UNIT-3
Sampling Theory

By
D.V. L. Prasanna
Associate Professor
Aditya Engineering College(A)

Contents

Introduction – Population and samples – Sampling distribution of Means and Variance – Central limit theorem (without proof)-Point and Interval estimations – Maximum error of estimate.

Sampling:

Sampling is a method that allows us to get information about the population based on the statistics from a subset of the population (sample), without having to investigate every individual.

Sampling Frame – It is a list of items or people forming a population from which the sample is taken.

Population(or Universe): A collection or aggregate or totality of persons, things or objects or statistical data under study is known as population.

Size of the Population: Number of units in the population is known as size of the population. It is denoted by N.

Finite Population: The number of units in the population is finite then it is known as finite population.

Infinite Population: : The number of units in the population is infinite then it is known as infinite population.

Sample: A portion(or subset) of the population is known as sample.

Size of the sample: Number of units in the sample is known as size of the sample. It is denoted by n.

Large sample: If the size of the sample is greater than 30 then it is known as large sample.

Small sample: If the size of the sample is less than 30 then it is known as small sample.

Parameter: The values that are obtained from the population data or a number that describes the property of the population is known as parameters

Statistic: The values that are obtained from the sample data or a number that describes the property of the sample is known as Statistics.

Probability Sampling: In probability sampling, every element of the population has an equal chance of being selected. Probability sampling gives us the best chance to create a sample that is truly representative of the population

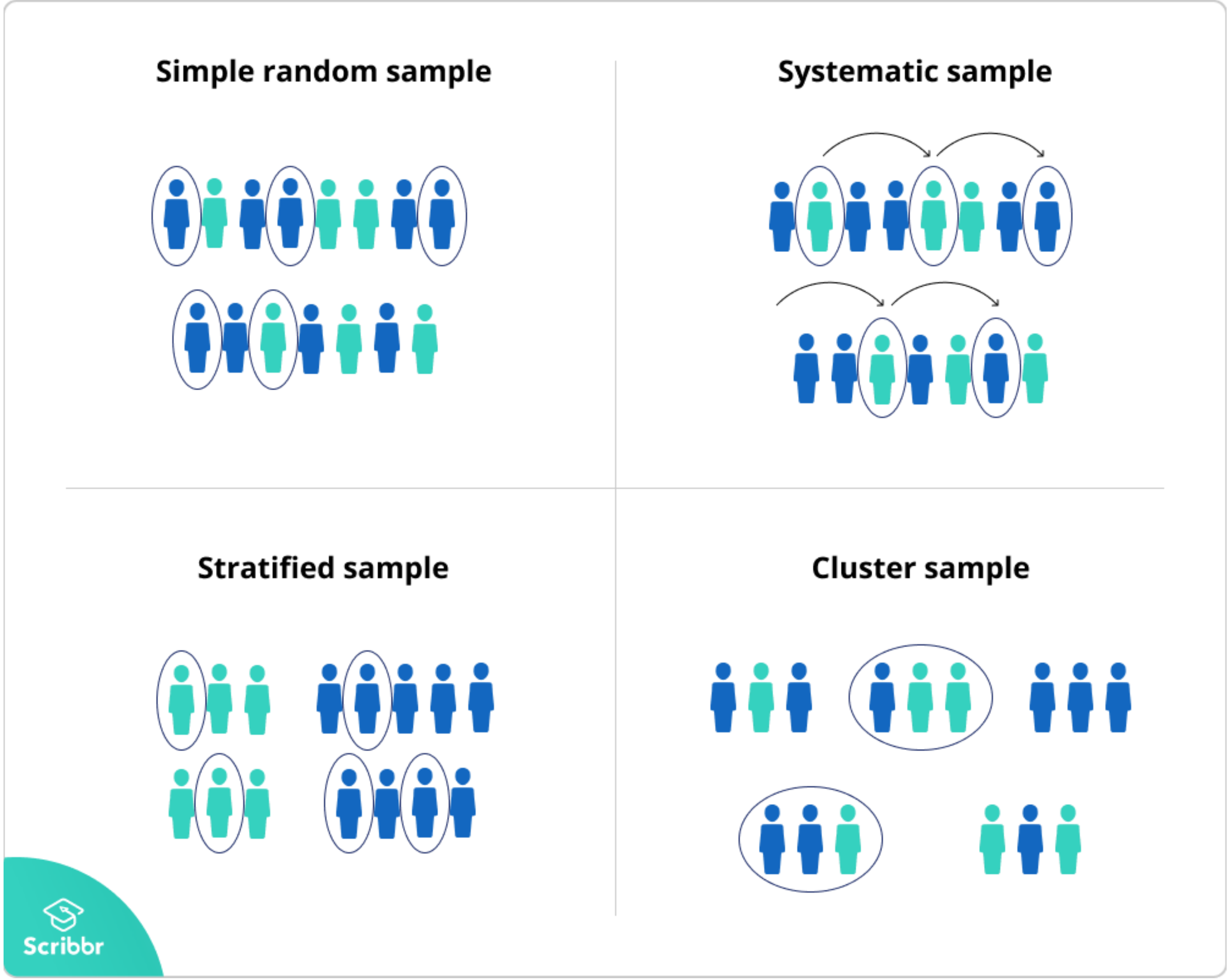
Non-Probability Sampling: In non-probability sampling, all elements do not have an equal chance of being selected. Consequently, there is a significant risk of ending up with a non-representative sample which does not produce generalizable results.

Probabilistic sampling methods:

1. Simple random sampling
2. Systematic sampling
3. Stratified sampling
4. Cluster sampling

Non-Probabilsitic sampling methods:

1. Purposive or Judgement sampling
2. Sequential sampling


Simple random sample
Systematic sample
Stratified sample
Cluster sample
Scribbr

Probabilistic sampling methods:

1. Simple random sampling

In a simple random sample, every member of the population has an equal chance of being included in the sample.. Your sampling frame should include the whole population. To conduct this type of sampling, tools like random number generators or other techniques are used that are based entirely on chance.

Ex: 1. selecting randomly 20 words from a dictionary is a random sample

2. Choosing 10 patients from a hospital in order to test the efficacy of a certain nely invented drug.

2. Systematic sampling or Quasi-random sampling:

In this method the sample is formed by some systematic manner by taking items at regular intervals.

All the units of the population are arranged in some order. Then from the first k items, one unit is selected at random. This unit and every kth unit of the serially listed population combined together constitute a systematic sample. This type of sampling is known as systematic sampling.

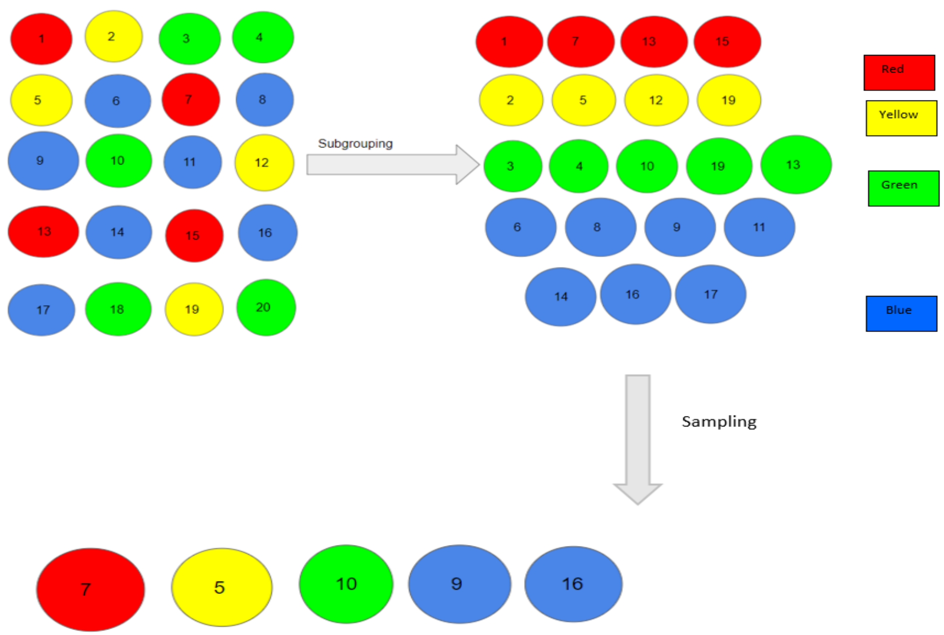
The difference between random sampling and systematic sampling is in case of random sampling all units are chosen randomly, whereas in case of systematic sampling only first unit is chosen at random.

3. Stratified sampling or stratified random sampling:

This method is useful when the population is heterogeneous.

In this type of sampling, the population is first sub-divided into several groups called strata according to some relevant characteristics(e.g. gender, age range, income bracket, job role). so that each stratum is more or less homogeneous. Each stratum is called sub population.

Then a small sample is select from each stratum at random. All the sub samples are combined to form the stratified sample which represents the population properly.



4. Cluster sampling

Cluster sampling also involves dividing the population into subgroups, but each subgroup should have similar characteristics to the whole sample. Instead of sampling individuals from each subgroup, you randomly select entire subgroups.

If it is practically possible, you might include every individual from each sampled cluster. If the clusters themselves are large, you can also sample individuals from within each cluster using one of the techniques above.

This method is good for dealing with large and dispersed populations, but there is more risk of error in the sample, as there could be substantial differences between clusters. It's difficult to guarantee that the sampled clusters are really representative of the whole population.

Non-probabilistic sampling methods:

1. **Purposive or Judgement sampling:**

When the choice the individual items of a sample entirely depends on the individual judgement of the investigator, it is called a purposive or Judgement sampling.

In this method, the members constituting the sample are chosen not according to some definite scientific procedure, but according to convenience and personal choice of the individual who selects the sample.

In this type, the investigator must have a good deal of experience and a thorough knowledge of the population.
Purposive selection is always subject to some kind of bias.
This method is suitable when the sample is small.

2. Sequential sampling:

It consists of a sequence of samples drawn one after another from the population depending on the results of the previous samples.
If the result of the first sample leads to a decision which is not acceptable, the lot from which the sample was drawn is rejected. Then a second ssample is drawnand as before if required, a third sample is drawn to arrive at a final decision to accept or reject the lot.
If the first sample is acceptable then there is no need to draw new sample.
It is widely used in Statistical quality control in factories engaged in mass production and other areas.

Population and Samples:

1. Sample mean:

The sample mean of a set of observations is the sum of observations divided by the number of observations in the data. It is denoted by $\bar{x}$

For n values in a set of data namely as $x_1, x_2, x_3, \ldots x_n$, the sample mean of data is given as:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \cdots\cdots + x_n}{n}$$

or

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

Population mean :-

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$$

Variance:

Variance is the average squared difference of the values from the mean

Variance reflects the degree of spread in the data set. The more spread the data, the larger the variance is in relation to the mean.

Variance formula for populations

Formula	Explanation
$$\sigma^2 = \frac{\Sigma (X - \mu)^2}{N}$$	•σ^2 = population variance •Σ = sum of… •X = each value •μ = population mean •N = number of values in the population

Variance formula for samples

Formula	Explanation
$s^2 = \frac{\Sigma (X - \bar{x})^2}{n - 1}$	• s^2 = sample variance • Σ = sum of… • X = each value • $\bar{x}$ = sample mean • n = number of values in the sample

Standard deviation:

Square root of variance is called standard deviation.

Standard deviation formula for Population:

Formula	Explanation
$\sigma = \sqrt{\dfrac{\Sigma (X-\mu)^2}{N}}$	•σ = population standard deviation •∑ = sum of… •*X* = each value •μ = population mean •*N* = number of values in the population

Standard deviation formula for samples:

Formula	Explanation
$s = \sqrt{\dfrac{\Sigma (X-\bar{x})^2}{n-1}}$	•*s* = sample standard deviation •∑ = sum of… •*X* = each value •x̅ = sample mean •*n* = number of values in the sample

Sampling Distributions: The probability distribution of a sample statistic is often called as sampling distribution of the statistic.

The standard deviation of the sampling distribution of a statistic is called **Standard Error(S.E)**

The mean of the sampling distribution of means, denoted by $\mu_{\bar{x}}$, is given by $E(\bar{X}) = \mu_{\bar{x}} = \mu$ where μ is the mean of the population.

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

If a population is infinite or if sampling is with replacement, then the variance of the sampling distribution of means, denoted by $\sigma_{\bar{x}}^2$ is given by $E[(\bar{X} - \mu)^2] = \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$

where σ^2 is the variance of the population.

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} \quad \text{(with replacement)}$$

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right) \quad \text{(without replace)}$$

If the population is of siqe N, if sampling is without replacement, and if the sample size is $n \le N$ then $\sigma_{\bar{x}}^{2} = \frac{\sigma^{2}}{n}\left(\frac{N-n}{N-1}\right)$

The factor $\left(\frac{N-n}{N-1}\right)$ is called the finite population correction factor, is close to 1 (and can be omitted for most practical purposes) unless the samples constitutes a substantial portion of the population.

Note: No. of samples of size n can be drawn from a population of size N

a) with replacement $= N^{n}$

b) without replacement $= {}^{N}C_{n}$.

Sampling distribution of means:-

$S_1 \to \bar{x}_1$

$S_2 \to \bar{x}_2$

$S_3 \to \bar{x}_3$

$S_4 \to \bar{x}_4$

$\vdots$

$S_n \to \bar{x}_n$

$\bar{X} = \{\bar{x}_1, \bar{x}_2, \bar{x}_3, \cdots, \bar{x}_n\}$

mean of sampling distbn of means

$$\mu_{\bar{x}} = \frac{\bar{x}_1 + \bar{x}_2 + \cdots + \bar{x}_n}{n} = \frac{\sum_{i=1}^{n} \bar{x}_i}{n}$$

$$\sigma^2_{\bar{x}} = \frac{(\bar{x}_1 - \mu_{\bar{x}})^2 + (\bar{x}_2 - \mu_{\bar{x}})^2 + \cdots + (\bar{x}_n - \mu_{\bar{x}})^2}{n}$$

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

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

mean of sampling distribution of means = mean of pop

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} \quad \text{(with replacement)}$$

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right) \quad \text{(without replacement)}$$

PROBLEMS:1. A population consists of five numbers 2,3,6,8 and 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 deviation of the sampling distribution of means

Solution: Given that N=5, n=2 and

i. Mean of the population

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

ii. Variance of the population

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N} = \frac{(2-6)^2 + (3-6)^2 + (6-6)^2 + (8-6)^2 + (11-6)^2}{5}$$

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

$$= 10.8$$

$$\therefore \sigma = 3.29$$

Sampling with replacement(infinite population):
The total number of samples with replacement is
$N^n = 5^2 = 25$

There 25 samples can be drawn

$$\begin{Bmatrix} (2,2) & (2,3) & (2,6) & (2,8) & (2.11) \\ (3,2) & (3,3) & (3,6) & (3,8) & (3,11) \\ (6,2) & (6,3) & (6.6) & (6,8) & (6,11) \\ (8,2) & (8,3) & (8,6) & (8,8) & (8,11) \\ (11,2) & `(11,3) & (11,6) & (11,8) & (11,11) \end{Bmatrix}$$

The sample means are

Sampling dist of means = $\bar{x}$

$$\begin{Bmatrix} 2 & 2.5 & 4 & 5 & 6.5 \\ 2.5 & 3 & 4.5 & 5.5 & 7 \\ 4 & 4.5 & 6 & 7.0 & 8.5 \\ 5 & 5.5 & 7 & 8 & 9.5 \\ 6.5 & 7 & 8.5 & 9.5 & 11 \end{Bmatrix}$$

iii. The mean of the sampling distribution of means is

$$\mu_{\bar{x}} = \frac{2 + 2.5 + 4 + 5 + 6.5 + - - - - - + 11}{25}$$

$$= \mathbf{6}$$

iv. The standard deviation of the sampling distribution of means

$$\sigma^2_{\bar{x}} = \frac{(2-6)^2 + (2.5-6)^2 + - - - - - + (11-6)^2}{25}$$

$$= 5.40$$

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

(or) $\sigma^2_{\bar{x}} = \frac{\sigma^2}{n} = \frac{10.8}{2} = 5.4$

1) (b) <u>without replacement</u>

No of samples of size 2 that can be drawn from population of size 5 without replacement $= N_{C_n}$

$$= 5_{C_2} = \frac{5!}{3!\,2!} = \frac{5 \times \cancel{4}^{2} \times \cancel{3!}}{\cancel{3!} \times \cancel{2}} = 10 \text{ samples}$$

population is 2, 3, 6, 8, 11   2, 3, 6, 8, 11

samples are (2,3) (2,6) (2,8) (2,11)
(3,6) (3,8) (3,11)
(6,8) (6,11) and (8,11)

Sampling distribution of means $(\bar{x})$ $= \begin{Bmatrix} 2.5, 4, 5, 6.5, \\ 4.5, 5.5, 7 \\ 7, 8.5, 9.5 \end{Bmatrix}$

Mean of sampling distribution of means $[\mu_{\bar{x}}]$

$$= (2.5+4+5+6.5+4.5+5.5+7+7+8.5+9.5)/10 = 6$$

variance of sampling distribution of means

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

$$= \frac{(2.5-6)^2 + (4-6)^2 + (5-6)^2 + (6.5-6)^2 + (4.5-6)^2 + (5.5-6)^2 + (7-6)^2 + (7-6)^2 + (8.5-6)^2 + (9.5-6)^2}{10}$$

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

standard deviation $\sigma_{\bar{x}} = \sqrt{4.05} = 2.01$

(or)

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right) = \frac{10.8 \times (5-2)}{2 \times (5-1)} = \frac{10.8 \times 3}{2 \times 4}$$

$$= 4.05.$$

2. A population consists of five numbers4, 8, 12, 16, 20, 24. Consider all possible samples of size two which can be drawn without 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 deviation of the sampling distribution of means

Solution: Given that N=6, n=2 and

i. Mean of the population

$$\mu = \sum \frac{x_i}{N} = \frac{4+8+12+16+20+24}{6} = \frac{84}{6} = 14$$

ii. Variance of the population

$$\sigma^2 = \sum \frac{(x_i - \bar{x})^2}{N} = \frac{(4-14)^2 + (8-14)^2 + (12-14)^2 + (16-14)^2 + (20-14)^2 + (24-14)^2}{6}$$

$$= \frac{100 + 36 + 4 + 4 + 36 + 100}{6}$$

$$= 46.67 \qquad \sigma = 3.29$$

Sampling without replacement (finite population):

The total number of samples without replacement is $N_{c_n} = 6_{c_2} = 15$

There 15 samples can be drawn

$$\begin{Bmatrix} (4,8) & (4,12) & (4,16) & (4,20) & (4,24) \\ (8,12) & (8,16) & (8,20) & (8,24) & \\ (12,16) & (12,20) & (12,24) & & \\ (16,20) & (16,24) & & & \\ (20,24) & & & & \end{Bmatrix}$$

The sample means are

$$\begin{Bmatrix} 6 & 8 & 10 & 12 & 14 \\ 10 & 12 & 14 & 16 & \\ 14 & 16 & 18 & & \\ 18 & 20 & & & \\ 22 & & & & \end{Bmatrix}$$

iii. The mean of the sampling distribution of means is

$$\mu_{\bar{x}} = \frac{6+8+10+12-----20+22}{15}$$

$$= \mathbf{14}$$

iv. The standard deviation of the sampling distribution of means

$$\sigma_{\bar{x}}^2 = \frac{(6-14)^2+(8-14)^2+----+(22-14)^2}{15}$$

$$= 18.67$$

$$\sigma_{\bar{x}} = 4.32$$

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

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

b) Calculate the mean and variance of the population

c) Find the mean and standard deviation of the sampling distribution of means.

Solution:

N=5,n=3

No. of samples that can be drawn without replacement= $N_{C_n} = 5_{C_3} = 10$

a) The samples are

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

b) Mean of the population (μ) $= \dfrac{\sum_{i=1}^{5} x_i}{N}$

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

Variance of population $= \dfrac{\sum (x_i-\mu)^2}{N}$

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

$$= \frac{360}{5} = 72$$

$S.D = \sqrt{72} = 8.485$

c) Sampling distribution of mean =

$\{6, 8, 12, 9, 13, 15, 10, 14, 16, 17\}$

Mean of sampling distribution of means $(\mu_{\bar{x}})$

$$= \frac{6+8+12+9+13+15+10+14+16+17}{10} = \frac{120}{10} = 12$$

variance of sampling distribution of means $(\sigma_{\bar{x}}^2)$

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

$= 13.3$

$\therefore \sigma_{\bar{x}} = \sqrt{13.3} = 3.65$

3, 6, 9, 15, 27

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

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

sampling distribution of means

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

Mean of sampling dist of means ($\mu_{\bar{x}}$)

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

$$\sigma_{\bar{x}}^2 = \frac{\sum_{i=1}^{10} (\bar{x}_i - \mu_{\bar{x}})^2}{n} = \frac{(6-12)^2 + (8-12)^2 + (12-12)^2 + \cdots + (16-12)^2 + (17-12)^2}{10}$$

$$= 13.3$$

$$\sigma_{\bar{x}} = \sqrt{13.3} = 3.65$$

Exercise Problems:

4. Samples of size 2 are taken from the population 1, 2, 3, 4, 5, 6. Which can be drawn without replacement? 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 deviation of the sampling distribution of means

Central limit theorem:

If $\overline{x}$ be the mean of a random sample of size n drawn from a population having mean μ and standard deviation σ, then the standardized sample mean is

$$z = \frac{\overline{x} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}$$

is a random variable whose distribution function approaches that of the standard normal distribution N(z;0,1) as $n \to \infty$

Standard error formulas:

Standard error of sample mean $\overline{x}$ or S.E of $\overline{x}$ = $\frac{\sigma}{\sqrt{n}}$

Standard error of sample proportion p or S.E of p = $\sqrt{\frac{PQ}{n}}$

Standard error of sample standard deviation = $\frac{\sigma}{\sqrt{2n}}$
(s)

4. S.E of difference of sample means i.e., $(\bar{x}_1 - \bar{x}_2)$

$$= \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

5. S.E of difference of proportions i.e., $(p_1 - p_2)$

$$= \sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}$$

6. S.E of difference of standard deviations $(s_1 - s_2)$

$$= \sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}$$

Note:

1) For a finite population of size N, when a sample is drawn without replacement

(i) S.E of sample mean $= \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$.

(ii) S.E of sample proportion $= \sqrt{\frac{PQ}{n}} \cdot \sqrt{\frac{N-n}{N-1}}$.

where $\sqrt{\frac{N-n}{N-1}}$ is known as finite population correction factor.

Sampling distribution of differences and sums:-

1. $\mu_{s_1 - s_2} = \mu_{s_1} - \mu_{s_2}$

2. $\sigma_{s_1 - s_2} = \sqrt{\sigma_{s_1}^2 + \sigma_{s_2}^2}$.

3. $\mu_{s_1 + s_2} = \mu_{s_1} + \mu_{s_2}$

4. $\sigma_{s_1 + s_2} = \sqrt{\sigma_{s_1}^2 + \sigma_{s_2}^2}$

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

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

7. $\mu_{p_1 - p_2} = \mu_{p_1} - \mu_{p_2} = p_1 - p_2$

8. $\sigma_{p_1 - p_2} = \sqrt{\sigma_{p_1}^2 + \sigma_{p_2}^2} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$

1) Let $u_1 = (3,7,8)$, $u_2 = (2,4)$. Find

a) μ_{u_1} b) μ_{u_2} c) Mean of sampling distribution of the difference of means $\mu_{u_1 - u_2}$

d) σ_{u_1} e) σ_{u_2}

f) the standard deviation of the differences of means $\sigma_{u_1 - u_2}$.

Soln:-

a) $\mu_{u_1} = \dfrac{3+7+8}{3}$

$= 18/3 = 6$

b) $\mu_{u_2} = \dfrac{2+4}{2} = \dfrac{6}{2} = 3$

$\therefore \mu_{u_2} = 3$

d) $u_1 = (3, 7, 8)$

$\mu_{u_1} = 6$

$$\sigma_{u_1}^2 = \frac{(3-6)^2 + (7-6)^2 + (8-6)^2}{3}.$$

$$= \frac{9+1+4}{3} = \frac{14}{3} = 4.667 \qquad \therefore \sigma_{u_1}^2 = 4.667$$

$$\sigma_{u_1} = \sqrt{4.667} = 2.16$$

e) $u_2 = (2, 4)$, $\mu_{u_2} = 3$

$$\sigma_{u_2}^2 = \frac{(2-3)^2 + (4-3)^2}{2}$$

$$= \frac{1+1}{2} = 1$$

$$\therefore \sigma_{u_2} = 1.$$

c) $u_1 = (3, 7, 8)$ $u_2 = (2, 4)$

$u_1 - u_2 = (1, -1, 5, 3, 6, 4)$

Mean of sampling distribution of differences $u_1 - u_2$

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

$$= \frac{18}{6} = 3$$

$$\therefore \mu_{u_1 - u_2} = 3$$

(or) $\mu_{u_1 - u_2} = \mu_{u_1} - \mu_{u_2}$

$= 6 - 3 = 3$.

$\therefore \mu_{u_1 - u_2} = 3$

f) Variance of sampling distribution of differences $u_1 - u_2$ i.e.,

$\mu_{u_1 - u_2} = 3$, $u_1 - u_2 = \{1, -1, 5, 3, 6, 4\}$

$$\sigma^2_{u_1-u_2} = \frac{(1-3)^2+(-1-3)^2+(5-3)^2+(3-3)^2+(6-3)^2+(4-3)^2}{6}$$

$$= \frac{4+16+4+0+9+1}{6}$$

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

Standard deviation $\sigma_{u_1-u_2} = \sqrt{\text{variance}}$

$$= \sqrt{5.667}$$

$$= 2.38$$

or $\sigma_{u_1-u_2} = \sqrt{\sigma_{u_1}^2+\sigma_{u_2}^2} = \sqrt{4.667+1} = \sqrt{5.667} = 2.38$

2) Find the value of the finite population correction factor for $n = 10$ & $N = 1000$.

Soln :- $N = 1000$

$n = 10$

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

$$= \frac{990}{999} = 0.991.$$

3) The variance of a population is 2. The size of the sample collected from the population is 169. What is standard error of mean?

Soln :- $\sigma^2 = 2$, $n = 169$

$$S.E \text{ of } \bar{x} = \sqrt{\frac{\sigma^2}{n}} = \sqrt{\frac{2}{169}} = 0.185$$

4) when a sample is taken from an infinite population what happen to the standard error of mean if the sample size is decreased from 800 to 200.

Soln :- standard error for sample mean $\bar{x} = \frac{\sigma}{\sqrt{n}}$

If $n_1 = 800$ $\quad S.E_1 = \frac{\sigma}{\sqrt{800}} = \frac{\sigma}{\sqrt{4 \times 200}} = \frac{\sigma}{2\sqrt{200}}$

If $n_2 = 200$

$$SE_2 = \frac{\sigma}{\sqrt{200}}$$

$$S.E_1 = \frac{1}{2}\left(\cdot \frac{\sigma}{\sqrt{200}}\right) = \frac{1}{2} \cdot SE_2$$

$$SE_2 = 2 \cdot SE_1$$

$\therefore$ If sample size decreased from 800 to 200, then S.E will be multiplied by 2.

5.What is the effect on standard error, if a sample is taken from an infinite population of sample size is increases from 400 to 900.

Solution:

$$n_1 = 400, n_2 = 900$$

S.E of sample mean $\bar{x} = \frac{\sigma}{\sqrt{n}}$

$$S.E_1 = \frac{\sigma}{\sqrt{n_1}} = \frac{\sigma}{\sqrt{400}} = \frac{\sigma}{20}$$

$$S.E_1 = \frac{\sigma}{20}$$

$$S.E_2 = \frac{\sigma}{\sqrt{n_2}} = \frac{\sigma}{\sqrt{900}} = \frac{\sigma}{30}$$

$$S.E_2 = \frac{\sigma}{30} = \frac{\left(\frac{\sigma}{20}\right)}{\left(\frac{3}{2}\right)} = \frac{S.E_1}{\left(\frac{3}{2}\right)}$$

Thus if the sample size is increased from 400 to 900, the standard error will be divided by $\frac{3}{2}$

Problems on central limit theorem:

1. The mean height of students in a college is 155cms and standard deviation is 15. what is the probability that the mean height of 36 students is less than 157 cms.

Solution: Given mean of the population (μ) = 155 cm

Standard deviation of the population (σ) = 15 cm

Sample size (n) = 36

Mean of sample ($\bar{x}$) = 157 cm

$$\text{Now } Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$= \frac{157 - 155}{\frac{15}{\sqrt{36}}}$$

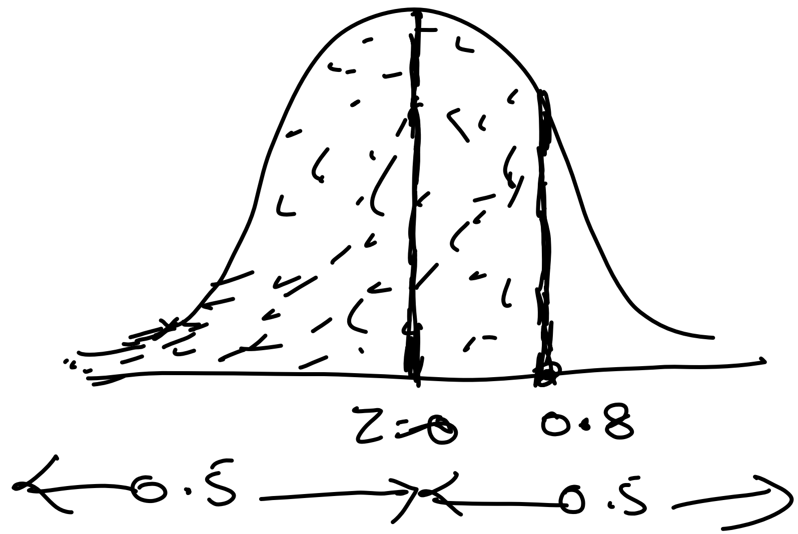
$$= 0.8$$

$$\therefore P(\overline{x} \leq 157) = P(Z < 0.8)$$

$$= 0.5 + P(0 \leq Z \leq 0.8)$$

$$= 0.5 + 0.2881$$

$$\therefore P(\overline{x} \leq 157) = 0.7881$$



2. If a 1-gallon can of paint covers on an average 513 square feet with a standard deviation of 31.5 square feet, what is the probability that the mean area covered by a sample of 40 of these 1-gallon cans will be anywhere from 510to 520 square feet?

Solution :-

$\mu = 513$ sq. feet

$\sigma = 31.5$ sq. feet.

$n = 40$

$P(510 \leq \bar{x} \leq 520) = ?$

If $\bar{x} = 510$, $z = \dfrac{\bar{x} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}$

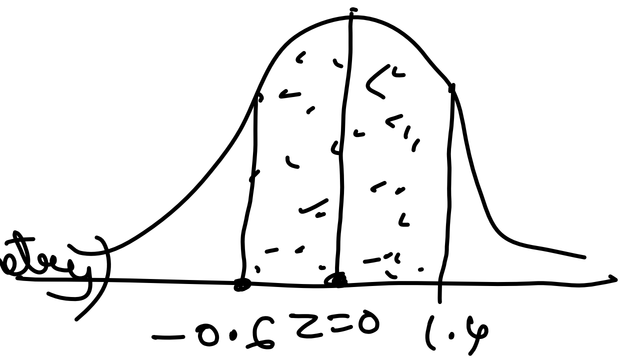
$= \dfrac{510 - 513}{\left(\frac{31.5}{\sqrt{40}}\right)} = -0.6$

$\bar{x} = 520, \quad z = \dfrac{520 - 513}{\left(\dfrac{31.5}{\sqrt{40}}\right)} = 1.4$

$P(510 \leq \bar{x} \leq 520) = P(-0.6 \leq z \leq 1.4)$

$= P(-0.6 \leq z \leq 0) + P(0 \leq z \leq 1.4)$

$= P(0 \leq z \leq 0.6) + P(0 \leq z \leq 1.4)$ ($\because$ due to symmetry)



$= 0.2258 + 0.4192$

$= 0.645$

$\therefore P(510 \leq \bar{x} \leq 520) = 0.645.$

3. The mean of certain normal population is equal to the standard error of the mean of the samples of 64 from that distribution. Find the probability that the mean of the sample size 36 will be negative.

Solution:– mean of population $= \mu$

S.E of sample mean $\bar{x} = \frac{\sigma}{\sqrt{n}}$ where $n = 64$.

Given that

mean of normal population $=$ S.E of $\bar{x}$

$$\mu = \frac{\sigma}{\sqrt{n}} \Rightarrow \mu = \frac{\sigma}{\sqrt{64}} \qquad P(\bar{x} < 0)$$

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

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \quad \text{where } n = 36$$

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

$$z = \frac{\bar{x} - \frac{\sigma}{8}}{\frac{\sigma}{6}} = \frac{\bar{x}}{\left(\frac{\sigma}{6}\right)} - \frac{\left(\frac{\sigma}{8}\right)}{\left(\frac{\sigma}{6}\right)}$$

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

$z = \bar{x}$ 5

$-\frac{2}{3}$

$$= \frac{6\bar{x}}{\sigma} - \frac{\cancel{\sigma}}{\cancel{8}_4} \times \frac{\cancel{6}^3}{\cancel{\sigma}}$$

$-8 \quad -7 \quad -6 \quad -5$

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

If $\bar{x}<0$, $z<$ 5

$$z = \frac{6\bar{x}}{\sigma} - 0.75$$

If $\bar{x} < 0$ then clearly z will be less than 0.75

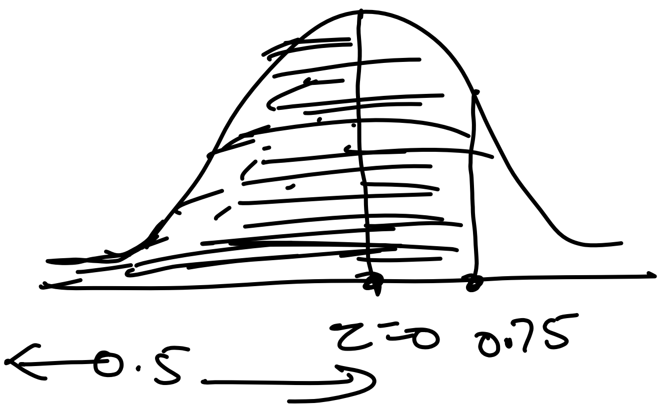
(or) If $z < 0.75$ then only $\bar{x}$ will be negative.

$$P(\bar{x} < 0) = P(z < 0.75)$$

$$= 0.5 + P(0 < z < 0.75)$$

$$= 0.5 + 0.2734$$

$$P(\bar{x} < 0) = 0.7734.$$



4. A normal population has a mean of 0.1 and standard deviation of 2.1 . Find the probability that the mean of a sample of size 900 will be negative.

Solution:

population mean $\mu = 0.1$

population standard deviation $= 2.1$

mean of sample $= \bar{x}$

sample size $n = 900$

we have

$$z = \frac{\bar{x} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}$$

$$z = \frac{\bar{x} - 0.1}{\left(\frac{2.1}{\sqrt{900}}\right)} = \frac{\bar{x} - 0.1}{0.07}$$

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

$\bar{x} - 0.1 = 0.07Z$

$\bar{x} = 0.07Z + 0.1$

probability that the sample mean $\bar{x}$ will be negative

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

$= P(0.07Z + 0.1 < 0)$

$= P(0.07Z < -0.1)$

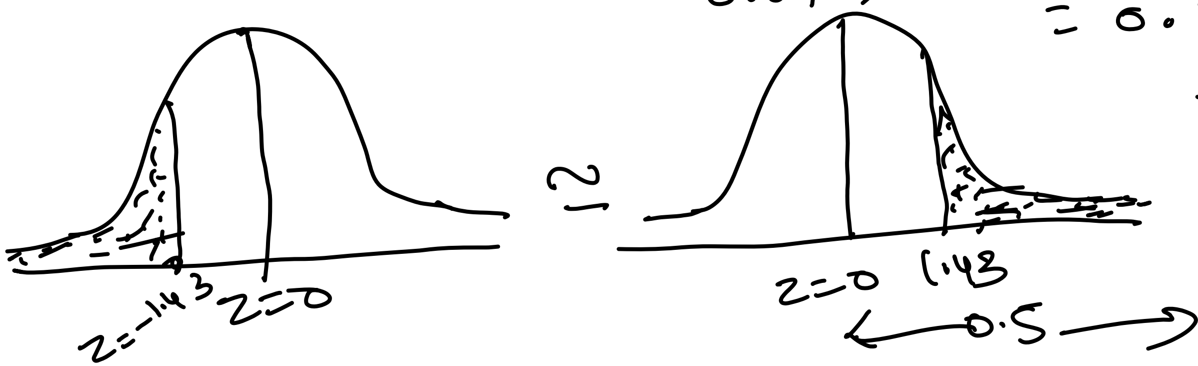
$= P\left(Z < \frac{-0.1}{0.07}\right) = P(Z < -1.43)$

$= 0.5 - P(0 \leq Z \leq 1.43)$

$= 0.5 - 0.4236$

$= 0.0764$

$\therefore P(\bar{x} < 0) = 0.0764$



5. A random sample of size 64 is taken from a normal population with mean51.4 and S.D 68. what is the probability that the mean of sample will

a) exceed 52.9 b) fall between 50.5 and 52.3 c) less than 50.6

6. If the mean breaking strength of copper wire is 575lbs, with a standard deviation of 8.3lbs. How large a sample must be used in order that there will be one chance in 100 that the mean breaking strength of the sample is less than 572lbs.

7)The mean voltage of a battery is 15 and S.D is 0.2. Find the probability that four such batteries connected in series will have a combined voltage of 60.8 or more volts.

Solution:

Let the mean voltage of batteries $A, B, C, D \; be \; \overline{X_A}, \overline{X_B}, \overline{X_C}, \overline{X_D}$

$$\mu_{\overline{X_A}+\overline{X_B}+\overline{X_C}+\overline{X_D}} = \mu_{\overline{X_A}} + \mu_{\overline{X_B}} + \mu_{\overline{X_C}} + \mu_{\overline{X_D}}$$
$$= 15 + 15 + 15 + 15 = 60$$

$$\sigma_{A+B+C+D} = \sqrt{\sigma_A^2 + \sigma_B^2 + \sigma_C^2 + \sigma_D^2}$$

$$= \sqrt{(0.2)^2 + (0.2)^2 + (0.2)^2 + (0.2)^2}$$
$$= \sqrt{4 \times (0.2)^2} = 0.4$$

Let X be the combined voltage of the series

$$when \; X = 60.8, z = \frac{X-\mu}{\sigma} = \frac{60.8-60}{0.4} = 0.2$$

$$p(X \geq 60.8) = p(z > 2) = 0.5 - 0.4772 = 0.0228$$

$= 0.5 - p(0 \leq z \leq 2)$

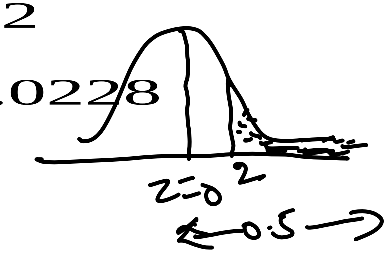


Table : Area under standard normal curve from 0 to $\frac{x-\mu}{\sigma}$

$\frac{x-\mu}{\sigma}$	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.0000	.0040	.0080	.0120	.0159	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1084	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1916	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2258	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2518	.2549
0.7	.2580	.2611	.2642	.2671	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4430	.4441
1.6	.4452	.4463	.4474	.4485	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4654	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4762	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4865	.4868	.4871	.4875	.5878	.4881	.4884	.4887	.4992
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946.	4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4980	.4980	.4981
2.9	.4981	.4982	.4983	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993

Estimation

Estimate:

An estimate is a statement made to find unknown population parameter

Estimator:

The method or rule to determine an unknown population parameter is called an **Estimator.** For example sample mean is an estimator of population mean because sample mean is a method of determining the population mean. A parameter can have many or 1,2 estimators. The estimators should be found so that they are very near to parameter values.

Statistical estimation:

The process by which we draw a conclusion about some measure of a population based on a sample value. The measures might be a variable , such as mean, S.D etc.

Types of estimation:

There are two types of estimates to determine the parameters of the population

1. Point estimation
2. Interval estimation

Point estimation:

An estimate of a population parameter given by a single number is called a **point estimate** of the parameter. If we say that a distance in 5.28 mts , we are giving a point estimate.

Interval estimation:

An estimate of a population parameters given by two numbers between which the parameter may be considered to lie is called an **interval estimate** of the parameter. The distance lie between 5.25 and 5.31 mts.

Unbiased estimator:

A statistic or point estimator $\hat{\theta}$ is said to be an unbiased estimator of the parameter θ if the mean of $\hat{\theta}$ is equal to the population parameter θ i.e.,

$$E(\hat{\theta}) = \theta$$

Properties of good estimator:

The estimator should be

a)consistent

b)efficient

c) sufficient

Maximum error:

$$E = Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \rightarrow \text{large samples}$$

$$E = t_{\alpha/2} \cdot \frac{S}{\sqrt{n}} \rightarrow \text{small samples}$$

$$E = z_{\alpha/2}\sqrt{\frac{PQ}{n}} \quad (\sigma \text{ unknown})$$

$$E = \frac{z_{\alpha/2}}{2\sqrt{n}} \quad [\text{Both } \sigma \text{ \& } P \text{ are unknown}]$$

σ-population S.D
n=sample size
P=population proportion
Q=1-P
E=maximum error
α is level of significance

for 99% $z_{\alpha/2}$ =2.58
for 98% $z_{\alpha/2}$ =2.33
for 9.73% $z_{\alpha/2}$ = 3
95% $z_{\alpha/2}$ =1.96
90% $z_{\alpha/2}$ =1.645

- Confidence interval for μ (for large samples n ≥ 30) σ known

large samples

$$\bar{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \text{ or } \left(\bar{x} - z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\right) \text{ or } (\bar{x} - E, \bar{x} + E)$$

Small samples $\rightarrow \left(\bar{x} - t_{\alpha/2}\cdot\frac{S}{\sqrt{n}}, \bar{x} + t_{\alpha/2}\frac{S}{\sqrt{n}}\right)$ or $(\bar{x} - E, \bar{x} + E)$

4. A random sample of 400 items is found to have mean 82 and standard deviation of 18. Find the maximum error of estimation at 95% confidence interval. Find the confidence limits for the mean if $\bar{x}$ =82.

solution:- Sample size $n = 400$

sample mean $\bar{x} = 82$

standard deviation $s = 18$

Maximum error $(E) = ?$

At 95% confidence $z_{\alpha/2} = 1.96$

$$E = z_{\alpha/2} \cdot \frac{s}{\sqrt{n}} = 1.96 \times \frac{18}{\sqrt{400}}$$

$$E = 1.76$$

95% confidence interval $= \left(\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right)$ or

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

$$= \left(82 - 1.96 \times \frac{18}{\sqrt{400}},\ 82 + 1.96 \times \frac{18}{\sqrt{400}}\right)$$

$$= (80.24,\ 83.76)$$

or $(\bar{x} - E, \bar{x} + E)$

$$= (82 - 1.76,\ 82 + 1.76)$$

$$= (80.24,\ 83.76)$$

2. Determine a 95% confidence interval for the mean of normal distribution with variance 0.25, using a sample of size 100 values with mean 212.3.

Solution: Given Sample size (n) = 100 variance $\sigma^2 = 0.25$

$\therefore$ Standard deviation of sample (σ) = $\sqrt{0.25} = 0.5$

Mean of sample ($\bar{x}$) = 212.3 and $Z_{\alpha/2} = 1.96$(for 95%)

$\therefore$ 95% Confidence interval = $\left(\bar{x} - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right)$

$$= \left(212.3 - 1.96 \cdot \frac{0.5}{\sqrt{100}}, 212.3 + 1.96 \cdot \frac{0.5}{\sqrt{100}} \right)$$

$$= (212.202, 212.398)$$

3. What is the size of the smallest sample required to estimate an unknown proportion to within a maximum error of 0.06 with at least 95% confidence.

Soln: sample size $n = ?$

maximum error $E = 0.06$

For 95% confidence, $z_{\alpha/2} = 1.96$

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

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

$$n = \left(\frac{z_{\alpha/2}}{2E}\right)^2$$

$$= \left(\frac{1.96}{2 \times 0.06}\right)^2$$

$$n = 266.78 \simeq 267$$

$\therefore$ Sample size $= 267$.

4. In a study of an automobile insurance a random sample of 80 body repair costs had a mean of Rs. 472.36 and the standard deviation of Rs. 62.35. If $\bar{x}$ is used as a point estimate to the true average repair costs, with what confidence we can assert that the maximum error doesn't exceed Rs. 10.

Solution: Given Sample size (n) = 80

Standard deviation of sample (s) = 62.35

Mean of sample ($\bar{x}$) = 472.36

Maximum Error(E)=10

$$\therefore E = Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$\Rightarrow Z_{\alpha/2} = \frac{E.\sqrt{n}}{\sigma} = \frac{10\sqrt{80}}{62.35} = 1.4345$$

The area when z=1.43 from the tables is 0.4536

$$\therefore \frac{\alpha}{2} = 0.4236 \quad \Rightarrow \alpha = 0.8472$$

$\therefore$ Confidence = $(1-\alpha)100\% = 84.72\%$

Hence we are 84.72% confidence that the maximum error is Rs. 10.

5.A random sample of size 100 has a standard deviation of 5. what can you say about the maximum error with 95%confidence.

6. The mean and standard deviation of population are 11,795 and 14,054 respectively. If n=50, find the 95% confidence interval for the mean.

7.What is the maximum error one can expect to make with the probability 0.90 when using the mean of a random sample of size n=64 to estimate the mean of the population with variance 2.56.

Solution:

Sample size n=64

Population variance $\sigma^2 = 2.56$

Standard deviation $\sigma = \sqrt{2.56} = 1.6$

$For\ .90\ probability\ \ z_{\alpha/2} = 1.645$

$$E = Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

8. It is desired to estimate the mean number of hours of continuous use until a certain computer will first require repairs. If it can be assumed σ=48 hours, how large a sample be needed so that one will be able to assert with 90% confidence that the sample mean is off by at most 10 hours.

Solution: $\sigma = 48$

For 90%, $Z_{\alpha/2} = 1.645$

$n = ?$

$E = 10$ hours

$$E = Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \Rightarrow \sqrt{n} = \frac{Z_{\alpha/2} \cdot \sigma}{E}$$

$$n = \left(\frac{Z_{\alpha/2} \cdot \sigma}{E}\right)^2 = \left(\frac{1.645 \times 48}{10}\right)^2 = 62.34 \simeq 62$$

9. Find 95% confidence limits for the mean of normality distributed population from which the following sample was taken 15,17,10,18,16,9,7,11,13,14.

Soln:-

$$\left(\bar{x} - z_{\alpha/2}\frac{s}{\sqrt{n}},\ \bar{x} + z_{\alpha/2}\frac{s}{\sqrt{n}}\right) \rightarrow \text{Confidence limits}$$

95% $z_{\alpha/2} = 1.96$

$n = 10$, $\bar{x} = \dfrac{15+17+10+18+16+9+7+11+13+14}{10}$ $\qquad \dfrac{\sum x_i}{n}$

$\therefore \bar{x} = 13$

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

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

$$s^2 = \frac{40}{3} \qquad \therefore s = 3.65$$

95% Confidence limits $= \left(\bar{x} - z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}\right)$

$$= \left(13 - 1.96 \times \frac{3.65}{\sqrt{10}}, 13 + 1.96 \times \frac{3.65}{\sqrt{10}}\right)$$

$$= (10.74, 15.26)$$

10 If we can assert with 95% that the maximum error is 0.05 and $P = 0.2$, find the size of the sample.

Solution: $P = 0.2$ $Q = 1 - P = 1 - 0.2 = 0.8$

For 95%, $z_{\alpha/2} = 1.96$,

Maximum Error $E = 0.05$

$$E = z_{\alpha/2} \cdot \sqrt{\frac{PQ}{n}} \Rightarrow \sqrt{n} = z_{\alpha/2} \frac{\sqrt{PQ}}{E}$$

$$n = \left(\frac{z_{\alpha/2}}{E}\right)^2 \times PQ = \left(\frac{1.96}{0.05}\right)^2 \times 0.2 \times 0.8 = 246.$$

$\therefore$ sample size $= 246$

Bayesian estimation:

In Bayesian estimation prior feelings about the possible values of μ are combined with the direct sample evidence which give the posterior distribution of μ approximately normally distributed with

mean $\mu_1 = \dfrac{n\bar{x}\sigma_0^2 + \mu_0\sigma^2}{n\sigma_0^2 + \sigma^2}$ and standard deviation $\sigma_1 = \sqrt{\dfrac{\sigma^2\sigma_0^2}{n\sigma_0^2 + \sigma^2}}$. In the computation

and μ_1 and σ_1, σ^2 is assumed to be known. When σ^2 is unknown which is generally the case, σ^2 is replaced by sample variance S^2 provided

$n \geq 30$(Large sample)

Bayesian interval for μ:

$(1-\alpha)$ 100% Bayesian interval for μ is given by

$$\left(\mu_1 - z_{\alpha/2}\cdot\sigma_1 \,,\, \mu_1 + z_{\alpha/2}\cdot\sigma_1\right)$$

11. A professor's feeling about the mean mark in the final examination in "probability" of a large group of students is expressed subjectively by normal distribution with $\mu_0 = 67.2\, and\, \sigma_0 = 1.5$

a) If the mean mark lies in the interval (65,70) determine the prior probability the professor should assign to the mean mark.

b) find the posterior mean and standard deviation $i.e., \mu_1\, and\, \sigma_1$ if the examinations are conducted on a random sample of 40 students yielding mean 74.9 and S.D 7.4. use S=7.4 as an estimate σ.

c) Determine the posterior probability which he will thus assign to the mean mark being in the interval (65,70) using results obtained in (b)

d) Construct 95% Bayesian interval for μ

Solution: $\mu_0 = 67.2$, $\sigma_0 = 1.5$

Sample size $n = 40$, Let x be the mean marks of the student

a) Prob that the mean mark lies in $(65, 70) = P(65 \le x \le 70)$

If $x = 65$, $Z = \frac{x - \mu_0}{\sigma_0}$

$$= \frac{65 - 67.2}{1.5} = -1.466 \simeq -1.47$$

If $x = 70$, $z = \frac{70 - 67.2}{1.5} = 1.866 \simeq 1.87$

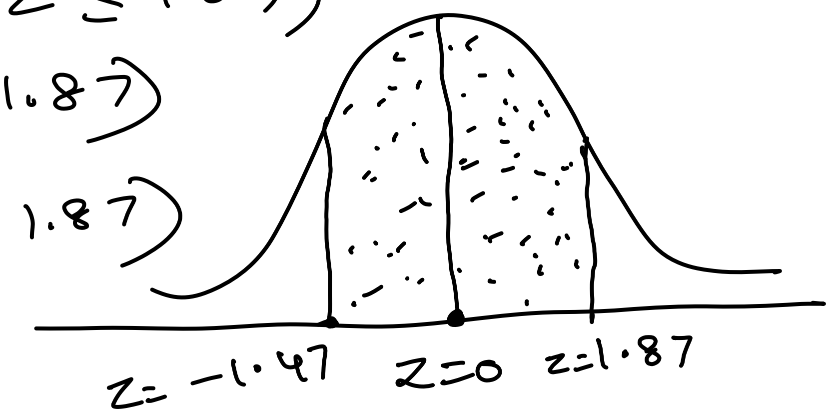
$\therefore P(65 \le x \le 70) = P(-1.47 \le z \le 1.87)$

$= P(-1.47 \le z \le 0) + P(0 \le z \le 1.87)$

$= P(0 \le z \le 1.47) + P(0 \le z \le 1.87)$

$= 0.4292 + 0.4693$

$= 0.8985$.



$\therefore$ the prior probability that the professor should assign to the mean mark $= 0.8985$

b) $\bar{x} = 74.9$, $n = 40$, $\mu_0 = 67.2$, $\sigma_0 = 1.5$

$\sigma = s = 7.4$

Posterior mean $(\mu_1) = \dfrac{n\bar{x}\sigma_0^2 + \mu_0\sigma^2}{n\sigma_0^2 + \sigma^2}$

$$= \frac{40 \times 74.9 \times (1.5)^2 + 67.2 \times (7.4)^2}{40 \times (1.5)^2 + (7.4)^2}$$

$$= 71.98 \simeq 72$$

standard deviation $(\sigma_1) = \sqrt{\dfrac{\sigma^2\sigma_0^2}{n\sigma_0^2 + \sigma^2}}$

$$= \sqrt{\frac{(7.4)^2 \times (1.5)^2}{40 \times (1.5)^2 + (7.4)^2}}$$

$$= 0.9225 \simeq 0.923$$

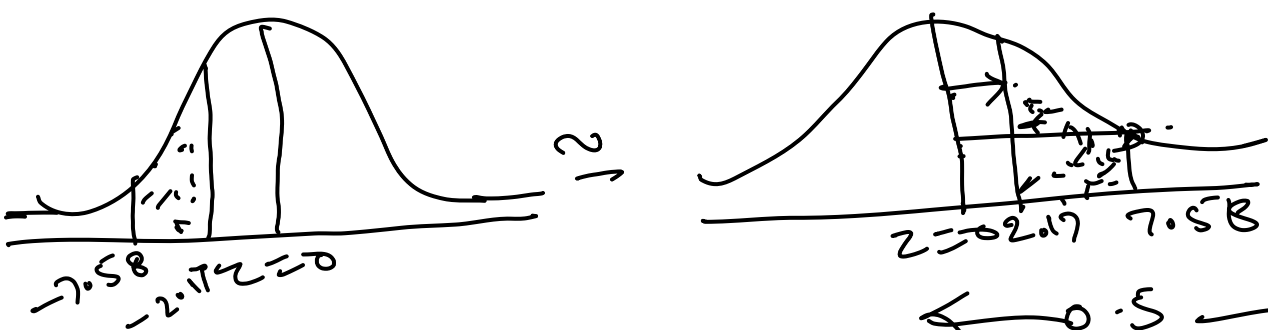
$\therefore \mu_1 = 72, \quad \sigma_1 = 0.923$

c) $(65, 70)$, $\mu_1 = 72$, $\sigma_1 = 0.923$

$x = 65$, $z = \dfrac{x - \mu_1}{\sigma_1} = \dfrac{65 - 72}{0.923} = -7.58$

$x = 70$, $z = \dfrac{70 - 72}{0.923} = -2.166 \simeq -2.17.$

$P(65 \le x \le 70) = P(-7.58 \le z \le -2.17)$



$= P(0 \le z \le 7.5) - P(0 \le z \le 2.17)$

$\quad 0.5 \qquad\qquad - 0.4850$

$=$

d) 95% Bayesian interval is

$$\left(\mu_1 - z_{\alpha/2} \cdot \sigma_1 \,,\ \mu_1 + z_{\alpha/2} \cdot \sigma_1\right)$$

For 95%, $z_{\alpha/2} = 1.96$, $\mu_1 = 71.987$, $\sigma_1 = 0.923$

$$= (71.987 - 1.96 \times 0.923 \,,\ 71.987 + 1.96 \times .923)$$

$$= (70.178 \,,\ 72.909)$$

12. The mean mark in Mathematics in common entrance that will vary from year to year. If this variation of the mean mark is expresses subjectively by normal distribution with mean $\mu_0 = 72 \, and \, \text{var}\, iance \, \sigma_0^2 = 5.76$

a) What probability can we assign to the actual mean mark being somewhere between 71.8 and 73.4 for the next years test.

b) Construct a 95% Bayesian interval for μ if the test is conducted for a random sample of 100 students from the next incoming class yielding a mean mark of 70 with S.D of 8

c) what posterior probability should we assign to the event part of (a)

Hint :

a) $P(71.8 \le x \le 73.4) \rightarrow$ prior probability $\quad z = \dfrac{x-\mu_0}{\sigma_0}$

b) $n = 100, \ \bar{x} = 70, \ s = 8$ i.e., $\sigma = 8$

c) Posterior prob $P(71.8 \le x \le 73.4) \longrightarrow \quad z = \dfrac{x-\mu_1}{\sigma_1}$

for b) & c)

First find μ_1 & σ_1 then probability

13. The mean of a random sample is an unbiased estimate of the mean of the population 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 the samples listed in (i) and assigning each sample a probability of 1/10. Prove that $\bar{x}$ is an unbiased estimate of θ. $E(\bar{x}) = \theta$

Solution:

Number of samples of size 3 that can be drawn without replacement from the population of size 5 = $5_{c_3} = \frac{5!}{3!.2!} = 10$

Let $\theta \rightarrow$ mean of the pop

$\bar{x} \rightarrow$ mean of the sample

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

$$Mean\ of\ the\ population\ (\theta) = \frac{3+6+9+15+27}{5}$$

$$= 12$$

$$Means\ of\ the\ samples\ 6,8,12,9,13,15,10,14,16,17$$

$\bar{x}$	6	8	12	9	13	15	10	14	16	17
$P(\bar{x})$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$

$$E(\bar{x}) = \sum_{i=1}^{10} \bar{x}_i P(\bar{x}_i)$$

$$E(\bar{x}) = 6 \times \frac{1}{10} + 8 \times \frac{1}{10} + 12 \times \frac{1}{10} + 13 \times \frac{1}{10} + 15 \times \frac{1}{10} + 10 \times \frac{1}{10}$$

$$+ 14 \times \frac{1}{10} + 16 \times \frac{1}{10} + 17 \times \frac{1}{10} = \frac{120}{10} = 12$$

θ = mean of population = 12

$\therefore E(\bar{x}) = \theta$

$\therefore \bar{x}$ is an unbiased estimate of θ.

14. Suppose that we observe a random variable having the binomial distribution and get x successes in n trials.

a) Show that $\frac{x}{n}$ is an unbiased estimate of the binomial parameter p.

we have to $E\left(\frac{x}{n}\right) = p$

b) Show that $\frac{x+1}{n+2}$ is not an unbiased estimate of the binomial parameter p $E\left(\frac{x+1}{n+2}\right) \neq p$

Solution:

Consider

a)
$$E\left(\frac{x}{n}\right) = \frac{1}{n} E(x)$$ [$\because$ $E(kx) = k\,E(x)$, k is constant]

$$= \frac{1}{n}.np = p$$ [$\because$ mean of Binomial dist $E(x) = \mu = np$]

$$E\left(\frac{x}{n}\right) = p$$

$$\therefore \frac{x}{n}\, is\, an\, unbiased\, estimate\, of\;\; p$$

b)
$$E\left(\frac{x+1}{n+2}\right) = E\left(\frac{x}{n+2} + \frac{1}{n+2}\right)$$

$$= \frac{1}{n+2} E(x) + E\left(\frac{1}{n+2}\right)$$ [$\because$ $E(ax+by) = a\,E(x) + b\,E(y)$]

$$= \frac{np}{n+2} + \frac{1}{n+2}$$ [$\because E(c) = c$, c is constant)

$$= \frac{np+1}{n+2} \neq p$$

$$E\left(\frac{x+1}{n+2}\right) \neq p$$

$$\therefore \frac{x+1}{n+2} \text{ is not an unbiased estimate of } p.$$