Regression Model - Project

Executive Summary The below analysis is to explore the relationship between a set of variables and miles per gallon (MPG) (outcome). In particular we are interested in the following:- "Is an automatic or manual transmission better for MPG" "Quantify the MPG difference between automatic and manual transmissions"

Loading and Pre-processing Data

```
library(datasets)
data(mtcars)
dt <- mtcars
for (i in 1:nrow(dt)){
  if (dt$am[i] == 0) {dt$trans[i] <- "Automatic"}
  else if (dt$am[i] == 1) {dt$trans[i] <- "Manual"}
}
```

Exploratary Data Analysis We begin the exploratory data analysis by comparing the automatic versus manual transmission based on the MPG. The data analysis figure can be found in the appendix. **Means test (ttest) to analyze mileage of automatic vs. manual transmission**

```
mpg_am <- dt[mtcars$am == 0,]$mpg
mpg_mt <- dt[mtcars$am == 1,]$mpg
t.test(mpg_am, mpg_mt)
```

```
##
##  Welch Two Sample t-test
##
## data:  mpg_am and mpg_mt
## 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
```

From the above we note that the p-value is 0.001374 which is neither 5% or 1%, hence the alternative hypothesis (the difference in means is not equal to 0) is true. The mean mileage of automatic transmission is 17.15 MPG and the manual transmission is 24.39 MPG. But, the above is based on the assumption that all the other factors(No of Cylinders, Weight, etc.) are the same or constant to both Automatic and Manual Transmission. So we need more detailed analysis to conclude. Hence we will explore more through regression models. **Simple Regression Analysis**

```
fit1 <- lm(mpg~am, data = dt)
summary(fit1)
```

```
##
## Call:
## lm(formula = mpg ~ am, data = dt)
##
## 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
```

From the above interpretation of coefficient and intercepts, we note that, on average, manual transmission cars have 7.245 MPGs more than automatic transmission.Without any additional variables,because of p-value, we can conclude that manual is better than automatic for fuel efficiency, however there could be other factors which affect the efficiency for which we need multi variate regression. **Mutli Variate Regression Analysis**

```
fit <- lm(mpg ~ . , data = dt)
fit.best <- step(fit, trace=0)
summary(fit.best)
```

```
##
## Call:
## lm(formula = mpg ~ wt + qsec + am, data = dt)
##
## 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 ***
## am            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
```

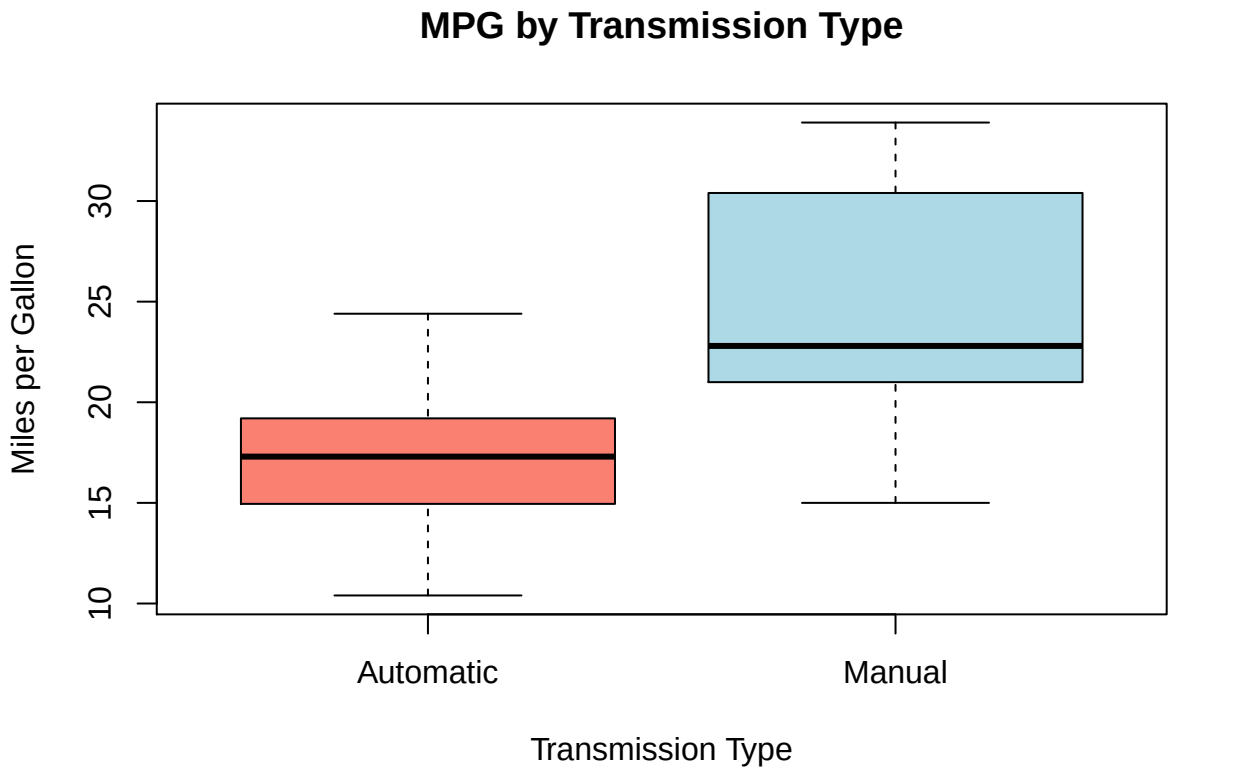
Using the more complex regression models, the multivariate analysis, we observe that in addition to transmission, weight of the vehicle as well as acceleration speed have the highest relation to explaining the variation in MPG. The adjusted R^2 is 84% which means that the model explains 84% of the variation in MPG indicating it is a robust and highly predictive model. **Conclusion** The question "Is an automatic or manual transmission better for MPG" is not answerable because, in the quantifying analysis we see that transmission cannot be the only parameter to gauge the MPG but the weight and qces have to be considered for the same.

Appendix

Exploratory Data Analysis

Comparing Automatic Vs. Manual Transmission

```
boxplot(mpg~trans, data = dt, col = c("salmon", "light blue"), xlab = "Transmission Type",
        ylab = "Miles per Gallon", main = "MPG by Transmission Type")
```



Best model residual plot

```
par(mfrow = c(2,2))
plot(fit.best)
```