Regression Models Course Project

martinvanelp

26 juli 2015

Transmission and MPG

Executive summary

The analysis of the available data leads to the conclusion that the kind of transmission does *not* significantly influence the miles per gallon use of a car, when correcting for other factors. Cars with manual transmission do use on average 7 MPG more than cars with automatic transmission, but this is primarily determined by the corresponding weight and gross horsepower of these cars.

Introduction

This paper explores the relationship between the kind of transmission of a car and its miles per gallon (MPG) use. For this we use our 1974 data collection (see Appendix for details).

Two questions are answered: * Is an automatic or manual transmission better for MPG? * Quantify the MPG difference between automatic and manual transmissions.

Is an automatic or manual transmission better for MPG?

To determine the influence of the transmission on the MPG of a car the first hurdle is to arrive at a model of other factors that influence this, like the weight of the car. Therefore the correlation matrix of the available data was investigated (see Appendix).

Step by step variables in the data were dropped to arrive at the most parsimonious model to answer the question. Investigation of the correlation matrix led to the conclusion that the V/S, 1/4 mile time and the number of forward gears seemed least relevant. These variables were therefore dropped from the model first. After that the carburetors were dropped because of the low significance and counter-intuitive alternating signs. In the resulting model the rear axle ratio and displacement were the most insignificant and therefore dropped. Then the number of cylinders were dropped as most insignificant to arrive at a model with the kind of transmission as a factor variable, and weight and gross horsepower as *significant* confounding variables. Lastly, the model with just the transmission as a factor variable was considered to explain the MPG.

The ANOVA test of these models confirms, with just a low P-value for the 2nd model, that weight and horsepower should be included an none of the other variables.

ANOVA

```
## Analysis of Variance Table
##
## Model 1: mpg ~ factor(am)
## Model 2: mpg ~ factor(am) + wt + hp
## Model 3: mpg ~ factor(am) + wt + hp + factor(cyl)
## Model 4: mpg ~ factor(am) + wt + drat + hp + disp + factor(cyl)
## Model 5: mpg ~ factor(am) + factor(carb) + wt + drat + hp + disp + factor(cyl)
## Model 6: mpg ~ factor(am) + factor(carb) + factor(gear) + vs + qsec +
##     wt + drat + hp + disp + factor(cyl)
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)
## 1     30 720.90
## 2     28 180.29  2    540.61 33.6748 2.839e-06 ***
## 3     26 151.03  2     29.27  1.8230    0.1956
## 4     24 150.10  2      0.93  0.0576    0.9442
## 5     19 132.74  5     17.36  0.4326    0.8189
## 6     15 120.40  4     12.34  0.3842    0.8165
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

However, when we look at the residual plots of the respective models, and a model with just weight and horsepower, it seems that information on the kind of transmission (like the other dropped variables) adds little explanatory value (see Appendix). The residuals hardly move compared with the most parsimonious model with just weight and horsepower.

The summary statistics of the model including transmission, weight and horsepower are shown beneath. That leads to the conclusion that the sign of the transmission factor variable is insignificant (P-value too high). These observations mean that the kind of transmission does not significantly influence MPG, when based on the analysis of this data collection.

Model with transmission, weight and horsepower

```
##
## Call:
## lm(formula = mpg ~ factor(am) + wt + hp, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4221 -1.7924 -0.3788  1.2249  5.5317
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 34.002875   2.642659  12.867 2.82e-13 ***
## factor(am)1  2.083710   1.376420   1.514 0.141268
## wt          -2.878575   0.904971  -3.181 0.003574 **
## hp          -0.037479   0.009605  -3.902 0.000546 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.538 on 28 degrees of freedom
## Multiple R-squared:  0.8399, Adjusted R-squared:  0.8227
## F-statistic: 48.96 on 3 and 28 DF,  p-value: 2.908e-11
```

Quantification of the difference between automatic and manual transmission

Although the previous answer concludes that the kind of transmission does not influence MPG significantly when correcting for other factors, the question could still remain what the difference is in the given data collection, disregarding other features of the cars.

Disregarding the other features of the cars gives a significant outcome (low P-value in the summary beneath). The cars with manual transmission in this sample have on average a 7 MPG higher use (17 + 7 MPG) than cars in the sample with an automatic transmission (17 MPG). This is the result of other characteristics of these cars: primarily the corresponding weight and gross horsepower of these cars.

Model with just transmission

```
##
## Call:
## lm(formula = mpg ~ factor(am), data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.3923 -3.0923 -0.2974  3.2439  9.5077
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   17.147      1.125  15.247 1.13e-15 ***
## factor(am)1    7.245      1.764   4.106 0.000285 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.902 on 30 degrees of freedom
## Multiple R-squared:  0.3598, Adjusted R-squared:  0.3385
## F-statistic: 16.86 on 1 and 30 DF,  p-value: 0.000285
```

Appendix

The data was extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973-74 models). The data contains 11 variables:

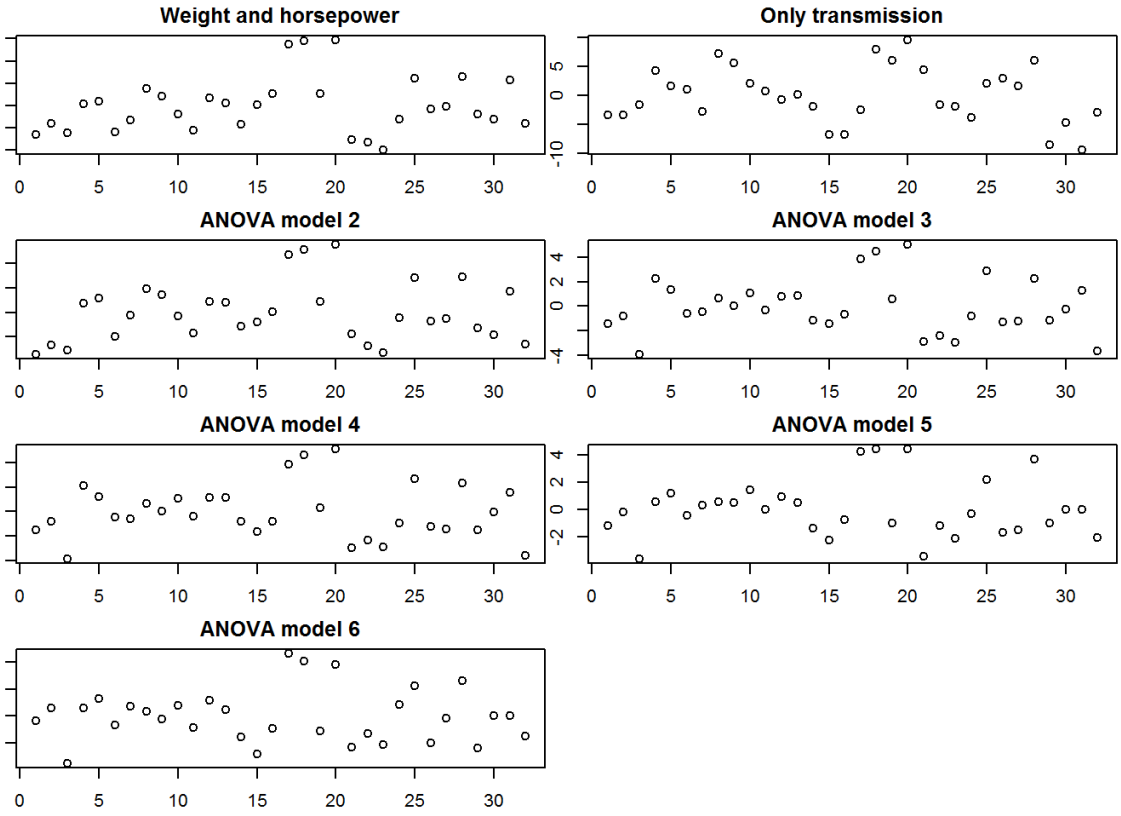
1. mpg - Miles/(US) gallon
2. cyl - Number of cylinders
3. disp - Displacement (cu.in.)
4. hp - Gross horsepower
5. drat - Rear axle ratio
6. wt - Weight (lb/1000)
7. qsec - 1/4 mile time
8. vs - V/S
9. am - Transmission (0 = automatic, 1 = manual)
10. gear - Number of forward gears
11. carb - Number of carburetors

Beneath an example of the data and the correlation matrix is given.

```
##                   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
```

```
##             mpg        cyl       disp         hp        drat         wt
## mpg   1.0000000 -0.8521620 -0.8475514 -0.7761684  0.68117191 -0.8676594
## cyl  -0.8521620  1.0000000  0.9020329  0.8324475 -0.69993811  0.7824958
## disp -0.8475514  0.9020329  1.0000000  0.7909486 -0.71021393  0.8879799
## hp   -0.7761684  0.8324475  0.7909486  1.0000000 -0.44875912  0.6587479
## drat  0.6811719 -0.6999381 -0.7102139 -0.4487591  1.00000000 -0.7124406
## wt   -0.8676594  0.7824958  0.8879799  0.6587479 -0.71244065  1.0000000
## qsec  0.4186840 -0.5912421 -0.4336979 -0.7082234  0.09120476 -0.1747159
## vs    0.6640389 -0.8108118 -0.7104159 -0.7230967  0.44027846 -0.5549157
## am    0.5998324 -0.5226070 -0.5912270 -0.2432043  0.71271113 -0.6924953
## gear  0.4802848 -0.4926866 -0.5555692 -0.1257043  0.69961013 -0.5832870
## carb -0.5509251  0.5269883  0.3949769  0.7498125 -0.09078980  0.4276059
##             qsec         vs          am       gear        carb
## mpg   0.41868403  0.6640389  0.59983243  0.4802848 -0.55092507
## cyl  -0.59124207 -0.8108118 -0.52260705 -0.4926866  0.52698829
## disp -0.43369788 -0.7104159 -0.59122704 -0.5555692  0.39497686
## hp   -0.70822339 -0.7230967 -0.24320426 -0.1257043  0.74981247
## drat  0.09120476  0.4402785  0.71271113  0.6996101 -0.09078980
## wt   -0.17471588 -0.5549157 -0.69249526 -0.5832870  0.42760594
## qsec  1.00000000  0.7445354 -0.22986086 -0.2126822 -0.65624923
## vs    0.74453544  1.0000000  0.16834512  0.2060233 -0.56960714
## am   -0.22986086  0.1683451  1.00000000  0.7940588  0.05753435
## gear -0.21268223  0.2060233  0.79405876  1.0000000  0.27407284
## carb -0.65624923 -0.5696071  0.05753435  0.2740728  1.00000000
```

Residual plots of the different models are as beneath.


Weight and horsepower
Only transmission
ANOVA model 2
ANOVA model 3
ANOVA model 4
ANOVA model 5
ANOVA model 6