MPG and Transmission

Steeg Pierce

5/29/2020

Executive Summary

The mtcars data does not appear to show a significant relationship between transmission and fuel efficiency (mpg) when accounting for a couple of confounding variables. We are not able to quantify any difference in mpg coming from the transmission type as the relationship is not deemed significant in a linear regression.

Exploring the Data

Just by taking a quick look at the data, we can see that the means of fuel efficiencies in miles per gallon for cars without automatic transmission (24.39) differs from that of cars with automatic transmission(17.15). This, however, does not control for all the other factors that might be confounding that mean nor does it account for the variability of the data.

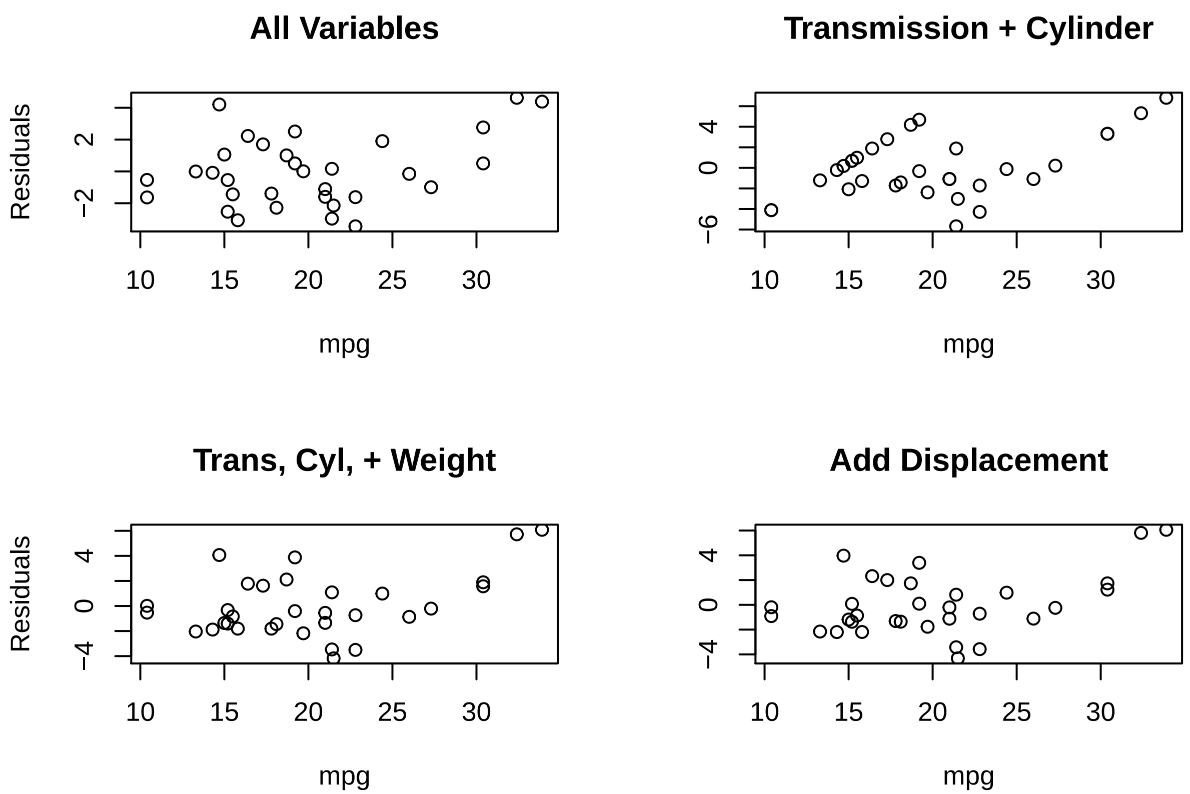
If we regress mpg against transmission type by itself, we get a significant result to the .001 level that says cars with manual transmissions get better fuel efficiency than those with automatic transmissions. When we add in the other variables, though, the p value becomes .234 (not significant). By doing an anova() test, we can see that the most significant variables for mpg are number of cylinders, engine displacement, and weight. As one might expect, the correlation between weight of the car, displacement of the engine, and the number of cylinders is quite high.

We can see the correlations here: Cylinder/Displacement: 0.902 Cylinder/Weight: 0.782 Displacement/Weight: 0.888

Summaries for some of the fits are in the appendix, along with the code.

Testing for Bias

Now that we have a couple significant variables, we can try testing for bias. Let's just take a look at residuals plotted across mpg.



From the plot, it appears that limiting the variables to transmission type and number of cylinders may add some bias, but adding weight to the mix corrects for the bias. We can double check that with an anova test and inded we find that the addition of addition of weight corrects bias, but including additional variables is unnecessary. (note: the anova test can be found in the appendix).

One other way we can look at it is the probability the transmission is manual as mpg rises using a binomial model. With a GLM, it appears the probability of a car having an automatic transmission declines by about 7% with an increase of one mpg, even controlling for cylinders and weight. The p-value, however, is .83. This result is clearly not significant. (summary of glm is also in appendix)

Analyzing the Results

When confounding variables are taken into account, there does not appear to be a relationship between fuel efficiency and transmission type from this data. Although there is a correlation between mpg and transmission type (1), it is nowhere near significant when accounting for weight and number of cylinders. The correlation we are seeing could be coming from compact cars with high mpg.

Appendix

Linear Regressions

```
singleFit <- lm(mpg ~ factor(am), mtcars)
allFit <- lm(mpg ~ ., mtcars)
trioFit <- lm(mpg ~ factor(am) + cyl + wt + disp, mtcars)
cwFit <- lm(mpg ~ factor(am) + cyl + wt, mtcars)
cylFit <- lm(mpg ~ factor(am) + cyl, mtcars)
co_cyl_disp <- cor(mtcars$cyl, mtcars$disp)
co_cyl_wt <- cor(mtcars$cyl, mtcars$wt)
co_disp_wt <- cor(mtcars$disp, mtcars$wt)
```

Model of MPG vs Transmission by itself:

```
##
## Call:
## lm(formula = mpg ~ factor(am), data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.3923 -3.0923 -0.2974  3.2439  9.5077
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   17.147      1.125  15.247 1.13e-15 ***
## factor(am)1    7.245      1.764   4.106 0.000285 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.902 on 30 degrees of freedom
## Multiple R-squared:  0.3598, Adjusted R-squared:  0.3385
## F-statistic: 16.86 on 1 and 30 DF,  p-value: 0.000285
```

MPG vs Transmission, Cylinders, and Weight:

```
##
## Call:
## lm(formula = mpg ~ factor(am) + cyl + wt, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.1735 -1.5340 -0.5386  1.5864  6.0812
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  39.4179     2.6415  14.923 7.42e-15 ***
## factor(am)1   0.1765     1.3045   0.135  0.89334
## cyl          -1.5102     0.4223  -3.576  0.00129 **
## wt           -3.1251     0.9109  -3.431  0.00189 **
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.612 on 28 degrees of freedom
## Multiple R-squared:  0.8303, Adjusted R-squared:  0.8122
## F-statistic: 45.68 on 3 and 28 DF,  p-value: 6.51e-11
```

MPG vs All Other Variables:

```
##
## Call:
## lm(formula = mpg ~ ., data = mtcars)
##
## 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) 12.30337   18.71788   0.657   0.5181
## 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
## vs           0.31776    2.10451   0.151   0.8814
## am           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
```

Bias Testing

```r
par(mfrow = c(2, 2))
mpg <- mtcars$mpg; residAll <- resid(allFit); residTrio <- resid(trioFit); residCW <- resid(cwFit)
residCyl <- resid(cylFit)
plot(mpg, residAll, main = 'All Variables', ylab = 'Residuals')
plot(mpg, residCyl, main = 'Transmission + Cylinder', ylab = '')
plot(mpg, residCW, main = 'Trans, Cyl, + Weight', ylab = 'Residuals')
plot(mpg, residTrio, main = 'Add Displacement', ylab = '')
fitTest <- anova(singleFit, cylFit, cwFit, trioFit, allFit)
gm <- glm(am ~ mpg + cyl + wt, family = 'binomial', data = mtcars)
```

ANOVA Test:

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

```
##
## Model 1: mpg ~ factor(am)
## Model 2: mpg ~ factor(am) + cyl
## Model 3: mpg ~ factor(am) + cyl + wt
## Model 4: mpg ~ factor(am) + cyl + wt + disp
## Model 5: mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear + carb
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)
## 1     30 720.90
## 2     29 271.36  1    449.53 64.0039 8.231e-08 ***
## 3     28 191.05  1     80.32 11.4351  0.002817 **
## 4     27 188.43  1      2.62  0.3732  0.547817
## 5     21 147.49  6     40.93  0.9713  0.468405
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

GLM Summary:

```
##
## Call:
## glm(formula = am ~ mpg + cyl + wt, family = "binomial", data = mtcars)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.97676  -0.25913  -0.03874   0.21014   1.92638
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 18.37047   14.50717   1.266   0.2054
## mpg         -0.07499    0.36600  -0.205   0.8377
## cyl          1.23755    0.88856   1.393   0.1637
## wt          -8.07603    3.30246  -2.445   0.0145 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 43.23  on 31  degrees of freedom
## Residual deviance: 14.69  on 28  degrees of freedom
## AIC: 22.69
##
## Number of Fisher Scoring iterations: 7
```