Course Project Regression Models by Philipp Paier

Executive Summary

In this project, the influence of transmission and other factors on fuel consumption of automobiles is investigated. Though manual transmission seems to lead to less fuel consumption, it will be shown, that when factors such as weight and horspower are taken into account, the difference is insignificant. **Notes:** Code has been developed and tested with **R 3.1.1** on **Windows Vista**.

Exploratory Analysis

```
## Loading required package: knitr
```

The dataset under observation has been extracted by the Motor Trend US magazine in 1974. It can be loaded into R using the following commands:

```
library(datasets)
data(mtcars)
```

The dataset consists of 32 observations and 11 variables, like Miles per Gallon, horsepower, weight and so on. A detailed list can be found by using the R command ?mtcars.

At first we have a look at some plots to see how the individual aspects might be related to eachother. The plots can be found in the Appendix as **Figure 1**. As we are especially interested in the influence of transmission on mpg, we also make a boxplot of mpg versus transmission. **Figure 2** in the Appendix shows the result. The figure strongly suggests that manual transmission leads to more Miles per Gallon, but also other factors, such as weight and horsepower seem to have an obvious impact on mpg.

Regression Models

Note:
Because analysis should fit on 2 pages, excessive R output (e.g. given by summary) is hidden. Instead the important values (such as p-values and R-squared) are embedded in the text.

First we have a look at a simple regression model, that is based on transmission as predictor only and interpret the resulting coefficients.

```
mtcars$am <- as.factor(mtcars$am)
fit1 <- lm(mpg ~ am , data=mtcars)
coefficients(fit1)
```

```
## (Intercept)         am1
##   17.147368    7.244939
```

The value of R-squared is ~0.36, which means that in this model, transmission accounts for pproximately 36% of the variance in mpg. The coefficients can be interpreted as follows: Manual transmission results in around 7.25 more mpg compared to automatic transmission on average and a small p-value (0.000285) indicates a high significance of transmission.

The exploratory data analysis has shown, that other variables also seem to correlate with mpg. We therefore make a model with weight and horsepower, as those make sense to influence mpg, which is also indicated by **Figure 1** in the Appendix.

```
fit3 <- lm(mpg ~ wt + hp, data=mtcars)
```

This model now explains around 82,6% of the variance in mpg and the p-values (both are below 0.01) show, that both variables are highly significant. Now we add transmission to the equation:

```
fit4 <- lm(mpg ~ wt + hp + as.factor(am), data=mtcars)
```

We now see that given horsepower and weight as factors, transmission doesn't have a significant impact on mpg (p-value of transmission is 0.14). We finally compare the model, that only uses weight and horsepower with one that uses all variables, by making an analysis of variance:

```
anova(fit3,lm(mpg ~ . , data=mtcars))
```

The p-value of the F-Test (~0.57) shows, that including more variables doesn't improve the model and we fail to reject the null hypothesis, saying that adding more variables has no impact. Consequently we use the weight/horsepower model as our final model.

Questions of Interest

Is an automatic or manual transmission better for MPG?

We have shown, that if we model our predictor for mpg, based on transmission only, manual transmission seems to be better for mpg. However we have also shown, that including the variables weight and horspower leads to the conclusion that transmission is insignificant.

Quantify the MPG difference between automatic and manual transmissions

Given the model based on transmission only, the mpg is around 7,25 higher for manual transmission than for automatic transmission.

Residual Plot, some Diagnostics and Uncertainty in the Model

An additional Residual Plot can be seen in **Figure 3** of the Appendix and illustrates the uncertainty in our Model. The graph on the upper left indicates a pattern, as the fitted values define a quadratic curve. This indicates, that our assumption of a linear model might be false. Actually we can see from the exploratory analysis, that mpg is probably not linear dependent from horsepower and weight. There also seem to be some leverage according to the graph on the lower right corner of the figure. Calling hatvalues shows that the Maserati Bora has the biggest leverage:

```
rownames(mtcars[which.max(hatvalues(fit3)),])
```

```
## [1] "Maserati Bora"
```

Though, the model explains almost 83% of the variance, there is some uncertainty in our model, especially, because there is strong evidence that the actual model isn't linear. Further analysis with nonlinear models is therefore recommended, but is out of scope of this report.

Appendix

Figure 1: How are variables related to eachother?

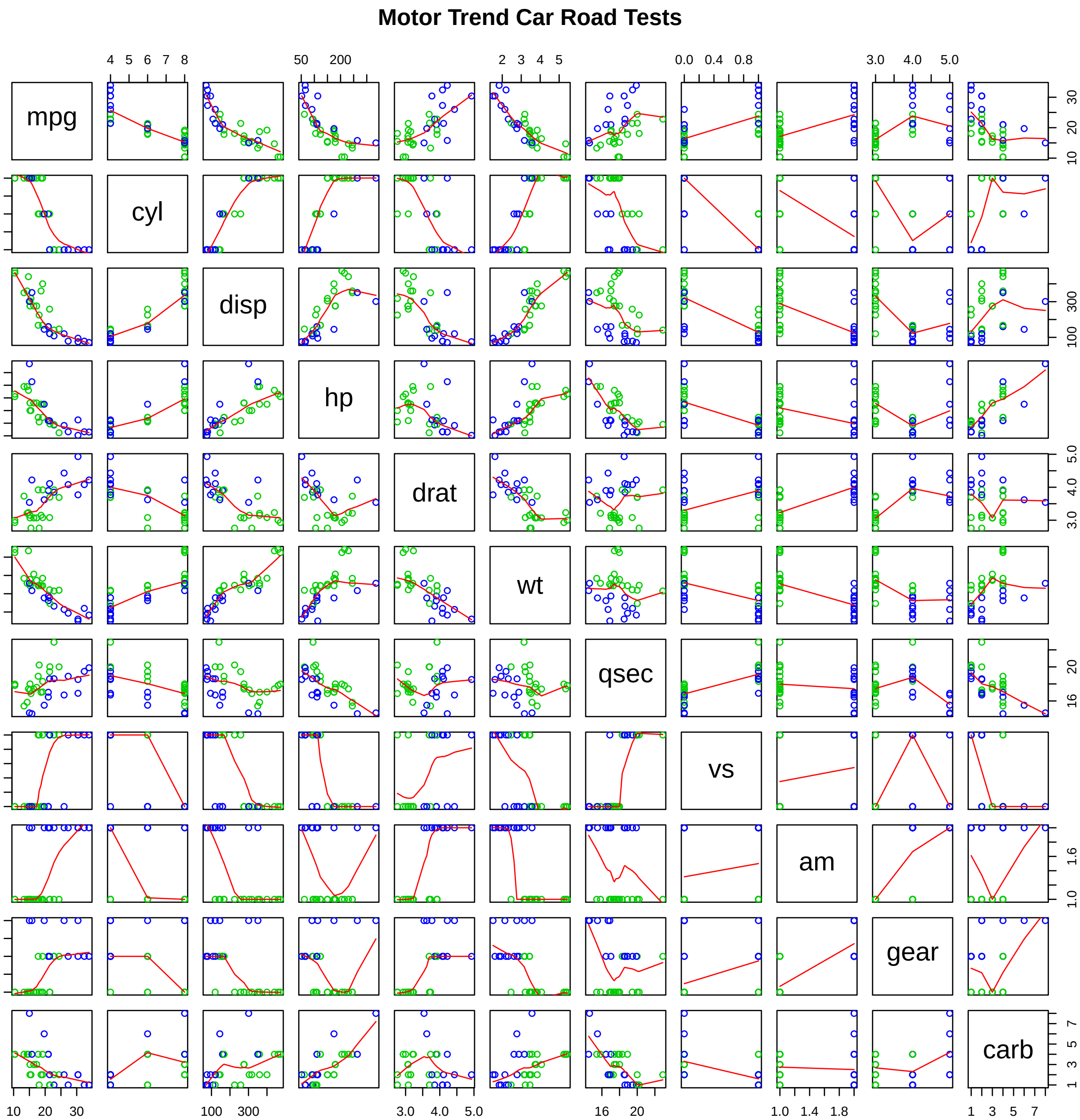

Figure 2: Transmission vs. Miles per Gallon

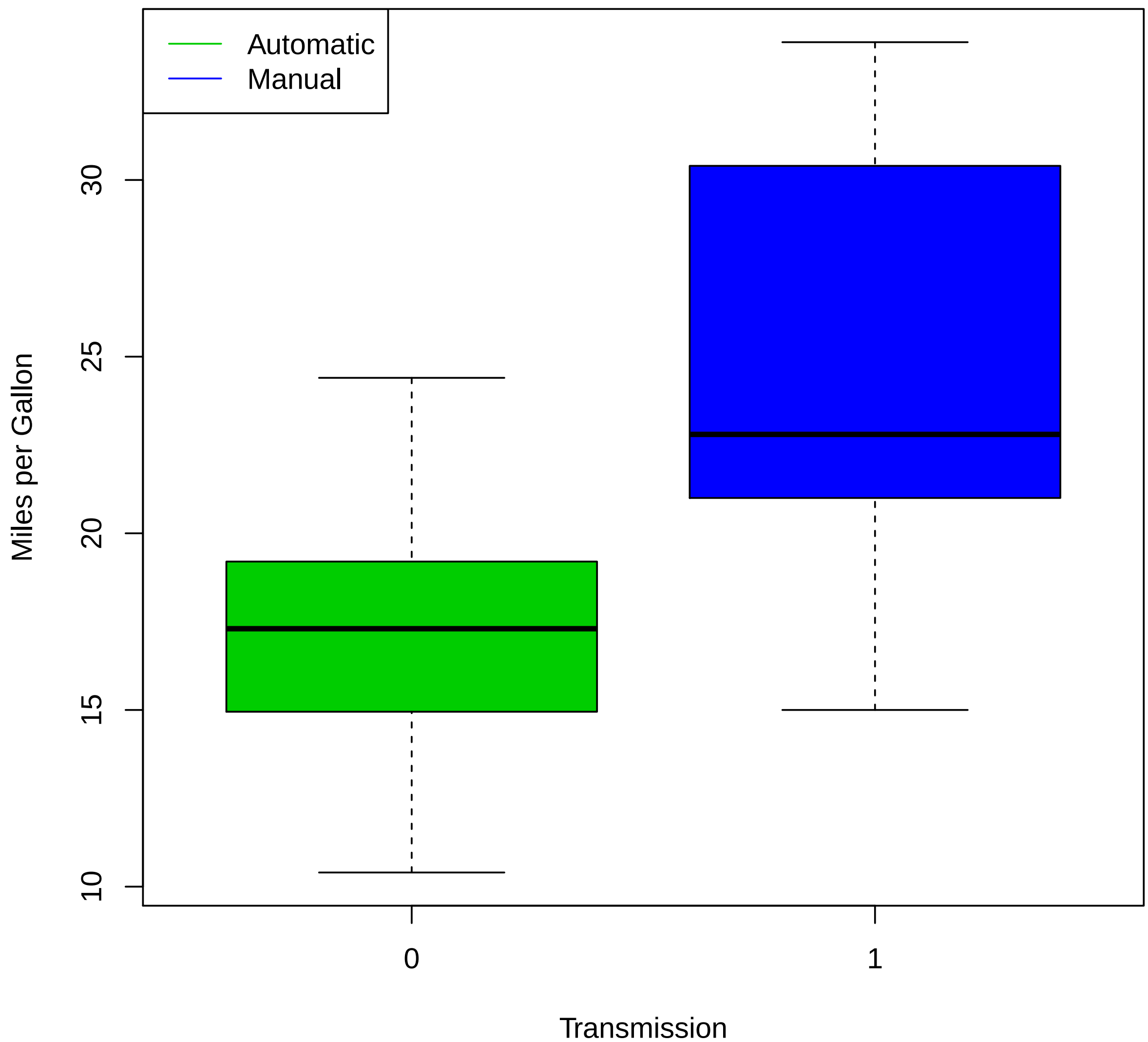

Figure 3: Residual Plot