Examples Chapter(8)
Sampling Methods and
the Central Limit Theorem

GOALS

1. Explain why a sample is the only feasible way to learn about a population.
2. Describe methods to select a sample.
3. Define and construct a sampling distribution of the sample mean.
4. Explain the central limit theorem.
5. Use the central limit theorem to find probabilities of selecting possible sample means from a specified population.

The Samples

Why Sample the Population?

1. To contact the whole population would be time consuming.
2. The cost of studying all the items in a population may be prohibitive.
3. The physical impossibility of checking all items in the population.
4. The destructive nature of some tests.
5. The sample results are adequate.

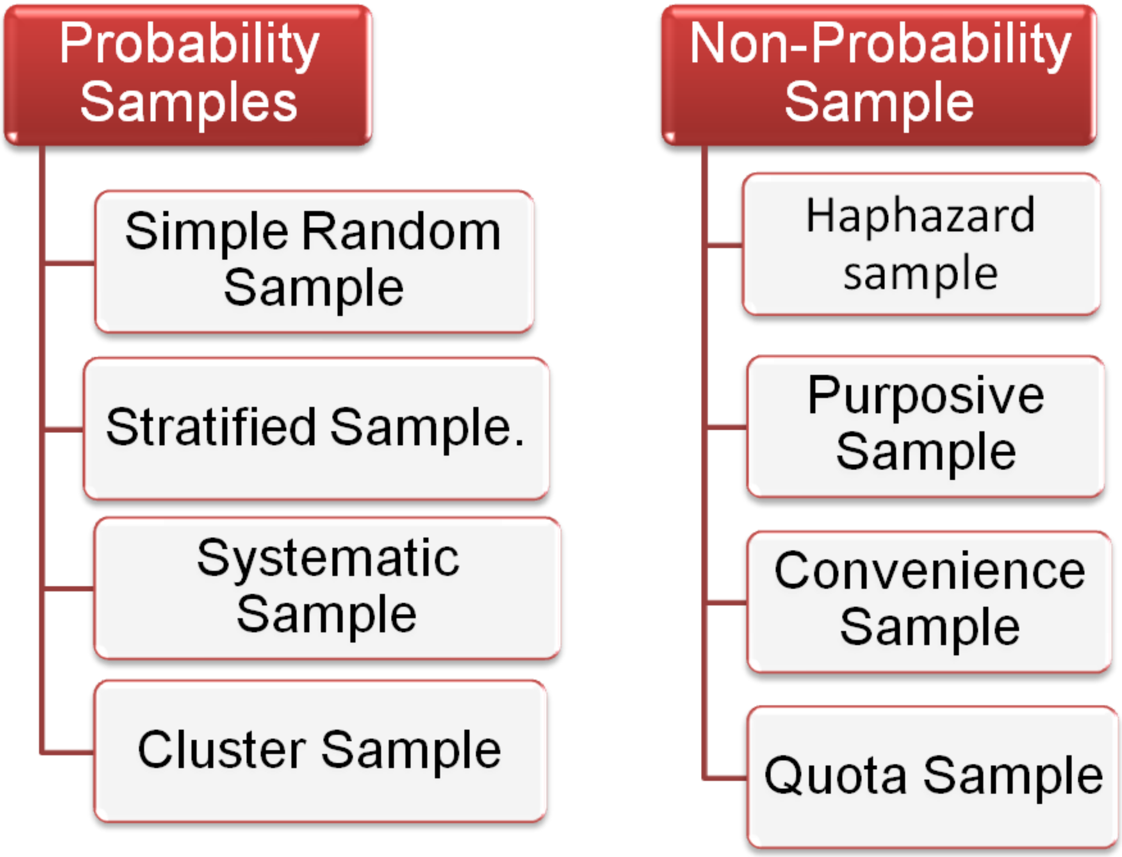
What is a Probability Sample?

A probability sample is a sample selected such that each item or person in the population being studied has a known likelihood of being included in the sample.

Four Most Commonly Used Probability Sampling Methods

1. Simple Random Sample
2. Systematic Random Sampling
3. Stratified Random Sampling
4. Cluster Sampling

Types of Samples



(1)Simple Random Sample (SRS):

Simple random sample: A sample selected so that each item or person in the population has the same chance of being included.

This can only be done by using Random Numbers Table, or any other randomization device, e.g. computers, lottery…etc.

In case that the population is homogeneous, the best sampling method to follow is the simple random sampling.

SRS is not a good design if the population under study is very **large or heterogeneous**.

How we used the Random Numbers Table? Look at the next

Example: We want selected SRS, if the

Population size (N) = 75 & a sample size (n) = 15.

Step (1): Assign number for each item in the frame from (1 to N) or (1 to 75).

Step (2): selected the starting point in the numbers table.

Suppose we starting at the intersection of (row 3 & column 2) and move horizontally and only reading two numbers according N (go horizontally or vertically) without repetition.

Step (3): We selected the items number:

71 , 57 , 18 , 37 , 22 , 75 , 65, 17 , 63 , 11,31 , 30 ,19 ,66,46

Note: We disregarded any number greater than N

Part of a table of random numbers

39634	62349	74088	65564	16379	19713	39153	69459	17986	24537		
14595	35050	40469	27478	44526	67331	93365	54526	22356	93208		
30734	71571	83722	79712	25775	65178	07763	82928	31131	30196		
64628	89126	91254	24090	25752	03091	39411	73146	06089	15630	42831	95113
43511	42082	15140	34733	68076	18292	69486	80468	80583	70361	41047	26792
78466	03395	17635	09697	82447	31405	00209	90404	99457	72570	42194	49043
24330	14939	09865	45906								

If we moving vertically the items a sample number:

71 , 57 ,18 , 26 , 11 , 37 , 3 ,61 , 40 , 47 , 46 , 22 , 25 , 44 , 35

39634	62349	74088	65564	16379	19713	39153	69459	17986	24537		
14595	35050	40469	27478	44526	67331	93365	54526	22356	93208		
30734	71571	83722	79712	25775	65178	07763	82928	31131	30196		
64628	89126	91254	24090	25752	03091	39411	73146	06089	15630	42831	95113
43511	42082	15140	34733	68076	18292	69486	80468	80583	70361	41047	26792
78466	03395	17635	09697	82447	31405	00209	90404	99457	72570	42194	49043
24330	14939	09865	45906								

(2)Systematic Random Sample

Systematic Random Sample: A random starting point is selected, and then every kth member of the population is selected.

Here the steps you need to follow in order to achieve a systematic random sample are:

1- Number the units in the population from (1 to N)
2- Determine the sample size (n).
3-Calculate the interval size, K = N/n.
4- Randomly select an integer between 1 to k.
5- Then take every kth unit from other intervals.

Example: Selection of Systematic Sample

N = 100

want n = 20

N/n = 5

select a random number from 1-5: chose 4

start with #4 and take every 5th unit

1	26	51	76
2	27	52	77
3	28	53	78
4	29	54	79
5	30	55	80
6	31	56	81
7	32	57	82
8	33	58	83
9	34	59	84
10	35	60	85
11	36	61	86
12	37	62	87
13	38	63	88
14	39	64	89
15	40	65	90
16	41	66	91
17	42	67	92
18	43	68	93
19	44	69	94
20	45	70	95
21	46	71	96
22	47	72	97
23	48	73	98
24	49	74	99
25	50	75	100

- The first unit, from the Table is number 4, then the units of the sample will be:

  4, 9, 14, 19, 24, 29…… 99.

- Systematic Sample has the advantage to SRS of good coverage of the population
- However, it has the same disadvantages.

(3)Stratified Sample

Stratified Random Sample: A population is divided into subgroups, called strata, and a sample is randomly selected from each stratum.

Stratified random sampling is used when we have heterogeneous population regarding some characteristics.

- The population is sub-divided into Strata. Each Stratum is homogeneous internally.

- A sub-sample is drawn from Each Stratum and the totality of the sub-samples constitutes the stratified sample size.
- How drawn? How many units from each stratum?
- The sub-sample can be drawn using: SRS or Systematic RS.
- The size of the sub-sample from each stratum can be determined through **: Proportional Allocation:**

Sub-sample = $\frac{\text{Size of stratum}}{\text{Size of population}}$ * sample size
= $(N_h / N) * n$

The sub-sample is taken Proportional to the size of the stratum. (Larger strata will have larger sub-sample).

Example: We want selected random sample form the following table

Sample size = 240

Stratum (Students)	**Freshman**	**Sophomore**	**Junior**	**Senior**
Size N_h	200	600	100	300

The Sub-Samples: (Proportional Allocation):

$n_1 = (200/1200) *240 = 40$

$n_2 = (600/1200) *240 = 120$

$n_3 = (100/1200) *240 = 20$

$n_4 = (300/1200) *240 = 60$

$n = 40 + 120 + 20 + 60 = 240$

(4) Multi- Stage or Cluster Sample:

Cluster sample: A population is divided into clusters using naturally occurring geographic or other boundaries. Then, clusters are randomly selected and a sample is collected by randomly selecting from each cluster.

The sample is selected into stages; each stage is composed of two steps:

- Sub-division into smaller Clusters.
- Selection of some clusters.

This goes on in stages till the ultimate unit which is indivisible, e.g. household or an individual.

An Example A study number of traffic accidents done to all Saudi Arabia:

Stage One:

1- Subdivide Saudi Arabia into cities, (25).

2- Select some cities, (say 10).

Stage Two:

1- Sub-divide each selected city into streets.

2- Select some streets from each selected city.

Review

Example (1) : Compute the mean, Variance, Standard deviation of the following population values:

6, 3, 5, 4, and 2 by using two deferent formulas

Solution:

i	X_i	X_i^2
1	**6**	**36**
2	**3**	**9**
3	**5**	**25**
4	**4**	**16**
5	**2**	**4**
Total	$\sum X_i = 20$	$\sum X_i^2 = 90$

$$\mu_x = \frac{\sum_{i=1}^{N} X_i}{N} = \frac{20}{5} = 4$$

$$\sigma^2 = \frac{1}{N}\left(\sum_{i=1}^{N} X_i^2 - \frac{\left(\sum_{i=1}^{N} X_i\right)^2}{N}\right) = \frac{\sum_{i=1}^{N} X_i^2}{N} - \mu_x^2 = \frac{90}{5} - (4)^2 = 18 - 16 = 2$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} X_i^2}{N} - \mu_x^2} = \sqrt{\frac{90}{5} - (4)^2} = \sqrt{18-16} = \sqrt{2} = 1.414$$

Summary

	Population	***Sample***
The mean	$\mu = \frac{\sum_{i=1}^{N} X_i}{N}$	$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$
Variance	$= \frac{\sum_{i=1}^{N} X^2{}_i}{N} - \mu_x^2$	$S^2 = \frac{\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1}$ $= \frac{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2}{n-1}$
Standard deviation	$\sigma_x = \sigma = \sqrt{\sigma_x^2}$	$S = \sqrt{S^2}$

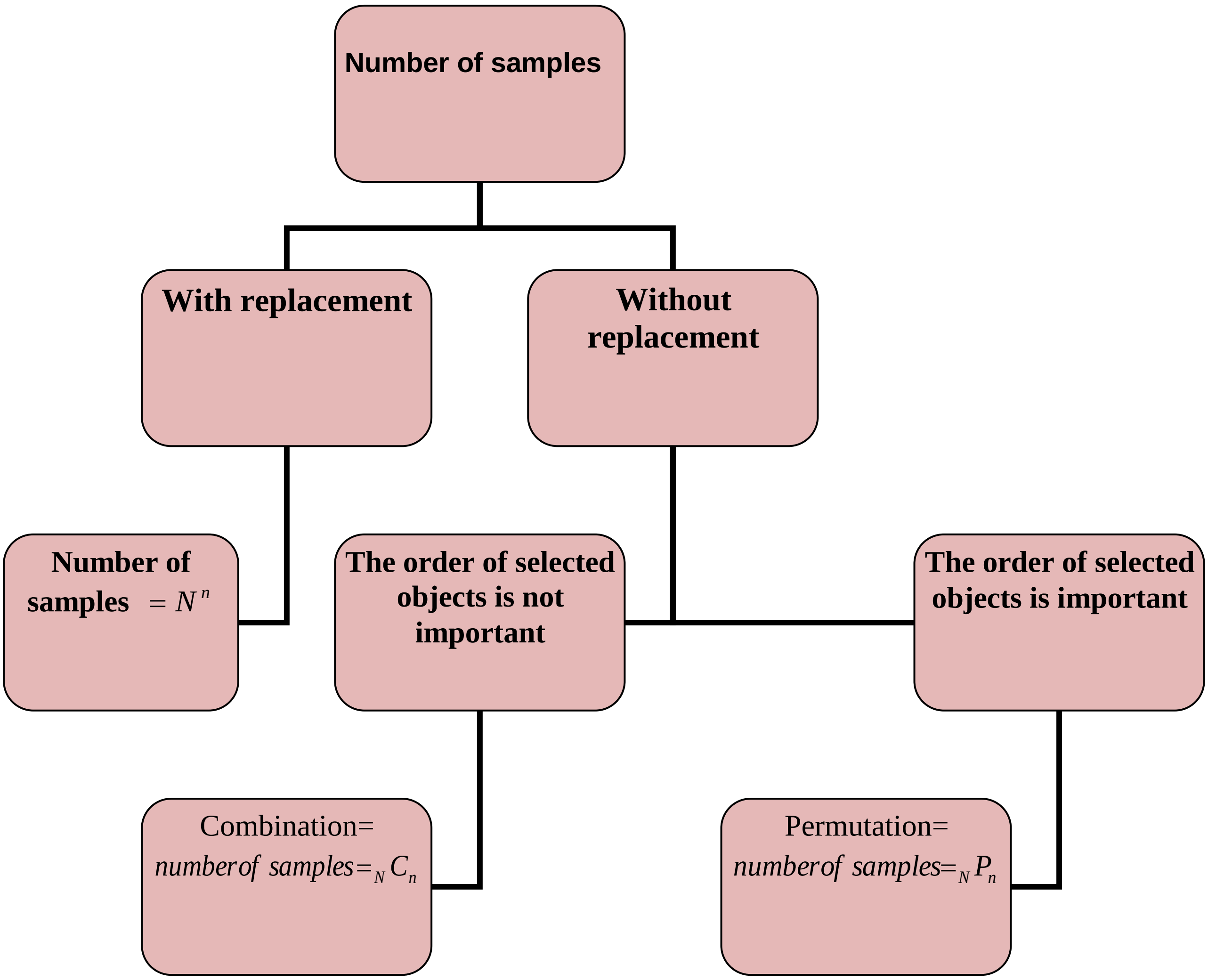
Sampling distribution:

Definition

Let $\hat{\theta}$ (read "theta hat") be any sample statistic; for example $\hat{\theta}$ could be the sample mean $\overline{X}$; the sample proportion P ; the sample variance S^2 and so on.

A sampling distribution of $\hat{\theta}$ is a probability distribution obtained by Listing all possible values that $\hat{\theta}$ could have from all possible samples of Size n and the corresponding probabilities of occurrence

How to draw sample from population



Case (1) the sampling with replacement:

$= K = N^n$ Number of samples

Case (2) the sampling without replacement:

(2-1) the order of selected objects is important

$$Number\ of\ samples = K = {}^{N}P_n = \frac{N!}{(N-n)!} = N(N-1)(N-2)(N-3)\ldots\ldots(N-n+1).$$

(2-2) The order of selected objects is not important

$$Number\ of\ samples = K = {}^{N}C_n = \frac{N!}{(N-n)!\,n!}.$$

$$= \frac{N(N-1)(N-2)(N-3)\ldots\qquad(N-n+1)\ldots\ldots}{n!}$$

Sampling distribution of the sample mean

Sampling distribution of the sample mean: A probability distribution of all possible sample means of a given sample size n .

We study three cases:

Case (1): The sampling distribution with replacement:

Example (2)

Refer to example (1) list all possible samples of 2 executives from the population and compute their means **(with replacement)**.**Find:**

1. How many different samples of 2 are possible?

2. The probability of selecting any of the possible simple random samples.

3. The mean of $\bar{X}_i$ (the mean of sampling distribution of the sample means), the variance of $\bar{X}_i$ (the variance of sampling distribution of the sample means), the standard deviation of $\bar{X}_i$ (the standard deviation of sampling distribution of the sample means).

Solution:

With replacement

1. Number of samples=k $= N^n = 5^2 = 25$

2. The probability of selecting any of the possible simple random samples

$$P = \frac{1}{N^n} = \frac{1}{25}$$

The items: 6, 3, 5, 4, 2

The possible samples:

Sample number(i)	**Samples**	$\bar{X}$
1	**6 ,6**	6
2	**6,3**	4.5
3	**6,5**	5.5
4	**6,4**	5
5	**6,2**	4
6	3,6	4.5
7	3,3	3
8	3,5	4
9	3,4	3.5
10	3,2	2.5
11	5 ,6	5.5
12	5 ,3	4
13	5,5	5
14	5, 4	4.5
15	5,2	3.5
16	4 ,6	5
17	4, 3	3.5
18	4 ,5	4.5
19	4,4	4
20	4,2	3
21	2 ,6	4
22	2, 3	2.5
23	2 ,5	3.5
24	2,4	3
25	2,2	2
		$\sum \bar{X} = 100$

$$\sum_{i=1}^{25} \bar{X}_i = 100$$

The mean of $\overline{\bar{X}_i} = \mu_{\bar{x}} = \frac{\sum_{i=1}^{25} \bar{X}_i}{k} = \frac{\bar{x}_1 + \bar{x}_2 . + \ldots\ldots\ldots. + \bar{x}_{25} ..}{K} = \frac{100}{25} = 4$

$\therefore \mu_{\bar{x}} = \mu_x$ **, In this case the statistic is an unbiased estimator of the parameter.**

The Variance of $\overline{X}_i$

Sample number (i)	Samples	$\overline{X}$	$\overline{X}_i^2$
1	6 ,6	6	36
2	6,3	4.5	20.25
3	6,5	5.5	30.25
4	6,4	5	25
5	6,2	4	16
6	3,6	4.5	20.25
7	3,3	3	9
8	3,5	4	16
9	3,4	3.5	12.25
10	3,2	2.5	6.25
11	5 ,6	5.5	30.25
12	5 ,3	4	16
13	5,5	5	25
14	5, 4	4.5	20.25
15=	5,2	3.5	12.25
16	4 ,6	5	25
17	4, 3	3.5	12.25
18	4 ,5	4.5	20.25
19	4,4	4	16
20	4,2	3	9
21	2 ,6	4	16
22	2, 3	2.5	6.25
23	2 ,5	3.5	12.25
24	2,4	3	9
25	2,2	2	4
		$\sum \overline{X} = 100$	$\sum \overline{X}^2 = 425$

$$\sigma_{\bar{x}}^2 = \frac{\sum_{i=1}^{k} \overline{X}_i^2}{k} - \mu_{\bar{x}}^2 = \frac{425}{25} - (4)^2 = 17 - 16 = 1$$

The Standard deviation of $\overline{X}_i$ =

$$\sigma_{\bar{x}} = \sqrt{\frac{\sum_{i=1}^{k} \overline{X}_i^2}{k} - \mu_{\bar{x}}^2} = \sqrt{\frac{425}{25} - (4)^2} = \sqrt{17 - 16} = \sqrt{1} = 1$$

$$\therefore \sigma_{\bar{x}}^2 = \frac{\sigma_x^2}{n} = \frac{2}{2} = 1 \qquad \therefore \sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{1.414}{\sqrt{2}} = \frac{1.414}{1.414} = 1$$

Conclusion: For case of sampling with replacement, the sample mean is a random variable with mean and variance given by:

$$\therefore \mu_{\bar{x}} = \mu_x$$

$$\sigma_{\bar{x}}^2 = \frac{\sigma_x^2}{n} \qquad \sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$$

Case (2): The sampling distribution without replacement:

Case (2-1): The order of selected objects is important.

Case (2-1): The order of selected objects is not important.

Example (3)

Refer to example (1) list all possible samples of 2 executives from the population and compute their means without replacement if:

- The order of selected objects is important.
- The order of selected objects is not important

Find:

1. How many different samples of 2 are possible?

2. The probability of selecting any of the possible simple random samples.

3. The mean of $\bar{X}_i$ (the mean of sampling distribution of the sample means), the variance of $\bar{X}_i$ (the variance of sampling distribution of the sample means), the standard deviation of $\bar{X}_i$ (the standard deviation of sampling distribution of the sample means).

Solution:

Case (2-1): Without replacement and the order of selected objects is important:

1- Number of samples

$$K={}^{N}P_n = \frac{N!}{(N-n)^!} = N(N-1)(N-2)(N-3)......(N-n+1).$$

$$={}_5P_2 = \frac{5^!}{(5-2)^!} = \frac{5^!}{3^!} = 20$$

1. **The probability of selecting any of the possible simple random samples** $P = \frac{1}{{}_N P_n} = \frac{1}{20}$

2. **The possible samples**

Sample number(i)	1	2	3	4	5	6	7	8	9	10
Samples	6,3	6,5	6,4	6,2	3,5	3,4	3,2	5,4	5,2	4,2
Sample number(i)	11	12	13	14	15	16	17	18	19	20
Samples	3,6	5,6	4,6	2,6	5,3	4,3	2 ,3	4,5	2,5	2,4

3. The mean & variance

Sample number(i)	1	2	3	4	5	6	7	8	9	10
Samples	6,3	6,5	6,4	6,2	3,5	3,4	3,2	5,4	5,2	4,2
$\bar{X}_i$	4.5	5.5	5	4	4	3.5	2.5	4.5	3.5	3
$\bar{X}^2_i$	20.25	30.25	25	16	16	12.25	6.25	20.25	12.25	9
Sample number(i)	11	12	13	14	15	16	17	18	19	20
Samples	3,6	5,6	4,6	2,6	5,3	4,3	2 ,3	4,5	2,5	2,4
$\bar{X}_i$	4.5	5.5	5	4	4	3.5	2.5	4.5	3.5	3
$\bar{X}^2_i$	20.25	30.25	25	16	16	12.25	6.25	20.25	12.25	9

$$\sum_{i=1}^{20} \bar{X}_i = 80 \qquad \sum_{i=1}^{20} \bar{X}^2_i = 335$$

The mean of $\bar{X}_i = \mu_{\bar{x}} = \frac{\sum_{i=1}^{20} \bar{X}_i}{k} = \frac{\bar{x}_1 + \bar{x}_2 . + ... + \bar{x}_{20} ..}{k} = \frac{80}{20} = 4$

$\therefore \mu_{\bar{x}} = \mu_x$ **, In this case the statistic is an unbiased estimator of the parameter**

The Variance of $\bar{X}_i$ **=**

$$\sigma_{\bar{x}}^2 = \frac{\sum_{i=1}^{k} \bar{X}_i^2}{k} - \mu_{\bar{x}}^2 = \frac{335}{20} - (4)^2 = 16.75 - 16 = 0.75$$

The Standard deviation of $\bar{X}_i$ **=**

$$\sigma_{\bar{x}} = \sqrt{\frac{\sum_{i=1}^{k} \bar{X}_i^2}{k} - \mu_{\bar{x}}^2} = \sqrt{\frac{335}{20} - (4)^2} = \sqrt{16.75 - 16} = \sqrt{0.75} = 0.87$$

$$\therefore \sigma_{\bar{x}}^2 = \frac{\sigma_x^2}{n}\left(\frac{N-n}{N-1}\right) = \frac{2}{2}\left(\frac{5-2}{5-1}\right) = (1)\left(\frac{3}{4}\right) = 0.75$$

$$\therefore \sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}\sqrt{\left(\frac{N-n}{N-1}\right)} = \frac{1.414}{\sqrt{2}}\sqrt{\left(\frac{5-2}{5-1}\right)} = \frac{1.414}{1.414}\sqrt{\left(\frac{3}{4}\right)} = \sqrt{0.75} = 0.87$$

Case (2-2):

Without replacement and the order of selected objects is not important:

1- **The Number of samples**

$$K = {}^{N}C_n = \frac{N!}{(N-n)^! \, n^!}.$$

$$= \frac{N(N-1)(N-2)(N-3)\ldots \qquad (N-n+1)\ldots\ldots}{n^!} = {}_5C_2 = \frac{5^!}{(5-2)^! \, 2^!} = 10$$

2. **The probability of selecting any of the possible simple random samples** $= P = \frac{1}{{}_NC_n} = \frac{1}{10}$

The possible samples

Sample number(i)	1	2	3	4	5	6	7	8	9	10
Samples	6,3	6,5	6,4	6,2	3,5	3,4	3,2	5,4	5,2	4,2

3. **The mean & variance**

Sample number(i)	1	2	3	4	5	6	7	8	9	10
Samples	6,3	6,5	6,4	6,2	3,5	3,4	3,2	5,4	5,2	4,2
$\bar{X}_i$	4.5	5.5	5	4	4	3.5	2.5	4.5	3.5	3
$\bar{X}^2_i$	20.25	30.25	25	16	16	12.25	6.25	20.25	12.25	9

$$\sum_{i=1}^{10} \bar{X}_i = 40 \qquad \sum_{i=1}^{10} \bar{X}^2_i = 167.5$$

The mean of $\bar{X}_i = \mu_{\bar{x}} = \frac{\sum_{i=1}^{10} \bar{X}_i}{k} = \frac{\bar{x}_1 + \bar{x}_2 + \ldots\ldots\ldots + \bar{x}_{10}}{k} = \frac{40}{10} = 4$

$\therefore \mu_{\bar{x}} = \mu_x$ **, In this case the statistic is an unbiased estimator of the parameter**

The Variance of $\bar{X}_i =$

$$\sigma_{\bar{x}}^2 = \frac{\sum_{i=1}^{k} \bar{X}_i^2}{k} - \mu_{\bar{x}}^2 = \frac{167.5}{10} - (4)^2 = 16.75 - 16 = 0.75$$

The Standard deviation of $\bar{X}_i =$

$$\sigma_{\bar{x}} = \sqrt{\frac{\sum_{i=1}^{k} \bar{X}_i^2}{k} - \mu_{\bar{x}}^2} = \sqrt{\frac{167.5}{10} - (4)^2} = \sqrt{16.75 - 16} = \sqrt{0.75} = 0.87$$

$$\therefore \sigma_{\bar{x}}^{2} = \frac{\sigma_x^2}{n}\left(\frac{N-n}{N-1}\right) = \frac{2}{2}\left(\frac{5-2}{5-1}\right) = (1)\left(\frac{3}{4}\right) = 0.75$$

$$\therefore \sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}\sqrt{\left(\frac{N-n}{N-1}\right)} = \frac{1.414}{\sqrt{2}}\sqrt{\left(\frac{5-2}{5-1}\right)} = \frac{1.414}{1.414}\sqrt{\left(\frac{3}{4}\right)} = \sqrt{0.75} = 0.87$$

Conclusion: For case of sampling without replacement, the sample mean is a random variable with mean and variance given by:

$$\therefore \mu_{\bar{x}} = \mu_x$$

$$\sigma_{\bar{x}}^{2} = \frac{\sigma_x^2}{n}\left(\frac{N-n}{N-1}\right) \quad \textbf{and} \quad \sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}\sqrt{\left(\frac{N-n}{N-1}\right)}$$

Summary

	Population	**Sample**	**Sampling distribution**
The mean	$\mu = \frac{\sum_{i=1}^{N} X_i}{N}$	$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$	$\mu_{\bar{x}} = \frac{\sum_{i=1}^{K} \bar{X}_i}{k} = \mu$
Variance	$= \frac{\sum_{i=1}^{N} X^2{}_i}{N} - \mu_x^{2}$	$S^2 = \frac{\sum_{i=1}^{n} X_i^{2} - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1}$ $= \frac{\sum_{i=1}^{n} X_i^{2} - n\bar{X}^2}{n-1}$	$\sigma_{\bar{x}}^{2} = \frac{\sum_{i=1}^{k} \bar{X}_i^2}{k} - \mu_{\bar{x}}^{2}$ With replacement $\sigma_{\bar{x}}^{2} = \frac{\sigma_x^2}{n}$ Without replacement $\sigma_{\bar{x}}^{2} = \frac{\sigma_x^2}{n}\left(\frac{N-n}{N-1}\right)$
Standard deviation	$\sigma_x = \sigma = \sqrt{\sigma_x^2}$	$S = \sqrt{S^2}$	$\sigma_{\bar{x}} = \sqrt{\sigma_{\bar{x}}^{2}}$ With replacement $\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$ Without replacement $\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$
Number of samples	**K**	**-**	**With replacement** $k = N^n$ **Without replacement** **Order important** $k = {}_N P_n$ **Order not important** $k = {}_N C_n$

Different examples

Example (4):

(A) Assume having population of 6 units, a random sample of size 2 is drawn from this population **with replacement**. The following information is :

$\sum \overline{X} = 234$ and $\sum \overline{X}^2 = 1681.5$

Find:

-The number of possible samples drawn from this population equal

-The mean of the sampling distribution of the sample mean

-The variance of the sampling distribution of the sample mean

Solution:

$$K = N^n = 6^2 = 36 \, , \mu_{\bar{x}} = \frac{\sum \overline{X}}{K} = \frac{234}{36} = 6.5 \, ,$$

$$\sigma^2{}_{\bar{x}} = \frac{\sum \overline{X}^2}{K} - (\mu_{\bar{x}})^2 = \frac{1681.5}{36} - 6.5^2 = 4.4583$$

☪☪☪☪☪☪

(B) Assume having population of 6 units, a random sample of size 2 is drawn from this population **without replacement and the order is important**. The following information is :

$\sum \overline{X} = 195$ and $\sum \overline{X}^2 = 1374.5$

Find:

-The number of possible samples drawn from this population equal

-The mean of the sampling distribution of the sample mean

-The variance of the sampling distribution of the sample mean

Solution:

$$K = {}^N P_n = {}^6 P_2 = 30 \, , \mu_{\bar{x}} = \frac{\sum \overline{X}}{K} = \frac{195}{30} = 6.5 \, ,$$

$$\sigma^2{}_{\bar{x}} = \frac{\sum \overline{X}^2}{K} - (\mu_{\bar{x}})^2 = \frac{1374.5}{30} - 6.5^2 = 3.5666$$

☪☪☪☪☪☪

(C) Assume having population of 6 units, a random sample of size 2 is drawn from this population **without replacement and the order is not important**. The following information is:

$\sum \overline{X} = 97.5$ and $\sum \overline{X}^2 = 687.25$

Find:

-The number of possible samples drawn from this population equal

-The mean of the sampling distribution of the sample mean

-The variance of the sampling distribution of the sample mean

Solution:

$K = {}^{N}C_n = {}^{6}C_2 = 15$, $\mu_{\bar{x}} = \frac{\sum \bar{X}}{K} = \frac{97.5}{15} = 6.5$,

$\sigma^2_{\bar{x}} = \frac{\sum \bar{X}^2}{K} - (\mu_{\bar{x}})^2 = \frac{687.25}{15} - 6.5^2 = 3.5666$

Example (5)

Random sample of size n were selected from populations with the means and variance given here. Find the number of samples, the mean and standard deviation of the sampling distribution of the sample mean in each case:

a. $N = 6 \quad n = 3 \quad \mu = 10 \quad \sigma^2 = 9 \quad (with\ replacement)$

b.

$N = 7 \quad n = 2 \quad \mu = 5 \quad \sigma^2 = 4 \quad (without\ replacement\ and\ the\ order\ is\ important)$

c.

$N = 8 \quad n = 3 \quad \mu = 120 \quad \sigma^2 = 16 \quad (without\ replacement\ and\ the\ order\ is\ not\ important)$

Solution:

a.

The number of samples $= k = N^n = 6^3 = 216$

$\mu_{\bar{x}} = \mu = 10$

$\sigma^2_{\bar{x}} = \frac{\sigma^2}{n} = \frac{9}{3} = 3$

$\sigma_{\bar{x}} = \sqrt{\sigma^2_{\bar{x}}} = \sqrt{3} = 1.73$

or

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{3}} = \frac{3}{1.73} = 1.73$

b.

The number of samples $= k = {}_N P_n = {}_7 P_2 = 7 \times 6 = 42$

$\mu_{\bar{x}} = \mu = 5$

$\sigma^2_{\bar{x}} = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right) = \frac{4}{2}\left(\frac{7-2}{7-1}\right) = 2\left(\frac{5}{6}\right) = \frac{10}{6} = 1.67$

$\sigma_{\bar{x}} = \sqrt{\sigma^2_{\bar{x}}} = \sqrt{1.67} = 1.29$

or

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\left(\sqrt{\frac{N-n}{N-1}}\right) = \frac{2}{\sqrt{2}}\left(\sqrt{\frac{7-2}{7-1}}\right) = \frac{2}{1.4142}\left(\sqrt{\frac{5}{6}}\right) = 1.4142\left(\sqrt{0.83}\right) = (1.4142)(0.91) = 1.29$

c.

The number of samples $= k = {}_N C_n = {}_8 C_3 = \frac{8 \times 7 \times 6}{3 \times 2 \times 1} = 56$

$\mu_{\bar{x}} = \mu = 120$

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right) = \frac{16}{3}\left(\frac{8-3}{8-1}\right) = 5.33\left(\frac{5}{7}\right) = 5.33(0.71) = 3.81$$

$$\sigma_{\bar{x}} = \sqrt{\sigma_{\bar{x}}^2} = \sqrt{3.81} = 1.95$$

or

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\left(\sqrt{\frac{N-n}{N-1}}\right) = \frac{4}{\sqrt{3}}\left(\sqrt{\frac{8-3}{8-1}}\right) = \frac{4}{1.73}\left(\sqrt{\frac{5}{7}}\right) = 2.312\left(\sqrt{0.71}\right) = (2.312)(0.84) = 1.95$$

Example (6)

Random sample of size 3 were selected from populations N. Find the number of samples:

a. With replacement.
b. Without replacement and the order is important.
c. Without replacement and the order is not important.

Solution:

a. The number of samples= $k = N^3$

b. The number of samples= $k = {}_N P_3 = N(N-1)(N-2)$

c. The number of samples= $k = {}_N C_3 = \frac{N(N-1)(N-2)}{3^!} = \frac{N(N-1)(N-2)}{6}$

Example (7)

Suppose that a random sample of size 81 from a non-normal distribution with mean 50 and variance 36(with replacement)

Find :

a. The mean of sampling distribution of the sample mean.
b. The standard error of the sampling distribution of the sample mean.

Solution:

a. $\mu_{\bar{x}} = \mu = 50$

b. $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{6}{9} = 0.67$

Example (8)

If $N = 6 \quad , \; n = 2$

Find

a. Number of samples and the probability of selecting any of the possible simple random samples in case of drawing with replacement.

b. Number of samples and the probability of selecting any of the possible simple random samples in case of drawing without replacement, order are important.

c. Number of samples and the probability of selecting any of the possible simple random samples in case of drawing without replacement, order are not important.

Solution

a.

$k = N^n = 6^2 = 36$

$P = \frac{1}{N^n} = \frac{1}{36}$

b.

$k = {}_N P_n = \frac{N!}{(N-n)!} = N(N-1) = {}_6 P_2 = 6 \times 5 = 30$

$P = \frac{1}{{}_N P_n} = \frac{1}{30}$

c.

$k = {}_N C_n = \frac{N!}{n!(N-n)!} = \frac{N(N-1)}{n!} = {}_6 C_2 = \frac{6!}{2!(6-2)!} = \frac{6 \times 5}{2 \times 1} = 15$

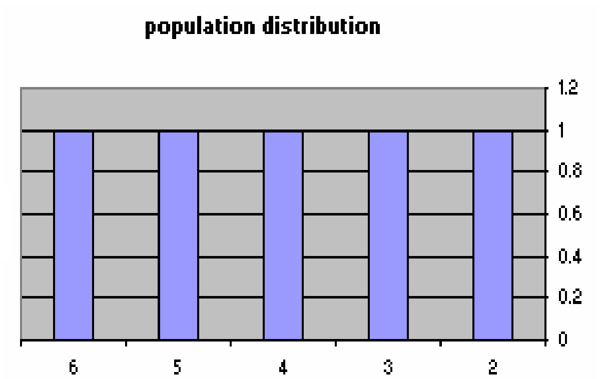
Note:

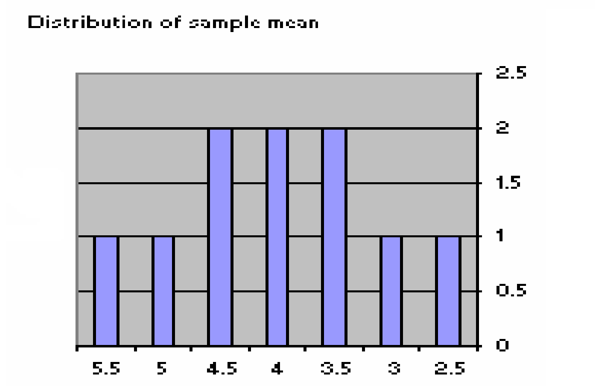
1. The mean of the sample means is exactly equal to the population mean : (ex.2 &3)

$$\mu_{\bar{x}} = \mu = 4$$

2. The dispersion (variation, spread) of the sampling distribution of sample means is less than the population distribution.(with out replacement)

3. The sampling distribution of sample means tends to become bell-shape and to approximate the normal probability distribution. (ex.3)





$\sigma_{\bar{X}} = +\sqrt{\sigma^2_{\bar{X}}}$ is the "standard deviation of the sampling distribution of the sample mean"; and always refer to it as the "standard error of the mean".

4- The factor $\dfrac{N-n}{N-1}$; is called the finite population correction factor and can be ignored if :

$\dfrac{n}{N} \leq 0.05$ (i.e. do not use the finite population correction unless the sample is more than 5 percent of the size of the population).

Some important theorems

Theorem1:

If the population follows a normal probability distribution, then for any sample size the sampling distribution of the sample mean will be also normal.

Suppose that $X \sim N\left(\mu, \sigma_X^2\right)$ **and let** $\bar{X}$ **be the mean of a sample of size** n **then** $\bar{X} \sim N\left(\mu, \sigma_{\bar{X}}^2\right)$ **where;**

$$\sigma_{\bar{X}}^2 = \begin{cases} \dfrac{\sigma_X^2}{n}; & \textit{in case of replacment} \\ \\ \dfrac{\sigma_X^2}{n}\left(\dfrac{N-n}{N-1}\right); & \textit{in case of without replacment} \end{cases}$$

Theorem2 (The Central limit Theorem):

If all sample of a particular size are selected from any population, the sampling distribution of the sample mean is approximately a normal distribution. This approximation improves with larger sample ($n \geq 30$).

For a population with a mean μ and a variance σ^2 the sampling distribution of the means of all possible samples of size n generated from the population will be approximately normally distributed with mean μ and standard deviation $\sigma/\sqrt{n}$ as the sample size n becomes larger **(**$n \geq 30$**).**

Facts:

(1) The expected value of the sample mean equal the population mean:

$$E\left(\bar{X}\right) = \mu_{\bar{X}} = \mu$$

Where $\mu_{\bar{X}}$ is the mean of the Sampling distribution of means.
And μ is the mean of the population.
(2) If a population is infinite or if sampling is with replacement, then the variance of the Sampling distribution of means is:

. $$\sigma_{\bar{X}}^{\ 2} = \frac{\sigma^2}{n}$$

(3) If a population is of size N, if sampling is without replacement, then the variance of the Sampling distribution of means is:

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

(4) Sampling distribution of sample means will be exactly normal if the population is normally distributed.

(5) If a population from which samples are taken is normally distributed with mean μ and variance σ^2 then the sample mean is normally distributed with mean μ and variance $\frac{\sigma^2}{n}$.

(If the population distribution is normal, Sampling distribution of the $\bar{X}$ will be exactly normal $\bar{X} \sim N(\mu, \sigma_{\bar{x}}^2)$).

(6) Suppose that X is a random variable having some distribution function with mean μ and variance σ^2. Then the sampling distribution of the mean approaches a normal distribution with mean μ and standard deviation $\sigma/\sqrt{n}$ as the **sample size** n becomes larger, irrespective of the shape of the original distribution (Central Limit Theorem).

(If the population distribution is non-normal, Sampling distribution of the $\bar{X}$ will be approximately normal by Central limit theorem $\bar{X} \approx N(\mu, \sigma_{\bar{x}}^2)$**).**

(7) As the sample sizes increases, the variability of each sampling distribution decreases so that they become increasingly more leptokurtic

Example (9)

Suppose that the random variable X represent the IQ (Intelligence quotient) score for students at certain university, and that $X \sim N(100, 75^2)$. A random sample of size 25 students is selected. Find the following:

1. Probability that the mean IQ score computed from the sample will be grater than 125.
2. Probability that the mean IQ score computed from the sample will be less than 80.
3. Probability that the mean IQ score computed from the sample will be between 70 and 130.

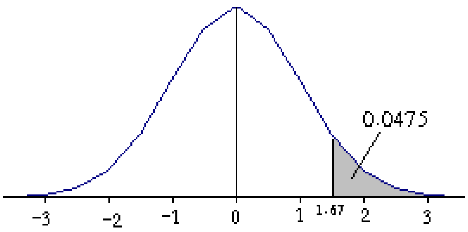
Solution:

1. Note that $\bar{X} \sim N(100, 225)$

   $\mu = 100$, $\sigma^2{}_{\bar{x}} = \frac{75^2}{25} = \frac{5625}{25} = 225$

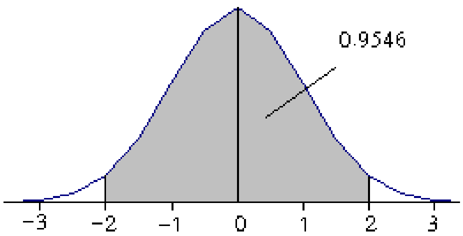
   $\sigma_{\bar{x}} = \sqrt{225} = 15$

   $P(\bar{X} > 125) = P\left(Z > \frac{125 - 100}{15} \right) = 0.0475$



2. $P(\bar{X} < 80) = P\left(Z < \frac{80 - 100}{15} \right) = \Phi(-1.33) = 0.0918$

3. $P(70 < \bar{X} < 130) = P\left(\frac{70 - 100}{15} < Z < \frac{130 - 100}{15} \right) = 0.9546$



Sampling distribution of the difference between two Independent samples means

(Normal populations)

Setting : (Dr. Amina Ali Saleh:

Suppose that there are two populations 1 and 2; and that X_1 is a random variable defined on the first population and that X_2 is another random variable defined on the second population. A random sample of size n_1 is selected from population 1; let this random sample be denoted by:

$$X_{11}, X_{12}, X_{13}, \ldots\ldots, X_{1n_1}$$

The sample mean of this sample is denoted by :

$$\bar{X}_1 = \frac{\sum X_{1i}}{n_1}$$

Similarly, another independent random sample of size n_2 is selected from population 2; let this random sample be denoted by:

$$X_{21}, X_{22}, X_{23}, \ldots\ldots, X_{2n_2}$$

The sample mean of this sample is denoted by:

$$\bar{X}_2 = \frac{\sum X_{2i}}{n_2}$$

We request the sampling distribution of the random Variable $\bar{X}_1 - \bar{X}_2$

The sampling distribution of the difference between two independent sample means can be obtained in a similar way as the sampling distribution of one sample mean discussed before.

Characteristics of the sampling distribution of the difference between two sample means $\bar{X}_1 - \bar{X}_2$:

$$\mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2$$

$$\sigma^2_{\bar{X}_1 - \bar{X}_2} = \sigma^2_{\bar{X}_1} + \sigma^2_{\bar{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

If $X_i \sim N(\mu_i, \sigma_i^2); \quad i = 1, 2$

Hence $\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_{\bar{X}_1 - \bar{X}_2}, \sigma^2_{\bar{X}_1 - \bar{X}_2}\right)$

$$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

Then

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1)$$

Example (10)

Assume there are two types of plant. The mean height of type 1 is 32 while the mean height of type 2 is 22. The variances of the two types are 60 and 70 respectively and the heights of both types are normally distributed. Two samples randomly selected from the two populations, from the population of type 1, 10 plants are selected and from the population of type 2, 14 plants are selected. What is the probability that the mean of the 10 plants of type 1 will exceed the mean of the 14 plants of type 2 by 5 or more?

Solution:

From the above information we find that:

$$\mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2$$
$$= 32 - 22 = 10$$

And;

$$\sigma^2_{\bar{X}_1 - \bar{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$
$$= \frac{60}{10} + \frac{70}{14} = 11 = (3.317)^2$$

And hence:

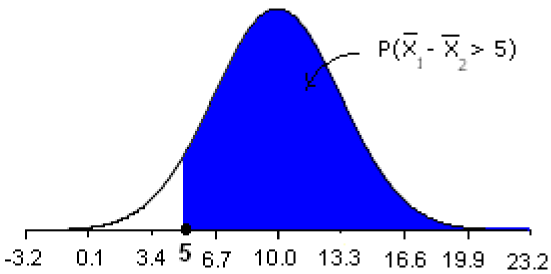
$$\bar{X}_1 - \bar{X}_2 \sim N(10, 11)$$

Then we request the following probability:

$$P(\bar{X}_1 - \bar{X}_2 > 5) = P\left(Z > \frac{5-10}{\sqrt{11}}\right)$$

$$= P(Z > -1.51) = 0.934$$

The sampling distribution is shown in the following figure; the area above 5 is shaded.



Example (11)

The electric light bulbs of manufacturer A have a mean lifetime of 1400 hours with a standard deviation of 200 hours, while those of manufacturer B have a mean lifetime of 1200 hours with a standard deviation of 100 hours. If random sample of 125 bulbs of each brand are tested (If the population distribution is normal), **Find:**

1. The probability that the brand A bulbs will have a mean lifetime which is at most 1350 hours.
2. The probability that the brand B bulbs will have a mean lifetime which is at most1220 hours.
3. The probability that the brand A bulbs will have a mean lifetime which is at least 160 hours more than the brand B bulbs.
4. The probability that the brand A bulbs will have a mean lifetime which is at least 250 hours more than the brand B bulbs

Solution:

1. The probability that the brand A bulbs will have a mean lifetime which is at most 1350 hours.

$$\mu_A = ,1400 \quad \sigma^2_{\bar{x}} = \frac{40000}{125} = 320$$

$$\sigma_{\bar{x}} = \sqrt{320} = 17.89$$

$$\mu_A = 1400\ hr \qquad \sigma_A = 200 \qquad n_A = 125$$

$\bar{X} \sim N(1400,320)$

$$P(\bar{X}_A \le 1350) = P\left(Z \le \frac{\bar{X}_A - \mu_A}{\frac{\sigma_A}{\sqrt{n_A}}} \right) = P\left(Z \le \frac{1350-1400}{\frac{200}{\sqrt{125}}} \right) = P\left(Z \le \frac{-50}{\frac{200}{11.18}} \right)$$

$$= P\left(Z \le \frac{-50}{17.89} \right) = P(Z \le -2.79) = 0.5 - \phi(-2.79) = 0.5 - 0.4974 = 0.0026$$

2. The probability that the brand B bulbs will have a mean lifetime which is at most1220 hours.

3. $\mu_B = ,1200 \quad \sigma^2{}_{\bar{x}} = \frac{10000}{125} = 80$

$\sigma_{\bar{x}} = \sqrt{80} = 8.94$

$\mu_B = 1200\ hr \qquad \sigma_B = 100\ hr \qquad n_B = 125$

$\bar{X} \sim N(1200,80)$

$$P(\bar{X}_B \le 1220) = P\left(Z \le \frac{\bar{X}_B - \mu_B}{\frac{\sigma_B}{\sqrt{n_B}}} \right) = P\left(Z \le \frac{1220-1200}{\frac{100}{\sqrt{125}}} \right) = P\left(Z \le \frac{20}{\frac{100}{11.18}} \right)$$

$$= P\left(Z \le \frac{20}{8.94} \right) = P(Z \le 2.24) = 0.5 + \phi(2.24) = 0.5 + 0.4875 = 0.9875$$

3. The probability that the brand A bulbs will have a mean lifetime which is at least 160 hours more than the brand B bulbs.

$$\bar{X}_A - \bar{X}_B = \mu_{\bar{X}_A - \bar{X}_B} = \mu_A - \mu_B = 1400 - 1200 = 200\,hr$$

$$\sigma_{\bar{X}_A - \bar{X}_B} = \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}} = \sqrt{\frac{40000}{125} + \frac{10000}{125}} = \sqrt{320+80} = \sqrt{400} = 20$$

$$\bar{X}_A - \bar{X}_B \sim N(200,20^2)$$

$$P(\bar{X}_A - \bar{X}_B > 160) = P\left(Z > \frac{(\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B)}{\sigma_{\bar{X}_A - \bar{X}_B}} \right) = P\left(Z > \frac{160-200}{20} \right) = P\left(Z > \frac{-40}{20} \right)$$

$$= P(Z > -2) = 0.5 + \phi(-2) = 0.5 + 0.4772 = 0.9772$$

4. The probability that the brand A bulbs will have a mean lifetime which is at least 250 hours more than the brand B bulbs**.**

$$\bar{X}_A - \bar{X}_B = \mu_{\bar{X}_A - \bar{X}_B} = \mu_A - \mu_B = 1400 - 1200 = 200\,hr$$

$$\sigma_{\bar{X}_A - \bar{X}_B} = \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}} = \sqrt{\frac{40000}{125} + \frac{10000}{125}} = \sqrt{320+80} = \sqrt{400} = 20$$

$$\bar{X}_A - \bar{X}_B \sim N(200,20^2)$$

$$P\left(\overline{X}_A - \overline{X}_B \geq 250\right) = P\left(Z \geq \frac{\left(\overline{X}_A - \overline{X}_B\right) - \left(\mu_A - \mu_B\right)}{\sigma_{\overline{X}_A - \overline{X}_B}}\right) = P\left(Z \geq \frac{250 - 200}{20}\right) = P\left(Z \geq \frac{50}{20}\right)$$
$$= P\left(Z \geq 2.5\right) = 0.5 - \phi\left(2.5\right) = 0.5 - 0.4938 = 0.0062$$

Sampling distribution of sample proportion

(Large samples)

If a random sample of n observations is selected form a binomial population with parameter p, then the sampling distribution of the sample proportion (P = x/n)

Will have a mean π and a variance $\frac{\pi(1-\pi)}{n}$

When the sample size n is large, the sampling distribution of p can be approximated by a normal distribution. The approximation will be adequate if $n \geq 100$, $0.05 < \pi < 0.95$,

$(n\pi \geq 5 \, , n(1-\pi) \geq 5)$ **If** $X \sim Bin(n, \pi)$ and n large

$$\therefore X \sim N(\mu, \sigma^2) \quad where \;\; \mu = n\pi \;\; , \;\; \sigma^2 = n\pi(1-\pi)$$

$$P \sim N(\mu_p, \sigma_p^2) \qquad where \;\; \mu_p = \pi \;\; , \;\; \sigma_p^2 = \frac{\pi(1-\pi)}{n}$$

Where,

$$the\, mean\; of\;\; samlling\, distribution\, of\;\; proportion = \mu_p = E(P) = E\left(\frac{X}{n}\right) = \frac{1}{n}E(X) = \frac{1}{n}(n\pi) = \pi$$

$$the\, \mathrm{var}\, iance\, of\;\; samlling\, distribution\, of\;\; proportion = \sigma_p^2 = Var(P)$$

$$= Var(\frac{X}{n}) = \frac{1}{n^2}Var(X) = \frac{1}{n^2}(n\pi(1-\pi)) = \frac{n\pi(1-\pi)}{n^2} = \frac{\pi(1-\pi)}{n}$$

Example (12)

In a survey, 500 parents in the USA were asked about the important of sports for boys and girls .of the parents interviewed, 300 agreed that the genders are equal and should have equal opportunities to participate in sports.

Find:

1. The sample proportion of parents who agree that the genders are equal and should have equal opportunities.
2. The true proportion in the population is equal to some unknown value that call π ,estimate the true proportion π

3. The mean of the sample distribution of proportion of parents who agree that the genders are equal and should have equal opportunities.
4. The variance of the sample distribution of proportion ,in case of sampling with replacement,

Solution:

1. $P = \frac{X}{n} = \frac{300}{500} = 0.6$

2. $\pi = 0.6$

3. $\mu_p = \frac{n\pi}{n} = \pi = 0.6$

4. $\sigma_p^2 = \frac{\pi(1-\pi)}{n} = \frac{0.6(0.4)}{500} = 0.00048$

$\sigma_p = 0.0219$

$P \dot{\sim} N(0.6, 0.00048)$

Example (13)

Suppose that $X \sim BIN(100.0.3))$. Using the normal approximation for the binomial probabilities find the following probability:

$1.P(P \geq 0.25),\ 2.P(P \leq 0.25)$

Solution:

$\mu_p = \pi = 0.3$, $\sigma_p^2 = \frac{\pi(1-\pi)}{n} = \frac{0.3 \times 0.7}{100} = 0.0021$, $\sigma_p = \sqrt{0.0021} = 0.0458$

$P \sim N(0.3, 0.0021)$

1.

$P(P \geq 0.25) \doteq P\left(Z \geq \frac{0.25 - 0.3}{0.0458}\right) = P\left(Z \geq \frac{-0.05}{0.0458}\right)$

$= P(Z \geq -1.09) =$

$= 0.5 + 0.3621 = 0.8621$

2.

$P(P \leq 0.25) \doteq P\left(Z \leq \frac{0.25 - 0.3}{0.0458}\right) = P\left(Z \leq \frac{-0.05}{0.0458}\right)$

$= P(Z \leq -1.09)$

$= 0.5 - 0.3621 = 0.1379$

Example(14)

The "Academic Center of Human Studies" has a small medical center which has limited facilities and in serious cases the patient should transfer to the nearest hospital for treatment, however this is consider as a rare case. Fifty-four percent of all students are in favor of improving the medical center. A random sample of 1000 students is selected, and asked if they in favor of improving the medical center. Find the following:

1. The probability that the percentage of students in favor of improving the medical center is less than 50%.

Solution:

For the population of all students, the proportion in favor of improving the medical center π ; is equal to 0.54. For a random sample of size 1000, we can verify that both $n\pi$ and $n(1-\pi)$ are greater than 5.

Then, the sample proportion of students in favor of improving the medical center follows approximately a normal distribution with mean and variance given by:

$\mu_P = 0.54$;

$$\sigma_P^2 = \frac{(0.54)(1-0.54)}{1000} = 0.0002484 = (0.0158)^2$$

hence;

$$P \sim N\left(0.54,\, 0.0158^2\right)$$

First: Finding the probability that the percentage of students in favor of improving the medical center is less than 50%, using the correction for continuity, this probability is computed as follows:

$$P(P < 0.5) \doteq P\left(Z \le \frac{0.5-0.54}{0.0158}\right) = P(Z \le -2.53) = 0.5 - 0.4943 = 0.0057$$

Remark :

1. The sampling distribution of P is completely described by both n and π .
2. In case of sampling without replacement, the variance of P becomes:

$$\sigma_P^2 = \frac{\pi(1-\pi)}{n}\left(\frac{N-n}{N-1}\right)$$

3. The sampling distribution of P is approximately normally distributed if n is fairly large($n \ge 100$) and π not close to 0 or 1 (0.05 < π < 0.95). A rule of thumb is that the approximation is good if both $n\,\pi$ and $n(1-\pi)$ are greater than or equal 5
4. As sample size an increases, the standard error of P is decreases.

Sampling distribution of the difference between two sample proportions

(Large and independent samples)

Theorem:

Given that P_1 and P_2 are sample proportions from two independent

$$P_1 - P_2 \sim N\left(\pi_1 - \pi_2, \frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}\right)$$

Provided that $n_i \geq 100$, $0.05 \leq \pi_i \leq 0.95$, $n_i \pi_i > 5$ and $n_i(1-\pi_i) > 5$ for $i = 1, 2.$ $\mu_{P_1 - P_2} = \pi_1 - \pi_2$

Example (15)

Assume that 0.8 of university graduates are able to pass ICDL test, while only 0.5 of secondary school graduates are able to pass the test. Suppose that 120 persons are sampled from the population of university graduates and that 125 persons are sampled from the population of secondary school graduates, and the ICDL test is held to all of them. What is the probability that $P_1 - P_2 \leq 0.2$?

Solution: find that:

$$\mu_{P_1 - P_2} = \pi_1 - \pi_2 = 0.8 - 0.5 = 0.3$$

And

$$\sigma^2_{P_1 - P_2} = \frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2} \qquad = \frac{(0.8)(0.2)}{120} + \frac{(0.5)(0.5)}{125}$$

$$0.0013 + 0.002 == 0.0033$$

$$\sigma_{p_1 - p_2} = 0.0574$$

Then;

$$P(P_1 - P_2 \leq 0.2) = P\left(Z \leq \frac{0.2 - 0.3}{0.0574}\right) = \left(Z \leq \frac{-0.1}{0.0574}\right)$$

$$= P(Z \leq -1.74) = 0.5 - 0.4591 = 0.0409$$