6.4.1

AI25BTECH11001 - ABHISEK MOHAPATRA

Question: Fit a straight line trend by the method of least squares and find the trend value for the year 2008 using the data from the given table

TABLE : Show yearly trend of production of goods in lakh tonnes

Year	Production (in lakh tonnes)
2001	30
2002	35
2003	36
2004	32
2005	37
2006	40

Solution: Equation of a straight line be:

$$y = mx + c \tag{1}$$

$$y = \begin{pmatrix} c & m \end{pmatrix} \begin{pmatrix} 1 \\ x \end{pmatrix} \tag{2}$$

Let this equation be

$$y = \mathbf{N}^\top \mathbf{X} \tag{3}$$

Let the given value of the years be a column vector $\mathbf{X_0}$ and the corresponding values of production be $\mathbf{D}$.

let $\mathbf{X} = \begin{pmatrix} (1)_{n\times 1} & \mathbf{X_0} \end{pmatrix}$.

let $\mathbf{X} = \begin{pmatrix} \mathbf{x_1} & \mathbf{x_2} & \dots & \mathbf{x_n} \end{pmatrix}^\top$ and $\mathbf{D} = \begin{pmatrix} y_1 & y_2 & \dots & y_n \end{pmatrix}^\top$

so sum of the square of error = e =

$$\Sigma |y_i - \mathbf{N}^\top \mathbf{x_i}|^2 \tag{4}$$

$$= \Sigma \left(y_i - \mathbf{N}^\top \mathbf{x_i} \right) \left(y_i - \mathbf{N}^\top \mathbf{x_i} \right) \tag{5}$$

$$= \Sigma \left((y_i)^2 - 2y_i^\top \mathbf{N}^\top \mathbf{x_i} + (\mathbf{N}^\top \mathbf{x_i})^2 \right) \tag{6}$$

for this to be minimum , $\nabla_{\mathbf{N}}$ e = 0

$$\nabla_{\mathbf{N}} e = \Sigma \left(-2y_i \mathbf{x_i} + 2 \left(\mathbf{N}^\top \mathbf{x_i} \right) \mathbf{x_i} \right) = 0 \tag{7}$$

$$\nabla_{\mathbf{N}} e = \Sigma \left(-2y_i \mathbf{x_i} + 2 \left(\mathbf{x_i} \mathbf{x_i}^\top \right) \mathbf{N} \right) = 0 \tag{8}$$

so,

$$\left(\Sigma \mathbf{x_i} \mathbf{x_i}^\top \right) \mathbf{N} = \Sigma y_i \mathbf{x_i} \tag{9}$$

Or,

$$\mathbf{N} = \left(\Sigma \mathbf{x_i} \mathbf{x_i}^\top \right)^{-1} \left(\Sigma y_i \mathbf{x_i} \right) \tag{10}$$

Or,

$$\mathbf{N} = \left(\mathbf{X}^{\top}\mathbf{X}\right)^{-1}\left(\mathbf{X}^{\top}\mathbf{D}\right) \tag{11}$$

Given,

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 2001 & 2002 & 2003 & 2004 & 2005 & 2006 \end{pmatrix}^{\top} \tag{12}$$

And,

$$\mathbf{D} = \begin{pmatrix} 30 & 35 & 36 & 32 & 37 & 40 \end{pmatrix}^{\top} \tag{13}$$

$$\mathbf{X}^{\top}\mathbf{D} = \begin{pmatrix} 30 + 35 + 36 + 32 + 37 + 40 \\ 2001 \times 30 + ... + 2006 \times 40 \end{pmatrix} = \begin{pmatrix} 210 \\ 420761 \end{pmatrix} \tag{14}$$

$$\left(\mathbf{X}^{\top}\mathbf{X}\right) = \begin{pmatrix} 6.0 & 12021.0 \\ 12021.0 & 24084091.0 \end{pmatrix} \tag{15}$$

$$\left(\mathbf{X}^{\top}\mathbf{X}\right)^{-1} = \begin{pmatrix} 229372.295 & -114.485714 \\ -114.485714 & 0.0571429 \end{pmatrix} \tag{16}$$

Putting the matrices,

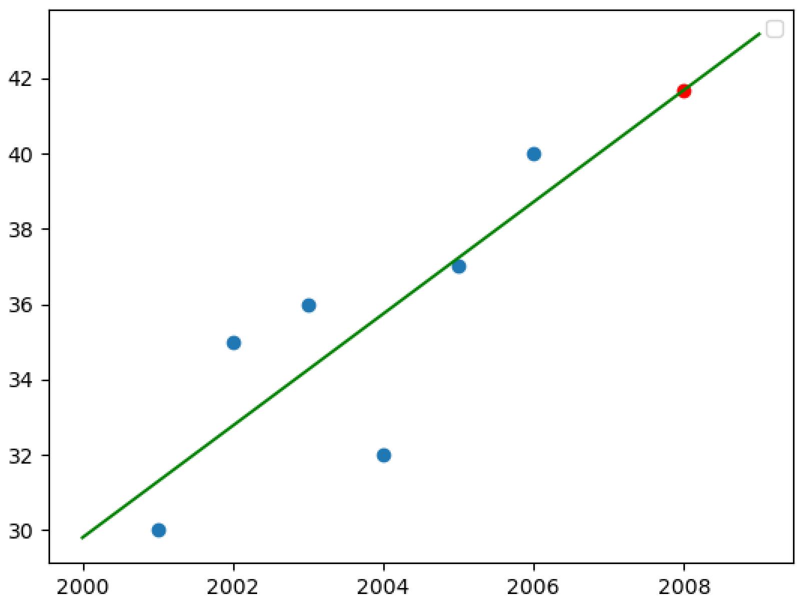
$$\mathbf{N} = \begin{pmatrix} 229372.295 & -114.485714 \\ -114.485714 & 0.0571429 \end{pmatrix} \begin{pmatrix} 210 \\ 420761 \end{pmatrix} = \begin{pmatrix} -2941.628571 \\ 1.485714 \end{pmatrix} \tag{17}$$

So,

$$y = \mathbf{N}^{\top} \begin{pmatrix} 1 \\ 2008 \end{pmatrix} = 41.685714 \tag{18}$$

Therefore, expected value is 41.685714.

Graph:


42
40
38
36
34
32
30
2000
2002
2004
2006
2008