Analysis of data (mtcars) using linear regression model

Samiul Azam

January 8, 2017

Executive summary

In this project I do exploratory data analysis of the 'mtcars' dataset to find answer of two target questions: (a) Is an automatic or manual transmission better for MPG? (b) Quantify the MPG difference between automatic and manual transmissions. Moreover, regression analysis has been applied to retrieve the relation between mpg with different car factors (weight, displacement, qsec, drat, horsepower). Inferential statistics is applied to figureout uncertainity.

Exploratory analysis of data

```r
data(mtcars) # Loading the data
summary(mtcars$mpg)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10.40   15.42   19.20   20.09   22.80   33.90
```

```r
summary(mtcars$wt)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.513   2.581   3.325   3.217   3.610   5.424
```

The data contains 32 observations. Each observation consists of 11 variables. Basic statistical measures of few important variables (mpg, wt) are provided above. In Figure 1, we see inverse relation between the MPG (mile per gallon) and wt (car weight). That means more weight means less MPG.

Answers of the questions

Figure 2 shows the scatter plot of 'mpg' for 'am' (automatic = 0 and manual = 1). From the figure, it seems to be very clear that cars with automatic transmission shows less mpg than the manual. The mean mpg for automatic transmission is 17.15 whereas manual transmission shows 24.39. However, a car's MPG depends on other factors also, such as weight, disp, number of cylinders, hp. We need car data where all other configuration are same except the transmission.

Linear regression and interpretation of the coefficients

Consider linear regression considering the variables mpg as response and wt + qsec as predictor.

```r
fit <- lm(mpg ~ wt + qsec, data = mtcars)
summary(fit)$coefficients
```

```
##               Estimate Std. Error    t value     Pr(>|t|)
## (Intercept) 19.746223  5.2520617   3.759709 7.650466e-04
## wt          -5.047982  0.4839974 -10.429771 2.518948e-11
## qsec         0.929198  0.2650173   3.506179 1.499883e-03
```

In the above we see the coefficients for the predictors. The interpretation is as follows: The model estimates an expected 5.05 decrease (negative) in standardized mpg for per unit (1000 lbs) increase in car weight holding the remaining variables as constant. The model estimates an expected 0.93 increase (positive) in standardized mpg for per unit (quarter mile time) increase in car qsec holding the remaining variables as constant.

Model Selection

For model selection, I apply ANOVA over different models.Following we see the output of the anova function in R. It shows all the models that i select in this analysis.

```
## Analysis of Variance Table
##
## Model 1: mpg ~ cyl
## Model 2: mpg ~ cyl + disp
## Model 3: mpg ~ cyl + disp + hp
## Model 4: mpg ~ cyl + disp + hp + drat
## Model 5: mpg ~ cyl + disp + hp + drat + wt
##   Res.Df    RSS Df Sum of Sq       F   Pr(>F)
## 1     30 308.33
## 2     29 270.74  1    37.594  5.8381 0.023005 *
## 3     28 261.37  1     9.371  1.4552 0.238553
## 4     27 244.90  1    16.467  2.5573 0.121871
## 5     26 167.43  1    77.476 12.0314 0.001837 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From above, we say that the model 5 is the best among them as it shows minimum RSS and very low P-value.

Residual Plots and diagnosis

If we see the residual plots (residual vs fitted value, scale location, residual vs leverage) of model 5, there is no systemmatic pattern inside them (more scattered and random). Moreover, most of the points fall on the line of normal Q-Q plot. So the model performs well enough.

Quantify uncertainity

I use t.test to find the confidence interval in the difference between mpg of auto and manual transmission.

```
## [1]  3.209684 11.280194
## attr(,"conf.level")
## [1] 0.95
```

I found that the confidence interval doesnt contain the zero (allways positive). That means the population mean of mpg in manual transmission is always higher.

Appendix of Figures

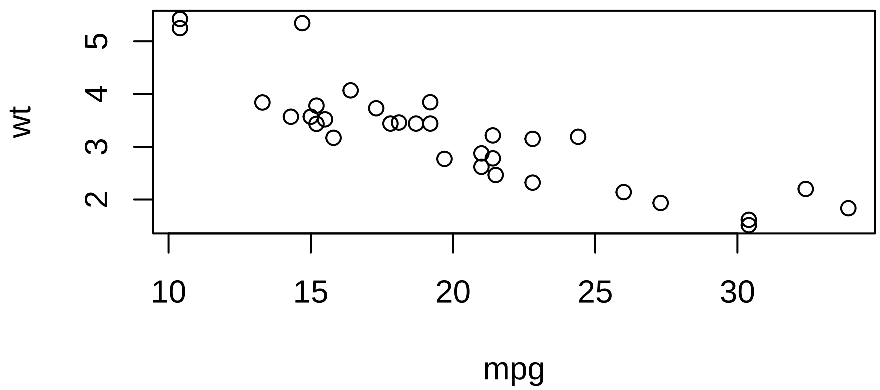

Figure 1: Scatter plot of mpg (Miles per gallon) and wt (car weight).

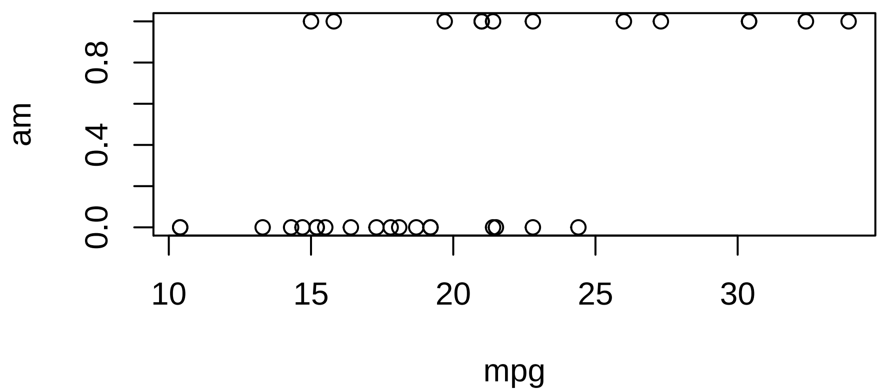

Figure 2: Scatter plot of mpg (Miles per gallon) and am (auto or manual). 0 means automatic and 1 means manual transmission.

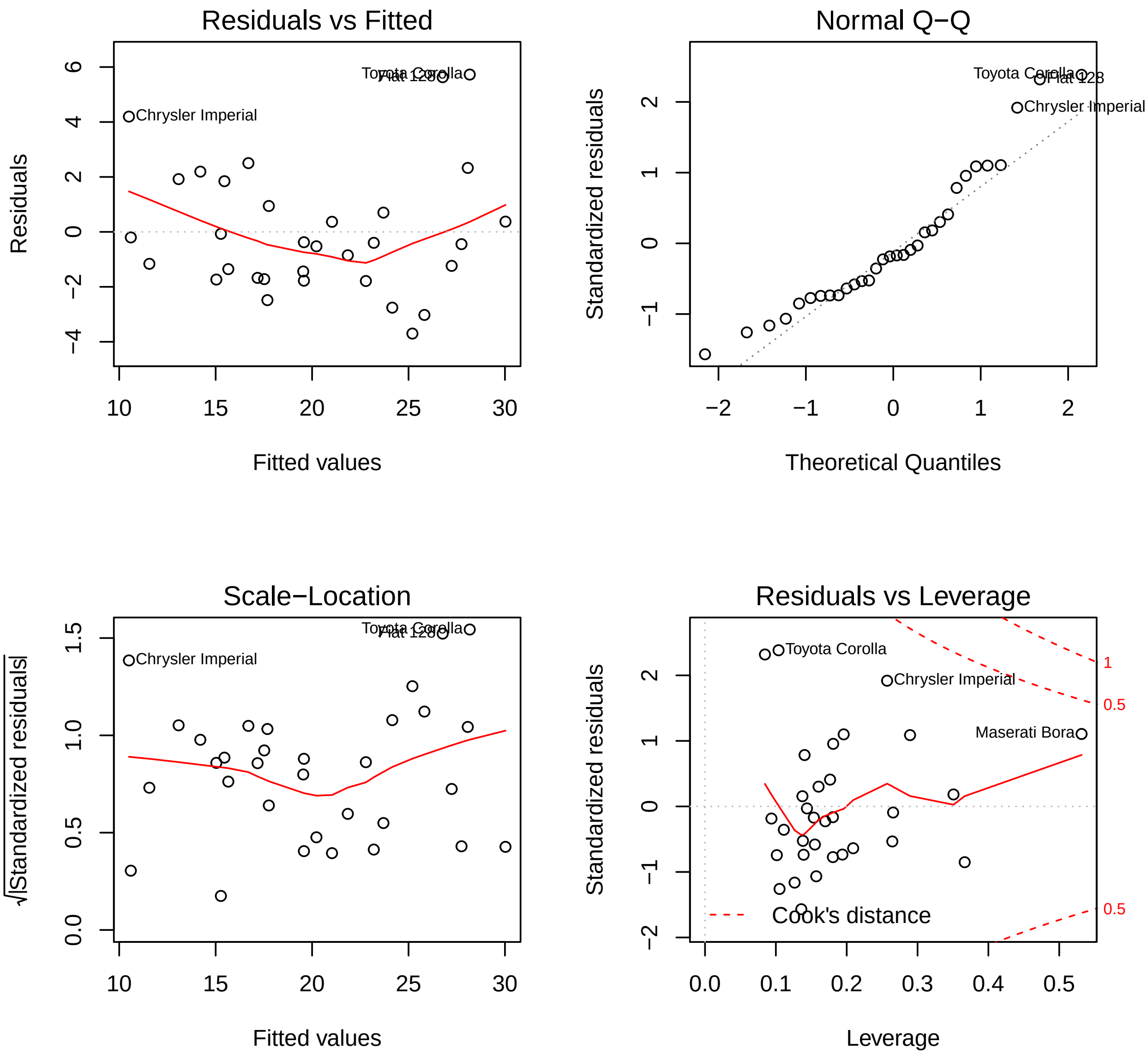

Figure 3: Residual plots.