Regression Model

Zheng Wu

June 21, 2015

Executive Summary

Motor Trend is a magazine about the automobile industry. Looking at a data set of a collection of cars, they are interested in exploring the relationship between a set of variables and miles per gallon (MPG) (outcome). They are particularly interested in the following two questions:

1."Is an automatic or manual transmission better for MPG"

2."Quantify the MPG difference between automatic and manual transmissions"

With hypothesis tetsing and linear regression, this report presents answers to the two questions of interest that there is a significant difference between the MPG for automatic and manual transmission cars.

Reading in data

```
data(mtcars)
```

Change am into a factor variable

```
mtcars$am <- as.factor(mtcars$am)
levels(mtcars$am) <- c("Automatic", "Manual")
```

Exploratory data analysis

A boxplot to get an initial idea of whether mpg for automatic and manual is different is shown in the appendix

Hypothesis testing of mean

```
auto <- mtcars[mtcars$am == "Automatic",]
manual <- mtcars[mtcars$am == "Manual",]
t.test(auto$mpg, manual$mpg)
```

```
##
##  Welch Two Sample t-test
##
## data:  auto$mpg and manual$mpg
## 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:
##  -11.280194  -3.209684
## sample estimates:
## mean of x mean of y
##  17.14737  24.39231
```

This suggests that manual transmission cars has better mpg in great confidence

Regression Models

Simple linear Regression

```
model <- lm(mpg ~ am, data = mtcars)
coef(model)
```

```
## (Intercept)    amManual
##   17.147368    7.244939
```

```
summary(model)$adj.r.squared
```

```
## [1] 0.3384589
```

The R^2 is not very good for this model, suggesting we need to take into account more variables and build a further linear regression model.

Further linear Regression using AIC

```
further.model <- step(lm(data = mtcars, mpg ~ .),trace=0,steps=12)
coef(further.model)
```

```
## (Intercept)          wt        qsec    amManual
##    9.617781   -3.916504    1.225886    2.935837
```

This suggest that wt, qsec, am are the most important variables in explaining the variation in mpg.

Final model

```
final.model <- lm(mpg~am + wt + qsec, data = mtcars)
anova(model, final.model)['Pr(>F)']
```

```
##     Pr(>F)
## 1
## 2 1.55e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

This suggests final.model is better than the initial simple model. The diagnostics plot are shown in the appendix.

Summary

```
summary(final.model)
```

```
##
## Call:
## lm(formula = mpg ~ am + wt + qsec, 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
## amManual      2.9358     1.4109   2.081 0.046716 *
## wt           -3.9165     0.7112  -5.507 6.95e-06 ***
## qsec          1.2259     0.2887   4.247 0.000216 ***
## ---
## 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
```

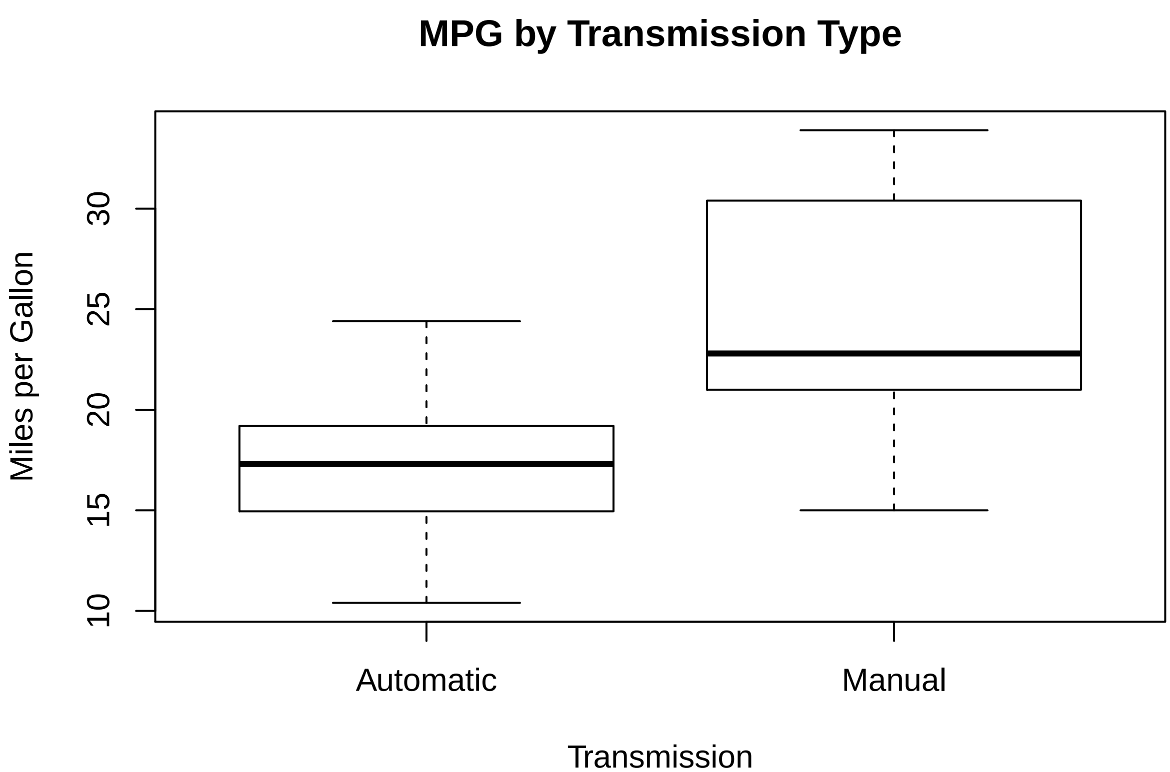
We see that manual transmission have about 2.94 higher MPG than automatic transmission cars taken into account the confound of wt and qsec. This is lower than if we did not adjust for wt and qsec. Therefore, we can answer the first question: Maunal transmission has better MPG.

To answer the second question, we would need more information about wt and qsec

Appendix

Exploratory Analysis plot

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



Diagnostic plots

```r
par(mfrow = c(2,2))
plot(final.model)
```