

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

```
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])$p.value
```

```
## [1] 0.001373638
```

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.

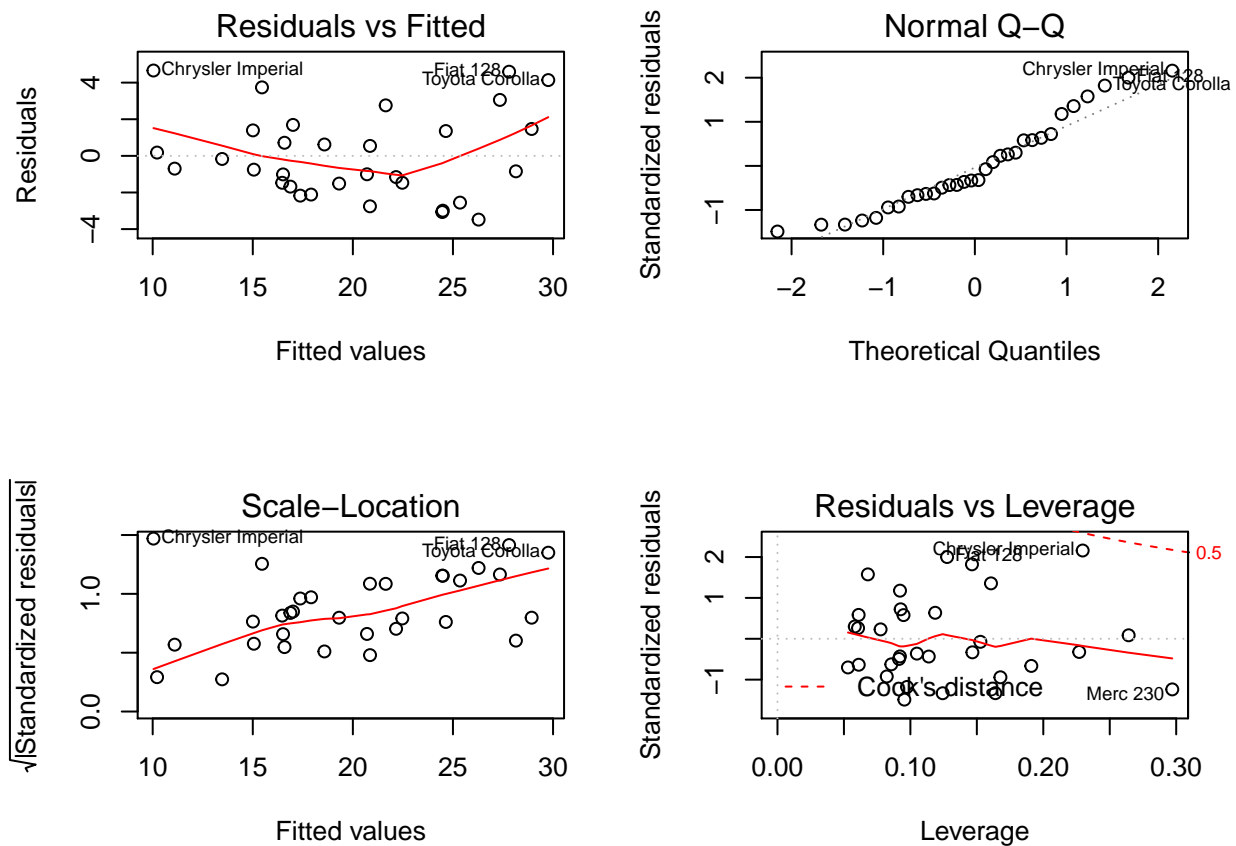
```
library(aod)
test <- wald.test(b = coef(fit), Sigma = vcov(fit), Terms = 4)
```

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 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)
```



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. The residual plots show that for the most part the residuals meet the assumptions for normality and homoskedasticity.