Effects of Vehicle Transmission Type on Fuel Efficiency

Executive Summary

Data analysis suggests that cars with manual transmission have better fuel efficiency compared to those with automatic transmission. Including weight in the regression model brought us to conclusion that this is actually to a great extend effect of the car's weight as a confounding variable. Cars with automatic transmission are heavier and have lower MPG than cars with manual transmission, so we cannot quantify how different is fuel efficiency taking into consideration just transmission type, in this case there is very small difference in MPG between automatic and manual cars. Adding the interaction between weight and transmission types in the model and comparing MPG between cars with manual and automatic transmission with same weight we see more significant positive difference. The last model where MPG is quantified with transmission type, weight, interaction between weight and transmission, and horse power as explanatory variables, shows significant positive difference of MPG between cars with manual and automatic transmission with same weight and horsepower. Generally cars with more horsepower have poorer fuel efficiency than those with less horsepower.

Introduction

The purpose of this analysis is to investigate which type of transmission is better in respect of fuel efficiency and to quantify how different is fuel efficiency between automatic and manual transmission. The analysis is done with data sample **mtcars** available in {datasets} package.

> 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).[1]

Exploratory Analysis

Dataset comprises of 32 observations on 11 variables. 6 of the variables are continuous numerical: fuel efficiency in miles per gallon (mpg), engine displacement in cubic inches (disp), gross horsepower(hp), rear axle ratio (drat), car's weight (wt) in 1000 pounds and quarter mile time (qsec). The rest 5 are categorical variables: number of cylinders (cyl), V/straight engine (vs), transmission type (am; 0-automatic, 1-manual), number of forward gears (gear) and number of carburetors (carb). In this phase variables were investigated plotting histograms, boxplots and qq plots for continuous numeric variables (Fig. 1) and bar plots for categorical variables(Fig 2).

Quarter mile time variable is approximately normally distributed; horse power and engine displacement show trend of right skew; fuel efficiency, weight and rear axle ratio show some elements of multimodality following the trend of the bell curve. There are outlier observations in horse power and quarter time variables.

In order to investigate the relationship between fuel efficiency and other parameters matrix of scatterplots is generated (Fig. 3). None of the graphs in the matrix show a perfectly linear relationship however, a line does reasonably estimate the average change in MPG for all 5 continuous variables.

Statistical Modeling

To relate fuel efficiency to transmission type we performed a standard simple and multivariate linear regression models. Model selection was performed on the basis of our exploratory analysis and questions of interest. Coefficients were estimated with ordinary least squares and standard errors were calculated using standard asymptotic approximations.

Reproducibility

All analyses performed in this manuscript are reproduced in the R markdown file mtcarsfinal.Rmd available on Github [https://github.com/CrazyFarang/RegressionModelsProject (https://github.com/CrazyFarang/RegressionModelsProject)].

Results

We first fit unadjusted regression model relating fuel efficiency and transmission type.

$$mpg = \beta_0 + \beta_1 * 1(am = \text{"Manual"}) + e$$

```
##               Estimate Std. Error t value  Pr(>|t|)
## (Intercept)     17.147      1.125  15.247 1.134e-15
## as.factor(am)1   7.245      1.764   4.106 2.850e-04
```

where $\beta_0 = 17.15$, 95% CI(14.85,19.44) is average value of miles per gallon for cars(MPG) with automatic transmission, $\beta_1 = 7.24$, 95% CI(3.64,10.85) is estimated expected change in MPG for cars with manual transmission and error term e represents all sources of unmeasured and unmodeled random variation in MPG (Fig. 4) In this model transmission type is statistically significant predictor of MPG (2.8502×10^{-4}), but the model has very poor power of prediction, it explains just 34 % of the variability in MPG. Residual analysis (Fig. 5) shows a pattern in residual versus dependent and fitted variable, almost normal distribution of residuals. Influence measures analysis doesn't point to high leverage observation in this model.

In order to quantify fuel efficiency we need to fit multivariable model. Second model we fit is adding vehicle's weight into initial simple regression model in order to show that weight is confounding variable high negatively correlated both with transmission type (r=-0.6925) as dependent variable and MPG as response variable (r=-0.8677). We fit the model

$$mpg = \beta_0 + \beta_1 \, 1(am = \text{"Manual"}) + \beta_2 wt + e,$$

```
##               Estimate Std. Error  t value  Pr(>|t|)
## (Intercept)   37.32155    3.0546  12.21799 5.843e-13
## as.factor(am)1 -0.02362   1.5456  -0.01528 9.879e-01
## wt            -5.35281    0.7882  -6.79081 1.867e-07
```

where $\beta_0 = 37.32$, 95% CI(31.07,43.57) is expected value of MPG for automatic cars with weight 0 pounds, $\beta_1 = -0.02$, 95% CI(-3.18,3.14) is estimated expected change in MPG for cars with manual transmission and weight 0 pounds($\beta_0 + \beta_1$ is expected value of MPG for cars with manual trnasmission and weight 0 pounds), $\beta_2 = -5.35$, 95% CI(-6.96,-3.74) is expected change in MPG for 1 lb/1000 change in weight for either automatic or manual transmission and error term e represents all sources of unmeasured and unmodeled random variation in MPG. (Fig. 6) Transmission type in this model is not longer statistically significant predictor of MPG (p=(0.99, the two regression lines for automatic and manual transmission almost overlap, so the only predictor of MPG is actually the confounding variable - weight. Cars with automatic transmission are heavier and have worse fuel efficiency than cars with manual transmission which have bigger MPG. This model explains 74 % of the variability in MPG, much better than the previous one, but that is not result of the

transmission type, rather than the weight. Residual analysis (Fig. 7) shows random trend in residual versus fitted values, distribution of residuals is not quite normal. Influence measures analysis doesn't point to high leverage observations in this model.

The third model adds interaction of weight and transmission to the second model.

$$mpg = b_0 + b_1 \, 1(am = \text{"Manual"}) + b_2 wt + b_3 \, 1(am = \text{"Manual"}) wt + e$$

```
##              Estimate Std. Error t value  Pr(>|t|)
## (Intercept)    31.416     3.0201  10.402  4.001e-11
## as.factor(am)1 14.878     4.2640   3.489  1.621e-03
## wt             -3.786     0.7856  -4.819  4.551e-05
## as.factor(am)1:wt -5.298  1.4447  -3.667  1.017e-03
```

Estimated values of regression coefficients and their 95% confidence intervals: b_0=31.42, 95% CI(25.23,37.6); b_1=14.88, 95% CI(6.14,23.61); b_2=-3.79, 95% CI(-5.4,-2.18); b_3=-5.3, 95% CI(-8.26,-2.34). b_0 is expected value of MPG for automatic cars with weight 0 pounds; $b_0 + b_1$ is expected value of MPG for cars with manual transmission and weight 0 pounds; b_2 is expected change in MPG for 1 lb/1000 change in weight of automatic cars; $b_2 + b_3$ is expected change in MPG for 1 lb/1000 change in weight of cars with manual transmission and e is everything we didn't measure. (Fig. 8) Interaction of transmission type and weight brings change in quantifying MPG, regression lines for the two types of transmission are not overlapping anymore, neither they are parallel. Transmission type in this model is statistically significant predictor of MPG (p=0.0016) and this model explains 82 % of the variability in MPG. Residual analysis (Fig. 9) shows random trend in the relationship of residuals and fitted values, distribution looks more like normal compared with the previous model. Influence measures analysis doesn't point to high leverage observations in this model.

Last we fit the model using forward selection method of p values [2], starting with model that includes just the transmission type of the vehicle. We also observe Akaike's information criterion (AIC) [3] for the fitted model, and it seems that corresponds with the p value selection. The idea is to keep transmission type statistically significant and add some other predictors that are good quantifiers of fuel efficiency (MPG). In the first step the selection is horse power, so we add logarithmic transformation of horse power variable (remember it was right skewed). In the second we select the weight, and since the weight is confounding variable, we additionally add the interaction between the weight and transmission type. So according this approach the model that best quantifies MPG keeping the transmission variable statistically significant is following (Fig. 10):

$$mpg = b_0 + b_1 \, 1(am = \text{"Manual"}) + b_2 \log 10(hp) + b_3 \, wt + b_4 1(am = \text{"Manual"}) * wt + e,$$

```
##              Estimate Std. Error t value  Pr(>|t|)
## (Intercept)    50.933     5.6876   8.955  1.434e-09
## as.factor(am)1 10.111     3.7174   2.720  1.128e-02
## log10(hp)     -11.769     3.0866  -3.813  7.245e-04
## wt             -2.162     0.7729  -2.798  9.374e-03
## as.factor(am)1:wt -3.105  1.3183  -2.355  2.603e-02
```

Estimated values of regression coefficients and their 95% confidence intervals: b_0=50.93, 95% CI(39.26,62.6); b_1=10.11, 95% CI(2.48,17.74); b_2=-11.77, 95% CI(-18.1,-5.44); b_3=-2.16, 95% CI(-3.75,-0.58); b_4=-3.1, 95% CI(-5.81,-0.4). b_0 is expected MPG of car with automatic transmission, weight 0 pounds and horse power 1, b_1 is expected change in MPG in cars with manual transmission, weight 0 pounds and horse power 1, b_2 is expected change in MPG with change of 1 unit in log base 10 of horse power for cars holding all other predictors constant, b_3 is expected change in MPG for automatic cars with one unit change (1lb/1000) of weight holding horse power constant, b_4 is expected change of change of MPG for manual cars with one unit change (1lb/1000) of weight holding horse power constant and e is error term comprising everything we didn't measure. Automatic cars have worse MPG compared to a manual car with same weight and horse power. Cars with more horsepower have poorer fuel efficiency than those with less horsepower. We observed a statistically significant (p=0.0113) association between MPG and transmission type and this model explains 88 % of the variability in MPG. Residuals (Fig. 11) show random pattern of variation, indicating a good fit for linear model, their distribution is approximately normal and they are homoscedastic.

The overall p value value of the model is significant (p=1.244 $\times$ 10^{-12}), telling us that regression as a whole is performing better than any set of random predictors, the model outperforms random noice as predictor.

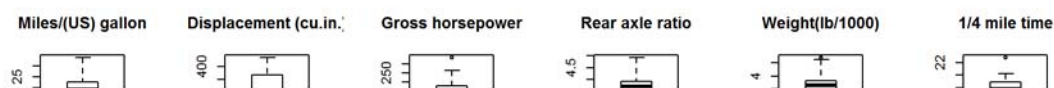Nested model analisys shows that every subsequent model has better perfomance than the previous one.

```
## Analysis of Variance Table
##
## Model 1: mpg ~ as.factor(am)
## Model 2: mpg ~ as.factor(am) + wt
## Model 3: mpg ~ as.factor(am) + wt + as.factor(am):wt
## Model 4: mpg ~ as.factor(am) + wt + log10(hp) + as.factor(am):wt
##   Res.Df RSS Df Sum of Sq    F  Pr(>F)
## 1     30 721
## 2     29 278  1      443 97.8  1.8e-10 ***
## 3     28 188  1       90 19.9  0.00013 ***
## 4     27 122  1       66 14.5  0.00072 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

References

1. Motor Trend Car Road Tests. R Documentation. [http://stat.ethz.ch/R-manual/R-patched/library/datasets/html/mtcars.html (http://stat.ethz.ch/R-manual/R-patched/library/datasets/html/mtcars.html)]
2. Forward- and Backward-Stepwise Selection. [book auth.] Robert Tibshirani & Jerome Friedman Trevor Hastie. The Elements of Statistical Learning; Data Mining, Inference and Prediction. s.l. : Springer, 2013. [http://statweb.stanford.edu/~tibs/ElemStatLearn/ (http://statweb.stanford.edu/~tibs/ElemStatLearn/)]
3. Akaike's Information Criterion. Wikipedia. [http://en.wikipedia.org/wiki/Akaike_information_criterion (http://en.wikipedia.org/wiki/Akaike_information_criterion).]
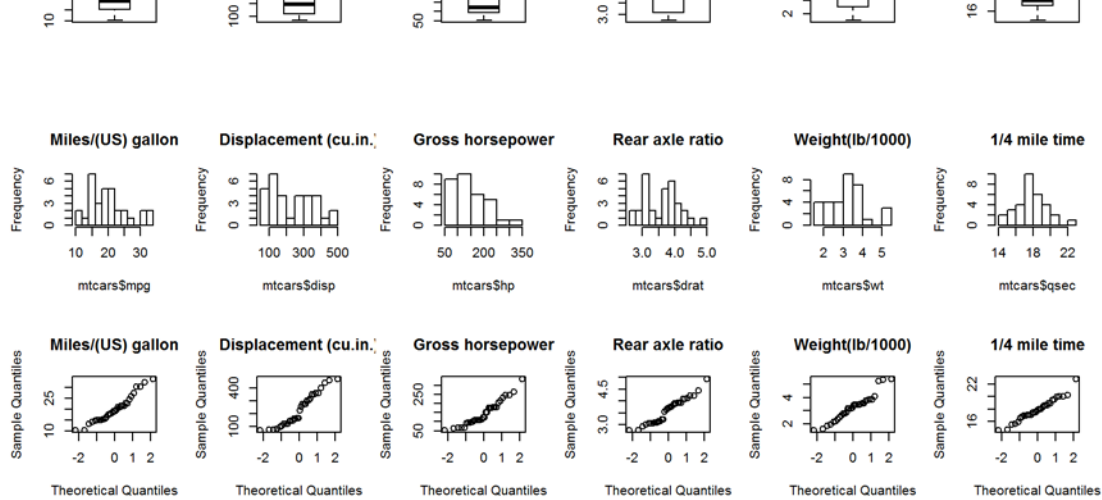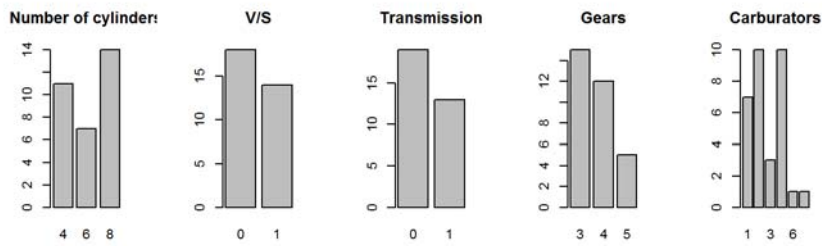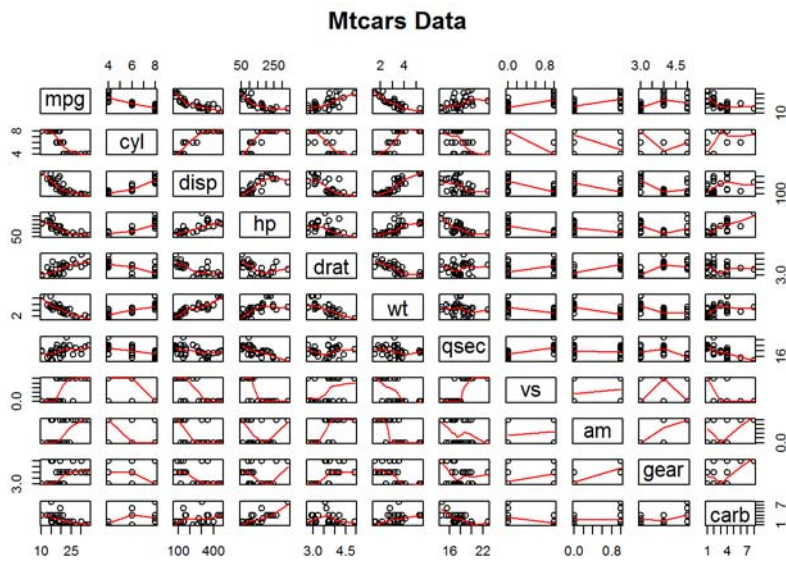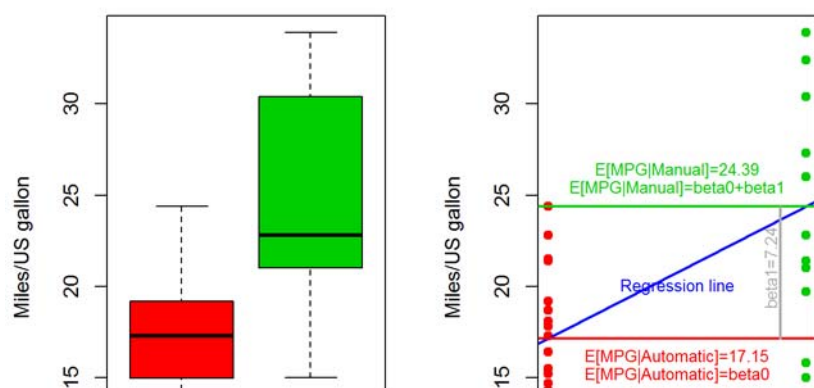
Apendix

Fig. 1

Fig. 2

Fig. 3

Mtcars Data

Fig. 4

Fuel Efficiency by Transmission

E[MPG|Manual]=24.39
E[MPG|Manual]=beta0+beta1

Regression line

E[MPG|Automatic]=17.15
E[MPG|Automatic]=beta0

10 ─
                Automatic   Manual   Automatic   Manual
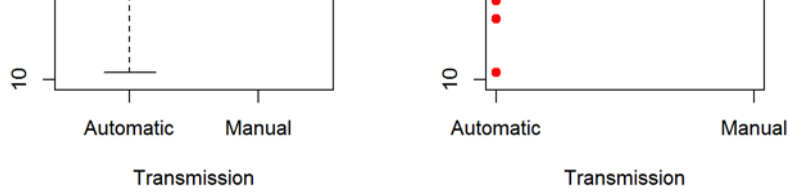
                    Transmission   Transmission

Fig.5

Fig. 6

Fuel Efficiency and Weight by Transmission
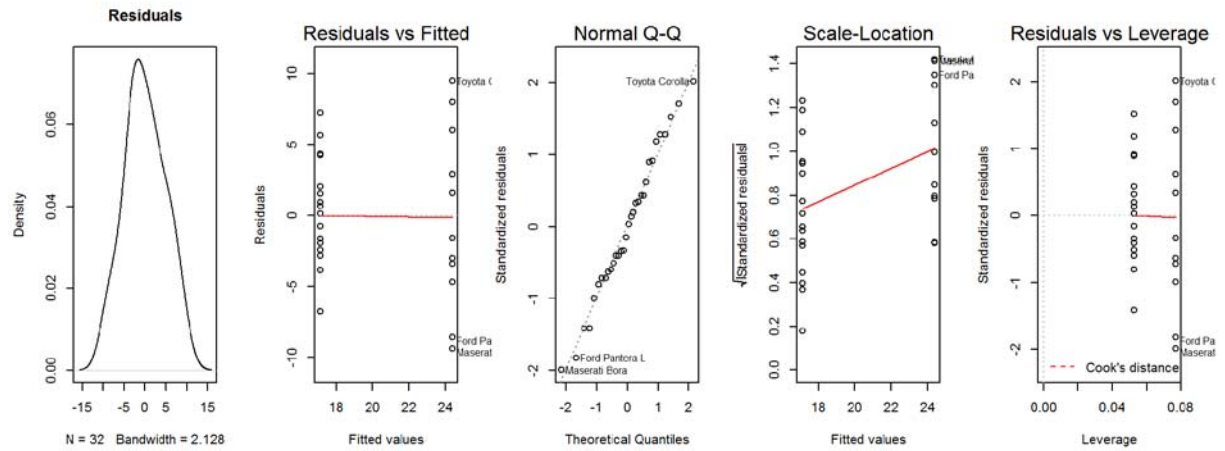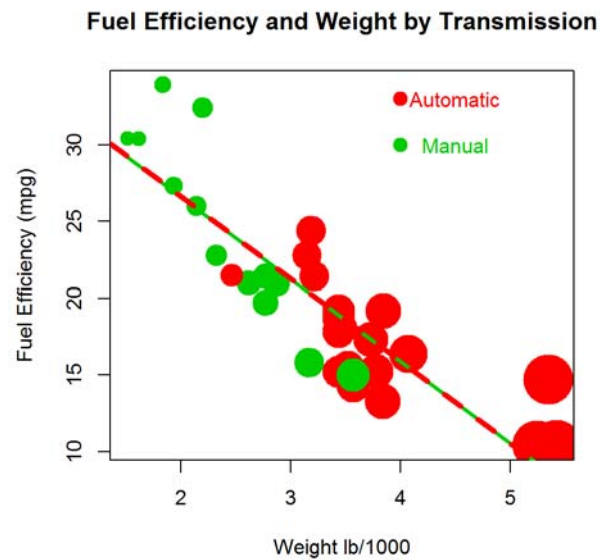
Fig. 7
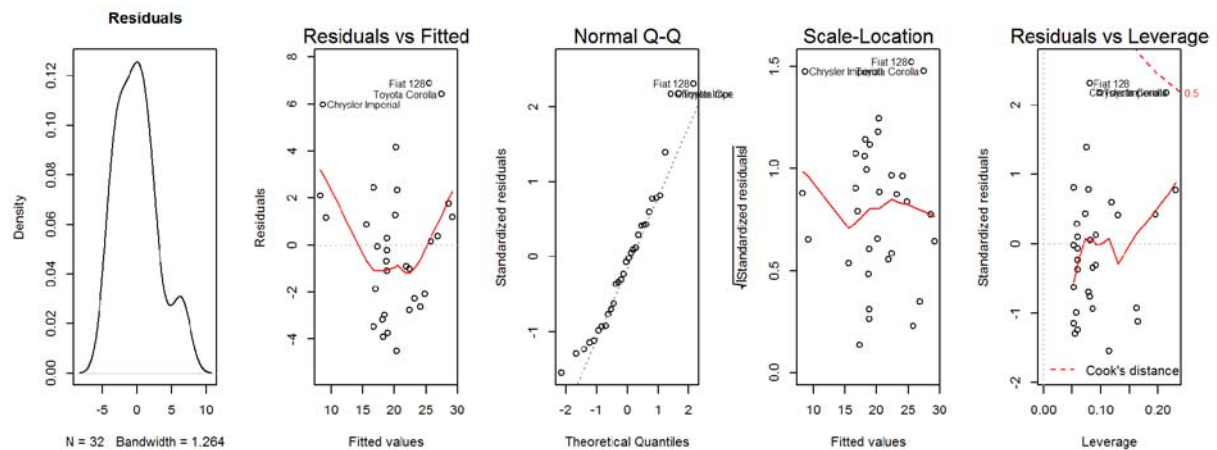
Fig. 8

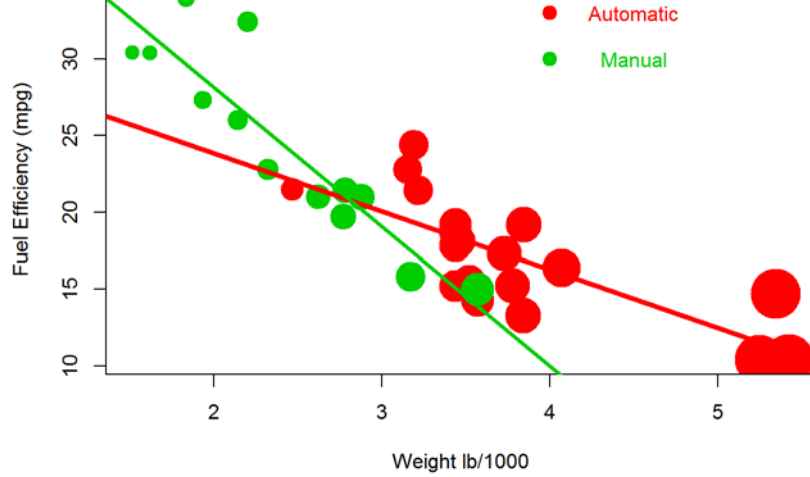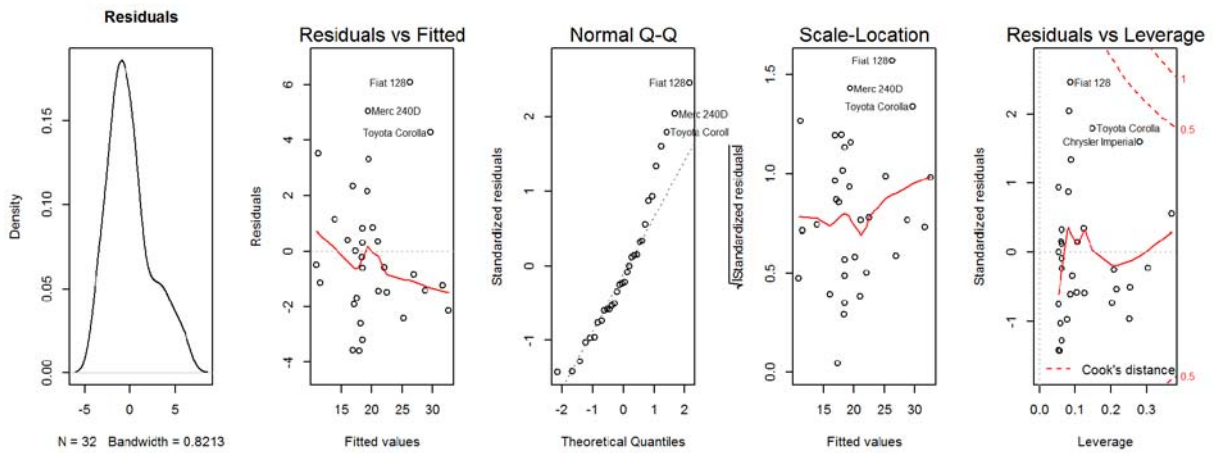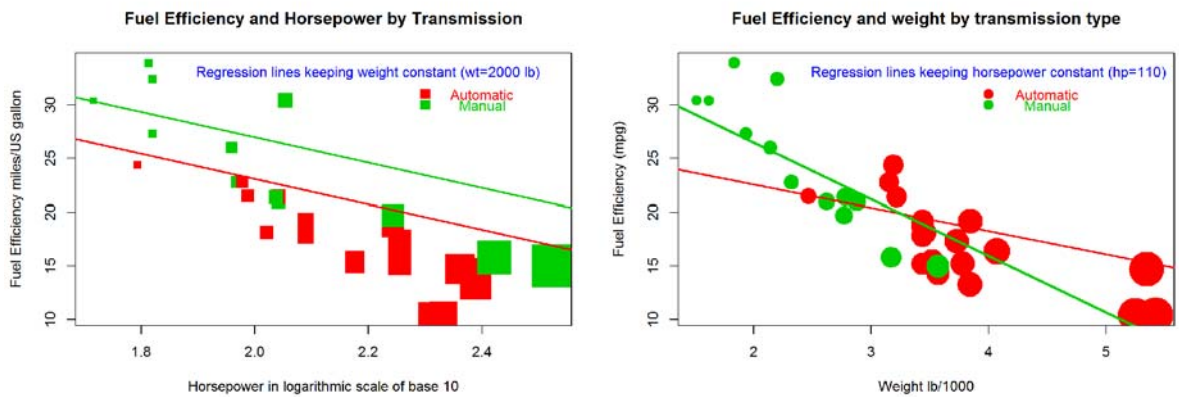## Fuel Efficiency and Weight by Transmission with Interaction

Fig. 9

Fig. 10

Fig. 11