Transmission Type fuel effieciency Analysis

Francois Schonken

26/10/2014

Synopsis

Using the Motor Trends data this report endeavours to find out whether automatic or manual transmission is better for MPG? This report will further more attempt to quantify the MPG difference between automatic and manual transmissions. In an effort to keep the report consise most R code is not show, please feel free to visit this GitHub Repo for the complete R Markdown code.

Data Dictionary

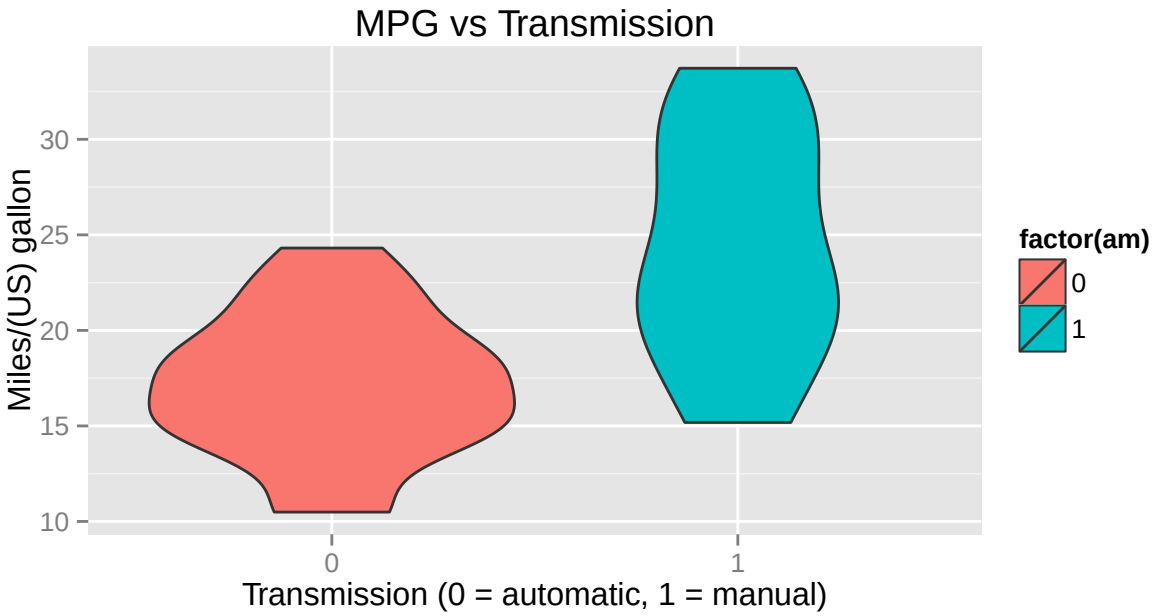
The Motor Trends data consists of 32 observations. For a full description of the data definition please refer to the Motor Trends Data Documentation.

Field	Definition	Field	Definition
mpg	Miles/(US) gallon	cyl	Number of cylinders
disp	Displacement (cu.in.)	hp	Gross horsepower
drat	Rear axle ratio	wt	Weight (lb/1000)
qsec	1/4 mile time in seconds	vs	V/S (0 = V-engine, 1 = straight engine)
am	Transmission (0 = automatic, 1 = manual)	gear	Number of forward gears
carb	Number of carburetors		

Exploring the Motor Trend data

Let's start by getting a feel for the dataset by running `summuary()` on a few of the more interesting fields followed by a violin plot to get an idea of the role transmission plays in MPG.

```
##     cars.mpg        cars.cyl      cars.disp        cars.hp
##  Min.   :10.4   Min.   :4.00   Min.   : 71.1   Min.   : 52.0
##  1st Qu.:15.4   1st Qu.:4.00   1st Qu.:120.8   1st Qu.: 96.5
##  Median :19.2   Median :6.00   Median :196.3   Median :123.0
##  Mean   :20.1   Mean   :6.19   Mean   :230.7   Mean   :146.7
##  3rd Qu.:22.8   3rd Qu.:8.00   3rd Qu.:326.0   3rd Qu.:180.0
##  Max.   :33.9   Max.   :8.00   Max.   :472.0   Max.   :335.0
##    cars.drat       cars.wt       cars.qsec      cars.carb
##  Min.   :2.76   Min.   :1.51   Min.   :14.5   Min.   :1.00
##  1st Qu.:3.08   1st Qu.:2.58   1st Qu.:16.9   1st Qu.:2.00
##  Median :3.69   Median :3.33   Median :17.7   Median :2.00
##  Mean   :3.60   Mean   :3.22   Mean   :17.8   Mean   :2.81
##  3rd Qu.:3.92   3rd Qu.:3.61   3rd Qu.:18.9   3rd Qu.:4.00
##  Max.   :4.93   Max.   :5.42   Max.   :22.9   Max.   :8.00
```



We find that the mean MPG for our Automatic transmisions is 17.1474 and 24.3923 for our Manual transmissions giving us a 7.2449 mean MPG difference with the Automatic transmission being less effiecient. This extremely simple comparison is just the beginning of our journey.

Regression Modeling

Next we iteratively run linear regressions starting with `summary(lm(mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear + carb, data = cars))$coef`, in each subsequent run we remove the entry (field) with the largest `Pr(>|t|)` score until we only have one left, namely `wt`. Next we use the `anova()` function to find the cut-off point in terms regression variables. We find that including the `gear` is not beneficial to our regression model and when it is not specified the Transmission type `am` scrapes in just under 5%.

Field	Pr(>F) for (mpg ~ wt, qsec, am)	Pr(>F) for (mpg ~ wt, qsec, am, gear)
wt	NA	NA
qsec	9.2862 × 10-4	0.0011
am	0.0467	0.0508
gear		0.8897

The fields `wt`, `qsec` and `am` feel like the right set of variables to be looking at but a linear coeficient against `qsec` seems odd, one would expect diminishing returns for fuel consumed as the `qsec` shrunk. To this end I introduce the variable `qsec.2.I` which reads `1/(qsec * qsec)`. The following two tables detail the coeficients of the respective linear regression models. We see that using the variable `qsec.2.I` as opposed to `qsec` yields a tighter fit to our data as evidenced by every single `Pr(>|t|)` almost being halved.

```
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept)    9.618     6.9596   1.382 1.779e-01
## wt            -3.917     0.7112  -5.507 6.953e-06
## qsec           1.226     0.2887   4.247 2.162e-04
## am             2.936     1.4109   2.081 4.672e-02
```

```
##              Estimate Std. Error t value  Pr(>|t|)
## (Intercept)    42.806     2.6020  16.451 6.347e-16
## wt             -3.845     0.6806  -5.649 4.719e-06
## qsec.2.I    -3603.694   767.4360  -4.696 6.370e-05
## am              3.191     1.3612   2.344 2.639e-02
```

Using the variable `qsec` we arive a an model that reads for each `wt` (lb/1000) increase MPG drops by -3.9165. For each additional second in `qsec` (1/4 mile time in seconds) MPG increases by 1.2259 and Manual transmissions yield a 2.9358 MPG benefit over Automatic transmissions.

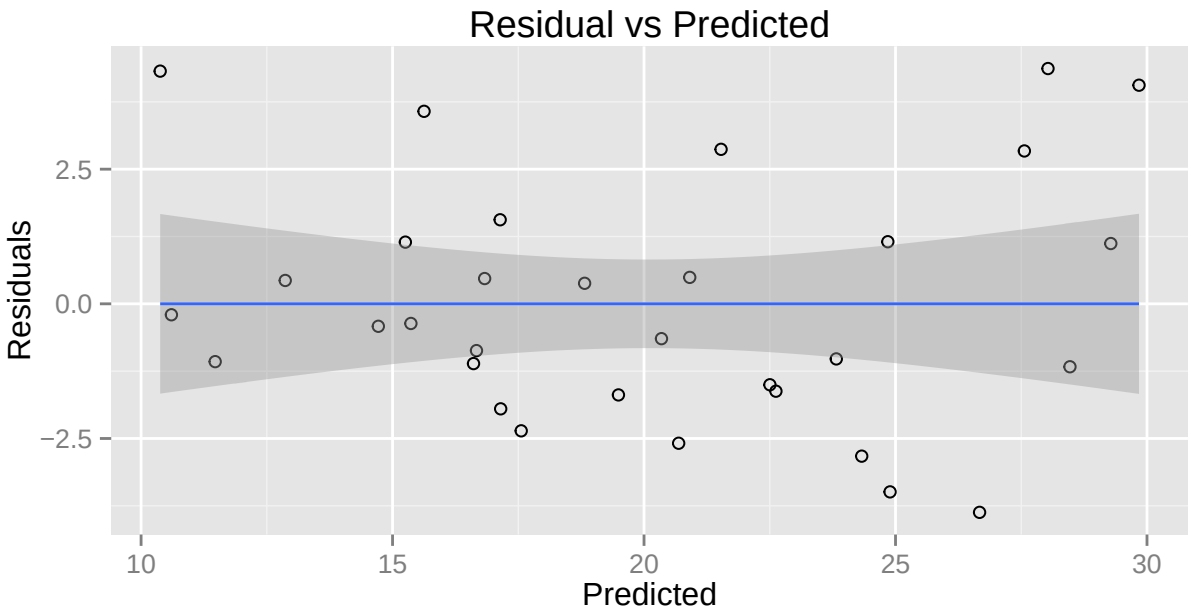
Confidence

The R^2 for the `qsec` and `qsec.2.I` regression models are 0.8497 and 0.8617 respectively. The R^2 for the `qsec.2.I` based regression model is marinally stronger (by 0.0121) and with this model explaining 0.8617 of the variability this is a highly predictive model.

Residuals

Lastly we will review the Residuals vs. Predicted of our regression using the `qsec.2.I` field.

```
fit <- lm(mpg ~ wt + qsec.2.I + am, data = cars)
```



The Residuals vs. Predicted look good for this model and I believe we can in good concience use it.

Conclusion

Using our `qsec.2.I` based regression model we find that Manual transmissions yeild better MPG in the amount of 3.1909 MPG.