Motor Trend Regression Analysis Report

Ajani Gyasi

Executive Summary

Our job is to analyze the `mtcars` data set to find key trends in the relationship between already determined variables and miles per gallon (mpg). This data set was extracted from the 1974 Motor Trend U.S. magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (R-Source: ?mtcars). We will gain insight on whether automatic or manual transmission better for MPG. And to broaden our analyses, we will quantify the MPG difference between automatic and manual transmissions. In order to do this, we will fit a few models to find the highest R-squared value, which will explain the percentage of the variance of the MPG variable. Let's explore if lighter weighted cars with manual transmission has lower MPG values than heavier weighted with automatic transmission.

Exploratory Analysis

Load the data in. Use pairs plot to show bivariate relationship between variables. Also, we changed some variables from numeric to factor class.

```
library(datasets)
data(mtcars)
attach(mtcars)
#change variables to factors
am <- as.factor(am)
vs <- as.factor(vs)
```

We determined that Transmission has a pretty good separation in its relationship between automatic and manual per MPG. Please refer to Appendix (Figure 1).

Regression Analysis

This model and our pairs plot will provide us a lens through the data set. Even though there are no significant coefficients through this model, it can still explain 81% of the variance of MPG variable. Please refer to Appendix (Figure 2).

```
model <- lm(mpg ~ ., data = mtcars)
summary(model)
```

Here, we fit a Transmission model. The Adjusted R-squared is very low (34%), which means we need to add other variables to our model.

```
amModel <- lm(mpg ~ am, data = mtcars)
summary(amModel)
```

This model gives us significant coefficients that we can use to boost our Adjusted R-squared. Even though, 83% is good, I think we can do better. We also took a closer look at the connection between MPG and weight via Transmission and noticed the trend of automatic cars weighing heavier than manual cars. This information will help us develop our next model. Please refer to Appendix (Figure 3).

```r
stepModel <- step(model)
summary(stepModel, direction = "backwards")
```

The previous model, gave us insight about the variables to model. Our scatter plot led us to an interaction between the two variables, weight and Transmission. We will include this interaction in our new model along with the variables our previous model provided. Notice, the boost in our Adjusted R-squared at 88% and the all of the coefficients are significant(based on 0.05 significant level), which indicates a pretty good fit to our model.

```r
amwtModel <- lm(mpg ~ wt + qsec + am + wt:am, data = mtcars)
summary(amwtModel)
```

```
##
## Call:
## lm(formula = mpg ~ wt + qsec + am + wt:am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.5076 -1.3801 -0.5588  1.0630  4.3684
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    9.723      5.899   1.648 0.110893
## wt            -2.937      0.666  -4.409 0.000149 ***
## qsec           1.017      0.252   4.035 0.000403 ***
## am            14.079      3.435   4.099 0.000341 ***
## wt:am         -4.141      1.197  -3.460 0.001809 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.084 on 27 degrees of freedom
## Multiple R-squared:  0.8959, Adjusted R-squared:  0.8804
## F-statistic: 58.06 on 4 and 27 DF,  p-value: 7.168e-13
```

Residual Plot Diagnostics

We set up a plot window in Appendix (Figure 4) for a residual analysis to further diagnose the model we chose as the best fit to our data.

Residuals vs. Fitted plot Shows no pattern, which satisfies the accuracy of independence assumption

Normal Q-Q plot The residuals show normal distribution, which means we met our normality assumption

Scale-Location plot The residuals show random distribution, which satifies the constant variance assumption

Residuals vs. Leverage Shows all values fall within the red 0.5 dotted-line range, which means no outliers present to distort the analysis.

We can conclude, that when weight and 1/4 mile time(qsec) remain constant, cars with manual transmission add 14.079 + (-4.141)*wt more MPG on average than cars with automatic transmission.

Appendix

Figure 1: Boxplot of MPG and Transmission

```r
boxplot(mpg ~ am, xlab = "Tranmission(0 = automatic, 1 = manual)"
        , ylab = "MPG", main = "MPG and Transmission")
```

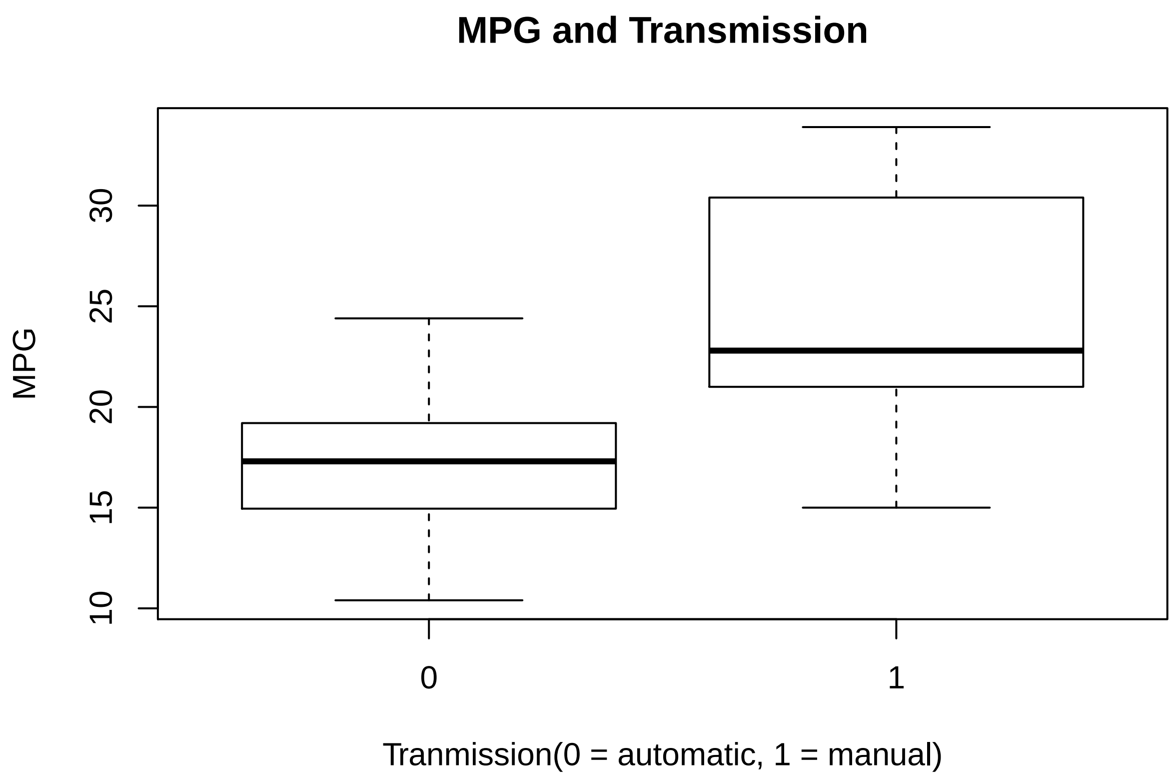


Figure 2: Pairs Plot of Motor Trend Test

```r
pairs(mtcars, panel = panel.smooth,
      main = "Motor Trend Car Road Test")
```

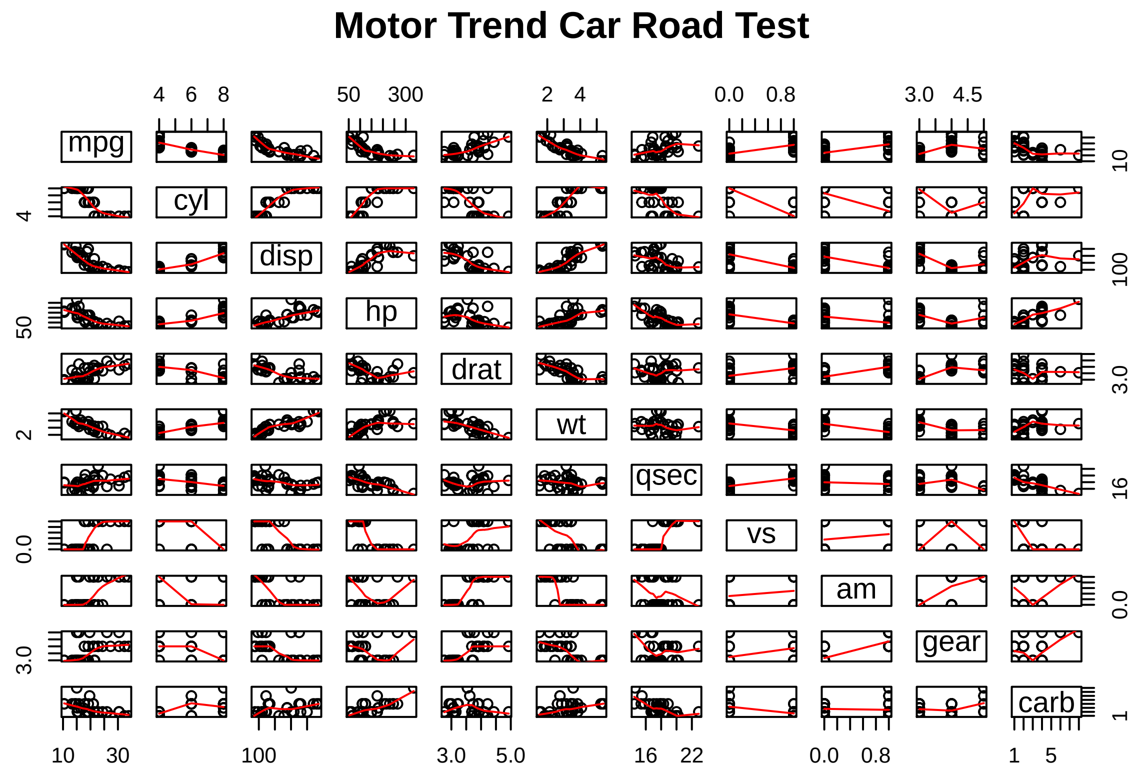


Figure 3: Scatter Plot of Relationship of MPG and Weight via Transmission

```r
library(ggplot2)
```

```
##
## Attaching package: 'ggplot2'
##
## The following object is masked from 'mtcars':
##
##     mpg
```

```r
ggplot(mtcars, aes(x = wt, y = mpg, color = am, group = am, height = 3, width = 3)) +
  geom_point() +
  facet_grid(. ~ am) + xlab("weight") +
  ggtitle("MPG vs. Weight via Tramission Types")
```

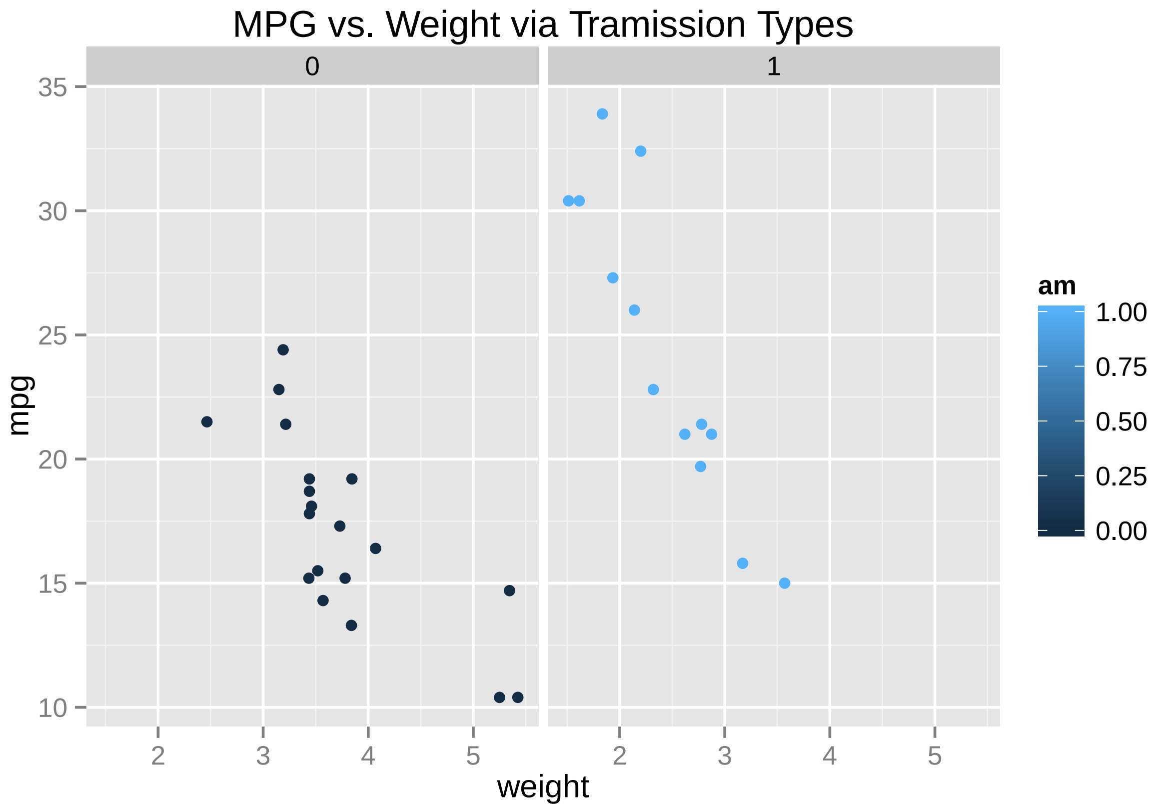
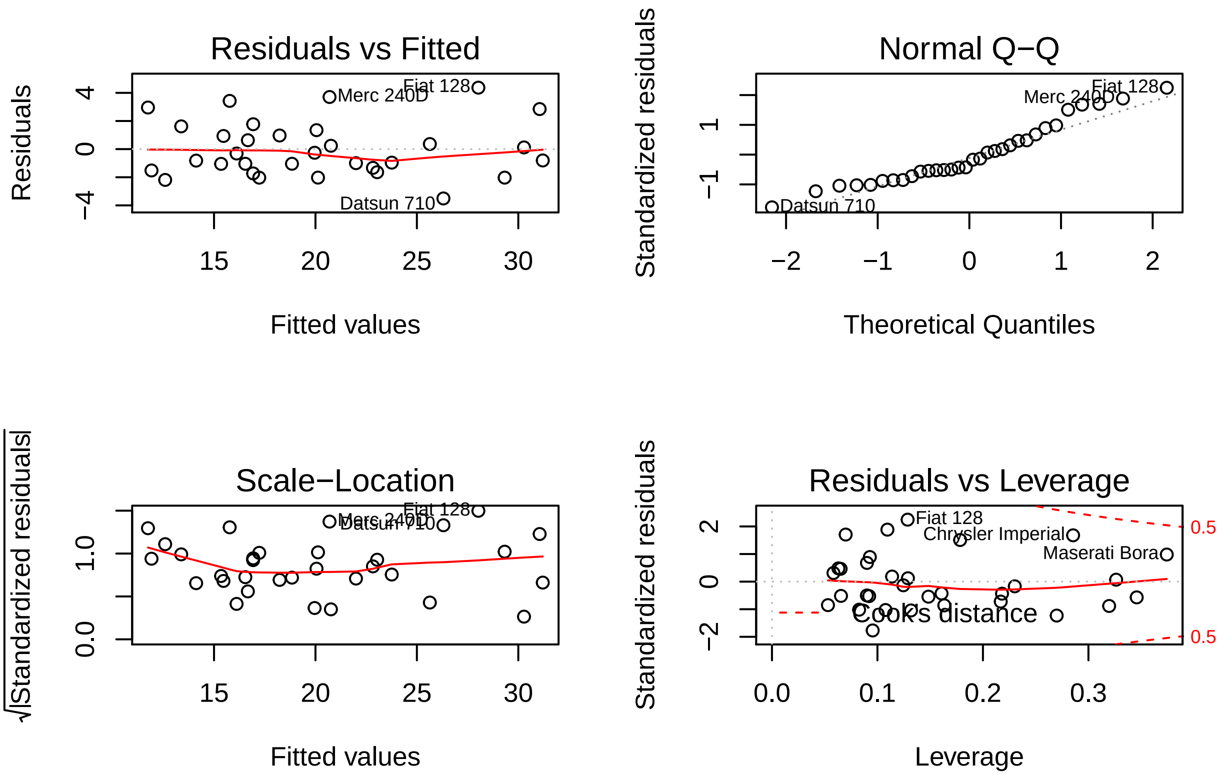


Figure 4 : Residual Plot

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


Residuals vs Fitted
Residuals
Fitted values
Merc 240D
Fiat 128
Datsun 710
Normal Q−Q
Standardized residuals
Theoretical Quantiles
Scale−Location
√|Standardized residuals|
Residuals vs Leverage
Leverage
Chrysler Imperial
Maserati Bora
Cook's distance
0.5