Clean Coal and Energy 清洁煤与能源, 2018, 6(1), 1-10
Published Online March 2018 in Hans. http://www.hanspub.org/journal/cce
https://doi.org/10.12677/cce.2018.61001

Hans汉斯

Analysis of Factors Affecting Fuel Efficiency

Fengjiao Yi*, Haoda Wang

School of Mathematical Sciences, Ocean University of China, Qingdao Shandong
Email: *yifengjiao_yfj@163.com, wanghaodawhd@163.com

Received: Jun. 21st, 2018; accepted: Jul. 4th, 2018; published: Jul. 11th, 2018

Abstract

With the rapid development of the automotive industry, energy and environmental issues have followed, and it is of great significance to improve the fuel efficiency of automotive engines. This article is based on vehicle displacement, horsepower, vehicle length, vehicle weight and other indicators, using fuel efficiency as a dependent variable, and vehicle index as an independent variable to establish a regression model. Focusing on the problem of multicollinearity in the model, we try to use the variable selection method, stepwise regression method, principal component regression method, Ridge regression and Lasso, and partial least squares method to improve the common multiple linear regression model, and compare various methods. The advantages and disadvantages are selected from the best methods.

Keywords

Fuel Efficiency, Regression Model, Regression Diagnosis, Multicollinearity

燃油效率影响因素分析

伊凤娇*，王浩达

中国海洋大学数学科学学院，山东 青岛
Email: *yifengjiao_yfj@163.com, wanghaodawhd@163.com

收稿日期：2018年6月21日；录用日期：2018年7月4日；发布日期：2018年7月11日

摘　要

随着汽车行业的迅猛发展，能源、环境问题也随之而来，提高汽车发动机的燃油效率具有非常重要的意义。本文主要从汽车排量、马力、车长、车重等指标出发，以燃油效率为因变量，以汽车指标为自变量建立回归模型。重点针对模型中存在的多重共线性问题，尝试使用变量选择法、逐步回归法、主成分回

*通讯作者。

文章引用: 伊凤娇, 王浩达. 燃油效率影响因素分析[J]. 清洁煤与能源, 2018, 6(1): 1-10.
DOI: 10.12677/cce.2018.61001

归法、岭回归和Lasso以及偏最小二乘法对普通多元线性回归模型做了改进，并比较各种方法的优劣性，从中选取最优的方法。

关键词

燃油效率，回归模型，回归诊断，多重共线性

1. 引言

燃油效率，定义为产生特定推力或马力使用的燃油所含的能量除以这份燃油所含的全部潜在能量，其评价指标为每加仑燃油或每千克燃油所能产生的公里数。

我国的石油消耗在过去 20 年里以每年 5%的速度增加[1]，然而石油作为一种不可再成能源，其总量却在持续减少，这更加剧了石油供给与需求之间的矛盾。近年来为了缓解石油供给的压力，除了寻找新的可替代能源之外从自身出发提高石油使用效率也成为了主要的研究问题。然而关于燃油效率的影响因素有哪些的问题，不同的时期，不同的学者做了不同的探讨[2] [3]。但是在研究过程中常常会出现所选因素相关性太强导致模型解释性很差的问题。本文主要从汽车本身的指标出发，探讨诸如排量、马力、车长、车重等对燃油效率的影响作用，同时消除模型的多重共线性，使得其更具有实际的解释价值。

2. 符号说明及数据预处理

2.1. 符号说明

本文选取了 32 种汽车的燃油效率数据(见表 1)，并对原始数据做了单位换算，使得更符合我们的习惯(ftp://ftp.wiley.com/public/sci_tech_med/introduction_linear_regression/)。

2.2. 数据预处理

观察数据可以发现，23 和 25 号汽车的扭力数据缺失(见表 2)，需要寻找合适的方法进行缺失值得填充。

通过原始数据的散点图矩阵(见图 1)可以发现变量间存在较强的相关关系，特别是 x_3 与 x_1、x_2，其散点图几乎在一条直线上，说明三者间具有很强的相关性，因此可以采用回归的方式填补缺失值。

将数据分为两部分，一部分包含所有的完整的行，一部分由 23 号和 25 号这两行组成。使用第一部分完整的数据集，以 x_3 为因变量，以 x_1、x_2 为自变量，建立回归模型。然后将 23 号和 25 号的 x_1和x_2 带入回归方程，计算得到 x_3 作为缺失值的估计。回归模型结果如下：

该回归结果(见图 2)的 F 检验显著，t 检验也均显著，R 方非常大。表明 x_3 的 88.84%可以由 x_1、x_2 决定。但是模型中也存在问题，比如 x_1、x_2 高度相关，多重共线性问题非常严重，但是考虑到我们只是问了寻找一个精确的模拟出缺失值而不是为了得到一个精确的可以解释的模型，所以这个问题可以暂时忽略。

最终我们利用回归方程 $x_3 = 42.3457x_1 + 0.5232x_2$ 补足 x_3 的取值，保证了数据集的完整性。

3. 多元线性回归模型

首先建立变量的相关系数矩阵(见表 3)，通过相关系数矩阵可以更直观的看出 y 与

Table 1. Symbol description
表 1. 符号说明

符号	含义	单位
y	燃油效率	千米/升
x_1	排量	升
x_2	马力	牛顿米
x_3	扭力	牛顿米
x_4	压缩比率	
x_5	后轴比率	
x_6	总车长	米
x_7	车宽	米
x_8	车重	千克
x_9	变速器类型	1——自动；0——手动

Table 2. Description of missing values (NA indicates missing values)
表 2. 缺失值说明(NA 表示缺失值)

标号	y	x_1	x_2	x_3
23	5.87286	6.6	250.675	NA
24	11.2926	1.6	101.625	112.465
25	10.4076	2.3	116.53	NA

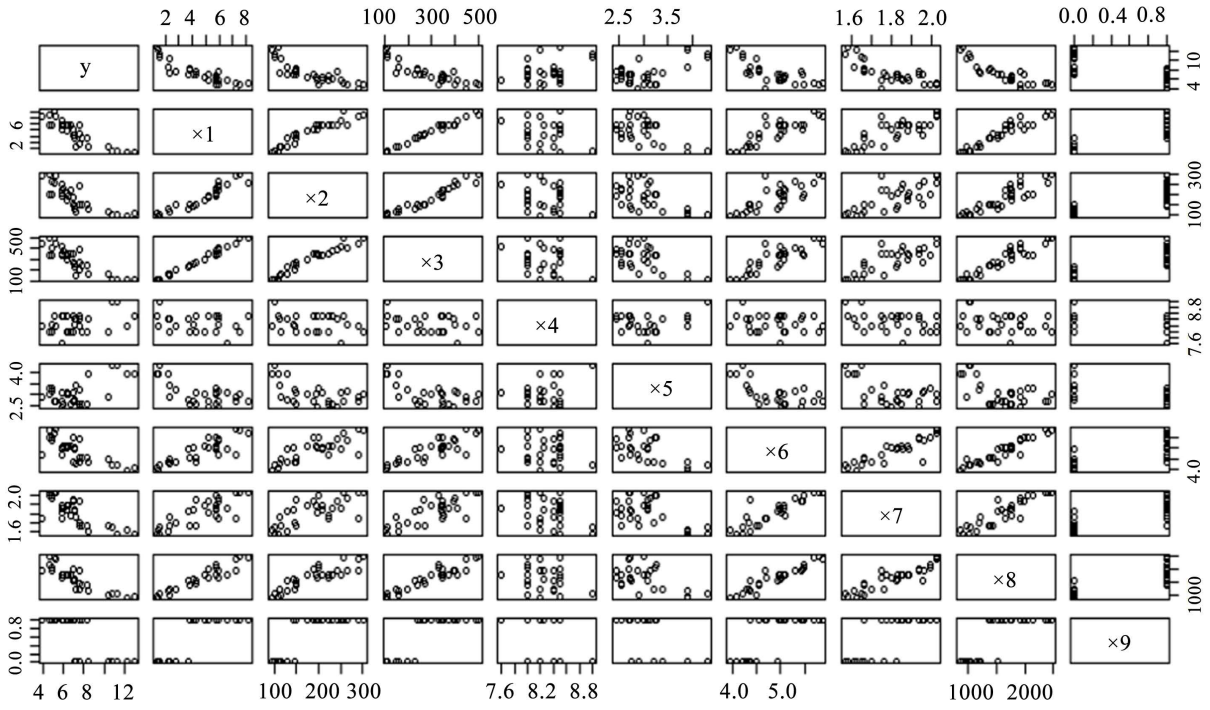

Figure 1. Scatterplot matrix
图 1. 散点图矩阵

```
Coefficients:
   Estimate Std. Error t value Pr(>|t|)
x1  42.3457     3.4052  12.436 6.39e-13 ***
x2   0.5232     0.0882   5.932 2.20e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 12.75 on 28 degrees of freedom
Multiple R-squared:  0.9985,    Adjusted R-squared:  0.9984
F-statistic:  9157 on 2 and 28 DF,  p-value: < 2.2e-16
```

Figure 2. Regression results (x_3 is dependent variable)
图 2. 回归结果(x_3 为因变量)

Table 3. Correlation coefficient matrix
表 3. 相关系数矩

	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
y	1.00	−0.88	−0.81	−0.86	0.36	0.59	−0.75	−0.77	−0.86
x_1	−0.88	1.00	0.95	0.99	−0.33	−0.63	0.86	0.80	0.95
x_2	−0.81	0.95	1.00	0.97	−0.29	−0.52	0.80	0.72	0.88
x_3	−0.86	0.99	0.97	1.00	−0.32	−0.63	0.86	0.79	0.94
x_4	0.36	−0.33	−0.29	−0.32	1.00	0.37	−0.26	−0.32	−0.28
x_5	0.59	−0.63	−0.52	−0.63	0.37	1.00	−0.55	−0.43	−0.54
x_6	−0.75	0.86	0.80	0.86	−0.26	−0.55	1.00	0.88	0.95
x_7	−0.77	0.80	0.72	0.79	−0.32	−0.43	0.88	1.00	0.90
x_8	−0.86	0.95	0.88	0.94	−0.28	−0.54	0.95	0.90	1.00

x_1、x_2、x_3、x_6、x_7、x_8 具有很明显的负相关关系。y 与 x_4、x_5 具有正相关关系。x_9 为定性变量，所以并没有计算。因此可以尝试拟合全模型。

根据回归结果(见图 3)，全模型的 F 检验显著，表明模型的线性关系显著，调整后的 R 方达到了 0.98，但是只有两个变量(x_6、x_8)通过了 t 检验，而且回归系数的符号与预期不符，比如从散点图矩阵和相关系数矩阵上可以得出 y 与 x_3 呈负相关关系，但是回归结果 x_3 的系数为正。

以上特点都是变量间具有多重共线性的经典特征，为了进一步验证模式是否存在多重共线性问题，计算了方差膨胀因子(见表 4)，发现其取值非常大，证实了这个问题。

此外，多重共线性在相关系数矩阵中也有所体现，例如 x_1 和 x_3 的相关系数竟然达到了 0.99，几乎是完全相关的。所以多重共线性成为了该模型中亟待解决的问题，否则该模型即使通过了显著性检验也无法进行很好的解释。

4. 多重共线性问题的解决

为了消除多重共线性，本部分分别使用了所有子集法、逐步回归法、岭回归和 lasso、主成分回归法、偏最小二乘回归法，并比较了集中方法的效果，从中选出了最优的方法。

4.1. 所有子集法

根据调整后的 R 方，选择了 R 方最大的三个自变量，分别是 x_5、x_6、x_8 重新拟合多元线性回归模型，计算回归系数和方差膨胀因子(见图 4)。

x_6 的回归系数符号仍然与实际相反， x_6、x_8 的方差膨胀因子依然很大(分别为 9.96 和 9.87)，变量选择后并没有消除多重共线性(见图 5)。

4.2. 逐步回归法

基于 AIC 和 BIC 的向后剔除法选择的结果都是 x_5、x_6、x_8，但是该变量子集已经被证明是无法消除多重共线问题的。基于 AIC 和 BIC 的向前选择法选择的结果都是 x_1，拟合一元线性回归并进行模型诊断后可以发现只保留一个变量(见图 6)。

```
Coefficients:
    Estimate Std. Error t value Pr(>|t|)
x1 -0.942790   0.857743  -1.099   0.2831
x2 -0.012419   0.016460  -0.754   0.4582
x3  0.020112   0.018654   1.078   0.2921
x4  0.474538   0.483155   0.982   0.3362
x5  1.231054   0.680639   1.809   0.0836 .
x6  2.994165   1.246072   2.403   0.0247 *
x7 -2.642442   2.895681  -0.913   0.3709
x8 -0.006104   0.002337  -2.612   0.0156 *
x9  0.527349   0.851102   0.620   0.5416
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.054 on 23 degrees of freedom
Multiple R-squared:  0.9858,    Adjusted R-squared:  0.9802
F-statistic:   177 on 9 and 23 DF,  p-value: < 2.2e-16
```

Figure 3. Regression results (y is dependent variable)
图 3. 回归结果(y 为因变量)

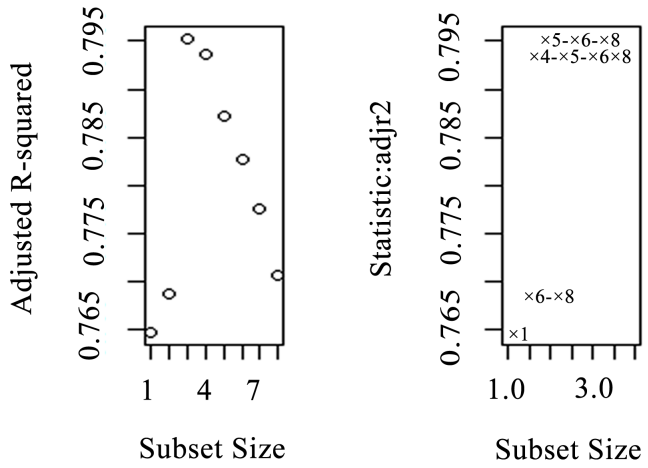

Figure 4. Plots of adj R^2 against subset size for the best subset of each size
图 4. 最佳子集的调整后 R 方

Table 4. Variance inflation factor
表 4. 方差膨胀因子

变量	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
VIF	539.1	296.4	994.5	461.8	127.9	1075.1	796.6	445.4	14.1

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.773875   3.991365   0.444   0.6601
x5           0.929261   0.425763   2.183   0.0376 *
x6           2.952884   1.098930   2.687   0.0120 *
x8          -0.007278   0.001327  -5.485 7.37e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.012 on 28 degrees of freedom
Multiple R-squared:  0.8151,    Adjusted R-squared:  0.7953
F-statistic: 41.14 on 3 and 28 DF,  p-value: 2.156e-10
```

Figure 5. Regression results of all possible subsets
图 5. 所有子集法回归结果

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  11.9378     0.5111   23.36  < 2e-16 ***
x1           -1.0223     0.1014  -10.09 3.74e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.085 on 30 degrees of freedom
Multiple R-squared:  0.7723,    Adjusted R-squared:  0.7647
F-statistic: 101.7 on 1 and 30 DF,  p-value: 3.743e-11
```

Figure 6. Regression results of stepwise methods
图 6. 逐步回归法回归结果

但是这样也有弊端，只保留一个变量，损失了其他信息。标准残差图中似乎蕴含着二次项关系，信息提取不充分，只用 x_1 不能完全解释 y (见图 7)。

4.3. 岭回归和 lasso

4.3.1. 岭回归

岭回归是一种专用于共线性数据分析的有偏估计回归方法，实质上是一种改良的最小二乘估计法，通过放弃最小二乘法的无偏性，以损失部分信息、降低精度为代价获得回归系数更为符合实际、更可靠的回归方法[4]。

最小二乘法求解回归系数的过程中需要考虑特征矩阵是否可逆的问题，引入岭回归就是为了解决这个问题。普通最小二乘法的回归系数矩阵的估计为 $\hat{B}=\left(X'X\right)^{-1}X'Y$,岭回归的参数估计为 $\hat{B}_k=\left(X'X+kI\right)^{-1}X'Y$

对变量做标准化(记为 sy, sx_1 等)处理后建模，最终选择了 6 个自变量，得到的回归结果为：

$$\text{sy}=-0.3549sx_1-0.0564sx_2-0.3197sx_3+0.0613sx_4+0.1040sx_5-0.1334sx_7-0.2422sx_8$$

用岭回归的方法得出的回归系数与实际预期相符合了。

4.3.2. Lasso

Lasso 算法改进了最小二乘法，在估计回归系数的同时可以达到变量选择的目的[5]。是受约束的最小二乘法，考虑 P 个自变量的回归模型，在 $\sum_{j=1}^{p}\left|\beta_j\right|\le s,\left(s\ge 0\right)$ 的约束条件下，使得残差平方和最小，

$$\min\sum_{i=1}^{n}\left\{y_i-\left(\beta_0+\beta_1x_{1i}+\cdots+\beta_px_{pi}\right)\right\}^2$$

使用拉格朗日乘数法

$$\min \sum_{i-1}^{n}\left\{y_i-\left(\beta_0+\beta_1 x_{1i}+\cdots+\beta_p x_{pi}\right)\right\}^2+\lambda\sum_{j=1}^{p}\left|\beta_j\right|, \lambda\geq 0$$

即：最小化残差的平方和加上对回归系数的绝对值的惩罚项。

当 s 非常大时，约束条件几乎不起作用，$\lambda=0$ 结果跟普通最小二乘法相同。

当 s 比较小时，λ 较大，回归结果中有的系数等于 0，我们就把等于 0 的系数对应的自变量删除，所以 Lasso 可以同时进行参数估计和变量选择。

K 折交叉验证是评价模型的一种常用方法，它把所有的观测数据大致分为 k 等份，然后轮流以其中的所有可能的 k−1 份为训练集，用来拟合数据，剩下的一份为测试集，一共计算 k 次，得到拟合测试集时的均方误差那样的 k 个指标再做平均，对于每个模型都做一遍，然后选择平均均方误差最小的模型[6]。根据交叉验证，最佳 λ 的简约模型是选择了 x_1 和 x_8 (见图 8)。

$$\mathrm{sy}=-0.5206sx_1-0.2287sx_8$$

Lasso 做参数估计的同时也起到了了变量选择作用，选出的简约模型保留了 x_1 和 x_8，符号也是正确的。

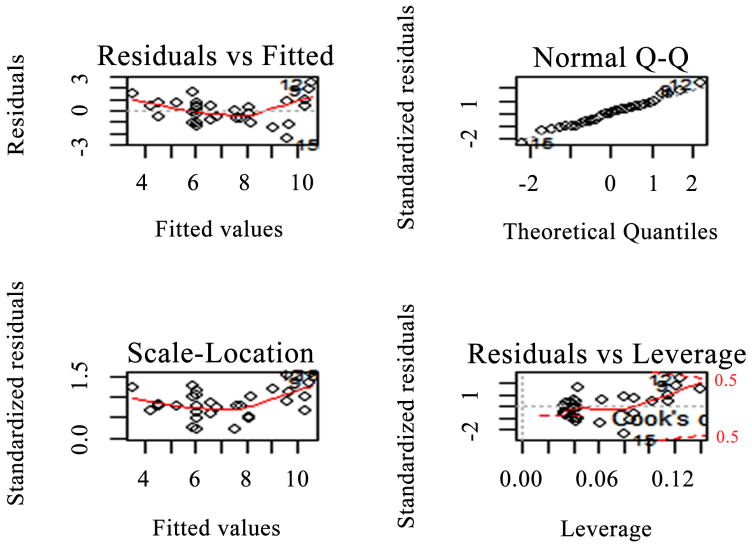

Figure 7. Model diagnostic plots
图 7. 模型诊断图

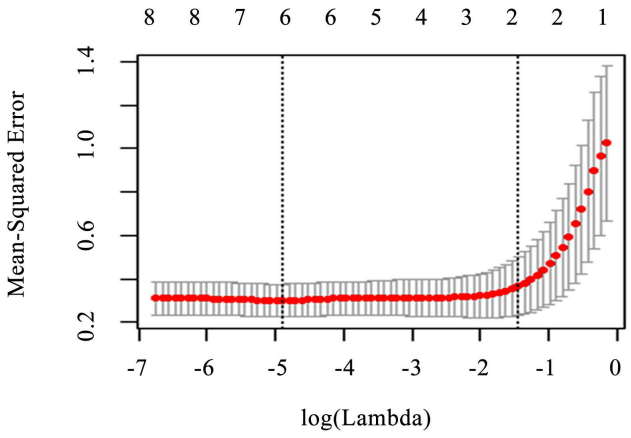

Figure 8. Cross validation
图 8. 交叉验证

4.4. 主成分回归法

4.4.1. 主成分回归

主成分回归是将主成分变量作为自变量，建立回归模型，具体的参数估计等跟普通最小二乘法无异，主成分只是原始变量的线性组合，组合后看作一些新的变量[7]。因为主成分是从标准化后的数据阵出发的，所以这里也先对数据做了标准化处理。

记主成分矩阵为 Z，原始变量矩阵为 X，因变量为 Y，则 $Z=XV$

建立 Y 和 Z 的回归方程为 $Y=Z\beta+\varepsilon$，估计的回归方程为 $\hat{Y}=Z\hat{\beta}$

再根据主成分与原始变量的关系，可以逆变换回带 $\hat{Y}=Z\hat{\beta}=XV\hat{\beta}\triangleq X\hat{\gamma}$，则建立了 Y 和 X 的回归方程，其中的回归系数 $\hat{\gamma}=V\hat{\beta}$。

为了消除单位和数量级的差异，对变量做了标准化处理，主成分分析(见图 9)和主成分回归(见图 10)结果如下：

由 $\hat{\gamma}=V\hat{\beta}$ 可求回归结果为：

$$\text{sy}=-0.1545sx_1-0.1481sx_2-0.1549sx_3+0.0616sx_4+0.1056sx_5-0.1465sx_6-0.0842sx_7-0.1526sx_8$$

4.4.2. 不完全主成分回归

不完全主成分回归删除了 V 中相应的有着很小的方差对回归贡献不大的的后几列，保留前 p 列，回归系数也只有 p 个了，回归系数的估计值为 $\hat{\gamma}_p=V_p\hat{\beta}_p$。

至于保留几列，p 等于多少，有不同的标准，需要谨慎对待。这里我们看到在主成分分析中，提取两个主成分时，累计贡献率达到了 81% (见图 11)，为了简约，我们就只保留了两个主成分。

```
Importance of components:
                          Comp.1    Comp.2     Comp.3     Comp.4     Comp.5     Comp.6      Comp.7      Comp.8
Standard deviation     2.2943695 1.0067608 0.86596811 0.66649632 0.46154271 0.21520760 0.116824833 0.07210470
Proportion of Variance 0.6792428 0.1307829 0.09676139 0.05731837 0.02748667 0.00597604 0.001761038 0.00067085
Cumulative Proportion  0.6792428 0.8100256 0.90678704 0.96410540 0.99159207 0.99756811 0.999329150 1.00000000

Loadings:
   Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8
x1 -0.418        -0.148 -0.158  0.148  0.517 -0.431  0.550
x2 -0.401        -0.120 -0.322  0.462 -0.599  0.295  0.243
x3 -0.419        -0.153 -0.149  0.230  0.100 -0.298 -0.792
x4  0.167  0.662 -0.682  0.251
x5  0.286  0.473  0.229 -0.778 -0.129        -0.128
x6 -0.396  0.164         0.105 -0.676 -0.457 -0.366
x7 -0.228  0.543  0.645  0.391  0.275
x8 -0.413  0.111        -0.130 -0.396  0.381  0.698
```

Figure 9. Principal component analysis
图 9. 主成分分析

```
Coefficients:
    Estimate Std. Error t value Pr(>|t|)
pr1  0.36954    0.03636  10.164 3.58e-10 ***
pr2  0.09799    0.08286   1.183   0.2486
pr3  0.11613    0.09633   1.205   0.2398
pr4  0.13191    0.12516   1.054   0.3024
pr5  0.05152    0.18074   0.285   0.7781
pr6 -0.81882    0.38762  -2.112   0.0452 *
pr7 -1.16037    0.71406  -1.625   0.1172
pr8 -1.01158    1.15692  -0.874   0.3906
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4719 on 24 degrees of freedom
Multiple R-squared:  0.8276,    Adjusted R-squared:  0.7701
F-statistic:  14.4 on 8 and 24 DF,  p-value: 1.866e-07
```

Figure 10. Principal component regression
图 10. 主成分回归

```
Coefficients:
    Estimate Std. Error t value Pr(>|t|)
pr1  0.36954    0.03900   9.476 1.58e-10 ***
pr2  0.09799    0.08887   1.103    0.279
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5061 on 30 degrees of freedom
Multiple R-squared:  0.7521,    Adjusted R-squared:  0.7356
F-statistic:  45.5 on 2 and 30 DF,  p-value: 8.214e-10
```

Figure 11. Incomplete principal component regression
图 11. 不完全主成分回归

回归结果为：

$$sy = -0.1565sx_1 - 0.1428sx_2 - 0.1554sx_3 + 0.1264sx_4 + 0.1519sx_5 - 0.1304sx_6 - 0.0310sx_7 - 0.1417sx_8$$

4.5. 偏最小二乘回归

偏最小二乘回归类似于主成分回归，主成分回归是在自变量中找到一些相互独立的主成分，主成分是原始变量的线性组合，用主成分代替原来的变量进行回归，以解决多重共线性问题。偏最小二乘回归则是先在因变量(如果有多个因变量)和自变量中各自寻找一个因子[8]，条件是这两个因子在其他可能的因子中最相关，然后在选中的这一对因子的正交空间中再选择一对最相关的因子，如此下去，直到这些因子有充分的代表性为止(可以用交叉验证)。

图 12 中，横坐标均表示自变量因子的个数。左图纵坐标表示 RMSEP (root mean squared error of prediction)，黑线表示留一法计算结果，红线表示调整后留一法计算结果，RMSEP 越小越好。中图纵坐标表示 MSEP (mean squared error of prediction)，黑线表示留一法计算结果，红线表示调整后留一法计算结果，MSEP 越小越好。右图纵坐标表示 R 方(R-squared)，R 方越大越好。

根据交叉验证的结果(见图 12)，无论是 RMSEP、MSEP 还是 R 方准则，都是因子个数为 2 时最好，但是一个因子和两个因子，三个准则几乎相差无几，为了简约，还是选择一个因子，回归结果如下：

$$sy = -0.1625sx_1 - 0.1492sx_2 - 0.1585sx_3 + 0.0660sx_4 + 0.11sx_5 - 0.1396sx_6 - 0.0594sx_7 - 0.1596sx_8$$

4.6. 比较

通过实例比较，我们发现在多重共线性严重存在的情况下，三种有偏回归方法(主成分回归；岭回归和 lasso，偏最小二乘回归)在建立模型和预测因变量方面优于普通回归模型，也优于变量选择和删除变量法。因为主成分是原始变量的线性组合，可以代表原始变量的信息，而且主成分之间没有相关性，所以可以避免多重共线性问题，偏最小二乘回归类似；岭回归和 lasso 对回归系数加入惩罚项，避免参数估计值过大，降低共线性的影响，lasso 的惩罚函数是绝对值形式，还可以剔除多重共线性强的变量。所有子集法和逐步回归法选择出的模型要么仍然存在多重共线性问题，要么只选择了 x_1 一个自变量，对 y 的信息解释不充分。删除变量法原则上是删除一些相关性大的、影响力小的自变量，但是缺乏客观的评价标准，需要专业的知识做支撑。

但这三种有偏方法建立的模型哪个更好，不同的问题和不同的判别标准，答案是不同的[9]。通常而言，在追求预测效果时，我们可以首先试用偏最小二乘回归方法，而要想对回归系数进行直观的控制时，可首选岭回归；而对某些综合因素特别关心需要把它拟合进回归方程时，可考虑选用主成分回归。不能偏心哪一种方法，回归效果要和实际情况相对照。

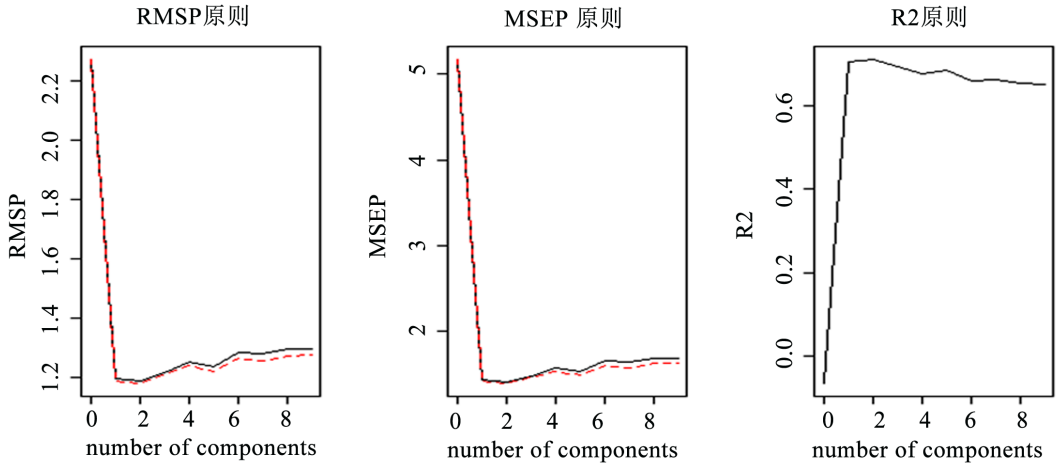

Figure 12. Partial least squares regression cross validation results
图 12. 偏最小二乘回归交叉验证结果

5. 结论

本文主要从汽车本身的指标出发，探讨诸如排量、马力、车长、车重等对燃油效率的影响作用。根据建模结果，燃油效率与排量、马力、扭力、车长、车宽、车重呈正相关关系，排量和车重对燃油效率的影响作用最大。压缩比率、后轴比率对燃油效率有正的影响，但是影响作用不大。变速器的类型未体现出对燃油效率的影响。另外，汽车的上述指标之间存在很强的相关关系，比如排量、马力、扭力之间，车长、车宽、车重之间相关系数几乎达到了 0.9，多重共线性是需要重点讨论的问题，通过五种方法的比较，主成分回归；岭回归和 lasso，偏最小二乘回归表现较好。

本文虽然得到了一些指导性的结论，但仍有很多不足之处。比如数据集来源于国外网站，没有纳入针对国内自主品牌汽车的数据，下一步，将考虑量化国产车诸多指标对燃油效率的影响。其次限于篇幅，本文未展开论述不同回归方法的理论内涵，将来会更加深入的探索每种方法的理论，并尝试优化改进算法。最后，在后续研究中，将尝试加入影响燃油效率的其他方面的因素，探讨诸如汽车先进节油技术、汽车保养状况、胎压、润滑油、驾驶操作规范水平等对燃油效率的影响。

参考文献

[1] 潘虹如. 燃油税改革: 小排量、新能源汽车受宠[J]. 标准生活, 2009(1): 41-43.

[2] 王登峰, 邓阳庆, 刘延林, 等. 汽车使用诸因素对燃油经济性的影响分析与试验研究[J]. 公路交通科技, 2008, 25(8): 150-153.

[3] 陈海涛. 汽车结构因素对燃油经济性的影响[J]. 公路与汽运, 2006(2): 1-4.

[4] 陈晓停, 曹兰杰, 汪金花. 岭估计在多项式曲面拟合 GNSS 高程中的应用[J]. 河北联合大学学报(自然科学版), 2016, 38(4): 1-6.

[5] 张秀秀, 王慧, 田双双, 等. 高维数据回归分析中基于 LASSO 的自变量选择[J]. 中国卫生统计, 2013, 30(6): 922-926.

[6] 龙泽海, 杨毅, 赵月丽. 基于 lasso 方法的银行对中小企业贷款供给意愿研究[J]. 金融与经济, 2017(3): 58-65.

[7] Massy, W.F. (1965) Principal Components Regression in Exploratory Statistical Research. *Publications of the American Statistical Association*, **60**, 234-256. https://doi.org/10.1080/01621459.1965.10480787

[8] 林育贤, 冯圣红. 基于 PCA 法的 BP 网络对冰蓄冷系统的空调负荷预测[J]. 建筑节能, 2016(7): 13-15.

[9] 龙泽海, 杨毅, 赵月丽. 基于 lasso 方法的银行对中小企业贷款供给意愿研究[J]. 金融与经济, 2017(3): 58-65.

知网检索的两种方式：

1. 打开知网页面 http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD
   下拉列表框选择：[ISSN]，输入期刊 ISSN：2334-332X，即可查询
2. 打开知网首页 http://cnki.net/
   左侧“国际文献总库”进入，输入文章标题，即可查询

投稿请点击：http://www.hanspub.org/Submission.aspx
期刊邮箱：cce@hanspub.org