# The Relationship Between Fuel Economy and Transmission Type

*May 10, 2015*

## Executive Summary

This project involves using the `mtcars` data in the **R** `datasets` package to explore the relationship between fuel economy and type of transmission (automatic vs. manual). The two questions of interest are: (a) is an automatic or manual transmission better for miles per gallon (MPG), and (b) quantify the MPG difference between automatic and manual transmissions.

## Exploratory Data Analysis

The `mtcars` data was extracted from the 1974 *Motor Trend* US magazine, and consists of fuel consumption and 10 aspects of automobile design and performance for 32 automobiles.

Figure 1 shows a boxplot of MPG by number of cylinders (a categorical proxy for the size of the car) by transmission type. The boxplot shows that manual transmissions have a higher MPG value than automatic transmissions, but the difference is greatest for 4-cylinder cars. Further exploratory analysis in Figure 2 shows that predictors that are negatively correlated with MPG are all indicators of a car's power or weight (number of cylinders, engine displacement, horsepower, number of carburators, weight). These predictors are also very highly intercorrelated, which suggests that multicollinearity will be a concern when fitting a regression model. To address this concern, a power-to-weight ratio variable was computed by dividing a car's horsepower by its weight. Figure 3 shows the predictors that are positively correlated with MPG.

## Regression Models

To explore the relationship between fuel economy and type of transmission while holding other variables constant, start by fitting a linear regression model using power-to-weight ratio and transmission type as predictor variables and MPG as the criterion variable. Then update the model by adding the remaining predictor variables in decreasing order of their correlation with MPG (drive ratio, engine shape, number of transmission speeds, and quarter mile time). The analysis of variance comparing the five models suggests that Model 3 is the final model, as no further statistically significant decrease in RSS is observed.

```
## Analysis of Variance Table
##
## Model 1: mpg ~ p2w + am
## Model 2: mpg ~ p2w + am + drat
## Model 3: mpg ~ p2w + am + drat + vs
## Model 4: mpg ~ p2w + am + drat + vs + gear
## Model 5: mpg ~ p2w + am + drat + vs + gear + qsec
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1     29 474.06
## 2     28 420.32  1    53.746 4.5296 0.043350 *
## 3     27 305.92  1   114.401 9.6415 0.004684 **
## 4     26 305.42  1     0.497 0.0419 0.839513
## 5     25 296.64  1     8.784 0.7403 0.397733
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

However, the Lotus Europa has a Cook's distance greater than 4 divided by the number of observations, suggesting that this data point may have a high influence on the regression model and should be examined more closely.

```
## Lotus Europa
##        0.527
```

Figure 4 shows that the slope of the regression line changes drastically depending on whether or not the Lotus Europa is included in the model. This data point has been removed from the final model, which is shown below. The diagnostic plots in Figure 5 do not give reason to suggest that the assumptions of the model (normality, linearity, homoscedasticity) have been violated.

```
##
## Call:
## lm(formula = mpg ~ p2w + am + drat + vs, data = df.car2)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -5.293 -1.919  0.414  1.501  6.410
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 13.00568    5.94080   2.189  0.03775 *
## p2w         -0.13390    0.04763  -2.811  0.00926 **
## amManual     5.01900    1.75387   2.862  0.00821 **
## drat         2.63258    1.72583   1.525  0.13923
## vsStraight   3.09877    1.61857   1.915  0.06662 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.09 on 26 degrees of freedom
## Multiple R-squared:  0.7557, Adjusted R-squared:  0.7181
## F-statistic:  20.1 on 4 and 26 DF,  p-value: 1.196e-07
```

Looking at the confidence intervals below, we can say with 95% confidence that having a manual transmission results in an estimated increase in fuel economy between 1.4 and 8.6 MPG, holding other predictors constant.

```
##    2.5 %   97.5 %
## 1.413873 8.624133
```

Finally, the variance inflation factors are all less than 4, suggesting that multicollinearity is not present in the final model.

```
##      p2w       am     drat       vs
## 1.742241 2.368918 2.754251 2.070603
```

## Conclusions

The findings of the current project suggest that: (a) a manual transmission is better for MPG than an automatic, but the difference is greatest for 4-cylinder cars, and (b) having a manual rather than an automatic transmission results in an estimated increase in fuel economy between 1.4 and 8.6 MPG.

# Appendix

**Figure 1: MPG by Number of Cylinders and Transmission Type**

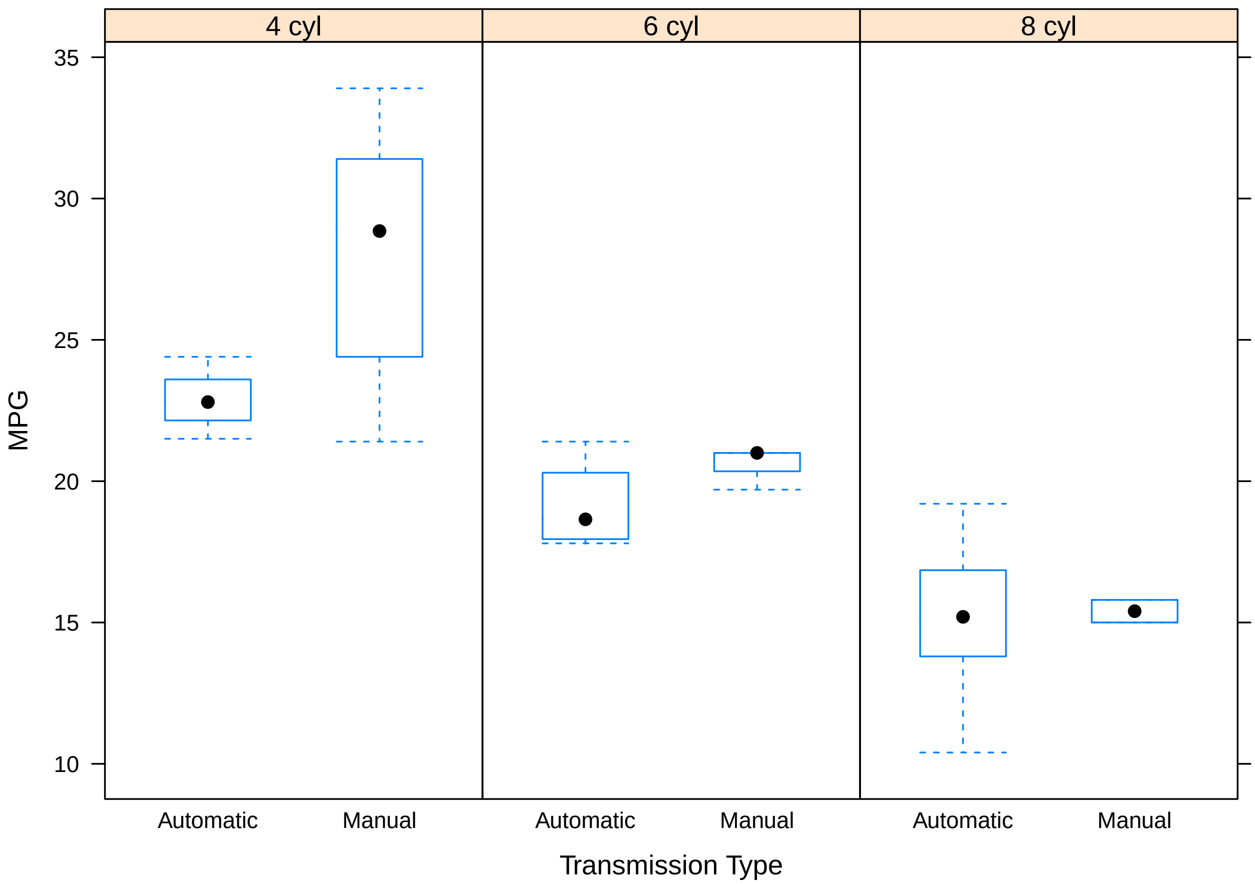


**Figure 2: Predictors Negatively Correlated With MPG**

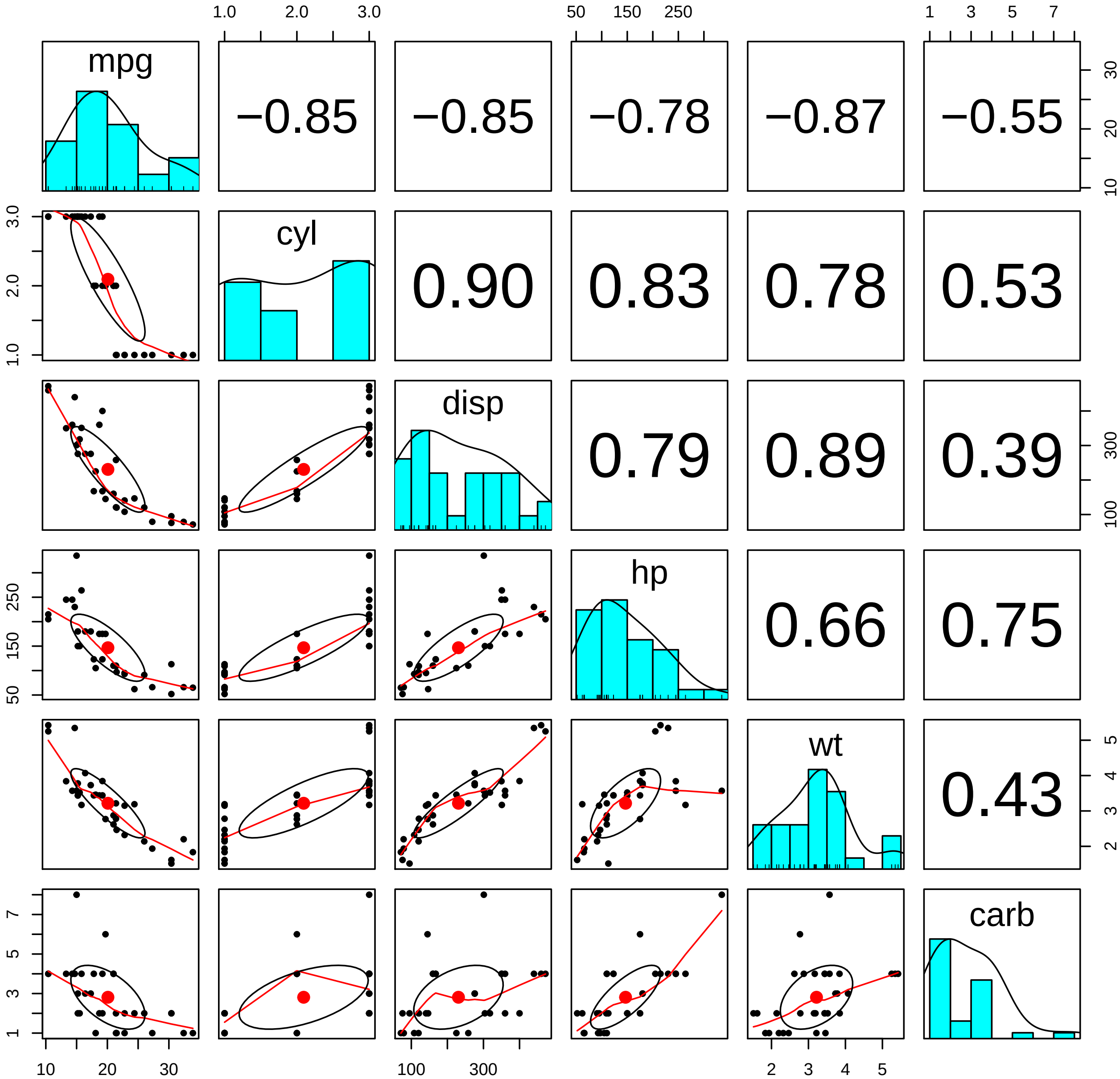
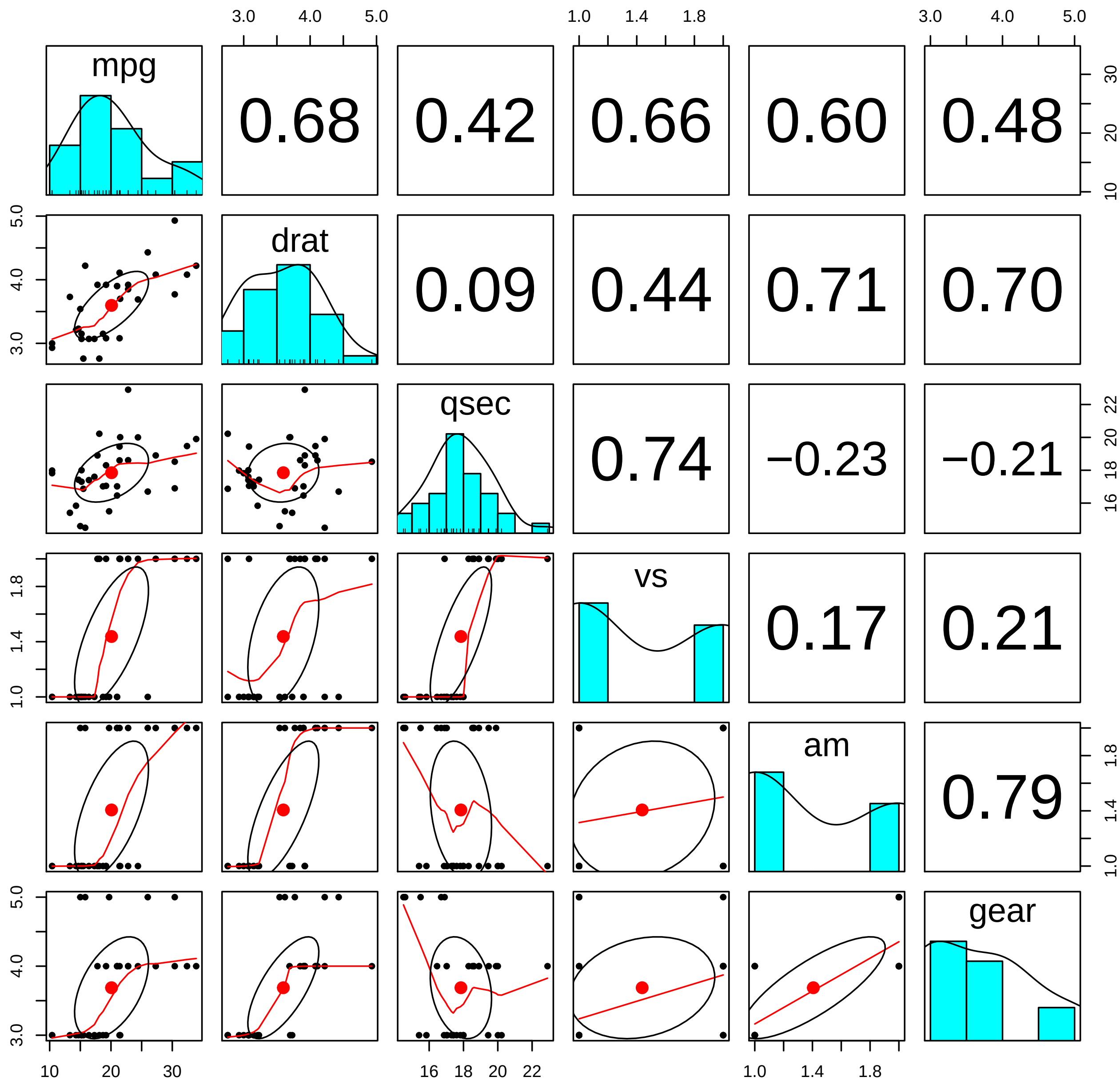


**Figure 3: Predictors Positively Correlated With MPG**



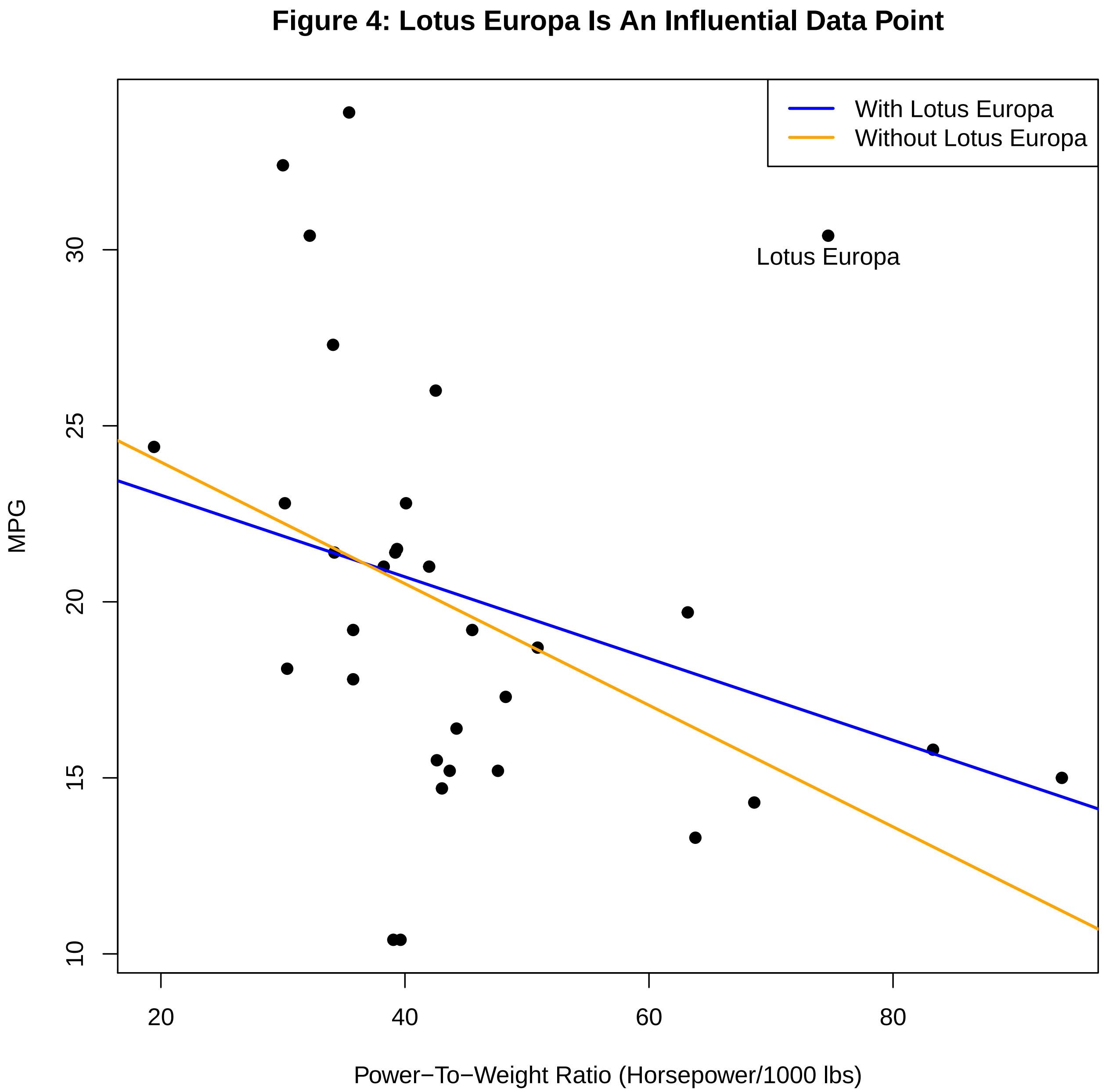


Figure 4: Lotus Europa Is An Influential Data Point
With Lotus Europa
Without Lotus Europa
Lotus Europa
MPG
10
15
20
25
30
20
40
60
80
Power−To−Weight Ratio (Horsepower/1000 lbs)

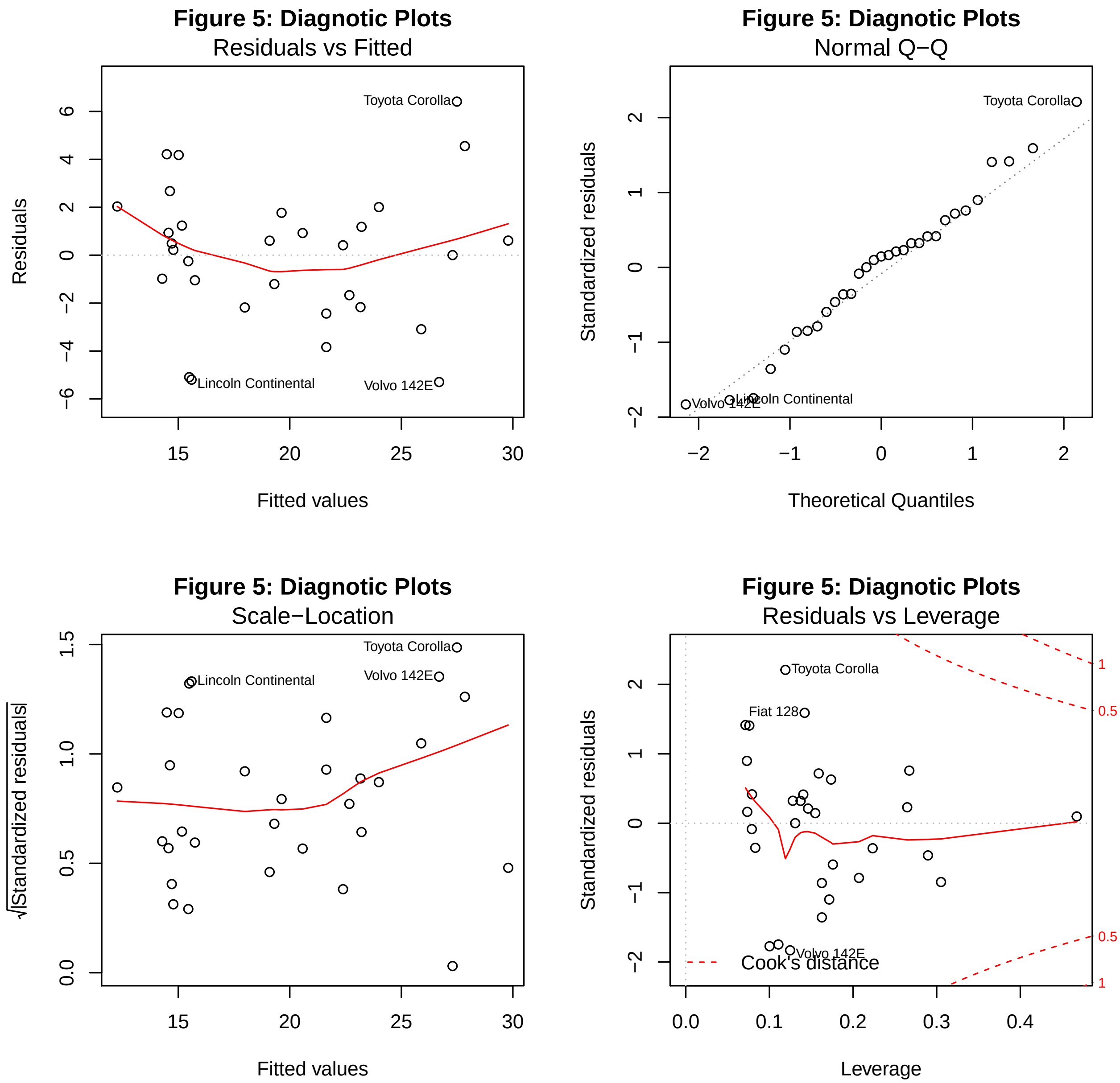


Figure 5: Diagnotic Plots
Residuals vs Fitted
Residuals
Fitted values
Toyota Corolla
Lincoln Continental
Volvo 142E
Figure 5: Diagnotic Plots
Normal Q−Q
Standardized residuals
Theoretical Quantiles
Toyota Corolla
Volvo 142E
Lincoln Continental
Figure 5: Diagnotic Plots
Scale−Location
√|Standardized residuals|
Fitted values
Toyota Corolla
Lincoln Continental
Volvo 142E
Figure 5: Diagnotic Plots
Residuals vs Leverage
Standardized residuals
Leverage
Toyota Corolla
Fiat 128
Volvo 142E
Cook's distance