Page No.
Date: / /

The Role of statistics in Engineering:-

→ Statistics is the science of data.

→ It deals with collection, Presentation, Analysis and used of data to make decisions, solve Problems and design Product and Process

Ex:- Hospital Emergency Department.

→ statistical method are used to help us describe and understand variability. Ex:- Ride bike/vechile.

classification:-

→ The Process of arranging the data into groups (or) Classes according to resemblances and similarities is technically called Classification.

Bases of classification:-

① Geographical :- Area wise (or) Regional

(ii) Chronological:- wrt occurrence of time.

(iii) Qualitative:- wrt character (or) attribute.

(iv) Quantitative:- wrt numerical values (or) magnitudes.

Variable:-

→ The characteristics, which is capable of direct quantitative measurement is called variable

Ex:- Height, weight etc.

Primary Data:-

→ Data collected directly from the source through methods like surveys, interviews, observations, experiments etc.

Ex:- Data collected from the experiments performed by reasearcher.

Secondary Data:-

→ Data has already been collected and is available for use by others.

Ex:- Informations from company reports.

Tabulation of Data :-

→ The statistical table may be defind as logical and systematic arrangement of statistical data in Rows and columns.

→ It is designed to simplify the presentation of data for the purpose of analysis and statistical inferences.

Procedure:-

(I) Table NO : Ta/01

(II) Title :

(III)

stub	caption							Total
items	sub-Head			sub-Head				
	c_1	c_2	c_3	c_1	c_2	c_3	c_4	
Total								

Foot Note: Source: MP/33/17/8/24

Q) Out of total no. of 2807 women, who were interview for employement in a textile factory 912 where from textile areas and the rest from non textile areas. Among them the married women who belongs to textile areas, 347 were having some work experience and 173 did not have work experience, while for non textile area the corresponding figures were 199 and 670 resp. The total number of women having no experience was 1841 of whom 311 resides in textile areas. Of the total no. of women, 1418 were unmarried and of these the number of women having experience in the textile and non textile areas was 254 and 166 resp. Tabulate the above information.

Page No.
Date: / /

1) Table No : Ta/02

(ii) Title: Employment of women in Textile & non Textile area

Stub	Textile			Non - Textile			Total		
Items	Emp.	Non emp.	Total	Emp.	Non Emp.	Total	Emp.	Non Emp.	Total
Married	347	173	520	199	670	869	546	843	1389
Unmarried	254	138	392	166	860	1026	420	998	1418
Total	601	311	912	365	1530	1895	966	1841	2807

Foot note:-

Source:- Book

P-2nd./15/7/25

Q) In a sample study about coffee habit in two town in the year of 2000, the following information was received
Town 'A' : Females were 40%, Total coffee drinkers were 45%, male non coffee drinkers were 20%.
Town 'B' : Males were 55%, male non coffee drinker 30% and female coffee drinkers were 15%.
Tabulate the above informations.

Ans→ Table No : Ta/02

Title:- Number of males or females between two town drink coffee or not.

Stub	Town A			Town B			Total
items	C.D.	N.C.D	Total	C.D	N.C.D	Total	
male	40%	20%	60%	25%	30%	55%	
female	5%	35%	40%	15%	30%	45%	
Total	45%	55%	100%	40%	60%	100%	

Foot note:-

Source - Q.P.

P. 2nd. 17/5/25

Page No.
Date: / /

2) In 1990 out of total 2000 students in a college, 1400 were for graduation and rest of the post graduation. Out of 1400 graduate student, 100 were girls, however in all there were 600 girls in the college. In 1995 no. of graduate students increased to 1700, out of which 250 were girls but the no. of PG students fall to 500 of which only 50 were boys. In 2000, out of 800 girls, 650 were for graduation, where as the total number of graduates were 2200. The no. of boys and girls in PG classes are equal. Represent the above in the tabular form also calculate the percentage increased in the no. of graduate student in 2000 as compared to 1990.

Ans→ Table No:- Ta/03

Title:- In 3 years the number graduate and post graduate students in a college.

Stub	Graduate			Post graduate			Total		
Items	Boys	Girls	Total	Boys	Girls	Total	Boy	Girl	Total
1990	1300	100	1400	100	500	600	1400	600	2000
1995	1450	250	1700	50	450	500	1500	700	2200
2000	1550	650	2200	150	150	300	1700	800	2500

Footnote:-

Source:- Q.P.

P.2nd. 21/7/25

Percentages

$$\frac{2200-1400}{1400} \times 100 = 57.1\%.$$

Date: / /

Stem and leaf :-

→ A stem and leaf diagram splits each number in a data set into a 'stem' (the leading digit) and 'leaf' (the last digit).

→ It helps in quickly seen the frequency and distribution of data values.

Example :-

Test score out of 100 of 12 students : 48, 52, 56, 59, 61, 63, 63, 64, 68, 70, 72, 75.

Sol^n :-

Stem = Ten digit place
leaf = unit place

stem	leaf
4	8
5	2 6 9
6	1 3 3 4 8
7	0 2 5

Q) 12.3, 12.7, 13.2, 13.5, 14.1, 14.6, 14.9, 15.0, 15.4, 15.6, 15.9, 16, 16.2, 16.4, 16.8, 17, 17.3, 17.6, 18, 18.5, 18.9

Sol^n :-

Stem	leaf
~~12~~	~~.3 .7~~
~~13~~	~~.2 .5~~
~~14~~	~~.1 .6 .9~~
~~15~~	~~.0 .4 .6 .9~~
~~16~~	~~.0 .2 .4 .8~~
~~17~~	~~.0~~

Ans :-

Stem	leaf
12	3 7
13	2 5
14	1 6 9
15	0 4 6 9
16	0 2 4 8
17	0 3 6
18	0 5 9

Page No.
Date: / /

Q) 1.758, 2.05, 2.114, 3, 3.1, 3.79, 4.0, 4.11, 5.9, 6.999

Ans :- ~~Stem~~ ~~leaf~~

Method-1 ~~stem~~

Round off the digits :

1.758 = 1.8
2.05 = 2.1
2.114 = 2.1
3 = 3.0
3.1 = 3.1
3.79 = 3.8
4.0 = 4.0
4.11 = 4.1
5.9 = 5.9
6.999 = 7.0

Stem	leaf
1	8
2	1 1
3	0 1 8
4	0 1
5	9
7	0

Method-2 :- (Direct)

Stem	leaf
1	758
2	050 114
3	000 100 790
4	000 110
5	900
6	999

H.W.

Example-3 In a trip organized by a college, there were 100 persons the average cost works out of Rs. 15.60 per head. There were 80 students each of whom paid Rs. 16 members of the teaching staff were charged at a higher rate. The number of servants was 6 (all males) and they were not charged. The number of ladies was 20% of the total of which two were ladies staff members.

Ans → Table ⊢ Ta/04

~~Table~~ Title: A trip organized by a college

Stub Items	Number of Persons	Rate charged (RS)	Total amount
Student	80	16.00	1,280
Teaching Staff	14	20	280
servant (all male)	6	0.00	0.00
Total	100	15.60	15.60

Example-7 In 2,000, out of total 1750 workers of a factory, 1200 were members of a trade union. The number of women employed was 200 of which 175 did not belong to a trade union. In 2002 the number of union workers increased to 1580 of which 1290 were men. On the other hand, the number of non-union workers fell down to 208, of which 180 were men.
In 2004, there were 1,800 employees who belong to a trade union and 50 who did not belong to a trade union, of all the employees in 2004, 300 were women of whom only 8 did not belong to trade union. Present the above data in a suitable form.

Table ⊢ Ta/05

Title ⊢

Stub Item	Trade union			NOT a union			Total		
	men	women	Total	men	women	Total	men	women	Total
2000	1175	25	1200	375	175	550	1550	200	1750
2002	1290	290	1580	180	28	208	1470	318	1788
2004	1508	~~220~~ 292	1800	42	8	50	1550	300	1850

Foot note ⊢ Source ⊢ g.P.

Example-9 In 2002, the number of workers in the trade union was 3450 of which 3200 were men. The number of non-trade union workers was 760 of which 330 were women.
In 2003, out of a total of 4,000 workers in a factory. 3,300 were members of a trade union. The number of women workers employed was 500 out of which 400 did not belong to any union. Present the following information in a suitable form.

Ans) Table ⊢ Ta/06

Title ⊢

Stub Item	Trade			Non trade			Total		
	men	women	Total	men	women	Total	men	woman	Total
2002	3200	250	3450	430	330	760	3630	580	4210
2003	3200	100	3300	300	400	700	3500	500	4000

Foot note ⊢ Source - g.P.

Example-11 A Survey of 370 students from Commerce faculty and 130 students from science faculty revealed that 180 students were studying for only C.A. Examination, 140 for only costing Examination and 80 for both C.A. and Costing Examinations. The rest had offered Part-time managment course. Of those studying for costing only, 13 were girls and 90 boys belong to commerce faculty. Out of 80 studying for both C.A. and Costing 72 were from commerce faculty amongest which 70 were boys. Amongst those who offered Part-time management course. 50 boys were from science faculty and 30 boys and 10 girls from Commerce faculty. In all there were 110 boys in science faculty.

Table -> Ta/04

Title ->

Stub Items	Stream Science B	G	T	Commerce B	G	T	Total Stud B	G	T
Management	50	10	60	30	10	40	80	20	100
CA			25			155			180
Costing		0	37	90	13	103			140
CA & Costing		.	8	70	2	72			80
Total	110	20	130			370			500

Histogram :-

→ A histogram is a graph containing a set of rectangles

→ Each being constructed to represent the size of the class intervals by it's width and the frequency in each class intervals by it's height.

Inclusive	Exclusive
$a \le x \le b$	$a \le x < b$
0 - 9	0 - 10
10 - 19	10 - 20
20 - 29	20 - 30

M -> 'kink' sign.

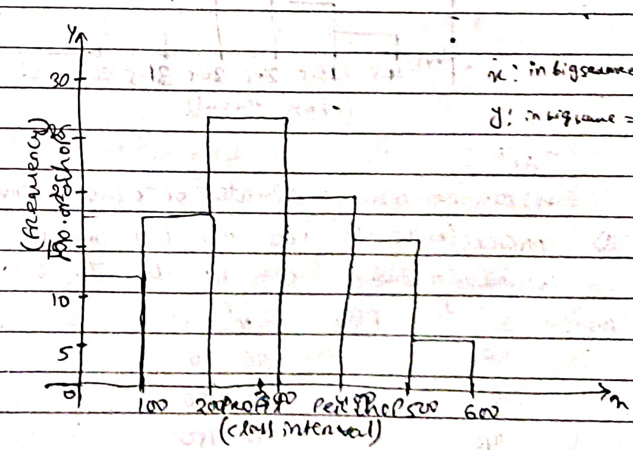
Type-1 Histogram :-

Histogram with equal class intervals :-

Example- The monthly Profit in rupees in 100 shops distributed as follows.

Profit Per shop	0 - 100	100 - 200	200 - 300	300 - 400	400 - 500	500 - 600
No. of shop	12	18	27	20	17	6

Ans ->



x : in big square = 1 ...
y : in big square = 5 ...

Page No.
Date: / /

Type-2:-

Histogram when class intervals are not continuous:-

Q)

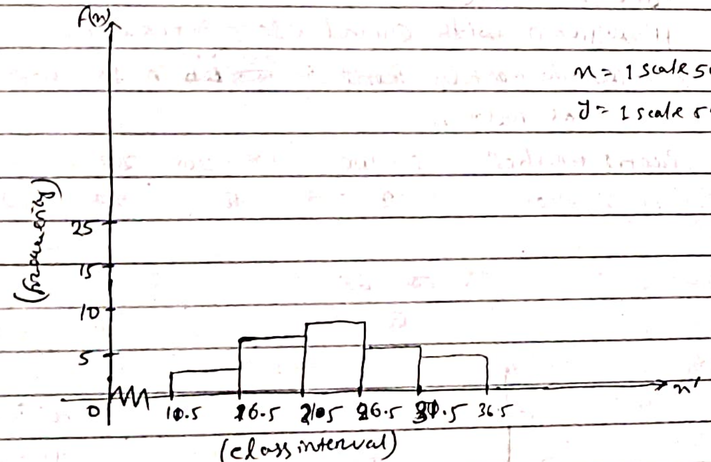
Class interval	12-16	17-21	22-26	27-31	32-36
Frequency	2	6	7	5	3

Ans→

x	$f(x)$	x'
12-16	2	11.5-16.5
17-21	6	16.5-21.5
22-26	7	21.5-26.5
27-31	5	26.5-31.5
32-36	3	31.5-36.5

formula
$h=1$ $\frac{h}{2}=0.5$
$\left(I-\frac{h}{2}, I+\frac{h}{2}\right)$

$x = 1$ scale 5 unit
$y = 1$ scale 5 unit



Type-3:-

Histogram when mid points of class interval are given:-

Q)

Marks (Mid Point)	100	120	140	160	180	200
Number of Students	5	6	4	7	5	3

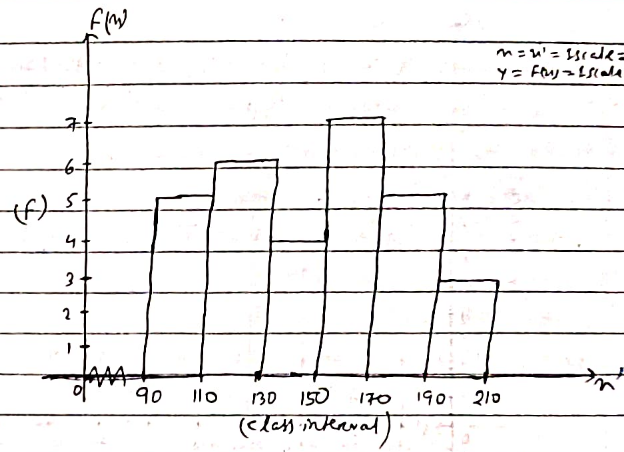
Ans→

x (mid)	$f(x)$	x'
100	5	90-110
120	6	110-130
140	4	130-150
160	7	150-170
180	5	170-190
200	3	190-210

$h=20$, $h/2=10$
$100-h/2$, $100+h/2$

Page No.
Date: / /

$x = x' = 1$ scale $= 20$ units
$y = f(x) = 1$ scale $= 1$ units



H.W:-

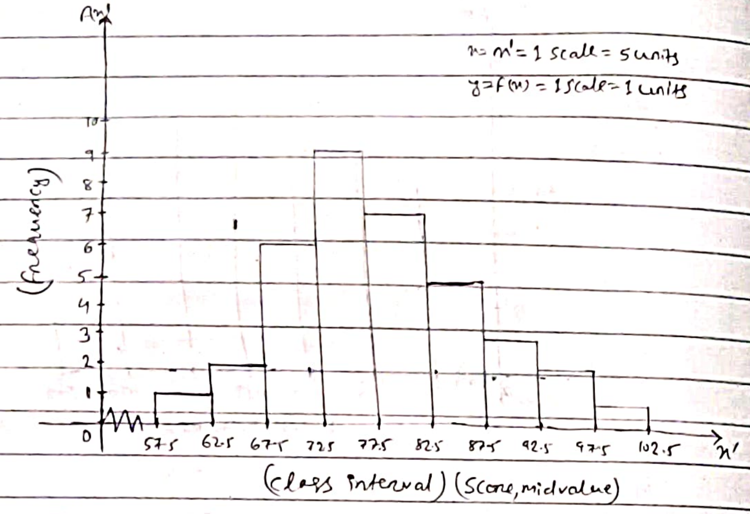
Q)

Score (mid value)	60	65	70	75	80	85	90	95	100
Frequency	1	2	6	9	7	5	3	2	1

Ans→

x	$f(x)$	x'
60	1	57.5-62.5
65	2	62.5-67.5
70	6	67.5-72.5
75	9	72.5-77.5
80	7	77.5-82.5
85	5	82.5-87.5
90	3	87.5-92.5
95	2	92.5-97.5
100	1	97.5-102.5

Page No.
Date: / /



Example-3 :

Studentary	Total			
	Male	Female	total	Paid
Student	62	18	80	1280
Teaching staff	12	2	14	280
Non teaching staff (servant)	6	0	6	0
Total	80	20	100	1560

Page No.
Date: / /

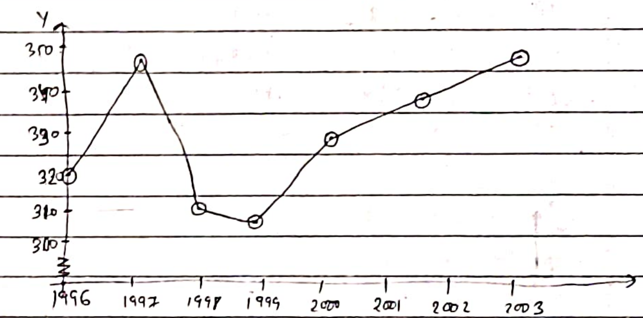
Time series graph:-

→ when we observe the values of a variable at different period of time, the series, so form is known as TSG.

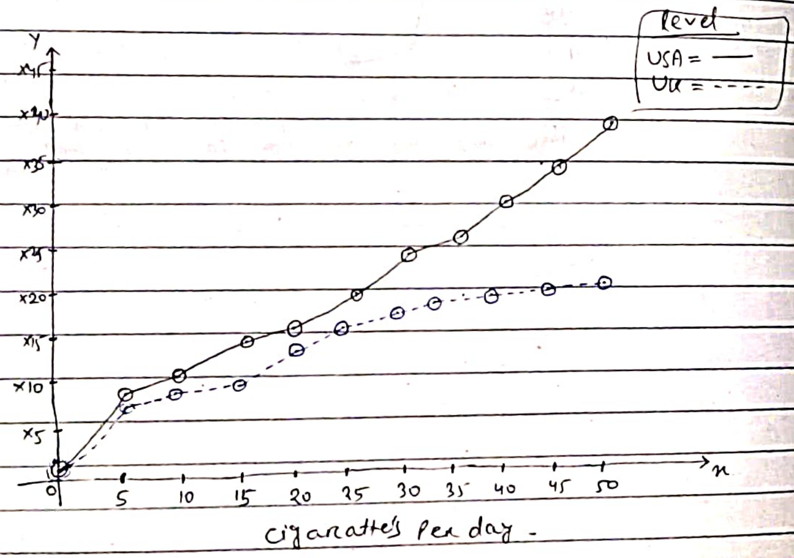
→ Take the time on x axis and the values of variables on Y-axis.

→ This is known as line graph.

Year	1996	1997	1998	1999	2000	2001	2002	2003
Percapita income	320	335.8	310.4	307.9	329.2	329.2	339.4	345.8



Cigaratte's per day	Increase of lung cancer USA	UK
0	x1	x1
5	x8	x7
10	x11	x8
15	x14	x9
20	x16	x13
25	x20	x16
30	x24	x17
35	x27	x18
40	x31	x19
45	x35	x20
50	x39	x21



Box Plot :- / whisker plot

→ A Box plot (or) whisker plot is a graphical representation in statistics that shows the distribution or database using 5 numbers of summeries :

→ The 5 numbers or summeries are :- minimum, 1st Quartile (Q_1), median (Q_2) 3rd quartile (Q_3), and maximum.

→ ~~out layer =~~

→ out lier = $[Q_1 - (1.5 \times IQR), Q_3 + (1.5 \times IQR)]$ (LB, UB)

· $IQR = Q_3 - Q_1$

→ IQR means Inter Quartile range.

→ 1st quartile / lower quartile : It is 25% of data.

→ median / 2nd quartile : It's 50% of data.

→ 3rd quartile / upper Q3 : It is 75% of data.

→ whisker : The line extend from the box to minimum and maximum is called whisker line.

Q) 4, 7, 8, 12, 17, 19, 25 Plot the box diagram.

Sol:- Arranging the data in Ascending / descending order

A.O = 4, 7, 8, 12, 17, 19, 25

minimum = 4 maximum = 25

median (Q_2) = 12

$Q_1 = 7$ $Q_3 = 19$

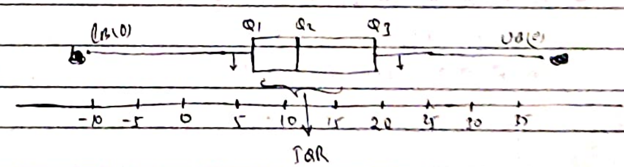
$IQR = Q_3 - Q_1 = 19 - 7 = 12$

outlier (LB) = $Q_1 - (1.5 \times IQR)$

$= 7 - (1.5 \times 12) = -11$

(UB) = $Q_3 + (1.5 \times IQR)$

$= 19 + (1.5 \times 12) = 37$.



⇒ It is a Positive skew Box plot.

Q) 40, 52, 55, 60, 70, 75, 85, 85, 90, 90, 92, 94, 95, 98, 100, 115, 125, 155.

Sol^n:- A.O. = 40, 52, 55, 60, 70, 75, 85, 85, 90, 90, 92, 94, 95, 98, 100, 115, 125, 155.

median = $\frac{90+90}{2} = 90 = Q_2$

$\boxed{\frac{18}{2} = 9 \,/\, 9+1 = 10}$

$Q_1 = 70$, $Q_3 = 98$

$IQR = 98 - 70 = 28$

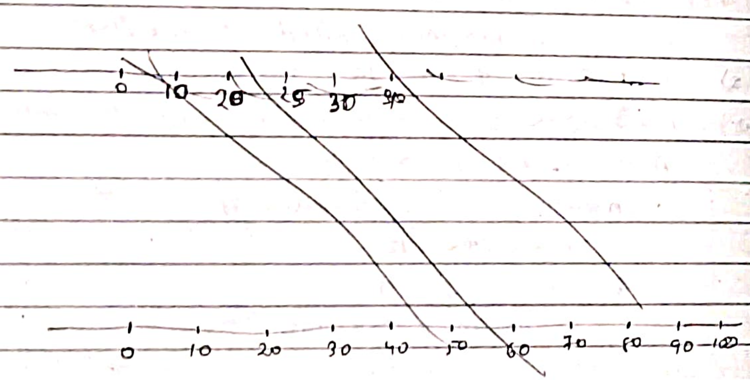
$LB(O) = Q_1 - (1.5 \times IQR)$

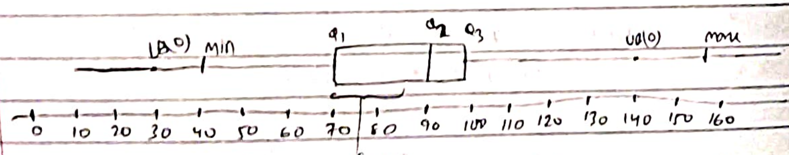
$= 70 - (1.5 \times 28) = 28$

$UB(O) = Q_3 + (1.5 \times IQR)$

$= 98 + (1.5 \times 28) = 140$

Minimum = 40, maximum = 155





→ It is a ~~positive~~ negative skew box plot.

Frequency distribution :-

→ Frequency distribution of a variable x in the order set {x, f}, where f is the frequency.

→ Frequency distribution are 2 types :-

(i) Discreate F.D.

(ii) Group F.D. / continuous F.D.

Discreate F.D :-

Q) From a Discreate F.D. from the following score : 15, 18, 16, 20, 25, 24, 25, 20, 16, 15, 18, 18, 16, 24, 15, 20, 24, 30, 27, 16, 24, 25, 26, 18, 27, 25, 28, 24, 24, 16, 18, 25, 20, 16, 15, 20, 27, 24, 28, 16.

Ans:-

x	Tally	f	C.F. (Cummulative Frequency)
15	||||	4	4
16	~~||||~~ |	6	10
18	~~||||~~ |	6	16
20	~~||||~~ |	6	22
24	~~||||~~	5	27
25	~~||||~~	5	32
27	|||	3	35
28	|||	3	38
29	|	1	29
30	|	1	40

Continuous frequency distributions

Q) Construct a continuous a F.D. table of water tax bill of 30 houses in a locality: 144, 184, 130, 195, 132, 134, 196, 114, 212, 174, 188, 210, 202, 145, 175, 154, 174, 178, 166, 146, 135, 115, 120, 114, 140, 188, 176, 166, 210, 208.

Soln. A.S.O:- 114, 114, 115, 120, 130, 132, 134, 135, 140, 144, 145, 146, 154, 166, 166, 174, 174, 175, 176, 178, 184, 188, 188, 195, 196, 202, 208, 210, 210, 212

min = 114, max = 212.

Range = 212 − 114 = 98

class size = 10 (say)

Class interval = $\frac{\text{Range}}{\text{class size}} = \frac{98}{10} = 9.8 \simeq 10$.

Class interval	Tally	f	c.f
114 − 124	IIII	4	4
125 − 135	IIII	4	8
136 − 146	IIII	4	12
147 − 157	I	1	13
158 − 168	II	2	15
169 − 179	~~IIII~~	5	20
180 − 190	III	3	23
191 − 201	II	2	25
202 − 212	~~IIII~~	5	30

Class size = difference between two numbers.

49

Q) The class mark of a frequency distributions are 6, 10, 14, 18, 22, 26, 30. Find the class size and the class interval.

Soln. class size = 4

$\frac{a+(a+4)}{2} = \frac{4+8}{2} = 6$

Range = 30 − 6 = 24

Class intervals = $\frac{\text{Range}}{\text{C.S}} = \frac{24}{4} = 6$

~~class interval~~	midterm	f	c.f.
4 − 8	6	1	1
8 − 12	10	1	2
12 − 16	14	1	3
16 − 20	18	1	4
20 − 24	22	1	5
24 − 28	26	1	6
28 − 32	30	1	7

Page No.
Date: / /

Q) The distance convered by 24 cars in two hour : 125, 140, 128, 108, 96, 149, 136, 112, 84, 123, 130, 120, 103, 89, 65, 103, 145, 97, 102, 87, 67, 78, 98, 126. Represent a cummutative frequency table using 60 as the lower of the first term and all the class having class size 15.

Sol^n: A.O:- 65, 67, 78, 84, 87, 89, 96, 97, 98, 102, 103, 103, 108, 112, 120, 123, 125, 126, 128, 130, 136, 140, 145, 149

Class size = 15

min = 65, max = 149, Range = 149 − 65 = 84

class interval = $\frac{84}{15}$ = 5.6

Class Interval	Tally	f	C.F
60 − 75	II	2	2
76 − 91	IIII	4	6
92 − 107	~~IIII~~ I	6	12
108 − 123	IIII	4	16
124 − 139	~~IIII~~	5	21
140 − 155	III	3	24

Page No.
Date: / /

Measure of Central Tendency :-

→ In statistics, a measure of central tendency is a single value that represents the centre (or) typical values of data sets.

→ It gives an idea about the average behaviour of data.

→ The types of measure of C.T:-

① Arithmatic mean (A.M).
② median
③ mode
④ Geometric mean (G.M).
⑤ Harmonic mean (H.M.).

① Arithmatic mean (A.M.) :-

→ The A.M. is the sum of all observations divided by the total number of observations.

Formula :-

(i) For n observation $x_1, x_2, \ldots, x_n$ the A.M. is $\bar{x} = \frac{x_1+x_2+\ldots+x_n}{n}$

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

(ii) For group data, $\bar{x} = \frac{x_1f_1+x_2f_2+\ldots+x_nf_n}{f_1+f_2+\ldots+f_n} = \frac{\sum_{i=1}^{n} x_if_i}{\sum f_i} = \frac{\sum x_if_i}{N}$

∵ f_i = frequency

Page No.
Date: / /

Q)

C.I.	Frequency
0-10	4
10-20	6
20-30	10
30-40	5

Find the A.M.

Ans)

C.I.	f	x	fx
0-10	4	5	20
10-20	6	15	90
20-30	10	25	250
30-40	5	35	175

$$A.M. = \frac{5\times4+15\times6+10\times25+5\times35}{4+6+10+5} = 21.4$$

Q) Find the A.M. of marks of 5 student 10, 30, 25, 20, 15.

Ans) $A.M. = \frac{10+30+25+15+20}{5} = \frac{100}{5} = 20$

② Step Deviation method :-

Formula

$$\boxed{\bar{x} = A + \frac{h}{N}\sum_{i=1} f_i d_i}$$

where A = Assume mean.

h = magnitude of class interval.

$d_i = \frac{x_i - A}{h}$

Page No.
Date: / /

Q)

C.I.	f	x_i	$d_i = \frac{x_i - A}{h}$	$f_i d_i$
0-8	8	4	-2	-16
8-16	7	12	-1	-7
16-24	16	20	0	0
24-32	24	28	1	24
32-40	24	36	2	48
40-48	7	44	3	21

Here A = 20

h = 8

$N = \sum f_i$

$\bar{x} = A + \frac{h}{N}\sum f_i d_i$

$= 20 + \frac{8}{86} \times 70$

$= 26.5$

Q) Calculate the mean by step deviation method

mark	No. of student (f_i)	x_i	$d_i = \frac{x_i - A}{h}$	$f_i d_i$
0-10	6	5	-3	-18
10-20	5	15	-2	-10
20-30	8	25	-1	-8
30-40	15	35 ✓	0	0
40-50	7	45	1	7
50-60	6	55	2	12
60-70	3	65	3	9
				-8

A = 35

h = 10

$\sum f_i d_i = -8$

$N = \sum f_i = 50$

$\bar{x} = A + \frac{h}{N} \times \sum f_i d_i$

$= 35 + \frac{10}{50} \times (-8)$

$= 33.4$.

Page No.
Date: / /

L.S Properties:-

① The sum of the squares of the deviation of a set of values is minimum when taken about mean.

② For frequency distribution x_i represents set of data and f_i represents correspondence frequency. where $i = 1, 2, \ldots$

Let $Z = \sum_{i=1}^{n} (x_i - A)^2 f_i$ be the sum of the square of deviation from any arbitrary point A.

③ We have to prove that Z is minimum when $A = \bar{x}$.

④ Applying the principle of maxima (or) minima from calculus, Z is minimum for $\frac{\partial Z}{\partial A} = 0$ and $\frac{\partial^2 Z}{\partial A^2} > 0$.

⑤ $\frac{\partial Z}{\partial A} = 0$

$\Rightarrow \frac{\partial}{\partial A} \sum_{i=1}^{n} f_i (x_i - A)^2 = 0$

$\Rightarrow -2 \sum_{i=1}^{n} f_i (x_i - A) = 0$

$\Rightarrow \sum_{i=1}^{n} f_i (x_i - A) = 0$

$\Rightarrow \sum_{i=1}^{n} f_i x_i - A \sum_{i=1}^{n} f_i = 0$

$\Rightarrow A \sum f_i = \sum f_i x_i$

$\Rightarrow A = \frac{\sum f_i x_i}{\sum f_i} = \frac{\sum f_i x_i}{N} = \boxed{\bar{x}}$

$\frac{\partial^2 Z}{\partial A^2} = \frac{\partial}{\partial A}\left(-2 \sum_{i=1}^{n} f_i (x_i - A)\right)$

$= (-2)(-1) \sum_{i=1}^{n} f_i = 2 \sum_{i=1}^{n} f_i = 2N > 0$.

Conclusion:- frequency will never be zero.

Property-2:-

→ Algebric sum of deviation of a value from their arithmatic mean is 0.

→ If x_i, $(i = 1, 2 \ldots n)$ be the data base and f_i is the corresponding frequency then,

$$\boxed{\sum_{i=1}^{n} f_i (x_i - \bar{x}) = 0}.$$

Page No.
Date: / /

Imp Property-3:-

→ Mean of the composite series if $\bar{x}_i$, $(i = 1, 2, \ldots n)$ are the mean of K-component series of n_i, $(i = 1, 2, \ldots n)$ respectively then the mean $\bar{X}$ of the composite series obtained by combining the components of series given by the formula,

$$\boxed{\bar{X} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2 + \cdots + n_n \bar{x}_n}{n_1 + n_2 + \cdots + n_n}}.$$

Q) The average weekly salary of female employee in a firm is 5200 rupees and that of male employee is getting 4200 rupees. The mean salary of all the employees is 5000 rupees. Find the percentage of female and male employee.

Sol^n:- $\bar{x}_1$ = avg. of female salary = 5200

$\bar{x}_2$ = " " male " = 4200

$\bar{x}$ = total salary avg = 5000

$\bar{x} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2}$

$\Rightarrow 5000 (n_1 + n_2) = n_1 \bar{x}_1 + n_2 \bar{x}_2$

$\Rightarrow 5000 (n_1 + n_2) = n_1 5200 + n_2 4200$

$\Rightarrow 5000 n_1 + 5000 n_2 = 5200 n_1 + 4200 n_2$

$\Rightarrow 800 n_2 = 200 n_1$

$\Rightarrow 4 n_2 = n_1$

$\Rightarrow \frac{n_1}{n_2} = \frac{4}{1}$

$\because n_1$ = no. of female

n_2 = no. of male

% of female employee $= \frac{4}{1+4} \times 100$

$= 80\%$.

% of male employee $= \frac{1}{4+1} \times 100$

$= 20\%$.

Page No.
Date: / /

Q) A farm of readymade garments made for men and women shirt. It's overall profit average is 6%. ~~sales~~, Profit in men shirt 8%. ~~profit in sale of women shirt~~ ~~and~~ womens shirt comprise 60% of output, what is the average profit for sale in women shirt.

Ans) ~~output is~~ %. The avg. output is 100%.
The output from women shirt is 60%. (n_1)
The output from men " " 40%. (n_2)
Given overall profit $(\bar{X}) = 6\%$
The profit over menshirt $(\bar{x}_2) = 8\%$.

$$\bar{X} = \frac{\bar{x}_1 n_1 + \bar{x}_2 n_2}{n_1 + n_2}$$

$$\Rightarrow 6\% = \frac{\bar{x}_1 60 + (8\%)(40\%)}{(60\%) + (40\%)}$$

$$\Rightarrow 6\% = \frac{60\% \cdot \bar{x}_1 + 320\%}{100\%}$$

$$\Rightarrow 600\% = 60\% \cdot \bar{x}_1 + 320\%$$

$$\Rightarrow 60\% \cdot \bar{x}_1 = 280\%$$

$$\Rightarrow \bar{x}_1 = \frac{28\cancel{0}}{6\cancel{0}} = 4.66\%$$

~~Weighted~~
~~Weighted mean~~
→ ~~Let WAB is the set~~

Page No.
Date: / /

Weighted mean:-

→ Let w_i be the weight attached to the items x_i.
→ Then we define weighted arithmatic mean =

$$\boxed{WAM = \frac{\sum w_i x_i}{\sum w_i}}$$

Q) Find the simple and weighted arithmatic mean of the 1st n natural numbers, the weights being the corresponding numbers.

Ans)

x_i	w_i	$w_i x_i$
1	1	1
2	2	4
3	3	9
4	4	16
⋮	⋮	⋮
n	n	n×n

$$AM = \frac{\sum x}{n} = \frac{1+2+3+\cdots n}{n}$$

$$= \frac{\cancel{n}(n+1)}{2 \times \cancel{n}} = \frac{n+1}{2}$$

$$W.A.M = \frac{1^2+2^2+\cdots+n^2}{1+2+3+\cdots+n}$$

$$\frac{\sum w_i x_i}{\sum w_i} = \frac{\frac{n(n+1)(2n+1)}{6}}{\frac{n(n+1)}{2}}$$

$$= \frac{\cancel{n(n+1)}(2n+1)}{\cancel{6}\,3} \times \frac{\cancel{2}}{\cancel{n(n+1)}}$$

$$= \frac{(2n+1)}{3}$$

Corrected mean:-

$$C.M = \frac{\text{Incorrect } \sum x - \text{Incorrect item} + \text{correct item}}{n}$$

Q) The mean mark score by 100 student was found to be 40. Later on it was discovered that a score of 53 was misread as 83. Find the correct mean.

Ans) ~~Incorrect~~ $\bar{x} = 40$

$n = 100$, $\bar{x} = \frac{\sum x}{n} \Rightarrow \sum x = n\bar{x}$

$\sum x = 40 \times 1000 = 4000$

$$C.M. = \frac{4000 - 83 + 53}{100} = \frac{4000 - 30}{100} = 39.7.$$

Page No.
Date: / /

Q) mean of 25 observations was found to be 78.4, but later on it is found that 96 is misread as 69, find correct mean.

Ans-> $c.m. = \frac{25 \times 78.4 - 69 + 96}{25} = 79.48$

Median (Md or M):-

-> Median is defined as the middle most or the central value of the variables in a set of observations, when the observations are arranged either in ascending order (or) in desecending order.

Formula (1):-

Data is ungrouped:-

case-1:- when n is odd, the median Md or $M = \left(\frac{n+1}{2}\right)^{th}$ term

case 2:- when n is even, then median $M_d = \frac{\left(\frac{n}{2}\right)^{th} \text{term} + \left(\frac{n}{2}+1\right)^{th} \text{term}}{2}$

Q) 31, 35, 27, 32, 43, 37, 41, 34, 28, 36, 44, 45, 42, 30.
Find the median.

Ans-> A.O:- 27, 28, 30, 31, 32, 34, 35, 36, 37, 41, 42, 43, 44, 45

$n = 14$

$\left(\frac{n}{2}\right)^{th}$ term $= 35$, $\left(\frac{n+1}{2}\right)^{th}$ term $= 36$

$\frac{35+36}{2} = 35.5$.

Q) 6, 10, 4, 3, 9, 11, 22 find the median.

Ans-> A.O:- 3, 4, 6, 9, 10, 11, 22

$n = 7$

$\left(\frac{n+1}{2}\right)^{th} = 4^{th}$ term $= 9$

Page No.
Date: / /

Case:- If n is variable which taken the values $x_1, x_2, \ldots, x_n$ with frequencies $f_1, f_2, \ldots, f_n$ resp. Then the median of the given data is,

$M = \left(\frac{n+1}{2}\right)^{th}$ of data.

Q)

n	f
20	4
25	16
40	6
50	8
80	6

Ans->

n	f	c.f.
20	4	4
25	16	20
40	6	26
50	8	34
80	6	40

$M = \left(\frac{n+1}{2}\right)^{th}$ of data $= \left(\frac{40+1}{2}\right)^{th}$ of data $= 20.5$ of data

Here median is $\underset{=}{M} = 40$.

Case:- The median for the continous frequency data distribution is

$M = l + \frac{h}{f}\left(\frac{N}{2} - C\right)$. where l = lower limit of median class

h = width of the class interval.

f = frequency of the median class.

N = submession of frequency.

C = Cummulative frequency of the class preceding to the median class.

Page No.
Date: / /

Q)

C.I.	f
0-10	14
10-20	23
20-30	27
30-40	21
40-50	15

Ans→

C.I.	f	C.f
0-10	14	14
10-20	23	37
20-30	27	64
30-40	21	85
40-50	15	100
	N=100	

$N = \frac{100}{2} = 50$

$64 > 50$

$l = 20$

$h = 10$

$f = 27$

$N = 100$

$c = 37$

$M = l + \frac{h}{f}\left(\frac{N}{2} - c\right)$

$= 20 + \frac{10}{27}\left(\frac{100}{2} - 37\right)$

$= 24.81$

Q)

C.I.	f	C.f
0-5	6	6
5-10	12	18
10-15	50	68
15-20	120	188
20-25	225	413
25-30	250	663
30-35	185	848
35-40	110	958
40-45	32	990
45-50	10	1000
	N=1000	

Page No.
Date: / /

$N = \frac{1000}{2} = 500$

$663 > 500$

$l = 25$

$h = 5$

$f = 250$

$N = 1000$

$C = 413$

$M = l + \frac{h}{f}\left(\frac{N}{2} - C\right)$

$= 25 + \frac{5}{250}\left(\frac{1000}{2} - 413\right)$

$= 26.74$

Q) The median distribution is 30 marks. Find the missing frequency.

Marks (C.I.)	No. of students (f)
0-10	10
10-20	?
20-30	25
30-40	30
40-50	?
50-60	10
	100

$\rightarrow f_1 = 25 - 10 = 15$

The number of students are 100 and find the missing frequency

Ans→

$N = 100$

$m = 30$

$l = 30$

$h = 10$, $f = 30$

$C = ?$

$m = l + \frac{h}{f}\left(\frac{N}{2} - c\right)$

$\Rightarrow 30 = 30 + \frac{10}{30}\left(\frac{100}{2} - c\right)$

$\Rightarrow c = 50$

$f_2 = 100 - 90 = 10$

Page No.
Date: / /

Partition values

→ The division of array series into more than two equal parts, the dividing places are known as Partition value.

→ 3 types:-

① Quartile
② Decile
③ Percentile

① Quartile

$Q_1 \quad Q_2 \quad Q_3$

$Q_1 = \frac{1}{4}^{th}$ data

$Q_2 = \frac{1}{2}^{th}$ data

$Q_3 = \frac{3}{4}^{th}$ data

$$Q_1 = l + \frac{h}{f}\left(\frac{N}{4} - c\right)$$
$$Q_2 = l + \frac{h}{f}\left(\frac{N}{2} - c\right)$$
$$Q_3 = l + \frac{h}{f}\left(\frac{3N}{4} - c\right)$$

② Deciles:-

$D_1 < D_2 < \ldots < D_9$

$\frac{iN}{10}$, $i = 1, 2, \ldots, 10$

$$D_i = l + \frac{h}{f}\left(\frac{iN}{10} - c\right)$$

③ Percentiles:-

100 equal parts divisions:-

$P_1 < P_2 < P_3 < \ldots < P_{99}$

$\frac{iN}{100}$, $i = 1, 2, \ldots$

$$P_i = l + \frac{h}{f}\left(\frac{iN}{100} - c\right)$$

Page No.
Date: / /

Q)

Mark	No. of student
0-10	4
10-20	10
20-30	30
30-40	40
40-50	47
50-60	50

Find the Quartiles, 10 Percentile, and 6 Decile.

Ans→

C.I.	f	c.f
0-10	4	4
10-20	10	14
20-30	30	44
30-40	40	84
40-50	47	131
50-60	50	181

For quartile (Q_1):-

$N = \frac{181}{4} = 45.25$

$l = 30$ $84 > 45.25$

$h = 10$

$f = 40$

$N = 181$

$c = 44$

For quartile (Q_2):-

$\frac{181}{2} = 90.5$ $131 > 90.5$

$l = 40$

$h = 10$

$f = 47$

$c = 84$

$N = 181$

For Q_3:-

$N = 181 = \frac{181 \times 3}{4} = 135.75$

$181 > 135.75$

$l = 50, h = 10, f = 50, N = 181, c = 131$

$\cancel{Q_1 = 40 + \frac{10}{47}\left(\frac{181}{4} - 84\right) = 31.75}$

$\cancel{Q_2 = 40 + \frac{10}{47}\left(\frac{181}{2} - 84\right) = 41.38}$

$\cancel{Q_3 = 40 + \frac{10}{47}\left(\frac{3 \times 181}{4} - 84\right) =}$

$Q_1 = 30 + \frac{10}{40}\left(\frac{181}{4} - 44\right) = 30.31$

$Q_2 = 40 + \frac{10}{47}\left(\frac{181}{2} - 84\right) = 41.38$

$Q_3 = 50 + \frac{10}{50}\left(\frac{181 \times 3}{4} - 131\right) = 50.95$

For 10 Percentiles

$N = 181$, $\frac{10 \times 181}{100} = 18.1$ / $44 > 18.1$

$l = 20$, $f = 30$, $181 = N$

$h = 10$, $c = 14$

$P_{10} = 20 + \frac{10}{30}\left(\frac{10 \times 181}{100} - 14\right) = 21.36$

For 6 Deciles:-

$N = 181$, $\frac{6 \times 181}{10} = 108.6$ / $131 > 108.6$

$l = 40, h = 10, f = 47, c = 84, N = 181$

$D_6 = 40 + \frac{10}{47}\left(\frac{6 \times 181}{10} - 84\right)$

$= 45.23$.

Page No.
Date: / /

Q) Find the range of the marks obtained by middle 80% of student.

Mark	No. of Students (c.f)
10 marks (or) less	4
20 marks or less	10
30 marks (or) less	30
40 marks (or) less	40
50 marks (or) less	47
60 marks (or) less	50

Ans →

Marks	C.F	F	C.I	C.I
≤ 10	4	4	0 - 10	0.5 - 10.5
≤ 20	10	6	11 - 20	10.5 - 20.5
≤ 30	30	20	21 - 30	20.5 - 30.5
≤ 40	40	10	31 - 40	30.5 - 40.5
≤ 50	47	7	41 - 50	40.5 - 50.5
≤ 60	50	3	51 - 60	50.5 - 60.5
	50			

P_{10} :

$N = 50$, $\frac{10 \times 50}{100} = 5$

$l = 10.5$, $h = 10$, $C = 4$, $f = 6$

$P_{10} = 10.5 + \frac{10}{6}(5 - 4) = 12.16$

P_{90} :-

$N = 50$, $\frac{90 \times 50}{100} = 45$

$l = 40.5$, $h = 10$, $C = 40$, $f = 7$,

$P_{90} = 40.5 + \frac{10}{7}(45 - 40)$

$= 47.64$

$P_{90} - P_{10} = 47.64 - 12.16 = 35.48$.

Page No.
Date: / /

Mode :- The value which is repeated maximum number of times is the mode of the series.

Ex:- 5, 6, 6, 6, 8, 9, 10 :

mode :- 6.

→ Discrete Frequency Distribution series:-

Mode is the value of the variable corresponding to the maximum frequency.

Ex:- size of the item 4 5 6 8

Frequency 4 5 6 9

modal value (mo) : 8

→ Mode by Grouping method

Q) ~~column (m) :~~ ~~Frequency~~

Q) ~~m :-~~

~~f :- 2 5 8 9 12 14 14~~

Q)

column (m) :	Frequency (f)
4	2
5	5
6	8
7	9
8	12
9	14
10	14
11	15
12	11
13	13

x	f (I)	II	III	IV	V	VI
4	2	7		15		
5	5		13		22	
6	8	17				29
7	9		21	35		
8	12	26			$\boxed{40}$ (8,9,10)	
9	14		$\boxed{28}$ (9,10)			$\boxed{43}$ (9,10,11)
10	14	$\boxed{29}$ (10,11)		$\boxed{40}$ (10,11,12)		
11	$\boxed{15}$ (11)		26		39	
12	11	24				
13	13					

Analysis table →

var → / cell ↓	4	5	6	7	8	9	10	11	12	13
I								1		
II							1	1		
III						1	1			
IV							1	1	1	
V					1	1	1			
VI						1	1	1		
Total					1	3	5	4	1	

Here 5 is the heighest.

So modal value $(M_o) = 10$ as the heighest weight is 5.

Q)

x	f (I)	II	III	IV	V	VI
10	10	22		37		
11	12		27		$\boxed{46}$ (11,12,13)	
12	15	$\boxed{34}$ (12,13)				$\boxed{54}$ (12,13,14)
13	19		$\boxed{39}$ (13,14)	$\boxed{47}$ (13,14,15)		
14	$\boxed{20}$ (14)	28			32	
15	8		12			15
16	4	7		15		
17	3		11			
18	8					

Analysis table:-

var → / cell	10	11	12	13	14	15	16	17	18
I					1				
II			1	1					
III				1	1				
IV				1	1	1			
V		1	1	1					
VI			1	1	1				
Total:-		1	3	5	4	1			

Here Modal value (M_o) is 13 as it has highest weight is 5.

④ Mode in continuous frequency distribution:-

Formula:- $\boxed{M_o = l + \frac{f_m - f_1}{2f_m - f_1 - f_2} \times h}$

where l = lower limit of modal class.

f_m = frequency of " "

f_1 = frequency preceding the modal class.

f_2 = frequency succeeding the " "

h = width of the modal class.

Page No.
Date: / /

Q)

C.I.	f
21-22	7
23-24	13
25-26	$\boxed{22}$
27-28	10
29-30	8

Ans - $f_m = 22$

$f_1 = 13$

$f_2 = 10$

$l = 25$

$h = 2$

$$M_o = l + \frac{f_m - f_1}{2f_m - f_1 - f_2} \times h$$

$$= 25 + \frac{22-13}{(2\times 22)-13-10} \times 2$$

$$= 25.85$$

Q)

C.I.	f
1-5	7
6-10	10
11-15	16
16-20	$\boxed{32}$
21-25	24

Ans - $f_m = 32$

$l = 16$

$f_1 = 16$

$f_2 = 24$

$h = 5$

$$M_o = l + \frac{f_m - f_1}{2f_m - f_1 - f_2} \times h$$

$$= 16 + \frac{32-16}{2\times 32-16-24} \times 5$$

$$= 19.33$$

Page No.
Date: / /

Emperical relationship between mean, median and mode +

$\boxed{Mode = 3\,median - 2\,mean}$

Q) If the value of the mean and mode is 66 and 60 respectively find the median.

Ans - Here mean = 66

mode = 60

Mode = 3median − 2mean

$\Rightarrow 60 = 3median - 2\times 66$

$\Rightarrow median = \frac{132+60}{3} = 64$

Geometric Mean:-

→ If $x_1, x_2, ..., x_n$ are n number of variables of x. None of them being zero, then the Geometric mean is defined as,

$$\boxed{G = (x_1, x_2 ... x_n)^{\frac{1}{n}}}$$

By taking logarithm of both side.

$$\log G = \frac{1}{n}[\log x_1 + \log x_2 + ... + \log x_n]$$

$$\log G = \frac{1}{n}\sum_{i=1}^{n} \log x_i$$

$$\boxed{G = \text{anti}\log\left[\frac{\sum_{i=1}^{n} \log x_i}{n}\right]}$$

→ In case of frequency distribution the Geometric mean is

$$G = (x_1^{f_1}\ x_2^{f_2} ....\ x_n^{f_n})^{\frac{1}{N}}$$

$N = \Sigma f_i$

$$\log G = \frac{1}{N}[f_1x_1\ f_2x_2 ... f_nx_n]$$

$$\log G = \frac{1}{N}\sum_{i=1}^{n} \log f_i x_i$$

$$\boxed{G = \text{anti}\log\left[\sum_{i=1}^{n} \frac{f_i \log x_i}{N}\right]}$$

Page No.
Date: / /

Q) Compute geometric mean of the data.
10, 110, 120, 50, 52, 80, 37, 60

Ans) $G_m = \text{antilog}\left[\frac{\sum \log x_i}{n}\right]$

x	$\log x$
10	1
110	2.04
120	2.07
50	1.69
52	1.71
80	1.90
37	1.56
60	1.77
	13.74

$\text{antilog}\left[\frac{\sum \log x_i}{n}\right]$

$= \text{antilog}\left(\frac{13.74}{8}\right)$

$= \text{antilog}(1.71)$

$\simeq 51.28$

Q) The geometric mean of 10 observation on a certain variable was calculated as 16.2. It was later discovered that one of the observation was wrongly written 12.9. Infact 21.9 then calculate correct geometric mean.

Ans) $(x_1, x_2 - -, x_n)^{1/10} = G$

$(x_1, x_2, \dots x_n)^{1/10} = 16.2$

$x_1, x_2, \dots x_n = (16.2)^{10}$

Let x_1 is wrong data $(x_1 = 12.9)$ $x_1' = 21.9$

$x_2, x_3 \dots x_{10} = \frac{(16.2)^{10}}{12.9}$

$x_1', x_2 \dots x_{10} = \frac{21.9 \times (16.2)^{10}}{12.9}$

$G' = (x_1', x_2 \dots x_{10})^{1/10} = \left[\frac{21.9 \times (16.2)^{10}}{12.9}\right]^{1/10}$

$= 17.08$

Formula: $\left[\frac{\text{Correct} \times (G)^{10}}{\text{wrong}}\right]^{1/10}$

Page No.
Date: / /

Harmonic mean (H): The harmonic mean (H) of n observation $x_i, i = 1, 2, \dots n$

i) $H = \frac{1}{\frac{1}{n}\sum\left(\frac{1}{x_i}\right)}$ ——— ①

→ In case of frequency distribution, the harmonic mean is

$H = \frac{1}{\frac{1}{N}\sum_{i=1}\left(\frac{f_i}{x_i}\right)}$ ——— ② $(\because N = \sum f_i)$

→ If $x_1, x_2, \dots x_n$ are n observations with the weight $w_1, w_2, \dots w_n$ respectively there weight harmonic mean

$H = \frac{1}{1/N \sum_{i=1}\left(\frac{w_i}{x_i}\right)}$ ——— ③

Total Price $= \frac{\text{Money}}{\text{quantity}}$

Speed $= \frac{\text{Distance}}{\text{Time}}$

work done perhour $= \frac{\text{Total work done}}{\text{time taken}}$

Q) Mineral water is sold at the rate of 8, 10, 12, and 15 rupees per liter in 4 different months. Assuming that equal amount are spent on water by a family in the 4 months. Find the average price in rupees per month.

Ans) $H = \frac{1}{\frac{1}{n}\sum_{i=1}\left(\frac{1}{x_i}\right)}$

$= \frac{1}{\frac{1}{4}\left(\frac{1}{8} + \frac{1}{10} + \frac{1}{12} + \frac{1}{15}\right)}$

$= 10.66$

Q) The arithmatic mean of 2 observation is 127.5 and the geometric mean is 60 then find the harmonic mean.

Ans) $\frac{a+b}{2} = 127.5$ $\quad (ab)^{1/2} = 60$

$a + b = 255$ $\quad ab = 3600$

$H.M = \frac{1}{\frac{1}{2}\left(\frac{1}{a} + \frac{1}{b}\right)}$ $\quad HM = \frac{GM^2}{AM}$

$HM = \frac{2}{\frac{1}{a} + \frac{1}{b}}$

$= \frac{2}{\frac{a+b}{ab}}$

$= \frac{2ab}{a+b} = \frac{2 \times 3600}{255} = 28.23$

$a + b = 255$, $a = \frac{3600}{b}$

$\frac{3600}{b} + b = 255$

$\Rightarrow 3600 + b^2 = 255b$

$\Rightarrow b^2 - 255b + 3600 = 0 \Rightarrow b = 15, a = 240$

$a + b = 255$ | $ab = 3600$

$b = 255 - a$ | $a(255 - a) = 3600$

$b = 15$ | $255a - a^2 = 3600$

$\Rightarrow a^2 - 255a + 3600 = 0$

$\Rightarrow a = 240$

Page No.
Date: / /

Measure of dispersion :-

→ The meaning of dispersion is scatteredness. It helps in finding out the variability of the data.

→ The first scene related to the limits within which the data form and the second takes into the account the amount, absolute or relative by which the value of the items differs from an average.

→ There are 4 types of dispersions :-

① Range dispersion.
② Quartile "
③ Mean "
④ Standard "

Range dispersion	Quartile dispersion
Range = $X_{max} - X_{min}$ $= L-S$	Q_1, Q_2, Q_3 $Q.D = \frac{Q_3-Q_1}{2}$
Coefficient Range $\frac{L-S}{L+S}$	Coefficient of $QD = \frac{Q_3-Q_1}{Q_3+Q_1}$

Find the coefficient of range dispersion :-

Mark	No. of Student
0-10	13
10-20	8
20-30	7
30-40	10
40-50	11
50-60	23
60-70	15
70-80	10

Find the range and coefficient

Range = L-S = 80 - 0 = 80

Coefficient $= \frac{80-0}{80+0} = 1$.

Q) If the first quartile is 104 quartile deviation 18. Find third quartile and coefficient of quartile deviation.

Ans. $Q_1 = 104$ $QD = 18$

$QD = \frac{Q_3-Q_1}{2}$ $= 2QD + Q_1$

$= 2 \times 18 + 104$

$\Rightarrow Q_3 = 140$

$QD = \frac{Q_3-Q_1}{Q_3+Q_1} = \frac{140-104}{140+104} = 0.1475409836.$

Page No.
Date: / /

Q)

x	f	c.f
0-10	20	20
10-20	45	65
20-30	85	150
30-40	160	310
40-50	70	380
50-60	55	435
60-70	35	470
70-80	30	500

From the above data calculate the % of workers getting wages more than 44 rupees and find the quartile deviation, the variable represent the wages in rupees frequency represent no. of workers.

Ans) 40 - 50 → 70 for 44 and more than $\frac{50-44}{10} \times 70 = 42$

$42 + 55 + 35 + 30 = 162$

$\boxed{\text{or If less than 44 then } \frac{44-40}{10} \times 72}$

% of workers getting more than 44 is $\frac{162}{500} \times 100 = 32.4$

$Q_1 = l + \frac{h}{f}\left(\frac{N}{4} - c\right)$

$= 20 + \frac{10}{85}\left(\frac{500}{4} - 65\right) = 27.05$

$Q_3 = l + \frac{h}{f}\left(\frac{3N}{4} - c\right) = 49.28$

$QD = \frac{49.28 - 27.05}{2} = 11.11$

Mean deviation for N observations.

$MD = \frac{1}{n}\sum_{i=1}^{n} |x_i - \bar{x}|$ (about mean)

$MD = \frac{1}{n}\sum_{i=1}^{n} |x_i - Md|$ (about median)

$MD = \frac{1}{n}\sum_{i=1}^{n} |x_i - Mo|$ (about mode)

Mean deviation for grouped data

$MD = \frac{1}{N}\sum_{i=1}^{n} f_i |x_i - \bar{x}|$ (about mean)

$MD = \frac{1}{N}\sum_{i=1}^{n} f_i |x_i - Md|$ (about median)

$MD = \frac{1}{N}\sum_{i=1}^{n} f_i |x_i - Mo|$ (about mode)

Sometime $d = |x_i - m|$

m = may be mean, median or mode

The coefficient of mean deviation

$C.M.D = \frac{MD}{\bar{x}}$ (about mean)

$C.M.D = \frac{MD}{median}$ (about median)

$C.M.D = \frac{MD}{mode}$ (mode)

Page No.
Date: / /

Q) Find coefficient of mean deviation.

c.I	f
0-10	8
10-20	12
20-30	10
30-40	8
40-50	3
50-60	2
60-70	7

By using step deviation method mean, mean deviation,
find M.D. and coefficient of M.D.

Q)

c.I	f	d_i	$f_i d_i$	$\lvert x_i - \bar{x} \rvert$	$f_i \lvert x_i - \bar{x} \rvert$
0-4	4	-2	-8	-7.2	28.2
4-8	6	-1	-6	-3.2	18.2
8-12	8	0	0	0.8	6.4
12-16	5	1	5	4.8	19
16-20	2	2	4	8.8	15.6
	$\Sigma f_i = 25$		$\Sigma f_i d_i = -5$		87.4

$\bar{x} = A + \frac{h}{N} \Sigma f_i d_i$

$= 10 + \frac{4}{25} \times -5 = 9.2$

$M.D = \frac{1}{N} \Sigma f_i \lvert x_i - \bar{x} \rvert$

$= \frac{1}{25} \times 87.4 = 3.84$

$C.M.D = \frac{M.D}{\bar{x}} = \frac{3.84}{9.2} = 0.417$

Page No.
Date: / /

Standard deviation:-

→ Standard deviation usually denoted by 'σ'.

→ It is the positive ~~square~~ root of the arithmatic mean of the square of the deviation of the given values of their arithmatic mean.

→ For the frequency distribution x_i, $i = 1, 2, \dots n$, at the set of data, f_i, $i = 1, 2, \dots, n$, set of respective frequency. then the sigma

$$S.D. = \sigma = \sqrt{\frac{1}{N} \Sigma f_i (x_i - \bar{x})^2} \quad \text{—— (1)}$$

where $\bar{x}$ is the arithmatic mean where $N = \Sigma f_i$

→ The square of standard deviation is called varience, and given by,

$$\sigma^2 = \frac{1}{N} \Sigma f_i (x_i - \bar{x})^2$$
$$var = \sigma^2 = \frac{1}{N} \Sigma f_i x_i^2 - \left(\frac{1}{N} \Sigma f_i x_i\right)^2 \quad \text{—— (2)}$$

→ Group mean square deviation denoted by S.

$$S = \sqrt{\frac{1}{N} \Sigma f_i (x_i - A)^2} \quad \text{—— (3)}$$

where A = arbitrary number
S^2 = mean square deviation

Varience of combined series:-

→ If n_1 and n_2 are the sizes, $\bar{x}_1$ and $\bar{x}_2$ are the corresponding arithmatic mean and σ_1 and σ_2 are the standard deviation of the two series., then the standard deviation σ of the combine series of size $n_1 + n_2$ is

$$\boxed{\sigma^2 = \frac{1}{n_1 + n_2} \left[n_1 (\sigma_1^2 + d_1^2) + n_2 (\sigma_2^2 + d_2^2) \right]}$$

where $d_1 = \bar{x}_1 - \bar{x}$
$d_2 = \bar{x}_2 - \bar{x}$

$$\left(\bar{x} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2} \right)$$
(combined mean)

Page No.
Date: / /

Q) Calculate mean and standard deviation for the following table keeping the edge distribution of 542 members of a club.

x_i	Age	No. of members (f)	d_i	$f d_i$	d^2	$f d^2$
25	20-30	3	-3	-9	9	27
35	30-40	61	-2	-122	4	244
45	40-50	132	-1	-132	1	132
55	50-60	153	0	0	0	0
65	60-70	140	1	140	1	140
75	70-80	51	2	102	4	204
85	80-90	2	3	6	9	18
		542		-15		765

$A = 55$,

$$d_i = \frac{25 - 55}{10} = -3$$

$$\bar{x} = A + \frac{h}{N} \Sigma f d_i = 55 + \frac{10}{542} \times (-15) = 54.72$$

$$\sigma^2 = h^2 \left[\frac{\Sigma f d^2}{N} - \left(\frac{\Sigma f d}{N} \right)^2 \right]$$
$$= 100 \left[\frac{765}{542} - \left(\frac{-15}{542} \right)^2 \right]$$
$$\boxed{\sigma = 11.87}$$

Coefficient of standard deviation $= \frac{\sigma}{\bar{x}} = \frac{11.87}{54.72} = 0.216$

Coefficient of variation $= \frac{\sigma}{\bar{x}} \times 100 = \frac{11.87}{54.72} \times 100 = 21.6$

Page No.
Date: / /

Q) Find out the mean and standard deviation of following data;

Age under	No of person dying	f	c.I
10	15	15	0-10
20	30	15	10-20
30	53	23	20-30
40	75	22	30-40
50	100	25	
60	110	10	
70	115	5	
80	125	10	

Sol:

c.I.	f	x	di	fidi	d²	fd²
0-10	15	5	-4	-60	16	240
10-20	15	15	-3	-45	9	135
20-30	23	25	-2	-46	4	92
30-40	22	35	-1	-22	1	22
40-50	25	45	0	0	0	0
50-60	10	55	1	10	1	10
60-70	5	65	2	10	4	20
70-80	10	75	3	30	9	90
	125			-123		609

$A = 45$

$d_i = \frac{x-A}{h} = \frac{5-45}{10} = -4$

$\bar{x} = A + \frac{h}{N}\Sigma f_i d_i = 45 + \frac{10}{125} \times (-123) = 35.16$

$\sigma^2 = h^2\left[\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2\right]$

$= 100\left[\frac{609}{125} - \left(\frac{-123}{125}\right)^2\right]$

$= 100\left[4.87 - 0.96\right]$

$\sigma^2 = 391$

$\sigma = \sqrt{391} = 19.77$.

$C.O.S.D. = \frac{\sigma}{\bar{x}} = \frac{19.77}{35.16} = 0.56$.

Page No.
Date: / /

Imp) Q) Initially there were 9 workers all being paid a uniform wage. Later a 10th worker added whose wage rate is 20 rupees less than from others. Compute :-

(i) the effect on mean wage.

(ii) Standard deviation of wage of each of the 9 workers be rupees 'C' and the wage of the 10th worker is given to be rupees 'C-20'.

Ans)

Sl. no	x (wage)	$x-\bar{x}$	$(x-\bar{x})^2$
1	C	2	4
2	C	2	4
3	C	2	4
4	C	2	4
5	C	2	4
6	C	2	4
7	C	2	4
8	C	2	4
9	C	2	4
10	C-20	-18	324
			360

(i) wage of 9 workers $= \frac{9C}{9} = C$

wage of 10 workers $= \frac{9C + C - 20}{10}$

$= C - 2$

effect on mean wage ; $C - (C-2)$

$= 2$

(ii) S.D. ⇒ $\sigma = \sqrt{\frac{1}{n}\Sigma(x-\bar{x})^2}$

$= \sqrt{\frac{360}{10}} = 6$.

Page No.
Date: / /

Q) The mean of two samples of size 50 and 100 resp. are 54.1 and 50.3 and the standard deviations are 8 and 7. Obtain the standard deviation of the sample size 150 by combining the two samples.

Ans→ Here $n_1 = 50$

$n_2 = 100$

$\bar{x}_1 = 54.1$, $\bar{x}_2 = 50.3$

$$\boxed{\sigma^2 = \frac{1}{n_1+n_2}\left[n_1(\sigma_1^2+d_1^2) + n_2(\sigma_2^2+d_2^2)\right]}$$

$d_1 = \bar{x}_1 - \bar{x} = 54.1 - 51.56 = 2.54$

$d_2 = \bar{x}_2 - \bar{x} = 50.3 - 51.56 = -1.26$

$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1+n_2} = \frac{50\times 54.1 + 100\times 50.3}{150} = 51.56$

$\sigma^2 = \frac{1}{50+100}\left[50(64+(2.54)^2) + 100(49+(-1.26)^2)\right]$

$\sigma^2 = \frac{1}{150}\left[50\times(70.45) + 100\times(50.58)\right]$

$\sigma^2 = \frac{1}{150}(3522.5 + 5058)$

$\sigma^2 = \frac{1}{150}\times 8580.5$

$\sigma^2 = 57.20$

$\boxed{\sigma = 7.56}$

Page No.
Date: / /

Q) For a set of 10 observation mean is 5, standard deviation is 2 and coefficient of variation is 60%, from the following this statement is false (or) true.

Ans→ the statement is false, bcz,

$\bar{x} = 5$, $\sigma = 2$, $C.V = \frac{2}{5}\times 100 = 40\%$. But given is 60% so it is false.

Q) If the observation is 10, mean is 12, and submission x^2 is 1530. then find coefficient of variation.

Ans→ $\bar{x} = 12$, $\Sigma x^2 = 1530$, $n = 10$

$\bar{x}^2 = 12^2 = 144$

$\sum_{i=1} \bar{x}^2 = 144\times 10 = 1440$

$\sigma^2 = \frac{\sum_{i=1}^{10} x^2 - \sum_{i=1} \bar{x}^2}{n} = \frac{1530-1440}{10} = \frac{90}{10} = 9$

$\boxed{\sigma = 3}$

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

$= \frac{3}{12}\times 100 = 25\%$

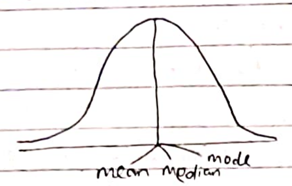
Skewness (SK)

Page No.

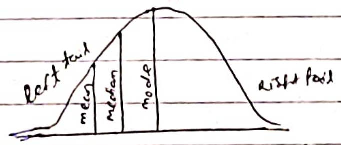
Date: / /

→ when mean = median = mode the distribution is called symmetrical.

→ Symmetric distribution when plotted on graph will give bell shape curve known as normal curve.

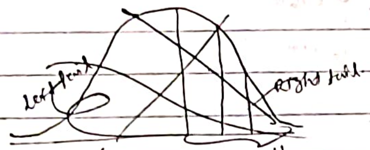


→ A distribution which is not symmetrical is said to be skew distribution.

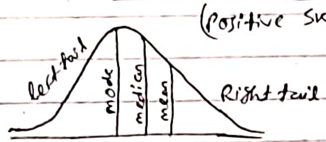


$\Rightarrow M < Md < Mo$

(Negative skewed)



(Positive skewed)



$\Rightarrow Mo < Md < M.$

(positive skewed)

Negative skewed distribution :-

→ If the curve has a long tail towards the left it is said to be negative skewed distribution.

Positive skewed distribution :-

→ If the curve has a long tail towards the right it is said to be Positive skewed distributed.

Page No.

Date: / /

→ Skewness refers to the asymmetric (or) lack of symmetry in the shape of a frequency distribution.

→ The skewness present when :-

① Arithmatic mean ≠ mode ≠ median / $AM \neq M_o \neq M_d$.

② $Q_3 - M_d \neq M_d - Q_1$

③ Frequency of either side of mode is not equal.

Measure of Skewness (SK) :-

→ measure of skewness are absolute skeweness, mean − mode.

$\boxed{SK = \text{mean} - \text{mode}}$

$SK > 0 \Rightarrow$ +ve skewed

$SK < 0 \Rightarrow$ −ve skewed.

Q) Find the standard deviation of 240, 260, 263, 251, 290, 245, 255, 288, 277 and 272. Find the coefficient of standard deviation also.

Ans)

x	$x-\bar{x}$	$(x_i-\bar{x})^2$
240	−24.1	580.81
260	−4.1	16.81
263	−1.1	1.21
251	−13.1	171.61
290	25.9	70.81
245	−19.1	364.81
255	−9.1	82.81
288	23.9	571.21
277	12.9	166.41
272	7.9	62.41
		2688.9

$\bar{x} = \frac{2641}{10} = 264.1$

$\sigma = \sqrt{\frac{\sum(x_i-\bar{x})^2}{n}} = \sqrt{\frac{2688.9}{10}} = \sqrt{268.89}$

$= 16.39$

$C.o.SD = \frac{\sigma}{\bar{x}} = \frac{16.39}{264.1}$

$= 0.0620.$

Relative measure (or) Coefficient of skewness -

-> There are many types of relative measure of skewness (or) coefficient of skewness.

-> One of the most used method is Karl Person's coefficient of skewness (or) Persian coefficient of skewness.

$$\boxed{SK_P = \frac{\text{mean} - \text{mode}}{\text{S.D.}}}$$

$$= \frac{\text{mean} - (3\,\text{median} - 2\,\text{mean})}{\text{S.D.}}$$

$$= \frac{3(\text{mean} - \text{median})}{\text{S.D.}}$$

$$= \boxed{\frac{3(\text{mean} - \text{median})}{\sigma}}$$

$$= \boxed{\frac{3(\bar{x} - M_d)}{\sigma}}$$

Q) Find the Persian coefficient of given data: 25, 15, 23, 40, 27, 25, 23, 25, 20.

Ans) mean $(\bar{x}) = \frac{223}{9} = 24.77$.

A.S.O = 15, 20, 23, 23, 25, 25, 25, 27, 40

median $= \left(\frac{9+1}{2}\right) = 5^{th}$ term $= 25$

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
15	-9.77	95.45
20	-4.77	22.75
23	-1.77	3.13
23	-1.77	3.13
25	0.23	0.052
25	0.23	0.052
25	0.23	0.052
27	2.23	4.97
40	15.23	231.95
		361.536

$$\sigma = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n}} = \sqrt{\frac{361.536}{9}} = \sqrt{40.17}$$
$$= 6.33$$

Q)

C.I	f
10-15	8
15-20	16
20-25	30
25-30	45
30-35	62
35-40	32
40-45	15
45-50	5

Find mean, mode, S.D. co of skewness.

Ans)

x	f	x_i	$d_i = \frac{x-A}{h}$	$f_i d_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
10-15	8	12.5	-4	-32	-17.47	305.20
15-20	16	17.5	-3	-48	-12.47	155.50
20-25	30	22.5	-2	-60	-7.47	55.80
25-30	45	27.5	-1	-45	-2.47	6.10
30-35	$\boxed{62}$	32.5	0	0	2.53	6.40
35-40	32	37.5	1	32	7.53	56.70
40-45	15	42.5	2	30	12.53	157.00
45-50	5	47.5	3	15	17.53	307.30
	213			-108		

A = 32.5

h = 5

N = 213

$\bar{x} = A + \frac{h}{N}\sum f_i d_i$

$= 32.5 + \frac{5}{213} \times (-108)$

$= 32.5 + (-2.53) = 29.97$

$f_i d_i^2$
128
144
130
45
0
35
60
45
574

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

$\sigma^2 = 25\left(\frac{574}{213} - \left(\frac{-108}{213}\right)^2\right)$

$\sigma^2 = 80.04$

$\sigma = \sqrt{80.04} = 8.94$

$= 7.8$

Page No.
Date: / /

For mode:

$l = 30$

$f_m = 62$

$f_1 = 45$

$f_2 = 32$

$h = 5$

$M_0 = l + \frac{f_m - f_1}{2f_m - f_1 - f_2} \times h$

$= 30 + \frac{62 - 45}{2\times 62 - 45 - 32} \times 5$

$= 31.8$

$SK = \frac{\text{mean} - \text{mode}}{S.D} = \frac{29.97 - 31.8}{7.8} = -0.235$

Moments :-

① Moments about mean / central momentum :- (μ_r).

$\mu_r = \frac{1}{N} \sum_{i=1}^{n} f_i (x_i - \bar{x})^r$

when $r=0$; $\mu_0 = \frac{1}{N} \sum f_i 1$

$= \frac{N}{N} = 1$

when $r=1$;

$\mu_1 = \frac{1}{N} \sum f_i (x_i - \bar{x})$

$= \frac{1}{N} [\sum f_i x_i - \bar{x} \sum f_i]$

$= \frac{\sum f_i x_i}{\sum f_i} - \bar{x} \frac{\sum f_i}{N}$

$= \bar{x} - \bar{x}\, \cancel{N}/\cancel{N} = \bar{x} - \bar{x} = 0$

when $r=2$,

$\mu_2 = \frac{1}{N} \sum f_i (x_i - \bar{x})^2$

$= \frac{h^2}{N} \cdot \sum f_i d^2$, $d = \frac{x - \bar{x}}{h}$

when $r=3$,

$\mu_3 = \frac{1}{N} \sum f_i (x_i - \bar{x})^3$

$= \frac{h^3}{N} \sum f_i d^3$.

when $r=4$,

$\mu_4 = \frac{1}{N} \sum f_i (x_i - \bar{x})^4$

$= \frac{h^4}{N} \sum f_i d^4$.

Page No.
Date: / /

② Moments about Arbitary number / Raw momentum (μ_r') :-

$\mu_r' = \frac{1}{N} \sum f_i (x_i - A)^r$

when $r=0$,

$\mu_0' = \frac{1}{N} \sum f_i \cdot 1 = 1$

when $r=1$,

$\mu_1' = \frac{1}{N} \sum f_i (x_i - A)$

$= \frac{1}{N} \sum f_i x_i - \frac{A \sum f_i}{N}$

$= \bar{x} - A$.

when $r=2$,

$\mu_2' = \frac{1}{N} \sum f_i (x_i - A)^2$

$= \frac{h^2}{N} \sum f_i d_i^2$, $d = \frac{x_i - A}{h}$.

when $r=3$,

$\mu_3' = \frac{1}{N} \sum f_i (x_i - A)^3$

$= \frac{h^3}{N} \sum f_i d_i^3$, $d = \frac{x_i - A}{h}$

when $r=4$,

$\mu_4' = \frac{1}{N} \sum f_i (x_i - A)^4$

$= \frac{h^4}{N} \sum f_i d_i^4$.

③ Moments about Origin (V_r) :-

$V_r = \frac{1}{N} \sum f_i (x_i - 0)^r$

when $r=0$,

$V_0 = \frac{1}{N} \sum f_i 1 = 1$

when $r=1$,

$V_1 = \frac{1}{N} \sum f_i (x_i - 0)$

$= \frac{1}{N} \sum f_i x_i - \frac{\sum f_i}{N} 0$

$= \bar{x}$.

when $r=2$,

$V_2 = \frac{1}{N} \sum f_i (x_i - 0)^2$

$= \frac{1}{N} \sum f_i x_i^2$.

when $r=3$,

$V_3 = \frac{1}{N} \sum f_i (x_i - 0)^3$

$= \frac{1}{N} \sum f_i x_i^3$.

Page No.
Date: / /

Relation between central momentum and Raw momentum

① $\mu_1 = 0$

② $\mu_2 = \mu_2' - (\mu_1')^2$

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

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

⑤ $\beta_1 = \frac{\mu_3^2}{\mu_2^3}$ (coefficient of skewness).

⑥ $\beta_2 = \frac{\mu_4}{\mu_2^2}$ (coefficient of kurtosis).

⑦ i) $\gamma_1 = \pm\sqrt{\beta_1}$ (Gamma coefficient)

(ii) $\gamma_2 = \beta_2 - 3$.

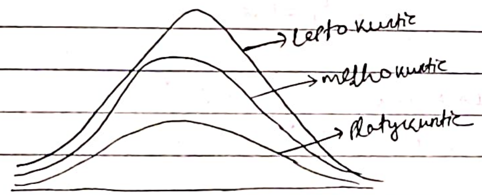
Kurtosis:-

→ Kurtosis are 3 types :-

(i) Leptokurtic : when $\beta_2 > 3$ and it is peaked data distribution.

(ii) Mesokurtic : when $\beta_2 = 3$ and it's normal data distribution.

(iii) Platykurtic : when $\beta_2 < 3$ and the data set is scattered (or) flat.



Q) Calculate the first 4 Raw momentum and also find β_1 and β_2.

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

Page No.
Date: / /

Ans)

x	f	$d = x - A$	fd	fd^2	fd^3	fd^4
0	1	−4	−4	16	−64	256
1	8	−3	−24	72	−216	648
2	28	−2	−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
	256		0	512	0	2816

here $A = 4$

$\Sigma f = N = 256$

$\mu_1' = \frac{1}{N}\Sigma f_i(x_i - A)$

$= \frac{1}{N}\Sigma f_i d = \frac{1}{256} \times 0 = 0$

$\mu_2' = \frac{1}{N}\Sigma f_i(x_i - A)^2$

$= \frac{1}{N}\Sigma f_i d^2 = \frac{512}{256} = 2$

$\mu_3' = \frac{1}{N}\Sigma f d^3 = 0$

$\mu_4' = \frac{1}{N}\Sigma f_i d^4 = \frac{2816}{256} = 11$

$\mu_1 = 0$

$\mu_2 = \mu_2' - (\mu_1')^2 = 2 - 0^2 = 2$

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

$= 0 - 3.2.0 + 2 = 0$

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

$= 11 - 4.0.0 + 6.2.0 - 3.0 = 11$

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

$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{11}{4} = 2.75$ (Platykurtic)

Page No.
Date: / /

Q) Find 4 Raw Moments, Coefficient of Kurtosis, coefficient of skewness.

C.I.	f
0-10	15
10-20	17
20-30	23
30-40	22
40-50	25
50-60	10
60-70	5
70-80	10

Ans.

C.I.	f	x_i	$d = \frac{x_i - A}{h}$	fd	fd^2	fd^3	fd^4
0-10	15	5	-3	-45	135	-405	1215
10-20	17	15	-2	-34	68	-136	272
20-30	23	25	-1	-23	23	-23	23
30-40	22	35	0	0	0	0	0
40-50	25	45	1	25	25	25	25
50-60	10	55	2	20	40	80	160
60-70	5	65	3	15	45	135	405
70-80	10	75	4	40	160	640	2560
	127			-2	496	316	4660

Here $A = 35$

$h = 10$

$\Sigma f = N = 127$

$\mu_1' = \frac{1}{N}\Sigma f_i(x_i - A) = \frac{1}{N}\Sigma f_i d = \frac{1}{127}\times(-2) = -0.0157$

$\mu_2' = \frac{1}{N}\Sigma f_i d^2 = \frac{496}{127} = 3.90$

$\mu_3' = \frac{1}{N}\Sigma f_i d^3 = \frac{316}{127} = 2.48$

$\mu_4' = \frac{1}{N}\Sigma f_i d^4 = \frac{4660}{127} = 36.69$

$\mu_1 = 0$

$\mu_2 = \mu_2' - (\mu_1')^2 = 3.90 - (0)^2 = 3.90$

$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2(\mu_1')^3 = 2.48 - 3\cdot(3.90)\cdot 0 + 2(0)^3 = 2.48$

$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'(\mu_1')^2 - 3(\mu_1')^4 = 36.69 - 4\cdot(2.48)\cdot 0 + 6\cdot(3.90)(0)^2$
$- 3\cdot(0)^4 = 36.69$

Coefficient of kurtosis $(\beta_2) = \frac{\mu_4}{\mu_2^2} = \frac{36.69}{(3.90)^2} = 2.41$

Coefficient of skewness $(\beta_1) = \frac{\mu_3^2}{\mu_2^3} = \frac{(2.48)^2}{(3.90)^3} = 0.10$.

Page No.
Date: / /

Correlation:-

→ Correlation analysis deals with the association between two (or) more variables.

→ Correlation analysis attempt to determine the degree ~~of~~ relationship betn variables.

Types of Correlation:-

(i) Positive Correlation :- If two variables X and Y moves in same direction i.e. if one variable increase and the other variable increase too (or) viceversa, then it is called +ve correlation.

(ii) Negative Correlation :- If two variables X and Y moves in opposite direction i.e. if one increase then other one fall down (or) vica versa, then it's called -ve correlation.

(iii) Perfect Correlation :- when both the variables X and Y increase and decreased simultaneously then it's called perfect correlation.

Remark

→ Correlation always varries between -1 to +1.

→ High degree Correlation : It happens when the degree lies between 0.75 to 1 (or) (-0.75) to (-1).

→ Moderate degree correlation : It happens when the degree lies between 0.25 to 0.75 (or) (-0.25) to (-0.75).

→ Low degree correlation :- It happens when the degree lies between 0 to 0.25 (or) 0 to (-0.25).

Pearson's method for Correlation :-

→ Pearson's method for Correlation is known as Persian correlation coefficient between two variables X and Y, which is denoted as $r(X,Y)$, is a numerical measure of linear relationship between the variables.

→ It is defined as the ratio of covarience between X and Y and the product of standard deviation of X and Y.

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

where $Cov(x,y) = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{n}$

$\sigma_x = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}}$ $\sigma_y = \sqrt{\frac{\Sigma(y-\bar{y})^2}{n}}$

Q) Find the degree of correlation between X and Y.

X	Y	$(x-\bar{x})$	$(y-\bar{y})$	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$
2	4	-3	-6	+18	9	36
3	7	-2	-3	+6	4	9
4	8	-1	-2	+2	1	4
5	9	0	-1	0	0	1
6	10	1	0	0	1	0
7	14	2	4	8	4	16
8	18	3	8	24	9	64
		0	0	58	28	130

$\bar{X} = \frac{35}{7} = 5$

$\bar{Y} = \frac{70}{7} = 10$

$r(x,y) = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2}\sqrt{\Sigma(y-\bar{y})^2}} = \frac{58}{\sqrt{28}\sqrt{130}} = \frac{58}{(5.29)(11.40)}$

$\therefore$ It is High degree correlation. $= \frac{58}{60.30} = 0.96$.

Q)

X	Y	$(x-\bar{x})$	$(y-\bar{y})$	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$
39	37	-24	-28	-888	576	784
65	53	2	-12	-24	4	144
62	58	-1	-7	7	1	49
90	86	27	21	567	729	441
62	62	-1	-3	3	1	9
75	68	12	3	36	144	9
25	60	-38	-5	190	1444	25
98	91	35	26	910	1225	676
36	51	-27	-14	378	729	196
78	84	15	19	285	225	361
		0	0	1464	5078	2694

$\bar{x} = \frac{630}{10} = 63$

$\bar{y} = \frac{650}{10} = 65$

$r(x,y) = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2}\sqrt{\Sigma(y-\bar{y})^2}} = \frac{1464}{\sqrt{5078}\sqrt{2694}} = \frac{1464}{(71.26)(51.90)}$

$= 0.39$

$\therefore$ It is moderate degree correlation.

Spearman's Rank method:-

→ To find the relation between two variables x and y the Spearman's Rank method for non-repeated values,

$\boxed{R = 1 - \frac{6\Sigma D^2}{N^3-N}}$, where D is the difference between R_1 and R_2. $\boxed{D = R_1 - R_2}$

Spearman's UnRanked method:-

$$\boxed{R = 1 - \frac{6\left[\Sigma D^2 + \frac{\Sigma m_i(m_i^2-1)}{12}\right]}{N^3-N}}$$

Q) In a fancy dress Competition two Judges accorded with the following rank to 8 Participant. Calculate the correlation of rank given by the Judges.

Sl. No.	R_1	R_2	$D = R_1 - R_2$	D^2
1	8	7	1	1
2	7	5	2	4
3	6	4	2	4
4	3	1	2	4
5	2	3	-1	1
6	1	2	-1	1
7	5	6	-1	1
8	4	8	-4	16
				32

$R = \left(1 - \frac{6 \times 32}{512-8}\right) = \left(1 - \frac{192}{504}\right) = 1 - 0.380 = 0.62/0.619$

$\therefore$ It is moderate degree correlation.

Sol Q)

S.No	Mark given by Judge1 (J1)	Mark given by Judge2 (J2)	R_1	R_2	D	D^2
1	50	12	3	5.5	-2.5	6.25
2	33	12	5	5.5	-0.5	0.25
3	40	24	4	1	3	9
4	10	6	10	8.5	1.5	2.25
5	15	15	8	4	4	16
6	15	4	8	10	-2	4
7	65	20	1	2	-1	1
8	24	9	6	7	-1	1
9	15	6	8	8.5	-0.5	0.25
10	57	18	2	3	-1	1
						41

$7^{th}+8^{th}+9^{th}/3 = 8,\ m_1=3$

$5^{th}+6^{th}/2 = 5.5,\ m_2=2$

$8^{th}+9^{th}/2 = 8.5,\ m_3=2$

$$R = 1-\frac{6\times 41}{1000-10} = 1-\frac{246}{990} = 0.751$$

$$R = 1-\frac{6\left[41+\frac{3(9-1)+2(4-1)+2(4-1)}{12}\right]}{10^3-10} = 0.733$$

Q)

No.	x	Y	R_1	R_2	D	D^2
1	68	62	4	5	-1	1
2	64	58	6	7	1	1
3	75	68	2.5	3.5	-1	1
4	50	45	9	10	-1	1
5	64	81	6	1	5	25
6	80	60	1	6	-5	25
7	75	68	2.5	3.5	-1	1
8	40	48	10	9	1	1
9	55	50	8	8	0	0
10	64	70	6	2	4	16
						72

$2^{th}+3^{th}/2 = 5/2 = 2.5 = m_1=2$

$2^{th}+5^{th}+6^{th}/3 = \cancel{\ }$ $9 = m_2=3$

3^{th} and $= 7/2 = 3.5 = m_3=2$

$$R = 1-\frac{6\left[72+\frac{2(4-1)+3(9-1)+2(4-1)}{12}\right]}{1000-10}$$

$$= 1-\frac{6[72+3]}{990}$$

$$= 0.545$$

Regression Analysis:-

→ Regression Analysis is a mathematical measure of the average relationship between two (or) more variables in terms of original units of data.

Simple Regression:-

→ The regression analysis confined to study of only two variable at a time is called simple Regression.

Multiple Regression:-

→ The regression analysis for studying more than two variable at a time is called Multiple Regression.

Line of Regression:-

→ A Line of regression is the line which keeps the best estimate of one variable X for any given variable or other variable Y.

I. Line of regression of Y on X.

$$Y-\bar{Y} = b_{yx}(X-\bar{X})$$

$$\because b_{yx} = \frac{cov(X,Y)}{\sigma_x^2} = \frac{\Sigma(X-\bar{X})(Y-\bar{Y})}{\Sigma(X-\bar{X})^2}$$

II. Line of regression of X on Y

$$X-\bar{X} = b_{xy}(Y-\bar{Y})$$

$$\because b_{xy} = \frac{cov(X,Y)}{\sigma_y^2} = \frac{\Sigma(X-\bar{X})(Y-\bar{Y})}{\Sigma(Y-\bar{Y})^2}.$$

List square method:-

→ The Principle of List square method which consists of minimizing the sum of the square (or) Reschedule on the error of estimate i.e. the deviation between the given observation value of the variables and their corresponding estimate value as given by the line of the best fit.

Page No.
Date: / /

Normal Equation

→ The system of equations required to be solved for obtaining the values of constant in the estimating equation is known as normal equation.

Line of the best fit

I. y on x

$y = ax + b$

Normal Equations

$\Sigma y = a\Sigma x + nb$

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

II. x on y

$x = ay + b$

Normal Equations

$\Sigma x = a\Sigma y + nb$

$\Sigma xy = a\Sigma y^2 + b\Sigma y$

Q) Calculate the line of regression by using LSM method.

x	y	xy	x^2	y^2
1	2	2	1	4
2	5	10	4	25
3	3	9	9	9
4	8	32	16	64
5	7	35	25	49
15	25	88	55	151

I. $y = ax + b$

$\Sigma y = a\Sigma x + nb \Rightarrow 25 = 15a + 5b$

$\Sigma xy = a\Sigma x^2 + b\Sigma x \Rightarrow 88 = 55a + 15b$

$a = 1.3,\ b = 1.1$

$\Rightarrow y = 1.3x + 1.1.$

II. $x = ay + b$

$\Sigma x = a\Sigma y + nb \Rightarrow 15 = 25a + 5b$

$\Sigma xy = a\Sigma y^2 + b\Sigma y \Rightarrow 88 = 151a + 25b$

$\Rightarrow a = 0.5,\ b = 0.5$

$\Rightarrow x = 0.5y + 0.5.$

Page No.
Date: / /

Q)

x	y	xy	x^2	y^2
0	2	0	0	4
1	1	1	1	1
2	3	6	4	9
3	2	6	9	4
4	4	16	16	16
5	3	15	25	9
6	5	30	36	25
21	20	74	91	68

I. $y = ax + b$

$\Sigma y = a\Sigma x + nb \Rightarrow 20 = 21a + 7b$

$\Sigma xy = a\Sigma x^2 + b\Sigma x \Rightarrow 74 = 91a + 21b$

$a = 0.5,\ b = 1.35$

$\Rightarrow y = 0.5x + 1.35$

II. $x = ay + b$

$\Sigma x = a\Sigma y + nb \Rightarrow 21 = 20a + 7b$

$\Sigma xy = a\Sigma y^2 + b\Sigma y \Rightarrow 74 = 68a + 20b$

$a = 1.28,\ b = -0.68$

$x = 1.28y - 0.68.$

Q)

x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
91	71	1	1	1	1	1
97	75	7	5	49	25	35
108	69	18	-1	324	1	-18
121	97	31	27	961	729	837
67	70	-23	0	529	0	0
124	91	34	21	1156	441	714
51	39	-39	-31	1521	961	1209
73	61	-17	-9	289	81	153
111	80	21	10	441	100	210
57	47	-33	-23	1089	529	759
900		0	0	6360	2868	3900

$\bar{x} = \frac{900}{10} = 90$

$\bar{y} = \frac{700}{10} = 70$

Page No.
Date: / /

$b_{yx} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2} = \frac{3900}{6360} = 0.61$

$b_{xy} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(y-\bar{y})^2} = \frac{3900}{2868} = 1.359$.

(i) $b_{xy} = \frac{3900}{2868} = 1.359$

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

$x - 90 = 1.359 \times (y-70)$

$x = 1.35y - 94.5 + 90$

$x = 1.35y - 4.5$.

(ii) $b_{yx} = \frac{3900}{6360} = 0.61$

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

$y - 70 = 0.61(x-90)$

$y = 0.61x - 54.9 + 70$

$y = 0.61x + 15.1$

Unbiased Estimate means $E(t) = \theta$

Q) t is a unbiased estimation defined by $t = \frac{2x_1 + x_2 + \lambda x_3}{5}$.

Find the value of λ where the parameter is mean.

Ans) $E(t) = \theta$, $\theta = \mu$

$E(t) = \mu$

$E\left[\frac{2x_1 + x_2 + \lambda x_3}{5}\right] = \mu$

$\Rightarrow \frac{1}{5}E(2x_1 + x_2 + \lambda x_3) = \mu$

$\Rightarrow 2E(x_1) + E(x_2) + \lambda E(x_3) = 5\mu$

$\Rightarrow 2\mu + \mu + \lambda\mu = 5\mu$

$\Rightarrow 3\mu + \lambda\mu = 5\mu$

$\Rightarrow \lambda\mu = 2\mu$

$\Rightarrow \boxed{\lambda = 2}$

Page No.
Date: / /

Q) Let $x_1, x_2, \ldots, x_n$ be the random samples from a normal population $N(\mu, 1)$. So that if $t = \frac{\Sigma x_i^2}{n}$ is an unbiased data then the estimate is $\mu^2 + 1$.

Page No.
Date: / /

Consistency Point Estimate :-

→ A statistics of $x_1, x_2, \ldots, x_n$ is consistent estimate of parameter θ if,

$$\lim_{n\to\infty} E(t) = \infty$$
and $\lim_{n\to\infty} var(t) = 0$.

Q) Show that normal distribution of the sample mean is consistent.

Sol) $\bar{x} = \frac{\Sigma x_i}{n}$, $N(\mu, \sigma^2)$

$\lim_{n\to\infty} E(t) = \mu$, $\lim_{n\to\infty} var(t) = 0$

$LHS = \lim_{n\to\infty} E(t) = \lim_{n\to\infty} E(\bar{x})$

$= \lim_{n\to\infty} E\left(\frac{\Sigma x_i}{n}\right) = \lim_{n\to\infty} \frac{1}{n} E(\Sigma x_i)$

$= \lim_{n\to\infty} \frac{1}{n} E(x_1, x_2, \ldots, x_n) = \lim_{n\to\infty} \frac{1}{n} (\mu_1, \mu_2, \ldots \mu_n)$

$= \lim_{n\to\infty} \frac{1}{\cancel{n}} \times \cancel{n}\mu = \mu$.

$\lim_{n\to\infty} var(t)$

$= \lim_{n\to\infty} \left(\frac{\sigma^2}{n}\right)$ $\quad \boxed{\sigma^2 = \Sigma(x_i - \bar{x})^2}$

$= 0$

$\therefore$ $N(\mu, \sigma^2)$ is consistent.

Efficiency Point Estimation :-

→ If 't' is the most efficient estimate of the parameter θ with varience s^2 and t_1 is any other estimate with varience s_1^2 then the efficient of t_1 is defined by $var(t) < var(t_1)$

(or) $\frac{s^2}{s_1^2} < 1$.

* $var(ax) = a^2 var(x)$.

Q) x_1, x_2, x_3 are the random samples. t_1 is defined as

$t_1 = \frac{x_1 + 2x_2 + x_3}{4}$

and $t_2 = \frac{x_1 + x_2 + x_3}{3}$, so that t_2 is more efficient than t_1.

Sol) $f(x_1, x_2, x_3)$

$t_1 = \frac{x_1 + 2x_2 + x_3}{4}$

$var(t_1) = var\left(\frac{x_1 + 2x_2 + x_3}{4}\right)$

$= \frac{1}{16} var(x_1) + 4 var(x_2) + var(x_3)$

$= \frac{1}{16} (\sigma^2 + 4\sigma^2 + \sigma^2)$

$= \frac{6}{16} \sigma^2$.

Page No.
Date: / /

$var(t_2) = \left(\frac{x_1 + x_2 + x_3}{3}\right)$

$= \frac{1}{9} var(x_1) + var(x_2) + var(x_3)$

$= \frac{1}{9} (\sigma^2 + \sigma^2 + \sigma^2) = \frac{1}{\cancel{9}} \times \cancel{3}\sigma^2 = \frac{1}{3}\sigma^2$.

$\frac{var(t_2)}{var(t_1)} = \frac{\frac{1}{3}\sigma^2}{\frac{6}{16}\sigma^2} = \frac{8}{9} < 1$.

$\therefore$ $var(t_2) < var(t_1)$.

$\therefore$ varience of (t_1) is more, so (t_2) is more efficient.

Sufficiency Point Estimation :-

→ Let $t(x_1, x_2, \ldots, x_n)$ is sufficient for θ if and only if the joint distribution for likelyhood function 'l' of the sample can be expressed as,

$$\boxed{l = g(t, \theta) \cdot h(x)}$$

Q) A random sample from a population with probability function $f(x, \theta) = \theta x^{\theta - 1}$, $0 < x < 1$ is sufficient.

Sol) $f(x, \theta) = \theta x^{\theta - 1}$

$f(x_1, \theta) = \theta x_1^{\theta - 1}$

$f(x_2, \theta) = \theta x_2^{\theta - 1}$

$f(x_n, \theta) = \theta x_n^{\theta - 1}$

$l = \prod_{i=1}^{n} f(x_i, \theta) = \theta^n x_1^{\theta-1} \cdot x_2^{\theta-2} \cdot x_3^{\theta-3} \ldots x_n^{\theta-1}$

$l = \theta^n (x_1 \cdot x_2 \cdot x_3 \ldots x_n)^{\theta - 1}$

$= \theta^n (\prod_{i=1}^{n} x_i)^{\theta - 1}$

$= \theta^n (\prod_{i=1}^{n} x_i)^{\theta} \cdot (\prod_{i=1}^{n} x_i)^{-1}$ $\quad \because t = t(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} x_i$

$= \theta^n t^\theta \cdot \frac{1}{\prod_{i=1}^{n} x_i}$

$\therefore$ Here $g(t, \theta) = \theta^n t^\theta$ and $h(x) = \frac{1}{\prod_{i=1}^{n} x_i}$.

Page No.
Date: / /

Q) If $x_1, x_2, \dots x_n$ are the random samples, the Bernoulli distribution function. $f(x,p) = p^x(1-p)^{1-x}$. so that p is sufficient.

Soln) $f(x,p) = p^x(1-p)^{1-x}$

$f(x_1,p) = p^{x_1}(1-p)^{1-x_1}$

$f(x_2,p) = p^{x_2}(1-p)^{1-x_2}$

$\vdots$

$f(x_n,p) = p^{x_n}(1-p)^{1-x_n}$

$l = \prod_{i=1}^{n} f(x_i,\theta) = p^{x_1+x_2+\dots+x_n}(1-p)^{1-x_1+1-x_2+1-x_3+\dots+1-x_n}$

$= p^{\sum x_i}(1-p)^{n-\sum x_i}$

$l = p^{\sum x_i}(1-p)^{n-\sum x_i}$ $t = \sum x_i$

$l = p^t(1-p)^{n-t} \cdot 1$

$= p^t(1-p)^{n-t} \cdot x^0$

$\therefore$ Here $g(t,p) = p^t(1-p)^{n-t}$, $h(x) = 1$.

Q) The Probability funn $f(x,\theta) = \frac{1}{\theta}e^{-\frac{x}{\theta}}$. (exponent distribution) so that exponent distribution is sufficient.

Page No.
Date: / /

MAXIMUM Likelyhood Estimate :- (MLE)

→ MLE method is a technique used for estimate two parameter of a given distribution using some observed data.

→ Let $x_1, x_2, \dots, x_n$ are random samples the maximum likelyhood function $L = \prod_{i=1}^{n} f(x_i,\theta)$ using the maximum (or) minimum principle calculate the MLE of θ.

→ Since $\log L$ is non decreasing and attend the extreme value so the equation $\frac{\partial}{\partial\theta}(\log L) = 0$. and $\frac{\partial^2}{\partial\theta^2}(\log L) < 0$. which is known as likelyhood equation for estimating θ.

Q) Let $x_1, x_2, \dots, x_n$ are the random samples from the normal distribution population, find MLE of μ and σ^2.

Soln) $N(\mu, \sigma^2)$

$f(x,\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

For $\theta = \mu$

$l = \prod_{i=1}^{n} f(x_i;\theta)$

$f(x_1,\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x_1-\mu}{\sigma}\right)^2}$

$\vdots$

$f(x_n,\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x_n-\mu}{\sigma}\right)^2}$

$l = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2}\left\{\left(\frac{x_1-\mu}{\sigma}\right)^2 + \dots + \left(\frac{x_n-\mu}{\sigma}\right)^2\right\}}$

$l = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2}$

$\log l = -n\log\sqrt{2\pi\sigma^2} - \frac{1}{2}\sum\left(\frac{x_i-\mu}{\sigma}\right)^2$

$\frac{d}{d\mu}\log l = 0 + \frac{1}{2}\sum 2\left(\frac{x_i-\mu}{\sigma}\right)\left(\frac{+1}{\sigma}\right)$

$= \sum\left(\frac{x_i-\mu}{\sigma^2}\right)$.

$\frac{d}{d\theta}\log L = 0$

$\Rightarrow \frac{d}{d\mu}\log l = 0$

$\Rightarrow \sum \frac{x_i-\mu}{\sigma^2} = 0$

$\Rightarrow \sum_{i=1}^{n} x_i - \sum_{i=1}^{n}\mu = 0$

$\Rightarrow \sum x_i - n\mu = 0$

$\Rightarrow \boxed{\mu = \frac{\sum x_i}{n}}$

Page No.
Date: / /

$$\frac{\partial^2}{\partial \mu^2}(\log l) = \frac{\partial}{\partial \mu}\left(\frac{d}{\partial \mu}\log l\right)$$
$$= \frac{\partial}{\partial \mu}\left[\frac{\Sigma(x_i-\mu)}{\sigma^2}\right]$$
$$= \Sigma\left(\frac{-1}{\sigma^2}\right) = \frac{-n}{\sigma^2} < 0$$

$\therefore$ for $\mu = \frac{\Sigma x_i}{n}$, MLE of μ.

Q) Find the MLE for σ^2.

Ans for $\theta = \sigma^2$

$$l = \prod_{i=1}^{n} f(x_i;\theta)$$
$$f(x_1,\theta) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{1}{2}\left(\frac{x_1-\mu}{\sigma}\right)^2}$$
$$f(x_n,\theta) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{1}{2}\left(\frac{x_n-\mu}{\sigma}\right)^2}$$
$$l = \left(\frac{1}{\sqrt{2\pi\sigma}}\right)^n e^{-\frac{1}{2}\left\{\left(\frac{x_1-\mu}{\sigma}\right)^2 + \dots + \left(\frac{x_n-\mu}{\sigma}\right)^2\right\}}$$
$$l = \left(\frac{1}{\sqrt{2\pi\sigma}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2}$$
$$\cancel{\log l = -n\log\sqrt{2\pi\sigma} - \frac{1}{2}\Sigma\left(\frac{x_i-\mu}{\sigma}\right)^2}$$

for $\theta = \sigma^2$

Suppose $\sigma^2 = t$

$$l = \left(\frac{1}{\sqrt{2\pi t}}\right)^n e^{-\frac{1}{2}\Sigma\frac{(x_i-\mu)^2}{t}}$$

$$\log l = \frac{-n}{2}[\log 2\pi + \log t] - \frac{1}{2}\Sigma\frac{(x_i-\mu)^2}{t}.$$

$$\boxed{\frac{d}{dt}\log l = 0 \qquad \frac{d^2}{dt^2}\log l < 0}$$

$$\frac{d}{dt}\left(\frac{-n}{2}[\log 2\pi + \log t] - \frac{1}{2}\frac{\Sigma(x_i-\mu)^2}{t}\right)$$

$$\Rightarrow -\frac{n}{2}\cdot\frac{1}{t} + \frac{1}{2}\Sigma(x_i-\mu)^2\cdot\left(\frac{1}{t^2}\right) = 0$$
$$\Rightarrow \frac{-n}{2t} + \frac{1}{2}\frac{\Sigma(x_i-\mu)^2}{t^2} = 0$$
$$\Rightarrow \frac{1}{t}\frac{\Sigma(x_i-\mu)^2}{t} = \frac{n}{t}$$
$$\Rightarrow \frac{\Sigma(x_i-\mu)^2}{t} = n$$
$$\Rightarrow \boxed{t = \frac{\Sigma(x_i-\mu)^2}{n}}$$

$$\frac{d^2}{dt^2}\log l = \frac{d}{dt}\left(\frac{d}{dt}\log l\right)$$
$$= \frac{d}{dt}\left[\frac{-n}{2t} + \frac{1}{2}\frac{\Sigma(x_i-\mu)^2}{t^2}\right]$$
$$= \frac{-n}{2}\left(\frac{-1}{t^2}\right) + \frac{1}{2}\Sigma(x_i-\mu)^2\left(\frac{-2}{t^3}\right)$$
$$= \frac{n}{2t^2} - \frac{\Sigma(x_i-\mu)^2}{t^3} = \frac{n}{2t^2} - \frac{tn}{t^3}$$
$$= \frac{-n}{2t^2} < 0.$$

Page No.
Date: / /

$\therefore$ For $\sigma^2 = \frac{\Sigma(x_i-\mu)^2}{n}$, MLE of σ^2.

H.W) Let $x_1, x_2, \dots x_n$ are the random samples of exponent distribution. Find MLE of θ.

Sol^n) $f(x,\theta) = \frac{1}{\theta}e^{-\frac{x}{\theta}}$.

Interval Estimate:-

$E(t) = \theta$ or $[c_1, c_2]$ → confidence interval

$E(\bar{x}) = \mu$ $\quad c_1 \leq \theta \leq c_2$

$E(s^2) = \sigma^2$ $\quad c_1 = t - s = t - SE(t) \times t_\alpha$ (interval state, standard state, critical value)

$96\% = t_\alpha$ $\quad c_2 = t + s = t + SE(t) \times t_\alpha$.

Condifence interval of mean.

Stata (t) will be $t \sim \underbrace{N(0,1)}_{\text{standardized normal distribution}} = Z$.

$$Z = \frac{t - E(t)}{SE(t)} \sim N(0,1) \qquad E(t) = \theta$$

Step 1:- Compute $\bar{x}$.

Step 2:- Select confidence level and find $t(\alpha)$ from the table.

Step 3:- Compute standard error of $\bar{x}$.

Case 1:- when σ is unknown standard error of $\bar{x}$ will be

$$SE(\bar{x}) = \frac{S}{\sqrt{n-1}}$$

where n = the size of the sample

S is the standard deviation of sampling.

Case 2:- when σ is known

$$SE(\bar{x}) = \frac{\sigma}{\sqrt{n}} \text{ (for population)}$$

Step 4:- Compute confidence interval

Case 1:- $[c_1, c_2] = \left[\bar{x} - \frac{S}{\sqrt{n-1}} \times t_\alpha,\ \bar{x} + \frac{S}{\sqrt{n-1}} \times t_\alpha\right]$

Case 2:- $[c_1, c_2] = \left[\bar{x} - \frac{\sigma}{\sqrt{n}} \times t_\alpha,\ \bar{x} + \frac{\sigma}{\sqrt{n}} \times t_\alpha\right]$

Q) A random sample of size 100 has mean 15, the population varience being 25. Find the interval estimate of the population mean with a confidence level of 99%. and 95%.

Page No.
Date: / /

Table of t_α

Confidence level	t_α
90%	1.64
95%	1.96
98%	2.33
99%	2.58
100%	3

Ans → $\bar{x} = 15,\ n = 100,\ \sigma^2 = 25,\ \sigma = 5$

For $t_\alpha = 99\%$.

$SE(\bar{x}) = \frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{100}} = 0.5$.

$[C_1, C_2] = [\bar{x} - \frac{\sigma}{\sqrt{n}} \times t_\alpha,\ \bar{x} + \frac{\sigma}{\sqrt{n}} \times t_\alpha]$

$= [15 - 2.58 \times 0.5,\ 15 + 2.58 \times 0.5]$

$= [13.71,\ 16.29]$

For $t_\alpha = 95\%$.

$[C_1, C_2] = [\bar{x} - \frac{\sigma}{\sqrt{n}} \times t_\alpha,\ \bar{x} + \frac{\sigma}{\sqrt{n}} \times t_\alpha]$

$= (15 - 1.96 \times 0.5,\ 15 + 1.96 \times 0.5]$

$= [14.02,\ 15.98]$

Interval estimate of Proportion:-

P = Population proportion

p = sample proportion

$E(p) = P$.

Step 1:-

① Standard error of sample population is

$SE(P) = \sqrt{\frac{PQ}{n}}$, P = known, $Q = 1-P$

n = sample size.

$SE(P) = \sqrt{\frac{pq}{n}}$ = P = unknown, p = sample proportion, $q = 1-p$

S-2

① $SE(p) = \sqrt{\frac{PQ}{n}} \times \sqrt{\frac{N-n}{N-1}}$ P = known, $Q = 1-P$

N = size of population

n = size of sample.

② $SE(P) = \sqrt{\frac{pq}{n}} \times \sqrt{\frac{N-n}{N-1}}$, P = unknown

Page No.
Date: / /

S-3:-

$[p - SE \times t_\alpha,\ p + SE \times t_\alpha]$

Q) In a random selection of 64 of 600 round crossing in a town, the mean no. of automobile accident per year was found 4.2 and sample standard deviation was 0.8. Construct 95% confidence interval for mean no. of automobile accident per crossing the year

Sol → $\bar{x} = 4.2$ $\qquad SE(\bar{x}) = \frac{s}{\sqrt{n-1}} \times \sqrt{\frac{N-n}{N-1}}$

$s = 0.8$ $\qquad = \frac{0.8}{\sqrt{64-1}} \times \sqrt{\frac{600-64}{600-1}}$

$t_\alpha = 1.96$ $\qquad = 0.095$

$n = 64$

$N = 600$

$[\bar{x} - SE(\bar{x}) \times t_\alpha,\ \bar{x} + SE(\bar{x}) \times t_\alpha]$

$= [4.2 - 1.96 \times 0.095,\ 4.2 + 1.96 \times 0.095]$

$= [4.01,\ 4.38]$

Q) A random sample of 800 units from a large consignment shows that 200 units were vanished. Find 95% confidence limit for the population proportion of damage units in the consignment.

Sol → $n = 800$

$p = \frac{200}{800} = 0.25$

$q = 1 - p$

$= 1 - \frac{1}{4} = 0.75$

$SE(P) = \sqrt{\frac{pq}{n}}$

$= \sqrt{\frac{0.25 \times 0.75}{800}}$

$= 0.0153$

$[p - SE \times t_\alpha,\ p + SE \times t_\alpha]$

$= [0.25 - 0.0153 \times 1.96,\ 0.25 + 0.0153 \times 1.96]$

$= [0.22,\ 0.279]$

Page No.
Date: / /

Page No.
Date: / /

Testing of Hypothesis :-

1) Hypothesis
2) Null Hypothesis (H_0)
3) Alternative Hypothesis (H_1)
4) level of significance (α), Z_α
5) SE (t).
6) Critical value. $Z = \frac{t - E(t)}{SE(t)} = \frac{t - 0}{SE(t)}$.
7) One tail and Two tail.
8) Type I and Type II Error.

~~Table~~
Critical value Table :

SI	α	O.T Z_α	T.T Z_α
1	0.1	1.28	1.65
2	0.05	1.64	1.96
3	0.01	2.33	2.58
4	0.001	3.10	3.29

Q) The mean height of 100 students is 64 and standard deviation 3. Test the statement that the mean height of population is 67 at 5% of level of significance.

Sol^n $H_0 : \mu = 67$ (Two Tailed)
$H_1 : \mu \neq 67$.
$\Rightarrow \mu > 67$ (or) $\mu < 67$.

$n = 100$, $\bar{x} = 64$, $S = 3$.

$\alpha = 5\% = 0.05$, $Z_\alpha = 1.96$

$$Z = \frac{t - E(t)}{SE(t)} = \frac{|\bar{x} - \mu|}{SE(\bar{x})}.$$

$$SE(\bar{x}) = \frac{S}{\sqrt{n-1}} = \frac{3}{\sqrt{99}} = 0.30.$$

$$Z = \frac{|64 - 67|}{0.30} = 10$$

$\boxed{\text{If } |Z| < |Z_\alpha| \Rightarrow H_0 \text{ will accept}}$ → It a general statement.

$|Z| = 10$ & $|Z_\alpha| = 1.96$

∴ H_0 is Rejected.

∴ The mean population is not 67.

Page No.

Date: / /

Q) A random sample of 400 male student is found to be a mean height of 171.38 cm. Can it be regarded as a sample from a large population with mean height 171.17 cm and standard deviation 3.30 with $\alpha = 0.05$.

Sol^n: $H_0: \mu = 171.17$ (Two Tailed)

$H_1: \mu \neq 171.17$

$\mu > 171.17$ (or) $\mu < 171.17$

$n = 400$, $\sigma = 3.30$, $\alpha = 0.05$

$Z_\alpha = \frac{s}{\sqrt{n}} = \frac{3.30}{\sqrt{400}} = 0.165$

Definitions :-

① **Hypothesis** :- A hypothesis is an educated guess (or) proposed explanation that can be tested through experiments (or) observations.

② **Null Hypothesis** :- It is a statement that assumes there is no effect (or) no difference between groups (or) variables. It serve as the default assumption to be tested in research.

③ **Alternative Hypothesis** :- It is a statement that suggests there is an effect (or) a difference between groups (or) variables. It is opposite of the null hypothesis.

④ **Level of significance** :- It is the probability of rejecting a true null hypothesis. It shows the risk of making a Type 1 error, usually denoted by α (alpha), such as 0.05 (or) 5%.

⑤ **Standard error of sample** :- It is the measure of how much a sample statistic (like the mean) is expected to vary from the true population value due to random sampling. It shows the accuracy of the sample estimate.

⑥ **Critical value** :- It is a number that helps decide whether to accept (or) reject the null hypothesis in a test.

⑦ **One-tail** :- It checks for an effect in one direction (either greater than or less than).

Page No.

Date: / /

⑧ **Two-tail** :- It checks for an effect in both direction (greater than or less than).

⑨ **Type-1 error** :- The mistake of rejecting the null hypothesis when it is actually true.

⑩ **Type-2 error** :- when the null hypothesis is accepted but it is actually false.

Q) The mean lifetime of a sample of 400 flourescent light tube produced by a company is found to be 1570 hours with a standard deviation of 150 hrs. Test the hypothesis that the mean life time of the ball produced by the company is 1600 hrs against the alternative hypothesis i.e. greater than 1600 hrs at 1% level of significance.

Ans: $H_0: \mu = 1600$ (one Tail)

$H_1: \mu > 1600$

$n = 400$

$\bar{x} = 1570$

$s = 150$

$SE(\bar{x}) = \frac{s}{\sqrt{n-1}} = \frac{150}{\sqrt{399}} = 7.51$

$\alpha = 1\%$

$Z = \frac{\bar{x} - \mu}{SE(\bar{x})} = \left|\frac{1570 - 1600}{7.51}\right| = 4$

$Z_\alpha = 2.33$

$Z > Z_\alpha \rightarrow$ Null hypothesis rejected.

Q) In a sample of 400 burners, there are 12 whose internal diameter are not within tolerance. Is this sufficient evidence for concluding that the manufacturing process is turning out more than 2% defective burners. Take $\alpha = 5\%$.

Ans: $n = 400$

$p = \frac{12}{400} = 0.03$

$P = 2\% = 0.02$, $\alpha = 5\%$

$H_0: P \leq 2\% = 0.02$ (one tail)

$H_1: P > 2\% = 0.02$

Page No.
Date: / /

$SE(P) = \sqrt{\frac{PQ}{n}}$, $Q = 1-P$.

$= \sqrt{\frac{0.02(1-0.02)}{400}}$

$= 0.007$

$Z = \frac{|0.03 - 0.02|}{0.0070} = 1.42$

$Z < Z_\alpha$ → Null hypothesis accepted.

→ Since $|Z|$ is less than Z_α as 1.4 is less than 1.645 so the null hypothesis is accepted which implise the process is not out of control.

Q) In a Random sample of 400 person from a large population 120 were female can it be said that male and female are in the ratio 5:3 in the population? $\alpha = 10\%$

Ans→ n (sample size) = 400

No. of female = 120

$p = \frac{120}{400} = 0.3$

$M:F = 5:3$, $T = 5+3 = 8$

$H_0: P = \frac{3}{8}$ (Two Tail)

$H_1: P \neq \frac{3}{8}$

$\alpha = 10\% = 0.10$

$Z_\alpha = \frac{5}{\sqrt{n}} = 1.65$ $SE(P) = \sqrt{\frac{PQ}{n}} = \sqrt{\frac{\frac{3}{8}(1-\frac{3}{8})}{400}} = 0.024$

$Z = \frac{|p-P|}{SE(P)} = \frac{|0.3 - 0.375|}{0.024} = 3.125$

∴ $3.125 \nless 1.65 \Rightarrow H_0$ is Rejected.

Page No.
Date: / /

Test of Significance of Difference betn two means?

→ Let A and B be two population with mean μ_1 and μ_2 and varience σ_1^2 and σ_2^2 resp.

→ Let's take two independent sample size n_1 and n_2 from the two population.

→ Let $\bar{x}_1$ and $\bar{x}_2$ are sample mean.

(i) To test the equality of two population mean i.e. to test wheather $\mu_1 = \mu_2$.

(ii) To test the significance of the difference between two independent sample mean i.e. $\bar{x}_1 - \bar{x}_2$.

Step 1:- Choose null hypothesis H_0 such that $H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$ (T-T)

$\left.\begin{array}{l}\mu_1 > \mu_2 \\ \mu_1 < \mu_2\end{array}\right\}$ O-T

Step 2:- Compute the standard error between the difference of two mean.

$SE(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

$SE(\bar{x}) = \sqrt{\frac{\sigma^2}{n}}$

$SE(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$

* Large popn: $SE(x) = \sqrt{\frac{S^2}{n}}$

Small popn: $SE(x) = \sqrt{\frac{S^2}{n-1}}$

Step 3:- $\boxed{Z = \frac{\bar{x}_1 - \bar{x}_2}{SE(\bar{x}_1 - \bar{x}_2)}}$

Step 4:- $\alpha =$, $Z_\alpha =$

Page No.
Date: / /

Q) A college conduct both day and night class. Indent to be identical. A sample of 100 day students yield examination result as $\bar{x}_1 = 72.4$ and $\sigma_1 = 14.8$. A sample of 200 night student yield the examination result as $\bar{x}_2 = 73.9$ & $\sigma_2 = 17.9$.
Are the two mean statistical equal at 10% level?

Soln Given $n_1 = 100$ $\quad n_2 = 200$

$\bar{x}_1 = 72.4$ $\quad \bar{x}_2 = 73.9$

$\sigma_1 = 14.8$ $\quad \sigma_2 = 17.9$

$$SE(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$
$$= \sqrt{\frac{(14.8)^2}{100} + \frac{(17.9)^2}{200}} = 1.94$$

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{SE(\bar{x}_1 - \bar{x}_2)} = \frac{72.4 - 73.9}{1.94} = 0.77.$$

$\alpha = 1.65$ $Z_\alpha = 0.77$. Hen H_0's accepted.

Q) A random samples of 1000 workers from South India show that their mean weigh are 47 rupees per week, with standard deviation 28 rupees. A random sample of 1500 workers from North India keeps a mean weigh 49 rupees per week, with the standard deviation 40 rupees. Is there any significance difference betn their mean with 5% level of significance?

Ans Given $\bar{x}_1 = 47$ $\quad S_1 = 28$ $\quad n_1 = 1000$

$\bar{x}_2 = 49$ $\quad S_2 = 40$ $\quad n_2 = 1500$

$\alpha = 5\% \Rightarrow Z_\alpha = 1.96$

$$SE(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} = \sqrt{\frac{(28)^2}{1000} + \frac{(40)^2}{1500}} = 1.36$$
$$Z = \left|\frac{\bar{x}_1 - \bar{x}_2}{SE(\bar{x}_1 - \bar{x}_2)}\right| = \left|\frac{-2}{1.36}\right| = 1.47$$

$\alpha = 1.96 > Z = 1.47$. Here H_0 is accepted.

Page No.
Date: / /

Q) ~~The yield of~~ wheat ~~from a district A~~

Test of significance for difference S.D. (σ)	Test of significance for difference proportion.
$H_0: \sigma_1 = \sigma_2$	$H_0: P_1 = P_2$
$H_1: \sigma_1 \neq \sigma_2$ (two tail).	$H_1: P_1 \neq P_2$
α = given; Z_α = Table	α = given, Z_α = Table.
$Z = \frac{S_1 - S_2}{SE(S_1 - S_2)}$, S_1 = sample S_1, S_2 = 2nd sample	$Z = \frac{P_1 - P_2}{SE(P_1 - P_2)}$
Case 1: $SE(S_1 - S_2) = \sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}$ n_1 = 1st, n_2 = 2nd	(i) $SE(P_1 - P_2) = \sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}$
Case 2: If σ_1 & σ_2 are not given	(ii) $SE(P_1 - P_2) = \sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_1}{n_2}}$
$SE(S_1 - S_2) = \sqrt{\frac{S_1^2}{2n_1} + \frac{S_2^2}{2n_2}}$	(iii) $SE(P_1 - P_2) = \sqrt{\hat{P}\hat{Q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$
	$\hat{P} = \frac{n_1P_1 + n_2P_2}{n_1 + n_2}$
	$\hat{Q} = 1 - \hat{P}$.

Q) In a large city A, 20% of a random sample of 900 school children had defective eye sight. In other large city B 15% of random sample of 1600 children are same defect. Is this difference betn the two proportion significant at 5% of level?

Ans $P_1 = 20\% = 0.20$.

$P_2 = 15\% = 0.15$

$n_1 = 900$ $\qquad \hat{P} = \frac{900 \times 0.20 + 1600 \times 0.15}{900 + 1600} = 0.168$

$n_2 = 1600$

$$SE(P_1 - P_2) = \sqrt{0.168 \times 0.832\left(\frac{1}{900} + \frac{1}{1600}\right)}$$
$$= 0.015$$
$$Z = \frac{0.20 - 0.15}{0.015} = 3.33$$

$\alpha = 5\%$, $Z_\alpha = 1.96$

$|Z| = 3.33$, $> 1.96 = Z_\alpha$

$\therefore H_0$ is rejected.

$\therefore$ There is a difference between the two proportions.

Page No.
Date: / /

Q) Random sample drawn from two countries given in the following data releating to the height of the adult.

~~height~~	Country A.	Country B
mean height	67.42	67.25
S.D.	2.58	2.50
sample size	1000	1200

Is there difference betn the S.D. at 5% significance level.

Soln $H_0: \sigma_1 = \sigma_2$

$H_1: \sigma_1 \neq \sigma_2$

$\alpha = 5\%$, $Z_\alpha = 1.96$

$n_1 = 1000$, $n_2 = 1200$

$S_1 = 2.58$, $S_2 = 2.50$

$SE(S_1 - S_2) = \sqrt{\frac{(2.58)^2}{2\times1000} + \frac{(2.50)^2}{2\times1200}} = 0.077$

$Z = \frac{2.58 - 2.50}{0.077} = 1.039$

$\Rightarrow Z = 1.039 < Z_\alpha = 1.96$

$\therefore H_0$ is accepted.

Q) A Company has head office Kolkata and branch at Mumbai. The personal director want to know if the workers at the two place would like the introduction of a new plan and a survey has conducted for this purpose. out of sample of 500 workers at Kolkata 62% favour the new brand. At Mumbai out of 400 worker 41% were against the plan. Is there any S.D. betn the two groups in their attitude towards the new plan at 5% level.

Page No.
Date: / /