Sampling Distribution

1 Population and Samples

Define (i) Population (ii) Sample (iii) Small Sampling
(iv) large Sampling.

Population:- The term Population refers to information of group of observations about which inferences are to be made Population size denote as 'N' represents the number of objects or observations in the population. Population may be finite or infinite depending upon N being finite or infinite.

Example: (i) Engineering Students in Andra Pradesh.
(ii) Budget of India.

Sample: The term Sample refers to a finite subset of the Population. Sample size is represented as 'n' denoting the number of objects or observation in the sample.

Example: 1) Engineering students of PVPSIT College
(2) Budget of Andhra Pradesh.

Small Sampling: The sample comprising of objects less than 30 (ie $n < 30$) it is known an Small Sampling

Large Sampling: The sample comprising of objects which are more than ($n \geq 30$) it is known as large Sampling.

~~Sampling Distribution of Mean (σ known)~~

Sample Mean: If $x_1, x_2 \dots x_n$ represent a random Sample of Size n, then the Sample Mean is defined by the statistic

$$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}$$

Sample Variance :- If $x_1, x_2 \dots x_n$ represent a random of size n, then the Sample Variance is defined by the statistics

$$S^2 \text{ (or) } (\sigma^2) = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}$$

Sample S.D : The sample S.D denoted by s, is the Positive Square root of the Sample Variance.

Sampling Distribution :- The Probability distribution of a statistic is called a Sampling Distribution.

If we draw a Sample of size n from a given finite Population of Size N, then the total number of Possible Samples is NC_n.

$$NC_n = \frac{N!}{n!\,(N-n)!} = k$$

Standard Error :- The standard deviation of the Sampling Distribution of a statistic is known as its Standard Error. and is denoted by (S.E). Thus the standard Error of the sampling distribution of means is Called standard Error of Means. It is used to assess the difference between the Expected and the observed values.

The standard Error of Means = $\frac{\sigma}{\sqrt{n}}$.

Sampling Distribution ~~the~~ of the Mean (σ known)

The Probability distribution of $\bar{x}$ is called the sampling distribution of the Mean. The sampling distribution of a statistic depends on the size of the Population, the size of the Samples, and the Method of Choosing the Samples.

Infinite Population :- suppose the samples are drawn from an infinite Population (or) Sampling is done with replacement, then

The Mean, $\mu_{\bar{x}} = \frac{\mu + \mu + \cdots + \mu}{n} = \mu$.

Variance $\sigma_{\bar{x}}^2 = \frac{\sigma^2 + \sigma^2 + \cdots + \sigma^2}{n} = \sigma^2$

S.D $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

The sampling distribution of $\bar{x}$ will be approximately normal with mean μ and Variance $\frac{\sigma^2}{n}$ Provided that the sample size is large.

Central Limit theorem :- If $\bar{x}$ is the Mean of a random sample of size n taken from a population with Mean μ and finite Variance σ^2, then the limiting form of the distribution of $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ as $n \to \infty$ is the standard normal Distribution.

Finite Population :- Consider a finite Population of size N with Mean μ and S.D σ. Draw all Possible Samples of size n without replacement from this Population.

Then the mean of the Sample distribution of means for $(N > n)$ is

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

The variance is $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$

and S.D is $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$

Here, the factor $\left(\frac{N-n}{N-1}\right)$ often called the finite Population Correction factor.

Normal Population :- Sampling distribution of $\bar{x}$ is normally distributed Even for Small samples of size $n < 30$ Provided sampling is from normal Population.

Non-normal Population :- Consider a Population with unknown (non-normal) distribution. Let the Population mean μ and Population variance σ be the both finite.

Note The Population is finite assume that the Population Size N is at least twice the sample size n. Draw all possible samples of size n. Then the sampling Distribution of $\bar{x}$ is approximately normally distributed with mean $\mu_{\bar{x}} = \mu$ and variance $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$ Provided the Sample Size is large (i.e $n \geq 30$)

Sampling Distributions of differences & Sums

Let μ_{S_1} & σ_{S_1} be the Mean & S.D of a Sampling distribution of statistic S_1 obtained by calculating S_1 for all possible samples of size n_1 drawn from Population 'A'. μ_{S_2} & σ_{S_2} be the Mean & S.D of Sampling distribution of statistic S_2 obtained by Calculating S_2 for all possible Samples of size n_2 drawn from another different Population B'.

A distribution of the differences $S_1 - S_2$ Called the Sampling distribution of differences of the statistics from the two Population A & B.

The Mean $\mu_{S_1-S_2}$ & the S.D $\sigma_{S_1-S_2}$ of the Sampling distribution of differences are given by

$$\mu_{S_1-S_2} = \mu_{S_1} - \mu_{S_2}$$

& $\sigma_{S_1-S_2} = \sqrt{\sigma_{S_1}^2 + \sigma_{S_2}^2}$ assuming that the Samples are independent.

Sampling distribution of sum of statistics has mean $\mu_{S_1+S_2}$ & S.D $\sigma_{S_1+S_2}$ given by $\mu_{S_1+S_2} = \mu_{S_1} + \mu_{S_2}$

& $\sigma_{S_1+S_2} = \sqrt{\sigma_{S_1}^2 + \sigma_{S_2}^2}$ assuming the Samples are independent

For infinite Population the Sampling distribution of sums of means by Mean $\mu_{\bar{x}_1+\bar{x}_2}$ & $\sigma_{\bar{x}_1+\bar{x}_2}$ given by

$$\mu_{\bar{x}_1+\bar{x}_2} = \mu_{\bar{x}_1} + \mu_{\bar{x}_2} = \mu_1 + \mu_2$$

$$\sigma_{\bar{x}_1+\bar{x}_2} = \sqrt{\sigma_{\bar{x}_1}^2 + \sigma_{\bar{x}_2}^2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Prob. Find the value of finite Population Correction factor for $n=10$ and $N=100$

Sol: Given $N =$ The size of the finite Population $= 1000$
$n =$ the size of the Sample $= 10$

$$\text{Correction factor} = \frac{N-n}{N-1} = \frac{1000-10}{1000-1} = \frac{990}{999} = 0.991$$

Prob: How many different Samples of size 2 can be chosen, from a finite Population of Size 25.

Sol: we can take $^{N}C_{n}$ Samples of size n from the Population of Size N.

Here $N=25$, $n=2$

we can take $^{25}C_{2} = 300$ Samples of size 2 from finite Population of Size 25.

Prob: A Population Consists of five numbers 2, 3, 6, 8, and 11. Consider all possible samples of size 2 that can be drawn with replacement from this Population find

(a) The Mean of the Population (b) The S.D of the Population

(c) the Mean of the Sampling distribution of Means and

(d) the S.D of the Sampling distribution of Means (i.e. The SE of Means)

(a) Mean of the Population

$$\mu = \frac{2+3+6+8+11}{5} = \frac{30}{5} = 6$$

(b) Variance of the Population (σ^2)

$$\sigma^2 = \Sigma \frac{(x_i - \bar{x})^2}{n}$$

$$= \frac{(2-6)^2 + (3-6)^2 + (6-6)^2 + (8-6)^2 + (11-6)^2}{5} = 10.8$$

For Sampling distribution of differences of Proportions.

$$\mu_{P_1-P_2} = \mu_{P_1} - \mu_{P_2} = P_1 - P_2$$

$$\sigma_{P_1-P_2} = \sqrt{\sigma_{P_1}^2 + \sigma_{P_2}^2} = \sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}$$

Prob: 3 Masses are measured as 62.34, 20.48, 35.97 kgs with S.D 0.54, 0.21, 0.46 kgs. Find the Mean & S.D of the Sum of the masses.

Let the 3 masses measured be $\bar{x}_1, \bar{x}_2, \bar{x}_3$.

The mean of the sum of the masses is

$$\mu_{\bar{x}_1+\bar{x}_2+\bar{x}_3} = \mu_{\bar{x}_1} + \mu_{\bar{x}_2} + \mu_{\bar{x}_3}$$
$$= 62.34 + 20.48 + 35.97$$
$$= 118.79.$$

$$\sigma_{\bar{x}_1+\bar{x}_2+\bar{x}_3} = \sqrt{\sigma_{\bar{x}_1}^2 + \sigma_{\bar{x}_2}^2 + \sigma_{\bar{x}_3}^2} = \sqrt{(0.54)^2 + (0.21)^2 + (0.46)^2}$$
$$= 0.74$$

Prob. Let $u_1 = (3, 7, 8)$ $u_2 = (2, 4)$. Find (a) μ_{u_1} (b) μ_{u_2} (c) Mean of the sampling distribution of the difference of means $\mu_{u_1-u_2}$ (d) σ_{u_1} (e) σ_{u_2} (f) the S.D of sampling distribution of the differences of means $\sigma_{u_1-u_2}$

$$\mu_{u_1} = \frac{3+7+8}{3} = \frac{18}{3} = 6$$

$$\mu_{u_2} = \frac{2+4}{2} = \frac{6}{2} = 3$$

$u_1 - u_2 = (1, -1, 5, 3, 6, 4)$ [i.e 3-2, 3-4, 7-2, 7-4, 8-2, 8-4]

$$\mu_{u_1-u_2} = \frac{1-1+5+3+6+4}{6} = \frac{18}{6} = 3$$

$$\sigma_{u_1}^2 = \sum_{i=1}^{N} \frac{(x_i - \mu_{u_1})^2}{N}$$

$$= \frac{(3-6)^2 + (7-6)^2 + (8-6)^2}{3}$$

$$= \frac{9+1+4}{3} = \frac{14}{3} = 4.667$$

$$\boxed{\sigma_{u_1} = 2.16}$$

$$\sigma_{u_2}^2 = \sum_{i=1}^{\infty} \frac{(x_i - \mu_{u_2})^2}{N}$$

$$= \frac{(2-3)^2 + (4-3)^2}{2} = \frac{1+1}{2} = \frac{2}{2} = 1$$

$$\sigma_{u_2} = 1$$

$$\sigma_{u_1-u_2}^2 = \sum_{i=1}^{N} \frac{(x_i - \mu_{u_1-u_2})^2}{N}$$

$$= \frac{(1-3)^2 + (-1-3)^2 + (5-3)^2 + (3-3)^2 + (6-3)^2 + (4-3)^2}{6}$$

$$= \frac{34}{6} = 5.667$$

$$\sigma_{u_1-u_2} = \sqrt{5.667} = 32.11 \times$$

(2) Let $u_1 = \{2,7,9\}$ $u_2 = \{3,8\}$ find

(i) μ_{u_1}, (ii) μ_{u_2} (iii) $\mu_{u_1+u_2}$ (iv) $\mu_{u_1-u_2}$, (v) σ_{u_1}, (vi) σ_{u_2}.

(vii) $\sigma_{u_1+u_2}$, (viii) $\sigma_{u_1-u_2}$ verify that $\mu_{u_1+u_2} = \mu_{u_1} + \mu_{u_2}$

& $\mu_{u_1-u_2} = \mu_{u_1} - \mu_{u_2}$

Sol. $u_1 = \{2,7,9\}$ $u_2 = \{3,8\}$

$$\mu_{u_1} = \frac{2+7+9}{3} = 6, \quad \mu_{u_2} = \frac{3+8}{2} = \frac{11}{2} = 5.5$$

$$u_1 + u_2 = \{5, 10, 10, 15, 12, 17\}$$

$$\mu_{u_1+u_2} = \frac{5+10+10+15+12+17}{6} = \frac{69}{6} = 11.5.$$

$u_1 - u_2 = \{-1, -6, 4, -1, 6, 1\}$

$\mu_{u_1-u_2} = \frac{-1-6+4-1+6+1}{6} = \frac{3}{6} = \frac{1}{2} = 0.5$

$\sigma_{u_1}^2 = \sum_{i=1}^{n} \frac{(x_i - \mu_{u_1})^2}{N} = \frac{(2-6)^2 + (7-6)^2 + (9-6)^2}{3}$

$= \frac{16+1+9}{3}$

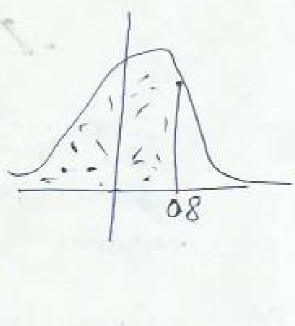
$= \frac{26}{3}$

Prob 3. Let $u_1 = \{1, 2, 3\}$ $u_2 = \{4, 5, 6\}$. find μ_1, μ_2, Mean of the Sampling distribution of the differences of Means $\mu_{u_1-u_2}$, σ_{u_1}, σ_{u_2}. S.D of the Sampling distribution of differences of Means $\sigma_{u_1-u_2}$.

Prob: The Mean height of students in a College is 155 cms and S.D is 15. what is the Probability that the Mean height of 36 students is less than 157 cms.

Sol. μ = Mean of the Population.
= Mean heights of the students of a College = 155 cms.
σ = S.D of Population = 15 cm.
n = Sample size = 36
$\bar{x}$ = Mean of Sample = 157 cms.

$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{157-155}{15/\sqrt{36}} = 0.8$



$P(\bar{x} < 157) = P(z < 0.8) = 0.5 + P(0 \le z \le 0.8)$

$= 0.5 + 0.2881$

$= 0.7881$

Thus the Probability that the Mean of height 36 students is less than 157 = 0.7881

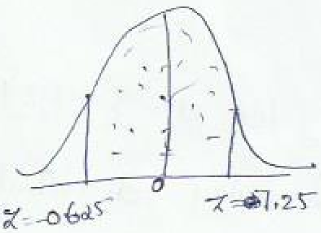
Prob. A random Sample of size 100 is taken from an infinite Population having the Mean $m = 76$ and the Variance $S^2 = 256$. what is the Probability that $\bar{x}$ will be between 75 & 78.

n = Sample size = 100

μ = Mean of the Population = 76

σ^2 = variance of the Population = 256

$\sigma = 16$.



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

If $\bar{x}_1 = 75$

$$z_1 = \frac{75-76}{16/\sqrt{100}} = -0.625$$

If $\bar{x}_2 = 78$ then $z_2 = \dfrac{\bar{x}_2 - \mu}{\sigma/\sqrt{n}} = \dfrac{78-76}{16/\sqrt{100}} = 1.25$

$$P(75 \le \bar{x} \le 78) = P(z_1 \le z \le z_2)$$
$$= P(-0.625 \le z \le 1.25)$$
$$= P(-0.625 \le z \le 0) + (0 \le z \le 1.25)$$
$$= 0.2334 + 0.3944$$
$$= 0.628.$$

Prob. A normal Population has a Mean 0.1 and S.D of 2.1 Find the Probability that Mean of Sample of size 900 will be negative.

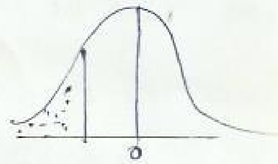
Sol. Given $\mu = 0.1$, $\sigma = 2.1$ and $n = 900$

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{\bar{x} - 0.1}{2.1/\sqrt{900}}$$

$$z = \frac{\bar{x}-0.1}{0.07}$$

$\bar{x} = 0.1+(0.07)z$ where $z \sim N(0,1)$

The required Probability that the Sample mean is negative is given by



$$P(\bar{x}<0) = P(0.1+0.07z<0)$$

$$= P(0.07z<-0.1)$$

$$P\left(z<\frac{-0.1}{0.07}\right)$$

$$P(z<-1.43)$$

$$= 0.5 - P(0<z<1.43)$$

$$= 0.5 - 0.4236$$

$$= 0.0764.$$

Prob. Samples of size 2 are taken from the Population 4, 8, 12, 16, 20, 24 without replacement. Find

(a) Mean of the Population

(b) Standard Deviation of Population

(c) The Mean of Sampling distribution of Means.

(d) S.D of Sampling distribution of Means.

Prob: Determine the Mean and S.D of Sampling distributions of variances for the Population 3, 7, 11, 15 with $n=2$ and the Sampling is with replacement.

... of the Population is 3, 6, 9, 15, 27

(i) List all possible samples of size 3 that can be taken without replacement from the finite population.

(ii) Calculate the mean of each of the sampling distribution of means

(iii) Find the S.D of sampling distribution of means.

Sol: The mean of the population is

$$\mu = \frac{3+6+9+15+27}{5} = \frac{60}{5} = 12$$

The SD of the population

$$\sigma = \sqrt{\frac{(3-12)^2+(6-12)^2+(9-12)^2+(15-12)^2+(27-12)^2}{5}}$$

$$= \sqrt{\frac{81+36+9+9+225}{5}} = \sqrt{72}$$

$$= 8.485$$

(a) Sampling without replacement (finite population) the total number of samples without replacement is, $^{N}C_n = {}^5C_3 = 10$.

(3,6,9) (3,6,15) (3,6,27) (3,9,15) (3,9,27)
(3,6,9) (3,6,15) (3,6,27) (6,15,27) (9,15,27)
(3,15,27) (6,9,15) (6,9,27)

Arithmetic mean for each of these samples

6	8	12	9	13
15	10	14	16	17

(b) Mean of sampling distribution of mean is

$$\mu_{\bar{x}} = \frac{6+8+12+9+13+15+10+14+16+17}{10}$$

$$\mu_{\bar{x}} = \frac{120}{10} = 12$$

(c) S.D of sampling distribution of means

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

$\sigma = 8.485$ $n=3$, $N=5$

$$\sigma_{\bar{x}} = \frac{8.485}{\sqrt{3}}\sqrt{\frac{5-3}{5-1}}$$

$$\sigma_{\bar{x}} = 3.46$$

Estimate :- An estimate is a statement made to find an unknown Population Parameter.

Estimator :- The Procedure or rule to determine an unknown Population Parameter is Called an Estimator.

Estimators are two types.

(a) Point estimation

(b) Interval estimation.

Statistical Estimation :- It is a Part of statistical inference where a Population Parameter is estimated from the Corresponding Sample statistics.

Point Estimation :- A Point estimation of a Parameter is a statistical estimation where the Parameter is estimated by a Single numerical value from sample data.

Definition :- A Point estimate of a Parameter θ is a single numerical value, which is Computed from a given Sample and serves as an approximation of the unknown Exact Value of the Parameter.

Definition :- A Point estimator is a statistic for estimating the Population Parameter θ and will be denoted by $\hat{\theta}$.

Properties of Estimation :- An estimator is not Expected to estimate the Population Parameter without error. An estimator should be close to the true value of unknown Parameter.