Transmission type effects on fuel economy

Alex Gaggin

Thursday, October 25, 2015

Summary

While intuitively it looks like manual transmission provides much better fuel economy, statistical analysis of specifications for 32 cars of 1973-74 year models shows that most of this effect caused by manufacturers' preference to install automatic transmissions on heavier cars, which are naturally driven less miles per gallon. Transmission type does influence fuel economy, but it's third source of fuel economy influence after weight and quarter mile time. As it's trumped by first two influencers, its size of effect is hard to measure confidently - the data shows that **switching to manual can save from 0.05 to 5.8 miles per gallon**.

Exploration of the Motor Trends dataset

Dataset has 32 observations on 11 variables and their descriptions are available in dataset's help page. By checking correlations, we see that weight variable affects the fuel economy the most - and by plotting mpg and weight against transmission types (see Appendix 1) we see that weight difference for transmissions is stronger than mpg difference.

```
cor(mtcars)[, "mpg"] %>% "["(-1) %>% abs %>% sort(decr=TRUE) %>% signif(2)
```

```
##   wt  cyl disp   hp drat   vs   am carb gear qsec
## 0.87 0.85 0.85 0.78 0.68 0.66 0.60 0.55 0.48 0.42
```

Significance of MPG difference for transmission types

```
t.test(mtcars[mtcars$am == 0, ]$mpg, mtcars[mtcars$am == 1, ]$mpg)$p.value
```

```
## [1] 0.001373638
```

Small p-value shows that we have to accept alternative hypothesis and presume there's a meaningful difference in mean mpgs between automatic and manual transmissions. Still, strong relation doesn't mean causation. Let's model if this relation still stands after other factors are taken into account.

Linear models

First, let's model dependance just for weight, and then find which of other variables is most correlated with the residuals. We add then this variable to the model as a second regressor.

```
fit1 <- lm(mpg ~ wt, mtcars)
mtcars[ ,-c(1, 6)] %>% cbind(resid(fit1)) %>% cor %>%
        "["(, 10) %>% "["(1:9) %>% abs %>% which.max %>% names
```

```
## [1] "qsec"
```

```r
fit2 <- lm(mpg ~ wt + qsec, mtcars)
summary(fit2)$coefficients
```

```
##              Estimate Std. Error    t value     Pr(>|t|)
## (Intercept) 19.746223  5.2520617   3.759709 7.650466e-04
## wt          -5.047982  0.4839974 -10.429771 2.518948e-11
## qsec         0.929198  0.2650173   3.506179 1.499883e-03
```

We see that time to quarter mile is also significant. It makes sense - presumably there should be two types of wasteful cars - large (heavy) ones, and sporty (with rapid acceleration to quarter mile) ones.
Reiterating the same logic, we build next two nested models.

```r
fit3 <- lm(mpg ~ wt + qsec + factor(am), mtcars)
fit4 <- lm(mpg ~ wt + qsec + factor(am) + carb, mtcars)
anova(fit1, fit2, fit3, fit4)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ wt
## Model 2: mpg ~ wt + qsec
## Model 3: mpg ~ wt + qsec + factor(am)
## Model 4: mpg ~ wt + qsec + factor(am) + carb
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)
## 1     30 278.32
## 2     29 195.46  1    82.858 13.8740 0.0009124 ***
## 3     28 169.29  1    26.178  4.3832 0.0458190 *
## 4     27 161.25  1     8.036  1.3456 0.2562120
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We see that three-variable model still significantly improved results, while including carb variable doesn't improve it further. Residuals for three-variable model don't have any apparent patterns in them (see Appendix 2), although they aren't ideally normal - a bit skewed. Still, we accept this model as the best linear model we could build on this dataset.

```r
fit3 %>% confint %>% signif(2)
```

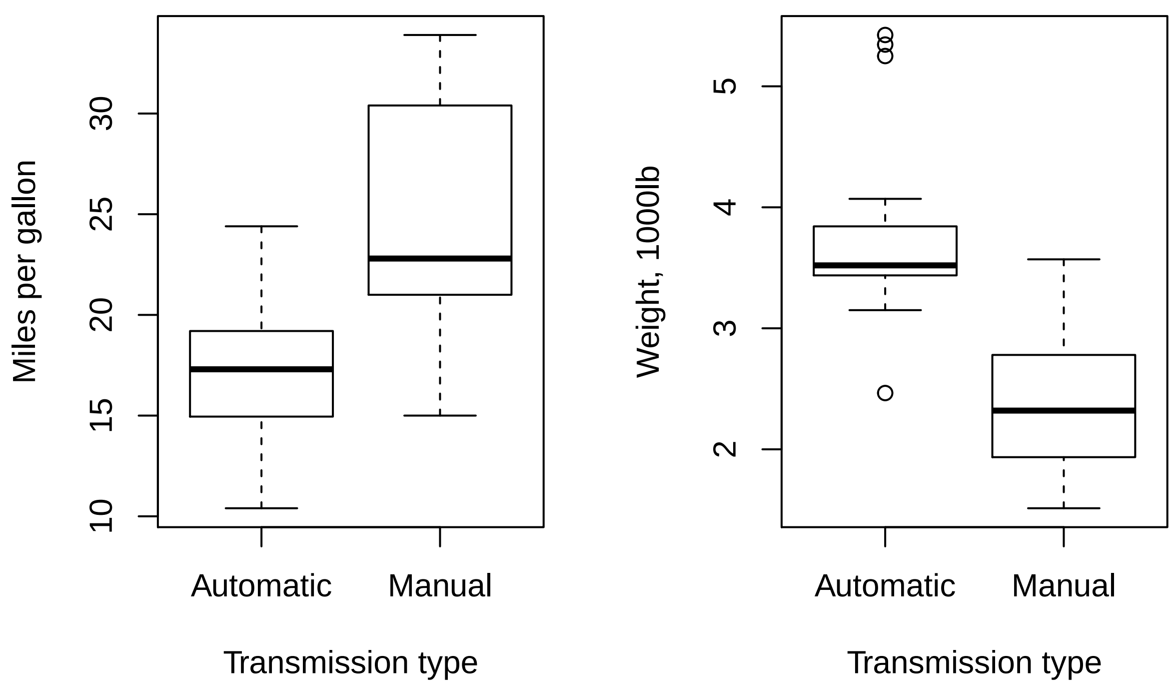
```
##              2.5 % 97.5 %
## (Intercept) -4.600   24.0
## wt          -5.400   -2.5
## qsec         0.630    1.8
## factor(am)1  0.046    5.8
```

We see that manual transmissions allow to get better mileage, although exact scale of that effect is hard to pin down, as it's trumped by bigger effects of cars' weights and acceleration capabilities expressed as the quarter mile time. It's barely fits 95% significance rule and can be anywhere between 0.05 and 5.8 miles per gallon.

Appendices

Appendix 1. Charts for mpg and weight dependance on transmission types

```
par(mfrow = c(1, 2))
amFactor <- factor(mtcars$am, levels=0:1, labels=c("Automatic", "Manual"))
plot(amFactor, mtcars$mpg, xlab="Transmission type", ylab="Miles per gallon")
plot(amFactor, mtcars$wt, xlab="Transmission type", ylab="Weight, 1000lb")
```



Appendix 2. Residuals analysis for the selected model

```r
par(mfrow = c(1, 2)); plot(fit3, which = 1:2)
```