Regression Models Course Project

Cliff Hayes

March 27, 2016

Executive Summary

This paper will analyze the `mtcars` dataset in an effort to answer the following questions:

1. Is an automatic or manual transmission better for MPG (miles per gallon)?
2. Can the MPG difference between automatic and manual transmissions be quantified?

The paper will explore 10 automobile design and performance variables in relation to MPG. It will be shown that a regression model can be built using 3 of these variables to quantify the MPG difference between automatic and manual transmissions which shows a manual transmission results in higher MPG. Specifically, it will be shown that a car having a manual transmission weighing 2500 lbs will result in an additional 3.7265 MPG when compared to a car having an automatic transmission when both cars have the same weight and acceleration.

Load, process and explore the data

```
library(datasets)
data(mtcars)
# What is contained in this dataset?
mtcars[1:3,]
```

```
##                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
```

Since we are looking to characterize the variable of `mpg` based on the transmission type of automatic or manual (variable `am`), let's convert the variable `am` from numeric to a factor variable. Next we can do a simple boxplot to show the difference in MPG based on transmission type with the mean mpg listed for each transmission type.

```
mtcars$am <- factor(mtcars$am,labels=c('Automatic','Manual'))
means <- aggregate(mpg ~ am, mtcars, mean)
means$mpg <- round(means$mpg, 2)
# See Appendix:Figure 1
```

This plot shows that manual transmission cars average higher mpg ratings than automatic. Let's consider the weight of the car based on its transmission type. It seems plausible that heavier cars would yield lower MPG and lighter cars would yield a higher MPG.

```
# See Appendix:Figure 2
```

MPG is affected by weight and transmission type. The scatterplot also shows us that cars with automatic transmissions tend to be heavier than cars with manual transmissions.

Regression modeling

For our initial regression model, we'll simply start with `mpg` as the dependent variable and show all the independent variables as regressors.

```
fit1 <- lm(mpg ~ ., mtcars)
summary(fit1) # results are hidden
```

Here we see the regressors of `wt` (-3.72), `qsec` (0.82) and `amManual` (2.52) affecting `mpg` the most, but the inclusion of all the variables renders only `wt` marginally statistically significant at 0.63. Let's see what model is produced using only the aforementioned three regressors.

```
fit2 <- lm(mpg ~ wt + qsec + am, mtcars)
summary(fit2) # results are hidden
```

This is looking pretty good. All three regressors have statistically significant effects upon the outcome of `mpg` with `wt` and `qsec` at < 0.001, and `am` at 0.48. The adjusted R-squared value is now 0.8336 which would explain 83% of the variance in MPG. Recall from our exploratory plot in Figure 2, we saw that weight and transmission type appeared correlated in their effect upon MPG. Let's look at that by adding a fourth regressor which is the interaction of `wt` and `am`.

```
fit3 <- lm(mpg ~ wt + qsec + am + wt:am, mtcars)
summary(fit3) # results are hidden
```

This model appears to be even stronger than our original three regressor model. With a residual error of 2.084, an adjusted R-squared of 0.8804 (explaining 88% of the variance in the MPG variable) and all four regressors being statistically significant, it looks like we should select this model for our analysis.

The model is telling us that when we hold `qsec` and `weight` constant, cars with a manual transmission will add 14.0479 + ((-4.141)*wt) MPG on average over cars with an automatic transmission. For instance, a car having a manual transmission weighing 2500 lbs will result in an additional 3.7265 MPG more than a car having an automatic transmission if both cars have the same `wt` and `qsec`.

Diagnostics

See Appendix:Figures 3 & 4

Figure 3 is a pairs plot showing distribution and correlation of the selected variables.

Figure 4 is a residuals plot showing:

1. The Residuals vs. Fitted plot shows homoscedasticity with no consistent pattern
2. The Normal Q-Q plot shows a normal distribution where the points fall near the line.
3. The Scale-Location plot shows random distribution which affirms constant variance.
4. The Residuals vs. Leverage plot shows all values between the 0.5 bands with no outliers.

Appendix: Figures

Figure 1 - boxplot `mpg` by `am`

```
library(ggplot2)
plot1 <- ggplot(mtcars, aes(am, mpg, fill = am)) + geom_boxplot() +
    scale_fill_discrete(name = "Transmission\nType") +
    geom_text(data = means, aes(label = mpg, y = mpg + 1))
plot1
```

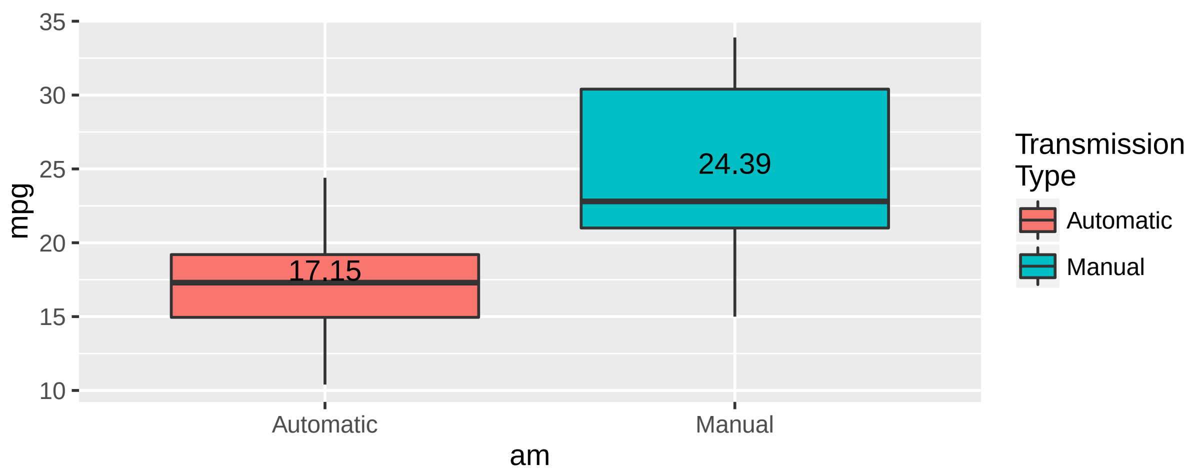

Figure 2 - scatterplot of `mpg` by `wt` also showing `am` by color

```
plot2 <- ggplot(mtcars, aes(wt, mpg, group = am, color = am, height=5, width=5)) +
    geom_point() + scale_colour_discrete(name = "Transmission\nType") +
    xlab("weight") + ggtitle("Scatterplot of MPG by weight and transmission type")
plot2
```

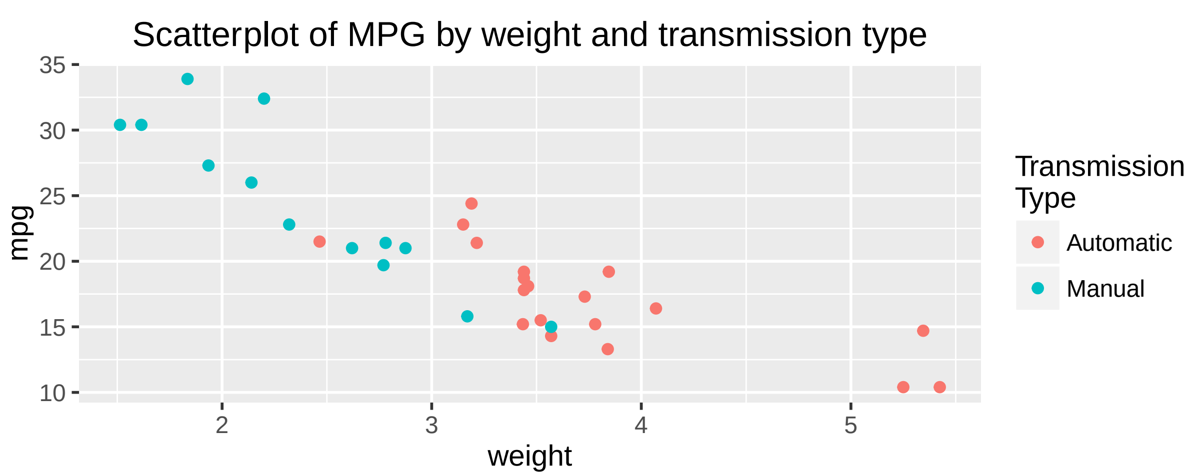

Figure 3 - Pair plot of selected variables

```
require(GGally)
mtcars_vars <- mtcars[, c(1,6,7,9)]
plot3 <- ggpairs(mtcars_vars)
plot3
```

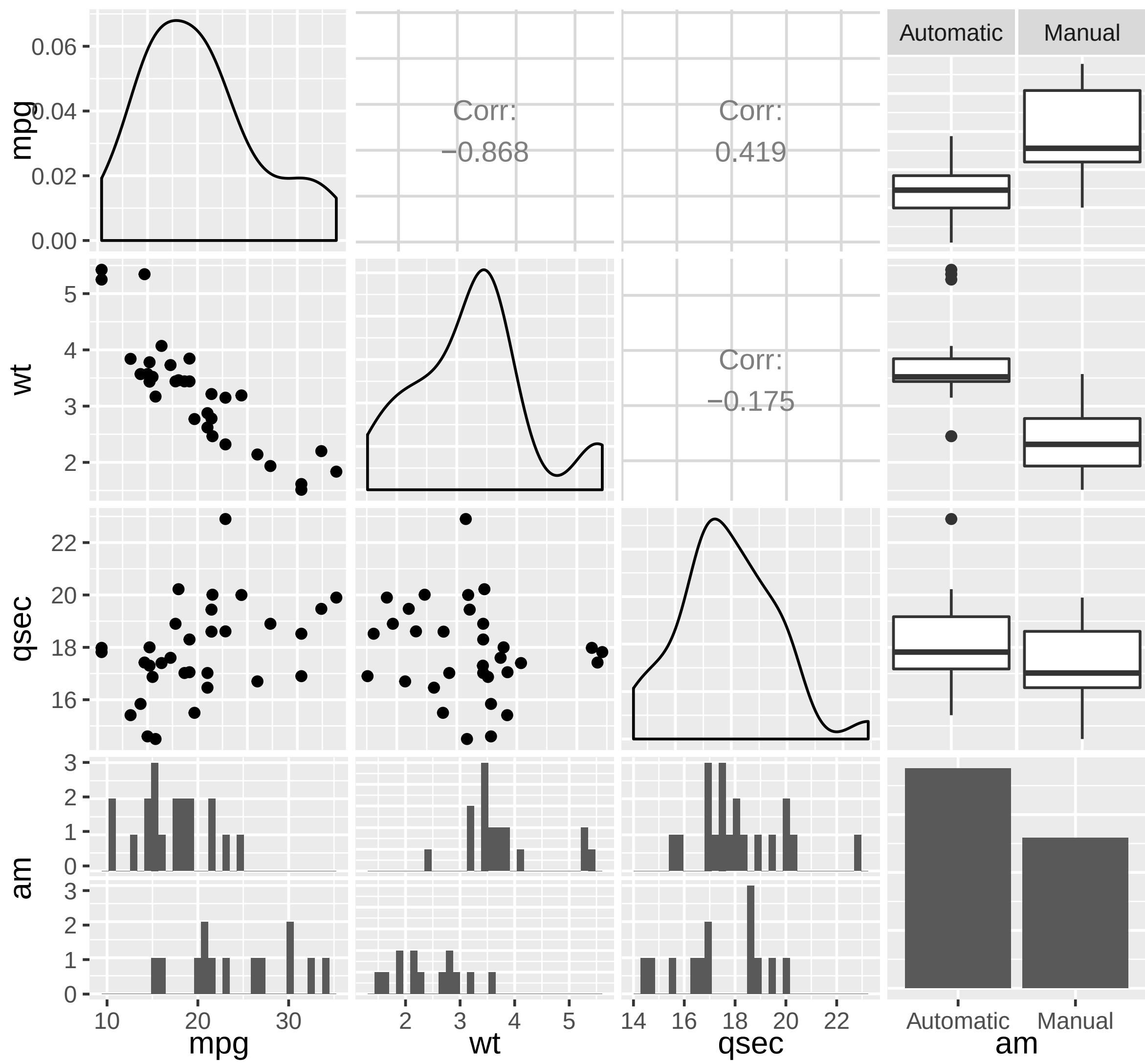

Figure 4 - Residual plot

```r
par(mfrow = c(2,2))
plot(fit3)
```