Motor Trend magazine - Data analysis of influence on MPG for Automatic vs. Manual Transmission.

by jmvilaverde

Thursday, June 18, 2015

Executive summary

With the linear regression model evaluated have concluded that, with 95% confidence, is estimated that a manual transmission results in a 0.0457303 to 5.8259441 increase in MPG comparing to use an automatic transmission.

1.Initial Exploratory Data Analysis

Structure from ?mtcars and values for factors. *(View Figure 1.Mtcars pairs comparation.)*

Format: A data frame with 32 observations on 11 variables.

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

Correlation between mpg and am is 0.5998324. (Closer to -1 or 1 is stronger relationship, when is 0 implies no linear relationship).

2.Model proposal and analysis

2.1.Analysis of Model Initial. *(View Figure C1.Figure Summary Model Initial.)*

$$mpg_i = \beta_0 + \beta_{am}am_i$$

- P-value intercept and coefficients: 1.133983e-15, 2.850207e-04 < 0.05 are good p-values.
- P-value Model: 2.850207e-04. < 0.05 is a good p-value for the model.
- R^2 value: 0.3597989. **This low value indicates that model Initial only fits 36% of the data.**

2.2.Analysis of model Complete. *(View Figure C2.Figure Summary Model Complete.)*

$$mpg = \beta_{cyl}cyl + \beta_{dips}disp + \beta_{hp}hp + \beta_{drat}drat + \beta_{wt}wt + \beta_{qsec}qsec + \beta_{vs}vs + \beta_{am}am + \beta_{gear}gear + \beta_{carb}carb$$

- P-value intercept and coefficients: 5.181244e-01, 9.160874e-01, 4.634887e-01, 3.349553e-01, 6.352779e-01, 6.325215e-02, 2.739413e-01, 8.814235e-01, 2.339897e-01, 6.652064e-01, 8.121787e-01 > 0.05 are bad p-values.
- P-value Model: 5.03445e-10. < 0.05 is a good p-value for the model.
- R^2 value: 0.8690158. This high value indicates that the model Complete is a good fit to the data.

The p-value for intercept and coefficients indicates that model Complete is not a good model for fit the data.

2.3.Propose and analysis of alternative model using step R function. *(View Figure C3.Figure Summary Model Step.)*

step(lm(mpg ~ ., data=mtcars)): lm, mpg ~ wt + qsec + am, mtcars

$$mpg = \beta_{wt}wt + \beta_{qsec}qsec + \beta_{am}am$$

- P-value intercept and coefficients: 1.779152e-01, 6.952711e-06, 2.161737e-04, 4.671551e-02 < 0.05 for coefficients are good p-value for the model.
- p-value Model: 2.038468e-12. < 0.05 is a good p-value for the model.
- R^2: 0.8496636. This high value indicates that the model Step is a good fit to the data.(1 means 100% fit).

This model have a p-value < 0.05 for model and coefficients, and a R^2 value near 85%. This model is accepted.

(View Figure 2. Model Step plot.)

(View Figure C4. Anova comparation between the models.)

Final Analysis based on information from Model Step.

Is an automatic or manual transmission better for MPG?

The manual transmission mean is higher than automatic transmission mean for Model Step, due to this analysis is determined that manual transmission is better for MPG.

Quantify the MPG difference between automatic and manual transmissions.

With 95% confidence, we estimate that a manual transmission results in a 0.0457303 to 5.8259441 increase in MPG comparing to use of automatic transmission for a car with the same weight (wt) and 1/4 mile time (qsec).

Apendix. (Note: Summary and Anova are included as Figures).
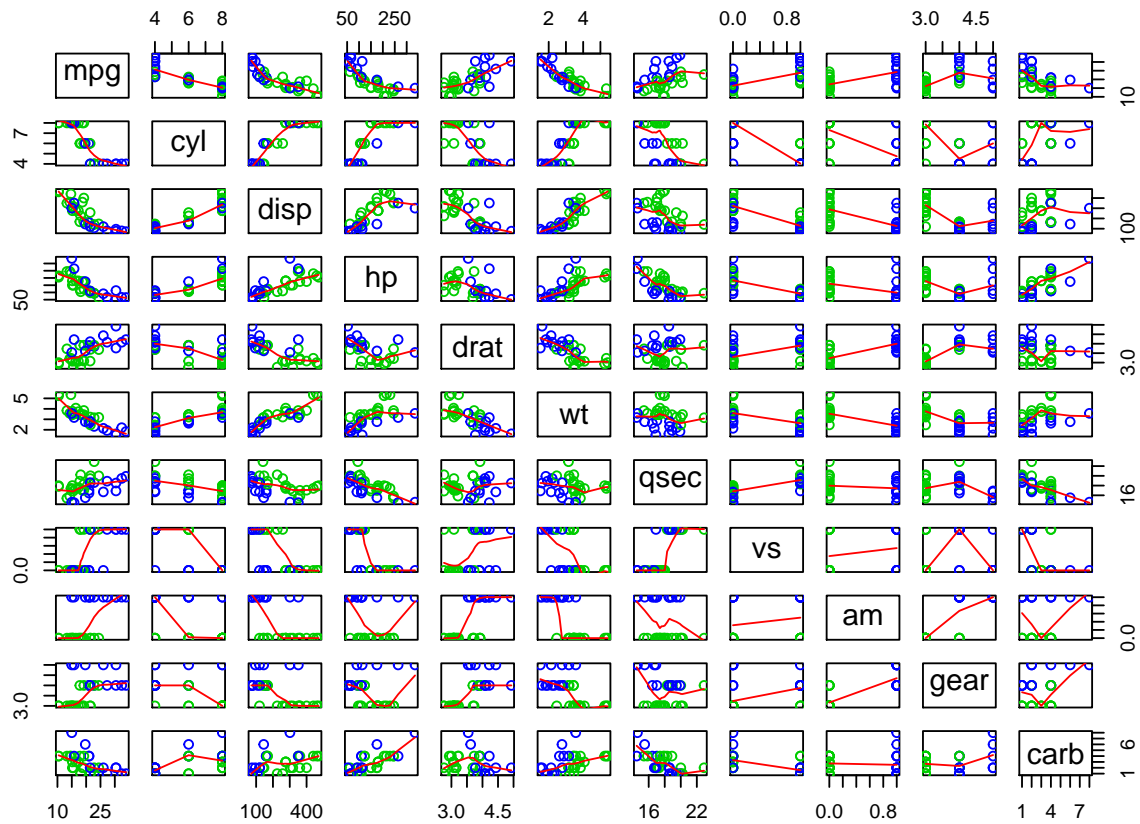
Figure 1. Mtcars pairs comparation.

Figure 2.Model Step Plot.
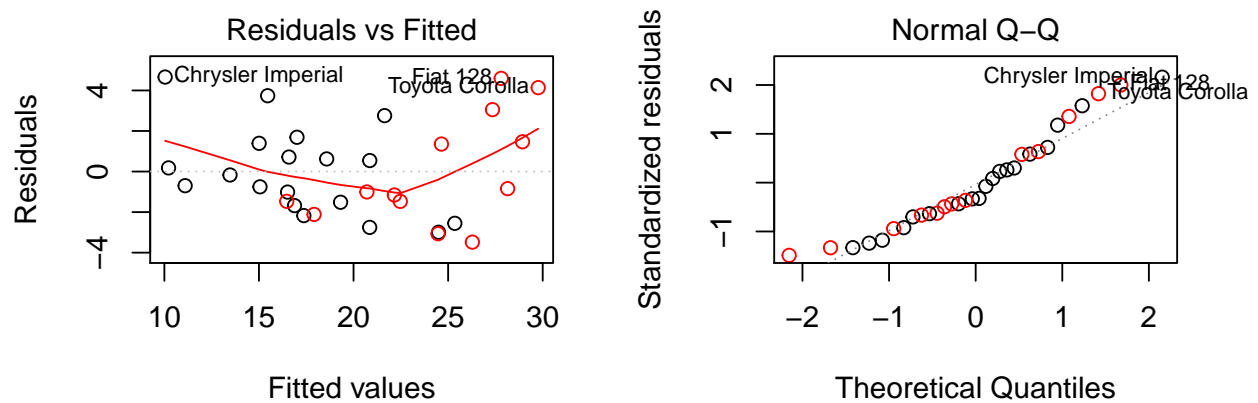
Figure 2.Model Step Plot.
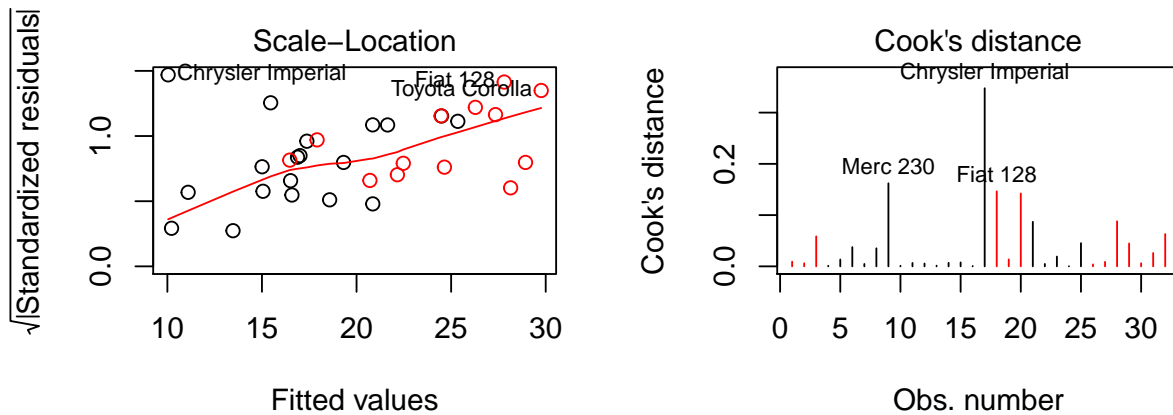

Figure C1.Figure Summary Model Initial.

```
##
## Call:
## lm(formula = mpg ~ 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 ***
## am             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
```

Figure C2.Figure Summary Model Complete.

```
##
## Call:
## lm(formula = mpg ~ ., data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4506 -1.6044 -0.1196  1.2193  4.6271
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.30337   18.71788   0.657   0.5181
## cyl         -0.11144    1.04502  -0.107   0.9161
```

```
## disp             0.01334      0.01786    0.747    0.4635
## hp              -0.02148      0.02177   -0.987    0.3350
## drat             0.78711      1.63537    0.481    0.6353
## wt              -3.71530      1.89441   -1.961    0.0633 .
## qsec             0.82104      0.73084    1.123    0.2739
## vs               0.31776      2.10451    0.151    0.8814
## am               2.52023      2.05665    1.225    0.2340
## gear             0.65541      1.49326    0.439    0.6652
## carb            -0.19942      0.82875   -0.241    0.8122
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.65 on 21 degrees of freedom
## Multiple R-squared:  0.869,  Adjusted R-squared:  0.8066
## F-statistic: 13.93 on 10 and 21 DF,  p-value: 3.793e-07
```

Figure C3.Figure Summary Model Step.

```
##
## Call:
## lm(formula = mpg ~ wt + qsec + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4811 -1.5555 -0.7257  1.4110  4.6610
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.6178     6.9596   1.382 0.177915
## wt           -3.9165     0.7112  -5.507 6.95e-06 ***
## qsec          1.2259     0.2887   4.247 0.000216 ***
## am            2.9358     1.4109   2.081 0.046716 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.459 on 28 degrees of freedom
## Multiple R-squared:  0.8497, Adjusted R-squared:  0.8336
## F-statistic: 52.75 on 3 and 28 DF,  p-value: 1.21e-11
```

Figure C4. Anova comparation between the models.

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ wt + qsec + am
## Model 3: mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear + carb
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)
## 1     30 720.90
## 2     28 169.29  2    551.61 39.2687 8.025e-08 ***
## 3     21 147.49  7     21.79  0.4432    0.8636
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```