
UNIT10 PLANE OF REGRESSION

Structure

- 10.1 Introduction
 - Objectives
- 10.2 Yule's Notation
- 10.3 Plane of Regression for three Variables
- 10.4 Properties of Residuals
- 10.5 Variance of the Residuals
- 10.6 Summary
- 10.7 Solutions / Answers

10.1 INTRODUCTION

In Unit 9, you learnt the concept of regression, linear regression, lines of regression and regression coefficient with its properties. Unit 9 was based on linear regression in which we were considering only two variables. In Unit 6 of MST-002, you studied the correlation that measures the linear relationship between two variables. In many situations, we are interested in studying the relationship among more than two variables in which one variable is influenced by many others. For example, production of any crop depends upon soil fertility, fertilizers used, irrigation methods, weather conditions, etc. Similarly, marks obtained by students in exam may depend upon their IQ, attendance in class, study at home, etc. In these type of situations, we study the multiple and partial correlations. In this unit you will study the basics of multiple and partial correlations that includes Yule's notations, plane of regression, residuals, properties of residuals and variance of residuals.

For the study of more than two variables, there would be need of much more notations in comparison to the notations used in Unit 9. These notations were given by G.U. Yule (1897). Yule's notations and residuals are described in Section 10.2. Plane of regression and normal equations are given in Section 10.3. Properties of residuals are explored in Section 10.4 whereas Section 10.5 explains the variance of residuals.

Objectives

After reading this unit, you would be able to

- define the Yule's notation;
- describe the plane of regression for three variable;
- explain the properties of residuals;
- describe the variance of the residuals;
- find out the lines of regression; and
- find out the estimates of dependent variable from the regression lines.

10.2 YULE'S NOTATION

Karl Pearson developed the theory of multiple and partial correlation for three variables and it was generalized by G.U. Yule (1897). Let us consider three random variables X_1 , X_2 and X_3 and \bar{X}_1 , \bar{X}_2 and \bar{X}_3 are their respective means. Then regression equation of X_1 on X_2 and X_3 is defined as

$$X_1 = a + b_{12.3}X_2 + b_{13.2}X_3 \quad \dots (1)$$

where, $b_{12.3}$ and $b_{13.2}$ are the partial regression coefficients of X_1 on X_2 and X_1 on X_3 keeping the effect of X_3 and X_2 fixed respectively.

Taking summation of equation (1) and dividing it by N , we get

$$\bar{X}_1 = a + b_{12.3}\bar{X}_2 + b_{13.2}\bar{X}_3 \quad \dots (2)$$

On subtracting equation (2) from equation (1), we get

$$X_1 - \bar{X}_1 = b_{12.3}(X_2 - \bar{X}_2) + b_{13.2}(X_3 - \bar{X}_3) \quad \dots (3)$$

Suppose $X_1 - \bar{X}_1 = x_1$, $X_2 - \bar{X}_2 = x_2$ and $X_3 - \bar{X}_3 = x_3$

Now, plane of regression x_1 on x_2 and x_3 (equation (3)) can be rewritten as

$$x_1 = b_{12.3}x_2 + b_{13.2}x_3 \quad \dots (4)$$

Right hand side of equation (2) is called the estimate of x_1 which is denoted by $x_{1.23}$. Thus,

$$x_{1.23} = b_{12.3}x_2 + b_{13.2}x_3 \quad \dots (5)$$

Error of estimate or residual is defined as

$$\begin{aligned} e_{1.23} &= x_1 - x_{1.23} \\ e_{1.23} &= x_1 - b_{12.3}x_2 - b_{13.2}x_3 \end{aligned} \quad \dots (6)$$

This residual is of order 2.

If we are considering n variables x_1, x_2, \dots, x_n , the equation of the plane of regression of x_1 on x_2, \dots, x_n is

$$x_1 = b_{12.3456\dots n}x_2 + b_{13.24\dots n}x_3 + \dots + b_{1n.23\dots(n-1)}x_n \quad \dots (7)$$

and error of estimate or residual is

$$x_{1.23\dots n} = x_1 - (b_{12.34\dots n}x_2 + b_{13.24\dots n}x_3 + \dots + b_{1n.23\dots(n-1)}x_n) \quad \dots (8)$$

Note: In above expressions we have used subscripts involving digits 1, 2, 3, ..., n and dot (.). Subscripts before the dot are known as the primary subscripts whereas the subscripts after the dot are called secondary subscripts.

The number of secondary subscripts decides the order of regression coefficient.

For example $b_{12.3}$ is the regression coefficient of order 1, $b_{12.34}$ is of order 2 and so on $b_{12.34\dots(n-1)}$ of order $(n-2)$.

Order in which secondary subscripts ($b_{12.34}$ or $b_{12.43}$) is immaterial but the order of primary subscripts is very important and decides the dependent and independent variables. In $b_{12.34}$, x_1 is dependent variable and x_2 is independent variable whereas in $b_{21.34}$, x_2 is dependent variable and x_1 is independent variable.

Order of residuals is determined by the number of secondary subscripts. For example $x_{1.23}$ is residual of order 2 while as $x_{1.234}$ is of order 3. Similarly, $x_{1.234\dots n}$ is residual of order $(n-1)$.

10.3 PLANE OF REGRESSION FOR THREE VARIABLES

Let us consider three variables x_1, x_2 and x_3 measured from their respective means. The regression equation of x_1 depends upon x_2 and x_3 is given by (from equation (4)).

$$x_1 = b_{12.3}x_2 + b_{13.2}x_3 \quad \dots (9)$$

If x_3 is considered as a constant then the graph of x_1 and x_2 is a straight line with slope $b_{12.3}$, similarly the graph of the x_1 and x_3 would be the straight line with slope $b_{13.2}$, if x_2 is considered as a constant.

According to the principle of least squares, we have to determine constants $b_{12.3}$ and $b_{13.2}$ in such a way that sum of squares of residuals is minimum i.e.

$$U = \sum (x_1 - b_{12.3}x_2 - b_{13.2}x_3)^2 = \sum e_{1.23}^2 \text{ is minimum.}$$

$$\text{here, } e_{1.23} = x_1 - b_{12.3}x_2 - b_{13.2}x_3 \quad \dots (10)$$

By the principle of maxima and minima, we take partial derivatives of U with respect to $b_{12.3}$ and $b_{13.2}$

Thus,

$$\frac{\partial U}{\partial b_{12.3}} = \frac{\partial U}{\partial b_{13.2}} = 0$$

Let us take

$$\frac{\partial U}{\partial b_{12.3}} = 0$$

$$\Rightarrow \sum 2(x_1 - b_{12.3}x_2 - b_{13.2}x_3)(-x_2) = 0$$

$$\Rightarrow \sum x_2(x_1 - b_{12.3}x_2 - b_{13.2}x_3) = 0 \quad \dots (11)$$

$$\begin{aligned} &\Rightarrow \sum (x_2 x_1 - b_{12.3} x_2^2 - b_{13.2} x_2 x_3) = 0 \\ &\Rightarrow b_{12.3} \sum x_2^2 + b_{13.2} \sum x_2 x_3 - \sum x_1 x_2 = 0 \end{aligned} \quad \dots (12)$$

Similarly,

$$\frac{\partial U}{\partial b_{13.2}} = 0 \Rightarrow \sum x_3 (x_1 - b_{12.3} x_2 - b_{13.2} x_3) = 0 \quad \dots (13)$$

$$\Rightarrow b_{12.3} \sum x_2 x_3 + b_{13.2} \sum x_3^2 - \sum x_1 x_3 = 0 \quad \dots (14)$$

As we know that

$$\begin{aligned} \sigma_i^2 &= \frac{1}{N} \sum (x_i - \bar{x}_i)^2 \quad (\text{for } i = 1, 2 \text{ and } 3) \\ &= \frac{1}{N} \sum x_i^2 - \bar{x}_i^2 \end{aligned}$$

Since, x_1 , x_2 and x_3 are measured from their means therefore

$\bar{x}_1 = \bar{x}_2 = \bar{x}_3 = 0$ then

$$\sigma_i^2 = \frac{1}{N} \sum x_i^2 \quad \dots (15)$$

Similarly, we can write (for $i \neq j = 1, 2, 3$)

$$\text{Cov}(x_i, x_j) = \frac{1}{N} \sum x_i x_j \quad \dots (16)$$

and consequently, using equations (15) and (16), correlation coefficient between x_i and x_j can be expressed as

$$r_{ij} = \frac{\text{Cov}(x_i, x_j)}{\sqrt{V(x_i)V(x_j)}} = \frac{\sum x_i x_j}{N \sigma_i \sigma_j} \Rightarrow \text{Cov}(x_i, x_j) = r_{ij} \sigma_i \sigma_j \quad \dots (17)$$

Dividing equations (12) and (14) by N provides

$$b_{12.3} \frac{1}{N} \sum x_2^2 + b_{13.2} \frac{1}{N} \sum x_2 x_3 - \frac{1}{N} \sum x_1 x_2 = 0 \quad \text{and} \quad \dots (18)$$

$$b_{12.3} \frac{1}{N} \sum x_2 x_3 + b_{13.2} \frac{1}{N} \sum x_3^2 - \frac{1}{N} \sum x_1 x_3 = 0 \quad \dots (19)$$

Using equations (15), (16) and (17) in equations (18) and (19), we have

$$\begin{aligned} &b_{12.3} \sigma_2^2 + b_{13.2} \text{Cov}(x_2, x_3) - \text{Cov}(x_1, x_2) = 0 \quad \text{From equation (18)} \\ &\Rightarrow b_{12.3} \sigma_2^2 + b_{13.2} r_{23} \sigma_2 \sigma_3 - r_{12} \sigma_1 \sigma_2 = 0 \\ &\Rightarrow \sigma_2 (b_{12.3} \sigma_2 + b_{13.2} r_{23} \sigma_3 - r_{12} \sigma_1) = 0 \\ &\Rightarrow b_{12.3} \sigma_2 + b_{13.2} r_{23} \sigma_3 - r_{12} \sigma_1 = 0 \end{aligned} \quad \dots (20)$$

Similarly,

$$b_{12.3} \text{Cov}(x_2, x_3) + b_{13.2} \sigma_3^2 - \text{Cov}(x_1, x_3) = 0 \quad \text{From equation (19)}$$

$$\begin{aligned}
&\Rightarrow b_{12.3}r_{23}\sigma_2\sigma_3 + b_{13.2}\sigma_3^2 - r_{13}\sigma_1\sigma_3 = 0 \\
&\Rightarrow \sigma_3(b_{12.3}r_{23}\sigma_2 + b_{13.2}\sigma_3 - r_{13}\sigma_1) = 0 \\
&\Rightarrow b_{12.3}r_{23}\sigma_2 + b_{13.2}\sigma_3 - r_{13}\sigma_1 = 0 \quad \dots (21)
\end{aligned}$$

where, r_{12} is the total correlation coefficient between x_1 and x_2 , r_{13} is the total correlation coefficient between x_1 and x_3 and similarly, r_{23} is the total correlation coefficient between x_2 and x_3 . Thus, we have two equations (20) and (21).

Solving the equations (20) and (21) for $b_{12.3}$ and $b_{13.2}$, we obtained

$$\begin{aligned}
b_{12.3} &= \frac{\begin{vmatrix} r_{12}\sigma_1 & r_{23}\sigma_3 \\ r_{13}\sigma_1 & \sigma_3 \end{vmatrix}}{\begin{vmatrix} \sigma_2 & r_{23}\sigma_3 \\ r_{23}\sigma_2 & \sigma_3 \end{vmatrix}} \\
b_{12.3} &= \frac{\sigma_1}{\sigma_2} \frac{\begin{vmatrix} r_{12} & r_{23} \\ r_{13} & 1 \end{vmatrix}}{\begin{vmatrix} 1 & r_{23} \\ r_{23} & 1 \end{vmatrix}} = \frac{\sigma_1(r_{12} - r_{13}r_{23})}{\sigma_2(1 - r_{23}^2)} \quad \dots (22)
\end{aligned}$$

Similarly,

$$b_{13.2} = \frac{\sigma_1}{\sigma_3} \frac{\begin{vmatrix} 1 & r_{12} \\ r_{23} & r_{13} \end{vmatrix}}{\begin{vmatrix} 1 & r_{23} \\ r_{23} & 1 \end{vmatrix}} = \frac{\sigma_1(r_{13} - r_{12}r_{23})}{\sigma_3(1 - r_{23}^2)} \quad \dots (23)$$

If we write

$$t = \begin{vmatrix} 1 & r_{12} & r_{13} \\ r_{21} & 1 & r_{23} \\ r_{31} & r_{32} & 1 \end{vmatrix} \quad \dots (24)$$

$b_{12.3}$ and $b_{13.2}$ can be written as

$$b_{12.3} = -\frac{\sigma_1}{\sigma_2} \frac{t_{12}}{t_{11}}$$

and

$$b_{13.2} = -\frac{\sigma_1}{\sigma_3} \frac{t_{13}}{t_{11}}$$

where, t_{ij} is the cofactor of the element in the i^{th} row and j^{th} column of t .

Substituting the values of $b_{12.3}$ and $b_{13.2}$ in equation (9), we get the equation of the plane of regression of x_1 on x_2 and x_3 as

$$x_1 = -\frac{\sigma_1}{\sigma_2} \frac{t_{12}}{t_{11}} x_2 - \frac{\sigma_1}{\sigma_3} \frac{t_{13}}{t_{11}} x_3 \quad \text{where, } a = 0$$

$$\Rightarrow \frac{x_1}{\sigma_1} t_{11} + \frac{x_2}{\sigma_2} t_{12} + \frac{x_3}{\sigma_3} t_{13} = 0 \quad \dots (25)$$

Similarly, the plane of regression of x_2 on x_1 and x_3 is given by

$$\Rightarrow \frac{x_1}{\sigma_1} t_{21} + \frac{x_2}{\sigma_2} t_{22} + \frac{x_3}{\sigma_3} t_{23} = 0 \quad \dots (26)$$

and the plane of regression of x_3 on x_1 and x_2 is

$$\Rightarrow \frac{x_1}{\sigma_1} t_{31} + \frac{x_2}{\sigma_2} t_{32} + \frac{x_3}{\sigma_3} t_{33} = 0 \quad \dots (27)$$

In general the plane of regression of x_i on the remaining variable x_j ($j \neq i = 1, 2, \dots, n$) is given by

$$\Rightarrow \frac{x_1}{\sigma_1} t_{i1} + \frac{x_2}{\sigma_2} t_{i2} + \dots + \frac{x_i}{\sigma_i} t_{ii} + \dots + \frac{x_n}{\sigma_n} t_{in} = 0; i = 1, 2, \dots, n \quad \dots (28)$$

Example 1: From the given data in the following table find out

- Least square regression equation of X_1 on X_2 and X_3 .
- Estimate the value of X_1 for $X_2 = 45$ and $X_3 = 8$.

X_1	1	2	3	4	5
X_2	3	4	5	6	7
X_3	4	5	6	7	8

Solution: (i) Here X_1, X_2 and X_3 are three random variables with their respective means \bar{X}_1, \bar{X}_2 and \bar{X}_3 .

Let $X_1 - \bar{X}_1 = x_1, X_2 - \bar{X}_2 = x_2$ and $X_3 - \bar{X}_3 = x_3$

Then linear regression equation of x_1 on x_2 and x_3 is

$$x_1 = b_{12.3} x_2 + b_{13.2} x_3$$

From equation (22) and (23), we have

$$b_{12.3} = \frac{\sigma_1 (r_{12} - r_{13} r_{23})}{\sigma_2 (1 - r_{23}^2)}$$

and,

$$b_{13.2} = \frac{\sigma_1 (r_{13} - r_{12} r_{23})}{\sigma_3 (1 - r_{23}^2)}$$

The value of $\sigma_1, \sigma_2, \sigma_3, r_{12}, r_{13}$ and r_{23} can be obtained through some calculations given in the following table:

S. No.	X ₁	X ₂	X ₃	x ₁ = X ₁ -5	x ₂ = X ₂ -6	x ₃ = X ₃ -7	(x ₁) ²	(x ₂) ²	(x ₃) ²	x ₁ x ₂	x ₁ x ₃	x ₂ x ₃
1	1	3	4	-4	-3	-3	16	9	9	12	12	9
2	3	5	5	-2	-1	-2	4	1	4	2	4	2
3	4	6	6	-1	0	-1	1	0	1	0	1	0
4	7	7	9	2	1	2	4	1	4	2	4	2
5	10	9	11	5	3	4	25	9	16	15	20	12
Total	25	30	35	0	0	0	50	20	34	31	41	25

$$\bar{X}_1 = \frac{\sum X_1}{N} = \frac{25}{5} = 5$$

$$\bar{X}_2 = \frac{\sum X_2}{N} = \frac{30}{5} = 6$$

$$\bar{X}_3 = \frac{\sum X_3}{N} = \frac{35}{5} = 7$$

$$\sigma_1^2 = \frac{1}{N} \sum x_1^2 \quad \text{from equation (15)}$$

$$= \frac{1}{5}(50) = 10$$

$$\Rightarrow \sigma_1 = \sqrt{10}$$

$$= 3.162$$

$$\sigma_2^2 = \frac{1}{N} \sum x_2^2$$

$$= \frac{1}{5}(20) = 4$$

$$\Rightarrow \sigma_2 = \sqrt{4} = 2$$

$$\sigma_3^2 = \frac{1}{N} \sum x_3^2$$

$$= \frac{1}{5}(34) = 6.8$$

$$\Rightarrow \sigma_3 = \sqrt{6.8}$$

$$= 2.608$$

$$r_{12} = \frac{\sum x_1 x_2}{N \sigma_1 \sigma_2} \quad \text{from equation (17)}$$

$$= \frac{31}{5 \times 3.162 \times 2} = 0.98$$

$$r_{13} = \frac{\sum x_1 x_3}{N \sigma_1 \sigma_3}$$

$$= \frac{41}{5 \times 3.162 \times 2.608} = 0.994$$

$$r_{23} = \frac{\sum x_2 x_3}{N \sigma_2 \sigma_3}$$

$$= \frac{25}{5 \times 2 \times 2.608} = 0.959$$

Now, we have

$$\begin{aligned} b_{12.3} &= \frac{\sigma_1 (r_{12} - r_{13} r_{23})}{\sigma_2 (1 - r_{23}^2)} \\ &= \frac{3.162 \times (0.98 - 0.994 \times 0.959)}{2 \times (1 - (0.959)^2)} = 0.527 \end{aligned}$$

$$\begin{aligned} b_{13.2} &= \frac{\sigma_1 (r_{13} - r_{12} r_{23})}{\sigma_3 (1 - r_{23}^2)} \\ &= \frac{3.162 \times (0.994 - 0.98 \times 0.959)}{2.608 \times (1 - (0.959)^2)} = 0.818 \end{aligned}$$

Thus, regression equation of x_1 on x_2 and x_3 is

$$x_1 = 5.276x_2 + 0.818x_3$$

After substituting the value of x_1 , x_2 and x_3 , we will get the following regression equation of X_1 on X_2 and X_3 is

$$(X_1 - 5) = 0.527 (X_2 - 6) + 0.818 (X_3 - 7)$$

$$\Rightarrow X_1 = -3.891 + 5.276 X_2 + 0.818 X_3$$

(ii) Substituting $X_2 = 45$ and $X_3 = 8$ in regression equation

$$\Rightarrow X_1 = -3.891 + 5.276 X_2 + 0.818 X_3$$

we get estimated value of X_1 i.e. $X_1 = 26.38$

Let us solve some exercises.

E1) For the data given in the following table find out

(i) Regression equation of X_1 on X_2 and X_3 .

(ii) Estimate the value of the value of X_1 for $X_2 = 6$ and $X_3 = 8$.

X_1	2	6	8	10
X_2	4	5	9	12
X_3	4	6	10	12

10.4 PROPERTIES OF RESIDUALS

Property 1: The sum of the product of a variate and a residual is zero if the subscript of the variate occurs among the secondary subscripts of the residual, i.e. $\sum x_2 e_{1.23} = 0$. Here, subscript of the variate x i.e. 2 is appearing in the second subscripts of the $e_{1.23}$.

Proof: If the regression equation of x_1 on x_2 and x_3 is

$$x_1 = b_{12.3}x_2 + b_{13.2}x_3$$

Here, x_1 , x_2 and x_3 are measured from their respective means.

Using equation (10) in equations (11) and (13) we have following normal equations

$$\sum x_2 e_{1.23} = 0 = \sum x_3 e_{1.23}$$

Similarly, normal equation for regression lines of x_2 on x_1 and x_3 & x_3 on x_2 and x_1 are

$$\sum x_1 e_{2.13} = 0 = \sum x_3 e_{2.13}$$

$$\sum x_2 e_{3.12} = 0 = \sum x_1 e_{3.12}$$

Property 2: The sum of the product of two residuals is zero provided all the subscripts, primary as well as secondary, are appearing among the secondary subscripts of second residual i.e. $\sum x_{3.2} e_{1.23} = 0$, since primary as well as secondary subscripts (3 and 2) of the first residual is appearing among the secondary subscripts of the second residual.

Proof: We have

$$\begin{aligned} \sum x_{3.2} e_{1.23} &= \sum (x_3 - b_{32}x_2) e_{1.23} \\ &= \sum (x_3 e_{1.23} - b_{32}x_2 e_{1.23}) = 0 \end{aligned}$$

(From Property 1: $\sum x_3 e_{1.23} = 0$ and $\sum x_2 e_{1.23} = 0$)

Similarly,

$$\sum x_{2.3} e_{1.23} = 0$$

Property 3: The sum of the product of any two residuals is unaltered if all the secondary subscript of the first occur among the secondary subscripts of the second and we omit any or all of the secondary subscripts of the first.

Proof: We have

$$\sum x_{1.2}e_{1.23} = \sum x_1e_{1.23},$$

i.e. Right hand side and left hand side are equal even

$$\begin{aligned}\sum x_{1.2}e_{1.23} &= \sum (x_1 - b_{12.3}x_2)e_{1.23} \\ &= \sum (x_1e_{1.23} - b_{12.3}x_2e_{1.23}) \\ &= \sum x_1e_{1.23}\end{aligned}$$

(From Property 1, $\sum x_2e_{1.23} = 0$)

Now let us do some little exercises.

E2) Show that $\sum x_2e_{1.23} = 0$.

E3) Show that $\sum x_3e_{1.23} = 0$.

10.5 VARIANCE OF THE RESIDUALS

Let x_1 , x_2 and x_3 be three random variables then plane of regression of x_1 on x_2 and x_3 is defined as

$$x_1 = a + b_{12.3}x_2 + b_{13.2}x_3$$

Since x_1 , x_2 and x_3 are measured from their respective means so

$\sum x_1 = \sum x_2 = \sum x_3 = 0$ and we get $a = 0$ and regression equation becomes

$$x_1 = b_{12.3}x_2 + b_{13.2}x_3$$

and error of the estimate or residual is (See Section 10.2)

$$e_{1.23} = x_1 - b_{12.3}x_2 - b_{13.2}x_3$$

Now the variance of the residual is denoted by $\sigma_{1.23}^2$ and defined as

$$\sigma_{1.23}^2 = \frac{1}{N} \sum (e_{1.23} - \bar{e}_{1.23})^2$$

$$\bar{e}_{1.23} = \bar{x}_1 - b_{12.3}\bar{x}_2 + b_{13.2}\bar{x}_3 = 0 \text{ because } \sum x_1 = \sum x_2 = \sum x_3 = 0$$

and

$$\begin{aligned}\sigma_{1.23}^2 &= \frac{1}{N} \sum e_{1.23}^2 \quad \dots(29) \\ &= \frac{1}{N} \sum (x_1 - b_{12.3}x_2 - b_{13.2}x_3)(x_1 - b_{12.3}x_2 - b_{13.2}x_3) \\ &= \frac{1}{N} \sum x_1(x_1 - b_{12.3}x_2 - b_{13.2}x_3)\end{aligned}$$

$$\begin{aligned}
& -\frac{1}{N} \sum b_{12.3} x_2 (x_1 - b_{12.3} x_2 - b_{13.2} x_3) \\
& -\frac{1}{N} \sum b_{13.2} x_3 (x_1 - b_{12.3} x_2 - b_{13.2} x_3) \\
\sigma_{1.23}^2 &= \frac{1}{N} \sum (x_1^2 - b_{12.3} x_1 x_2 - b_{13.2} x_1 x_3) \\
& -\frac{1}{N} \sum (b_{12.3} x_2 x_1 - b_{12.3}^2 x_2^2 - b_{13.2} b_{12.3} x_2 x_3) \\
& -\frac{1}{N} \sum (b_{13.2} x_3 x_1 - b_{13.2} b_{12.3} x_3 x_2 - b_{13.2}^2 x_3^2)
\end{aligned}$$

We know that

$$b_{12.3} \sum x_2^2 + b_{13.2} \sum x_2 x_3 - \sum x_1 x_2 = 0$$

and

$$b_{12.3} \sum x_2 x_3 + b_{13.2} \sum x_3^2 - \sum x_1 x_3 = 0$$

(see equations (12) and (14) of Section 10.3)

Therefore,

$$\begin{aligned}
\sigma_{1.23}^2 &= \frac{1}{N} \sum x_1^2 - b_{12.3} \frac{1}{N} \sum x_1 x_2 - b_{13.2} \frac{1}{N} \sum x_1 x_3 \\
\sigma_{1.23}^2 &= \sigma_1^2 - b_{12.3} r_{12} \sigma_1 \sigma_2 - b_{13.2} r_{13} \sigma_1 \sigma_3 \\
\sigma_{1.23}^2 &= \sigma_1^2 - \frac{r_{12} - r_{13} r_{23}}{1 - r_{23}^2} \cdot \sigma_1^2 r_{12} - \frac{r_{13} - r_{12} r_{23}}{1 - r_{23}^2} \cdot \sigma_1^2 r_{13} \\
\sigma_{1.23}^2 &= \frac{\sigma_1^2}{1 - r_{23}^2} (1 - r_{23}^2 - r_{12}^2 - r_{13}^2 + 2r_{12} r_{23} r_{13}) \quad \dots (30)
\end{aligned}$$

10.6 SUMMARY

In this unit, we have discussed:

1. The Yule's notation for trivariate distribution;
2. The plane of regression for trivariate distribution;
3. How to get normal equations for the regression equation of x_1 on x_2 and x_3 ;
4. The properties of residuals;
5. The variance of residuals; and
6. How to find the estimates of dependent variable of regression equations of three variables.

10.7 SOLUTIONS / ANSWERS

E1) (i) Here X_1 , X_2 and X_3 are three random variables with their respective means \bar{X}_1 , \bar{X}_2 and \bar{X}_3 .

Let $X_1 - \bar{X}_1 = x_1$, $X_2 - \bar{X}_2 = x_2$ and $X_3 - \bar{X}_3 = x_3$

Then linear regression equation of x_1 on x_2 and x_3 is

$$x_1 = b_{12.3}x_2 + b_{13.2}x_3$$

From equation (22) and (23), we have

$$b_{12.3} = \frac{\sigma_1(r_{12} - r_{13}r_{23})}{\sigma_2(1 - r_{23}^2)}$$

and

$$b_{13.2} = \frac{\sigma_1(r_{13} - r_{12}r_{23})}{\sigma_3(1 - r_{23}^2)}$$

σ_1 , σ_2 , σ_3 , r_{12} , r_{13} and r_{23} can be obtained through the following table.

S. No.	X_1	X_2	X_3	$x_1 =$ $X_1 - 6.5$	$x_2 =$ $X_2 - 7.5$	$x_3 =$ $X_3 - 8$	$(x_1)^2$	$(x_2)^2$	$(x_3)^2$	x_1x_2	x_1x_3	x_2x_3
1	2	4	4	-4.5	-3.5	-4	20.25	12.25	16	15.75	18	14
2	6	5	6	-0.5	-2.5	-2	0.25	6.25	4	1.25	1	5
3	8	9	10	1.5	1.5	2	2.25	2.25	4	2.25	3	3
4	10	12	12	3.5	4.5	4	12.25	20.25	16	15.75	14	18
Total	26	30	32	0	0	0	35	41	40	35	36	40

$$\bar{X}_1 = \frac{\sum X_1}{N} = \frac{26}{4} = 6.5$$

$$\bar{X}_2 = \frac{\sum X_2}{N} = \frac{30}{4} = 7.5$$

$$\bar{X}_3 = \frac{\sum X_3}{N} = \frac{32}{4} = 8$$

$$\sigma_1^2 = \frac{1}{N} \sum x_1^2 \text{ from equation (15)}$$

$$= \frac{1}{4}(35) = 8.75$$

$$\Rightarrow \sigma_1 = \sqrt{35} = 2.958$$

$$\begin{aligned}\sigma_2^2 &= \frac{1}{N} \sum x_2^2 \\ &= \frac{1}{4}(41) = 10.25 \\ \Rightarrow \sigma_2 &= \sqrt{10.25} = 3.202\end{aligned}$$

$$\begin{aligned}\sigma_3^2 &= \frac{1}{N} \sum x_3^2 \\ &= \frac{1}{4}(40) = 10\end{aligned}$$

$$\Rightarrow \sigma_3 = \sqrt{10} = 3.162$$

$$\begin{aligned}r_{12} &= \frac{\sum x_1 x_2}{N \sigma_1 \sigma_2} \quad \text{from equation (17)} \\ &= \frac{35}{4 \times 2.958 \times 3.202} = 0.924\end{aligned}$$

$$\begin{aligned}r_{13} &= \frac{\sum x_1 x_3}{N \sigma_1 \sigma_3} \\ &= \frac{36}{4 \times 2.958 \times 3.162} = 0.962\end{aligned}$$

$$\begin{aligned}r_{23} &= \frac{\sum x_2 x_3}{N \sigma_2 \sigma_3} \\ &= \frac{40}{4 \times 3.202 \times 3.162} = 0.988\end{aligned}$$

Now, we have

$$\begin{aligned}b_{12.3} &= \frac{\sigma_1 (r_{12} - r_{13} r_{23})}{\sigma_2 (1 - r_{23}^2)} \\ &= \frac{2.958 \times (0.924 - 0.962 \times 0.988)}{3.202 \times \{1 - (0.988)^2\}} = -1\end{aligned}$$

$$\begin{aligned}b_{13.2} &= \frac{\sigma_1 (r_{13} - r_{12} r_{23})}{\sigma_3 (1 - r_{23}^2)} \\ &= \frac{2.958 \times (0.962 - 0.924 \times 0.988)}{3.162 \times \{1 - (0.988)^2\}} = 1.9\end{aligned}$$

Thus, regression equation of x_1 on x_2 and x_3 is

$$x_1 = -x_2 + 1.9x_3$$

After substituting the value of x_1 , x_2 and x_3 , we will get the following regression equation of X_1 on X_2 and X_3 is

$$(X_1 - 6.5) = -(X_2 - 7.5) + 1.9(X_3 - 8)$$

$$\Rightarrow X_1 = -1.2 - X_2 + 1.9X_3$$

(ii) Substituting $X_2 = 6$ and $X_3 = 8$ in regression equation

$$\Rightarrow X_1 = -1.2 - X_2 + 1.9X_3$$

we get estimated value of X_1 i.e. $X_1 = 8$

E2) **Hint:** According to the property 1: $\sum x_2 e_{1.23} = 0$, since subscript of the variate x_2 i.e. 2 is appearing in the second subscript of $e_{1.23}$ i.e. in 23.

E3) **Hint:** According to the property 1: $\sum x_3 e_{1.23} = 0$, since subscript of the variate x_3 i.e. 3 is appearing in the second subscript of $e_{1.23}$ i.e. in 23.