Sampling and Estimation

Population and Sample
simple random Sampling (with replacement &
without replacement)

Sampling distributions of means (σ known)
Sampling distributions of mean (σ unknown)
$\hookrightarrow$ t distribution

Sampling distributions of Variance (σ unknown)
$\rightarrow \chi^2$ distribution

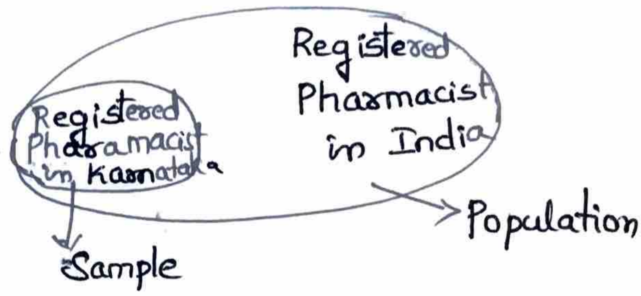
Estimation - Maximum Likelihood Estimation.

SAMPLING AND ESTIMATION

POPULATION AND SAMPLING

Suppose we want to study and analyse data concerning a large group of individuals or objects, such as heights of men in a city, no of defective bolts produced in a factory in a given week, marks in mathematics of all students who appeared for a university examination in a certain year, such entire group is called the population or universe.

Small part of this group is sample.



Any selection of individuals grouped together by a common feature can be said as Population

Population

- **Finite Population**
  Ex:- No of students in our College
- **Infinite Population**
  Ex:- No of germs in the patient's body
- **Existent Population**
  use unit is available in solid form
  Ex:- Books, Students
- **Hypothetical Population**
  whose unit is not available in solid form
  Ex:- out come of tossing a coin

Sample

* Significant portion of a population, not an entire population.

Sample

- Small Sample size (less than 30 Sample size)
  - Student's t Test
  - χ^2 test
  - F Test
- Large Sample Size 30 or more than 30 Sample Size
  - Large Sample Test
  - Z-test

$\boxed{\text{Sampling}}$ ⇒ Process of learning about the population

- Probability or Random Sampling Methods
  - Simple Random Sampling
    - Lottery Method
    - Table of Random no
  - Restricted Random Sampling
- Non-Probability or Non-Random Sampling Methods
  - Judgement Sampling
  - Quota Sampling
  - Convenience Sampling

The process of sampling involves 3 elements

* Selecting the Sample
* Collecting the information
* Making an inference about the population

Sampling without replacement | With replacement.

Box contains 6 Red, 4 Green, 5 - Blue } 15 marbles in Box

Say I pick 1 Blue marble &

$P(B) = 5/15$.

No I dont put back the Blue marble =

$P(R) = 6/14$, $P(B) = 5/13$

$$P(\text{Getting a red, blue, green without replacement}) = \frac{6}{15} \times \frac{4}{14} \times \frac{5}{13}$$

$$\text{With replacement} = \frac{6}{15} \times \frac{4}{15} \times \frac{5}{15}.$$

SAMPLING DISTRIBUTION OF MEANS

Suppose we draw all possible samples of certain size N from a population and find the mean $\bar{X}$ of each of these samples. The frequency distribution of these means is called the sampling distribution of means.

Let the population be finite with size N_P. Let μ and σ be its mean & standard deviation respectively. Then for the sampling distribution of means the mean & the S.D [or standard error] denoted by $\mu_{\bar{X}}$ and $\sigma_{\bar{X}}$ are given by

$$\boxed{\mu_{\bar{X}} = \mu} \quad \text{——— (1)}$$

$$\boxed{\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{N}}\sqrt{\frac{N_P - N}{N_P - 1}}} \quad \text{——— (2)}$$

If the population is infinite (or if the sampling is with replacement), then

$$\boxed{\mu_{\bar{X}} = \mu} \quad \text{——— (3)} \quad \because \frac{N_P - N}{N_P - 1} \to 1$$

$$\boxed{\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{N}}} \quad \text{——— (4)} \quad \text{if } N_P \to \infty$$

In a data concerned with finite population, the size of the population has to be specified. If the size of the population is not explicitly mentioned, we take that the population is infinite.

It can be proved that for large values of N (≥ 30), the sampling distribution of mean is approximately a Normal distribution for which the sample mean $\bar{X}$ is the random variable. If the population itself is normally distributed, the sampling distribution of means is a normal distribution even for small values of N (<30). Accordingly, the standard normal variate for the distribution of mean is given by

$$\boxed{Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}}}$$

$\mu_{\bar{X}} = \mu$ in the cases of both finite & infinite populations,

$$\boxed{Z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}}}$$

If the population is finite with its size N_p known before hand, $\sigma_{\bar{X}}$ is computed by using formula (2). In all other cases, $\sigma_{\bar{X}}$ is computed by using formula (4)

PROBLEMS

1. A population consists of the four numbers 3, 7, 11, 15. Consider all possible samples of size two which can be drawn with replacement from this population. Find (i) The population mean (ii) the population standard deviation (iii) the mean of the sampling distribution of means & (iv) the standard error of means.

Solⁿ :- Given population 3, 7, 11, 15

$$\text{Mean } \mu = \frac{1}{4}(3+7+11+15) = 9$$

$$\text{Variance } \sigma^2 = \frac{1}{4}\left[(3-9)^2 + (7-9)^2 + (11-9)^2 + (15-9)^2\right]$$

$$= \frac{1}{4}(36+4+4+36) = 20$$

$\sigma = \sqrt{20} = 4.472$

The possible samples of size $N = 2$ which can be drawn with replacement from the given population are (3,3) (3,7) (3,11) (3,15)

(7,3) (7,7) (7,11) (7,15)

(11,3) (11,7) (11,11) (11,15)

(15,3) (15,7) (15,11) (15,15)

which are 16 in no. The means of these samples are

3, 5, 7, 9

5, 7, 9, 11

7, 9, 11, 13

9, 11, 13, 15

The frequency distribution of means is

Mean (x_i)	No of items (f_i)
3	1
5	2
7	3
9	4
11	3
13	2
15	1

$$\mu_{\bar{X}} = \frac{\sum f_i x_i}{\sum f_i} = \frac{1}{16}(3+10+21+36+33+26+15)$$

$$= 9$$

$$\sigma_{\bar{x}}^2 = \frac{\sum f_i x_i^2}{\sum f_i} - (\mu_{\bar{x}})^2$$

$$= \cancel{\frac{1}{16}\left[3+20+63+146+99\right.}$$

$$= 10$$

$\therefore$ standard error is $\sigma_{\bar{X}} = \sqrt{10} = 3.162$

Thus, for the given sample distribution of mean, $\mu_{\bar{X}} = 9$ & standard error is $\sigma_{\bar{X}} = \sqrt{10}$.

* observe $\mu_{\bar{X}} = 9 = \mu$

$$\sigma_{\bar{X}}^2 = 10 = \frac{\sigma^2}{N}$$

(2) Find the mean and the standard error in the sampling distributions of means for the sampling considered in ex. 1, but without replacement.

Sol^n:- The possible samples of size $N = 2$ which can be drawn without replacement from the given population are

(3,7) (3,11) (3,15) (7,11) (7,15) (11,15)

The means are 5, 7, 9, 9, 11, 13

$$\mu_{\bar{X}} = \frac{1}{6}(5+7+9+9+11+13) = 9$$

$$\sigma_{\bar{X}}^2 = \frac{1}{6}\left[(5-9)^2 + (7-9)^2 + (9-9)^2 + (9-9)^2 + (11-9)^2 + (13-9)^2\right]$$

$$= 20/3$$

$$\sigma_{\bar{X}} = \sqrt{\frac{20}{3}} = 2.582$$

* observe $N_p = 4,\ N = 2,\ \sigma = \sqrt{20}$

$$\sigma_{\bar{X}} = \frac{\sqrt{20}}{\sqrt{2}}\sqrt{\frac{4-2}{4-1}} = \frac{\sqrt{20}}{\cancel{\sqrt{2}}} \times \frac{\cancel{\sqrt{2}}}{\sqrt{3}} = \sqrt{\frac{20}{3}}$$

3. A population consists of the five no 2, 3, 6, 8, 11. Consider all possible samples of size two which can be drawn with replacement from this population. Find (i) the mean of the population (ii) the standard deviation of the population (iii) the mean of the sampling distribution of means (iv) the standard error of means.

4. Repeat the above problem for the case where the sampling is without replacement.

5. The weights of 3000 workers in a factory are normally distributed with mean 68 kgs and standard deviation 3 kgs. If 80 samples consisting of 25 workers each are obtained, what would be the mean and standard deviation of the sampling distribution of means if sampling were done (a) with replacement, (b) without replacement?
In how many samples will the mean is likely to be (i) between 66.8 & 68.3 kgs and (ii) less than 66.4 kgs?

Given $N_p = 3000$ & $N = 25$, $\mu = 68$, $\sigma = 3$

sampling with replacement, $\mu_{\bar{X}} = \mu = 68$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{N}} = \frac{3}{\sqrt{25}} = 0.6$$

Sampling without replacement

$\mu_{\bar{X}} = \mu = 68$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{N}}\sqrt{\frac{N_p - N}{N_p - 1}} = \frac{3}{\sqrt{25}}\sqrt{\frac{3000-25}{3000-1}}$$

$$= \frac{3}{5} \times 0.996 = 0.5976 \simeq 0.6$$

Thus $\mu_{\bar{X}}$ and $\sigma_{\bar{X}}$ have the same value in both cases.

Given Population is Normally distributed, the sample distribution of mean is also taken to be Normally distributed.

$\mu_{\bar{X}} = 68$, $\sigma_{\bar{X}} = 0.6$

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{\bar{X} - 68}{0.6}$$

For $\bar{X} = 66.8$, $Z = -2$

For $\bar{X} = 68.3$, $Z = 0.6$

For $\bar{X} = 66.4$, $Z = -2.67$

$$P(66.8 < \bar{X} < 68.3) = P(-2 < Z < 0.6)$$

$$= A(2) + A(0.6) = 0.4772 + 0.1915$$

$$= 0.6687$$

$\therefore$ In 80 samples, the expected no of samples having mean b/w 66.8 & 68.3 is 0.6687×80

$\cong 53$

$$P(\overline{X} < 66.4) = P(Z < -2.67)$$
$$= 0.5 - P(0 < Z < 2.67)$$
$$= 0.5 - 0.4962 = 0.0038$$

Accordingly, in 80 samples, the expected no of Samples having means less than 66.4 kg is

$0.0038 \times 80 \simeq 0.304$

6. It is guaranteed that a 4-litre can of a wall paint covers 57 square meters on the average with a standard deviation of 3.5 sq. mts. Find the probability that the total area covered by a sample of 40 of these 4-litre cans will be between 2200 and 2300 square meters.

$$\mu = 57, \quad \sigma = 3.5 \qquad N = 40$$

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{N}} = \frac{3.5}{\sqrt{40}} = 0.55$$

$$Z = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}} = \frac{\overline{X} - 57}{0.55}$$

For $\overline{X} = \frac{2200}{40} = 55$, $\quad Z = \frac{55 - 57}{0.55} = -3.64$

For $\overline{X} = \frac{2300}{40} = 57.5$, $\quad Z = \frac{57.5 - 57}{0.55} = 0.91$

$$\therefore P(55 < \overline{X} < 57.5) = P(-3.64 < Z < 0.91)$$
$$= P(0 < Z < 3.64) + P(0 < Z < 0.91)$$
$$= 0.4998 + 0.316$$
$$= 0.8158$$

Hence, it is about 82% certain that the sample covers a total area of 2200 to 2300 sq mts.

7. If the mean of an infinite population is 575 with standard deviation of 8.3, how large a sample must be used in order that there be one chance in 100 that the mean of the sample is less than 572?

$$\mu = 575, \quad \sigma = 8.3 \qquad N = ?$$

$$P(\bar{X} < 572) = 1/100 = 0.01$$

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{\bar{X} - 575}{\sigma/\sqrt{N}}$$

$$= \frac{\sqrt{N}[\bar{X} - 575]}{8.3}$$

For $\bar{X} = 572$, $\quad Z = \frac{-3\sqrt{N}}{8.3} = -(0.361)\sqrt{N}$

$$P(\bar{X} < 572) = 0.01$$

$$\Rightarrow 0.01 = P(Z < -0.361\sqrt{N})$$

$$0.01 = 0.5 - P(0 < Z < 0.361\sqrt{N})$$

$$P(0 < Z < 0.361\sqrt{N}) = 0.5 - 0.01$$

$$= 0.49$$

$$A(0.361\sqrt{N}) = 0.49$$

$$= A(2.35)$$

$$\Rightarrow 0.361\sqrt{N} = 2.35$$

$$\sqrt{N} = 6.51$$

$$N = 42.38$$

$\therefore$ The required sample size is 43 or above