Motor Trend Data Analysis Report by: Josué Lavandeira

Executive Summary

This report analyzes the dataset `mtcars` and explores the relationship between some of the cars' characteristics with their MPG performance. The study shows that the MPG that a car averages vary in direct correlation to the car's transmission type, it also shows that this correlation becomes less significant when the factors `cyl`, `hp` and `wt` are added to the model. The model that takes these four factors as predictors for a car's MPG performance, results the best model to predict a car's MPG, accounting for about 84% of the variability of this measure.

Data loading and transformation

First, we load the data set `mtcars` and observe the variables we need to change from `numeric` class to `factor` class.

```
data(mtcars)
str(mtcars) ## Results hidden
```

Now we transform the necessary variables to factors

```
mtcars$cyl <- factor(mtcars$cyl)
mtcars$vs <- factor(mtcars$vs)
mtcars$gear <- factor(mtcars$gear)
mtcars$carb <- factor(mtcars$carb)
mtcars$am <- factor(mtcars$am,labels=c('Automatic','Manual'))
```

And now we can observe how the data has changed:

```
str(mtcars) ## Results hidden
```

Exploratory data analysis

First we have to analyze the correlations between the observed variables by plotting all relationships between variables in the `mtcars` dataset (**Appendix: Figure 1**), then we need to quantify this correlations by using linear models. We also need to establish the effect that the transmission has on the cars' MPG performance by doing a box plot of the `mpg` variable for each of the transmission types (**Appendix: Figure 2**).

Regression Analysis

We need to create our base linear model that represents the relation between the `mpg` and the rest of the variables, and then we have to find the best model fit, and then we compare the two models. Based on the initial observations of the pairs plot where several variables show a visible correlation with `mpg`, we need to build a model with all the variables as predictors, and then do a selection to detect significant predictors for the best possible model. We use the step method for this which will help us select our best model fit.

```
basemod <- lm(mpg ~ ., data = mtcars)
bestmod <- step(basemod, direction = "both")
summary(bestmod) ## Results hidden
anova(basemod, bestmod) ## Results hidden
```

The best model fit obtained by the step method, includes the variables `am`, `cyl`, `hp` and `wt`. The results show that the model explains about 84% of the variability of the MPG performance of the car. By looking the values on the `anova` table, we can observe that the p-value is highly significant, so we reject the null hypothesis, which states that the variables `am`, `cyl`, `hp` and `wt` don't contribute significantly to the accuracy of the model.

Inference

We test the null hypothesis that claims the car's transmission has no significant effect on the cars MPG performance (assuming the MPG has a normal distribution). We use the t-Test as our tool.

```
bm_ttest <- t.test(mpg ~ am)
bm_ttest$p.value ## Results hidden
bm_ttest$estimate ## Results hidden
```

Since the p-value is 0.00137, we reject our null hypothesis and we can say that the car's transmission has a direct impact in the car's MPG performance. On average, a car with a manual transmission will give 7.25 MPG more than automatic transmitted cars.

```
ammod <-lm(mpg ~ am, data=mtcars)
summary(ammod) ## Results hidden
```

We can see that a car gives on average 17.147 MPG with an automatic transmission, while **cars with manual transmissions give on average 7.245 MPG more than cars with automatic transmissions**. The change in transmission type (`am`) **can only account for about 33% of the variance of MPG (`mpg`)**.

Residual Analysis and Diagnostics

Now we may create the residual plots (**Appendix: Figure 3**) from which we can make the following observations:

1. The Residuals vs. Fitted plot does not show a consistent pattern, this supports the accuracy of the independence assumption.
2. The Normal Q-Q plot clearly shows the residuals are normally distributed, as all values fall close to the line.
3. The Scale-Location plot shows points scattered in a constant pattern, which indicates a constant variance.
4. There are some outliers in the plots. But if we test the Dfbetas, we find that none of the values is gretaer than 1, meaning no observation affects significantly the estimate of a variable's regression coefficient.

```
sum((abs(dfbetas(bestmod)))>1)
## [1] 0
```

Conclusions

We know now car's transmission does seem to directly affect a car's MPG performance, but it can only account for about 33% of the MPG performance variance confidently. It's best to use a model that takes into account the number of engine cylinders, maximum horse power and the weight of the car along with the transmission to predict a car's MPG performance, this model will account for about 84% of the MPG performance behavior of the car. Additionally:

- Cars equipped with a manual transmission give more miles per gallon than cars equipped with automatic transmissions by a factor of 1.81. This is true for when variables `cyl`, `hp` and `wt` are accounted for.

- The total miles per gallon a car can give will decrease by a factor 2.5 for every 1000 lb increase in the car's weight. This is true for when variables `cyl`, `hp` and `am` are accounted for.

- Cars that have 6 and 8 cylinder engines give less miles per gallon than cars with 4 cylinder engines by factors of 2.16 and 3 respectively. This is true for when variables `am`, `hp` and `wt` are accounted for.

- When a car's hp is increased, the car's mileage per gallon will be reduced by a factor of 0.32, this appears to be negligible, yet we must remember this is only true for when variables `cyl`, `am` and `wt` are accounted for.

Appendix

Figure 1.

Boxplot of MPG vs. Transmission

```
boxplot(mpg ~ am, xlab="Transmission", ylab="MPG", main="Boxplot of MPG vs Transmission",
data=mtcars, names=c("Manual", "Automatic"), col=c("cyan","red"))
```

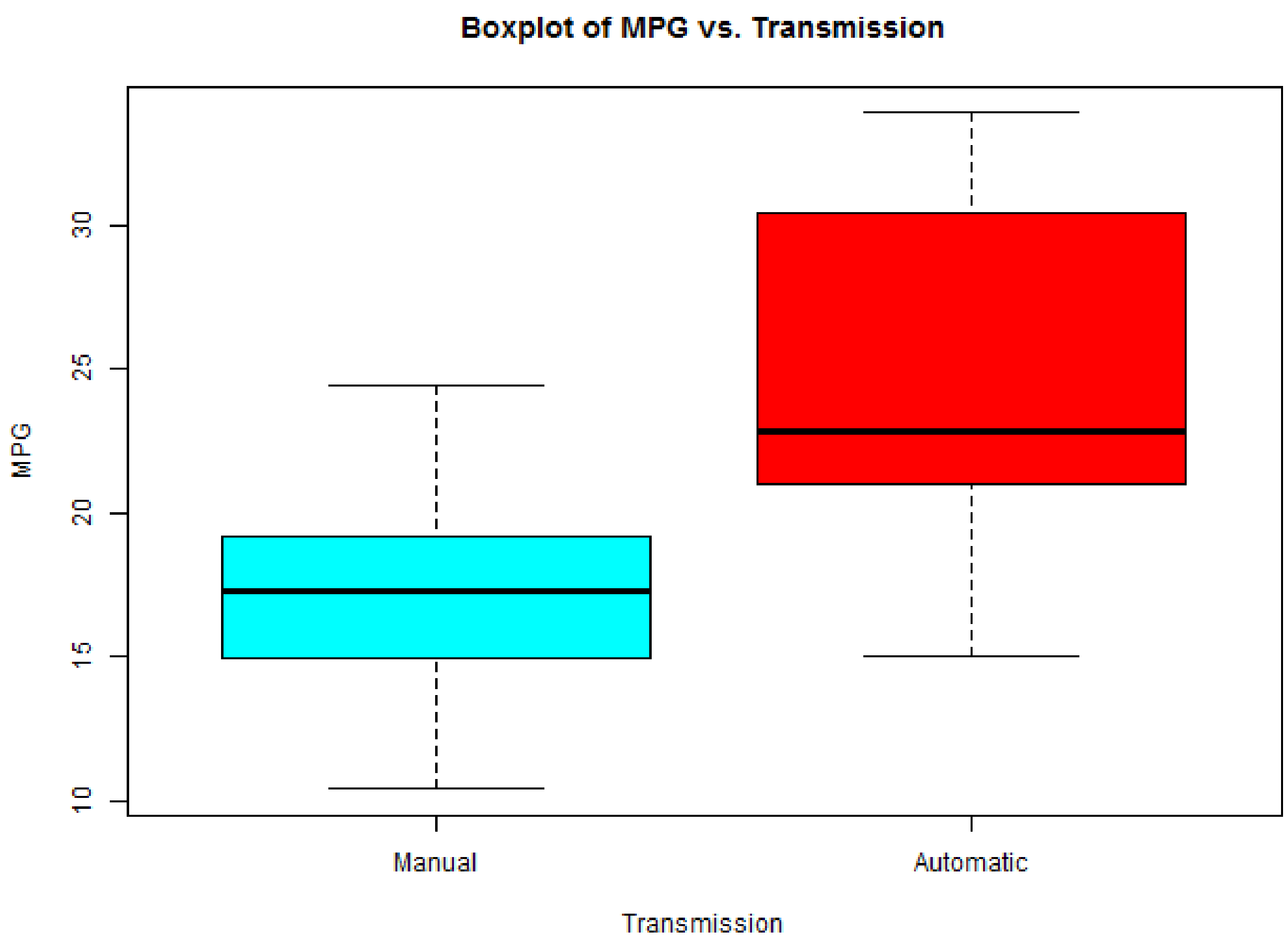

Figure 2.

Paired Variables in Motor Trend Car Observations

```
pairs(mtcars, panel=panel.smooth, main="Paired Variables in Motor Trend Car Observations")
```

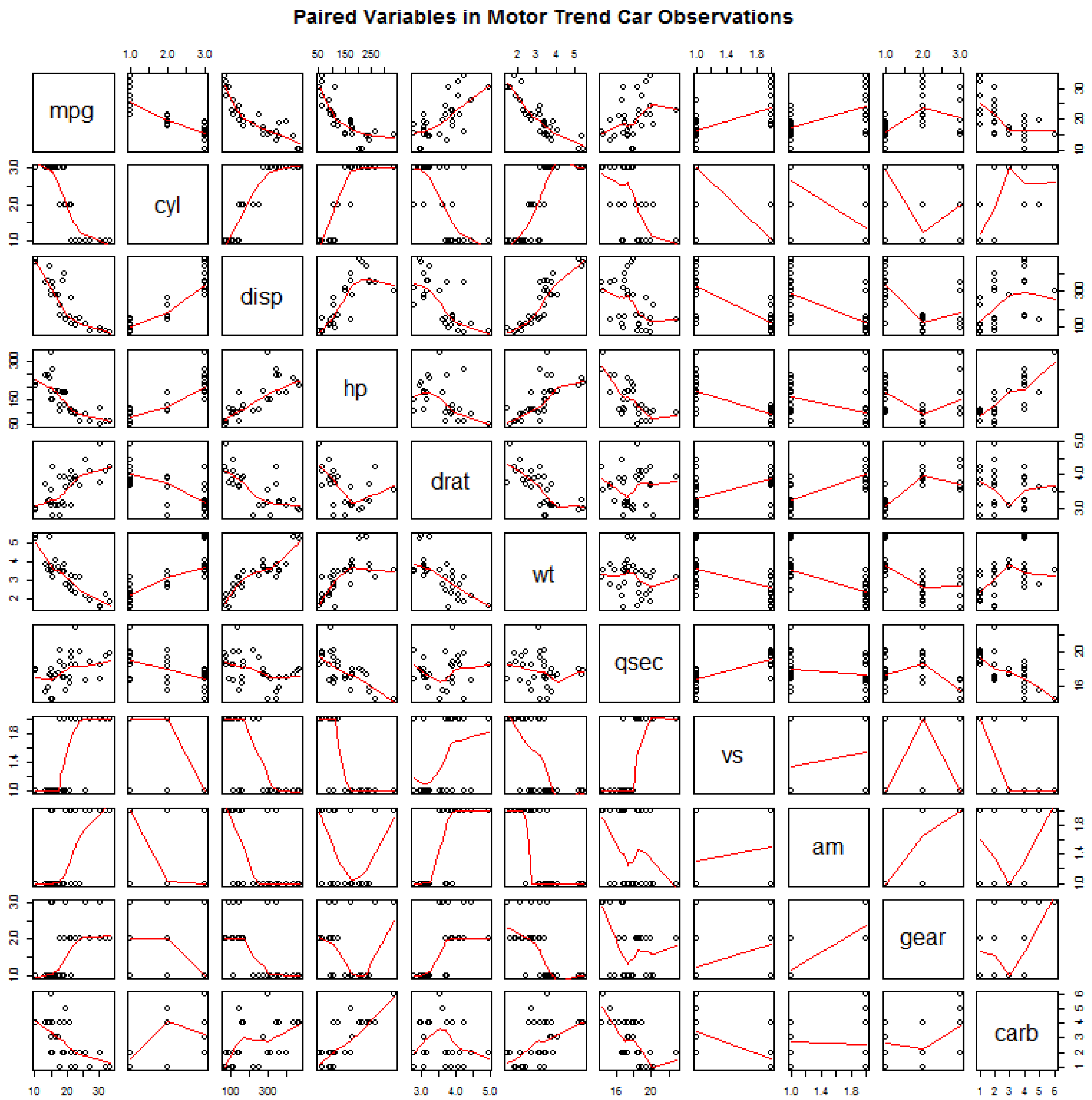

Figure 3.

Best model plots

```
par(mfrow = c(2,2))
plot(bestmod)
```