Analysis of MPG Based on Transmission Type

Kate Prendergast

Executive Summary

In this report, we analyzed the effect of certain factors on MPG. In particular, we looked at the difference between manual and automatic transmissions, and we attempted to quantify this difference. We used the mtcars dataset for this analysis. We found that there was not a significant difference in MPG between manual and automatic transmissions when other factors were held constant. Weight of the vehicle seemed to be a more important factor to focus on.

Note that for the 'am' variable that designates the transmission type, 0 = automatic and 1 = manual.

Analysis

First we load the data and take a look at the first few rows. For our first piece of analysis, we fit a model that just analyzes the change in MPG against transmission type.

```
data("mtcars")
head(mtcars, 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
```

```
fit1 <- lm(mpg ~ factor(am), data = mtcars)
summary(fit1)$coef
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 17.147368   1.124603 15.247492 1.133983e-15
## factor(am)1  7.244939   1.764422  4.106127 2.850207e-04
```

```
confint(fit1)
```

```
##                2.5 %   97.5 %
## (Intercept) 14.85062 19.44411
## factor(am)1  3.64151 10.84837
```

We see that there is a noticeable increase in estimated MPG (7.24) when the transmission type is manual (1) over automatic. Our 95% confidence interval for the difference (3.64 to 10.85) does not include 0, so the difference seems significant. However, intuition tells us that other factors have a significant impact on MPG. Two such factors are the weight of the car and the size of the engine (wt and disp, respectively). We fit another model including these 2 variables so that we can see the change in MPG relative to transmission type while weight and engine size are held constant.

```
fit2 <- lm(mpg ~ factor(am) + wt + disp, 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
## wt          -3.27904388 1.32750927 -2.4700723 1.986658e-02
## disp        -0.01780491 0.00937465 -1.8992613 6.787740e-02
```

```r
confint(fit2)
```

```
##                   2.5 %      97.5 %
## (Intercept) 28.03782444 41.31399731
## factor(am)1 -2.86275907  3.21820735
## wt          -5.99832334 -0.55976441
## disp        -0.03700801  0.00139819
```

Now we see that the increase in MPG for manual transmission is much smaller (0.178) when weight and engine size are held constant. In fact our 95% confidence interval for the difference (-2.86 to 3.22) includes 0, so we cannot say for sure whether manual transmission has a significant effect on MPG. Weight seems to be a more significant driver of MPG.

Next, we compare our two models to determine if the inclusion of weight and engine size creates a better model.

```r
anova(fit1, fit2)
```

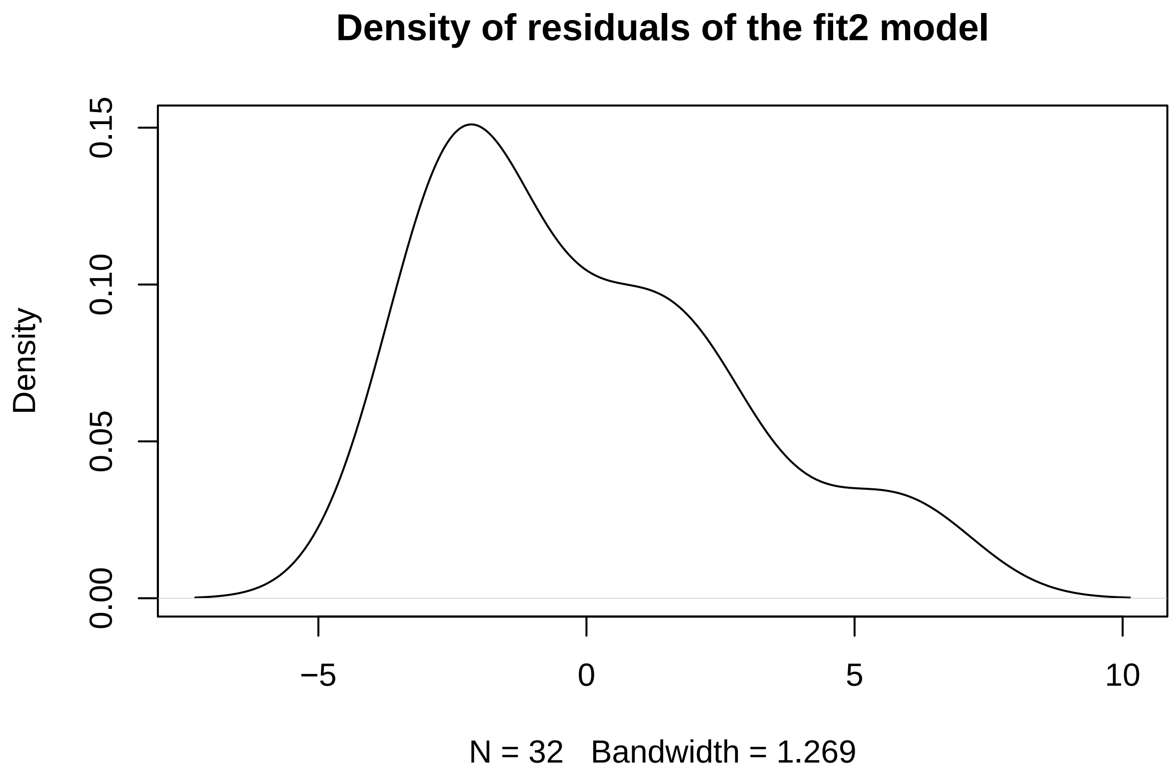
```
## Analysis of Variance Table
##
## Model 1: mpg ~ factor(am)
## Model 2: mpg ~ factor(am) + wt + disp
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     30 720.90
## 2     28 246.56  2    474.34 26.934 2.996e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Our p-value of 2.996e-07 indicates that there is a significant difference in including the two additional variables in our second model.

Finally, we plot the density of the residuals of our second model. (See the appendix for the plot results.) Though it's not a perfect normal distribution, the residuals do seem to generally be normally distributed, and the deviations may be attributable to the small size of the data set. However, this does shed some level of doubt on our model. Ideally, more data would be collected and analyzed to draw a firmer conclusion.

Appendix

```r
plot(density(resid(fit2)), main = "Density of residuals of the fit2 model")
```



Density of residuals of the fit2 model
Density
0.00
0.05
0.10
0.15
−5
0
5
10
N = 32   Bandwidth = 1.269