

# Regression Models - Motor Trend

## Executive Summary

Using the mtcars dataset answer the following questions:

“Is an automatic or manual transmission better for MPG?”

“Quantify the MPG difference between automatic and manual transmissions”

Variables included in mtcars: \* mpg: Miles per (US) gallon \* cyl: Number of cylinders \* disp: displacement (cu.in.) \* hp: Gross horsepower \* drat: Rear axle ratio \* wt: Weight (1000 lbs) \* qsec: 1/4 mile time \* vs: Engine shape (0 = V-shaped, 1= straight) \* am: Transmission type (0 = automatic, 1 = manual) \* gear: Number of forward gears \* carb: Number of carburetors

Data Source: Henderson and Velleman (1981), Building multiple regression models interactively. Biometrics, 37, 391–411.

Process: \* Quick look at the dataset and some data cleaning \* Exploratory data analysis and plots \* Model creation and analysis \* Answering the initial questions

```
FALSE -- Attaching packages ----- tidyverse 1.2.1 --
FALSE v ggplot2 3.2.1      v purrr  0.3.2
FALSE v tibble  2.1.3      v dplyr  0.8.3
FALSE v tidyr   0.8.3      v stringr 1.4.0
FALSE v readr   1.3.1      v forcats 0.4.0

FALSE -- Conflicts ----- tidyverse_conflicts() --
FALSE x dplyr::filter() masks stats::filter()
FALSE x dplyr::lag()    masks stats::lag()
```

## Exploratory data analysis

The mtcars dataset contains data extracted from the 1974 Motor Trend US magazine and comprises of fuel consumption and 10 aspects of automobile design and performance for 32 automobiles.

```
summary(mtcars)
```

```
str(mtcars)
```

## Visualizations

Please see appendix

### Plot 1

We can see that manual transmission engines have generally have higher miles per gallon as it sits to the right of automatic transmission engines.

### Plot 2

Manual has a higher range but less variance than Automatic transmission

### Plot 3

From the pairs plot (see appendix below) we see that cyl, disp, drat and wt influence mpg. These variables will be used in the model.

## Modelling

We will test three models and compare the results based on their  $R^2$  and analysis of their residuals.

```
model_1 <- lm(mpg~.,data=mtrend)
```

```
model_2 <- lm(mpg~cyl+disp+drat+wt,data=mtrend)
```

Analysis of Model 1 & 2

From the F-statistic, this model is statistically significant. In addition, there is no pattern in the residuals and the points lie within the cook's distance line. This means that we do not need to remove any influential outliers.

The type of transmission of the model (automatic vs manual) interacts with a lot of the other variables. We would need to include an interaction term to assess its effects on mpg.

Adding an interaction term between weight and transmission type since weight is the most statistically significant variable. #####Final Model #Adding an interaction term and simplifying the model for greater interpretation

```
model_3 <- lm(mpg~cyl+(am*wt),data=mtrend)
```

With the final model (results in appendix) we have similarly good residuals as per model 1 and 2 but an 86% adjusted  $r^2$  and a lower p-value on the F-statistic.

## Interpreting Coefficients

- Intercept: On average the miles per gallon under automatic transmission when the number of cylinders and weight is 0 is 34 miles per gallon .
- Cyl: Adding 1 unit of cylinders will decrease the miles per gallon by 1.1814
- Weight: Adding 1000(lbs) of weight in a car will decrease the mpg by 2.37 mpg
- amManual: Under manual transmission and for a cylinder count of 0 and weight of 0, mpg is 11.9385 higher compared to automatic transmission
- amManual:wt: The effect of weight on miles per gallon of a car decreases by 4.1974 under manual transmission compared to automatic transmission.

## Answers to questions

- Is an automatic or manual transmission better for mpg? The coefficient of the amManual variable in all three models suggests that a manual transmission has a positive correlation with higher mpg levels compared to automatic transmission.
- Quantify the MPG difference between automatic and manual transmissions Based on model 3, Under manual transmission and for a cylinder count of 0 and weight of 0, mpg is 11.9385 higher compared to automatic transmission given the assumption of normally distributed errors.

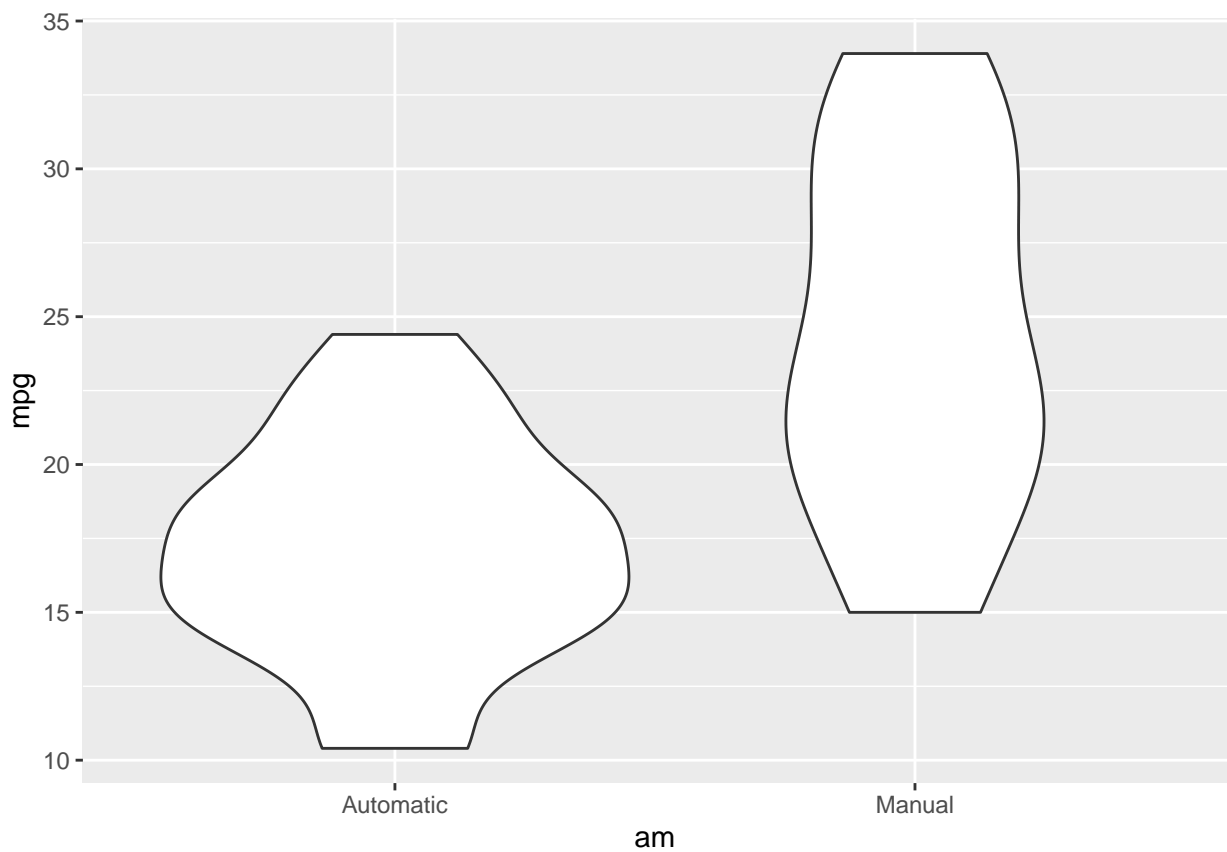
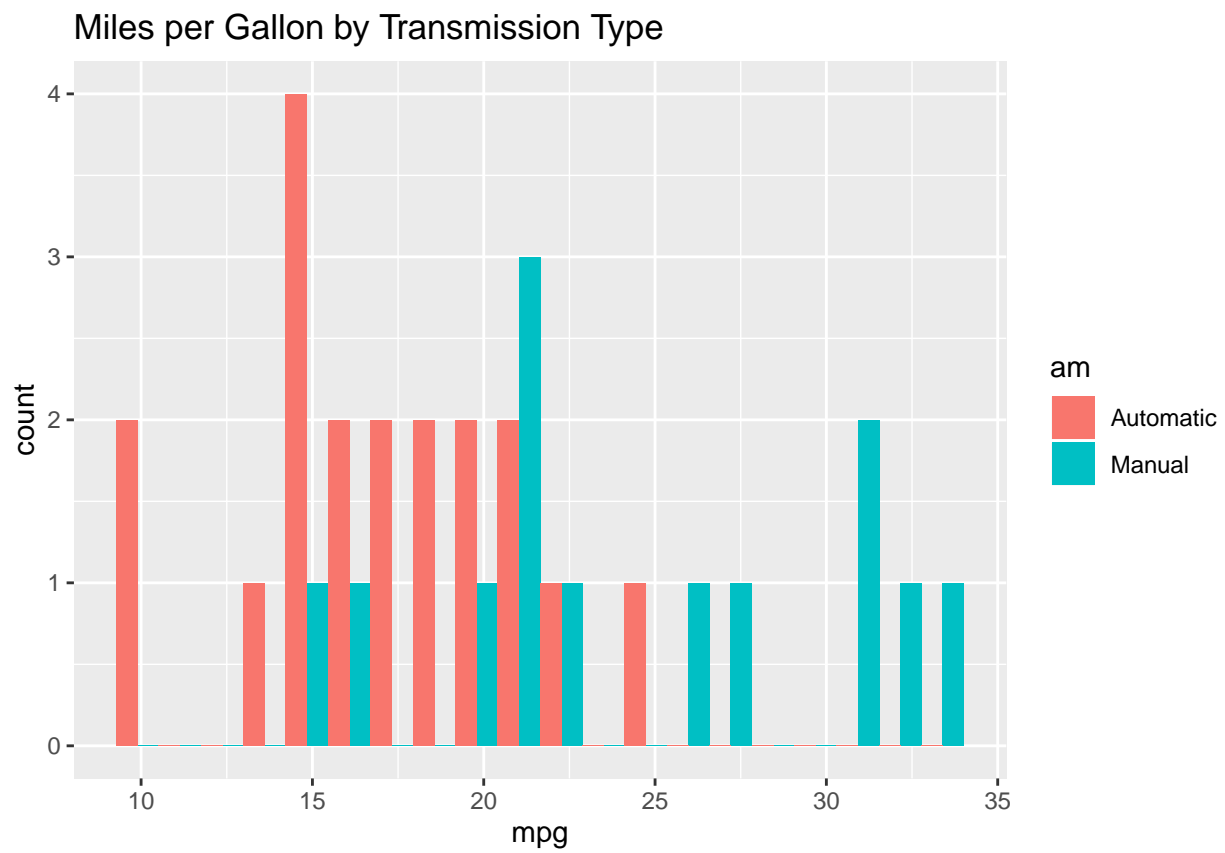
## End of Report

Please see appendix for plots and summary statistics

## Appendix

Plots

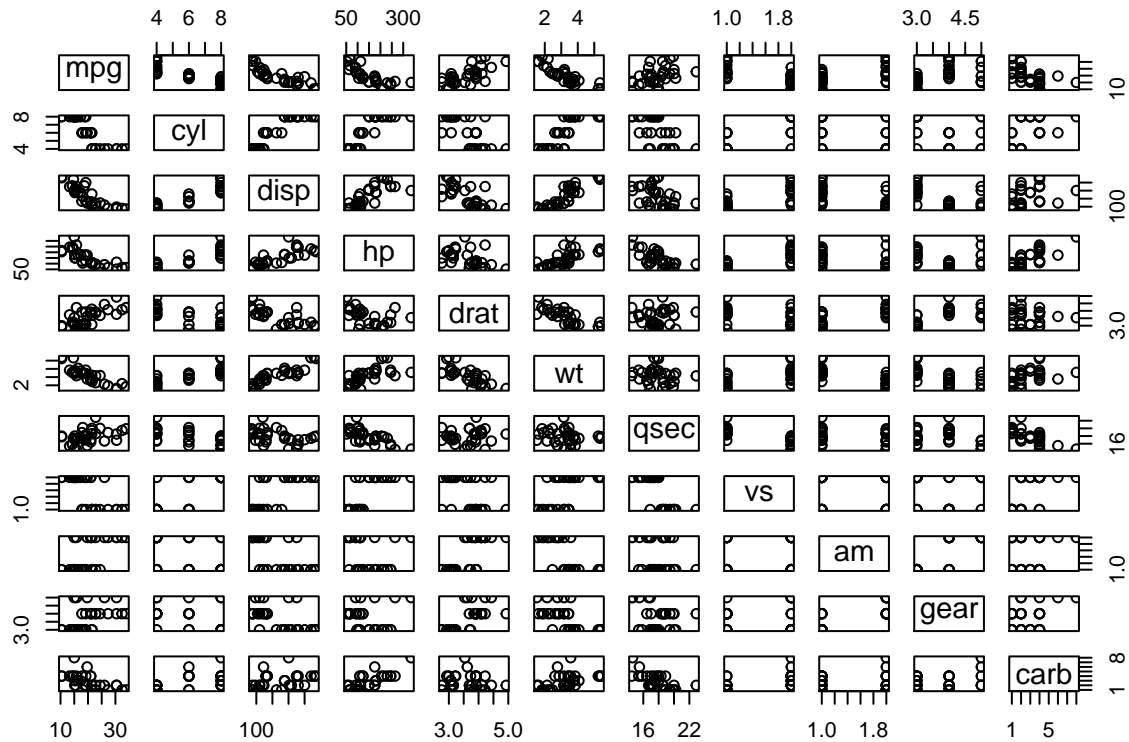
1



#2

```
#### Pairs #3
```

```
pairs(mtrend)
```



```
#### Model 1 Summary
```

```
summary(model_1)
```

```
##
## Call:
## lm(formula = mpg ~ ., data = mtrend)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.4506 -1.6044 -0.1196  1.2193  4.6271
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.62114    19.02842   0.663  0.5144
## cyl         -0.11144     1.04502  -0.107  0.9161
## disp         0.01334     0.01786   0.747  0.4635
## hp          -0.02148     0.02177  -0.987  0.3350
## drat         0.78711     1.63537   0.481  0.6353
## wt          -3.71530     1.89441  -1.961  0.0633
## qsec         0.82104     0.73084   1.123  0.2739
## vsV-engine  -0.31776     2.10451  -0.151  0.8814
## amManual     2.52023     2.05665   1.225  0.2340
## gear         0.65541     1.49326   0.439  0.6652
```

```
## carb          -0.19942    0.82875  -0.241   0.8122
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.65 on 21 degrees of freedom
## Multiple R-squared:  0.869, Adjusted R-squared:  0.8066
## F-statistic: 13.93 on 10 and 21 DF,  p-value: 3.793e-07
```

## Model 2 Summary

```
summary(model_2)
```

```
##
## Call:
## lm(formula = mpg ~ cyl + disp + drat + wt, data = mtrend)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.4067 -1.4096 -0.4954  1.3346  6.0729
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  41.160271   7.304325   5.635 5.56e-06 ***
## cyl          -1.786074   0.634821  -2.814 0.00903 **
## disp         0.007472   0.012062   0.619 0.54080
## drat        -0.010492   1.337929  -0.008 0.99380
## wt          -3.638075   1.102460  -3.300 0.00272 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.642 on 27 degrees of freedom
## Multiple R-squared:  0.8326, Adjusted R-squared:  0.8078
## F-statistic: 33.57 on 4 and 27 DF,  p-value: 4.057e-10
```

## Chosen Model Summary Statistics and plot

```
summary(model_3)
```

```
##
## Call:
## lm(formula = mpg ~ cyl + (am * wt), data = mtrend)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.4621 -1.4913 -0.7879  1.3959  5.3499
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  34.2830    2.7965  12.259 1.52e-12 ***
## cyl          -1.1814    0.3803  -3.106 0.00442 **
## amManual     11.9385    3.8453   3.105 0.00444 **
## wt          -2.3689    0.8244  -2.874 0.00782 **
## amManual:wt  -4.1974    1.3115  -3.200 0.00350 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##  
## Residual standard error: 2.265 on 27 degrees of freedom  
## Multiple R-squared:  0.877, Adjusted R-squared:  0.8588  
## F-statistic: 48.13 on 4 and 27 DF,  p-value: 6.643e-12
```

```
plot(model_3)
```

