mtcars - Regression Exercise

Quick Inspection of Data Set

Firstly we have a quick loot at the data set to set a sense of the size of the data set, and the names of the variables

```
tail(mtcars)
```

```
##                 mpg cyl  disp  hp drat    wt qsec vs am gear carb
## Porsche 914-2  26.0   4 120.3  91 4.43 2.140 16.7  0  1    5    2
## Lotus Europa   30.4   4  95.1 113 3.77 1.513 16.9  1  1    5    2
## Ford Pantera L 15.8   8 351.0 264 4.22 3.170 14.5  0  1    5    4
## Ferrari Dino   19.7   6 145.0 175 3.62 2.770 15.5  0  1    5    6
## Maserati Bora  15.0   8 301.0 335 3.54 3.570 14.6  0  1    5    8
## Volvo 142E     21.4   4 121.0 109 4.11 2.780 18.6  1  1    4    2
```

To find out more about this data set, simpy type `help(mtcars)` to get the help file.

Exercise 1 : Simple Linear Regression

In this exercise, fit a simple linear model, where weight (***wt***) is the independent variable, and miles per gallon (***mpg***) is the dependent variable. Call the fitted model `Fit1`.

```
Fit1 = lm(mpg ~ wt,data=mtcars)
Fit1
```

```
##
## Call:
## lm(formula = mpg ~ wt, data = mtcars)
##
## Coefficients:
## (Intercept)           wt
##      37.285       -5.344
```

More information about the model can be found using the `summary()` command.

```
summary(Fit1)
```

```
##
## Call:
## lm(formula = mpg ~ wt, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5432 -2.3647 -0.1252  1.4096  6.8727
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  37.2851     1.8776  19.858  < 2e-16 ***
## wt           -5.3445     0.5591  -9.559 1.29e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.046 on 30 degrees of freedom
## Multiple R-squared:  0.7528, Adjusted R-squared:  0.7446
## F-statistic: 91.38 on 1 and 30 DF,  p-value: 1.294e-10
```

Exercise 2 : Multiple Linear Regression

In this exercise, fit a simple linear model, where weight (***wt***) and number of cylinders (***cyl*** are the independent variable, and miles per gallon (***mpg***) is the dependent variable. Call the fitted model `Fit2`.

```
Fit2 = lm(mpg ~ wt+cyl,data=mtcars)
Fit2
```

```
##
## Call:
## lm(formula = mpg ~ wt + cyl, data = mtcars)
##
## Coefficients:
## (Intercept)           wt          cyl
##      39.686       -3.191       -1.508
```

```
summary(Fit2)
```

```
##
## Call:
## lm(formula = mpg ~ wt + cyl, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.2893 -1.5512 -0.4684  1.5743  6.1004
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  39.6863     1.7150  23.141  < 2e-16 ***
## wt           -3.1910     0.7569  -4.216 0.000222 ***
## cyl          -1.5078     0.4147  -3.636 0.001064 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.568 on 29 degrees of freedom
## Multiple R-squared:  0.8302, Adjusted R-squared:  0.8185
## F-statistic: 70.91 on 2 and 29 DF,  p-value: 6.809e-12
```

Exercise 3 : Confidence Intervals for Regression Estimates

Using the regression model from the previous exercise (i.e. `Fit2`), compute 95% and 99% confidence intervals for the regression estimates.

```
confint(Fit2)
```

```
##                 2.5 %     97.5 %
## (Intercept) 36.178725 43.1937976
## wt          -4.739020 -1.6429245
## cyl         -2.355928 -0.6596622
```

```
confint(Fit2, conf.level=0.99)
```

```
##                 2.5 %     97.5 %
## (Intercept) 36.178725 43.1937976
## wt          -4.739020 -1.6429245
## cyl         -2.355928 -0.6596622
```

Exercise 4 : Akaike Information Criterion

When comparing two or more candidate models, the model with lowest AIC value is considered the best model. Use the AIC to compare `Fit1` and `Fit2`.

```
AIC(Fit1)
```

```
## [1] 166.0294
```

```
AIC(Fit2)
```

```
## [1] 156.0101
```

Exercise 5 : ANOVA tables for Regression Models

Construct the ANOVA tables for the fitted models computed previously.

```
anova(Fit1)
```

```
## Analysis of Variance Table
##
## Response: mpg
##           Df Sum Sq Mean Sq F value    Pr(>F)
## wt         1 847.73  847.73  91.375 1.294e-10 ***
## Residuals 30 278.32    9.28
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(Fit2)
```

```
## Analysis of Variance Table
##
## Response: mpg
##           Df Sum Sq Mean Sq F value    Pr(>F)
## wt         1 847.73  847.73  128.60 3.535e-12 ***
## cyl        1  87.15   87.15   13.22  0.001064 **
## Residuals 29 191.17    6.59
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```