Mileage: Is Automatic better than Manual Transmission?

Talha Mahmood Chaudhry

8/31/2017

Executive Summary

It has long been a subject of debate whether Automatic Transmission has lower mileage than Manual Transmission. We here at *Motor Trend* want to illuminate our readers with our proposition and the resulting analysis. Using the **mtcars** dataset in R, it is determined that mileage increases by 7.245 for manual from automatic. However, upon closer inspection and varoance testing, it is determined that changes in mileage are more explained by number of Cylinders and Weight of the car. Thus, although it does seem that Manual has better mileage than Automatic, the other factors have more impact on mileage than transmission type.

Data Structure and Processing[1]

Here's a look at a few lines of the data. For a documentaion about the data click here.

```
##                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
```

The features of the data are as follows:
1. **mpg** Miles/(US) gallon 7. **qsec** 1/4 mile time
2. **cyl** Number of cylinders 8. **vs** V/S
3. **disp** Displacement (cu.in.) 9. **am** Transmission (0 = automatic, 1 = manual)
4. **hp** Gross horsepower 10. **gear** Number of forward gears
5. **drat** Rear axle ratio 11. **carb** Number of carburetors
6. **wt** Weight (1000 lbs)

For our analysis, we can safely disregard "qsec" as a predictor, since its a determinant of acceleration. Whereas "cyl", "vs", "am", "gear" and "carb" are categorical variables and are treated as factors.

Exploratory Analysis

Looking at the boxplot (Fig 1 Appendix) depicting mileage (mpg) grouped by transmission it can be observed that Manual Transmission does have a better mileage than Automatic Transmission. The pairs plot shows that the features of the dataset have some correlation with each other. This will have to be taken into consideration when building models.(Appendix - Fig 2)

Inference

A two sample t-test was performed to test this against the null hypothesis that mean mileage for both types of Transmission are the same.

[1] Click here for R markdown file

The p-value, 0.00137, is small enough that the null hypothesis can be rejected in favor of the alternative. The means, 24.39 mpg for Manual Transmission and 17.15 mpg for Automatic Transmission, correspond with what was observed in the boxplot.

Regression Analysis

Using R's *step* function an optimal model is generated:

```
mtcars2 <- dat[,-7] # remove qsec as a feature
initmdl <- lm(mpg ~ ., data = mtcars2) # model mpg against all other variables
bestmdl <- step(initmdl, direction = "both") # determine step's best model
```

The model generated is mpg ~ cyl + hp + wt + am. The coefficients and their p-values are as follows:

```
## (Intercept)        cyl6        cyl8          hp          wt    amManual
##      33.708      -3.031      -2.164      -0.032      -2.497       1.809
```

```
## [1] "P-Values"
```

```
## (Intercept)        cyl6        cyl8          hp          wt    amManual
##       0.000       0.041       0.352       0.027       0.009       0.206
```

Notice, that the variable "am", signifiying Transmission type, is represented as *amManual*, which suggests that mileage increases by 1.81 when change from Automatic to Manual, keeping all else constant. However, the **p-value**, 0.206 is greater than $\alpha = 0.5$ significant level, which means the null hypothesis that the coefficient of am is zero cannot be confidently rejected.
The best model was then tested against two other models. One, base model, is of the formula mpg ~ am, the second, fit3 has the same formula as the best model excluding the variable **am**. Results of the ANOVA (Analysis of Variance) Test are as follows:

```
basemdl <- lm(mpg ~ am, mtcars2)
fit = update(bestmdl, mpg ~ cyl + hp + wt)
vartest <- anova(basemdl, fit, bestmdl)
```

The p-value, 6.7009102×10^{-9}, for model fit is below our threshold of 0.05. Here are the coefficients

```
##                Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 35.84599532  2.0410191 17.562793 2.670320e-16
## cyl6        -3.35902490  1.4016697 -2.396445 2.374718e-02
## cyl8        -3.18588444  2.1704753 -1.467828 1.537047e-01
## hp          -0.02311981  0.0119522 -1.934357 6.361269e-02
## wt          -3.18140405  0.7196010 -4.421067 1.441756e-04
```

The p-values suggest that most of the coefficients are significant. The **wt** coefficient suggests that for an increase of 1000 lbs, mileage decreases by -3.181 with all else constant. The intercept is mileage for a 4-cylinder car with everything else accounted for. Mileage decreases if cylinders increase. Any changes in Horsepower, hp, seems to have a minute effect.

Residual diagnostics of the model fit suggest that the residulas are quasi-normally distributed, while the outliers have not much significant leverage. (See Appendix - Fig 3) *****

Appendix

Figure 1

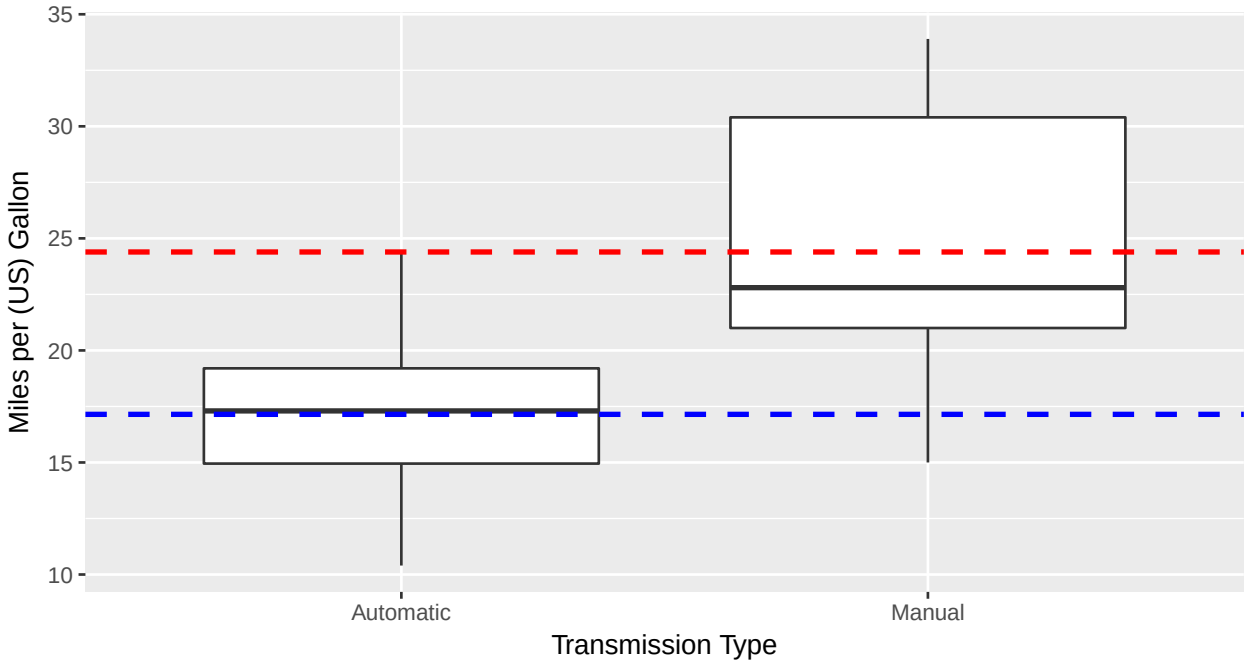
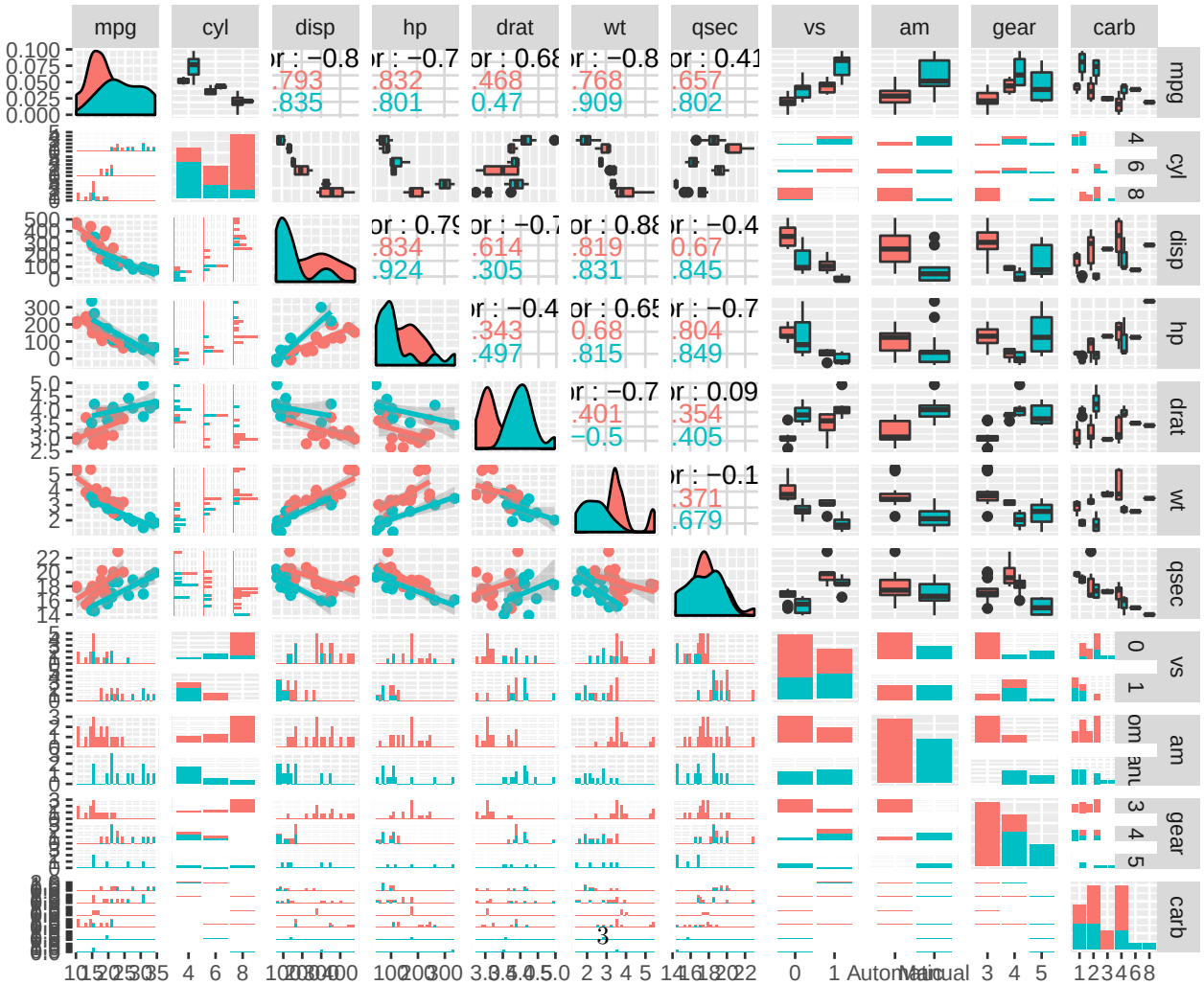
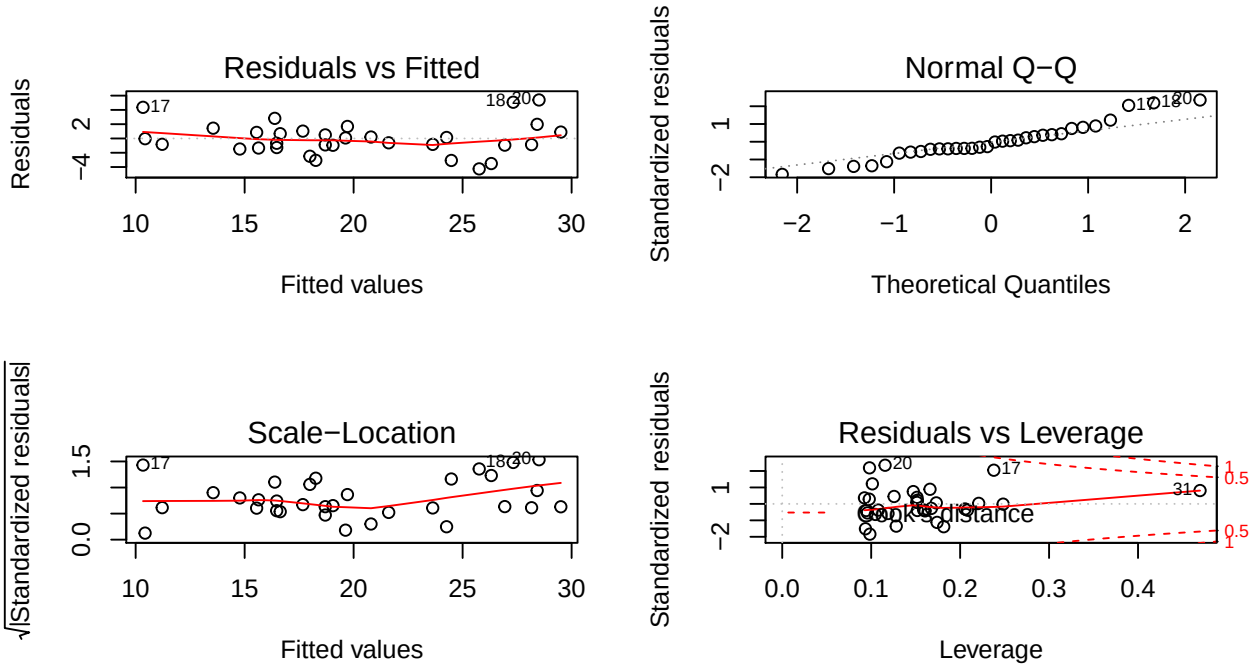


Fig 2






```
##
##  Welch Two Sample t-test
##
## data:  dat$mpg[dat$am == "Manual"] and dat$mpg[dat$am == "Automatic"]
## t = 3.7671, df = 18.332, p-value = 0.001374
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   3.209684 11.280194
## sample estimates:
## mean of x mean of y
##  24.39231  17.14737
```