，使用前后两小时数据平均值进行了插值补全。使用2小时内操作变量的平均值作为辛烷值操作变量数据插入附件一的样本数据中。

将超出范围的异常值删除后，对缺失数据进行插值填补，以2#催化汽油进装置流量的缺失数据为例，选取SPSS数据分析软件对2#催化汽油进装置流量的4项缺失数据填补，填补效果良好。图5.2是插值前后的对比图，2#催化汽油进装置流量数据填补完成后，整体数据呈平稳趋势，保持在40-60之间，无异常值，插值效果好，可以采用。

表4.2 插值前后数据对比

|  |  |
| --- | --- |
| 插值前 | 插值后 |
| 53.86215 | 53.86215 |
| 46.35702 | 46.35702 |
| 44.51688 | 44.51688 |
| 0\* | 52.38405 |
| 0\* | 52.38405 |
| 0\* | 52.38405 |
| 0\* | 52.38405 |
| 48.84124 | 48.84124 |
| 51.11994 | 51.11994 |
| 51.83154 | 51.83154 |
| 46.70276 | 46.70276 |

注：灰底上标\*的为缺失值

|  |  |
| --- | --- |
|  |  |
| 图4.2 插值前后对比 | |

（处理了多少数据，有什么有益效果）



5. 问题二：寻找建模主要变量、建立辛烷值损失预测模型

5.1问题分析

问题二要求从367个变量中选取30个以下具有代表性、独立性的的主要变量。附件一共给出了7个原料性质、2个再生吸附性质、2个产品性质以及354个操作变量共367个建立辛烷值损失模型，采用先降维后建模的方法，忽略次要因素，分析影响模型的主要因素，拟采用因子分析进行降维求解。

要求将选取的主要变量作为因变量，辛烷值的损失作为自变量，建立预测模型。主要变量对辛烷值损失量有不同的贡献值，拟采用多元回归和BP神经网络的方法进行预测。

（考虑加点东西）

5.2模型的建立

因子分析：

因子分析法的核心就是对若干综合指标进行因子分析并提取公共因子，再以每个因子的方差贡献率作为权术与该因子的得分乘数之和构造得分函数。

本文在建立辛烷值损失模型过程中采用的因子分析的数学模型如下

|  |  |
| --- | --- |
|  | （5.1） |

上式中，向量是原始数据变量，是公共因子，即选择的主要变量，是公共因子的系数，称为因子载荷矩阵，称为因子载荷，是第个原有变量在第个因子的载荷，是的特殊因子，是不能被前个公共因子包含的部分。

多元回归模型：

多元线性回归描述的是一个变量受到多个不同变量的影响的模型，建立多元线性回归模型时，为了保证回归模型具有优良的解释能力和预测效果，应首先注意自变量选择，其准则是：

a) 自变量对因变量必须具有显著的影响，并呈密切的线性相关；

b) 自变量与因变量之间的线性相关必须是真实的，而不是形式上的；

c) 自变量之间应具有一定的互斥性，即自变量之间的相关程度不应高于自变量与因变量之间的相关程度；

d) 自变量应具有完整的统计数据，其预测值容易确定。

多元线性回归模型的一般表达式如下：

|  |  |
| --- | --- |
|  | （5.2） |

式中为被解释变量（因变量），为解释变量（自变量），为解释变量的数目，为回归参数矩阵，为随机扰动。

5.3模型求解与分析

因子分析模型求解与分析

（思考加点东西）

因子分析法的求解步骤如下：

1. 数据标准化，将数据进行无量纲化处理，消除单位的影响。

数据标准化处理方法如下：

|  |  |
| --- | --- |
|  | （5.2） |

式中为标准化处理后的数据，为原始数据，为原始数据均值，为原始变量数据的标准差。

1. 确定待分析的原始变量是否适合进行因子分析。

本文采用KMO检验和巴特利特（Bartlett）球形检验。当KMO>0.5,巴特利特球形检验的相伴概率值小于0.05时，适合因子分析。运用SPSS数据分析软件进行KMO和巴特利特检验。其结果如下：

表5.1 KMO和巴特利特检验

|  |  |  |
| --- | --- | --- |
| KMO取样适切性量数 | 0.774 | |
| 巴特利特球形度检验 | 近似卡方 | 20039.552 |
| 自由度 | 946 |
| 显著性 | 0.000 |

如表5.1所示，KMO值为0.774大于0.5，巴特利特球形检验的相伴概率值近似为0.000，在5%的显著性水平下拒绝原假设，认为原始变量间存在关联性，因此原有指标适合做因子分析。

1. 确定因子的个数

使用Matlab建立原始变量的相关系数矩阵，求取特征值和特征向量，以碎石图的形式表示，部分具体结果如下：（具体见附件1）

表5.2 解释方差总和

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 因子 | 特征值 | 方差百分比 | 累计百分比 | 因子 | 特征值 | 方差百分比 | 累计百分比 |
| 1 | 113.500 | 32.615 | 32.615 | 21 | 2.369 | 0.681 | 84.962 |
| 2 | 39.230 | 11.273 | 43.888 | 22 | 2.266 | 0.651 | 85.614 |
| 3 | 23.422 | 6.730 | 50.618 | 23 | 1.861 | 0.535 | 86.148 |
| 4 | 20.393 | 5.860 | 56.479 | 24 | 1.827 | 0.525 | 86.673 |
| 5 | 14.136 | 4.062 | 60.540 | 25 | 1.768 | 0.508 | 87.182 |
| 6 | 11.885 | 3.415 | 63.956 | 26 | 1.680 | 0.483 | 87.664 |
| 7 | 9.848 | 2.830 | 66.786 | 27 | 1.522 | 0.437 | 88.102 |
| 8 | 8.299 | 2.385 | 69.170 | 28 | 1.467 | 0.422 | 88.523 |
| 9 | 7.282 | 2.092 | 71.263 | 29 | 1.415 | 0.407 | 88.930 |
| 10 | 6.469 | 1.859 | 73.122 | 30 | 1.391 | 0.400 | 89.329 |
| 11 | 5.896 | 1.694 | 74.816 | 31 | 1.244 | 0.357 | 89.687 |
| 12 | 5.073 | 1.458 | 76.274 | 32 | 1.233 | 0.354 | 90.041 |
| 13 | 4.849 | 1.393 | 77.667 | 33 | 1.228 | 0.353 | 90.394 |
| 14 | 4.232 | 1.216 | 78.883 | 34 | 1.151 | 0.331 | 90.725 |
| 15 | 3.991 | 1.147 | 80.030 | 35 | 1.110 | 0.319 | 91.044 |
| 16 | 3.618 | 1.040 | 81.070 | 36 | 1.072 | 0.308 | 91.352 |
| 17 | 3.095 | 0.889 | 81.959 | 37 | 1.030 | 0.296 | 91.648 |
| 18 | 2.864 | 0.823 | 82.782 | 38 | 0.984 | 0.283 | 91.931 |
| 19 | 2.820 | 0.810 | 83.593 | 39 | 0.938 | 0.270 | 92.200 |
| 20 | 2.398 | 0.689 | 84.282 | 40 | 0.926 | 0.266 | 92.466 |

表5.2显示，共有37个因子的特征值大于1，基于过程中内定取特征值大于1的原则，初步筛选1至37个因子，方差积累量为91.648%，表明前37个因子的信息可以代表原始数据。问题二要求选取30个以下的主要变量，第22个因子的方差累计量为85.641%，需要对第22个因子到第27个因子的特征值和方差累计量的变化趋势进行分析，确定选取的因子数量。

图5.1 因子碎石图

由图5.1的因子碎石图分析可知，因子1至因子10特征值变化剧烈，从因子10至因子40特征值变化趋于平缓，对因子5到因子40碎石图进行放大，观察特征值的变化趋势，结合表5.1的特征值及方差的百分比，选择1至26号因子作为主要变量，共保留26个公共因子。

1. 旋转因子载荷矩阵

本文采用主成分分析法对因子载荷矩阵进行估计，同时为了更好地了解每个公共因子的意义，便于对公共因子命名和解释，采用最大方差法将因子载荷矩阵旋转，针对旋转后的因子载荷矩阵进行分析。表5.3是部分旋转因子载荷矩阵：（具体见附件1）

表5.3 旋转因子载荷矩阵

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 因 子 | | | | | | | |
| 1 | 2 | 3 | 4 | … | 24 | 25 | 26 |
| K-103A排气压力 | -0.895 | -0.304 | -0.193 | -0.149 | … | 0 | -0.023 | -0.024 |
| K-103A排气温度 | -0.897 | -0.295 | -0.198 | -0.151 | … | 0 | -0.024 | -0.024 |
| D-123压力 | -0.905 | -0.208 | -0.172 | -0.166 | … | -0 | -0.025 | -0.006 |
| E-206壳程出口管温度 | -0.922 | -0.199 | -0.094 | -0.082 | … | -0 | -0.008 | -0.011 |
| 反吹氢气压力 | 0.985 | 0.031 | 0.013 | 0.046 | … | -0 | -0.025 | 0.016 |
| ME-103进出口差压 | 0.984 | 0.042 | 0.019 | 0.047 | … | -0 | -0.025 | 0.016 |
| 过滤器ME-101出口温度 | 0.976 | 0.151 | -0.004 | 0.041 | … | -0 | -0.029 | 0.015 |
| … | … | … | … | … | … | … | … | … |
| D-102温度 | 0.001 | 0.097 | 0.134 | 0.138 | … | 1 | 0.123 | 0.179 |
| 原料进装置流量累计 | 0.092 | -0.132 | -0.14 | 0.047 | … | 0 | 0.717 | 0.038 |
| 燃料气进装置压力 | -0.169 | -0.157 | -0.121 | -0.028 | … | 0 | 0.343 | -0.204 |
| 反吹氢气温度 | 0.364 | -0.163 | 0.177 | 0.042 | … | 0 | -0.45 | -0.079 |
| 再生器下部温度 | -0.345 | -0.204 | 0.034 | -0.038 | … | -0 | -0.029 | -0.799 |
| 再生烟气氧含量 | -0.336 | -0.206 | -0.021 | -0.034 | … | -0 | -0.029 | -0.8 |

表5.2共26个公共因子，368个原始变量（？），分析每个公共因子对应的旋转后的因子系数，发现有很明显地两极分化，可以更清晰地确定每个公共因子的对应变量。通过对每个公共因子系数绝对值大小的统计，确认原始变量对每个公共因子的贡献值，公共因子和原始变量的对应关系如下：

表5.4 公共因子对应原始变量个数

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 因子 | 变量个数 | 因子 | 变量个数 | 因子 | 变量个数 |
| 1 | 143 | 10 | 8 | 19 | 4 |
| 2 | 35 | 11 | 7 | 20 | 4 |
| 3 | 30 | 12 | 8 | 21 | 2 |
| 4 | 14 | 13 | 5 | 22 | 1 |
| 5 | 11 | 14 | 4 | 23 | 4 |
| 6 | 9 | 15 | 3 | 24 | 1 |
| 7 | 13 | 16 | 1 | 25 | 3 |
| 8 | 13 | 17 | 5 | 26 | 2 |
| 9 | 13 | 18 | 4 |  |  |

表5.4表明，前9个公共因子对应的原始变量较多，其中公共因子1包含143个原始变量，占比最大，公共因子14、公共因子25各对应三个原始变量，公共因子21、公共因子26各对应2个原始变量，公共因子16、公共因子22和公共因子24都只对应1个原始变量。下表是26个公共因子对应的具体原始变量及主要因素，由主要因素指向每个公共因子的意义，用于解释和命名。

表5.5 公共因子对应主要变量

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 因子 | 变量 | 主要因素 | 因子 | 变量 | 主要因素 |
| 1 | D-123压力 | 与预加氢反应器有关 | 14 | 再生进料罐/LH差压 | 与反应温度压力有关 |
| E-206壳程出口管温度 | 硫含量,μg/g |
| ... | ... |
| K-102B进气压力 | R-101顶反应产物出口管温度 |
| 1.1步骤PIC2401B.OP | 反应器上部温度 |
| 2 | 8.0MPa氢气至循环氢压缩机入口 | 与预加氢反应器有关 | 15 | E-101管程出口总管压力 | 与反应有关 |
| 密度(20℃),kg/m³ | 冷氮气过滤器ME-114差压 |
| ... | … |
| K-103B排气温度 | 原料进装置温度 |
| D-124压力 | 原料泵出口流量 |
| ... | ... | ... | ... | ... | ... |
| 12 | 反应器顶部压力 | 与加氢脱硫反应器有关 | 25 | 净化风进装置流量 | 与物料消耗有关 |
| D105压力 | 循环水进装置流量 |
| 还原器压力 | 1.0MPa蒸汽进装置温度 |
| F-101辐射室底部压力 | D-123蒸汽出口流量 |
| 13 | K101机出口压力 | 与原料有关 | 26 | 再生器下部温度  再生烟气氧含量 | 与再生器有关 |
| E-101管程入口总管压力 |
| D-101脱水包液位 |
| D104去稳定塔流量 |

（4）计算每一个样本的因子得分

公共因子反映各原始变量的相关关系，使用公共因子代表原始变量，需要使用原始变量的观测值来计算各个公共因子的得分，其数学模型为：

|  |  |
| --- | --- |
|  | （5.3） |

式中，为标准化后的数据，为第个原始变量对第个因子的得分。

选取部分因子得分矩阵如下：（具体见附件1）

表5.6 因子得分系数矩阵

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 因 子 | | | | | | | |
| 1 | 2 | 3 | 4 | … | 24 | 25 | 26 |
| 硫含量,μg/g | -0.001 | -0.013 | 0.008 | -0.005 | … | -0.014 | -0.018 | -0.002 |
| 辛烷值RON | 0.000 | -0.012 | -0.013 | -0.002 | … | 0.036 | -0.038 | -0.031 |
| 饱和烃,v%（烷烃+环烷烃） | -0.005 | 0.005 | -0.001 | -0.003 | … | 0.008 | 0.022 | -0.015 |
| 烯烃,v% | 0.008 | -0.011 | -0.001 | 0.004 | … | 0.016 | 0.012 | 0.009 |
| 芳烃,v% | -0.010 | 0.015 | 0.005 | -0.004 | … | -0.062 | -0.084 | 0.014 |
| 溴值,gBr/100g | 0.002 | -0.017 | -0.003 | -0.014 | … | -0.016 | -0.009 | 0.029 |
| 密度(20℃),kg/m³ | -0.002 | 0.008 | 0.009 | 0.001 | … | 0.020 | -0.109 | -0.024 |
| … | … | … | … | … | … | … | … | … |
| 汽油产品去气分累积流量 | 0.002 | 0.011 | 0.006 | -0.093 | … | 0.017 | -0.012 | 0.022 |
| 8.0MPa氢气至循环氢压缩机入口 | 0.000 | -0.029 | 0.001 | 0.003 | … | 0.034 | 0.005 | 0.000 |
| 8.0MPa氢气至循环氢压缩机入口 | -0.011 | 0.000 | 0.000 | -0.002 | … | 0.007 | -0.008 | 0.005 |
| 8.0MPa氢气至反吹氢压缩机出口 | 0.002 | 0.011 | 0.006 | -0.093 | … | 0.017 | -0.012 | 0.022 |
| 8.0MPa氢气至反吹氢压缩机出口 | 0.000 | -0.029 | 0.001 | 0.003 | … | 0.034 | 0.005 | 0.000 |
| D101原料缓冲罐压力 | -0.011 | 0.000 | 0.000 | -0.002 | … | 0.007 | -0.008 | 0.005 |

将表5.5的系数带入（5.3）中可以得到对应的26个公共因子的函数，同时根据各个公共因子的方差百分比计算公共因子的综合得分：

|  |  |
| --- | --- |
|  | （5.4） |

多元回归模型求解及分析

对辛烷值损失预测模型采用多元线性分析模型进行预测，预测目标是辛烷值的损失值，辛烷值损失值的影响因素是因子分析确定的26个公共因子。通过对预处理的321条样本数据进行多元线性回归，得到的部分预测数据如下：（具体见附件2）

表5.6 多元线性回归

|  |  |  |  |
| --- | --- | --- | --- |
| 真实值 | 预测值 | 残差 | 相对误差 |
| 1.38 | 1.35513 | -0.02487 | 0.01802 |
| 1.18 | 1.27855 | 0.09855 | 0.08351 |
| 1.38 | 1.30974 | -0.07026 | 0.05091 |
| 1.38 | 1.31217 | -0.06783 | 0.04915 |
| 1.28 | 1.35871 | 0.07871 | 0.06149 |
| 1.41 | 1.34358 | -0.06642 | 0.04710 |
| 1.20 | 1.31609 | 0.11609 | 0.09674 |
| 1.30 | 1.32514 | 0.02514 | 0.01934 |
| 1.10 | 1.34217 | 0.24217 | 0.22016 |
| 1.40 | 1.38911 | -0.01089 | 0.00778 |
| … | … | … | … |
| 1.15 | 1.25113 | 0.10113 | 0.08794 |
| 1.15 | 1.21824 | 0.06824 | 0.05934 |
| 1.15 | 1.10056 | -0.04944 | 0.04299 |
| 1.05 | 1.12883 | 0.07883 | 0.07508 |
| 1.25 | 1.12419 | -0.12581 | 0.10065 |
| 1.15 | 1.20300 | 0.05300 | 0.04608 |
| 1.25 | 1.13387 | -0.11613 | 0.09291 |
| 1.15 | 1.25231 | 0.10231 | 0.08897 |
| 1.25 | 1.24231 | -0.00769 | 0.00615 |
| 1.35 | 1.27309 | -0.07691 | 0.05697 |
| 1.28 | 1.33931 | 0.05931 | 0.04634 |
| 1.25 | 1.26374 | 0.01374 | 0.01099 |



图5.2 残差图

残差是真实值与预测值的差，残差图是以自变量为横坐标，残差为纵坐标做出的散点图。通过残差所提供的信息，分析数据的可靠性、周期性或其他干扰，可用于分析关于误差项假定的合理性以及线性回归关系的假定的可能性。残差平方和反应除去自变量和因变量的线性关系外，所有其他因素影响的总和，残差平方和越小，说明预测越准确。通过对残差的分析，共有321个残差，其中大于0.6的有4个，大于0.5的有8个，大于1的只有一个，有97.5078%的残差落在[-0.5 0.5]区间内，且此区间范围内所有散点随机分布，没有固定的趋势，不存在异方差的情况，说明模型选择存在合理性。

相对误差定义为绝对误差与约定真值的比值，即

|  |  |
| --- | --- |
|  | （） |

对表5.6的相对误差进行统计分析，发现相对误差大于0.3的有18个，相对误差在0.3以内的占比较重，为94.3925%，可以认为多元线性回归模型具有一定合理性。

（相关指数要不要说）

相关指数用来衡量回归模型的拟合效果，越大，残差平方和越小，回归模型的拟合效果越好，相关指数的公式如下：

|  |  |
| --- | --- |
|  | （） |

式中是第个数据，是回归值，是数据均值。

F检验法（如何假设的）

得到的回归参数矩阵如下：

|  |  |  |  |
| --- | --- | --- | --- |
| 系数 | 因变量 | 系数 | 因变量 |
| 0.073945 | 常数项 | 824.2377 | 14 |
| 15759.58 | 1 | -205.375 | 15 |
| 6749.597 | 2 | -190.192 | 16 |
| -993.655 | 3 | 669.5576 | 17 |
| 820.6517 | 4 | 2592.382 | 18 |
| -1894.63 | 5 | -4341.61 | 19 |
| 1220.185 | 6 | -1573.5 | 20 |
| 866.8945 | 7 | -255.626 | 21 |
| 1628.206 | 8 | -327.865 | 22 |
| 4194.193 | 9 | 1309.526 | 23 |
| -445.591 | 10 | -449.963 | 24 |
| -1308.49 | 11 | 2177.433 | 25 |
| -1102.15 | 12 | 1142.167 | 26 |
| -3675.99 | 13 |  |  |

6. （有具体写“问题三：具体问题”，否则“问题三的建模与求解”）

问题三（问题的具体分析），（我们拟采用的方法），其建模求解流程框图如下：

图7.1 问题三建模求解流程

据图对建模求解的具体方法进行说明。

表7.1（问题三的数据处理结果）

|  |  |  |  |
| --- | --- | --- | --- |
|  | 项目1 | 项目2 | 项目3 |
| 种类1 |  |  |  |
| 种类2 |  |  |  |
| 种类3 |  |  |  |

对问题求解的结果进行分析。

8. 模型评价与改进

8.1 模型的优点

1. 这里写模型的优点一。

2. 这里写模型的优点二。

3. 这里写模型的优点三。

8.2 模型的缺点

1. 这里写模型的缺点一。

2. 这里写模型的缺点二。

8.3 模型的改进与推广

本文提出的模型和方法可推广应用于XXXXXXXXX：

1. 本文模型在推广过程的普遍适用性（属优点）。
2. 本文模型在对同一领域的其他研究的参考意义和价值（属优点）。
3. 本文模型在推广过程中可能遇到的困难和不足之处（属缺点）。

参考文献

[1]左苏. 基于主成分回归模型的工程项目成本预测[D].扬州大学,2014.

附录

程序1：XXXXX程序

程序2：XXXXX程序

程序3：XXXXX程序

程序4：XXXXX程序