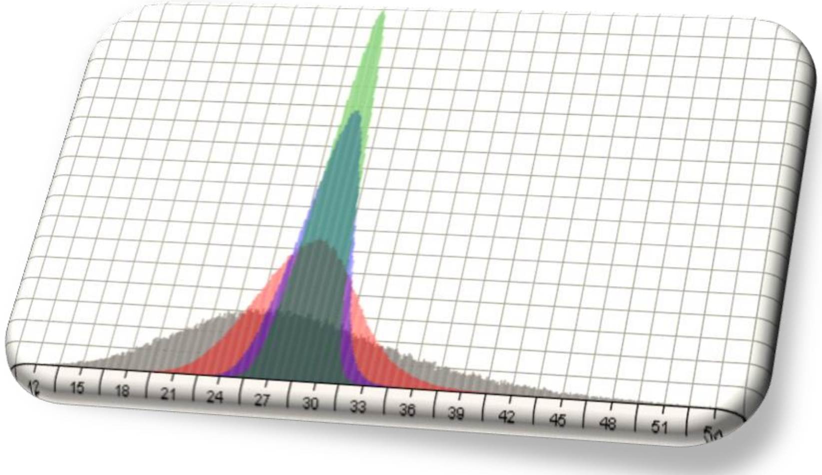
SAMPLING THEORY



Basic Definitions

- **Average(Mean) :** For ungrouped data $\mu = \frac{\sum x_i}{n}$ **or** For grouped data $\frac{\sum f_i \cdot x_i}{\sum f_i}$
- **Standard Deviation:** $\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$
- **Population or Universe :** The group of individuals from which we draw data for a study
- **Sample :** Finite subset of the population
- **Sampling :** The process of selecting a sample from a population
- **Sample size :** The number of individuals in a sample
- **Parameter :** The statistical constants such as mean S.D. of the population
- **Statistics:** The statistical constants such as mean S.D. of the sample

- **Symbols which are used for population and sample:** $\rightarrow$

Population	Sample
Parameter	Statistics
Population size (N)	Sample size (n)
Population Mean (μ)	Sample Mean $(\mu_{\bar{x}})$
Population S.D (σ)	Sample S.D $(\sigma_{\bar{x}}\ \ or\ s)$

- **Sampling distribution:** Let us consider a population of size N and let us draw all possible samples of a given size n. For each of these samples, we compute a statistic (i.e., sample mean, sample variance, sample proportion etc..). The value of the statistic may vary from sample to sample.
- "The probability distribution of values of the statistic for different samples of the same size is called **sampling distribution of the statistic".**
- When we obtain a distribution of mean, it is called Sampling distribution of mean and when we obtain a distribution of proportion, it is called Sampling distribution of proportion
- **Standard error :** The standard deviation of sampling distribution is called the standard error (S.E)

- The relation between mean of the sampling distribution and population mean: $\boldsymbol{\mu_{\bar{x}} = \mu}$
- The relation between variance of the sampling distribution and population variance:
  - **With replacement** $\rightarrow \boldsymbol{\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}}$ **without replacement** $\rightarrow \boldsymbol{\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)}$

Example: Consider a population consisting of four numbers 3, 7, 11, 15. Consider all possible samples of size 2. Find i) mean of the population (μ) ii) S.D. of the population (σ)

iii) mean and variance of the sample distribution ($\mu_{\bar{x}}, \sigma_{\bar{x}}$) with replacement and show that $\mu_{\bar{x}} = \mu$, and $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$ also iv) prove that $\mu_{\bar{x}} = \mu$, and $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$ in the case of without replacement.

- i) Population mean : N=4

$$\mu = \frac{3+7+11+15}{4} = 9$$

- ii) Population S.D. :

$$\sigma^2 = \frac{[(3-9)^2+(7-9)^2+(11-9)^2+(15-9)^2]}{4} = 20$$

$$\sigma = 4.472$$

- All possible samples of size 2 with replacement :{(3,3), (3,7), (3,11), (3,15), (7,3), (7,7), (7,11), (7,15), (11,3), (11,7), (11,11), (11,15)}
- Sample means={3,5,7,9,5,7,9,11,7,9,11,13,9,11,13,15}
- Distribution of the sample means

x	3	5	7	9	11	13	15
f	1	2	3	4	3	2	1

- Mean of the sample means : $\mu_{\bar{x}} = \frac{\sum(f \cdot x)}{\sum f} = 9;$
- $\sigma^2 = \mu_{\bar{x}^2} - (\mu_{\bar{x}})^2 = \frac{\sum(f \cdot x^2)}{\sum f} - (\mu_{\bar{x}})^2 = \frac{1456}{16} - 9^2 = 10$

- All possible samples of size 2 without replacement :{(3,7), (3,11), (3,15), (7,3),(7,11), (7,15), (11,3),(11,7), (11,15),(15,3),(15,7),(15,11)}
- Sample means={5,7,9,11,13} (Example: Mean of (3,7) is (3+7)/2=5)
- Distribution of the sample means

x	**5**	**7**	**9**	**11**	**13**
f	2	2	4	2	2

- Mean of the sample means : $\mu_{\bar{x}} = \frac{\sum(f \cdot x)}{\sum f} = \frac{108}{12} = 9;$
- Variance of the sample distribution : $\sigma^2 = \mu_{\bar{x}^2} - (\mu_{\bar{x}})^2 = \frac{1052}{12} - (9)^2 = \frac{80}{12} = \frac{20}{3}$

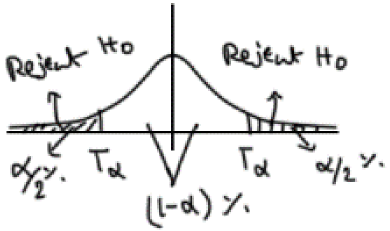
- Mean with or without replacement : $\mu_{\bar{x}} = 9 = \mu,$
- Variance with replacement : $\sigma_{\bar{x}}^2 = \frac{20}{2} = 10 = \frac{\sigma^2}{n}$
- Variance without replacement : $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right) = 10\left(\frac{4-2}{4-1}\right) = \frac{20}{3}$

- **Hypothesis :** An assumption or concession made for the sake of argument
- **Null Hypothesis(H_0):** A tentative assumption is made about the parameter for the sake of testing.
- **Alternative Hypothesis (H_1) :** The hypothesis which is the opposite of what is stated in the null hypothesis.
- **Types of Errors:**

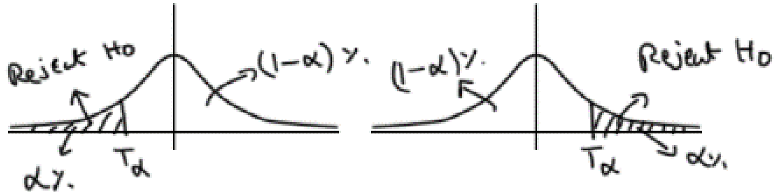
Types of Errors:		**Decision**	
		Accept H_0	**Reject H_0**
Fact	**H_0 is True**	Correct Decision	Type I Error
	H_0 is False	Type II Error	Correct Decision

- **Test statistic** is the statistic based on whose distribution the test is conducted.

- **Critical region :** The test procedure divides the possible values of the test statistic into two regions namely an acceptance region for H_0 and a rejection region for H_0. The region where H_0 is rejected is known as the critical region
- **Level of Significance (LoS) :** The probability of rejecting H_0 when it is true. Usually we take 5% and 1% LoS.
- **One/Two tailed test :** The nature of the critical region depends on the alternative hypothesis H_1
  - For example if $H_0: \mu = \mu_0$ and if $H_1: \mu < \mu_0$ then we use chose the critical region from one-tailed test(left)
    if $H_1: \mu > \mu_0$ then we use chose the critical region from one-tailed test(right)
    if $H_1: \mu \neq \mu_0$ then we use chose the critical region from two-tailed test



Two-tailed Test



One-tailed Test

- **Test procedure**

  The steps in the application of a statistical test procedure for testing a null hypothesis are as follows:
  - Setting up the null hypothesis.
  - Setting up the alternative hypothesis.
  - Identifying the test statistic.
  - Setting a suitable level of significance such as 1% or 5%.(Default 5%)
  - Identifying the critical region.
  - Making decision based on calculated value T and critical value T_α.
  - Accept H_0 if $|T| < |T_\alpha|$ or Reject H_0 if $|T| > |T_\alpha|$

Central Limit Theorem

The sampling distribution of the mean approaches standard normal distribution as size of the sample increases, i.e., standardized sample mean is given by

$$z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$$

Z-test

- For large samples **(n ≥ 30),** most of the sampling distributions tend to normality, and so, the test may be based on normal distribution.
- Critical values for z test and confidence limits for z

Tests	1% Level	5% Level
Two tailed	$\lvert Z_\alpha \rvert = 2.58$	$\lvert Z_\alpha \rvert = 1.96$
Right tailed	Z_α=2.33	$Z_\alpha = 1.645$
Left tailed	$Z_\alpha = -2.33$	$Z_\alpha = -1.645$

95% Confidence limits for μ

$$\boxed{\left(\bar{x} - 1.96\,\frac{\sigma}{\sqrt{n}},\quad \bar{x} + 1.96\,\frac{\sigma}{\sqrt{n}}\right)}$$

99% Confidence limits for μ

$$\boxed{\left(\bar{x} - 2.58\,\frac{\sigma}{\sqrt{n}},\quad \bar{x} + 2.58\,\frac{\sigma}{\sqrt{n}}\right)}$$