Is an automatic or manual transmission better for MPG?

Lukas Nick

August 24 2014

Executive summary

Questions

This report tries to answer two questions:

1. Is an automatic transmission better for MPG?
2. Quantify the MPG difference between automatic and manual transmissions.

Data

I try to answer these questions based on the dataset mtcars, which has 32 cars and the following variables:

```
##  [1] "mpg"  "cyl"  "disp" "hp"   "drat" "wt"   "qsec" "vs"   "am"   "gear"
## [11] "carb"
```

Exploratory analysis

Let's check how the variables correlate with mpg:

```
##         cyl       disp     hp        drat      wt        qsec    vs
## corr    -0.8522   -0.8476  -0.7762   0.6812    -0.8677   0.4187  0.664
## p-value 6.113e-10 9.38e-10 1.788e-07 1.776e-05 1.294e-10 0.01708 3.416e-05
##         am       gear     carb
## corr    0.5998   0.4803   -0.5509
## p-value 0.000285 0.005401 0.001084
```

In order to find variables that we should adjust for, find out which variables `am` depends on.

The categorial variables:

```
##           cyl     vs     gear      carb
## n.vars    2       2      2         2
## n.cases   32      32     32        32
## statistic 8.741   0.9069 20.94     6.237
## parameter 2       1      2         5
## approx.ok FALSE   TRUE   FALSE     FALSE
## p.value   0.01265 0.3409 2.831e-05 0.2838
## call      NULL    NULL   NULL      NULL
```

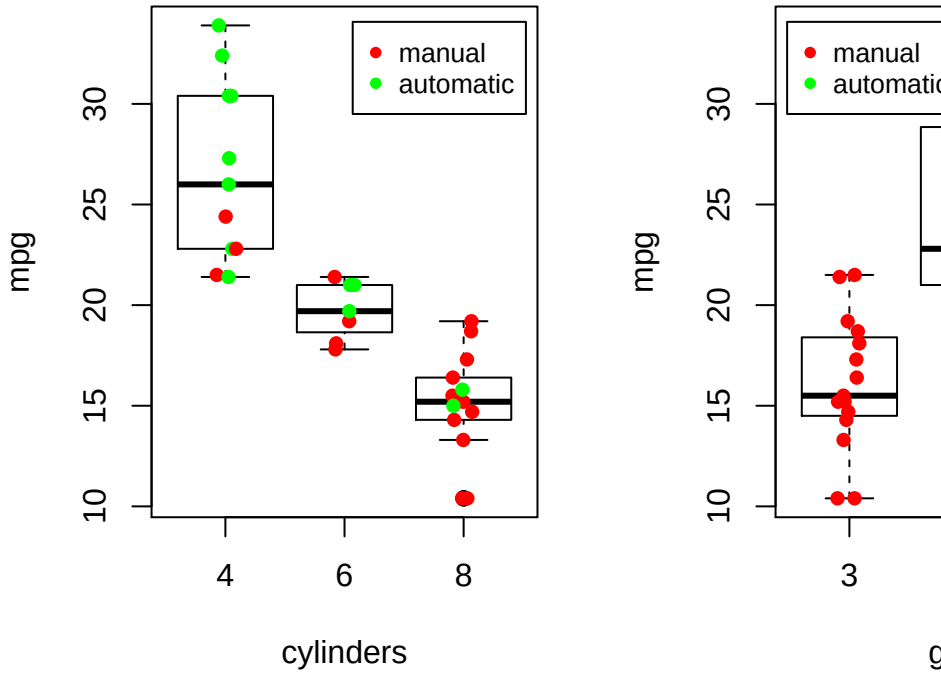
So the variables `cyl` and `gear` differ, depending on `am`.

Influence of the numerical variables:

```
##      disp.p.value hp.p.value drat.p.value wt.p.value qsec.p.value
## [1,] 0.00023      0.221      5.267e-06    6.272e-06  0.2093
```

So the variables`disp`, `drat`, and `wt` differ depending on `am`.

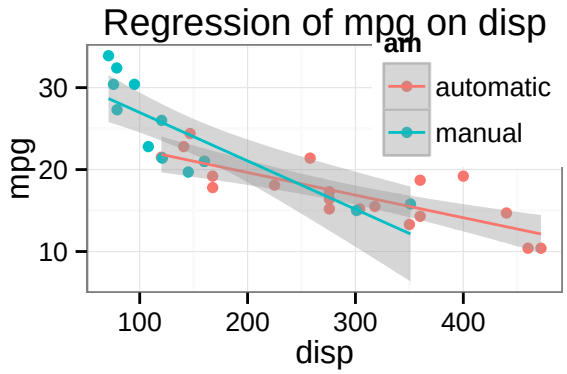
Let's look at some plots to get a feeling of the data distributions.
Since - in theory, and also according to above t-tests - the number of cylinder might have an influence on the relationship between transmission and mpg, plot mpg for 4, 6, and 8 cylinders separately and color the data points

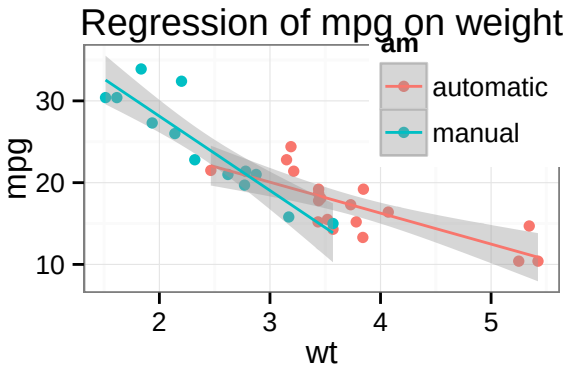


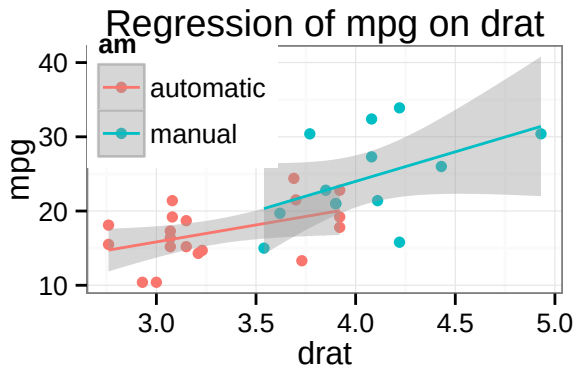
by transmission type. Do the same for `gear`.

Now let's *look* at the relationships between `mpg`, `am` and the numerical variables *disp*, *drat*, and *wt*:

```
## Loading required package: grid
```







Regression results: manual transmission cars have better mpg than automatic transmission cars

Simple linear regression

Above boxplots (see fig. XX) suggested that manual transmission cars have higher mpg values than automatic transmission cars. A simple linear regression of `mpg` on `am` confirms this:

```
## $coefficients
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept)   17.147      1.125  15.247 1.134e-15
## ammanual       7.245      1.764   4.106 2.850e-04
```

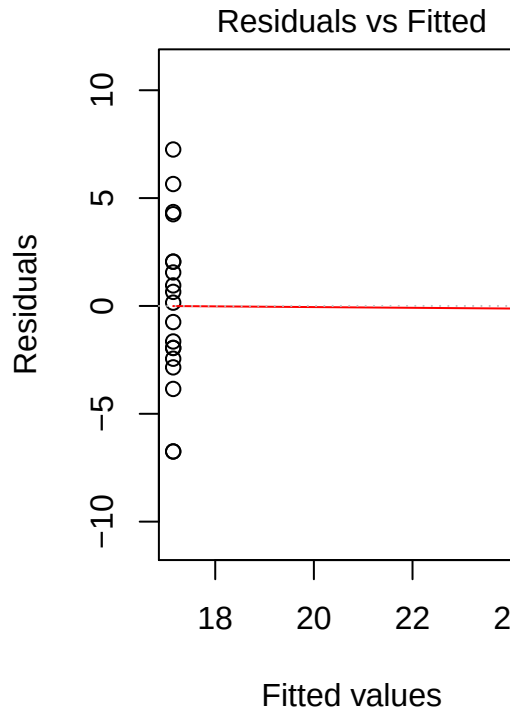
The mean of mpg for the **manual transmission cars is 7.2 miles higher than for the automatical transmission cars**, a significant difference (p=0.0003).

Confidence intervals

With 95% confidence we can estimate that difference between automatic and manual transmission cars is between 3.65 and 10.85 miles per gallon.

```
##                    lower bound upper bound
## mean automatic          14.851       19.44
## manual - automatic       3.642       10.85
```

Diagnostics for the simple regression.



The residuals of the simple regression seem to be random and normally distributed:

No data point has an overly strong influence:

```
range(hatvalues(fit))
```

```
## [1] 0.05263 0.07692
```

Multiple regression: adjusting for weight unveils a complex relationship between transmission and mpg

Car weight is related to both transmission and mpg. Looking at the three plots relating `disp`, `drat`, and `wt` to `mpg` (see fig. XX), the difference between automatic and manual transmission seemed to be the greatest in the case of weight. So let's first adjust for weight:

```
## $coefficients
##             Estimate Std. Error  t value  Pr(>|t|)
## (Intercept) 37.32155     3.0546 12.21799 5.843e-13
## ammanual    -0.02362     1.5456 -0.01528 9.879e-01
## wt          -5.35281     0.7882 -6.79081 1.867e-07
```

This model explains much more variance than the model that includes transmission alone (75% as opposed to 36%).
Adjusting for weight, we see that difference in transmission disappears.

Let's include the interaction term:

```
## $coefficients
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept)   31.416     3.0201  10.402 4.001e-11
## ammanual      14.878     4.2640   3.489 1.621e-03
## wt            -3.786     0.7856  -4.819 4.551e-05
## ammanual:wt   -5.298     1.4447  -3.667 1.017e-03
```

Now we see that there is in fact a difference between automatic and manual transmission: the mpg of both manual and automatic transmission cars drops the heavier the car, but the mpg drop is much steeper for manual than for automatic transmission cars.
Concretely:

- if the weight of an automatic transmission car goes up by 1000lbs, the expected mpg drops by roughly 3.8 miles.
- if the weight of a manual transmission car goes up by 1000lbs, the expected mpg drops only roughly 9.1 miles – more than twice as fast.

But remember: the mean mpg of manual transmission cars is still significantly higher than for automatic transmission cars.

This model with the interaction explains 83% of variance (as opposed to 75% for the model without the interaction term).

Including cylinders: significant, but little change

Based on the relationships established in the exploratory section, `cyl` have an influence on both `mpg` and transmission. Figure XX shows a similar relationship between transmission, cylinders, and mpg as between transmission, weight, and mpg.

Including cylinders to the interaction (`mpg ~ am*wt + cyl`) makes a significant, but small difference: it adds only 5% explained variance. This small additional value of including cylinders is also confirmed by an anova between the models:

```
fit1 <- lm(mpg ~ am, data=cars)
fit2 <- update(fit1, mpg ~ am + wt, data=cars)
fit3 <- update(fit1, mpg ~ am + wt + am*wt, data=cars)
fit4 <- update(fit1, mpg ~ am + wt + am*wt + cyl, data=cars)
anova(fit1, fit2, fit3, fit4)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + wt
## Model 3: mpg ~ am + wt + am:wt
## Model 4: mpg ~ am + wt + cyl + am:wt
```

```
##   Res.Df RSS Df Sum of Sq     F  Pr(>F)
## 1     30 721
## 2     29 278  1       443 83.39 1.4e-09 ***
## 3     28 188  1        90 17.02 0.00034 ***
## 4     26 138  2        50  4.71 0.01794 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Cylinder is of course related with weight, so it makes sense that it doesn't explain much more.

Conclusions

Using linear regression models, we tried to answer whether manual or automatic transmission cars have higher mpg, and how big the difference is.

A single variable linear regression shows that the mpg of manual transmission cars is with 95% confidence between 3.6 and 10.8 miles higher.

This result is qualified by controlling for weight: the difference in between manual and automatic transmission cars is smaller for heavier cars.

Limitations

Not included in this report are the influences of other variables of the dataset. Some exploratory results suggested that they wouldn't change the relationship of transmission and mpg greatly, but there is some room for further research.

Applying regression to variables like number of cylinders violate the assumption of normality, so the results are somewhat questionable. Still, the visual analysis (looking at the plots) corroborates the regression results.