Regression Models Course Project

Abhishek Kumar

22 September 2020

Data Processing

In `mtcars` dataset, the variable `mpg` represents the Miles per (US) gallon and the `am` variable represents the tranmission where 0 represent the *automatic* and 1 represent the *manual* transmission.

```
library(datasets); library(ggplot2)
data("mtcars"); dat <- mtcars; str(mtcars)
dat$cyl <- as.factor(dat$cyl); dat$vs <- as.factor(dat$vs)
dat$am <- as.factor(dat$am); dat$gear <- as.factor(dat$gear)
dat$carb <- as.factor(dat$carb)
```

Exploratory Analysis

```
t.test(mpg~am, dat)$p.value
```

```
## [1] 0.001373638
```

The t-test suggests that data provide sufficient evidence in favour of alternative hypothesis i.e. there are significant differences as p-value (0.001) is less than 0.05.

Regression Analysis

- **Model Selection:** The `step()` function in R uses the AIC for selection of best models i.e. with higher AIC values.Here, I will use a backward model selection strategy because it will prevent loss of any variable as we will start from a model which include all the predictors.

```
mdl.full <- lm(mpg~., data = dat)
mdl.step <- step(mdl.full, direction = "backward", trace = FALSE)
summary(mdl.step)$coef
```

```
##                Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 33.70832390 2.60488618 12.940421 7.733392e-13
## cyl6        -3.03134449 1.40728351 -2.154040 4.068272e-02
## cyl8        -2.16367532 2.28425172 -0.947214 3.522509e-01
## hp          -0.03210943 0.01369257 -2.345025 2.693461e-02
## wt          -2.49682942 0.88558779 -2.819404 9.081408e-03
## am1          1.80921138 1.39630450  1.295714 2.064597e-01
```

- **Coeffiecient Interpretation:** *This `mdl.step` model suggests that If all other variables are held constant, the `mpg` is expected to be 1.809 times higher for manual tranmission (`am1`) than the automatic transmission (`am0`) on average.*

- **Quantification of Uncertainity:**

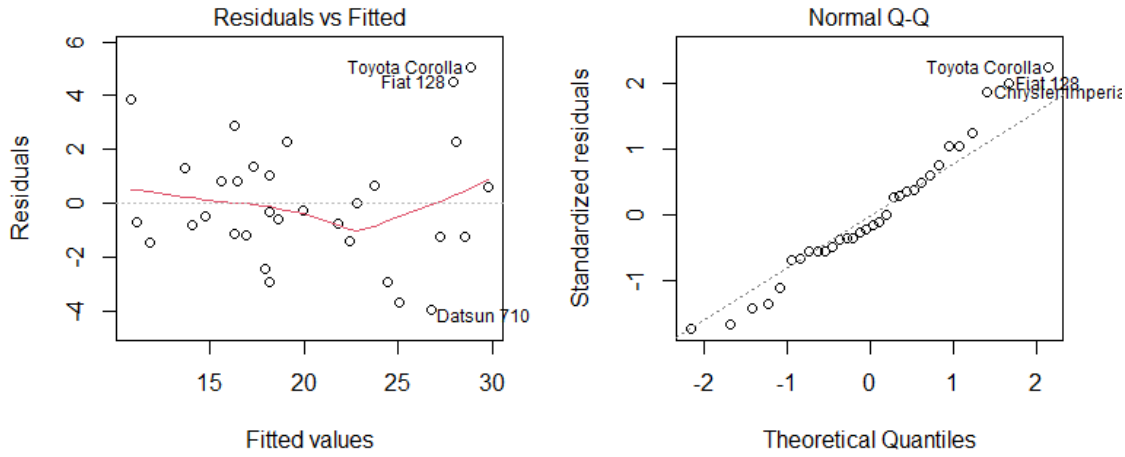
```
confint(mdl.step, parm = "am1", level = 0.95)
```

```
##          2.5 %   97.5 %
## am1 -1.060934 4.679356
```

Thus, our step model (`mdl.step`) suuggests that If all other variables are held constant,the `mpg` is expected to be 1.061 times lower to 4.679 times higher for manual tranmission (`am1`) than the automatic transmission (`am0`) on average at 95% significance level. since this is a relatively broader interval, the uncertainity in prediction can be moderate to high.

- **Model Diagnostics:**

```
par(mfrow = c(1, 2)); plot(mdl.step, which = 1:2)
```



The above plots illustrates the model assumptions. The residuals vs Fitted plot suggests moderate linear association as 3 points can be considered as ouliers. The assumption of normaility is also fairly met since some points at the upper end are way from the normal line.

Executive Summary

The miles per gallon (mpg) is an important parameter of fuel economy. Our analysis suggests that mpg is associated with number of cylinder (cyl), engine capacity horse power (`hp`), weight (`wt`) and transmission type (`am`). Although the transmission type is not a significant predictor of `mpg`, it has influence on `mpg`. Statistical evidence supports that manual tranmission is more likely to have higher `mpg` on average, however, there is moderate to large uncertainity in the coefficients suggesting that predictions may not be accurate.