28/7/22

Engineering statics

origin and development of statistics :-

→ simply it is known as 'political state'.

Concept of statistical population and a sam

Having car.

20,00,000 = population of int

Prize = Value

200 cars = sample.

avg value = parameter - define characte the popu

200 numbers = Sample data

avg of the data = statistic

Basic Definations:

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

Stastics always gives us approximate not exact value (Drawback)

→ Descriptive statics : took sample and collect the data, and analyzing.

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

→ Statistics is a collection of methods for collecting, displaying, analyzing and drawing conclusions from data.

→ Descriptive st. :

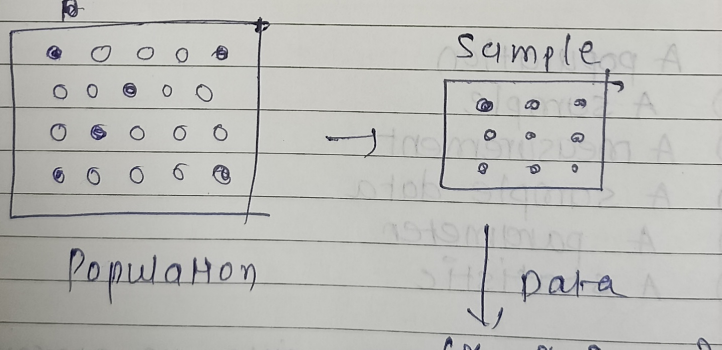
→ Inferential st. : drawing ..... (pictorial)

★ Data : Quantitative and qualitative (categorical)

↓ Prize of the car

|| Brand | Features of the car

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



Par. ← Stats.

μ ← $\bar{x}$

σ^2 ← s^2

r ← s

p ← p

* Importance and scope of statistics :-

→ ~~statics~~ 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 charaterization, performan 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 ~ test of significanc
→ The Indian Meteorological dep - Weather report (forcast).
→ Disaster management.
→ Statistics and psychology and Education.
→ statistics and War.

☆ 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 ~~applied~~ capable of quantitative measurement.

(2) statistics does not study individuals
→ we need group data or ~~can~~ ~~comp~~ comparizon data.

(3) statistical laws are not exact
→ It is predictions of all the patters or ~~dem~~ followed by the demands. [probablity].

(4) statistics is liable to be misused.
→ It may use or interpret wrongly or also be given by wrong person.
→ It ~~s~~ must be used by experts.
methods are the most dangerous tools in the hands of the inexperts.

1/8/'22

☆ 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~~ value having 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.

(1) Nominal scale of measurement:

→ Ex Where do you live?
Options: 1 - Suburbs
2 - City
3 - Town

→ ~~for~~ for ans, we are consideri
it as a non numerical but for
researcher, ~~it~~ is numerical.

→ While capturing nominal data, resea
conduct analysis based on the asso
labels.

→ Nominal data

- Nominal with order
  - → some of N.data can be sub categorised in order.
  - Ex → Much cold, cold, Hot, ...
- Nominal without order
  - → Ex male, Female
- Dichoto
  - → complete opposite the each other.
  - ex yes & N

② Ordinal scale of measurement :

→ as a Variable measurement scale
→ Used to depict non-mathematical ideas such as frequency.

→ Feedback (Websites with emoji feedback
(~~a~~ satisfied are you our services)
(should be in order)
(unsatisfied to satisfied)

→ An ordinal scale of measurement represents an ordered ~~ser~~ series of relationships or rank order.

→ These Values can be added or substracted but cannot multiply or ~~devi~~ divided.

⇒ Relation Between Nominal and ordinal :-

→ 1) The order of Variables is ~~a~~ 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 convinient for researcher.

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 so data does not exist.
- in interval, zero has meaning (0°c) temp

⇒ Data points on the interval scale have same difference between them.

④ Ratio scale of measurement :-

- Includes properties from all four scales of measurement.

- Data in the ratio scale can be added, substracted, divided and multiplied.

Types of Data

- **Quantative**
  - → 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
        - **Interval**
          - Numbers with known diff. between variables, such as time
        - **Ratio**
          - Numbers that have measurable intervals where diff. can be determined, such as hight or weight..
- **Qualitative**
  - → non numerical
    - **Nominal**
      - Naming Variables
    - **Ordinal**
      - The order of Values.

CHAP - 1

?/8/22

☆ 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 descriptio of the entire data in distribution.

⇒ The mean :-

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

$$\boxed{\bar{x} = \frac{\Sigma x}{n}}$$

⇒ Population mean of a set of N population data is the number μ defined by the formula.

$$\mu = \frac{\Sigma x}{N}$$

(5, 17)

+6 \ / −6

11

→ Ex

x	0	1	2	3	4
f	3	6	6	3	5

→ $\bar{x} = \frac{3+6+6+3+5}{5} =$

NAME Dhyey Bhagat
STD. DIV. ROLL NO.
SUBJECT
SCHOOL

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

☆ The median :-

24.8 22.8 24.6 192.5 25.2 18.5 23.7

avg = 47.44 ≈ 47.4

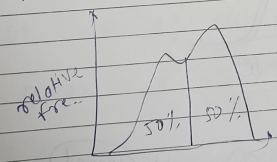
→ Accending arde

18.5 22.8 23.7 24.6 24.8 25.2 192.5

→ If it is even numbers then form of the

$\frac{4+6}{2} = 4.5$ → Median

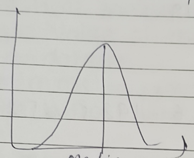
→ Sample median $\tilde{x}$.



Median

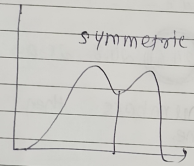
There are four types of cure

Relati
tve



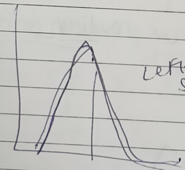
Median
symmetric data

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



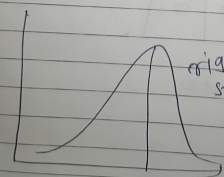
symmetric skewed
sqwead
data

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



Left skewed
sqwead data

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

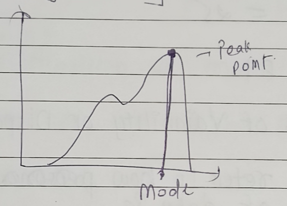


right skewed
~~sqwead~~
data

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

* The mode :-

→ The sample mode of the data that frequently occuring 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

→ Mean $= \bar{x} = \frac{\Sigma x}{N}$

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

$\bar{x} = 41.45$

~~median~~ =

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

median = 25

→ mode = 21

Measures of Vaiability or Dispersion:

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

→ There measures of the variability
- Range → between Highest and lowest
- variance
- standard deviation

The Variance and the standard Deviation

⇒ Sample Variance:

$$s^2 = \frac{\Sigma (x - \bar{x})^2}{n-1}$$

→ by algebra is equivalent to the formula,

$$s^2 = \frac{\Sigma x^2 - \frac{1}{n}(\Sigma x)^2}{n-1}$$

⇒ Sample standard deviation:

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

⇒ If we consider it for population,

$s = \sigma$ & $s^2 = \sigma^2$

(Imp notations)

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

$$\Rightarrow \sigma^2 = \frac{\Sigma (\cancel{\mu}\, x - \mu)^2}{N}$$

⇒ Data set II :

46 37 40 33 42 36 40 47 34 45

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

$\bar{x} = 40$

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

$$s = \sqrt{24.889} = 4.98 \approx 4.99$$

★ Key Points :-

→ The range, the S.D, and the Variance each give a quantitative ans to the que " How variable are the datas

Unit -3 ⇒ Data Analysis

★ Random Variable :

→ Assumes Numerial values associate with the random outcome

→ Each Numerical Value to each variable.

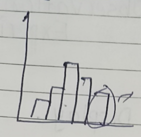
⇒ Types of Random Va.

Discreter	Contineous
→ Nouber of sales	– Time
→ Number of calles	– length
	– Weight

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

→ OBSERVATIONAL STUDY

★ Frequency Histograms :-



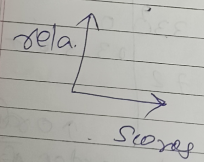
→ S & L dia. is not practical for large data sets, so we can go for this type of diagr...

★ Relative Frequency Histograms :-

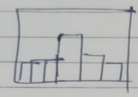
→ Five students scored in the 80s.
→ If there are 30 students in the statics class then,

R.F.D $= \frac{5}{30}$ %.

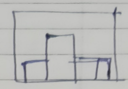
→



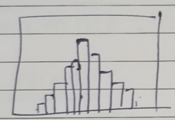
→ When is the size n of a S.



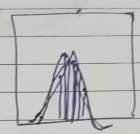
(b) Medium sample



(a) small sample



(c) large sample

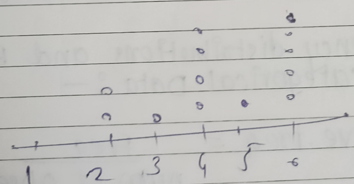


(d) Very Large sample

Ex we find area under the curve.

(It is freq— but with minor size) that's why it is looking ~~size~~ smooth.

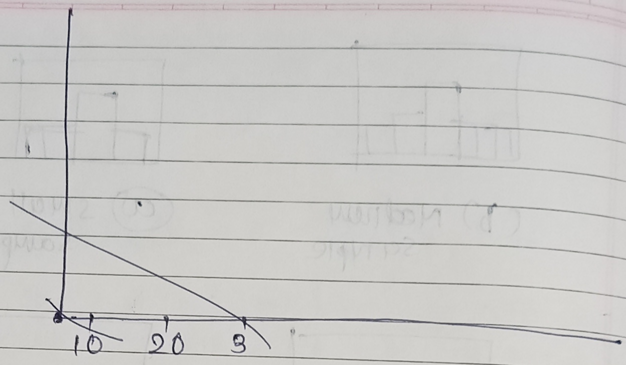
Dotplots for Numerical Data:-



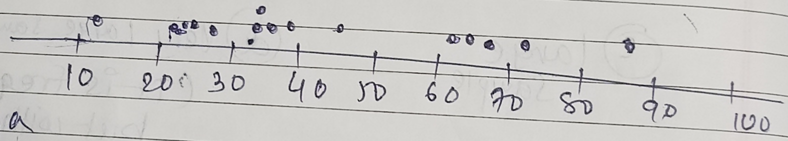
When to Use: Small numerical data sets.

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

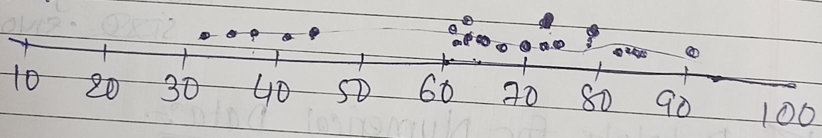
Numerical :



Texas



California

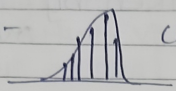


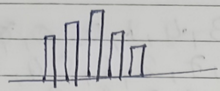
22 aug'22

★ Frequency distributions and Bar charts for categorical Data :-

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

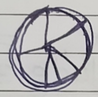
☆ Bar charts :-

Hist -  connected.

Bar 

When to Use : Categorical Data

☆ Pie charts :-

 depending on the area, dividing By 360°...

→ When to Use :- ~~Cato~~ 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.

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

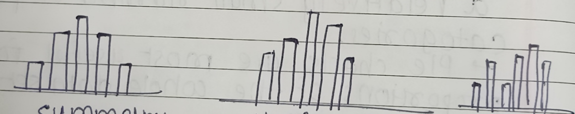
2, 2, 0, 0, 5, 8, 3, 4, 1, 0, 0, ~~1~~, 7, 1, 5, 4, 0, 4, 0, 1, 8, 9, 7, 0, 1, 7, 2, 5, 5, 4, 3, 3, 0, 0, 2, 5, 1, 3, 0, 1, 0, 2, 4, 5, 0, 5, 7, 5, 1.

(1) at least 1 day of leave,
⇒ = 50 - 12 =

(2) Between 3 and 5, ⇒ 5
(we have to include 3, 4, 5)

(3) More than 5 days ⇒ 9 ⇒ 17

⇒



symmetry   Almost symmetry   No symm.

☆ Correlation :-

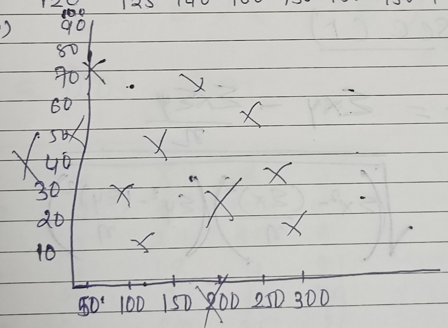
• Finding the relationship between two quantitative variables without being able to infer casual relationships.

☆ Scatter diagram :-

- One variable is called independent (x) and the second is called dependent(y)
- Points are not joined

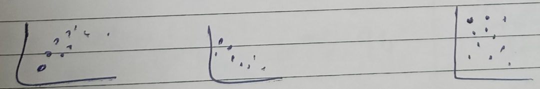
Example:

x-axis ⇒ Wt (kg)	67	69	85	83	74	81	97	92	114	85
y-axis ⇒ SBP (mHg)	120	125	140	160	130	180	150	140	200	130



→ Types of

① Positive ② Nager ③ No relation



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 assosiation.
- The value of ~~1~~ r ranges between (-1) to (+1).

stron Interme weak weak Inter. strong.

-1 -0.75 -0.25 0 0.25 0.75 +1

Perfect correlati. No relati. Perfect corr.

⇒ SCC (r)

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

Example

Age x	Weight y	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

$$r = \frac{461 - \frac{(41)(66)}{6}}{\sqrt{\left(291 - \frac{(41)^2}{6}\right)\left(742 - \frac{(66)^2}{6}\right)}} = \frac{10}{\sqrt{(244.3)(21)}}$$

$= \cancel{0.03}$ 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
$\Sigma x = 32$	$\Sigma y = 32$	129	230	204

$$r = \frac{129 - 170.67}{\sqrt{(59.33)(33.33)}} = -0.94$$

Indirect Strong correlation

Coefficient of correlation :

Pearson correlation (r),

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

$\bar{x}$ = mean of x
$\bar{y}$ =
x_i, y_i = sample

→ b/w -1 to 1

-1 → strong -ve

1 → strong positive

If 0 → No relation

26/08

Partial correlation :-

→ It is a measure of the strength and direction of a linear relationship between two continuous variables whilst ~~for~~ controlling for the effect of one or more other continuous variables. [Also known as 'covariates' or 'control' variable]

Ex 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,

$$(r_{xy.z}) = \frac{(r_{xy} - r_{yz} * r_{zx})}{\sqrt{(1 - r^2_{yz})(1 - r^2_{zx})}}$$

rechecks

x	y	z	$(x_i-\bar{x})$	$(y_i-\bar{y})$	$(z_i-\bar{z})$	$(x_i-\bar{x})(y_i-\bar{y})$	(yz)	(xz)
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	-9.33	-12.44	12
13	51	37	-6	7.11	4	-42.66	-28.44	-24
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	51.1	71.54	140
27	71	43	8	27.11	10	216.88	271.10	80
171	395	297				363.62	373.58	452

$\bar{x} = \frac{171}{9} = 19$

$\bar{y} = \frac{395}{9} = 43.89$

$\bar{z} = \frac{297}{9} = 33$

$r_{xy} = 0.53$

$r_{yz} = 0.34$

$r_{xz} = 0.80$

$(r_{xy.z}) = \cancel{0}.46$

$(x_i-\bar{x})^2$	$(y_i-\bar{y})^2$	$(z_i-\bar{z})^2$
81	221.71	256
36	118.59	100
0	8.35	144
9	9.67	16
36	50.55	16
4	0.01	64
16	166.15	36
100	26.11	196
64	734.95	100
346	1336.09	928

~~r_{xy} =~~

Multiple correlation :-

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

⇒ formula :

$$(R_{x.yz}) = \sqrt{\frac{(r^2_{xz} + r^2_{yz} - 2r_{xy}\, r_{yz}\, r_{zx})}{(1 - r^2_{xy})}}$$

⇒

Subj.	A. A (x)	A (y)	I (z)	$(x_i - \bar{x})$, $\bar{x}=15.3$	$(y_i - \bar{y})$, $\bar{y}=4.8$	$(z_i - \bar{z})$, $\bar{z}=25.\ldots$
1	15	6	25	-0.3	1.2	-0.8
2	18	3	29	2.7	-1.8	3.2
3	13	8	27	-2.3	3.2	1.2
4	14	6	24	-1.3	1.2	-1.8
5	19	2	30	3.7	-2.8	4.2
6	11	3	21	-4.3	-1.8	-4.8
7	17	4	26	1.7	-0.8	0.2
8	20	4	31	4.7	-0.8	5.2
9	10	5	20	-5.3	0.2	-5.8
10	16	7	25	0.7	2.2	-0.8
	153	48	258			

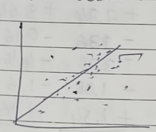
Find Partial ~~&~~ & Multiple correlation

28 aug

Regression Analyses :-

→ Technique concerned ....

→ Regression goes one step beyond correla[tion] 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.

$$\beta_1 = r \times \frac{S_y}{S_x}$$

Regression :-

Correlation and Regression :

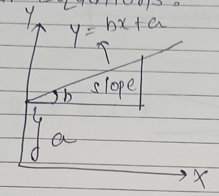
→ Regression minimizes the residuals.

$$\boxed{b_1 = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}}}$$

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

$$\hat{y} = a + bx$$

✗ Linear Equations :-



✗ Numerical

S.No.	Age(x)	Weight(y)	xy	x^2
1	7	12	84	49
2	6	8	48	36
3	8	12	96	64
4	5	10	50	25
5	6	11	66	36
6	9	13	117	81
	41	66	461	291

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

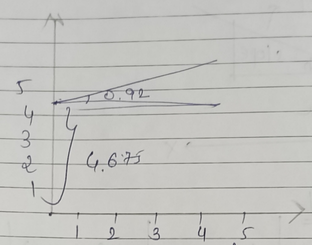
⇒

$$b_1 = \frac{461 - \frac{2706}{6}}{291 - 280.167} = \frac{10}{10.83} = 0.923$$

$$\Rightarrow \cancel{y = bx + a}$$
$$\cancel{= 0.923(8.5) +}$$

$$\Rightarrow \hat{y} = 11 + 0.923\,(8.5 - 6.83)$$
$$= 12.54.$$

$$y_{(x)} = 4.675 + 0.92x$$



② ⟹

Age (x)	B.P (y)	xy	x^2
20	120	2400	400
43	128	5504	1849
63	141	8883	3969
26	126	3276	676
53	134	7102	2809
31	128	3968	961
58	136	7888	3364
46	132	6072	2116
58	140	8120	3364
70	144	10080	4900
46	128	5888	2116
53	136	7208	2809
60	146	8760	3600
20	124	2480	400
63	143	9009	3969
43	130	5590	1849
26	124	3224	676
19	121	2299	361
31	126	3780	961
23	123	2829	529

$\Sigma x = 852$ $\Sigma y = 2630$ $\Sigma xy = 114360$. $\Sigma x^2 = 39230$

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

$$b = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{n}}{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}$$

$x = 25$ years = ?

Hypotheses testing (1 sept / 2021)

→ Types of Hypotheses.

- Null Hy... (H_0)
- Alternative Hy (H_a)
  - → contradictory to the null Hypoth

- Reject H_0 (and therefore accept H_a) or
- Fail to reject H_0 (and therefor fail to accept H_a).

- The null Hypothesis be an assertion containing an equal sign,
  - can have any one of the three forms "<", ">", "≠"

Example (1)

Null Hy. $H_0 : \mu = 127.5$
Alter. $H_a : \mu > 127.5$

(2) $H_0 : \mu = 8$
$H_a : \mu \neq 8$ (Kam bhi Nai Honi chahiye and zyada bhi nai Honi 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 $H_a : \mu < \mu_0$ then reject H_0 if $\bar{x}$ is far to the left of μ_0.
- If H_a " " " $H_a : \mu > \mu_0$ then reject H_0 if $\bar{x}$ is far to the right of μ_0.
- If H_a " " " $H_a : \mu \neq \mu_0$ the reject H_0 if $\bar{x}$ is far away from H_0 in either direction.

2nd Sept

Ex-3

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ | $\mu_{\bar{x}} = \mu = 8.0$

⟹ ~~It~~ in Ex-2, it is known that... the population is normally d... $\sigma = 0.15g$, and suppose that the test of hypotheses $H_0: \mu = 8$ vs $H_a: \mu \neq 8$. with a sample of size 5. Construct the rejection region for the test for the choice $\alpha = 0.10$.

⟹ Test statistics

- t-distribution | student distri..
- z-distribution (Here z-distn)
- chi^2 - distribution
- F-distribution

→ ~~dr~~ $\alpha/2$ is 0.05 & df $= \infty$ (cause not given so consider infinite) (d of freedom)

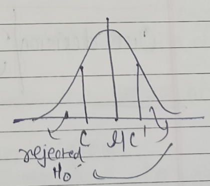
$H_0: \mu = 8.0$

$H_a = \mu \neq 80$

$\boxed{Z_{0.05} = 1.645}$

$\boxed{c = \mu_{\bar{x}} - (1.645)(\sigma_{\bar{x}})}$

$\boxed{c' = \mu_{\bar{x}} + (1.645)(\sigma_{\bar{x}})}$



$c = 7.89$

$c' = 8.11$

→ So, ~~stat~~ state the statement ~~so~~ which ~~st~~ Hypothesis will be accepted.

Statement:

At level of significance $\alpha = 0.01$ & taking a sample of size 5, if the value of μ is between 7.89 and 8.11 will be accepted, or if it is less than 7.89 & more than 8.11, we will be rejecting null hypothesis.

⇒ If H_a has the $\mu \neq \mu_0$ - two tailed test

$\mu < \mu_0$ - left tailed test

$\mu > \mu_0$ - right tailed test

the last two - one tailed test

Two types of errors,

		True state of Nature	
		H_0 is ✓	H_0 is ✗ true
Our decision	Do not reject H_0	Correct decision	Type II error
	Reject H_0	Type I error	correct decision

error is dangerous.

5 Sept

* State the null and Alternate hypothesis

$H_0 : \mu = \mu_0$

(a) ~~H_0:~~ $\mu_0 = 74.5°C$ / $H_0 : \mu = \mu_0$

$H_a : \mu > \mu_0$

(b) $\mu_0 = 145$ Pounds

$H_0 : \mu = \mu_0$

$H_a : \mu > \mu_0$

(c) $\mu_0 = \$ 14,756$

$H_0 : \mu = \mu_0$

$H_a : \mu > \mu_0$

(d) $\mu_0 = \$ 82.53$

$H_0 : \mu = \mu_0$

$H_a : \mu \neq \mu_0$

(e) $\mu = 69.4$

$H_0 : \mu = \cancel{\mu_0}\ 69.4$ acres

$H_a : \mu < 69.4$ acres.

→ (Unit lagane ki zarurat Nai he)

→ Proper method:

Null Hypothesis, $H_0 : \mu = 74.5$

Alternative Hypothesis, $H_a : \mu > 74.5$

a) ~~N.H~~ $H_0: \mu = 38.2$
A.H $H_a: \mu < 38.2$

.H, $H_0: \mu = 58,291$
.H, $H_a: \mu \neq 58,291$

.H, $H_0: \mu = 133$
H, $H_a: \mu > 133$

.H, $H_0: \mu = 161.9$
.H, $H_a: \mu \neq 161.9$

.H, $H_0: \mu = 42.8$
.H, $H_a: \mu > 42.8$

Sampling Distribution

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

The mean and standard Deviation of the sample :-

$\bar{X}$ ~~mean~~ → random variable
$\mu_{\bar{X}}$ = mean^2
$\sigma_{\bar{X}}$ = standard deviation

with the replacement of size two,
2 samples

Probability distribution,
[152, 156, 160, 164]

mean $\bar{x}$ = 2 — Sample

152, 156	152, 160	~~152, 160~~	152, 152
156, 160	160, 152	~~160, 156~~	156, 156
160, 164	152, 164	~~[illegible]~~	160, 160
164, 160	164, 152		164, 164
160, 156	156, 164		
156, 152	164, 156		

→ Sample mean

$$\mu_{\bar{x}} = \Sigma \bar{x} p(\bar{x})$$

$\bar{x}$ =	152	154	156	158	160	162	164
$P(\bar{x})$ =	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{1}{16}$

$$\mu_{\bar{x}} = \left(152 \times \frac{1}{16}\right) + 154\left(\frac{2}{16}\right) + \ldots 164\left(\frac{1}{16}\right)$$

$$\boxed{\mu_{\bar{x}} = 158}$$

$$\sigma_{\bar{x}} = \Sigma \bar{x}^2 (p\bar{x})$$
$$= 24974$$

$$\sigma_{\bar{x}} = \sqrt{\Sigma \bar{x}^2 p(\bar{x}) - \mu_{\bar{x}}^2}$$

$$= \sqrt{10}$$

$\Rightarrow$ Mean $= \dfrac{152 + 156 + 160 + 164}{4}$

$= 158$

→ Population mean = sample mean.

→ standard deviation $= \sqrt{20}$

→ so from values

sample mean $\mu_{\bar{x}} = \mu$ (Population mean) & $\sigma_{\bar{x}}$ (sample standard dev.) $= \dfrac{\sigma}{\sqrt{n}}$ (population standard deviation)

$\bar{X}$ & $\bar{x}$ stands for individual value it takes.

8 sept

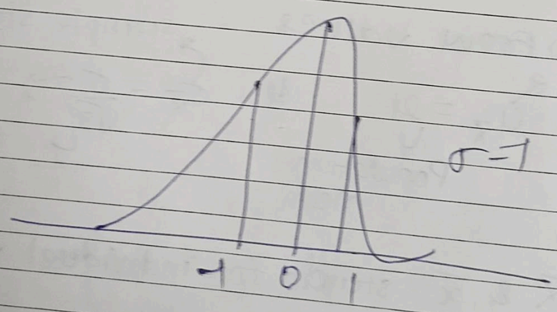
$\mu = \$13,525$
$\sigma = \$4,180$
Samples of size - 100.

→ mean $\mu_{\bar{x}} = ?$
standard deviation $\sigma_{\bar{x}}$?
Sample mean $\bar{x} = ?$

$\mu_{\bar{x}} = \mu = \$13,525$

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

* Standard Normal Random Variable



- Desity curve
- Cumulative Normal probability = ?

★ Find the probabilities ind, where as always z denote 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 $\Rightarrow$ Find the probabilities indicated.

a) $P(Z>1.60)$

b) $P(Z>-1.02)$

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

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

③ a) $P(0.5<Z<1.57)$

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

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

$\Rightarrow P(Z<0.09) - P(Z<-2.55)$

$=$

* Std. Nor. random Var. 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 ∞.

$\Rightarrow$ z is st. No. ra. Var.

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

(a) $P(X < 14)$

$$= P\left(\cancel{\frac{0-10}{2.5}} < Z < \frac{14-10}{2.5}\right)$$

$$= P(\cancel{-4.0} < Z < 1.6)$$

$=$

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

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

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

$\mu = 37,500$ miles
$\sigma = 4500$ mile.

$P(30,000 < Z < 40,000)$

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

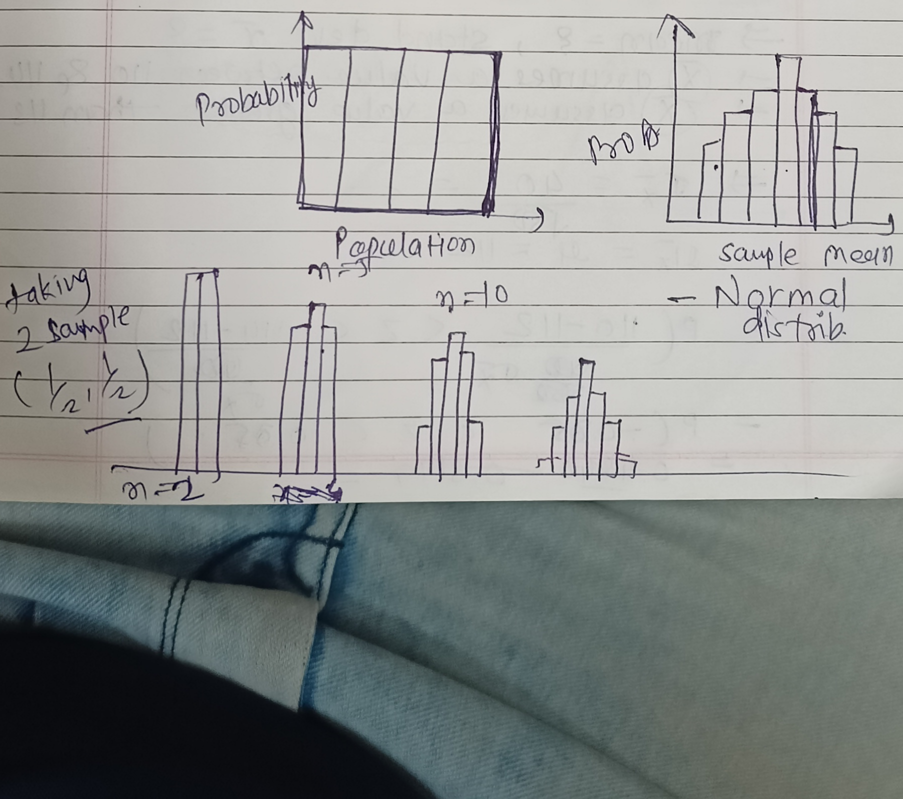
$= P(-1.67 < Z < 0.57)$

$= (\quad) - (\quad)$

$= \_\_\_\_$

The Sampling Distribution of the sample mean :-

⇒ The central Limit theorem :-



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

(9 Sept)

Imp

(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 μ and standard deviation the population standard deviation σ,

(2) $\bar{x}$, the mean of the measurements in a sample of size n, the..
$$\mu_{\bar{x}} = \mu \quad \& \quad \sigma_{\bar{x}} = \sigma/\sqrt{n}.$$

Example
$n = 50$
$\mu = 112$
$\sigma = 40$

⇒ mean = ? , stand. devi $\bar{x}$ = ?
⇒ $\bar{X}$ assumes a value between 110 & 114.
⇒ $\bar{X}$ assumes a value greater than 113

⇒ $\sigma_{\bar{x}} = \frac{40}{\sqrt{50}} =$
$\mu_{\bar{x}} = \mu = 112$

⇒ $P\left(\frac{110-112}{\sigma_{\bar{x}}} < z < \frac{114-112}{\sigma_{\bar{x}}}\right)$
$= P(-0.35 < z < 0.35)$
$= 0.4840 - 0.5199 =$

$\Rightarrow \quad Z > 113$

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

$\Rightarrow \quad P(Z > 0.18)$

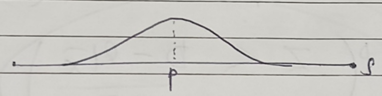
$\Rightarrow 1 - P(Z < 0.18)$

$= 1 - 0.5714$

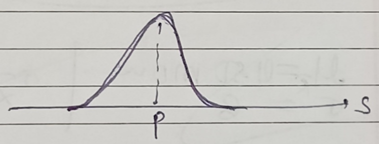
$= 0.4286$

★ Normally Distributed Populations :-

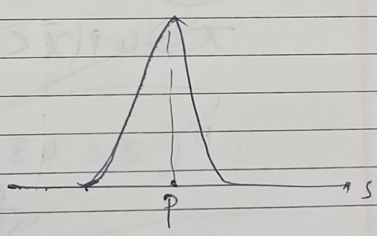
Popu. Distri



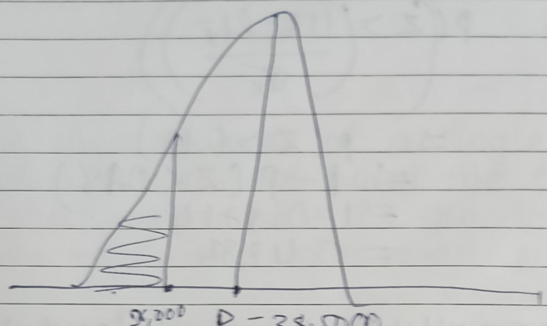
sampling dis. of $\bar{x}$
with $n = 5$



Sampli...
$n = 30$



*



$\mu = 38,5000.$

$< 36,000$

$\sigma = 2500$ mile.
$\mu = 38,500$ $\qquad P(\bar{x} < 36)$

$$P\left(Z < \frac{D - \mu_{\bar{x}}}{\sigma_{\bar{x}}}\right)$$

*

$\mu_{\bar{x}} = 50$ mon. $\quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$
$\sigma = 6$

$$P(\bar{x} < 48) = P\left(Z < \frac{D - \mu_{\bar{x}}}{\sigma_{\bar{x}}}\right)$$

$$P\left(Z < \frac{48 - 50}{\frac{6}{\sqrt{30}}}\right)$$

$\mu = 50$
$\sigma = 6$

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

$= P(Z < -0.333)$
$= 0.3707$

$n = 36$

(b) $\rightarrow$ $Z(\bar{X} < 48) = P\left(Z < \frac{48-50}{\left(\frac{6}{6}\right)}\right)$

$= P(Z < -2.00)$
$= 0.0228$

2 Sept

Hypotheses Testing

(most of the portion already completed)

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

Std. Test stats for large sample Hp. Test concerning a single population mean.

If σ is known : $Z = \dfrac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$

If σ is unknown : $Z = \dfrac{\bar{x}-\mu_0}{S/\sqrt{n}}$

- The test statistic has the standard normal distribution.

Ex It is hoped that a newly developed pain reliever will more quickly produce perceptible 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_0

$H_0 : \mu = 3.5$
vs. $H_a : \mu < 3.5$ @ $\alpha = 0.05$

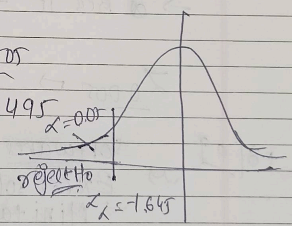
$$z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

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

→ so H_a ~~is~~ has "<" so it's left-tailed test.

→ $z_\alpha = z_{0.05} = -1.645$

	Z		A
z_1	-1.64	A_1	0.0505
z_x	z_x	A_x	0.05
z_2	-1.65	A_2	0.0495



Interpolarity.

$$\frac{z_1 - z_2}{A_1 - A_2} = \frac{z_1 - z_x}{A_1 - A_x}$$

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

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

Ex 2

$\mu = 8.1$
$\sigma = 0.22$
$n = 30$
$\bar{x} = 8.2$
$S = 0.25$
$\alpha = 1\%$

, $H_0 : \mu_0 = 8.1$
$H_a : \mu_0 \neq 8.1$

$$Z_\alpha = Z_{0.01} = \frac{8.2 - 8.1}{0.25/\sqrt{30}}$$

→ Bcs it ≠, $\alpha = \frac{0.01}{2} = 0.005$

$Z_{0.005}$ →

6/09

Softwares
→ Excel solver
→ Mini tab
→ Origin pro,

Solver

Files → Op. → Add in → Excel Ad in.
in the above bar
then Data & Select

Solver is used for nonlinear regression Analysis