Automatic vs Manual Transmission: A Regression Analysis on mtcars Dataset

vionnet

Executive Summary

The purpose of this analysis is to examine whether automatic transmission is better than manual transmission for miles per gallon (MPG) using the mtcars dataset in R. The mtcars dataset is the motor trend car road tests data for 32 automobiles from the 1974 Motor Trend US magazine. Our analysis on this data shows that on average, the manual transmission performs better than the automatic transmission with 7.24 more MPGs. After adjusting for confonding variables and interactions between variables, our regression model, which includes transmission, horsepower, weight, and the interaction between transmission and weight as predictors, shows that the manual transmission performs better than the automatic transmission with 11.55 more MPGs.

There are several limitations to our analysis: the dataset used in analysis only contains 32 observations. The sample may not be representative. Moreoever, data was extracted from the 1974 magazine. Car performance has improved a lot since then. We may draw a different conclusion if using more recent car performance data.

Exploratory Data Analysis

After getting the data, and creating factor variables within the data, we perform a few exploratory analysis of the data. The "MPG by Transmission Type" boxplots in Appendix show that the median MGP of the automatic transmission is lower than the median, as well as the lower quantile of the MGP of the manual transmission. They also show that the automatic transmission have more consistent MPGs than the manual transmission. The boxplots suggest differences between the two transmission types.

The "Overall MPG Distribution" plot in Appendix shows that MPG is roughly normally distributed.

We divide the mtcars dataset into two groups based on transmission types. The automatic transmission dataset contains 19 observations, and the manual transmission dataset contains 13 observations.

The "Distribution of MPG by Transmission" plot in Appendix shows that MPG is roughly normally distributed for the automatic transmission and somewhat normally distributed for the manual transmission.

From these plots, we assume independent and random sample, normal distribution, as well as unequal variance in the two sub-datasets, and perform the Welch two sample t-test.

```
#Perform the Welch Two Sample t-test
t.test(mpg ~ am, paired = FALSE, var.equal = FALSE, data = mtcars)
```

```
##
##  Welch Two Sample t-test
##
## data:  mpg by am
## t = -3.767, df = 18.33, p-value = 0.001374
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.28  -3.21
## sample estimates:
## mean in group Automatic    mean in group Manual
##                   17.15                   24.39
```

P-value from the above test is small (< 0.05), hence we reject the null hypothesis, and claim that there is a significant difference in means of MPGs for the two transmissions. The difference in means is 7.24 MPGs.

Model Selection

If we only include am in our model, we see from the R^2 value that only 36% of the total variation can be explain by the model. This suggests the existence of confonding variables.

```
fit1 <- lm(mpg~am, data=mtcars)
summary(fit1)$r.squared
```

```
## [1] 0.3598
```

After trials and errors (details omitted due to report length), we find that hp and wt are the confonding variables.

We would also like to examine potential interactions between am and wt, and interactions between am and hp. When incorporating the interaction betwwen am and wt into the model, we get that their interaction is somewhat significant (p-value is $0.0197 < 0.05$). When incorporating the interaction between am and hp into the model, the ANOVA test shows that it is not significant.

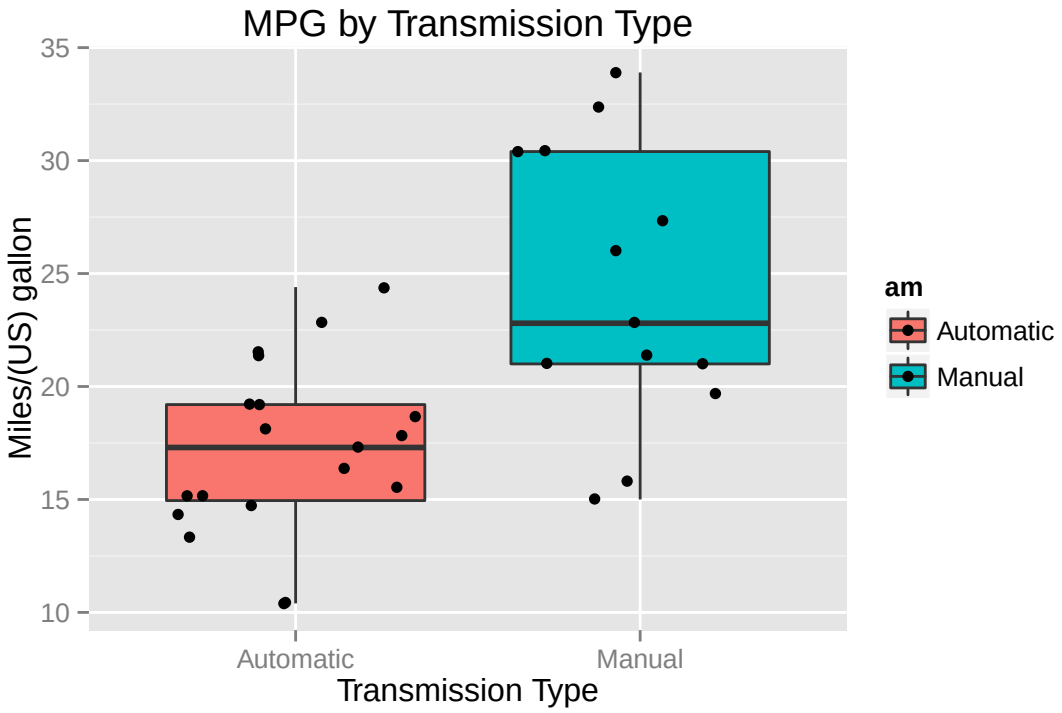
```
## Analysis of Variance Table
##
## Model 1: mpg ~ am + hp + wt + am * wt
## Model 2: mpg ~ am + hp + wt + am:wt + am:hp
##   Res.Df RSS Df Sum of Sq    F Pr(>F)
## 1     27 147
## 2     26 136  1      10.9 2.09   0.16
```

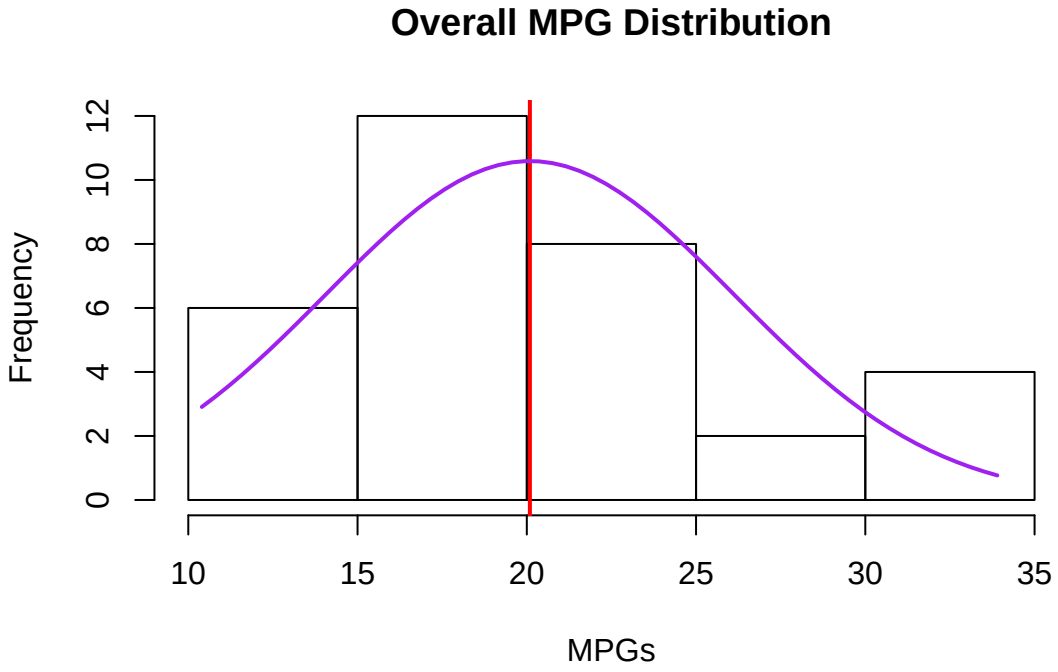
Before settling down on the linear regression model with am, hp, wt, and am*wt as predictors, let us examine its residual plots displayed in the Appendix. From these plots, we see that residuals from the model demostrate homoscedasticity and normality, which are desired results.

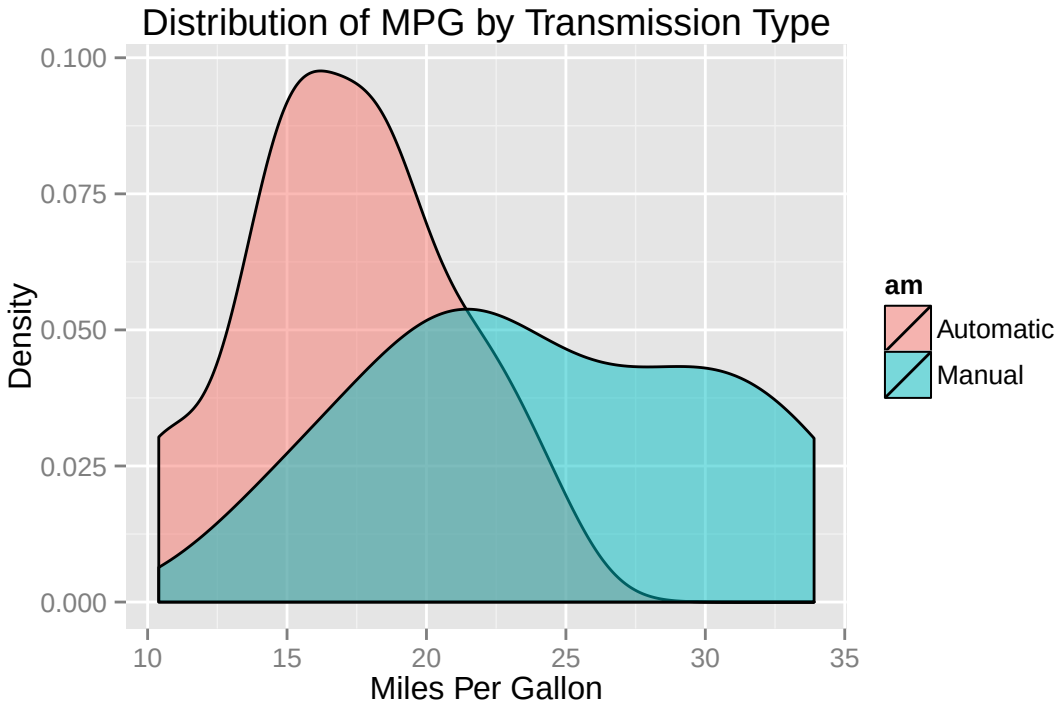
Summary below shows that our model explains 87% of the variance. On average, the manual transmission has 11.55 more MPGs than the automatic transmission.

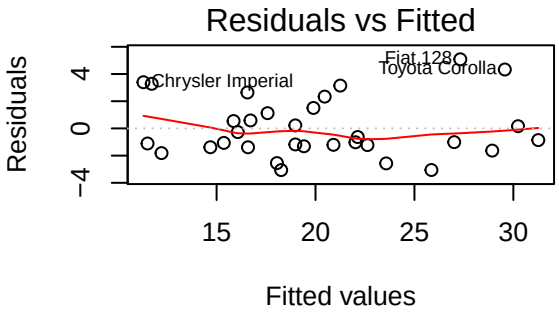
```
##
## Call:
## lm(formula = mpg ~ am + hp + wt + am * wt, data = mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.064 -1.332 -0.935  1.218  5.082
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  30.9473     2.7234   11.36  8.5e-12 ***
## amManual     11.5548     4.0233    2.87   0.0078 **
## hp           -0.0270     0.0098   -2.75   0.0105 *
## wt           -2.5156     0.8445   -2.98   0.0061 **
## amManual:wt  -3.5779     1.4428   -2.48   0.0197 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.33 on 27 degrees of freedom
## Multiple R-squared:  0.87,   Adjusted R-squared:  0.85
## F-statistic:   45 on 4 and 27 DF,  p-value: 1.45e-11
```

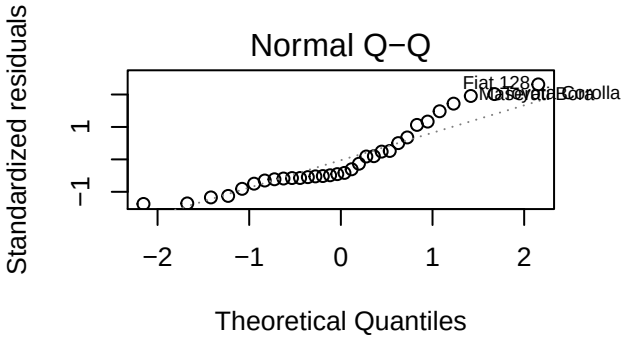
Appendix

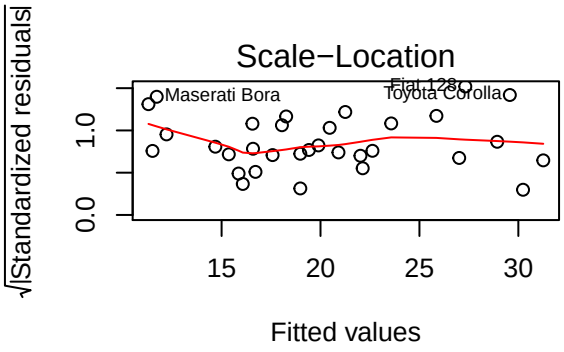

MPG by Transmission Type
Miles/(US) gallon
35
30
25
20
15
10
Automatic
Manual
Transmission Type
am
Automatic
Manual

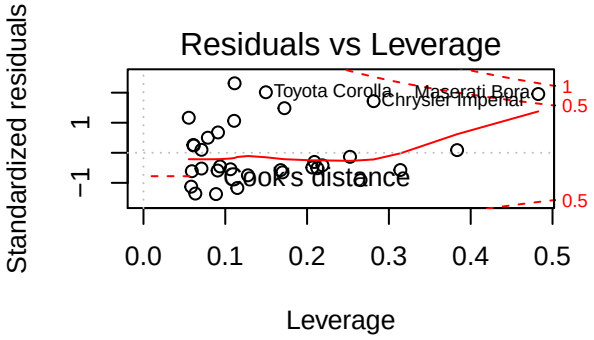

Overall MPG Distribution
Frequency
0
2
4
6
8
10
12
10
15
20
25
30
35
MPGs


Distribution of MPG by Transmission Type
Density
0.100
0.075
0.050
0.025
0.000
10
15
20
25
30
35
Miles Per Gallon
am
Automatic
Manual


Residuals vs Fitted
Residuals
Chrysler Imperial
Fiat 128
Toyota Corolla
4
0
−4
15
20
25
30
Fitted values


Normal Q−Q
Standardized residuals
Fiat 128
1
−1
−2
−1
0
1
2
Theoretical Quantiles


Scale−Location
√|Standardized residuals|
Maserati Bora
Fiat 128
Toyota Corolla
1.0
0.0
15
20
25
30
Fitted values


Residuals vs Leverage
Standardized residuals
Toyota Corolla
Maserati Bora
Chrysler Imperial
Cook's distance
1
0.5
0.5
1
−1
0.0
0.1
0.2
0.3
0.4
0.5
Leverage