statistics Part – 2

Main objective*: *Inferential statistics

will work on sample and estimate on population

we will identify the sample mean based on that we will estimate the population mean

we will identify the sample median based on that we will estimate the population med

we will identify the sample mode based on that we will estimate the population mode

we will identify the sample var based on that we will estimate the population var

statistic

what ever we are identifying on sample is called as statistic

ex: sample mean, sample median called as statistic

parameter

what ever we are estimating on population is called as parameter

ex: population mean, population median called as parameter

how to choose the sample:

assume that in our class 100 *members are there*

$$different\ types\ of\ students\ (Boys - Girls, age, height, weight, areas)$$

$$10\ members$$

idea: based on height

- *age group*
- *areas*
- *gender*
- *language, qualification*
- *randomly*
- *age, gender, lang, qulaification*
- *area*
- 1,11,21,31

Simple random sample: we are just selecting randomly

Systematic random sample : we are slecting randomly but in some order

Stratified random sample: based on age, gender,

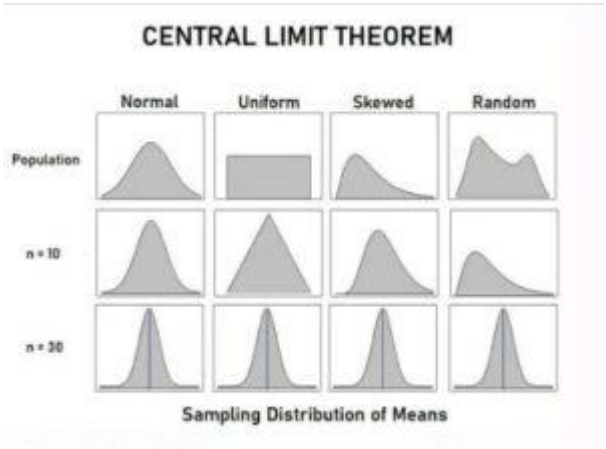
first we will divide into some groups

then we will select it

this groups are called as starta

cluster random sample: Based on geographical (area)

- *assume that we have* 1000 *observations are there*
- *we are drawing a distribution plot*
- *assume that the distribution plot does not follows Normal distribution*
- *we are dividing into* 10 *groups (samples)*
- *each sample group has* 100 *observations*
- 10 *groups are :* $s_1, s_2, s_3, \ldots . s_{10}$
- *now calculate sample group means*
- $\overline{s_1}, \overline{s_2}, \overline{s_3}, \ldots \ldots, \overline{s_{10}}$
- *now plot the distribution plot on sample means*
- *if this sample means we have minimum* 30
- *then the distribution follow Normal distribution*
- *this is called as Central Limit Theorem*



- $Satistics\ part-1: we\ are\ talking\ about\ raw\ observation\ (x_1, x_2 \ldots . x_n)$
  - $statisics\ part-2: we\ will\ talk\ about\ sample\ means(\bar{s_1}, \bar{s_2}, \bar{s_3}, \ldots \ldots, \bar{s_n})$

$part-1: we\ learednd\ mean\ , sd, z\ score$

$here\ also\ we\ need\ to\ calculate\ some\ metrics$

$Mean$

$\boldsymbol{x_1, x_2, x_3, \ldots . x_{1000}} = \boldsymbol{mean1} = \boldsymbol{u}$

$\overline{(\bar{s_1} + \bar{s_2} + \bar{s_3}, \ldots \ldots, + \overline{\bar{s_{10}}})}$/n = mean2

$X =$ 1,2,3,4,5,6,7,8,9,10,11,12

$group-1:$ 1,2,3,4

$\bar{s_1} =$

$group-2:$ 5,6,7,8

$\bar{s_2} =$

$group-3:$ 9,10,11,12

$\bar{s_3} =$

$$\frac{\left[\frac{1+2+3+4}{4} + \frac{5+6+7+8}{4} + \frac{9+10+11+12}{4}\right]}{3} = \frac{1+2+3+4+5+6+7+8+9+10+11+12}{12}$$

Mean same

Standard deviation

how much an observation is deviated fro mean point

how much an sample mean is deviated fro mean point

ex: ameerper hostel students sleeping hour 12 hours average

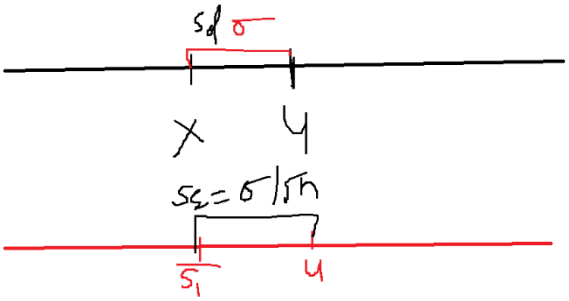
$sachin = 6$

$raj = 14 \quad \frac{14+6}{2} = 10$ ***sample mean***

Standard error

how much an sample mean is deviated fro mean point

$$SE = \frac{\sigma}{\sqrt{n}} \quad where\ n = Number\ of\ observations$$



$z - score$

$$\frac{\bar{x} - u}{\frac{\sigma}{\sqrt{n}}}$$

$\boldsymbol{part-1}$	$\boldsymbol{part-2}$
$\boldsymbol{x}$	$\bar{x}$
$\boldsymbol{u}$	$\boldsymbol{u}$
$\boldsymbol{\sigma}$	$\frac{\boldsymbol{\sigma}}{\sqrt{\boldsymbol{n}}}$
$\frac{x-\boldsymbol{u}}{\boldsymbol{\sigma}}$	$\frac{\bar{x}-\boldsymbol{u}}{\frac{\boldsymbol{\sigma}}{\sqrt{\boldsymbol{n}}}}$