Statistics

Introduction

- Statistics is the area of science that deals with collection, organization, analysis, and interpretation of data.
- It also deals with methods and techniques that can be used to draw conclusions about the characteristics of a large number of data points--commonly called a population.
- By using a smaller subset of the entire data.

Data Representation

There are two types of Data Representation:

- Graphical Representation
- Numerical Representation

Numerical Representation

A fundamental concept in summary statistics is that of a central value for a set of observations and the extent to which the central value characterizes the whole set of data. Measures of central value such as the mean or median must be coupled with measures of data dispersion.

- Measures of Central Tendencies :

1) Arithmetic Mean	**2) Geometric Mean**
3) Harmonic Mean	**4) Mode**
5) Median	

Measures of Central Tendencies

Measures Of Central Tendency	Formula
Arithmetic Mean	**For individual Data** $A.M. = \bar{x} = \frac{\sum x_i}{n}$ **For Frequency Distribution** $A.M. = \bar{x} = \left(\frac{\sum f_i x_i}{n}\right)$
Variance	**For individual Data** $\sigma^2 = \frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2$ **For Frequency Distribution** $\sigma^2 = \frac{\sum f_i x_i^2}{n} - \left(\frac{\sum f_i x_i}{n}\right)^2$
Standard Deviation	$\sigma = \sqrt{\sigma^2}$
Coefficient Of Variation	$C.V. = \frac{\sigma}{\bar{x}} * 100$

Example:

1) Find standard deviation of data 15,10,05.

Solution:

X	**15**	**10**	**5**	**Σ(X)=30**
X^2	**225**	**100**	**25**	**Σ (X^2)=350**

$$\bar{x} = \frac{\sum x}{3} = \frac{30}{3} = 10$$

$$\sigma = \sqrt{\frac{x^2}{3} - \left(\bar{x}\right)^2} = \sqrt{\frac{350}{3} - 100} = 4.082$$

Example:

2) Find the coefficient of variation for the data 1,3,5,7,9 is :

Solution:

$$\text{Coefficient of Variation} = \frac{\sigma}{A.M.}\times 100$$

$$A.M. = \frac{1+3+5+7+9}{5} = \frac{25}{5} = 5$$

$$\sigma = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2} = \sqrt{\frac{1+9+25+49+81}{5} - \left(\frac{25}{5}\right)^2}$$

$$= \sqrt{\frac{165}{5} - 25} = \sqrt{8} = 2.8284$$

$$\text{Coefficient of Variation} = \frac{\sigma}{A.M.}\times 100 = \frac{2.8284}{5}\times 100 = 56.56$$

Moments

❑ Moments about Origin OR Central Moments :

The r^{th} moment about the mean of a distribution is denoted by μ_r and is given by,

$$\mu_r = \frac{1}{N}\sum f\left(x-\bar{x}\right)^r$$

where $\bar{x}$ is A.M. of the distribution.

Note that : $\mu_0 = 1, \mu_1 = 0$ and

$\mu_2 =$ Second Central Moment about Mean is Variance.

Moments

❑ Moments about any number 'A' OR Raw Moments :

The r^{th} moment about arbitrary number 'A' of a distribution is denoted by μ_r' and is given by,

$$\mu_r' = \frac{1}{N}\sum f\left(x - A\right)^r$$

where 'A' is any arbitrary number

Note that : $\mu_0' = 1$

$$\mu_1' = \bar{x} - A$$

Relation Between μ_r and μ_r'

$$\mu_2 = \mu_2' - (\mu_1')^2$$

$$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2(\mu_1')^3$$

$$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'(\mu_1')^2 - 3(\mu_1')^4$$

Skewness

Skewness signifies departure from symmetry. We study skewness to have an idea about the shape of the curve which we draw with the given data.

▪ If the frequency curve stretches to the right as in Figure (a) then the distribution is right skewed or is said to have positive skewness.

▪If the frequency curve stretches to the left as in Fig (b) then the distribution is left skewed or said to have negative skewness.

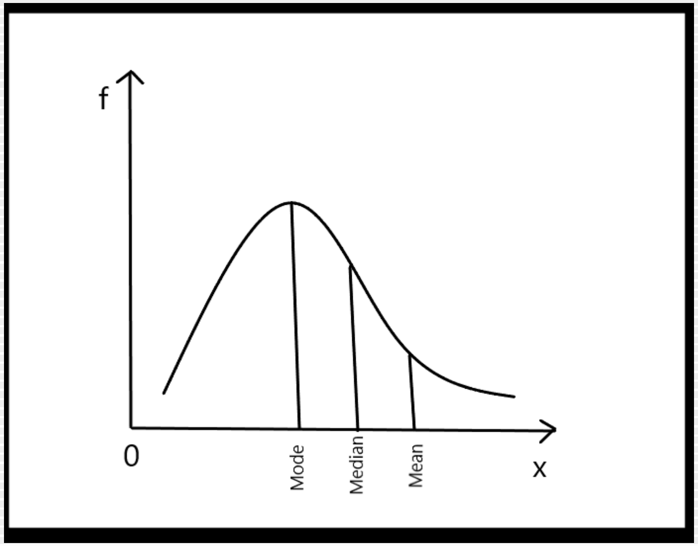


Figure (a)

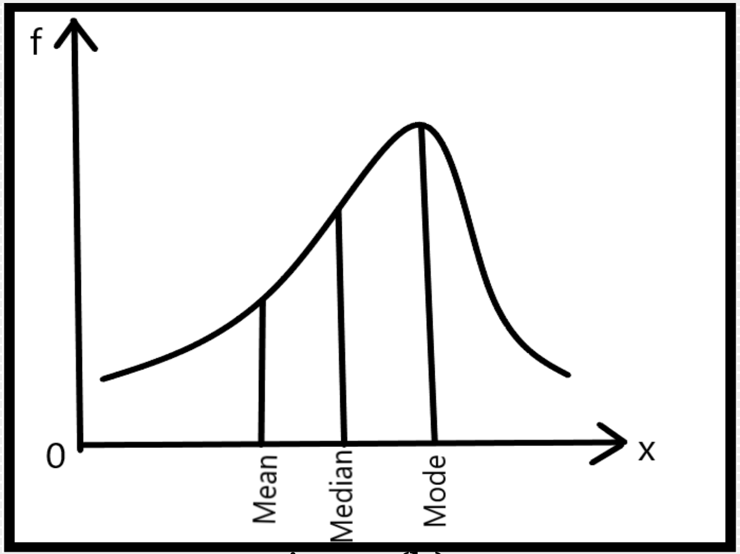


Figure (b)

Skewness :

The different measures of skewness are:

(i) Skewness $= \dfrac{3(Mean - Median)}{Std.Deviation}$

(ii) Coefficient of Skewness, $\beta_1 = \dfrac{(\mu_3)^2}{(\mu_2)^3}$

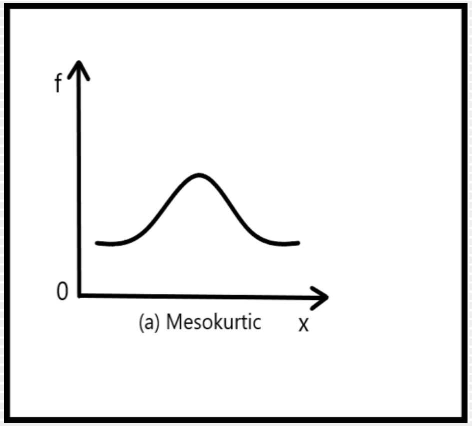
Kurtosis

Kurtosis gives an idea of the flatness or peakedness of the curve. It is measured by the coefficient of kurtosis given by ,

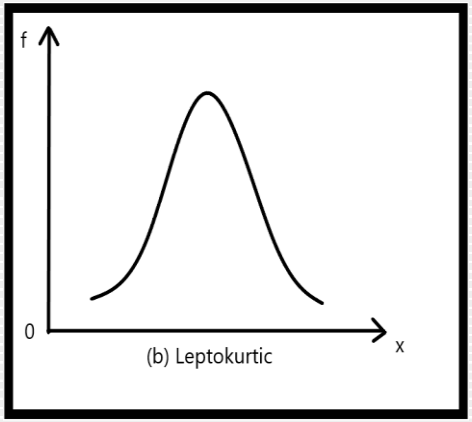
$$\beta_2 = \frac{\mu_4}{(\mu_2)^2}$$

The curve of Fig. (a) which is neither flat nor peaked is called the normal curve or Mesokurtic curve. For a normal distribution, $\beta_2 = 3$.

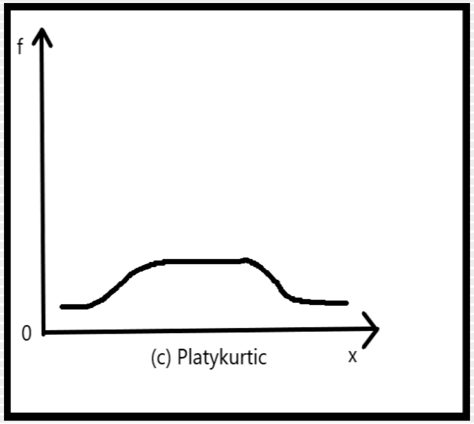
The curve which is flatter than the normal curve is called Platykurtic and that of Fig.(b) which is more peaked is called Leptokurtic. For Platykurtic Curves $\beta_2 < 3$, for Leptokurtic curves $\beta_2 > 3$.



(a) Mesokurtic



(b) Leptokurtic



(c) Platykurtic

Multiple Choice Questions

1) If the data is presented in the form of frequency distribution then 1st moment μ_1 about the arithmetic mean $\bar{x}$ about the distribution is $\left(N = \sum f\right)$

[A] 1 [B] σ^2 [C] 0 [D] $\frac{1}{N}\sum f\left(x-\bar{x}\right)^3$

2) If μ_1' and μ_2' are the first two moments of the distribution about certain number then second moment of the distribution about the arithmetic mean is given by

[A] $\mu_2' - \left(\mu_1'\right)^2$ [B] $2\mu_2' - \mu_1'$ [C] $\mu_2' + \left(\mu_1'\right)$ [D] $\mu_2' + 2\left(\mu_1'\right)$

3) If be the first moment of the distribution about any number A then arithmetic mean is given by
[A] $\mu_1' + A$ [B] μ_1' [C] $\mu_1' - A$ [D] $\mu_1' A$

4) Second moment μ_2 about mean is
[A] Mean [B] Standard deviation [C] Variance [D] Mean deviation

5) Coefficient of skewness β_1 is given by

[A] $\frac{\mu_2^3}{\mu_3^2}$ [B] $\frac{\mu_1^2}{\mu_2^3}$ [C] $\frac{\mu_2^2}{\mu_3^2}$ [D] $\frac{\mu_3^2}{\mu_2^3}$

Multiple Choice Questions

6) For a distance coefficient of kurtosis 2.5, this distribution is

[A] Leptokurtic [B] Mesokurtic [C] Platykurtic [D] None of these

7) For a distribution coefficient of kurtosis this distribution is

[A] Leptokurtic [B] Mesokurtic [C] Platykurtic [D] None of these

8) The first four moments of a distribution about the mean are 0, 16, -64 and 162. Standard deviation of a distribution is

[A] 21 [B] 12 [C] 16 [D] 4

9) The Standard deviation and Arithmetic Mean of three distributions x, y, z are as follow:

Arithmetic Mean	Standard deviation
X=18.0	5.4
Y=22.5	4.5
Z=24.0	6.0

The more stable distribution is

[A] x [B] y [C] z [D] x and z

Examples

Q. The first four moments of a distribution about 2 are 1, 2.5, 5.5, 16. Calculate the first four moments about mean, A.M, S.D, coefficient of skewness and coefficient of kurtosis. **(May 2014)**

Solution:

Given $\mu_1' = 1, \mu_2' = 2.5, \mu_3' = 5.5, \mu_4' = 16$

To use the relations between μ_r' & μ_r

$$\mu_1 = 0$$

$$\mu_2 = \mu_2' - \left(\mu_1'\right)^2 = 2.5 - 1 = 1.5$$

$$\mu_3 = \mu_3' - 3\mu_1'\mu_2' + 2\left(\mu_1'\right)^3 = 5.5 - 3*2.5 + 2 = 0$$

$$\mu_4 = \mu_4' - 4\mu_1'\mu_3' + 6\left(\mu_1'\right)^2\mu_2' - 3\left(\mu_1'\right)^4 = 16 - 4*5.5 + 6*2.5 - 3 = 6$$

$$S.D. = \sqrt{\mu_2} = 1.2247$$

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = 0$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = 2.666$$

Correlation

To measure the intensity or degree of linear relationship between two variables , Karl Pearson developed a formula called correlation coefficient .

Correlation coefficient between two variables x and y denoted by r (x,y) is defined as

$$r(x,y) = \frac{\operatorname{cov}(x,y)}{\sigma_x \sigma_y}.$$

In a bivariate distribution if (x_i, y_i) take the values $(x_1, y_1), (x_2, y_2) \ldots\ldots\ldots\ldots (x_n, y_n)$

$$\operatorname{cov}(x,y) = \frac{1}{n}\sum (x_i - \bar{x})(y_i - \bar{y})$$

where $\bar{x}, \bar{y}$ are arithmetic mean for x and y series respectively.

Regression Lines

Regression line of y on x :

$$y-\bar{y}=r\frac{\sigma_y}{\sigma_x}(x-\bar{x})=b_{yx}(x-\bar{x})$$

Regression line of x on y :

$$x-\bar{x}=r\frac{\sigma_x}{\sigma_y}(y-\bar{y})=b_{xy}(y-\bar{y})$$

Multiple Choice Questions

1) Slope of regression line of y on x is

[A] $r(x,y)$ [B] $r\frac{\sigma_y}{\sigma_x}$ [C] $\frac{\sigma_y}{\sigma_x}$ [D] $r\frac{\sigma_x}{\sigma_y}$

2) Slope of regression line of x on y is

[A] $r(x,y)$ [B] $r\frac{\sigma_x}{\sigma_y}$ [C] $\frac{\sigma_x}{\sigma_y}$ [D] $r\frac{\sigma_y}{\sigma_x}$

3) If b_{xy} and b_{yx} are the regression coefficient x on y and y on x respectively then the coefficient of correlation $r(x,y)$ is given by

[A] $\sqrt{b_{xy}+b_{yx}}$ [B] $\sqrt{\frac{b_{xy}}{b_{yx}}}$ [C] $b_{xy}b_{yx}$ [D] $\sqrt{b_{xy}b_{yx}}$

4) If $\sum xy = 1242, \bar{x} = -5.1, \bar{y} = -10, n = 10,$ then Cov(x,y) is

[A] 67.4 [B] 83.9 [C] 58.5 [D] 73.2

5) If the two regression coefficients are 0.16 and 4 then the correlation coefficient is

[A] 0.08 [B] -0.8 [C] 0.8 [D] 0.64

Multiple Choice Questions

6) If $\sum xy = 2800, \bar{x} = 16, \bar{y} = 16, n = 10$ of x is 36 and variance of y is 25 then correlation coefficient is equal to

[A] 0.95 [B] 0.73 [C] 0.8 [D] 0.65

7) Coefficient of correlation between the variables x and y is 0.8 and their covariance is 20, the variance of x is 16. Standard deviation of y is

[A] 6.75 [B] 6.25 [C] 7.5 [D] 8.25

8) The regression lines are 9x+y=15 and 4x+y=5. Correlation r(x,y) is given by

[A] 0.444 [B] -0.11 [C] 0.663 [D] 0.7

9) For a given set of Bivariate data $\bar{x} = 53.2, \bar{y} = 27.9$, Regression coefficient of y on x= -1.5. By using line of regression y on x the most probable value of y when x is 60 is

[A] 157.7 [B] 137.7 [C] 197.7 [D] 217.7

10) Given the following data $\bar{x} = 36, \bar{y} = 85, \sigma_x = 11, \sigma_y = 8, r = 0.66.$ By using line of regression x on y, the most probable value of x when y=75 is

[A] 29.143 [B] 24.325 [C] 31.453 [D] 26.925

Examples

Q1. Calculate the coefficient of correlation for the following data. **(May - 2014)**

X	1	2	3	4	5	6	7	8	9
y	9	8	10	12	11	13	14	16	15

Solution: Given

x	y	xy	y^2	x^2
1	9	9	81	1
2	8	16	64	4
3	10	30	100	9
4	12	48	144	16
5	11	55	121	25
6	13	78	169	36
7	14	98	196	49
8	16	128	256	64
9	15	135	255	81

$\therefore$

$$\sum X = 45, \sum Y = 108, \sum XY = 597, \sum X^2 = 285, \sum Y^2 = 1386$$

Here n = 9

$$\overline{X} = \frac{\sum X}{n} = \frac{45}{9} = 5, \overline{Y} = \frac{\sum Y}{n} = \frac{108}{99} = 12$$

$$\sigma_x{}^2 = \frac{\sum X^2}{n} - (\overline{X})^2 = 6.67, \sigma_y{}^2 = \frac{\sum Y^2}{n} - (\overline{Y})^2 = 10$$

Examples

$$\operatorname{cov}(x, y) = \frac{\sum XY}{n} - \overline{X}\overline{Y} = 6.333$$

$$r(x, y) = \frac{\operatorname{cov}(x, y)}{\sigma_x . \sigma_y} = 0.775$$

Examples

Q2.Calculate the coefficient of correlation form the following information: **(May 2016)**

$$\sum x = 40, \sum x^2 = 190, \sum y^2 = 200, \sum xy = 150, \sum y = 40, n = 10$$

Solution : $$r(x,y) = \frac{\text{cov}(x,y)}{\sigma_x \sigma_y}$$

$$\text{cov}(x,y) = \frac{1}{n}\sum xy - \bar{x}\,\bar{y}$$

$$\bar{x} = \frac{\sum x}{n} = \frac{40}{10} = 4, \bar{y} = 4$$

$$\sigma_x = \sqrt{\frac{190}{10} - 16} = 5.477, \quad \sigma_y = \sqrt{\frac{200}{10} - 16} = 6.325$$

Cov(x,y) =15 - 16 = -1

$$r(x,y) = \frac{-1}{34.64} = -0.2886$$

Examples

Q3.The regression equation are given by $8x-10y+66=0, 40x-18y=214$.The value of variance of x is 9.

Find i) The mean values of x &y

ii) The correlation coefficient between x & y **(May 2015)**

iii)S.D of y

Solution: Given $8x-10y+66=0, 40x-18y=214$

Since both the regression lines pass through the point $(\bar{X}, \bar{Y})$

$$8\bar{x}-10\bar{y}+66=0$$

$$40\bar{x}-18\bar{y}=214$$

By solving these equations we get, $\bar{X}=13, \bar{Y}=17$

Since $y=\frac{8}{10}x+\frac{66}{10}=0.8x+6.6$

$$x=\frac{214}{40}+\frac{18}{40}y=0.45y+5.35$$

Therefore $b_{yx}=0.8, b_{xy}=0.45$

$$r=b_{yx}b_{xy}=0.6$$

Also , $\sigma^2{}_x=9 \Rightarrow \sigma_x=3$

$$b_{yx}=\frac{r\sigma_y}{\sigma_x} \Rightarrow \frac{b_{yx}\sigma_x}{r}=\sigma_y=4$$

,

Practice Questions

Q1. Line of regression y on x is 8x-10y+66=0. Line of regression x on y is 40x-18y-214=0. The value of variance of x is 9. Find the standard deviation of y .

Q2. Find correlation coefficient (r)for the following data:

X	12	15	18	20
y	12	16	20	25

Q3. Two lines of regression are given by 5y-8x+17=0 and 2y-5x+14=0 . If , find

a) The mean value of x and y b) ${\sigma^2}_x$ c)Coefficient of correlation between x and y.

Q4. Find the first four moments about mean for the following distribution . Also find β_1 and β_2.

Marks	0-10	10-20	20-30	30-40
Number of workers	6	26	47	15

Q5. First four moments of a distribution about value 5 are-4,22,-117 and 560.Obtain the first four central moments and coefficient of skewness and kurtosis.

Q6. First four moments about the working mean 3.5 of a distribution are 0.0375,0.4546,0.0609 & 0.5074.Calculate the first four moment about the mean. Also calculate the coefficient of skewness.

Curve Fitting

Fitting Straight Line:

Let $(x_i, y_i); i = 1,2,3,......n$ be the observed values of (x , y).

To fit the straight line,

y=ax+b

using least square criteria $(x_1, y_1), (x_2, y_2), (x_3, y_3)...............(x_n, y_n)$ are the observed values.

Consider the equations $a\sum x + nb = \sum y..............(1)$

$$a\sum x^2 + b\sum x = \sum xy.......(2)$$

Solving (1) and (2) , we determine the values of a and b which gives the straight line y=ax+b , best fit for the given data.

Curve Fitting

Fitting Second Degree Parabola :

Consider the parabola of the form, $y = ax^2 + bx + c$

to the set of observed values $(x_i, y_i); i = 1,2,3,.......n.$

Now consider the equations

$$a\sum x^2 + b\sum x + nc = \sum y.........(1)$$

$$a\sum x^3 + b\sum x^2 + c\sum x = \sum xy.........(2)$$

$$a\sum x^4 + b\sum x^3 + c\sum x^2 = \sum x^2 y.........(3)$$

Equations (1),(2) and (3) are three simultaneous equations in three unknowns a, b, c. Solving these equations we determine a, b, c which gives best fit parabola for the given data .

Multiple Choice Questions

1) Least square fit for the straight line $y = ax+b$ to the data

x	1	2	3
y	5	7	9

is

(a) $y=2x+4$ (b) $y=2x-3$ (c) $y=2x+3$ (d) $y=3x-4$

2) Least square fit for the straight line $x = ay+b$ to the data

y	0	1	2
x	2	5	8

is

(a) $x=3y-1$ (b) $x=3y+1$ (c) $x=3y+2$ (d) $x=3y-4$

Multiple Choice Questions

3) For least square fit of the straight line y = ax+b to the data

x	0	1	2
y	-1	1	3

the normal equations are

(a) 3a+3b=3
5a+3b=7

(b) 3a+3b=3
3a+5b=7

(c) 3a+3b=3
5a+7b=3

(d) 3a+3b=7
5a+3b=3.

4) Least square fit for the curve $y = ax^b$ to the data

x	2	4	6
y	2	16	54

is

(a) $y = \frac{1}{4}x^3$

(b) $y = \frac{1}{4}x^2$

(c) $y = 2x^3$

(d) $y = \frac{1}{2}x^3$

Multiple Choice Questions

5) For the least square fit for the parabola $y = ax^2 + bx + c$ to the data

X	0	1	2
Y	2	2	4

the normal equations are

(a) 5a+3b+3c=8
9a+5b+3c=10
17a+9b+5c=18

(b) 5a+3b+3c=18
9a+5b+3c=8
17a+9b+5c=10

(c) 17a+3b+3c=8
9a+17b+3c=10
17a+9b+17c=18

(d) 5a+3b+3c=0
9a+5b+3c=0
17a+9b+5c=0

6) For the least square fit for the parabola $x = ay^2 + by + c$ to the data

Y	1	2	3
x	3	7	13

the normal equations are

(a) 3a+6b+3c=23
36a+3b+6c=56
98a+36b+3c=148

(b) 14a+6b+3c=0
36a+14b+6c=0
98a+36b+14c=0

(c) 14a+6b+3c=23
36a+14b+6c=56
98a+36b+14c=148

(d) 14a+6b+3c=148
36a+14b+6c=23
98a+36b+14c=56

Example

Q. Fit a straight line of the form y = mx+c to the following data , by using the method of least squares.

x	0	1	2	3	4	5	6	7
Y	-5	-3	-1	1	3	5	7	9

Solution : Preparing the table as

x	y	xy	X^2
0	-5	0	0
1	-3	-3	1
2	-1	-2	4
3	1	3	9
4	3	12	16
5	5	25	25
6	7	42	36
7	9	63	49
$\sum x = 28$	$\sum y = 16$	$\sum xy = 140$	$\sum x^2 = 140$

Example

Here n=8 (Total number of points)

Consider the equations $$a\sum x + nb = \sum y \ldots\ldots\ldots\ldots\ldots(1)$$
$$a\sum x^2 + b\sum x = \sum xy \ldots\ldots\ldots(2)$$

Substituting values from table in equations (1) and (2) we get ,

28m+8c=16 | 140m+28c=140

OR 7m+2c=4 ……(3) | OR 5m+c=5 ………..(4)

Solving (3) and (4) we get ,

m=2, c=-5.

Hence the equation of the straight line is ,

y = 2x-5

Example

Q1. Fit a straight line of the form X=aY+b to the following data by the least square method

(Dec. 2019)

x	2	5	8	11	17	20
y	2	3	4	5	7	8

Q2. Fit a straight line of the form Y=aX+b to the following data by the least square method

(May 2019)

x	1	3	4	5	6	8
y	-3	1	3	5	7	11

University Questions

Q1. First four moments of a distribution about value 4 are -1.5, 17, -30 and 108.Obtain the first four central moments, mean, S.D and coefficient of skewness and kurtosis. (May – 2015)

Solution: Given $\mu_1' = -1.5, \mu_2' = 17, \mu_3' = -30, \mu_4' = 108$

We know the relations between μ_r & μ_r'

$$\mu_1 = 0, \quad \mu_2 = \mu_2' - \left(\mu_1'\right)^2 = 17 - 4 = 12.5$$

$$\mu_3 = \mu_3' - 3\mu_1'\mu_2' + 2\left(\mu_1'\right)^3 = -30 - 3(17)(-1.5) + 2(-1.5)^3 = 113.25$$

$$\mu_4 = \mu_4' - 4\mu_1'\mu_3' + 6\left(\mu_1'\right)^2\mu_2' - 3\left(\mu_1'\right)^4$$

$$= 108 - 4(-30)(-1.5) + 6 * 17(-1.5)^2 - 3(-1.5)^4 = 142.3125$$

$$S.D. = \sqrt{\mu_2} = 3.5355$$

Also, $\beta_1 = \dfrac{\mu_3^{\,2}}{\mu_2^{\,3}} = 6.5667$

$$\beta_2 = \frac{\mu_4}{\mu_2^{\,2}} = 11.385$$

University Questions

Q2. First four moments of a distribution about value 44.5 are -0.4, 2.99, -0.08 and 27.63. Obtain the first four central moments, mean, S.D and coefficient of skewness and kurtosis. **(Nov 2014)**

Solution: Given

$$\mu_1' = -0.4, \mu_2' = 2.99, \mu_3' = -0.08, \mu_4' = 27.63, A = 44.5$$

$$\overline{X} = \mu_1' + A = -0.4 + 44.5 = 44.1$$

Now $\mu_1 = 0$

$$\mu_2 = \mu_2' - \left(\mu_1'\right)^2 = 2.99 - (0.4)^2 = 2.83$$

$$\mu_3 = \mu_3' - 3\mu_1'\mu_2' + 2\left(\mu_1'\right)^3 = -0.08 - +(1.2)2.99 - 2.4 = 3.38$$

$$\mu_4 = \mu_4' - 4\mu_1'\mu_3' + 6\left(\mu_1'\right)^2\mu_2' - 3\left(\mu_1'\right)^4$$

$$= 27.63 - 1.92 + 1.9136 = 29.2748$$

Also, $$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(3.38)^2}{(2.83)^3} = 0.50405$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{29.2748}{(2.28)^2} = 5.63150$$

University Questions

Q3. First four moments of a distribution about value 5 are 2, 20, 40 and 50.Obtain the first four central moments, mean, S.D and coefficient of skewness and kurtosis. **(Nov – 2015)**

Solution: Given

$$\mu_1' = 2, \mu_2' = 20, \mu_3' = 40, \mu_4' = 50$$

Now, $\mu_1 = 0$

$$\mu_2 = \mu_2' - \left(\mu_1'\right)^2 = 20 - 4 = 16$$

$$\mu_3 = \mu_3' - 3\mu_1'\mu_2' + 2\left(\mu_1'\right)^3 = 40 - 120 + 16 = -64$$

$$\mu_4 = \mu_4' - 4\mu_1'\mu_3' + 6\left(\mu_1'\right)^2\mu_2' - 3\left(\mu_1'\right)^4$$

$$= 50 - 960 + 320 - 48 = -638$$

$$S.D = \sqrt{\mu_2} = 4$$

Also , $\beta_1 = \dfrac{\mu_3^2}{\mu_2^3} = 1$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = -204922$$

University Questions

Q4. Calculate the coefficient of correlation between the marks obtained by 8 students in Mathematics and Statistics from the following table.

Students	A	B	C	D	E	F	G	H
Mathematics (x)	25	30	32	35	37	40	42	45
Statistics(y)	8	10	15	17	20	22	24	25

Solution :

X	25	30	32	35	37	40	42	45	286
Y	8	10	15	17	20	22	24	25	**141**
X^2	625	900	1024	1225	1369	1600	1764	2025	**10532**
Y^2	64	100	225	289	400	484	576	625	**2763**
XY	200	300	480	595	740	880	1008	1125	**5328**

University Questions

$$\bar{x} = \frac{\sum x}{8} = \frac{286}{8} = 35.75$$

$$\bar{y} = \frac{\sum y}{8} = \frac{141}{8} = 17.625$$

$$Cov(x,y) = \frac{\sum xy}{n} - \bar{x}\,\bar{y} = \frac{5328}{8} - (35.75 \times 17.625) = 35.9063$$

$$Var(x) = \frac{\sum x^2}{n} - \bar{x}^2 = \frac{10532}{8} - (35.75)^2 = 38.4375$$

$$Var(y) = \frac{\sum y^2}{n} - \bar{y}^2 = \frac{2763}{8} - (17.625)^2 = 34.4375$$

$$r(x,y) = \frac{Cov(x,y)}{\sigma_x \sigma_y} = 0.9826$$