

一、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。我国高速公路的设计者和建设者为延长高速公路路面寿命做了大量的试验和调查研究，获得了一批宝贵的数据。

由于我国建设高速公路的历史不长，对高速公路路面寿命的客观规律的了解还不太全面、深入（例如不同地区的高速公路路面对抗水损害性能、高温性能、低温性能的要求是否可能不同？工程技术人员对一些现象的解释还不令人十分信服等等）。因此需要对现有的试验数据进行研究分析，从中寻找提高高速公路路面质量的改进方案以延长高速公路路面的寿命。

给出一批目前被认为对高速公路路面的质量有较大影响或有影响的因素的测试数据，另外关于高速公路路面质量的四个试验指标也给出了测试数据。

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？如果有数量关系，请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你们认为影响高速公路路面质量的最重要的和比较重要的因素之间比较精确的数学模型，你们选择这些因素的理由是什么？根据模型你们认为采用什么样的方案可以提高高速公路路面的质量。

三．若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，你们有何见解？

四．沥青（在数据表格中混合料类型中含 25 是普通沥青，是其他数字如 13、16 或 20 的是改性沥青）、碎石质量对高速公路路面抗水损害性能、高温性能、低温性能究竟有没有影响，你们的观点是什么？根据是什么？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、低温性能究竟有没有影响，你们的观点是什么？根据又是什么？

五．根据你们对数据分析的结果，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？你们对高速公路建设部门有什么建议？

二、问题分析

问题一：要检查四个指标之间的数量关系可以从相关或回归下手，本题的四个检验指标在实际中为检验路面的不同方面，它们平行存在，因此回归方法不适用，重点在相关性。考察少数指标的相关性首先应当观察指标之间的散点图，但四个指标在两两交互散点图中分布混乱，没有线形或常见非线性曲线关系。计算四个指标之间的 pearson 相关系数和 spearman 相关系数都小于 0.5，没有明显的相关性。

而灰色关联分析体现一种模糊关联的思想，它衡量了两个序列的整体相似性[1]。首先对四个检验路面质量指标的序列分别做一次累减，然后在累减后的四个指标序列中任取两列，计算这两列的所有局部关联系数，通过折线段上局部关联系数均值的计算再取绝对值，就可以克服没有明显线性关系和常见曲线关系而导致貌似无相关的弱点，反映出复杂数据的相关性。

问题二：要找出影响高速公路路面质量的最重要和比较重要的因素，就应该从检验高速公路质量的指标入手，考察公路质量在接受质量检验时，运用逐步回归法就能够找出对质量检验指标起显著性作用的质量指标，即是重要的指标，在这些表现在显著的重要指标中，能够综合解释多个公路质量检验指标和重点解释

某个公路质量检验指标的公路质量指标为最重要指标，其次为较重要指标。逐步回归中的偏相关系数，就可以精确反映它们之间的关系。

问题三：以最小空隙率作为联结路面压实度上界和级配组成的桥梁，把通过 4.75mm 直径筛孔的细集料形状看成是规格不同的球体，并用两个大小不同的球体简化模型，得出了当两球体积相差很大时的最小空隙率和最大压实度上界，且此时的空隙率不随小球的直径的改变而变化。

问题四：首先根据专家建议认为区分沥青的指标是改性沥青与否，碎石质量是由岩性来衡量。要知道沥青、碎石质量对高速公路路面抗水损害性能、高温性能、低温性能究竟有没有影响，就应该比较改性沥青与普通沥青在这些检验指标上的区别，比较石灰岩、玄武岩和其他岩性在这几项检验指标上的区别。应当使用方差分析，为了得到精确的模型，使用 GLM 广义方差分析来进行方差分析。可得到期间是否有差别，以及差别的具体大小。

要考虑不同厂家、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、低温性能究竟有没有影响，就应当采用两因素交互作用的方差分析。

问题五：根据前面的分析结果，综合给出政策建议。

三、基本假设

1. 对于所有数值型缺失数据，通过 Litter' s MCAR 检验，认为是随机缺失^[1]，采用 EM 缺失值算法填补；
2. 随机误差服从古典假设；
3. 通过筛孔 4.75 以上的是粗料，4.75-0.075 之间是细料，0.075 以下视为矿粉；
4. 混合料类型中含 25 的一律视为普通沥青，其余混合料类型为改性性沥青；
5. 沥青只分为普通沥青和改性沥青的区别；
6. 碎石质量是按照岩性区分的；
7. 碎石形状都视为球体。

四、符号说明

- t_i ：高速公路路面质量第 i 各性能指标的关联度序列
 x_0 ：待求关联度序列 0
 x_1 ：待求关联度序列 1
 n ：元素个数
 x_2 ： x_0 的一阶差分序列

x_3 : x_1 的一阶差分序列

M: 折线段数

ξ_i : 每段折线上的关联系数

τ_{ij} ($1 \leq i \neq j \leq 4$): 高速公路路面指标序列 t_i 和 t_j 的关联度

Z_i ($i=1, 2, \dots, 10$): 影响高速公路路面质量的因素指标

Y_i ($i=1, 2, 3, 4$): 高速公路路面质量的检验指标

r_{ij} : 相关系数

V_i : 方差贡献

五、问题求解

5.0 数据处理

1. 异常数据处理

对每个指标进行描述性观察, 观察其最大值、最小值、均值和散点图等, 发现拉弯应变的第 226 个观测数据为 31450, 是均值的 10 倍左右。因此认为是数据录入时小数点位置错误, 故将其值缩小 10 倍。

2. 缺失值处理

《2007 年 C 题 附件 1 配合比报告统计》中有大量的缺失指, 若果全部删除可能损失大量的信息, 因此应当设法填补缺失值。

数据的缺失可以分为完全随机缺失、随机缺失和非随机缺失。经过 Litter ‘MCAR 检验, p 值小于 0.05, 认为《2007 年 C 题 附件 1 配合比报告统计》中的缺失值情况属于随机缺失, 即有缺失值的变量其缺失情况的发生与数据集中其他无缺失值的变量有关。这种情况下, 缺失值不仅会导致损失信息, 更可能导致分析结论发生偏差。针对随机缺失, 我们运用 SPSS 软件的 Missing Value Analysis 菜单用 EM 算法估计出连续型数值变量的缺失值。

EM (Expectation-Maximization, 期望最大化) 迭代方法主要用于不完整数据下的贝叶斯网络参数学习^[2]。在进行贝叶斯网络参数学习过程中, 数据集包括变量集和训练样本集, 每个数据集是一组变量值, 不同的数据集可能包含不同的变量, 这就意味着数据集可能是不完整的, 其中有的可能并不包含所有变量的值。当训练样本集完整时可以采用 MAP 或 MLE 进行参数估计, 但是当样本数据不完整时, 如当父节点集的样本数目为 0 时, MLE 算法就会产生溢出问题, 从而无法计算。EM 算法是一种从不完整数据集中计算 MLE 估计和 MAP 的通用算法, Lauritzen 已经提出 EM 算法可以应用于数据缺失情况下的贝叶斯网络的参数学习。EM 算法的主要思想是给出一个初始估计参数值, 然后不断地去修正它, 使得 MLE 或 MAP 的值为最大, 即最大化函数 $E[\ln p(Y/\theta)]$, 其中 Y 为全部训练样本 (可观测和不可观测变量的数据), P 为条件概率。

3. 通过筛孔百分率

在《2007 年 C 题 附件 1 配合比报告统计》中, 通过筛选百分率有 13 个指标, 为了便于分析, 将这三个指标综合为粗细比率指标。根据相关专业资料^[3], 4.75mm 是区分碎石粗细的分界点, 大于 4.75mm 的是粗石, 小于 4.75mm 的是细

石。将通过筛孔百分率大于 4.75mm 的粗石比重除以小于 4.75mm 的细石比重得到粗细比率指标。

5. 1 问题一

5.1.1 模型建立:

把 TSR、S0、车辙和弯拉应变四个检验路面质量指标分别看成四个待求关联度的序列 t_1, t_2, t_3, t_4 ，再把这四个序列两两组合，求灰色关联度。

设 x_0 、 x_1 表示两个待求关联度序列 t_i, t_j ($1 \leq i \neq j \leq 4$)，其元素个数都为 n 。

1. 作一次累减生成，得：

$$x_2(i) = x_0(i+1) - x_0(i), x_3(i) = x_1(i+1) - x_1(i);$$

2. 计算所有局部关联系数：

① 先记折线段数为 $m=n-1$ ，即 x_0 、 x_1 各点的连线；

② 计算每段折线上的关联系数， $\xi_i, i = 1, 2, \dots, n-1$ ；

$$\text{若 } x_2(i) \neq 0, \text{ 记 } \omega = \frac{x_3(i)}{x_2(i)}, \text{ 则 } \xi_i = \frac{\omega(\omega^2 + 1)}{\omega^4 + 1};$$

$$\text{若 } x_2(i) = 0 \text{ 且 } x_3(i) \neq 0, \text{ 则 } \xi_i = 0;$$

$$\text{若 } x_2(i) = 0 \text{ 且 } x_3(i) = 0, \text{ 则 } \xi_i = 0 \text{ 且 } m = m-1$$

3. 求关联系数的均值并取绝对值得到第 i 个序列与第 j 个序列的关联度：

$$\tau_{ij} = \left| \frac{1}{m} \sum_{i=1}^{n-1} \xi_i \right|$$

5.1.2 问题一模型求解

把 TSR、S0、车辙、弯拉应变四个指标两两组合，分别用上面的模型求其灰色关联度值，可得下表：

表 1：四个检验路面质量指标的灰色关联度

	TSR	S0	车辙	弯拉应变
TSR		0.1627	0.0005	0.0073
S0			0.0037	0.0012
车辙				0.0917
弯拉应变				

故：

$\tau_{12}=0.1627$, 表示TSR和S0之间关联度为0.1627;

$\tau_{13}=0.0005$, 表示TSR和车辙之间的关联度为0.0005;

$\tau_{14}=0.0073$, 表示TSR和弯拉应变的关联度为0.0073;

$\tau_{23}=0.0037$, 表示S0和车辙的关联度为0.0037;

$\tau_{24}=0.0012$, 表示S0和拉弯应变的关联度为0.0012;

$\tau_{34}=0.0917$, 表示车辙和拉弯应变的关联度为0.0917;

根据文献^[4], 计算关联度的真正目的是用其考察因素之间的关联性, 它是通过两序列的关联性来体现的。两序列的关联性是指两序列所连折线的整体相似性, 而折线的整体相似性是以局部的相似性为基础的, 因此用局部相似程度的均值来度量整体的相似性, 即两因素的关联度。这样局部相似性(即两线段的相似性)的度量对关联度的计算起决定性作用。而这样计算出来的灰关联度值大的两序列组合, 就代表它们的相对关联性也较大。

结论: TSR 和 S0 的灰关联度最大, 车辙和弯拉应变的灰关联度次之, 余下的四对组合灰关联度相对较小。所以我们可以认为, 指标 TSR 和指标 S0、指标车辙和指标弯拉应变, 这两对组合指标应该是分别受一些相同因素控制和影响的, 从而可以和第二问联系起来。算法流程图详见附件。

5. 2 问题二

5. 2. 1 模型建立

将油石比、粗细比率、空隙率、矿料间隙率、饱和度、粉胶比、毛体积密度、最大理论密度、%Gmm(最初)、%Gmm(最大)这十个因素指标分别对 TSR、S0、车辙和拉弯应变四个检验路面质量指标进行逐步回归, 找出影响高速公路质量的最重要和比较重要的因素。

逐步回归分析的基本思想是有进有出: 先确定包含若干个自变量(影响总体评价的因素 $Z_i, 1 \leq i \leq m$)的初始集合; 从集合外的自变量中引入一个对因变量路面质量检验指标(Y)影响最大的变量进入回归方程; 对已选入回归方程中的变量逐个进行检验, 当原引入的变量由于后面变量的引入而变得不再显著时, 将其剔除。引入一个变量或剔除一个变量, 都为逐步回归的一步, 每一步都要进行 F 检验, 以确保每次引入新变量之前回归方程中只包含对因变量影响显著的变量。反复这个过程, 直到既无显著变量引入回归方程、也无不显著变量从回归方程中剔除为止^[5]。

设某个公路质量检验指标为因变量 Y, 有 $m=10$ 个自变量 Z (路面质量指标), 把 Y 看作是第 11 个自变量, 即 $Y = Z_{m+1}$:

$$Z_{1K}, Z_{2K} \cdots \cdots Z_{mK}, Z_{(m+1)K} \quad (K = 1, 2 \quad N)$$

第一步：计算相关系数增广矩阵：

$$r_{ij} = \frac{\sum_{k=1}^N (Z_{ik} - \bar{Z}_i)(Z_{jk} - \bar{Z}_j)}{S_i S_j}, i, j = 1, 2, \dots, 10, 1,$$

$$\text{其中, } \bar{Z}_i = \frac{1}{N} \sum_{k=1}^N X_{ik}, S_i = \sqrt{\sum_{k=1}^N (X_{ik} - \bar{X}_i)^2} \quad (1)$$

相关系数排列成矩阵为：

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{1m+1} \\ r_{21} & r_{22} & r_{2m+1} \\ \dots & \dots & \dots \\ r_{m1} & r_{m2} & r_{mm+1} \\ r_{m+11} & r_{m+12} & r_{m+1m+1} \end{bmatrix} \quad (2)$$

这是对称矩阵。| r_{ij} | ≤ 1 ，当 $i=j$ 时， $r_{ij}=1$ 。

第二步：计算方差贡献

$$v_i = \frac{r_{m+1i}^* r_{m+1i}}{r_{ij}}, 1 \leq i \leq m \quad (3)$$

第三步：剔除变量。从所有 $V_i < 0$ 的 V_i 中选出 $V_{\min} = \min\{|V_i|\}$ ， V_{\min} 对应的变量为 Z_i 。检验变量 Z_i 的显著性，计算 F_2^* 值：

$$F_2^* = |v_{\min}| \phi / r_{m+1m+1}$$

其中： $\phi = N - K_2 - 1$ （自由度， k_2 为选入回归方程中变量的个数）。根据 F_2^* 值确定是否剔除变量。若 $F_2^* < F_2$ ，则剔除对应自变量 Z_i (F_2 为一定置信水平下的 F 检验值，为剔除变量时的检验值)。自由度加 1，按式(4)计算新的相关系数增广矩阵 R ，转第二步继续计算回归方程中变量的方差贡献，考虑是否还有变量需要剔除。若 $F_2^* \geq F_2$ ，则无变量可剔除，引入变量，按式(3)计算不在回归方程中变量的方差贡献，转第四步。

凡是引入或剔除第 k 个变量，对系数矩阵均按式(4)进行计算。在第 s 步回归后计算第 $s+1$ 步的系数矩阵的公式为

$$r_{ij}^{(s+1)} = \begin{cases} r_{ij}^{(s)} / r_{kk}^{(s)}, i = k, j \neq k \\ r_{ij}^{(s)} - r_{ik}^{(s)} r_{kj}^{(s)} / r_{kk}^{(s)}, i \neq k, j \neq k \\ -r_{ij}^{(s)} / r_{kk}^{(s)}, i \neq k, j = k \\ 1/r_{kk}^{(s)}, i = k, j = k \end{cases} \quad (4)$$

第四步：引入变量。从所有 $V_i > 0$ 的 V_i 中选出 $V_{\max} = \max \{V_i\}$ ， V_{\max} 对应的变量为 Z_i 。检验变量 Z_i 的显著性，计算 F_1^* 值 $F_1^* = V_{\max} (\Phi - 1) / (r_{m+1m+1} - V_{\max})$ 。若 $F_1^* \geq F_1$ (F_1 为一定置信水平下的 F 检验值，为引入变量时的检验值)，则引入变量 X_i 。自由度减 1，按式 (4) 计算新的相关系数矩阵 R ，按式 (3) 计算回归方程中所有变量的方差贡献，转第三步。若 $F_1^* < F_1$ ，则无变量可以引入。

在逐步筛选的过程中，每选入或剔除一个变量前，都要对变量进行显著性检验。 F 检验水平的给定(即上述 F_1 、 F_2 的给定)应视具体问题而定，一般情况下，若使最终的回归方程包含更多的自变量， F 水平不易过高。

上述过程一直进行到无变量可剔可选为止。即如果第 N 步的 $F_2^* \geq F_2$ (回归方程中无变量可剔除)，而第 $N+1$ 步的 $F_1^* < F_1$ (无新变量可引入)时，逐步回归结束。

假定当筛选结束时，逐步回归共进行 N 步，并引入 1 个变量 $Z_{k1}, Z_{k2}, \dots, Z_{ki}$ ，回归系数：

$$b_{k_i} = r_{k_i, m+1} * S_{m+1} / S_{k_i}, b_0 = \bar{Y} - \sum_{i=1}^l b_{k_i} \bar{Z}_{k_i}$$

k_i 是回归方程中与原自变量对应的下标标号， S_{k_i} 与 S_{m+1} 已在第一步中求出。于是得原变量的回归方程为：

$$\hat{y} = b_0 + b_{k_1} Z_{k_1} + b_{k_2} Z_{k_2} + \dots + b_{k_i} Z_{k_i}。$$

5.2.2 模型求解

根据上述模型，在 SPSS 软件的 Regression-Linear 中选择 stepwise 回归法，将十个公路质量指标对四个检验公路质量指标分别回归，剔除不重要的影响因素，得到检验路面质量指标和影响因素之间的逐步回归线形模型：

$$\begin{cases} \text{TSR} = -0.182 * \% \text{Gmm}(\text{最大}) + 0.161 * \text{粗细比率} \\ \text{S0} = 0.96 * \text{油石比}(\%) - 0.82 * \text{矿料间隙率} - 0.385 * \% \text{Gmm}(\text{最初}) + 0.205 * \text{最大理论密度} \\ \text{车辙} = 0.322 * \text{矿料间隙率} + 0.317 * \text{饱和度} - 0.167 * \% \text{Gmm}(\text{最初}) + 0.142 * \text{最大理论密度} \\ \text{拉弯应变} = 2.019 * \text{饱和度} - 1.467 * \text{矿料间隙率} + 1.056 * \text{空隙率} + 0.180 * \text{粗细比率} \end{cases}$$

由各回归方程的偏相关系数可知，矿料间隙率是对车辙影响最大的指标，是对拉弯应变和 S0 影响第二大的指标；

饱和度是对拉弯应变影响最大的指标，是对车辙影响第二大的指标；

以上两个检验公路质量的指标都主要由矿料间隙率和饱和度决定，这与第一问这两个检验指标的相关性较高一致；

%GMM (最大) 是影响 TSR 的首要指标；

%GMM (最初)、最大理论密度、粗细比率同时能够显著影响两个路面检验指

标，是比较重要的指标；

空隙率仅是对拉弯应变影响第三大的指标。

结论：矿料间隙率、饱和度、%GMM（最大）为最重要因素，油石比、%Gmm（最初）、粗细比率、最大理论密度、空隙率为较重要因素。

提高高速公路路面质量的方案：

（1）首先矿料间隙率可以显著改变车辙、拉弯应变，还能兼顾到 S0 增强抗水性能。矿料间隙率应该是沥青混合料配合比设计的最重要参数。矿料间隙率过大或过小都会对沥青混合料的路用性能产生不利影响。矿料间隙率过小则沥青混合料的剩余空隙率和沥青用量过小，沥青混合料耐久性较差，抗疲劳能力弱，使用寿命短。在实际施工时，导致部分矿料颗粒的表面未被沥青完全裹覆，混合料过于干涩，施工和易性差。有水分作用时，沥青与矿料容易剥离，使混合料松散、解体；矿料间隙率过大，则沥青用量过大、细集料用量偏多，这对沥青混合料路用性能的影响既有有利的方面，又有不利的方面。有利的一面是沥青混合料的抗疲劳性能较好，不易出现疲劳开裂；不利的一面是沥青混合料的高温稳定性差，容易出现车辙、拥包、推挤等形式的病害。在模型分析中，发现在现有数据里面提高矿料间隙率可以显著提高车辙指标，拉弯应变和 S0，说明该指标还应当在一定范围内提高，会对路面质量改善起到显著作用。

（2）重点提高饱和度，稍微提高空隙率可以显著加强拉弯应变、车辙。饱和度是反映有效沥青占矿料间隙率提及的百分比。在控制空隙率的条件下，提高矿料间隙率就可以提高饱和度，这和上面的结论保持一致。而在实际中实施的关键是，在提高矿料间隙率时，要提提高的有效沥青占总体积的百分比，要控制空隙占总体积的百分比

（3）注重提高压实度可以显著提高 TSR 和 S0 抗水性指标。在土基填筑过程中，土在最佳含水量下压实是最有效、最经济的方法。对于过干的土可以适当洒水，对于过湿的土可以翻晒或拌合吸水材料。提高压实度，对土进行翻晒或拌合吸水材料。

（4）有效提高油石比增强 TSR 抗水性。利用压实热拌沥青混合料的空隙率和矿料间隙率等体积特性，考虑集料吸收沥青的情况，重点提高“有效油石体积比”。

（5）谨慎通过粗细比率的改变提高 TSR 抗水性能和抗低温性能。粗细比率应当谨慎控制太高太低了都不好，在模型中，适当的提高粗细比率，增加粗石的比重，会对路面质量有显著改善。级配会影响到公路质量的很多方面，应当谨慎、充分利用。

5.3 问题三模型的建立与求解

路面压实度在数学意义下的上界对 应着混合料最小的空气空隙率，在把碎石均视为球体，并且混合料具有相同的油石比情况下，那么混合料的最小空隙率就是由混合料的级配决定的，所以求路面压实度的上界与级配组成之间的数量关系就转化为求混合料的最小空隙率与级配组成之间的数量关系。

设级配为：

表 2：通过筛孔百分率(%)

31.5	26.5	19.0	16.0	13.2	9.5	4.75	2.36	1.18	0.60	0.30	0.15	0.075
x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}

经过差分，得：

$$\Delta x_i = x_{i+1} - x_i (i = 1, 2 \dots 12)$$

$$\Delta x_{13} = x_{13}$$

其中：

Δx_i ：等 i 个筛孔尺寸范围内通过的集料质量比例，如

Δx_1 ：筛孔直径 26.5~31.5 通过的集料质量比例

Δx_2 ：筛孔直径 19.0~26.5 通过的集料质量比例

.....

Δx_{13} ：筛孔直径 0~0.075 通过的集料质量比例

又由于同种岩性的集料密度是一样的，故这样的质量比例可以转化为体积比例

$$\Delta x_1 : \Delta x_2 : \dots : \Delta x_{13} = v_1 : v_2 : \dots : v_{13}$$

假设 1、碎石的形状为球形

假设 2、在同一个筛孔尺寸范围内，碎石取同一球体形状，该球体直径为筛孔尺寸范围两个边界值的均值，如：

$$d_1 = \frac{31.5 + 26.5}{2} = 29$$

所以得到 13 种形状的球体，对应的直径分别为：

$$d_1, d_2, \dots, d_{13}$$

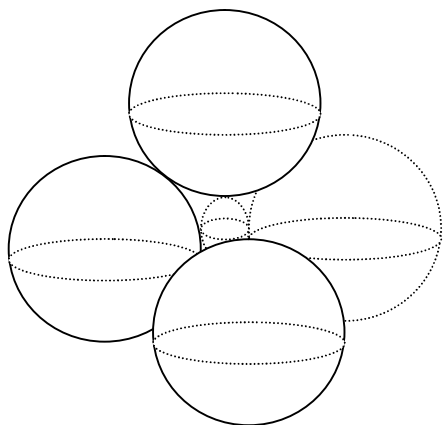
在假设 2 的前提下，我们可以得到同一个筛孔尺寸范围内的球的个数比例：

$$n_i = \frac{\Delta V_i}{\frac{4}{3} \times \pi d_i^3} (i = 1, 2 \dots 13)$$

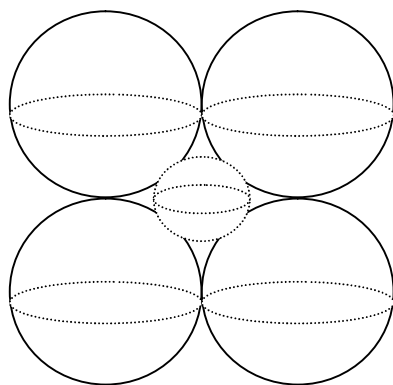
5.3.1 理论上的思路：

此问所对应的数学问题就是把数目一定的大小不同的 13 种球投入一个正方体箱子中，怎样的投放方式可以使正方体箱子所剩的空隙体积最小，我们只考虑组成这种投放方式的“基本单元”，有下面三种最主要的“基本单元”：

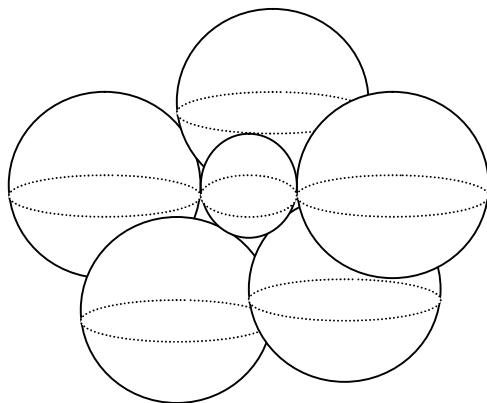
1、三个大球夹一个小球图



2、四个大球夹一个小球图

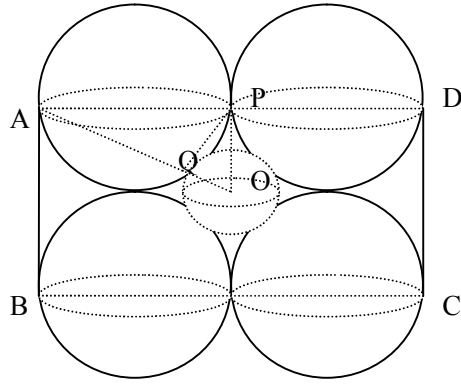


3、五个大球夹一个小球图



这里我们只对第二种的“基本单元”求其最小空隙体积，按此思路可类推到任何一种情形。

解决思路：



(一) 分 4 个大球和 1 个小球，其摆放状态如图所示。设四个大球的球边界点为 A、B、C、D，小球球心为 O，球 A 和球 D 相切于点 P，OA 交球 A 于点 Q。易知，ABCD 为四个大球的纵切面外边界。记大球半径 R_m 、小球半径 R_s ，则：

$$OP = R_m, OA = \sqrt{AP^2 + OP^2} = \sqrt{5}R_m$$

$$R_s \leq OQ = OA - AQ = (\sqrt{5} - \sqrt{AP^2 - PQ^2})R_m \approx 0.447R_m \quad (\text{因为 } AQ = 2PQ)$$

未填上小球时大球间的空隙体积为：

$$V_{\text{空隙}0} = [V_{\text{长方体}} - 8 \times \frac{1}{4} V_{\text{大球}}] = [4 \times 2 \times 2 - \frac{8\pi}{3}] / 2 \times R_m^3 \approx 3.81R_m^3$$

其中的长方体指的是 8 个小球放成两排，每排 4 个，即如图的后方（纸面背面）还有相同的 4 个球。另外的 4 个球也有如 ABCD 的点，记为 EFGH，则长方体就是 ABCD-EFGH

填上小球后的空隙体积为：

$$V_{\text{空隙}} = [V_{\text{空隙}0} - V_{\text{小球}}] = [3.81 - \frac{4\pi(0.447)^3}{3}] R_m^3 \approx 3.437R_m^3,$$

$$\text{故有 } V_{\text{空隙}0} \approx 1.109 V_{\text{空隙}}$$

设大球个数：小球个数 = $N : M$ （即大球 N 个，小球 M 个）。

当 $N \geq 4M$ 时，空隙体积和为：

$$V_{\text{总空隙}} = M \times V_{\text{空隙}} + (N - 4M) \times V_{\text{空隙}0} \div 4 = (N \times 1.109 / 4 - 0.109M) V_{\text{空隙}}$$

N 越大，总空隙体积越大，反之 N 越小，总空隙体积越小，最小为 $N = 4M$ 时达到。

$$\text{Min}_{N \geq 4M} (V_{\text{总空隙}}) = M V_{\text{空隙}}$$

当 $N < 4M$ 时，空隙体积和为：

$$V_{\text{总空隙}} = [\frac{N}{4}] \times V_{\text{空隙}} + (M - [\frac{N}{4}]) \times V_{\text{空隙}0},$$

其中 $V_{\text{空隙}0}'$ 表示 4 个小球在一起时的空隙体积。

$$V_{\text{空隙}0}' = 3.81Rs^3 \approx 0.34Rm^3 = 0.1V_{\text{空隙}}$$

所以有

$$V_{\text{总空隙}} = \left[\frac{N}{4}\right] \times V_{\text{空隙}} + (M - \left[\frac{N}{4}\right]) \times 0.1V_{\text{空隙}}$$

显然，当 $N=4M$ 时，总空隙达到最小（因为 $M - \left[\frac{N}{4}\right] \geq 0$ ）

所以，我们的结论是总空隙在级配比恰好为 4: 1 时，总空隙最小，即压实度最大。

（二）、大球仍是 4 个一组，小球 a 个， $a > 1$ 。

这时，

$$V_{\text{空隙}} = V_{\text{空隙}0} - a \times V_{\text{小球}} = V_{\text{空隙}0} - \frac{a \times 4\pi \times Rs^3}{3}$$

假设仍旧有 N 个大球、 M 个小球，

$$V_{\text{总空隙}} = \left[\frac{N}{4}\right] \times V_{\text{空隙}} + (M - \left[\frac{N}{4}\right] \times a) \times V_{\text{空隙}0}'$$

与（一）相同， $N: M=4: a$ 时总空隙最小，压实度最大。

（三）、大球 n 个，小球 m_1, m_2, \dots, m_{13} 个

这时，和上面分析的一样，我们有大球和各种小球的组合为级配比组合

$$\Delta x_1 : \Delta x_2 : \dots : \Delta x_{13} = m_1 : m_2 : \dots : m_k$$

时，总空隙最小，从而压实度最大。

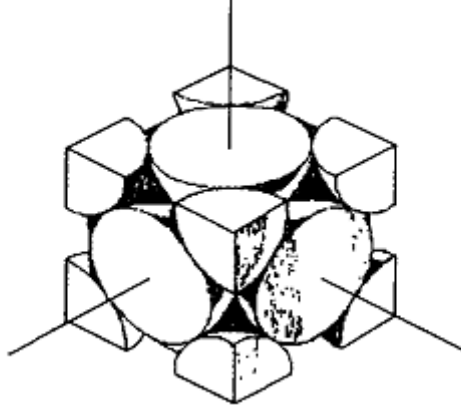
因此，当级配比和混合料本身的各颗粒体积大小比相同时，总空隙体积最小，从而压实度最大。

这只是理论上的解决思路，但由于实际应用中各粒径颗粒不可能是完全的球形，所以这种理论无法直接应用到实际中去。下面我们给出一种简化模型。

5.3.2 简化模型——考虑只有两种球的情况

当级配组成中两种球的直径之比大于一定值时，大球之间形成的空隙可以由大量小球来填充。

一个重要结论：对于形状简单的同一粒径的球形集料而言，存在一个 25.95% 的最小空隙率^[6]。



此结论中的最小空隙率与球体的直径无关。根据这个重要的结论我们就可以构造出两个球情况下的最小空隙率 ξ 。

1. 最小空隙率与级配之间的数量关系

如果大球和小球的半径大小既定，则在上图的切块里大球与小球的数量就既定，不妨设上图切块中包含了 $n_{\text{大}}$ 个大球和 $n_{\text{小}}$ 个小球，令 $k = \frac{n_{\text{大}}}{n_{\text{小}}}$ 。

在这个切块中：

$$V_{\text{空}} = V_{\text{块}} - V_{\text{大}} - V_{\text{小}}$$

其中， $V_{\text{空}}$ 是此切块的空隙体积， $V_{\text{块}}$ 是切块的体积， $V_{\text{大}}$ 是切块中大球的总体积， $V_{\text{小}}$ 是切块中的小球的总体积。

依据图形可以看出，这个切块的边长跟大球半径有如下关系：

$$L_{\text{块}} = (4 \times r_{\text{大}}) \cos(\pi/4)$$

并且切块相当于包括 4 个大球，而小球就应该包含了 $4k$ 个，可以算出：

$$V_{\text{块}} = L_{\text{块}}^3 = [(4r_{\text{大}}) \cos(\pi/4)]^3$$

$$V_{\text{大}} = 4 \times \frac{4}{3} \pi r_{\text{大}}^3$$

$$V_{\text{小}} = 4k \times (\frac{4}{3} \pi r_{\text{小}}^3)$$

上述三式代入 $V_{\text{空}} = V_{\text{块}} - V_{\text{大}} - V_{\text{小}}$ 得到如下结果：

$$V_{\text{空}} = 16\sqrt{2}r_{\text{大}}^3 - \frac{16}{3}\pi r_{\text{大}}^3 - \frac{16}{3}k\pi r_{\text{小}}^3$$

现在就来考虑两球的级配比例 $\frac{\Delta x_{\text{大}}}{\Delta x_{\text{小}}}$ 和最小空隙率之间的数量关系：

因为 $\frac{\Delta x_{\text{大}}}{\Delta x_{\text{小}}} = \frac{n_{\text{大}}}{n_{\text{小}}}$ ，所以：

(1) 若 $\frac{\Delta x_{\text{大}}}{\Delta x_{\text{小}}} = \frac{n_{\text{大}}}{n_{\text{小}}} = k$ ，则相当于二者的混合恰好能够形成上图的结构而无多余的

的球，这样，空隙率就为：

$$\xi = \frac{V_{\text{空}}}{V_{\text{块}}} = \frac{16\sqrt{2}r_{\text{大}}^3 - \frac{16}{3}\pi r_{\text{大}}^3 - \frac{16}{3}k\pi r_{\text{小}}^3}{16\sqrt{2}r_{\text{大}}^3} = \frac{(1 - \frac{\sqrt{2}}{6})r_{\text{大}}^3 - \frac{\sqrt{2}}{6}k\pi r_{\text{小}}^3}{r_{\text{大}}^3}$$

(2) 若 $\frac{\Delta x_{\text{大}}}{\Delta x_{\text{小}}} = \frac{n_{\text{大}}}{n_{\text{小}}} > k$ ，即是说把这些大小球拼凑成这种一个个切块，大球有剩。

所以可以将总体积分割成由小球数量决定的大球数量所对应的切块，和由多余的大球也排成这种形式的切块（只不过没有小球填充在里面）。从而空隙率：

$$\begin{aligned} \xi &= \frac{V_{\text{空}}}{V_{\text{块1}} + V_{\text{块2}}} = \frac{\frac{kn_{\text{小}}}{4}(2\sqrt{2}r_{\text{大}})^3 - \frac{kn_{\text{小}}}{4}\frac{4\pi r_{\text{大}}^3}{3} - n_{\text{小}}\frac{4\pi r_{\text{小}}^3}{3} + 25.95\%(n_{\text{大}} - kn_{\text{小}})\frac{4\pi r_{\text{大}}^3}{3}}{\frac{kn_{\text{小}}}{4} \times (2\sqrt{2}r_{\text{大}})^3 + (n_{\text{大}} - kn_{\text{小}})\frac{4\pi r_{\text{大}}^3}{3}} \\ &= \frac{4\sqrt{2}kr_{\text{大}}^3 - \frac{4}{3}\pi(kr_{\text{大}}^3 + r_{\text{小}}^3) + 25.95\%(\frac{n_{\text{大}}}{n_{\text{小}}} - k)\frac{4\pi r_{\text{大}}^3}{3}}{4\sqrt{2}kr_{\text{大}}^3 + (\frac{n_{\text{大}}}{n_{\text{小}}} - k)\frac{4\pi r_{\text{大}}^3}{3}} \\ &= \frac{4\sqrt{2}kr_{\text{大}}^3 - \frac{4}{3}\pi(kr_{\text{大}}^3 + r_{\text{小}}^3) + 25.95\%(\frac{\Delta x_{\text{大}}}{\Delta x_{\text{小}}} - k)\frac{4\pi r_{\text{大}}^3}{3}}{4\sqrt{2}kr_{\text{大}}^3 + (\frac{n_{\text{大}}}{n_{\text{小}}} - k)\frac{4\pi r_{\text{大}}^3}{3}} \end{aligned}$$

(3) 若 $\frac{\Delta x_{\text{大}}}{\Delta x_{\text{小}}} = \frac{n_{\text{大}}}{n_{\text{小}}} < k$ ，同理可以算得：

$$\xi = \frac{4\sqrt{2}\frac{\Delta x_{\text{大}}}{\Delta x_{\text{小}}}r_{\text{大}}^3 - \frac{4}{3}\pi\frac{\Delta x_{\text{大}}}{\Delta x_{\text{小}}}(r_{\text{大}}^3 + \frac{1}{k}r_{\text{小}}^3) + \frac{4}{3} \times 25.95\%\pi(1 - k\frac{\Delta x_{\text{大}}}{\Delta x_{\text{小}}})r_{\text{小}}^3}{4\sqrt{2}\frac{\Delta x_{\text{大}}}{\Delta x_{\text{小}}}r_{\text{大}}^3 + \frac{4}{3}\pi(1 - k\frac{\Delta x_{\text{大}}}{\Delta x_{\text{小}}})r_{\text{小}}^3}$$

5.4 问题四

5.4.1 模型建立

对第一个问题使用 GLM (Generalized Linear Model 广义方差分析或广义线性模型) 对影响检验路面质量指标的路面质量指标进行广义线性回归和方差分析。当数据不满足均衡设计, 则应该使用 GLM 过程的线性模型结构。

一个方差分析模型可以表示为一个线性模型, 这个模型是一个方程式, 它预报响应变量为参数和设计变量的线性函数。一般地,

$$y_i = \beta_0 x_{0i} + \beta_1 x_{1i} + \dots + \beta_k x_{ki} + \varepsilon_i \quad i=1,2,\dots,n$$

其中 y_i 是第 i 个观测的响应, β_k 是待估的未知参数, 而 x_{ij} 是设计变量。在方差分析中设计变量是一些指示变量, 也就是它们总是等于 0 或 1。

最简单的模型是对所有数据拟合唯一的一个均值。在这种情况下, 仅有一个参数 β_0 和一个设计变量 x_{0i} , 它的值总是 1:

$$y_i = \beta_0 x_{0i} + \varepsilon_i = \beta_0 + \varepsilon_i \quad i=1,2,\dots,n$$

β_0 的最小二乘估计是 y_i 的均值。这个最简单的模型是所有较复杂模型的基础,

而且所有较大模型都是与这简单模型比较。在线性模型的参数表示中, β_0 通常称为截距 (intercept)。

单因子访查分析模型对于分类变量的每个水平通过引入一个指示变量来表示。假设变量 A 有 4 个水平, 每个水平有两个观测。这个指示变量被创建如下:

表 3: 指示变量

截距	A1	A2	A3	A4
1	1	0	0	0
1	1	0	0	0
1	0	1	0	0
1	0	1	0	0
1	0	0	1	0
1	0	0	1	0
1	0	0	0	1
1	0	0	0	1

此例对应的线性模型是

$$y_i = \beta_0 + \beta_1 A1_i + \beta_2 A2_i + \beta_3 A3_i + \beta_4 A4_i$$

为了构造交互的或嵌套的效应, 用户可以简单地把主效应列相乘得到所有组合。这些内容将在 GLM 过程的 GLM 模型的参数化法中介绍。

1. 线性假设

当模型被表示为线性模型结构时，假设检验表示为参数的线性函数。例如，你可能想要检验： $\beta_2 - \beta_3 = 0$ 。一般地，关于线性假设的系数是一组 L_i ：

$$H_0 : L_0\beta_0 + L_1\beta_1 + \dots + L_k\beta_k = 0$$

这些线性函数的一部分可组成一个检验。这些检验用矩阵方程可表示为

$$H_0 : L\beta = 0$$

对每个线性假设，能够构造使用与该假设的平方和。这些平方和或者作为这些估计的二次型：

$$SS(L\beta = 0) = (Lb)'[L(X'X)^{-1}L']^{-1}(Lb)$$

或者等价地作为由零假设所约束模型的误差平方和（SSE）的增量：

$$SS(L\beta = 0) = SSE(\text{约束模型}) - SSE(\text{全模型})$$

被计算，然后用这个 SS 除以相应的自由度，并用它作为 F 统计量的分子。

5.4.2 模型求解

1. 沥青、碎石质量对高速公路路面性能的影响

（1）沥青质量对检验高速公路路面质量指标的影响

首先规定按质量沥青划分为普通沥青和改性沥青，不再按其型号详分。将沥青改性与否作为各公路质量检验指标的单影响因素在 SPSS 中做 GLM (Generalized Linear Model)，进行方差分析和广义线性回归模型。分析结果显示：沥青改性与否对高速公路路面抗水性能无影响，但是改性的沥青可以十分显著的提高高速公路路面高温性和低温型。具体分析如下：

I、改性沥青和普通沥青在对高速公路路面抗水性能无明显影响。在改性沥青与否对 TSR 和 SO 的 GLM 模型分析中，P 值分别为 0.373 和 0.495，均大于 0.05（见表 3 和表 4），故均不能拒绝原假设，广义线性回归模型结果中应变量基本全部由常数项解释，说明沥青改进与否对抗水性能的改善确实没有影响，改进高速公路路面的抗水性能另有因素。

表 4：改性沥青和普通沥青对 TSR 的 GLM 分析

Dependent Variable: TSR(%)				
变异来源	自由度	均方MS	F	P值
校正的模型	1	8.403	.795	.373
常数项	1	1533934.047	145113.148	.000
改进沥青还是普通沥青	1	8.403	.795	.373
误差	287	10.571		
合计	289			
校正合计	288			

表 5：改性沥青和普通沥青对 S0 的 GLM 分析

Dependent Variable: S0(%)

变异来源	自由度	均方MS	F	P值
校正模型	1	41.206	3.815	.052
常数项	1	1700862.165	157453.450	.000
改性沥青还是普通沥青	1	41.206	3.815	.052
误差	287	10.802		
合计	289			
校正合计	288			

II、改性沥青较普通沥青可以十分显著的提高高速公路路面高温性。在该指标对车辙的 GLM 模型分析中，F 值为 118.252 拒绝原假设，分析结果如下：

表 6：改性沥青和普通沥青对车辙的 GLM 分析

Dependent Variable: 车辙

变异来源	自由度	均方MS	F	P值
校正的模型	1	240819327.158	118.252	.000
常数项	1	2600968927.951	1277.181	.000
改性沥青还是普通沥青	1	240819327.158	118.252	.000
误差	287	2036491.892		
合计	289			
校正合计	288			

并且有如下精确关系，在其它影响车辙因素不变的情况下，改性路面相对普通路面可以平均使车辙指标平均显著上升 2141.207。

$$\text{车辙} = 2447.840 + 2141.207 * \text{改性}$$

$$\text{标准差: (171.797) (196.904)}$$

$$\text{T 值: (14.248) (10.874)}$$

III、改性沥青较普通沥青可以十分显著的提高高速公路路面低温性能。在该指标对拉弯应变的 GLM 模型分析中，F 值为 42.607 拒绝原假设，分析结果如下。

表 7：改性沥青和普通沥青对弯拉应变的 GLM 分析

Dependent Variable: 弯拉应变

变异系数	自由度	均方MS	F	P值
校正的模型	1	4718892.426	42.607	.000
常数项	1	1761893613.872	15908.135	.000
改进沥青还是普通沥青	1	4718892.426	42.607	.000
误差	287	110754.254		
合计	289			
校正合计	288			

并且有如下精确关系，在其它影响拉弯应变因素不变的情况下，改性路面相对普通路面可以平均使拉弯应变指标平均显著上升 299.732。

$$\text{拉弯应变} = 2745.963 + 299.732 * \text{改性}$$

$$\text{标准差: (40.064) (45.919)}$$

T 值: (68.539) (6.527)

(2) 碎石质量对检验高速公路路面质量指标的影响

碎石质量按照岩性分为三类,即石灰岩、玄武岩、和其他种类。将其作为各公路质量检验指标的单影响因素在 SPSS 中做 GLM,进行方差分析和广义线形回归模型。分析结果显示,碎石质量可以十分明显的改变路面质量的抗高温和抗低温性能,以玄武岩的性能最优,其他性能次之,石灰岩性能最差。碎石质量可以较显著的改变 TSR 指标反映出的抗水性能,但不能改变 S0 指标所反映的抗水性能。具体分析如下。

I、碎石质量对 TSR 抗水性能指标有显著性影响。在该指标对 TSR 的 GLM 模型分析中, F 值为 3.257 , P 值为 0.04 小于 0.05 的显著性水平,拒绝原假设,见表 7。但结果的显著性不是很强。

表 8: 碎石质量对 TSR 的 GLM 分析

Dependent Variable: TSR(%)					
变异来源	自由度	MS均方	F	P值	检验效能 ^a
校正的模型	2	39.322	3.257	.040	.616
常数项	1	1103579.914	91404.027	.000	1.000
碎石质量	2	39.322	3.257	.040	.616
误差	234	12.074			
合计	237				
校正的合计	236				

a. Computed using alpha = .05

II、碎石质量对 S0 抗水性能指标无显著性影响。在岩性对 S0 的 GLM 模型分析中, F 值均不能拒绝原假设, 广义线形回归模型结果中应变变量基本全部由常数项解释, 说明岩性对 S0 抗水性能的改善确实没有影响, 改进高速公路路面的抗水性能另有因素。分析结果如下:

表 9: 碎石质量对 S0 的 GLM 分析

Dependent Variable: S0(%)					
变异来源	自由度	MS均方	F	P值	检验效能 ^a
误差	2	32.104	2.887	.058	.561
常数项	1	1229844.236	110586.6	.000	1.000
碎石质量	2	32.104	2.887	.058	.561
误差	234	11.121			
合计	237				
校正的误差	236				

a. Computed using alpha = .05

III、碎石质量可以十分显著的改变高速公路路面高温性。在该指标对车辙的 GLM 模型分析中, F 值为 30.153。拒绝原假设, 分析结果如下。并且有如下精确分析关系, 在其它影响车辙的因素不变的情况下, 如从使用其它岩性的碎石改用于石灰岩可使车辙指标平均下降 277.065, 如从使用其它岩性的碎石改用于玄武岩可使车辙指标平均上升 1407.708。

表 10：碎石质量对车辙的 GLM 分析

Dependent Variable: 车辙

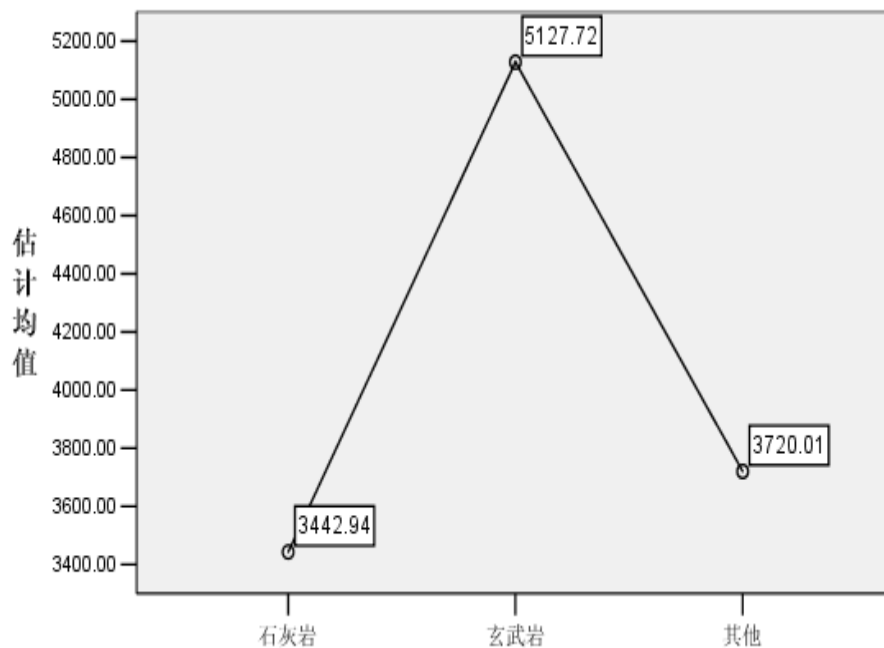
变异来源	自由度	MS均方	F	P值	检验效能 ^a
校正的模型	2	70721416.32	30.153	.000	1.000
常数项	1	2564207511	1093.277	.000	1.000
碎石质量	2	70721416.32	30.153	.000	1.000
误差	234	2345431.603			
合计	237				
校正的误差	236				

a. Computed using alpha = .05

$$\text{车辙} = 3720.01 - 277.065 \times \text{石灰岩} + 1407.708 \times \text{玄武岩}$$

标准差: (300.348) (328.195) (1407.708)
T 值: (12.386) (-8.44) (4.052)

图 1：不同碎石质量的车辙指标的估计均值比较图



IV、碎石质量可以十分显著的改变高速公路路面低温性能。在该指标对拉弯应变的 GLM 模型分析中，F 值为 16.335。拒绝原假设，分析结果略。并且有如下精确分析关系，在其它影响拉弯应变的因素不变的情况下，如从使用其它岩性的碎石改用为石灰岩可使车辙指标平均下降 80.473，如从使用其它岩性的碎石改用为玄武岩可使车辙指标平均上升 205.375。

表 11：碎石质量对拉弯应变的 GLM 分析

Dependent Variable: 弯拉应变

变异来源	自由度	MS均方	F	P值	检验效能 ^a
校正的模型	2	2004383.820	16.335	.000	1.000
常数项	1	1370004971	11165.147	.000	1.000
碎石质量	2	2004383.820	16.335	.000	1.000
误差	234	122703.711			
合计	237				
校正的误差	236				

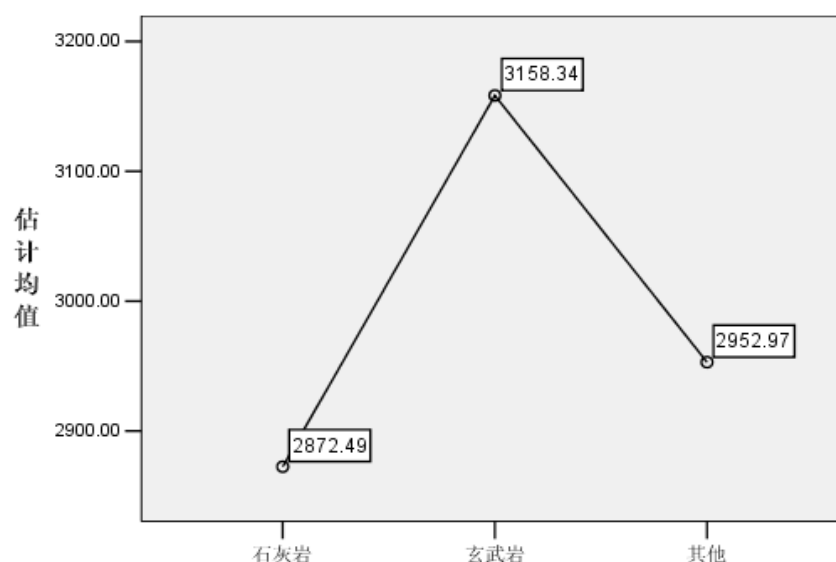
a. Computed using alpha = .05

拉弯应变 = 2952.976 - 80.473 * 石灰岩 + 205.375 * 玄武岩

标准差: (68.698) (75.067) (79.454)

T 值: (42.985) (-1.702) (2.585)

图 2：不同碎石质量拉弯应变的估计均值比较图



2. 不同产地、不同厂家但型号、类型相同的沥青、碎石质量对检验公路路面指标的影响

根据沥青、碎石的型号和类型，我们将混合料分为六种组合，如下表：

表 12：沥青、碎石组合表

	石灰岩	玄武岩	其他岩性
改性	1 改性石灰岩 (82 例)	2 改性玄武岩 (75 例)	3 改性其他岩性 (17 例)
普通	4 普通石灰岩 (51 例)	5 普通玄武岩 (2 例)	6 其他岩性 (9 例)

对每种组合的混合料，分析其对高速路面质量的影响时，需要考虑到“石料产地”和“沥青厂家”这两个因素的影响，因此我们做双因素方差分析。进一步

地，由于“石料产地”和“沥青厂家”这两个因素搭配在一起可能会产生交互作用，故最终选择运用 GLM（广义因素方差分析）做出交互作用的双因素方差分析。

由于厂家的数据分布十分离散，因此仅取部分频数大于 10 的六家厂家进行比较分析，他们是韩国 SK（42 例），路安特（11 例），韩国 SK（22 例），江阴克斯密（12 例），盘锦（16 例），江阴宝利（13 例）。

将所有产地按照中国地区划分，数据中出现产地被划分到华东、华中、华北、西南、东北、西北以及其他地区中。各地区的频数如下表，按频数高低主要考察华东地区、华中地区、其他地区。在各指标的分析中，只分析频数最高的几个：华东（124 例）、华中（76 例）、西北（11 例）、华北（10 例）和其他即国外地区（17 例）。

（1）改性石灰岩

I、将产地、厂家及其交互项对 TSR 做双因素方差分析。由输出结果可知，GLM 的模型 P 值为 0.036 小于 0.05 有统计学意义。用于检验“产地”和“厂家”的 P 值分别为 0.285 和 0.101，都小于 0.05，拒绝原假设，表明不同的产地和厂家之间对 TSR 没有显著差异。

表 13：改性石灰岩——产地和厂家双因素对 TSR 的 GLM 分析

Dependent Variable: TSR(%)					
变异来源	自由度	均方	F	P 值	检验效能 ^a
校正的模型	5	42.165	3.011	.036	.742
常数项	1	145238.785	10370.118	.000	1.000
产地*厂家	0
产地	2	18.821	1.344	.285	.254
厂家	2	36.377	2.597	.101	.454
误差	19	14.006			
合计	25				
校正的合计	24				

a. Computed using alpha = .05

尽管从整体上看，不同的厂家对 TSR 没有明显差异，但是可以用 LSD 方法两两比较方法考察，厂家之间的区别。

Bonferroni (LSD) 方法是一种两两比较方法。适合于如下情况：如果存在明确的对照组，要进行的试验正性研究，即计划好的某两个或几个组间（和对照组）的比较。

从下表可见，韩国 SK 对路安特、克拉玛依的均值比较 P 值分别是 0.039、0.015，路安特和克拉玛依对江阴克斯密的均值比较 P 值分别为 0.035、0.019，都小于 0.05，因此可以判断它们之间存在显著差别。

表 14：改性石灰岩——厂家的 TRS 均值的两两比较

Dependent Variable: TSR(%)
LSD

(I) 沥青厂家	(J) 沥青厂家	均值差异	标准差	P值
韩国SK	路安特	5.0875*	2.29174	.039
	克拉玛依	4.8569*	1.81848	.015
	江阴科斯密	-.9125	2.29174	.695
路安特	韩国SK	-5.0875*	2.29174	.039
	克拉玛依	-.2306	2.24890	.919
	江阴科斯密	-6.0000*	2.64627	.035
克拉玛依	韩国SK	-4.8569*	1.81848	.015
	路安特	.2306	2.24890	.919
	江阴科斯密	-5.7694*	2.24890	.019
江阴科斯密	韩国SK	.9125	2.29174	.695
	路安特	6.0000*	2.64627	.035
	克拉玛依	5.7694*	2.24890	.019

Based on observed means.

*. The mean difference is significant at the .05 level.

同样，用LSD方法两两考察产地的差异，如下表。发现华中地区对西北地区、其他地区即国外的P值小于0.05，均指显著性偏高。

表 15：改性石灰岩——地区的 TRS 均值的两两比较

Dependent Variable: TSR(%)
LSD

(I) 产地	(J) 产地	均值差异	标准差	P值
华东	华中	-4.5513	2.39705	.073
	西北	2.4987	1.84705	.192
	其他	3.8821	2.39705	.122
华中	华东	4.5513	2.39705	.073
	西北	7.0500*	2.64627	.015
	其他	8.4333*	3.05565	.012
西北	华东	-2.4987	1.84705	.192
	华中	-7.0500*	2.64627	.015
	其他	1.3833	2.64627	.607
其他	华东	-3.8821	2.39705	.122
	华中	-8.4333*	3.05565	.012
	西北	-1.3833	2.64627	.607

Based on observed means.

*. The mean difference is significant at the .05 level.

II、将产地、厂家及其交互项对 S0 做双因素方差分析。由输出结果可知，GLM 的模型 P 值为大于 0.05 没有统计学意义。“产地”和“厂家”及其交互效应都对没有 S0 没有影响。

表 16：改性石灰岩——产地和厂家双因素对 S0 的 GLM 分析

Dependent Variable: S0(%)

变异来源	自由度	均方	F	P 值	检验效能 ^a
校正的模型	5	35.987	1.826	.156	.496
常数项	1	160267.180	8132.766	.000	1.000
产地*厂家	0
产地	2	44.120	2.239	.134	.399
厂家	2	3.185	.162	.852	.071
误差	19	19.706			
合计	25				
校正的合计	24				

a. Computed using alpha = .05

尽管整体模型不显著，无法考察通过GLM的两因素方差分析考察不同厂家、产地对四项检验路面质量的指标，但还是可以用LSD方法，进行两两比较，结果如下表。可见，任意两个厂家的该指标都没有显著差别。

表 17：改性石灰岩——厂家的 S0 均值的两两比较

Dependent Variable: S0(%)

LSD

(I) 沥青厂家	(J) 沥青厂家	均值差异	标准差	P 值
韩国SK	路安特	1.2375	2.71843	.654
	克拉玛依	4.2094	2.15706	.066
	江阴科斯密	.0375	2.71843	.989
路安特	韩国SK	-1.2375	2.71843	.654
	克拉玛依	2.9719	2.66762	.279
	江阴科斯密	-1.2000	3.13898	.706
克拉玛依	韩国SK	-4.2094	2.15706	.066
	路安特	-2.9719	2.66762	.279
	江阴科斯密	-4.1719	2.66762	.134
江阴科斯密	韩国SK	-.0375	2.71843	.989
	路安特	1.2000	3.13898	.706
	克拉玛依	4.1719	2.66762	.134

Based on observed means.

但是，其他即国外地区的该项指标对华东、华中地区比较的P值分别为0.034、0.010，都显著偏低。

表 18：改性石灰岩——地区的 S0 均值的两两比较

Dependent Variable: S0(%)

LSD

(I) 产地	(J) 产地	均值差异	标准差	P值
华东	华中	-3.8846	2.84335	.188
	西北	1.5024	2.19095	.501
	其他	6.4821*	2.84335	.034
华中	华东	3.8846	2.84335	.188
	西北	5.3870	3.13898	.102
	其他	10.3667*	3.62458	.010
西北	华东	-1.5024	2.19095	.501
	华中	-5.3870	3.13898	.102
	其他	4.9796	3.13898	.129
其他	华东	-6.4821*	2.84335	.034
	华中	-10.3667*	3.62458	.010
	西北	-4.9796	3.13898	.129

Based on observed means.

*. The mean difference is significant at the .05 level.

III、将产地、厂家及其交互项对车辙做双因素方差分析。由输出结果可知，GLM 的模型 P 值为小于 0.05 有统计学意义。但用于检验“产地”和“厂家”的对车辙的影响的 P 值都大于 0.05，没有显著差异。

表 19：改性石灰岩——产地和厂家双因素对车辙的 GLM 分析

Dependent Variable: 车辙

变异来源	自由度	均方	F	P值	检验效能 ^a
校正的模型	5	10024577.03	7.764	.000	.994
常数	1	294989564.0	228.467	.000	1.000
产地*厂家	0
产地	2	2139205.788	1.657	.217	.305
厂家	2	2881803.312	2.232	.135	.398
误差	19	1291169.463			
合计	25				
校正的误差	24				

a. Computed using alpha = .05

在用LSD方法两两比较均值差异，如下表。可见，韩国SK和路安特对克拉玛依的车辙指标均指比较的P值均小于0.05，显著偏高。

表 20：改性石灰岩——厂家的车辙均值的两两比较

Dependent Variable: 车辙

LSD

(I) 沥青厂家	(J) 沥青厂家	均值差异	标准差	P值
韩国SK	路安特	-1017.6250	695.83658	.160
	克拉玛依	2342.6250*	552.14079	.000
	江阴科斯密	-536.6250	695.83658	.450
路安特	韩国SK	1017.6250	695.83658	.160
	克拉玛依	3360.2500*	682.82914	.000
	江阴科斯密	481.0000	803.48288	.556
克拉玛依	韩国SK	-2342.6250*	552.14079	.000
	路安特	-3360.2500*	682.82914	.000
	江阴科斯密	-2879.2500*	682.82914	.000
江阴科斯密	韩国SK	536.6250	695.83658	.450
	路安特	-481.0000	803.48288	.556
	克拉玛依	2879.2500*	682.82914	.000

Based on observed means.

*. The mean difference is significant at the .05 level.

分地区比较，华东地区对西北、其他地区的车辙均值比较的P值为分别为0.000、0.001均小于0.05，显著偏高。华中地区对西北、其他地区的车辙均值比较的P值均为0.001均小于0.05，显著偏高。

表 21：改性石灰岩——地区的车辙均值的两两比较

Dependent Variable: 车辙

LSD

(I) 产地	(J) 产地	均值差异	标准差	P值
华东	华中	-620.0769	727.81217	.405
	西北	2455.7564*	560.81694	.000
	其他	2933.2564*	727.81217	.001
华中	华东	620.0769	727.81217	.405
	西北	3075.8333*	803.48288	.001
	其他	3553.3333*	927.78211	.001
西北	华东	-2455.7564*	560.81694	.000
	华中	-3075.8333*	803.48288	.001
	其他	477.5000	803.48288	.559
其他	华东	-2933.2564*	727.81217	.001
	华中	-3553.3333*	927.78211	.001
	西北	-477.5000	803.48288	.559

Based on observed means.

*. The mean difference is significant at the .05 level.

IV、将产地、厂家及其交互项对拉弯应变做双因素方差分析。由输出结果可知，GLM 的模型 P 值为大于 0.05 没有统计学意义。用于检验“产地”和“厂家”的对拉弯应变都没有显著差异。

表 22：改性石灰岩——产地和厂家双因素对拉弯应变的 GLM 分析

Dependent Variable: 弯拉应变

变异来源	自由度	均方	F	P值	检验效能 ^a
校正的模型	5	30472.203	.195	.961	.085
常数项	1	174786346.8	1119.908	.000	1.000
产地*厂商	0
产地	2	65842.798	.422	.662	.108
厂商	2	10616.783	.068	.934	.059
误差	19	156072.084			
合计	25				
校正的合计	24				

a. Computed using alpha = .05

用LSD法进行两两比较，所有的P值均大于0.05，任何两个厂家或产地都没有显著性差别。

表 23：改性石灰岩——厂家的拉均应变均值的两两比较

Dependent Variable: 弯拉应变

LSD

(I) 沥青厂家	(J) 沥青厂家	均值差异	标准差	P值
韩国SK	路安特	44.0086	241.92361	.858
	克拉玛依	27.3298	191.96446	.888
	江阴科斯密	86.2086	241.92361	.726
路安特	韩国SK	-44.0086	241.92361	.858
	克拉玛依	-16.6788	237.40127	.945
	江阴科斯密	42.2000	279.34932	.882
克拉玛依	韩国SK	-27.3298	191.96446	.888
	路安特	16.6788	237.40127	.945
	江阴科斯密	58.8788	237.40127	.807
江阴科斯密	韩国SK	-86.2086	241.92361	.726
	路安特	-42.2000	279.34932	.882
	克拉玛依	-58.8788	237.40127	.807

Based on observed means.

表 24：改性石灰岩——地区的拉均应变均值的两两比较

Dependent Variable: 弯拉应变

LSD

(I) 产地	(J) 产地	均值差异	标准差	P值
华东	华中	-18.1666	253.04065	.944
	西北	-93.6686	194.98092	.636
	其他	161.4449	253.04065	.531
华中	华东	18.1666	253.04065	.944
	西北	-75.5020	279.34932	.790
	其他	179.6115	322.56481	.584
西北	华东	93.6686	194.98092	.636
	华中	75.5020	279.34932	.790
	其他	255.1135	279.34932	.373
其他	华东	-161.4449	253.04065	.531
	华中	-179.6115	322.56481	.584
	西北	-255.1135	279.34932	.373

Based on observed means.

(2) 改性玄武岩

I、将产地、厂家及其交互项对 TSR 做双因素方差分析。由输出结果可知，GLM 的模型 P 值为 0.797 大于 0.05 没有统计学意义。不同的产地和厂家之间对 TSR 都没有显著差异。

表 25：改性玄武岩——产地和厂家双因素对 TSR 的 GLM 分析

Dependent Variable: TSR(%)

变异来源	自由度	均方	F	P值	检验效能 ^a
校正的模型	8	7.224	.560	.797	.190
常数项	1	113088.519	8761.948	.000	1.000
产地*厂家	1	.024	.002	.966	.050
产地	2	15.483	1.200	.323	.230
厂家	3	1.922	.149	.929	.073
误差	19	12.907			
合计	28				
校正的合计	27				

a. Computed using alpha = .05

II、将产地、厂家及其交互项对 S0 做双因素方差分析。由输出结果可知，GLM 的模型 P 值为大于 0.05 没有统计学意义。“产地”和“厂家”及其交互效应都对没有 S0 没有影响。

表 26：改性玄武岩——产地和厂家双因素对 S0 的 GLM 分析

Dependent Variable: S0(%)

变异来源	自由度	均方	F	P值	检验效能 ^a
校正的模型	8	12.921	1.236	.332	.410
常数项	1	128083.904	12248.599	.000	1.000
厂家*产地	1	13.803	1.320	.265	.194
产地	2	5.227	.500	.614	.120
厂家	3	1.699	.163	.920	.075
误差	19	10.457			
合计	28				
校正的合计	27				

a. Computed using alpha = .05

III、将产地、厂家及其交互项对车辙做双因素方差分析。由输出结果可知，GLM 的模型 P 值为小于 0.05 有统计学意义。但用于检验“产地”和“厂家”的对车辙的影响的 P 值都大于 0.05，没有显著差异。

表 27：改性玄武岩——产地和厂家双因素对车辙的 GLM 分析

Dependent Variable: 车辙

变异来源	自由度	均方	F	P值	检验效能 ^a
校正的模型	8	4432247.317	3.101	.020	.856
常数项	1	276486368.6	193.470	.000	1.000
产地*厂家	1	1263363.553	.884	.359	.145
产地	2	441754.938	.309	.738	.092
厂家	3	1828631.537	1.280	.310	.287
误差	19	1429092.588			
合计	28				
校正的合计	27				

a. Computed using alpha = .05

IV、将产地、厂家及其交互项对拉弯应变做双因素方差分析。由输出结果可知，GLM 的模型 P 值为大于 0.05 没有统计学意义。用于检验“产地”和“厂家”对拉弯应变影响的 P 值都大于 0.05，没有显著差异。

表 28：改性玄武岩——产地和厂家双因素对弯拉应变的 GLM 分析

Dependent Variable: 弯拉应变

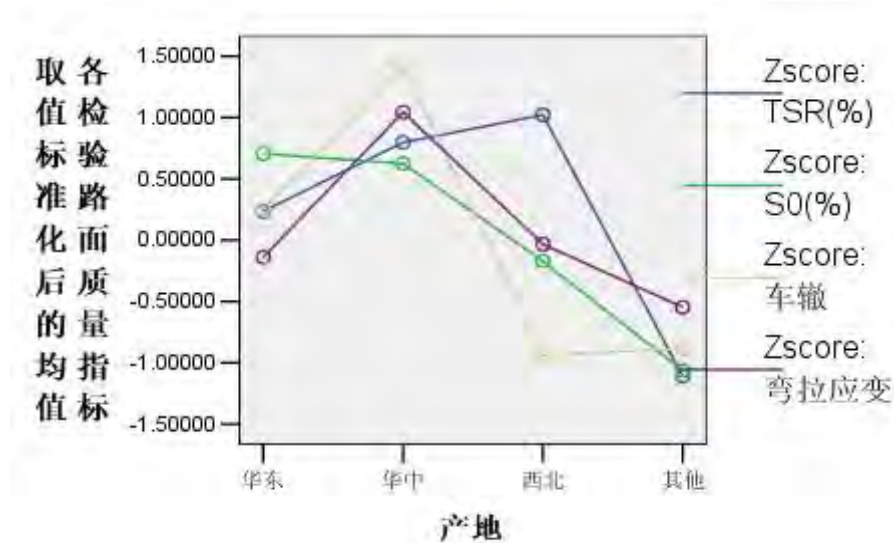
变异来源	自由度	均方	F	P值	检验效能 ^a
校正的模型	8	171820.565	.740	.656	.246
常数项	1	159776081.0	688.589	.000	1.000
产地*厂家	1	370999.976	1.599	.221	.225
产地	2	50526.240	.218	.806	.079
厂家	3	72410.706	.312	.816	.099
误差	19	232034.186			
合计	28				
校正的合计	27				

a. Computed using alpha = .05

(3) 改性其他岩性

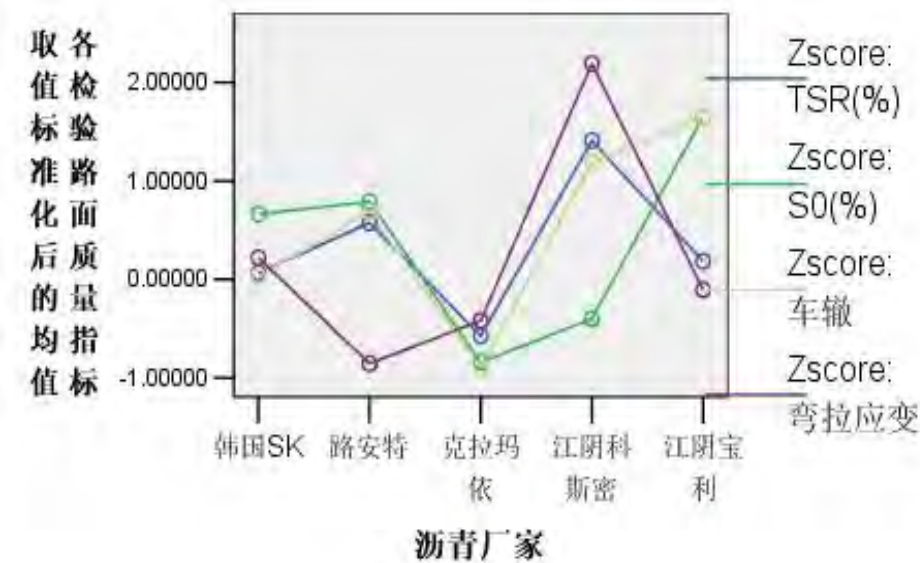
由于划分到此类的有效数据只有 9 个，如此少的数据已经不适宜合作双因素方差分析，因此仅能比较均值，获得粗略的观察。由下图观察发现，华中地区的各指标整体水平都很高，华东地区次之，而国外地区的各指标整体水平都很低。

图 3：改性其他岩性——不同产地的各路路面质量检验指标标准化均值比较图



由下图观察发现，江阴科斯密厂家除 S0 抗水性指标以外的各项路面质量检验指标都很高，克拉玛依的各项指标都很低。韩国 SK 公司的指标表现一般。

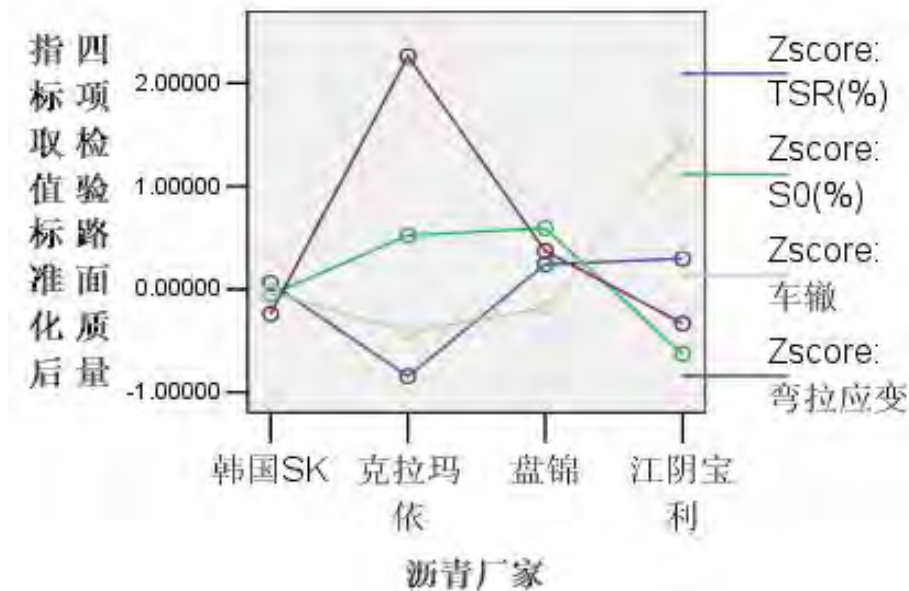
图 4：改性其他岩性——不同厂家的各路面质量检验指标标准化均值比较图



(4) 普通石灰岩

此时有 23 个有效数据，有 19 个数据的厂家是韩国 SK，20 个数据的地区是来自华东和华北地区。根据数据条件首先对韩国 SK 和克拉玛依的四项指标作图形比较。

图 5：普通石灰岩——不同厂家的各路面质量检验指标标准化均值比较图



再对华东地区和华北地区的四项指标作单因素方差检验，如下表，各项指标的 Levene' s 方差齐检验的 P 值分别为 0.106、0.604、0.764、0.361 均大于 0.05，可见方差是齐的，因此观察方差齐时的 T 检验的 P 值。可见，TSR 和拉弯应变的 P 值为 0.02 和 0.027，拒绝原假设，即华东地区和华北地区的 TSR 抗水性指标和抗低温指标有显著差异。

表 29：华东地区和华北地区的四项路面质量检验指标的均值 T 检验

		方差齐次检验		T检验				
		F	P值	t	自由度	P值(双尾)	均值差别	标准差差别
TSR(%)	方差齐时	2.888	.106	-2.550	18	.020	-3.80296	1.49136
	方差不齐时			-2.420	12.302	.032	-3.80296	1.57133
S0(%)	方差齐时	.279	.604	-1.968	18	.065	-2.08303	1.05854
	方差不齐时			-1.927	15.474	.072	-2.08303	1.08070
车辙	方差齐时	.093	.764	-.431	18	.672	132.37287	307.44331
	方差不齐时			-.427	16.632	.675	132.37287	309.89470
弯拉应变	方差齐时	.880	.361	-2.404	18	.027	119.63207	49.76888
	方差不齐时			-2.339	14.835	.034	119.63207	51.13911

5.5 问题五

TSR、S0、车辙、弯拉应变这四个指标测试的是混合料的质量，即公路建造原材料质量。这只能作为衡量路面质量的一部分。路面质量除了受原材料质量影响外，应该路面设计、施工质量等方面的影响。

1. 抗水性指标的改进。不论是TSR抗水性指标还是S0抗水性指标，在分析第二问各路面质量指标对与抗水性能指标的逐步回归结果都显示出常数项过大，见如下模型，抗水性能指标不能很好的被十个衡量路面质量的指标所解释，这可能是因为现有的TSR和S0不适合我国的实际情况。作为检验路面质量重要方面的指标，必须加以改善。

$$\begin{cases}
 \text{TSR} = 275.267 * (\text{Constant}) - 1.965 * \% \text{Gmm}(\text{最大}) + 0.903 * \text{粗细比率} \\
 \text{S0} = 216.398 * (\text{Constant}) - 1.809 * \% \text{Gmm}(\text{最初}) + 3.745 * \text{油石比} - 1.515 * \text{矿料间隙率} + 12.171 * \text{最大理论密度} \\
 \text{车辙} = 12897.827 * (\text{Constant}) + 304.623 * \text{矿料间隙率} + 143.752 * \text{饱和度} - 402.187 * \% \text{Gmm}(\text{最初}) + 4305.258 * \text{最大理论密度} \\
 \text{拉弯应变} = -11241.6 * (\text{Constant}) + 192.322 * \text{饱和度} + 1116.017 * \text{空隙率} - 291.756 * \text{矿料间隙率} + 108.86 * \text{粗细比率}
 \end{cases}$$

2. 检验路面质量其他方面指标的建立。除了这四个指标以外，公路质量还可有其他方面衡量，比如耐用性，抗疲劳性等，这些指标有待开发。在逐步回归时，有的质量指标例如空隙率、毛体积密度、粉胶比等都没有对解释四项检验路面质量指标起到较好的作用，但是作为公路质量的重要方面，有一定需要有新的路面质量检验指标来衡量特定的方面。

3. 原材料的质量控制指标。在第四问中沥青、碎石质量对高速公路路面抗水损害性能、高温性能、低温性能的影响分析中发现原材料对路面质量有着重要影响，因此考虑原材料产地、厂家、型号等因素的综合指标，对把握好质量控制的源头有着至关重要的作用。

4. 施工质量主要就是原材料质量和施工队伍对施工指标的检测严格度，原材料质量除了上述四个指标外，应该还包括直接衡量路面寿命的疲劳抗裂、抗老化以及衡量路面质量的抗滑指标。而检测严格度是个主观性指标，只能要求公路建设部门严把质量关，严格遵守部颁标准要求办事。

5. 指标体系的完善。一个好的路面质量检验指标体系应该是可以全面反映公路质量各个方面，而且这个指标体系应该是有效的，各指标应该各尽其能反映公

路质量的不同方面，在第一问指标之间的相互关联中我们发现，同样是衡量抗水性能的指标TSR和S0之间存在着较大的相关性无可厚非，但是却都不能利用现有的路面质量指标，可见其缺乏效率。

六、模型的评价

本文解决了利用路面质量指标和检验路面的指标，解决了公路改进的问题：主要的优点有：

1. 灰色关联度分析较好的描述两序列相似程度有效识别两序列间的局部相似性；
2. 运用逐步回归得到了精确的路面检验指标与路面质量指标之间的数量关系；
3. 经过对问题的抽象得到了理论上的级配比和混合料本身颗粒大小成对应关系的结论；
4. 运用 GLM 广义线性模型进行方差分析，不仅讨论了指标之间的显著性区别，还给出了具体差异大小；
5. 运用 EM 缺失值填补方法，填补缺失值，最大限度地避免了缺失值带来的信息损失。

存在的问题：

- 1、灰色关联度是一种模糊的概念，至今没有详细完整的理论支撑，故而它只能作为一种参考指标；
- 2、但是理论模型的假设条件偏强，现实中混合料的颗粒形状不可能完全是球形，所以现实情况还有待商榷。

参考文献

- [1] 张文彤, SPSS11 统计分析教程: 北京, 北京希望电子出版社, 2002. 321-328
- [2] 张少中, 章锦文, 张志勇, 韩美君, 王秀坤, 面向大规模数据集的贝叶斯网络参数学习算法, 计算机应用, 2006. 7: 1689-1691
- [3] 关于灰色关联度计算方法的研究 李宏艳 山东工商学院数学学院 系统工程与电子技术 2004年9月第26卷第9期
- [4] 何晓群, 刘文卿, 应用回归分析: 北京, 中国人民大学出版社, 2001. 126- 129
- [5] 赵芳林, 牛雅莉, 应用逐步回归分析的方法建立教学评估模型, 科技信息, 2007. 3: 141-123, 2007
- [6] 康爱红, 杨荣臻, 严城, 窦如初, 最大密度和最大嵌结力集料级配的设计, 华东公路 2003. 5
- [7] JTG F 40-2004 公路沥青路面施工技术规范 (附条文说明)

附录

问题一的程序:

```
function resul=hguanlian(x0,x1)
n0=length(x0);
n1=length(x1);
if n0~=n1
    disp('The size of x0 &x1 must be same');
end
x2=zeros(1,n0-1);
x3=zeros(1,n0-1);
kthe=zeros(1,n0-1);
for i=1:n0-1
    x2(i)=x0(i+1)-x0(i);
    x3(i)=x1(i+1)-x1(i);
end
m=n0-1;
for i=1:m
    if x2(i)~=0
        omega=x3(i)/x2(i);
        kthe(i)=omega*(omega*omega+1)/(omega^4+1);
    elseif x2(i)==0&x3(i)~=0
        kthe(i)=0;
    elseif x2(i)==0&x3(i)==0
        kthe(i)=0;
        m=m-1;
    end
end
end
if n0==1
    kthel=1;
elseif n0>1
    kthel=sum(kthe)/m;
end
resul=kthel;

function res=lhgl(x)
res12=hguanlian(x(:,1),x(:,2));
res13=hguanlian(x(:,1),x(:,3));
res14=hguanlian(x(:,1),x(:,4));
res23=hguanlian(x(:,2),x(:,3));
res24=hguanlian(x(:,2),x(:,4));
res34=hguanlian(x(:,3),x(:,4));
res=[res12 res13 res14 res23 res24 res34];
```

问题一的算法流程图：

