Date __/__/__ Page No.: ____

Explaination of Predition Model

Hypothesis Case -

Let's assume that the outbreak of the Covid-19 crisis started with the report of 1 case initially (on day 1).
Now, let's assume that the sick person infects 2 other, making the growth rate = 2
So now these 2 people will affect 2 more people and with no of days passing we'll have 2^n more cases (where n = no. of days)

Hence, the total no. of cases (y) comes out to be

$$y = a b^x$$

Here, y = total cases
a = Initial reported cases
b = Growth rate
x = No. of days.

Pratical Case -

The above hypothetical formulae can predict an ideal case of crisis growth. However, in pratical world, It is not necessary that growth rate of the virus is constant for every person/day.

In simpler words, it is not necessary that if person 1 transmits virus to 2 people than person-2 also transmits virus to exactly 2 people (growth rate can be different)

In the present scenario we have the data of number of cases every day. Which means we have y and x.
Using this we've to find the value of a & b, so that the average growth factor of the pandemic can be used to predict no. of cases in future.

The Question : How?

Linear Regression

Linear regression is a statistical model which can help. The model gives the value of a and b for the following form of eqn

$$y_0 = a_0 + b_0 x_0$$

We thus have created our hypothesis to the following form -

$$y = ab^x$$
$$\Rightarrow \log y = \log(ab^x)$$
$$\Rightarrow \log y = \log a + x \log b$$

Applying this to regression formulae we can use

$$\underset{\log y}{\underset{\uparrow}{y_0}} = \underset{\log a}{\underset{\uparrow}{a_0}} + \underset{\log b}{\underset{\uparrow}{b_0}}\ \underset{x}{\underset{\uparrow}{x_0}}$$

Linear Regression Algorithm

To estimate values of a & b, we have to find a_0, b_0

$$y_0 = a_0 + b_0 x_0 \quad - (i)$$

$$x_0 y_0 = x_0 a_0 + b_0 x_0^2 \quad - (ii)$$

Summing up the whole data (different values of y_0 & x_0) we will get

$$eq^n (i) \text{ as} \Rightarrow \quad \Sigma y_0 = na + b_0 \Sigma x_0$$

$$eq^n (ii) \text{ as} \Rightarrow \quad \Sigma(x_0 y_0) = a_0 \Sigma x_0 + b_0 \Sigma(x_0^2)$$

Substituting the values we'll get —

$$b_0 = \frac{n\Sigma(x_0 y_0) - \Sigma x_0 \Sigma y_0}{n \Sigma x_0^2 - \Sigma x_0 \Sigma x_0}$$

$$a_0 = \frac{\Sigma y_0 - b\Sigma x_0}{n}$$

In our code we've stored the available data in these variables y_0 (total cases (every day)) and x_0 (day number).

Hence, we got the value of a_0 & b_0

Now, we took

$$a_0 = \log a \Rightarrow a = e^{a_0}$$

$$b_0 = \log b \Rightarrow b = e^{b_0}$$

Now, this value of growth rate can be used to predict the future cases using $\boxed{y = a b^x}$