CHAPTER # 10

Hashim Farooqi

SAMPLING AND SAMPLING DISTRIBUTION

- **Population**
  A group of all possible elements or objects are called population, for example, Human Population, the total number of students in college. The number of elements involved in population is called size of the population. It is denoted by N.

- **Finite Population**
  A population said to be finite if it consists of a finite or fixed number of elements for example, All university students In Pakistan, the weights of all students enrolled at Punjab University.

- **Infinite Population**
  A population said to be infinite if there is not limit to the number of elements. For example, All heights between 2 and 3 meters.

- **Existent Population**
  A population which consists of concrete objects is called an existent population.

- **Hypothetical Population**
  A population which does not contain concrete objects or items is called hypothetical population.

- **Sample**
  Representation small part of a population is called sample. The number of elements desired in sample is called sample size. It is denoted by n.

- **Sampling**
  Technique of selecting a true sample is called sampling. Sampling is broadly (mostly) distributed into two classes.
  a) Probability or Random Sampling
  b) Non-probability or Non-random Sampling

  a) Probability or Random Sampling
  Technique of sampling where every sampling unit is selected untirely at random, therefore every sampling unit have same chances of selection in the sample, the probability involved in the selection of sampling unit such a technique is called probability sampling.
  Some Important probability samplings are
  1. Simple random sampling
  2. Stratified sampling
  3. Systematic sampling
  4. Cluster sampling
  5. Multistage and Multiphase sampling

  b) Non-probability or Non-random Sampling
  In non probability sampling, the selection of the elements is not base on probability theory but the personal judgment plays a significant role in the selection of the sample the examples of non probability sampling are.
  1. Judgment or Purposive Sampling
  2. Quota Sampling

- **Sampling With Replacement (W.R)**
  Sampling is said to be with replacement if the selected unit is replaced to the population before selecting the next unit, thus sampling unit can be selected more than once. For example, the sampling with replacement is Just like Prize bond scheme.

The number of possible samples of size "n" from a population of size "N" using this technique will be N^n. If we have a population containing 6 elements and like to draw all possible sample of size 2 talking with replacement sampling then number of possible sample will become 36 i.e. 6^2.

- **Sampling Without Replacement (W.O.R)**

Sampling is said to be without replacement if the selected unit is not replaced to the population before selecting the next unit, thus sampling unit can never be selected more than once. For example, the sampling without replacement is Just like Committee System.
The number of possible sample of size "n" from a population of size "N" is obtained by using following formula

No. of Possible samples = $^{N}C_n = \binom{N}{n} = \frac{N!}{(N-n)!n!}$

If for example, we have N=5 and n=2, the no. of possible samples will be 10 i.e.
$^{5}C_2 = \frac{5!}{(5-2)!2!} = 10.$

- **Parameter**

Numerical information or values drawn from population are called parameter. These are fixed numbers. It is usually denoted by Greek or capital letters. For example, population mean μ, and standard deviation σ.

- **Statistic**

Numerical information or values drawn from sample are called statistic. It vary from sample to sample from the same population. It is denoted by Roman or Small letters. For example, sample mean X and sample standard deviation S.

- **Sampling Units**

A basic element or object which we select for a sample are called sampling units. For example, if we want to measure the average height of college students are sampling units.

- **Sampling Frame**

The complete list of all possible sampling units is called a "frame".

- **Census**

Complete enumeration of similar and dissimilar units is termed as census.

- **Sample Survey**

In a sample Survey, enumeration is limited to only a part, or a sample select from the population.

- **Preference to Sample Survey Over Complete Survey**

We prefer sample survey to complete survey due to
1) Reduced cost which we incur on sample
2) Greater speed in presenting the result
3) Greater scope of inquiry
4) Greater accuracy

- **Sampling Error**

The difference b/w parameter and statistic due to small size of sample is called sampling error. It can be reduced by increasing the sample size to a sufficient level.
Sampling Error = $\bar{x} - \mu$

Where $\bar{x}$ = Sample Mean μ = Population Mean

- **Non-Sampling Error**

The non-sampling error is those errors that arise due to defective sampling frame or information not being provided correctly. For example, income, Sale, Production and Age etc are not coated correctly in the most of the cases.

- **Sampling Bias**

Bias is a cumulative component of error which arises due to defective selection of the sample or negligence of the investigator. Errors due to bias increase with an increase in the size of sample.

- **Standard Error**

The standard deviation of a sampling distribution of statistic is called standard error (abbreviated to S.E).

$$S.E(\bar{x}) = \frac{\sigma}{\sqrt{n}}$$

- **Sampling Distribution**

Frequency distribution of statistics from all samples is called sampling distribution. For example, sampling distributions of sample mean or sample distribution of sample variance.

- **Simple Random Sampling**

Technique of sampling where every sampling unit is selected at random from a homogeneous population that every sampling unit have equal chances of selection in sample and every part of population have similar characteristics. For example, Random number table or lottery Method.

- **Stratified Sampling**

When a population has highly variable material, the simple random sampling fails to give accurate results. In this case our population is heterogeneous which is divided into homogenous subgroups called strata. Then a sample is selected separately from each strata at random and the combined into a single sample. This method is called stratified random sampling.

- **Systematic Sampling**

Systematic sampling is a method of selecting a sample that calls for taking every Kth element in the population. The first unit in the sample is selected at random from first 1 to K units the population and the every Kth unit is included in the sample.

- **Cluster Sampling**

Cluster sampling a method of selection a sample in which population is divided into natural groups , such as household , agricultural forms, etc. which are called cluster and taking these clusters as sampling units, a sample is draw at random.

- **Quota Sampling**

Quota sampling is method of selecting a sample of convenience with certain controls to avoid some of the more serious biases involved in talking those most conveniently available. In those method quotas are setup example, by specifying number of interviews from urban and rural, males and females etc.

- **Sampling Distribution**
  Frequency distribution of statistics from all samples is called sampling distribution. For example, sampling distributions of sample mean or sample distribution of sample variance.

Population Size = N

Population = X

Sample Size = n

Population Mean = $\mu = \frac{\sum x}{N}$

Population Variance = $\sigma^2 = \frac{\sum (x-\mu)}{N}$

Population Standard Deviation = $\sigma = \sqrt{\frac{\sum (x-\mu)}{N}}$

Population Proportion = $P = \frac{X}{N}$ Where X is represent the number of even, odd or specific number.

Sample Proportion = $P = \frac{\sum x}{n}$

Sample Mean = $\bar{x} = \frac{\sum x}{n}$

Biased Sample Variance = S^2 = $\frac{\sum (x-\bar{x})}{n}$

Biased Sample Standard Deviation = S = $\sqrt{\frac{\sum (x-\bar{x})}{n}}$

Unbiased Sample Variance = s^2 = $\frac{\sum (x-\bar{x})}{n-1}$

Unbiased Sample Standard Deviation = s = $\sqrt{\frac{\sum (x-\bar{x})}{n-1}}$

Sample Draw with Replacement = N^n

Sample Draw with Out Replacement = ${}^{N}C_n$ = $\binom{N}{n}$ = $\frac{N!}{(N-n)!n!}$

- **Sampling Distribution of Mean**

1) **Mean of the sampling distribution of** $\bar{x}$

$$\mu_{\bar{x}} = E(\bar{x}) = \sum \bar{x} f(\bar{x})$$

2) **Variance of the sampling distribution of** $\bar{x}$

$$\sigma_{\bar{x}}^2 = \left[E(\bar{x})^2 - \left(E(\bar{x}) \right)^2 \right] = \left[\sum \bar{x}^2 f(\bar{x}) - \left(\sum \bar{x} f(\bar{x}) \right)^2 \right]$$

3) $S.E(\bar{x})$ **=Standard Deviation of the sampling distribution of** $\bar{x}$

$$\sigma_{\bar{x}} = \sqrt{E(\bar{x})^2 - \left(E(\bar{x}) \right)^2} = \sqrt{\sum \bar{x}^2 f(\bar{x}) - \left(\sum \bar{x} f(\bar{x}) \right)^2}$$

4) **Population Mean =** $\mu = \dfrac{\sum x}{N}$

5) **Population Variance =** $\sigma^2 = \left[\dfrac{\sum x^2}{N} - \left(\dfrac{\sum x}{N} \right)^2 \right]$

6) **Population Standard deviation =** $\sigma = \sqrt{\dfrac{\sum x^2}{N} - \left(\dfrac{\sum x}{N} \right)^2}$

<u>Verification for With Replacement</u> (W.R)

a. $\mu_{\bar{x}} = \mu$

b. $\sigma_{\bar{x}}^2 = \dfrac{\sigma^2}{n}$

c. $\sigma_{\bar{x}} = S.E(\bar{x}) = \dfrac{\sigma}{\sqrt{n}}$

<u>Verification for With Out Replacement</u> (W.O.R)

a. $\mu_{\bar{x}} = \mu$

b. $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} . \frac{N-n}{N-1}$

c. $\sigma_{\bar{x}} = S.E(\bar{x}) = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$

- **Sampling Distribution of Difference b/w two means** ($\bar{x}_1 - \bar{x}_2$)

1) **Mean of the sampling distribution of** $\bar{x}_1 - \bar{x}_2$

$$\mu_{\bar{x}_1 - \bar{x}_2} = E(\bar{x}_1 - \bar{x}_2) = \sum (\bar{x}_1 - \bar{x}_2) f(\bar{x}_1 - \bar{x}_2)$$

2) **Variance of the sampling distribution of** $\bar{x}_1 - \bar{x}_2$

$$\sigma_{\bar{x}_1 - \bar{x}_2}^2 = \left[E(\bar{x}_1 - \bar{x}_2)^2 - \left(E(\bar{x}_1 - \bar{x}_2) \right)^2 \right] = \left[\sum (\bar{x}_1 - \bar{x}_2)^2 f(\bar{x}_1 - \bar{x}_2) - \left(\sum (\bar{x}_1 - \bar{x}_2) f(\bar{x}_1 - \bar{x}_2) \right)^2 \right]$$

3) $S.E(\bar{x}_1 - \bar{x}_2)$ **=Standard Deviation of the sampling distribution of** $\bar{x}_1 - \bar{x}_2$

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{E(\bar{x}_1 - \bar{x}_2)^2 - \left(E(\bar{x}_1 - \bar{x}_2) \right)^2} = \sqrt{\sum (\bar{x}_1 - \bar{x}_2)^2 f(\bar{x}_1 - \bar{x}_2) - \left(\sum (\bar{x}_1 - \bar{x}_2) f(\bar{x}_1 - \bar{x}_2) \right)^2}$$

4) **Population Mean** $x_1 = \mu_1 = \frac{\sum x_1}{N_1}$

5) **Population Mean** $x_2 = \mu_2 = \frac{\sum x_2}{N_2}$

6) **Population Variance** $x_1 = \sigma_1^2 = \left[\frac{\sum x_1^2}{N_1} - \left(\frac{\sum x_1}{N_1} \right)^2 \right]$

7) **Population Variance** $x_2 = \sigma_2^2 = \left[\frac{\sum x_2^2}{N_2} - \left(\frac{\sum x_2}{N_2} \right)^2 \right]$

8) Population Standard deviation $x_1 = \sigma_1 = \sqrt{\frac{\sum x_1^2}{N_1} - \left(\frac{\sum x_1}{N_1}\right)^2}$

9) Population Standard deviation $x_2 = \sigma_2 = \sqrt{\frac{\sum x_2^2}{N_2} - \left(\frac{\sum x_2}{N_2}\right)^2}$

<u>Verification for With Replacement</u> (W.R)

a. $\mu_{\bar{x}_1-\bar{x}_2} = \mu_1 - \mu_2$

b. $\sigma^2_{\bar{x}_1-\bar{x}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$

c. $\sigma_{\bar{x}_1-\bar{x}_2} = S.E(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

<u>Verification for With Out Replacement</u> (W.O.R)

a. $\mu_{\bar{x}_1-\bar{x}_2} = \mu_1 - \mu_2$

b. $\sigma^2_{\bar{x}_1-\bar{x}_2} = \frac{\sigma_1^2}{n_1}\left(\frac{N_1-n_1}{N_1-1}\right) + \frac{\sigma_2^2}{n_2}\left(\frac{N_2-n_2}{N_2-1}\right)$

c. $\sigma_{\bar{x}_1-\bar{x}_2} = S.E(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{\sigma_1^2}{n_1}\left(\frac{N_1-n_1}{N_1-1}\right) + \frac{\sigma_2^2}{n_2}\left(\frac{N_2-n_2}{N_2-1}\right)}$

- **Sampling Distribution of Sample Proportion** ($\hat{P}$)

1) Mean of the sampling distribution of $\hat{P}$

$$\mu_{\hat{P}} = E(\hat{P}) = \sum \hat{P} f(\hat{P})$$

2) Variance of the sampling distribution of $\hat{P}$

$$\sigma^2_{\hat{P}} = \left[E(\hat{P})^2 - \left(E(\hat{P}) \right)^2 \right] = \left[\sum \hat{P}^2 f(\hat{P}) - \left(\sum \hat{P} f(\hat{P}) \right)^2 \right]$$

3) $S.E(\hat{P})$ **=Standard Deviation of the sampling distribution of** $\hat{P}$

$$\sigma_{\hat{P}} = \sqrt{E(\hat{P})^2 - \left(E(\hat{P}) \right)^2} = \sqrt{\sum \hat{P}^2 f(\hat{P}) - \left(\sum \hat{P} f(\hat{P}) \right)^2}$$

4) Population Mean = $P = \dfrac{X}{N}$ Where X is represent the number of even, odd or specific number.

Verification for With Replacement (W.R)

a. $\mu_{\hat{P}} = P$

b. $\sigma^2_{\hat{P}} = \frac{Pq}{n}$ where $q = 1 - P$

c. $\sigma_{\hat{P}} = S.E(\hat{P}) = \sqrt{\frac{Pq}{n}}$

Verification for With Out Replacement (W.O.R)

a. $\mu_{\hat{P}} = P$

b. $\sigma^2_{\hat{P}} = \frac{Pq}{n} \cdot \frac{N-n}{N-1}$

c. $\sigma_{\hat{P}} = S.E(\hat{P}) = \sqrt{\frac{Pq}{n} \cdot \frac{N-n}{N-1}}$

- **Sampling Distribution of Difference b/w two Proportion ($\hat{P}_1 - \hat{P}_2$)**

1) **Mean of the sampling distribution of** $\hat{P}_1 - \hat{P}_2$

$$\mu_{\hat{P}_1-\hat{P}_2} = E(\hat{P}_1 - \hat{P}_2) = \sum (\hat{P}_1 - \hat{P}_2) f(\hat{P}_1 - \hat{P}_2)$$

2) **Variance of the sampling distribution of** $\hat{P}_1 - \hat{P}_2$

$$\sigma^2_{\hat{P}_1-\hat{P}_2} = \left[E(\hat{P}_1 - \hat{P}_2)^2 - \left(E(\hat{P}_1 - \hat{P}_2) \right)^2 \right] = \left[\sum (\hat{P}_1 - \hat{P}_2)^2 f(\hat{P}_1 - \hat{P}_2) - \left(\sum (\hat{P}_1 - \hat{P}_2) f(\hat{P}_1 - \hat{P}_2) \right)^2 \right]$$

3) $S.E(\hat{P}_1 - \hat{P}_2)$ **=Standard Deviation of the sampling distribution of** $\hat{P}_1 - \hat{P}_2$

$$\sigma_{\hat{P}_1-\hat{P}_2} = \sqrt{E(\hat{P}_1 - \hat{P}_2)^2 - \left(E(\hat{P}_1 - \hat{P}_2) \right)^2} = \sqrt{\sum (\hat{P}_1 - \hat{P}_2)^2 f(\hat{P}_1 - \hat{P}_2) - \left(\sum (\hat{P}_1 - \hat{P}_2) f(\hat{P}_1 - \hat{P}_2) \right)^2}$$

4) **Population Mean** $x_1 = P_1 = \frac{X_1}{N_1}$

Where x_1 is represent the number of even, odd or specific number

5) **Population Mean** x_2 = $P_2 = \dfrac{X_2}{N_2}$

Where x_2 is represent the number of even, odd or specific number

<u>Verification for With Replacement</u> (W.R)

a. $\mu_{\hat{P}_1-\hat{P}_2} = P_1 - P_2$

b. $\sigma^2_{\hat{P}_1-\hat{P}_2} = \dfrac{P_1q_1}{n_1} + \dfrac{P_2q_2}{n_2}$ where $q_1 = 1 - P_1$, $q_2 = 1 - P_2$

c. $\sigma_{\hat{P}_1-\hat{P}_2} = S.E(\hat{P}_1 - \hat{P}_2) = \sqrt{\dfrac{P_1q_1}{n_1} + \dfrac{P_2q_2}{n_2}}$

<u>Verification for With Out Replacement</u> (W.O.R)

a. $\mu_{\hat{P}_1-\hat{P}_2} = P_1 - P_2$

b. $\sigma^2_{\hat{P}_1-\hat{P}_2} = \dfrac{P_1q_1}{n_1}\left(\dfrac{N_1-n_1}{N_1-1}\right) + \dfrac{P_2q_2}{n_2}\left(\dfrac{N_2-n_2}{N_2-1}\right)$

c. $\sigma_{\hat{P}_1-\hat{P}_2} = S.E(\hat{P}_1 - \hat{P}_2) = \sqrt{\dfrac{P_1q_1}{n_1}\left(\dfrac{N_1-n_1}{N_1-1}\right) + \dfrac{P_2q_2}{n_2}\left(\dfrac{N_2-n_2}{N_2-1}\right)}$

- **<u>Sampling Distribution of Biased Variance</u> (S^2)**

1) Mean of the sampling distribution of S^2

$$\mu_{S^2} = E(S^2) = \sum S^2 f(S^2)$$

2) Variance of the sampling distribution of S^2

$$\sigma^2_{S^2} = \left[E(S^2)^2 - \left(E(S^2)\right)^2\right] = \left[\sum (S^2)^2 f(S^2) - \left(\sum S^2 f(S^2)\right)^2\right]$$

3) $S.E(S^2)$ **=Standard Deviation of the sampling distribution of** S^2

$$\sigma_{S^2} = \sqrt{E(S^2)^2 - \left(E(S^2)\right)^2} = \sqrt{\sum (S^2)^2 f(S^2) - \left(\sum S^2 f(S^2)\right)^2}$$

4) Population Mean = $\mu = \frac{\sum x}{N}$

5) Population Variance = $\sigma^2 = \left[\frac{\sum x^2}{N} - \left(\frac{\sum x}{N}\right)^2\right]$

6) Population Standard deviation = $\sigma = \sqrt{\frac{\sum x^2}{N} - \left(\frac{\sum x}{N}\right)^2}$

Verification

$$\mu_{S^2} = E(S^2) \neq \sigma^2$$

- **Sampling Distribution of Un-Biased Variance** **(** s^2 **)**

1) Mean of the sampling distribution of s^2

$\mu_{s^2} = E(s^2) = \sum s^2 f(s^2)$

2) Variance of the sampling distribution of s^2

$$\sigma^2_{s^2} = \left[E(s^2)^2 - \left(E(s^2) \right)^2 \right] = \left[\sum (s^2)^2 f(s^2) - \left(\sum s^2 f(s^2) \right)^2 \right]$$

3) $S.E(s^2)$ **=Standard Deviation of the sampling distribution of** s^2

$$\sigma_{s^2} = \sqrt{E(s^2)^2 - \left(E(s^2) \right)^2} = \sqrt{\sum (s^2)^2 f(s^2) - \left(\sum s^2 f(s^2) \right)^2}$$

4) Population Mean = $\mu = \dfrac{\sum x}{N}$

5) Population Variance = $\sigma^2 = \left[\dfrac{\sum x^2}{N} - \left(\dfrac{\sum x}{N} \right)^2 \right]$

6) Population Standard deviation = $\sigma = \sqrt{\dfrac{\sum x^2}{N} - \left(\dfrac{\sum x}{N} \right)^2}$

<u>Verification</u>

$\mu_{s^2} = E(s^2) = \sigma^2$