Analysis of MPG by Manual vs. Automatic

Andrew Witherspoon

12/12/2018

In this analysis we will use data extracted from the 1974 Motor Trend Magaizne to analyze fuel consumption (MPG) comparing automatic to manual transmissions. This data set provides fuel consumption, as well as 10 other variables (aspects of design and performance) for 32 automobiles.

The first question we will attempt to answer is, “is an automatic or manual transmission better for fuel consuption?” So let’s start by using just those two variables to construct a quick linear regression model, without an intercept.

MODEL FITTING

```
rownames(mtcars) <- NULL
mtcars$cyl <- factor(mtcars$cyl)
mtcars$vs <- factor(mtcars$vs)
mtcars$am <- factor(mtcars$am)
fit1 <- lm(mpg ~ factor(am) -1, data = mtcars)
summary(fit1)$coef
```

```
##             Estimate Std. Error  t value     Pr(>|t|)
## factor(am)0 17.14737   1.124603 15.24749 1.133983e-15
## factor(am)1 24.39231   1.359578 17.94109 1.376283e-17
```

This initial model indicates that from out data set, cars with automatic transmissions have an average of 17 mpg, while cars with manual transmissions get an average of 24 mpg. This isn’t particularly instersting, as the type of cars that have manual transmission may be bigger heavier cars, with bigger more powerful engines.

Let’s do another model and include vehicle weight and engine displacement as regressor variables.

```
fit2 <- lm(mpg ~ factor(am) + disp + wt, data = mtcars)
summary(fit2)$coef
```

```
##                Estimate Std. Error    t value     Pr(>|t|)
## (Intercept) 34.67591088 3.24060891 10.7004306 2.115200e-11
## factor(am)1  0.17772414 1.48431586  0.1197347 9.055483e-01
## disp        -0.01780491 0.00937465 -1.8992613 6.787740e-02
## wt          -3.27904388 1.32750927 -2.4700723 1.986658e-02
```

The results here show that when we account for engine size and vehicle weight, there is almost no discernable difference between automatic and manual transmission’s affect on mpg. A manual transmission may account for less than a .2 mpg improvement, but even that is not a statistically significant differance (note the high p-value).

Following the same logic, engine power is likely a factor in mpg. Let’s look at the average horsepower for automatic and manual transmissions.

```
avgs <- aggregate(disp ~ am, mtcars, mean)
names(avgs) <- c("Transmission (0 = automatic, 1 = manual)", "Average HP")
avgs
```

```
##   Transmission (0 = automatic, 1 = manual) Average HP
## 1                                       0   290.3789
## 2                                       1   143.5308
```

So we see that autmatic transmissions are used on cars that on average, have more powerful engines. Using horsepower as a regressor will help ensure we are measuring the mpg difference in transmission type while accounting for the difference in engine power.

```
fit3 <- lm(mpg ~ factor(am) + hp, data = mtcars)
summary(fit3)$coef
```

```
##                Estimate  Std. Error    t value     Pr(>|t|)
## (Intercept) 26.5849137 1.425094292 18.654845 1.073954e-17
```

```
## factor(am)1  5.2770853 1.079540576  4.888270 3.460318e-05
## hp          -0.0588878 0.007856745 -7.495191 2.920375e-08
```

Now we see there is a significant difference between automatic and manual transmissions. When accounting for engine horsepower, a manual transmission appears to get significantly better gas milage.

Since vehicle weight is also highly correlated with mpg (see pairs plot in Plot Appendix), let's do one more model with both horsepower and weight as additional regressors to accompany transmission type.

```
fit4 <- lm(mpg ~ factor(am) + hp + wt, data = mtcars)
summary(fit4)$coef
```

```
##                Estimate  Std. Error   t value     Pr(>|t|)
## (Intercept) 34.00287512 2.642659337 12.866916 2.824030e-13
## factor(am)1  2.08371013 1.376420152  1.513862 1.412682e-01
## hp          -0.03747873 0.009605422 -3.901830 5.464023e-04
## wt          -2.87857541 0.904970538 -3.180850 3.574031e-03
```

MODEL SELECTION

```
AIC(fit1,fit2, fit3, fit4)
```

```
##      df      AIC
## fit1  3 196.4844
## fit2  5 166.1514
## fit3  4 164.0061
## fit4  5 156.1348
```

Model fit4 has the absolute lowest AIC, but the automatic/manual coefficient is insignificant, so even though it might work well as a prediction model, it does not do well to answer the question we are trying to answer (it relys more heavily on other variables to predict mpg). The next best AIC is fit3, for which the transmission type coefficient is significant.

Next, we will look at the Residual Plot for each model. These are available in the Plot Appendix. Again, we are drawn to fit3, as this model shows almost no pattern. Model fit3 will be our model. Let's take one more look at the coefficients:

```
summary(fit3)$coef
```

```
##               Estimate  Std. Error   t value     Pr(>|t|)
## (Intercept) 26.5849137 1.425094292 18.654845 1.073954e-17
## factor(am)1  5.2770853 1.079540576  4.888270 3.460318e-05
## hp          -0.0588878 0.007856745 -7.495191 2.920375e-08
```

SUMMARY / QUESTIONS ANSWERED

Is an automatic or manual transmission better for MPG?

Based on our chosen model, which uses horsepower as a regressor, a manual transmission is better for MPG.

Quantify the MPG difference between automatic and manual transmissions

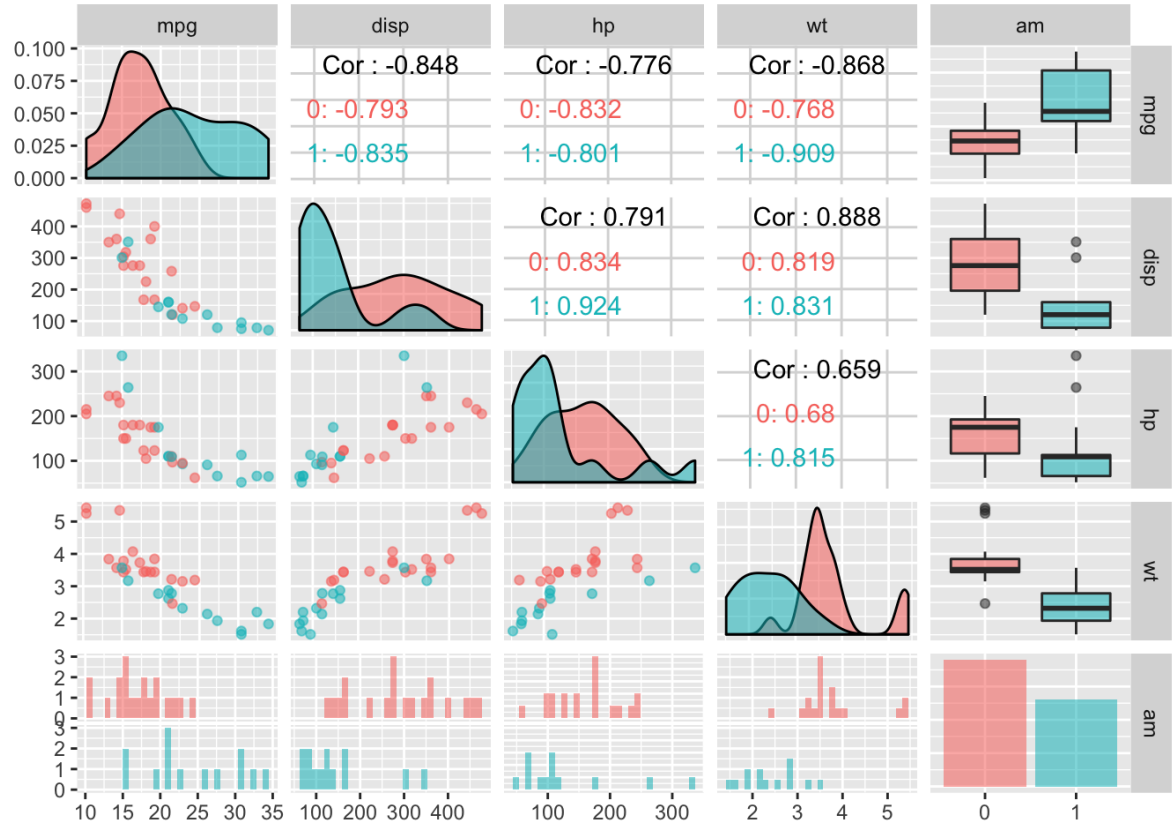
At a given level of engine horsepower, a manual transmission is expected to have a 5 MPG improvement over an automatic transmission.

We are working with a small sample size, and it is difficult to know if we our data is anaomolistic, or if there are some other factors in that weren't available in the data that could have provided more insight. The estimated improvement in MPG is just that, an estimate; however, it appears a fairly safe assumption that a manual transmission does does improve fuel efficiency.

PLOT APPENDIX

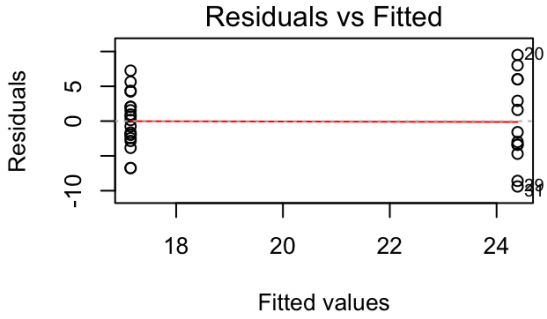
Pairs Plot:

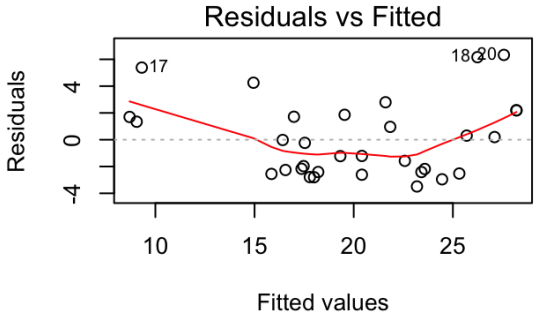
```
library(GGally)
ggpairs(mtcars[,c(1,3,4,6,9)], aes(colour = am, alpha = 0.4))
```

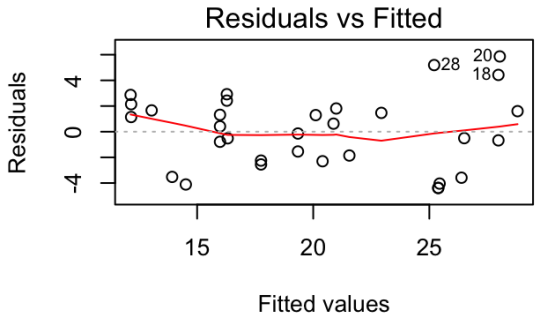


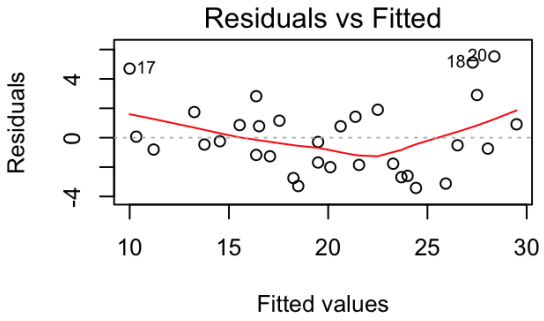
Reisdual Plots: *top left: fit1, top right, fit2, bottom left: fit3, bottom right fit4*

```
par(mfrow = c(2,2))
plot(fit1, which = 1)
plot(fit2, which = 1)
plot(fit3, which = 1)
plot(fit4, which = 1)
```










```
par(mfrow = c(1,1))
```