

## 1 问题重述

高速公路路面的寿命对降低高速公路的运行成本、保障运输安全有着极其重要的意义。其施工工艺比较简单：以分别达到各自标准的沥青、矿粉、碎石做原料，按一定的比例组成为混合料，经过充分拌和后，比较均匀地铺在已经造好的基层上、再经多次碾压路面以达到设计的压实度，最后经过一段时间让沥青路面降温就可以交付使用。

题中首先给出了一些影响公路路面质量的指标：油石比、筛孔通过率、空隙率 (VV)、矿料间隙率 (VMA)、饱和度 (VFA)、粉胶比 (DP)、毛体积密度、最大理论密度、%Gmm(最初)和%Gmm (最大)。然后又给出了高速公路路面质量的抗水损害性能、高温性能、低温性能的四个指标：冻融劈裂强度比(TSR(%))、浸水马歇尔稳定度比(S0(%))、车辙、弯拉应变。

现有一批关于原料和上述指标的数据，题中要我们对这些数据进行充分地分析、研究，以获取尽可能多的提高高速公路路面质量的有用信息，进而解决如下问题：

一、描述高速公路路面质量的抗水损害性能、高温性能、低温性能的四个指标之间有没有数量关系？如果有，建立它们之间的数学模型。

二、建立描述高速公路路面的抗水损害性能、高温性能、低温性能的四个质量指标和影响高速公路路面质量的最重要和比较重要的因素之间比较精确的数学模型，并给出选择这些因素的理由，进而根据上述模型给出可以提高高速公路路面质量的方案。

三、从理论上探讨集料的筛孔通过率（即级配组成，指混合料中粗细不同集料的构成比例）与路面压实度的上界（数学意义下的上界）之间的数量关系。

四、判断沥青、碎石质量对高速公路路面抗水损害性能、高温性能、低温性能有没有影响，以及不同厂家、不同产地的但型号相同、类型相同的沥青、碎石质量对高速公路路面抗水损害性能、高温性能、低温性能有没有影响。

五、根据对数据分析的结果，判断现在测试高速公路路面质量的试验项目中是否有重要的遗漏，进而向高速公路建设部门给出我们的建议。

## 2 问题分析

### 1.1 对附件中数据的分析

从附件一中不难看出，表格中存在一些缺失数据，显然如何处理这些数据会直接影响到模型的优劣。处理缺失数据的方法主要有两种，一种是忽略法，即直接忽略缺失值而利用剩余的值进行分析；另一种是推断填补法，即利用数据集中其它完整的数据，通过回归分析、贝叶斯计算公式或决策树等预测方法，推断出该位置最可能的取值。由于推断填补法采用的回归分析、贝叶斯公式或决策树等

方法都带有一定的误差，从而会在我们的多元线性回归模型中引入新的误差，影响模型的精确度，同时考虑到表中还存在大量的完整数据，因此在模型的建立与求解过程中，我们不采用推断填补方法，而是直接忽略这些缺失数据，当然在忽略的时候要忽略该缺失位置所在行的数据。

由于附件一中给出的是测量数据，误差在所难免，但是一些明显的异常值会对结果产生质的影响，因此在建模时首先要对这些值进行处理，对某些明显存在问题的数据，我们采用了将其所在行的数据忽略的方法。例如附件一中第 226 行的车辙指标的数据为 31450，通过与其他行数据的比较，我们发现该数据在数量级上存在误差，对结果影响太大，因此在建模时我们首先在数据集中去除了该行数据。

## 1.2 对问题一的理解

问题一要我们判断高速公路路面质量的抗水损害性能、高温性能、低温性能的四个指标之间有没有数量关系，并给出相应数学模型。即要判断附件一中最后四列值间的数量关系，这显然是一个多元回归问题。为了使结果更加细致、全面，我们首先判断任两个指标间的数量关系，然后判断任意三个指标间的数量关系，最后再判断四个指标间的数量关系。为了更加直观的表述指标间的数量关系，我们可以定义一个度量来衡量其数量关系。

在判断两个指标间的数量关系时，可以通过计算其相关系数判断其相关性，相关系数即作为衡量其数量关系的度量。

在判断三个指标间的数量关系时，我们可以采用多元回归分析的方法，由于有效数据量较小，这里我们采用多元线性回归模型进行分析。在模型的建立过程中，任取四个指标中的一个指标作为因变量，在其余三个指标中任取两个指标作为自变量，进而建立我们需要的多元线性回归模型，显然对每个因变量需要建立三个模型，以判断其与其它任意两个指标的数量关系，当因变量取遍所有的指标时即可得到每个指标与其它任意两个指标的关系。为了得到任意三个指标间的最合适关系，我们引入残差平方和，通过对残差平方和的改进引入了一个衡量因变量指标和自变量指标间数量关系表达式优劣的度量，我们称之为拟和度，拟和度越小说明残差平方和越小，即说明因变量指标和自变量指标间数量关系表达式越精确，以此作为衡量标准可以得出三个指标间最优的关系表达式。

在判断四个指标间的数量关系时，任取四个指标的一个指标作为因变量，其余三个指标作为自变量，进而建立相应的多元线性回归模型，当因变量取遍所有的指标时即可得到每个指标与其它三个指标的关系，最后通过拟和度来衡量因变量指标和自变量指标间数量关系表达式的优劣，进而找到四个指标间最优的关系表达式。

显然，数量间的关系不仅仅指线性关系，而在上述求解模型中我们只讨论了

指标间的线性关系，为了使结果更加全面，我们在模型的进一步讨论中讨论了指标间的非线性关系。

### 1.3 对问题二的理解

问题二要我们建立描述高速公路路面的抗水损害性能、高温性能、低温性能的四个质量指标和影响高速公路路面质量的最重要和比较重要的因素之间比较精确的数学模型，并根据上述模型给出可以提高高速公路路面质量的方案。

由于题中没有给出沥青、石料对质量指标影响的量值，因此在此问题的求解中，我们首先假设沥青、矿石质量对质量指标的影响可以忽略不计，进而建立每个质量指标与前面因素之间的多元回归模型，通过计算得出每个自变量的偏回归平方和，偏回归平方和可以衡量每个自变量在多元线性回归中所起的作用，从而根据每个自变量的偏回归平方和可以找到影响该指标的最重要和比较重要的因素。由于通过筛口百分率这一因素的特殊性，我们分别采用两种方法进行了建模：一种是取其中一列作为该因素的代表进行回归分析；另一种是将该因素所有列都参与回归分析。

当然忽略沥青、石料对质量指标的影响是建立上述模型的需要，但是其影响是客观存在的，因此在讨论了上述模型后，在模型的进一步讨论中，我们还分析了这两个因素的影响，即在附件一中找了几组特殊的数据，分别考虑了固定沥青、石料等信息通过分析因素差值对质量指标差值的影响，找到了影响质量指标的最重要和比较重要的因素，当然此模型一个很大的缺点就是数据量太小，结果不够精确，但可以作为第一个模型的补充和完善。

### 1.4 对问题三的理解

问集料的筛孔通过率是指集料中能够通过直径大小不同的各种筛孔的各部分的质量占集料总质量的比例，即混合料中粗细不同集料的构成比例。这个指标反映的是集料的级配组成。路面压实度的上界（数学意义下的上界）是指经过无数次的压实，在实际情况下可能达到的最大压实度，也就是说即使压实密度达到最大，试件内部还是有可能有空隙。

实际中路面压实度的上界达不到 100%，这是由两个因素导致的。

1) 集料是由各种不同形状、不同大小的碎石颗粒组成的混合物，理论上不可能使填充率达到 100%，总会有空隙存在。

2) 沥青是集料粘合剂，它的性质(比如流动性差)以及用量决定了沥青不可能填充完集料间隙，而且沥青冷却也会造成空隙。

给定一组筛孔通过率就相当于给定一个堆积颗粒系统的级配分布，根据此级配分布总存在一个最优堆积结构使得空隙率最小，此时路面压实度达到它在数学意义下的上界，即不能再压，否则球体就会“破裂”。

根据上面的分析,问题三就转化为求紧密堆积颗粒系统中颗粒级配组成与堆积率之间的数量关系。

表征颗粒堆积状态的基本参数有空隙率、堆积率、配位数、空隙(孔隙)分布等。其中:

空隙率 --- 空隙率为颗粒体中空隙所占的容积率。

堆积率 --- 堆积率  $\lambda$  表征颗粒体中固体颗粒所占的容积率

配位数 --- 每个颗粒和周围其他颗粒接触点的数目。

为了简化问题,先做如下假设:

1) 集料里的碎石颗粒都是球体。

2) 考虑到当筛孔孔径小到一定程度时,通过该筛孔的集料可以看成是连续的填充空间,即空隙率为 0;国际上以 2.36mm 为区分粗细集料的标准,所以我们假设粒径小于 2.36mm 的集料对空间的填充度为 100%,即不考虑粒径小于 2.36mm 的集料对路面压实度的影响。下面我们只考虑粒径较大的颗粒的最优堆积情况。

3) 假设颗粒粒径是不连续的,即只考虑不连续粒径颗粒的堆积问题。

解题思路:我们先从只有一种颗粒的体系入手,通过理论计算得到了最大堆积率,再考虑体系包含多种颗粒的情况。由于包含大量多尺度颗粒的体系,其级配组成与堆积率的关系还不清楚,所以我们通过查阅资料,决定先考虑几种特殊情况,比如细颗粒的尺寸和数量刚好能填满粗颗粒的空隙。然后再对一般情况应用已有的结果进行讨论,最后得出了集料级配组成与颗粒堆积密度的关系。

## 1.5 对问题四的理解

为了更好的理解该问题,我们先对该问题涉及到的一些概念进行分析:

(1) “碎石质量”中的质量不是物理学中重量的概念,而是指等级质量,即衡量好坏的一类标准,例如质量好、一般、差等。

(2) “不同厂家、不同产地的但型号相同、类型相同的沥青、碎石”理解为:不同厂家的但型号相同的沥青和不同产地的但类型相同的碎石,即沥青分厂家和型号而碎石分产地和类型。

求解该问题主要是解决两个问题,下面分别对这两个问题进行分析:

第一、判断沥青、碎石质量对高速公路路面抗水损害性能、高温性能、低温性能有没有影响。判断对性能的影响即要判断对性能四个指标的影响,显然这是一个判断因素对试验结果影响的问题,而方差分析法就是一种根据试验结果进行分析,鉴别各个有关因素对试验结果影响的有效方法。考虑到此问题涉及到沥青、碎石质量两个因素,且数据不具备重复的条件,因此我们可以采用双因素无重复试验的方差分析法进行求解。

在分析数据时,我们认为:此问题中的沥青只考虑普通沥青和改性沥青两种

类型，即沥青因素仅有两个水平；由于碎石质量由厂家和型号共同决定，因此我们认为厂家和型号均相同时才算一个水平。这样就可以建立起双因素试验的方差分析法所需要的数据集，进而建立模型解决该问题。

第二、判断不同厂家的但型号相同的沥青和不同产地的但类型相同的碎石对高速公路路面抗水损害性能、高温性能、低温性能有没有影响。这仍然是一个判断因素对试验结果影响的问题，同样可以采用方差分析法。但是通过对数据进行分析我们发现，可用作双因素试验方差分析法的数据几乎没有，因此双因素试验的方差分析法不可行，鉴于此，我们采用了单因素试验方差分析法，即独立的考虑不同厂家但型号相同的沥青和不同产地但类型相同的碎石对公路路面性能的影响。

## 1.6 对问题五的理解

问题五要我们根据对数据分析的结果，判断现在测试高速公路路面质量的试验项目中是否有重要的遗漏，进而向高速公路建设部门给出我们的建议。

从题目中所给的数据可知，现在测试高速公路路面质量的试验项目都是基于如何延长高速公路路面寿命做的，虽然使用寿命对成本和保障运输安全有极其重要的意义，但我们觉得实际中仅考虑这些是远远不够的，因为公路还有很多其他的重要属性，而这些属性又是由不同的因素决定的，因此通过查阅资料，我们找到了几个对路面质量影响比较大而试验项目中又不是很重视的因素，如果将这些因素加入到试验项目，会使试验变得更具体、全面。

## 3 模型假设

1. 假设不考虑基层质量对路面质量的影响；
2. 假设附件中给的数据是可信的；
3. 假设沥青技术标准中的气候分区对我们分析问题没有影响；
4. 假设弯拉应变的单位对分析问题没有影响，即可以不考虑；

## 4 符号说明

$y_1^{(j)} (j=1,2,\cdots,290)$ ：附件一中冻融劈裂强度比(TSR(%))指标在第  $j$  行的取值。

$y_2^{(j)} (j=1,2,\cdots,290)$ ：附件一中浸水马歇尔稳定度比(S0(%))指标在第  $j$  行的取值。

$y_3^{(j)} (j=1,2,\cdots,290)$ ：附件一中车辙指标在第  $j$  行的取值。

$y_4^{(j)} (j=1,2,\cdots,290)$ ：附件一中弯拉应变指标在第  $j$  行的取值。

$\overline{y_i}(i=1,2,3,4)$ : 表示对应的指标所在列的均值。

$\widetilde{y_i}^{(j)}(i=1,2,3,4; j=1,2,\dots,290-k)$ : 表示对应的指标去除要忽略的行后按原次序重新排列后第  $j$  行的取值, 其中  $k$  为要忽略的行的数目。

$S_{LX}$ : 拟和度。

$x_i^{(j)}(i=1,3,4,\dots,10; j=1,2,\dots,290)$ : 依次表示附件一中油石比、VV、VMA、VFA、DP、毛体积密度、最大理论密度、%Gmm(最初)和%Gmm(最大)在第  $j$  行的取值。

$x_{2,i}^{(j)}(i=1,2,\dots,13; j=1,2,\dots,290)$  表示附件一中通过筛孔百分率一栏中第  $i$  列第  $j$  行的取值。

$\widetilde{x_i}^{(j)}(i=1,3,4,\dots,10; j=1,2,\dots,290-k)$  和  $\widetilde{x_{2,i}}^{(j)}(i=1,2,\dots,13; j=1,2,\dots,290-k)$ : 分别代表相应数据在忽略某些缺失数据后重组的数据, 其中  $k$  为要忽略的行的数目。

## 5 模型建立与求解

本部分, 我们将建立该题目的数学模型并给出各问题的求解。

### 5.1 问题一的求解

为了使结果更加细致、全面, 我们将分三步求解该问题: 首先判断任两个指标间的数量关系, 然后判断任意三个指标间的数量关系, 最后再判断四个指标间的数量关系。

#### 5.1.1 任两个指标间的数量关系

我们采用相关系数作为衡量两个指标间数量关系的度量, 下面以冻融劈裂强度比(TSR(%))指标和浸水马歇尔稳定度比(S0(%))指标间的数量关系为例给出我们的模型。

设  $y_i^{(j)}(i=1,2,3,4; j=1,2,\dots,290)$  依次表示附件一中冻融劈裂强度比指标、浸水马歇尔稳定度比指标、车辙指标和弯拉应变指标在第  $j$  行的取值。

首先对附件一中的数据进行初步处理: 首先去除一些异常值(见 1.2 中对异常

值的分析), 对剩余的数据, 若第  $j$  行数据中,  $y_1^{(j)}$  或者  $y_2^{(j)}$  有缺失值, 则忽略第  $j$  行的数据, 即在相关系数的求解过程中不使用此行的数据, 然后根据忽略的行数重新记录这两列的数据, 但次序不变, 依次设为  $\widetilde{y}_1^{(j)} (j=1, 2, \dots, 290-k)$  和  $\widetilde{y}_2^{(j)} (j=1, 2, \dots, 290-k)$ , 其中  $k$  为要忽略的行的数目。则对二维子样  $(\widetilde{y}_1^{(j)}, \widetilde{y}_2^{(j)}) (j=1, 2, \dots, 290-k)$ , 有

$$\text{均值: } \overline{y_1} = \frac{1}{290-k} \sum_{i=1}^{290-k} \widetilde{y}_1^{(i)}, \quad \overline{y_2} = \frac{1}{290-k} \sum_{i=1}^{290-k} \widetilde{y}_2^{(i)}$$

$$\text{方差: } S_{y_1}^2 = \frac{1}{290-k} \sum_{i=1}^{290-k} (\widetilde{y}_1^{(i)} - \overline{y_1})^2, \quad S_{y_2}^2 = \frac{1}{290-k} \sum_{i=1}^{290-k} (\widetilde{y}_2^{(i)} - \overline{y_2})^2$$

$$\text{协方差: } S_{y_1 y_2} = \frac{1}{290-k} \sum_{i=1}^{290-k} (\widetilde{y}_1^{(i)} - \overline{y_1})(\widetilde{y}_2^{(i)} - \overline{y_2})$$

因此, 子样  $(y_1^{(j)}, y_2^{(j)}) (j=1, 2, \dots, 290)$  的相关系数为:

$$R_{y_1 y_2} = \frac{S_{y_1 y_2}}{S_{y_1} \cdot S_{y_2}}$$

代入附件一中的数据, 即可求得冻融劈裂强度比(TSR(%))指标和浸水马歇尔稳定度比(S0(%))指标间的相关系数, 对于表中的缺损值, 我们采用直接忽略的方法, 这一点在 1.1(对附件中数据的分析)一节中已经做了详细的说明, 这里不再赘述。同理可以求得四个指标中任意两个指标间的相关系数(见下表)。

表 5.1

	冻融劈裂强度比	浸水马歇尔稳定度比	车 辙	弯拉应变
冻融劈裂强度比	1	0.30189	0.15289	0.10122
浸水马歇尔稳定度比	0.30189	1	0.3076	0.2039
车 辙	0.15289	0.3076	1	0.30846
弯拉应变	0.10122	0.2039	0.30846	1

由表 5.1 可以看出: 任意两个指标间的相关系数都较小, 但均不为 0, 即任意两个指标的数量之间存在一定的联系, 但是联系都不太明显。

### 5.1.2 任意三个指标间的数量关系

在分析三个指标间的数量关系时, 可以采用多元回归模型, 这里我们采用多元线性回归模型进行求解。模型基本思路如下: 任取四个指标中的一个指标作为

因变量，在其余三个指标中任取两个指标作为自变量，进而建立我们需要的多元线性回归模型，显然对每个因变量需要建立三个模型，以判断其与其它任意两个指标的数量关系，当因变量取遍所有的指标时即可得到每个指标与其它任意两个指标的关系。

设  $y_i^{(j)} (i=1,2,3,4; j=1,2,\dots,290)$  依次表示附件一中冻融劈裂强度比指标、浸水马歇尔稳定度比指标、车辙指标和弯拉应变指标在第  $j$  行的取值。

首先对附件一中的数据进行初步处理：首先去除一些异常值(见 1.1 中对异常值的分析)，对剩余的数据，若三个指标中有一个的第  $j$  行数据为缺失值，则三个指标均忽略第  $j$  行的数据，即在模型求解过程中不使用此行的数据，然后根据忽略的行数重新记录这两列的数据，但次序不变，依次设为  $\widetilde{y}_i^{(j)} (i=1,2,3,4; j=1,2,\dots,290-k)$ ，其中  $k$  为要忽略的行的数目。

取  $\widetilde{y}_i^{(j)} (i=1,2,3,4; j=1,2,\dots,290-k)$  作为因变量， $\widetilde{y}_l^{(j)} (l \neq i; j=1,2,\dots,290-k)$  和  $\widetilde{y}_m^{(j)} (m \neq i, m \neq l; j=1,2,\dots,290-k)$  作为自变量，则对这  $290-k$  组数据  $(y_i^{(j)}, y_l^{(j)}, y_m^{(j)})(j=1,2,\dots,290-k)$  有：

$$\begin{cases} y_i^{(1)} = \beta_0 + \beta_1 y_l^{(1)} + \beta_2 y_m^{(1)} + \varepsilon_1 \\ y_i^{(2)} = \beta_0 + \beta_1 y_l^{(2)} + \beta_2 y_m^{(2)} + \varepsilon_2 \\ \dots \quad \dots \quad \dots \\ y_i^{(290-k)} = \beta_0 + \beta_1 y_l^{(290-k)} + \beta_2 y_m^{(290-k)} + \varepsilon_{290-k} \end{cases} \quad (5.1)$$

其中：  $\beta_1, \beta_2$  为待估参数。

模型(5.1)可以表示成矩阵形式：

$$Y = X\beta + \varepsilon \quad (5.2)$$

其中：  $Y = (y_i^{(1)}, \dots, y_i^{(290-k)})^T$ ，  $\beta = (\beta_1, \beta_2)^T$ ，  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_{290-k})^T$

$$X = \begin{pmatrix} 1 & y_l^{(1)} & y_m^{(1)} \\ 1 & y_l^{(2)} & y_m^{(2)} \\ \dots & \dots & \dots \\ 1 & y_l^{(290-k)} & y_m^{(290-k)} \end{pmatrix}$$

方程(5.2)的最小二乘解为：  $\hat{\beta} = (X^T X)^{-1} X^T Y$



$Y$  关于  $X$  的多元线性回归方程为:  $\hat{Y} = X\hat{\beta}$

$$\text{残差平方和: } S_L = \sum_j (y_i^{(j)} - \widehat{y_i^{(j)}})^2 \quad (5.3)$$

由于残差平方和一定程度上反映了回归模型的显著性,但四个指标间数值的数量级差异较大,使得残差平方和不能很好的反映模型的好坏,因此我们通过改进残差平方和得到了一个衡量因变量和自变量间数量关系表达式优劣的一个度量---拟和度(记为  $S_{LX}$ ), 与(5.3)式对应的拟和度定义如下:

$$S_{LX} = \sqrt{\frac{S_L}{290-k}} = \sqrt{\frac{\sum_j (\frac{y_i^{(j)} - \widehat{y_i^{(j)}}}{y_i^{(j)}})^2}{290-k}}$$

显然, 拟和度越小说明残差平方和越小, 即说明因变量指标和自变量指标间数量关系表达式越精确。

下面我们以考虑第 1, 2, 3 个指标间的关系为例对拟和度的应用作一个详细的说明, 根据模型首先建立每个指标与其他两个指标间的回归方程, 并计算相应的拟和度, 最小的拟和度对应的回归方程即为我们要找的三个指标间的关系。

经过计算, 我们发现  $i=2, l=1, m=3$  时的拟和度最小, 为 0.032576, 此时有:

$$\widehat{y_2} = 67.546 + 0.23509y_1 + 0.00057425y_3$$

因此, 第 1, 2, 3 个指标间的关系式可表示为:

$$\widehat{y_2} = 67.546 + 0.23509y_1 + 0.00057425y_3$$

达到此式的拟和度为 0.032576。

同理, 可以得出任意三个指标间的关系表达式及相应的拟和度。

### 5.1.3 四个指标间的数量关系

分析四个指标间数量关系的模型基本思路如下: 任取四个指标的一个指标作为因变量, 其余三个指标作为自变量, 进而建立相应的多元线性回归模型, 当因变量取遍所有的指标时即可得到每个指标与其它三个指标的关系, 最后通过拟和度来衡量因变量和自变量间数量关系表达式的优劣。

由于该模型与 5.1.2 中的模型仅在自变量的数量上存在差异, 因此不再赘述, 仅给出计算结果, 即每个指标与其余三个指标间的拟和度(见下表)。

表 5.2

指标	1-2,3,4	2-1,3,4	3-1,2,4	4-1,2,3
拟和度	0.038	0.0325	0.5879	0.18

注: i-j,k,l 对应的的相关度表示第 i 个指标与其余三个指标间的拟和度

由表 5.2 可以看出：当用第 1、3、4 个指标表示第 2 个指标时，拟和度最小，因此这四个指标间的关系表达式可表示为：

$$\hat{y}_2 = 67.6979 + 0.2366y_1 + 0.0006y_3 - 0.0001y_4$$

相应的拟和度为 0.0325。

## 5.2 问题二的求解

由于题中没有给出沥青、石料对质量指标影响的量值，因此在此问题的求解过程中，我们首先假设沥青、矿石质量对质量指标的影响可以忽略不计，进而建立每个质量指标与前面因素之间的多元回归模型，通过计算得出每个自变量的偏回归平方和，偏回归平方和可以衡量每个自变量在多元线性回归中所起的作用，从而根据每个自变量的偏回归平方和可以找到影响该指标的最重要和比较重要的因素。

设  $x_i^{(j)} (i=1,3,4,\dots,10; j=1,2,\dots,290)$  依次表示附件一中油石比、VV、VMA、VFA、DP、毛体积密度、最大理论密度、%Gmm(最初)和%Gmm(最大)在第  $j$  行的取值； $x_{2,i}^{(j)} (i=1,2,\dots,13; j=1,2,\dots,290)$  表示附件一中通过筛孔百分率一栏中第  $i$  列第  $j$  行的取值。

建立模型前要先对附件一中的数据进行初步处理：首先去除一些异常值(见 1.1 中对异常值的分析)，对剩余的数据，若因变量或自变量有一个缺少第  $j$  行的数据，则因变量和自变量均忽略第  $j$  行的数据，即在模型的建立过程中不使用此行的数据，然后根据忽略的行数重新记录这些数据，但次序不变。  
 $\widetilde{x_i^{(j)}} (i=1,3,4,\dots,10; j=1,2,\dots,290-k)$  和  $\widetilde{x_{2,i}^{(j)}} (i=1,2,\dots,13; j=1,2,\dots,290-k)$  分别代表相应数据在忽略某些缺失数据后重组的数据，其中  $k$  为要忽略的行的数目。

下面建立公路路面质量指标与影响因素间的多元回归模型，考虑到通过筛孔百分率这一因素的特殊性，我们分两种情况进行了建模：一种是取其中一列作为该因素的代表进行回归；另一种是将该因素所有列都参与回归分析。

### 5.2.1 通过筛孔百分率因素中的每一列都参与回归分析

在这种情况下，由于筛孔百分率因素中的前四列数据相差不大而且缺损严重，因此在建立模型时，我们舍弃了这四列的数据，下面给出我们的模型：

取定一个公路质量指标  $y_i^{(j)} (i=1,2,3,4)$ ，对  $290-k$  组数据

$(y_i^{(j)}, x_1^{(j)}, x_{2,5}^{(j)}, x_{2,6}^{(j)}, \dots, x_{2,10}^{(j)}, x_3^{(j)}, \dots, x_{13}^{(j)})(j=1, 2, \dots, 290-k)$  有:

$$\begin{cases} y_i^{(1)} = \beta_0 + \beta_1 x_1^{(1)} + \beta_{2,5} x_{2,5}^{(1)} + \dots + \beta_{2,13} x_{2,13}^{(1)} + \dots + \beta_{10} x_{10}^{(1)} + \varepsilon_1 \\ y_i^{(2)} = \beta_0 + \beta_1 x_1^{(2)} + \beta_2 x_{2,5}^{(2)} + \dots + \beta_{2,13} x_{2,13}^{(2)} + \dots + \beta_{10} x_{10}^{(2)} + \varepsilon_2 \\ \dots \quad \dots \quad \dots \\ y_i^{(290-k)} = \beta_0 + \beta_1 x_1^{(290-k)} + \beta_2 x_{2,k}^{(290-k)} + \dots + \beta_{2,13} x_{2,13}^{(290-k)} + \dots + \beta_{10} x_{10}^{(290-k)} + \varepsilon_{290-k} \end{cases}$$

对上式利用最小二乘法进行求解即可得所求的多元线性回归方程,不妨设为  $\hat{Y}_i = X\hat{\beta}$ , 下面讨论每个自变量在多元线性回归中所起的作用。

由回归方程易得回归平方和:  $U = \sum_j (\overline{y_i^{(j)}} - \widehat{y_i^{(j)}})^2$ , 因为回归平方和  $U$  是所有自变量对  $y_i$  变化的总的的影响,所以自变量在总的回归中所起的作用可根据它在  $U$  中的影响大小来衡量,即去掉一个变量后,看回归平方和减少的是否明显,减少的数值越大,说明该变量在回归中所起的作用越大;反之,该变量的作用就越小。我们把取消一个自变量后回归平方和减少的数值称为  $y_i$  对这个自变量的偏回归平方和,记作  $P_i$ , 即

$$P_i = U - U'$$

其中,  $U'$  是去掉一个自变量所引起线性变化的回归平方和。

一般偏回归平方和可按下式计算:  $P_i = \frac{b_i^2}{c_{ii}}$

其中,  $c_{ii}$  是自变量的正规方程组系数矩阵的逆矩阵  $C = (c_{ij})$  中的元素;  $b_i$  是回归方程的回归系数。

下面给出每个自变量对应不同质量指标时的  $P_i$  值(见下表):

表 5.3

	$x_1$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$x_8$	$x_9$	$x_{10}$
TSR	48564	0.9193	123.19	51.201	0.0045 768	0.09412 5	10.93	16.299	1.4064
SO	17733	7.9779	8.9331	10.448	0.0417 05	3.5868	2.139	15881	6.8353
车辙	2.6103 e+008	1.6666 e+006	2.4485 e+006	8.8201 e+007	9179.5	1.7489e +006	1.5478e +006	4.7442e +008	3.2199e +005

弯拉应变	2.3926 e+008	7415.3	1.9758 e+005	1.7798 e+005	477.21	7205.2	1.2093e +005	1.6626e +007	17685
	$x_{2,5}$	$x_{2,6}$	$x_{2,7}$	$x_{2,8}$	$x_{2,9}$	$x_{2,10}$	$x_{2,11}$	$x_{2,12}$	$x_{2,13}$
TSR	304.41	4.1602	5.7793	10.778	0.3247 7	0.04238 3	0.12542	0.10177	0.21627
SO	11852	0.7492 4	8.7514	4.7122	2.9151	1.2136	0.94553	4.412	13.084
车辙	1.1078 e+009	2.394e +005	7.3908 e+005	1.1703 e+005	4.8413 e+005	7347.4	6.2208e +005	1.9628e +006	2.4236e +006
弯拉应变	1.0266 e+007	79548	44361	13.954	3405.9	26322	63885	1.0068e +005	7934.2

由表 5.3 可以看出：

对 TSR 影响最大的因素为第一列，即油石比，以下依次孔径为 13.2(由于忽略了筛孔百分率的前四列，此即为最大孔径)时的集料通过率、VMA；

对 SO 影响最大的因素为油石比，以下依次为%Gmm(最初)、孔径为 13.2 时集料的通过率；

对车辙影响最大的因素为孔径为 13.2 时集料的通过率，以下依次为%Gmm(最初)、油石比；

对弯拉应变影响最大的因素为油石比，以下依次为%Gmm(最初)、孔径为 13.2 时集料的通过率。

综上，影响路面质量最大的因素为油石比，比较重要的因素有孔径为 13.2(由于忽略了筛孔百分率的前四列，此即为最大孔径)时的集料通过率、%Gmm(最初)。

### 5.2.2 取通过筛孔百分率因素中的某一列作代表进行回归分析

在这种情况下，为了结果的精确性，我们对每一列都进行了分析，即每一列都作代表分别进行了回归分析。这里不妨设第 m 列作为代表。

取定一个公路质量指标  $y_i^{(j)} (i=1,2,3,4)$ ，对  $290-k$  组数据  $(y_i^{(j)}, x_1^{(j)}, x_{2,k}^{(j)}, \dots, x_{13}^{(j)}) (j=1,2,\dots,290-k)$  有：

$$\begin{cases} y_i^{(1)} = \beta_0 + \beta_1 y_1^{(1)} + \beta_2 y_{2,m}^{(1)} + \beta_3 y_3^{(1)} + \dots + \beta_{10} y_{10}^{(1)} + \varepsilon_1 \\ y_i^{(2)} = \beta_0 + \beta_1 x_1^{(2)} + \beta_2 x_{2,m}^{(2)} + \beta_3 y_3^{(2)} + \dots + \beta_{10} x_{10}^{(2)} + \varepsilon_2 \\ \dots \quad \dots \quad \dots \\ y_i^{(290-k)} = \beta_0 + \beta_1 x_1^{(290-k)} + \beta_2 x_{2,m}^{(290-k)} + \beta_3 y_3^{(290-k)} + \dots + \beta_{10} x_{10}^{(290-k)} + \varepsilon_{290-k} \end{cases}$$

考虑到通过筛孔百分率因素一栏中列数很多，这里我们只选用国际上通用的粗集料和细集料的分界线 2.36 一列进行求解，于是上述方程可化为：

$$\begin{cases} y_i^{(1)} = \beta_0 + \beta_1 y_1^{(1)} + \beta_2 y_{2,8}^{(1)} + \beta_3 y_3^{(1)} + \cdots + \beta_{10} y_{10}^{(1)} + \varepsilon_1 \\ y_i^{(2)} = \beta_0 + \beta_1 x_1^{(2)} + \beta_2 x_{2,8}^{(2)} + \beta_3 y_3^{(2)} + \cdots + \beta_{10} x_{10}^{(2)} + \varepsilon_2 \\ \dots \quad \dots \quad \dots \\ y_i^{(290-k)} = \beta_0 + \beta_1 x_1^{(290-k)} + \beta_2 x_{2,8}^{(290-k)} + \beta_3 y_3^{(290-k)} + \cdots + \beta_{10} x_{10}^{(290-k)} + \varepsilon_{290-k} \end{cases}$$

根据 5.2.1 中的分析，要衡量因素对质量指标的影响，只需比较自变量的偏回归平方和  $P_i$ ，下面给出每个自变量对应不同质量指标时的  $P_i$  值(见下表)：

表 5.4

	$x_1$	$x_{2,8}$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$x_8$	$x_9$	$x_{10}$
TS	57062	532.84	0.035	208.31	39.092	0.016	0.3037	15.938	307.94	0.56496
R			278			991				
S	35406	11952	0.061	54.184	8.5692	0.361	0.0081	0.04909	4734.5	12.599
O			03			63	833	2		
车	1.6367e	2.974e+	9455.	2.9405e	5.2218e	28278	76877	6663.4	1.4334e	1.0632e
辙	+007	008	3	+006	+007				+009	+006
弯	4.6786e	4.4393e	27.27	1.8627e	4.5915e	3499.	2623.9	2.594e	9.2546e	30244
拉	+008	+006	3	+006	+005	1		+005	+006	
应										
变										

由表 5.4 可以看出：

对 TSR 影响最大的因素为油石比，以下依次为通过筛孔百分率、%Gmm(最初)；

对 SO 影响最大的因素为%Gmm(最初)，以下依次为油石比、通过筛孔百分率；

对车辙影响最大的因素为%Gmm(最初)，以下依次为通过筛孔百分率、VFA 油石比；

对弯拉应变影响最大的因素为油石比，以下依次为%Gmm(最初)、通过筛孔百分率。

### 5.2.3 提高公路路面质量的方案

这里不妨采用 5.2.1 中的模型。5.2.1 中已经给出了对每个公路质量指标  $y_i^{(j)} (i=1,2,3,4)$  影响最大和比较大的因素，我们发现影响最大的两个因素的偏回

归平方和一般都比其它因素的偏回归平方和大很多, 因此我们取对每个质量指标影响最大的前两个因素建立回归模型, 经计算可得相应的线性回归方程:

$$\begin{cases} \hat{y}_1 = 86.113 + 0.67729x_1 - 0.051243x_{2.5} \\ \hat{y}_2 = 174.52 + 0.85282x_1 - 1.0356x_9 \\ \hat{y}_3 = 19377 + 32.684x_{2.5} - 216.79x_9 \\ \hat{y}_4 = -2593.2 + 0.15791x_1 + 65.135x_9 \end{cases}$$

由上式可知, 每个质量指标主要由某几个因素决定, 因此根据此模型, 我们可以制定提高高速公路路面质量的方案: 即独立考虑每个质量指标, 根据上式给出的关系, 易得要使该指标变大的措施。如要使 TSR 变大, 只需要使油石比变大或使通过孔径为 13.2 集料百分比变小。

### 5.3 问题三的求解

我们先从只有一种颗粒的体系入手, 通过理论计算得到了最大堆积率, 再考虑体系包含多种颗粒的情况。由于包含大量多尺度颗粒的体系, 其级配组成与堆积率的关系还不清楚, 所以我们通过查阅资料, 决定先考虑几种特殊情况, 比如细颗粒的尺寸和数量刚好能填满粗颗粒的空隙。然后再对一般情况应用已有的结果进行讨论, 最后得出了集料级配组成与颗粒堆积密度的关系。

#### 5.3.1 单尺度颗粒体系的堆积

单尺度颗粒体系的最紧密堆积方式如图 5.1 所示:

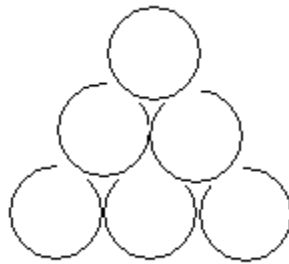


图 5.1 等径颗粒最紧密堆积方式二维视图

晶体学上已经证明, 这种堆积方式是堆积率最大(空隙率最小)的一种堆积方式, 并且这种堆积方式的周期性单元是面心结构, 如图 5.2 所示:

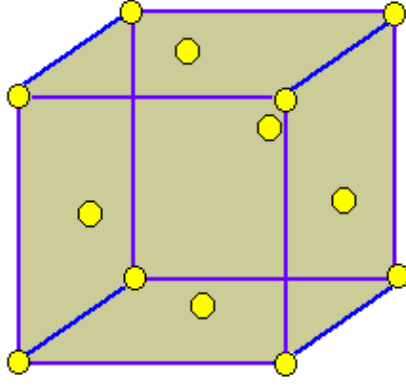


图 5.2 面心结构

设每个球的半径为  $r$ ，立方体边长为 1，则  $r = \frac{\sqrt{2}}{4}$ ；八个顶角上的球在立方体内的体积占球自身体积的  $\frac{1}{8}$ ，六个面上的球在立方体内的体积占球自身体积的  $\frac{1}{2}$ ，故球体体积对立方体体积的占有率为

$$8 \times \frac{1}{8} \times \frac{4}{3} \pi \left( \frac{\sqrt{2}}{4} \right)^3 + 6 \times \frac{1}{2} \times \frac{4}{3} \pi \left( \frac{\sqrt{2}}{4} \right)^3 = 74.05\%$$

结论：单尺度粒径颗粒的堆积率与颗粒大小无关，而与颗粒堆积方式有关，只要按上述密排结构堆积，不管颗粒多大或多小，堆积率都为 74.05%。

### 5.3.1 对理想球状颗粒的堆积计算

理想球状颗粒是指颗粒都是球体，表面光滑，相互之间没相对移动的干扰。本文讨论的集料颗粒堆积密度是指集料中固体物质体积占整个集料总体积（包括空隙体积量）的份数，以小数表示，用下式计算：

$$D = 1 - \frac{\varepsilon}{100}$$

式中： $D$ ：堆积密度

$\varepsilon$ ：空隙体积含量，%

在由多粒级组分组成的集料中，若颗粒间隙都刚好能够被下一级小颗粒填满，便能达到最大的堆积密度，随着填充颗粒级数的增加，堆积密度会接近  $D = 1$  的极限值。若各粒级的原始堆积密度  $D_0$  都相等，则混合颗粒所能达到的堆积密度与粒级份数之间的关系为：

$$D_{\max}(n) = 1 - \left( \frac{\varepsilon_0}{100} \right)^n = 1 - (1 - D_0)^n \quad (1)$$

式中： $D_{\max}(n)$ ： $n$ 级组分混合集料堆积密度；

$D_0$ : 各粒级组分的堆积密度;

$\varepsilon_0$ : 各粒级组分的空隙体积含量, %

$n$ : 粒级组分数。

图5.3是用方程(1)计算得出的不同原始堆积密度的颗粒级数与理论上所能达到的最大堆积密度间的关系。从中可以看出, 原始堆积密度越大, 为达到同样的混合集料堆积密度所需的粒级越少。

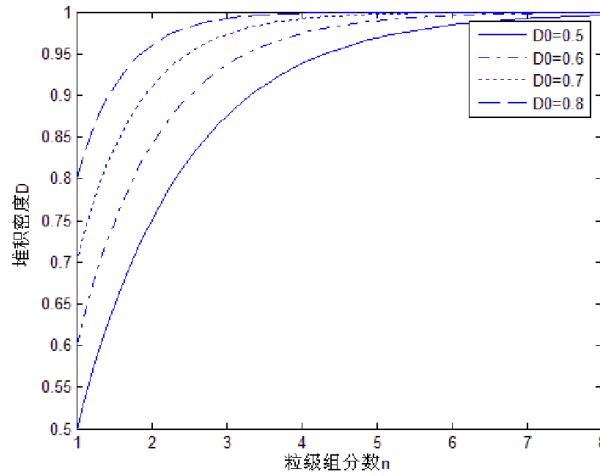


图5.3 不同原始堆积密度的多粒级混合粉体  
中粒级组分数与堆积密度D 的关系

达到最大堆积密度时粒径为  $X_i$  的粒级的含量即为最佳含量  $S_{i,opt}$ 。此最佳含量是以相关粒级  $i$  的固体物质含量与集料物质总量之比来表示。因为固体物质含量与堆积密度成比例, 在各粒级原始堆积密度  $D_0$  相等的前提下, 各粒级组分的最优含量  $S_{i,opt}$  可由方程(1)得出的最大堆积密度  $D_{max}$  进行计算, 这里各级填充组分的固体物质含量取决于  $1 \sim (i-1)$  粒级粗颗粒间的空隙含量。

$$S_{i,opt} = \frac{i\text{粒级的固体物质含量}}{\text{集料固体物质总量}} = \frac{\left(\frac{\varepsilon}{100}\right)^{i-1} \cdot D_0}{D_{max}(n)} = \frac{(1-D_0)^{i-1} \cdot D_0}{D_{max}(n)} \quad (2)$$

$$\sum_{i=1}^n S_i = 1$$

$S_{i,opt}$  为  $n$  粒级混合集料达到最大堆积密度时第  $i$  粒级的固体物质含量份数 ( $S_1$  是最大颗粒组分数,  $S_n$  是最小颗粒组分数。  $i+1$  级固体颗粒是  $i$  级固体颗粒的填



充组分)。

图5.4为由方程(2)计算得出的达到最大堆积密度时粒径为 $X_i$ 粒级的最佳体积含量分数 $S_{i,opt}$ 。图中4个混合颗粒都由8个粒度组成,只是原始堆积密度不同。图2显示,填充细集料的需要量很少,其体积含量分数随粒径减小依次呈指数关系下降,初始密度越大,下降的程度越大。

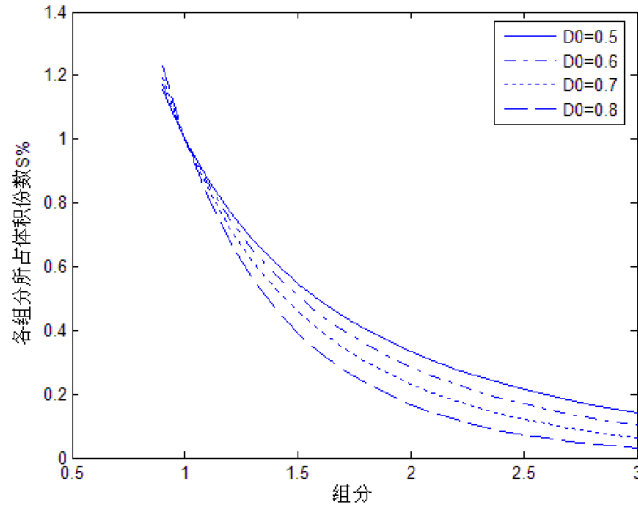


图5.4 不同初始堆积密度D0的8组分混合粉体

各粒度组分 $X_i$ 的最佳体积含量份数 $S_{i,opt}$

### 5.3.3 实际多尺度颗粒的堆积

实际中,粗细颗粒的搭配不可能刚好使得细颗粒的尺寸完全符合上述理想状态的尺寸,这时便会将粗颗粒架开。另外,对集料总体积的影响不仅是粗细颗粒的体积含量,更主要的是它们的含量比例。若细颗粒含量太少,在粗颗粒之间必然留有未被填满的空隙;反之,所有粗颗粒的空隙都被填满之后,剩余的细颗粒会将粗颗粒挤开。

计算步骤:

1. 确定单一粒径颗粒空隙体积份数 $K$ 。对颗粒粒度分布进行分级,每一粒度都作为单一粒径颗粒看待。参数 $K$ 即为每一粒度的空隙体积含量与固体物质体积含量之比。而等粒径球体颗粒的理论堆积密度视其排列方式而定。这一点在前面做过讨论。

$$K = \frac{\varepsilon}{100 - \varepsilon}$$

式中:  $K$ : 单一粒径颗粒空隙体积份数;

$\varepsilon$ : 空隙体积含量, %;

2. 将各粒级按粒径大小顺序排列, 计算多粒级混合颗粒的空隙体积份数 $h$ 。计算基本方程是以粗细两组分组成的混合颗粒为基础, 首先计算在两种特殊情况下的堆积密度的变化。

第一种类型: 细颗粒刚好能填满粗颗粒间隙, 混合颗粒的空隙含量等于粗颗粒空隙体积含量减去细颗粒固体物质体积含量, 此时粗颗粒作为“基础颗粒”考虑, 空隙体积份数的计算为:

$$h_1 = \frac{\text{粗颗粒空隙体积}-\text{细颗粒固体体积}}{\text{粗颗粒固体体积}+\text{细颗粒固体体积}}$$

设粗颗粒固体体积为1, 则有如下关系:

$$h_1 = \frac{K_2 - \frac{S_1}{1-S_1}}{1 + \frac{S_1}{1-S_1}} = K_2 - (K_2 + 1) \cdot S_2$$

其中, 等式右边下标为“1”表示细颗粒的量, 下标为“2”表示粗颗粒的量。

$h_1$ : 两粒级混合颗粒空隙体积份数;

$S_1$ : 细颗粒固体物质体积含量份数。

第二种类型: 细颗粒份数大于粗颗粒空隙体积含量。此时混合颗粒的空隙体积含量等于细颗粒的空隙体积含量, 所以细颗粒成为“基础颗粒”。空隙体积份数为:

$$h_2 = \frac{\text{细颗粒空隙体积}}{\text{粗颗粒固体体积}+\text{细颗粒固体体积}}$$

设细颗粒固体体积为1, 则细颗粒空隙体积为 $K_1$ , 而粗颗粒固体体积 $V_2$ 与粗颗粒固体体积含量份数 $S_2$ , 于是可得如下关系:

$$h_s = \frac{K_1}{\frac{S_2}{1-S_2} + 1} = K_1 - K_1 \cdot S_2 a_{si}$$

$$h_s = K_0 - \sum_{i=1}^n a_{si} \cdot S_i$$

其中等式右边下标为“1”表示细颗粒的量, 下标为“2”表示粗颗粒的量。

第三种类型: 除以上两种特殊情况外, 常常发生的是由于细颗粒尺寸不合适, 不能填入粗颗粒间隙中, 由于这种“颗粒干扰作用”增大了总体积量, 也提高了

空隙体积份数，计算时将这个因素考虑在系数 $a$ 中。由基础颗粒的空隙体积份数 $K$ 减去其他各级填充颗粒的相应固体体积份数 $S_i$ ，即可得出 $h_s$ 值，同时考虑相应的

的系数 $a_{si}$ ，从而有：
$$h_s = K_0 - \sum_{i=1}^n a_{si} \cdot S_i$$

$h_s$ ：基础颗粒级的空隙体积份数；

$K_0$ ：单一粒径颗粒空隙体积份数；

$S_i$ ： $i$ 粒级所占整个固体体积份数

对每一粒级都分别作一次基础颗粒，计算出相应的 $a_{si}$ 和 $S_i$ ，按顺序汇总列表，找出基础颗粒空隙体积份数最大（ $h_s$ 值）的粒级作为集料堆积密度 $D$ 的计算依据，按下式可得出给定集料级配组成的颗粒堆积密度为：

$$D = \frac{1}{1 + h_{\max}}$$

下面举例计算，设一个由5个粒级组成的球形颗粒混合集料，各粒级的空隙体积份数都为最紧密堆积（四面体型）的空隙体积份数，为 $K_0 = 0.7405$ ，各级粒径分别为0.25、0.5、1.00、2.00和4.00mm。各自占总体积的比例为24%，25%，19%，17%，15%，按照上述方法最后计算得出 $D = 0.91$ 。

## 5.4 问题四的求解

求解该问题主要是解决两个问题，下面分别对这两个问题进行建模求解：

### 5.4.1 判断沥青、碎石质量对高速公路路面性能的影响

显然这是一个判断因素对试验结果影响的问题，考虑到此问题涉及到沥青、碎石质量两个因素，且数据不具备重复的条件，因此我们可以采用双因素无重复试验的方差分析法进行求解。但从题中所给数据不难看出，并不是所有的数据均可用作双因素试验的方差分析法，这就需要在建模前先对数据进行一下处理，找出可以利用的数据进而建立模型。

设沥青为因素 $A$ ，碎石质量为因素 $B$ ，且因素 $A$ 有 $r$ 个水平 $A_1, \dots, A_r$ ，因素 $B$ 有 $s$ 个水平 $B_1, \dots, B_s$ 。由于此问题中的沥青只考虑普通沥青和改性沥青两种类型，因此因素 $A$ 仅有两个水平，即 $r = 2$ ；由于碎石质量由厂家和型号共同决定，因此我们认为厂家和型号均相同时才算因素 $B$ 的一个水平，这样就可以建立起双因

素试验的方差分析法所需要的数据集(见下表)。

表 5.5

因素 $B$ 因素 $A$	$B_1$	$B_2$	$\cdots$	$B_s$
$A_1$	$z_{11}^{(i)}$	$z_{12}^{(i)}$	$\cdots$	$z_{1s}^{(i)}$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
$A_r$	$z_{r1}^{(i)}$	$z_{r2}^{(i)}$	$\cdots$	$z_{rs}^{(i)}$

其中  $z_{lk}^{(i)} (i=1,2,3,4;l=1,2,\cdots,r;k=1,2,\cdots,s)$  表示水平  $A_l$  和水平  $B_k$  在第  $i$  个性能指标处的取值。

由表 5.5 可求得方差分析表如下：

表 5.6

方差来源	平方和	自由度	均 方	F 比
因素 $A$	$S_A$	$r-1$	$\overline{S_A} = \frac{S_A}{r-1}$	$F_A = \overline{S_A} / \overline{S_E}$
因素 $B$	$S_B$	$s-1$	$\overline{S_B} = \frac{S_B}{s-1}$	$F_B = \overline{S_B} / \overline{S_E}$
误 差	$S_E$	$(r-1)(s-1)$	$\overline{S_E} = \frac{S_E}{(r-1)(s-1)}$	
总 和	$S_T$	$rs-1$		

其中：
$$T_{..} = \sum_{l=1}^r \sum_{k=1}^s Z_{lk}^{(i)}, \quad T_{l.} = \sum_{k=1}^s Z_{lk}^{(i)}, (l=1,2,\cdots,r)$$

$$T_{.k} = \sum_{l=1}^r Z_{lk}^{(i)}, (k=1,2,\cdots,s)$$

$$S_T = \sum_{l=1}^r \sum_{k=1}^s (Z_{lk}^{(i)})^2 - \frac{T_{..}^2}{rs}$$

$$S_A = \frac{1}{s} \sum_{l=1}^r T_{l.}^2 - \frac{T_{..}^2}{rs}$$

$$S_B = \frac{1}{r} \sum_{k=1}^s T_{.k}^2 - \frac{T_{..}^2}{rs}$$

$$S_E = S_T - S_A - S_B$$

根据假设检验理论，取显著性水平为  $\alpha$ ，则有：

若  $F_A = \frac{\overline{S_A}}{\overline{S_E}} \geq F_{\alpha}(r-1, (r-1)(s-1))$  成立，则说明因素  $A$  在置信水平  $1-\alpha$  下对试验结果有影响；

若  $F_B = \frac{\overline{S_B}}{\overline{S_E}} \geq F_{\alpha}(s-1, (r-1)(s-1))$  成立，则说明因素  $B$  在置信水平  $1-\alpha$  下对试验结果有影响。

在上述模型中代入我们筛选出的数据集，即可判断出沥青、碎石质量对公路路面质量指标的影响，下面以沥青、碎石质量对 TSR 的影响为例给出我们的结果，经计算可得沥青、碎石质量影响 TSR 的方差分析表(见下表)。

表 5.7 沥青、碎石质量影响 TSR 的方差分析表

方差来源	平方和	自由度	均 方	F 比
碎石质量	230.26	14	16.447	1.0067
沥 青	0.3	1	0.3	0.01836
误 差	228.73	14	16.338	
总 和	459.29	29		

取置信度为 90%，则有

$$F_{\alpha}(r-1,(r-1)(s-1))=2.2806, \quad F_{\alpha}(s-1,(r-1)(s-1))=3.1362,$$

即有  $F_A=1.0067 < 2.2806=F_{\alpha}(r-1,(r-1)(s-1))$ ,  $F_B=0.01836 < 3.1362=F_{\alpha}(s-1,(r-1)(s-1))$

因此，根据假设检验理论，沥青和碎石质量在置信度为 90%下均对 TSR 无显著影响。

同理可得：置信度为 90%时，

- 沥青和碎石质量都对 SO 没有显著影响；
- 沥青对车辙有显著影响，而碎石质量对车辙没有显著影响；
- 沥青对和碎石质量均对弯拉应变有显著影响。

### 5.4.2 判断不同厂家的但型号相同的沥青和不同产地的但类型相同的碎石对高速公路路面性能的影响

这仍然是一个判断因素对试验结果影响的问题，同样可以采用方差分析法。但是通过对数据进行分析我们发现，可用作双因素试验方差分析法的数据几乎没有，因此双因素试验的方差分析法不可行，鉴于此，我们采用了单因素试验方差分析法，即独立的考虑不同厂家但型号相同的沥青和不同产地但类型相同的碎石对公路路面性能的影响。

下面我们以沥青厂作为因素，TSR 取值作为试验结果为例进行讨论，假设通过分析附件一中数据得到沥青厂因素(记为因素 A)有 s 个水平  $A_1, \dots, A_s$ ，且在水平  $A_j(j=1,2,\dots,s)$  下，TSR 有  $n_j$  个值，记为  $z_{1j}, \dots, z_{n_jj}$ 。这样我们就得到方差分析所需的数据表(见下表)。

表 5.8

水平	$B_1$	$B_2$	...	$B_s$
----	-------	-------	-----	-------

观察结果				
	$z_{11}$	$z_{12}$	$\dots$	$z_{1s}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	$z_{n_1 1}$	$z_{n_2 2}$	$\dots$	$z_{n_s s}$

进而可以求得单因素试验方差分析表如下：

表 5.9

方差来源	平方和	自由度	均 方	F 比
因素 $A$	$S_A$	$s-1$	$\overline{S_A} = \frac{S_A}{s-1}$	$F = \overline{S_A} / \overline{S_E}$
误 差	$S_E$	$n-s$	$\overline{S_E} = \frac{S_E}{n-s}$	
总 和	$S_T$	$n-1$		

其中： $T_{..} = \sum_{j=1}^s \sum_{i=1}^{n_j} z_{ij}$ ， $T_{.j} = \sum_{i=1}^{n_j} z_{ij}$ ，( $j=1, 2, \dots, s$ )

$$S_T = \sum_{j=1}^s \sum_{i=1}^{n_j} z_{ij}^2 - \frac{T_{..}^2}{n}$$

$$S_A = \sum_{j=1}^s \frac{T_{.j}^2}{n_j} - \frac{T_{..}^2}{n}$$

$$S_E = S_T - S_A$$

根据假设检验理论，取显著性水平为  $\alpha$ ，则有：

若  $F = \frac{\overline{S_A}}{\overline{S_E}} \geq F_{\alpha}(s-1, n-s)$  成立，则说明因素  $A$  在置信水平  $1-\alpha$  下对试验结果有影响。

下面给出我们的计算结果，首先给出方差分析表(见下表)：

表 5.10 沥青厂影响 TSR 的方差分析表

方差来源	平方和	自由度	均 方	F 比
沥青厂	18.246	3	6.082	0.95153
误 差	57.526	9	6.3918	
总 和	75.772	12		

取置信度为 90%，则有  $F_{\alpha}(s-1, n-s) = 2.8129$ ，即有  $F = 0.95153 < 2.8129 = F_{\alpha}(s-1, n-s)$ ，

因此，根据假设检验理论，沥青厂在置信度为 90% 下对 TSR 无显著影响。

同理可得：置信度为 90% 时，

沥青厂对 SO、车辙、弯拉应变均没有显著影响；

碎石产地对 TSR、SO 无显著影响；

碎石产地对车辙、弯拉应变有显著影响。

## 5.5 问题五的求解

从题目中所给的数据可知，现在测试高速公路路面质量的试验项目都是基于如何延长高速公路路面寿命做的，虽然使用寿命对成本和保障运输安全有极其重要的意义，但我们觉得实际中仅考虑这些是远远不够的，因为公路还有很多其他的重要属性，而这些属性又是由不同的因素决定的，因此通过查阅资料，我们找到了几个对路面质量影响比较大而试验项目中又不是很重视的因素，现介绍如下：

### 1、石料的磨光值、磨耗值、冲击值

题目中的数据仅是对石料的产地、种类作了测试，而石料的某些具体特性并没有得到关注，经过查阅资料，我们发现石料的磨光值对路面的抗滑性有至关重要的影响，磨光值高的石料，在轮胎的长期作用下，能长时间保持其良好的微观构造，而磨耗值和冲击值是评价石料抵抗摩擦、撞击和剪切等综合作用的指标，因此这三个指标都非常重要，应该给与精确的试验。

### 2、路面厚度

题目中并未对路面的厚度进行测验，而路面厚度不仅对路面的使用寿命有重要影响，而且对路面降低噪声也有重要的作用，因此应该对该指标进行精确的试验。

当然，影响路面质量的因素还有很多，这里我们仅以上面两个因素为例进行分析，进而完善试验项目。

## 6 结果分析

### 6.1 问题一结果分析

问题一我们通过拟和度给出了四个指标间的关系表达式，但我们发现在计算回归方程时误差有点大，拟和度就是我们根据残差平方和引入的概念，一定程度上反映了误差的大小，因此我们觉得四个指标间的关系并不是很明显。下面给出四个指标关系表达式的误差分布图：

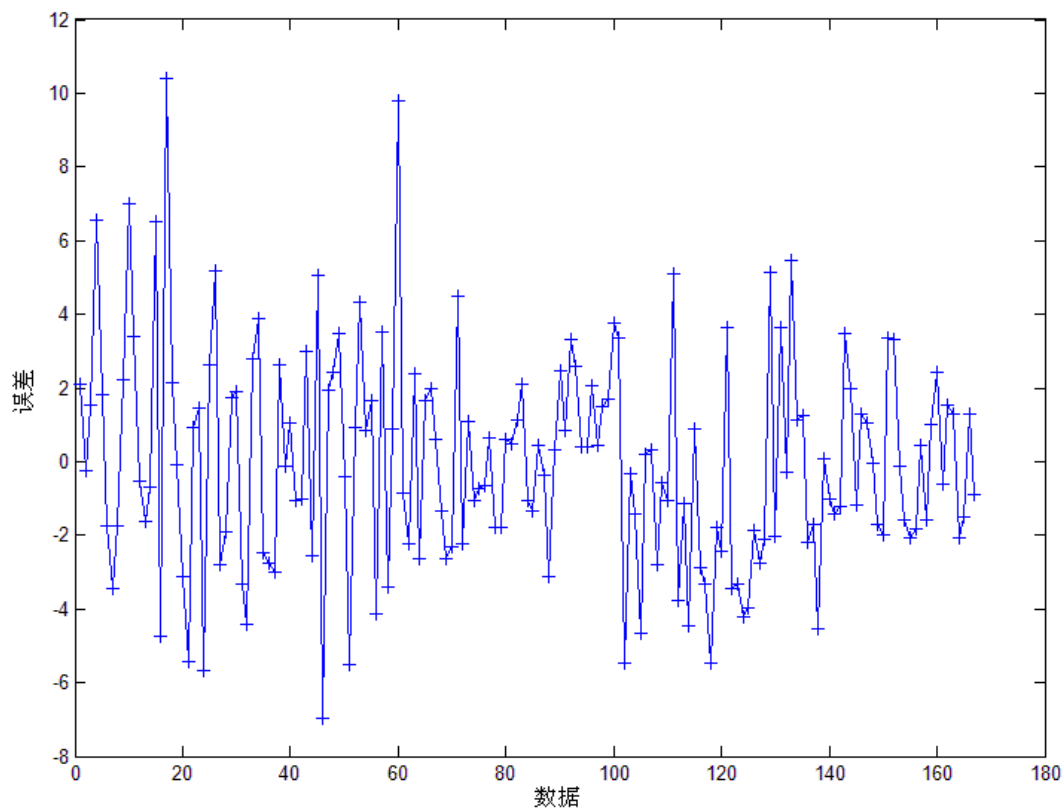


图 6.1

即过计算，可得均方根误差为 2.953，而第 2 个指标的数量级为 100 左右，即此表达式的误差达到了 3%左右，因此误差还是比较大的，所以我们认为这四个指标间的关系不是很明显。

## 6.2 问题二结果分析

从 5.2.1 和 5.2.2 两个模型的结果来看，影响公路路面质量的最重要的因素比较固定，当然 5.2.2 模型中我们还取了其他孔径进行了回归分析，结果与 5.2.2 中的结果很相近，这就在理论上证明了结果的正确性。

从结果中我们知道影响公路路面质量最重要的几个因素为油石比，通过筛孔百分率和%Gmm(最初)。其中石油比和通过筛孔百分率表示混合料的质量，这里的质量是衡量优劣的一种度量，而不是物理意义上的重量。显然路面质量要受混合料质量的制约，而混合料最重要的指标也就是石油比和石料的大小体积比率，因此油石比和通过筛孔百分率对公路路面质量的影响至关重要；对于%Gmm(最初)，我们理解试件中的空隙最初就来源于此，因为内部空隙一旦存在，就几乎不可能在消失，虽然空隙可以随着碾压而不断减少，但一般不会消失，因此如果最初状态时试件内空隙较小，那么空隙会一直小下去，因此%Gmm(最初)也是影响公路路面质量的一个重要因素。

对于其他因素，其  $P_i$  值与上述三个因素的  $P_i$  值相差太大，同时从这些因素的



定义来看，它们大多可以由其他因素表示，这样它们的作用最终还是归结到别的因素上，因此它们对公路路面质量的影响不是很大。

## 6.2 问题四结果分析

问题四分别利用双因素无重复试验的方差分析法和单因素试验方差分析法对问题进行了求解，显然这两个方法都要建立在对数据的选取上，因此得出有影响的结论则说明有影响，如果得出没影响的结论则可能是由于数据取得不合理或者数据量太少，即如果得出没影响的结论并不能肯定没有影响。

## 7 模型的进一步讨论

### 7.1 问题一中非线性回归模型

在问题一的求解过程中，我们仅讨论了指标间的线性关系，结果表明指标间的线性关系是很弱的，因此我们考虑它们之间是否存在某种非线性关系，为此我们还建立了它们之间的非线性回归模型，因为如果回归多项式次数很高的话那么参数将会非常多，而题中给的数据较少，无法满足多次多项式回归模型的求解需求，因此我们主要研究了二次回归和三角函数回归模型。

下面给出我们的一组计算结果，在考虑第 1, 2, 3 个指标间的关系时，采用二次多项式回归模型，经过计算发现在用第 1, 3 个指标表示第 2 个指标时拟和度达到最小，此时这三个指标间的关系式可表示为：

$$\hat{y}_2 = -120.85 + 4.5612y_1 + 0.001598y_3 - 0.025093y_1^2 + 1.6832 \times 10^{-6}y_1y_3 - 1.3099 \times 10^{-7}y_3^2$$

对应的拟和度为：0.031596。

为了能直观的比较线性回归模型与二次回归模型，我们将线性回归模型的结果列出：

$$\text{表达式: } \hat{y}_2 = 67.546 + 0.23509y_1 + 0.00057425y_3$$

拟和度：0.032576。

### 7.2 问题二中考虑沥青和石料影响的模型

在问题二中我们假设不考虑沥青和石料的影响，但是这两个因素的影响是客观存在的，而且问题四中还让我们单独讨论该问题，可见直接忽略这两个因素是不太合理的。

为了能更准确的找到影响质量指标的最重要和比较重要的因素，我们可以先排除沥青和石料的影响，在附件一中找沥青和石料特征均一致的数据，然后考虑有差异的因素的差值与质量指标差值的关系，进而可以找到影响质量指标的最重要的因素，由于屏蔽了沥青和石料的影响，这个结果在理论上应该比忽略沥青和石料所得的结果精确，但满足此条件的数据量太小，这又会大大降低结果的精确

度，尽管如此，此模型可以作为问题二求解模型的一种补充和完善。

## 8 模型优缺点

- 1、模型具有广泛性，可以应用到其他处理数据关系的问题中；
- 2、模型具有较强的理论依据；
- 3、模型得到的大多数结论与其物理意义相符合；
- 4、由于缺失数据较多且不规律，导致模型误差较大；

## 9 参考文献

- [1] 王正明，易东云，测量数据建模与参数估计，国防科技大学出版社，1996
- [2] 孙德栋等，沥青路面设计与施工技术，黄河水利出版社，2003
- [3] 吴翊，吴孟达等，数学建模的理论与实践，国防科技大学出版社，1999
- [4] 梁晋文等，误差理论与数据处理，中国计量出版社，2001
- [5] 盛骤等，概率论与数理统计，高等教育出版社，2001
- [6] 梁之舜等，概率论与数理统计，高等教育出版社，1998
- [7] 丁振良等，误差理论与数据处理，哈尔滨工业大学出版社，2002
- [8] 张志涌等，精通 MATLAB 6.5 版，北京航空航天大学出版社，2003
- [9] 杨青云，数据处理方法，冶金工业出版社，1993
- [10] 魏娜等，紧密堆积水泥浆体系的堆积率计算，石油钻探技术，2007，35(4)
- [11] 黄柏宗，紧密堆积理论的微观机理及模型设计，石油钻探技术，2007，35(1)