Regression Models Course Project

Sandeep Anand

April 23, 2017

Executive Summary

Analysis for *Motor Trend* and Showing the dependence of MPG on specific covariates

Two Important analysis that needs to be done are below

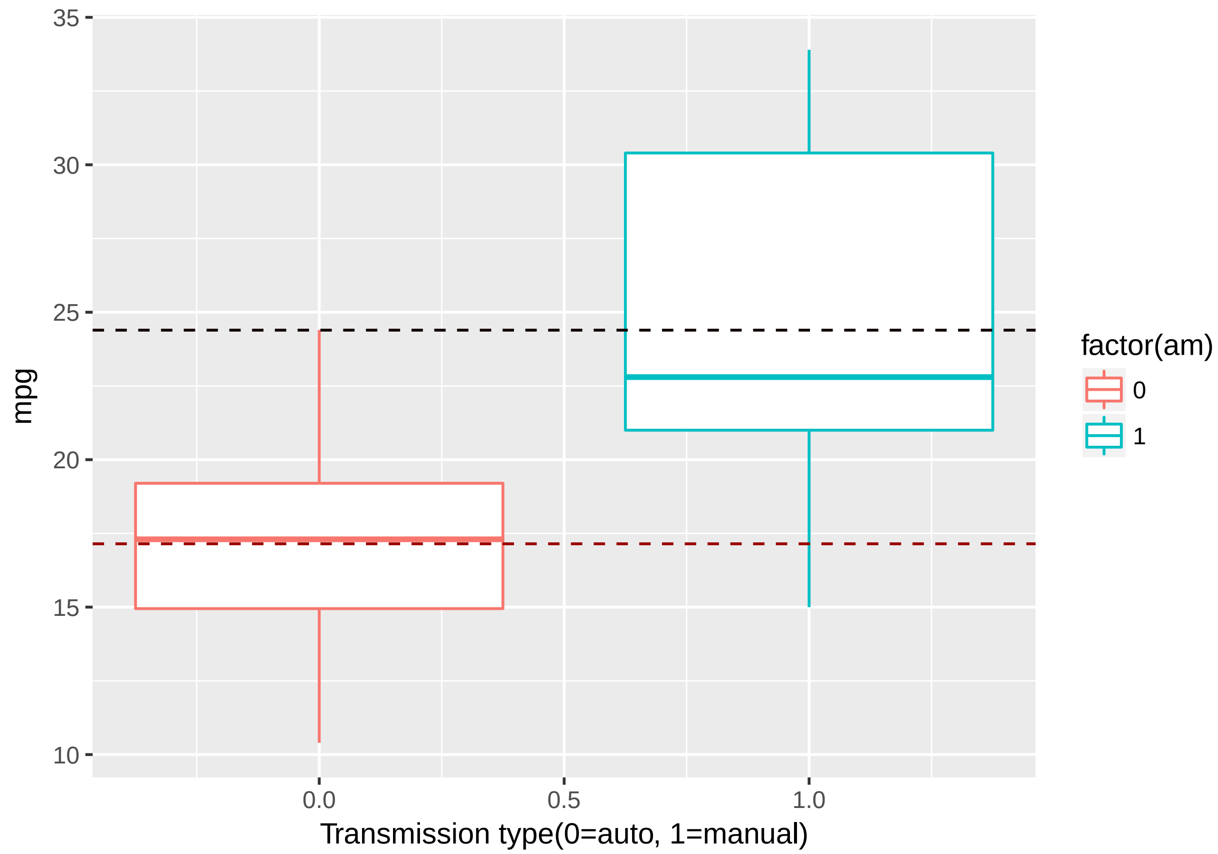
- Is an automatic or manual transmission better for MPG
- Quantify the MPG difference between automatic and manual transmissions

Step-1

- Finding the Coefficients
- Determining which transmission has better MPG
- Comparing the Means of both the transmissions below
- t-test to check the p-value and the means as well

```
##   Transmission  meanval
## 1    Automatic 17.14737
## 2       Manual 24.39231
```

```
##
##  Welch Two Sample t-test
##
## data:  auto$mpg and manual$mpg
## 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 of x mean of y
##  17.14737  24.39231
```



Result-1

- The Box-plot shows that the mean of mpg for manual transmission cars are better than auto transmissions
- The t test result also shows that , p value = 0.0013 , ie < 0.05 , so we can reject the Null Hypothesis and hence Automatic Transmission shows lower MPG than manual transmissions, though there are other covariates that we would need to see if they have any particular relation in deacrease in the mpg .

Now analyzing this based on different Models we can fit .

1st Model Fit

- **Fitting A linear Model with response as MPG and Factoring the am variable**

Trying Fitting Linear Model here

```r
fitlm<-lm(mpg~factor(am)-1,data = mtcars)
summary(fitlm)
```

```
##
## Call:
## lm(formula = mpg ~ factor(am) - 1, 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|)
## factor(am)0   17.147      1.125   15.25 1.13e-15 ***
## factor(am)1   24.392      1.360   17.94  < 2e-16 ***
## ---
## 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:  0.9487, Adjusted R-squared:  0.9452
## F-statistic: 277.2 on 2 and 30 DF,  p-value: < 2.2e-16
```

Conclusions from the Linear Model Fit

- Both the Transmission values seem to be important and significant
- Mean that are the coefficients reflect the fact that the response mpg is more dependent on am=1 , ie, Manual transmission
- The Above is a multivariate analysis where we consider both factors am=0 and am=1
- $$R^2 = -.95$$

  ie, "Goodness of Fit", so it seems that 95% of the variance of my response ie, mpg is explained by my model , but is that actually true ?
    - To answer the above question we would need to feed in other covariates and check

2nd Model Fit

- **Fitting A Multivariate linear Model with response as MPG and all the other covariates present**
- **Fitting A Multivariate linear Model with response as MPG and specific dependent covariates present** ### Conclusions from the plots below :
- Seems this model is following normality with most of the points for the Residuals
- The Slope Coefficients of the factored am[0,1] show the high dependence of "mpg" on Transmission modes
- From the Plot-2, it is clear the autocorreclation plot show no obvious trend of failure, and fitted values vs. Residuals appear quite random
- From the Best Fitting model with covariates weight, horsepower and Transmission mode, It is clear than there is high dependence of mpg on am.
- With a Very low p-value for the best fit model, we can reject the Null Hypothesis
- The below model with high R^2 value explain most of the variance **Summaried of Two Multivariate models compared**

```
## lm(formula = mpg ~ . + factor(am) - am - 1, data = mtcars)
```

```
##                Estimate  Std. Error    t value   Pr(>|t|)
## cyl         -0.11144048  1.04502336 -0.1066392 0.91608738
## disp         0.01333524  0.01785750  0.7467585 0.46348865
## hp          -0.02148212  0.02176858 -0.9868407 0.33495531
## drat         0.78711097  1.63537307  0.4813036 0.63527790
## wt          -3.71530393  1.89441430 -1.9611887 0.06325215
## qsec         0.82104075  0.73084480  1.1234133 0.27394127
## vs           0.31776281  2.10450861  0.1509915 0.88142347
```

```
## gear          0.65541302  1.49325996  0.4389142 0.66520643
## carb         -0.19941925  0.82875250 -0.2406258 0.81217871
## factor(am)0 12.30337416 18.71788443  0.6573058 0.51812440
## factor(am)1 14.82360104 18.35265170  0.8077090 0.42831299
```

```
## [1] 0.9894964
```

```
## lm(formula = mpg ~ am + wt + hp, data = mtcars)
```

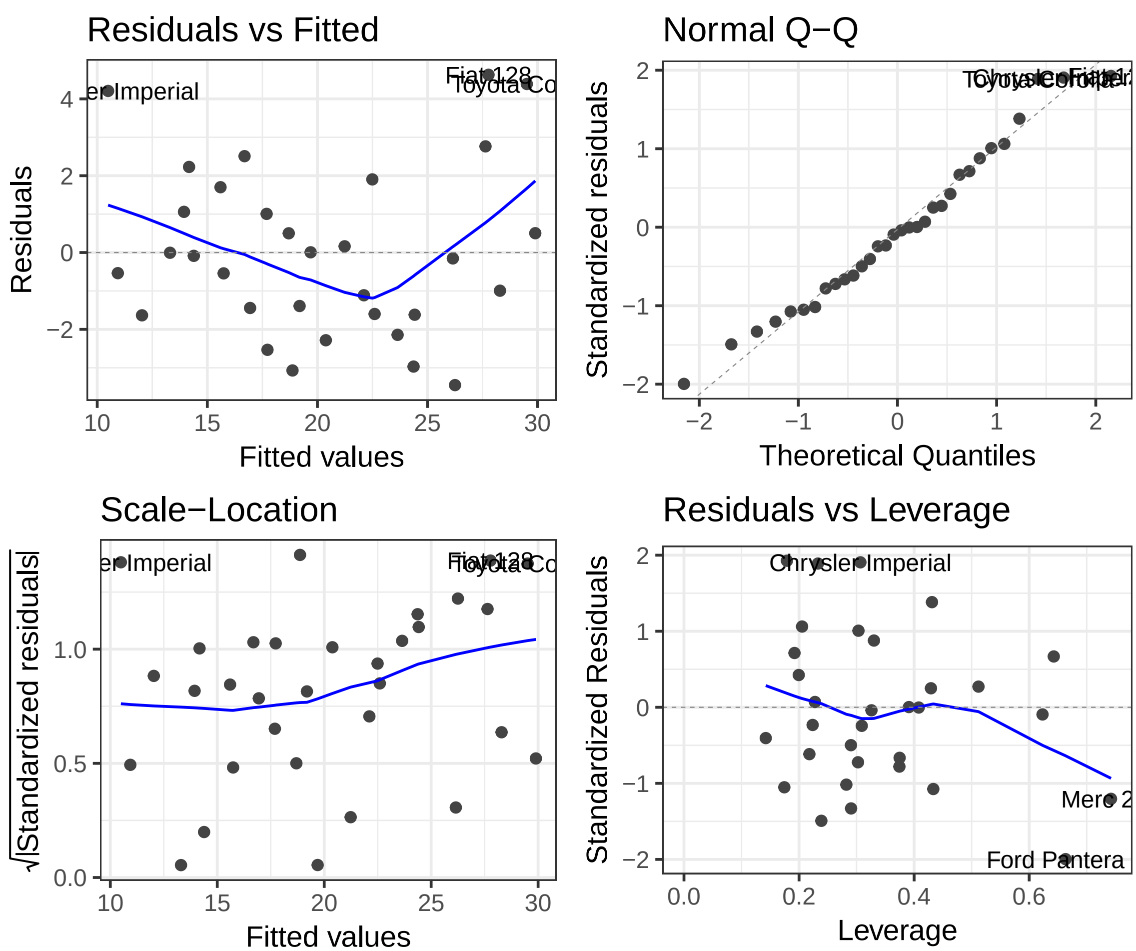
```
##                Estimate  Std. Error   t value     Pr(>|t|)
## (Intercept) 34.00287512 2.642659337 12.866916 2.824030e-13
## am           2.08371013 1.376420152  1.513862 1.412682e-01
## wt          -2.87857541 0.904970538 -3.180850 3.574031e-03
## hp          -0.03747873 0.009605422 -3.901830 5.464023e-04
```

```
## [1] 0.8398903
```

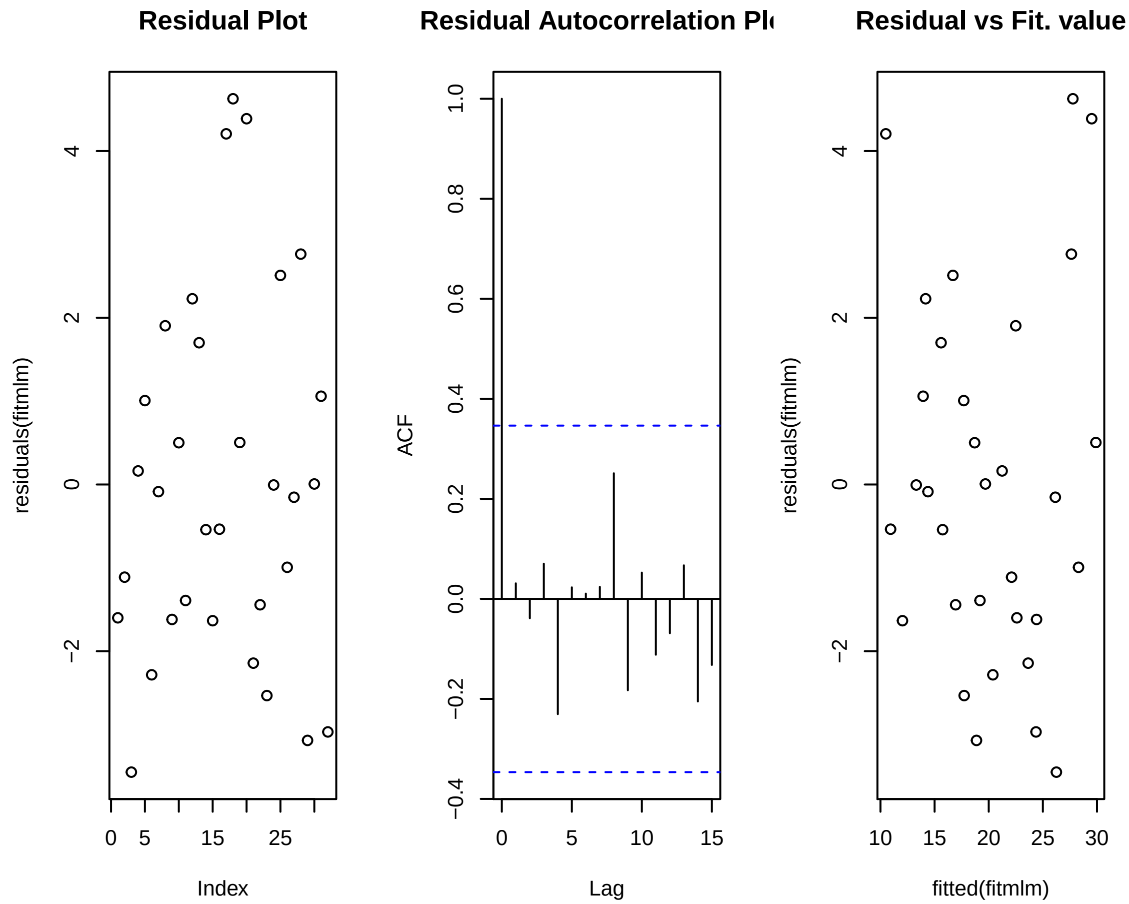
Appendix

plot-1

```
## Warning: package 'ggfortify' was built under R version 3.3.3
```



plot-2



Residual Plot
Residual Autocorrelation Plo
Residual vs Fit. value
residuals(fitmlm)
−2
0
2
4
0
5
15
25
Index
ACF
−0.4
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
Lag
residuals(fitmlm)
−2
0
2
4
10
15
20
25
30
fitted(fitmlm)