Coursera: Linear Regression Project - mtcars

Cliff Weaver

October 18, 2015

Executive Summary

The mtcars dataset is analyzed to evaluate the relationship between a set of variables and miles per gallon (MPG). The data was extracted from the 1974 Motor Trend Magazine providing the MPG and other automobile variables for 32 automobiles (1973–74 models). Regression models and exploratory data analyses to mainly explore how automatic and manual transmissions features affect MPG. The analysis supports the conclusion that lighter cars with a manual transmission offer better fuel economy as measured by MPG when compared to heavier cars with an automatic transmissions. #Problem Statement Motor Trend Magazine is 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

The Data

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). A table with information about each variable can be found in the appendix (*Figure 1 - Data Definitions*). Here is a peek of the data after some variables have been transformed to factors.

```
##                    mpg    wt cyl  hp  qsec        am       vs
## Mazda RX4         21.0 2.620   6 110 16.46    Manual       V8
## Mazda RX4 Wag     21.0 2.875   6 110 17.02    Manual       V8
## Datsun 710        22.8 2.320   4  93 18.61    Manual Straight
## Hornet 4 Drive    21.4 3.215   6 110 19.44 Automatic Straight
## Hornet Sportabout 18.7 3.440   8 175 17.02 Automatic       V8
## Valiant           18.1 3.460   6 105 20.22 Automatic Straight
```

Exploratory Data Analysis

Plotting the relationship between all the variables of the mtcars dataset we observe the variables cyl, disp, hp, drat, wt, vs and am have a correlation with mpg (*Figure 2 - Pairs Plot*). Additionally, referencing the boxplot in *Figure 3 - Boxplot Transmission v MPG*, evidence suggests there is a relatively large difference between the mpg mean between automatic and manual transmission cars. The average MPG for cars with automatic transmissions is 17.17 MPG compared to the average for cars with manual transmissions: 24.39 MPG. (See *Figure 4 - MPG Means*)

Model Selection

Multilinear regression analysis is an effective tool to assess how the car's transmission type (automatic/manual) effects its fuel efficiency while controlling other variables. To set a baseline model, we first include only the transmission type as the variable as predictors pf mpg. (See *Figure 5 - Linear Regression - Single Variable*)

```
##              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
```

Because the very small p-value from the model above, the null hypothesis is rejected and we conclude there is linear correlation between the predictor variable am and mpg. The model summary returns an adjusted R^2 that explains only 34% of the variance.

Because more variables in this dataset likely have linear correlations with mpg, a multivariable regression model will be used to identify the variables to include in the final model - a stepwise regression function, step(). This adds and removes independent variables to the model until it finds the combination of independent variables that minimizes the AIC of the model.

```
summary(fit_step)$coef
```

```
##                Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 33.70832390 2.60488618 12.940421 7.733392e-13
## cyl6        -3.03134449 1.40728351 -2.154040 4.068272e-02
## cyl8        -2.16367532 2.28425172 -0.947214 3.522509e-01
## hp          -0.03210943 0.01369257 -2.345025 2.693461e-02
## wt          -2.49682942 0.88558779 -2.819404 9.081408e-03
## amManual     1.80921138 1.39630450  1.295714 2.064597e-01
```

The step regression suggests that cyl, hp, wt and am explain most of the variation with mpg. Comparing this suggested model with the first model where we analyzed mpg and am we find substantial evidence the step model is superior (See *Figure 6 - Best Fit Model*):

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ cyl + hp + wt + am
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     30 720.90
## 2     26 151.03  4    569.87 24.527 1.688e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residuals examination

In *Figure 7 - Residuals* the residual plots of our regression model are presented. From this information, it can be determined:

- The points in the Residuals vs. Fitted plot seem to be randomly scattered on the plot and verify the independence condition. (There is a bit of a curve to the residual plot that it departs slightly from normality. The residuals for the Chrysler Imperial, Fiat 128, and Toyota Corolla are called out because they exert some influence on the shape of the curve.)
- The Normal Q-Q plot consists of the points which mostly fall on the line indicating that the residuals are normally distributed.
- The Scale-Location plot consists of points scattered in a constant band pattern, indicating constant variance.
- There are some distinct points of interest (outliers or leverage points) in the top right of the plots.

Conclusions

Based on the analysis we can conclude:

- Cars with Manual transmission get 1.8 more miles per gallon compared to cars with Automatic transmission. (1.8 adjusted for hp, cyl, and wt).
- mpg will decrease by 2.5 for every 1000 lb increase in wt.
- mpg decreases negligibly (only 0.32) with every increase of 10 in hp.
- If number of cylinders, cyl increases from 4 to 6 and 8, mpg will decrease by a factor of 3 and 2.2 respectively (adjusted by hp, wt, and am).

Appendix

Figure 1: Data Definitions

Variable Name	Description
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
vs	V/S (V - V Engine; S - Straight Engine)
am	Transmission (0 = automatic, 1 = manual)
gear	Number of forward gears
carb	Number of carburetors

Figure 2 - Pairs Plot

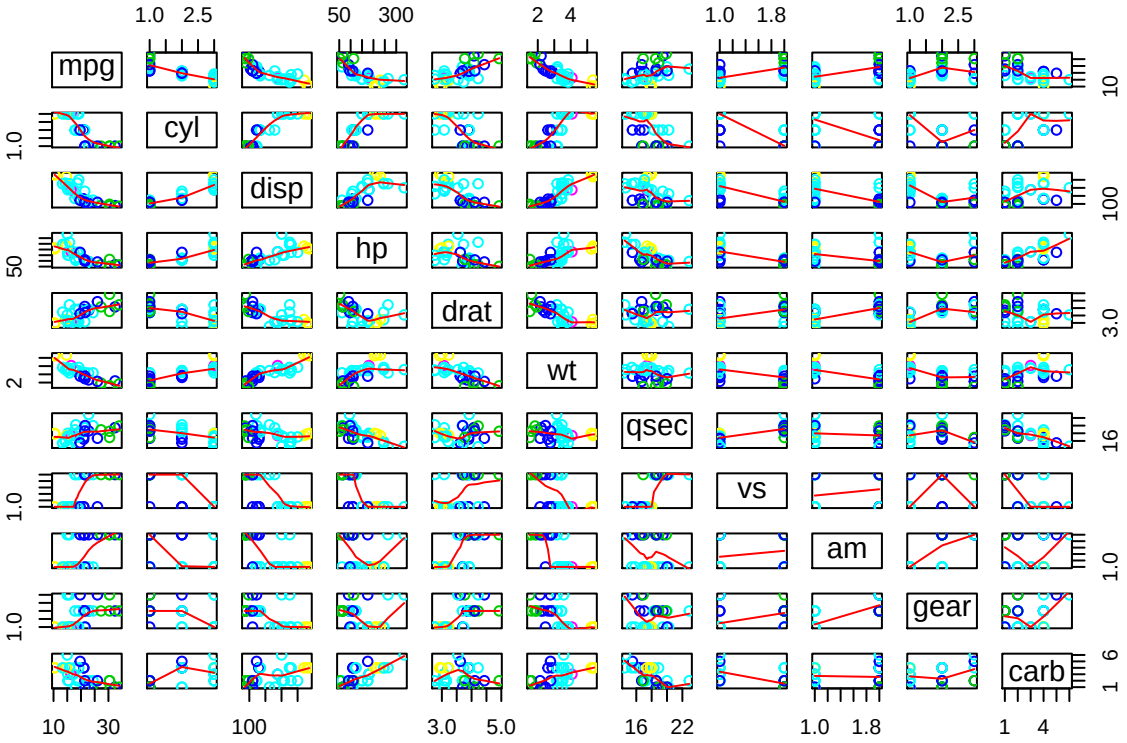


Figure 3 - Boxplot Transmission v MPG

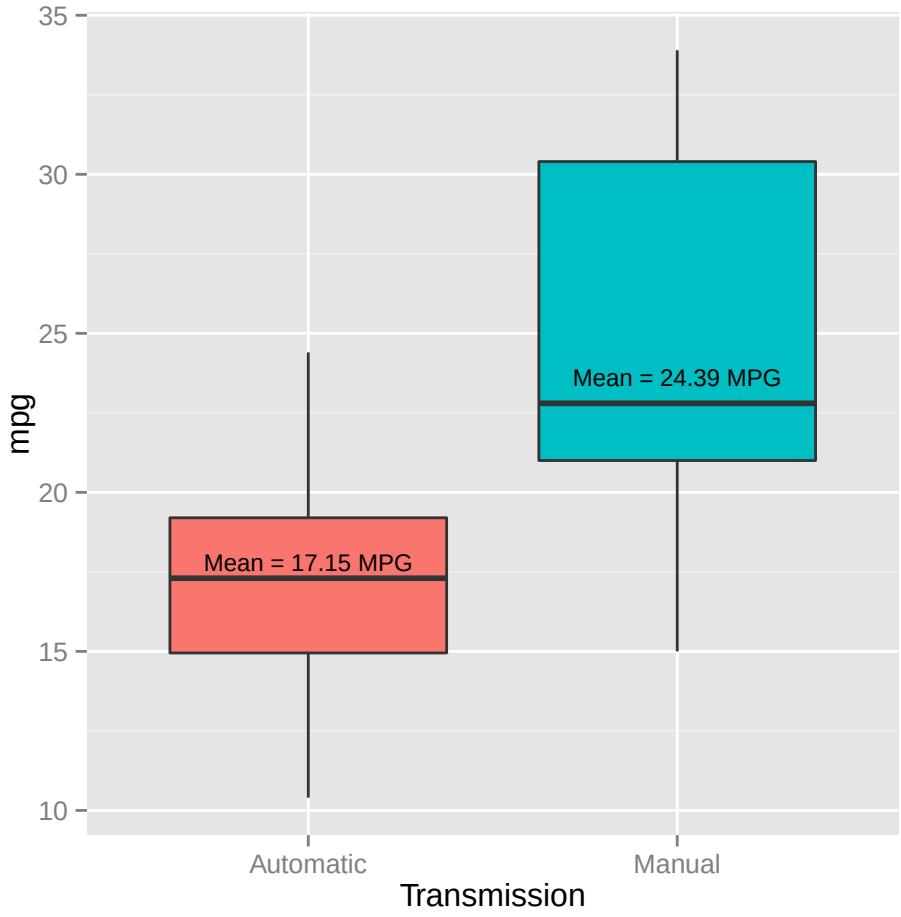


Figure 4 - MPG Means

```
## : Automatic
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10.40   14.95   17.30   17.15   19.20   24.40
## ------------------------------------------------------------
## : Manual
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   15.00   21.00   22.80   24.39   30.40   33.90
```

Figure 5 - Linear Regression - Single Variable

```
##              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
```

```
## [1] 0.3384589
```

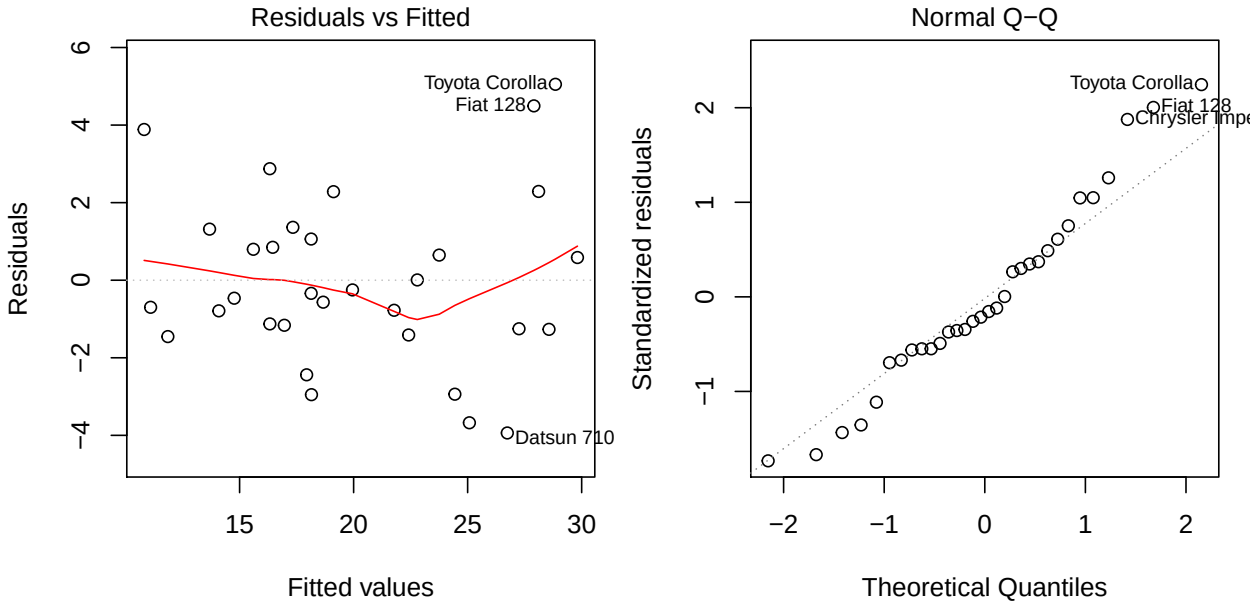
Figure 6 - Best Fit Model

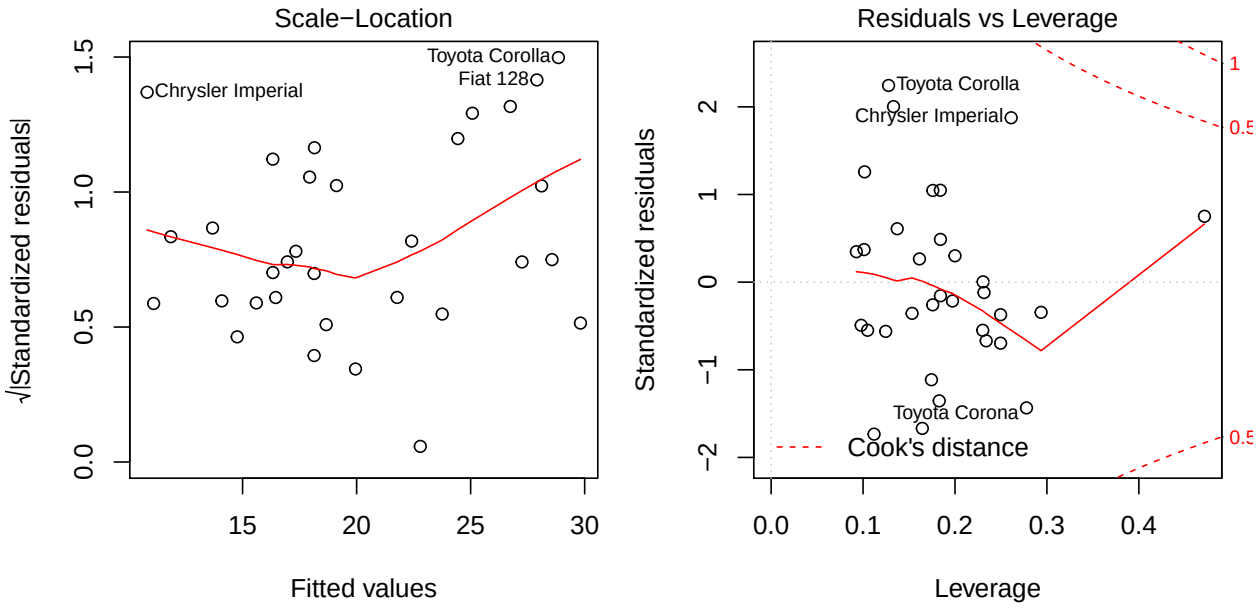
```
summary(fit_best)$coef
```

```
##                Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 33.70832390 2.60488618 12.940421 7.733392e-13
## cyl6        -3.03134449 1.40728351 -2.154040 4.068272e-02
## cyl8        -2.16367532 2.28425172 -0.947214 3.522509e-01
## hp          -0.03210943 0.01369257 -2.345025 2.693461e-02
## wt          -2.49682942 0.88558779 -2.819404 9.081408e-03
## amManual     1.80921138 1.39630450  1.295714 2.064597e-01
```

Figure 7 - Residuals

```
par(mfrow = c(1,2), mar = c(4, 4, 2, 1))
plot(fit_best)
```




Scale−Location
√|Standardized residuals|
Fitted values
Chrysler Imperial
Toyota Corolla
Fiat 128
Residuals vs Leverage
Standardized residuals
Leverage
Toyota Corolla
Chrysler Imperial
Toyota Corona
Cook's distance