# STP598-Assignment 1

*Hao Wang*
*hwang306*

*September 13, 2017*

## 1 Question 1

Load the data first, with a brief inspection.

```
mydata <- read.csv("https://raw.githubusercontent.com/haowang666/Computational-Stats/mas
summary(mydata)
```

```
##      HtVol             Male              CT              Age
##  Min.   : 112.2   Min.   :0.0000   Min.   :0.000   Min.   : 13.0
##  1st Qu.: 340.7   1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:110.5
##  Median : 539.9   Median :1.0000   Median :1.000   Median :174.5
##  Mean   : 535.5   Mean   :0.6034   Mean   :0.569   Mean   :156.8
##  3rd Qu.: 680.7   3rd Qu.:1.0000   3rd Qu.:1.000   3rd Qu.:203.2
##  Max.   :1340.2   Max.   :1.0000   Max.   :1.000   Max.   :359.0
##        Ht               Wt              BMI             BSA
##  Min.   : 71.0   Min.   :  7.90   Min.   :13.51   Min.   :0.390
##  1st Qu.:128.2   1st Qu.: 28.85   1st Qu.:16.62   1st Qu.:1.005
##  Median :157.2   Median : 54.30   Median :21.37   Median :1.550
##  Mean   :148.0   Mean   : 53.60   Mean   :22.72   Mean   :1.458
##  3rd Qu.:165.7   3rd Qu.: 71.90   3rd Qu.:27.07   3rd Qu.:1.810
##  Max.   :186.0   Max.   :139.20   Max.   :46.51   Max.   :2.590
```

### 1.1 1.a Predictive model

I build a simple linear model first

$$\hat{HtVol} = \beta_0 + \beta_1 Male + \beta_2 Age + \beta_3 Ht + \beta_4 Wt$$

Use `lm` function to obtain the LS coefficients

```
lm <- lm(HtVol ~ Male + Age + Ht + Wt, data = mydata)
summary(lm)
```

```
##
## Call:
```

```
## lm(formula = HtVol ~ Male + Age + Ht + Wt, data = mydata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -209.56  -57.02    2.58   41.83  199.03
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -263.6665    90.2773  -2.921  0.00512 **
## Male          41.3395    23.1534   1.785  0.07991 .
## Age           -0.3373     0.3797  -0.888  0.37835
## Ht             3.3829     0.9713   3.483  0.00100 **
## Wt             6.0891     0.6795   8.962 3.37e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 82.28 on 53 degrees of freedom
## Multiple R-squared:  0.8904, Adjusted R-squared:  0.8821
## F-statistic: 107.6 on 4 and 53 DF,  p-value: < 2.2e-16
```

With the `lm` function, the fitting line is

$$\hat{HtVol} = -263.67 + 41.34 * Male - 0.34 * Age + 3.38 * Ht + 0.68 * Wt$$

I use package 'L1pack' to perform least absolute deviation regression

Unlike Least Squares, Least Absolute Deviation minimize

$$S = \sum_{i=1}^{n} |y_i - f(x_i)|$$

```
library("L1pack")
lad  <-  lad(HtVol ~ Male + Age + Ht + Wt, data = mydata)
summary(lad)
```

```
## Call:
## lad(formula = HtVol ~ Male + Age + Ht + Wt, data = mydata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -133.32  -37.07    0.00   36.72  339.75
##
## Coefficients:
##             Estimate  Std.Error Z value   p-value
## (Intercept) -213.7607    0.7758 -275.5348    0.0000
```

```
## Male          48.1662    0.1990  242.0777    0.0000
## Age            0.1834    0.0033   56.2063    0.0000
## Ht             2.9940    0.0083  358.7031    0.0000
## Wt             4.4133    0.0058  755.8221    0.0000
##
## Degrees of freedom: 58 total; 53 residual
## Scale estimate: 80.61168
## Log-likelihood: -332.7006 on 6 degrees of freedom
```

The fitting line is

$$\hat{HtVol} = -213.76 + 48.17 * Male + 0.18 * Age + 2.99 * Ht + 4.41 * Wt$$

## 1.2 1.b

Steps are very similar, just change the explainatory variables

```
lm <- lm(HtVol ~ Male + Age + BMI + BSA, data = mydata)
summary(lm)
```

```
##
## Call:
## lm(formula = HtVol ~ Male + Age + BMI + BSA, data = mydata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -184.474  -51.125   -2.295   35.538  265.924
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -121.9456    41.7940  -2.918  0.00516 **
## Male          37.0077    24.0400   1.539  0.12965
## Age           -0.6737     0.3966  -1.699  0.09520 .
## BMI           -5.2994     2.9600  -1.790  0.07912 .
## BSA          590.6002    73.5415   8.031 9.99e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 84.36 on 53 degrees of freedom
## Multiple R-squared:  0.8848, Adjusted R-squared:  0.8761
## F-statistic: 101.8 on 4 and 53 DF,  p-value: < 2.2e-16
```

The linear fitting is

$$\hat{HtVol} =$$