SIMPLE LINEAR REGRESSION Part 3

DR NISHITA GUPTA
DEPARTMENT OF COMMERCE
SRI AUROBINDO COLLEGE

Using Assumed Mean for Calculating Regression Coefficient

- As we are aware that regression coefficients are independent of change of origin (but not of scale). As such, if the values of X and Y are large or the actual mean is not an integer, we can also use an assumed mean to compute regression coefficient.
- Substituting the values of X and Y by U and V respectively in formula,
- **Example**
- U = X – A , where A = assumed mean for variable X
- V = Y – B , where B = assumed mean for variable Y
- $b_{YX} = b_{VU}$ and $b_{XY} = b_{UV}$

$$bYX = bVU = \frac{n\sum UV - (\sum U)(\sum V)}{n\sum U^2 - (\sum U)^2}$$

$$bXY = bUV = \frac{n\sum UV - (\sum U)(\sum V)}{n\sum V^2 - (\sum V)^2}$$

- Calculate the regression coefficient and regression equation of Y on X based on the following data:

X	40	38	35	42	30
Y	30	35	40	36	29

X	Y	Deviation from assumed mean A = 35, U = X – 35	Deviation from assumed mean B = 30, V = Y- 30	U^2	V^2	UV
40	30	40 – 35 = 5	30 – 30 = 0	25	0	0
38	35	38 – 35 = 3	35 – 30 = 5	9	25	15
35	40	35 – 35 = 0	40 – 30 = 10	0	100	0
42	36	42 – 35 = 7	36 – 30 = 6	49	36	42
30	29	30 – 35 = -5	29 – 30 = -1	25	1	5
$\sum X=185$	$\sum Y=170$	$\sum U = 10$	$\sum V = 20$	$\sum U^2 = 108$	$\sum V^2 = 162$	$\sum UV = 62$

- Regression coefficient of Y on X:
- $b_{YX} = b_{VU} = \dfrac{n\sum UV - (\sum U)(\sum V)}{n\sum U^2 - (\sum U)^2}$

- $b_{VU} = \dfrac{5(62) - (10)(20)}{5(108) - (10)^2}$
- $= \dfrac{310 - 200}{540 - 100}$
- $= \dfrac{110}{440}$
- $b_{VU} = 0.25$

- $\bar{X} = \frac{\sum X}{n.} = \frac{185}{5} = 37$
- $\bar{Y} = \frac{\sum Y}{n.} = \frac{170}{5} = 34$
- Regression equation of Y on X
- $Y-\bar{Y} = b_{YX}(X-\bar{X})$
- Y – 34 = 0.25(X – 37)
- Y = 0.25X + 24.75

Determining the line of regression of Y on X and that of X on Y out of the given two regression lines

- To find the line of regression of Y on X or X on Y out of the given two regression lines, we follow the following steps:
- **Step 1**. Choose any one of the two regression lines as the line of regression of Y on X and the other as the line of regression of X on Y.
- **Step 2**. Find two regression coefficients b_{YX} and b_{XY} .
- **Step 3**. Compute the product $b_{YX} . b_{XY}$.
- **Step 4**. If product of b_{YX} and b_{XY} is less than or equal to 1, then the assumption made in step 1 is correct, otherwise the assumption is wrong.
- I.E, If $b_{YX} . b_{XY} \leq 1$, assumption made in step 1 is correct. If $b_{YX} . b_{XY} > 1$, assumption made in step 1 is incorrect

- Ex-Out of the following two regression lines, find the line of regression of
- Y on X:
- $X + 2Y = 5$ and $2X + 3Y = 8$
- We are given two regression equations
- $X + 2Y = 5$ ...... (1)
- $2X + 3Y = 8$ ......(2)
- Let us arbitrarily assume that Equation (1) is the line of regression of Y on X and consequently Equation (2) is the line of regression of X on Y. Our assumption will be correct if we show that $b_{YX} \cdot b_{XY} \leq 1$
- Solving Equation (1) for Y in terms of X, we obtain
  $X + 2Y = 5$
- $2Y = 5 - X$
- $Y = \frac{5}{2} - \frac{X}{2}$
- Now this equation is in the form of $Y = a + bX$

- Where a = intercept = 5/2
- And b = Slope or $b_{YX} = -1/2$
- Similarly, solving Equation (2) for X in terms of Y, we obtain
- 2X + 3Y = 8
- 2X = 8 – 3Y
- $X = \frac{8}{2} - \frac{3Y}{2}$
- $X = 4 - \frac{3Y}{2}$
- Now this equation is in the form of X = c + dY
- Where c = intercept = 4
- And d = Slope = $b_{XY} = -3/2$

- $b_{YX}=-1/2$ and $b_{XY}=-3/2$

- $b_{YX} \cdot b_{XY} = -1/2 \times -3/2$
- $= 3/4 \leq 1$, which is true

- Hence our assumption is correct that the equation X + 2Y =5 is the line of regression of Y on X and 2X + 3Y = 8 is the line of regression of X on Y.