Project 1

Briggs, Thomas

7FEB24

This is a presentation of hypothesis testing and linear regression, using the data set mtcars. We will be exploring mtcars to find if there is a difference between automatic and manual transmissions, and if one type of car gets better miles per gallon than the other

```
data(mtcars)
summary(mtcars)
```

```
##       mpg             cyl             disp             hp
##  Min.   :10.40   Min.   :4.000   Min.   : 71.1   Min.   : 52.0
##  1st Qu.:15.43   1st Qu.:4.000   1st Qu.:120.8   1st Qu.: 96.5
##  Median :19.20   Median :6.000   Median :196.3   Median :123.0
##  Mean   :20.09   Mean   :6.188   Mean   :230.7   Mean   :146.7
##  3rd Qu.:22.80   3rd Qu.:8.000   3rd Qu.:326.0   3rd Qu.:180.0
##  Max.   :33.90   Max.   :8.000   Max.   :472.0   Max.   :335.0
##       drat             wt             qsec             vs
##  Min.   :2.760   Min.   :1.513   Min.   :14.50   Min.   :0.0000
##  1st Qu.:3.080   1st Qu.:2.581   1st Qu.:16.89   1st Qu.:0.0000
##  Median :3.695   Median :3.325   Median :17.71   Median :0.0000
##  Mean   :3.597   Mean   :3.217   Mean   :17.85   Mean   :0.4375
##  3rd Qu.:3.920   3rd Qu.:3.610   3rd Qu.:18.90   3rd Qu.:1.0000
##  Max.   :4.930   Max.   :5.424   Max.   :22.90   Max.   :1.0000
##        am              gear            carb
##  Min.   :0.0000   Min.   :3.000   Min.   :1.000
##  1st Qu.:0.0000   1st Qu.:3.000   1st Qu.:2.000
##  Median :0.0000   Median :4.000   Median :2.000
##  Mean   :0.4062   Mean   :3.688   Mean   :2.812
##  3rd Qu.:1.0000   3rd Qu.:4.000   3rd Qu.:4.000
##  Max.   :1.0000   Max.   :5.000   Max.   :8.000
```

```
head(mtcars)
```

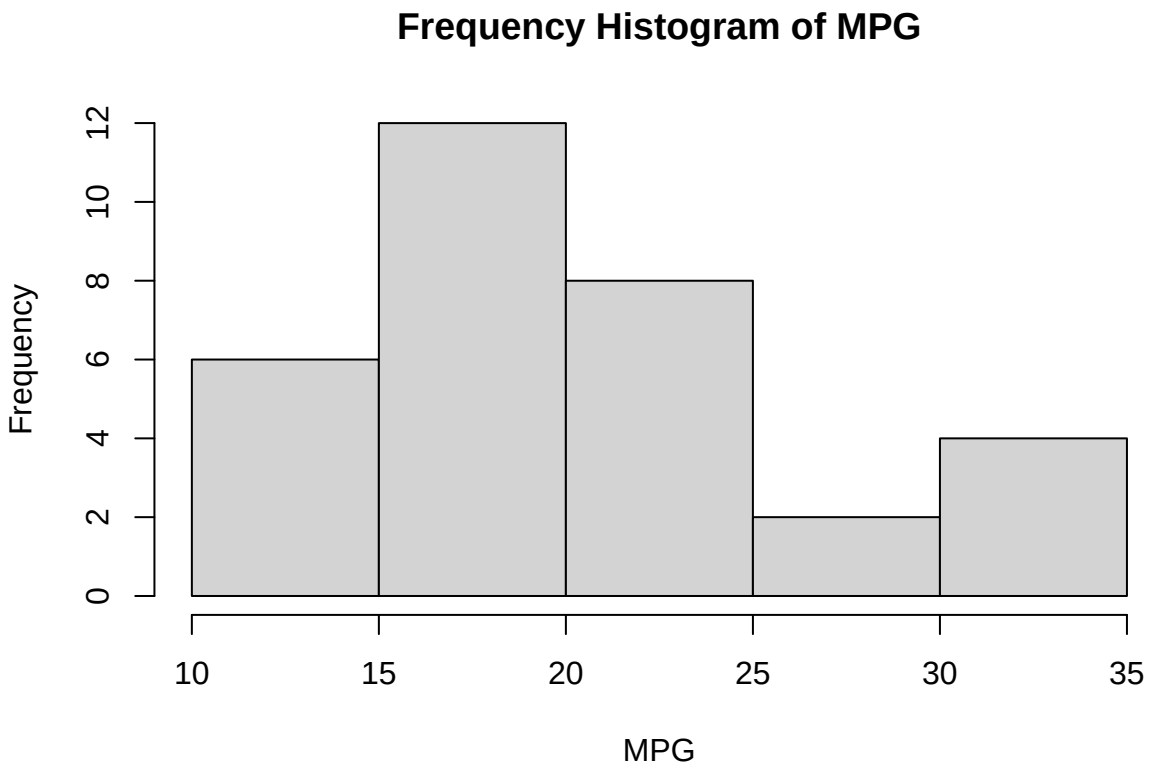
```
##                   mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

Exploring the Data

Histogram

First, we want to see if the variable mpg has a semblance of normal distribution. We will do that with a frequency histogram.

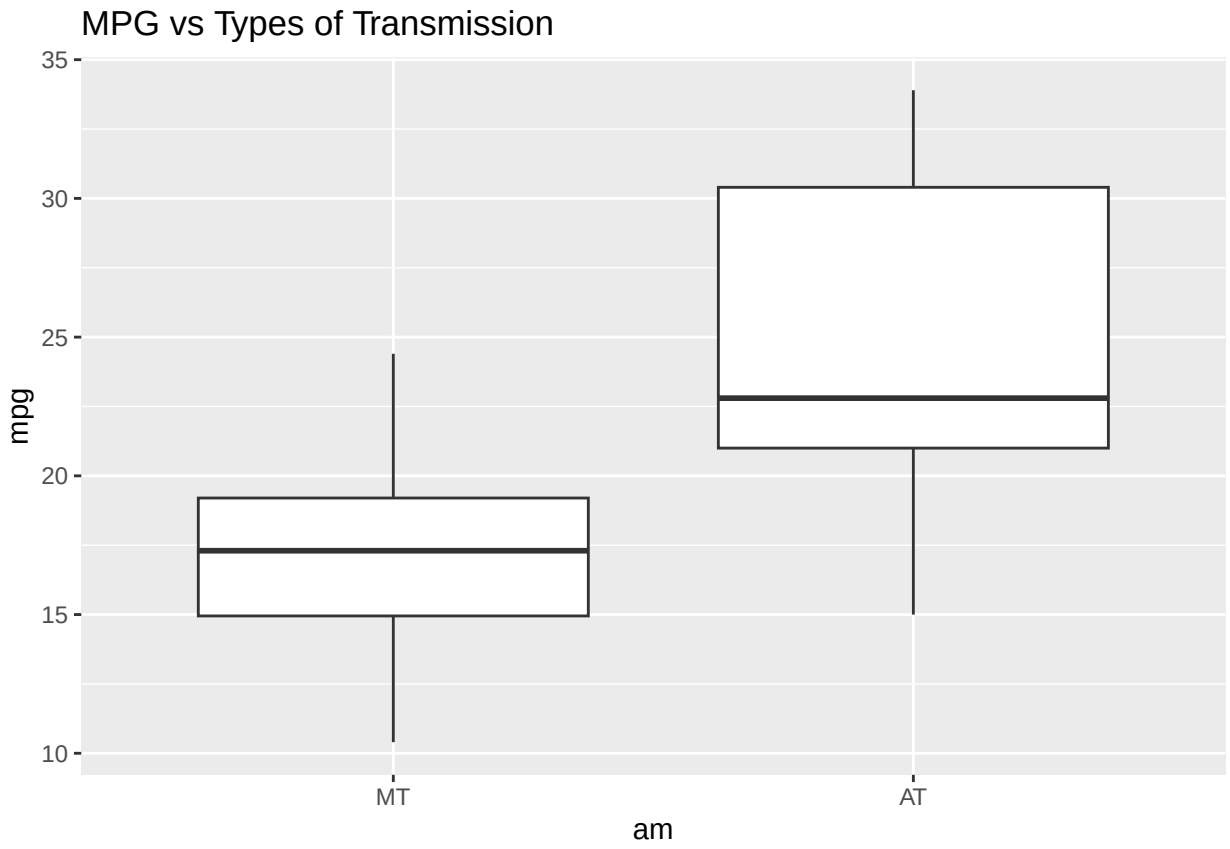
```
hist(mtcars$mpg,breaks = 5,xlab="MPG", main = "Frequency Histogram of MPG")
```



Boxplot

Next, we want to see a comparison between manual and automatic transmission and mpg. We do that with a box plot of mpg vs am

```
input<-mtcars
input$am <- as.factor(input$am)
levels(input$am) <-c("MT", "AT")
ggplot(input, aes(x=am, y=mpg)) + geom_boxplot() + ggtitle(label = "MPG vs Types of Transmission")
```

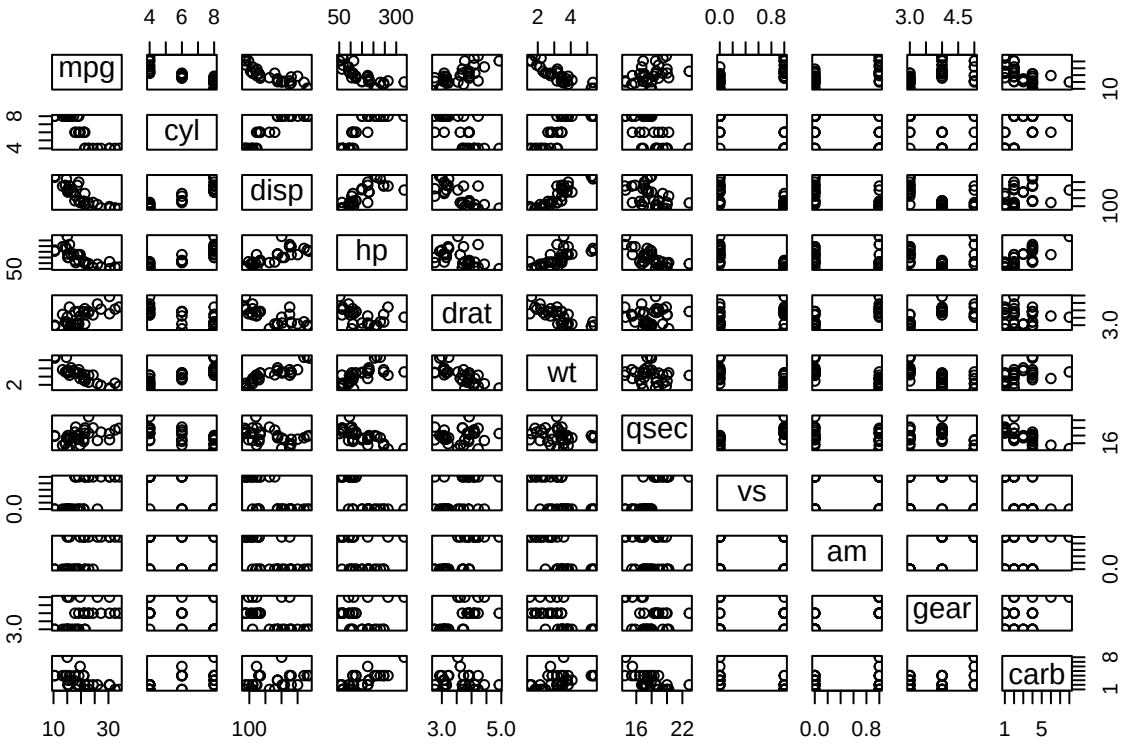


This plot indicates to us that manual is better for mpg than automatic

Pair Plot

Finally, we use a pair plot to show relationships between mpg and all the variables in the data set. We want to see how strongly other variables effect mpg.

```
pairs(mpg ~ ., data = mtcars)
```



Our pair plot shows a strong relationship between mpg and am (transmission), wt (weight), and qsec (acceleration over a quarter mile)

Testing to Determine Which Transmission is Better

While our box plot seems to suggest that manual transmission is better for miles per gallon than automatic, we want to try and prove it. We can do that with a t-test.

```
set.seed(1234)
t.test(input$mpg~input$am)
```

```
##
##  Welch Two Sample t-test
##
## data:  input$mpg by input$am
## t = -3.7671, df = 18.332, p-value = 0.001374
## alternative hypothesis: true difference in means between group MT and group AT is not equal to 0
## 95 percent confidence interval:
##  -11.280194  -3.209684
## sample estimates:
## mean in group MT mean in group AT
##         17.14737         24.39231
```

The mean for our automatic transmission group is 17.14737, and the mean for our manual transmission group is 24.39231.
Our p-value is 0.001374, which means our results are significant.
Therefore, we reject the null hypothesis and say that manual transmission is better than automatic transmission for miles per gallon

Regression

Single

We next want to create a simple linear regression model with type of transmission as the predictor variable, and mpg as the outcome

```
shape_1 <- lm(formula = mpg ~ am, data = mtcars)
summary(shape_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 ***
## am             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
```

This linear model uses only one predictor, and it indicates that our model is significant.
The p-values for coefficients are small and the t values are high.
Both of these results indicate significance. However, our R^2 value is 0.36, which means that only about 36% of the variation can be explained by the model.

Multiple Regression Models

Next we use multiple variables in our model, to show additional relationships and helps us to reduce multicollinearity.

```
shape_2 <- lm(formula = mpg ~ am + wt, data = mtcars)
shape_3 <- lm(formula = mpg ~ am + wt + qsec, data = mtcars)
anova(shape_1, shape_2, shape_3)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + wt
## Model 3: mpg ~ am + wt + qsec
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     30 720.90
## 2     29 278.32  1    442.58 73.203 2.673e-09 ***
## 3     28 169.29  1    109.03 18.034 0.0002162 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The analysis of variance shows us that the addition of other variables can improve the model. We also see a decrease in residual sum of squares and in our p-values.
This indicates that our last model (shape_3) is our best choice.

Best Model

Here is a visual of shape_3

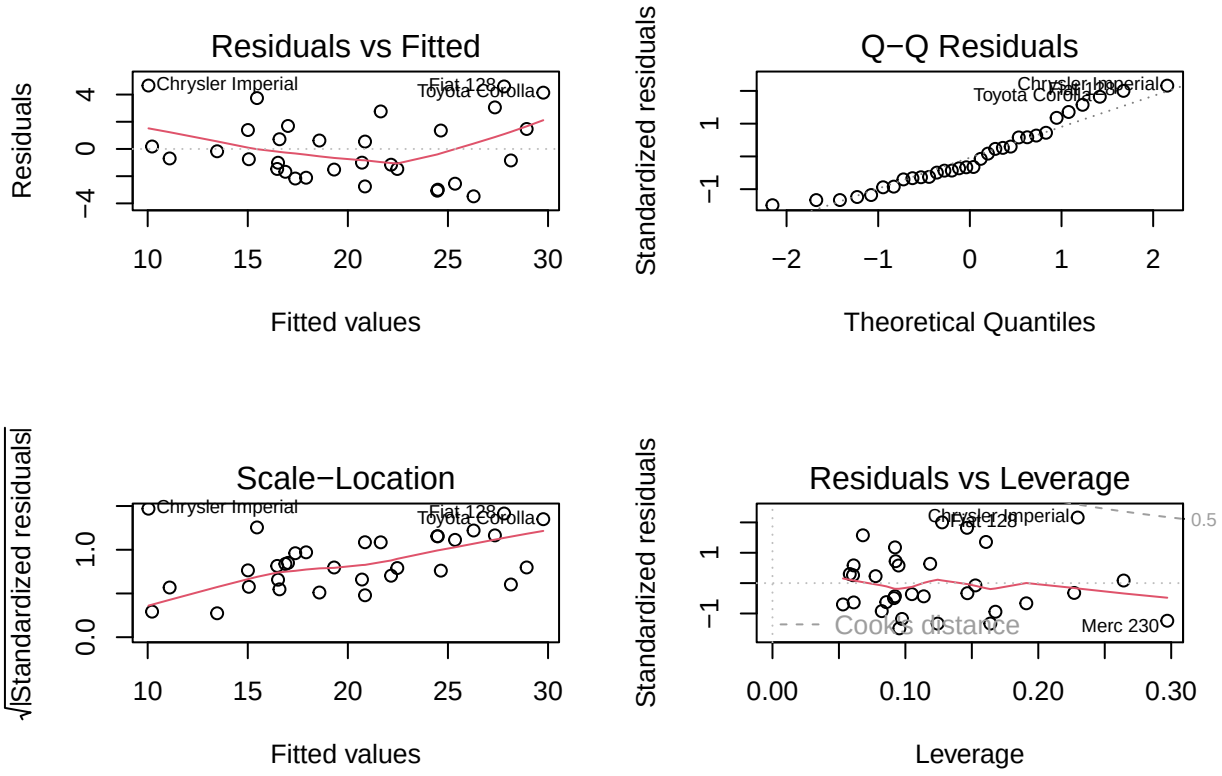
```
summary(shape_3)
```

```
##
## 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
## am            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
```

The most drastic difference between this model and our first model is the R^2 value, which has improved from just 36% to 85%

Residual Plots

```
par(mfrow = c(2,2))
plot(shape_3)
```



Conclusions

From our tests and analysis we can make several conclusions
* Manual transmission is better than automatic transmission for miles per gallon
* The three variables with the biggest influences on miles per gallon are wt, qsec, and am
* The difference in mpg between automatic and manual is about 3