Motor Trend Analysis

CVava, June 12th, 2015

Executive Summary

The purpose of this analysis is to find out whether cars with automatic or manual transmission have a better MPG and if true by how much. We have based our analysis on the data extracted from the *1974 Motor Trend US* magazine and the results reflect the reality of this limited sample.

With confidence level better than 95% we estimate that a car from our data set with manual transmission has a higher **mpg** than a car with automatic transmission if all other things remain the same.

With confidence level of 95% based on our selected model we estimate that switching from automatic to manual transmission we expect a 2 increase in **mpg** (more precisely in the interval [-1.3, 5.19]).

Major limitations. This analysis is based only on car models from 1973 and 1974, there are very few samples, and the multivariate model is linear. An unusual factor limiting the value of this analysis is the complex economic and political environment of 1973-1974 including the **Oil Crisis of 1973** creating massive gas price inflation and shortages leading to a higher demand for more efficient cars while the automatic transmission was still under development.

Exploratory Data Analysis

The analysis is based on the **mtcars** set containing 32 observations of 11 variables (shown below with their meaning and type). Some variables will be treated as discrete due to their very low number of possible values.

Variable	Meaning	Type (# of levels)
mpg	Miles/(US) gallon	continuous
cyl	Number of cylinders	discrete (3)
disp	Displacement (cu.in.)	continuous
hp	Gross horsepower	continuous
drat	Rear axle ratio	continuous
wt	Weight (lb/1000)	continuous
qsec	1/4 mile time	continuous
vs	(V engine or Straight engine)	discrete (2)
am	Transmission (0 = automatic, 1 = manual)	discrete (2)
gear	Number of forward gears	discrete (3)
carb	Number of carburetors	discrete (6)

Figure 1 from the Appendix shows the relations between all variables. *Carburetors* are represented by symbols (1 circle, 2 square, 3 diamond, 4 triangle), *cylinders* by outside color (4 black, 6 green, 8 cyan), *gear* by inside color (3 black, 4 red, 5 green), *engine type* by line width (thin vs thick), and *transmission type* by symbol size (manual shown as smaller).

Outliers. We'll eliminate the samples with 6 and 8 carburetors since each has a single observation and an outlier effect (for 8 carburetors average **mpg** is 15.1 vs average for the rest of 24).

Figure 1 shows a large triangle with thin green line and cyan fill representing an unique car (straight engine, automatic, 5 gear, 8 cylinders, 4 carburetors) far away from all the other samples with common parameters.

Meaningful Contributions. All variables seem to have some effect on **mpg**. However, some variables appear highly correlated as **disp - cyl** (0.91), **disp - wt** (0.9) and **cyl - wt** (0.8) and may be eliminated.

Model Design

As *Figure 1* shows, most high **mpg** cars have 4 cylinders, 3 gears, S engine, 1 or 2 carburetors, low weight, high *1/4 mile time* and manual transmission. To execute a comparison over the whole potential range of all variables we have to build a model.

After fitting three models (**Model A** - all variables, **Model B** - exclude **disp** and **vs**, **Model C** - exclude **disp**, **vs** and **cyl**) and doing the anova comparison we found large p-values (A - B as 0.28, B - C as 0.63).

Figure 2 from the Appendix shows the residual plot of **Model C**. For the limited number of samples the spread appears acceptable. *Figure 3* shows the histogram of residuals "suggesting" the bell curve but with an elevated right half. To understand this slight aberration we have to remember the complex economical and political environment of 1973-1974 creating massive gas price inflation and shortages leading to a higher demand for more efficient cars (Toyota Corolla, Fiat 128, Chrysler Imperial, Merc 240D). *Figure 4* shows the same residuals making more obvious the effect of these unusual samples.

We have decided to retain **model C** for our analysis. It's R^2 value of 0.877 is acceptable for such a small sample and is very close to the value 0.89 of **model A** (based on all variables). Residuals' distribution is acceptable and the histogram is the closest to a bell curve from any other linear model we have tested.

```
##                    Estimate  Std. Error      t value  Pr(>|t|)
## (Intercept)    7.879918e+00 12.92428922   0.609698335 0.5486026
## hp             3.348066e-05  0.02350232   0.001424568 0.9988768
## drat           1.995366e+00  1.85206077   1.077376302 0.2935334
## wt            -1.996376e+00  1.29778763  -1.538291454 0.1389099
## qsec           4.609215e-01  0.58190757   0.792087099 0.4371666
## am             1.954077e+00  1.96529582   0.994291333 0.3314006
## gear           1.753221e+00  1.52433034   1.150158402 0.2630084
## carb          -1.381505e+00  0.82483154  -1.674893362 0.1087840
```

The coefficients tell us for example that for each increase of 1000lb in **wt** we should expect to see an **mpg** decrease of -2 and for each increase in the number of **gears** we should expect an **mpg** increase of 1.8.

Is an automatic or manual transmission better for MPG

A simplistic analysis (average **mpg** of all cars with or without automatic transmission) shows that a car with automatic transmission (**am** = 0) has a lower average **mpg** (17.1) than with manual transmission (26.7).

The 95% confidence interval for the difference between the **mpg** for cars with automatic and manual transmission is [3.2 11.3] which implies again that the automatic transmission has a lower **mpg**.

From the retained model we can extract the coefficient for the **am** parameter which implies that switching from automatic to manual transmission we should expect a 2 increase in **mpg**. This value would be the best estimation if not for the limitations already mentioned.

The MPG difference between automatic and manual transmissions

The **MPG** differential can be estimated roughly from the **mpg** averages of cars with manual and automatic transmission as 9.51.

From the statistical analysis of the two data sets (manual and automatic transmission) we can infer the 95% confidence interval for the difference between the **mpg**s [3.2 11.3] as a better estimate.

Using the retained model we can derive the confidence interval for the **mpg** difference knowing that the **mpg** increase has a mean of 2 and a standard deviation of 1.97. With the same confidence level of 95% we can predict that the difference will be somewhere in the interval [-1.3, 5.19].
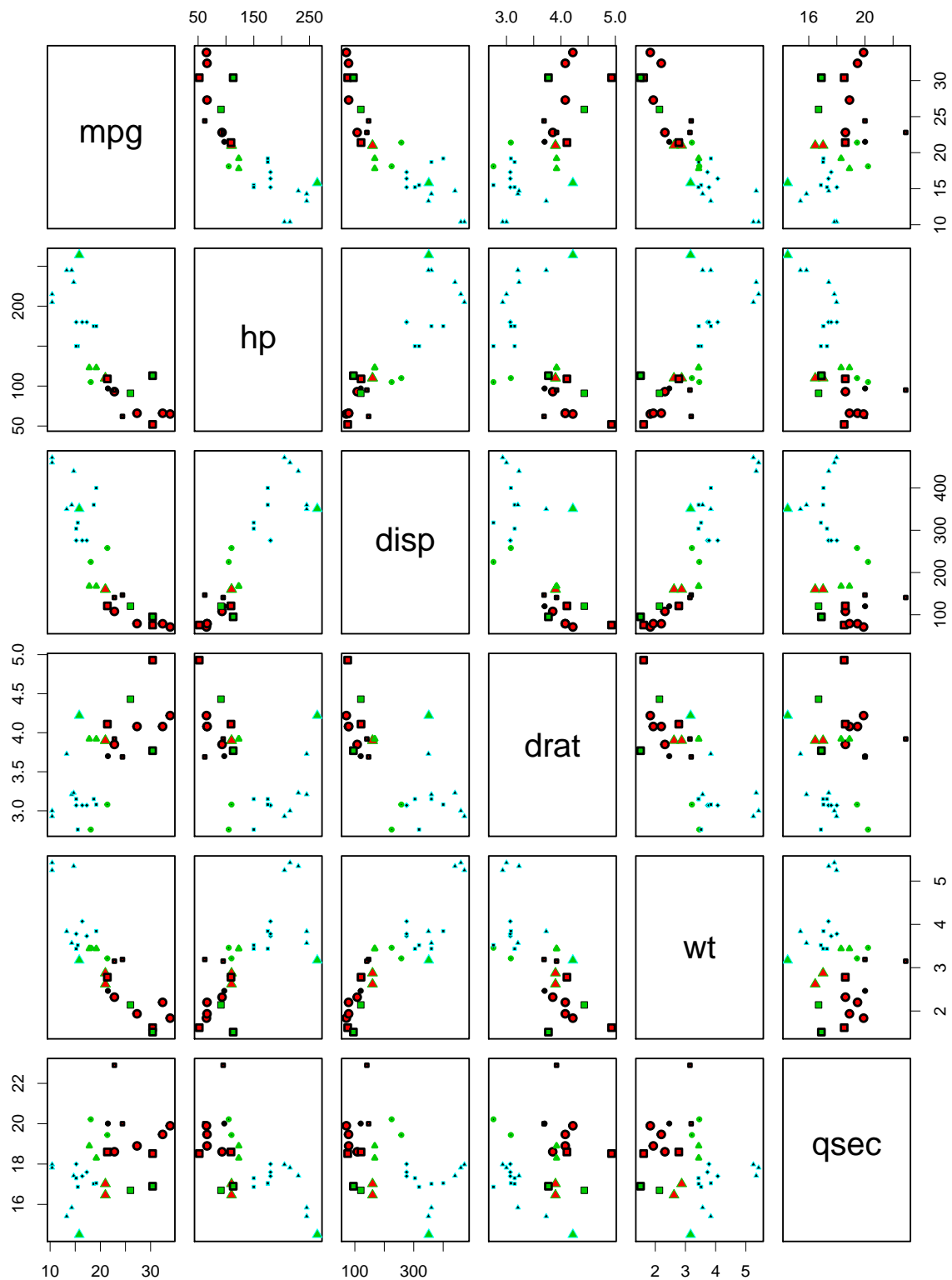
Fig 1. Exploratory Visualization of mtcars Data Set
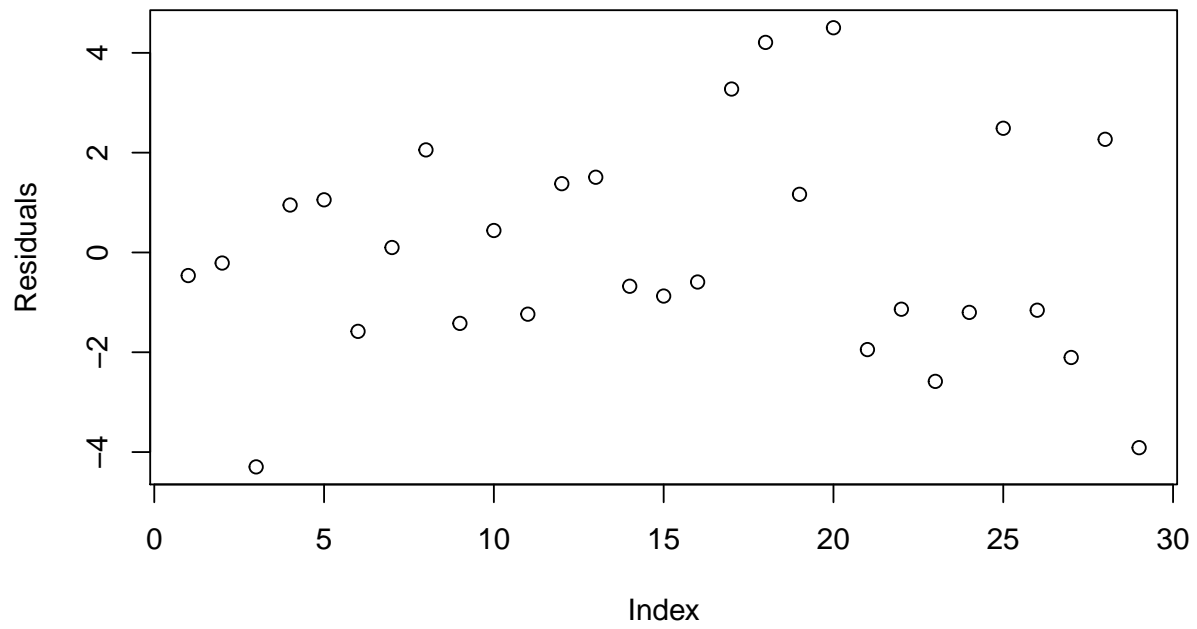
Fig 2. Residuals Plot of Retained Model
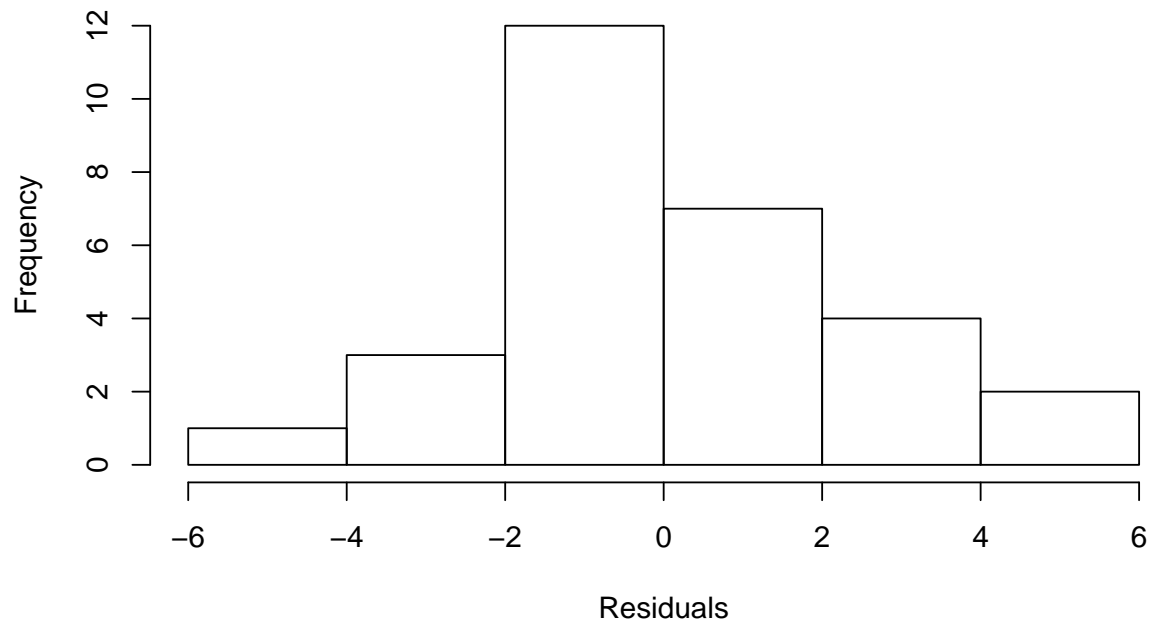

Fig 3. Residuals Histogram of Retained Model
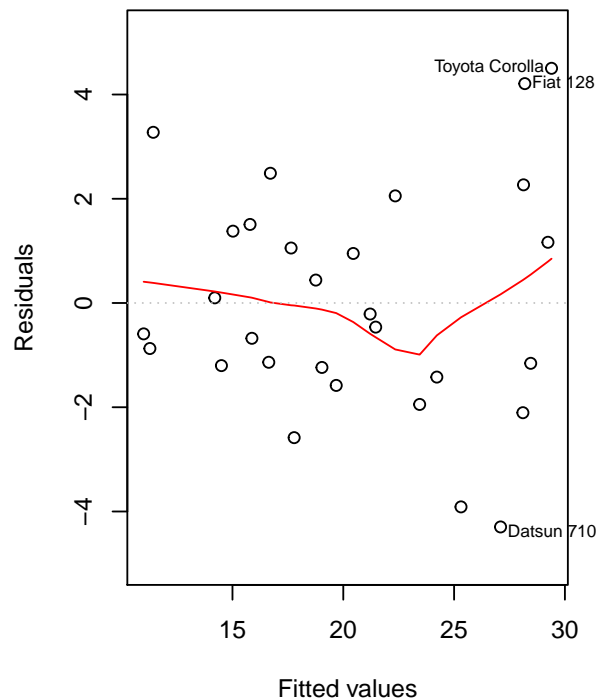
Fig 4. Residuals of Retained Model
Residuals vs Fitted

Fig 4. Residuals of Retained Model
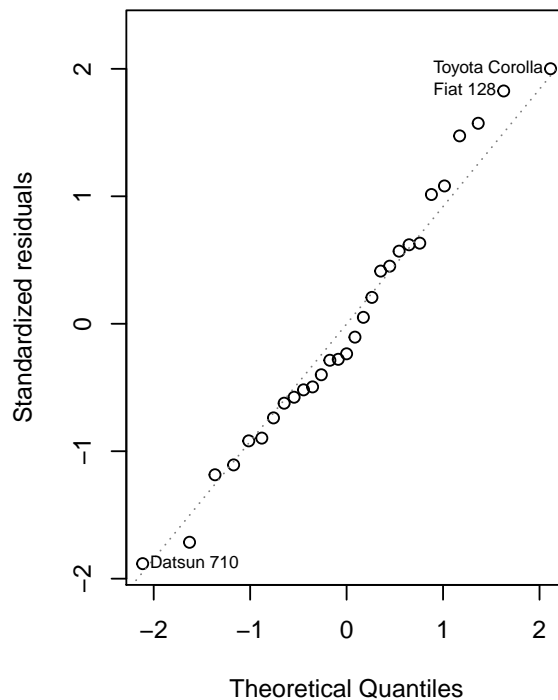Normal Q–Q

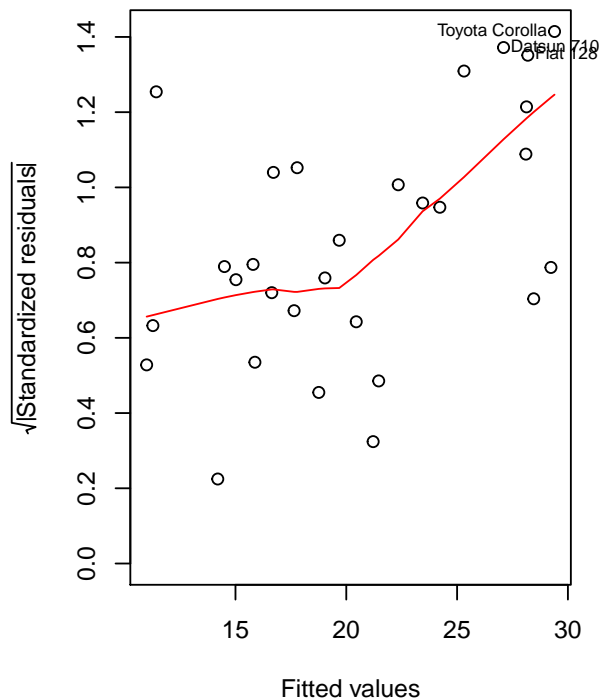**Fig 4. Residuals of Retained Model**
Scale–Location

Fig 4. Residuals of Retained Model
Residuals vs Leverage