Analysing the “mtcars” data base

The “mtcars” data base is analysed for investigating: (1) the effect of automatic or manual transmission on MPG, and (2) quantifying the MPG difference between automatic and manual transmissions. For this purpose the best parsimonious model is found based on the eliminating the least significant variables from the model and choosing a model with highest adjusted R^2.The role of different factors including the “am” factor (automatic/manual) on the MPG of the vehichles is studied. The observations and reasonings are supported by figures and tables.

Fitting a Multi-variable Linear Regression Model

The effect of different factorson the performance of the cars (MPG) is studied by fitting a multivariable linear regression model to the data. Starting from a complete model, the best parsimonious model is found by identifying the most significant variables on the MPG and the model with highest adjusted R^2. At the first step the “drat” parameter with highest P-value (0.9876) is distinguished as the least important parameter and eliminated, which has resulted in increase in the adjusted R^2 value and P-values of the remaining parameters of the model (the scattered plot for MPG versus this parameter can be found in the appendix). In the next step parameters “gear” and “carb” are identified as factors with small effect on MPG and deleted from the model resulting in the increase in the mentioned criteria. Finally, the best model first order linear model is found as the linear combination of “cyl”, “hp”, “wt”, “vs”, and “am” parameters with R^2 = 0.8418, F-Statistics = 28.49 (6 and 25 DF), and 5.064e-10 P-value. The coefficients of the chosen parameters and the residual plots for this model are as shown below.

```
##
## Call:
## lm(formula = mpg ~ . - drat - gear - carb - disp - qsec, data = mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -3.67  -1.66  -0.43   1.33   5.47
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  33.2416     5.4853    6.06  2.1e-06 ***
## cyl          -0.4018     0.7936   -0.51    0.617
## hp           -0.0259     0.0139   -1.87    0.073 .
## wt           -2.5433     0.9351   -2.72    0.011 *
## vs            1.1707     1.8128    0.65    0.524
## am            1.9757     1.6483    1.20    0.241
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.54 on 26 degrees of freedom
## Multiple R-squared:  0.851,  Adjusted R-squared:  0.823
## F-statistic: 29.8 on 5 and 26 DF,  p-value: 5.57e-10
```

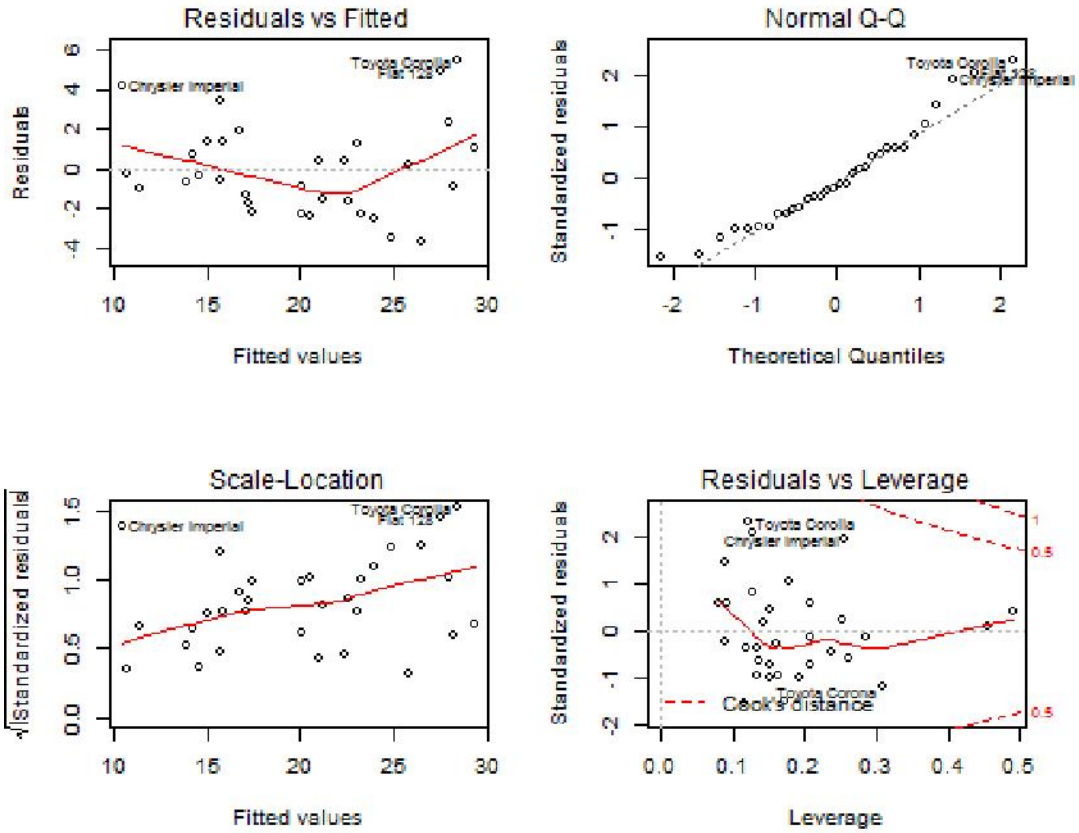

Figure 1. Residuals plots

It can be obsrved that the residuals are distributed almost normally, but since the sample size is small, justifying other effects such as homoescedacity can not be done easily. The model suggests that the manual transmission can considerably decrease the fuel consumption. Taking the intercept equal to zero for the mentioned model, the coefficient of the “am” factor (automatic/manual) will be 2.70384, which suggest that on average the manual transmision will result in 2.70384 more MPG. There are meaningful relations between variables. For example cars with more cylanders are supposed to have higher horse power and recognizing these relationships can help to simplifying the model.

The pairwise relationships of the variables are shown by Figure 2 in the appendix. It can be observed that the parameters "disp", "hp", and "cyl" are highly correlated and increase in each of them can result in highe values for the others. Therefore, between the remaining parameters "hp" and "cyl" one can be eliminated withput significant decrease in model's ability on representing the data. If the factor "cyl" be eliminated from the model, the p_value of the parameters will significantly increase in exchange to slight change in the R^2 value. Further simplifying the model by eliminating the least remaining important parmater "vs" will result in the following model:

```
##
## Call:
## lm(formula = mpg ~ . - drat - gear - carb - disp - qsec - cyl -
##     vs, data = mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.422 -1.792 -0.379  1.225  5.532
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 34.00288    2.64266   12.87  2.8e-13 ***
## hp          -0.03748    0.00961   -3.90  0.00055 ***
## wt          -2.87858    0.90497   -3.18  0.00357 **
## am           2.08371    1.37642    1.51  0.14127
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.54 on 28 degrees of freedom
## Multiple R-squared:  0.84,   Adjusted R-squared:  0.823
## F-statistic:   49 on 3 and 28 DF,  p-value: 2.91e-11
```

If only the P-value criteria is chosen for selecting the parmaters of the model, based on P-values and 5% significance level parameter "am" can be neglected, and the new model will be consist of all signifcant parameters with slightly less R^2 value:

```
##
## Call:
## lm(formula = mpg ~ . - drat - gear - carb - disp - qsec - cyl -
##     vs - am, data = mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.941 -1.600 -0.182  1.050  5.854
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 37.22727    1.59879   23.28  < 2e-16 ***
## hp          -0.03177    0.00903   -3.52   0.0015 **
## wt          -3.87783    0.63273   -6.13  1.1e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.59 on 29 degrees of freedom
## Multiple R-squared:  0.827,  Adjusted R-squared:  0.815
## F-statistic: 69.2 on 2 and 29 DF,  p-value: 9.11e-12
```

This study shows that on average the manual transmission will result in higher millage and less fuel consumption, but the type of transmission is not the only factor that is dominating the changes in “mpg” variable. It can be distinguished form the Figure 2 that the higher “am” (manual transmission) means less “wt” and higher “disp”, whcich suggests that the reason that manual transmission has resulted in higher MPG could be that the lighter and smaller cars have this type of transmission.

Appendix

```
pairs(mtcars)
```

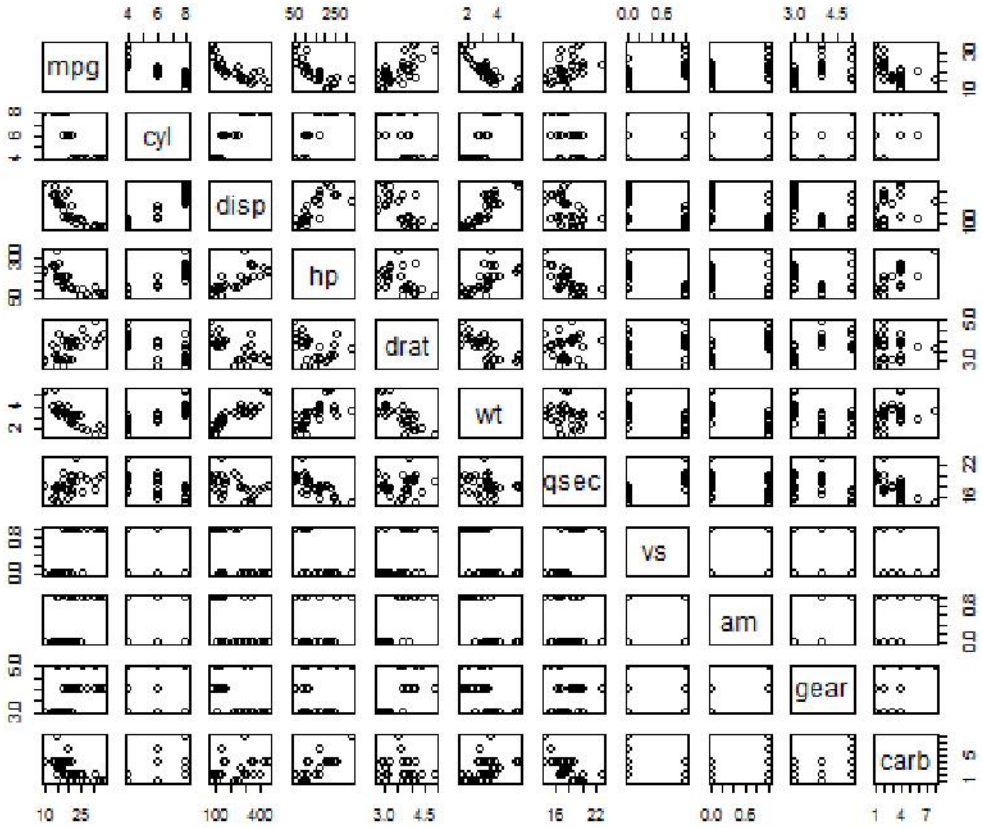

Figure 2. Model variables' paired plot