<u>Regression</u>

16. What do you mean by regression ?

Solution : Regression is the average relationship between two variables and from this average relationship the average value of one variable is estimated corresponding to a given value of the other. This process is known as simple regression.

There are two types of variables in regression analysis. One is dependent variable and the other is independent variable.

The value of the dependent variable is to be predicted. It is also known as explained variable. An independent variable is the variable which influences the value of the other variable. It is also known as explanator.

17. What are the lines of regression ?

Solution : Line of regression is the line which gives the best estimate of one variable for any specific value of the other variable. For bivariate distribution we have two lines of regression :

(i) The regression line of Y on X – this gives the best estimated value of Y corresponding to a given value of X.

(ii) The regression line of X on Y – this gives the best estimated value of X corresponding to a given value of Y.

There are always two lines of regression since each variable may be treated as the dependent as well as the independent variable.

18. What are regression equations ? Explain them.

Solution : The regression equations express the regression lines.

(i) Regression equation of Y on X is

$Y - \bar{Y} = b_{yx}(X - \bar{X})$, where b_{yx} = Regression co-efficient of Y on X.

$$= \frac{Cov(X,Y)}{{\sigma_x}^2} = r.\frac{\sigma_y}{\sigma_x} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2} = \frac{n\sum xy - (\sum x)(\sum y)}{n\sum x^2 - (\sum x)^2}$$

(ii) Regression equation of X on Y is

$X - \bar{X} = b_{xy}(Y - \bar{Y})$, where b_{xy} = Regression co-efficient of X on Y.

$$= \frac{Cov(X,Y)}{{\sigma_y}^2} = r.\frac{\sigma_x}{\sigma_y} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(y-\bar{y})^2} = \frac{n\sum xy-(\sum x)(\sum y)}{n\sum y^2-(\sum y)^2}$$

Note : For actual mean in decimal, the above formulae may be considered as :

$$b_{yx} = \frac{n\sum uv-(\sum u)(\sum v)}{n\sum u^2-(\sum u)^2};$$

$$b_{xy} = \frac{n\sum uv-(\sum u)(\sum v)}{n\sum v^2-(\sum v)^2}$$

where, $u = x - A$ & $v = y - B$; A and B are assumed means.

Properties of Regression Co-efficients

19. The geometric mean of the regression co-efficient is the correlation co-efficient. **(Property No. 1) (Important)**

Solution : We know that

$$b_{yx} = r.\frac{\sigma_y}{\sigma_x} \quad \& \quad b_{xy} = r.\frac{\sigma_x}{\sigma_y}$$

$$\therefore \quad b_{yx} \,.\, b_{xy} = r.\frac{\sigma_y}{\sigma_x} \,.\, r.\frac{\sigma_x}{\sigma_y}$$

$$\Rightarrow b_{yx} \,.\, b_{xy} = r^2$$

$$\Rightarrow r = \pm\sqrt{b_{yx}b_{xy}}$$

Note : If both b_{yx} and b_{xy} are positive then r is also positive and if both b_{yx} and b_{xy} are negative then r is also negative. Since, $\sigma_x > 0$, $\sigma_y > 0$, the sign of each r, b_{yx}, b_{xy} depends on $Cov(X,Y)$.

20. If one of the regression co-efficient is greater than 1, then other must be less than 1. **(Property No. 2) (Important)**

Solution : Let us consider, $b_{yx} > 0$

⇨ $\frac{1}{b_{yx}} < 1$

Also, $b_{yx} \,.\, b_{xy} = r^2$

⇨ $b_{yx} \,.\, b_{xy} \leq 1$, since $-1 \leq r \leq 1$, therefore $r^2 \leq 1$

⇨ $b_{xy} \leq \frac{1}{b_{yx}} < 1$

Hence, $b_{xy} < 1$.

21. Regression co-efficients are independent of the change of origin but not of scale. **(Property No. 3) (Important)**

Solution : Let us consider, X and Y are the original variables and after changing origin and scale new variables are :

$$u = \frac{X-a}{h} \quad \text{and} \quad v = \frac{Y-b}{k}$$

where, a, b, h, k are all constants and $h > 0$ and $k > 0$.Since, correlation co-efficients are independent of change of origin and scale, $r_{xy} = r_{uv}$. Again, SD is independent of change in origin but not of scale.

∴ $\sigma_x = h\sigma_u$ and $\sigma_y = h\sigma_v$

Now, we have, $b_{yx} = r_{xy}.\frac{\sigma_y}{\sigma_x} = r_{uv}.\frac{k\sigma_v}{h\sigma_u} = \frac{k}{h} b_{vu}$

Also, we have, $b_{xy} = r_{xy}.\frac{\sigma_x}{\sigma_y} = r_{uv}.\frac{k\sigma_u}{h\sigma_v} = \frac{h}{k} b_{uv}$

Hence the proof.

22. The arithmetic mean of the regression co-efficients is greater than and equal to correlation co-efficient, provided $r > 0$. **(Property No. 3) (Important)**

Solution : We want to show that

$$\frac{b_{yx}+b_{xy}}{2} \geq r$$

⇨ $b_{yx} + b_{xy} \geq 2r$

⇨ $r.\frac{\sigma_y}{\sigma_x} + r.\frac{\sigma_x}{\sigma_y} \geq 2r$

⇨ $\frac{\sigma_y}{\sigma_x} + \frac{\sigma_x}{\sigma_y} \geq 2$

⇨ ${\sigma_y}^2 + {\sigma_x}^2 \geq 2\sigma_x\sigma_y$

⇨ ${\sigma_y}^2 + {\sigma_x}^2 - 2\sigma_x\sigma_y \geq 0$

⇨ $(\sigma_y - \sigma_x)^2 \geq 0$ which is always true.

Hence, the proof.

Note : Both the regression lines passes through the point $(\bar{x}, \bar{y})$. **(Important for objective question)**