Regression Project

Gerald Belton

Executive Summary

In this project we analyze the *mtcars* dataset with the goal of answering the question "Is an automatic or manual transmission better for MPG?" We also try to quantify that answer. After adjusting for confounding variables (wt, qsec) we determine that a manual transmission increases gas mileage by approximately 2.9 MPG.

Exploratory Data Analysis

```
data(mtcars)
str(mtcars)
```

The "am" variable contains 0 for automatic transmission, 1 for manual. Let's change that to a factor for ease of handling.

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

Initially, we plot the relationships between the variables. See Figure 1 in the appendix. It appears that variables such as "disp", "hp", and "qsec" are correlated with mpg. We also created a box plot (Figure 2) which appears to show that the manual transmission delivers higher mpg than the manual transmission.

Regression Analysis

We initially build a model using all of the available variables.

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

In this initial model, the adjusted R^2 value is .8066, so the model can explain about 81% of the variation in mpg. However, the P values for all of the variables are above 0.05 showing that the results are not significant.

Next we use the "step" method to run lm multiple times to build multiple regression models and select the best variables.

```
adjustedModel <- step(carsModel)
```

```
summary(adjustedModel)
```

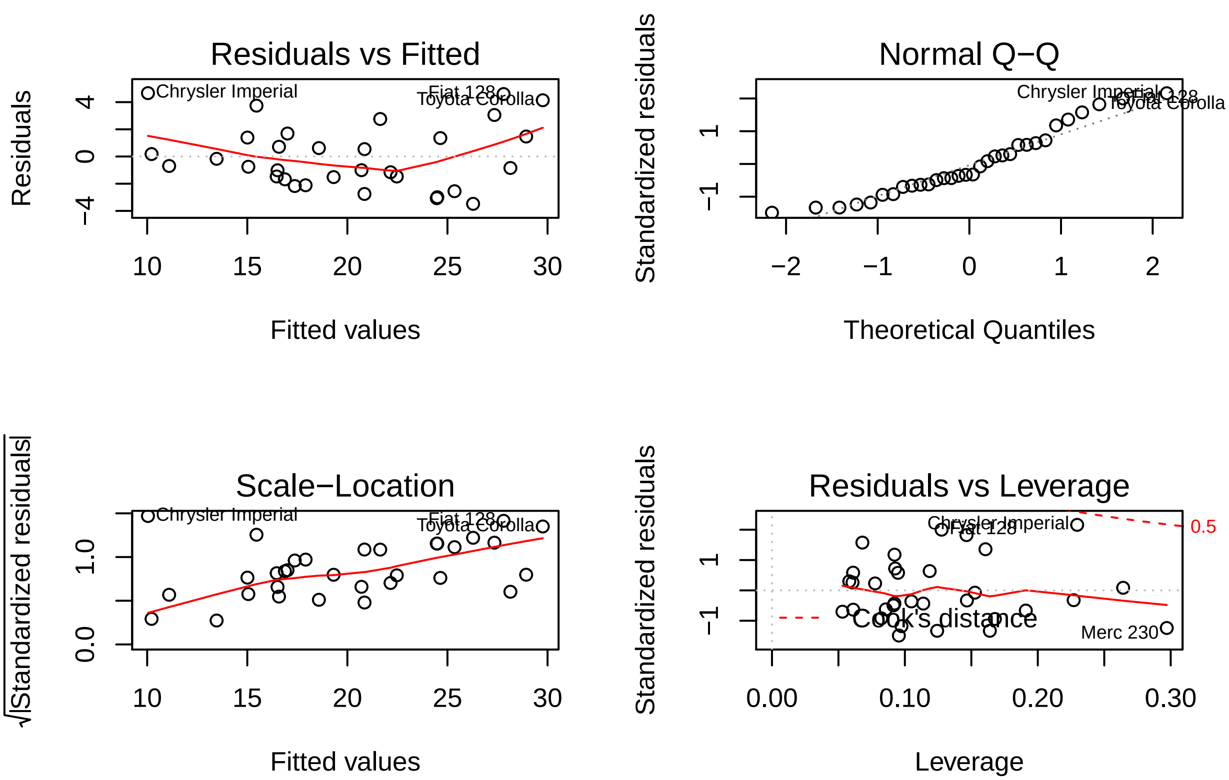
```
##
## Call:
## lm(formula = mpg ~ wt + qsec + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4811 -1.5555 -0.7257  1.4110  4.6610
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.6178     6.9596   1.382 0.177915
## wt           -3.9165     0.7112  -5.507 6.95e-06 ***
## qsec          1.2259     0.2887   4.247 0.000216 ***
```

```
## amManual      2.9358     1.4109   2.081 0.046716 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.459 on 28 degrees of freedom
## Multiple R-squared:  0.8497, Adjusted R-squared:  0.8336
## F-statistic: 52.75 on 3 and 28 DF,  p-value: 1.21e-11
```

The step method has created a model using weight, quarter-mile times, and transmission type. The adjusted R^2 value is now 0.83, so the model accounts for 83% of the variation in MPG. The model coefficients indicate that for every 1,000 pounds of additional weight, the MPG drops by 3.9 MPG. For every additional second that it takes to run the quarter-mile, the MPG increases by 1.2. A manual transmission will, all else being equal, give a 2.9 MPG increase.And finally, the p-values for all of the variables are below 0.05.

Residuals

```r
par(mfrow=c(2, 2))
plot(adjustedModel)
```

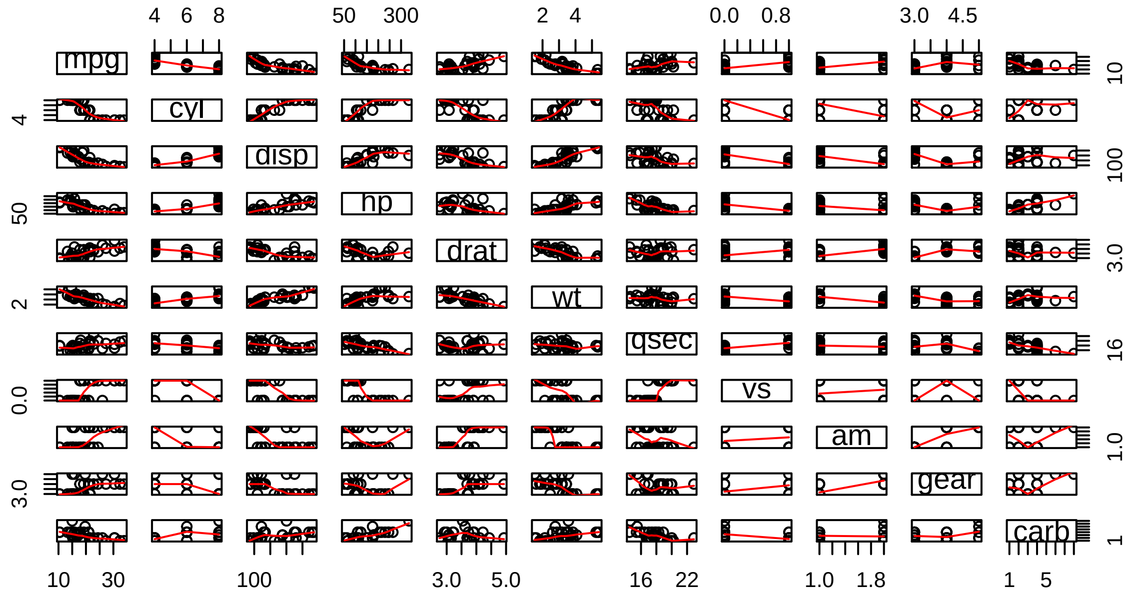


Residuals vs Fitted shows that there is no trend among the residuals, indicating that they are independent. The Normal Q-Q plot the points fall mostly along the line indicating that the residuals are normally distributed.

Appendix

```r
require(stats); require(graphics)
pairs(mtcars, panel = panel.smooth, main = "Figure 1: Scatterplot Matrix")
```

Figure 1: Scatterplot Matrix




```r
boxplot(mpg~am, data = mtcars,
        xlab = "Transmission",
        ylab = "Miles per Gallon",
        main = "Figure 2: MPG by Transmission Type")
```

Figure 2: MPG by Transmission Type