# Unit 1

Linear Regression

Simple Linear Regression

- Regression analysis makes use of mathematical models to describe relationships.
- For example, suppose that height was the only determinant of body weight.
- If we were to plot height (the independent or 'predictor' variable) as a function of body weight (the dependent or 'outcome' variable), we might see a very linear relationship, as illustrated below.
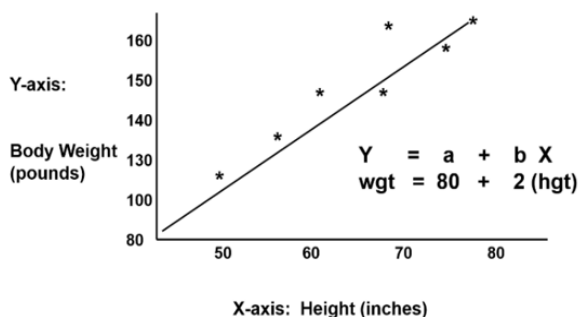
We could also describe this relationship with the equation for a line, Y = a + b(x), where 'a' is the Y-intercept and 'b' is the slope of the line. We could use the equation to predict weight if we knew an individual's height. In this example, if an individual was 70 inches tall, we would predict his weight to be:

Weight = 80 + 2 x (70) = 220 lbs.

The "a" or "B0" is the y intercept and it represents the predicted value of Y when X=0.

In contrast the slope, referred to as "b" (or beta 1), quantifies the average change in Y relative to a 1 unit change in Y.

Simple linear equation: $y = \beta_0 + \beta_1 x_1$

One dependent variable and one independent variable form simple linear prediction equation.

# Multiple Linear Regression Analysis

Multiple linear regression analysis is an extension of simple linear regression analysis which enables us to assess the association between two or more independent variables and a single continuous dependent variable.

This is a very useful procedure for identifying and adjusting for confounding. To provide an intuitive understanding of how multiple linear regression does this, consider the following hypothetical example.

Suppose an investigator had developed a scoring system that enabled her to predict an individual's body mass index (BMI) based on information about what they ate and how much. The investigator wanted to test this new "diet score" to determine how closely it was associated with actual measurements of BMI. Information is collected from a small sample of subjects in order to compute their "diet score," and the weight and height of each subject is measured in order to compute their BMI. The graph below shows the relationship between the new "diet score" and BMI, and it suggests that the "diet score" is not a very good predictor, (i.e., there is little if any association between the two.

"Diet Score"

While this is discouraging, the investigator considers that it is possible that confounding by age and/or gender is masking the true relationship between "diet score" and BMI. She first identifies which subjects who are older than 20 years old, and it turns out that the younger subjects and older subjects are clustered in the scatter plot, as shown in the figure below.

"Diet Score"
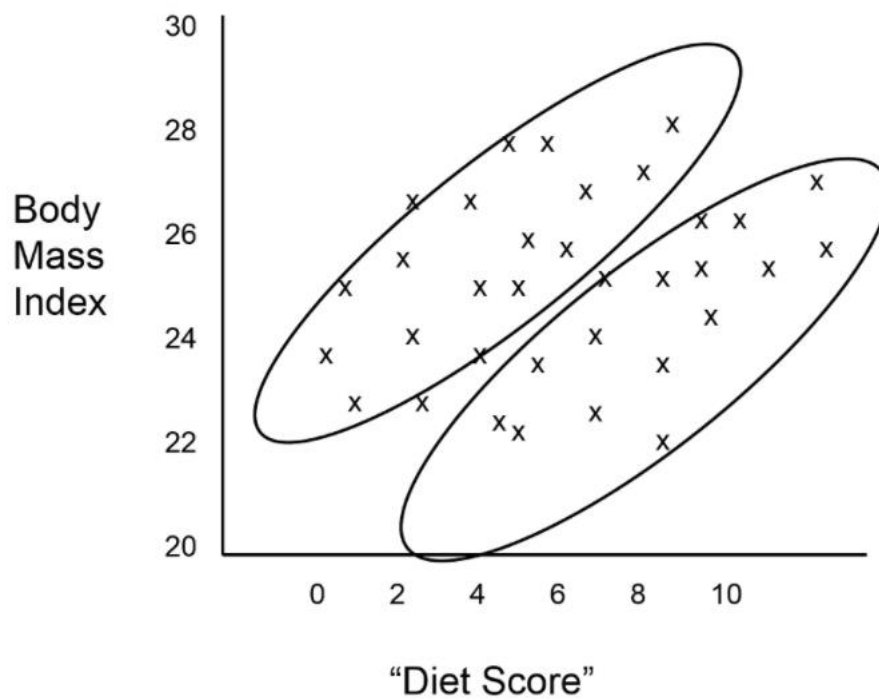
The investigator suspected that gender might also be a confounding factor, and when she identified males and females, the graph looked like this:

"Diet Score"
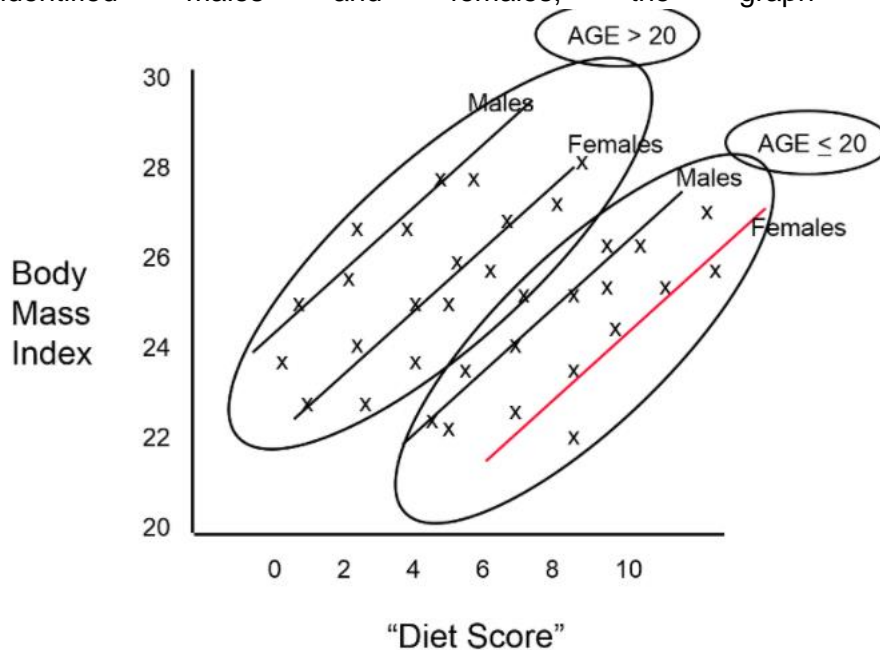
These findings indicate that both age and gender have an impact on BMI, because the older group has higher BMIs than the younger group, while males consistently have higher BMIs than the females. In addition, age and gender are also associated with "diet score," which is the "exposure" of interest, because diet scores are not equally distributed by gender or by age. In other words, both age and gender meet the criteria to be confounders. We can also see (in this very hypothetical example) that there is a striking linear relationship between "diet score" and BMI within each of the four age and gender groups. In other words, it is only after "taking into account" these two confounding variables that we can see that there really is a relationship between diet score and BMI. The true relationship was confounded by these other factors.

When analyzing data, it is always easier to deal with numeric data instead of text. For example, when dealing with dichotomous data the number "1" might conveniently indicate that the characteristic is present instead of "yes" or "true," and the number "0" conveniently indicates "no" or "false". It is also best to organize data sets so that the information for individual subjects is listed in a row, and the columns contain the variables. Consequently, in this scenario my data set would perhaps look something like this:

Subject ID	Diet Score	Male	Age>20	BMI
A	4	0	1	27
B	7	1	1	29
C	6	1	0	23
D	2	0	0	20
E	3	0	1	21
etc.	...	...	...	...

With data is are coded in this fashion, matrix math can be used to find the coefficients for each variable that led to the best "fit" of the data. For the hypothetical example we are considering here, multiple linear regression analysis could be used to compute the coefficients, and these could be used to describe the relationships in the graph mathematically with the following equation:
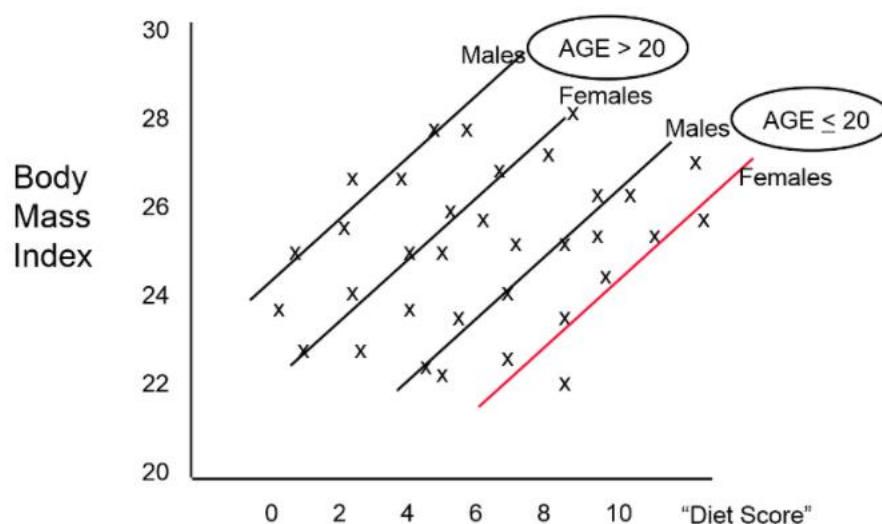
$$BMI = 18.0 + 1.5 \text{ (diet score)} + 1.6 \text{ (male)} + 4.2 \text{ (age>20)}$$

Note that the numbers in **red** are the coefficients that the analysis provided.

The Independent Effect of Each Independent Variable

In a sense, the equation above is a prediction of what an individual's BMI will be based on their diet score, gender and age group. The equation has an intercept of 18.0, meaning that I start with a baseline value of 18. I then multiply 1.5 x (diet score); I multiply 1.6 x (male) and multiply 4.2 x (age>20). But remember that in the data base I coded male as 1 for males and as 0 for females; for age group I coded it as 1 if the subject was older than 20 and coded it as 0 if they were less than 20.

Besides being useful for describing the relationships and making predictions, this mathematical description provides a powerful mean of **controlling for confounding**. For example, *the coefficient 1.5 for diet score indicates that for each additional point in diet score, I must add 1.5 units to my prediction*, regardless of whether it is a male or female or an adult or a child. In other words, the equation has quantified the association of diet score on BMI **independent of** (i.e., controlling for) gender and age group. Similarly, it means that I should add 1.6 units to my prediction if the individual is a male, regardless of there age and diet score. And I should add 4.2 to my prediction if the person is over age 20, regardless of their diet score or gender. As a result, the regression analysis has enabled us to dissect out the independent (unconfounded) association of each factor with the outcome of interest. The equation describes the graphical representation of this data, shown below.

$$BMI = 18.0 + 1.5 \text{ (diet score)} + 1.6 \text{ (if male)} + 4.2 \text{ (if adult)}$$

$$Y = a + b_1 X_1 + b_2 X_2 + b_3 X_3$$

The figure above and the equation enable us to see the impact of each of the independent variables after controlling for confounding. The equation is a mathematical expression of what

we see in the figure, and the coefficients for each variable describe an unconfounded measure of the association of each variable with the outcome.

Hence Multiple Linear Regression equation is

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

More than one independent variables are used to predict one dependent variable and form the multiple linear regression equation.
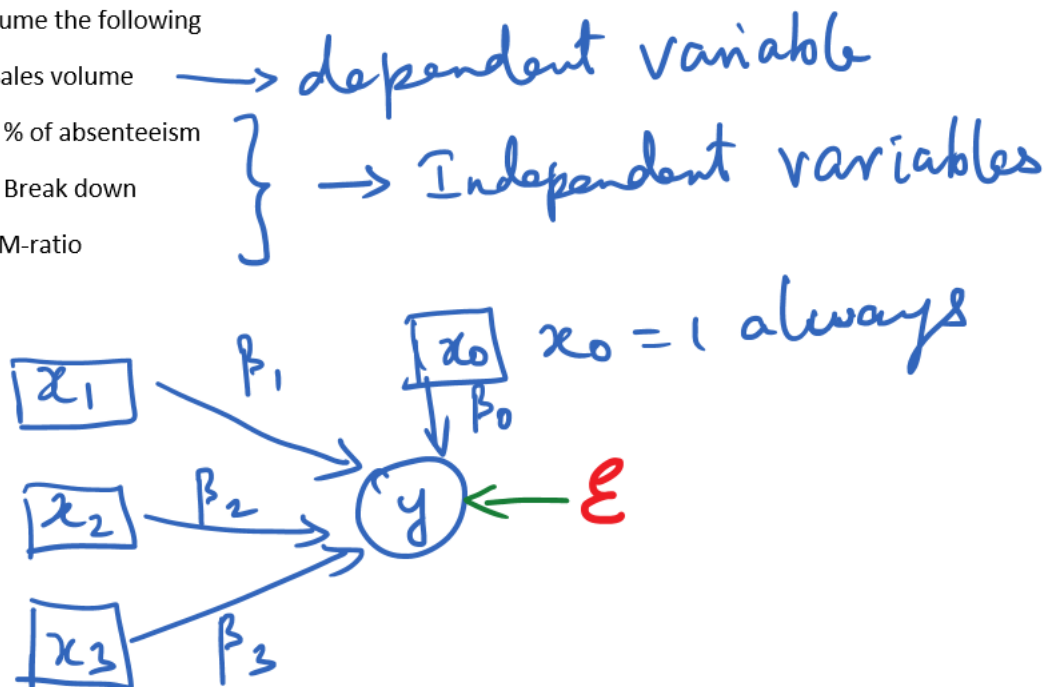
1.MLR Discussion:

Assume the following

Y=Sales volume $\longrightarrow$ dependent variable

X_1= % of absenteeism

X_2= Break down $\left.\begin{array}{l}\end{array}\right\} \longrightarrow$ Independent variables

X_3=M-ratio

$x_0 = 1$ always

$\mathcal{E}$ is error term which is the difference between actual and predicted y value.

$$y = \beta_0 x_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \mathcal{E}$$

From this, we can have

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}_{P \times 1} \qquad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}_{(P+1) \times 1}$$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \cdots + \beta_p x_p + \epsilon$$

So p variables are used to predict y values. Assume all the p variables are having n number of values.

Dependent var $\quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$

Independent var

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_2 & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{i1} & x_{i2} & \cdots & x_{ip} \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

$$\therefore \quad y_i = \beta_0 + \beta_1 x_{i1} + \cdots \beta_p x_{pp} + \epsilon$$

$$y_i = E[y_i \mid x_\cdot = x_{i1}, x_{i2} \cdots x_i] + \epsilon_i$$

$$y_i = \hat{y}_i + \epsilon_i$$
$$\epsilon_i = y_i - \hat{y}_i$$
$$\text{Error} = \text{actual} - \text{predicted}$$

Expanding MLR formula into matrix format,

$$y_i = \beta_0 + \beta_1 x_1 + \cdots \beta_p x_{ip} + \varepsilon_i$$

$$y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} \quad - \quad \beta_p x_{1p} + \varepsilon_1$$

$$y_2 = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} - - - \beta_p x_{2p} + \varepsilon_2$$

$$\vdots$$

$$y_p = \beta_0 + \beta_1 x_{p1} + \beta_2 x_{p2} - - - \beta_p x_{pp} + \varepsilon_p$$

$$y_n = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} \cdots \beta_p x_{np} + \varepsilon_n$$

$$
\begin{bmatrix} y_1 \\ y_2 \\ y_i \\ \vdots \\ y_n \end{bmatrix}_{n \times 1}
=
\begin{bmatrix}
1 & x_{11} & x_{12} & \cdots & x_{1p} \\
1 & x_{21} & x_{22} & \cdots & x_{2p} \\
1 & x_{i1} & x_{i2} & \cdots & x_{ip} \\
\vdots & & & & \\
1 & x_{n1} & x_{n2} & \cdots & x_{np}
\end{bmatrix}_{n \times (p+1)}
\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}_{(p+1) \times 1}
+
\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n \times 1}
$$

Design matrix   Regression co-efficient   Error term
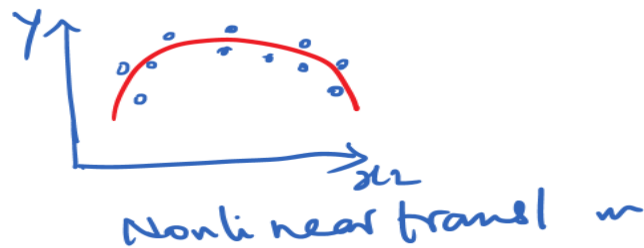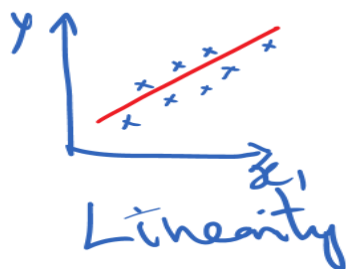
$$y = x\beta + \varepsilon \qquad ---- \;\; ①$$

MLR :

$$y_{n \times 1} = X_{n \times (p+1)} \cdot \beta_{(p+1) \times 1} + \varepsilon_{n \times 1}$$
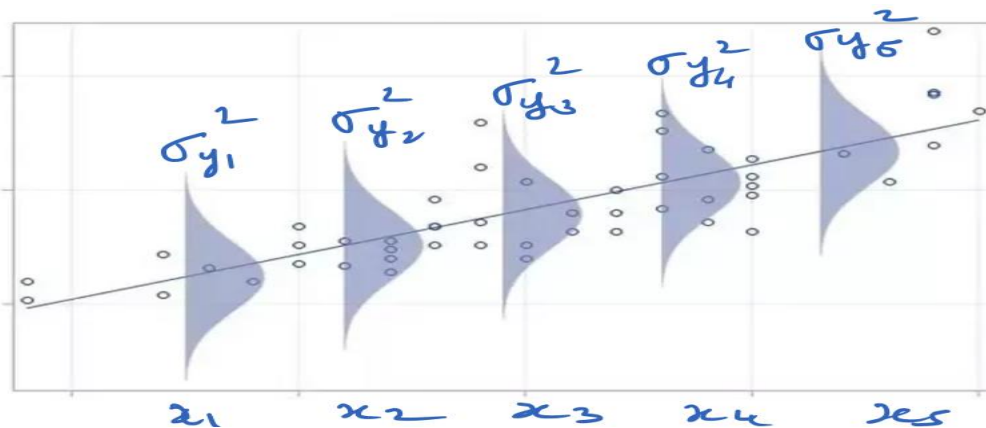
Design matrix   Regression co-efficient   Error term

2. **Assumptions**

1. Linearity
2. Homoskedasticity or equal Y variance across the values of X
3. Uncorrelated error terms
4. Normality of the error

1. Linear relationship between dependent and independent variable

2. Homoskedasticity or equal Y variance across the values of X

When the variance of dependent variable computed based on each independent variable, they are all equal.
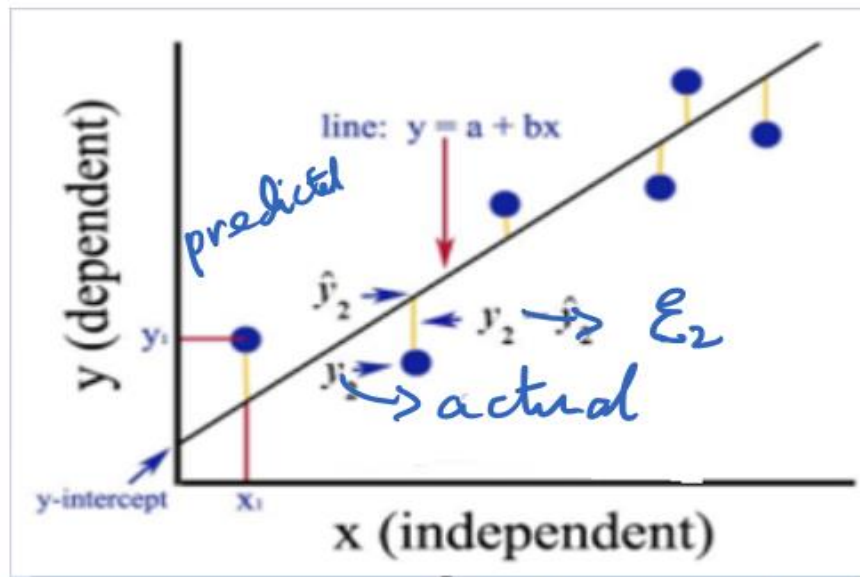
$$\sigma_{y_1}^2 = \sigma_{y_2}^2 = \sigma_{y_3}^2 = \sigma_4^2 = \sigma_5^2$$

In general,

$$\sigma_{y_1}^2 = \sigma_{y_2}^2 \cdots = \sigma_{y_n}^2 \quad \text{for}$$

n independent variables.

3. Uncorrelated error terms

4.

$$\mathcal{E} = actual - predicted$$

$$\mathcal{E}_i = y_{2i} - \hat{y}_{2i}$$

$$Cor(y_{2i} - \hat{y}_{2i}) = \begin{bmatrix} \mathcal{E}_1^2 & -\text{-}\text{-}\text{-}\text{-} & 0 \\ 0 & \mathcal{E}_2^2 & \text{-}\text{-}\text{-}\text{-}0 \\ 0 & \vdots & \ddots & \mathcal{E}_n^2 \end{bmatrix}$$

4. Normality of the error

$$\mathcal{E}_i \sim N(0, \sigma_i^2)$$

$$\sigma^2 = \sigma_y^2 = \sigma_{y_1}^2 = \sigma_{y_2}^2 = \cdots \sigma_{y_n}^2$$

3. Estimation of Model Parameter:

(B). Estimation of model parameter

$$MLR : y = x\beta + \varepsilon$$

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} \cdots + \beta_p x_{ip} + \varepsilon_i$$

$$\varepsilon_i = y_i - \sum_{j=0}^{p} \beta_j x_{ij}$$

SSE

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \left[y_i - \sum_{j=0}^{p} \beta_j x_{ij} \right]^2$$

Select β_j in such a way that, Sum square error (SSE) can be reduced.

$$\frac{\partial SSE}{\partial \beta} = 0 \quad \text{Subject to} \quad \frac{\partial^2 SSE}{\partial \beta_j \partial \beta_t} > 0$$

We know

$$E = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} \qquad y = x\beta + \varepsilon$$

$$\boxed{\varepsilon = y - x\beta}$$

$$SSE = \varepsilon^T \cdot \varepsilon = (y - x\beta)^T \cdot (y - x\beta)$$

Select only the β terms,

$$y^T y - y^T x \beta - (x\beta)^T y + (x\beta)^T x \beta$$

$\hat{\beta}$ and β are same

$$- y^T x \beta - x^T y \beta + x^T x \beta^2$$

$$- x^T y \beta - x^T y \beta + x^T x \beta^2 \qquad \because a^T b = b^T a$$

$$-2 y x^T \beta + x^T x \beta^2$$

$$\frac{\partial SSE}{\partial \beta} = -2 y x^T + 2 x^T x \beta = 0$$

$$-2 x^T (y - x\beta) = 0$$

$$-x^T (y - x\beta) = 0$$

$$-x^T y - x^T x \beta = 0$$

$$x^T x \beta = x^T y \quad ——①$$

pre multiply ① by $(x^T x)^{-1}$

$$(x^T x)^{-1} (x^T x) \beta = (x^T x)^{-1} x^T y$$

$$\boxed{\hat{\beta} = (x^T x)^{-1} x^T y}$$

Problem 1:

$$y = \begin{vmatrix} 10 \\ 20 \\ 30 \\ 40 \\ 50 \end{vmatrix} \qquad x = \begin{bmatrix} 1 & 5 \\ 1 & 7 \\ 1 & 10 \\ 1 & 12 \\ 1 & 20 \end{bmatrix}$$

$$5 \times 1 \qquad\qquad\qquad 5 \times 2$$

Calculate $\hat{\beta} = (x^T x)^{-1} x^T y$

Sol: ①

$$x^T x = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 5 & 7 & 10 & 12 & 20 \end{bmatrix} \begin{bmatrix} 1 & 5 \\ 1 & 7 \\ 1 & 10 \\ 1 & 12 \\ 1 & 20 \end{bmatrix}$$

$$2 \times 5 \qquad\qquad 5 \times 2$$

$$x^T x = \begin{bmatrix} 5 & 54 \\ 54 & 718 \end{bmatrix}$$

② $(x^T x)^{-1} = \dfrac{1}{|x^T x|} \, adj(x^T x)$

$$|X^T X| = \begin{vmatrix} 5 & 54 \\ 54 & 718 \end{vmatrix} = (5 \times 718) - 54^2$$

$$= 3590 - 2916$$

$$= 674$$

$$adj(x^T x) = \begin{bmatrix} 718 & -54 \\ -54 & 5 \end{bmatrix}$$

$$(x^T x)^{-1} = \frac{1}{674} \begin{bmatrix} 718 & -54 \\ -54 & 5 \end{bmatrix}$$

$$= \begin{bmatrix} 1.07 & -0.08 \\ -0.08 & 0.007 \end{bmatrix}$$

③ $x^T y = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 5 & 7 & 10 & 12 & 20 \end{bmatrix} \begin{bmatrix} 10 \\ 20 \\ 30 \\ 40 \\ 50 \end{bmatrix}$

$2 \times 5 \qquad 5 \times 1$

$$x^T y = \begin{bmatrix} 150 \\ 1970 \end{bmatrix}$$

$$\hat{\beta} = (x^T x)^{-1} x^T y$$

$$= \begin{bmatrix} 1.07 & -0.0801 \\ -0.0801 & 0.007 \end{bmatrix} \begin{bmatrix} 650 \\ 1970 \end{bmatrix}$$

$2 \times 2 \qquad 2 \times 1$

$$\hat{\beta} = \begin{bmatrix} 2.7 \\ 1.775 \end{bmatrix} \begin{matrix} \beta_0 \\ \beta_1 \end{matrix}$$

Now we have β_0 and β_1,

we know $y = \beta_0 + x_1 \beta_1 + \varepsilon$

So, substitute β_0, β_1 — the estimated parameters.

$$\therefore \quad y = 2.7 + 1.775 x_1 + \varepsilon$$

$$\hat{y} = 2.7 + 1.775 x_1$$

$$\hat{\varepsilon} = y - \hat{y}$$

$$\hat{\varepsilon} = \begin{bmatrix} 10 \\ 20 \\ 30 \\ 60 \\ 10 \end{bmatrix} - \begin{bmatrix} 2.7 + 1.775 \times 5 \\ 2.7 + 1.775 \times 7 \\ 2.7 + 1.775 \times 10 \\ 2.7 + 1.775 \times 12 \\ 2.7 + 1.775 \times 20 \end{bmatrix}$$

$$\hat{\varepsilon} = \begin{bmatrix} -1.575 \\ 4.875 \\ 9.55 \\ 16 \\ 18.5 \end{bmatrix}$$

Sampling distribution of Regression Coefficient

$$\hat{\beta} = (x^T x)^{-1} x^T y \; ; \; X_1 = (x_{11} \cdots x_{1p})$$

$$X = \begin{bmatrix} 1 & x_{11} & - & - & - & x_{1p} \\ 1 & x_{21} & & - & - & x_{2p} \\ \vdots & \vdots & & & & \vdots \\ 1 & x_{n1} & - & & - & x_{np} \end{bmatrix}$$

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n \times 1}$$

$$\hat{\beta}_{(p+1) \times 1} \quad \rightarrow \text{each sample with}$$
$$\text{different } \hat{\beta}$$

$$\hat{\beta}_{(p+1) \times 1} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_p \end{bmatrix}$$

What is the distribution of $\hat{\beta}$?

To know the distribution, we required to know expectation of $\hat{\beta}$ $\;$ and $cov(\hat{\beta})$

Mean:
$$E[\hat{\beta}] = E\left[(x^T x)^{-1} x^T y \right]$$

(from eq-1)

$$= E\left[(x^T x)^{-1} x^T (x \beta + \varepsilon) \right]$$

$$= \beta + (x^T x)^{-1} x^T * 0$$

$$E[\hat{\beta}] = \beta$$

unbiased estimation

$$= \left[(x^T x)^{-1} x^T x \beta + (x^T x)^{-1} x^T \varepsilon \right]$$

$$= E \left[I \beta + (x^T x)^{-1} x^T \varepsilon \right]$$

$$= E[\beta] + (x^T x)^{-1} x^T E[\varepsilon]$$

Consider

$$Cov(\hat{\beta}) = E \left[\{ \hat{\beta} - E[\hat{\beta}] \} \{ \hat{\beta} - E[\hat{\beta}] \}^T \right]$$

$$\hat{\beta}_{(p+1) \times 1} \quad ; \quad Cov(\hat{\beta})_{(p+1) \times (p+1)}$$

$$Cov(\hat{\beta}) = E \left[(\hat{\beta} - \beta) \cdot (\hat{\beta} - \beta)^T \right]$$

We know $\hat{\beta} = (x^T x)^{-1} x^T y$

$$\hat{\beta} - \beta$$

$$\therefore \hat{\beta} - \beta = (x^T x)^{-1} x^T y - \beta$$

$$= (x^T x)^{-1} x^T (x\beta + \varepsilon) - \beta$$

$$= (x^T x)^{-1} x^T x \beta$$

$$+ (x^T x)^{-1} x^T \varepsilon - \beta$$

$$\hat{\beta} - \beta = (\beta + (x^T x)^{-1} x^T \varepsilon) - \beta$$

$$= (x^T x)^{-1} x^T \varepsilon$$

$$(\hat{\beta} - \beta)^T = \left[(x^T x)^{-1} x^T \varepsilon \right]^T$$

$$= \varepsilon^T x (x^T x)^{-1}$$

$$E\left[(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T \right] = E\left[\left((x^T x)^{-1} x^T \varepsilon \cdot \right. \right.$$

$$\left. \left. \left(\varepsilon^T x (x^T x)^{-1} \right) \right) \right]$$

$$= (x^T x)^{-1} x^T E[\varepsilon \varepsilon^T] x (x^T x)^{-1}$$

$$= (x^T x)^{-1} x^T (\sigma^2 I) x (x^T x)^{-1}$$

$$= \sigma^2 (x^T x)^{-1} x^T x (x^T x)^{-1}$$

$$= \sigma^2 (x^T x)^{-1} = Cov(\hat{\beta})$$

$$= s_e^2 (x^T x)^{-1}$$

$$\sigma^2 = s_e^2 = \frac{SSE}{n-(p+1)}$$

Confidence Interval

$\hat{\beta}$ is a random variable
with mean β and
$Cov \ \sigma^2 (x^T x)^{-1}$

$$\therefore \ \hat{\beta} \sim N_{p+1}[\beta, \ \sigma^2 (x^T x)^{-1}]$$

$$X \sim N_p(\mu, \Sigma) \quad ; \quad \mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_p \end{bmatrix}$$

$$\bar{X} \sim N_p(\mu, \Sigma/n) \quad ; \quad \bar{X} = \begin{bmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_p \end{bmatrix}$$

$$P\left\{ n(\bar{x} - \mu)^T S^{-1} (\bar{x} - \mu) \right.$$
$$\left. \leq \frac{(n-1)p}{n-p} F_{p, n-p} \right\} = 1 - \alpha$$

Confidence region
w.r.t pop. mean vector

$$\left\{ (\hat{\beta} - \beta)^T (x^T x)(\hat{\beta} - \beta) \leq \right.$$
$$\left. S_a^2 (p+1) F_{p+1, n-p-1}(\alpha) \right\}$$
$$= 1 - \alpha$$

Confidence region for β

$$\hat{\beta}_j \rightarrow \frac{\hat{\beta}_j - E(\hat{\beta}_j)}{SE(\hat{\beta}_j)} = \frac{\hat{\beta}_j - \beta_j}{SE(\hat{\beta}_j)}$$

$$Cov(\hat{\beta}) = S_e^2 (x^T x)^{-1}$$

$$\text{Correlation}: \quad S_e^2 \begin{array}{c} 0 \\ 1 \\ \vdots \\ P \end{array} \begin{bmatrix} C_{00} & C_{01} & \cdots & C_{0p} \\ C_{10} & C_{11} & \cdots & C_{1p} \\ \vdots & & & \\ C_{p0} & \cdots & \cdots & C_{pp} \end{bmatrix} \begin{array}{c} 0 \quad 1 \quad \quad P \end{array}$$

$$\hat{\beta}_j \sim \text{univariate normal}$$

$$\bar{x} \rightarrow \frac{\bar{x} - E(\bar{x})}{SE(\bar{x})} = \frac{\bar{x} - \mu}{s/\sqrt{n}} \sim t_{n-1}$$
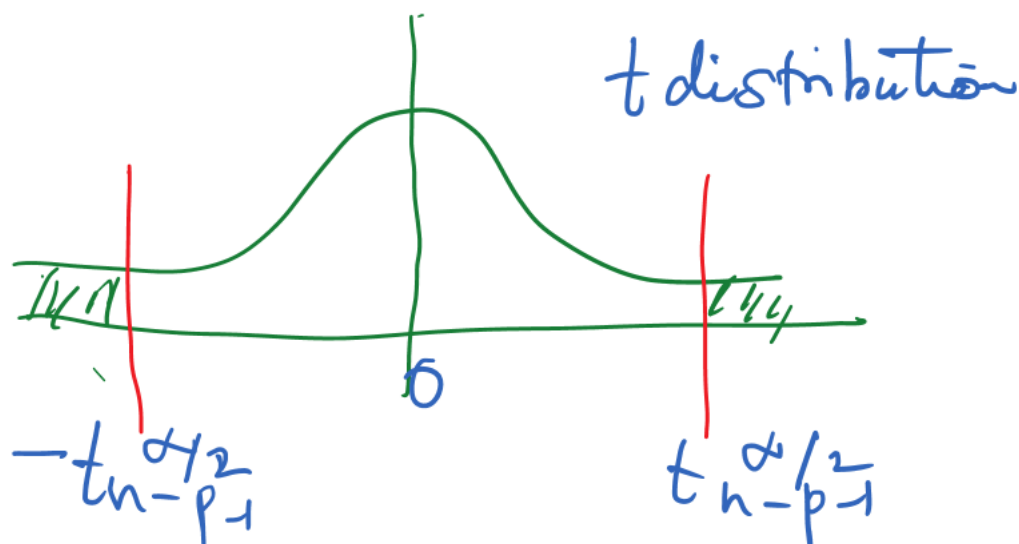
t distribution
with $n-1$ d.f

$$SE(\hat{\beta}_j) = \sqrt{S_e^2 \cdot C_{jj}}$$

$$\therefore \quad Cor(\hat{\beta}) = \frac{\hat{\beta_j} - \beta_j}{\sqrt{Se^2 \, c_{jj}}}$$

$$y = \beta_0 + \beta_1 x_1 + \cdots \beta_j x_j + \cdots \beta_p x_p + \varepsilon$$

$$\frac{\hat{\beta_j} - \beta_j}{\sqrt{Se^2 \, c_{jj}}} \sim t_{n-p-1}$$

t distribution

$$-t_{n-p-1}^{\alpha/2} \qquad\qquad t_{n-p-1}^{\alpha/2}$$

$$-t_{n-p-1}^{(\alpha/2)} \leq \frac{\hat{\beta_j} - \beta_j}{\sqrt{Se^2 \, c_{jj}}} \leq t_{n-p-1}^{(\alpha/2)}$$

$$\hat{\beta_\delta} - t_{n-p-1}^{(\alpha/2)} \sqrt{s^2 c_{j\delta}} \le \beta_j$$

$$\le \hat{\beta_\delta} + t_{n-p-1}^{(\alpha/2)} \sqrt{s^2 c_{j\delta}}$$

Case 1:

Suppose if your confidence interval is -1.5 to 1.5

$$\text{and } \beta_j = 4, \quad \hat{\beta} = 3$$

$$\text{then } 3\text{-}1.5 <= 4 <= 3\text{+}1.5$$

$$1.5 <= 4 <= 4.5$$

So, the predicted coefficient is in acceptable range.

Case 2:

Suppose if your confidence interval is -1.5 to 1.5

$$\text{and } \beta_j = 4, \quad \hat{\beta} = 6$$

$$\text{then } \mathbf{6\text{-}1.5 <=} 4 <= 6\text{+}1.5 \text{ -> not satisfied}$$

$$\mathbf{4.5 <=} 4 <= 7.5$$

So, the predicted coefficient is not in acceptable range.

Problem 1: compute $\hat{\beta}$ for the following data

I	Y=sales	X_1=% absent	X_2=Breaking
1	100	9	62
2	110	8	58
3	105	7	64
4	94	14	60
5	95	12	63
6	99	10	57
7	104	7	55
8	108	4	56
9	105	6	59
10	98	5	61
11	105	7	57
12	110	6	60