承 诺 书

我们仔细阅读了大学生数学建模竞赛的竞赛规则.

我们完全明白，在竞赛开始后参赛队员不能以任何方式（包括电话、电子邮件、网上咨询等）与队外的任何人（包括指导教师）研究、讨论与赛题有关的问题。

我们知道，抄袭别人的成果是违反竞赛规则的, 如果引用别人的成果或其他公开的资料（包括网上查到的资料），必须按照规定的参考文献的表述方式在正文引用处和参考文献中明确列出。

我们郑重承诺，严格遵守竞赛规则，以保证竞赛的公正、公平性。如有违反竞赛规则的行为，我们将受到严肃处理。

我们的题目是：基于保理业务中授信机制评估和违约风险预测的数学模型分析

参赛年级是（一年级，二年级以上，研究生）： 二年级以上

所属学院（请填写完整的全名，可填多个）： 计算机

参赛队员姓名学号：1. 杨旭（学号：2016060202023）

(打印并签名) 2. 江科成（学号：2016060202011）

3. 吴嘉津（学号：2016020906006）

指导教师或指导教师组负责人 (有的话打印)： 王志勇

是否愿意参加国内赛（是，否）： 是

日期：2018 年 5 月 20 日

报名队号（请查阅《2018校内赛报名队信息-0517》后填写）：

H113

**2018电子科技大学大学生数学建模竞赛**

**编 号 专 用 页**

**报名队号（请查阅《2018校内赛报名队信息-051》后填写）：**

H113

评阅记录：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 评  阅  人 |  |  |  |  |
| 评  分 |  |  |  |  |
| 备  注 |  |  |  |  |

基于保理业务中授信机制评估和违约风险预测的数学模型分析

摘要

本文针保理业务中的客户授信评估问题，通过对样本数据多个属性进行分析，找数据规律，建立了客户违约预测模型，以及数据完整和缺失情况下的授信额度估算模型，最终得到一个良好的客户授信评估方案。

针对建立违约预测模型的问题，我们需要利用聚类分析的方法来解决。首先通过H-K算法判断出题目所给的样本集为非线性可分。其次使用SVM（支持向量机）算法对非线性样本集进行训练，得到分类器。我们对所给数据进行异常排除后的数据作为样本数据进行反复迭代，训练，最后通过我们VSM算法得到这些数据的分类，经过与样本数据的比较检验，然后通过样本数逐渐增加进行了5次训练，最终发现分类器的正确率趋于稳定，于是我们得到的违约模型判断不违约客户不违约的正确率为0.9323，判断违约客户为违约客户的正确率为0.81，有很高的准确率。

针对建立授信机制的模型，由于影响（变量）因素有28个所以我们首先采用主成分析法，通过SPSS分析，我们发现KMO值为0.49，不适合主成分分析，R=0.958说明数据适合多元回归分析并且拟合度非常高，因此我们用SPSS对样本数据进行多元回归，排除V11变量，得到多元回归函数表达式即计算授信得分的表达式。

并通过检验，模型的正确率为0.91.然后将所给要预测的80个数据带到模型一中得到违约风险的预测结果。

针对带有缺失值的授信机制评估和风险预测，我们建立了三种模型分别为E-M，回归，多重替换法，通过SPSS自带的分析功能填补缺失，异常的值，最后代入前两个违约预测以及授信评估模型中检验在不同缺失数据情况下模型的准确率，经过计算数据在缺失少于个值时三种模型对违约预测，授信评估的值分别都处于90%，90%以上，但当数据大量缺失后三个模型对违约预测，授信评估的值分别为70%，90%左右，说明大量数据缺失对正确违约预测的影响很小。综合来看多重替换模型结果最优。

对于问题四的，我们基于前三问得到的模型，

关键字 SPSS SVM（支持向量机） 多元回归分析 E-M（期望值最大化方法

） 多重替代法

**一、问题重述**

最近一些年以来，无论国内贸易还是国际贸易，赊销结算方式日渐盛行，以国际贸易为例，信用证的使用率已经降至16%，在发达国家已降至10%以下，赊销基本上取代了信用证成为主流结算方式。在赊销贸易下，产生了企业对应收账款的管理和融资需求，这便是保理业务发展的基础。

随着社会经济的蓬勃发展，保理行业需求越来越大。保理公司需要根据客户提供的数据进行授信额度的计算，包括违约的可能性等等。客户进件过程中资料不齐全导致无法完整提取财务信息，按现有额度授信公式计算需还原财务报表，工作量过于繁杂且误差较大。基于已完结的保理数据文件解决以下问题：

1.分析数据文件中已完结的数据，找出规律，建立违约预测模型，以预测客户违约的概率。

2.用完整的数据建立授信额度估算模型，并对数据文件中最后80个客户进行违约预测。

3.建立数据有残缺情况下的确定授信额度的数学模型，并对客户违约进行分析。

4.写一份不超过2页的非技术报告给公司管理层。

**二、问题分析**

2.1问题一分析

在题目所给的数据中，前284727个客户样本数据，客户提供的数据一共有28项，客户的类型分成违约和不违约两个类别。我们将客户的数据抽象成28维空间中的点，违约客户和不违约客户所代表的点属于不同的类别，问题就转化为我们如何对在这28维空间的点进行分类。因此问题转化成空间维度为28维的二分聚类分析问题，目标是找到能有效对客户是否违约进行预测的判别函数，即28维空间中的分类判别函数，即分类超平面。我们需要先判断样本集是否是线性可分的，因此我们可以先利用Ho-Kashvao算法（HK算法）判断样本集是否线性可分，若线性可分，则可得出判别函数，若不可分，则HK可以检测出来。如果HK算法检测出样本集是非线性可分的，我们则使用支持向量机算法（SVM）对非线性可分的样本集进行训练，最终得到分类器。得到分类器以后，我们将使用分类器对测试样本集进行测试，以得出训练出的分类器的正确率。对于一个提供完整数据的客户，我们可以将客户数据带入分类器，则可得出我们能有多大的把握判断客户是否违约。

2.2问题二分析

我们需要在数据完整的情况下建立授信额度估算模型。对于授信额度的估算，我们可以考虑使用多元线性回归分析或者主成分回归分析。考虑到客户数据的特征为28个，变量较多，因而我们首先考虑主成分回归分析是否可行，若主成分回归分析对此题目的情况不适用，我们将会考虑使用多元线性回归分析。以客户提供的28个数据来估算授信额度，通过样本集进行计算，我们可以给出多元线性回归函数或者主成分回归函数。对于一个提供完整数据的客户，我们可以使用该多元线性回归或者主成分回归函数以客户提供的28个数据为前提估算出授信额度。然后将客户数据送入问题一种训练出的分类器进行分类，预测客户是否会违约。

2.3问题三分析

在实际情况下，客户提供的数据可能不完整，若要建立数据有残缺情形下的确定授信额度的数学模型并对客户违约进行分析，我们需要预测出客户没有提供的数据。对于数据残缺的原因，我们考虑了数据缺失的三种情况：一是数据原本就是缺失的、二是客户提供的数据是错误的，因而认为数据是缺失的、三是客户提供的数据是不重要的。对于如何预测残缺值，我们可以考虑使用主成分回归分析、多元线性回归分析、多重替换方法、期望值最大化方法（EM算法）来使用已知数据对客户缺失的数据进行预测插补，利用上述方法对残缺值进行预测并进行插补，得出完整数据。对缺失值进行插补后，我们利用问题一和问题二中得出的分类器和回归函数并利用插补后的完整数据对客户的授信额度进行预测和分析客户是否违约。我们将对上述插补缺失值的方法都加以使用并比较，考虑到客户缺失的数据个数会有不同情况，因此，在不同缺失数据个数的情况下，我们利用测试样本集对不同缺失值插补方法进行测试并比较，并得出不同方法在不同情况下预测授信额度和判断客户是否违约的正确率，以得出哪一种方法的正确率更高。

**三、变量符号说明**

|  |  |  |
| --- | --- | --- |
| **变量符号** | **变量含义** | **单位** |
|  | 训练样本集 |  |
|  | 类别号 |  |
|  | 第i种类别 |  |
|  | 增广的权向量 |  |
|  | 增广的样本向量 |  |
|  | 线性判别函数 |  |
|  | 样本数据x的伪逆 |  |
|  |  |  |
|  | 分类函数 |  |
|  | 函数间隔 |  |
|  | 几何间隔 |  |
|  | 拉格朗日乘子 |  |
|  | 一个样本实际属于 |  |
|  | 把一个样本判断成属于 |  |
|  | 多重可决系数 |  |
|  | 常数项 |  |
|  | 回归系数 |  |
|  | 随机误差 |  |
| *Vi(i=1,2,…27)* | 第i种评估参数 |  |

**四、基本假设**

1.

2.

3.

**五、模型的建立与求解**

5.1保理业务中客户授信额度评估和违约风险的成因及识别

5.1.1保理业务中客户违约风险的成因

保理业务中，客户违约风险的产生主要是因为在此业务中存在这大量的委托代理关系，并由当事方之间的信息不对成导致的，保理业务主要涉及到两方即保理商和卖方将其现在或未来的基于与买方订立的货物销售/服务合同所产生的应收账款准让给保理商，由保理商向其提供资金融通/买方资信评估/销售账户管理/信用风险担保/账款催收等一系列服务的综合金融服务方式。它是商业贸易中以脱手/赊账方式结算货款时，卖方为了强化应收账款管理增加流动性而采用的一种委托代理商管理应收账款的做法。在这层委托代理关系的形成和建立过程中，当时双方存在这大量的信息不对称问题，并因相关主题的利己行为和逆向选择而产生了大量的道德风险，这是保理业务中客户产生违约的主要诱因。

5.1.2保理业务中客户授信额度评估和违约风险的识别

深入剖析保理业务中决定授信额度的因素对于了解、识别、评估和管理违约风险具有重要作用，基于以上对保理业务中客户违约风险的成因的认知，我们可以从客户的基本信息、信贷情况、工作业务三方面的信息对其进行授信额度评估以及判断是否违：

1. 基本信息：客户的户籍所在地、婚姻状况、年龄、学历、固有资产。
2. 信贷情况：客户的有效信用卡数、卡均使用率、月均还款占比、近5年内贷款逾期次数、征信五级分类次级以上级别贷款、贷款类查询记录（含本人）、第三方征信风险得分、贷款涉及金融机构类型。
3. 工作业务：新开金融类账户数量、业务所处阶段、申请人控股比例、申请人出资方式、合伙人数量、申请人公检法记录、经营合同风险性质、出租率、退租率、毛利率、利息保障倍数、月均银行账户资金留存、大额进出账交易频率、交易对手数量、同一交易对手频次。

5.1.3保理业务客户授信额度评估和违约风险识别的备选指标体系

基于保理业务的客户授信额度评估和违约风险的识别，本文从定量的方面，选择客户基本信息、信贷情况、工作业务三个方面的28个指标作为备选的风险指标，构建客户授信额度评估和违约风险预测指标体系。如表1所示：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 类型 | 指标编号 | 指标名称 | 类型 | 指标编号 | 指标名称 |
| 基本信息 | V1 | 户籍所在地 | 工作业务 | V15 | 业务所处阶段 |
| V2 | 婚姻状况 | V16 | 申请人控股比例 |
| V3 | 年龄 | V17 | 申请人出资方式 |
| V4 | 学历 | V18 | 合伙人数量 |
| V5 | 固有资产 | V19 | 申请人公检法记录 |
| 信贷情况 | V6 | 有效信用卡数 | V20 | 经营合同风险性质 |
| V7 | 卡均使用率 | V21 | 出租率 |
| V8 | 月均还款占比 | V22 | 退租率 |
| V9 | 近5年内贷款逾期次数 | V23 | 毛利率 |
| V10 | 征信五级分类次级以上级别贷款 | V24 | 利息保障倍数 |
| V11 | 贷款类查询记录（含本人） | V25 | 月均银行账户资金留存 |
| V12 | 第三方征信风险得分 | V26 | 大额进出账交易频率v |
| V13 | 贷款涉及金融机构类型 | V27 | 交易对手数量 |
| 工作业务 | V14 | 新开金融类账户数量 | V28 | 同一交易对手频次。 |

5.2数据预处理

表 1 客户授信额度评估和违约风险预测指标

实际上获取的数据往往因为各种原因而出现异常情况，所以要偏离期望值的异常数据。对题目中所给的样本案例的初始训练子集，采用拉依达准则法（3法），把偏离均值三倍标准差范围以上的异常数据删去，经预处理后的数据可以进行实证预测评价。

5.3基于H-K算法和SVM（支持向量机）的违约预测模型

为了建立客户的违约预测模型，我们需要对题目中给出的数据进行分析，把数据中违约和不违约的数据进行聚类分析，首先判断出数据是否为线性可分的数据，然后再分别利用SVM（支持向量机）算法求解出分隔函数，最后再对客户违约的概率进行求解。

5.3.1利用H-K算法判断数据是否线性可分

通过对题目的分析，我们需要将数据分成两类：违约样本和不违约样本。H-K算法是一种能给出分类判别函数的聚类分析算法，它即是在最小均方误差准则下求得权矢量。相对于其他的算法，优点在于：他适用于线性可分和非线性可分得情况，对于线性可分的情况，能够给出最优权矢量，对于非线性可分得情况，能够判别出来并退出迭代过程。

我们首先对样本数据进行处理[1]：对每组完整数据进行增广化和符号化。设为样本集（训练集），且训练样本之间相互独立，n为训练样本个数，为样本对应的类别号。用表示类别，所以两个类别为当时，；时，。定义为增广的样本向量，定义为增广的权向量。具体带入题中所给数据可得如下两种形式的数据：

对于没有违约的数据：

对于违约的数据：

所有数据经过处理后得到的数据见附件1。

将样本向量和权向量增广后，问题就转化为如何利用训练样本集寻求满足如下条件的权向量。

如果存在一个w使得样本集中的全部样本满足上式，则称样本集为线性可分的，称为线性判别函数。

H-K算法具体步骤为(源代码见附件2)：

①对规范化后的数据x求其伪逆：

②置初始值。设置判断标志flag=0

③计算：

④如果的各分量停止变为正值，并且不为0，则标志flag=1，退出循环，数据线性不可分。

如果=0，则输出并且退出循环，此即为迭代计算出的最终权向量。

如果>0,则

⑤循环次数k+1，回到第③步继续循环。

由以上步骤便可以得出数据的线性可分与否：我们通过利用MATLAB运行算法，结果标志flag为1，表明题目所给样本集是非线性可分的。

5.3.2利用SVM建立违约预测模型

支持向量机( Support Vector Machine，SVM)是一种二类分类模型，其基本模型定义为特征空间上的间隔最大的线性分类器，其基本思想为寻求一个最优

分类超平面，平面中原数据中的两类样本能够被最大化地分开，这两类样本尽可能的达到最大间隔距离。

（1）线性分类

我们考虑的是一个两类（违约与不违约）的分类问题，数据点用x来表示，这是一个n维向量，中的T代表转置，而类别用y来表示，可以取1或者-1，分别代表两个不同的类。一个线性分类器的学习目标就是要在 n 维的数据空间中找到一个分类超平面，其方程可以表示为：

这个分类超平面把两种不同类的数据点（y不同）分隔开来，在超平面一边的数据点所对应的y全是−1，而在另一边全是1。

接着，我们可以令分类函数为：

显然，如图1所示：如果，那么x是位于超平面上的点。若满足，其对应的y等于−1，而则对应y=1的数据点。

（2）函数间隔和几何间隔

一般而言，一个点距离超平面的远近可以表示为分类预测的确信或准确程度。在超平面确定的情况下， 能够相对的表示点到距离超平面的远近，对于用于测试的样本，计算测试样本的，由于测试样本的类别值已知，因而可以使用的符号与类标记的符号是否一致表示分类是否正确，所以，可以用量 的正负性来判定或表示分类的正确性和确信度。

我们定义函数间隔为：

接着，我们定义超平面 关于训练数据集的函数间隔为超平面关于中所有样本点的函数间隔最小值，其中，是特征，是结果标签，表示第i个样本，有：

上述定义的函数间隔虽然可以表示分类预测的正确性和确信度，但如果成比例的改变和，函数间隔的值也会成比例改变。因此我们利用超平面的单位向量引入几何间隔：

几何间隔就是函数间隔除以，而且函数间隔实际上就是，而几何间隔才是直观上的点到超平面的距离。

(3)最大间隔分类器

按照我们前面的分析，对一个数据点进行分类，当它的间隔越大的时候，分类的可信度越大。对于一个包含n个点的数据集，我们可以很自然地定义它的间隔为所有这n个点的间隔值中最小的那个。于是，为了使得分类的可信度高，我们希望所选择的超平面能够最大化这个间隔值。

函数间隔可以在超平面保持不变的情况下被取得任意大，而几何间隔则没有这个问题，因为它除上了这个分母，所以缩放和的时候的值是不会改变的，它只随着超平面的变动而变动，因此，这是更加合适的一个间隔。

这样一来，我们的最大间隔分类器的目标函数可以定义为：

根据数据集到超平面的函数间隔的定义，还需要满足一些条件：

其中(等价于)，处于方便推导和优化的目的，我们可以令，此时，上述的目标函数转化为：

通过求解这个问题，我们就可以找到一个间隔最大的分类函数，如下图所示，中间的红色线条是最优超平面，另外两条线到红线的距离都是等于的 (便是 上文所定义的几何距离，当令时，便为，而我们上面得到的目标函数便是在相应的约束条件下，要最大化这个值)：

通过最大化距离，我们使得该分类器对数据进行分类时具有了最大的可信度，从而设计决策最优分类超平面。

（4）从原始问题到对偶问题的求解

之前得到的目标函数为：

即：

（1）

通过给每一个约束条件加上一个拉格朗日乘值，即引入拉格朗日乘子，我们便可以通过拉格朗日函数将约束条件融和到目标函数里去：

然后我们令：

而当所有约束条件都满足时，则有 亦即我们最初要最小化的量。如果约束条件没有得到满足， 等于无穷大，自然不会是我们所要求的最小值。现在的目标函数变成了：

这里用表示这个问题的最优值，这个问题和我们最初的问题是等价的。对目标函数进行KKT条件的验证，发现满足。于是我们便可以把最小和最大的位置交换：

求解对偶问题可以得到:

因此分类函数为:

（5）线性不可分情况

在线性不可分的情况下，支持向量机通过某种事先选择的非线性映射 (核函数) 将输入变量映射到一个高维特征空间，在这个空间中构造最优分类超平面。但对于样本的多个特征无法有效分析数据特征，故使用高斯（RBF）核函数进行映射：

使用高斯核函数映射到高维空间后我们的分类函数为：

其中由以下的约束条件得出：

至此，我们成功求出样本的分类函数，即此样本的超平面。通过超平面即可对未知数据进行类别的判断。

5.3.3支持向量机模型的算法流程

我们进行5次训练和预测，将每一次选定的样本划分为2份（训练样本集和预测样本集），样本(训练样本量，预测样本量)个数分别为：(10000+366，80000+100），（20000+366，80000+100），（50000+366，80000+100），（100000+366，80000+100），（200000+366，80000+100），其中每个样本集均是从题目所给的样本集中随机选取，具体的算法步骤如下：

（1）针对每份训练样本，操作如下：

①用当前样本中的训练样本集训练SVM模型，求解式子(1)中的二次规划问题。

②依据训练样本集的训练结果，构建与预测样本集相对应的 SVM 分类函数:

③根据分类函数式子对预测样本集中每个样本 进行测试，令的目标类别标识为，测试类别标识为:

④计算预测样本集的测试准确率，判断违约的准确率和判断不违约的准确率分别记为和:

（3）构建 SVM 模型。用核函数和得出的模型参数来训练 SVM，求得私募股权投资风险预测中的支持向量，构造 SVM 私募股权投资风险的预测函数。

（4） 风险预测。运用构造的 SVM 私募股权投资风险预测函数的分类率，估算出私募股权投资项目的风险状况，根据风险状况采取相应的处置措施。

5.3.4单个类别概率的计算

为了预测一个样本属于某一个类别的概率，我们需要通过概率公式来计算。由上文可知，两个类别为，设他们表示一个样本实际属于，分别表示把一个样本判断成属于。以为例，一个样本实际属于的概率就可以分为两种情况：

其中，表示样本在判断为违约情况下违约的概率，表示样本在判断为不违约的情况下违约的概率。均可用实际样本求出。

得到的5次结果如下：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 第一次 | 不违约 | 违约 | 第二次 | 不违约 | 违约 |
| 训练样本量 | 10000 | 366 | 训练样本量 | 20000 | 366 |
| 预测样本量 | 80000 | 100 | 预测样本量 | 80000 | 100 |
| 正确率 | 0.8756 | 0.8756 | 正确率 | 0.9283 | 0.64 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 第三次 | 不违约 | 违约 | 第四次 | 不违约 | 违约 |
| 训练样本量 | 50000 | 366 | 训练样本量 | 100000 | 366 |
| 预测样本量 | 80000 | 100 | 预测样本量 | 80000 | 100 |
| 正确率 | 0.9386 | 0.63 | 正确率 | 0.9302 | 0.83 |

表 2 五次训练以及预测样本得到的结果

|  |  |  |
| --- | --- | --- |
| 第五次 | 不违约 | 违约 |
| 训练样本量 | 200000 | 366 |
| 预测样本量 | 80000 | 100 |
| 正确率 | 0.9323 | 0.81 |

其中，每份样本中第一个正确率为，表示对不违约样本预测的正确率，第二个正确率为，表示对违约样本预测的正确率。从五次实验来看，随着样本数量的增加，两个正确率趋于稳定值。

5.4基于多元线性回归的授信额度估算模型

为了建立客户授信额度估算模型，我们需要对题中所给的样本进行分析。其中，最主要的分析方法有两种：主成分分析法和多元线性回归法。通过KMO检验和Bartlett球形检验，选出最佳方法进行计算分析最终得出模型。

5.4.1 KMO检验和Bartlett球形检验

KMO（Kaiser-Meyer-Olkin)检验统计量是用于比较变量间简单相关系数和偏相关系数的指标。Bartlett's球状检验用于检验相关阵中各变量间的相关性，是否为单位阵，即检验各个变量是否各自独立。因子分析前，首先进行KMO检验和巴特利球体检验。

KMO检验用于检查变量间的相关性和偏相关性，取值在0~1之前。KMO统计量越接近于1，变量间的相关性越强，偏相关性越弱，因子分析的效果越好。实际分析中，KMO统计量在0.7以上时效果比较好；当KMO统计量在0.5以下，此时不适合应用因子分析法，应考虑重新设计变量结构或者采用其他统计分析方法。

如果变量间彼此独立，则无法从中提取公因子，也就无法应用因子分析法。Bartlett球形检验判断如果相关阵是单位阵，则各变量独立因子分析法无效。由SPSS检验结果显示Sig.<0.05（即p值<0.05）时，说明各变量间具有相关性，因子分析有效。

我们通过SPSS软件，对所给样本进行KMO和Bartlett球形检验，得到如下结果(表2）：

|  |  |  |
| --- | --- | --- |
| KMO 取样适切性量数 | | 0.490 |
| 巴特利特球形度检验 | 近似卡方 | 96.681 |
| 自由度 | 378 |
| 显著性 | 1.000 |
| 表 3 KMO 和巴特利特检验 | | |

接着分别对样本数据进行主成分分析中碎石图的绘制和多元线性回归分析中值（多重可决系数，它是在因变量的总变化中，由回归方程解释的变动(回归平方和)所占的比重，越大，回归方各对样本数据点拟合的程度越强，所有自变量与因变量的关系越密切。）的计算得到如下两图（图4，表3）

从KMO值<0.5，以及碎石图中特征值大于1的变量个数占绝大多数所以因子分析法，主成分分析法不适用于本题样本。由R=0.958说明多远线性回归的拟合优度非常高，因此我们采用多元回归法建立授信模型。

5.4.2多元线性回归模型的建立

多元回归分析预测法，是指通过对两个或两个以上的自变量与一个因变量的相关分析，建立预测模型进行预测的方法。当自变量与因变量之间存在线性关系时，称为多元线性回归分析。

一元线性回归是一个主要影响因素作为自变量来解释因变量的变化，在现实问题研究中，因变量的变化往往受几个重要因素的影响，此时就需要用两个或两个以上的影响因素作为自变量来解释因变量的变化，这就是多元回归亦称多重回归。当多个自变量与因变量之间是线性关系时，所进行的回归分析就是多元线性回归。

设y为自变量，为自变量，并且自变量与因变量之间为线性关系时，则多元线性回归模型为：

其中，为常数项，为回归系数，为固定时，每增加一个单位对y的效应，即对y的偏回归系数；同理为,固定时，每增加一个单位对y的效应，即，对y的偏回归系数，之后同理。如果两个自变量同一个因变量y呈线相关时，可用二元线性回归模型描述为：

如果有n组样本，记为。那么这个多元线性回归，将会组成一个矩阵，令：

那么，多元线性回归方程矩阵形式为：。

其中：代表随机误差，其中随机误差分为：可解释的误差和不可解释的误差，随机误差必须满足以下四个条件，多元线性方程才有意义。

1.服从正态分布：随机误差必须是服从正态分部的随机变量。

2.无偏性假设：期望值为0。

3.同共方差性假设：所有的随机误差变量方差都相等

4.独立性假设：所有的随机误差变量都相互独立，可以用协方差解释。

至此，我们完成了线性回归模型的建立。之后利用SPSS软件对样本数据进行多元线性回归分析得到结果。

5.4.3多元线性回归结果分析

多元性回归模型与一元线性回归模型一样，在得到参数的最小二乘法的估计值之后，也需要进行必要的检验与评价，以决定模型是否可以应用。

（1）结构分析1：输入/除去的变量



表 5 输入/除去的变量

由于开始选择的是“逐步”法，逐步法是“向前”和“向后”的结合体，从结果可以看出，最先进入“线性回归模型”的是V2，建立了模型1。紧随其后的是V7建立了模型2，接着一共建立了27个模型。模型中此方法有个概率值，当小于等于0.05时，进入“线性回归模型”（最先进入模型的，相关性最强，关系最为密切）当大于等0.1时，从“线性模型中”剔除。

（2）结果分析2：模型汇总和Anova表



表 6 27个模型汇总（部分，全表见附件3）

从图7中可以看出，有27个模型，（模型1到模型27）从R27拟合优度来看，模型27的拟合优度明显比其他模型要好一些（为0.917）

表 7 Anova表（部分，全表见附件4）



从“Anova”表中，可以看出“模型27”中的“回归平方和”为16346877340.000，“残差平方和”为1470605687.000，由于总平方和=回归平方和+残差平方和，由于残差平方和(即指随即误差，不可解释的误差）几乎是“回归平方和”的十分之一，所以此线性回归模型几乎解释了总平方和。

根据后面的“F统计量”的概率值为0.00，由于0.00<0.01，随着“自变量”的引入，其显著性概率值均远小于0.01，所以可以显著地拒绝总体回归系数为0的原假设，通过ANOVA方差分析表可以看出“:授信额度”与除自变量V1(贷款类查询记录（含本人）)外其他27个变量之间存在着线性关系。

接下来进一步分析线性关系的强弱。

（3）结果分析3：已排除变量

表 8 已排除的变量（部分，全表见附件5）



从“已排除的变量”表中，可以看出：各模型各变量处V11外的T检的概率值都大于“0.05”所以V11不能够引入“线性回归模型”，必须剔除。

（4）结果分析4：系数表

从“系数” 表中可以看出，多元线性回归方程应该为：



表9 系数表（部分，全表见附件6）

常数项的sig为（0.00<0.1) 所以常数项具备显著性.

（5）结果分析5：残差

从上图可以得知：大部分自变量的残差都符合正态分布，更加支持“模型27”

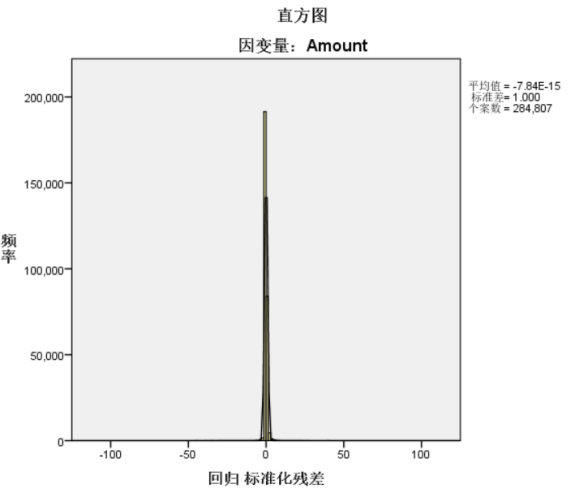


图 5回归标准化残差与频率直方图

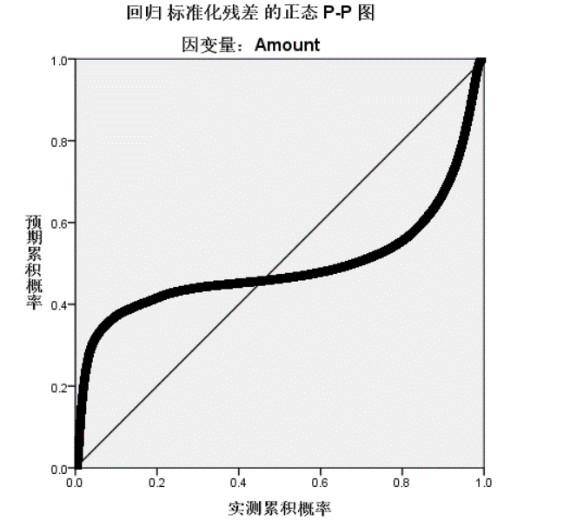


图 6回归标准化残差的正态P-P图

5.4.4 多元线性回归方程模型的检验

模型检验算法步骤如下：

1. 将所有样本（sum）的数据值（V1-V28）带入算法得到所有数据的授信预测结果。
2. 设置阈值为20，即预测值和真实值只差的绝对值小于等于20，此时计数器count加1。
3. 准确率即为count/sum。

我们通过MATLAB计算得到准确率的结果为0.91说明我们建立的多元回归模型正确率很高。

5.4.5对条目未知客户违约的预测

利用样本数据训练，将creditdata.xlsx中最后80个class条目未知的客户（ID从U01到U80的数据）的信息（V1-V28）代入到第一问SVM模型中，我们得到所有未知客户违约风险预测结果如下表：。

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 未知客户违约风险预测结果 | | | | | | | |
| U01 | 1 | U21 | 1 | U41 | 1 | U61 | 0 |
| U02 | 0 | U22 | 0 | U42 | 0 | U62 | 1 |
| U03 | 1 | U23 | 1 | U43 | 0 | U63 | 1 |
| U04 | 0 | U24 | 1 | U44 | 0 | U64 | 0 |
| U05 | 0 | U25 | 0 | U45 | 0 | U65 | 0 |
| U06 | 1 | U26 | 0 | U46 | 0 | U66 | 0 |
| U07 | 1 | U27 | 0 | U47 | 0 | U67 | 0 |
| U08 | 0 | U28 | 0 | U48 | 1 | U68 | 0 |
| U09 | 1 | U29 | 0 | U49 | 0 | U69 | 0 |
| U10 | 1 | U30 | 0 | U50 | 0 | U70 | 1 |
| U11 | 0 | U31 | 0 | U51 | 0 | U71 | 0 |
| U12 | 1 | U32 | 0 | U52 | 0 | U72 | 1 |
| U13 | 0 | U33 | 1 | U53 | 0 | U73 | 0 |
| U14 | 0 | U34 | 0 | U54 | 1 | U74 | 0 |
| U15 | 1 | U35 | 0 | U55 | 0 | U75 | 1 |
| U16 | 1 | U36 | 1 | U56 | 0 | U76 | 0 |
| U17 | 1 | U37 | 0 | U57 | 0 | U77 | 0 |
| U18 | 0 | U38 | 0 | U58 | 0 | U78 | 0 |
| U19 | 0 | U39 | 0 | U59 | 0 | U79 | 1 |
| U20 | 1 | U40 | 0 | U60 | 0 | U80 | 0 |

5.5插补残缺数据的授信额度数学模型

在保理业务的实际过程中，由于各种原因，客户提供的数据往往不完整，导致我们无法利用之前求出的模型对客户的授信额度进行评估，所以需要通过新的方式建立合作的模型填补残缺数据，然后才能完成评估。

通过思考，我们得出数据残缺的原因有三种：

* 原本数据就是残缺的
* 数据是不重要的
* 数据是错误的，因而认为数据是残缺的

在这个例子中，V1-V28的数据都认为是对结果有影响的，因而都是重要的数据，所以残缺数据的原因共两个：

* 数据遗漏导致缺失
* 数据错误而认为其残缺

对于数据是否错误的判断，我们先对训练样本集进行考察，找出训练样本中的数值过大或者过小的情况，这一些拥有过大或者过小的数值的样本取到训练样本集的0.3%，随后根据这些小概率的样本设置一个范围，数据在范围内我们认为数据是合理的，数据在范围外我们认为数据是错误的。

5.5.1数据缺失的处理方法

（1）EM（Expectation-maximization，期望最大）算法填补缺失值

期望最大算法是一种从不完全数据或有数据丢失的数据集（存在隐含变量）中求解概率模型参数的最大似然估计方法。首先对缺失数据进行预测，反推缺失数据与已知数据的关系，再由这个关系结合已知数据推出缺失数据，再由缺失数据推出这个关系，如此反复，直到缺失数据收敛。

（2）多元线性回归预测残缺值

多元线性回归预测残缺值是把缺失属性作为因变量，其他相关属性作为自变量，利用他们之间的关系建立回归模型的来预测缺失值，以此完成缺失值插补的方法。

（3）多重替化法

用一系列可能的值来替换每一个缺失值，以反映被替换的缺失值的不确定性。然后，用标准的统计分析过程把来自于各个数据集的统计结果进行综合，得到总体参数的估计值。

5.5.2缺失数据的处理模型

（1）E-M算法填补缺失值

首先设缺失数据为Q，Z表示缺失数据与已知数据的关系，且Z为隐藏变量。初始化分布参数，重复以下步骤直到收敛：

* E步骤：根据参数初始值或上一次迭代的模型参数来计算出隐性变量的后验概率，其实就是隐性变量的期望。作为隐藏变量的现估计值：
* M步骤：将似然函数最大化以获得新的参数值：

其中，为的某种分布。

通过这两个步骤不断的迭代，就可以得到使似然函数最大化的参数了。

（2）多元线性回归预测残缺值

利用多元线性回归模型进行缺失值插补的步骤如下：

* 第一步，选择变量建立多元线性回归模型：

设y为因变量，表示缺失值所在的属性，选出为自变量，表示与缺失值所在的属性相关的属性，则多元线性回归模型为：

其中，为参数，服从零均值、相互独立且同方差服从正态分布等经典假定。

* 第二步，计算参数的估计值：

利用实际问题中的训练数据，采用普通最小二乘法可以求出对参数的估计值。使用最小平方法的好处是所有样本点预测误差平方和会最小。

求出了参数估计值即可得到多元回归方程：

* 第三步，带入各个自变量对应的属性值，就能得到缺失值。

（3）多重替代法

多重替代的思想来源于贝叶斯估计，认为待插补的值是随机的，它的值来自于已观测到的值。具体实践上通常是估计出待插补的值，然后再加上不同的噪声， 形成多组可选插补值。根据某种选择依据，选取最合适的插补值。多重插补方法分为三个步骤：

* 第一步，为每个空值产生一套可能的插补值， 这些值反映了无响应模型的不确定性；每个值都可以被用来插补数据集中的缺失值， 产生若干个完整数据集合。
* 第二步，每个插补数据集合都用针对完整数据集的统计方法进行统计分析。
* 第三步，对来自各个插补数据集的结果， 根据评分函数进行选择，产生最终的插补值。

5.5.3模型及其检验指标

我们采用五重交叉验证来检验分类效果。所谓五重交叉验证就是将所有的已知数据均分为5份，每一封都独立的建立一个模型，并用其余的数据进行验证。交叉验证的一个优点是即使数据量很小也能达到很好的效果。

模型检验指标为模型预测准确率和类预测准确率。分别定义如下：

其中，代表一个类“整个数据集中预测正确的记录个数”和“某个类中预测正确的记录个数”。

5.5.4结果及分析

通过SPSS软件自带的缺失值处理功能，选择相对应的三种模型进行分析和计算，得到的结果如下表：

表 10 三个模型分类情况的比较

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 训练样本  数量 | 缺失属性  数量 | 缺失情况 | 模型一 | | 模型二 | | 模型三 | |
| R1 | R2 | R1 | R2 | R1 | R2 |
| 100000 | 3 | 很少 | 0.945 | 0.92 | 0.915 | 0.935 | 0.94 | 0.95 |
| 6 | 一般 | 0.825 | 0.92 | 0.815 | 0.915 | 0.825 | 0.93 |
| 9 | 严重 | 0.74 | 0.92 | 0.785 | 0.905 | 0.75 | 0.89 |

其中，模型一：E-M模型，模型二：多元线性回归模型，模型三：多重替代模型，模型预测准确率，：类预测准确率。

结果分析：从表中的实验数据来看，当缺失数据很少，或一般时，三种模型准确率相差不大，预测授信额度时模型一的准确率最高，预测是否违约时，模型三的准确率最高。当缺失数据量很大时，就预测是否违约来说，三种模型效果依然差不多。

5.6关于客户授信评估问题的报告

尊敬的领导：

您好！我们是电子科技大学的学生。我们对保理业务中有关客户授信额度的计算和违约风险评估建立了有效的数学模型，并通过模型检验得到高度的有效性。我们希望贵公司能采用我们的模型以完善对客户的授信机制。接下来我将全面介绍我们的研究成果。

首先我们通过查阅资料，认识到影响客户授信评分以及违约风险的因素可分为这三类：基本信息（如年龄，婚姻等），信贷情况（如有效信用卡数，卡均使用率等），工作业务（如新开金融类账户数量，业务所处阶段等）。结合贵公司提供的数据我们确定了28个影响因素。

首先是违约风险的预测模型，我们建立了VSM(支持向量机)模型，支持向量机方法是建立在统计学习理论的VC维理论和结构风险最小原理基础上的，根据有限的样本信息在模型的复杂性（即对特定训练样本的学习精度）和学习能力（即无错误地识别任意样本的能力）之间寻求最佳折中，以求获得最好的推广能力 。 由于客户只存在违约或不违约两种情形，因此结果只有两种。贵公司提供了二十八万数据，我们对这些数据进行异常排除后的数据作为样本数据进行反复迭代，训练，最后通过我们VSM算法得到这些数据的分类，经过与样本数据的比较检验，得到违约情况下判违的正确率为0.9323，不违约情况判不违的概率是0.81，因此说明我们的模型有很高的准确率。

针对授信得分的估计，由于通过我们对数据进行SPSS（一种数据分析软件）分析得到所有变量间相关性极差，因此我们建立了多元回归模型。将信用得分视为因变量，其他所有影响因素为自变量，建立多个变量之间线性或非线性数学模型数量关系式并利用样本数据进行分析。最后我们得到一个用于预测授信得分的回归函数，然后我们将所有样本带到模型中进行检验，正确率为91.8%，说明我们的模型有很高的可靠性。

最后考虑到客户所给的数据有缺失情况（包括异常数据），我们建立了三种模型分别为E-M，回归，多重替换法，通过SPSS自带的分析功能填补缺失，异常的值，最后代入前两个违约预测以及授信评估模型中检验在不同缺失数据情况下模型的准确率，经过计算数据在缺失少于个值时三种模型对违约预测，授信评估的值分别都处于90%，90%以上，但当数据大量缺失后三个模型对违约预测，授信评估的值分别为70%，90%左右，说明大量数据缺失对正确违约预测的影响很小。综合来看多重替换模型结果最优。

以上别是我们对于保理业务中有关客户授信额度的计算和违约风险评估建立的数学模型，感谢您的阅读，望这些模型能被借鉴。

此致

敬礼!

**七、模型的评价与改进方向**

7.1模型的评价

基于问题一，二，三，建立出的相关模型，我们估计出了授信额度以及判断是否违约，通过检验，我们模型的正确率基本在0.95左右，有很高的可信度，因此可用于实际运用中。

7.2模型的改进方向

1、实

。

**八、参考文献**

[1]田永军.新颖的模式分类算法MatMHKS及在乳腺癌计算机辅助诊断中的应用.南京航空航天大学,2005.

**九、附件清单**

附件1一

附件2 医

附件3 其

**十、附件**

**附件1**増广符号化后的数据

**附件2** H-K算法MATLAB源代码

**附件3**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **模型摘要** | | | | | | | | |
| **模型** | **R** | | **R 方** | **调整后 R 方** | | | **标准估算的误差** | |
| **1** | **.531** | | **.282** | **.282** | | | **211.880973999999980** | |
| **2** | **.664** | | **.440** | **.440** | | | **187.131337399999980** | |
| **3** | **.768** | | **.590** | **.590** | | | **160.249019200000000** | |
| **4** | **.839** | | **.705** | **.705** | | | **135.915968599999990** | |
| **5** | **.870** | | **.757** | **.757** | | | **123.407219700000000** | |
| **6** | **.896** | | **.803** | **.803** | | | **110.955125500000000** | |
| **7** | **.921** | | **.848** | **.848** | | | **97.616251160000000** | |
| **8** | **.928** | | **.860** | **.860** | | | **93.462831410000000** | |
| **9** | **.934** | | **.872** | **.872** | | | **89.623649790000000** | |
| **10** | **.939** | | **.882** | **.882** | | | **85.835214020000000** | |
| **11** | **.945** | | **.893** | **.893** | | | **81.994806440000000** | |
| **12** | **.950** | | **.902** | **.902** | | | **78.187686200000000** | |
| **13** | **.952** | | **.906** | **.906** | | | **76.489386680000000** | |
| **14** | **.954** | | **.910** | **.910** | | | **75.189033760000000** | |
| **15** | **.955** | | **.912** | **.912** | | | **74.231009740000000** | |
| **16** | **.956** | | **.914** | **.914** | | | **73.401519730000000** | |
| **17** | **.957** | | **.915** | **.915** | | | **72.857990300000000** | |
| **18** | **.957** | | **.916** | **.916** | | | **72.367369090000000** | |
| **19** | **.958** | | **.917** | **.917** | | | **72.007426080000000** | |
| **20** | **.958** | | **.917** | **.917** | | | **71.961822490000000** | |
| **21** | **.958** | | **.917** | **.917** | | | **71.922359650000000** | |
| **22** | **.958** | | **.917** | **.917** | | | **71.899246400000000** | |
| **23** | **.958** | | **.917** | **.917** | | | **71.887180350000010** | |
| **24** | **.958** | | **.917** | **.917** | | | **71.875780990000000** | |
| **25** | **.958** | | **.917** | **.917** | | | **71.869254600000000** | |
| **26** | **.958** | | **.917** | **.917** | | | **71.864901020000000** | |
| **27** | **.958** | | **.917** | **.917** | | | **71.861146250000000** | |
| **附件4：ANOVA** | | | | | | | | | | |
| **模型** | | **平方和** | | | **自由度** | **均方** | | **F** | | **显著性** |
| **1** | **回归** | **5031576329.000** | | | **1** | **5031576329.000** | | **112077.941** | | **.000** |
| **残差** | **12785906700.000** | | | **284805** | **44893.547** | |  | |  |
| **总计** | **17817483020.000** | | | **284806** |  | |  | |  |
| **2** | **回归** | **7844177415.000** | | | **2** | **3922088708.000** | | **112001.637** | | **.000** |
| **残差** | **9973305610.000** | | | **284804** | **35018.137** | |  | |  |
| **总计** | **17817483020.000** | | | **284806** |  | |  | |  |
| **3** | **回归** | **10503813720.000** | | | **3** | **3501271238.000** | | **136343.675** | | **.000** |
| **残差** | **7313669310.000** | | | **284803** | **25679.748** | |  | |  |
| **总计** | **17817483020.000** | | | **284806** |  | |  | |  |
| **4** | **回归** | **12556292810.000** | | | **4** | **3139073203.000** | | **169926.251** | | **.000** |
| **残差** | **5261190214.000** | | | **284802** | **18473.151** | |  | |  |
| **总计** | **17817483020.000** | | | **284806** |  | |  | |  |
| **5** | **回归** | **13480151230.000** | | | **5** | **2696030246.000** | | **177028.677** | | **.000** |
| **残差** | **4337331793.000** | | | **284801** | **15229.342** | |  | |  |
| **总计** | **17817483020.000** | | | **284806** |  | |  | |  |
| **6** | **回归** | **14311298870.000** | | | **6** | **2385216478.000** | | **193746.142** | | **.000** |
| **残差** | **3506184157.000** | | | **284800** | **12311.040** | |  | |  |
| **总计** | **17817483020.000** | | | **284806** |  | |  | |  |
| **7** | **回归** | **15103652580.000** | | | **7** | **2157664654.000** | | **226432.988** | | **.000** |
| **残差** | **2713830445.000** | | | **284799** | **9528.932** | |  | |  |
| **总计** | **17817483020.000** | | | **284806** |  | |  | |  |
| **8** | **回归** | **15329686810.000** | | | **8** | **1916210851.000** | | **219364.036** | | **.000** |
| **残差** | **2487796213.000** | | | **284798** | **8735.301** | |  | |  |
| **总计** | **17817483020.000** | | | **284806** |  | |  | |  |
| **9** | **回归** | **15529880000.000** | | | **9** | **1725542222.000** | | **214822.783** | | **.000** |
| **残差** | **2287603025.000** | | | **284797** | **8032.399** | |  | |  |
| **总计** | **17817483020.000** | | | **284806** |  | |  | |  |
| **10** | **回归** | **15719196100.000** | | | **10** | **1571919610.000** | | **213353.290** | | **.000** |
| **残差** | **2098286923.000** | | | **284796** | **7367.684** | |  | |  |
| **总计** | **17817483020.000** | | | **284806** |  | |  | |  |
| **11** | **回归** | **15902764010.000** | | | **11** | **1445705819.000** | | **215034.052** | | **.000** |
| **残差** | **1914719015.000** | | | **284795** | **6723.148** | |  | |  |
| **总计** | **17817483020.000** | | | **284806** |  | |  | |  |
| **12** | **回归** | **16076447800.000** | | | **12** | **1339703983.000** | | **219145.283** | | **.000** |
| **残差** | **1741035225.000** | | | **284794** | **6113.314** | |  | |  |
| **总计** | **17817483020.000** | | | **284806** |  | |  | |  |
| **13** | **回归** | **16151265620.000** | | | **13** | **1242405047.000** | | **212354.198** | | **.000** |
| **残差** | **1666217409.000** | | | **284793** | **5850.626** | |  | |  |
| **总计** | **17817483020.000** | | | **284806** |  | |  | |  |
| **14** | **回归** | **16207442550.000** | | | **14** | **1157674468.000** | | **204775.242** | | **.000** |
| **残差** | **1610040472.000** | | | **284792** | **5653.391** | |  | |  |
| **总计** | **17817483020.000** | | | **284806** |  | |  | |  |
| **15** | **回归** | **16248215470.000** | | | **15** | **1083214364.000** | | **196581.966** | | **.000** |
| **残差** | **1569267559.000** | | | **284791** | **5510.243** | |  | |  |
| **总计** | **17817483020.000** | | | **284806** |  | |  | |  |
| **16** | **回归** | **16283096280.000** | | | **16** | **1017693517.000** | | **188889.103** | | **.000** |
| **残差** | **1534386748.000** | | | **284790** | **5387.783** | |  | |  |
| **总计** | **17817483020.000** | | | **284806** |  | |  | |  |
| **17** | **回归** | **16305741350.000** | | | **17** | **959161255.800** | | **180691.304** | | **.000** |
| **残差** | **1511741676.000** | | | **284789** | **5308.287** | |  | |  |
| **总计** | **17817483020.000** | | | **284806** |  | |  | |  |
| **18** | **回归** | **16326037990.000** | | | **18** | **907002110.300** | | **173189.967** | | **.000** |
| **残差** | **1491445040.000** | | | **284788** | **5237.036** | |  | |  |
| **总计** | **17817483020.000** | | | **284806** |  | |  | |  |
| **19** | **回归** | **16340842660.000** | | | **19** | **860044350.700** | | **165869.400** | | **.000** |
| **残差** | **1476640362.000** | | | **284787** | **5185.069** | |  | |  |
| **总计** | **17817483020.000** | | | **284806** |  | |  | |  |
| **20** | **回归** | **16342717610.000** | | | **20** | **817135880.700** | | **157793.814** | | **.000** |
| **残差** | **1474765411.000** | | | **284786** | **5178.504** | |  | |  |
| **总计** | **17817483020.000** | | | **284806** |  | |  | |  |
| **21** | **回归** | **16344339820.000** | | | **21** | **778301896.400** | | **150459.715** | | **.000** |
| **残差** | **1473143200.000** | | | **284785** | **5172.826** | |  | |  |
| **总计** | **17817483020.000** | | | **284806** |  | |  | |  |
| **22** | **回归** | **16345291670.000** | | | **22** | **742967803.300** | | **143721.360** | | **.000** |
| **残差** | **1472191353.000** | | | **284784** | **5169.502** | |  | |  |
| **总计** | **17817483020.000** | | | **284806** |  | |  | |  |
| **23** | **回归** | **16345790920.000** | | | **23** | **710686561.800** | | **137522.958** | | **.000** |
| **残差** | **1471692104.000** | | | **284783** | **5167.767** | |  | |  |
| **总计** | **17817483020.000** | | | **284806** |  | |  | |  |
| **24** | **回归** | **16346262790.000** | | | **24** | **681094283.000** | | **131838.448** | | **.000** |
| **残差** | **1471220234.000** | | | **284782** | **5166.128** | |  | |  |
| **总计** | **17817483020.000** | | | **284806** |  | |  | |  |
| **25** | **回归** | **16346535120.000** | | | **25** | **653861404.800** | | **126590.007** | | **.000** |
| **残差** | **1470947904.000** | | | **284781** | **5165.190** | |  | |  |
| **总计** | **17817483020.000** | | | **284806** |  | |  | |  |
| **26** | **回归** | **16346718490.000** | | | **26** | **628719941.900** | | **121737.274** | | **.000** |
| **残差** | **1470764535.000** | | | **284780** | **5164.564** | |  | |  |
| **总计** | **17817483020.000** | | | **284806** |  | |  | |  |
| **27** | **回归** | **16346877340.000** | | | **27** | **605439901.400** | | **117241.876** | | **.000** |
| **残差** | **1470605687.000** | | | **284779** | **5164.024** | |  | |  |
| **总计** | **17817483020.000** | | | **284806** |  | |  | |  |

其它部分程序清单和简要说明

表 25：其它部分程序清单

|  |  |  |  |
| --- | --- | --- | --- |
| 程序名称 | 功能 | 输入参数 | 输出参数 |
| getOutPatient.m | 生成病人流 | 无 | 门诊病人信息 |
| displayMessage.m | 在matlab中展示原始数据 | 数据矩阵 | 原始数据信息 |