Transmission type effects on fuel economy

Alex Gaggin

Thursday, October 15, 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

```
data(mtcars); dim(mtcars); head(mtcars)
```

```
## [1] 32 11
```

```
##                    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
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

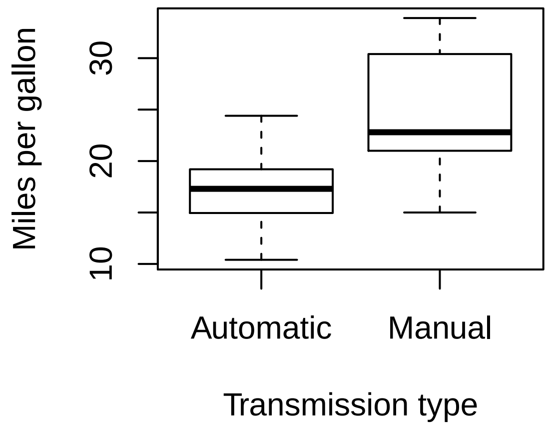
Dataset has 32 observations on 11 variables and their descriptions are available in dataset's help page. Let's see what variables affect mpg the most.

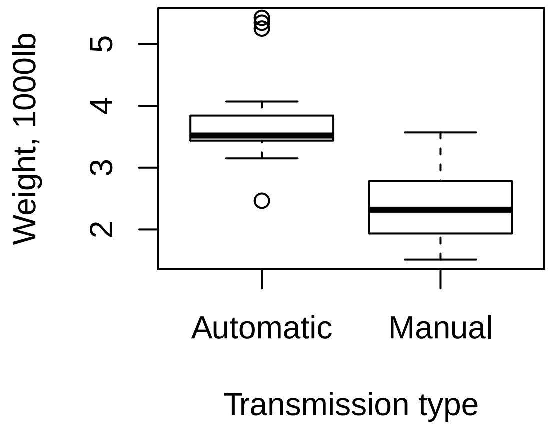
```
cor(mtcars)[, "mpg"] %>% "["(-1) %>% abs %>% sort(decr=TRUE) %>% signif(2)
```

```
##   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
```

We are interested in effect of transmission type on fuel economy, so let's plot it, and also how weight (most correlated to mpg variable) depends on transmission type as well.

```
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")
```





Significance of MPG difference for transmission types

We've seen the difference visually, now let's quantify it.

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

```
##
##  Welch Two Sample t-test
##
## data:  mtcars[mtcars$am == 0, ]$mpg and mtcars[mtcars$am == 1, ]$mpg
## 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:
##  -11.280194  -3.209684
## sample estimates:
## mean of x mean of y
##  17.14737  24.39231
```

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.

Strong relation doesn't mean causation though, and the chart above has shown that transmission types is even more different in weights than in fuel economy. 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.

```r
fit1 <- lm(mpg ~ wt, mtcars)
summary(fit1)$coefficients
```

```
##               Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 37.285126   1.877627 19.857575 8.241799e-19
## wt          -5.344472   0.559101 -9.559044 1.293959e-10
```

Let's see what variable correlate the most with the residuals and include it to the model.

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

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

```
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.

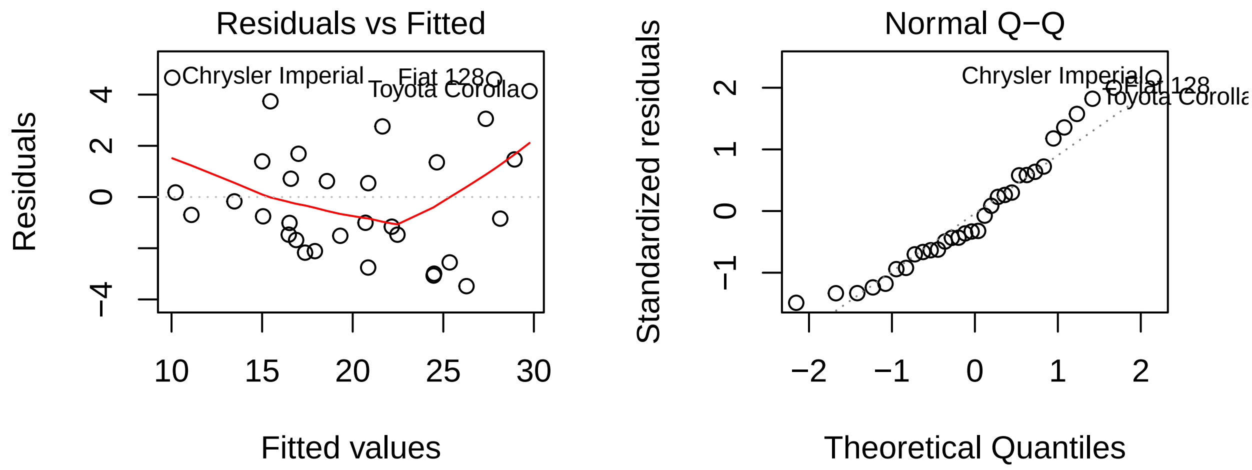
Following the same logic, we build next two nested models.

```
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. Let's check its residuals.

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



Residuals don't have apparent patterns in them, 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.

```
summary(fit3)$coefficients
```

```
##               Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)   9.617781  6.9595930  1.381946 1.779152e-01
## wt           -3.916504  0.7112016 -5.506882 6.952711e-06
## qsec          1.225886  0.2886696  4.246676 2.161737e-04
## factor(am)1   2.935837  1.4109045  2.080819 4.671551e-02
```

```
summary(fit3)$r.squared
```

```
## [1] 0.8496636
```

```
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.