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。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。

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，对这批数据中所涉及的指标解释如下：

1．油石比。为混合料中沥青与集料（即矿粉、碎石）的质量比，被普遍认为是影响路面性能的关键指标。

2．筛孔通过率。即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（显然筛孔直径越大则通过的混合料就越多），是反映集料粗细程度及大小搭配情况的指标。例如通过直径为 0.6mm 圆孔的集料部分占集料的百分比可能是 10.1%；通过直径为 4.75mm 圆孔的集料部分占集料的百分比可能是 43.9%；通过直径为 19mm 圆孔的集料部分占集料的百分比可能是 99%.

3．VV：空隙率。为混合料经碾压后达到设计压实状态时，其中空隙体积占总体积的百分比。

4．VMA：矿料间隙率。为混合料经碾压后达到设计压实状态时，混合料中“有效沥青(指进入路面的沥青，而非施工中使用的沥青原料的全体)体积＋空隙体积”占总体积的百分比。

5．VFA：饱和度。等于(VMA-VV)/VMA，单位是％，反映有效沥青体积占 VMA 体积的百分率。

6．DP：粉胶比。混合料中矿粉质量与沥青质量的比。

7．毛体积密度。为达到设计压实状态的试件的密度。

8．最大理论密度。为混合料理论上的最大密度，即如果不含任何空隙时混合料的密度。VV＝100－毛体积密度/最大理论密度×100。

9．%Gmm(最初)。混合料在最初压实状态的压实度。压实度＝试件密度/最大理论密度×100。压实状态分为最初压实状态、设计压实状态和最大压实状态。最初压实状态的压实度为经过少数几次碾压后的压实度。

10．%Gmm(最大)。混合料试件成型时达最大压实状态时的压实度。一般应达到或略超过设计压实度。

关于高速公路路面质量的试验指标有以下四种，解释如下：

1．TSR（%），冻融劈裂强度比，反映混合料抗水损害性能的指标，越大越好。

2．S0(%)。浸水马歇尔稳定度比，反映混合料抗水损害性能的指标，越大越好。上述两种测试方法和测试指标经常同时采用，从数据看二者不完全一致。

3．车辙。动稳定度，为车辙试验中试件变形稳定时，荷载作用次数与变形深度的比值，单位是“次/mm”，反映混合料的抗车辙性能（高温性能），越大越好。

4．弯拉应变。为低温小梁弯曲试验时的破坏应变。反映混合料的低温变形

能力，越大越好。

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？如果有数量关系，建立它们之间的数学模型。

问题二．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我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，陈述选择这些因素的理由。根据模型讨论采用什么样的方案可以提高高速公路路面的质量。

问题三．若从理论上探讨集料的筛孔通过率与路面压实度的上界（数学意义下的上界）之间的数量关系，有何见解？

问题四．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，建立自己的观点，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，建立自己的观点，根据又是什么？

问题五．根据对数据分析的结果，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？对高速公路建设部门有什么建议？

二、问题分析

问题一：要求验证 TSR（%）、S0(%)、车辙和弯拉应变四个路面质量指标是否存在数量关系，并在存在数量关系的基础上，建立这四个指标之间相互关系的模型。在做实际分析的时候，首先应该做变量的散点图，确认变量分布特征后再进行统计分析。

作相关分析的方法有多种，首先想到的是考虑变量之间是否线性相关，线性相关分析有以下几种：Pearson 相关分析，但该方法的前提是要求变量服从正态分布；Spearman 秩相关函数与 Kandall 偏秩相关函数，适合于定序变量或不满足正态分布假设的等间隔数据。并且 Kandall 偏秩相关函数是在两个变量之间存在相关性的前提下，考虑两个变量的相关是否是由于第三个变量的存在。如果 Pearson 相关分析或 Spearman 秩相关函数分析结果表明，四个变量两两之间互不相关，则就没有使用 Kandall 偏秩相关函数的必要。而且这三个方法只适合用于两个序列之间的相关分析。

上述的三种相关性分析有可能出现这样的结果，四个变量两两互不相关。如果出现这样的情况，在考虑采用多元线性回归分析方法，分析一个变量与其余变量之间的线性关系。这样的分析方法也可能存在非线性相关的情况，进一步采用非线性分析方法，一般采用对数或者指数的非线性分析方法。

问题二：建立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，并分析选择那些因素的原因，同时还要给出提高高速公路路面的质量的方案。在建立模型之前，首先要分析“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”与路面质量之间的关系，因为“有较大影响或有影响的因素”是十个随机变量，路面质量的试验指标有四个随机变量 ，研究他们之间的关系，可以采用典型相关性分析。

首先因为典型相关性分析针对的是多元变量对多元变量之间的相关关系，而本题就属于此情况，而且在分析的过程当中，通过对典型相关系数的显著性检验，可以区分因素组中那些为重要因素，哪些是比较重要的因素，那些因素的影响可以忽略不计。在本题中，典型相关性分析就是研究两组随机变量因素组（油石比，

筛孔通过率，VV 空隙率，VMA 矿料间隙率等）与指标组（TSR 冻融劈裂强度比，SO 浸水马歇尔稳定度比，车辙和弯拉应变）之间相关关系。

在典型相关性分析之后，将重要因素与比较重要的因素提取出来，然后分析指标值与这些因素之间的关系。在进行相关行分析的过程当中，相关系数确定，由此可以确定指标值与因素变量之间的模型。

问题三：理论上探讨集料的筛孔通过率与路面压实度的上界之间的数量关系，并给出自己的见解。在做本问题之前，首先需要做一定的假设，假设混合料中小砂石是圆球形的，且都为半径相同的圆球形（半径有一定的上限限制）。问题就转化为在一个合适的密闭容器内进行砂石填充问题，也就是如何填充，才能使密度最大，也就是说填充后使得该密闭容器里的空隙最小，我们考虑到著名的“开普勒猜想”球体填充问题，本问题可以在此基础上作进一步理论上的探讨。

问题四要求做两部分的工作：

问题四（1）：这是一个定性分析问题。首先将沥青按照类型分为两类，型号为 25 的普通沥青和其他型号的改性沥青。碎石质量按照石料岩性的不同进行分类，然后将沥青类型与岩石类型进行组合，可以得到多种不同的组合，根据不同的组合，比较高速公路路面质量的四个指标变量之间的参数（例如，均值，方差），也可以采用区间的方法进行比较，目前一种有效的比较样本的是否存在显著差异是通过秩和检验法作的。这种方法简单

问题四（2）：这个问题是对问题四（1）的进一步细化，他把沥青的型号也考虑进来，在沥青型号、种类和根据岩石岩性分类的矿石质量进行组合，对于每种组合，考虑生产厂家与石料产地的不同与路面质量的影响。同样可以采用问题四（1）的比较方法进行比较，但是由于生产厂家和石料产地多而且基本各异。可以简单通过观察数据解决。最简单的方法就是采用均值比较法。

问题五要求分析问题的数据结果，并有根据的讨论测试高速公路路面质量的试验项目中有没有遗漏，并根据文中的结果的讨论和分析给高速公路建设部门写出一份建议。

三、模型假设

（1）假定问题中所给的数据真实可靠。

（2）由于沙石的体积不可能任意小，所以我们规定最小砂石的形状无差异，一律认为是半径相同球体。

（3）不考虑气候分区对本问题的影响

四、符号说明

$X_1, X_2 \cdots X_{10}$：表示影响高速公路路面质量的 10 个因素变量（油石比，筛孔通过率，VV 空隙率，VMA 矿料间隙率等）

$X_{11}, X_{12}, X_{13}, X_{14}$：表示高速公路路面质量的 4 个指标变量（TSR，SO，车辙和弯拉应变）

R：表示指标变量之间的 Pearson 相关系数

R_1：表示变量组 $X^{(1)}=(X_1,X_2,X_3,\cdots X_{10})$ 之间的 $X^{(2)}=(X_{11},X_{12},X_{13},X_{14})$ 相关矩阵

r：文中假设的小砂石的（球体）的半径

N：文中假设的小砂石的（球体）的个数

L：密闭正方体容器的边长

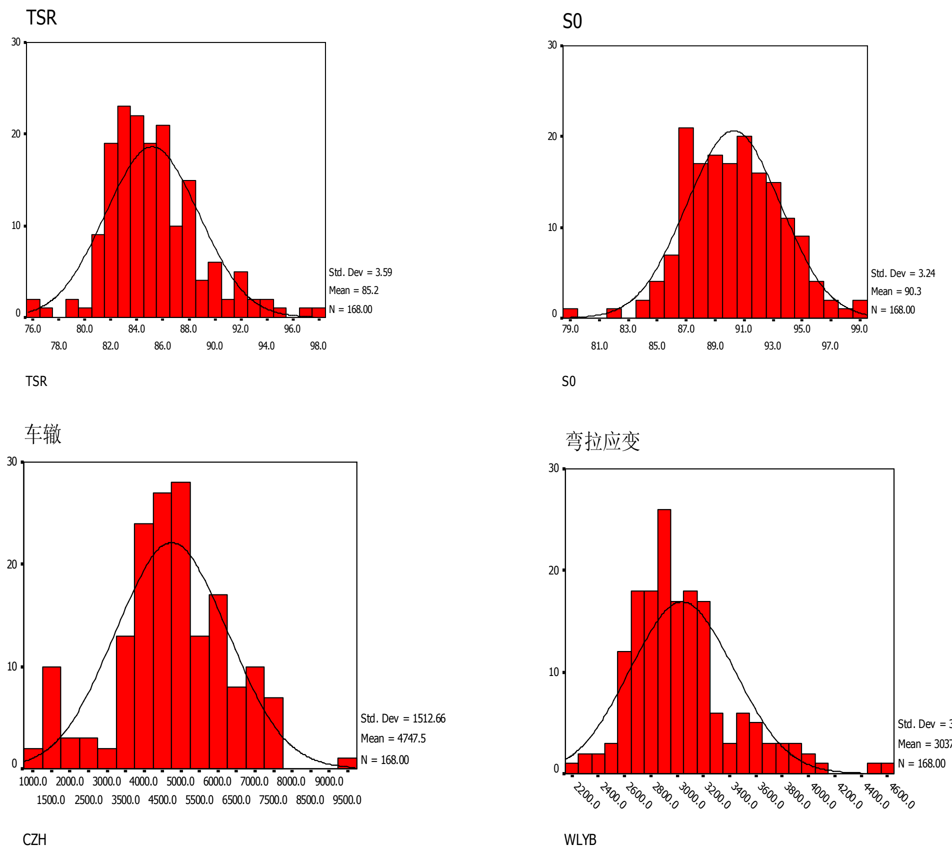
$A_{11},A_{12},A_{21},A_{22}$：表示沥青，碎石质量的不同组合，如：$i=1,2$ 表示普通沥青与改性沥青，$j=1,2$ 表示石灰岩与玄武岩。A_{11} 表示组合为：普通沥青，石灰岩

五、模型建立及求解

5.1 问题一：描述描述高速公路路面质量的的四个指标之间有没有数量关系？如果有数量关系，建立它们之间的数学模型。

5.1.1 指标变量正态分布的检验

在分析影响高速公路路面质量的四个指标变量满足某种分布时，首先分别画出四个指标变量的直方图如图（1）所示。



图（1）四个指标变量的二位直方图和正态分布曲线

因为正态分布随机变量是较广泛的存在的，从图（1）观察可以看到，四个

指标变量的数据大致落入一定区间之中，分布有可能呈正态性，但是还要考虑这四个指标变量（$X_{11},X_{12},X_{13},X_{14}$）是否真的服从正态分布，这也是对四个指标变量进行相关性分析的必要前提，有效的正态分布检验法有两种“峰度、偏度检验法”和“夏皮罗——威尔科法”[2]。

这里采用峰度、偏度检验法。峰度、偏度检验法适合于样本容量大于 100，对于本问题，有二百多个样本数据，显然适合采用此方法。通过在（$\alpha=0.01$）下进行检验，四个指标变量（$X_{11},X_{12},X_{13},X_{14}$）为正态分布的假设成立，可以认为数据是来自正态分布的总体。均值和方差见下表（1）所示：

X_{11}	0.4197	0.1663
X_{12}	0.5598	0.1654
X_{13}	0.4396	0.1784
X_{14}	0.3422	0.1712

表（1）指标变量的均值与方差

5.1.2 Pearson 线性相关分析：

由 5.1.1 中的分布检验可知四个指标变量（$X_{11},X_{12},X_{13},X_{14}$）都服从正态分布，而 Pearson 线性相关系数分析的前提就是要分析的变量服从正态分布。因此可以用 Pearson 线性相关系数描述两个变量之间的相关性。

Pearson 线性相关系数模型：

$$R=\frac{\sigma_{xy}^{2}}{\sigma_x\bullet\sigma_y}=\frac{\dfrac{\sum\left(x-\overline{x}\right)\left(y-\overline{y}\right)}{n}}{\sqrt{\dfrac{\sum(x-\overline{x})^2}{n}}\bullet\sqrt{\dfrac{\sum(y-\overline{y})^2}{n}}}=\frac{\sum\left(x-\overline{x}\right)\left(y-\overline{y}\right)}{\sqrt{\sum(x-\overline{x})^2}\bullet\sqrt{(y-\overline{y})^2}} \tag{1}$$

经过一定简化后得到公式：

$$R=\frac{n\sum xy-\left(\sum x\right)\left(\sum y\right)}{\sqrt{n\sum x^2-\left(\sum x\right)^2}\bullet\sqrt{n\sum y^2-\left(\sum y\right)^2}} \tag{2}$$

其中：R 为相关系数，n 为变量中数值的个数，X 与 Y 为两个随机变量。

对于本题数据，用 SPASS 统计软件处理后的结果如表（2）所示：

	STR	SO	车辙	弯拉应变
STR	1.000	0.304	0.149	0.100
SO	0.304	1.000	0.300	0.202

车辙	0.149	0.300	1.000	0.202
弯拉应变	0.100	0.202	0.309	1.000

表（2）Pearson 线性相关系数表

观察数据可以发现，相关系数一般都比较小，这说明四个指标之间的线性相关性非常小，通不过显著性检验，可以忽略不计。同时与采用 Spearman 秩相关函数作出的结果进行比较，发现所得数据结果大致一样，由此，不必要再进行 Kandall 偏秩相关函数的分析，同时可以得出以下重要结论。

结论一：四个路面质量指标（TSR（%）、SO(%)、车辙和弯拉应变）两两之间不是与低于 0.05 的概率线性相关的。

该结论说明，已知单个指标变量不能线性导出另一个指标变量的值。这也从一方面验证了题目中这样的陈述：两种测试方法和测试指标经常同时采用，从数据看二者不完全一致。

5.1.3 多元线性回归方法[4]：

多元线性回归模型的一般形式为：

$$y = a_0 + a_1x_1 + a_2x_2 + \cdots a_px_p + \varepsilon \qquad (3)$$

写成矩阵行式： $\vec{y} = XA + \vec{\varepsilon}$

其中：$a_0, a_1, \cdots a_p$ 是 $p+1$ 个未知参数，称为回归系数，y 为因变量，$x_1, x_2 \cdots x_p$ 是个可以精确测量的并可控制的一般变量。

多元线性回归主要描述一个变量与多个变量之间的线性关系，该模型的建立需要以下假设前提：

（1）自变量 $x_1, x_2 \cdots x_p$ 是确定性变量，且要求 $rk(X) = p+1 < n$，即 X 是满秩矩阵。从结论一可知，本题中 X 中的变量互不线性相关，由此可以保证 X 是一个满秩阵。满足假设（1）的要求。

（2）随机误差具有 0 均值和同方差，即

$$\begin{cases} E(\varepsilon_i) = 0 \\ \operatorname{cov}(\varepsilon_i, \varepsilon_j) = \begin{cases} \sigma^2 & i = j \\ 0 & i \neq j \end{cases} \end{cases}$$

其中：$i, j = 1, 2, \cdots n$，表明随机误差项 ε 的协方差假定表明误差项在不同样本点之间是独立的，不存在序列相关。

在本题中：采用对四个路面质量指标（$X_{11}, X_{12}, X_{13}, X_{14}$）进行多元线性回归分析，解得回归系数矩阵如下：

0.229 0.283 0.05316 0.2508

第一项为随机数，所作的是 TSR 与其它三个指标之间的线形关系。

对回归系数进行显著性检验，在可信度为 95%的清况下假设成立，这说明之四个指标之间只存在很弱的线性相关，即由其中的三项指标值无法通过线性计算得出第四个指标值。由此得出以下重要结论。

结论二：每个路面质量指标与其它三个指标是以低于的概率（0.05）线形

相关的。

这个结论说明，在测得三个指标的数值后，无法经过线性推导得出另外一个指标值，这也从一定的客观程度上说明衡量高速公路的路面指标质量时，四个指标都要进行测量的原因。

其次再考虑他们之间的非线性关系，一般采用对数或者指数关系去验证变量之间是否满足非线性关系。另外的做法就是采用非线性多项式拟合。经实际的计算，发现多项式中高次方的系数不为零。一般意义上讲，多项式拟和存在高次方的系数不为零时基本没有实际研究的意义。所以在文中关于这方面的内容不作过多的讨论。

5.2 问题二：建立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，并分析选择那些因素的原因？同时给出提高高速公路路面的质量的方案。

由问题二的分析，我们可以知道本题的目标就是要经过典型相关性分析[3]，研究因素组（油石比，筛孔通过率，VV 空隙率，VMA 矿料间隙率等，在文中用变量 $X_1, X_2, \cdots X_{10}$ 表示）与指标组（TSR 冻融劈裂强度比，SO 浸水马歇尔稳定度比，车辙和弯拉应变，在文中用变量 $X_{11}, X_{12}, X_{13}, X_{14}$ 表示）之间的相关关系。并通过计算出来典型相关系数的显著性检验，得出影响公路路面与重要因素和比较重要因素之间的数学模型。

5.2.1 数据处理：

首先在做典型相关性之前，由于数据之间量纲不统一，因此需要对数据进行无量钢化处理，这样可以使数据在公平的原则上进行处理，避免量纲不统一造成不切实际的结果。无量处理的一般方法是："标准化"处理法；极值处理法；功效系数法。在本文中，标准化处理采用"标准化"处理法，即：$x_{ij}^{*} = \dfrac{x_{ij} - \min x_i}{\max x_i - \min x_i}$

其中 x_{ij}^{*} 为标准观测值，$\min x_i$ 为第 i 项变量的最小值，$\max x_i$ 为第 i 项的最大值。其中 $i = 1, 2, \cdots 14, j = 1, 2, \cdots n$ 。

其次是对同过筛孔百分率的刻画，筛孔大小不同，通过的百分率不同。为了把此因素也像其它因素一样同等处理。我们采用两个随机变量处理，即用它的均值和方差来表征。也就是把 X_2 分为 X_2^1 和 X_2^2 两个随机变量来处理。

5.2.2 典型相关性分析的步骤：

要比较的两组随机变量：因素组 $X^{(1)} = (X_1, X_2, X_3, \cdots X_{10})'$，指标组 $X^{(2)} = (X_{11}, X_{12}, X_{13}, X_{14})'$，则原始资料矩阵为：

$$X=\begin{bmatrix}X^{(1)}\\X^{(2)}\end{bmatrix}=\begin{bmatrix}x_{11} & x_{12} & x_{1n}\\ x_{21} & x_{22} & x_{2n}\\ \vdots & \vdots & \vdots\\ x_{10,1} & x_{10,2} & x_{10,n}\\ x_{11,1} & x_{11,2} & x_{11,n}\\ \vdots & \vdots & \vdots\\ x_{14,1} & x_{14,2} & x_{14,n}\end{bmatrix} \quad (4)$$

（1）、计算相关矩阵

将矩阵分解为 $R_1=\begin{bmatrix}R_{11} & R_{12}\\ R_{21} & R_{22}\end{bmatrix}$，其中 R_{11}，R_{22} 分别为 $X^{(1)}$ 和 $X^{(2)}$ 的自相关系数阵，

$R_{12}=R_{21}^{'}$ 为第一组与第二组变量的互相关函数阵。

做两组变量的线性组合即：

$$U=l_1X_1+l_2X_2+\cdots l_{10}X_{10}\triangleq l^{'}X^{(1)} \quad (5)$$

$$V=m_1X_{11}+m_2X_{12}+m_3X_{13}+m_3X_{14}\triangleq m^{'}X^{(2)} \quad (6)$$

其中：$l=(l_1,l_2\cdots l_{10})^{'},m=(m_1,m_2,m_3,m_4)^{'}$ 为任意非零常数向量。

可以求出：$\rho_{UV}=Cov(U,V)=\dfrac{l^{'}R_{12}m}{\sqrt{l^{'}R_{11}l}\sqrt{m^{'}R_{22}m}}$

我们希望寻求 l 和 m，使得 ρ_{UV} 最大。

求得相关系数矩阵比较大，将其放在附录一中。

（2）、求典型相关系数及典型变量

根据数学分析中条件极值的求法引入 lagrange 乘数法求解结果如下：$\widehat{A}l=\lambda^2 l,\widehat{B}m=\lambda^2 m$。其中 $\widehat{A}=R_{11}^{-1}R_{12}R_{22}^{-1}R_{21}$，$\widehat{B}=R_{22}^{-1}R_{21}R_{11}^{-1}R_{12}$，求 $\widehat{B}=R_{22}^{-1}R_{21}R_{11}^{-1}R_{12}$ 的特征值和相应的特征向量（因为 $\widehat{B}$ 比 $\widehat{A}$ 的阶数低），从而得到典型的相关系数和典型变量如表（3）所示：

序号	典型相关系数	典型变量的系数
1	0.5822	2.0107 -2.0271 -0.4586 -0.3063 0.0249 0.5150 0.2517 -0.6707 0.0958 -0.1046 -0.3917 0.5173 0.2491 -0.2961 -0.7974
2	0.5007	-2.5361 1.2726 -1.0644 -0.1972 -0.6597 0.4217 0.3225 -0.3242 0.0321 0.5195 -0.1580 0.3087 -0.3858 -0.7608 0.2595
3	0.3899	1.9371 -0.6287 -1.0996 -0.0998 1.9543 -0.2043 0.0388

		0.2347 -0.6405 0.7176 0.0201 0.3583 -0.9019 0.7243 -0.1430
4	0.2551	-0.1305 -0.3780 -1.4576 0.2568 1.4122 -0.2088 0.6245 -0.2249 0.1456 -0.4582 0.2284 -0.7298 -0.3808 0.0151 -0.5299

表（3）典型相关系数与典型变量

（3）、典型系数的相关性检验

检验$\widehat{\lambda}_1$，在本题中$p_1=10, p_2=4$，为此计算 ：

$$\begin{cases} \Lambda_0=(1-\widehat{\lambda}_1^2)(1-\widehat{\lambda}_2^2)(1-\widehat{\lambda}_3^2)(1-\widehat{\lambda}_4^2) \\ \ln\Lambda_0 \\ Q_0=-\left[n-1-\dfrac{1}{2}(p_1+p_2+1)\right]\ln\Lambda_0 \end{cases}$$

查$p_1\times p_2$个自由度的χ^2分线个布表得：例如$\chi_\alpha^2(44)\approx\begin{cases}60.481 & \sigma=0.05\\ 68.701 & \sigma=0.01\end{cases}$

比较$\chi_{0.05}^2$与$\mathrm{Q_i}$的值，当$\mathrm{Q_i}\succ\chi_{0.05}^2$时说明存在典型相关系数与典型变量，否则不存在。以此类推其它典型相关系数的检验。

在本问题中，没有对数据进行相关整理的情况下，既不对影响因素进行分类的前提下，典型相关系数的计算结果如表（4）所示：

Λ_i	$\mathrm{Q_i}$	df(自由度）	$\chi_{0.05}^2$（临界值）
0.3927	53.2776	44	60.481
0.5941	29.6802	30	43.772
0.7928	13.2315	18	28.8693
0.9349	3.8346	8	15.5073

表（4）典型相关系数显著性检验表

由表（3）表（4）可知：

与 TSR 主要相关的因素是：油石比，筛孔通过率，矿料间隙率，DP 和毛体积密度。

与 SO 主要相关的因素是：油石比，筛孔通过率，矿料间隙率，Gmm（最初），Gmm（最大）。

与车辙主要相关的因素是：油石比，筛孔通过率，矿料间隙率，Gmm（最初），Gmm（最大）。

与指标弯拉应变主要相关的因素是：油石比，筛孔通过率，矿料间隙率。

由此可知：油石比，筛孔通过率，矿料间隙率同时影响四个试验指标。有理由认为公路路面质量的最重要的因素为油石比，筛孔通过率，矿料间隙率。其它为比较重要的因素。

采用多元线性回归得到它们之间的模型如下：

$$
\begin{aligned}
x_{11} &= 0.259x_1 - 0.370x_2^1 + 0.152x_2^2 + 0.222x_4 - 0.183x_6 + 0.101x_7 \\
x_{12} &= 0.524x_1 + 0.156x_2^1 - 0.135x_2^2 - 0.267x_4 + 0.373x_8 - 0.316x_9 \\
x_{13} &= 0.249x_1 + 0.754x_2^1 - 0.0725x_2^2 + 0.253x_4 + 0.849x_8 + 1.381x_9 \\
x_{14} &= -0.985x_1 - 0.548x_2^1 + 0.488x_2^2 - 0.425x_4
\end{aligned} \tag{7}
$$

5.2.3 改进的模型

由于样本的范围很广，其中包括十几个厂家和产地，不同类型和型号。因此增加了样本数据的随机性，降低了样本数据的规律性。因此上面的模型建立的数学关系式并不是很明显。某些对产品质量有着重要影响 但题中并没有给出的因素给样本数据带来很大的“噪声”。例如产品的石料材质，并导致模型准确性下降。

为提高模型的准确性，可以通过降低这些因素的影响实现。根据石料岩性的不同，将其分为石灰岩与玄武岩，分别建立模型，找出因素和产品指标的关系。

计算方法和第一种模型一样，只是数据的范围作了限定（产品的石料材质相同）。

得到典型的相关系数和典型变量如表（5）所示：

序号	典型相关系数	典型变量的系数
1	0.9131	4.0645 -4.1648 -0.0482 -0.0144 0.0000 0.0683 0.0080 -0.1509 0.0004 -0.0006 -0.0300 0.0692 0.0077 -0.0130 -0.2535
2	0.8506	-3.8041 1.9089 -1.5966 -0.2958 -0.9895 0.6326 0.4838 -0.4863 0.0481 0.7793 -0.2370 0.4630 -0.5787 -1.1412 0.3892
3	0.5839	2.9057 -0.9431 -1.6494 -0.1497 2.9314 -0.3065 0.0582 0.3520 -0.9607 1.0764 0.0302 0.5374 -1.3529 1.0865 -0.2145
4	0.5002	-0.1958 -0.5670 -2.1864 0.3852 2.1183 -0.3132 0.9368 -0.3373 0.2184 -0.6873 0.3426 -1.0947 -0.5712 0.0227 -0.7949

表（5）改进模型的典型相关系数与典型变量

典型相关系数的计算结果如表（6）所示：

Λ_i	Q_i	df(自由度）	$\chi^2_{0.05}$(临界值）
0.3927	82.9264	44	60.481
0.5941	46.1973	30	43.772
0.7928	20.5947	18	28.8693
0.9349	5.9686	8	15.5073

表（6）典型相关性的显著性检验结果

采用多元线性回归得到它们之间的模型如下：式（8）为石灰岩类型，（9）为玄武岩类型。

$$\begin{aligned}
x_{11} &= 0.191 + 0.049x_1 + 0.318x_2^1 - 0.115x_2^2 + 0.043x_5 - 0.116x_7 - 0.036x_8 \\
x_{12} &= 0.538 + 0.551x_1 + 0.318x_2^1 - 0.479x_2^2 - 0.468x_5 + 0.567x_9 - 0.318x_{10} \\
x_{13} &= -0.428 + 0.702x_1 + 1.029x_2^1 - 0.196x_2^2 + 0.429x_8 - 0.433x_9 + 0.108x_{10} \\
x_{14} &= 0.704 + 0.183x_1 - 0.972x_2^1 + 0.612x_2^2
\end{aligned} \tag{8}$$

$$\begin{aligned}
\mathrm{x}_{11} &= 0.105\mathrm{x}_1 + 1.655\mathrm{x}_2^1 - 1.458\mathrm{x}_2^2 + 0.840\mathrm{x}_4 \text{-} 0.535\mathrm{x}_6 + 0.363\mathrm{x}_7 \\
\mathrm{x}_{12} &= 1.311\mathrm{x}_1 - 2.391\mathrm{x}_2^1 + 2.499\mathrm{x}_2^2 \text{-} 0.960\mathrm{x}_4 + 0.697\mathrm{x}_8 \text{-} 0.684\mathrm{x}_9 \\
\mathrm{x}_{13} &= 0.151\mathrm{x}_1 + 0.135\mathrm{x}_2^1 \text{-} 0.454\mathrm{x}_2^2 + 0.373\mathrm{x}_4 \text{-} 0.05\mathrm{x}_8 + 0.206\mathrm{x}_9 \\
\mathrm{x}_{14} &= 0.102\mathrm{x}_1 + 0.757\mathrm{x}_2^1 \text{-} 0.390\mathrm{x}_2^2 + 0.589\mathrm{x}_4
\end{aligned} \tag{9}$$

采用分类处理的方法进行典型相关性检验和多元线性回归，分类后的检验结果较为显著，而且再进行多元线性回归后的模型更为精确。这三种模型的检验结果见附录三。

5.3 问题三：理论上探讨集料的筛孔通过率与路面压实度的上界之间的数量关系，并给出自己的见解。

5.3.1 沙石填充

首先考虑如何装箱：

在装箱时，一般的方法就是先装最大的沙石，再装次大的沙石，依次类推，最后，除了最小的沙石之外，其他沙石都已经装入密闭容器的清况下，此时密度的大小就受最小沙石的半径和排列的制约了。这种现象在现实中是存在的。如果沙石的半径可以任意小，则密闭容器在理论上是可以装满的。这时探讨集料的筛孔通过率与路面压实度的上界就没有必要了。因为密闭容器内的空隙总会被任意小的沙石填满，密度就会非常贴近于 1。问题是沙石粒不可能任意小。因此密闭容器的空隙也就不可能被任意填充。也就说最小沙石的半径与密度的关系极为密切。

其次确定最小沙石半径：

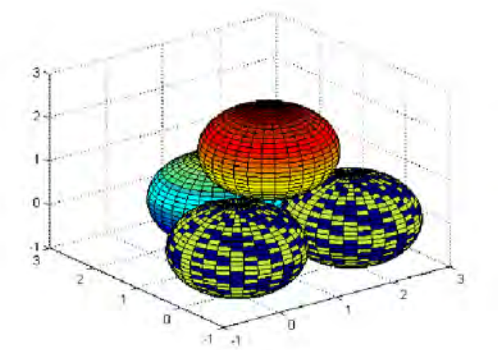
在确定沙石半径时，除了最小的沙石半径要考虑外，不考虑其他的沙石形状。最小的沙石半径及数量通过以下处理确定：最小的圆孔直径为 0.075mm,而且由数据知道当圆孔直径为 0.075mm 的平均筛孔通过率。此时将 0.075mm 作为最小沙石的直径。

现在除最小沙石之外所有的石块都已装入的情况下，密闭容器内的空隙就由最小沙石来进行填充，问题是这在理论上是可能实现的。现在剩下的问题就可用“开普勒猜想”来处理了。

5.3.2 开普勒猜想

在对这个问题的分析上，将问题归结为密闭容器的沙石填充问题。我们想到了著名的“开普勒猜想”，“开普勒猜想”即球体填充问题，它是一个数学问题：设箱子的容量为L，球的半径为r，求得数量为N，那么有$\left(N\times4\pi r^{3}\div3\right)\div L\prec1$，式子的左边看成为密度，球体填充问题就是要这个密度的上确界。他考虑得“最密”装箱方法：一个长为 2 的正方体，分别以他的八个顶点及六个面的中心为球心以（$\frac{\sqrt{2}}{2}$）为半径作球，因此在这个正方体内，球的体积便有 4 个整球的体积。八个角，每个角有 8 分之一个球，六个面，每个面有半个球，密度如下：

$$\left(4\times4\pi(\frac{\sqrt{2}}{2})^{3}\div3\right)\div2^{3}=\pi\div3\times\frac{\sqrt{2}}{2}=0.74048$$



图（2）球体在容器里的排列分布图

也就是说，开普勒认为球体装箱的密度上确界为$\frac{\pi}{18}$，并以$\frac{\pi}{18}$为最密。

按照开普勒设计的排列方式，就是先在下面排好，然后其它沙粒放到下面排好的空隙上，比如说：将三个球体排成一个三角形的，第四个球体放到三个球体组成的空隙上，这样排下去，密度为最大。

设最小沙石所占的比例为$a\%$，则它按照开普勒设想的装箱方法填充，按照正四面体（或者说金字塔形）的方式进行排列。可以测的不同的比例下，可以得出不同的密度：

$$\%\mathrm{Gmm}=1-a\%\times\left(1-0.74048\right)$$

该式的意义是指，最小沙粒比例所占开普勒空隙率的比例，也就是说，空隙被最小沙石占去的单位意义下的空间。以下是几组不同的$a\%$对应的密度值。

a	4	6	8	10
$\%Gmm$	98.96	98．44	97.92	74.48

5.4 问题四：检验沥青、碎石质量对高速公路路面质量的影响，检验不同厂家、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质量究竟有没有影响，根据结果建立自己的观点。

首先对石料岩性的数据进行观察，发现石料岩性分为石灰岩，玄武岩，石灰岩与玄武岩的结合，安山岩，砂岩，花岗岩，石榴石等，但是除了石灰岩与玄武岩外，其他石料岩性的数据都比较少，不足以通过它们去检验对高速公路路面质量的影响，因此为了简化问题，并使分析切实可信，所以不对数据量较少的石料岩性进行分析。只考虑石灰岩，玄武岩这两种石料岩性。

其次考虑沥青的分类，由于题目中已经对沥青进行了分类（数据表格中混合料类型中含 25 的是普通沥青，其他数字如 13、16 或 20 的是改型沥青），所以就采用题目中的分类方法，将沥青简单分为普通沥青和改型沥青两种。

有了碎石质量（石料岩性）与沥青的分类，又观察附件一中的数据， 可以将沥青与碎石质量进行组合。有四种组合：普通沥青，石灰岩；普通沥青，玄武岩；改性沥青，石灰岩；改性沥青，玄武岩。分别用 $A_{11}, A_{12}, A_{21}, A_{22}$ 表示。同过分类可以很清晰地对数据进行比较处理，在分析的基础上进行总结讨论。

5.4.1 秩和检验

如何进行数据之间的比较，以确定不同的沥青，碎石质量对高速公路路面质量的影响。秩和检验是最强的非参数检验之一，而且是一种有效的、且使用方便的检验方法，它一般被用来检验两个样本有无显著的质量差异。

但是秩和检验有一个重要前提，就是两个样本之间相互独立。且两个样本大致服从同一类型的分布。在问题一的求解过程当中，我们已经得出四个指标变量（$X_{11}, X_{12}, X_{13}, X_{14}$）均服从正态分布，由正态分布的性质可知：正态分布上不同各点之间是相互独立的 。因此可以对组合不同的质量指标进行秩和检验。

秩和检验的一般步骤：

（1）求出两个样本的总体均值 μ_1和μ_2，检验的假设是

$$H_0: \mu_1 = \mu_2, H_1: \mu_1 \neq \mu_2$$

（2）将数据按照从大到小的次序排列，得到对应于 n_1 个样本的秩和 r_1。

（3）当 H_0 为真时，求出

$$E(R_1) = \frac{1}{2} n_1 \left(n_1 + n_2 + 1 \right)$$

$$D(R_1) = \frac{1}{12} n_1 n_2 \left(n_1 + n_2 + 1 \right)$$

当 H_0 为真时近似有：$R_1 \sim N\left(E(R_1), D(R_1) \right)$

拒绝域为：$\dfrac{\left|r_1 - E(R_1)\right|}{\sqrt{D(R_1)}} \geq z_{0.025} = 1.96$

由秩和检验的一般步骤可以对四种不同组合分别进行两两之间的秩和检验，用 matlab 编程结果如表（7）——表（10）:

TSR	石灰岩，普通沥青	石灰岩，改性沥青	玄武岩，普通沥青	玄武岩，改性沥青
石灰岩，普通沥青	—	2.2021	1.4502	3.0336
石灰岩，改性沥青	2.2021	—	0.8051	1.2838
玄武岩，普通沥青	1.4502	0.8051	—	0.5528
玄武岩，改性沥青	3.0366	1.2828	0.5528	—

表（7）不同组合之间关于 TSR 指标的差异程度

从数据上来看，只要数据小于 1.96，说明通过了检验，也就是说这两种组合之间对于该目标无显著差异。但是一旦数据大于 1.96，这说明假设不成立，且这两个组合关于该项指标是有显著差异的。例如：石灰岩普通沥青，石灰岩改性沥青之间的 2.2021 已经大于 1.96，这说明石灰岩普通沥青与石灰岩改性沥青之间是有显著差异的。

同样也可得到关于其他指标变量的不同组合之间的关系：

SO	石灰岩，普通沥青	石灰岩，改性沥青	玄武岩，普通沥青	玄武岩，改性沥青
石灰岩，普通沥青	—	0.2753	1.6734	0.2053
石灰岩，改性沥青	0.2753	—	0.85557	0.5338
玄武岩，普通沥青	1.0734	0.8557	—	0.9214
玄武岩，改性沥青	0.2053	0.5338	0.9214	—

表（8）不同组合之间关于 SO 指标的差异程度

观察可知，数据全没有经过秩和检验，这说明不同沥青碎石质量关于 SO 这个指标都是没有显著差异的。也就是说不同的结合并不影响高速公路路面质量的 SO 指标值。

车辙	石灰岩，普通沥青	石灰岩，改性沥青	玄武岩，普通沥青	玄武岩，改性沥青
石灰岩，普通沥青	—	5.6678	0.2196	7.2398
石灰岩，改性沥青	5.6678	—	1.6533	4.0719
玄武岩，普通沥青	0.2196	1.6533	—	2.0768

玄武岩，改性沥青	7.2398	4.0719	2.0768	—

表（9）不同组合之间关于车辙指标的差异程度

弯拉应变	石灰岩，普通沥青	石灰岩，改性沥青	玄武岩，普通沥青	玄武岩，改性沥青
石灰岩，普通沥青	—	—	—	—
石灰岩，改性沥青	—	—	—	2.08014
玄武岩，普通沥青	—	—	—	—
玄武岩，改性沥青	—	2.08014	—	—

表（10）不同组合之间关于弯拉应变指标的差异程度

表中存在“—”说明无数据，在做关于弯拉应变的秩和检验中，由于数据不足，所以无法检验不同组合的显著差异情况。

由这四组表中的数据观察可以得出如下结论 ：

对于石灰石$\begin{cases}\text{普通沥青}\\\text{改性沥青}\end{cases}$，他们在关于指标 TSR 和车辙来说存在显著性差异。也就是说石灰石与普通沥青组合或者石灰石与改性沥青组合可以影响高速公路路面质量的两个指标值：TSR 和车辙。

关于玄武岩$\begin{cases}\text{改性沥青}\\\text{普通沥青}\end{cases}$，他们之间的组合仅仅关于指标值车辙存在显著性的差异。

同样对于普通沥青$\begin{cases}\text{石灰岩}\\\text{玄武岩}\end{cases}$无论关于那个指标值都无显著性差异。这也说明，只要使用了普通的沥青，无论采用何种岩石，对路面的质量影响应该没有太大差别。

改性沥青$\begin{cases}\text{石灰石}\\\text{玄武岩}\end{cases}$关于指标值车辙和弯拉应变的显著性差异比较大。这也说明在采用改性沥青的情况下，使用不同的岩石料主要影响路面质量的车辙和弯拉应变指标。

也可明显地分析到，使用改性沥青比普通沥青对路面质量的影响要大。

5.4.2 考虑不同厂家、不同岩料产地对路面质量的影响

与问题四（1）的分析相似，沥青不仅分为普通沥青和改性沥青，还要根据沥青的型号进行分类，观察数据沥青的型号有 70#，90#，110#，SBS 大致四种型号，还有多种厂家和石料产地等。如果将沥青型号、混合料分类、岩石岩性生

产厂家和石料产地进行自由组合分别讨论，这将是一个非常耗费人力和时间的事。而且对不同组合进行分析的方法相同。因此全面地去考虑将是不太可能的事，而且也没有这样考虑的必要，因为只要分析一种或两种组合，能够得出不同厂家，不同产地的相同型号和类型的沥青和碎石质量与高速公路路面的抗水损害性能、高温性能、低温性能是否有影响即可。

在实际解决问题时，我们可以挑选 70#的改型沥青，碎石质量的岩石岩性为石灰岩，固定这个组合，然后讨论不同厂家，不同石料产地对高速公路路面质量的不用指标的影响。

最直观的比较方法就是采用均值相比法：设一种生产厂家与石料产地的数据有 c 组，另外一组厂家与石料产地的数据有 d 组，比较组织不同的两组数据，就是看两组数据指标的平均值，看不同厂家不同产地但型号相同、类型相同的沥青碎石质量对路面质量的影响，就可以直观通过均值比较得出结论。

通过比较足够多的生产厂家与石料产地在路面质量上的差异，看是否遵循一定的规律。

5.5 问题五：分析数据结果，讨论测试高速公路路面质量的试验项目中有没有遗漏，给高速公路建设部门写出一份建议。

给高速公路建设部门的建议

尊敬的领导：

鉴于目前高速公路在国家政治、交通，经济、国防等各方面的重要性，我们对高速公路的路面质量评价指标和影响路面质量的各个因素进行讨论分析，找出了影响高速公路路面质量只要因素和非主要因素，同时也刻画了质量指标和重要因素与次重要因素之间的比较精确的模型。

同时还根据沥青（普通沥青与改性沥青）与碎石质量（石灰岩与玄武岩）的不同组合分析对路面质量的影响，得到一些结论。更进一步的，针对不同厂家，不同岩料产地生产的相同型号的沥青和碎石质量，与路面质量指标进行分析，讨论了生产厂家与岩料产地对路面质量指标的影响。结合我们建立的模型整理分析讨论给出以下几条建议：

建议（1）：检测高速公路路面质量的四个指标没有存在明显的关系，因此对路面质量进行检测时，四个指标需要一起测量。

建议（2）：影响高速公路路面质量的各个指标与影响质量指标的因素进行分析时，可以建立比较精确的模型，可以根据改变模型中的参数来改变指标值，或者需要提高公路路面质量的某一指标时，从与该指标影响大的因素下手。

建议（3）：采用改性沥青的高速公路路面质量明显比使用普通沥青的德路面质量高。因此，可在成本允许的条件下，使用改性沥青来提高公路寿命。

在检测公路路面质量时，不仅与这些工艺质量配比存在很大关系，还与公路上行驶的车载重量有很大关系。不同的级别公路由不同的车载承受能力，如果车载重量超出公路的承受能力，势必会对公路的质量造成很大的影响。因此在进行路面检测时，车载承受能力也是一个重要的指标。

三个关心公路建设的学生

2007 年 11 月 22 日

六、模型的评价与改进

6.1 模型的改进

问题一中，已经证明高速公路路面质量的四项试验指标以概率 95%不存在线性关系。因此如何有效的找到四项试验指标的非线性关系成为本问的难点。通过寻找具有正交化的试验数据可以比较好的解决这个问题。即：对于大样本中找尽可能多的样本，其中除了两个随机变量的值不同外，其它随机变量的值相同或近似相同。因此可找到两两随机变量数学关系。通过这一关系最终找到他们的非线性关系。

6.2 模型的评价

模型的优点：

1、模型二中采用典型相关性分析，这种方法能够分析多元之间的相关性，而且能够突出自变量对因变量的影响。

2、模型三中运用“开普勒猜想”，巧妙地解决了砂石填充与充实度的理论上界问题。

模型缺点：

把最小石块作为球体处理，不可避免的产生误差。

七、参考文献：

[1] http://www.jpszx.com/read.php?u=&tid=19781，论文撰写时的统计学问题.，2004.

[2] 盛骤等，概率论与数理统计，第三版，高等教育出版社，2003 ,P250~P253

[3] 于秀林等，多元统计分析，第一版，中国统计出版社，2003，P216~P255

[4] 何晓群，现代统计分析方法与应用，第一版，中国人民大学出版社，P102~P123

[5] 苏金明，SPASS 统计软件应用及开发指南，电子工业出版社，2004

[6] http://www.mikekong.net/Maths/maths-frame.php，球体填充问题