Week 1 Notes in RegMod – Least Square

```
> x <- c(3,6,9,12, 15)

> mean(x)
[1] 9

> x- mean(x)
[1] -6 -3  0  3  6 #Data is centered with 0 as middle

> (x- mean(x))^2
[1] 36  9  0  9 36 #Squaring the centered data
                   increases the spread but the middle
                   is still 0
> sum(x)
 [1] 45            #Sum of the original data

> sum(x- mean(x))
[1] 0              #Sum of centered data is 0

> sum((x- mean(x))^2)
[1] 90             #Sum of the squared centered data is
                   90
```

Let's replace the mean with something less

```
> mean_less <- mean(x)-2

> mean_less
[1] 7

> x- mean_less
[1] -4 -1  2  5  8

> (x- mean_less)^2
[1] 16  1  4 25 64

> sum(x)
 [1] 45

> sum(x- mean_less)
[1] 10

> sum((x- mean_less)^2)
[1] 110
```

Original data

```
> mean(x)
[1] 9

> x- mean(x)
[1] -6 -3  0  3  6


> (x- mean(x))^2
[1] 36  9  0  9 36




> sum(x- mean(x))
[1] 0

> sum((x- mean(x))^2)
[1] 90
```

```
> mean_more <- mean(x)+2

> mean_more
[1] 11

> x - mean_more
[1] -8 -5 -2  1  4

> (x - mean_more)^2
[1] 64 25  4  1 16

> sum(x - mean_more)
[1] -10

> sum((x - mean_more)^2)
[1] 110
```

```
> mean_less <- mean(x)-2

> mean_less
[1] 7

> x- mean_less
[1] -4 -1  2  5  8

> (x- mean_less)^2
[1] 16  1  4 25 64

> sum(x- mean_less)
[1] 10

> sum((x- mean_less)^2)
[1] 110
```

When you get a number greater than the mean as the middle the sum is negative: When you ge a number less than the mean as the middle the sum is positive.

Compared to the real mean, a larger or smaller number that replaces the mean as the middle results in a larger squared value

Least Square

```
> sum((x- mean(x))^2)
[1] 90

> sum((x - mean_more)^2)
[1] 110

> sum((x- mean_less)^2)
[1] 110
```

Mean Squared Error / Sum Squared Error

slope = 0.64

```
> yc <- galton$child - mean(galton$child)
> xc <- galton$parent - mean(galton$parent)
> beta <- 0.64
> mse <- mean((yc - beta * xc)^2 )
> mse
[1] 5.00042
> sse <- sum((yc - beta * xc)^2 )
> sse
[1] 4640.39
```

slope > 0.64

```
> beta <- 0.68
> mse <- mean((yc - beta
* xc)^2 )
> mse
[1] 5.00392
> sse <- sum((yc - beta *
xc)^2 )
> sse
[1] 4643.638
```

```
> beta <- 0.74
> mse <- mean((yc - beta
* xc)^2 )
> mse
[1] 5.028316
> sse <- sum((yc - beta *
xc)^2 )
> sse
[1] 4666.278
```

slope < 0.64

```
> beta <- 0.60
> mse <- mean((yc - beta
* xc)^2 )
> mse
[1] 5.007132
> sse <- sum((yc - beta *
xc)^2 )
> sse
[1] 4646.618
```

```
> beta <- 0.54
> mse <- mean((yc - beta
* xc)^2 )
> mse
[1] 5.036346
> sse <- sum((yc - beta *
xc)^2 )
> sse
[1] 4673.729
```

mean squared error
is closely related to the
y-intercept <- mean(yc) - beta1c * mean(xc)