

## Peer-graded Assignment: Regression Models Course Project

Simon Grasdal

2024-12-15

### Executive Summary

This is a report analyzing the mtcars dataset provided by R Studio. The purpose is to explore the relationship(s) that exist between transmission type and the Miles Per Gallon (MPG) extracted from the 1974 Motor Trend US magazine. The sample is comprised of 32 cars (1973-74 models) and explored 10 aspects of each vehicle including but not limited to: number of cylinders (cyl), weight (wt), and horsepower (hp). In this analysis, regression models are used to explore how **automatic** (am = 0) and **manual** (am = 1) transmissions impact **MPG**. T-test shows that the performance difference between cars with automatic and manual transmission. And it is about 7 MPG more for cars with manual transmission than those with automatic transmission. Then, several linear regression models are fitted and one with highest Adjusted R-squared value is selected. So, given that weight and 1/4 mile time are held constant, manual transmitted cars are  $14.079 + (-4.141) \text{ weight}$  more **MPG** on average better than automatic transmitted cars. Thus, cars that are lighter in weight with a manual transmission and cars that are heavier in weight with an automatic transmission will have higher MPG values.

### Analysis

#### Exploratory Data Analysis

```
##           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
## [1] 32 11
```

Our cursory look at the data tells us that there is a difference between the two transmission types. In the boxplot depicted in **Appendix Fig. 1** we can see that the manual transmission vehicles have higher **MPG** values. Likewise the pairwise correlations shown in **Appendix Fig.2** show correlations between other such variables like weight, displacement, and horsepower.

## Inference

To test if the automatic and manual transmissions are sufficiently different we use a two-sample t-test. The null hypothesis for this test is that there is no significant difference between the two transmissions types. The alternate hypothesis states that there is a difference.

```
## [1] 0.001373638
## mean in group 0 mean in group 1
##      17.14737      24.39231
```

Our result indicates that there is a significant difference in **MPG** between the two kinds of transmission types ( $T = 24.39$ ,  $p < .05$ ) thus rejecting the null hypothesis. This means that there is a significant difference in **MPG** between automatic and manual transmission types.

## Regression

First, we must fit the full model without any interference. The results indicate Residual standard error as 2.833 on 15 degrees of freedom. And the Adjusted R-squared value is 0.779, which means that the model can explain about 78% of the variance of the MPG variable. However, none of the coefficients are significant at 0.05 significant level. However, we can see that **weight** (wt) is quite close to significance ( $p = .06$ ). To confirm our suspicions we run a stepwise model of the regression.

This model is “mpg ~ wt + qsec + am”. It has the Residual standard error as 2.459 on 28 degrees of freedom. And the Adjusted R-squared value is 0.8336, which means that the model can explain about 83% of the variance of the MPG variable. All of the coefficients are significant at 0.05 significant level. According to the scatter plot (**Appendix Fig. 3**), it indicates that there appear to be an interaction term between “wt” variable and “am” variable, since automatic cars tend to weigh heavier than manual cars. Thereby, following model including the interaction term is generated:

This model has the Residual standard error as 2.084 on 27 degrees of freedom. And the Adjusted R-squared value is 0.8804, which means that the model can explain about 88% of the variance of the MPG variable. All of the coefficients are significant at 0.05 significant level, which is pretty good. Next, the simple model is fitted with MPG as the outcome variable and Transmission as the predictor variable.

It shows that on average, a car has 17.147 mpg with automatic transmission, and if it is manual transmission, 7.245 mpg is increased. This model has the Residual standard error as 4.902 on 30 degrees of freedom. And the Adjusted R-squared value is 0.3385, which means that the model can explain about 34% of the variance of the MPG variable. The low Adjusted R-squared value also indicates that other variables should be added to the model.

Finally, the final model is selected: mpg ~ wt + qsec + am + wt:am.

##		Estimate	Std. Error	t value	Pr(> t )
##	(Intercept)	9.723053	5.8990407	1.648243	0.1108925394
##	wt	-2.936531	0.6660253	-4.409038	0.0001488947
##	qsec	1.016974	0.2520152	4.035366	0.0004030165
##	am	14.079428	3.4352512	4.098515	0.0003408693
##	wt:am	-4.141376	1.1968119	-3.460340	0.0018085763

Thus, the result shows that when “wt” (weight lb/1000) and “qsec” (1/4 mile time) remain constant, cars with manual transmission add  $14.079 + (-4.141) \cdot \text{wt}$  more MPG (miles per gallon) on average than cars with automatic transmission. That is, a manual transmitted car that weighs 2000 lbs have 5.797 more MPG than an automatic transmitted car that has both the same weight and 1/4 mile time.

### Analysis of Residuals

According to the residuals plots shown in **Appendix Fig. 4**, the following assumptions can be verified concerning our model. 1. The Residuals vs. Fitted plot shows no consistent pattern, supporting the accuracy of the independence assumption. 2. The Normal Q-Q plot indicates that the residuals are normally distributed because the points lie closely to the line. 3. The Scale-Location plot confirms the constant variance assumption, as the points are randomly distributed. 4. The Residuals vs. Leverage argues that no outliers are present, as all values fall well within the 0.5 bands.

## Appendix

Figure 1

Boxplot of MPG vs. Transmission Type

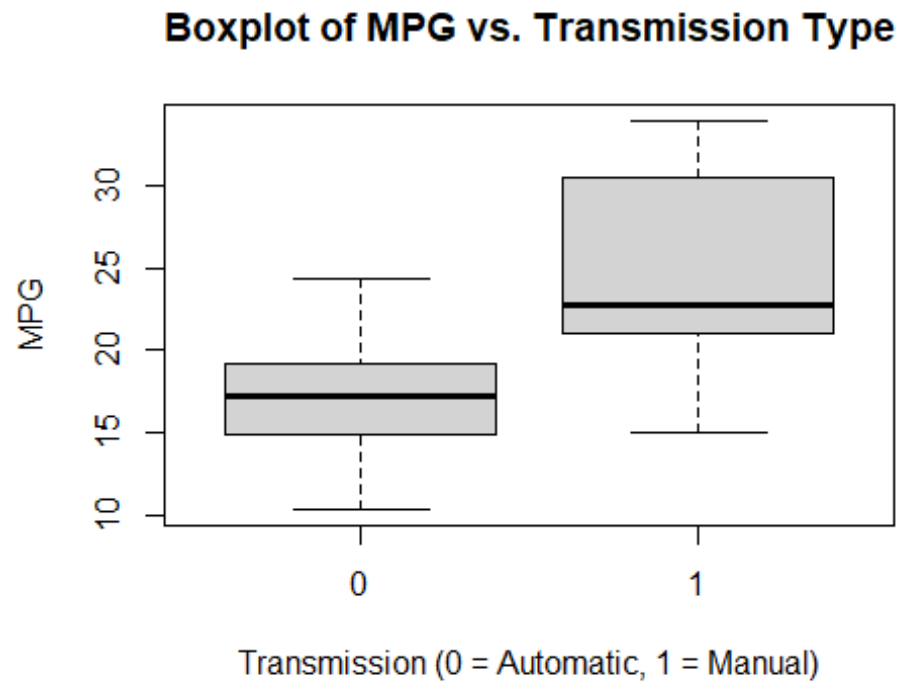


Figure 2

Pairwise Correlation Matrix

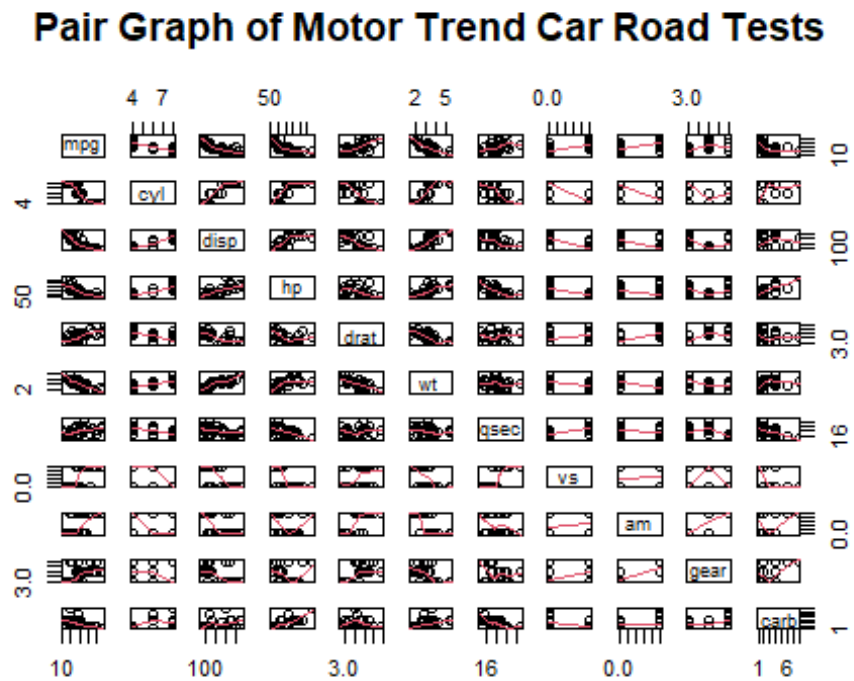


Figure 3

Scatterplot of MPG vs. Weight by Transmission

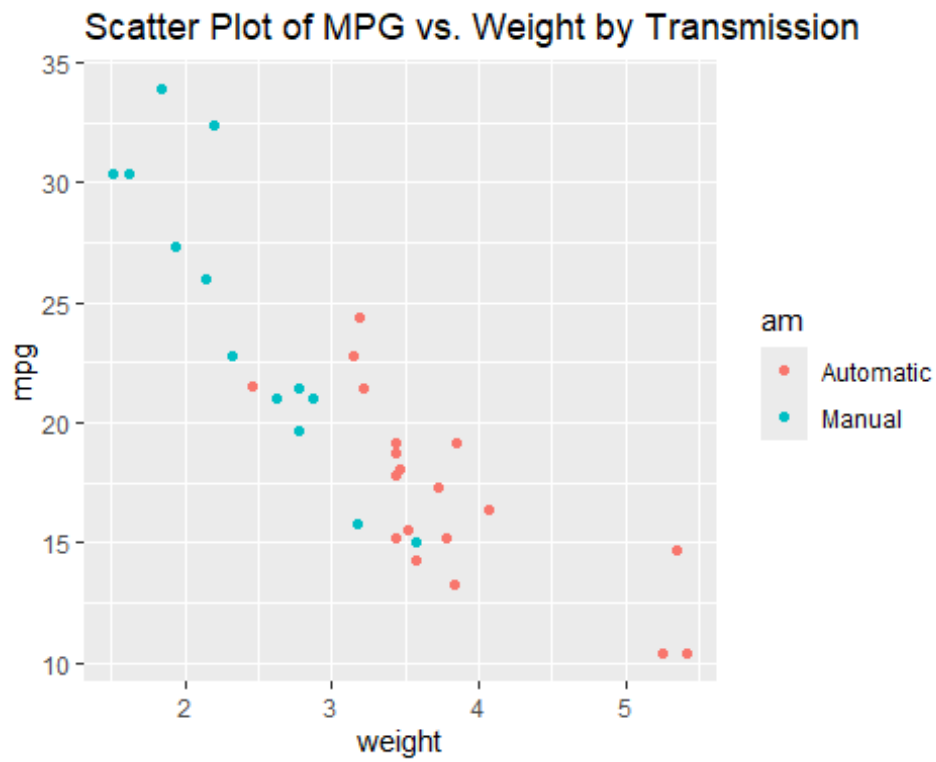


Figure 4  
Residuals

