---

title: "Week 4 Project -- Regression Analaysis"

author: "Dhruv Singh"

date: "February 16, 2020"

output:

pdf\_document: default

html\_document: default

---

## PART 0: SETUP

echo settings for embedding code

```{r setup, include=FALSE}

knitr::opts\_chunk$set(echo = TRUE)

```

Setting Directory

```{r dir}

getwd()

setwd("C:/Dhruv/misc/data/R\_7\_regression\_models/wk4\_logistic\_reg\_poisson\_reg")

```

[1] "C:/Dhruv/misc/data/R\_7\_regression\_models/wk4\_logistic\_reg\_poisson\_reg"

## Step 1: Coefficients

Loading and checking mtcars data

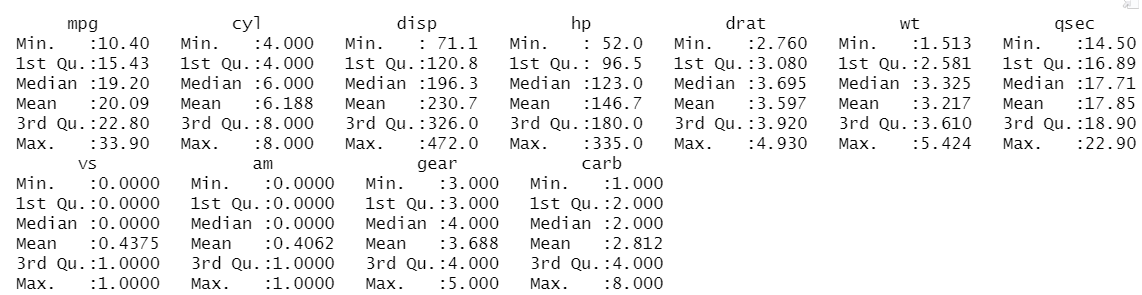
```{r mtcars}

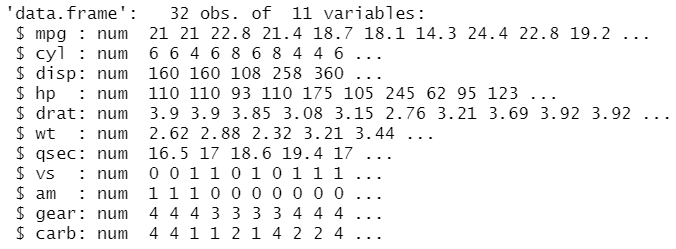
data("mtcars")

summary(mtcars)

str(mtcars)

```





```{r am}

fit <- lm(mpg ~ am, mtcars)

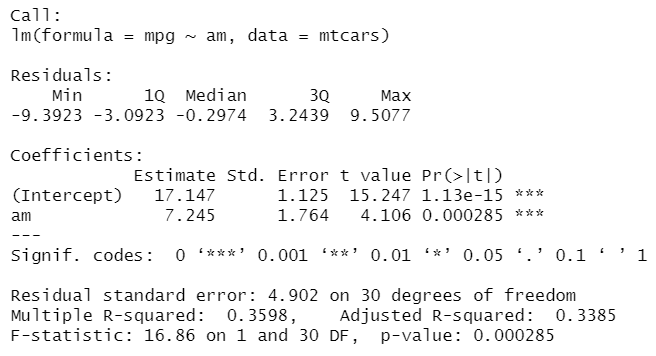
summary(fit)

# a simple two variable regregression reveals that am has a significant bearing on mpg

# binary input, 0: automatic, 1: manual

# manual is related to 7 more miles per gallon on average

```



## Step 2: Exploratory data analysis

```{r eda}

library(ggplot2)

# plotting mpg against wt

p1 <- ggplot(mtcars, aes(x = mpg)) + geom\_bar()

p1 + facet\_wrap(~am)

# from the graph below it appears that on average, manual cars yield higher miles per gallon than automatic, counterintuitively

```



## Step 3: Model fitting

```{r regression model}

# model 1

fit1 <- lm(mpg ~ am, mtcars)

summary(fit1)

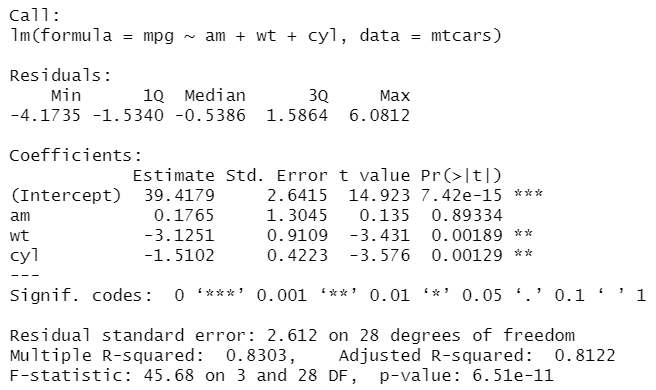
# model 2, seems to explain away the change attributable to am

# and instead attributes it to weight, and cylinders

fit2 <- lm(mpg ~ am+wt+cyl, mtcars)

summary(fit2)

```



## Step 4: questions of interest

```{r }

# thus we can see that after controlling for other related variables such as weight and cylinders

# the size of the effect of automatic vs manual reduces, and is no longer significant

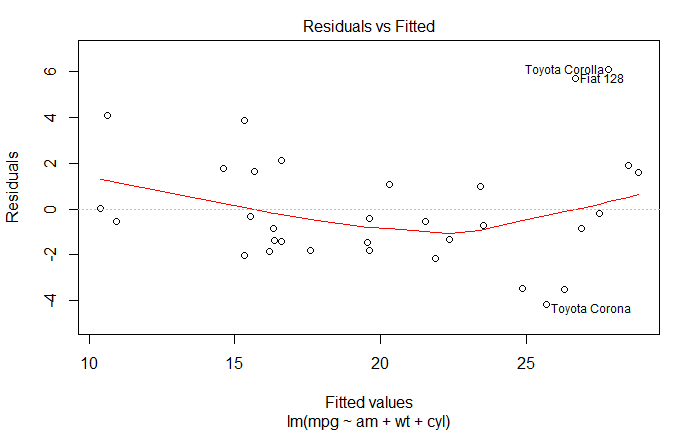
```

## Step 5: residual plot

```{r residual plot}

plot(fit2, which = 1)

```

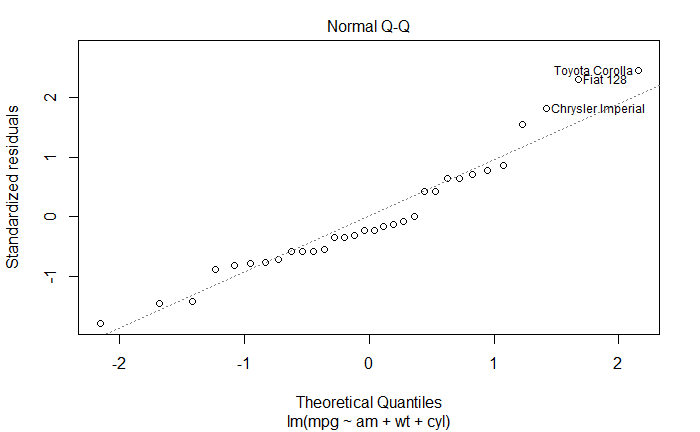


## Step 5: diagnostic plot

```{r residual plot}

plot(fit2, which = 2)

```



## Step 6: inference, uncertainty

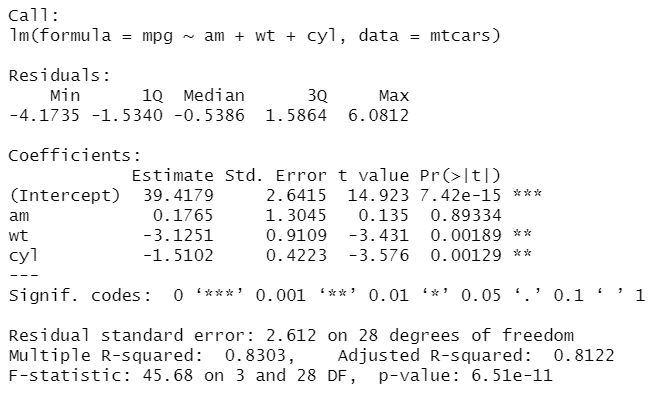
```{r inference}

summary(fit2)

# std. error of am is 1.3 and is larger than its coefficient of 0.179

# which is clearly indication that the am predictor is not significant

```



## Step 7: Report Length Criteria - 2 pages or more

## Step 8: Executive Summary

```{r executive summary}

# The model fit summary and related diagnostics are a clear indication that in order to select our predictor variables

# carefully, we can turn to a variety of methods.

# Some of these include factor analysis, as a form of unsupervised learning.

# but also vif factors, to indicate which coefficients have a larger or smaller effect on the outcome

# and helps parse out autocorrelation, that is within model correlations between coefficients.

```

## Step 9: Rmd, knitr

```{r rmd knitr}

# code all written in rmd, as visible by the code chunks

# knitr package on available for installation on system

# however, i have used it before and have published to rpubs for prev assignments

# from a diff machine.

# thanks!

```