Sampling Theorem:

Statistics part-2:

* Types of sampling
* Point estimates
* Interval estimates
* Central Limit theorem
* Hypothesis testing
* Z-test
* T-test
* F-test
* Chi-square test
* Confidence intervals
* Anova

Generally we works only on samples instead of population

1. Time complexity
2. Money
3. Human resource

When we find mean on samples , we estimate the mean value on population

Inferential statistics

Will work on samples will estimate the population

1 2 3 4 5 6 7 8 ….. 100

I asked you to pick 10 members how you will pick

1. Randomly : picking of 10 ============== > Random sampling
2. Every 10 the person 1, 11, 21,31 ======== > Systematic random sampling
3. Based on gender/ income/age ========= > Cluster sampling
4. Area : geography : Stratified sampling

Point estimates vs Parameter estimate:

We are finding mean on samples : Point estimate

Based on sample mean will try find population means, this population mean is called

Parameter estimate

We are finding <mean/median/mode/std/var> on samples : Point estimate

Based on sample <mean/median/mode/> will try find population <mean/median/mode>, this population <mean/median/mode/> is called Parameter estimate

There are some observation are there

From those observations will try to draw distribution plot , might be Normal or might not be Normal

Suppose the distribution does not follows normality

Then consider some sample sizes from those observations

Example:

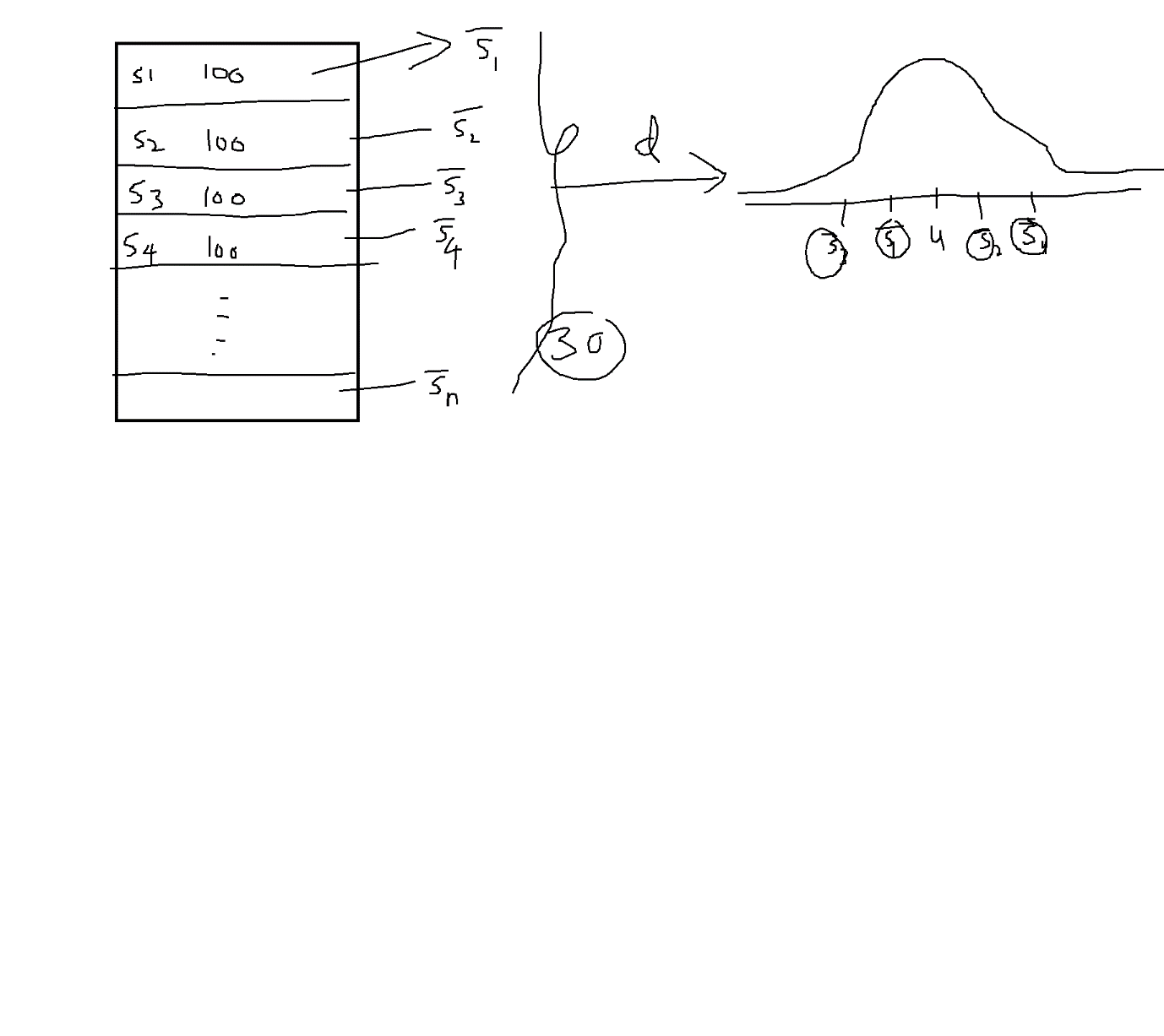
We have 1000 observations

We divided those 1000 observations into 10 parts (sample size)

Each sample size have 100 observations

Now calculate mean of those 10 parts

How many means : 10 means



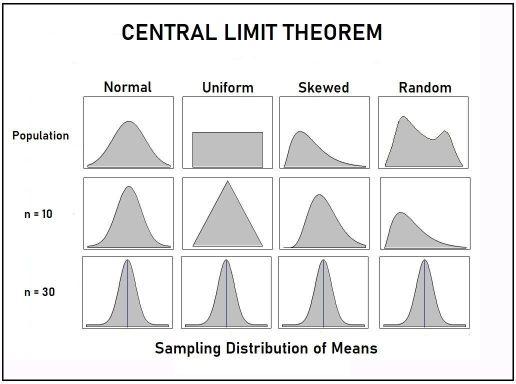
Central Limit Theorem:

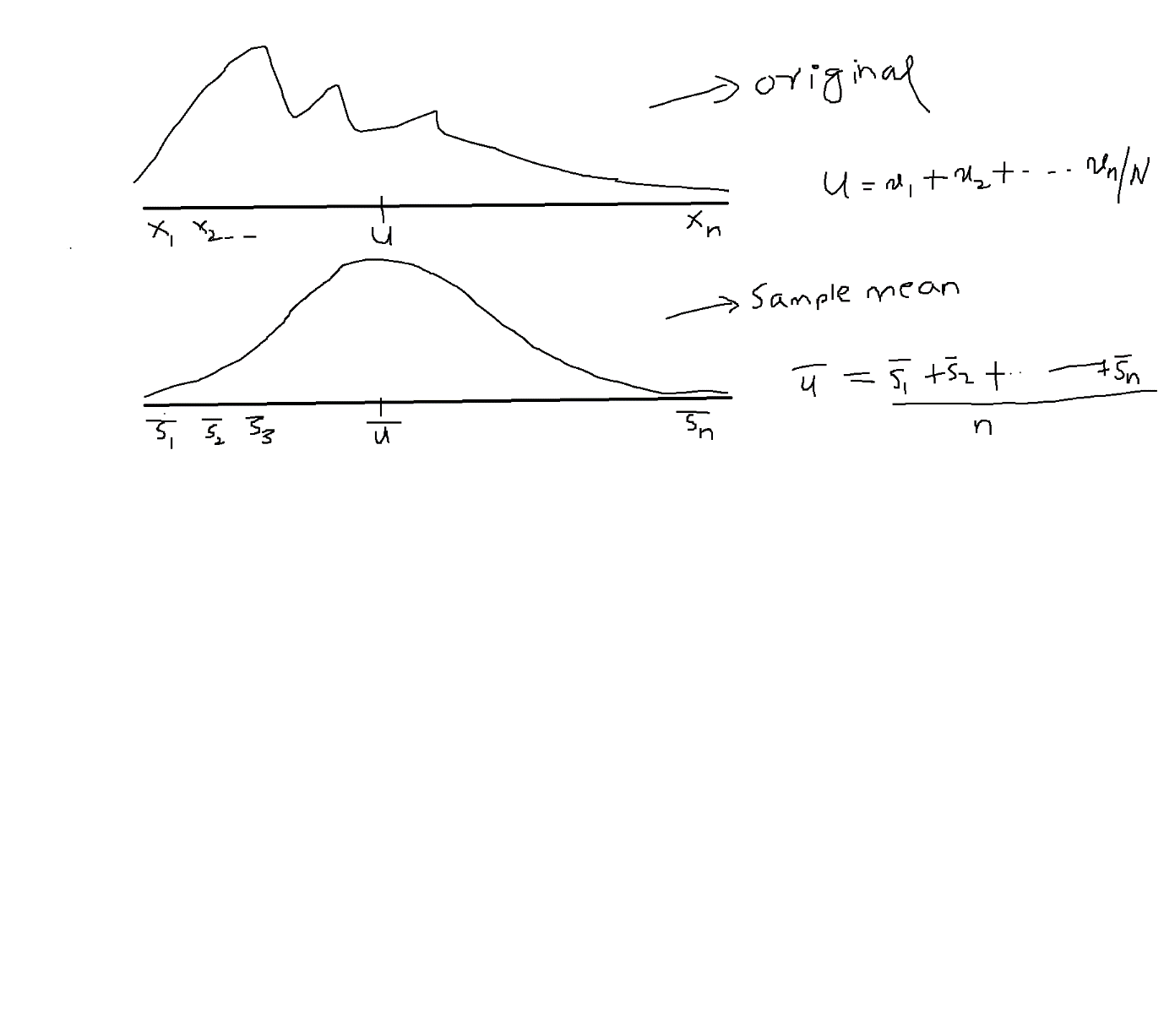
1. If data observations does not follow Normality : triangular/Rectangular
2. Divide data into min 30 parts (sample sizes)

,

1. Calculate mean of those 30 parts

1. Draw the distribution on those 30 sample means : sample distribution of means
2. It follows normal distribution





Original samples: ,

Original mean=

Sample sizes : ,

Means of means : Mean of Sample means= =

Standard error: (SE) (sample means)

How much one sample mean is deviated from mean of sample means

How much one is deviated from

Standard deviation (: (original observations)

How much one observation(x) is deviated from mean of observations (

Observations ============= sample means

Mean of observations ========= Mean of sample means (Mean of Means)

Sd ========================= SE

| Originals | Sample distribution |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

Confidence Interval: CI

Goal:

Find the mean for samples and estimate for population

Find the median for sample and estimate for population

What ever we are finding on samples : Point estimate

Qn)

You are finding mean on sample , and you are estimating population mean

Is both will match or not ?

In a school 10 years old school :

2013 to 2023

2013 first class students avg sleep time:12 PE

All 10 years frst class : 12 hours Parameter estimate

I divided data into 3 parts

Part1 ============= mean ========= 10 == Population mean

Part2 ============= mean ========= 12 == Population mean

Part3 ============ mean ========== 8 == Population mean

Interval estimate:

Point estimates never equal to Population parameters

Sample mean(Point Estimates) never equal to population mean(Parameters)

One sample mean is 8 but population mean is 10

One sample mean is 12 but population mean is 10

8 is single value : point estimate

Based on 8 we made some interval 8+4 and 8-4 : [4,12]

The new interval [4,12] capturing value 10

The new interval based on point estimate is called Interval estimates

point estimate not equal to parameter estimate...so will add interval

How can we choose the interval:

Outlier exist

Q3+ (IQR) Q1- (IQR)

Interval:

PE+ k\*SE

PE-k\*SE

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |