Motor Trend Cars Regression Model

Maurício Collaça

Nov 22, 2017

Executive Summary

Motor Trend magazine is about the automobile industry and they are interested in answer the following two questions:

- "Is an automatic or manual transmission better for MPG"
- "Quantify the MPG difference between automatic and manual transmissions"

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).

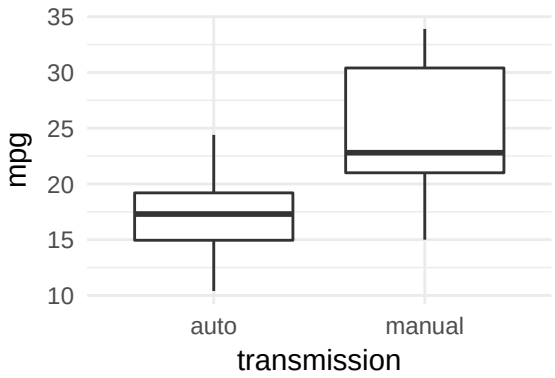
The following sections details the process and conclusion that the manual transmission is better for gas milleage and quantify the difference.

Exploratory Data Analysis

mtcars is a data frame with 32 observations on 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 (1000 lbs)
[, 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

The following boxplot suggests that manual transmission is better for miles per gallon.



One can also see that the higher average also implies in much higher variance.

Regression Model

The strategy for model selection is to perform an analysis of variance (ANOVA) of nested models starting from the base model `mpg ~ am` then adding regressors in the descending order of their correlation with `mpg`.

```
   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
```

The following model ANOVA suggests the first three models are adequate because they return p-values less than 0.05.

```
Analysis of Variance Table

Model  1: mpg ~ am
Model  2: mpg ~ am + wt
Model  3: mpg ~ am + wt + cyl
Model  4: mpg ~ am + wt + cyl + disp
Model  5: mpg ~ am + wt + cyl + disp + hp
Model  6: mpg ~ am + wt + cyl + disp + hp + drat
Model  7: mpg ~ am + wt + cyl + disp + hp + drat + vs
Model  8: mpg ~ am + wt + cyl + disp + hp + drat + vs + carb
Model  9: mpg ~ am + wt + cyl + disp + hp + drat + vs + carb + gear
Model 10: mpg ~ am + wt + cyl + disp + hp + drat + vs + carb + gear + qsec
   Res.Df    RSS Df Sum of Sq       F    Pr(>F)
1      30 720.90
2      29 278.32  1    442.58 55.1371 2.129e-06 ***
3      27 182.97  2     95.35  5.9395   0.01259 *
4      26 182.87  1      0.10  0.0123   0.91304
5      25 150.41  1     32.46  4.0440   0.06266 .
6      24 150.10  1      0.31  0.0384   0.84726
7      23 142.66  1      7.45  0.9275   0.35078
8      18 125.62  5     17.04  0.4245   0.82451
9      16 121.64  2      3.97  0.2475   0.78385
10     15 120.40  1      1.24  0.1546   0.69967
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Inspecting coefficients of the first 3 models, one can see the slope sign inversion. Also, the first model produces more significant, i.e. lower p-values for β_0 and β_1 coefficients, suggesting the additional regressors aren't good.

```
             Estimate Std. Error   t value     Pr(>|t|)
(Intercept) 17.147368   1.124603 15.247492 1.133983e-15
ammanual     7.244939   1.764422  4.106127 2.850207e-04
```

```
               Estimate Std. Error     t value     Pr(>|t|)
(Intercept) 37.32155131  3.0546385 12.21799285 5.843477e-13
ammanual    -0.02361522  1.5456453 -0.01527855 9.879146e-01
wt          -5.35281145  0.7882438 -6.79080719 1.867415e-07
```

```
              Estimate Std. Error    t value     Pr(>|t|)
(Intercept) 33.7535920  2.8134831 11.9970836 2.495549e-12
ammanual     0.1501031  1.3002231  0.1154441 9.089474e-01
wt          -3.1495978  0.9080495 -3.4685309 1.770987e-03
cyl6        -4.2573185  1.4112394 -3.0167231 5.514697e-03
cyl8        -6.0791189  1.6837131 -3.6105432 1.227964e-03
```

Examining the residual plots in the Appendix, one can confirm that the best model that explains the relatioship between the miles per galon and transmission is the model 1 because of is horizontal line in the Residuals vs Fitted plot and the diagonal fit in the Normal Q-Q plot.

The confidence intervals for the intercept and slope of the model 1 are:

```
               2.5 %   97.5 %
(Intercept) 14.85062 19.44411
ammanual     3.64151 10.84837
```

Conclusions

Interpreting the model 1 coefficients one can see that the manual transmission is better for the gas milleage by increasing 7.24 miles per gallon:

- The intercept coefficient tells that the average miles per galon for automatic transmission is 17.15, with a high significance of its p-value $1.1339835 \times 10^{-15}$ being less than the α level of 0.05.
- The slope coefficient `ammanual` tells that the increase in the average miles per galon for manual transmission is 7.24 with a high significance of its pvalue 2.8502074×10^{-4} being less than the α level of 0.05.

Appendix

Residual plots for models 1, 2 and 3