MPG Difference Explained by Automatic/Manual Transmissions

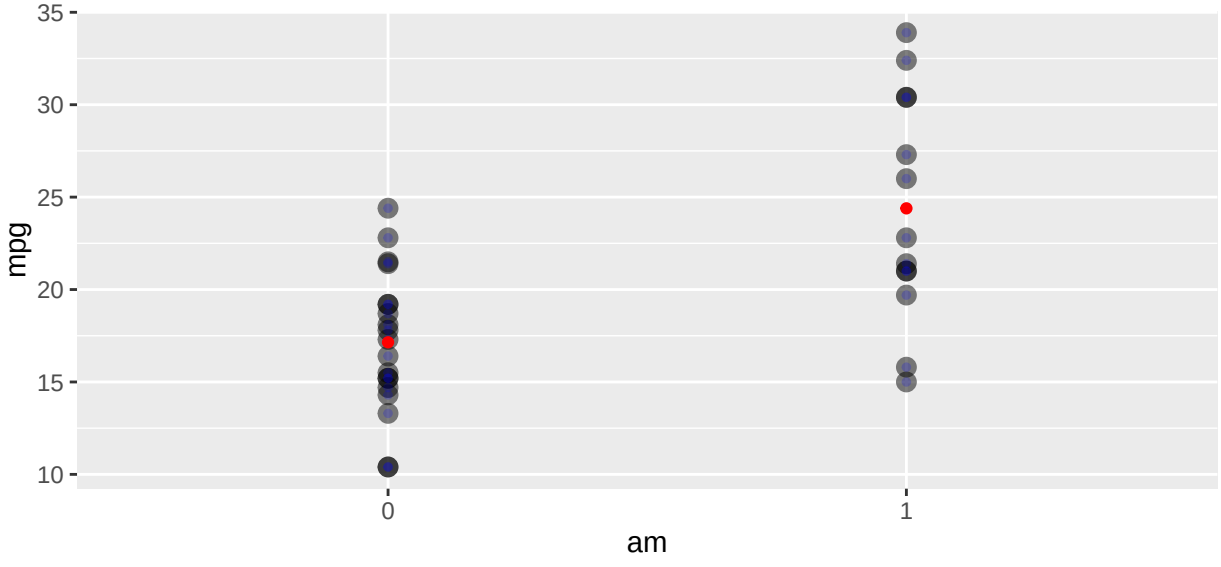
Executive Summary
In this report, we attempted to determine whether an automatic or manual transmission in a car results in better miles per gallon (mpg). We used the "mtcars" dataset to study if there is any relationship between the automatic/manual transmissions and mpg. The data contain 32 observations with 11 variables. Our initial study shows that there is a positive linear relationship between mpg performance and the automatic/manual transmissions. However, based on our residual analysis, heteroskedasticity seems to exist and we conclude that the mpg difference cannot be quantified between automatic and manual transmissions from the models we created from the "mtcars" dataset.

Exploratory Data Analysis
First, we conduct a quick review of the dataset called mtcars. From the help view, the following explanation of each variables was obtained.

```
require(datasets); data(mtcars); ?mtcars
```

[, 1] mpg Miles/(US) gallon
[, 2] cyl Number of cylinders
[, 3] disp Displacement (cu.in.)
[, 4] hp Gross horsepower
[, 5] drat Rear axle ratio
[, 6] wt Weight (1000 lbs)
[, 7] qsec 1/4 mile time
[, 8] vs V/S
[, 9] am Transmission (0 = automatic, 1 = manual)
[,10] gear Number of forward gears
[,11] carb Number of carburetors

Next we plot a graph showing miles per gallon on y-coordinate and automatic/manual transmissions as a factor on x-coordinate. The red dots indicate where the mean mpg values are in each group. It is clear that the mean mpg for the automatic transmission group (0) is lower than that of the manual transmission group (1).

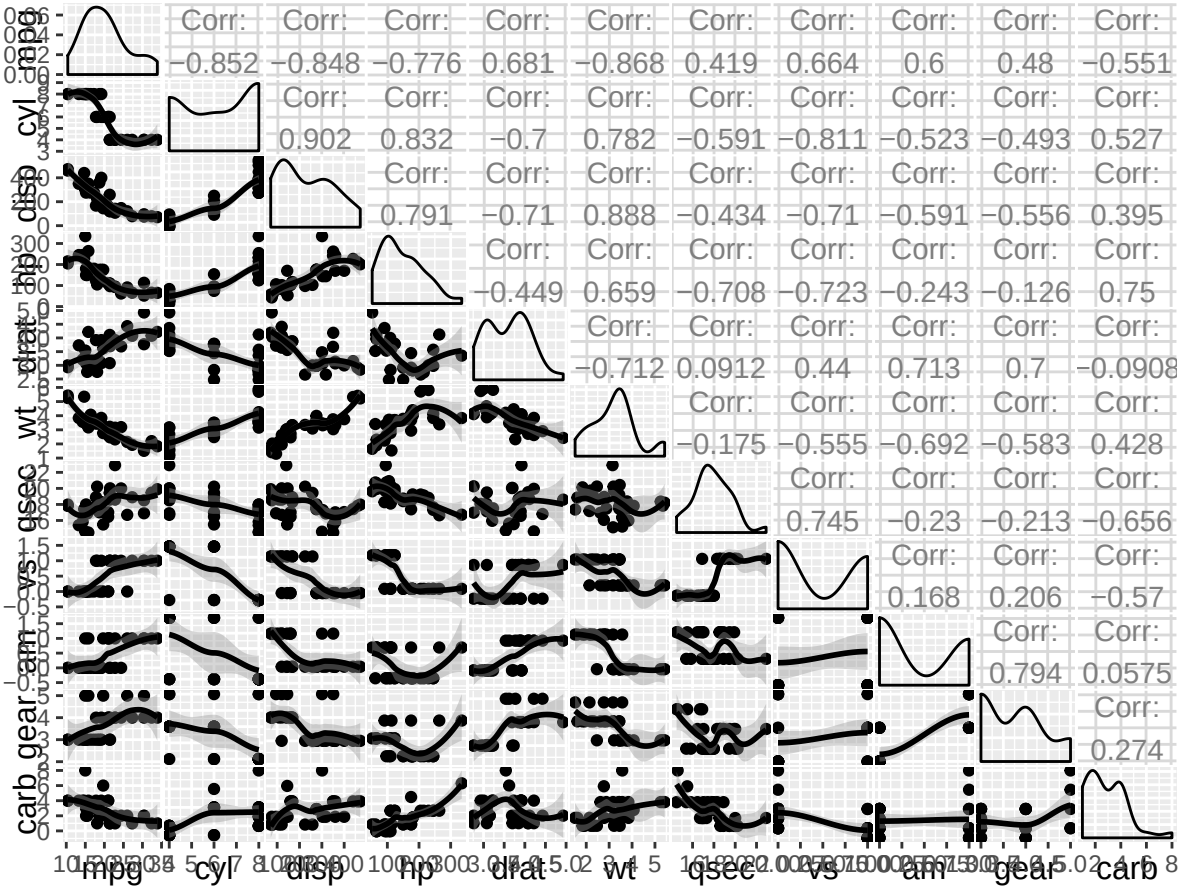


The following table contains the coefficients of the linear model of mpg regressed by am in the mtcars dataset. The slope coefficient tells us that when the transmission becomes automatic from manual, miles per gallon is expected to increase by 7.245. Also, this model's t value is 4.106. The probability of a data point observed to be as large as this t value is 0.000285, which is less than 95% significant level. Therefore, the null hypothesis stating that there is no linear relationship between mpg and am, is rejected. Finally, we also calculated the 95% confidence interval for the slope. Based on these numbers, we can say with 95% confidence that when a car's transmission changes from automatic to manual, its miles per gallon is estimated to increase from 3.642 to 10.848.

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 17.147368   1.124603 15.247492 1.133983e-15
## am           7.244939   1.764422  4.106127 2.850207e-04
```

```
## [1]  3.64151 10.84837
```

The following matrix of plots help us understand the relationships between each variable. According to this matrix, variables that have strong positive correlation with mpg are drat, vs and am, while wt has the strongest negative correlation with mpg.



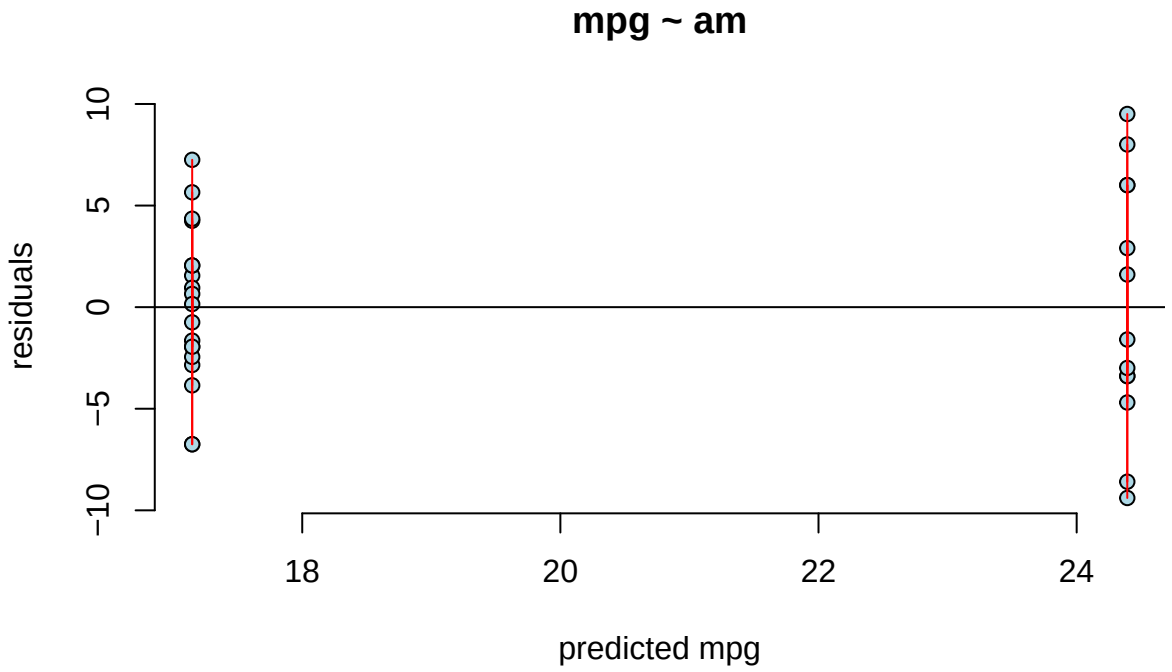
Model Selection

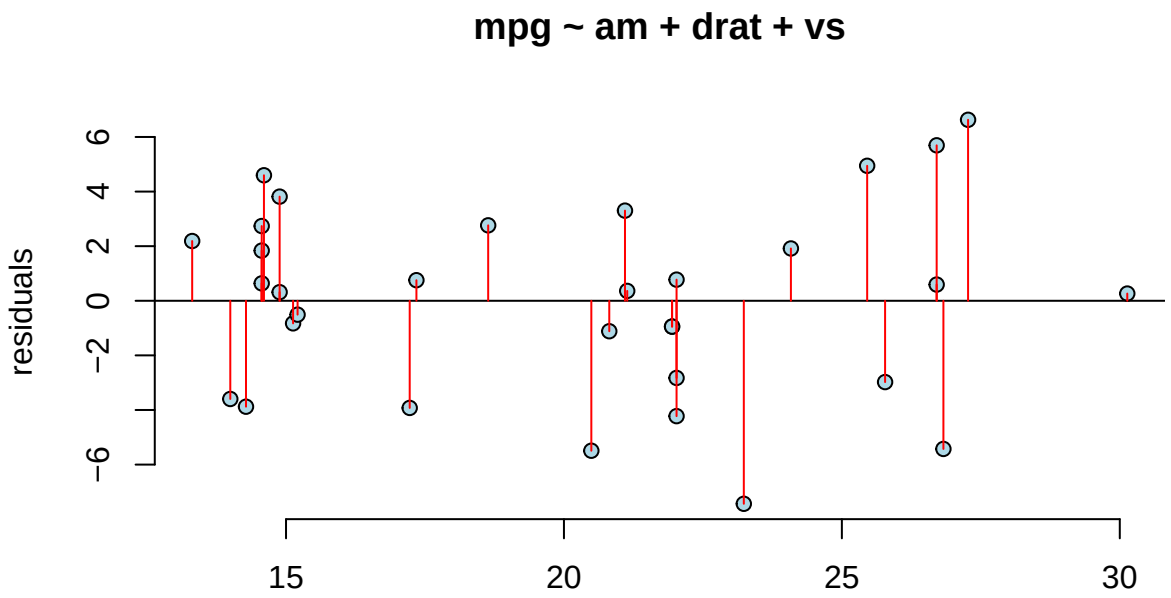
In order to select the right model, we conduct a nested model testing. Based on the strong correlation observed above, drat, vs and wt are chosen to create nested models. The first model includes only am as a regressor and the estimand is mpg. The second model uses am, drat and vs to predict mpg. Finally we add wt as a regressor to create the third model. According to the test results, all added variables are necessary.

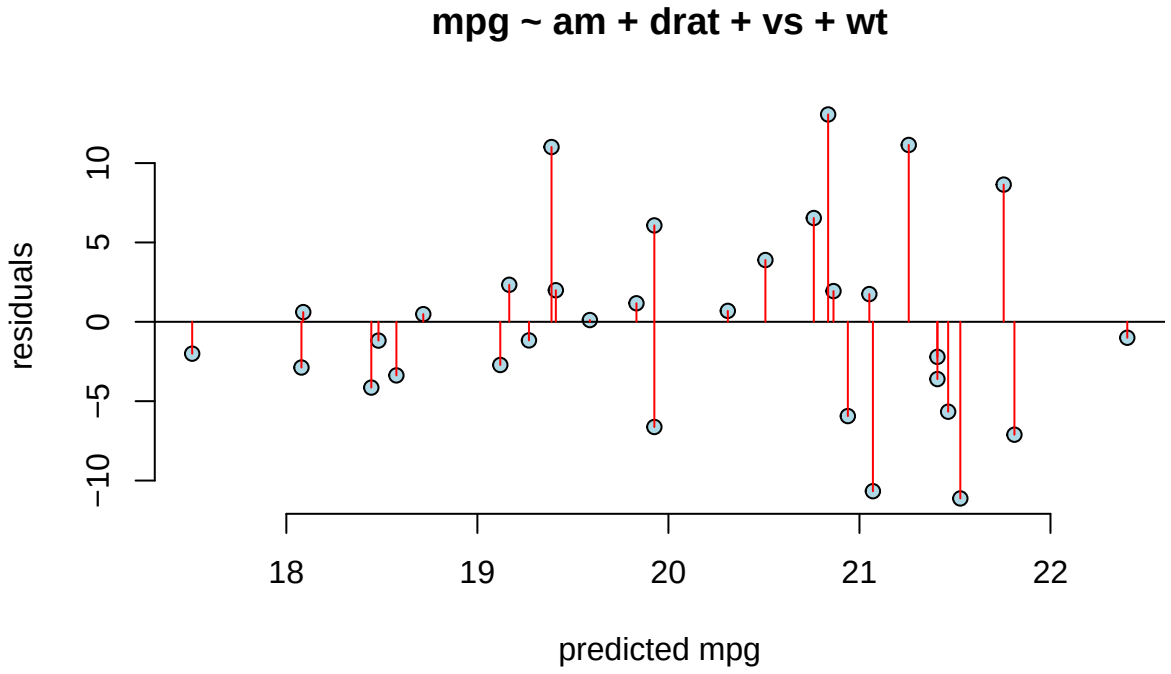
```
## Analysis of Variance Table
```

```
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + drat + vs
## Model 3: mpg ~ am + drat + vs + wt
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     30 720.90
## 2     28 339.99  2    380.91 23.941 1.046e-06 ***
## 3     27 214.79  1    125.20 15.738 0.0004828 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Before, we decide to choose the model with four variables, am, drat, vs and wt for our study, we plot residuals of the three models. All three residual plots show residul value growing as x value gets larger.







Conclusion

The residual plots from all three models above appear to show heteroskedasticity. Existence of heteroskedsticity can invalidate statistical tests of significance that assume that the modelling errors are uncorrelated and uniform. Therefore, it is not possible with this dataset to quantify the mpg difference between automatic and manual transmissions.