9.3

November 20, 2021

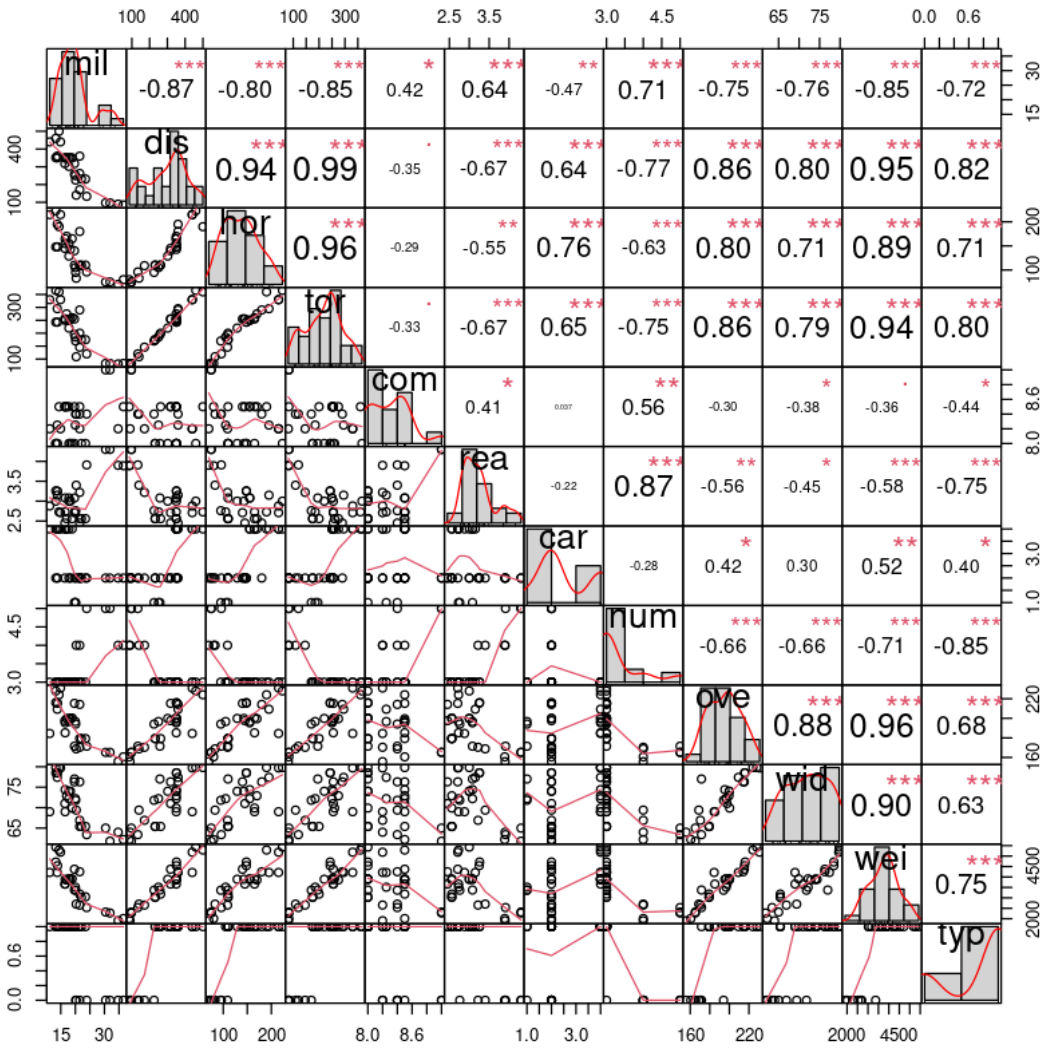
(1)

The correlation matrix of predictors and pairwise scatter plots are as follow:

[2]:
```
library(Hmisc)
```

[4]:
```
library(PerformanceAnalytics)
```

[7]:
```
data_gas<-read.table("table9.17.txt",
                     head = TRUE,
                     sep = "\t")
colnames(data_gas)<-c("mil","dis","hor","tor","com","rea","car","num","ove","wid","wei","typ")
chart.Correlation(data_gas, histogram=TRUE, pch=19)
```



From the plot above, high correlations are found between predictors, especially the correlation coefficients larger than 0.8.

[8]:
```
res <- cor(data_gas[2:12])
ev <- eigen(as.matrix(res))
cl <- sqrt(max(ev$values)/min(ev$values))
ev
```

```
eigen() decomposition
$values
 [1] 7.702574847 1.403077880 0.773435643 0.577055424 0.211498935 0.141941470
 [7] 0.095142049 0.050092536 0.033266309 0.008417705 0.003497202
```

```
$vectors
            [,1]         [,2]        [,3]         [,4]         [,5]         [,6]
 [1,] -0.3529639 -0.112431387  0.03114403 -0.006932422 -0.026272973 -0.09512815
 [2,] -0.3299718 -0.260762001  0.07836539 -0.194970349  0.142783457 -0.23889898
 [3,] -0.3510109 -0.139829772  0.04294522 -0.004153543  0.084990459 -0.18488343
 [4,]  0.1610427 -0.552726480  0.11863260  0.785849610 -0.096920435  0.09122188
 [5,]  0.2663779 -0.346997347 -0.43309789 -0.352178691 -0.516283052  0.07200995
 [6,] -0.2047881 -0.548146807  0.41844801 -0.380746710  0.007176897  0.38287792
 [7,]  0.3040550 -0.352222407 -0.22122179 -0.134117215  0.050372348 -0.57691563
 [8,] -0.3232988 -0.078466513 -0.36961713  0.180329365  0.200485930 -0.20407455
 [9,] -0.3026624  0.006019985 -0.54645511  0.094905101 -0.106514020  0.51959464
[10,] -0.3446125 -0.100475266 -0.26679114  0.040652506  0.028959499 -0.14008874
[11,] -0.3117090  0.181885175  0.24279993  0.119155548 -0.800493659 -0.27479473
             [,7]        [,8]        [,9]        [,10]        [,11]
 [1,]  0.26787382 -0.25888638 -0.49677393 -0.290946296  0.617904045
 [2,]  0.34910433  0.05057424  0.65243209  0.290811120  0.258528596
 [3,]  0.35518667 -0.06800437 -0.03290868 -0.466442937 -0.681570251
 [4,]  0.09287761 -0.06188507  0.06292276  0.051311641  0.012735988
 [5,]  0.06450059 -0.43886854  0.13804308 -0.086127357 -0.045372936
 [6,] -0.37681067  0.16574908 -0.13359309 -0.004651702 -0.059626414
 [7,] -0.02079064  0.55944398 -0.24949398 -0.055978181  0.049028663
 [8,] -0.67496023 -0.15486222  0.25287357 -0.294111256  0.091346835
 [9,]  0.19659254  0.52415223  0.01482782 -0.055178229  0.052597726
[10,] -0.06284718 -0.20261712 -0.39402290  0.714256660 -0.259679096
[11,] -0.16382124  0.22167146  0.06274209  0.017189710 -0.009773591
```

[9]:
```
cl
```

46.930758582079

The condition index calculated is larger than 15, which means there exists collinearity.

[12]:
```
k_2 <- c()
for (i in 1:length(ev$values)) {
  evn <- ev$values[1:i]
  k_2[i] <- sqrt(max(evn)/min(evn))
}
names(k_2) <- c("k1","k2","k3", "k4", "k5", "k6","k7","k8","k9","k10","k11")
```

[17]:
```
print(round(k_2, 2))
```

```
   k1    k2    k3    k4    k5    k6    k7    k8    k9   k10   k11
 1.00  2.34  3.16  3.65  6.03  7.37  9.00 12.40 15.22 30.25 46.93
```

The results show that K9, K10, and K11 are larger than 15 , which means there are three sets of collinearity. The following equation can be drawn:

$$\lambda_9 = 0.497\widetilde{X_1} - 0.652\widetilde{X_2} + 0.033\widetilde{X_3} - 0.063\widetilde{X_4} - 0.138\widetilde{X_5} + 0.134\widetilde{X_6} + 0.249\widetilde{X_7} -$$

$0.253\ \widetilde{X_8} - 0.015\widetilde{X_9} + 0.394\widetilde{X_{10}} - 0.063\widetilde{X_{11}}$
This equation can be simplified into: $+0.652\ \overline{X_2} + 0.138\widetilde{X_5} + 0.253\overline{X_8} = 0.497\overline{X_1} + 0.134\overline{X_6} + 0.249\widetilde{X_7} + 0.394\widetilde{X_{10}}$
X1, X2, X4, X5, X6, X7, X8 and X10 are varaibles involved in this set of collinearity.

$$\lambda_{10} = -0.291\widetilde{X_1} + 0.291\widetilde{X_2} - 0.466\widetilde{X_3} + 0.051\widetilde{X_4} - 0.086\widetilde{X_5} - 0.005\widetilde{X_6} - 0.056\widetilde{X_7} - 0.294\widetilde{X_8} - 0.055\widetilde{X_9} + 0.714\widetilde{X_{10}} + 0$$

This equation can be simplified into: $0.291\widetilde{X_1} + 0.466\widetilde{X_3} + 0.294\widetilde{X_8} = 0.291\widetilde{X_2} + 0.714\widetilde{X_{10}}$
X1, X2, X3, X8 and X10 are variables involved in this set of collinearity.

$$\lambda_{11} = 0.618\widetilde{X_1} + 0.259\widetilde{X_2} - 0.682\overline{X_3} + 0.013\widetilde{X_4} - 0.045\overline{X_5} - 0.060\widetilde{X_6} + 0.049\widetilde{X_7} + 0.091\widetilde{X_8} + 0.053\widetilde{X_9} - 0.260\widetilde{X_{10}} - 0.0$$

This equation can be simplified into: $0.682\widetilde{X_3} + 0.260\widetilde{X_{10}} = 0.618\widetilde{X_1} + 0.259\widetilde{X_2}$
X1, X2, X3 and X10 are varaibles involved in this set of collinearity.

[18]:
```
library("car")
```

```
Loading required package: carData
```

[20]:
```
lm <- lm(mil ~ dis + hor + tor + com + rea + car + num + ove + wid + wei + typ,
         data = data_gas)
step(lm,direction="both")
```

```
Start:  AIC=78.96
mil ~ dis + hor + tor + com + rea + car + num + ove + wid + wei +
    typ

       Df Sum of Sq    RSS    AIC
- typ   1     0.409 187.78 77.022
- car   1     0.580 187.95 77.049
- com   1     1.914 189.28 77.261
- hor   1     7.092 194.46 78.071
- wei   1     8.083 195.45 78.223
- num   1    11.043 198.41 78.674
<none>              187.37 78.956
- wid   1    15.816 203.18 79.387
- tor   1    18.297 205.66 79.752
- dis   1    18.412 205.78 79.768
- ove   1    21.409 208.78 80.202
- rea   1    37.260 224.63 82.397
```

```
Step:  AIC=77.02
mil ~ dis + hor + tor + com + rea + car + num + ove + wid + wei

       Df Sum of Sq    RSS    AIC
- car   1     0.493 188.27 75.100
- com   1     1.879 189.66 75.320
- hor   1     7.243 195.02 76.157
- wei   1     8.222 196.00 76.307
<none>              187.78 77.022
- num   1    14.011 201.79 77.181
- wid   1    16.916 204.69 77.609
- dis   1    18.019 205.80 77.771
- tor   1    18.293 206.07 77.810
- ove   1    21.630 209.41 78.292
+ typ   1     0.409 187.37 78.956
- rea   1    37.701 225.48 80.511

Step:  AIC=75.1
mil ~ dis + hor + tor + com + rea + num + ove + wid + wei

       Df Sum of Sq    RSS    AIC
- com   1     2.868 191.14 73.554
- hor   1     7.860 196.13 74.327
- wei   1     9.574 197.84 74.589
<none>              188.27 75.100
- num   1    14.356 202.63 75.305
- wid   1    18.087 206.36 75.852
- tor   1    18.459 206.73 75.906
- dis   1    18.605 206.88 75.928
- ove   1    22.510 210.78 76.489
+ car   1     0.493 187.78 77.022
+ typ   1     0.323 187.95 77.049
- rea   1    40.422 228.69 78.935

Step:  AIC=73.55
mil ~ dis + hor + tor + rea + num + ove + wid + wei

       Df Sum of Sq    RSS    AIC
- num   1    11.523 202.66 73.310
- hor   1    11.642 202.78 73.328
<none>              191.14 73.554
- wei   1    14.369 205.51 73.728
- dis   1    16.991 208.13 74.109
- wid   1    18.206 209.34 74.284
- tor   1    22.075 213.21 74.833
+ com   1     2.868 188.27 75.100
+ car   1     1.482 189.66 75.320
+ typ   1     0.225 190.91 75.519
```

```
- ove   1    29.281 220.42 75.830
- rea   1    41.482 232.62 77.446

Step:  AIC=73.31
mil ~ dis + hor + tor + rea + ove + wid + wei

       Df Sum of Sq    RSS    AIC
- dis   1     8.662 211.32 72.566
- wid   1     9.077 211.74 72.625
- hor   1    10.603 213.26 72.840
<none>              202.66 73.310
- tor   1    15.088 217.75 73.464
+ num   1    11.523 191.14 73.554
- wei   1    19.461 222.12 74.061
+ typ   1     2.872 199.79 74.882
- ove   1    27.555 230.22 75.135
+ car   1     0.858 201.80 75.183
+ com   1     0.035 202.63 75.305
- rea   1    34.481 237.14 76.024

Step:  AIC=72.57
mil ~ hor + tor + rea + ove + wid + wei

       Df Sum of Sq    RSS    AIC
- wid   1     6.076 217.40 71.416
- tor   1     6.535 217.86 71.479
- hor   1     6.851 218.17 71.523
<none>              211.32 72.566
+ dis   1     8.662 202.66 73.310
+ num   1     3.194 208.13 74.109
+ typ   1     0.475 210.85 74.498
+ car   1     0.144 211.18 74.545
+ com   1     0.119 211.20 74.549
- rea   1    38.878 250.20 75.632
- wei   1    54.518 265.84 77.451
- ove   1    55.836 267.16 77.599

Step:  AIC=71.42
mil ~ hor + tor + rea + ove + wei

       Df Sum of Sq    RSS    AIC
- hor   1     4.464 221.86 70.026
- tor   1     6.132 223.53 70.251
<none>              217.40 71.416
+ wid   1     6.076 211.32 72.566
+ dis   1     5.662 211.74 72.625
+ com   1     0.751 216.65 73.312
+ num   1     0.518 216.88 73.345
```

```
+ typ   1     0.498 216.90 73.347
+ car   1     0.079 217.32 73.405
- rea   1    35.359 252.76 73.937
- ove   1    53.735 271.13 76.043
- wei   1    87.485 304.88 79.562

Step:  AIC=70.03
mil ~ tor + rea + ove + wei

       Df Sum of Sq    RSS    AIC
- tor   1     1.687 223.55 68.253
<none>              221.86 70.026
+ hor   1     4.464 217.40 71.416
+ wid   1     3.690 218.17 71.523
+ dis   1     3.632 218.23 71.531
+ com   1     1.742 220.12 71.789
+ typ   1     1.072 220.79 71.881
+ num   1     1.052 220.81 71.883
+ car   1     0.515 221.35 71.956
- rea   1    34.227 256.09 72.330
- ove   1    50.385 272.25 74.166
- wei   1    84.653 306.52 77.722

Step:  AIC=68.25
mil ~ rea + ove + wei

       Df Sum of Sq    RSS    AIC
<none>              223.55 68.253
+ wid   1     5.033 218.52 69.570
+ com   1     2.712 220.84 69.887
+ tor   1     1.687 221.86 70.026
+ typ   1     1.327 222.22 70.075
+ num   1     0.631 222.92 70.168
+ car   1     0.130 223.42 70.236
+ hor   1     0.019 223.53 70.251
+ dis   1     0.006 223.55 70.253
- rea   1    36.597 260.15 70.802
- ove   1    53.108 276.66 72.648
- wei   1   194.715 418.27 85.048


Call:
lm(formula = mil ~ rea + ove + wei, data = data_gas)

Coefficients:
(Intercept)          rea          ove          wei
   4.494972     2.607338     0.218119    -0.009482
```

[21]:
```
vif(lm)
```

dis 128.834831785374 **hor** 43.9210627083326 **tor** 160.436093199685 **com** 2.05783404509763 **rea** 7.78075023365483 **car** 5.32671388203533 **num** 11.7350383346771 **ove** 20.5858100195177 **wid** 9.41944903851847 **wei** 85.6757547639356 **typ** 5.14254675291154

(4)

From the result, Displacement, Horsepower, Torque, Number of Transmission Speeds, Overall Length, and Weight are variables affected by collinearity.