Problem Set I

student ID : JinWoo Kim

Question 4, 5

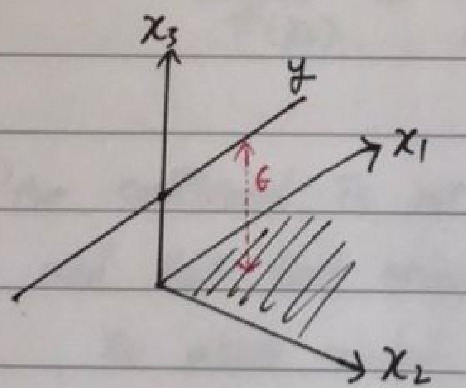
⟨Question 4⟩

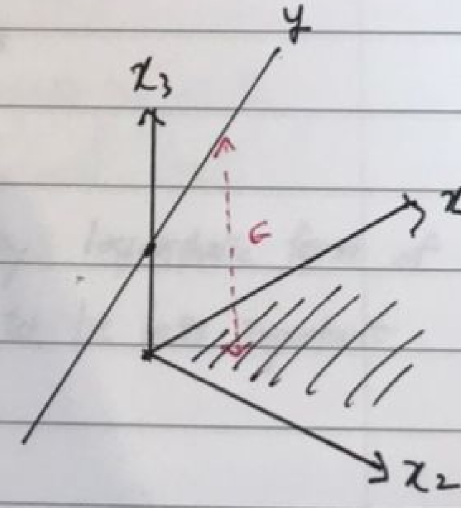
(1) Suppose, the multiple linear regression has only 3 explanatory variables.

ex) $y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$.

In a vector space, when a particular explanatory variable's unit changed, (thousands Kgs to millions Kgs) It means multyplying a variable by 1000.

Slope of the variable is changing drastically.





∴ the residuals will be bigger

Also, increase in unit means that variance of x_3 (observed data) is bigger. It affects that difference between y_i and predicted value is bigger.

(2) Suppose, the multiple linear regression model is

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$$

the coefficient of x_3 will be in logarithmic form.

$$y = \beta_1 x_1 + \beta_2 x_3 + \log(\beta_3) x_3 + \epsilon$$

when we want to know if other coefficient would be changed by β_3, we can get it using partial derivitive.

$$\frac{\partial(y)}{\partial(\beta_3)} = \frac{1}{\beta_3 \cdot \ln 10} \cdot x_3$$

Other coefficiend is not affected by logarithmic form of a coefficie[nt] and that form encourages x_3 to be less important to above model.

<Question 5>

(1) If the true relationship between X and Y is linear,

It's prediction value is $\bar{y}$

It has only RSS. ($RSS = TSS$)

For cubic regression,

$$\underbrace{(y_i - \bar{y})}_{} = (\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i)$$

differere between true Y and prection value Y.

It is not enaugh to tell whice model has lower RSS.
That's because a model well-fitted to our model
will have lower RSS. But we don't kow which
model fittest.

(2) We need more information
"we don't know how far it is from linear"
means that we could not explain how well the
true model fitted our data.

~~We need th~~

if we know the variance of error term

$$S_e^2 = \frac{RSS}{n-k}$$

We can guess which Rss is bigger than other one.