title: “Linear Regression on mtcars” author: “Alexander Bruns” date: “Friday, June 12, 2015” output: html_document: default

Summary

This study reviews the effect of 10 parameters on the gas mileage of 32 cars. The impact of each factor was investigated and modeled. It was found that the transmission type has no significant effect on the gas mileage. There are gasoline saving autmomatic automobiles availible as well as gasoline saving manual automobile. However, these findings only apply to this dataset, because it consists of very fewe observations and besides some exotic cars high hp cars are included; thus they do not represent the average car market.

Effect of transmission type on mpg based on the *mtcars* dataset

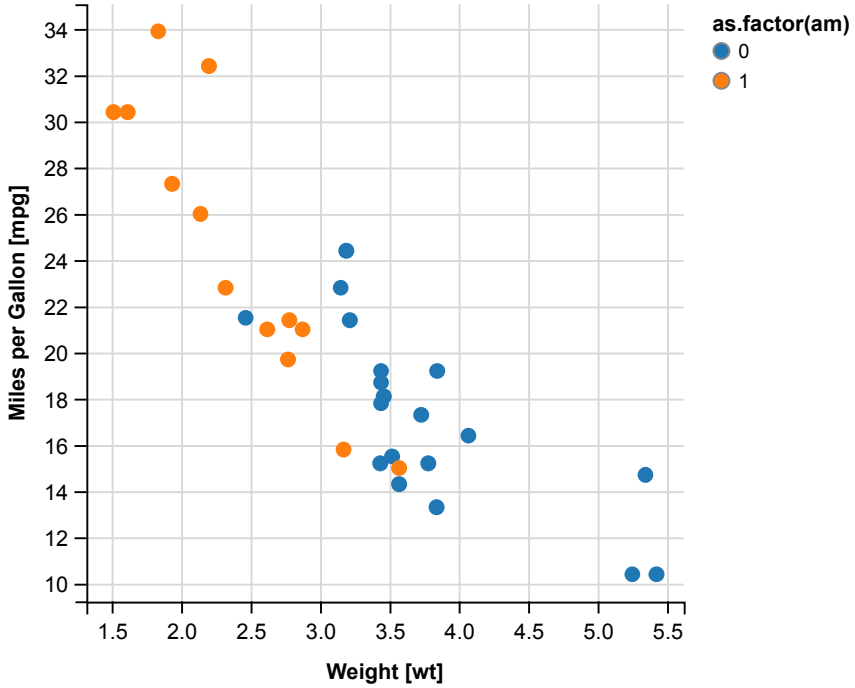
The present study aims to investigate the possible effect of the transmission type in 32 cars represented in the *mtcars* dataset among other variables.

This report is divided into three parts: Exploratory data analysis, Model selection and Model validation.

Exploratory Data Analysis

The exploratory data analysis (Appendix A) showed a strong correlation between mpg and weight, displacement and cylinders (R > 0.8). Displacement and cylinders are confounded as usually the displacement increases with cylinder number. Both are correlated with R > 0.9. The correlation between *mpg* and transmission type is about 0.6 which indicates that mpg depends on various factors.

Graph 1 reveals the relation between Weight, Transmission type and Miles per Gallon.



Model Selection

This section describes the selection of the best model for the question to how extent affects the transmission type the *mpg*. As shown in the previous section, only the variables with the highest correlation coefficient

were taken into account. The selection of the best model is done by means of a likelihood-ratio-test.

```
## Analysis of Variance Table
##
## Model 1: mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear + carb
## Model 2: mpg ~ wt + am + cyl
## Model 3: mpg ~ wt + am + cyl + wt * cyl
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1     21 147.49
## 2     28 191.05 -7   -43.553 0.8858 0.53445
## 3     27 154.31  1    36.736 5.2304 0.03268 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The result of the ANOVA table shows that the addition of an interaction term increases the accuray. Therefore the linear model *fit_interact* will be used the calculate prediction and confidence intervals.

Model Evaluation

The following table shows the summary of the linear model. As it can be seen, the transmission type has no significant effect on mpg. Weight [wt] and cylinders [cyl] are significant regressors. In fact, the gas mileage *does not* depent on the transmission type.

```
##
## Call:
## lm(formula = mpg ~ wt + am + cyl + wt * cyl, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.7887 -1.3812 -0.5268  1.3155  5.3945
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  57.0173     7.3507   7.757 2.43e-08 ***
## wt           -9.5003     2.6492  -3.586 0.001308 **
## am           -0.8619     1.2622  -0.683 0.500524
## cyl          -4.0111     1.0594  -3.786 0.000777 ***
## wt:cyl        0.8858     0.3494   2.535 0.017337 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.391 on 27 degrees of freedom
## Multiple R-squared:  0.863,  Adjusted R-squared:  0.8427
## F-statistic: 42.51 on 4 and 27 DF,  p-value: 2.815e-11
```

Residuals (see **Annex C**) are in a good approximation normally distributed. No autocorrelation nor heteroskedasticity could be observed. However, two car models - the Toyota Corolla and the Fiat 128 - show greater residuals tha expectetd. This might be due to the fact that these small cars were engineered to save gasoline, opposite to the majority of the investigated population.

APPENDIX A - Getting and Cleaning the *“mtcars”* Data

Requiered packages:

```
library(dplyr)
library(lattice)
library(datasets)
library(psych)
```

Getting and cleaning the data

```
data(mtcars)
```

APPENDIX B - Exploratory Data Analysis

Inspection for relationship between dependent and independent variables

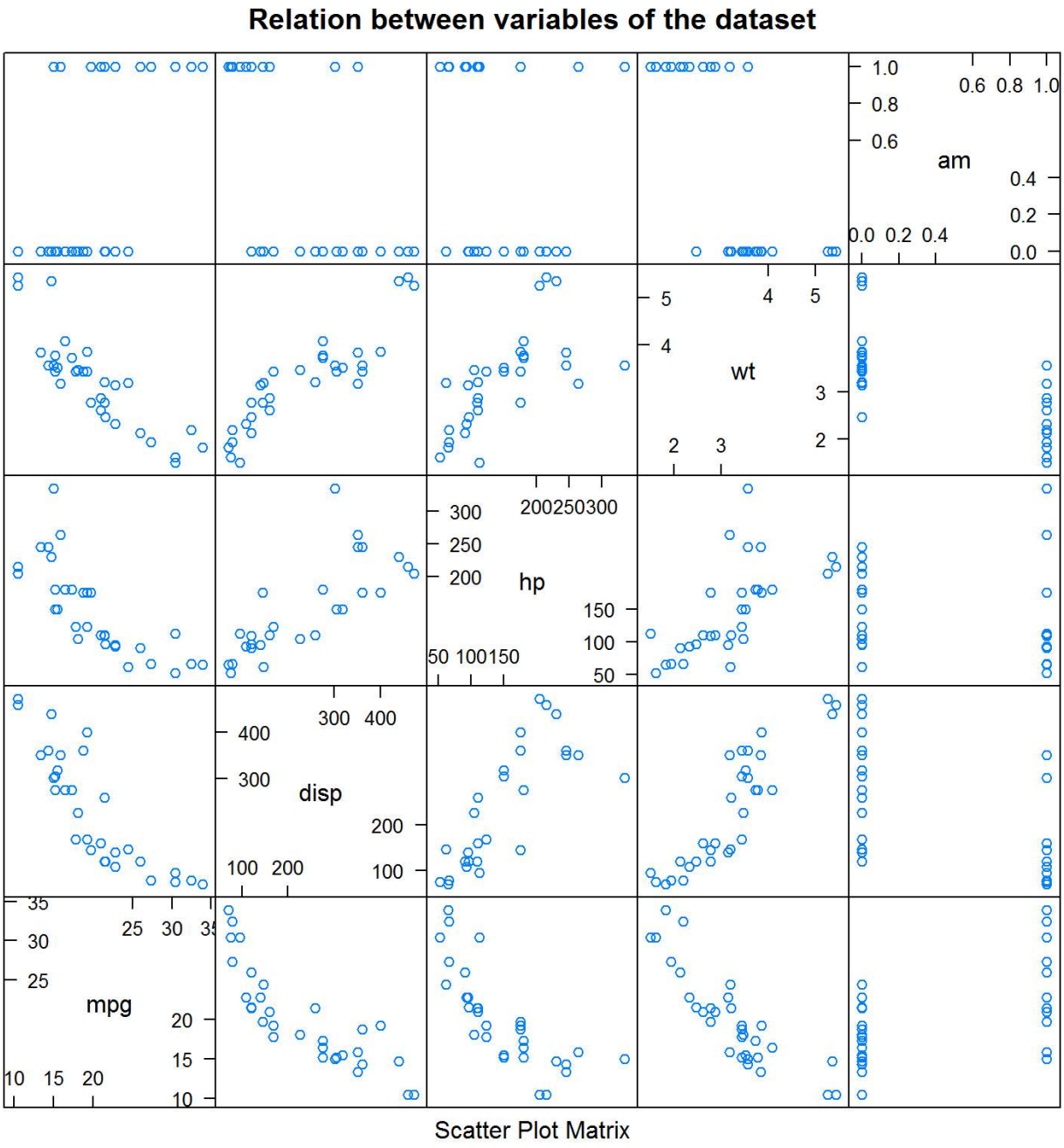
The dependent variable is *mpg* while the independent variables are weight *wt*, transmission type *am*, cylinder displacement *disp* and horsepower *hp*.

Firstly, the correlations between all variables are investigated.

```
corr.test(mtcars)$r[,1]
```

```
##        mpg        cyl       disp         hp       drat         wt
##  1.0000000 -0.8521620 -0.8475514 -0.7761684  0.6811719 -0.8676594
##       qsec         vs         am       gear       carb
##  0.4186840  0.6640389  0.5998324  0.4802848 -0.5509251
```

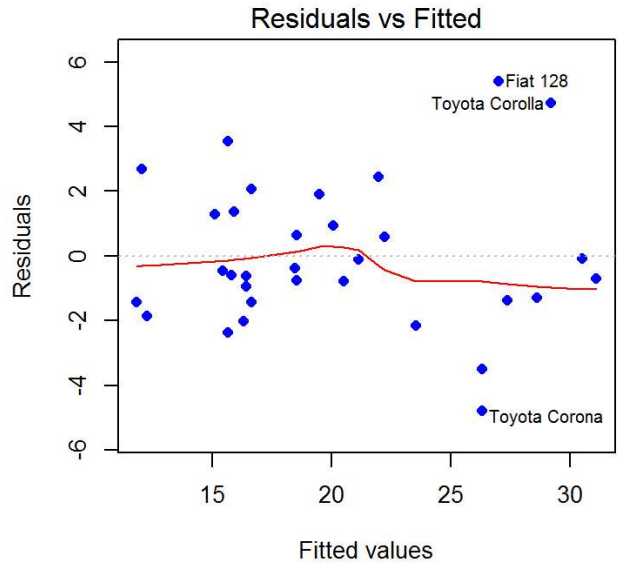
Then, all variables with a correlation with *mpg* higher than 0.6 are visualized in the following plot in order to examine the type of the correlation.

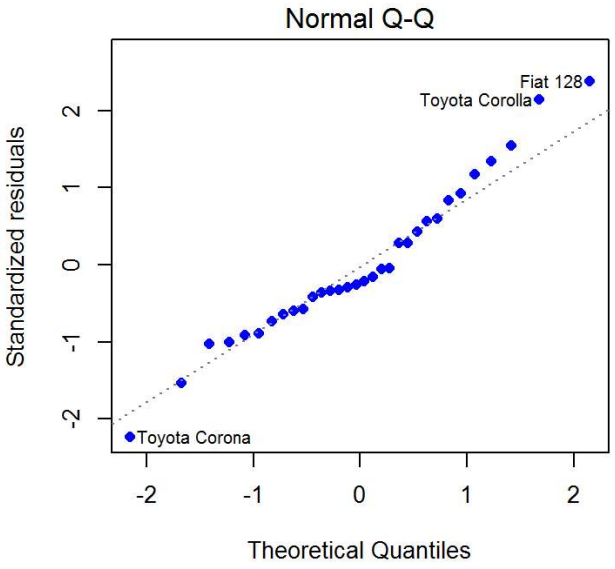


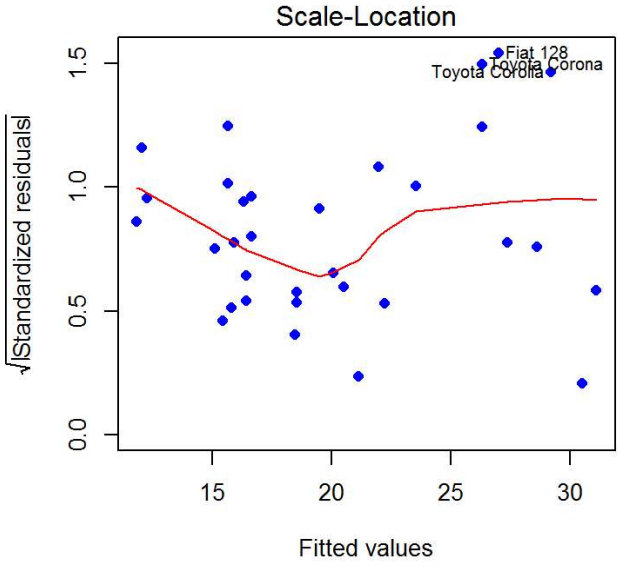
APPENDIX C - Residuals

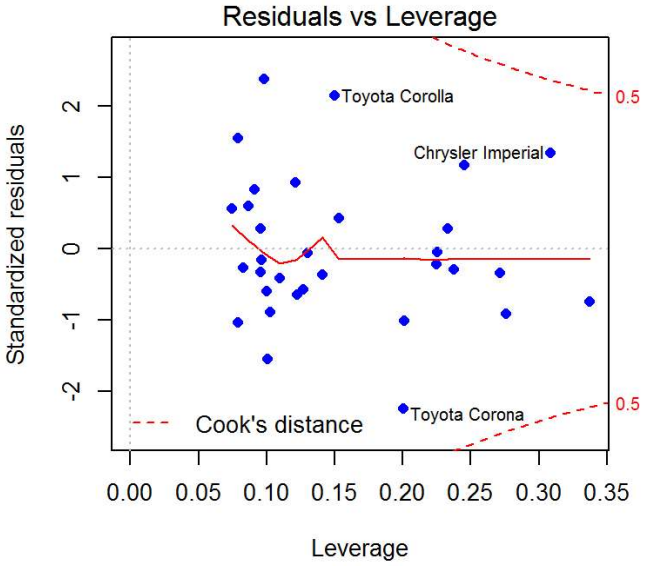
The residuals for the model were examined. The following four methods were utilized in order to determine the goodness of fit.

- Residuals vs Fitted
- QQ-Plot to examine normal distribution of residuals
- Standarized residuals
- Residuals vs Leverage









Residuals are in a good approximation normally distributed. No autocorrelation nor heteroskedasticity could be observed. However, two car models - the Toyota Corolla and the Fiat 128 - show greater residuals tha expectetd. This might be due to the fact that these small cars were engineered to save gasoline, opposite to the majority of the investigated population.