中国石油大学（华东）
CHINA UNIVERSITY OF PETROLEUM

《数学建模》期末作业

题　目：婴幼儿奶粉产品研究

姓名	李选	张子锜	曹佳慧	张世琛
学号	1808010202	1808010201	1808010203	1804030401
班级	计科 1802	计科 1802	计科 1802	计科 1802

2019 年 6 月 10 日

目录

摘要

食品安全一直是社会关注的焦点，而对于婴幼儿口粮—奶粉的选择问题，人们更是慎之又慎。面对市面上种类繁多的奶粉产品，为了给家长和商家提供合理可行的参考建议，本文将以评价量为因变量，研究各个变量与评价量的关系，采用统计分析、插值与拟合、文本挖掘和灰度分析等方法研究以下四个奶粉方面的问题。

问题一：通过收集团购价和评价量两方面的数据，并进行单变量分布分析，求出中位数、平均数和偏度系数等相关系数。因为我们无法精确的得到相关商家的销售量信息，而选取评价量可以从侧面反映顾客对于一个产品的关注度，再加上价格分析，更具备说服力。

问题二选取非数据变量——品牌、奶源场地和适用年龄和数据变量——团购价和商品毛重作为自变量，以评价量为因变量，分别绘制系列直方图和饼状图并作深层次分析，给出初步建议。

问题三：既然评价量具有很高的参考价值，为研究影响其变化的具体因素，采用数据挖掘中的有关文字挖掘技巧处理文本，将文字语言转化成为合理的数据，可得到对评价量影响程度排名前三的因素。

问题四：结合以上三个问题所得的结果，进行分析，给出进一步的合理建议。

关键词

数据挖掘 文本挖掘 统计分析 灰度分析 插值与拟合

一、问题重述

随着互联网的发展和经济全球化，市场上出现越来越多的品牌可供选择。而对于广大家长父母，在婴幼儿奶粉的选择上，也不再仅仅局限于国产品牌，电商和海淘的兴起提供了更多品牌。面对五花八门的奶粉品牌，如何选择是一个令人头痛的问题。

我们在某电商平台收集了不同品牌奶粉的销售信息，由于商家对销售量保密，无法得到精确的销售量信息，但由于评价量可以从一个侧面反映顾客对产品的关注度，所以对所给数据进行以下方面的分析，将有利于提出有益于商家和消费者的合理建议。

1. 选取重要变量，进行单变量分析。
2. 以评价量为因变量，分析其他变量和评价量之间的关系。
3. 以评价量为因变量，研究影响评价量的重要因素。
4. 根据以上分析，分别给商家和消费者提出建议。

二、问题分析

针对问题一，我们认为团购价和评价量属于重要变量，在收集相关数据后，我们通过 MATLAB 计算出平均值、方差、中位数等统计量，然后根据计算结果分别分析团购价和评价量。

针对问题二，我们将从非数据变量和数据变量两方面进行分析。对于非数据变量——品牌、奶源场地和适用年龄，即没有数据时,我们通过 MATLAB 画直方图进行直观分析，探究影响评价量深层次的原因。对于数据变量——商品毛重和团购价，采用插值和拟合的数学方法进行分析，得到相关意见。

针对问题三，研究影响评价量的因素,可以置换为求取灰关联度。因此，我们采用灰色系统中灰色关联度分析法研究各个变量对评价量的影响程度，得到它们的影响大小的相关排序,得出影响程度前三的分别为商品毛重、商品品牌和商品团购价。这和大众购买商品考虑的因素具有相同的原因，即考虑量入为出和品牌效应。总得说来，灰色关联度分析能够较好解决此类问题。

针对问题四,我们将给商家和消费者提供合理的参考意见。基于以上研究成果，我们可以从市场需求量和产品结构两个方面为商家出谋划策，占领消费市场，把握产品结构将会促进商家提高盈利率。对于消费者来说，既要考虑其个人自身情况，又要权衡奶粉的性价比，从品牌效应、场地和价格等方面提出合理化建议。

三、模型假设和符号说明

1. 假设奶粉数据来源真实精确，具有良好的代表性。
2. 忽略或假设商家对评价量的干预控制情况。
3. 假设控制变量法可用，即分析一种变量对评价量的影响时，其他变量保持不变。
4. 假设评价量可以作为评价奶粉品质的标准。

S(x)	三次样条插值函数
$a_{i,j}$	影响值
w_{ik}	权值函数
$tf_k(d_i)$	频率
t_k	特征
N_k	文本数
γ_k	贡献值
X_n	贡献值
ρ	分辨系数
r_i	关联度

四、模型的建立

（一）、问题一

4.1.1 问题一的分析

选取团购价和评价量这两个指标，通过 MATLAB 画出图形，计算每个指标的平均值、众数、中位数和方差，根据计算结果分析指标。选取具有代表意义的指标并进行对比分析，得出结论。

4.1.2 问题一的求解

从大众选择奶粉的心理来看，我们认为较为重要的是价格和评价量。而由 MATLAB 程序可以得知团购价均值和评价量均值分别为 367、15800，它们的中位数分别为 254、330.5，标准差分别为 377、72870，极差分别为 2590、683010。根据以上数据，汇出如下统计直方图:

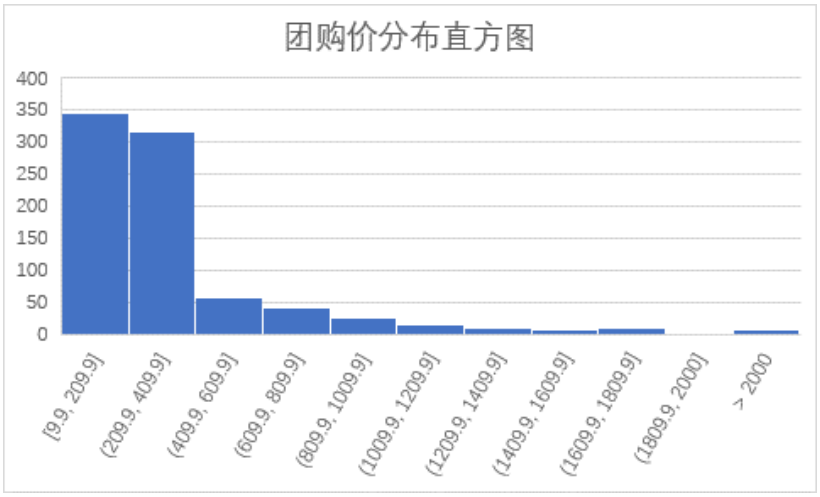

图 4.1.2.1 团购价分布直方图

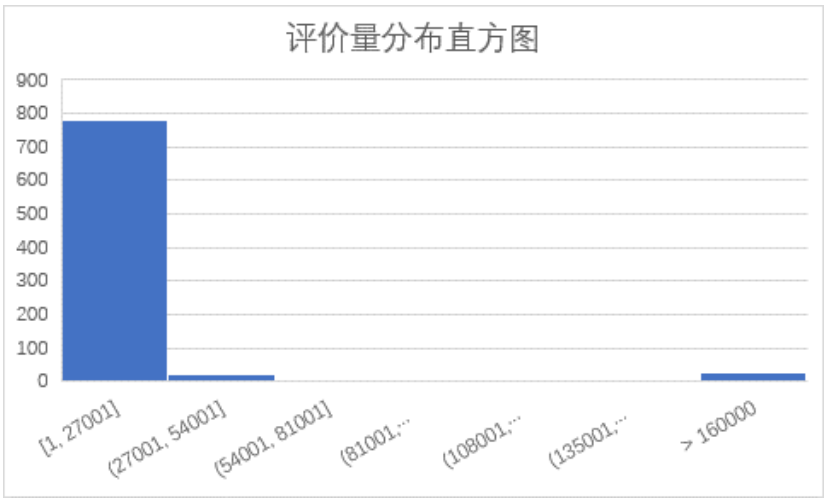

图 4.1.2.2 评价量分布直方图

对于团购价，由计算得到的均值可得，奶粉的平均价格为 367 元，而由直方图的分布来看，奶粉的价格大部分在 1000 元以下，较少部分超过 1000 元。

对于评价量，由直方图可看出评价量主要集中在 100000 以下，评价量越高

的奶粉，评价人数越少，可以看出人们更倾向于选择购买人数较多的奶粉。进一步分析，我们可知随着奶粉的价格越来越高，评价量也越来越少，说明大众对于价格的承受能力也应该考虑在内。

(二)、问题二

4.2.1 问题二的分析

以评价量为因变量，探究其他变量和评价量的关系，我们打算分为两个方面。第一，对于非数据变量——品牌、奶源场地和适用年龄对评价量的影响，利用MATLAB 绘出以以上因素为自变量、评价量为因变量的直方图;第二，对于数据变量——团购价和商品毛重等与评价量的关系，采用插值和拟合的方法进行分析求解，研究其变化规律。

4.2.2 问题二的模型建立与求解

4.2.2.1 非数据变量与评价量的关系

采用控制变量的方法，改变某一个变量,其余变量保持不变，分析评价量和该变量之间的关系。

1. 奶粉品牌与评价量的关系

分类统计数据，考虑主流的奶粉品牌，将其余出现次数较少的品牌合为一类，求取每一类的平均值，绘出以下直方图：

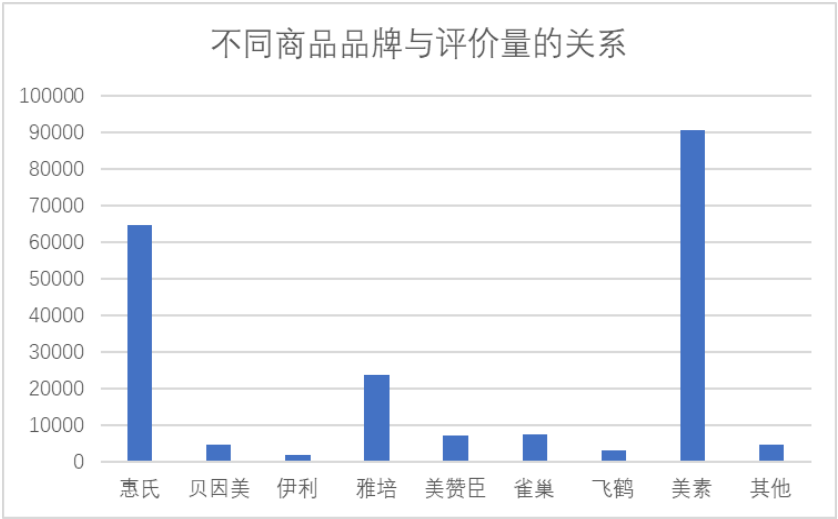

图 4.2.2.1.1 不同商品品牌和评价量的直方图

从该图可看出，评价量前三的品牌是美素、惠氏和雅培，这些都是知名度较高的奶粉品牌，我们猜测消费者可能更愿意选择大品牌和购买人数多的品牌。

2. 奶源场地与评价量的关系

奶源产地共有八处，而由于各个地区的地理环境、生产水平和奶牛品种等不同，所生产出的奶粉品质也存在差异。分析奶源原产地与评价量的关系，我们得到如下的直方图：

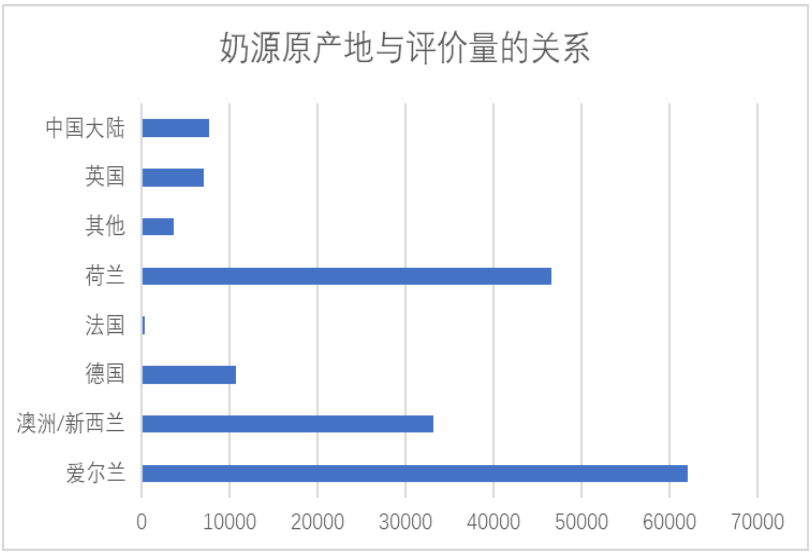

图 4.2.2.1.2 不同奶源场地和评价量的直方图

由该直方图可知，奶源场地在爱尔兰的奶粉评价量最高，其次在荷兰，再其次在澳洲和新西兰，最差的是法国。而我们也可看出，国产奶粉在国内的市场并不能说是炙手可热，联想到之前的“毒奶粉”事件，我们认为国产奶粉要想更进一步地占领国内市场，还有一段路要走。

3. 适用年龄与评价量的关系

从婴幼儿奶粉的角度来看，奶粉的适用人群一般是 0 到 6 岁的孩童绘制出奶粉适用人群饼状图如下：

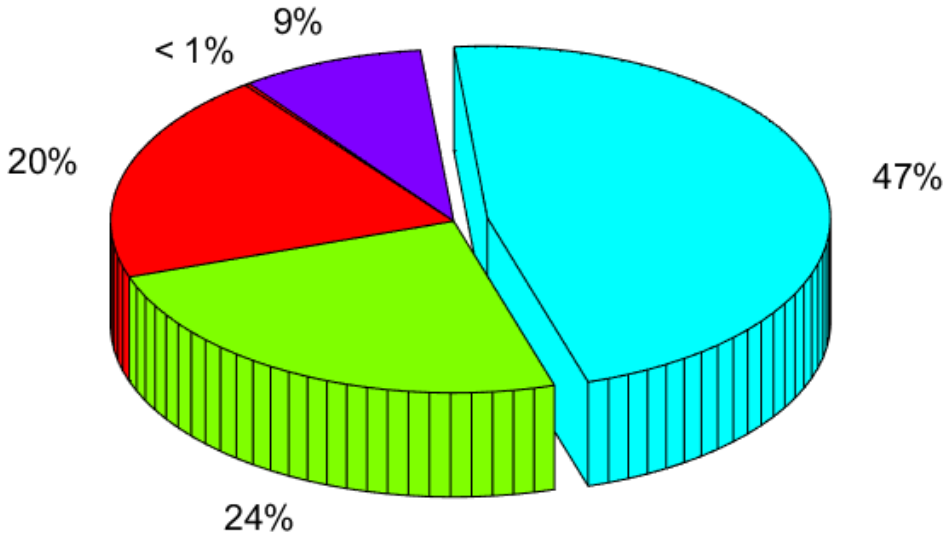

图 4.2.2.1.3 不同年龄段评价量的饼状图

从该图可看出，年龄在 1-3 岁和 3-6 岁的评价量占比较大，即说明 1-3 岁和 3-6 岁的奶粉市场需求较大，年龄在 6 岁及其以上的评价量最小。所以我们建议商家参照此年龄比率，增大针对年龄在 1-3 岁和 3-6 岁的奶粉供应，并区分不同年龄段孩子对奶粉的需求。

4. 包装单位与评价量的关系

分析奶粉的数据，我们发现奶粉在包装质地上也有所区别，分为袋装、盒装、桶装和箔装，根据包装单位的不同，绘制出如下直方图：

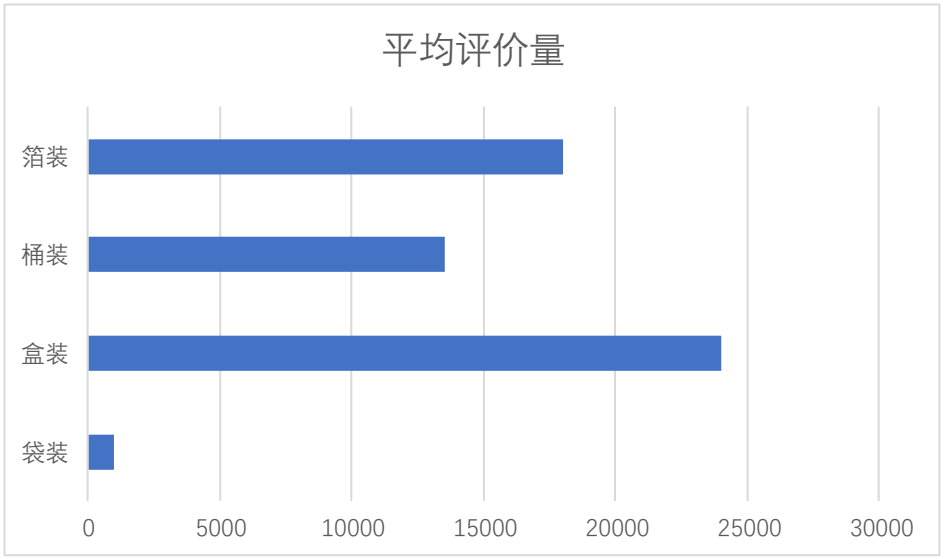

图 4.2.2.1.4 不同包装单位和评价量的直方图

由此我们可以看出，最受消费者青睐的是盒装奶粉，袋装奶粉不受欢迎。所以，商家在奶粉的外包装上也要加以选择，迎合消费者的心理，让自家奶粉更有市场。

5. 配方与评价量的关系

通过比对不同奶粉之间的参数差别，我们发现不同奶粉的配方可能也不一样，分为常规配方、特殊配方和有机三类，绘制出如下直方图：

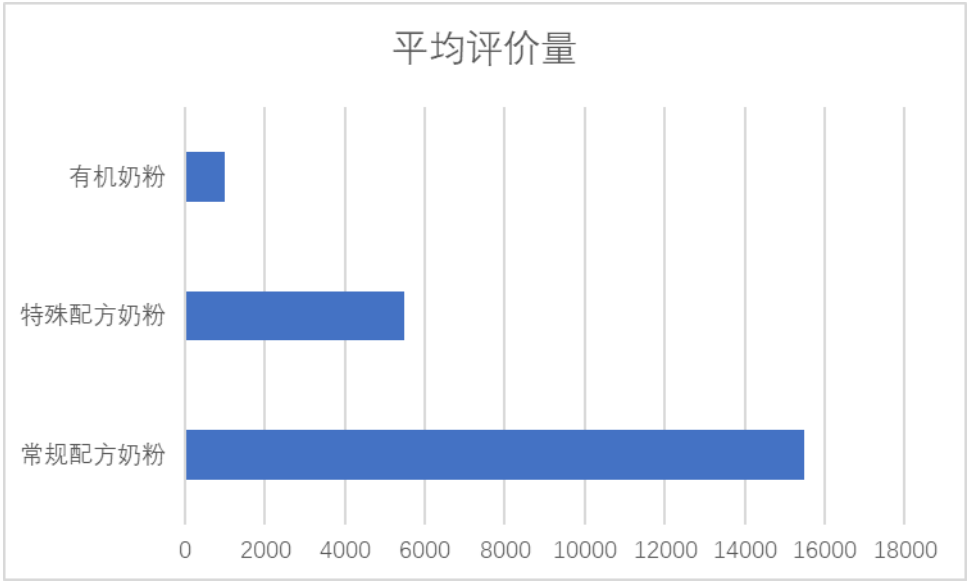

图 4.2.2.1.5 不同配方和评价量的直方图

我们可以看出，配方的不同也会影响奶粉的评价量，而图中可知，常规配方奶粉最受消费者欢迎，有机奶粉反而不太有市场，说明市场对于一般奶粉的接受度较高。

6. 段位与评价量的关系

查询资料我们了解到，一般奶粉分为三个段位，一段适合 0-6 个月的宝宝，二段适合 6-12 个月的宝宝，三段适合 12 个月以上的宝宝。本数据中还有四段奶粉，即适合 12 个月以上的宝宝。根据不同品牌的奶粉段数绘出如下直方图：

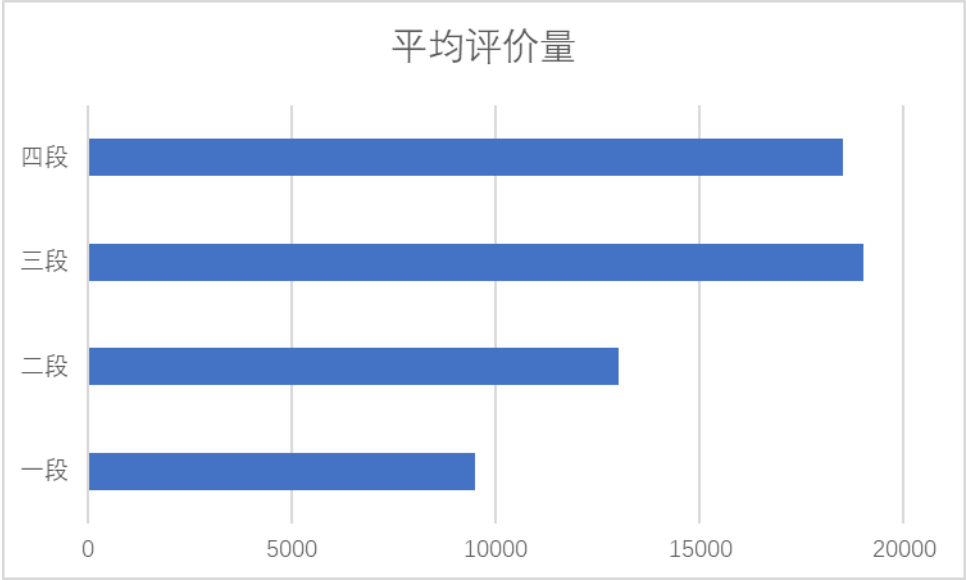

图 4.2.2.1.6 不同段位和评价量的直方图

我们可以从图中得知，大部分家长选择 3、4 段奶粉，这可能是与我国传统提倡母乳喂养有关，所以，我们建议商家加大月龄大的宝宝奶粉份额。

4.2.2.2 数据变量与评价量的关系

1．数据插值

假设给定的 n 个数据点$(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)$的观测值都是准确的，为了寻求它们所反映的关系，求解一条严格通过数据点的曲线，用它来进行分析和预测，这种方法通常称为插值法。在这类问题中，求解的关键在于选取一条哪种类型的曲线作为插值函数。由于多项式曲线是函数曲线中较为简单的曲线，因此首先考虑选取多项式函数作为插值函数来进行求解——多项式插值。

（1）多项式插值

事实上，对于给定的 n 个数据点$(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)$，总可以唯一确定一条 n-1 次多项式曲线$y = a_0 + a_1x + \cdots + a_{n-1}x^{n-1} = \sum_{i=0}^{n-1} a_i x_i$。n 个数据点都在曲线上，则有

$$\begin{cases} a_0 + a_1x_1 + \cdots + a_{n-1}x_1^{n-1} = y_1 \\ a_0 + a_1x_2 + \cdots + a_{n-1}x_2^{n-1} = y_2 \\ \vdots \\ a_0 + a_1x_n + \cdots + a_{n-1}x_n^{n-1} = y_n \end{cases} \tag{1}$$

即

$$\begin{pmatrix} 1 & x_1 & \dots & x_1^{n-1} \\ 1 & x_2 & \dots & x_2^{n-1} \\ \vdots & \vdots & & \vdots \\ 1 & x_n & \dots & x_n^{n-1} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{n-1} \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \tag{2}$$

令

$$\mathrm{A}=\begin{pmatrix} 1 & x_1 & \dots & x_1^{n-1} \\ 1 & x_2 & \dots & x_2^{n-1} \\ \vdots & \vdots & & \vdots \\ 1 & x_n & \dots & x_n^{n-1} \end{pmatrix}, \quad \mathrm{X}=\begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{n-1} \end{pmatrix}, \ \mathrm{Y}=\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

于是所求的多项式系数为方程 AX=Y 的解。由于系数矩阵的转置A^T为范德蒙矩阵，即$|A^T|$=A=$\prod_{1\le i<j\le n}(x_j-x_i)\neq 0$，根据克拉默法可知，方程组 AX=Y 有唯一的一组解 X=$(a_0, a_1, \dots, a_{n-1})^T = A^{-1}y$，再令 p=$(a_{n-1}, a_{n-2}, \dots, a_0)$，利用 MATLAB 提供的计算以向量 p=$(a_{n-1}, a_{n-2}, \dots, a_0)$为系数的多项式值的命令 polyval，可以求得多项式函数任一点的值。

（2）分段线性插值

当每个子区间上为一次多项式插值时，相应的插值称为分段线性插值。几何上为相邻两个数据点间用直线连接，此时，估计的中间值落在数据点之间的直线上。当然，当数据点个数增加和它们之间的距离减小时，线性插值就更加准确。

（3）三次样条插值

如果不采用直线连接数据点，而采用某些平滑、变化平缓的曲线来拟合数据点。最常用的方法是用一个 3 次多项式，来对相继数据点之间的各段建模，使其满足相邻两个 3 次多项式在节点处 1 阶、2 阶导数都相等，这样可以确定内部各段上的 3 次多项式，并且多项式通过节点的斜率和曲率是连续的，而第一个和最后一个多项式必须附加其他约束条件，使其得以确定。这种插值被称为三次样条插值或样条插值。

事实上，对于给定的 n 个数据点$(x_1,y_1),(x_2,y_2),\dots,(x_n,y_n)$，设函数 S(n)在每个子区间$[x_i, x_{i+1}]$上为一个三次多项式函数$S_i$(x)= $a_i x^3 + b_i x^2 + c_i x + d_i$ $(i = 1,2, \dots, n-1)$，满足：

① $S_{i-1}(x_i)=S_i(x_i) \quad (i = 2, \dots, n-1)$；

② $S'_{i-1}(x_i)=S'_i(x_i) \quad (i = 2, \dots, n-1)$；

③ $S''_{i-1}(x_i)=S''_i(x_i)\quad(i=2,\dots,n-1)$。

在端点处满足：

① $S_1(x_1)=y_1, S_n(x_n)=y_n$;

② $S'_1(x_1)=\alpha_1, S'_n(x_n)=\beta_1$（$\alpha_1,\beta_1$为端点的一阶导数值）；

③ $S''_1(x_1)=S''_n(x_n)$ 。

利用以上条件可以建立确定各段三次多项式相应的大型线性方程组（通常为三角对数模型）。理论上可以证明此方程组有唯一解，从而确定可以由 n 个三次多项式组成的分段三次多项式函数 S(x)——三次样条插值函数。

2. 数据拟合

对于已知的关于自变量和因变量的一组数据$(x_1,y_1),(x_2,y_2),\dots,(x_n,y_n)$，寻找一个合适类型的函数 y=f(x)（如线性函数 y=$ax+b$，多项式函数 $y=a_0+a_1x+\cdots+a_{n-1}x^{n-1}$，指数函数 y=$x^{ax+b}$等），使其在观测点$x_1,x_2,\dots,x_n$处取的 f($x_1$), f($x_2$), …,f($x_n$)与观察值$y_1,y_2,\dots,y_n$在某种衡量尺度上最接近，从而可用 y=f(x)作为由观测数据所反映的规律的近似表示，此问题在数学上被称为最佳曲线拟合问题。

从几何意义上来看，最佳曲线拟合问题等价于确定一条平面曲线（类型给定），使它和实验数据点“最接近”。这里并不要求曲线严格通过每个已知数据点，但在总体上要求曲线在各数据点处的取值与已知观测值之间的总体误差最小，这种方法的求解过程通常称为数据拟合，其实质是多元函数的求极值问题。一般先观察散点图来确定曲线的类型，不过散点图都是相关关系的粗略表示，有时候散点图可能与几种曲线都很接近，这时建立相应的经验函数都是合理的，但由于选择不同的曲线，得到同一个问题的多个不同经验的函数，怎样从这些函数中选择最优的一个。

（1）最小二乘法

对于已知的一组数据$(x_1,y_1),(x_2,y_2),\dots,(x_n,y_n)$，首先设定某一类型的函数 y=f(x)，然后确定函数中的参数，使得在各点处的偏差$r_i=f(x_i)-y_i$ (i=1,2,…,n-1)的平方和$\sum_{i=1}^{n}r_i^2$最小，这种根据偏差平方和最小的条件确定参数的方法叫做最小二乘法。

最小二乘法中，函数 f(x)的选取是非常重要的，但同时又比较困难。通常可根据相关问题的经验来选取，在进行实验分析；针对某些实际问题，往往要对问

题进行深入研究，分析问题的总体特征，确定问题求解的数学模型，在进行数据拟合。

对于拟合目标函数通常选取一组线性无关的简单函数类（又称为拟合基函数）$\varphi_1(x),\varphi_2(x),\dots,\varphi_m(x)$的线性组合

$$f(x)=a_1\varphi_1(x)+a_2\varphi_2(x)+\cdots+a_m\varphi_m(x)=\sum_{i=1}^{m}\varphi_i(x)\ (m\le n) \tag{3}$$

通过最小二乘法求出待定常数$a_i(i=1,2,3,\dots,m)$。当基函数为幂函数类$1,x,x^2,\dots,x^m$时，相应的拟合称为多项式拟合；当基函数为指数函数类$e^{\lambda_1 x},e^{\lambda_2 x},\dots,e^{\lambda_m x}$时，相应的拟合称为指数拟合；当基函数为三角函数类$\sin x,\cos x,\sin 2x,\cos 2x,\dots,\sin mx,\cos mx$时，相应的拟合称为三角拟合。当拟合函数设定之后，最小拟合问题就转化为多元函数的最小值问题：

$$\min_{a_i(i=1,2,3,\dots,m)\in R}\sum_{k=1}^{n}\left(\sum_{i=1}^{m}\varphi_i(x)-y_k\right)^2=\min_{a_i(i=1,2,3,\dots,m)\in R}\Phi(a_1,a_2,\dots,a_m) \tag{4}$$

根据多元函数取得极值的必要条件，求得驻点满足的方程（又称为法方程组），有$\frac{\partial\Phi}{\partial a_i}=0$(i=1,2,3,…,m)，即：

$$\begin{cases}\sum_{k=1}^{m}\varphi_1(x_k)\left(\sum_{i=1}^{m}a_i\varphi_i(x_k)-y_k\right)=0\\ \sum_{k=1}^{m}\varphi_2(x_k)\left(\sum_{i=1}^{m}a_i\varphi_i(x_k)-y_k\right)=0\\ \vdots\\ \sum_{k=1}^{n}\varphi_n(x_k)\left(\sum_{i=1}^{m}a_i\varphi_i(x_k)-y_k\right)=0\end{cases} \tag{5}$$

记

$$X=\begin{pmatrix}a_0\\a_1\\\vdots\\a_{n-1}\end{pmatrix},\quad A=\begin{pmatrix}\varphi_1(x_1)&\varphi_1(x_2)&\dots&\varphi_1(x_n)\\\varphi_2(x_1)&\varphi_2(x_2)&\dots&\varphi_1(x_n)\\\vdots&\vdots&&\vdots\\\varphi_m(x_1)&\varphi_m(x_2)&\dots&\varphi_m(x_n)\end{pmatrix},\ Y=\begin{pmatrix}y_1\\y_2\\\vdots\\y_n\end{pmatrix}$$

则可将法方程组(5)表示成矩阵形式

$$GG^TP=GY \tag{6}$$

可以证明当基函数$\varphi_1(x)$, $\varphi_2(x)$, …, $\varphi_n(x)$线性无关时，方程(6)中系数矩阵GG^T可逆，所以法方程组有唯一的一组解

$$P = (GG^T)^{-1}GY \tag{7}$$

从而求得最小二乘拟合函数$f(x) = a_1\varphi_1(x) + a_2\varphi_2(x) + \cdots + a_m\varphi_m(x)$

（2）多项式曲线拟合

如果拟合基函数为幂函数类：$1, x, x^2, \ldots, x^m$，则拟合目标函数为一个 m 次多项式函数$y = a_0 + a_1x + \cdots + a_{m-1}x^{m-1} + a_mx^m$。根据最小二乘法的思想，问题归结为 m+1 元函数

$$Q(a_0, a_1, \ldots, a_m) = \sum_{i=1}^{n}\left(\sum_{j=0}^{m} a_j x_i^j - y_i\right)^2$$

的最小值问题，利用多元函数取极值的条件

$$\frac{\partial Q(a_0, a_1, \ldots, a_m)}{\partial a_k} = 0\ (k = 0,1, \ldots, m) \tag{8}$$

得到法方程组

$$\sum_{i=1}^{n}\left(\sum_{j=0}^{m} a_j x_i^j - y_i\right)x_i^k = 0\ (k = 0,1, \ldots, m) \tag{9}$$

求解此方程可以求得拟合多项式的系数$a = (a_m, a_{m-1}, \ldots, a_0)^T$，从而求得关于已知数据点的 m 次拟合多项式函数。

（3）其他类型的拟合

当拟合基函数选取其他类函数时，利用最小二乘法进行拟合就得到相应类型的拟合曲线。常用的拟合函数还有指数函数类$y = \sum_{n=1}^{m} c_n G^{\lambda_n x}$、三角函数类$y = \sum_{n=1}^{m}(a_n \cos nx + a_n \sin nx)$等。对于实际问题需要进行曲线拟合时，到底选择哪一类函数，一方面往往根据经验来做出适当的选择。比如，根据以往统计的相关数据来拟合预测某一地区的气温或降雨量等变化，需要利用周期函数类（三角函数）来拟合；另一方面，多数问题无经验可循，在这种情况下，往往要对问题进行深入研究分析，找出问题的整体规律，确定相应的目标函数类，这样拟合出来的曲线才能比较准确地反映数据点的变化规律，依次进行预测才有意义。

3．商品毛重与评价量的关系

通过 MATLAB 三种插值结果来看，线性插值的效果较为理想，如下图 6.2.2

（1）所示：

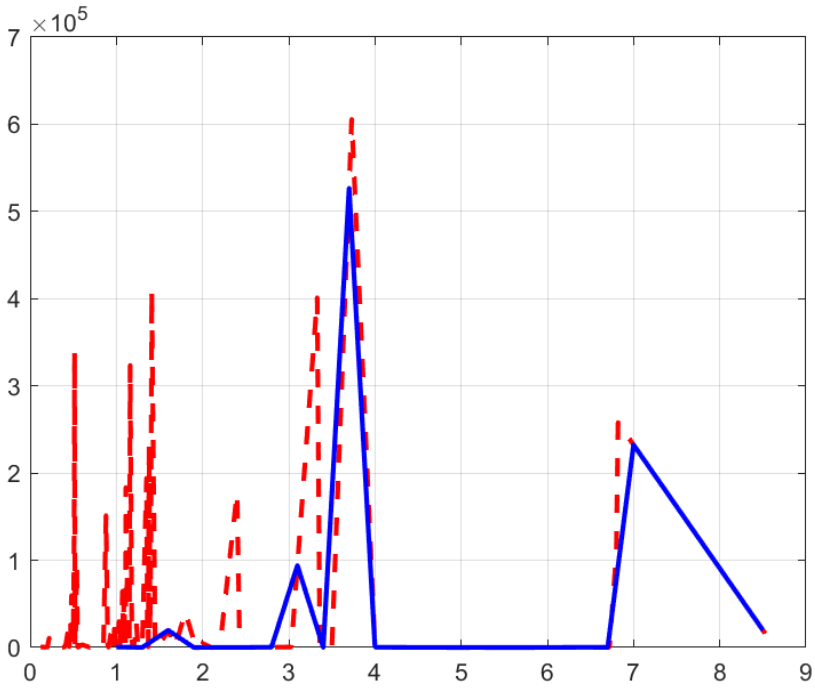

图 4.2.2.2.1 商品毛重与评价量线性插值结果

（红线:原始数据　蓝线:线性插值）

从结果来看，插值效果较好。分析图 6.2.2（1）可知，商品毛重对评价量的影响主要分布在 0-4kg，图形的顶点 3.7kg 左右，也就是说商品毛重在 3.7kg 左右最受消费者喜爱。

4. 团购价与评价量的关系

通过 MATLAB 对团购价与评价量线性插值、三次插值、样条插值、最邻近插值结果如图 6.2.2（2）所示：

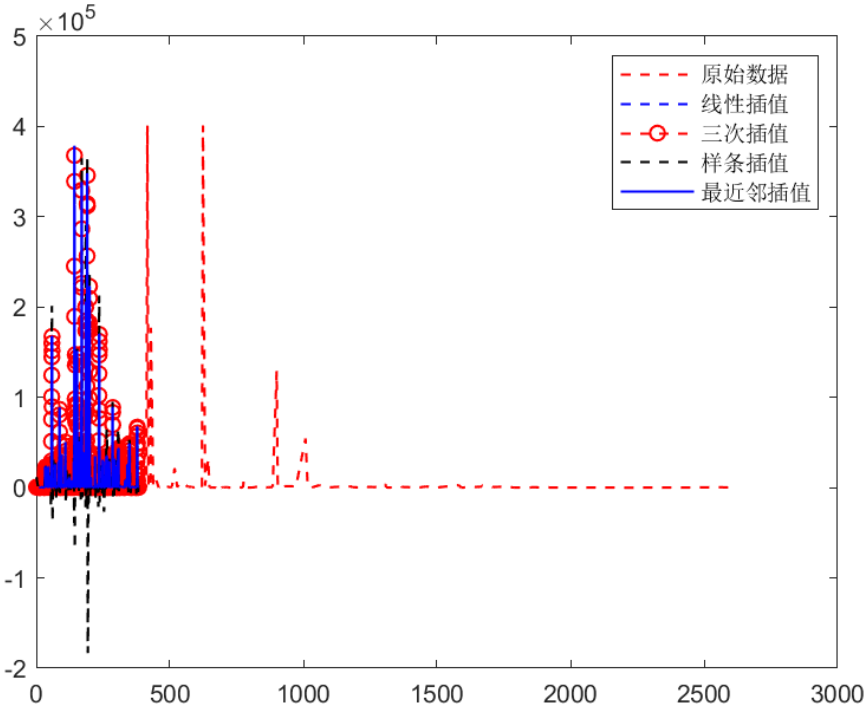

图 4.2.2.2.2 团购价与评价量的插值结果

团购价与评价量的拟合结果如图 6.2.2（3）所示：

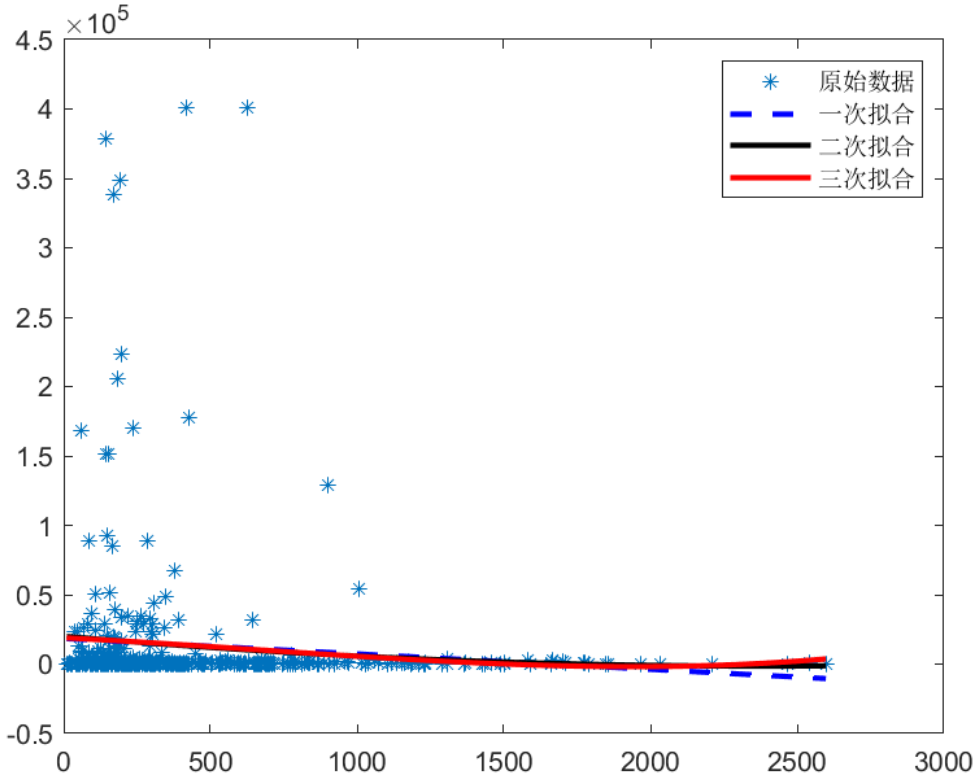

图 4.2.2.2.3 团购价与评价量的拟合结果

上图中最明显的红线是三次拟合曲线方程为：

$$y = 18941.4170 - 10.5071x - 0.0057x^2 + 2.8793 \times 10^{-6}x^3 \quad (10)$$

分析图 6.2.2（2）可知团购价在 0-500 元是大众价格，是最大市场需求人群所能承受的价格区间，而 100-200 元和 500-600 元之间评价量最大，说明处于这个价位的奶粉最受市场欢迎。但是处于高价位 1000-2500 元即属于奢侈品的奶粉也有少部分的市场需求量，却微乎其微，这部分可能主要用于富人购买。总之，销售商和生产商既要考虑普通大众的价格承受能力，又要满足富人对奶粉的高标准和低需求，不同价位的奶粉商品产品结构应该有所调整，不同价位的奶粉商品应该符合普通大众和富人阶层的人群比例。

(三)、问题三

4.3.1 问题三的分析

如何有效利用所给奶粉数据文本的信息来解决问题呢？

研究影响评价量的因素，可以置换为求取灰关联度。因此，我们采用灰色系统中灰色关联度分析法研究各个变量对评价量的影响程度，得到它们的影响大小的相关排序。由于商品名称、奶源产地、国产或进口、适用年龄、包装单位、配方、分类、段位都以文字语言描述为主，不便于问题研究，但是对评价量的影响较大。为了解决这个难题，我们采取两个方式处理文字语言，将其转化为 846*11 的数据矩阵，然后通过统计学方法分析数量关系。最后以评价量为因变量，通过 MATALB 绘出其他变量和评价量的条形图，综合分析其他变量和评价量的关系。

4.3.2 文字语言转数据处理

（1）方法一

奶粉数据文本共有 846 条信息，在这 846 条信息中由于奶粉的品牌、场地和国产或进口等的不同，导致奶粉的评价量也不一样。人为引入影响值 a_{ij} 代替文字语言，定义影响值:

a_{ij}=随机变量 X_i 的某个类别 j 出现的频率/随机变量 X_i 出现的总数 N(11)

式中: N 是固定值 846。

于是，所有的文字语言都可以用影响值代替，得到 846*11 的数据矩阵。

（2）方法二

1.文本挖掘的基本概念

1.1 数据挖掘

近年来，数据挖掘引起了信息产业界的极大关注，其主要原因是存在大量数据，可以广泛使用，并且迫切需要将这些数据转换成有用的信息和知识。获取的信息和知识可以广泛用于各种应用，包括商务管理，生产控制，市场分析，工程设计和科学探索等。

数据挖掘是人工智能和数据库领域研究的热点问题，所谓数据挖掘是指从数据库的大量数据中揭示出隐含的、先前未知的并有潜在价值的信息的非平凡过程。数据挖掘是一种决策支持过程，它主要基于人工智能、机器学习、模式识别、统计学、数据库、可视化技术等，高度自动化地分析企业的数据，作出归纳性的推理，从中挖掘出潜在的模式，帮助决策者调整市场策略，减少风险，作出正确的决策。知识发现过程由以下三个阶段组成：①数据准备；②数据挖掘；③结果表达和解释。数据挖掘可以与用户或知识库交互。

数据挖掘是通过分析每个数据，从大量数据中寻找其规律的技术，主要有数据准备、规律寻找和规律表示三个步骤。数据准备是从相关的数据源中选取所需的数据并整合成用于数据挖掘的数据集;规律寻找是用某种方法将数据集所含的规律找出来；规律表示是尽可能以用户可理解的方式（如可视化）将找出的规律表示出来。数据挖掘的任务有关联分析、聚类分析、分类分析、异常分析、特异群组分析和演变分析等。

1.2 文本挖掘

文本挖掘是抽取有效、新颖、有用、可理解的、散布在文本文件中的有价值知识，并且利用这些知识更好地组织信息的过程。

文本挖掘是信息挖掘的一个研究分支，用于基于文本信息的知识发现。文本挖掘利用智能算法，如神经网络、基于案例的推理、可能性推理等，并结合文字处理技术，分析大量的非结构化文本源（如文档、电子表格、客户电子邮件、问题查

询、网页等），抽取或标记关键字概念、文字间的关系，并按照内容对文档进行分类，获取有用的知识和信息。

文本挖掘是一个多学科混杂的领域，涵盖了多种技术，包括数据挖掘技术、信息抽取、信息检索，机器学习、自然语言处理、计算语言学、统计数据分析、线性几何、概率理论甚至还有图论。

按照文本挖掘的对象可把文本挖掘分类为:基于单文档的数据挖掘和基于文档集的数据挖掘。由于本问题只给出一个 txt 文本，故本文只考虑单文档的数据挖掘的情况。

2.向量空间模型(VSM 模型)

向量空间模型把对文本内容的处理简化为向量空间中的向量运算，并且它以空间上的相似度表达语义的相似度,直观易懂。当文档被表示为文档空间的向量,就可以通过计算向量之间的相似性来度量文档间的相似性。文本处理中最常用的相似性度量方式是余弦距离。基本思想是使用词袋法(Bagof-Word)表示文本,这种表示法的一个关键假设,就是文章中词条出现的先后次序是无关紧要的,每个特征词对应特征空间的一维,将文本表示成欧氏空间的一个向量。它的核心概念可以描述如下:

特征项:组成文档的字、词、句子等。$Document = D(t_{1,}t_{2,}...,t_{k,}...t_{n})$，其中$t_k$表示第 k 个特征项,作为一个维度。

特征项的权重:在一个文本中,每个特征项都被赋予一个权重,以表示特征项在该文本中的重要程度。

向量空间模型(VSM, Vector Space Model): 在舍弃了各个特征项之间的顺序信息之后,一个文本就表示成向量,即特征空间的一个点。如文本d_i的表示:

$V(d_i) = (w_{i1,}w_{i2,}...w_{ik,}...w_{in})$其中，$w_{ik} = f(t_k, c_j)$为权值函数,反映特征$t_k$决定文档$d_i$是否属于类$c_j$的重要性。

相似度(similarity) :对于所有文档都可映射到此文本向量空间，从而将文档信息的匹配问题转化为向量空间中的矢量匹配问题。n 维空间中点的距离用向量之间的余弦夹角来度量,也即表示了文档间的相似程度。假设目标文档为 U,未知文档为V_i,夹角越小说明文档的相似度越高。相似度计算公式如下:

$$similarity(U,V_i) = \cos(U,V_i) = \frac{U * V_i}{U \bigcirc V_i} = \frac{\sum_{k=1}^{m} w_{ik} \bigcirc w_k}{\sqrt{\sum_{k=1}^{m} w_{ik}^2 \sum_{k=1}^{m} w_k^2}} \quad (12)$$

权重通常是特征项在文档中所出现频率的函数，用

$$w_{ik} = tf_k(d_i)(14)$$

表示特征

$$w_{ik} = \begin{cases} 1, if tf_k(d_i) > 0 \\ 0, otherwise \end{cases}(13)$$

在文档 d_i 中出现的频率，权重函数有多种:

a. 最简单的布尔模型:

$$w_{ik} = \begin{cases} 1, if tf_k(d_i) > 0 \\ 0, otherwise \end{cases}(13)$$

b. 词频型:

$$w_{ik} = tf_k(d_i)(14)$$

c. 平方根型:

$$w_{ik} = tf_k(d_i)^{\frac{1}{2}}(15)$$

d. 对数型:

$$w_{ik} = \lg(tf_k(d_i)+1)(16)$$

e. TF-IDF 公式:

$$w_{ik} = tf_k(d_i)\lg(\frac{N}{N_K}+0.5)(17)$$

比较著名的权值函数是由 Salton 在 1988 年提出的 TF-IDF 公式，N 为训练文本总数，N_K 为训练文本集中出现词条 t_k 的文本数。在本文研究中选取 TF-IDF 公式，并且研究只有一个文本的情况，即 $\frac{N}{N_K}$ = 1, 故公式(17)简化为:

$$w_{ik} = tf_k(d_i)\lg(1.5)(18)$$

在这里只有一一个文本，所以 i= 1.

根据 TF-IDF 公式计算所得权值 w_{ik}，我们引入特征项 t_k 的贡献值 γ_k，表示贡献程度, 即

γ_k = w_{ik}。于是奶粉数据 txt 文本所有的信息都能用上, 得到 846X11 的数据矩阵，然后将其标准化，即:

$$x_{ij}^* = \frac{x_{ij} - \bar{\mu}_i}{\sigma_i} \quad (i = 1, 2\cdots n; j = 1, 2, \cdots p)$$

其中 $\bar{\mu}_i = \dfrac{\sum_{j=1}^{p} x_{ij}}{p}$ $\sigma_i = \sqrt{\dfrac{1}{p-1}\sum_{j=1}^{p}\left(x_{ij}-\bar{\mu}_i\right)^2}$，在这里 n=11， p= 846。

五、模型建立与求解

5.1 灰色系统的概念

信息不完全的系统称为灰色系统。信息不完全一般指：系统因素不完全明确；因素关系不完全清楚；系统结构不完全知道；系统的作用原理不完全明了。对于灰色系统，通过已知信息来研究和预测未知领域从而达到了解整个系统的目的。灰色系统理论与概率论、模糊数学一起成为研究不确定性系统三种常用方法，具有能够利用“少数据”建模寻求现实规律的良好特性，克服了数据不足或系统周期短的矛盾。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。

5.2 灰色关联分析

灰色系统理论提出了对各子系统进行灰色关联度分析的概念，意图透过一定的方法，去寻求系统中各子系统（或因素）之间的数值关系。因此，灰色关联度分析对于一个系统发展变化态势提供了量化的度量，非常适合动态历程分析。

灰色关联分析主要对态势发展变化的分析，也就是对系统动态发展过程的量化分析。灰色关联分析方法的基本思想是根据序列曲线的几何形状的相似程度来判断是否紧密，曲线越接近，相应序列之间的关联度就越大，反之越小。统计分析中的相关分析等方法是研究个因素之间的关联程度的一种有效方法，但它往往是需要大量的统计数据，计算量大，而且可能会出现反常的情况。为此，针对灰色系统中采用关联度分析的方法来研究相应的问题。以下分六个步骤进行灰关联分析，如下:

1.根据评价目的的确定评价指标体系，收集评价数据

根据 7.2 处理的结果，得到 846 X 11 的纯数字矩阵 B。$X_1, X_2, \ldots X_n$ 分别表示商品品牌贡献值、商品毛重、奶源产地贡献值、国产或进口贡献值、适用年龄贡献值、包装单位贡献值、配方贡献值、分类贡献值、段位贡献值、团购价(n= 10)的评价对象。矩阵如下:

$$B=\left(X_1,X_2,\cdots,X_n\right)=\begin{pmatrix} x_1(1) & x_2(1) & \cdots & x_n(1) \\ x_1(2) & x_2(2) & \dots & x_n(2) \\ \vdots & \vdots & & \vdots \\ x_1(p) & x_2(p) & \dots & x_n(p) \end{pmatrix}(p=846)$$

其中

$$X_i=(x_i(1),x_i(2),...x_i(p))^T \text{。}$$

2.确定参考数据列

参考数据列应该是一个比较理想的比较标准，可以以各指标的最优值(或最劣值)构成参考数据列，也可以根据评价目的选择其他参照值。记作

$$X_0=(x_0(1),x_0(2),...x_0(p))^T$$

3.对指标数据进行无量纲化

无量纲化后的数据序列仍记为$(X_0,X_1,X_2,...X_n)$,形成如下矩阵:

$$\left(X_0,X_1,\cdots,X_n\right)=\begin{pmatrix} x_0(1) & x_1(1) & \cdots & x_n(1) \\ x_0(2) & x_1(2) & \dots & x_n(2) \\ \vdots & \vdots & & \vdots \\ x_0(p) & x_1(p) & \dots & x_n(p) \end{pmatrix}$$

4.逐个计算每个被评价对象指标序列(比较序列)与参考序列对应元素的绝对差值，即:

$$\Delta_i(k)=\left|x_0(k)-x_i(k)\right|(k=1,2,\cdots,p;i=1,2,\cdots,n)(20)$$

5.确定

$$\min_i\min_k\left|x_0(k)-x_i(k)\right|(21)\text{与}\max_i\max_k\left|x_0(k)-x_i(k)\right|(22)$$

6.计算关联系数

由5可计算每个比较序列与参考序列对应元素的关联系数。即

$$\zeta_i(k)=\frac{\min_i\min_k\left|x_0(k)-x_i(k)\right|+\rho\max_i\max_k\left|x_0(k)-x_i(k)\right|}{\Delta_i(k)+\rho\max_i\max_k\left|x_0(k)-x_i(k)\right|}(23)$$

式中 ρ 分辨系数，在(0,1)内取值，若 ρ 越小，关联系数间差异越大，区分能力越强。通常取 ρ= 0.5。

7.计算关联度

对各个评价对象(比较序列)分别计算其个指标与参考序列对应元素的关联系数的均值，以反映各评价对象与参考序列的关联关系，并称其为关联度，记为:

$$\gamma_i=\frac{1}{p}\sum_{k=1}^{p}\varepsilon_i(k)(24)$$

5.3 问题三求解结果分析

通过 MATLAB 求解，可得到评价量与商品品牌贡献值、商品毛重、奶源产地贡献值、国产或进口贡献值、适用年龄贡献值、包装单位贡献值、配方贡献值、分类贡献值、段位贡献值、团购价的关联度，分别用 $r_1,r_2,...r_{10}$ 表示，如表 1:

关联度	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	γ_7	γ_8	γ_9	γ_{10}
数值	0.6652	0.8311	0.4791	0.5545	0.4607	0.4419	0.3680	0.4122	0.5172	0.8068

并得到评价量与其他变量的关联的的直方图，如图 5.3.1 可知

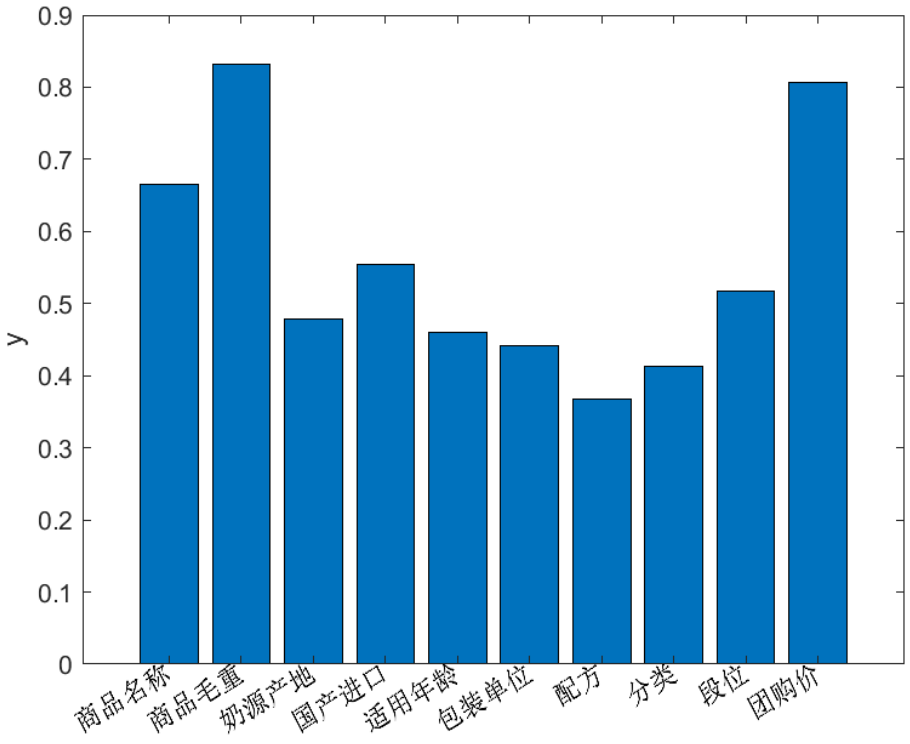

图 5.3.1

从图 7.3.3 可知，各个评价对象与评价量的影响程度从大到小排序可得:商品毛重>团购价>商品名称>国产或进口>段位>奶源场地>适用年龄>包装单位>分类>配方。从关联度结果来看,对评价量影响程度处于前三名的评价对象分别为商品毛重、团购价和商品名称，当然国产或进口、段位和奶源场地等也很重要。

5.4 问题三模型评价

针对问题三，研究其他变量对评价量的影响程度，和灰关联度不谋而合。因此，我们采用灰色系统中灰色关联度分析法研究各个变量对评价量的影响程度，得到它们的影响大致的排序,影响程度前三的分别为商品毛重、商品品牌和商品团购价，这和大众购买商品考虑的因素具有相同的原因，即考虑量入为出和品牌效应。总得说来，灰色关联度分析能够较好解决此类问题。

问题三通过数据挖掘中的文本挖掘处理奶粉数据，然后通过灰关联度分析得到较好的预期效果。

但文本挖掘的方法有待改进，可以采用其他数据处理的办法，使结果更加客观公正。

文本挖掘中 TF-IDF 公式适合文本很多的情况，处理单文本需要改进。还可以采用方法一的影响值来分析问题三，综合比较差异点。

5.5 问题四

5.5.1 给商家的建议

从市场需求量和产品结构两个方面为商家出谋划策，占领消费市场，把握产品结构将会促进商家提高盈利率。

首先要考虑的问题是奶粉市场，如何快速占领市场。更具体地说，市场上哪些商品更受欢迎？通过对问题三的研究，我们发现商品的受欢迎程度受到很多因素的影响，最重要的商品是毛重，其次是商品品牌和价格。同时，根据问题一和问题二，了解到商品品牌、奶源地、国产或进口、适用年龄、分段、包装单位、毛重、价格等因素对评价量的影响，最终得出影响奶粉市场的主要因素。

我们建议商家注意奶粉的包装方式，关注不同年龄段奶粉的需求，根据市场供求关系制定合理的价格。其次,根据不同年龄段的需求控制奶粉产品的比例,满足不同阶层的人的需求,适合大众的奶粉多生产,满足富人阶层的奶粉少生产,严格控制商品的重量和质量，逐渐淘汰重量在 4kg 到 4.7kg 的奶粉产品。

5.5.2 给消费者的建议

对于消费者来说，既要考虑自身情况，又要权衡奶粉的性价比，我们从品牌效应、场地和价格等方面提出合理化建议。

对于消费者来说，要根据实际情况选择合适的奶粉。每个孩子年龄不一样，需要针对性的奶制品，不能够随大众，应该考虑孩子自身的情况。同时，价格在 500-600 元和 100-200 元的奶粉最受欢迎，所以有孩子的消费者可以考虑。奶粉品牌做得比爱尔兰和美素，注重品牌效应的消费者可以考虑。奶源场地和是否是国产奶粉也是衡量奶粉受欢迎程度的重要指标。奶源地在荷兰和爱尔兰被优先考虑，进口奶粉比国产奶粉更受欢迎。

六、模型评价

6.1 模型优点

采用统计分析、插值和拟合和灰关联度分析的方法，合理的解决了本文研究的主要问题，取得了满意的成果。.

问题一进行单变量分布分析，分别研究团购价和评价量,求出中位数、平均数和偏度系数等相关系数，并且取得了预期的效果。

问题二选取非数据变量——品牌、奶源场地和适用年龄和数据变量——团购

价和商品毛重作为自变量，以评价量为因变量，通过直方图直观分析以及插值和拟合方法的灵活应用，得到各个变量和评价量的关系。

问题三采用数据挖掘中的有关文字挖掘技巧处理文本，将文字语言转化成为合理的数据，然后通过灰关联度分析得到较好的预期效果。

6.2 模型缺点

问题二中商品毛重拟合效果不尽人意，需要改进数学方法。

问题三中文本挖掘的方法有待改进，可以采用其他数据处理的办法，使结果更加客观公正。

6.3 模型改进

文本挖掘中 TF-IDF 公式适合文本很多的情况，处理单文本需要改进。还可以采用方法一的影响值来分析问题三，综合比较差异点。

6.4 模型应用与推广

统计分析法通过对研究对象的规模、速度、范围、程度等数量关系的分析研究，认识和揭示事物间的相互关系、变化规律和发展趋势，达到对事物的正确解释和预测。世间任何事物都有质和量两个方面，认识事物的本质时必须掌握事物的量的规律。数学已渗透到一切科技领域，使科技日趋量化，电子计算的推广和应用，量度设计和计算技术的改进和发展，已形成数量研究法，这已成为自然科学和社会科学研究中不可缺少的研究法。统计分析是研究大数据常用的方法之一，也应用到工程、经济和物理学等方面，均取得较好的效果，可以这样说没有统计就没有现代科技。

在离散数据的基础上补插连续函数，使得这条连续曲线通过全部给定的离散数据点。 插值是离散函数逼近的重要方法，利用它可通过函数在有限个点处的取值状况，估算出函数在其他点处的近似值。拟合就是把平面上一系列的点，用一条光滑的曲线连接起来。因为这条曲线有无数种可能，从而有各种拟合方法。拟合的曲线一般可以用函数表示，根据这个函数的不同有不同的拟合名字。 插值和拟合应用广泛，借助插值和拟合，通过数值方法可以近似求解任意曲线所围成的面积和曲线长度,在天气预测方面也非常准确。

灰色系统理论提出了对各子系统进行灰色关联度分析的概念，意图透过一定的方法，去寻求系统中各子系统（或因素）之间的数值关系。因此，灰色关联度分析对于一个系统发展变化态势提供了量化的度量，非常适合动态历程分析。灰关联度分析适用于生态系统、社会系统、经济系统以及农业系统，应用范围非常广泛，同时还弥补了统计分析的缺陷。

本文正是通过统计分析、插值和拟合以及灰关联度分析取得了较好的结果，这说明这些数学方法得到广泛认可。

七、参考文献

[1]杜天玉，蔡振雄，王吉等. 最小二乘法及其在 Matlab 中的应用[EB/OL]. 北京：中国科技论文在线 [2007-11-05]
[2]王中旺. 浅谈 MATLAB 在统计学实验教学中的应用[J]. 神州(上旬刊), 2017, 000(014):153.
[3]袁溪，数据挖掘技术及其应用[J].科技资讯. NO. 10 2010:p22, p24
[4]赵芳，马玉磊，浅析数据挖掘技术的发展及应用[J]，科技信息: P64
[5]Banchs R E，李亚宁. 用 MATLAB 做文本挖掘[J]. 国外科技新书评介, 2014, 000(006):P.16-16.
[6]柯宏发,陈永光，周广涛.电子装备干扰目标的灰关联预测分配模型[J].兵工学报，2008, 29 (2) :281-286.
[7]唐丽芳，贾冬青，孟庆鹏. 用 MATLAB 实现灰色预测 GM(1,1)模型[J]. 沧州师范学院学报, 2008, 24(2):35-37.

八、工作分工情况

（写明每名同学分工情况，并给自己打分，最高 10 分）

姓名	分工	分数
张世琛	编程	9
李选	建模	9
张子锜	写作、搜集资料	9
曹佳慧	写作、搜集资料	9

九、附程序

xiangguan.m

```
clc;
clear;
A=xlsread('data','A2:K847')
a=size(A,1);
b=size(A,2);
MX=max(A);
MN=min(A);
f1=1;
f2=1;
for j=1:b
    for i=1:a
        A(i,j)=(A(i,j)-MN(j))/(MX(j)-MN(j));
    end
end
A1=[A(:,1) A(:,1)   A(:,1) A(:,1) A(:,1) A(:,1) A(:,1) A(:,1) A(:,1) A(:,1)]
```

```
ping=abs(A(:,2:11)-A1);
p=0.5;
MX=max(ping)
MX=max(MX)
MN=min(ping)
MN=min(MN)
ping=(MN.*ones(a,b-1)+p*MX.*ones(a,b-1))./(ping+p*MX.*ones(a,b-1));
ping1=sum(ping,1)./a;
result=ping1
x=1:10;
y=result
figure
bar(x',y','stacked');
ylabel('y');
colormap(cool);
set(gca,'XTick',[1 2 3 4 5 6 7 8 9 10]);
set(gca,'XTicklabel',{'商品名称','商品毛重','奶源产地','国产进口','适用年龄','包装单位','配方','
分类','段位','团购价'});
h=gca;
th=fun(h,30);
clear
clc
figure
data=[168 207 396 73 2];
label={'0-0.5','0.5-1','1-3','3-6','6+'};
explode=[0 0 1 0 0 ];
bili=data/sum(data);
baifenbi=num2str(bili'*100,'%1.2f');
baifenbi=[repmat(blanks(2),length(data),1),baifenbi,repmat('%',length(data),1)];
baifenbi=cellstr(baifenbi);
Label=strcat(label,baifenbi');
pie(data,explode,Label);
colormap jet;
pie3(data,explode);
colormap hsv
```

tu1.m

```
clear;clc;
tuangou=xlsread('tuangou');
ping1=xlsread('ping1');
result=zeros(383,1);
tuan=zeros(383,1);
a=tuangou(1,1);
```

```
tuan(1,1)=a;
k=0;
for i=1:846
    if tuangou(i,1)==a
        k=k+1;
    end
end
for i=1:k
    result(1,1)=result(1,1)+ping1(i,1);
end
result(1,1)=result(1,1)./k;
for l=2:383
    a=tuangou(k+1,1);
    tuan(l,1)=a;
    j=k;
    for i=j+1:846
        if tuangou(i,1)==a
            k=k+1;
        end
    end
    for i=j+1:k
        result(l,1)=result(l,1)+ping1(i,1);
    end
    result(l,1)=result(l,1)./(k-j)
end

xx=tuan';
yx=result';
format short
hold off
xxi=1:0.3:383;
f1=interp1(xx,yx,xxi,'linear');
f2=interp1(xx,yx,xxi,'pchip');
f3=interp1(xx,yx,xxi,'spline');
f4=interp1(xx,yx,xxi,'nearext');
plot(xx,yx,'r--','linewidth',1)

hold on
plot(xxi,f1,'b--','linewidth',1)
plot(xxi,f2,'ro--','linewidth',1)
plot(xxi,f3,'k--','linewidth',1)
plot(xxi,f4,'b','linewidth',1)
legend('原始数据','线性插值','三次插值','样条插值','最近邻插值')
figure
```

```
x=tuan';
y=result';
a1=polyfit(x,y,1);
a2=polyfit(x,y,2);
a3=polyfit(x,y,3);
x1=[9.9:0.05:2598];
y1=a1(2)+a1(1)*x1;
y2=a2(3)+a2(2)*x1+a2(1).*x1.*x1;
y3=a3(1).*x1.*x1.*x1+a3(2).*x1.*x1+a3(3)*x1+a3(4);
plot(x,y,'*')
hold on
plot(x1,y1,'b--',x1,y2,'k',x1,y3,'r-','linewidth',2);
legend('原始数据','一次拟合','二次拟合','三次拟合')
```

tu2.m

```
clc
clear all
maozhong=xlsread('mao');
ping=xlsread('ping2');
result=zeros(105,1);
mao=zeros(105,1);
a=maozhong(1,1)
mao(1,1)=a;
k=0;
for i=1:846
    if maozhong(i,1)==a
        k=k+1;
    end
end;
for i=1:k
    result(1,1)=result(1,1)+ping(i,1);
end
result(1,1)=result(1,1)./k;
for l=2:105
    a=maozhong(k+1,1);
    mao(l,1)=a;
    j=k;
    for i=j+1:846
        if maozhong(i,1)==a
            k=k+1;
        end
    end
    for i=j+1:k
```

```
            result(l,1)=result(l,1)+ping(i,1);
        end;
        result(l,1)=result(l,1)./(k-j);
end
xx=mao';
yx=result';
format short
hold off
xxi=1:0.3:105;
f1=interp1(xx,yx,xxi,'linear');
plot(xx,yx,'r--','linewidth',2)
hold on
plot(xxi,f1,'b-','linewidth',2)
grid on
```

problem0.m

```
clc
clear all
data=textread('data.txt');
high=data(:,1);
high=high(:);
weight=data(:,2);
weight=weight(:);
shuju=[high weight];
jun_zhi=mean(shuju)
zhongweishu=median(shuju)
biaozhuncha=std(shuju)
jicha=range(shuju)
```

fun.m

```
function th=fun(h,rot)
if nargin==3
        x=[now-.7 now-.3 now];
        y=[20 35 15];
        figure
        plot(x,y,'-');
        datetick (x', 0, keepticks')
        h=gca;
        set(h,'position',[0.13 0.35 0.775 0.55]);
        rot=90;
end
if nargin==1
        rot=90;
```

```
end
a=get(h,'XTickLabel');
set(h,'XTickLabel',[]);
b=get(h,'XTick');
c=get(h, 'YTick');
if rot<180
    th=text(b, repmat (c(1)-.1*(c(2)-c(1)), length(b), 1), a,
'HorizontalAlignment','right','fontsize',10,'fontweight','bold', 'rotation',rot);
else
    th=text(b, repmat (c(1)-.1*(c(2)-c(1)), length(b), 1), a,
'HorizontalAlignment','left','fontsize',10,'fontweight','bold', 'rotation',rot);
end
end
```

zhuang.m

```
clc
clear all
[data1,data]=xlsread('zhuang.xls');
data=char(data);
data=abs(data);
a=zeros(846,7);
c=846.*11./log10(1.5)./10000;
reault=zeros(846,1);
for i=1:846
    a(i,1)=34955;
    a(i,2)=35013;
end
k=0;
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        k=k+1;
    end
end
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        result(i,1)=k./c;
    end
end
for i=1:846
    a(i,1)=30418;
    a(i,2)=35013;
end
k=0;
```

```
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        k=k+1;
    end
end
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        result(i,1)=k./c;
    end
end

for i=1:846
    a(i,1)=26742;
    a(i,2)=35013;
end
k=0;
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        k=k+1;
    end
end
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        result(i,1)=k./c;
    end
end

for i=1:846
    a(i,1)=31665;
    a(i,2)=35013;
end
k=0;
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        k=k+1;
    end
end
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        result(i,1)=k./c;
    end
end
xlswrite('zhuang1.xlsx',result,1,'A1:A846')
```

yang.m

```
clc
clear all
[data1,data]=xlsread('yang.xls');
data=char(data);
data=abs(data);
a=zeros(846,7);
c=846.*11./log10(1.5)./10000;
reault=zeros(846,1);
for i=1:846
    a(i,1)=29275;
    a(i,2)=22902;
end
k=0;
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        k=k+1;
    end
end
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        result(i,1)=k./c;
    end
end
for i=1:846
    a(i,1)=32650;
    a(i,2)=22902;
end
k=0;
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        k=k+1;
    end
end
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        result(i,1)=k./c;
    end
end
xlswrite('yang1.xlsx',result,1,'A1:A846')
```

peifang.m

```
clc
clear all
[data1,data]=xlsread('peifang.xls');
data=char(data);
data=abs(data);
a=zeros(846,7);
c=846.*11./log10(1.5)./10000;
result=zeros(846,1);
for i=1:846
    a(i,1)=24120;
    a(i,2)=35268;
end
k=0;
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        k=k+1;
    end
end
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        result(i,1)=k./c;
    end
end
for i=1:846
    a(i,1)=29305;
    a(i,2)=27530;
end
k=0;
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        k=k+1;
    end
end
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        result(i,1)=k./c;
    end
end

for i=1:846
    a(i,1)=26377;
    a(i,2)=26426;
end
k=0;
```

```
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        k=k+1;
    end
end
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        result(i,1)=k./c;
    end
end

for i=1:846
    a(i,1)=54;
    a(i,2)=23681;
end
k=0;
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        k=k+1;
    end
end
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        result(i,1)=k./c;
    end
end

for i=1:846
    a(i,1)=24503;
    a(i,2)=22269;
end
k=0;
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        k=k+1;
    end
end
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        result(i,1)=k./c;
    end
end
for i=1:846
    a(i,1)=27861;
```

```
    a(i,2)=22269;
end
k=0;
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        k=k+1;
    end
end
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        result(i,1)=k./c;
    end
end

for i=1:846
    a(i,1)=20854;
    a(i,2)=23427;
end
k=0;
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        k=k+1;
    end
end
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        result(i,1)=k./c;
    end
end
xlswrite('peifang1.xlsx',result,1,'A1:A846')
```

jinkou.m

```
clc
clear all
[data1,data]=xlsread('jinkou.xls');
data=char(data);
data=abs(data);
a=zeros(846,7);
c=846.*11./log10(1.5)./10000;
reault=zeros(846,1);
for i=1:846
    a(i,1)=36827;
    a(i,2)=21475;
```

```
end
k=0;
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        k=k+1;
    end
end
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        result(i,1)=k./c;
    end
end
for i=1:846
    a(i,1)=22269;
    a(i,2)=20135;
end
k=0;
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        k=k+1;
    end
end
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        result(i,1)=k./c;
    end
end
xlswrite('jinkou1.xlsx',result,1,'A2:A847')
```

ming.m

```
clc
clear all
[data1,data]=xlsread('ming.xls');
data=char(data);
data=abs(data);
a=zeros(846,7);
c=846.*11./log10(1.5)./10000;
result=zeros(846,1);
for i=1:846
    a(i,1)=24800;
    a(i,2)=27663;
end
k=0;
```

```
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        k=k+1;
    end
end
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        result(i,1)=k./c;
    end
end
for i=1:846
    a(i,1)=36125;
    a(i,2)=22240;
end
k=0;
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        k=k+1;
    end
end
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        result(i,1)=k./c;
    end
end

for i=1:846
    a(i,1)=20234;
    a(i,2)=21003;
end
k=0;
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        k=k+1;
    end
end
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        result(i,1)=k./c;
    end
end

for i=1:846
    a(i,1)=38597;
```

```
    a(i,2)=22521;
end
k=0;
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        k=k+1;
    end
end
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        result(i,1)=k./c;
    end
end

for i=1:846
    a(i,1)=32654;
    a(i,2)=36190;
end
k=0;
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        k=k+1;
    end
end
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        result(i,1)=k./c;
    end
end
for i=1:846
    a(i,1)=38592;
    a(i,2)=24034;
end
k=0;
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        k=k+1;
    end
end
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        result(i,1)=k./c;
    end
end
```

```
for i=1:846
    a(i,1)=39134;
    a(i,2)= 40548;
end
k=0;
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        k=k+1;
    end
end
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        result(i,1)=k./c;
    end
end


for i=1:846
    a(i,1)=23436;
    a(i,2)= 36798;
end
k=0;
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        k=k+1;
    end
end
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        result(i,1)=k./c;
    end
end

for i=1:846
    a(i,1)=21531;
    a(i,2)= 20048;
end
k=0;
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        k=k+1;
    end
end
```

```
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        result(i,1)=k./c;
    end
end

for i=1:846
    a(i,1)=21512;
    a(i,2)=29983;
end
k=0;
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        k=k+1;
    end
end
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        result(i,1)=k./c;
    end
end




for i=1:846
    a(i,1)=32654;
    a(i,2)=32032;
end
k=0;
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        k=k+1;
    end
end
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        result(i,1)=k./c;
    end
end


for i=1:846
```

```
    a(i,1)=22810 ;
    a(i,2)=32654;
end
k=0;
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        k=k+1;
    end
end
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        result(i,1)=k./c;
    end
end

for i=1:846
    a(i,1)=22307 ;
    a(i,2)= 20803;
end
k=0;
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        k=k+1;
    end
end
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        result(i,1)=k./c;
    end
end

for i=1:846
    a(i,1)=38597 ;
    a(i,2)= 22763;
end
k=0;
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        k=k+1;
    end
end
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        result(i,1)=k./c;
```

```
    end
end

for i=1:846
    a(i,1)=24481 ;
    a(i,2)=    23453;
end
k=0;
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        k=k+1;
    end
end
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        result(i,1)=k./c;
    end
end


for i=1:846
    a(i,1)=26126;
    a(i,2)=    19968;
end
k=0;
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        k=k+1;
    end
end
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        result(i,1)=k./c;
    end
end


for i=1:846
    a(i,1)=28595;
    a(i,2)= 20248;
end
k=0;
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
```

```
            k=k+1;
        end
end
for i=1:846
        if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
            result(i,1)=k./c;
        end
end



for i=1:846
        a(i,1)=32650 ;
        a(i,2)=32650;
end
k=0;
for i=1:846
        if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
            k=k+1;
        end
end
for i=1:846
        if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
            result(i,1)=k./c;
        end
end

for i=1:846
        a(i,1)=35834;
        a(i,2)= 20248;
end
k=0;
for i=1:846
        if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
            k=k+1;
        end
end
for i=1:846
        if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
            result(i,1)=k./c;
        end
end
```

```
for i=1:846
    a(i,1)=29305;
    a(i,2)= 31119;
end
k=0;
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        k=k+1;
    end
end
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        result(i,1)=k./c;
    end
end
xlswrite('ming1.xlsx',result,1,'A1:A846')
```

nian.m

```
clc
clear all
[data1,data]=xlsread('nian.xls');
data=char(data);
data=abs(data)
a=zeros(846,7);
c=846.*11./log10(1.5)./10000;
result=zeros(846,1);
for i=1:846
    a(i,1)=49;
    a(i,2)=45;
end
k=0;
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        k=k+1;
    end
end
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        result(i,1)=k./c;
    end
end
for i=1:846
    a(i,1)=48;
```

```
    a(i,2)=46;
end
k=0;
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        k=k+1;
    end
end
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        result(i,1)=k./c;
    end
end

for i=1:846
    a(i,1)=51;
    a(i,2)=45;
end
k=0;
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        k=k+1;
    end
end
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        result(i,1)=k./c;
    end
end

for i=1:846
    a(i,1)=48;
    a(i,2)=45;
end
k=0;
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        k=k+1;
    end
end
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        result(i,1)=k./c;
    end
```

```
end

for i=1:846
    a(i,1)=24503;
    a(i,2)=22269;
end
k=0;
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        k=k+1;
    end
end
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        result(i,1)=k./c;
    end
end
for i=1:846
    a(i,1)=27861;
    a(i,2)=22269;
end
k=0;
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        k=k+1;
    end
end
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        result(i,1)=k./c;
    end
end

for i=1:846
    a(i,1)=20854;
    a(i,2)=23427;
end
k=0;
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        k=k+1;
    end
end
for i=1:846
```

```
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        result(i,1)=k./c;
    end
end
xlswrite('nian1.xlsx',result,1,'A1:A846')
```

duan.m

```
clc
clear all
[data1,data]=xlsread('duan.xls');
data=char(data);
data=abs(data);
a=zeros(846,7);
c=846.*11./log10(1.5)./10000;
reault=zeros(846,1);
for i=1:846
    a(i,1)=49;
    a(i,2)=27573;
end
k=0;
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        k=k+1;
    end
end
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        result(i,1)=k./c;
    end
end
for i=1:846
    a(i,1)=50;
    a(i,2)=27573;
end
k=0;
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        k=k+1;
    end
end
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        result(i,1)=k./c;
```

```
    end
end

for i=1:846
    a(i,1)=51;
    a(i,2)=27573;
end
k=0;
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        k=k+1;
    end
end
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        result(i,1)=k./c;
    end
end

for i=1:846
    a(i,1)=52;
    a(i,2)=27573;
end
k=0;
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        k=k+1;
    end
end
for i=1:846
    if data(i,1)==a(i,1)&&data(i,2)==a(i,2)
        result(i,1)=k./c;
    end
end
xlswrite('duan1.xlsx',result,1,'A1:A846')
```