Matrix Approach

- **Simple Linear Regression Model.** The least square regression line was obtained by solving the following normal equations:

$$
\begin{aligned}
nb_0 + b_1 \sum X_i &= \sum Y_i \\
b_0 \sum X_i + b_i \sum X_i^2 &= \sum X_i Y_i
\end{aligned}
$$

in matrix terms, these 2 equations can be expressed as: $\boldsymbol{X'Xb} = \boldsymbol{X'Y}$.

Hence, the solution is given by, $\boldsymbol{b} = \boldsymbol{(X'X)^{-1}X'Y}$.

- **Properties:**

  1. Fitted values: $\hat{\boldsymbol{Y}} = \boldsymbol{Xb}$

  2. Hat Matrix: $\boldsymbol{H} = \boldsymbol{X(X'X)^{-1}X'} = \boldsymbol{H^2}$ $\quad\Rightarrow \hat{\boldsymbol{Y}} = \boldsymbol{HY}$

  3. Residuals: $\boldsymbol{e} = \boldsymbol{Y} - \hat{\boldsymbol{Y}}$ $\quad\Rightarrow \boldsymbol{e} = \boldsymbol{Y} - \boldsymbol{HY} = (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y}$.

  4. Variance of Error Terms: $\boldsymbol{\sigma^2(e)} = \sigma^2(\boldsymbol{I} - \boldsymbol{H})$ $\quad\Rightarrow \boldsymbol{s^2(e)} = MSE(\boldsymbol{I} - \boldsymbol{H})$.

- Matrix calculations in R

```
A=matrix(1:6,nrow=2)
B=matrix(c(-1,0,2,1,4,3),nrow=2)
A
B
A+B   # Matrix addition = addition of corresponding entries
A-B
t(A) # transpose of A
D=t(A)%*%A # To do matrix multiplication, use %*%
E=A%*%t(A) # Note that E is not equal to D.
E.inv=solve(E) # Computing the inverse of a square matrix
I=E.inv%*%E # The product of E and it's inverse is the Identity matrix
```

- Consider the Toluca Company example. Obtain the least-square regression line using the matrix approach.

```
data=read.csv("Toluca.csv",header=T)
attach(data)
x=size; y=hours
n=length(x)
t(y)%*%y # Sum of squares of yi's
X=matrix(c(rep(1,n),x),ncol=2) # The design matrix X
t(X)%*%X
t(X)%*%y
solve(t(X)%*%X)
B=solve(t(X)%*%X)%*%(t(X)%*%y) # Estimates of the regression coefficients
coefficients(lm(y~x)) # Note that you get these results

fitted=X%*%B
```