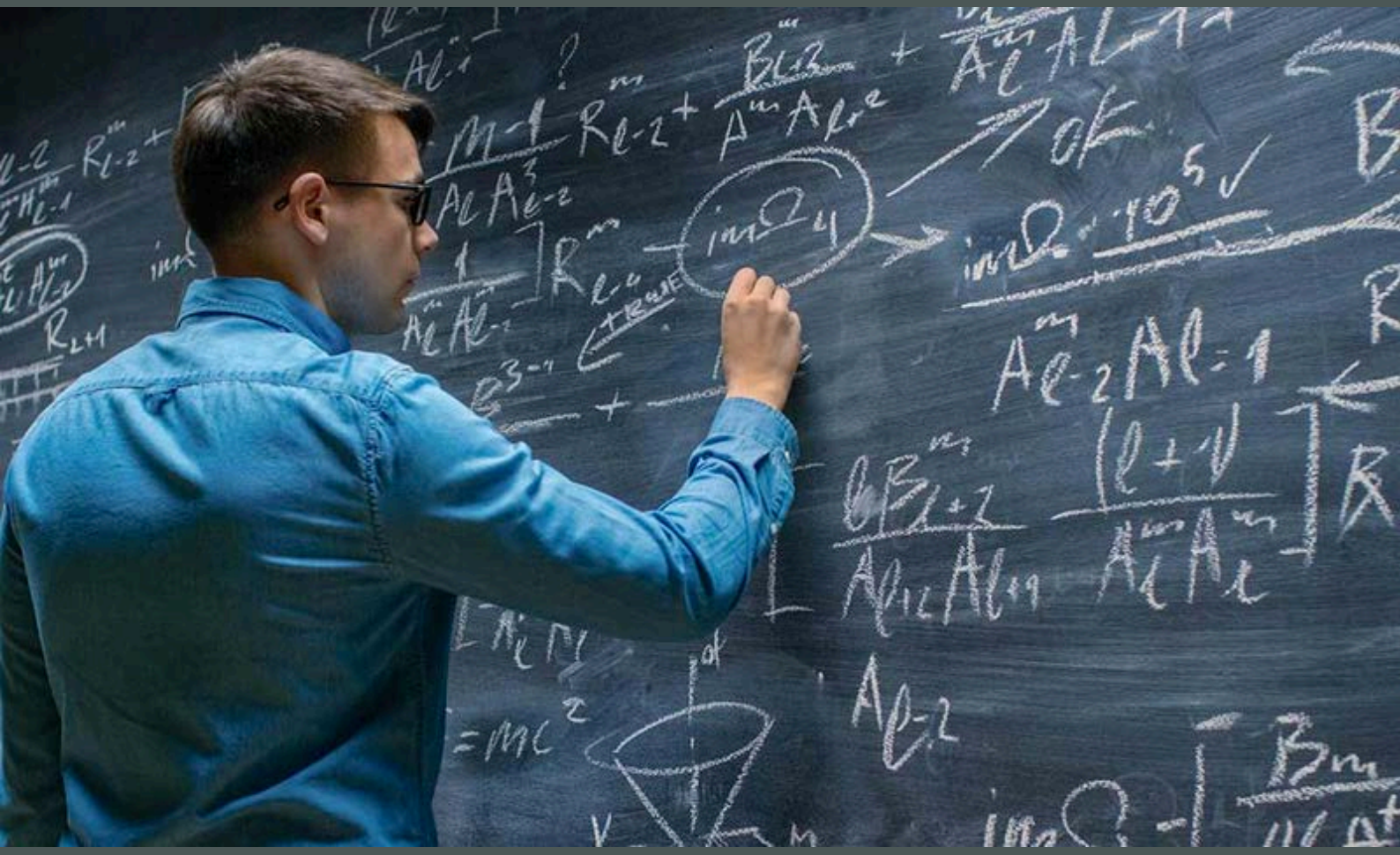
FREE BTECH NOTES

Advanced Engineering Mathematics

BY PROENGR TEAM



- Preparatory Books for Professional Engineering exams
- Unlimited Free Mock Tests
- Free Notes for BTECH and PE Exams
- Blog

94105 24390 PROENGR.IN INFO@PROENGR.IN

STUDY DESK TO LICENSE – YOUR COMPLETE EXAM COMPANION

STUDY MATERIAL FOR PROFESSIONAL ENGINEERING EXAMS

- Electrical Supervisor License Exams
- Boiler Operation Engineer (BOE) Exams
- Mine Manager Certificate/PSU Exams
- Energy Manager Exam
- PE Exams

BEST SELLER

AVAILABLE ON:
AMAZON
FLIPKART
POTHI

FREE STUFF
UNLIMITED MOCK TESTS
FREE NOTES FOR BTECH AND PE EXAMS



94105 24390
PROENGR.IN
INFO@PROENGR.IN

Books | Mock Tests
Free Notes | Blog

STATISTICAL TECHNIQUES - I

ARITHMETIC MEAN

For ungrouped data

If the variable x takes the n values $x_1, x_2 \ldots x_n$, the arithmetic mean

$$x = \frac{x_1 + x_2 \ldots\ldots\ldots x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

For Grouped Data

If the frequencies $f_1, f_2 \ldots\ldots\ldots f_n$ are given corresponding to the values $x_1, x_2 \ldots x_n$, then A.M. is given by

$$x = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} = \frac{1}{N}\sum_{i=1}^{n} f_i x_i \left[\text{where } N = \sum_{i=1}^{n} f_i\right]$$

MEDIAN

$$\text{Median} = I_1 + \frac{\frac{N}{2} - c.f}{f} xi$$

where i_1= lower limit of median class

c.f = cumulative frequency of the class

f = frequency of median class, and

i = size of class interval of the median class.

When total frequencies is odd and equal to say N, then value of (N + 1)/2 is taken.

MODE

Mathematical Formula. for a continuous series of a unimodal frequency distribution.

$$\text{Mode} = l_1 + \frac{f - f_{m-1}}{2f - f_{m-1} - f_{m+1}} xi$$

where l_1= lower limit of the modal class

f = frequency of the modal class

i = size of the modal class

f_{m-1}= frequency of the class preceding the modal class

f_{m+1} = frequency of the class following the modal class.

If $2f = f_{m-1} + f_{m+1}$, then above formula fails and mode is calculated with the following approximate formula

$$\text{Mode} = l_1 + \frac{f_{m-1}}{f_{m-1} + f_{m+1}} xi$$

RELATIONSHIP BETWEEN MEAN, MEDIAN AND MODE

- In case of a symmetrical distribution the mean, median and mode coincide with each other.
- In case of a moderately asymmetrical distribution, these three measures are related as Mean – Mode = 3(Mean - Median) or Mode = 3 Median - 2 Mean.

Example

The following table gives the population of males of different age groups of U.K. and India at the time of the census of 1931. Compare the average age of males in the two countries and account for the difference, if any

Age Group (Years)	***U.K. (In Lakhs)***	***India (In Lakhs)***
0-5	*18*	*214*
5-10	*19*	*258*
10-15	*20*	*222*
15-20	*18*	*157*
20-25	*16*	*145*
25-30	*14*	*161*
30-40	*27*	*257*
40-50	*25*	*184*
50-60	*19*	*120*
Above 60	*17*	*100*

Solution

Age Grope (Years)	Mid value (x or m)	$d = x_i - 275$	U K f_1	f_1d	India f_2	f_2d
0-5	2.5	-25.0	18	-450	214	-5350.0
5-10	7.5	-20.0	19	-380	258	-5160.0
10-15	12.5	-15.0	20	-300	222	-3330.0
15-20	17.5	-10.0	18	-180	157	-1570.0
20-25	22.5	-5.0	16	-18	145	-725.0
25-30	27.5	0	14	0	161	0
30-40	35.0	7.5	27	202.5	257	1927.5
40-50	45.0	17.5	25	437.5	180	3220.0
50-60	55.0	27.5	19	422.5	124	3300.0
Above 60	65.0	37.5	17	637.5	100	4750.0
			193	**410**	**1818**	**-3937.5**

Average age of males in U.K.

$$= a + \frac{\Sigma fd}{\Sigma f} = 27.5 + \frac{410}{193} = 27.5 + 2.12 = 29.62 \text{ years}$$

Average age of males in India

$$= 275 + \frac{(-3937.5)}{1818} = 27.5 - 2.17 = \text{years}$$

$\therefore$ Average age of males in U.K. is higher than that of males in India .

Example

Find the median, lower quartile and upper quartile of the following data

Wages (In Rs.)	***No. of Workers***
0-10	*22*
10-20	*38*
20-30	*46*
30-40	*35*
40-50	*20*

Solution

Class	f	c.f
0-10	22	22
10-20	38	60
20-30	46	106
30-40	35	141
40-50	20	161

Here N=161.

(i) Median = size of $\left[\frac{161}{2}\right]^{th}$ item =size of 80.5^{th} item.

$\therefore$ $\text{Median} = 1 + \frac{\frac{N}{2} - cf}{f} \text{ x } i = 20\frac{80.5 - 60}{46} \text{ x } 10 = 24.46$

(ii) For lower quartile, Q_1

$$\frac{N}{4} = \frac{161}{4} = 40.25$$

$\therefore$ Q_1 lies in the class 10-20

$$Q_1 = 1_1 + \frac{\frac{N}{4} - cf}{f} \text{ x } i = 10 + \frac{40.25 - 22}{38} \text{ x } 10 = ₹\ 14.8$$

(iii) For upper quartile, Q_3

$$\frac{3N}{4} = \frac{3x161}{4} = \frac{488}{4} = 120.75$$

$\therefore$ Q_3 lies in the class 30-40

$$Q_3 = 1_1 + \frac{\frac{3N}{4} - cf}{f} \text{ x } i = 30 + \frac{120.25 - 106}{35} \text{ x } 10 = ₹\ 34.21$$

Example

Calculate mode from the following data:

Marks	***No. of Students***
Above 0	*80*
,, 10	*77*
,, 20	*72*
,, 30	*65*
,, 40	*55*
,, 50	*43*
,, 60	*28*
,, 70	*16*
,, 80	*10*
,, 90	*8*
,, 100	*0*

Solution

Marks	No. of Students
0-10	3
10-20	5
20-30	7
30-40	10
40-50	13
50-60	15
60-70	12
70-80	6
80-90	2
90-100	8

Max. frequency is 15.

$\therefore$ Model class is 50-60.

$\therefore$ $\text{Mode} = l_1 + \dfrac{f - f_{m-1}}{2f - f_{m-1} - f_{m+1}} \times i$

$$= 50 + \frac{15-12}{30-12-12} \times 10 = 50 + \frac{3 \times 10}{6} = 55$$

STANDARD DEVIATION

It is defined as the square root of the arithmetic mean of the squared deviation, all deviations measured from the actual mean $\bar{x}$. the standard deviation is conveniently represented by Greek letter σ.

$$\sigma = \sqrt{\frac{\sum f\left(x - \bar{x}\right)^2}{\sum f}}$$

also $$\sigma = \sqrt{\frac{1}{N}\sum f_i d^i 2 - \left\{\frac{1}{N}\sum f_i d_i\right\}^2}$$

The square of standard deviation i.e. σ^2 is defined as variance and is given by

$$\sigma^2 = \frac{\sum f\left(x - \bar{x}\right)^2}{\sum f}$$

Example

The following table gives a set of observations x from frequency f.

x	*8.53*	*8.54*	*8.55*	*8.56*	*8.57*	*8.58*	*8.59*	*8.60*
f	*1*	*4*	*8*	*10*	*9*	*4*	*3*	*1*

Determine The Mean And The Standard Deviation.

Solution

The calculation are shown in the following table. The arbitrary mean is assumed at 8.56.

x	**f**	**d = x-5.56**	**fd**	**fd^2**
8.53	1	-0.03	-0.03	0.0009
8.54	4	-0.02	-0.08	0.0016
8.55	8	0.01	-0.08	0.0008
8.56	10	0.00	0.00	0.00
8.57	9	0.01	0.09	0.0009
8.58	4	0.02	0.08	0.0016
8.59	3	0.03	0.09	0.0027
8.60	1	0.04	0.04	0.0016
Total	**40**		**0.11**	**0.0101**

Mean $\bar{x} = 8.56 + 0.00275 = 8.56275$

$$\sigma = \sqrt{\frac{0.0101}{40} - \left(\frac{0.11}{40}\right)^2} = 0.0156$$

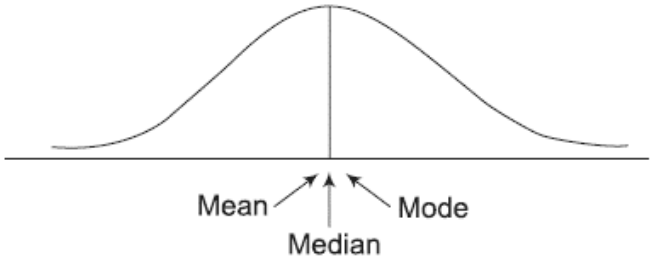
Coefficient of Variation. coefficient of variation is the ratio of S.D to the mean expressed in the percentage. Symbolically,

$$C.V. = \frac{\sigma}{\bar{x}} \times 100$$

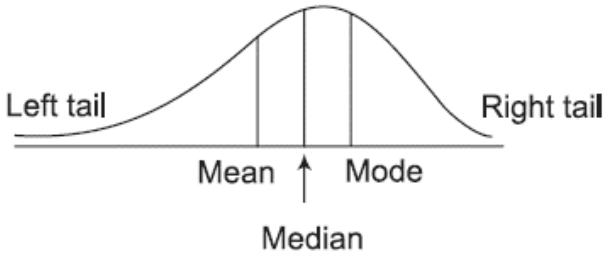
SKEWNESS

Skewness is a measure that refers to the extent of symmetry or asymmetry in a distribution. A distribution is said to be symmetrical when its mean, median, and mode are equal, and the frequencies are symmetrically distributed about the mean.

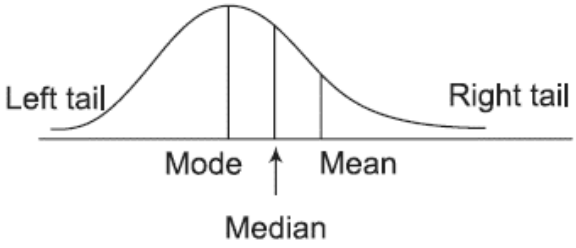
A symmetrical distribution when plotted on a graph will give a perfectly bell-shaped curve which is known as a normal curve as shown in given figure.



A distribution is said to be asymmetrical or skewed when the mean, median, and mode are not equal, i.e., the mean median, and mode do not coincide. If the curve has a longer tail towards the left, it is said to be a negatively skewed distribution (Fig. a). If the curve has a longer tail towards the right, it is said to be positively skewed (Fig. b).



Negatively skewed distribution



Positively skewed distribution

Skewness gives an idea of the nature and degree of concentration of observations about the mean.

Measures of skewness

A measure of skewness gives the extent and direction of skewness of a distribution. These measures can be absolute or relative. The absolute measures are also known as measures of skewness.

Absolute skewness = Mean – Mode

If the value of the mean is **greater** than the mode, the skewness will be positive and if the value of the mean is **less** than the mode, the skewness will be negative.

The relative measures of skewness is called the **coefficient of skewness**.

Karl Pearson's Coefficient of Skewness

Karl Pearson's coefficient of skewness denoted by S_k, is given by

$$S_k = \frac{\text{mean - mode}}{\text{standard deviation}} = \frac{\text{mean - mode}}{\sigma}$$

When the mode is **ill-defined** (*Example: The maximum frequency occurs at the end of the frequency distribution)* and the distribution is moderately skewed, the averages have the following relationship:

$$S_k = \frac{\text{mean - mode}}{\text{standard deviation}} = \frac{\text{mean - (3 median - 2 mean)}}{\sigma} = \frac{3(\text{mean-median})}{\sigma}$$

Coefficient of skewness based on third moment

$$\gamma_1 = \sqrt{\beta_1}$$

where $\beta_1 = \mu_3^2 \mid \mu_2^2$

This gives the simplest measure of skewness.

Since β, is always positive number, so the sign of skewness is given by the sign of μ_3.

Example

Calculate Karl Pearson's coefficient of skewness from the following data:

Wages (₹)	*10-15*	*15-20*	*20-25*	*25-30*	*30-35*	*35-40*	*40-45*	*45-50*
No. of workers	*8*	*16*	*30*	*45*	*62*	*32*	*15*	*6*

Solution

Let a = 32.5 be the assumed mean and h = 5 be the width of the class interval.

$$d = \frac{x-a}{h} = \frac{x-32.5}{5}$$

Wages (₹)	No. of workers (f)	Mid value (x)	$d = \frac{x-32.5}{5}$	d^2	fd	fd^2
10-15	8	12.5	-4	16	-32	128
15-20	16	17.5	-3	9	-48	144
20-25	30	22.5	-2	4	-60	120
25-30	45	27.5	-1	1	-45	45
30-35	62	32.5	0	0	0	0

35-40	32	37.5	1	1	32	32
40-45	15	42.5	2	4	30	60
45-50	6	47.5	3	9	18	54
$\Sigma f = 214, \Sigma fd = -105; \Sigma fd^2 = 583$						

$$N = \sum f = 214$$

$$\bar{x} = a + h\frac{\sum fd}{N} = 32.55\left(\frac{-105}{214}\right) = 30.05$$

$$\sigma = h\sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2} = 5\sqrt{\frac{583}{214} - \left(\frac{105}{214}\right)^2} = 7.88$$

Since the maximum frequency is 32, the mode lies in the interval 30–35.

Here, l = 30, h = 5, $f_m = 62$, $f_1 = 45$, $f_2 = 32$

$$Mode = l + h\left(\frac{f_m - f_1}{2f_m - f_1 - f_2}\right) = 30 + 5\left[\frac{62 - 45}{2(62) - 45 - 32}\right] = 31.81$$

Now $$S_k = \frac{\text{mean - mode}}{\sigma} = \frac{30.05 - 31.81}{7.88} = -0.223$$

Example

Calculate Karl Pearson's coefficient of skewness from the following data:

Weekly wages	*40-50*	*50-60*	*60-70*	*70-80*	*80-90*	*90-100*	*100-110*	*110-120*	*120-130*	*130-140*
No. of workers	*5*	*6*	*8*	*10*	*25*	*30*	*36*	*50*	*60*	*70*

Solution

The maximum frequency 70 occurs at the end of the frequency distribution. Hence, the mode is ill-defined and Karl Pearson's coefficient of skewness is obtained using the median. Let a = 85 be the assumed mean and h = 10 be the width of the class interval.

$$d = \frac{x - a}{h} = \frac{x - 85}{10}$$

Weekly wages	No. of workers (f)	Mid value (x)	$d = \frac{x-85}{10}$	d^2	fd	fd^2	CF
40-50	5	45	-4	16	-20	80	5
50-60	6	55	-3	9	-18	54	11
60-70	8	65	-2	4	-16	32	19
70-80	10	75	-1	1	-10	10	29
80-90	25	85	0	0	0	0	54
90-100	30	95	1	1	30	30	84
100-110	36	105	2	4	72	144	120
110-120	50	115	3	9	150	450	170
120-130	60	125	4	16	240	960	230
130-140	70	135	5	25	350	1750	300
$\Sigma f = 300; \Sigma fd = 778; \Sigma fd^2 = 3510$							

$$N = \sum f = 300$$

$$\bar{x} = a + h\left(\frac{\sum fd}{N}\right)$$

$$= 85 + 10\left(\frac{778}{300}\right)$$

$$= 110.93$$

$$\sigma = h\sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2}$$

$$= 10\sqrt{\frac{3510}{300} - \left(\frac{778}{300}\right)^2}$$

$$= 22.304$$

N/2 = 300/2 = 150

The cumulative frequency just greater than 150 is 170 and the corresponding class 110 – 120 is the median class.

Here, l = 110, h = 10, f = 50, c = 120

$$Medium = l + \frac{h}{f}\left(\frac{N}{2} - c\right) = 110 + \frac{10}{50}(150 - 120) = 116$$

Now skewness

$$S_k = \frac{3(mean - median)}{22.304} = \frac{3(110.93 - 116)}{22.304} = -0.682$$

Problem

Calculate the Karl Pearson coefficient for the following distribution.

Group	*1-5*	*6-10*	*11-15*	*16-20*	*21-25*	*26-30*	*31-35*
Frequency	*3*	*4*	*68*	*30*	*10*	*6*	*2*

Answer: 0.31

***Hint**: Here class interval is not of regular type. We have to change it.*

Group	*0.5-5.5*	*5.5-10.5*	*10.5-15.5*	*15.5-20.5*	*20.5-25.5*	*25.5-30.5*	*30.5-35.5*
Frequency	*3*	*4*	*68*	*30*	*10*	*6*	*2*

MOMENTS

Hence, **moment about the mean** are

$$\mu_1' = \frac{1}{N}\sum fd$$

$$\mu_2' = \frac{1}{N}\sum fd^2$$

$$\mu_3' = \frac{1}{N}\sum fd^3$$

$$\mu_4' = \frac{1}{N}\sum fd^4$$

where $d = \frac{m-a}{h} or \frac{x-a}{h}$

where m or x are mid values, "h" is interval and "a" is assumed value as per given data.

Central moments are

$$\mu_1 = 0$$

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

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

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

KURTOSIS

By measures of Kurtosis, we check the peakness of a distribution and its measure is given by

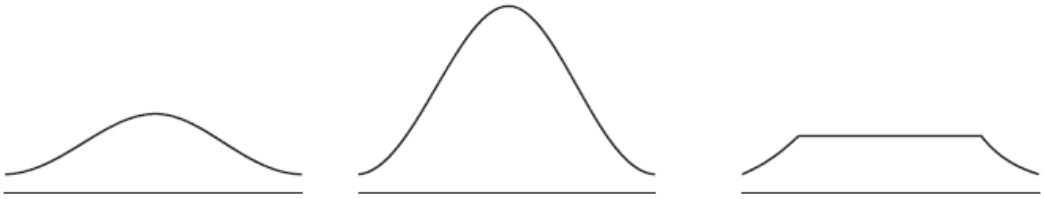
$\gamma_2 = \beta_2 - 3$ defines the excess of kurtosis

where $\beta_2 = \mu_4 \mid \mu_2^2$

If $\beta_2 = 3$, then curve is called normal curve or mesokurtic curve.

$\beta_2 > 3$, then curve is called leptokurtic

$\beta_2 < 3$, then curve platykurtic.



(a) Mesokurtic (b) Leptokurtic (c) Platykurtic

Example (UPTU 2006, VTU 2003, UTU 2013, 5 marks)

The first moments about the working mean 28.5 of a distribution are 0.294, 7.144, 42.409 and 454.98. Calculate the moment about mean, Also, evaluate β_1, β_2, and comment upon the skewness and kurtosis of the distribution.

Solution

First moments about 28.5 are given as

$\mu_1' = 0.294$

$\mu_2' = 7.144$

$\mu_3' = 42.409$

and $\mu_4' = 454.98$

Now $$\mu_1' = \frac{1}{N}\sum f(x - 28.5) = \frac{1}{N}\sum fx - 28.5 = \bar{x} - 28.5$$

$\therefore$ $$0.294 = \bar{x} - 28.5 \Rightarrow \bar{x} = 28.794$$

Now $\mu_1 = 0$

$\mu_2 = \mu_2' - \mu_1'^2 = 7.144 - (0.294)^2 = 7.058$

$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3 = 42.409 - 3(7.144)(0.294) + 2(0.294)^3 = 36.151$

$$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'(\mu_1')^2 - 3(\mu_1')^4 = 408.738$$

Therefore $$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{36.151^2}{7.058^3} = 3.717$$ (positive as sign of μ_3 is positive)

This shows a considerable skewness of the distribution.

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{408.738}{7.058^2} = 8.205 > 3$$

Hence distribution is leptokurtic.

Problem

Compute skewness and kurtosis, if the first four moments of a frequency distribution f(x) about the value x = 4 are respectively 1, 4, 10 and 45.

Answer: Skewness = 0; Kurtosis = 2.9

Example (UTU 2015, 5 marks)

From the following data, calculate the first four moments about the mean:

x	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
f	*1*	*8*	*28*	*56*	*70*	*56*	*28*	*8*	*1*

Solution

Let a = 4, d = x - a = x - 4

x	f	d = x - 4	fd	fd^2	fd^3	fd^4
0	1	-4	-4	16	-64	256
1	8	-3	-24	72	-216	648
2	28	-3	-56	112	-224	448
3	56	-1	-56	56	-56	56
4	70	0	0	0	0	0
5	56	1	56	56	56	56
6	28	2	56	112	224	448
7	8	3	24	72	216	648
8	1	4	4	16	64	256
	N = 256		0	512	0	2816

Now various moments are

$$\mu_1' = \frac{1}{N}\sum fd = 0$$

$$\mu_2' = \frac{1}{N}\sum fd^2 = \frac{512}{256} = 2$$

$$\mu_3' = \frac{1}{N}\sum fd^3 = 0$$

$$\mu_4' = \frac{1}{N}\sum fd^4 = \frac{2816}{256} = 11$$

Problem

Calculate the first three moments about 90 from the following frequency distribution.

x	*82*	*88*	*90*	*91*	*92*	*95*	*97*
f	*7*	*11*	*15*	*8*	*4*	*3*	*2*

Answer: -6.6, 13.98, -51.42

Hint: d = x - 90

Example

Find the first four central moments for the following distribution:

Size (x)	*0-5*	*5-10*	*10-15*	*15-20*	*20-25*
frequency (f)	*1*	*2*	*6*	*7*	*4*

Solution

Let a = 12.5

$$d = \frac{m-a}{h} = \frac{m-12.5}{5}$$

where m is mid value.

size	mid value (m)	f	$d = \frac{m-12.5}{5}$	fd	fd^2	fd^3	fd^4
0-5	2.5	1	-2	-2	4	-8	16
5-10	7.5	2	-1	-2	2	-2	2
10-15	12.5	6	0	0	0	0	0
15-20	17.5	7	1	7	7	7	7
20-25	22.5	4	2	8	16	32	64
		N = 20		Σfd = 11	$\Sigma fd^2 = 29$	$\Sigma fd^3 = 29$	$\Sigma fd^4 = 89$

Now $$\mu_1' = \frac{\sum fd}{N} h = \frac{11}{20} x5 = 2.75$$

$$\mu_2' = \frac{\sum fd^2}{N} h^2 = \frac{29}{20} x5^2 = 36.25$$

$$\mu_3' = \frac{\sum fd^3}{N} h^3 = \frac{29}{20} x5^3 = 181.25$$

$$\mu_4' = \frac{\sum fd^4}{N} h^4 = \frac{89}{20} x5^4 = 2781.25$$

Central moments

$\mu_1 = 0$

$\mu_2 = \mu_2' - \mu_1'^2 = 36.25 - (2.75)^2 = -28.6875$

$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3 = 181.25 - 3(2.75)(36.25) + 2(2.75)^3 = -76.2187$

$$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'(\mu_1')^2 - 3(\mu_1')^4 = 2260.7696$$

Problem (UTU 2016, 10 marks)

Calculate first four central moments and kurtosis for the following frequency distribution:

Class	*0-10*	*10-20*	*20-30*	*30-40*
Frequency	*1*	*3*	*4*	*2*

Answer: 0, -81, -144, 14817; $\beta_2 = \mu_4/\mu_2^2 = 2.26$

Example (UTU 2017, 5 marks)

For a distribution, the mean is 10, variance is 16, γ_1 is 1 and β_2 is 4. Find the first four moments about origin.

Solution

Given $\bar{x}$ = 10, variance = 16 $\Rightarrow \mu_2 = 16$

Also $$\gamma_1 = 1 = \sqrt{\beta_1} \Rightarrow \beta = 1 \Rightarrow \frac{\mu_3^2}{\mu_2^3} = 1$$

$\therefore$ $$\mu_3^2 = \mu_2^3 = 16^3 = (4^3)^2$$

$\therefore$ $\mu_3 = 64$

Also $$\beta_2 = 4 \Rightarrow \frac{\mu_4}{\mu_2^2} = 4 \Rightarrow \mu_4 = 4\mu_2^2 = 4x16^2 = 1024$$

Problem

Given the following information: Mean = 10, variance = 16 and $\sqrt{\beta_1} = 1$. Obtain the first three moments about the origin.

Answer: 10, 16, 1544

Problem

If $\beta_1 = +1$, $\beta_4 = 4$ and variance = 9, find the value of μ_3 and μ_4. Also comment upon nature of distribution.

Answer: $\mu_3 = 27$; $\mu_4 = 324$; It is positively skewed and leptokurtic.

CORRELATION

Correlation is the relationship that exists between two or more variables. Two variables are said to be correlated if a change in one variable affects a change in the other variable.

Some examples of such a relationship are as follows:

- Relationship between heights and weights.
- Relationship between price and demand of commodity.
- Relationship between rainfall and yield of crops.
- Relationship between age of husband and age of wife.

Karl Pearson's coefficient of correlation

The coefficient of correlation is the measure of correlation between two random variables X and Y, and is denoted by r.

$$r = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2 \sum (y-\bar{y})^2}}$$

Simplifying $$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\sum x^2 - \frac{(\sum x)^2}{n}} \sqrt{\sum y^2 - \frac{(\sum y)^2}{n}}}$$

Example

Calculate the correlation coefficient between x and y using the following data:

x	*2*	*4*	*5*	*6*	*8*	*11*
y	*18*	*12*	*10*	*8*	*7*	*5*

Solution

n = 6

x	y	x^2	y^2	xy
2	18	4	324	36
4	12	16	144	48
5	10	25	100	50
6	8	36	64	48
8	7	64	49	56
11	5	121	25	55
$\Sigma x = 36$	$\Sigma y = 60$	$\Sigma x^2 = 266$	$\Sigma y^2 = 706$	$\Sigma xy = 293$

Now $$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\sum x^2 - \frac{(\sum x)^2}{n}}\sqrt{\sum y^2 - \frac{(\sum y)^2}{n}}} = \frac{293 - \frac{(36)(60)}{6}}{\sqrt{266 - \frac{36^2}{6}}\sqrt{706 - \frac{60^2}{6}}} = -0.9203$$

Problem

Calculate the coefficient of correlation between x and y for the following data:

x	*1*	*3*	*4*	*5*	*7*	*8*	*10*
y	*2*	*6*	*8*	*10*	*14*	*16*	*20*

Answer: 1

RANK CORRELATION

Let a group of n individuals be arranged in order of merit with respect to some characteristics. The same group would give a different order (rank) for different characteristics. Considering the orders corresponding to two characteristics A and B, the correlation between these n pairs of ranks is called the rank correlation in the characteristics A and B for that group of individuals.

Spearman's Rank Correlation Coefficient

$$r = 1 - \frac{6\sum d^2}{n(n^2 - 1)}$$

Example

Ten participants in a contest are ranked by two judges as follows:

x	*1*	*3*	*7*	*5*	*4*	*6*	*2*	*10*	*9*	*8*
y	*3*	*1*	*4*	*5*	*6*	*9*	*7*	*8*	*10*	*2*

Calculate the rank correlation coefficient.

Solution

n = 10

Rank by first judge (x)	Rank by second judge (y)	d = x - y	d^2
1	3	-2	4
3	1	2	4
7	4	3	9
5	5	0	0
4	6	-2	4
6	9	-3	9
2	7	-5	25
10	8	2	4
9	10	-1	1
8	2	6	36
		$\Sigma d = 0$	$\Sigma d^2 = 96$

Now $$r = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6(96)}{10(10^2-1)} = 0.418$$

Example

Ten students got the following percentage of marks in mathematics and physics:

Maths (x)	*8*	*36*	*98*	*25*	*75*	*82*	*92*	*62*	*65*	*35*
Physics (y)	*84*	*51*	*91*	*60*	*68*	*62*	*86*	*58*	*35*	*49*

Find the rank correlation coefficient.

STUDY DESK TO LICENSE – YOUR COMPLETE EXAM COMPANION

STUDY MATERIAL FOR PROFESSIONAL ENGINEERING EXAMS

- Electrical Supervisor License Exams
- Boiler Operation Engineer (BOE) Exams
- Mine Manager Certificate/PSU Exams
- Energy Manager Exam
- PE Exams

BEST SELLER

AVAILABLE ON:
AMAZON
FLIPKART
POTHI

FREE STUFF
UNLIMITED MOCK TESTS
FREE NOTES FOR BTECH AND PE EXAMS



94105 24390
PROENGR.IN
INFO@PROENGR.IN

Books | Mock Tests
Free Notes | Blog

Solution

Here ranks are not given. We have to compute it from given data.

n = 10

x	y	Rank in maths (x_i)	Rank in Physics (y_i)	$d = x_i - y_i$	d^2
8	84	10	3	7	49
36	51	7	8	-1	1
98	91	1	1	0	0
25	60	9	6	3	9
75	68	4	4	0	0
82	62	3	5	-2	4
92	86	2	2	0	0
62	58	6	7	-1	1
65	35	5	10	-5	25
35	49	8	9	-1	1
				$\Sigma d = 0$	$\Sigma d^2 = 90$

Now $$r = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6(90)}{10(10^2-1)} = 0.455$$

Example (UTU 2017, 10 marks)

Obtain the rank correlation coefficient for the following data:

x	*68*	*64*	*75*	*50*	*64*	*80*	*75*	*40*	*55*	*64*
y	*62*	*58*	*68*	*45*	*81*	*60*	*68*	*48*	*50*	*70*

Solution

We will first convert this data to ranks obtained.

x	y	Rank x_i	Rank y_i	$d = x_i - y_i$	d^2
68	62	4	5	-1	1
64	58	6	7	-1	1
75	68	2.5	3.5	-1	1
50	45	9	10	-1	1

64	81	6	1	5	25
80	60	1	6	-5	25
75	68	2.5	3.5	-1	1
40	48	10	9	1	1
55	50	8	8	0	0
64	70	6	2	4	16
				$\Sigma d = 0$	$\Sigma d^2 = 72$

In X - series. 75 occurs two times, the ranks should be alloted are 2 and 3.

Therefore the average of 2 and 3 i.e. 2.5 is given to both.

Next value 68. so rank = 4

64 occurs 3 times. The ranks should be 5, 6 and 7. So rank = (5 + 6 + 7)/3 = 6

To Σd^2, we add $\frac{m(m^2-1)}{12}$ for each value repeated. So, for 75, m = 2 and for 64, m = 3.

So, for x series, correlation is $\frac{2(4-1)}{12}+\frac{3(9-1)}{12}=\frac{5}{2}$

In y series, 68 occurs twice, so rank = (3 + 4)/2 = 3.5

68 occurs twice, so m = 2

So, for y series, correlation = $\frac{2(4-1)}{12}=\frac{1}{2}$

$$\therefore \quad r=1-\frac{6\left(\sum d^2+\frac{5}{2}+\frac{1}{2}\right)}{n(n^2-1)}=1-\frac{6(72+3)}{10x99}=0.545$$

Problem

Calculate the coefficient of correlation of ranks obtained by 10 students of a class in Hindi and English:

Hindi	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
English	*3*	*8*	*1*	*7*	*10*	*2*	*9*	*4*	*6*	*5*

Answer: 0.103

REGRESSION

Regression is defined as a method of estimating the value of one variable when that of the other is known and the variables are correlated.

Line of regression of y on x

It is the line which gives the best estimate for the values of y for any given values of x.

The regression equation of **y on x** is given by

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

It is also written as

$$y = a + bx$$

Line of regression of x on y

It is the line which gives the best estimate for the values of x for any given values of y.

The regression equation for **x on y** is given by

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

It is also written as

$$x = a + by$$

Regression coefficients

The slope b of the line of regression of y on x is also called the **coefficient of regression of y on x**. It represents the increment in the value of y corresponding to a unit change in the value of x.

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

where b_{yx} = Regression coefficient of y on x.

Similarly $$b_{xy} = r\frac{\sigma_x}{\sigma_y}$$

where b_{xy} = Regression coefficient of x on y.

Expressions for Regression Coefficients

b_{yx}	b_{xy}
$b_{yx} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2}$	$b_{xy} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(y-\bar{y})^2}$
$b_{yx} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}}$	$b_{xy} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum y^2 - \frac{(\sum y)^2}{n}}$

Properties of regression coefficient

- The coefficient of correlation is the geometric mean of the coefficients of regression, i.e. $r = \sqrt{b_{yx} b_{xy}}$.
- If one of the regression coefficients is greater than one, the other must be less than one.
- The arithmetic mean of regression coefficients is greater than or equal to the coefficient of correlation.
- Regression Coefficients are independent of the change of origin but not of scale.
- Both regression coefficients will have the same sign i.e., either both are positive or both are negative.
- The sign of correlation is same as that of the regression coefficients, i.e., $r > 0$ if $b_{xy} > 0$ and $b_{yx} > 0$; and $r < 0$ if $b_{xy} < 0$ and $b_{yx} < 0$.

Example

Calculate linear regression coefficients from the following:

x	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
y	*3*	*7*	*10*	*12*	*14*	*17*	*20*	*24*

Solution

x	y	x^2	y^2	xy
1	3	1	9	3
2	7	4	49	14
3	10	9	100	30
4	12	16	144	48

5	14	25	196	70
6	17	36	289	102
7	20	49	400	140
8	24	64	576	192
$\Sigma x = 36$	$\Sigma y = 107$	$\Sigma x^2 = 204$	$\Sigma y^2 = 1763$	$\Sigma xy = 599$

Putting tabulated values and n = 8

We get $$b_{yx} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{\left(\sum x\right)^2}{n}} = 2.7976$$

and $$b_{xy} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum y^2 - \frac{\left(\sum y\right)^2}{n}} = 0.3540$$

Example

Obtain the line of regression of y on x for the data given below

x	*1.53*	*1.78*	*2.60*	*2.95*	*3.42*
y	*33.50*	*36.30*	*40.00*	*45.80*	*53.50*

Solution

n = 5

x	y	x^2	xy
1.53	33.50	2.3409	51.255
1.78	36.30	2.1684	64.614
2.60	40.00	6.76	104
2.95	45.80	8.7025	135.11
3.42	53.50	11.6964	182.97
$\Sigma x = 12.28$	$\Sigma y = 209.1$	$\Sigma x^2 = 32.6682$	$\Sigma xy = 537.949$

Putting tabulated values and n = 5

We get $$b_{yx} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{\left(\sum x\right)^2}{n}} = 9.726$$

The line of regression of y on x is given by

$$y - \bar{y} = b_{yx}(x - \bar{x}) \qquad (1)$$

Mean $$\bar{x} = \frac{\sum x}{n} = \frac{12.28}{5} = 2.456$$

and $$\bar{y} = \frac{\sum y}{n} = \frac{209.1}{5} = 41.82$$

From (1) $$y - 41.82 = 9.726(x - 2.456) = 9.726x - 23.887$$

$\therefore$ $$y = 17.932 + 9.726x$$

which is the required line of regression of y on x.

Problem

From the following data, obtain the two regression lines and correlation coefficient.

x	*6*	*2*	*10*	*4*	*8*
y	*9*	*11*	*5*	*8*	*7*

Answer: b_{yx} = -0.65; b_{xy} = -1.3; r = -0.9192 (negative sign as both b_{yx} and b_{xy} are negatives)

y = -0.65x + 11.9 and x = -1.3y + 16.4

Example (AKTU 2018, 2020, 2021, 7 marks)

In a partially destroyed laboratory record of an analysis of a correlation data, the following results only are eligible:

Variance of x = 9

Regression equations: 8x – 10y + 66 = 0, 40x – 18y = 214.

What were (a) the mean values of x and y (b) the standard deviation of y and the co-efficient of correlation between x and y.

Solution

(a) As both lines of regression pass through $(\bar{x}, \bar{y})$, therefore

$$8\bar{x} - 10\bar{y} + 66 = 0 \quad (1)$$

and $$40\bar{x} - 18\bar{y} - 214 = 0 \quad (2)$$

Solving, we get

$$\bar{x} = 13; \bar{y} = 17$$

(b) variance of x = $\sigma_x^2 = 9$

$\therefore$ $\sigma_x = 3$

The equations of lines of regression can be written as

$$y = 0.8x + 6.6$$

and $$x = 0.45y + 5.35$$

$\therefore$ $$r\frac{\sigma_y}{\sigma_x} = 0.8 \quad (3)$$

and $$r\frac{\sigma_x}{\sigma_y} = 0.45 \quad (4)$$

From (3) x (4)

$$r^2 = 0.8 \text{ x } 0.45 = 0.36$$

$\therefore$ r = 0.6 [positive sign is taken as regression coefficients are +ve]

From (3) $$\sigma_y = \frac{0.8\sigma_x}{r} = \frac{0.8x3}{0.6} = 4$$

Example (UTU 2013, AKTU 2019, 7 marks)

The equations of two lines of regression obtained in a correlation analysis of 60 observations are

5x = 6y + 24 and 1000, y = 768x - 3608.

What is the coefficient of correlation? Show that the ratio of coefficient of variability of x to that of y is 5/24. What is the ratio of variances of x and y?

Solution

Regression line of x on y is

$$5x = 6y + 24$$

i.e. $$x = (6/5)y + (24/5)$$

$\therefore$ $b_{xy} = 6/5 \quad (1)$

Regression line of y on x is

$1000y = 768x - 3608$

i.e. $y = 0.768x - 3.608$

$\therefore$ $b_{yx} = 0.768$ (2)

From (1) and (2)

$$r\frac{\sigma_x}{\sigma_y} = b_{xy} = \frac{6}{5} \quad (3)$$

and $$r\frac{\sigma_y}{\sigma_x} = b_{yx} = 0.768 \quad (4)$$

From (3) x (4)

$r^2 = 0.9216$

$\therefore$ $r = \pm 0.96 = +0.96$ [positive sign is taken as b_{xy} and b_{yx} both are +ve]

From (4)/(3)

$$\frac{\sigma_x^2}{\sigma_y^2} = \frac{6}{5x0.768} = 1.5625$$

$\therefore$ $$\frac{\sigma_x}{\sigma_y} = 1.25 = \frac{5}{4} \quad (5)$$

Since the regression lines pass through the point $(\bar{x}, \bar{y})$, we have

$$5\bar{x} = 6\bar{y} + 24$$

and $$1000\bar{y} = 768x - 3608$$

Solving the above equations for $\bar{x}$ and $\bar{y}$, we get

$$\bar{x} = 6; \bar{y} = 1$$

Coefficient of variability of x = $\frac{\sigma_x}{\bar{x}}$

Coefficient of variability of y = $\frac{\sigma_y}{\bar{y}}$

Required ratio = $\frac{\sigma_x}{\bar{x}} x \frac{\bar{y}}{\sigma_y} = \frac{\bar{y}}{\bar{x}}\left(\frac{\sigma_x}{\sigma_y}\right) = \frac{1}{6} x \frac{5}{4} = \frac{5}{24}$

Problem

The regression lines of y on x and x on y are respectively y = ax + b and x = cy + d. Show that

$$\frac{\sigma_y}{\sigma_x} = \sqrt{\frac{a}{c}}$$

$$\bar{x} = \frac{bc + d}{1 - ac}$$

and $$\bar{y} = \frac{ad + b}{1 - ac}$$

Example

From the following results, obtain the two regression equations and estimate the yield when the rainfall is 29 cm and the rainfall, when the yield is 600 kg:

	Yield (kg)	*Rainfall (cm)*
Mean	*508.4*	*26.7*
SD	*36.8*	*4.6*

The coefficient of correlation between yield and rainfall is 0.52.

Soution

Let rainfall in cm is denoted by x and yield in kg is denoted by y.

$$\bar{x} = 26.7; \bar{y} = 508.4; \sigma_x = 4.6; \sigma_y = 36.8; r = 0.52$$

Now $$b_{yx} = r\frac{\sigma_y}{\sigma_x} = 0.52\left(\frac{36.8}{4.6}\right) = 4.16$$

and $$b_{xy} = r\frac{\sigma_x}{\sigma_y} = 0.52\left(\frac{4.6}{36.8}\right) = 0.065$$

The equation of the line of regression of y on x is

$$y - \bar{y} = b_{yx}(x - \bar{x})$$

$\therefore$ y - 508.4 = 4.16(x - 26.7)

giving y = 4.16x + 397.328

The equation of the line of regression of x on y is

$$x - \bar{x} = b_{xy}(y - \bar{y})$$

$\therefore$ x - 26.7 = 0.065(y - 508.4)

giving x = 0.065y - 6.346

Estimated yield when the rainfall is 29 cm

STUDY DESK TO LICENSE – YOUR COMPLETE EXAM COMPANION

STUDY MATERIAL FOR PROFESSIONAL ENGINEERING EXAMS

- Electrical Supervisor License Exams
- Boiler Operation Engineer (BOE) Exams
- Mine Manager Certificate/PSU Exams
- Energy Manager Exam
- PE Exams

BEST SELLER

AVAILABLE ON:
AMAZON
FLIPKART
POTHI

FREE STUFF
UNLIMITED MOCK TESTS
FREE NOTES FOR BTECH AND PE EXAMS



94105 24390
PROENGR.IN
INFO@PROENGR.IN

Books | Mock Tests
Free Notes | Blog

$$y = 4.16(29) + 397.328 = 517.968 \text{ kg}$$

Estimated rainfall when the yield is 600 kg

$$x = 0.065(600) - 6.346 = 32.654 \text{ cm}$$

Problem

The following results were obtained from records of age (x) and systolic blood pressure (y) of a group of 10 men:

	x	*y*
Mean	*53*	*142*
variance	*130*	*165*

and $\Sigma(x - \bar{x})(y - \bar{y}) = 1220$

Find the approximate regression equation and use it to estimate the blood pressure of a man whose age is 45.

Answer: regression line of y on x; y = 0.935x + 92.445

Putting given value, required blood pressure = 134.52

Hint: Variance = σ^2

CURVE FITTING

Method of least squares (Fit a straight line)

(1) Let the equation of straight line be y = a + bx.
(2) From the given data we calculate Σx, Σy, Σx^2, and Σxy.
(3) Putting above values in normal equations

$$\Sigma y = na + b\Sigma x \quad (1)$$

$$\Sigma xy = a\Sigma x + b\Sigma x^2 \quad (2)$$

(4) Solving equations (1) and (2) for a and b.
(5) Putting the values of a and b in y = a + bx we get equation of straight line of best fit.

Example (UTU 2016, 5 marks)

Fit the straight line that best fits the following data:

x	*1*	*2*	*3*	*4*	*5*
y	*14*	*27*	*40*	*55*	*68*

Solution

Lets first form the table.

x	y	xy	x^2
1	14	14	1
2	27	54	4
3	40	120	9
4	55	220	16
5	68	340	25
$\Sigma x = 15$	$\Sigma y = 204$	$\Sigma xy = 748$	$\Sigma x^2 = 55$

Normal equations are

$$\Sigma y = na + b\Sigma x \quad (1)$$

$$\Sigma xy = a\Sigma x + b\Sigma x^2 \quad (2)$$

Putting tabulated values in (1) and (2)

$$204 = 5a + 15\,b$$

$$748 = 15a + 55b$$

Solving $a = 13.6;\ b = 0$

The equation of line of best fit will be

$$y = 13.6x$$

Problem

Fit a straight line to the following data:

x	*1*	*2*	*3*	*4*	*6*	*8*
y	*2.4*	*3*	*3.6*	*4*	*5*	*6*

Answer: $y = 1.9764 + 0.5059x$

Method of least squares (Fit a parabola)

(1) Let the equation of the parabola be $y = a + bx + cx^2$.

(2) From the given data we calculate Σx, Σy, Σx^2, Σx^3, Σx^4, Σxy and $\Sigma x^2 y$.

(3) Putting above values in normal equations

$$\Sigma y = na + b\Sigma x + c\Sigma x^2 \quad (1)$$

$$\Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma x^3 \quad (2)$$

$$\Sigma x^2 y = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4 \quad (3)$$

(1) Solving equations (i), (ii) and (iii) for a,b and c.

(2) Putting the values of a and b in y = a + bx + cx2 we get equation of parabola of best fit.

Example (UTU 2015, 5 marks)

Fit a second degree parabola in the following data by least square method:

X	*0*	*1*	*2*	*3*	*4*
Y	*1*	*4*	*10*	*17*	*30*

Solution

x	y	x^2	x^3	x^4	xy	x^2y
0	1	0	0	0	0	0
1	4	1	1	1	4	4
2	10	4	8	16	20	40
3	17	9	27	81	51	153
4	30	16	64	256	120	480
$\Sigma x = 10$	$\Sigma y = 62$	$\Sigma x^2 = 30$	$\Sigma x^3 = 100$	$\Sigma x^4 = 354$	$\Sigma xy = 195$	$\Sigma x^2y = 677$

Here n = 5

Substituting values in eqns.

$$\Sigma y = na + b\Sigma x + c\Sigma x^2$$
$$\Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma x^3$$
$$\Sigma x^2y = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4$$

We get

$$62 = 5a + 10b + 30c \quad (1)$$
$$195 = 10a + 30b + 100c \quad (2)$$
$$677 = 30a + 100b + 354c \quad (3)$$

Solving eqns. (1), (2) and (3), we get

$$a = 1.2, b = 1.1 \text{ and } c = 1.5$$

Hence the required second degree parabola is

$$y = 1.2 + 1.1x + 1.5x^2$$

Example

By the method of least squares, fit a parabola to the following data:

x	*1*	*2*	*3*	*4*	*5*
y	*5*	*12*	*26*	*60*	*97*

Answer: $y = 10.4 - 11.0857x + 5.7143x^2$

Fitting of logarithmic and exponential equations

Let the equation is $y = ab^x$.

Taking logrithm on both sides of the equation

$$\log_e y = \log_e a + x\log_e b$$

Putting $\log_e y = Y$, $\log_e a = A$, $x = X$ and $\log_e b = B$

We get $Y = A + BX$

This is linear equation in X and Y. The normal equations are

$$\Sigma Y = nA + B\Sigma X$$

and $$\Sigma XY = A\Sigma X + B\Sigma X^2$$

Solving these equations, A and B, and, hence, a and b can be found. The best fitting exponential curve is obtained by substituting the values of a and b in the equation $y = ab^x$.

Similarly, the best fitting exponential curves for the relation $y = ax^b$ and $y = aeb^x$ can be obtained.

Example

Fit a curve of the form $y = ae^{bx}$ *to the following data:*

x	*1*	*3*	*5*	*7*	*9*
y	*115*	*105*	*95*	*85*	*80*

Solution

Taking logarithm on both the sides

$$\log_e y = \log_e a + bx\log_e e = \log_e a + bx$$

Putting $\log_e y = Y$; $\log_e a = A$, $b = B$, $x = X$

Then $Y = A + BX$

This is linear equation in X and Y. The normal equations are

$$\Sigma Y = nA + B\Sigma X \quad (1)$$

and $$\Sigma XY = A\Sigma X + B\Sigma X^2 \quad (2)$$

Here n = 5

x	y	X	Y	X^2	XY
1	115	1	4.7449	1	4.7449
3	105	3	4.6539	9	13.9617
5	95	5	4.5539	25	22.7695
7	85	7	4.4427	49	31.0989
9	80	9	4.3820	81	39.438
		$\Sigma X = 25$	$\Sigma Y = 22.7774$	$\Sigma X^2 = 165$	$\Sigma XY = 112.013$

Substituting these values in Eqs (1) and (2),

$$22.7774 = 5\,A + 25\,B \quad (1)$$

$$112.013 = 25A + 165B \quad (2)$$

Solving Eqs (3) and (4),

$$A = 4.7897;\ B = -0.0469$$

$$\log_e a = A$$

$\therefore$ $\log_e a = 4\ 7897$

$$a = 120.2653$$

and $b = B = -0.0469$

Hence, the required equation of the curve is

$$y = 120.2653e^{-0.0469x}$$

Fitting of the curve $y = ax + bx^2$

Normal equations are

$$\Sigma xy = a\Sigma x^2 + b\Sigma x^3$$

$$\Sigma x^2 y = a\Sigma x^3 + b\Sigma x^4$$

Fitting of the curve $y = ax + (b/x)$

Normal equations are

$$\Sigma xy = nb + a\Sigma x^2$$

$$\Sigma(y/x) = na + b\Sigma(1/x^2)$$

Fitting of the curve $y = (a/x) + b\sqrt{x}$

$$\Sigma(y/x) = a\Sigma(1/x^2) + b\Sigma(1/\sqrt{x})$$

$$\Sigma y\sqrt{x} = a\Sigma(1/x) + b\Sigma x$$

Example (UPTU 2008, UTU 2013, 10 marks)

Obtain the least square fit of the form $f(t) = ae^{-3t} + be^{-2t}$ for the data.

t	*0.1*	*0.2*	*0.3*	*0.4*
f(t)	*0.76*	*0.58*	*0.44*	*0.35*

Solution

Normal equations of given curve are

$$\Sigma f(t).e^{-3t} = a\Sigma e^{-6t} + b\Sigma e^{-5t} \quad (1)$$

and $$\Sigma f(t).e^{-2t} = a\Sigma e^{-5t} + b\Sigma e^{-4t} \quad (2)$$

Let us make a table.

t	f(t)	e^{-4t}	e^{-5t}	e^{-6t}	$f(t)e^{-2t}$	$f(t)e^{-3t}$
0.1	0.76	0.6703	0.6065	0.5488	0.6222	0.5630
0.2	0.58	0.4493	0.3679	0.3012	0.3888	0.3183
0.3	0.44	0.3012	0.2231	0.1653	0.2415	0.1789
0.4	0.35	0.2019	0.1353	0.0907	0.1573	0.1054
		Σe^{-4t} = 1.6227	Σe^{-5t} = 1.3328	Σe^{-6t} = 1.106	$\Sigma f(t)e^{-2t}$ = 1.4098	$\Sigma f(t)e^{-3t}$ = 1.1656

From (1) and (2)

$$1.106a + 1.3328b = 1.1656 \quad (3)$$

and $$1.3328a + 1.6227b = 1.4098 \quad (4)$$

Solving $a = 0.6778$; $b = 0.3121$

Hence, the equation is

$$f(t) = 0.6778e^{-3t} + 0.3121e^{-2t}$$

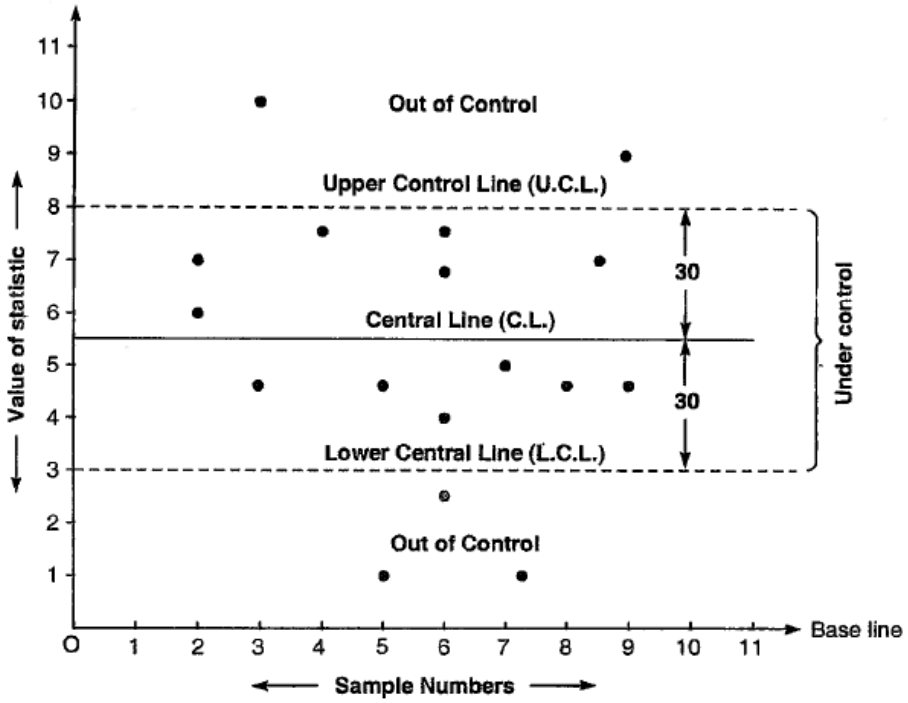
CONTROL CHARTS

A control chart is a graphical representation of the collected information. It detects the variation in processing and warns if there is any departure from the specified tolerance limits.

A control chart has **three horizontal lines** starting from the right hand side of the vertical line and parallel to the base line of the chart. The vertical line represents the quality statistic of each

sample. The base line shows the sample number scale (or sub-group scale). The three horizontal lines are known as control lines. These are:

- **Central line (C.L).** It passes through the middle of the chart and is parallel to the base line. It represents the prescribed standard quality of the product in the process.
- **Upper Control Limit (U.C.L).** It is shown in the chart by a dotted line that passes through the chart above and parallel to the central line and represents the upper limit of tolerance.
- **Lower Control Limit (L.C.L).** It is shown in the chart by a dotted line that passes through the chart below and parallel to the central line. It represents the lower limit of tolerance. This line indicates the maximum lowest limit of variation in the process that can reasonably attributed to chance causes. If a sample point falls beyond and outside this limit, it indicates that a variation in the process has occurred due to assignable causes. Thus, the process is deemed to be out of control.



Thus, a control chart contains two control limits:

- Upper Control Limit (U.C.L.)
- Lower Control Limit (L.C.L.)

Sample mean $\bar{x} = \dfrac{x_1 + x_2 + ... + x_n}{n} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$

Sample range $R = x_{max} - x_{min}$

$$\bar{\bar{x}} = \frac{\bar{x}_1 + \bar{x}_2 + ... + \bar{x}_k}{k} = \frac{1}{k}\sum_{i=1}^{k} x_i$$

$$\bar{R} = \frac{R_1 + R_2 + ... + R_k}{k} = \frac{1}{k}\sum_{i=1}^{k} R_i$$

For $\bar{X}$ chart

Central line = $\bar{\bar{x}}$ when tolerance limits are not given

= μ when tolerance limits are given

where $\mu = \frac{1}{2}[LCL + UCL]$

Now, LCL for $\bar{X}$ chart = $\bar{\bar{x}} - A_2\bar{R}$

UCL for $\bar{X}$ chart = $\bar{\bar{x}} + A_2\bar{R}$

A_2 depends on sample size n.

LCL for R chart = $D_3\bar{R}$

UCL for R chart = $D_4\bar{R}$

D_3 and D_4 depend on sample size.

Example (UTU 2017, 10 marks)

The following are the mean lengths and ranges of lengths of a finished product from 10 samples each of size 5. The specification limits for length are 200 ± 5 cm. Construct $\bar{X}$ and R-chart and examine whether the process is under control and state your recommendations.

Sample No.	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Mean $\bar{X}$	*201*	*198*	*202*	*200*	*203*	*204*	*199*	*196*	*199*	*201*
Range R	*5*	*0*	*7*	*3*	*3*	*7*	*2*	*8*	*5*	*6*

Assume for n = 5, A_2 = 0.58, D_4 = 2.11 and D_3 = 0

Solution

Control limits for $\bar{X}$ chart

CL = 200 [Tolerance limit are given, ∴ μ = 200]

$$UCL_{\bar{X}} = \bar{\bar{x}} + A_2\bar{R} = \mu + A_2\bar{R}$$
$$LCL_{\bar{X}} = \bar{\bar{x}} - A_2\bar{R} = \mu - A_2\bar{R}$$

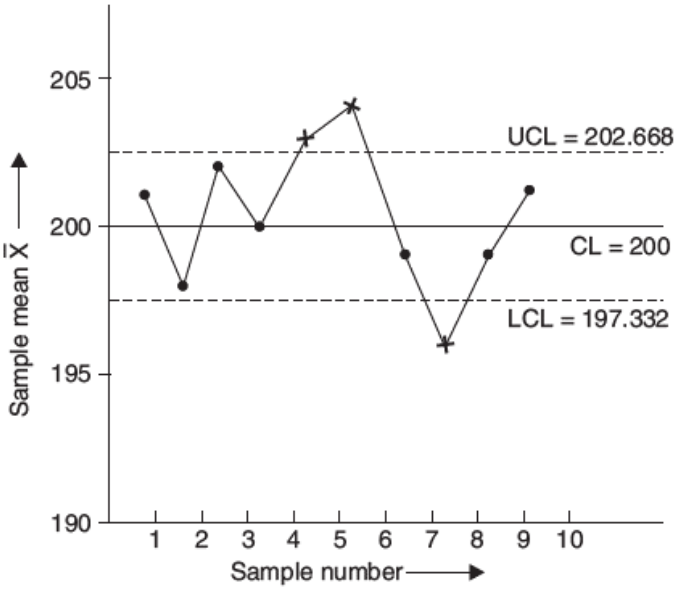
where $\bar{R} = \frac{R_1 + R_2 + ... + R_{10}}{10} = \frac{46}{10} = 4.6$

Then
$$UCL_{\bar{X}} = 200 + (0.58x4.6) = 202.668$$
$$LCL_{\bar{X}} = 200 - (0.58x4.6) = 197.332$$

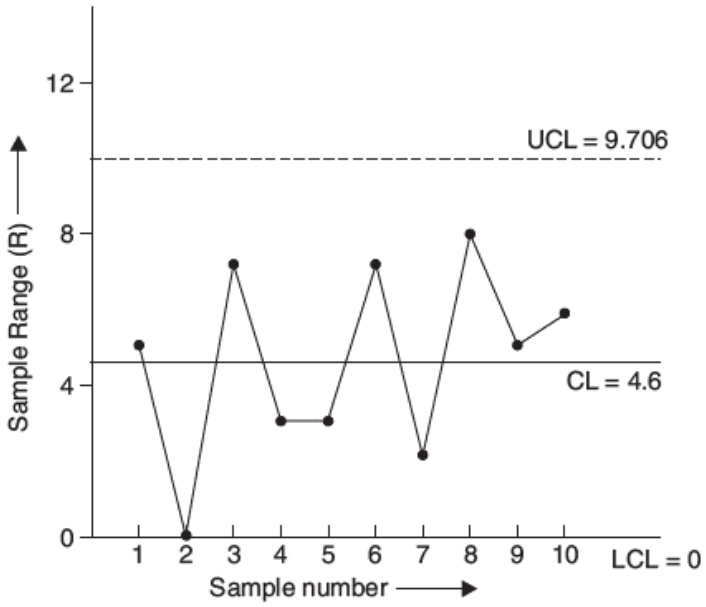
Control limits for R chart

$$CL = \bar{R} = 4.6$$
$$UCL_R = D_4\bar{R} = 2.11x4.6 = 9.706$$
$$LCL_R = D_3\bar{R} = 0x4.6 = 0$$

The $\bar{X}$ and R-charts



X-Chart



R-Chart

It is noted that all points lie within the control limits on the R chart. Hence the process variability is under control. But in X-chart, *points corresponding to sample number 5, 6, and 8 lie outside the control limits*. Therefore the process is **not in statistical control**. The process should be halted and it is recommended to check for any assignable causes. Fluctuation will remain until these causes, if found, are removed.

Problem

The following data show the values of sample mean ($\bar{x}$) and range (R) for 10 samples for size 6 each. Calculate the values for central line and the control limits for Mean-chart and Range-chart. Draw the control charts and comment on the state of control:

Sample No.	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Mean $\bar{x}$	*43*	*49*	*37*	*44*	*45*	*37*	*51*	*46*	*43*	*47*
Range R	*5*	*6*	*5*	*7*	*7*	*4*	*8*	*6*	*4*	*6*

For n = 6, $A_2 = 0.483$, $D_3 = 0$, $D_4 = 2.004$

Answer: For $\bar{X}$ *chart, CL =* $\bar{x} = 44.2$, $UCL_{\bar{X}} = 47.0014; LCL_{\bar{X}} = 41.3986$

For R chart, $UCL_R = 11.6232; LCL_R = 0$

ASSIGNMENT

MEASURES OF CENTRAL TENDENCY

Q.1. (AKTU 2019, 2 marks): Formula of measure of Kurtosis β_2 = ____

Q.2. (AKTU 2019, 2 marks): The first three central moments of a distribution are 0, 15, -31. Find the moment coefficient of skewness.

Answer: -0.53

Q.3. (UTU 2013, 5 marks): From the following data, calculate the first four moments about the mean:

Class	0-10	10-20	20-30	30-40	40-50	50-60	60-70
Frequency	4	8	11	15	11	7	4

Answer: 0, 256.56, 120.70, 151658.35

Q.4. (UTU 2015, 5 marks): From the following data, calculate the first four moments about the mean:

x	0	1	2	3	4	5	6	7	8
f	1	8	28	56	70	56	28	8	1

Answer: Solved in this module.

Q.5. (UTU 2016, 5 marks): Find the Pearson's coefficients β_1, β_2, γ_1, γ_2 for Poisson distribution.

Q.6. (UTU 2017, 10 marks): For a distribution, the mean is 10, variance is 16, γ_1 is 1 and β_2 is 4. Find the first four moments about origin.

Answer: Solved in this module.

Q.7. (UTU 2013, 5 marks): The first moments about the working mean 28.5 of a distribution are 0.294, 7.144, 42.409 and 454.98. Calculate the moment about mean, Also, evaluate β_1, β_2, and comment upon the skewness and kurtosis of the distribution.

Answer: Solved in this module.

Q.8. (UTU 2016, 10 marks): Find out the kurtosis of the data given below:

Class interval	0-10	10-20	20-30	30-40
Frequency	1	3	4	2

Answer: See module.

CURVE FITTING

Q.9. (AKTU 2018, 2020, 2 marks): Write the normal equation for the curve dy/dx = (a/x) + bx

Q.10. (AKTU 2019, 2 marks): Find the normal equation of the curve $ax + bx^2$.

Answer: Normal equations are

$$\Sigma xy = a\Sigma x^2 + b\Sigma x^3$$

$$\Sigma x^2 y = a\Sigma x^3 + b\Sigma x^4$$

Q.11. (AKTU 2021, 2 marks): Write the normal equation for the curve $y = \frac{a}{x} + b\sqrt{x}$

Answer:

$\Sigma(y/x) = a\Sigma(1/x^2) + b\Sigma(1/\sqrt{x})$

$\Sigma y\sqrt{x} = a\Sigma(1/x) + b\Sigma x$

Q.12. (JNTUH 2019, 2 marks): Construct normal equations for fit a straight line by method of least squares.

Q.13. (BPUT 2020, 6 marks): Fit a straight line y = a + bx to the following data by the method of least square:

x	0	1	2	3
y	1	2	1	10

Q.14. (JNTUH 2018, 10 marks): Estimate y at x = 5 by fitting a least squares curve of the form $y = \frac{b}{x(x-a)}$ to the following data

x	3.6	4.8	6.0	7.2	8.4	9.6	10.8
y	0.83	0.31	0.17	0.10	0.07	0.05	0.04

Q.15. (UTU 2013, 10 marks): Obtain the least square fit of the form $f(t) = ae^{-3t} + be^{-2t}$ for the data.

t	0.1	0.2	0.3	0.4
f(t)	0.76	0.58	0.44	0.35

Answer: Solved in this module.

Q.16. (UTU 2015, 5 marks): Fit the curve $y = ax^b$ to the following data:

X	1	2	3	4	5	6
Y	1200	900	600	200	110	50

Answer: $y = 2035x^{-1.7494}$

Q.17. (UPTU 2010, UTU 2017, 10 marks): Find the normal equations and hence fit the curve of the form ax + (b/x) for the following data:

x	1	2	3	4	5	6	7	8
y	5.4	6.2	8.2	10.3	12.6	14.8	17.2	19.5

Answer: $y = 2.39188x + \frac{2.98195}{x}$

Q.18. (UTU 2015, 5 marks): Fit a second degree parabola in the following data by least square method:

X	0	1	2	3	4
Y	1	4	10	17	30

Answer: Solved in this module.

Q.19. (UTU 2016, 5 marks): Fit the straight line that best fits the following data:

x	1	2	3	4	5
y	14	27	40	55	68

Answer: Solved in this module.

Q.20. (UTU 2017, 10 marks): Draw a trend line by the method of semi averages to the following data:

Years	1991	1992	1993	1994	1995	1996	1997	1998	1999
Sales	32	38	50	48	54	42	48	72	82

Q.21. (UTU 2018, 10 marks): Find the least square fit of the form $y = ax^2 + bx + c$ to the following data:

x	10	12	15	23	20
y	14	17	23	25	21

Answer: $y = 0.07x^2 + 3.03x - 8.89$

REGRESSION ANALYSIS

Q.22. (AKTU 2018, 2020, 2021, 7 marks): In a partially distributed laboratory record of an analysis of a correlation data, the following results are legible:

Variance of x = 9

Regression equation:

$8x - 10y + 66 = 0$

$40x - 18y = 214$

What were (i) the mean of x and y (ii) standard deviation of y and the coefficient of x and y.

Q.23. (UTU 2013, AKTU 2019, 7 marks): The equation of two regression lines, obtained in a correlation analysis of 60 observations are:

$5x = 6y + 24$

$1000y = 768x - 3608$

What is the correlation coefficient? Show that the ratio of coefficient of variability of x to that of y is 5/24. What is the ratio of variances of x and y?

Answer: Solved in this module.

Q.24. (BPUT 2020, 16 marks): Find the correlation coefficient and the equation of the lines of regression for the following values of x and y.

x	1	2	3	4	5
y	2	5	3	8	7

Answer: r = 0.8

y on x; $x = 1.3x + 1.1$ and x on y; $x = 0.5y + 0.5$

Q.25. (UTU 2015, 5 marks): The two regression equations of the variables x and y are

$x = 19.13 - 0.87y$

and $y = 11.64 - 0.50x$

Find regression coefficient and also regression line coefficient.

Answer: $b_{yx} = -0.50$; $b_{xy} = -0.87$; r = -0.66 (negative because b_{yx} and b_{xy} both are negatives)

Q.26. (UTU 2017, 10 marks): Obtain the rank correlation coefficient for the following data:

x	68	64	75	50	64	80	75	40	55	64
y	62	58	68	45	81	60	68	48	50	70

Answer: Solved in this module.

Q.27. (UTU 2015, 5 marks): The following data regarding the heights (y) and weights (x) of 100 college students are given:

$\Sigma x = 15000$; $\Sigma x^2 = 227{,}2500$; $\Sigma y = 6800$; $\Sigma y^2 = 463025$; $\Sigma xy = 102{,}2250$

Find the coefficient of correlation between height and weight and also the equation of regression of height and weight.

Answer: $y = 0.1x + 53$; $x = 3.6y - 94.8$

Q.28. (UTU 2016, 5 marks): If $r_{12} = 0.25$, $r_{13} = 0.35$, $r_{23} = 0.45$ then find r_{213} (multiple correlation coefficient).

Q.29. (UTU 2018, 10 marks): Find the regression line of y on x and x on y for the following data:

x	1	3	4	6	8	9	11	14
y	1	2	4	4	5	7	8	9

Answer: $x = 0.5 + 1.5y$, $y = 0.64x + 0.55$

Q.30. (UTU 2018, 10 marks): Find the coefficient of correlation between the values of x and y from the following data:

x	1	3	5	7	8	10
y	8	12	15	17	18	29

Answer: 0.988

Q.31. (UTU 2013, 10 marks): Intelligence test of two groups of boys and girls gave the following results:

	Means	SD	Size
Girls	84	10	121
Boys	81	12	81

Is the difference in mean scores significant? Also test for SD.

Answer: A similar problem is solved in this module.

Q.32. (UTU 2013, 10 marks): It is desired to compare three hospitals with regards to the number of deaths per month. A sample of death records were selected from the records of each hospital and the number of deaths were given below. From these data, suggest the difference in the number of the deaths per month among three hospitals.

Hospitals		
A	B	C
3	6	7
4	3	3
3	3	4
5	4	6
0	4	5

STATISTICAL QUALITY CONTROL

Q.33. (UTU 2014, 10 marks): A quality control engineer of a tyre company collected the following data. For each of 12 batches of 1000 tyres, he tested 5 tyres and recorded the following results with $\overline{x}$ and R measured in thousands of kilometres.

Batch	1	2	3	4	5	6	7	8	9	10	11	12
$\overline{x}$	51.5	49.7	50	50.7	50.7	50.6	49.8	51.1	50.2	50.4	50.6	50.7
R	1.5	1.6	1.8	0.1	0.9	2.1	0.3	0.8	2.3	1.3	2	2.1

Draw the control chart. Is the production process in control? Explain. Take $A_2 = 0.577$

Answer: For $\overline{X}$ chart, CL = 50.5, UCL = 51.31, LCL = 49.69

Q.34. (UTU 2017, 10 marks): The following are the mean lengths and ranges of lengths of a finished product from 10 samples each of size 5. The specification limits for length are 200 ± 5 cm. Construct $\overline{x}$ and R chart and examine whether the process is under control and state your recommendations.

Sample No.	1	2	3	4	5	6	7	8	9	10
Mean x	201	198	202	200	203	204	100	196	199	201
Range R	5	0	7	3	3	7	2	8	5	6

Answer: Solved in this module.