Regression Models Course Project

Nadia Stavisky

1 September 2019

Introduction

The Motor Trend, a magazine about the automobile industry. Looking at a data set of a collection of cars, they are interested in exploring the relationship between a set of variables and miles per gallon (MPG) (outcome). They are particularly interested in the following two questions:

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

Motor Trend Car Road Tests - 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).

Source

Henderson and Velleman (1981), Building multiple regression models interactively. Biometrics, 37, 391-411. A data frame with 32 observations on 11 (numeric) variables.

Exploratory data analysis

We are interested in learning relationship beween mpg and other variables.

Is an automatic or manual transmission better for MPG?

We will perform t-tests to compare two groups (car mpg performance for automatic and manual transmission type). ###MPG summary stats by transmission type (see Appendix 2).

- mpg summary for cars with automatic transmission:

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10.40   14.95   17.30   17.15   19.20   24.40
```

- mpg summary for cars with manual transmission:

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   15.00   21.00   22.80   24.39   30.40   33.90
```

Welch Two Sample t-test on if the automatic and manual averages differ, we are searching for any differences between the means rather than if one is specifically less than or greater than the other.

```
##
##  Welch Two Sample t-test
##
## data:  mpg by am
## 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 in group automatic    mean in group manual
##                17.14737                24.39231
```

The results of the t-test show p-value < 0.01 supporting the alternative hypothesis that "true difference in means is not equal to 0".

MPG difference between automatic and manual transmissions.

To quantify the MPG difference between automatic and manual transmissions let's explore the relationship between a set of variables and miles per gallon (MPG) (outcome):

Consider the correlation matrix for the quantitative predictor variables and response variable (Correlation matrix can be seen at Appendix1.):

```
##             mpg          am
## mpg   1.0000000  0.59983243
## cyl  -0.8521620 -0.52260705
## disp -0.8475514 -0.59122704
## hp   -0.7761684 -0.24320426
## drat  0.6811719  0.71271113
## wt   -0.8676594 -0.69249526
## qsec  0.4186840 -0.22986086
## vs    0.6640389  0.16834512
## am    0.5998324  1.00000000
## gear  0.4802848  0.79405876
## carb -0.5509251  0.05753435
```

From the correlation matrix results we can see that the quantitative variables such as "hp" and "wt" are most correlated (negatively) with "mpg", also there is strong correlation between transmission type and some other variables such as cyl, disp, drat, wt, gear.

We will perform variable sellection procedure by choosing best model by AIC in a Stepwise Algorithm: We will use stepwise regression

```
##
## Call:
## lm(formula = mpg ~ cyl + hp + wt + am, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9387 -1.2560 -0.4013  1.1253  5.0513
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 33.70832    2.60489  12.940 7.73e-13 ***
## cyl6        -3.03134    1.40728  -2.154  0.04068 *
## cyl8        -2.16368    2.28425  -0.947  0.35225
## hp          -0.03211    0.01369  -2.345  0.02693 *
## wt          -2.49683    0.88559  -2.819  0.00908 **
## ammanual     1.80921    1.39630   1.296  0.20646
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.41 on 26 degrees of freedom
## Multiple R-squared:  0.8659, Adjusted R-squared:  0.8401
## F-statistic: 33.57 on 5 and 26 DF,  p-value: 1.506e-10
```

Model summary that was selected by AIC in a Stepwise Algorith variables gives us next performance mettrics:
1. Residual standard error: tells us that the actual mpg deviate from the true regression line described in the model by approximately 2.4101196 Miles/gallon on average. the RSE decreased that proof the model

improvement. In the mtcars data set, the mean value of mpg is approximately 20.090625 units, and so the percentage error is 11.9962402% 2. Multiple R-squared and Adjusted R-squared tells us proportion of variation in the mpg has been explained by this mode. The R-squared decreased due to the number of variables decresed. Adjusted R-squared increased the model with the largest adjusted R2 will have only correct variables and no noise variables. 3. F-statistic: $33.57 > 1$, provides evidence against the null hypothesis H0 and suggests that at least one of the variables must be related to mpg.
4. p-value: 1.506e-10 p-value associated with the F-statistic is much smaller then 0.05, we can infer that there is an association between the predictor and the response. We can reject the null hypothesis-that stands on there is no relationship to exist between X and Y.

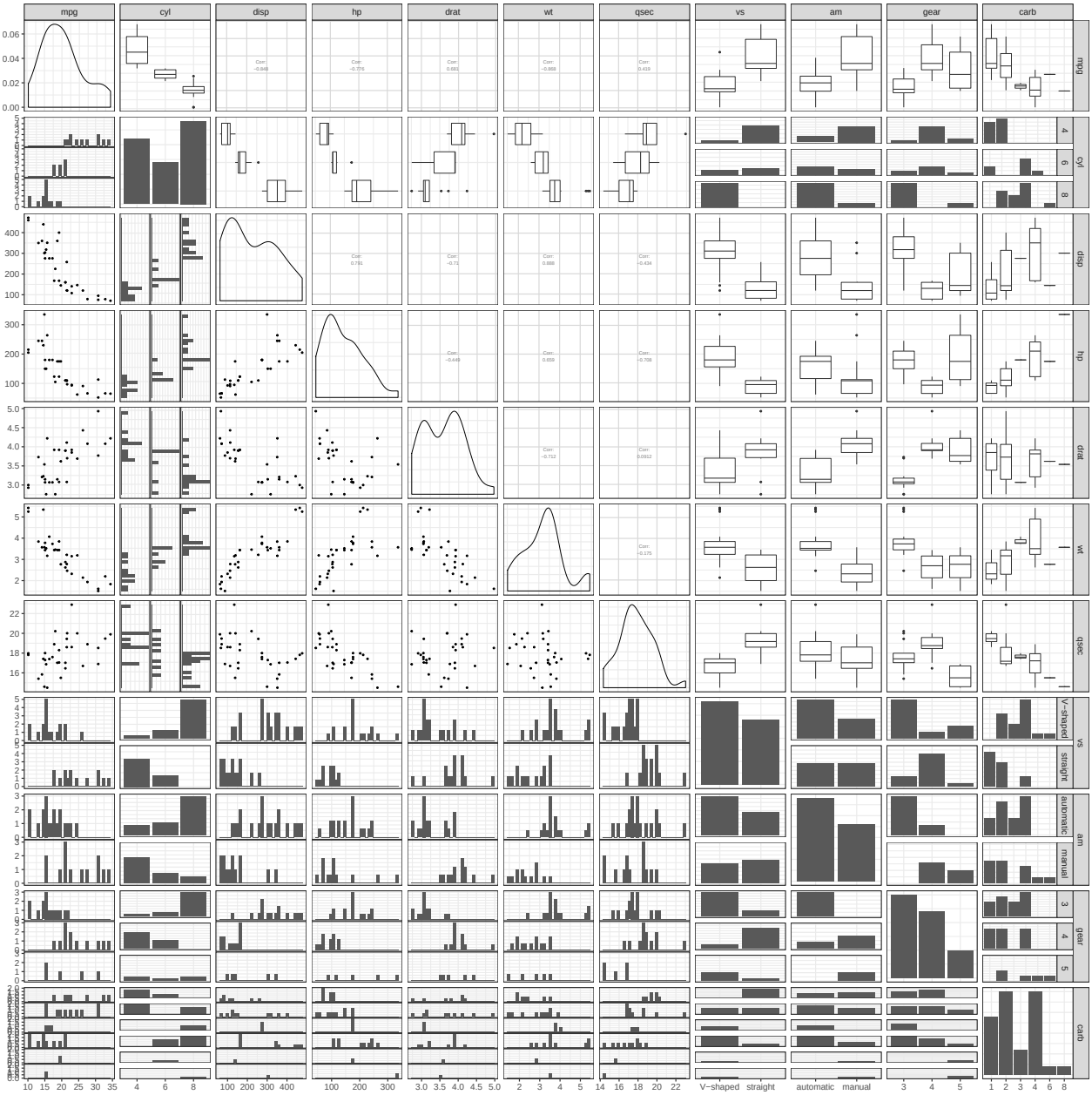
Diagnostic plots (Appendix 3)

1. Residuals vs Fitted - does not show any pattern, linear model explained relationship between predictors (Number of cylinders, Gross horsepower, Weight (1000 lbs), Transmission) and respond (Miles/(US) gallon);
2. Normal Q-Q - residuals quite well follow the line, while we can see some observations off (17, 18, 20).
3. Scale-Location - almost all residuals spread equaly along the range of predictors, no signs of homoscedasticity.
4. Residuals vs Leverage - there is no influential or cases, all cases are inside of the Cook's distance lines.

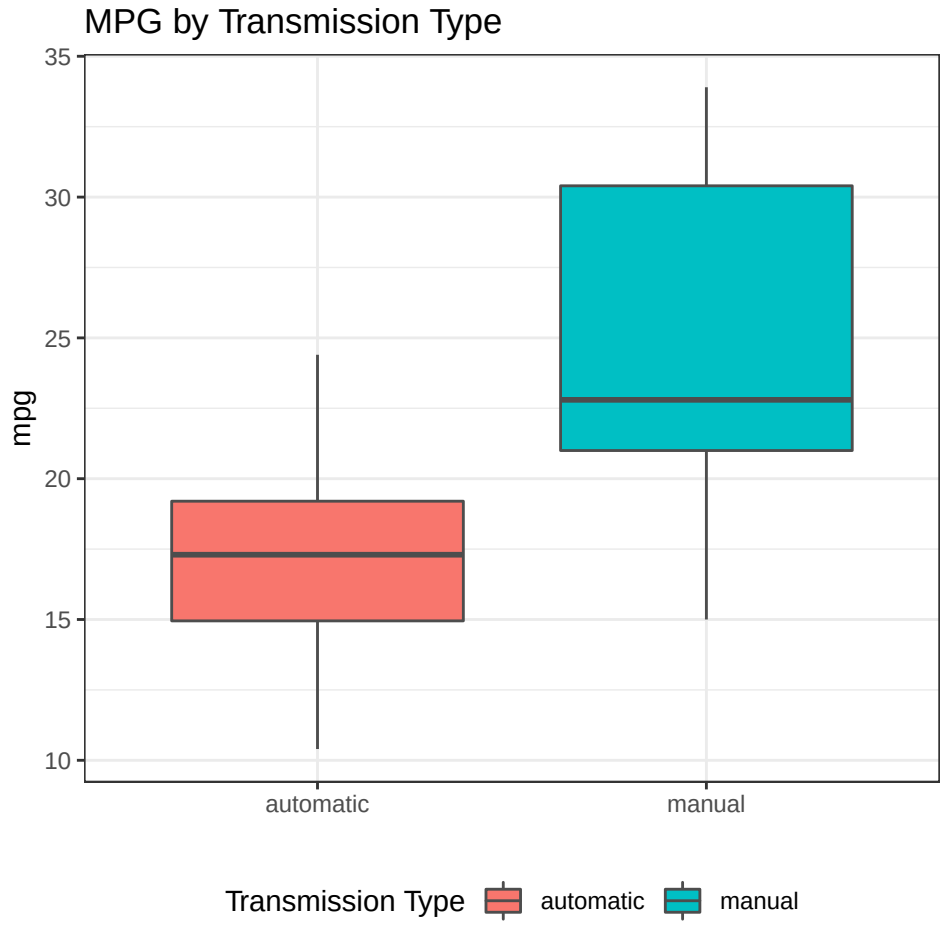
Conclusion

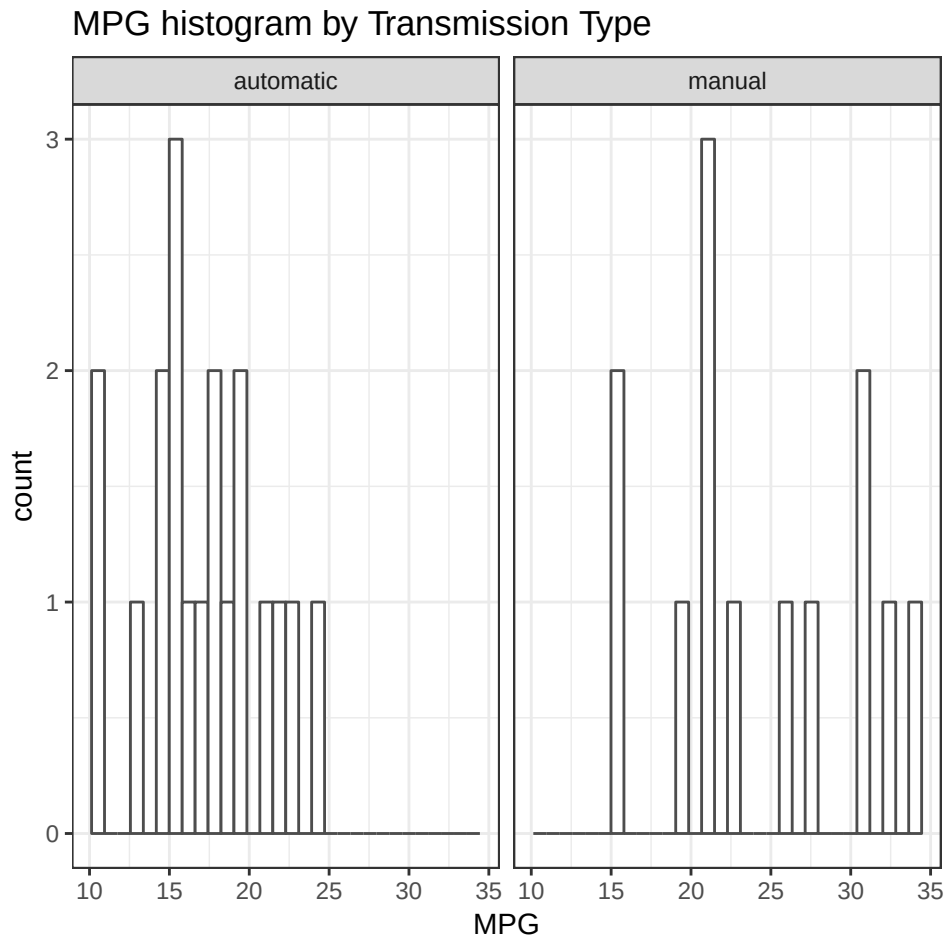
Out of all variables the best Multiple linear Regression was fitted that describes relationship between predictors (Number of cylinders, Gross horsepower, Weight (1000 lbs), Transmission) and respond (Miles/(US). Model performance and residual diagnostics confirmed a good fit. Model estimates that on average, manual transmission is better than automatic transmission by 1.8 mpg. However, transmission type is not the only factor accounting for MPG, cylinders, horsepower, and weitght are the important factors in affecting the MPG.

Appendix 1. Exploratory data analysis. Correlation between mtcars variables.



Appendix 2. Exploratory data analysis. Correlation between mpg and am variables.




MPG histogram by Transmission Type
automatic
manual
count
0
1
2
3
10
15
20
25
30
35
MPG

Appendix 3. Diagnostic plots.