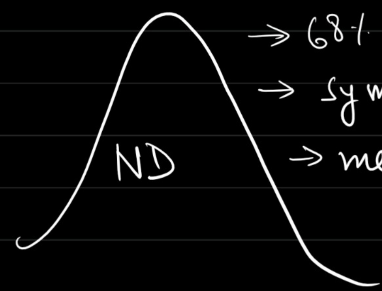
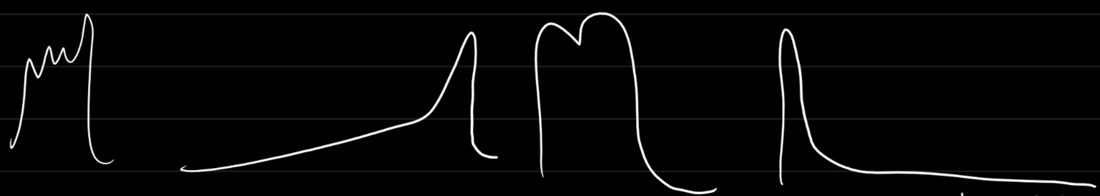
Central limit theorem



→ 68% - 95% - 99.7%
→ symmetrical
→ mean = median = mode



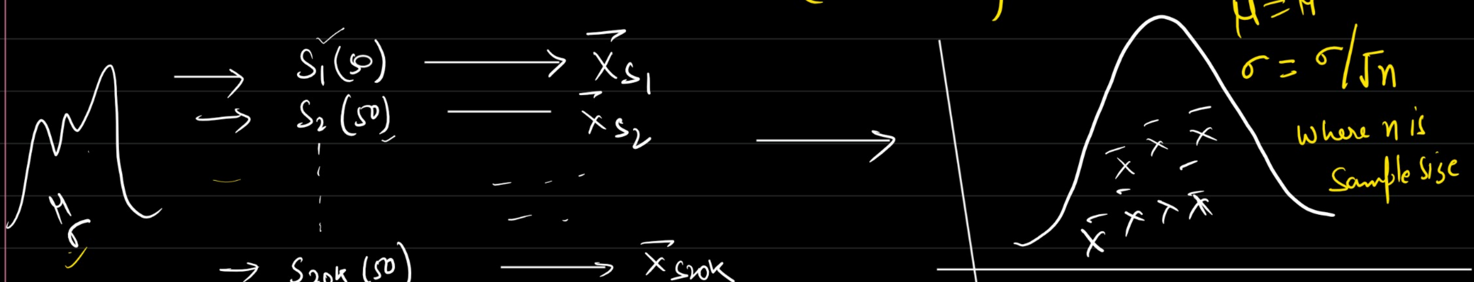
→ No information about irregular | haphazard distribution

* CLT

→ The CLT states that if you have a population with a mean μ and standard deviation σ and take sufficiently large no of random samples from the population with replacement, then the distribution of sample means will be approximately Normally distributed.

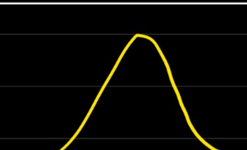
replacement → members can be repeated in samples.

→ Sampling mean of a population (μ, σ) will approximately be a Normal distribution $(\mu, \sigma/\sqrt{n})$



$\mu = \mu$
$\sigma = \sigma/\sqrt{n}$
where n is sample size

CLT → Sampling distribution of mean =



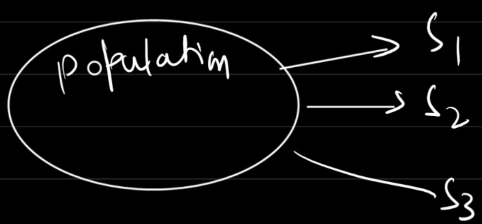
$\mu = \mu$
$\sigma = \sigma/\sqrt{n}$

population (μ, σ) $\xrightarrow[\text{of sample}]{\text{large no}}$ Sampling mean → Plot
↳ ND $\mu = \mu$, $\sigma = \sigma/\sqrt{n}$

Two Conditions of CLT

1. The no of Samples should be large.
2. The sample size should be greater than or equals to 30. (Excep the pop distn which is already a Normal distr)

Standard error $= \sigma / \sqrt{n}$



will the std/var of sample be same as of population?
→ No. In sample you have subset of population

$\frac{\sigma}{\sqrt{n}}$ → Higher the Sample Size, SE will be low

$SE \propto \frac{1}{n}$ SE ↓ n ↑

* Due to Variability/variance (Change in) of sample as compared to population, SE is used.
→ To minimise variability.

Q you have a population with $\mu = 100$ & std dev $\sigma = 20$, If you have sample size 50 from this population. what is prob that sample mean will be less than 105. (Since Sample is mentioned think of CLT)

→ $\mu = 100$, $\sigma = 20$, $n = 50$ $\bar{x} = 105$

$$Z_{score} = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{105 - 100}{20/\sqrt{50}} = \frac{5 \cdot \sqrt{2}}{4}$$

$$= \frac{5 \times 1.414}{4} = 1.7675$$

$Z_{score} = 1.7675$

in terms of prob $\Rightarrow 0.8944$ (from z table)