

## Engineering statics

23/7/22

Origin and development of statistics :-

→ simply it is known as 'political state'.

Concept of statistical population and a sample  
Having ~~car~~.

20,00,000 = population of int

Prize = Value

200 ~~cars~~ = Sample.

Avg value = parameter. - define character of the popu

200 numbers = Sample data

\* Avg of the data = statistic

### Basic Definitions:

- 1) A population
- 2) A sample
- 3) A measurement
- 4) A sample data
- 5) A parameter
- 6) A statistic

Statistics always gives us approximate  
not exact exact value (Drawback)

→ Descriptive statics : <sup>took</sup> sample and collect  
the data, and analyzing.

→ Inferential statics : After Analyzing  
to have a conclusion is called ...

$$T = 0.4 \times 249.87 + 0.6 \times 333.132$$

$$T = 252.28^{\circ}\text{C}$$

$$P = 755 \text{ kPa}$$

$$P = 783 \text{ kPa}$$

Page No. \_\_\_\_\_  
Date \_\_\_\_\_

- statistics is a collection of methods for collecting, displaying, analyzing and drawing conclusions from data.
- Descriptive st. :  $\rightarrow$  ~~with the help of~~
- Inferential st. :  $\rightarrow$  drawing ... (pictorial)

$\star$  Data  $\Rightarrow$  ~~are~~ Quantitative and qualitative  
 (Categorical)  
 ↓  
 size of the body / Features of car

- ⇒ the relationship between a population of interest and a sample drawn from that population.

•	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○

Population

Sample	A	(1)
•	○	(2)
○	○	(3)
○	○	(4)

$$(x_1, x_2, \dots, x_n)$$

Par.

$\mu$

$\sigma^2$

$P$

stats. -

$\bar{x}$

$s^2$

$P$

## \* Importance and scope of statistics :-

- Statistics and planning :-
- Statistics and economics :-
- Statistics and business :-
- Statistics and industry :-
  - It is very widely used in 'Quality control' in production Engineering.
- Insurance is a vast industry. There are hundreds of insurance.
- The premium of insurance is based on the statistics. (Basis of the data is Time)
- Statistics and Mathematics :-
  - Use of statistics in research can lead to summarization, Proper characterization; performance and description of the outcome of the research.  
(We use ANOVA - Analysis of Variance)
- Besides this, medical area would be less effective without the research.
- For Example : Which country did the most test of covid-19.
- Medical science. [T-test w-test of significance]
- The Indian Meteorological dep - Weather report (forecast).
- Disaster Management.
- Statistics and psychology and Education.
- Statistics and War.

$$T = 0.4 \times 249.87 + 0.6 \times 232.37$$
$$T = 242.89^{\circ}\text{C}$$

333.132  
15 kPa  
1.2a

Page No. \_\_\_\_\_  
Date \_\_\_\_\_



## Limitations of statistics :-

- ① → Statistics is not suited to the study of qualitative phenomenon.  
→ Statistics being a science dealing with a set of numerical data, is applicable to the study of only those subjects of enquiry which are capable of quantitative measurement.
- ② Statistics does not study individuals.  
→ We need group data or comparison data.
- ③ Statistical laws are not exact.  
→ It is predictions of all the patterns or demands followed by the demands.  
[Probability]
- ④ Statistics is liable to be misused.  
→ It may use or interpret wrongly or also be given by wrong person.  
Methods → It must be used by experts.  
are their most dangerous tools in the hands of the experts.

1/8/22

[Page No.]  
[Date]

### \* Scales of measurement :-

→ Sc. of " refer to ways in which variables/numbers are defined and categorized.

→ There are 4 types :-

Nominal

ordinal

interval

ratio

### \* Properties of measurement :-

→ Identity : Identity is a Unique meaning.

→ Magnitude : have an ordered relationship to one another.

→ Equal intervals : the data points along the scale are equal.

→ A minimum value of zero : The scale has a true zero point.

→ But it is not true in case of weight measurement.

$$T = 0.4 \times 274.87 + 0.6$$

$$T = 242.29^{\circ}\text{C}$$

Page No. \_\_\_\_\_  
Date \_\_\_\_\_

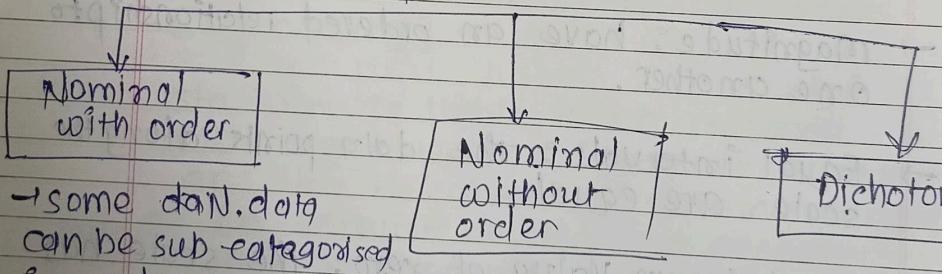
### ① Nominal scale of measurement:

- Ex Where do you live?  
Options: (1) suburbs  
(2) city  
(3) Town

→ Now for ans, we are considering it as a non numerical but for researcher, it's numerical.

- While capturing nominal data; researchers conduct analysis based on the associated labels.

#### Nominal data



→ Some dan. data can be sub-categorised in order.

→ Ex male

→ complete opposite to each other.

Female

Ex Yes & No

→ Ex Much cold,  
cold,  
Hot,  
~

## ② Ordinal scale of measurement:

- as a Variable measurement scale
- Used to depict non-mathematical ideas such as frequency.
- Feedback (Websites with emoji feedback)
  - (How satisfied are you our services)
  - (Should be in order)
  - (Unsatisfied to satisfied)
- An ordinal scale of measurement represents an ordered series of relationships or rank order.
- These values can be added or subtracted but cannot multiply or divide.

## ⇒ Relation between Nominal and ordinal:

- 1) The order of variables is of prime importance and so is the labeling.
- 2) This is where ordinal scale is a step above nominal scale.
- 3) Analyzing results based on the numerical will be convenient for researcher.

$$T = 0.4 \times 219.87 + 0.6 \times 233.27$$

$$T = 242.29^{\circ}\text{C}$$

15 kPa  
kPa

### 3) Interval scale of measurement :-

- Contains properties of nominal and ordered data, but the difference between data is quantified.
- In this, the number zero is an existing variable.
- In the ordinal, zero is so data does not exist.
- In interval, zero has meaning ( $0^{\circ}\text{C}$ ) - temp.
- Data points on the interval scale have same difference between them.

### 4) Ratio scale of measurement :-

- Includes properties of all four scales of measurement.
- Data in the ratio scale can be added, subtracted, divided and multiplied.

## Types of Data

Quantitative

→ numbers

Discrete

→ Whole numbers that can't be broken down  
→ ex [number of items]

Continuous

→ Numbers that can be broken down such as height, weight

Qualitative

→ non numerical

Nominal

↓  
Naming Variables

Ordinal

↓  
The order of values.

Interval

↓  
Numbers with known diff. between variables, such as time

Ratio

↓  
Numbers that have measurable intervals where diff. can be determined, such as height or weight..

CHAP - |

$\star$  The central tendency is stated as "the statistical measure that represents the single value of the entire distribution or a dataset."  
 → It aims to provide an accurate description of the entire data in distribution.

$\Rightarrow$  The mean :-

→ The first measure of central location is the usual "avg" that is familiar to everyone!

$$\bar{x} = \frac{\sum x}{n}$$

$\Rightarrow$  Population mean of a set of  $N$  population data is the number  $\bar{x}$  defined by the formula.

$$\bar{x} = \frac{\sum x}{N}$$

$$\text{Ex: } \begin{array}{c|ccccc} x & 0 & 1 & 2 & 3 & 4 \\ f & 3 & 6 & 6 & 3 & 5 \end{array}$$

$$\therefore \bar{x} = \frac{0+1+2+3+4}{5} =$$

18/12/22



NAME	Daray Bhogat	ROLL NO.	
STD.		Div.	
SUBJECT		Date	
SCHOOL			

$$\bar{x} = \frac{\sum x}{n} = \frac{0 \times 3 + 1 \times 6 + 2 \times 6 + 3 \times 3 + 4 \times 1}{19}$$

X The median :-

24.8 22.8 24.6 192.5 25.2 18.5 23.7

$$\text{avg} = 29.44 \approx 29.44$$

→ Ascending order

18.5 22.8 23.7 24.6 24.8 25.0 192.5

→ If it is even numbers then  
from of the

$$\frac{0+6}{2} = 0.5$$

→ Median

⇒ Sample median X.



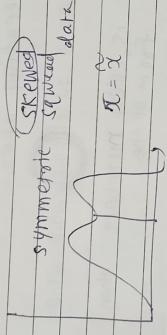
Median

There are four types of data

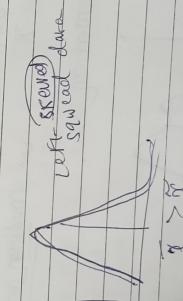
$$\bar{x} = \tilde{x}$$



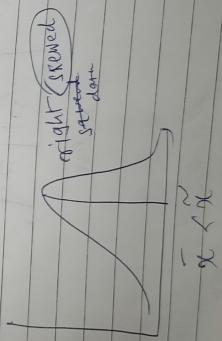
symmetric data



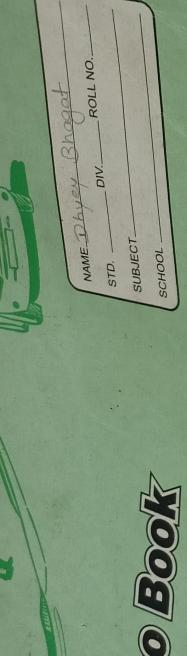
symmetric skewed data



$$\bar{x} > \tilde{x}$$

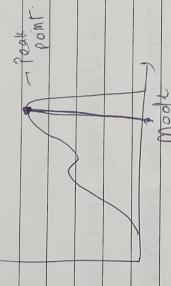


$$\bar{x} < \tilde{x}$$



### \* The mode :-

→ The sample mode of the data that frequently occurring data.



→ Usually we have one mean and one median but in the case of mode we can have repeated one.

→ Problem !

→ Mean, Median & Mode

21 23 23 54 67 21 25 21 54 72 75

$$\text{Mean} = \bar{x} = \frac{\sum x}{N}$$

$$\bar{x} = 21 + 23 + 23 + 54 + 67 + 21 + 25 + 21 + 54 + 72 + 75 \\ + 72 + 75 \\ \hline 11$$

$$\bar{x} = 41.45$$

$$T = 0.4 \times 219.87 + 0.6$$

$$\therefore T = 242.29^{\circ}\text{C}$$

$$P_{1\text{ sat}} = \frac{246}{25}$$

$$P_{2\text{ sat}} = \frac{246}{25}$$

DOMS	Page No.
Date	

Page No.	
Date	

median =

21, 21, 21, 23, 23, 25, 34, 54, 67,  
72, 75

median = 25

→ mode = 21

### ★ Measures of Variability or Dispersion:

→ Variability refers to how ~~personal~~ spread out a group of data is.

- There measures of the variability
- o Range → <sup>between</sup> Highest and lowest
  - o Variance
  - o Standard deviation

T = 0

T = 242.82

$$P_{\text{sat}} = 28.45 \text{ kPa}$$

$$P_1 = 23.383 \text{ kPa}$$

## \* The Variance and the standard Deviation

⇒ Sample Variance:

$$S^2 = \frac{\sum (x - \bar{x})^2}{n-1}$$

→ by algebra is equivalent to the formula,

$$S^2 = \frac{\sum x^2 - \frac{1}{n} (\sum x)^2}{n-1}$$

⇒ Sample standard deviation:

$$S = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}} = \sqrt{\frac{\sum x^2 - \frac{1}{n} (\sum x)^2}{n-1}}$$

→ If we consider it for population,

$$\sigma = \tau \quad \& \quad \sigma^2 = \tau^2$$

(Imp. notations)

$$\Rightarrow \tau = \sqrt{\frac{\sum (x - \mu)^2}{N}}$$

$$\Rightarrow \tau^2 = \frac{\sum (x - \mu)^2}{N}$$

⇒ Data set II:

x	$\bar{x}$	$(x-\bar{x})$	$(x-\bar{x})^2$
46	6	36	
37	-3	9	
40	0	0	
33	-7	49	
42	2	4	
36	-4	16	
40	0	0	
47	7	49	
34	-6	36	
45	5	25	
			224

$$\bar{x} = 40$$

$$S^2 = \text{Variance} = \frac{\sum (x-\bar{x})^2}{n-1} = \frac{224}{9} = 24.89$$

$$S = \sqrt{24.89} = 4.98 \approx 4.99$$

\* Key Points :-

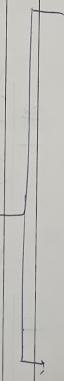
→ The range, the s.d and the variance give a quantitative ans to the que 'How variable are the data'

### Unit -3 ⇒ Data Analysis

\* Random Variable:

- Assumes Numerical values associate with the random outcome
- Each Numerical value to each variable.

⇒ Types of Random Var.



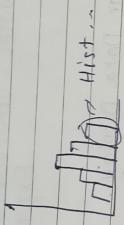
Discrete

- Number of Sales → Time
- Number of calls → length
- Weight

→ Data Collection is an imp step in the data analysis process

→ OBSERVATIONAL STUDY

## \* Frequency Histograms :-



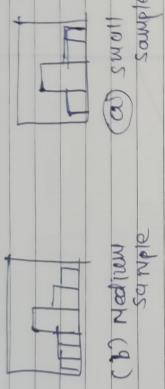
→ 5 & 1 dia. is not practical for large data sets, so we can go for this type of diagram.

## \* Relative Frequency Histograms :-

→ Five students scored in the 80s  
→ If there are 30 students in the  
Statistics class then,  
 $R.F. \% = \frac{\sum}{30} \times 100$



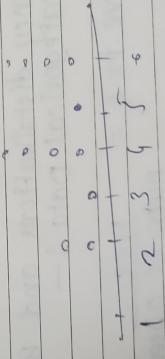
→ When is the size n of a set



Ex we find area under the curve.

- ② Large Sample
  - ③ Very Large Sample
- (It's freq. but with n/marsize)  
that's why it is looking smooth,

### Dotplots for Numerical Data

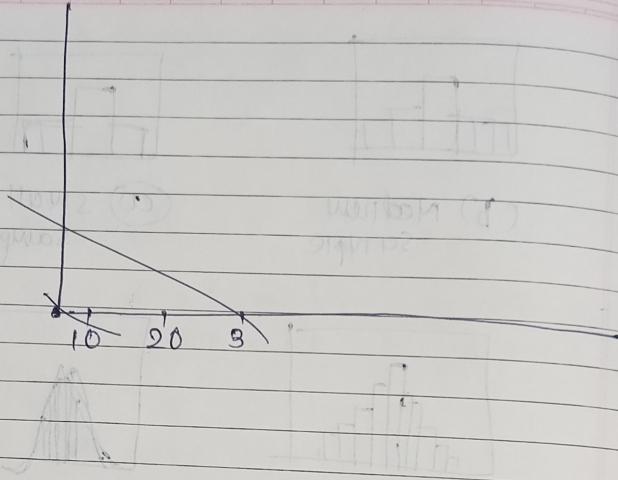


When to Use : Small numerical data sets.

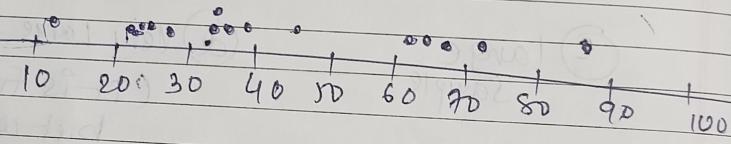
Cali : 64 41 44 31 37 73 72 68 35 39 81 90

Page No. \_\_\_\_\_  
Date \_\_\_\_\_

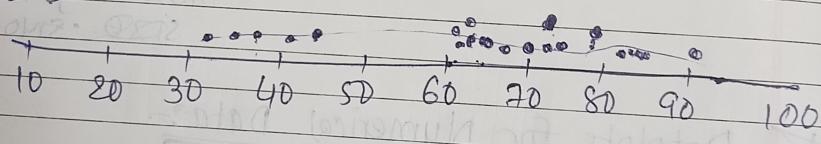
Numerical :-



Texas



California



22 Aug 92

\* Frequency distributions and Bar charts for categorical Data :-

relative freq. =  $\frac{\text{freq.}}{\text{number of observations in the data}}$

## \* Bar charts :-

Hist - connected.

Bar



## \* When to Use : Categorical Data

## \* Pie charts :-



depending on the area,  
dividing By  $360^\circ$ ...

→ When to Use :- Categorical data with  
a relatively small number of possible  
categories.

→ Pie charts are most useful for illustrating  
proportions of the whole data set for various  
categories.

Display Data represent the number of days of sick leave taken by each of 50 workers over last 6 weeks.

2, 2, 0, 0, 5, 8, 3, 4, 1, 0, 0, 1, 1, 5, 1, 2, 1, 5, 0, 4, 0, 1, 8, 9, 7, 0, 1, 7, 2, 5, 5, 4, 3, 3, 0, 0, 2, 5,

1, 3, 0, 1, 0, 2, 4, 5, 0, 3, 5, 5, 1, 1, 3, 0, 1, 0, 2, 4, 5, 0, 3, 5, 5, 1,

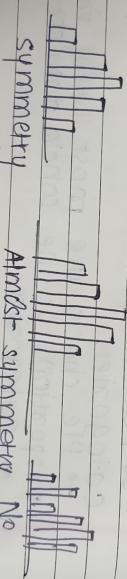
① At least 1 day off leave.

~~2~~ = 50 - 12 = 38

② Between 3 and 5  
we have to include 3, 4, 5

③ More than 5 days.  $\Rightarrow$  9

⇒



Symmetry Am not symmetry No symm..

**★ Correlation :-**

- Finding the relationship between two quantitative variables without being able to infer causal relationships.

**Scatter diagram :-**

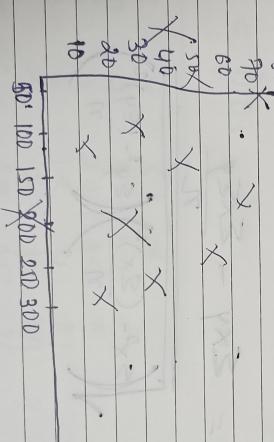
 One variable is called

Independent (X) and the second is called dependent  
points are not joined

Example:

WT  $\Rightarrow$  67 69 85 83 74 81 97 92 114 85  
(kg)

SBP (mmHg) 120 125 140 160 130 180 150 140 200 130



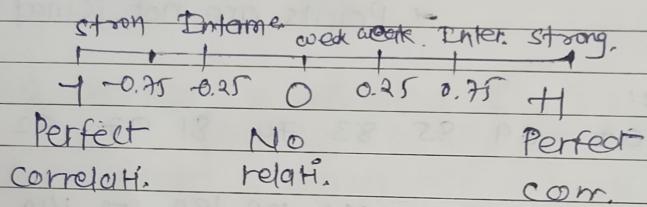
→ Types of

- Positive
- Negative
- None



## \* correlation coefficient:

- Simple correlation coeff. ( $r$ ),
- measures the Nature and strength between two variables of the quantitative type
- the sign of  $r$  denotes the Nature of association.
- The value of  $r$  ranges between (-1) to (+1).



⇒ Sec (r)

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{n}$$

$$\sqrt{\left( \frac{\sum x^2 - (\sum x)^2}{n} \right) \left( \frac{\sum y^2 - (\sum y)^2}{n} \right)}$$

Example

Age	Weight	$XY$	$x^2$	$y^2$
7	12	84	49	144
6	8	48	36	64
8	12	96	64	144
5	10	50	25	100
6	11	66	36	121
9	13	117	81	169
41	66	461	291	742

Page No.

Date

$$r = \frac{461 - \frac{(41)(66)}{6}}{10}$$

$$\sqrt{\left(291 - \frac{(41)^2}{6}\right) \left(742 - \frac{(66)^2}{6}\right)} \cdot \sqrt{(244, 3)(21)}$$

~~0.03~~      0.75

X	Y	$XY$	$x^2$	$y^2$
10	2	20	100	4
8	3	24	64	9
2	9	18	4	81
1	7	7	1	49
5	6	30	25	36
6	5	30	36	25
$\Sigma X = 32$	$\Sigma Y = 33$	$\Sigma XY = 129$	$\Sigma x^2 = 230$	$\Sigma y^2 = 204$

$$r = \frac{129 - \frac{(32)(33)}{6}}{\sqrt{(39, 33)(33, 33)}} = -0.94 \quad \begin{array}{l} \text{Indirect} \\ \text{Strong} \\ \text{correlation} \end{array}$$

## \* Coefficient of correlation :

Pearson correlation ( $r$ ),

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}}$$

$\bar{x}$  = mean of  $x$

$\bar{y}$  =

$x_i, y_i$  = sample

$\rightarrow$   $-1$  to  $1$

strong      strong positive  
weak      positive  
 $\Rightarrow 0 \rightarrow$  No relation

26/08

## \* Partial correlation :

→ It is a measure of the strength and direction of a linear relationship between two continuous variables whilst controlling for the effect of one or more other continuous variables.

[Also known as 'covariates' or 'control variables']

relationship

→ B/w Fertilizer & crop yield, keeping the weather conditions constant.

→ " " anxiety level & academic achievement, while controlling the intelligence.

→ Certain Assumptions;

→ Partial correlation coefficient formula;

$$\rho_{xy,z} = \frac{\rho_{xy} - \rho_{yz} * \rho_{zx}}{\sqrt{(1 - \rho^2_{yz})(1 - \rho^2_{zx})}}$$

$$\sqrt{(1 - \rho^2_{yz})(1 - \rho^2_{zx})}$$

see back

x	y	z	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(z_i - \bar{z})$	$(x_i - \bar{x})(y_i - \bar{y})$	$(y_i - \bar{y})(z_i - \bar{z})$	$(x_i - \bar{x})(z_i - \bar{z})$	$\rho_{xy,z}$
10	29	17	-9	-14.89	-16	134.01	238.24	144	
13	33	23	-6	-10.89	-10	65.34	108.9	60	
19	41	21	0	-2.89	-12	0	34.68	0	
16	47	29	-3	3.11	-4	-0.93	-12.44	-24	
13	51	37	-6	9.11	4	-0.26	-18.66	-28.44	
21	43	41	2	-0.08	8	-0.16	-0.64	16	
23	31	39	4	-12.89	6	-51.56	-309.36	24	
29	49	47	10	5.11	14	57.1	41.54	140	
27	71	73	8	29.01	10	28.0	271.10	80	
17	395	293				216.88	363.62	383.58	452

$$\bar{x} = \frac{171}{9} = 19 \quad \begin{matrix} (\bar{x}_i - \bar{x})^2 & (\bar{y}_i - \bar{y})^2 & (\bar{z}_i - \bar{z})^2 \end{matrix}$$

$$\rho_{xy} = 0.53 \quad \bar{y} = \frac{395}{9} = 43.89 \quad \begin{matrix} 81 & 221.21 & 359.256 \\ 36 & 118.59 & 100 \\ 0 & 8.35 & 144 \end{matrix}$$

$$\rho_{yz} = 0.34 \quad \bar{z} = \frac{297}{9} = 33 \quad \begin{matrix} 9 & 9.67 & 16 \\ 36 & 50.55 & 16 \\ 4 & 0.01 & 64 \end{matrix}$$

$$\rho_{xz} = 0.80 \quad \begin{matrix} 16 & 166.15 & 36 \\ 100 & 26.11 & 196 \\ 64 & 934.95 & 100 \end{matrix}$$

$$\rho_{xy,z} = 0.46 \quad \begin{matrix} 346 & 1336.09 & 928 \end{matrix}$$

## Multiple Correlation

- It deals with the situation in which the correlation b/w three or more variables are required to be found.

⇒ formula:

$$(R_{xyz}) = \sqrt{\frac{(r_{xz}^2 + r_{yz}^2 - 2r_{xy}r_{yz}r_{xz})}{(1 - r_{xy}^2)}}$$

Subj.	A <sub>1</sub>	A <sub>2</sub>	A <sub>3</sub>	T <sub>(A<sub>1</sub>-A<sub>2</sub>)</sub>	T <sub>(A<sub>2</sub>-A<sub>3</sub>)</sub>	T <sub>(A<sub>1</sub>-A<sub>3</sub>)</sub>
1	15	8	25	-0.3	-1.2	-0.8
2	18	3	20	0.9	-1.6	9.2
3	13	8	27	-2.3	3.2	1.2
4	14	6	24	-1.3	1.2	-1.8
5	19	2	30	3.2	-2.8	4.2
6	11	3	21	-4.3	-1.8	-4.8
7	17	4	26	1.9	-0.8	0.2
8	20	4	31	0.9	-0.8	5.2
9	10	5	20	-5.3	0.2	-5.8
10	16	7	25	0.9	2.2	-0.8
	153	48	258			

→ Find Partial & Multiple correlation

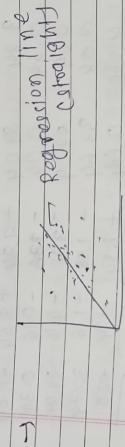
08 aug

Page No. \_\_\_\_\_  
Date \_\_\_\_\_

### \* Regression Analyses :-

→ Technique concerned ...

→ Regression goes one step beyond correlation in identifying the relationship between two variables.



→ The slope is often called the re. coff. and the intercept the reg. constant.

→ The slope can be defined as,

$$B_1 = \frac{\sum xy}{\sum x}$$

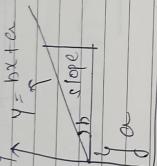
### \* Regression :-

#### Correlation and Regression

→ Regression minimizes the residuals

$$\begin{aligned} b_1 &= \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} & y &= \bar{y} + b(x - \bar{x}) \\ & & & \hat{y} = a + bx \end{aligned}$$

\* Linear Equations :-



\* Numerical

S.No.	Age(x)	Weight(y)	x <sup>2</sup>
1	4	84	16
2	6	98	36
3	8	12	64
4	5	10	25
5	6	11	36
6	9	13	81
	44	66	291

→ What is predicted weight when the age is 8.5 years.

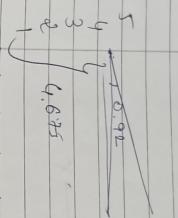
$$\Rightarrow b_1 = \frac{461 - 2706}{6} = \frac{10}{10.83} = 0.923$$

$$y = b_1 x + b_0$$

$$\Rightarrow y = 11 + 0.923(8.5) + 6.83 \\ = 12.54.$$

$$y(m) = 40.675 + 0.92x$$

→



Age	B. P.	$\frac{g}{kg}$	$x^2$
20	(4)	100	400
26	108	2400	400
42	108	5504	1849
63	141	8883	3969
67	126	3896	4489
53	134	2102	2809
31	128	3968	961
58	136	9888	3364
46	132	6632	2116
58	140	8100	3364
70	144	10080	4900
46	128	5888	2116
53	136	9208	2809
60	146	8760	3600
20	124	2480	400
63	143	9009	3969
43	130	5390	1849
26	124	3224	876
19	(2)	2999	16
31	146	3780	961
23	123	2829	529

$$\begin{aligned} Sx = 82 & \quad Sy = 2630 \quad Sxy = 114360 \\ Sx^2 = 39230 & \end{aligned}$$

$$y = \bar{y} + b(x - \bar{x})$$

a.

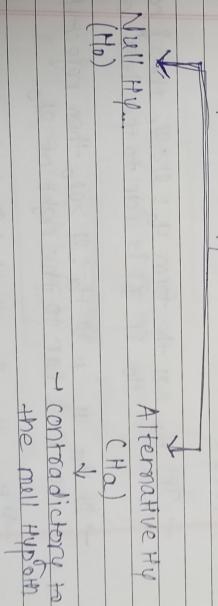
$$b = \frac{\sum xy - \bar{x}\bar{y}}{n}$$

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

$$\text{SL} = 154 \text{ years?}$$

## Hypotheses testing (1 sept / 2021)

→ Types of hypotheses.



- Reject  $H_0$  and therefore accept  $H_1$  or
- Fail to reject  $H_0$  (and therefore fail to ~~accept~~  $H_1$ ).

The null hypothesis is one an assertion containing an equal sign,

Or

Can have one of the three forms

$\leq$ ,  $\geq$ ,  $\neq$

### Example ①

$$\begin{aligned} \text{Null Hypothesis} &: H_0 : \bar{x} = 12.75 \\ \text{After} &: H_a : \bar{x} \neq 12.75 \end{aligned}$$

Q:  $\bar{x} = ?$

Kam bhi Nai Hani chahiye aur  
zyada bhi nai Hani chahiye.

\* The criterion for judging between  $H_0$  and  $H_a$  based on the sample data is: if the value of  $\bar{x}$  would be highly unlikely.

If  $H_a$  has the form  $H_a: \bar{x} < 0$  then reject  $H_0$  if  $\bar{x}$  is far to the left of  $\bar{x}_0$ .

If  $H_a: \bar{x} > 0$  then reject  $H_0$  if  $\bar{x}$  is far to the right of  $\bar{x}_0$ .

If  $H_a: \bar{x} \neq 0$  then reject  $H_0$  if  $\bar{x}$  is far away from  $\bar{x}_0$  in either direction.

2nd Sept

Ex. 9

$$\left| \bar{x} = \frac{\sum x}{n} = 8.0 \right|$$

In Ex. 2, it is known that the population is normally distributed and suppose that the test of hypotheses  $H_0: \mu = 8$  vs  $H_a: \mu \neq 8$  with sample size  $n$ . Construct the rejection region for the test for the choice  $\alpha=0.10$ .

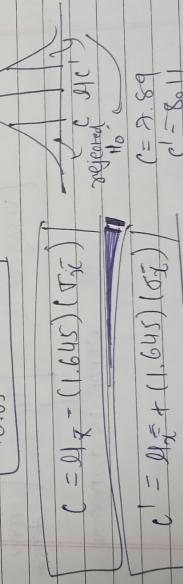
→ Test statistics

(Type I error) t-distribution  
z-distribution  
 $\chi^2$ -distribution  
F-distribution

t-distribution | student distribution  
z-distribution | normal distribution  
 $\chi^2$ -distribution | Chi-square distribution  
F-distribution | Cauchie not given

→  $\alpha/2$  is 0.05 &  $\beta = 0.10$  (or reject)  
 $H_0: \mu = 8.0$   
 $H_a: \mu \neq 8.0$

$$Z_{0.05} = 1.645$$



→ So, state the statement & which hypothesis will be accepted.

Statement:

At level of significance  $\alpha = 0.1$  & taking a sample size 5 if the value of  $\bar{y}$  is ~~for~~ between 7.89 and 8.11 will be accepted at  $\alpha$  if it is less than 7.89 & more than 8.11 ~~and~~ we will be rejecting null hypothesis.

$\Rightarrow$   $H_0 \neq H_a$  has the  $\alpha \neq 0.0$  - two tailed test

$\left\{ \begin{array}{l} \text{if } \bar{y} < 8.0 \rightarrow \text{left tailed test} \\ \text{if } \bar{y} > 8.0 \rightarrow \text{right tailed test} \end{array} \right.$   
the last two one tailed test

Two types of errors,

Our decision	True state of $H_0$	Take state of $H_0$
Do not reject $H_0$	$H_0$ is $\bar{y}$	$H_0$ is $\bar{y}$
Reject $H_0$	$H_0$ is $\bar{y}$	Type I error

Type II error

The type I error is dangerous.

1) If we fail to reject  $H_0$  even though  $H_0$  is false then we commit Type II error.

5 Sept

\* State the null and alternate hypothesis

$$H_0: \bar{U} = U_0$$

$$H_{0.05}: \bar{U} = 24.5 \text{ cm} \quad | \quad H_0: \bar{U} = U_0$$

(a)

$$H_0: \bar{U} = U_0$$

$$H_a: \bar{U} > U_0$$

(b)

$$H_0: \bar{U} = U_0$$

$$H_a: \bar{U} > U_0$$

(c)

$$H_0: \bar{U} = \$14,356$$

$$H_0: \bar{U} = U_0$$

$$H_a: \bar{U} > U_0$$

(d)

$$H_0: \bar{U} = \$14,356$$

$$H_0: \bar{U} = U_0$$

$$H_a: \bar{U} \neq U_0$$

(e)

$$H_0: \bar{U} = 89.4$$

$$H_0: \bar{U} = U_0$$

$$H_a: \bar{U} \neq U_0$$

→ (Unit lagane ki zarurat Naihe)  
→ proper method.

$$\left. \begin{array}{l} \text{Null Hypothesis, } H_0: \bar{U} = 24.5 \\ \text{Alternative Hypothesis, } H_a: \bar{U} > 24.5 \end{array} \right\}$$

a)  $H_0: \bar{X} = 38.2$   
 $H_A: \bar{X} < 38.2$

$H_0: \bar{X} = 58.291$   
 $H_A: \bar{X} \neq 58.291$

$H_0: \bar{X} = 13.3$   
 $H_A: \bar{X} > 13.3$

$H_0: \bar{X} = 161.9$   
 $H_A: \bar{X} \neq 161.9$

$H_0: \bar{X} = 42.8$   
 $H_A: \bar{X} > 42.8$

### Sampling Distribution

since a sample is random, every statistic is a random variable, it varies from sample to sample in a way that cannot be predicted with certainty.

The mean and standard Deviation of the sample :-

$\bar{X}$  is a random variable

$\mu_{\bar{X}} = \text{mean}$   
 $\sigma_{\bar{X}} = \text{standard deviation}$

with replacement of size two.

X  
2 samples

Probability distribution,

mean	Sample
$\bar{x}_1 = 152, 156, 160, 164$	
$\bar{x}_2 = 152, 156, 152, 152$	
$\bar{x}_3 = 156, 160, 160, 152$	
$\bar{x}_4 = 160, 164, 164, 164$	
$\bar{x}_5 = 164, 160, 164, 164$	
$\bar{x}_6 = 160, 156, 156, 164$	
$\bar{x}_7 = 164, 152, 164, 164$	
$\bar{x}_8 = 156, 152, 164, 156$	

→ Sample mean

$$M_{\bar{X}} = S \bar{X} p(\bar{X})$$

80

$\bar{X} = 152, 154, 156, 158, 160, 162, 164, 166$

$$P(\bar{X}) = \frac{1}{16}, \frac{3}{16}, \frac{4}{16}, \frac{3}{16}, \frac{2}{16}, \frac{1}{16}$$

$$M_{\bar{X}} = (152 \times \frac{1}{16}) + (154 \times \frac{3}{16}) +$$

$$+ (156 \times \frac{4}{16}) +$$

$$+ (158 \times \frac{3}{16}) +$$

$$+ (160 \times \frac{2}{16}) +$$

$$+ (162 \times \frac{1}{16})$$

.

.

.

$$\boxed{M_{\bar{X}} = 158}$$

$$\sigma_x = \sqrt{\sum (x_i - \bar{x})^2}$$

$$= \sqrt{10}$$

$$\Rightarrow \text{mean} = \frac{152 + 156 + 160 + 164}{4}$$

$$= 158$$

$\Leftarrow$  population mean = sample mean.

$\rightarrow$  Standard Deviation =  $\sqrt{20}$

$\Leftarrow$  so from values  $\Rightarrow$  sample standard dev.  
 sample mean  $\bar{x}_1 = 21$ ,  $\bar{x}_2 = 15$ ,  $\bar{x}_3 = 16$ ,  $\bar{x}_4 = 16$   
 population standard deviation  
 population mean

~~$x$~~  &  $\bar{x}$  stands for individual value & total.

8 sept

Page No.

Date

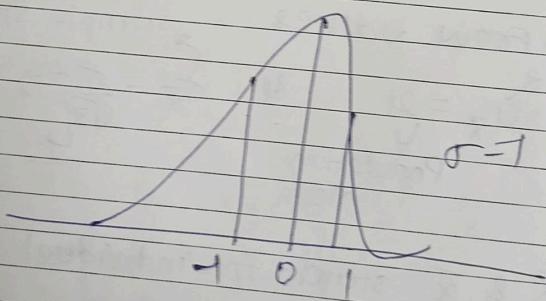
$\mu = \$13,525$   
 $\sigma = \$4,180$   
Samples of size - 100.

→ mean  $\bar{\mu}_x = ?$   
standard deviation  $\sigma_x = ?$   
sample mean  $\bar{x} = ?$

$$\bar{\mu}_x = \mu = \$13,525$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} =$$

\* standard Normal Random Variable



→ Density curve

→ Cumulative Normal probability = ?

16.59158 A  
1.25326 B  
364931 C  
2665574 D  
33.424

Page No. \_\_\_\_\_  
Date \_\_\_\_\_



Find the probabilities indicated, where  $Z$  always denotes a standard Normal Random Variable,

a)  $P(Z < 1.48)$

b)  $P(Z < -0.25)$

$\Rightarrow 0.9306$

(From Z table)

$\Rightarrow 0.4013$

② Find the probabilities indicated.

a)  $P(Z > 1.60)$

b)  $P(Z > -1.02)$

$\Rightarrow P(Z > 1.60) = P(Z < -1.02) = P(Z \leq -1.02)$

$P(Z > -1.02) = 1 - P(Z \leq -1.02)$

③

a)  $P(0.5 < Z < 1.57)$

b)  $P(-2.55 < Z < 0.09)$

$\Rightarrow P(Z < 1.57) - P(Z < 0.50)$

$= P(Z < 0.09) - P(Z < -2.55)$

=

\* Sta. Nor. random vari.  $Z$  is  
a normally distributed random  
variable with mean  $\mu = 0$  & standard  
deviation  $\sigma = 1$

$$P(a < X < b) = P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right)$$

$\Rightarrow a$  can be any decimal number  
or  $-\infty$

$\Rightarrow b$  can be any decimal number  
or  $\infty$ .

$Z$  is sta. No. var.

Numerical  $\mu = 10$  &  
 $\sigma = 2.5$

(a)  $P(X < 14)$

$$= P\left(\frac{14-10}{2.5} < Z < \frac{14-10}{2.5}\right)$$

$$= P(-0.8 < Z < 1.6)$$

=

(b)  $P(8 < X < 14)$

$$= P\left(\frac{8-10}{2.5} < Z < \frac{14-10}{2.5}\right)$$

$$= P(-0.8 < Z < 1.6)$$

$\bar{x} = 37,500 \text{ miles}$   
 $s = 4500 \text{ miles}$

$\rightarrow P(30,000 < z < 40,000)$

$\rightarrow P\left(\frac{30,000 - 37,500}{4500} < z < \frac{40,000 - 37,500}{4500}\right)$

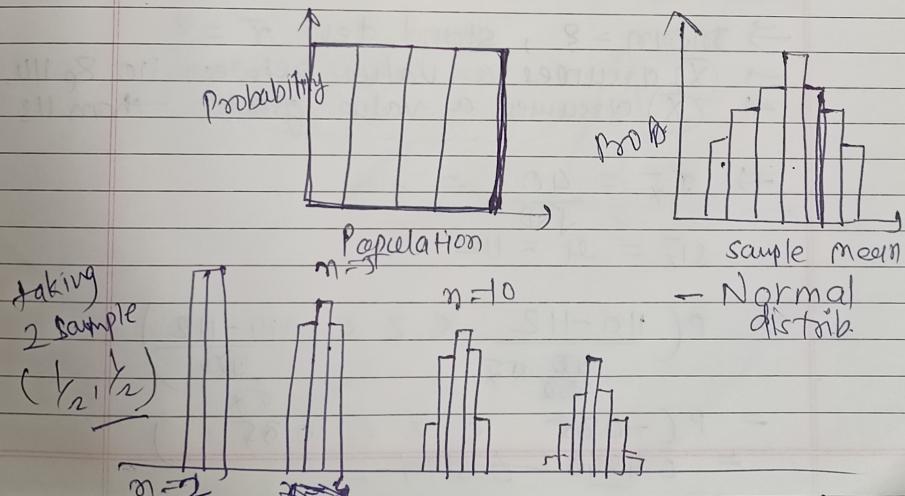
$= P(-1.67 < z < 0.56)$

$\approx 0.605$

$= ( ) - ( )$

$\star$  the sampling distribution of the sample mean :-

$\rightarrow$  The central limit theorem :-



Mean,  $\mu_{\bar{x}} = \mu$   
standard devia  $\sigma_{\bar{x}} = \sigma / \sqrt{n}$

(9 Sept)

Input

- ①  $x$ , the measurement of a single element selected at random from the population, the distribution of  $x$  is the distribution of the population, with mean  $\mu$  and standard deviation the population standard deviation  $\sigma$ ,
- ②  $\bar{x}$ , the mean of the measurements in a sample of size  $n$ , the  $\mu_{\bar{x}} = \mu$   $\sigma_{\bar{x}} = \sigma / \sqrt{n}$ .

Example

$$n = 50$$

$$\mu = 112$$

$$\sigma = 40$$

$\Rightarrow$  mean = ?, stand. devi  $\sigma = ?$

$\Rightarrow$   $\bar{x}$  assumes a value between 110 & 114.

$$\Rightarrow \sigma_{\bar{x}} = \frac{40}{\sqrt{50}} = \sim$$

$$\mu_{\bar{x}} = \mu = 112$$

$$\Rightarrow P\left(\frac{110 - 112}{\frac{40}{\sqrt{50}} \sigma_{\bar{x}}} < z < \frac{114 - 112}{\frac{40}{\sqrt{50}} \sigma_{\bar{x}}}\right)$$

$$= P(-0.85 < z < 0.85) \\ = 0.4840 - 0.3199 =$$

$$\Rightarrow Z > \underline{113}$$

$$P\left(Z > \frac{113 - \bar{x}}{\sigma_x}\right)$$

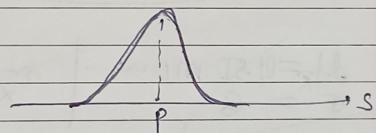
$$\begin{aligned} &\Rightarrow P(Z > 0.18) \\ &\Rightarrow 1 - P(Z < 0.18) \\ &= 1 - 0.5714 \\ &= 0.4286 \end{aligned}$$

### \* Normally Distributed Populations :-

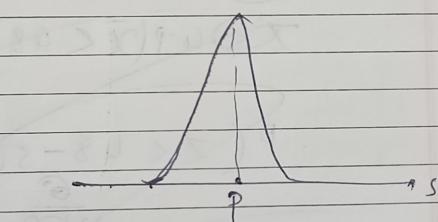
Popul. Distri

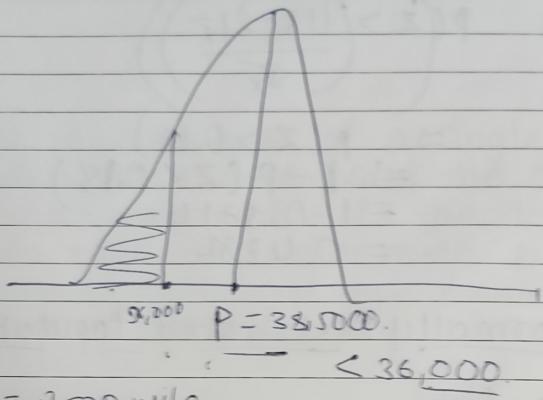


Sampling dis. of  $\bar{x}$   
with  $n = 35$



Sampling  
 $n = 30$





$\mu = 36,000$  miles.

$\sigma = 38,500$

$P(\bar{x} < 36)$

$$P\left(\bar{X} < \frac{\bar{D} - \mu}{\sigma/\sqrt{n}}\right)$$

$\star$

$$\mu_{\bar{x}} = 48 \text{ miles}$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

~~$$P(\bar{x} < 48) = P\left(\bar{z} < \frac{48 - 48}{\sigma/\sqrt{n}}\right)$$~~

~~$$P\left(\bar{z} < \frac{48 - 48}{\sigma/\sqrt{n}}\right)$$~~

$$\star \quad \begin{aligned} \sigma &= SD \\ \sigma &= 6 \end{aligned}$$

$$\textcircled{a} \quad Z(X < 48) = P\left(Z < \frac{48 - 50}{6}\right)$$
$$= P(Z < -0.333)$$
$$= 0.3707$$

$$\textcircled{b} \rightarrow \begin{aligned} n &= 36 \\ Z(\bar{X} < 48) &= P\left(Z < \frac{48 - 50}{\frac{\sigma}{\sqrt{n}}}\right) \\ &= P(Z < -2.00) \\ &= 0.0228 \end{aligned}$$

2 Sept

Page No. \_\_\_\_\_  
Date \_\_\_\_\_

## Hypotheses Testing

(Most of the portion already completed)

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

Sta. Test stats for large sample H.p. Tat w/

If  $\sigma$  is known:  $Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$  a single population mean

If  $\sigma$  is unknown:  $Z = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$

→ The test statistic has the standard normal distribution.

It is hoped that a newly developed pain reliever will more quickly produce perceptual reduction in pain to patients after minor surgeries than a standard pain reliever. The standard pain reliever is known to bring relief in an avg of 3.5 mins with standard deviation 2.1 mins. To test whether the new pain reliever works more quickly than the standard one

H<sub>0</sub>

$H_0: \bar{x} = 3.5$

$H_a: \bar{x} < 3.5 @ \alpha = 0.05$

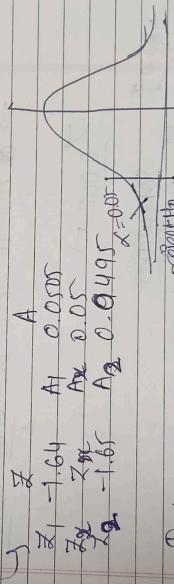
$$Z = \frac{\bar{X} - 3.5}{S/\sqrt{n}}$$

leaving

$$= \frac{3.1 - 3.5}{1.5/\sqrt{50}} = -1.886$$

→ So  $H_a$  has " $<$ " so it's left-tailed test.

$$\rightarrow Z_k = Z_{0.05} = -1.645$$



Interpolation,

$$Z_k = -1.645$$

$$\frac{Z_1 - Z_2}{A_1 - A_2} = \frac{Z_1 - Z_k}{A_1 - A_K}$$

$$\frac{-1.64 - (-1.65)}{0.0505 - 0.0495} = \frac{-1.64 - Z_k}{0.0505 - 0.05}$$

$$\boxed{Z_k = -1.645}$$

Ex 2

$$U = 8.1$$

$$\sigma = 0.22$$

$$n = 30$$

~~$$Z_{0.05}$$~~

$$x = 8.2$$

$$S = 0.25$$

$$\alpha = 1\%$$

$$Z_{0.05} =$$

$$Z_{0.001} = \frac{8.2 - 8.1}{0.25/\sqrt{30}}$$

~~Z<sub>0.05</sub>~~

$$\rightarrow \text{Bcs } H_0 \neq, \alpha = \frac{0.01 - 0.005}{2}$$

$$\underline{Z_{0.005}} \rightarrow$$

6/09

Software

→ Excel solver

→ Minitab

→ Origin Pro,

Solver

File → Op. → Add in → Excel Add in.

then Data in the above bar

& Select

Solver is used for nonlinear regression analysis