

Variance formula

$$\sigma^2 = \frac{\sum (x - \bar{x})^2}{n}$$

Standard deviation formula

$$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{n}} = SD$$

① Find the mean, median & variance of

a) 44, 46, 48, 45, 47

b) 34, 46, 59, 39, 52

Step 1: Sort the values

44, 45, 46, 47, 48

Step 1: Sort the values

34, 39, 46, 52, 59

Step 2: Find the mean

$$\begin{aligned}\text{mean} = \bar{X} &= \frac{\sum x}{n} \\ &= \frac{230}{5} = \underline{\underline{46}}\end{aligned}$$

Step 2: Find the mean

$$\begin{aligned}\bar{X} &= \frac{\sum x}{n} \\ &= \frac{230}{5} \\ &= \underline{\underline{46}}\end{aligned}$$

Step 3: Find the median

median = 46

Step 3: Find the median

median = 46

Step 4: Find the variance

$$\begin{aligned}\sigma^2 &= \frac{\sum (x - \bar{x})^2}{n} \\ &= \frac{((-2) + (-1) + 0 + 1 + 2)^2}{5} \\ &= \frac{4 + 1 + 0 + 1 + 4}{5} \\ &= \frac{10}{5} = \underline{\underline{2}}\end{aligned}$$

Step 4: Find the variance

$$\begin{aligned}\sigma^2 &= \frac{\sum (x - \bar{x})^2}{n} \\ &= \frac{(12 + (-7) + 0 + 6 + 13)^2}{5} \\ &= \frac{144 + 49 + 0 + 36 + 169}{5} \\ &= \frac{398}{5} = \underline{\underline{79.6}}\end{aligned}$$

CALCULATING OUTLIERS

① 7 3 35 14 9 7 8 12 2 (odd set)

Step 1 :- Sort the values

2 3 7 7 8 9 12 14 35

Step 2 :- Find the median

$$\text{median} = \underline{\underline{8}}$$

Step 3 :- Calculate Q_1 & Q_3

2 3 7 7 8 9 12 14 35

$$Q_2 = 8$$

$$Q_1 = \frac{3+7}{2} = \frac{10}{2} = \underline{\underline{5}}$$

$$Q_3 = \frac{12+14}{2} = \frac{26}{2} = \underline{\underline{13}}$$

Step 4 :- Calculate IQR (Interquartile Range)

$$IQR = Q_3 - Q_1$$

$$= 13 - 5$$

$$= \underline{\underline{8}}$$

Step 5 :- Calculate upper Bound & lower Bound

$$\text{upper Bound} = Q_3 + 1.5 * IQR$$

$$= 13 + 1.5 * 8$$

$$= 13 + 12$$

$$= \underline{\underline{25}}$$

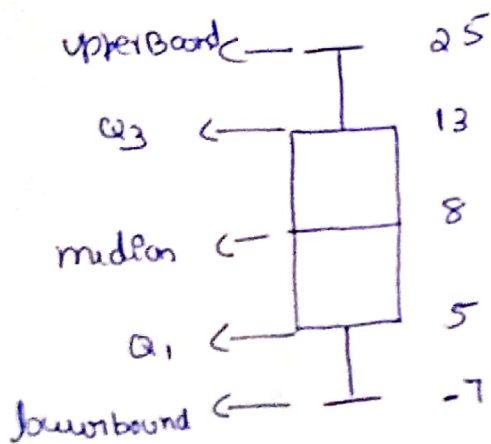
$$\text{lower bound} = Q_1 - 1.5 \times IQR$$

$$= 5 - 1.5 \times 8$$

$$= 5 - 12$$

$$= -7$$

step 6:- Identify the outliers



Therefore outliers in the data set are 35

Q 12 5 22 30 7 36 14 42 53 25 (even set)

step 1:- sort the values

5 7 12 14 22 25 30 36 42 53

step 2:- find the median

$$\frac{22 + 25}{2} = \underline{\underline{23.5}}$$

step 3:- calculate Q_1 & Q_3

5 7 12 14 22 25 30 36 42 53

$$Q_1 = 12$$

$$Q_3 = 36$$

Step 4 :- calculate IQR (Inter quartile Range)

$$IQR = Q_3 - Q_1$$

$$= 36 - 12$$

$$= \underline{\underline{24}}$$

Step 5 :- calculate upper Bound & Lower Bound

$$\text{Upper Bound} = Q_3 + 1.5 * IQR$$

$$= 36 + 1.5 * 24$$

$$= 36 + 36$$

$$= \underline{\underline{72}}$$

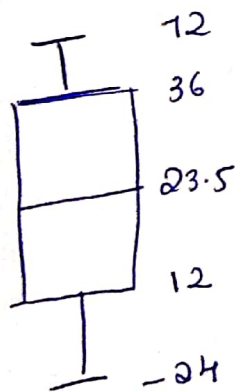
$$\text{Lower Bound} = Q_1 - 1.5 * IQR$$

$$= 12 - 1.5 * 24$$

$$= 12 - 36$$

$$= \underline{\underline{-24}}$$

Step 6 :- Identify the outliers



Therefore there are no outliers in the dataset.

APPLY OLS ALGORITHM AND FIND THE RELATIONSHIP b/w VARIABLES

X & Y [SIMPLE LINEAR REGRESSION]

① SLNO	1	2	3	4	5	6	7	8	9
X	14	16	14	15	16	15	13	12	13
Y	40	45	42	44	48	46	48	40	52

Step 1:- Find mean of x & y

$$\bar{x} = \frac{\sum x}{n} = \frac{128}{9} = \underline{\underline{14.2}}$$

$$\bar{y} = \frac{\sum y}{n} = \frac{405}{9} = \underline{\underline{45}}$$

Step 2:- Slope formula

$$m = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2}$$

Step 3 :- calculate the values to substitute into formula.

x	y	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})^2$	$(x - \bar{x})(y - \bar{y})$
14	40	-0.2	-5	0.04	1
16	45	1.8	0	3.24	0
14	42	-0.2	-3	0.04	0.6
15	44	0.8	-1	0.64	-0.8
16	48	1.8	3	3.24	5.4
15	46	0.8	1	0.64	0.8
13	48	-1.2	3	1.44	-3.6
12	40	-2.2	-5	4.84	11
13	52	-1.2	7	1.44	-8.4
				<u>15.56</u>	<u>6</u>

Step 4:- substitute into formula

$$m = \frac{6}{15.56}$$

$$m = \underline{\underline{0.38}}$$

Step 5 :- find the constant

$\bar{y} = mx + C \Rightarrow$ straight line formula

$$\bar{y} - m\bar{x} = C$$

$$\Rightarrow C = \bar{y} - m\bar{x}$$

$$C = 45 - (0.38) \times 14.2$$

$$C = 45 - 5.396$$

$$C = \underline{\underline{39.604}} = \bar{y} = \underline{\underline{0.38x + 39.604}}$$

Step 6 :- relation between x & y is

$$y - \bar{y} = m(x - \bar{x})$$

$$y - 45 = 0.38(x - 14.2)$$

$$y = 0.38x - 5.396 + 45$$

$$\underline{\underline{y = 0.38x - 5.396}}$$

MULTI LINEAR REGRESSION

For the given dataset apply multilinear regression analysis.
If age = 45, income = 40.4, savings = ?

<u>SLNO</u>	<u>age (x_1)</u>	<u>income (x_2)</u>	<u>savings (y)</u>
1	50	40.8	12.6
2	46	40.	14
3	52	44	10
4	44	42	16
5	43	38	14

Step 1 :- Identify the dependent & independent variables
dependent variable = Savings

Independent variable :- age, income.

Step 1 :- list down the formulas

→ straight line formula = $y = mx + c$

$$\rightarrow \hat{y} = a + b_1 x_1 + b_2 x_2 + \dots + b_n x_n + \epsilon_0$$

\downarrow m Slope \swarrow Regression parameters \downarrow error

$$\rightarrow a = \bar{y} - b\bar{x}$$

$$\rightarrow b = \frac{\text{covariance}(x, y)}{\text{var}(x, y)}$$

Step 2: Find the values

Sl no age (x_1) income (x_2) savings (y)

1
2
3
4
5

Step 2:- find the values

<u>y</u>	<u>x₁</u>	<u>x₂</u>	<u>x₁²</u>	<u>x₂²</u>	<u>x₁y</u>	<u>Σy</u>	<u>Σx₁x₂</u>
12.6	50	40.8	2500	1664.64	630	514.08	2040
14	46	40	2116	1600	644	560	1840
10	52	44	2704	1936	520	440	2288
16	44	42	1936	1764	704	672	1848
14	43	38	1849	1444	602	532	1634
<u>66.6</u>	<u>235</u>	<u>204.8</u>	<u>11105</u>	<u>8408.64</u>	<u>3100</u>	<u>2718.08</u>	<u>9650</u>

$$1] \sum x_1^2 = \sum x_1^2 - \frac{(\sum x_1)^2}{n} = 11105 - \frac{235 \times 235}{5} = \underline{\underline{60}}$$

$$2] \sum x_2^2 = \sum x_2^2 - \frac{(\sum x_2)^2}{n} = 8408.64 - \frac{(204.8)^2}{5} = \underline{\underline{20.0320}}$$

$$3] \sum x_1 y = \sum x_1 y - \frac{\sum x_1 \sum y}{n} = 3100 - \frac{235 \times 66.6}{5} = \underline{\underline{-30.2}}$$

$$4] \sum x_2 y = \sum x_2 y - \frac{\sum x_2 \sum y}{n} = 2718.08 - \frac{204.8 \times 66.6}{5} = \underline{\underline{-9.8560}}$$

$$5] \sum x_1 x_2 = \sum x_1 x_2 - \frac{(\sum x_1)(\sum x_2)}{n} = 9650 - \frac{(235)(204.8)}{5} = \underline{\underline{24.4}}$$

Step 3:- write down the main formulas & substitute the above values into it

$$b_1 = \frac{(\sum x_2^2)(\sum x_1 y) - \sum x_1 x_2 \sum x_2 y}{\sum x_1^2 \sum x_2^2 - (\sum x_1 x_2)^2}$$

~~$$b_1 = \frac{(\sum x_1 y) - (\sum x_1)(\sum y)}{(\sum x_1^2) - (\sum x_1)^2}$$~~

$$b_1 = \frac{(20.0320)(-30.2) - (24.4)(-9.8560)}{(60)(20.0320) - (24.4)^2}$$

$$= \frac{-604.9664 - (-240.4864)}{1201.92 - 595.36}$$

$$= \frac{-364.48}{606.56} = -0.6009$$

$$b_2 = \frac{(\sum x_1^2)(\sum x_2 y) - \sum x_1 x_2 \sum x_1 y}{(\sum x_2^2)(\sum x_1^2) - (\sum x_1 x_2)^2}$$

$$= \frac{(60)(-9.8560) - (24.4)(-30.2)}{(20.0320)(60) - (24.4)^2}$$

$$= 0.2399$$

step 4 :- calculate a

$$\bar{y} = a + b_1 \bar{x}_1 + b_2 \bar{x}_2 \quad \left[\bar{x}_1 = \frac{\sum x_1}{n} \quad \bar{x}_2 = \frac{\sum x_2}{n} \right]$$

$$a = \bar{y} - b_1 \bar{x}_1 - b_2 \bar{x}_2 \quad \left[\bar{y} = \frac{\sum y}{n} \right]$$

$$a = 13.32 - (-0.6009)(47) + (0.2399)(40.96) = \cancel{31.5586}$$

$$31.7360$$

Step 5 :- find the \bar{y} (predictable)

$$\bar{y} = a + b_1 \bar{x}_1 + b_2 \bar{x}_2$$

$$= 31.7360 + (-0.6009)(47) + (0.2399)(4096)$$

$$= \underline{\underline{13.32}}$$

Step 6 :- calculate if age = 45, income = 45.4 savings = ?

$$y = a + b_1 x_1 + b_2 x_2$$

$$y = 31.7360 + (-0.6009)(45) + (0.2399)(45.4)$$

$$y = \underline{\underline{14.6274}}$$