Automatic v. Manual Transmission: Which is Better for MPG?

Ilan Grapel

March 2, 2017

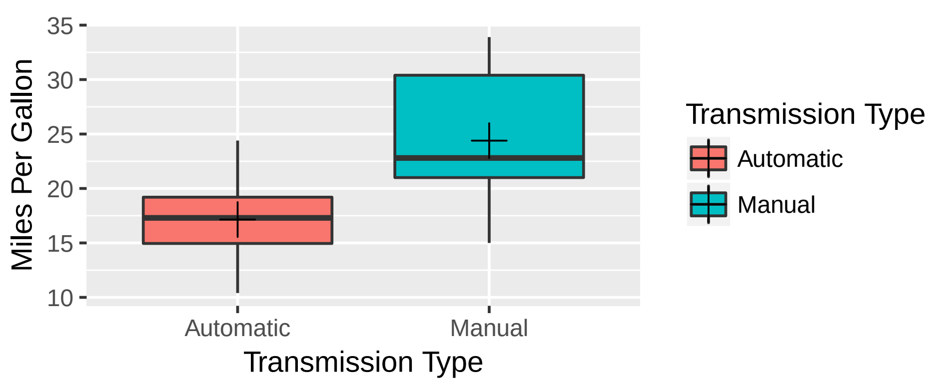
Executive Summary: This report examines whether automatic or manual transmission is better for MPG.To this end, the report will perform explanatory (§ I) and regression analyses (§ II) to determine if there is a difference in MPG for automatic versus manual transmissions. The report will investigate some regression plots to assess the selected model (§ III). Finally, the report will seek to quantify the mpg difference, if any, between the two transmissions (§ IV).

I. Explanatory Analyses: The report will rely on the "mtcars" data set. The `str(mtcars)` command in R reveals that there are 32 observations of 11 variables. The relvant variables for this analysis are the mpg variable (our response variable) and the am variable (our explanatory variable), which according to the documentation page is coded as a dummy variable (0 = automatic, 1 = manual).

a. Descriptive Statistics: According to `table(mtcars$am)` there are 19 automatic cars and 13 manual cars in our data set. According to `summary(mtcars$mpg)`, the five-number summary and mean for the cars' mpg is:

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10.40   15.42   19.20   20.09   22.80   33.90
```

b. Graphical Plots: This report will first examine side-by-side box plots comparing automatic to manual transmissions.



The manual cars appear to have more MPG overall. The box plots show that 75% of manual cars are more energy efficient than 75% of automatic cars. There is some overlap for manual cars in their first quartile.

To examine the affect of covariates, the report will produce three dot plots, which color dots according to other covariates (displacement, weight, and horsepower)[1].The colored box plots show some patterns, indicating that the covariates may be responsible for some of the difference in mpg.

II. Regression Models: The report will utilize a variety of regression models to evaluate the relationship between transmission and mpg. Through an ANOVA analysis, the report will determine essential covariates to include in the model. Finally, the report will look at select residual plots to ensure that our model is apt.

a. OLS Linear Regression: First, we will examine mpg regressed against transmission type[2]:

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    17.15       1.12   15.25        0
## am              7.24       1.76    4.11        0
```

In this case, the intercept estimate, 17.15 refers to the mean mpg value of cars with automatic transmission. The am estimate coefficient, 7.24, refers to the increase in the mean mpg value for manual transmission cars in comparison to automatic. In other words, the manual cars in the data set have a mean of about 24.39.

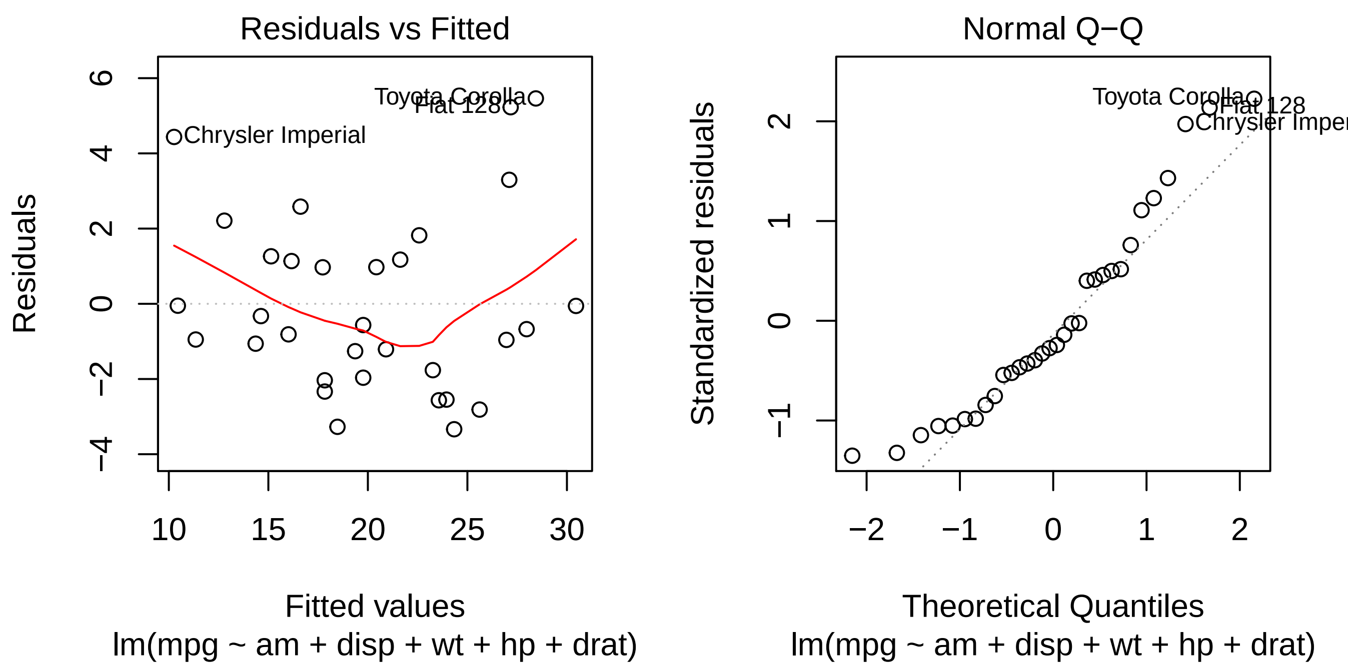
b. Multi-Variable Regression with all Covariates: Now, the report will try to improve its regression model by taking into consideration the covariates. Our colored dot plots, in graphical analyses above, indicate that covariates may affect the coefficient value of ourlinear regression.

We will perform a multi-variate regression analysis with all the 11 covariates included[3]. When holding the other variables constant, there is only a 2.5 mpg mean increase when switching from automatic to manual. Moreover, the t-statistics for this slope is 1.12, which corresponds to a P-value of 0.23. Thus, we do not have sufficient evidence to conclude that there is a difference in mpg under this model.

c. Nested ANOVA Analysis: A nested ANOVA analysis may allow us to find a better fitting model[4].

According to ANOVA, fit5 includes informative covariates (with a P-value of .016), while fit7 and fit9 include unnecessary covariates. Thus, we will refer to fit5 for the rest of our analysis[5]. This model, with an r^2 of .84, can explain 84% of the variation in mpg.

III. Residual Plots and Diagnostics for the Fit5 Model: The report will examine some basic residual plots to ensure that it has not missed any unusual patterns.



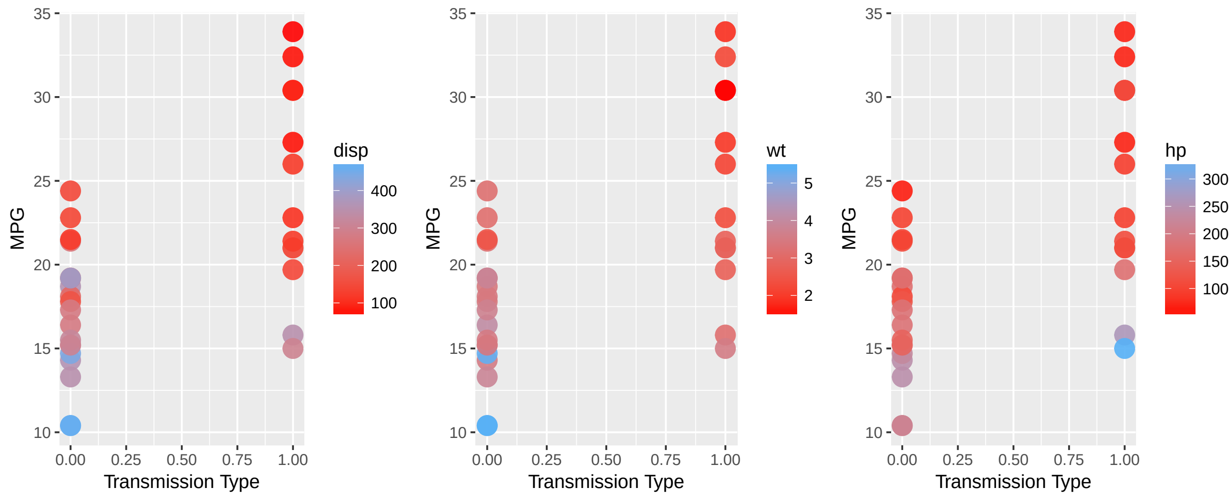
Overall, these residual plots do not show any unusual patters. Potential outliers are the Fiat 128, Toyota Corolla, and Chrysler Imperial. It may be worthwhile to investigate these data points further.

IV. Conclusion: This section will try to quantify the answer to the question - Is there a difference in mpg efficiency between automatic and manual transmissions?

The Fit5 model informs us that a switch to manual transmission increases the mean mpg by 1.6, holding displacement, weight, horsepower, and rear axle ratio constant. This coefficient has a t-value of only 1.03, which corresponds to a P-value of .31. Consequently, the data does not offer convincing evidence that a manual transmission is better for mpg.

APPENDIX

1.



2.

```r
fit_lin<- lm(mpg~am, mtcars)
round(summary(fit_lin)$coef,2)
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    17.15       1.12   15.25        0
## am              7.24       1.76    4.11        0
```

3.

```r
fit <- lm(mpg ~ . , data = mtcars)
summary(fit)$coef
```

```
##                Estimate  Std. Error    t value   Pr(>|t|)
## (Intercept) 12.30337416 18.71788443  0.6573058 0.51812440
## cyl         -0.11144048  1.04502336 -0.1066392 0.91608738
## disp         0.01333524  0.01785750  0.7467585 0.46348865
## hp          -0.02148212  0.02176858 -0.9868407 0.33495531
## drat         0.78711097  1.63537307  0.4813036 0.63527790
## wt          -3.71530393  1.89441430 -1.9611887 0.06325215
## qsec         0.82104075  0.73084480  1.1234133 0.27394127
## vs           0.31776281  2.10450861  0.1509915 0.88142347
## am           2.52022689  2.05665055  1.2254035 0.23398971
## gear         0.65541302  1.49325996  0.4389142 0.66520643
## carb        -0.19941925  0.82875250 -0.2406258 0.81217871
```

4.

```r
fit <- lm(mpg ~ . , data = mtcars)
fit1 <- lm(mpg ~ am, data = mtcars)
fit3 <- update(fit, mpg ~ am + disp + wt)
fit5 <- update(fit, mpg ~ am + disp + wt + hp+drat)
fit7<- update(fit, mpg ~ am + disp + wt + hp+drat+gear+carb)
fit9<- update(fit, mpg ~ am + disp + wt + hp+drat+gear+carb+cyl+vs)
anova(fit1, fit3, fit5,fit7,fit9)
```

```
## Analysis of Variance Table
```

```
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + disp + wt
## Model 3: mpg ~ am + disp + wt + hp + drat
## Model 4: mpg ~ am + disp + wt + hp + drat + gear + carb
## Model 5: mpg ~ am + disp + wt + hp + drat + gear + carb + cyl + vs
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)
## 1     30 720.90
## 2     28 246.56  2    474.34 33.3704 2.174e-07 ***
## 3     26 175.67  2     70.89  4.9871   0.01636 *
## 4     24 163.53  2     12.14  0.8542   0.43927
## 5     22 156.36  2      7.17  0.5042   0.61078
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

5.

```r
summary(fit5)
```

```
##
## Call:
## lm(formula = mpg ~ am + disp + wt + hp + drat, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.3357 -1.8147 -0.6196  1.1967  5.4609
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 29.744091   6.312978   4.712  7.2e-05 ***
## am           1.636571   1.588684   1.030  0.31243
## disp         0.004746   0.010830   0.438  0.66488
## wt          -3.020731   1.165647  -2.591  0.01547 *
## hp          -0.039701   0.012529  -3.169  0.00389 **
## drat         1.148012   1.449191   0.792  0.43543
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.599 on 26 degrees of freedom
## Multiple R-squared:  0.844,  Adjusted R-squared:  0.814
## F-statistic: 28.13 on 5 and 26 DF,  p-value: 1.037e-09
```