Analysis of Car Performance

EXECUTIVE SUMMARY

The aim of this project was to answer 2 questions: (1) Is an automatic or manual transmission better for miles per gallon (MPG)? (2) Quantify the MPG difference between automatic and manual transmissions.

To answer these questions, a regression analysis was conducted on a dataset which contains data on the fuel consumption and other aspects of automobile design and performance for 32 automobile models from 1973 to 1974. The analysis assumed that MPG (Y) is a function of various automobile features (X).

The regression analysis showed that manual transmission automobiles had a higher MPG, holding weight, displacement, and acceleration (as measured by qsec) constant. Manual transmission automobiles had 23.24 MPG on average, while automatic transmission automobiles had only 9.20 MPG on average.

INTRODUCTION

This analysis aimed to explore the relationship between a set of automobile features and MPG. Specifically, the goal was to answer the following questions: (1) Is an automatic or manual transmission better for MPG? (2) Quantify the MPG difference between automatic and manual transmissions.

The mtcars dataset was used in this analysis. This dataset contained 32 observations of 11 automobile features, taken from automobile models from 1973 to 1974. In this case, the dependent variable is MPG, or Miles per Gallon. This is a measure of fuel efficiency. The explanatory variable of interest is AM, a binary variable which signals whether an automobile has automatic (am = 0) or manual (am = 1) transmission. A visual examination of the data suggests that manual transmission automobiles have a higher MPG (See Appendix, Figure 1).

The remaining explanatory variables are: cyl (number of cylinders), disp (displacement), hp (gross horsepower), drat (rear axle ratio), wt (weight), qsec (quarter mile per second), vs (engine cylinder configuration); gear (number of forward gears), and carb (number of carburetors). Exploratory data analysis suggests that:

1) As number of cylinders increased, MPG decreased (See Appendix, Fig. 2).
2) MPG has a negative relationship with disp (Appendix Fig. 3), hp (Appendix, Fig. 4) and wt (Appendix, Fig. 5).
3) MPG has a positive relationship with drat (Appendix, Fig. 6) and qsec (Appendix, Fig. 7).
4) Engines with a straight line configuration (vs = 1) have higher MPGs (Appendix, Fig. 8).
5) Automobiles with 4 gears have the highest MPGs (Appendix, Fig. 9).
6) MPG decreased as the number of carburetors increased (Appendix, Fig. 10).

However, the possibility that there is an interaction effect between am and another explanatory variable cannot be disregarded. In fact, interaction plots hinted at the presence of interaction effects between am and cyl (Appendix, Fig. 11), hp (Appendix, Fig. 12), qsec (Appendix, Fig. 13), vs (Appendix, Fig. 14), and carb (Appendix, Fig. 15).

As part of the exploratory data analysis, the following variables were converted from numeric to factor variables: am, cyl, vs, gear and carb.

MODELS AND ANALYSIS

As an initial analysis, linear models were estimated using each of the explanatory variables. While all of the models had significant estimates, the r-squared values ranged from 0.18 (for qsec) to 0.75 (for wt). (See Appendix)

Next, the possibility of an interaction between am and another variable should be addressed. The analysis of variance method was used to determine possible interaction effects. In this case, there were 2 significant interactions: (1) am with wt, and (2) am with disp. (See Appendix)

After this, the regression model was selected using the nested models approach. Specifically, nested likelihood ratio tests were used to determine whether the added variables were necessary for the model. Thus, mpg was first regressed on am, and thereafter adding an additional term. The first additional term to be considered was the interaction between am and wt, followed by the interaction between am and disp. Thereafter, each of the remaining variables were additively included in the model.

The ANOVA method showed that the only necessary terms for the model were am, wt, disp, qsec, and the interaction term for am and wt. The resulting model is:

```
model <- lm(mpg ~ am * wt  + disp + qsec, data = mtcars)
tidy(model)
```

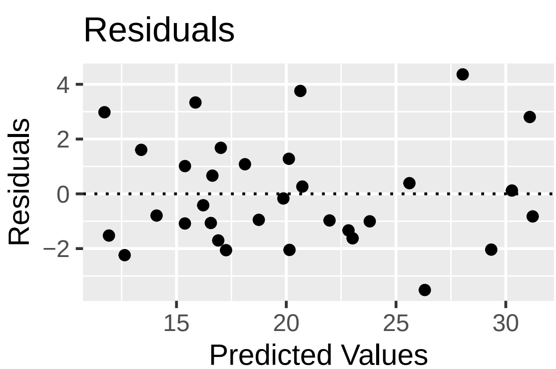
```
## # A tibble: 6 x 5
##   term        estimate std.error statistic  p.value
##   <chr>          <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)  9.20      7.12        1.29  0.208
## 2 am1         14.0       3.51        4.00  0.000465
## 3 wt          -3.06      1.10       -2.77  0.0101
## 4 disp         0.00127   0.00923     0.138 0.891
## 5 qsec         1.05      0.353       2.97  0.00627
## 6 am1:wt      -4.10      1.25       -3.29  0.00287
```

The model generated an r-squared of 0.876, which implies that the model is able to account for almost 88% of the variation in the data. The coefficients for am1, wt, qsec, and the interaction term were all statistically significant.

Based on this model, manual transmission automobiles have a higher mean MPG than automatic transmission automobiles, holding weight, displacement, and acceleration as measured by qsec constant. The average MPG of automatic transmission automobiles was 9.20 (with a 95% confidence interval of -5.44 to 23.83), while manual transmission automobiles was 23.24 (with a 95% confidence interval of 8.00 to 38.46).

Every additional 1000lbs of weight, however, lowered the average MPG by 3.06 (with a 95% confidence interval of -5.32 to -0.79), holding the other variables constant. For manual transmission automobiles, there is an additional decrease of 4.10 for every 1000lbs of additional weight. MPG increased by 1.05 (with a 95% confidence interval of 0.32 to 1.78) for every unit increase in acceleration as measured by qsec, holding the other variables constant.

The residual plot below shows a random pattern, suggesting a good fit of the linear model.



SUMMARY

In order to determine whether automatic of manual transmission was better for MPG, a regression analysis using the mtcars dataset was conducted. Based on the results of the analysis, manual transmission automobiles had on average a higher MPG than automatic transmission automobiles, holding other variables constant.

This conclusion however is limited to the scope of this analysis, i.e., to the automobile models included in the dataset.

APPENDIX

EXPLORATORY DATA ANALYSIS PLOTS

Figure 1

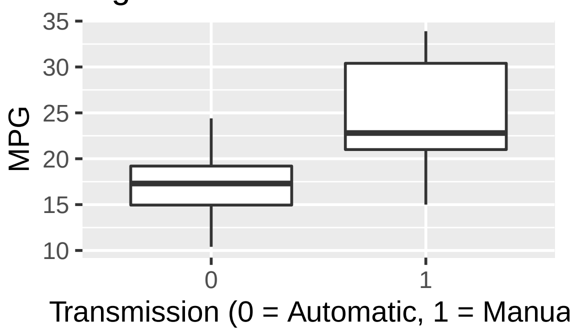


Figure 2

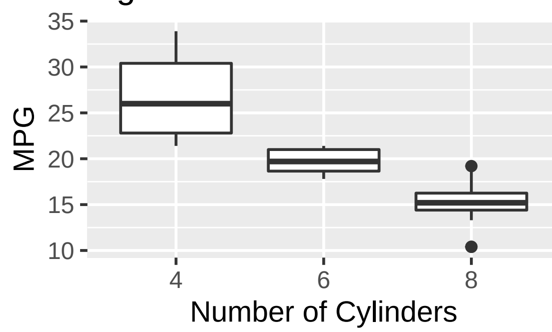


Figure 3

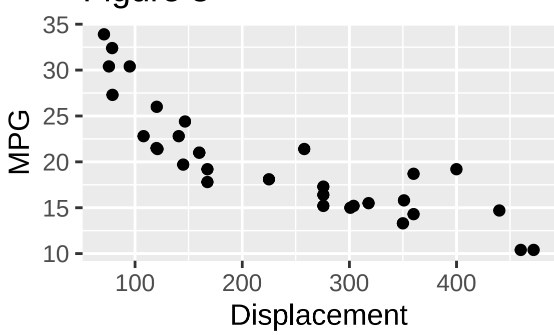


Figure 4

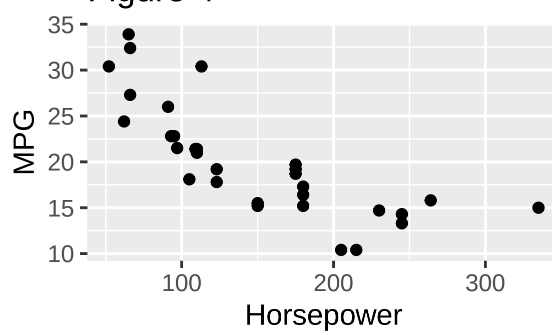


Figure 5

Figure 6

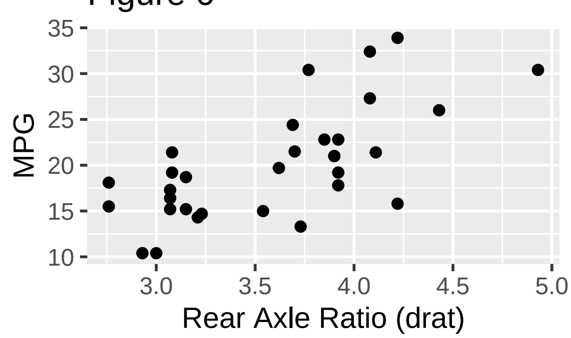


Figure 7

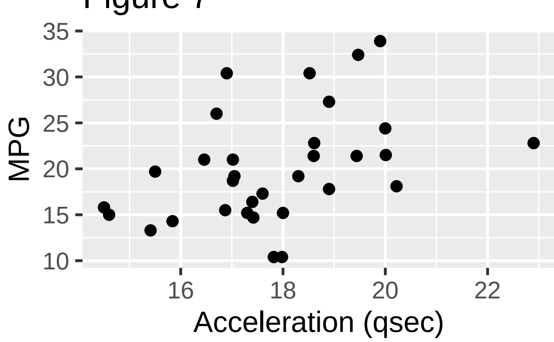


Figure 8

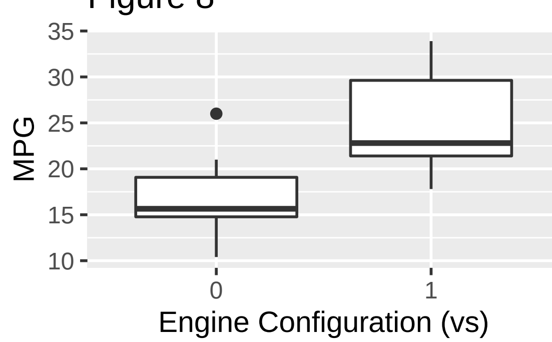
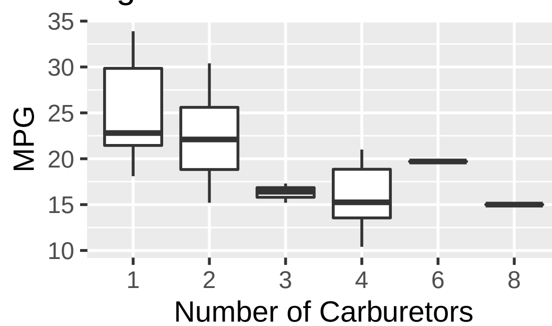


Figure 9

Figure 10



CHECKING FOR INTERACTION VARIABLES

Figure 11

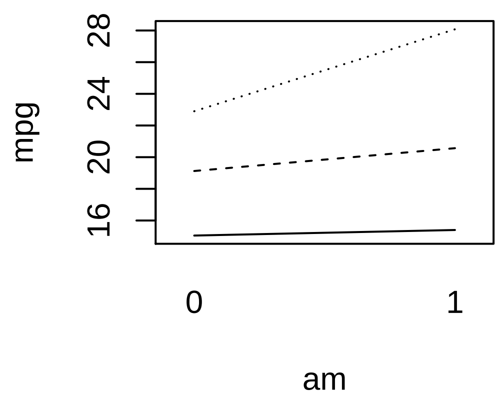


Figure 12

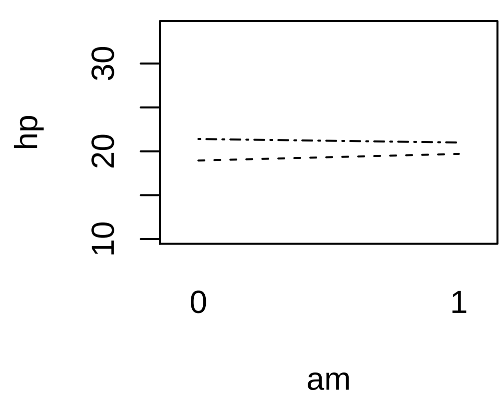


Figure 13

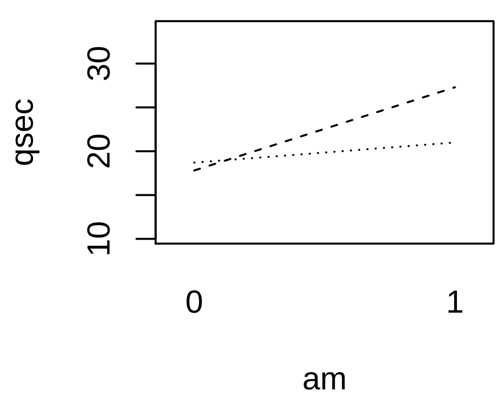
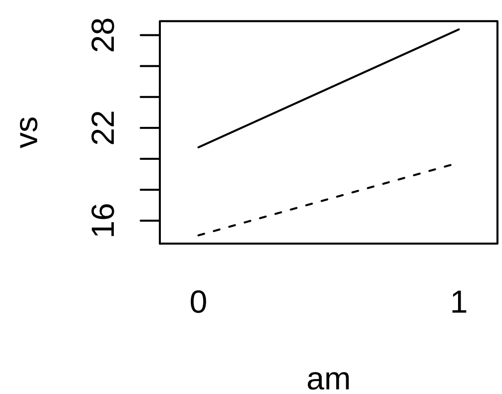
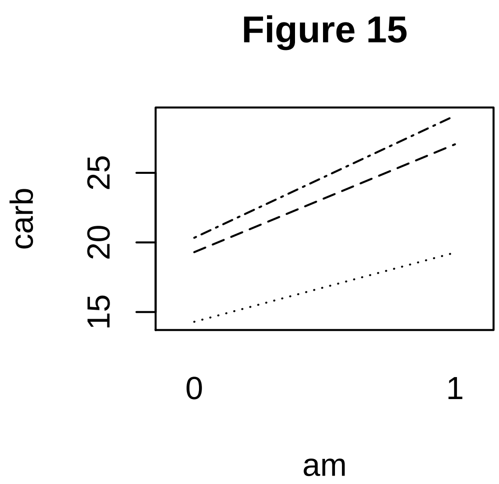


Figure 14





CHECKING R-SQUARED, INDIVIDUAL MODELS

```
##    Variable R-Squared
## 1       cyl 0.7324601
## 2      disp 0.7183433
## 3        hp 0.6024373
## 4      drat 0.4639952
## 5        wt 0.7528328
## 6      qsec 0.1752963
## 7        vs 0.4409477
## 8        am 0.3597989
## 9      gear 0.4291500
## 10     carb 0.4445294
```

LOOKING FOR INTERACTIONS, ANALYSIS OF VARIANCE

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## wt           1  847.7   847.7  126.25 6.92e-12 ***
## am           1    0.0     0.0    0.00  0.98556
## wt:am        1   90.3    90.3   13.45  0.00102 **
## Residuals   28  188.0     6.7
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## disp         1  808.9   808.9  95.729 1.55e-10 ***
## am           1   16.9    16.9   1.997   0.1686
## disp:am      1   63.7    63.7   7.537   0.0104 *
## Residuals   28  236.6     8.4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```