Regression Course Assignment

Karl Konz

November 6, 2016

Executive Summary

This project aims to answer the question is automatic or manual transmissions better for MPG and quantify the MPG difference between automatic and manual transmissions

Exploratory Analysis

Let's first examine the structure of the data

```
str(mtcars)
```

```
## 'data.frame':    32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
##  $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
##  $ qsec: num  16.5 17 18.6 19.4 17 ...
##  $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
##  $ am  : num  1 1 1 0 0 0 0 0 0 0 ...
##  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

Next, let's conduct a t test for the mean MPG between automatic and manual transmitions to see if there is a difference.

```
options(warn=-1)
set.seed(1234)
mtcars$am <- as.character(mtcars$am)
t.test(subset(mtcars,mtcars$am=="1",select=c(mpg))[,1]
       , subset(mtcars,mtcars$am=="0",select=c(mpg))[,1])
```

```
##
##  Welch Two Sample t-test
##
## data:  subset(mtcars, mtcars$am == "1", select = c(mpg))[, 1] and subset(mtcars, mtcars$am == "0", s
## t = 3.7671, df = 18.332, p-value = 0.001374
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   3.209684 11.280194
## sample estimates:
## mean of x mean of y
##  24.39231  17.14737
```

We can see that from above the t test for automatic and manual transmissions in the mtcars data yeilds a p value below the level of significance of $\alpha = .05$. This suggests we have evidence to reject the null hypothesis that the mean difference between the 2 transmission types are 0. We can therefore reject the null hypothesis for this test. Next we will use Stepwise comparison using AIC to select the final model.

```
fit <- lm(mpg ~ ., data = mtcars);fit <- step(fit)
```

Next test if there is a significant difference if automatic or manual transmissions term in this model performing a wald test.

```
library(aod)
wald.test(b = coef(fit), Sigma = vcov(fit), Terms = 4)
```

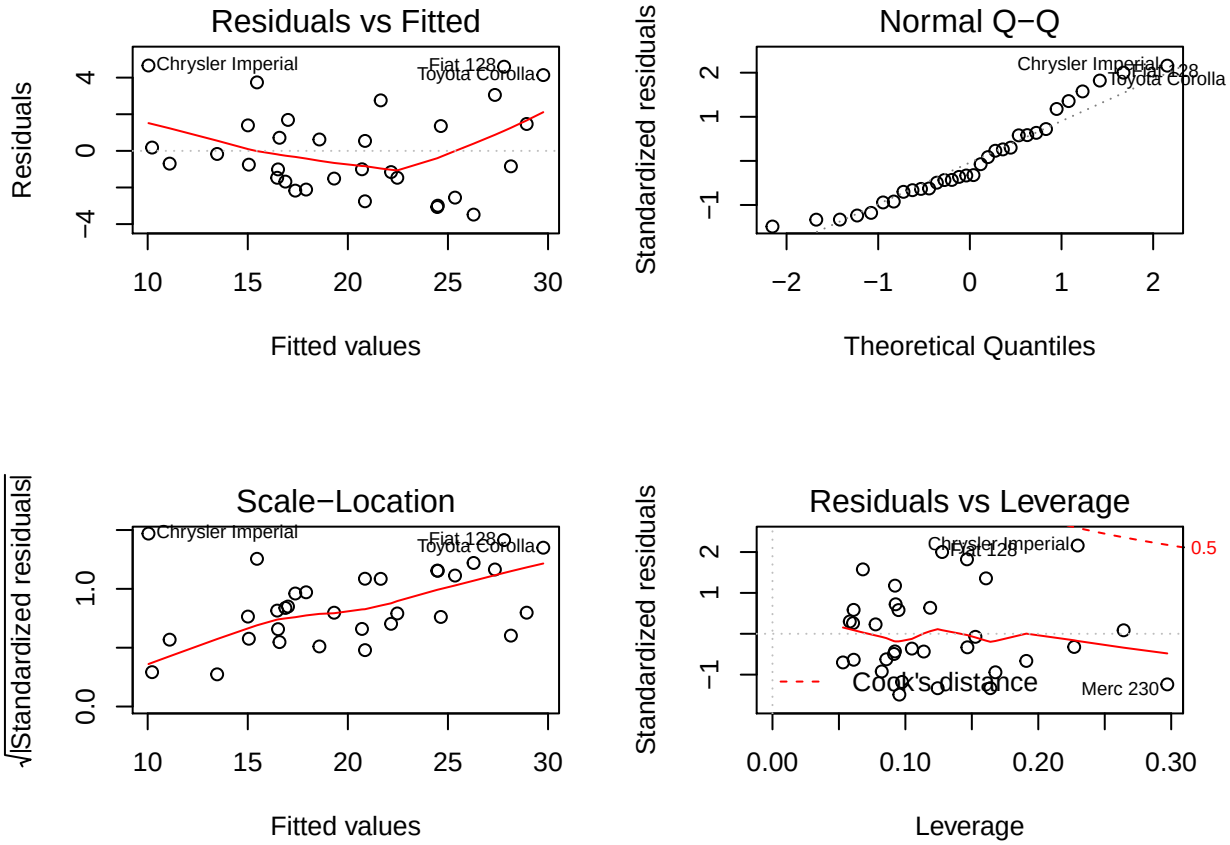
```
## Wald test:
## ----------
##
## Chi-squared test:
## X2 = 4.3, df = 1, P(> X2) = 0.037
```

From the output above we can see there is a slight but significant, at a 5% level of significance, for the am term in this model. Finally we will look at residual plots and coefficients for the final model called fit.

```
summary(fit)$coef
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)  9.617781  6.9595930  1.381946 1.779152e-01
## wt          -3.916504  0.7112016 -5.506882 6.952711e-06
## qsec         1.225886  0.2886696  4.246676 2.161737e-04
## am1          2.935837  1.4109045  2.080819 4.671551e-02
```

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



Conclusion

This analysis shows there is evidence of a difference in the MPG between automatic and manual transmissions. Based on the final model the MPG is roughly 2.94 higher for manual transmissions than for automatics when holding the other important predictors constant. Doing a simple t test showed a difference of about 7 between the 2 transmission types. The residual plots show that the residuals meet the assumptions for normality and homoskedasticity.