Transmission Types And Miles Per Gallon Of Cars

Author: DRC

Executive Summary

In this exploration, a set of car data (mtcars from the R datasets) was studied in order to determine a relationship between the miles per gallon (MPG) and the transmission types of the vehicles. I specifically focused on the following two questions:

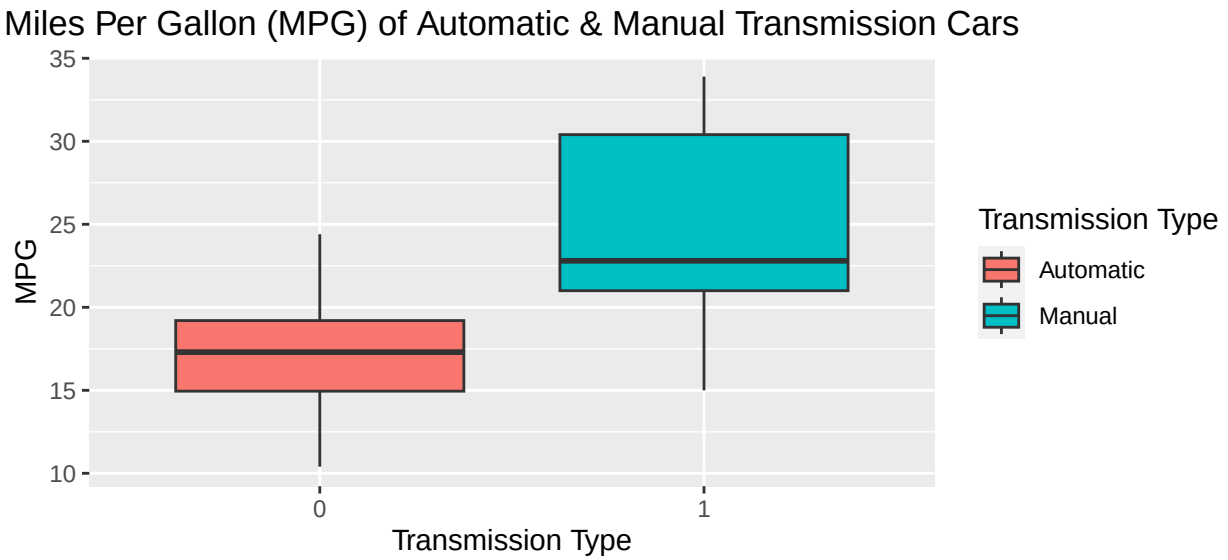
1. Is an automatic or manual transmission better for MPG?
2. Can we predict the transmission type when given the MPG?

While it can be concluded that in general manual transmission cars have higher MPG when compared to ones with automatic transmissions, it is difficult to predict the transmission type when given the MPG. This is due to the binary nature of the outcome data; both linear and logistic models had larger residual values for cars with MPG between 15 and 25. The limitations of both models are exposed for cars with MPGs inside of this range. Because of this, it is difficult to trust the predicted labels produced by both models, particularly for cars with 15 to 25 MPG.

Below the data is explored and modeled.

Exploring The Data

First, box plots were made to see how cars with manual transmissions compared to cars with automatic transmissions with respect to miles per gallon (MPG).



As can be seen from the box plots, manual transmissions appear to have higher MPGs than automatic transmissions do. To confirm that the two groups (automatic and manual transmissions cars) have different means, a t-test was performed.

```r
t.test(mpg ~ am, paired=FALSE, var.equal=FALSE, data=mtcars_df)
```

```
##
##  Welch Two Sample t-test
##
## data:  mpg by am
## t = -3.7671, df = 18.332, p-value = 0.001374
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -11.280194  -3.209684
## sample estimates:
## mean in group 0 mean in group 1
##        17.14737        24.39231
```

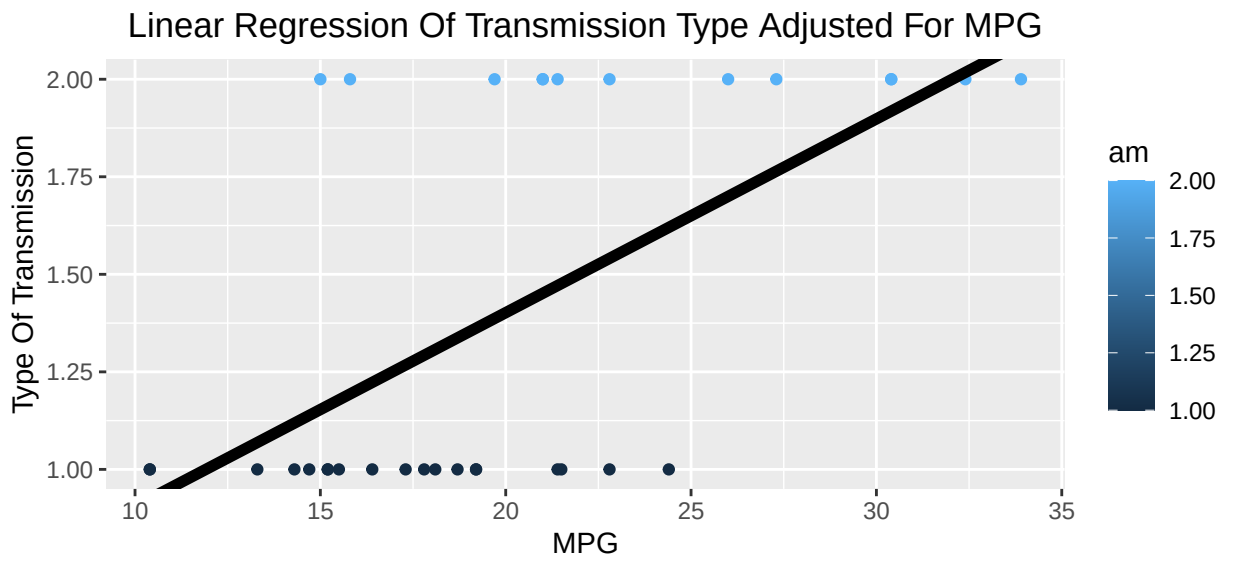
Because the p-value is so small (less than 0.05), we reject the null hypothesis in favor of the alternative. The expected MPGs for automatic and manual transmission cars are not the same. Because our confidence interval for the difference in expected values is negative, we are 95% confident that automatic cars have approximately 3 to 11 MPGs lower than manual cars.

A Linear Relationship?

With transmission type as the outcome, a linear relationship with MPG was created. As can be seen below, the linear model (the black line) does not predict our binary set of outcomes very well.

```r
mtcars_df$am <- as.numeric(mtcars_df$am)
linRegCars <- lm(am ~ mpg, data=mtcars_df)
ggplot(mtcars_df, aes(x=mpg, y=am, color=am)) +
    geom_point() +
    xlab("MPG") +
    ylab("Type Of Transmission") +
    geom_abline(intercept = summary(linRegCars)$coef[1], slope = summary(linRegCars)$coef[2], size=2) +
    ggtitle("Linear Regression Of Transmission Type Adjusted For MPG") +
    ggeasy::easy_center_title()
```

```
## Warning: Using ‘size‘ aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use ‘linewidth‘ instead.
```

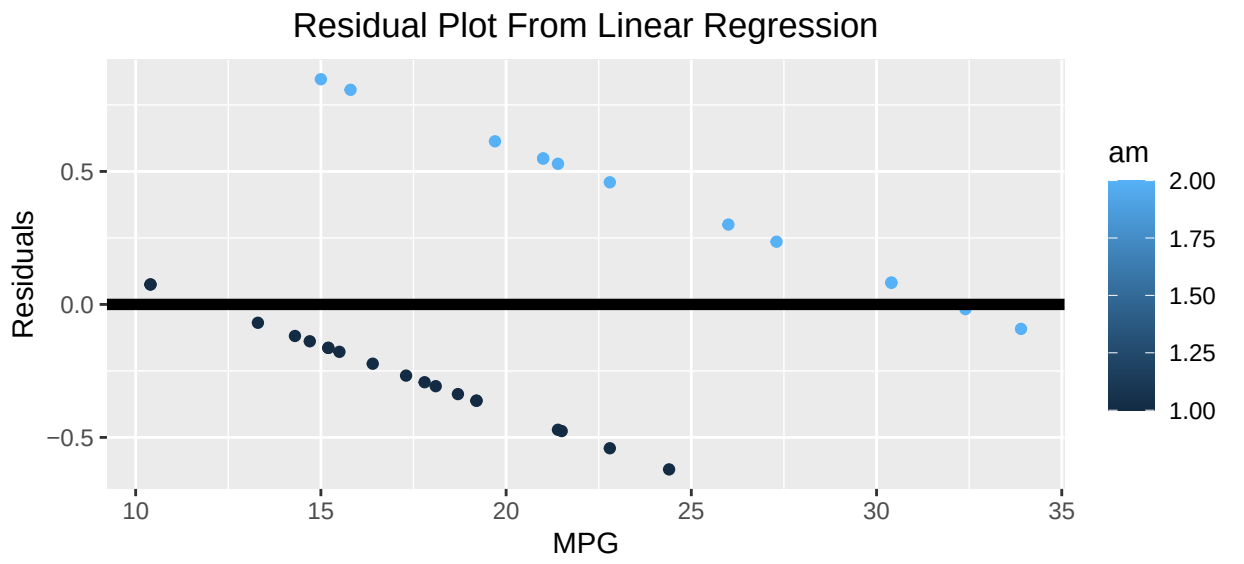



```
summary(linRegCars)$r.squared
```

```
## [1] 0.3597989
```

The r-squared value explains what proportion of the variance in transmission type can be explained by the variance in MPG. This value suggests that the data did not fit the linear regression model very well. To further illustrate this point, below is a residual plot showing the difference between the model's expected transmission type and the actual transmission type.

```
e <- resid(linRegCars)
ggplot(mtcars_df, aes(x=mpg, y=e, color=am)) +
    geom_point() +
    geom_hline(yintercept=0, size=2) +
    xlab("MPG") +
    ylab("Residuals") +
    ggtitle("Residual Plot From Linear Regression") +
    ggeasy::easy_center_title()
```



A Logistic Relationship?

Logistic models are used to predict binary output data, such as transmission type. Below, the data is fit to a logistic model.

```
mtcars_df <- data.frame(mtcars)
mtcars_df$am <- as.factor(mtcars_df$am)

logRegCars <- glm(am ~ mpg, family="binomial", data=mtcars_df)
exp(summary(logRegCars)$coef)
```

```
##                Estimate Std. Error     z value Pr(>|z|)
## (Intercept) 0.001355579  10.500697  0.06030812 1.004993
## mpg         1.359379288   1.121696 14.49049961 1.007535
```

```
exp(confint(logRegCars))
```

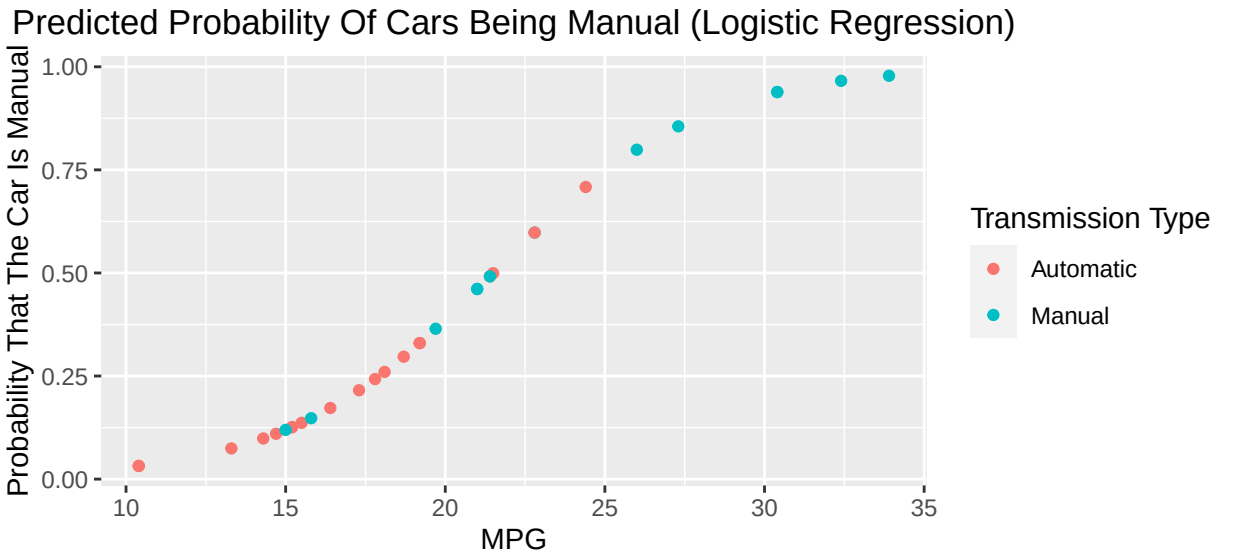
```
##                    2.5 %     97.5 %
## (Intercept) 4.425443e-06 0.06255158
## mpg         1.129764e+00 1.79946863
```

The mpg coefficient suggests that there is a nearly 36% increase in the odds of the probability the car has a manual transmission per unit increase in MPG (holding all other variables fixed). We can be 95% confident that the odds of the probability that the car has manual transmission increase between approximately 12% and 80% per unit increase in MPG (holding all other variables fixed).

Below is a chart showing the predicted probabilities that the car has a manual transmission for each MPG value represented in the data.

```
ggplot(mtcars_df, aes(x=mpg, y=logRegCars$fitted, color=am)) +
    geom_point() +
    xlab("MPG") +
    ylab("Probability That The Car Is Manual") +
```

```
    ggtitle("Predicted Probability Of Cars Being Manual (Logistic Regression)") +
    scale_color_discrete(name = "Transmission Type", labels = c("Automatic", "Manual")) +
    ggeasy::easy_center_title()
```



Below is a chart showing the residuals from the predicted odds of probability that the car has a manual transmission for each MPG value represented in the data. Similar to the linear model, the logistic model has some difficulty predicting for cars with MPGs in the middle of the range of MPG values represented in the data.

```
e <- resid(logRegCars)
ggplot(mtcars_df, aes(x=mpg, y=e, color=am)) +
    geom_point() +
    geom_hline(yintercept=0, size=2) +
    xlab("MPG") +
    ylab("Residuals") +
    ggtitle("Residual Plot From Logistic Regression") +
    scale_color_discrete(name = "Transmission Type", labels = c("Automatic", "Manual")) +
    ggeasy::easy_center_title()
```


Residual Plot From Logistic Regression
Residuals
MPG
2
1
0
−1
10
15
20
25
30
35
Transmission Type
Automatic
Manual