MOTOR TRENDS DATA - REGRESSION MODELS PROJECT

===

Executive Summary

The data (*mtcars*) that we are about to look at, was extracted from the 1974 Motor Trend US Magazine. It comprises fuel consumption along with 10 attributes and performances (**Appendix 2**) for 32 car models over the two years from 1973-74. We use regression analysis and data analysis to determine whether an automatic or manual transmission car is better for the MPG and quantify the difference. A simple regression model that looks at the MPG yield, ignoring all other variables other than the **am**, that the *transmission* of the car gave us an increased MPG of 7.245 if using a manual vs. automatic transmission. However the Adjusted R-squared value was only 0.3385, i.e. a poor fit. The null hypothesis (H0) that both automatic and manual vehicles yield the same MPG also failed.

Several multiple regression fitted models were tested and the conclusion was that a combination of factors/variables such as the displacement, the number of *cylinders*, the *weight* of the car, the *transmission* and the *horsepower* that were the best indicators of the MPG yields overall. The results showed that if the same car such as the Mazda RX4 was used, there was approximately 1.3 MPG increase when using the manual vs. automatic transmissions.

##EXPLORATORY DATA ANALYSIS Firstly, a standard boxplot also created (Please see **Appendix 1**) to show whether a single variable such as *transmission* could show whether there was a difference in MPG yield between the two transmission types. This clearly indicated that there was a difference and the manual transmission had an edge. To ensure the conclusion was correct, a second test was performed as follows.

A simple regression model was used to determine whether a single variable like the transmission of the vehicles could show whether automatic transmission vehicles were better than manual ones or vice versa. The following was used:

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

At face value, there appears to be a 7.245 increase in MPG yield in cars using manual transmissions over automatic ones. Yet the Adjusted R-squared fit was only 33.85%, indicating a poor fit and proving that a single variable such as *transmission* was not a good predictor of fuel efficiency.

##T-TESTING Next a T-test and hypothesis testing was used to infer whether the MPG of automatic and manual vehicles was the same, and if other variables played a part in the MPG yield. We set the null hypothesis (*H0*) as both automatic and manual vehicles yielded the same MPG. The alternate Hypothesis (*H1*) being that there is a significant enough difference between the two types of transmission used with other variables too playing an important part. The following was used:

```
test <- t.test(mpg ~ am, data=mtcars); test$p.value
```

[1] 0.001373638

The p-value was **0.0137**, therefore the null hypothesis *H0* was rejected.

##SELECTING THE VARIABLES TO USE

```
all <- lm(mpg~.,data=mtcars); summary(aov(all))
```

This showed us that the *cyl*, *disp*, *hp*, *drat*, *wt* and *am* were the most likely variables to use. We then tested how the *cyl*, *disp*, *hp*, *drat*, *wt* and *am* variables impact the MPG yield

##FITTING MULTIPLE REGRESSION MODELS The last part of the testing involved fitting regression models using the 6 variables that could have an impact on the MPG yield. This is done by fitting multiple regression models:

- model1 <- lm(mpg ~ cyl, data = mtcars) Adjusted R-squared: 0.7171
- model2 <- lm(mpg ~ wt, data = mtcars) Adjusted R-squared: 0.7446
- model3 <- lm(mpg ~ cyl + wt, data = mtcars) Adjusted R-squared: **0.8185**

- model4 <- lm(mpg ~ disp, data = mtcars) Adjusted R-squared: 0.7090
- model5 <- lm(mpg ~ disp + cyl, data = mtcars) Adjusted R-squared: 0.7430
- model6 <- lm(mpg ~ disp + cyl + wt, data = mtcars) Adjusted R-squared: **0.8147**
- model7 <- lm(disp + cyl + wt + am, data = mtcars) Adjusted R-squared: 0.8079
- model8 <- lm(disp + cyl + wt + am + hp, data = mtcars) Adjusted R-squared: **0.8273**
- model9 <- lm(disp + cyl + wt + am + drat, data = mtcars) Adjusted R-squared: **0.8211**

The results of the above fitted models indicated the following:

The strongest fit from the above models were model3, model6, model7, model8 and model9. The two models had the strongest fit thereby indicating a strong correlation to the MPG however, were models **8** and **9**. However adding drat, that is, rear axle ratio did not add that much value to the argument as it pulled the Adjusted R-squared down. When hp was added (but drat excluded), the model had the strongest explanation of MPG yields.

Therefore, the overall results lead us to the conclusion that all other factors remaining the same, the displacement, the number of cylinders, the weight of the car, the transmission and the horsepower combined were the best indicators of the MPG yield of a vehicle, not the type of transmission type on its own.

If a standard Mazda RX4 is used as an example, using model8, using the data in **Appendix 3**,

- 21+0.1225708(160)+(-1.1063798*4*6)+(-3.3026230*1*2.620)+(1.55649163*1*)+(-0.0276002*1*10)

gives us 22.3 vs. 21.0 MPG using a manual transmission over an automatic one.

###Appendix 1### [id]: https://github.com/aksheyp/Regression-Models/blob/master/Figures/MPG vs Transmission Boxplot Large.png "Boxplot of MPG vs Transmission"

```
boxplot(mpg ~ am, xlab="Transmission (0 = Automatic, 1 = Manual)", ylab="Miles Pe
r Gallon", col =terrain.colors(2))
title(main="Miles Per Gallon vs. Type of Transmission")
```

###Appendix 2### List of cars and attributes [id] https://github.com/aksheyp/Regression-Models/blob/master/Figures/CarsAndVariables.png "List of cars and their attributes"

```
data("mtcars")
mtcars
```

###Appendix 3### Model 8 evaluation [idt] https://github.com/aksheyp/Regression-Models/blob/master/Figures/Appendix 3.PNG "Model 8"

```
model8 <- lm(mpg  ~ disp + cyl + wt + am + hp, data = mtcars)
summary(model8)$coef
```