Regression Analysis of Mtcars

Araf

8/25/2020

Executive Summary

Motor Trend, a magazine about the automobile industry are interested in exploring the relationship between a set of variables and miles per gallon (MPG) (outcome). They are particularly interested in the following two questions:

"Is an automatic or manual transmission better for MPG" "Quantify the MPG difference between automatic and manual transmissions"

Automatic and manual transmissions have significant difference of MPG. By fitting a logistic regression model and getting the odds interpretation, we can state that there is ~36% probability for transmission to be manual if we see 1 unit of increase in miles per gallon (mpg).

Exploratory Data Analysis

There are 32 rows with 11 columns. All the variables are numerical. However, cyl, vs, am, gear and crab need closer look.

Let's see the pairs plot.

```
pairs(mtcars)
```

Relationship between Transmission & MPG

```
library(ggplot2)
ggplot(mtcars, aes(x = am, y = mpg, group = 1, col = am)) +
        geom_boxplot() + geom_jitter() +
        geom_point() +
        ggtitle("MPG by Transmission Type") +
        xlab("Transmission") +
        ylab("MPG")
```

Clearly, there is difference of mpg by transmission types. Now, we will build a linear model to quantify the impact of transmission type on mpg.

```
fit <- lm(mpg ~ am, data = mtcars)
fit$coefficients
```

```
## (Intercept)          am
##   17.147368    7.244939
```

F-Statistic for the model is significant which signals a true impact of transmission on mpg. Coefficient of 7.245 means for manual transmission, mpg increases by 7.245 compared to automatic transmission.

However, let's look at the plot:

```
par(mar=c(1,1,1,1))
plot(mtcars$mpg, mtcars$am)
```

It is quite clear a linear regression line does not suite the question in hand as we have two separate group of points. In this problem, logistic regression fits the context better.

```
fit2 <- glm(am ~ mpg, data = mtcars, family = 'binomial')
exp(fit2$coefficients)
```

```
## (Intercept)         mpg
## 0.001355579 1.359379288
```

Coefficient Interpretation: There is ~36% probability for transmission to be manual if we see 1 unit of increase in miles per gallon (mpg).

Model Selection

Two distinct set of data points for mpg by am signals that Logistic Regression is a better fitted model for this problem.

Residual Plot & Diagnostics

```
par(mfrow = c(2,2))
plot(fit2)
```

Inferences

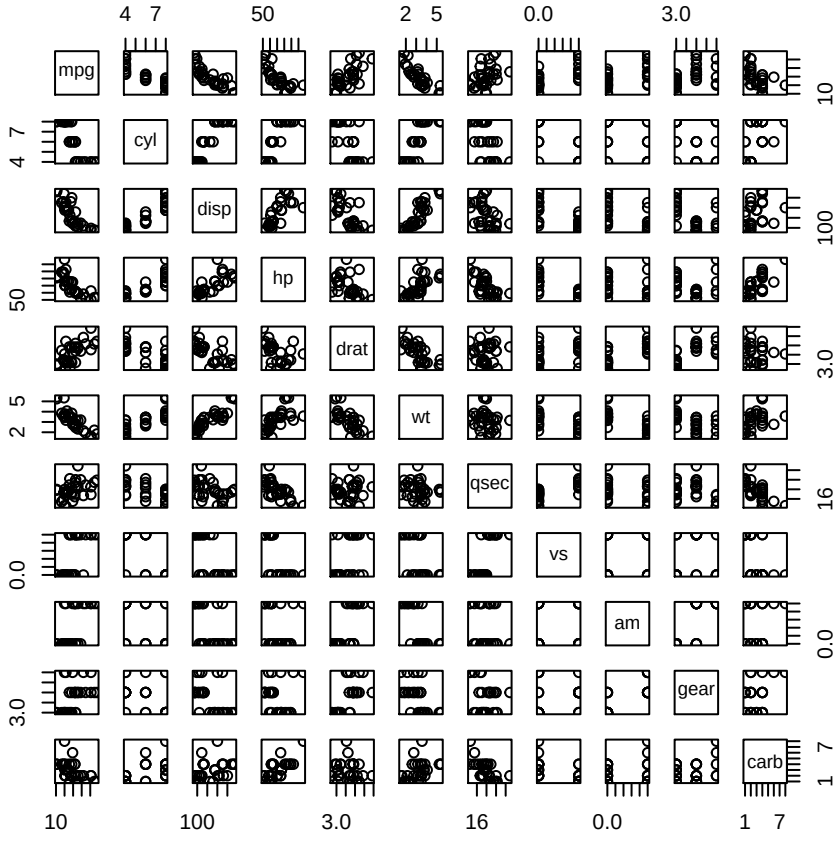
Our inferences from this sample dataset is bound to have some errors. To be precise, mpg variable of our logistic regression has a Std. Error of 0.1148. With a 95% confidence, we can estimate the coefficient to be between 0.307-0.1148 & 0.307+0.1148.

Conclusions

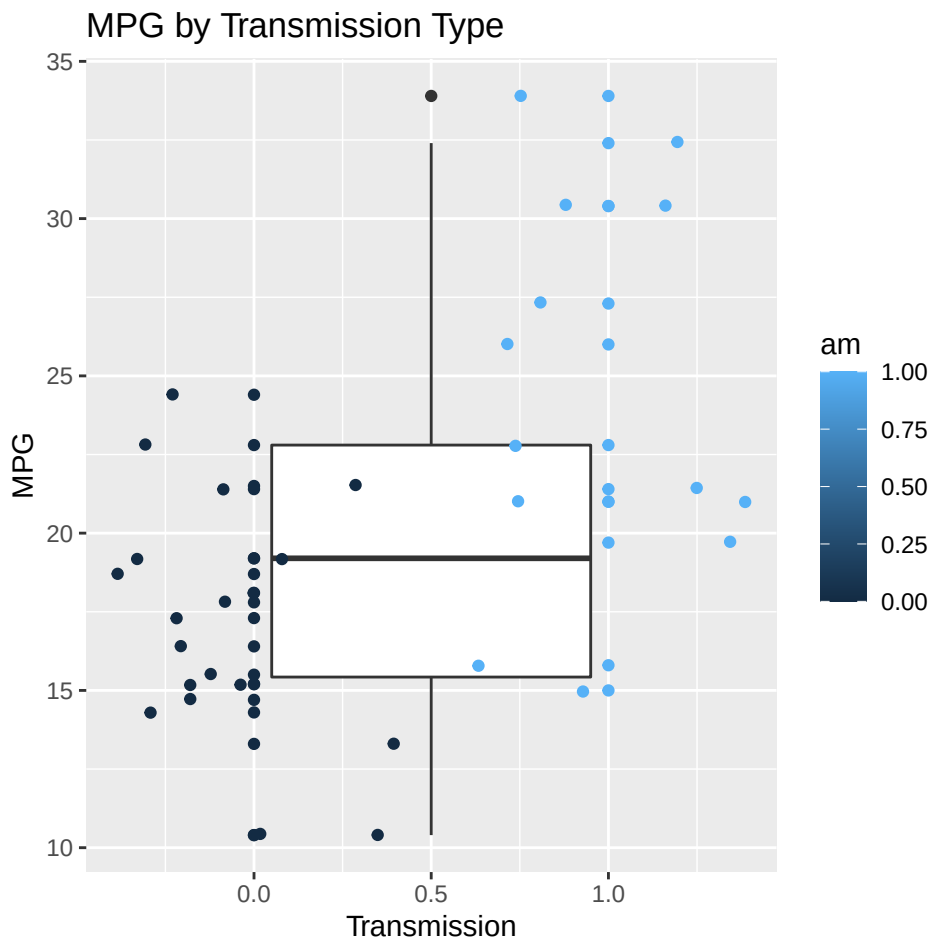
Automatic and manual transmissions have significant difference of MPG. By fitting a logistic regression model and getting the odds interpretation, we can state that there is ~36% probability for transmission to be manual if we see 1 unit of increase in miles per gallon (mpg).

Appendix

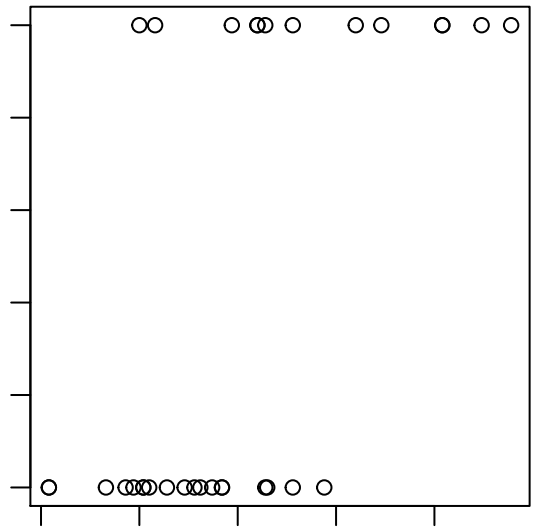
Pair Plot



Box Plot


MPG by Transmission Type
MPG
35
30
25
20
15
10
0.0
0.5
1.0
Transmission
am
1.00
0.75
0.50
0.25
0.00

Problem with Linear Regression



Residual Plots