UNIT - IV

Population, Parameter, Sample, Statistics, Sampling, Estimation, Test of significance, Test of hypothesis.

Population:

- Population [(or) universe] is an agregatte (or) totality of statistical data forming a subject of investigation.

Eg:- 1. The population of heights of Indian

2. The population of Nationalized bank in India.

- The no of observations in the population is defined to be the size of the population. It may be finite (or) infinite

Size of the population is denoted by 'N'.

Eg:- 1. If their are 600 students in the school that we classify according to Blood Type, we say that we have a population of size 600.

Sampling:-

- Most of the times, study of entire population may not be possible, to carry out. And Hence a part alone is selected from the given population. A portion of the population which is examined with a view to determing the population characteristics is called sample. i.e, a sample is a subset of population and the no of objects in the sample is called size of the sample & denoted by n.

- The process of selection of a sample is called sampeling.

- Thus, in estimating the characteristics of population, instead of enumerating entire population,

only the individual in the sample are examined. Then the sample characteristics are utilized to approximately estimate the population. The error involved in such approximation is known as sampling error.

Classification of Samples:-

- Samples are classified into 2 ways

1. Large sample :- If the size of a sample $n \geq 30$, the sample is said to be large sample.

2. Small sample:- If the size of sample $n < 30$, the sample is said to small (or) exact sample.

- In sampeling with replacement, each member of the population may be choosen more than one's (since the member is replaced in the population).
- Thus sampling from finite population with replacement can be considered theoritically as sampling from (with) infinte population.

Illy the sampling from finite population without replacement can be considered theoritically as sampling from finite population only.

Parameters and statistics:-

- Parameter is a statistical measure based on all units (or observations) of a population. Statistic (Samp statistic) is a statistical measure based on all uni selected in a sample.

Eg:- The statistical constants of the population namely Mean (μ), Variance (σ^2) are the usually reff to as parameters, statistical measures computed fr the sample observations alone eg:- Sample Mean ($\bar{x}$)

Sample variance (s^2) etc.. are usually to as statistics.

• In other words the mean, mode, median, S.D, variance measures of the population are called parameters and the measures obtain from the sample of population are called statistics.

Note:-

• Let us consider a finite population of size 'N' & random samples of size 'n'

1. The total number of samples with replacement is N^n

2. The total no of samples without replacement is N_{C_n}.

Sample Mean:-

If $x_1, x_2, \ldots x_n$ represent a random sample of size n, then the sample mean is defined by the statistic $\bar{x} = \sum_{i=1}^{n} x_i / n$.

Sample variance:-

If $x_1, x_2, \ldots x_n$ represent a random samples of size n, then the sample variance is defined by the statistic $s^2 = \frac{\sum_{i=1}^{n} (x_i - (\bar{x})^2}{n-1} = \frac{n(\sum x_i^2 - (\sum x_i)^2)}{n(n-1)}$

Eg:- A comparison of coffee prices at 4 randomly selected stores in a city showed increases from the previous month of 12, 15, 17 & 20 Rs per 1kg. Find the variance of this sample of price ↑.

Sol:- Sample Mean $\bar{x} = \frac{12+15+17+20}{4} = 16$

Sample Variance = $\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$

$$= \frac{(12-16)^2 + (15-16)^2 + (17-16)^2 + (20-16)^2}{3}$$

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

$$= \frac{16+1+16+1}{3} = \frac{34}{3} = 11.33$$

Standard Error of a statistic:-

- It is the standard deviation of the sampling distribution of the statistic.
- Standard error of sample mean = $\frac{\sigma}{\sqrt{n}}$ 90%
- Standard error of sample proportion = $\sqrt{\frac{PQ}{n}}$, $Q = 1-P$
- Standard error of sample S.D = $\frac{\sigma}{\sqrt{2n}}$

[Define population]

A population consists of 5 numbers 2, 3, 6, 8 & 11. Consider all possible samples of size 2 which can be drawn with replacement from this population. Find

a. The mean of the population

b. S.D of the population

c. The mean of sampling distribution of means &

d. The S.D of the sampling distribution of means

a. Mean of the population

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

b. Variance of the population

$$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n}$$

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

$$= \frac{(-4)^2 + (-3)^2 + (0)^2 + (2)^2 + (5)^2}{5}$$

$$= 10.8$$

$\cancel{S.D = \sqrt{\sigma}}$ $\quad \sigma = \sqrt{10.8}$

$$S.D = 3.29$$

c. Sampling with replacement (infinite population)

The total no of samples with replacement is

$N^n = 5^2 = 25$ [$\because$ N = population size
n = sample size]

They are (2, 2), (2, 3), (2, 6), (2, 8), (2, 11)
(3, 2), (3, 3), (3, 6), (3, 8), (3, 11)
(6, 2), (6, 3), (6, 6), (6, 8), (6, 11)
(8, 2), (8, 3), (8, 6), (8, 8), (8, 11)
(11, 2), (11, 3), (11, 6), (11, 8), (11, 11)

The sample means are

25 sample means

2	2.5	4	5	6.5
2.5	3	4.5	5.5	7
4	4.5	6	7	8.5
5	5.5	7	8	9.5
6.5	7	8.5	9.5	11

— (1)

$= \frac{150}{25} = 6$

d. The variance $\sigma^2_{\bar{x}}$ of the sampling distribution of means is obtained by subtracting the mean 6 from each number in eqn (1) and squaring the result, adding all 25 members thus obtain, and diving by 25.

$$\therefore \sigma^2_{\bar{x}} = \frac{\sum_{i=1}^{n} (x_i - \mu_{\bar{x}})^2}{25}$$

$$= \frac{(2-6)^2 + (2.5-6)^2 + (4-6)^2 + \dots + (25-6)^2 + (11-6)^2}{25}$$

$$= \frac{135}{25} = 5.40$$

$$\sigma_{\bar{x}} = \sqrt{5.40} = 2.32$$

(or) $\sigma^2_{\bar{x}} = \frac{\sigma^2}{n} \Rightarrow \sigma_x = \frac{\sigma}{\sqrt{2}} = \frac{3.29}{\sqrt{2}}$

n ↓ size of the sample

$= 2.326$

Imp Note:-

1. $\mu_{\bar{x}}$, $\sigma^2_{\bar{x}}$ are the mean of the sampling distribution of means and variance of the sampling distribution of means

2. For finite population (without replacement)

- The mean of sampling distribution of means

$\mu_{\bar{x}} = \mu$ where $E(\bar{x}) = \mu$

- The variance of sampling distribution of means

$\sigma^2_{\bar{x}} = \frac{\sigma^2}{n}\left[N-n/N-1\right]$ n = size of the sample.

N = size of the population

3. The factor $\left(\frac{N-n}{N-1}\right)$ is called a finite population correction factor.

4. For infinite population with replacement

a. The mean of sampling distribution means

$\mu_{\bar{x}} = \mu$ i.e $E(\bar{x}) = \mu$

b. The variance of sampling distribution of means

$\sigma^2_{\bar{x}} = \frac{\sigma^2}{n}$ n = size of sample.

Find the value of population correction factor, if

$n = 10$, $N = 1000$.

Given that

size of sample $n = 10$

size of population $N = 1000$

Now, finite population correction factor

$$= \frac{N-n}{N-1} = \frac{1000-10}{1000-1}$$

$$= \frac{990}{999}$$

$$= 0.9909$$

What is the value of correction factor, if $n = 5$, $N = 200$.

Given that

$n = 5$

$N = 200$

finite population correction factor

$$= \frac{N-n}{N-1} = \frac{200-5}{200-1}$$

$$= 0.9798$$

A sample is collected from the items produced by a factor. The sample size is 81 and S.D of the population is 0.3. Find the standard error of the sampling distribution of means. (S.Error = S.D)

Given that

$n = 81$ (sample size)

population S.D $\sigma = 0.3$

S.Error (or) S.D of the sampling distribution of mean $= \frac{\sigma}{\sqrt{n}} = \frac{0.3}{\sqrt{81}} = \frac{0.3}{9} = 0.0333$

The variance of population is 2 and size of ~~popul-ation~~ sample is 169. What sample collected is the S.Error of mean.

Given that

$n = 169$, $\sigma^2 = 2 \Rightarrow \sigma = \sqrt{2}$

S.Error of mean $= \frac{\sigma}{\sqrt{n}} = \frac{\sqrt{2}}{\sqrt{169}} = 0.108$

A population consists of 5 numbers 2, 3, 6, 8 & 11. Consider all possible samples of size 2 which can be drawn without replacements from this population. Find

a. The mean of the population

b. S.D of the population

c. The mean of sampling distribution of means

d. The S.D of the sampling distribution of means.

a. Mean of the population

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

b. Variance of the population

$$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n}$$

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

$$= 10.8$$

$$\sigma = \sqrt{10.8} = 3.29$$

c. Sampling without replacement (finite population)

The total no of samples without replacement is

$$^{N}C_{n} = {}^{5}C_{2} = 10 \text{ (samples of size 2)}$$

The 10 samples are

(2, 3), (2, 6), (2, 8), (2, 11)
(3, 6), (3, 8), (3, 11)
(6, 8), (6, 11)
(8, 11)

The selection (2, 3) is considered same as (3, 2)

Means of 10 samples are

(1) —— { 2.5 4 5 6.5
4.5 5.5 7
7 8.5
9.5 }

The mean of sampling distribution of means

$$\mu_{\bar{x}} = 2.5 + 4 + 5 + 6.5 + 4.5 + 5.5 + 7 + 7 + 8.5 + 9.5 / 10$$

$= 6$

d. Variance of sampling distribution of means

$$\sigma^2_{\bar{x}} = \frac{(2.5-6)^2 + (4-6)^2 + \dots + (8.5-6)^2 + (9.5-6)^2}{10}$$

$$= 4.05$$

$$\sigma = \sqrt{4.05}$$

$$= 2.012$$

If the population is 3, 6, 9, 15, 27 a) list all possible samples of size 3 that can be taken without replacement.

b. Calculate the mean of each of the sampling distribution.

c. S.D of sampling distribution of means.

a. Sampling without replacement (finite population)

The total no of samples without replacement of size 3 is $^{N}C_{n} = {}^{5}C_{3} = 10$

The 10 samples are

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

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

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

(6, 15, 27)

b. Mean of sampling distribution

6	8	12
9	13	15
10	14	16
17		

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

c. The variance of sampling distribution of mean

$$\sigma^2_{\bar{x}} = \frac{(6-12)^2 + (8-12)^2 + (12-12)^2 + (9-12)^2 + (13-12)^2 + (15-12)^2 + (10-12)^2 + (14-12)^2 + (16-12)^2 + (17-12)^2}{10}$$

$= 12$

$\sigma = \sqrt{12} = 3.46.$

Central limit theory:-

• If $\bar{x}$ is the mean of the sample of size n drawn from a population with mean (μ) and S.D (σ) then the standardized sample mean

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

is a random variable whose distribution function approaches that of the standard normal distribution as $n \to \infty$

The mean height of students in a college is 155cms & S.D is 15. What is the probability that the mean height of 36 students is less than 157cms.

Here μ = mean of the population = 155cm

σ = S.D of the population = 15

n = Sample size = 36

$\bar{x}$ = mean of sample = 157

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

Now $P(X < 157) = P(Z < 0.8)$

$= 0.5 + P(0 < Z < 0.8)$

$= 0.5 + A(0.8)$

$= 0.5 + 0.2881$

$= 0.7881$

A random sample of size 100 is taken from infinite population having mean $\mu = 76$ & $\sigma^2 = 256$. What is the probability that $\bar{x}$ will be b/w 75 & 78.

Given $n = 100$

$\mu = 76$

$\sigma = \sqrt{256} = 16$

$\bar{x}_1 = 75$

$$Z_1 = \frac{x_1 - \mu}{\sigma} = \frac{75 - 76}{16/10} = -\frac{1}{16/10} = -0.625$$

$$Z_2 = \frac{x_2 - \mu}{\sigma} = \frac{78 - 76}{16/10} = 1.25$$

$P(75 < X < 78) = P(-0.625 < Z < 1.25)$

$= A(0.625) + A(1.25)$

$= 0.6268$

WP

→ If a one Gallon can of paint covers on an average 513 sq.ft with a S.D of 31.5 sq.ft, what is the probability that the mean area covered by a sample of 40 of this one gallon cans will be anywhere from 510 sq.ft to 520.

The mean of certain normal population is equal to the S.D (or) standard error of the mean of the samples of 64 from that distribution. Find the sample size of 36 will be negative. ($\because n = 64$)

The standard error of mean $= \frac{\sigma}{\sqrt{n}} = \frac{\sigma}{\sqrt{64}} = \frac{\sigma}{8}$

Given that mean of the population $\mu =$ Standard error of the mean of the sample size 64

$$\mu = \frac{\sigma}{8}$$

W.K.T $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

$$z = \frac{\bar{x} - \sigma/8}{\frac{\sigma}{\sqrt{36}}} = \frac{\bar{x} - \frac{\sigma}{8}}{\sigma/6} = \frac{6\bar{x} - 6\frac{\sigma}{8}}{\sigma}$$

$$z = \frac{6\bar{x}}{\sigma} - \frac{6\sigma}{8\sigma} = \frac{6\bar{x}}{\sigma} - \frac{3}{4}$$

when $\bar{x}$ is -ve

i.e $\bar{x} < 0$

$6\bar{x} < 0$

$\frac{6\bar{x}}{\sigma} < 0$

$\frac{6\bar{x}}{\sigma} - \frac{3}{4} < -\frac{3}{4}$

$z < -0.75$

Estimation:-

1. Estimate: An estimate is a statement made to find an unknown population parameter.

2. Estimator:- The procedure (or) rule to determine an unknown population parameter is called an estimator.

Eg:- Sample mean is an estimator of population mean because sample mean is a method of determining the population mean.

A parameter can have 1 (or) 2 many estimaters.

Basically, their are 2 types of estimates to determine the statistic of the population parameters namely

a. point estimation

b. interval estimation

Point & Interval Estimation:-

- If an estimate of the population parameter is given by a single value, then the estimate is called point estimation of the parameter. But if an estimate of the population parameter is given by 2 different values b/w which the parameter may be considered, to lie, then the estimate is called the interval estimation of the parameter.

Eg:- If the height of the student is measured as 162cm then the measurement gives a point estimation, if the height is given as (163 ± 3.5) cm, then the height lies b/w 159.5 cm and 166.6cm, then the measurement gives an interval estimation.

• A point estimator is a statistic for estimating the population parameter 'θ' and will be denoted by '$\hat{\theta}$'.

→ An estimator is not expected to estimate the population parameter without error. An estimator should be closed to the true value of unknown parameter. ~~and~~

Unbiased Estimator:-

• A statistic (or) point estimator '$\hat{\theta}$' is said to be an unbiased estimator of the parameter 'θ', if $E(\hat{\theta}) = \theta$

Properties of estimators:-

• A good estimator is one which is as close to the true value of parameter.

• The imp properties of a good estimator are 1. consistency
2. unbiasedness
3. efficiency
4. sufficiency

Confidence Interval for μ, σ known (for large sample):-

• If $\bar{x}$ is the mean of the random sample of size n from the population with known variance σ^2, $(1-\alpha)100\%$ confidence interval for μ is given

$\bar{x} - z_{\alpha/2}\, \sigma/\sqrt{n} < \mu < \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ where $z_{\alpha/2}$ is the 'z' value leaving an area of $\alpha/2$ to the right.

• So, the maximum error estimate E with $1-\alpha$ probability is given by $E = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

• In this case sample size $n = \left(\frac{z_{\alpha/2}\, \sigma}{E} \right)^2$

Confidence interval for μ, σ unknown (for small sample)

- If $\bar{x}$, s are the mean & S.D of random sample from a normal population with unknown variance σ^2, $(1-\alpha)100\%$ confidence interval for μ is

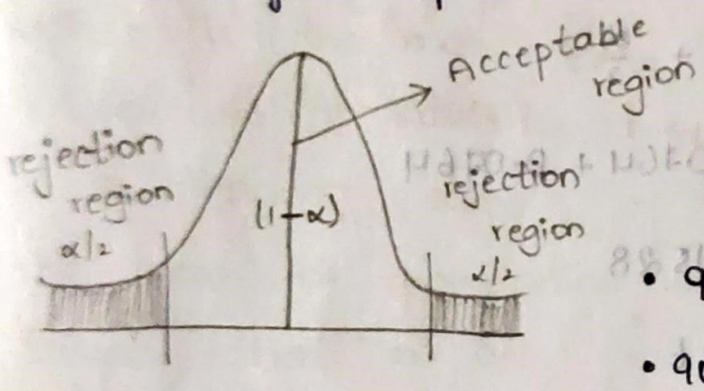
$$\bar{x} - t_{\alpha/2}\, s/\sqrt{n} < \mu < \bar{x} + t_{\alpha/2}\, s/\sqrt{n}$$

where $t_{\alpha/2}$ is the t value with $v = (n-1)$ degrees of freedom leaving an area of $\alpha/2$ to the right.

$\therefore$ The maximum error of estimate for small samples is given by $\epsilon = t_{\alpha/2}\, s/\sqrt{n}$ where n = sample size

s = S.D of sample.

Note:- (Large sample)



- At 95% confidence $z_{\alpha/2} = 1.96$
- 99% confidence, $z_{\alpha/2} = 2.58$
- 90% confidence, $z_{\alpha/2} = 1.64$

In a study of an automobile insurance a random sample of 80 body repairs cost had a mean Rs 472.36 and a S.D of Rs 62.35. If $\bar{x}$ is used as a point estimate to the true average repair cost, with what confidence we can assert that the maximum error doesnot exceed Rs 10.

Size of a random sample $n = 80$

The mean of random sample $\bar{x}$ = Rs 472.36

S.D σ = Rs 62.35

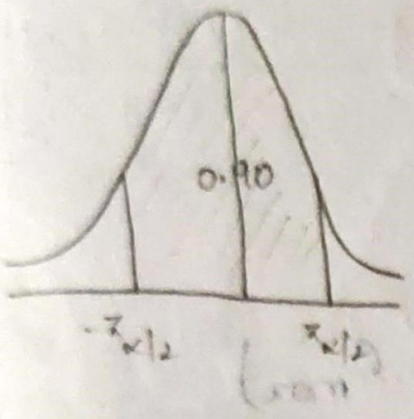
Max error of estimate = $\epsilon_{max} = 10$

assert with 90% confidence that the sample mean is off atmost 10 hrs.

It is given that $\epsilon_{max} = 10$hrs

S.D $\sigma = 48$hrs

And also given that 90% of confidence.



$$P(-z_{\alpha/2} < z < z_{\alpha/2})$$

$$\Rightarrow P(0 < z < z_{\alpha/2}) = \frac{0.90}{2} = 0.45$$

from the table (area to point)

$$z_{\alpha/2} = 1.64$$

$$\therefore \text{sample size } n = \left(\frac{z_{\alpha/2}\,\sigma}{E}\right)^2 = \left(\frac{1.64 \times 48}{10}\right)^2$$

$$= 61.96$$

What is the max error one can expect to make with probability 0.90 when using the mean of a random sample of size $n = 64$ to estimate the mean of population with $\sigma^2 = 2.56$.

Given that $n = 64$

Probability $= 0.90$

$\sigma^2 = 2.56 \Rightarrow \sigma = 1.6$

Confidence limit = 90%.

$$\therefore z_{\alpha/2} = 1.645$$

$$\therefore \epsilon_{max} = z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 1.645 \times \frac{1.6}{\sqrt{64}} = 0.328$$

A random sample of size 100 is taken from a population with $\sigma = 5.1$. Given that the S.M is 21.6. Construct a 95% confidence interval for the population mean μ.

Given that $n = 100$

σ = S.D of population = 5.1

S.M $\bar{x}$ = ~~at 95%~~ 21.6

At 95% confidence, $Z_{\alpha/2} = 1.96$ (from the table)

$$\epsilon_{max} = Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$= 1.96 \times \frac{5.1}{\sqrt{100}}$$

$$= 0.9996$$

Confidence interval for $\mu = (\bar{x} \pm \epsilon_{max})$

$$= (\bar{x} - \epsilon, \bar{x} + \epsilon)$$

$$= (21.6 - 0.9996, 21.6 + 0.9996)$$

$$= (20.6, 22.6)$$

Find 95% confidence limit for mean of a normality distribution population from which the following 15, 17, 10, 18, 16, 9, 7, 11, 13, 14.

$\bar{x} = 13$

S.V $$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

$$= \frac{(15-13)^2 + (17-13)^2 + (10-13)^2 + (18-13)^2 + (16-13)^2 + (9-13)^2 + (7-13)^2 + (11-13)^2 + (13-13)^2 + (14-13)^2}{9}$$

$s^2 = 13.33$

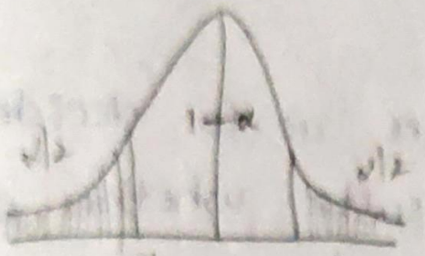
$s = \sqrt{13.33} = 3.65$

Given that confidence is 95%.

i.e $(1-\alpha)100\% = 95\%$

$1-\alpha = 0.95$

$\alpha = 0.05$ $\quad \therefore \frac{\alpha}{2} = 0.025$



Degree of freedom

$V = n-1$

$= 10-1 = 9$

from the table At $V = 9$ & $\frac{\alpha}{2} = 0.025$

$t_{\alpha/2} = 2.262$

$\therefore$ Confidence interval for $\mu = \bar{x} \pm t_{\alpha/2}$

$= 13 \pm 2.262$

$= \cancel{15.262}\ 10.738$, $\cancel{11.262}\ 15.262$

Test of Hypothesis:-

- The main object of the sampling theory is the study of test of hypothesis (or) test of significance.
- In many circumstence, we are to make decisions about population in basis of only sample information. For example on the basis of sample data, a drug chemist is to decide whether a new drug is really effective in curing the diesease.

Statistical Hypothesis:-

- In many circumstence, to arrive at decisions about the population on the basis of sample information.
- We make assumption (or) guesses about the