Regression Models Course Project

EXECUTIVE SUMMARY

In this report, we will examine the mtcars data set and explore how miles per gallon (MPG) is affected by different variables. In particularly, we will answer the following two questions: - 1 - Is an automatic or manual transmission better for MPG, and - 2 - Quantify the MPG difference between automatic and manual transmissions.

Load the mtcars dataset and display the internal structure of the variables.

```
knitr::opts_chunk$set(echo = TRUE, fig.path = 'figures/')
library(ggplot2)
data(mtcars)
head (mtcars, 1)
```

```
##           mpg cyl disp  hp drat   wt  qsec vs am gear carb
## Mazda RX4  21   6  160 110  3.9 2.62 16.46  0  1    4    4
```

EXPLORATORY DATA ANALYSIS

Perform some basic exploratory and inferential data analysis of the data to study the relationship between transmission type (manual or automatic) and automobile fuel consumption in miles per gallon (MPG).

```
by(data = mtcars$mpg,
   INDICES = list(factor(mtcars$am, labels = c("Automatic", "Manual"))), summary)
```

```
## : Automatic
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10.40   14.95   17.30   17.15   19.20   24.40
## ------------------------------------------------------------
## : Manual
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   15.00   21.00   22.80   24.39   30.40   33.90
```

Impact of Transmission Type on Fuel Consumption FigureS A.1 in the Appendix section plot the relationship between transmission type and fuel consumption in automobiles.

REGRESSION ANALYSIS

Let's now test this hypothesis with a Simple Linear Regression Test.

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

```
##
## 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
```

The p-value is less than 0.0003, so we will not reject the hypothesis. However, the R-squared value for this test is only ~= .35, suggesting that only a third or so of variance in MPG can be attributed to transmission type alone. Let's perform an Analysis of Variance for the data

```r
T_variance_analysis <- aov(mpg ~ ., data = mtcars)
summary(T_variance_analysis)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## cyl          1  817.7   817.7 116.425 5.03e-10 ***
## disp         1   37.6    37.6   5.353  0.03091 *
## hp           1    9.4     9.4   1.334  0.26103
## drat         1   16.5    16.5   2.345  0.14064
## wt           1   77.5    77.5  11.031  0.00324 **
## qsec         1    3.9     3.9   0.562  0.46166
## vs           1    0.1     0.1   0.018  0.89317
## am           1   14.5    14.5   2.061  0.16586
## gear         1    1.0     1.0   0.138  0.71365
## carb         1    0.4     0.4   0.058  0.81218
## Residuals   21  147.5     7.0
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
T_multivar <- lm(mpg ~ cyl + disp + wt + am, data = mtcars)
summary(T_multivar)
```

```
##
## Call:
## lm(formula = mpg ~ cyl + disp + wt + am, data = mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.318 -1.362 -0.479  1.354  6.059
##
```

```
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 40.898313   3.601540  11.356 8.68e-12 ***
## cyl         -1.784173   0.618192  -2.886  0.00758 **
## disp         0.007404   0.012081   0.613  0.54509
## wt          -3.583425   1.186504  -3.020  0.00547 **
## am           0.129066   1.321512   0.098  0.92292
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.642 on 27 degrees of freedom
## Multiple R-squared:  0.8327, Adjusted R-squared:  0.8079
## F-statistic: 33.59 on 4 and 27 DF,  p-value: 4.038e-10
```

From the above Analysis of Variance, we can look for p-values of less than 0.5. This gives us cyl, disp, and wt to consider in addition to transmission type (am)

```r
T_multivar <- lm(mpg ~ cyl + disp + wt + am, data = mtcars)
summary(T_multivar)
```

```
##
## Call:
## lm(formula = mpg ~ cyl + disp + wt + am, data = mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.318 -1.362 -0.479  1.354  6.059
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 40.898313   3.601540  11.356 8.68e-12 ***
## cyl         -1.784173   0.618192  -2.886  0.00758 **
## disp         0.007404   0.012081   0.613  0.54509
## wt          -3.583425   1.186504  -3.020  0.00547 **
## am           0.129066   1.321512   0.098  0.92292
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.642 on 27 degrees of freedom
## Multiple R-squared:  0.8327, Adjusted R-squared:  0.8079
## F-statistic: 33.59 on 4 and 27 DF,  p-value: 4.038e-10
```

This Multivariable Regression test now gives us an R-squared value of over .83, suggesting that 83% or more of variance can be explained by the multivariable model. P-values for cyl (number of cylinders) and weight are below 0.5, suggesting that these are confounding variables in the relation between car Transmission Type and Miles per Gallon.

ANALYSIS OF RESIDUALS

Figures A.2 The "Residuals vs Fitted" plot here shows us that the residuals are homoscedastic. We can also see that they are normally distributed, with the exception of a few outliers.

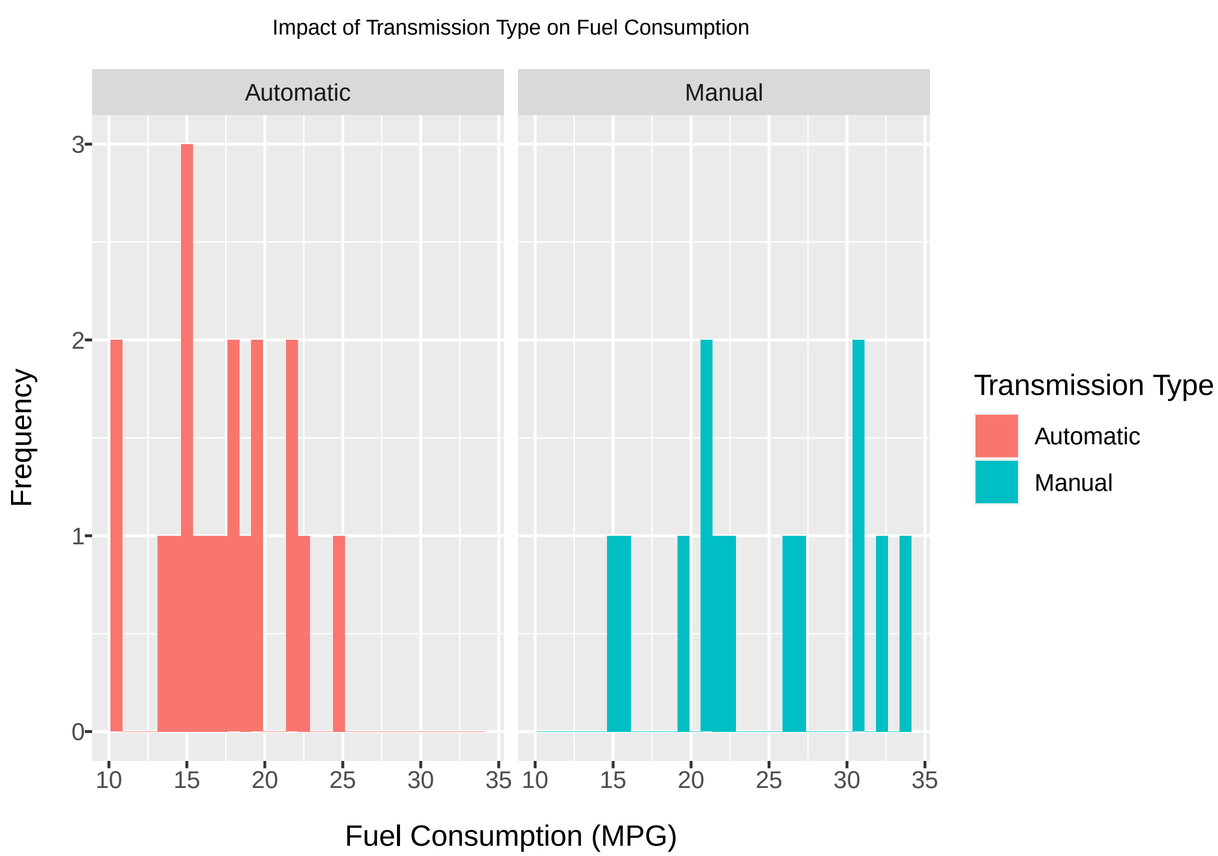
CONCLUSION

Automobiles with a manual transmission yield better gas mileage than vehicles with an automatic transmission.
Using the final multivariable regression model it's conclude that 'wt' and 'cyl'are confounding variables in the relationship between 'am and 'mpg' and that manual transmission cars on average have 1.55 miles per gallon more than automatic cars.

APPENDIX

A.1 Plot Impact of Transmission Type on Fuel Consumption

```
g <- ggplot(data = mtcars, aes(x = mpg, y = ..count..))
g <- g + geom_histogram(binwidth = 0.75,
                        aes(fill = factor(am, labels = c("Automatic", "Manual"))))
g <- g + facet_grid(. ~ factor(am, labels = c("Automatic", "Manual")))
g <- g + scale_colour_discrete(name = "Transmission Type")
g <- g + scale_fill_discrete(name = "Transmission Type")
g <- g + xlab("Fuel Consumption (MPG)") + ylab("Frequency") + theme(plot.title = element_text(size = 8,
           axis.text.x = element_text(angle = 0,hjust = 0.5,vjust = 0.5,margin = margin(b = 10, unit =
           axis.text.y = element_text(angle = 0, hjust = 0.5, vjust = 0.5, margin = margin(l = 10, unit
g
```



A.2 Residuals Plot

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