Peer Graded Assignment: Regression Models Course Project

Abhishek Anand

October24, 2019

Executive Summary

Motor Trend is a magazine about the automobile industry. It is interested in exploring the relationship between a set of variables and miles per gallon (MPG) (outcome), particularly:

- “Is an automatic or manual transmission better for MPG”
- “Quantify the MPG difference between automatic and manual transmissions”

Using a data set from Motor Trend Magazine along with linear regression and hypothesis testing, it can be concluded that there is a significant difference between the MPG of automatic and manual transmission cars.

To quantify the MPG difference between automatic and manual transmission cars, a linear regression model that took into account the weight, the type of transmission and the acceleration (qsec) was used. Controlling for these factors, manual transmission cars have a better fuel efficiency of 2.94 MPG more than automatic transmission cars.

Load Necessary Libraries

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.2.5
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.2.5
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
##     filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##     intersect, setdiff, setequal, union
```

Reading the Data

```
data(mtcars)
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 ...
```

Processing the Data

Convert “am” from numeric variable to a factor variable. “AT” = Automatic Transmission, “MT” = Manual Transmission.

```
mtcars$am<-as.factor(mtcars$am)
levels(mtcars$am)<-c("AT", "MT")
```

Exploratory Data Analysis

Mean of Automatic and Manual Transmission cars:

```
aggregate(mpg~am, data=mtcars, mean)
```

```
##   am      mpg
## 1 AT 17.14737
## 2 MT 24.39231
```

The mean MPG of manual transmission cars is 7.245 MPG higher than that of automatic transmission cars. Is this difference significant?

Running a t-test:

```
atData<-mtcars[mtcars$am == "AT",]
mtData<-mtcars[mtcars$am == "MT",]
t.test(atData$mpg, mtData$mpg)
```

```
##
##  Welch Two Sample t-test
##
## data:  atData$mpg and mtData$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
```

The p-value of the t-test is 0.001374, which falls within the 95% confidence interval. Hence, controlling for all other variables, there is a significant difference between the mean MPG of automatic and manual cars.

Histogram of the mpg for AT and MT cars

```
ggplot(data = mtcars, aes(mpg)) + geom_histogram() + facet_grid(.~am) + labs(x = "Miles per Gallon", y = "Frequency", title = "MPG Histogram for AT and MT cars")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

Boxplot of the mpg for AT and MT cars

```
ggplot(data = mtcars, aes(am,mpg)) + geom_boxplot() + labs(x= "Transmission", y = "MPG", title = "MPG: AT vs MT")
```

Correlations:

```
corr <- select(mtcars, mpg,cyl,disp,wt,qsec, am)
```

```
pairs(corr)
```

Linear Models

Model 1: Regress mpg against am

```
fit_1 <-lm(mpg~am, data = mtcars)
summary(fit_1)
```

```
##
## Call:
## lm(formula = mpg ~ am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.3923 -3.0923 -0.2974  3.2439  9.5077
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   17.147      1.125  15.247 1.13e-15 ***
## amMT           7.245      1.764   4.106 0.000285 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.902 on 30 degrees of freedom
## Multiple R-squared:  0.3598, Adjusted R-squared:  0.3385
## F-statistic: 16.86 on 1 and 30 DF,  p-value: 0.000285
```

From this simple linear regression model of mpg against am, manual transmission cars have 7.24 MPG more than automatic transmission cars. The R^2 value of this model is 0.3598, meaning that it only explains 35.98% of the variance.

Model 2: Using the Step Function

```
fit_2 = step(lm(data = mtcars, mpg ~ .),trace=0,steps=10000)
summary(fit_2)
```

```
##
## 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 ***
## amMT          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
```

This model uses a step algorithm to pick the variables that affect the mpg of cars the most. From the model, the weight, acceleration as well as the transmission mode affect the mpg of the car the most.

Based on this multivariate regression model, a manual transmission car has a fuel efficiency of 2.94 MPG higher than that of automatic transmission cars. The adjusted R^2 of the model is 0.834, meaning that 83% of the variance in mpg can be explained by the model.

ANOVA of the 2 Models

```
fit_step<-lm(mpg~ am + wt + qsec, data = mtcars)
anova(fit_1, fit_step)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + wt + qsec
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
```

```
## 1     30 720.90
## 2     28 169.29  2    551.61 45.618 1.55e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-value indicates that we should reject the null hypothesis that the means from both models are the same. That is, the weight and acceleration of the car have a significant impact on it's MPG.

Conclusion

In conclusion, holding the weight and acceleration (qsec) of the car constant, manual transmission cars offer 2.94 MPG better fuel efficiency.

Appendix

Model Residuals

```
par(mfrow = c(2,2))
plot(fit_2)
```

From the plot of the residuals, we observe that while there are a few outliers, there is nothing significant that skews the data.