Regression Model Assignment

Bodhi Goswami

10/27/2021

Executive Summary

This report was prepared to explore relations of Miles per Gallon (mpg) of cars with respect to pertinent parameters a car might have. This included weight of the car, gross horsepower, type of transmission (automatic or manual), number of cylinders etc. The data set built in in R called *mtcars* was used to obtain conclusions.

Data set Overview

The data set mtcars contains info from *1974 Motor Trend US magazine* and contains performance chart of 32 vehicles from 1973 models. Below is a preview of 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
```

Exploratory Analysis

The coefficients used for Automatic/Manual transmission is *am* in the data above. The values set are 0 for automatic and 1 for manual transmission. Upon a simple boxplot shown in Figure 1, we can infer the mean milage for manual transmission cars are higher than automatic transmission cars. The points shown in Purple are the individual points plotted along with the boxplot for confirmation of the same.

Model Selection for Inference on MPG v/s Transmission Type

While Binary and Poisson models do not suit to be a good fit since this mpg is a quantifiable outcome with an upper bound per this data. The two best suited models, in my opinion, for selection were Liner Model and Generalized Linear Model.

A linear model was fit as shown below to infer impact on milage due to all the vehicular parameters in the given dataset.

```
fit<-lm(mpg~. -1,data=mtcars)
summary(fit) $coef
```

```
##         Estimate Std. Error     t value   Pr(>|t|)
## cyl   0.35082641 0.76292423  0.45984438 0.65014009
## disp  0.01354278 0.01762273  0.76848373 0.45037109
## hp   -0.02054767 0.02143989 -0.95838513 0.34828334
## drat  1.24158213 1.46276742  0.84878985 0.40513967
```

```
## wt   -3.82613150 1.86238084 -2.05443023 0.05200271
## qsec  1.19139689 0.45942323  2.59324480 0.01659185
## vs    0.18972068 2.06824861  0.09173011 0.92774262
## am    2.83222230 1.97512820  1.43394353 0.16564985
## gear  1.05426253 1.34668717  0.78285629 0.44205756
## carb -0.26321386 0.81235653 -0.32401273 0.74898869
```

Under the Estimate column, this shows the estimated of impact each of these parameters have on one unit increase in mpg. For instance, one unit increase in gross horsepower would result in 2% reduction in milage. Since this is an inverse relation, higher number of milage can be achieved with lesser horsepower. On the contrary, higher number of gears will result in higher milage per the summary output. Each additional ton weight of car will reduce the Totale milage by about 4 miles/gallon. These are very interesting facts obtained from the linear model.

As an attempt to fit liner model for mpg v/s our regressor of interest (transmission) code below was used. Inferences follow.

```
lm.fit<-lm(mpg~as.factor(am), data=mtcars)
mtcars$am<-as.factor(mtcars$am)
summary(lm.fit)$coef
```

```
##                 Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)    17.147368   1.124603 15.247492 1.133983e-15
## as.factor(am)1  7.244939   1.764422  4.106127 2.850207e-04
```

This code sets am(0) which is automatic transmission and gives an Estimate mpg of 17.14. Same can be observed in the boxplot on Figure 1. For the manual transmission the Estimated miage is 17.14+7.24, which is about 7.24 mpg more than automatic. Hence from the linear model it can be inferred manual transmission cars serve higher milage per the given data.

A generalized linear model fit was also performed to verify if the inference we obtained is accurate.

```
glm.fit<-glm(mpg~as.factor(am), data=mtcars)
summary(glm.fit)$coef
```

```
##                 Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)    17.147368   1.124603 15.247492 1.133983e-15
## as.factor(am)1  7.244939   1.764422  4.106127 2.850207e-04
```

As expected, manual transmission cars are shown to have a higher milage of 7.24. This proves parity between Linear and Generalized Linear Model for this dataset to find mpg v/s transmission.

Residual Plot and Diagonostics

A resudual boxplot was done for mpg outcome upon the two regressors of transmissions. It was obtained the the residuals are much higher for Manual Transmission compared to Automatic Transmission. Please see plot shown on Apprendix Figure 2.

Conclusion

From the linear model discussed above it can be concluded that manual transmission cars provide higher milage than automatic transmission cars. Approximate difference in mpg for manual transmission cars was found to be around 7.2 miles per gallon.

Appendix

```
## Loading required package: ggplot2
```

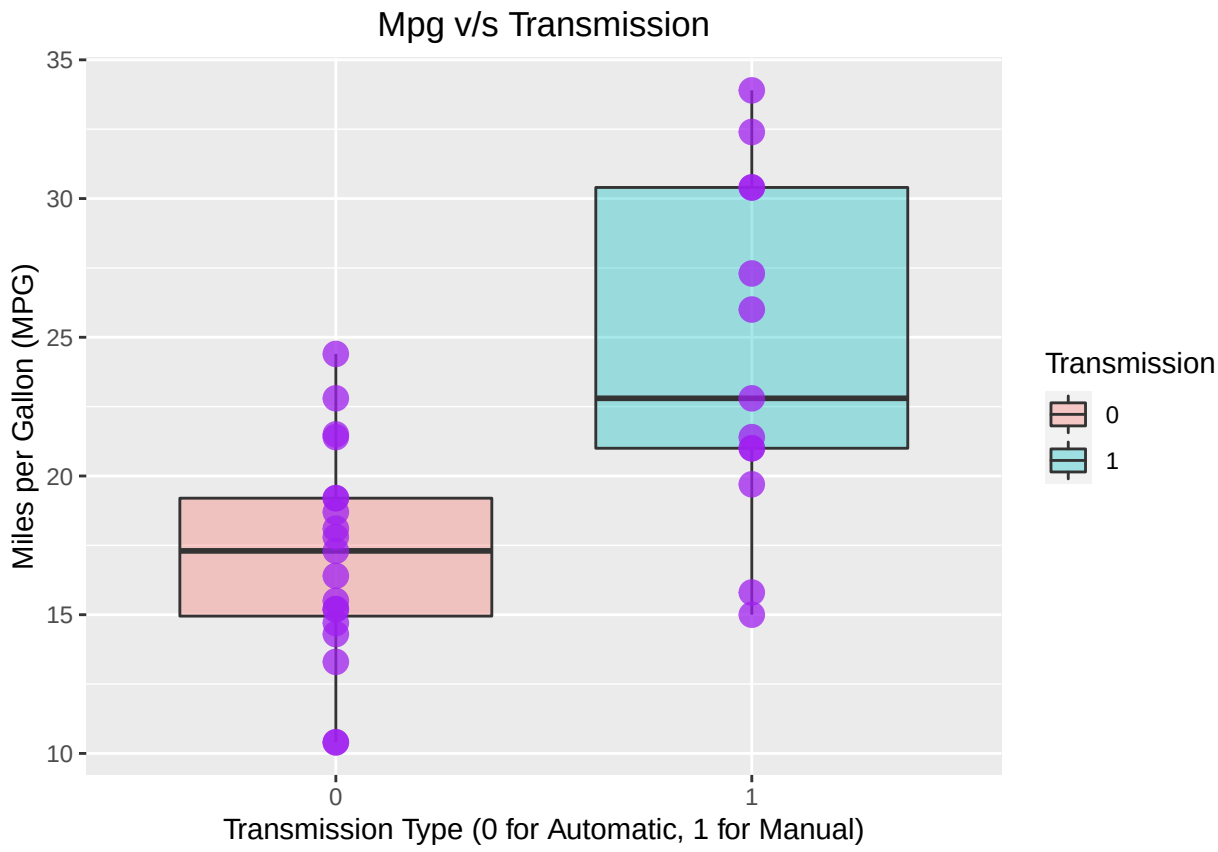

Fig 1: Plot Showing Higher Average MPG values for Manual Transmission

```r
am.fit<-lm(mpg~am, data=mtcars)
plot(resid(am.fit)~mtcars$am,pch=19,col="lightblue",
     xlab = "Transmission Type (0 for Automatic, 1 for Manual)",
     ylab = "Residuals")
abline(h=0, col="red", lwd=3)
```

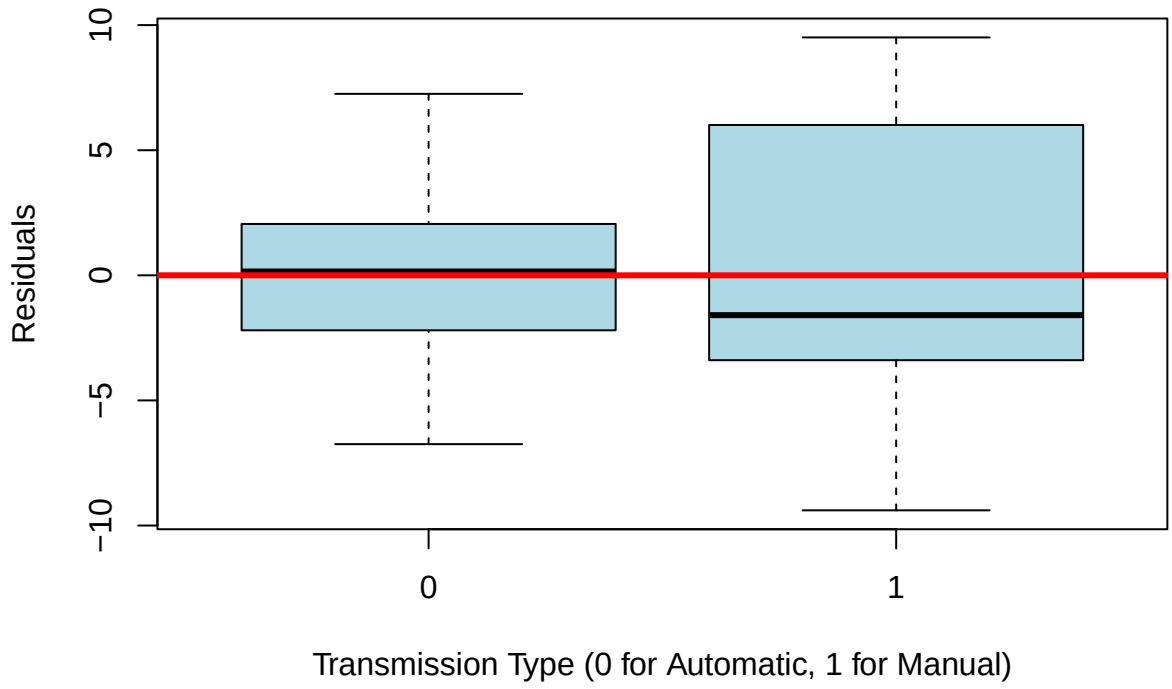

Fig 2: BoxPlot of Residuals