UNIT- II

* Sampling distribution And Estimation

Population:-
The collection of all objects under observations is called Population. It may be finite or Infinite.
→ The size of the population denoted by 'N'.

Sample:-
The part of the population is called sample
→ The size of the sample is denoted by 'n'.

Parameters:-
The statical constants such as population mean, variance, S.D and proportion. obtained from the population are called parameters.
→ These are denoted by μ, σ^2, σ and P.

Statistics:-
The stastical quantities obtained from the sample such as mean, variance, S.D and proportion are called statistics.
→ These are denoted by $\bar{x}, s^2$, s and p.

Population	Sample
$x_1, x_2 \ldots\ldots, x_n$	$x_1, x_2, \ldots\ldots, x_n$
$\mu = \frac{\sum_{i=1}^{n} x_n}{N}$	$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$
$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{N}$	$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}$
$\sigma = \sqrt{\sigma^2}$	$s = \sqrt{s^2}$

Sampling:-

The process of choosing samples from a population is called sampling.

Sampling Distribution:-

The set of all possible samples are called Sampling Distribution.

→ There are two type of sampling distributions.

i) Sampling without replacment
ii) Sampling with replacment

Note 1:- In sampling without replacment $^{N}C_{n}$ samples can be drawn from finite population

Note 2:- Sampling with replacment N^{n} samples can be drawn from infinite population.

The Sampling Distribution of Mean (σ-known):-

Suppose all possible samples of size n taken from the population size N.

i.e $x_1, x_2, x_3, \ldots, x_n$ are samples taken from the population and

i.e $\bar{x}_1, \bar{x}_2, \bar{x}_3, \ldots, \bar{x}_n$ are the means of each sample

(i) Mean of sampling distribution of means

$$\mu_{\bar{x}} = \frac{\sum_{i=1} \bar{x}_i}{n}$$

(ii) Variance of sampling distribution of mean

$$\sigma_{\bar{x}}^2 = \frac{\sum_{i=1}^{n} (\bar{x}_i - \mu_{\bar{x}})^2}{n}$$

(iii) $SD = \sqrt{\sigma^2_{\bar{x}}}$

Note 1) Suppose the samples are drawn from the infinite population (with replacment) then

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

(ii) $\sigma^2_{\bar{x}} = \frac{\sigma^2}{n}$

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

Note 2) Consider a finite population (without replacement)

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

(ii) $\sigma^2_{\bar{x}} = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$

Here $\frac{N-n}{N-1}$ is called population (or) correction factor.

Standard Error of Mean:-

The standard deviation of sampling distribution of statistics is called standard error and it is given by $\frac{\sigma}{\sqrt{n}}$.

$\because \frac{\sigma}{\sqrt{n}}$ is called standard error of mean.

Q) → A population cosists of 5 members 2, 3, 6, 8, and 11. consider all possible samples of size '2' which can be drawn with replacment from the population the find (i) the mean of the population

(ii) the S.D of the population.

(iii) the mean of the sampling distributions of mean

(iv) the S.D of " " "

$\because$ $N = 5$ $n = 2$

(i) Mean $\mu = \frac{2+3+6+8+11}{5} = \frac{30}{5} = 6$ $\quad \mu = \sum_{i=1}^{n} \frac{x_i}{N}$

(ii) $\sigma^2 = \sum_{i=1}^{5} \frac{(x_i - \mu)^2}{N}$

$= \frac{1}{5} \left[(2-6)^2 + (3-6)^2 + (6-6)^2 + (8-6)^2 + (11-6)^2 \right]$

$= \frac{1}{5} [16+9+0+4+25]$

$= \frac{1}{5} [54] = \frac{54}{5}$

$\sigma^2 = \frac{54}{5} = 10.8$

$\sigma = \sqrt{10.8}$

$S.D = 3.28$

(iii) Sampling with replacement

The total no. of sampling with replacement and size 2

$n=2$

$N=5,\ n=2$

No. of samples $= 5^2 = 25$

Samples $= \{(2,2),(2,3)(2,6)(2,8)(2,11)\ (3,2)(3,3)(3,6)\ (3,8)(3,11)(6,2),$
$(6,3)(6,6)\ (6,8)(6,11)\ (8,2)(8,3)(8,6)(8,8)(8,11)(11,2)$
$(11,3)(11,6)(11,8)(11,11)\}$

Means are $= \{2, 2.5, 4, 5, 6.5, 2.5, 3, 4.5, 5.5, 7, 4, 4.5, 6, 7,$
$8.5, 5, 5.5, 7, 8, 9.5, 6.5, 7, 8.5, 9.5, 11\}$

$\mu_{\bar{x}} = \sum_{i=1}^{25} \frac{\bar{x}_i}{25}$

$= \frac{1}{25} (2 + 2.5 + \dots + 11\}$

$= \frac{1}{25} (150)$

$\mu_{\bar{x}} = 6$

iv) $\sigma_{\bar{x}}^2 = \sum_{i=1}^{25} \frac{(\bar{x}_i - \mu_{\bar{x}})^2}{25}$

$= \frac{1}{25}\left[(2-6)^2 + (2.5)^2 + (4-6)^2 + \text{-------------} + (9-6)^2 + (8.5-6)^2 + (9.5-6)^2 + (11-6)^2\right\}$

$= \frac{1}{25}(135) = 5.4$

$\sigma_{\bar{x}} = \sqrt{\sigma_{\bar{x}}^2} = \sqrt{5.4} = 2.32$

Verification:-

$\mu = \mu_{\bar{x}} = 6$

$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} = \frac{10.8}{2} = 5.4$

=) Solve the above pblm ① without replacement

(iii) There no.of samples which can be drawn from given population with size 2 without replacement is

$^{N}C_{n} = {^5C_2} = \frac{5!}{2!(5-2)!} = 10$

The samples = $\{(2,3)\ (2,6)\ (2,8)\ (2,11)\ (3,6)\ (3,8)\ (3,11)\ (6,8)\ (6,11)\ (8,11)\}$

$= \{2.5, 4, 5, 6.5, 4.5, 5.5, 7, 7, 8.5, 9.5\}$

The mean of sampling distribution of mean.

$\mu_{\bar{x}} = \sum_{i=1}^{10} \frac{\bar{x}_i}{10}$

$= \frac{2.5+4+5 \text{-------} +9.5}{10}$

$\mu_{\bar{x}} = 6$

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

iv) $\sigma_{\bar{x}}^2 = \sum_{i=1}^{10} \frac{(\bar{x}_i - \mu_{\bar{x}})^2}{10}$

$= \frac{1}{10}\left\{(2.5-6)^2 + (4-6)^2 + (5-6)^2 + (6.5-6)^2 + \text{----} \right.$

$\left. \text{----} + (8.5-6)^2 + (9.5-6)^2\right\}$

$\sigma_{\bar{x}}^2 = 4.05$

$S.D = \sigma_{\bar{x}} = 2.01$

<u>verification:-</u>

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

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

$= \frac{10.8}{2}\left(\frac{5-2}{5-1}\right)$

$4.05 = 4.05$

21/10/24

If the population is 3, 6, 9, 15, 27 then

a) list all the possible samples of size 3 that can be taken without replacement from finite population

b) Calculate the mean of sampling distribution of means.

c) find the S.D of sampling distribution of means.

Sol:- a) $N=5$ $n=3$

without replacement $= {}^{N}C_{n}$

$$= {}^{5}C_{3} = \frac{5!}{(5-3)!\,3!}$$

$$= 10$$

The samples are = { (3,6,9), (3,6,15), (3,6,27)
(3,9,15) (3,9,27) (3,15,27)
(6,9,15) (6,15,27) (9,15,27)
(6,9,27) }

samples mean = { 6, 8, 12, 9, 13, 15, 10, 14, 16, 17 }

(b) The mean of sampling distribution mean $\mu_{\bar{x}} = \sum_{i=1}^{10} \frac{x_i}{10}$

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

$$= \frac{120}{10} = 12$$

(c) $\sigma_{\bar{x}}^2 = \sum_{i=1}^{10} \frac{(\bar{x}_i - \mu_{\bar{x}})^2}{10} \Rightarrow \frac{1}{10}\{(6-12)^2 + (8-12)^2 + (12-12)^2 + \cdots$

$$\cdots + (16-12)^2 + (17-12)^2\}$$

$$= \frac{1}{10}[36+16+0+9+1+9+4+4+16+25]$$

$$= 12$$

④ What is the value of correction factor if $n=5$, $N=200$

$$\text{correction factor} = \frac{N-n}{n-1}$$

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

$$= \frac{195}{199}$$

$$= 0.9798$$

⑤ The variance population is 2. The size of the sample collected from the population is 169. what is the standard error.

given $\sigma^2 = 2$
$n = 169$

$$\text{standard error} = \frac{\sigma}{\sqrt{n}}$$

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

$\sigma^2_{\bar{x}} = \frac{\sigma^2}{n}$

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

$$= \frac{\sqrt{2}}{\sqrt{169}}$$

$$= \frac{\sqrt{2}}{13}$$

$$= 0.108$$

<u>Central limit theorem</u> :—

If $\bar{x}$ is the mean of large sample size and $n (\geq 30)$ taken from a population having mean μ. and variance σ^2 then the limiting form of the distribution $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ is the standard normal distribution.

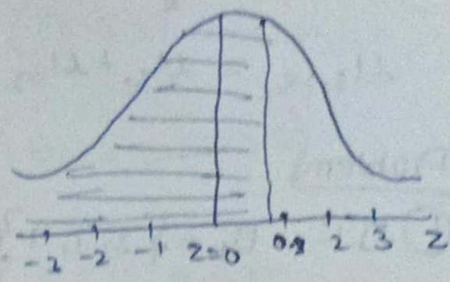
① The mean height of a student in a clg is 155cm and $S.D = 15$. what is the probability that the mean height of 36 students is less than 157 cm.

sol:- $\mu = 155$ $S.D = 15$ $n = 36$ $P(\bar{x} \leq 157) = ?$

we have C.L.T $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$

when $\bar{x} = 157 \Rightarrow z = \dfrac{157 - 155}{15/\sqrt{36}}$

$z = 0.8$



$P(\bar{x} < 157) = P(z < 0.8)$

$= 0.5 + A(0.8)$

$= 0.5 + 0.2881$

$= 0.7881$

② A normal population has a mean of 0.1 and S.D of 2.1. Find the probability that mean of sample size 900 will be negative.

Sol:- $\mu = 0.1$ S.D $= 2.1$ $n = 900$

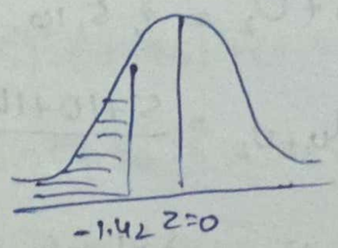
$P(\bar{x} < 0) = ?$

By C.L.T we have $z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

$= \dfrac{0 - 0.1}{2.1/\sqrt{900}}$

$z = -1.428$

$P(\bar{x} < 0) = P(z < -1.42)$



$= 0.5 - A(-1.42)$

$= 0.5 - A(1.42)$

$= 0.5 - 0.4222$

$= 0.078$

Sampling distribution of differences & sums

The sampling distribution of differences are given by

1) $\mu_{S_1 - S_2} = \mu_{S_1} - \mu_{S_2}$ and $\sigma_{S_1 - S_2} = \sqrt{\sigma^2_{S_1} + \sigma^2_{S_2}}$

where μ_{S_1} & σ_{S_1} are the mean & S.D of statistic S_1 from population A.

where μ_{S_2} & σ_{S_2} are the mean & S.D of statistic S_2 from population B.

2) Sampling distribution of sum's is given by

$\mu_{s_1+s_2} = \mu_{s_1} + \mu_{s_2}$ and $\sigma_{s_1+s_2} = \sqrt{\sigma_{s_1}^2 + \sigma_{s_2}^2}$

Problems

① If $U_1 = \{2, 7, 9\}$, $U_2 = \{3, 8\}$ then find

a) μ_{U_1}

b) μ_{U_2}

c) $\mu_{U_1+U_2}$

d) $\mu_{U_1-U_2}$

e) σ_{U_1}

f) σ_{U_2}

g) $\sigma_{U_1+U_2}$

h) $\sigma_{U_1-U_2}$

verification:-

(i) $\mu_{U_1+U_2} = \mu_{U_1} + \mu_{U_2}$

(ii) $\mu_{U_1-U_2} = \mu_{U_1} - \mu_{U_2}$

(iii) $\sigma_{U_1+U_2} = \sqrt{\sigma^2_{U_1} + \sigma^2_{U_2}}$

ⓐ $\mu_{U_1} = \dfrac{2+7+9}{3} = 6$

ⓑ $\mu_{U_2} = \dfrac{3+8}{2} = 5.5$

ⓒ $U_1 + U_2 = \{5, 10, 10, 15, 12, 17\}$

$\mu_{U_1+U_2} = \dfrac{5+10+10+15+12+17}{6} = 11.5$

d) $U_1 - U_2 = \{-1, -6, 4, -1, 6, 1\}$

$\mu_{U_1-U_2} = \dfrac{-1-6+4-1+6+1}{6} = \dfrac{3}{6} = \dfrac{1}{2} = 0.5$

e) $\sigma_{U_1} = \sigma^2_{U_1} = \dfrac{\sum_{i=1}^{3}(x_i - \mu_{U_1})^2}{3} = \frac{1}{3}\left((2-6)^2 + (7-6)^2 + (9-6)^2\right)$

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

$\sigma^2_{U_1} = \dfrac{\cancel{26}}{\cancel{3}}$ 8.66

$\sigma_{U_1} = \sqrt{8.66}$

f) $\sigma_{U_2} \Rightarrow \sqrt{\sigma^2_{U_2}} = \sum_{i=1}^{2} \frac{(x_i - \mu_{U_2})^2}{2} \Rightarrow \frac{1}{2}\left((3-5.5)^2 + (8.5-5)^2\right)$

$= \frac{1}{2}\left(6.25 + 12.25\right)$

$\sigma^2_{U_2} = \frac{1}{2}(18.5) = 9.25$

$\sigma_{U_2} = \sqrt{9.25} = 3.041$

g) $\sigma_{U_1+U_2} \Rightarrow U_1 + U_2 = \{5, 10, 10, 15, 12, 17\}$

$\sigma_{U_1+U_2} = \sqrt{\sigma_{U_1+U_2}^2} = \frac{1}{6}\sum_{i=1}^{6}(x_i - \mu_{U_1+U_2})^2$

$= \frac{1}{6}\left[(5-11.5)^2 + (10-11.5)^2 + (10-11.5)^2 + (15-11.5)^2 + (12-11.5)^2 + (17-11.5)^2\right]$

$= \frac{1}{6}\left[42.25 + 2.25 + 2.25 + 12.25 + 0.25 + 30.25\right]$

$= \frac{1}{6}[89.5] = 14.916$

$\sigma_{U_1+U_2} = \sqrt{14.916} = 3.8621$

h) $\sigma_{U_1-U_2} = U_1 - U_2 = \{-1, -6, 4, -1, 6, 1\} \Rightarrow \sqrt{\sigma^2_{U_1-U_2}} = \frac{1}{6}\sum_{i=1}^{6}(x_i - \mu_{U_1-U_2})^2$

$= \frac{1}{6}\left[(-1-0.5)^2 + (-6-0.5)^2 + (4-0.5)^2 + (-1-0.5)^2 + (6-0.5)^2 + (1-0.5)^2\right]$

$= \frac{1}{6}\left[2.25 + 42.25 + 12.25 + 2.25 + 30.25 + 0.25\right]$

$= \frac{1}{6}[89.5] \Rightarrow 14.916 \Rightarrow \sigma_{U_1-U_2} = \sqrt{14.916} = 3.8621$

(i) $\mu_{U_1+U_2} = \mu_{U_1} + \mu_{U_2}$

$11.5 = 6 + 5$

$11.5 = 11.5$

(iii) $\sigma_{U_1+U_2} = \sqrt{\sigma^2_{U_1} + \sigma^2_{U_2}}$

$3.8621 = \sqrt{8.66 + 9.25}$

$3.8621 = \sqrt{17.91} = 4.22$

(ii) $\mu_{U_1-U_2} = \mu_{U_1} - \mu_{U_2}$

$0.5 = 6 - 5.5$

$0.5 = 0.5$

Estimation:-

The procedure of estimating a population parameter by using sample information is called estimation.

Estimator:- ($\hat{\theta}$)

The statistic which is used to determine an unknown population parameter is called Estimator.

Ex: $\bar{x}$ is estimator of μ.

properties of good Estimator:-

An estimator is said to be good estimator if it contains

(i) unbaisedness
(ii) Consistency
(iii) Suffiency and Effiency.

We have two types of estimations
(i) Point Estimation
(ii) Interval Estimation

1) **Point Estimation:-** An estimation is said to be point estimation if we can estimate population parameter by a particular value.

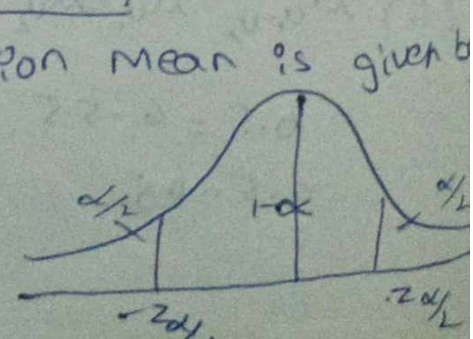
2) **Interval Estimation:-** An estimation is said to be Interval Estimation if we can estimate population parameter in a particular Range

Maximum error of estimate of mean:-

The maximum error for population Mean is given by

$$|\bar{x}-\mu|$$

we have CLT $\boxed{z=\dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}}}$



$$P\left(-z_{\alpha/2} \le z \le z_{\alpha/2}\right) = 1-\alpha$$

$$P\left(-z_{\frac{\alpha}{2}} \le \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \le z_{\alpha/2}\right) = 1-\alpha$$

$$P\left(-z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \le \bar{x}-\mu \le z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

$\because |x|<a$

$-a<x<a$

$$P\left(|x-\mu| \le z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

$$\boxed{E = z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}}$$

Confidence interval for μ:-

If $\bar{x}$ is the sample mean of the sample size 'n' taken from a population with variance 'σ^2' then $(1-\alpha)100\%$ confidence interval for μ is given by $(\bar{x}-E, \bar{x}+E)$ where $E = z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$

problems:-

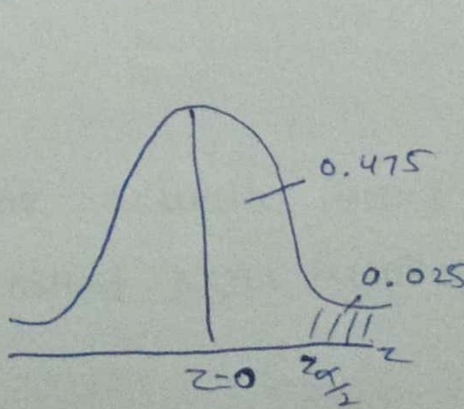
① A random sample of size $n=100$ is taken from a population with $\sigma=5.1$. Given that the sample mean is $\bar{x}=21.6$. construct a 95% confidence interval for a population mean μ.

sol:

we have confidence interval for μ.

is $(\bar{x}-E, \bar{x}+E) \rightarrow$ ①

where $E = z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$



given $\alpha = 5\%$

$\alpha = 0.05$

$\alpha/2 = 0.025$

from diagram

$P(0 \le z \le z_{\alpha/2}) = 0.475$

$z_{\frac{\alpha}{2}} = 1.96$

$E = z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$

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

$E = 0.9$

from ①

confidence interval

$= (21.6 - 0.9, 21.6 + 0.9)$

$= (20.7, 22.51)$