Lab 1:

The following data represent the number of days absent per year in a population of the employees of a small community hospital: 8, 3, 1, 11, 4 and 7.

(a) Find the population mean and population standard deviation.

(b) Consider all possible samples of size two i.e., n = 2 which can be drawn with simple random sampling with replacement (SRSWR) and sampling without replacement (SRSWOR)

(c) Compute mean for each possible samples for both sampling plans

(d) Draw histogram of distribution of population values and sampling distribution of means. Comment on the shape of the distribution of population values and shape of sampling distribution of means.

(e) Find mean of these samples means and verify that population mean is equals to

mean of the sample means i.e., $E(\overline{x})=\mu$ for both sampling plans

(f) Find the standard error of means for each sample and verify following results.

$S.D.(\overline{X}) = \frac{\sigma^2}{\sqrt{n}}$ *for SRSWR*

$S.D.(\overline{X}) = \frac{\sigma^2}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$ *for SRSWOR*

<u>Solution:</u>

Data,

X= Random variable X = No. of days an employee of a small community hospital remains absent in a year.

a) Given data: 8, 3, 1, 11, 4 and 7

Enter the data in the excel sheet, cells from B5:B10 with the title X in B4 cell.

The population mean is calculated by using the formula:

Mean(μ)= AVERAGE(B5:B10)

=5.6667

And, the population standard deviation is calculated as:

SD(σ)= STDEV.P(B5:B10)

=3.3500

b) Considering the possible samples of size two i.e., n=2 which was drawn with simple random sampling with replacement(SRSWR) and sampling without replacement(SRSWOR):

No. of possible samples using SRSWOR= Ncn=15

In excel we used the exact formula: COMBIN(6,2)

No. of possible samples using SRSWR= N^n

=6^2

=36

	First obs.	Second obs.		
Sample No.	x1	x2	Sample mean	Sample SD
1	8	3	5.5	3.5355
2	8	1	4.5	4.9497
3	8	11	9.5	2.1213
4	8	4	6	2.8284
5	8	7	7.5	0.7071
6	3	1	2	1.4142
7	3	11	7	5.6569
8	3	4	3.5	0.7071
9	3	7	5	2.8284
10	1	11	6	7.0711
11	1	4	2.5	2.1213
12	1	7	4	4.2426
13	11	4	7.5	4.9497
14	11	7	9	2.8284
15	4	7	5.5	2.1213

Sampling without replacement (SRSWOR) :

Sampling with replacement (SRSWR) :

Sample No.	First obs x1	Second obs x2	sample mean	sample sd
1	8	8	8	0.0000
2	8	3	5.5	3.5355
3	8	1	4.5	4.9497
4	8	11	9.5	2.1213
5	8	4	6	2.8284

6	8	7	7.5	0.7071
7	3	8	5.5	3.5355
8	3	3	3	0.0000
9	3	1	2	1.4142
10	3	11	7	5.6569
11	3	4	3.5	0.7071
12	3	7	5	2.8284
13	1	8	4.5	4.9497
14	1	3	2	1.4142
15	1	1	1	0.0000
16	1	11	6	7.0711
17	1	4	2.5	2.1213
18	1	7	4	4.2426
19	11	8	9.5	2.1213
20	11	3	7	5.6569
21	11	1	6	7.0711
22	11	11	11	0.0000
23	11	4	7.5	4.9497
24	11	7	9	2.8284
25	4	8	6	2.8284
26	4	3	3.5	0.7071
27	4	1	2.5	2.1213
28	4	11	7.5	4.9497
29	4	4	4	0.0000

30	4	7	5.5	2.1213
31	7	8	7.5	0.7071
32	7	3	5	2.8284
33	7	1	4	4.2426
34	7	11	9	2.8284
35	7	4	5.5	2.1213
36	7	7	7	0.0000

c) *Computing the mean for each possible samples for both sampling plans Sampling without replacement(SRSWOR):*

$E(\overline{x})$= *AVERAGE(E22:E36)*

=5.6667

Where the data range E22:E36 is sample mean of SRSWOR.

Sampling with replacement(SRSWR):

$E(\overline{x})$= *AVERAGE(E41:E76)*

=5.6667

Where the data range E41:E76 is sample mean of SRSWR.

***Conclusion**: The expected value of sample mean is equals to population means for both sampling plans i.e. SRSWOR and SRSWR. It means that sample mean is an unbiased estimator of population mean.*

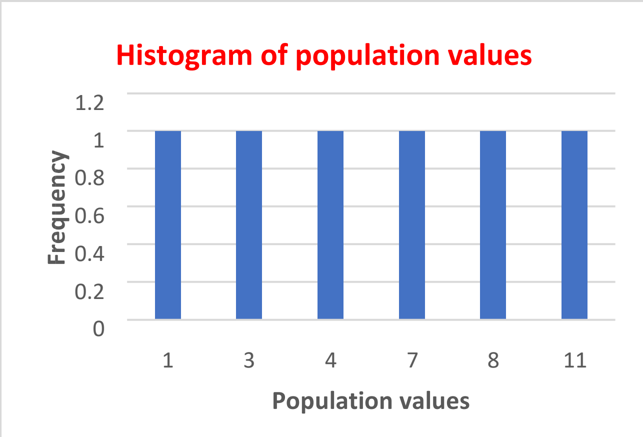
d) *Histogram of distribution of population values and sampling distribution of means.*

Histogram of Population values:

The frequency table is represented as:

X	*f*
1	1
3	1
4	1
7	1
8	1
11	1

Histogram represented by above frequency table is:



Distribution of population values is uniform.

Histogram of sample means (SRSWOR):

Range is calculated as: R= MAX(E22:E36)-MIN(E22:E36) No. of classes

*(k)= 1+3.3221*LOG10(n) where "n" is sample size*

*k= 1+3.3221*LOG10(15)*

=4.9071

Class width(h)= R/k

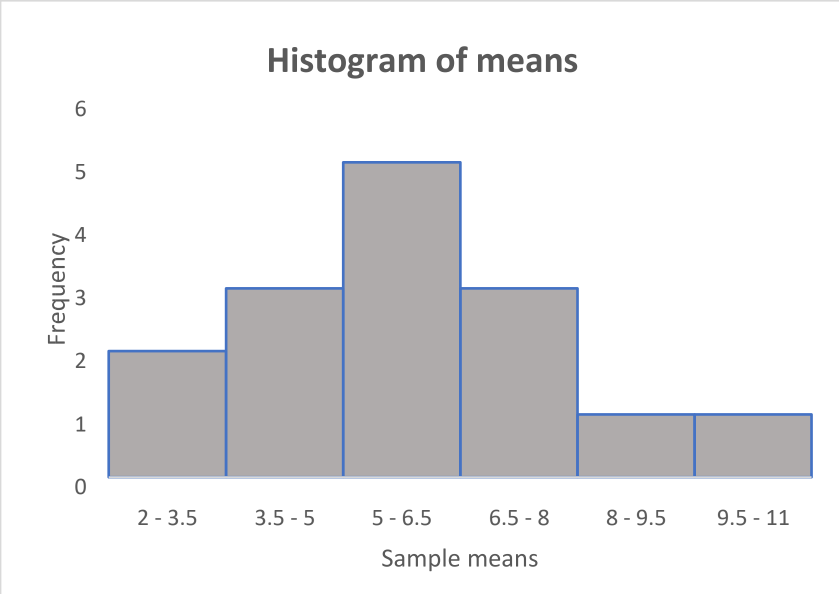
= C100/C101

=1.5284

From above information we construct the table below:

Class	*Bin Limit*	*Frequency*
2 - 3.5	*3.4*	*2*
3.5 - 5	*4.9*	*3*
5 - 6.5	*6.4*	*5*
6.5 - 8	*7.9*	*3*
8 - 9.5	*9.4*	*1*
9.5 - 11	*10.9*	*1*
	Total	*15*

Histogram represented by above table:



Distribution of sample mean is approximately normal.

Histogram of sample means (SRSWR):

Range is calculated as: R= MAX(E41:E76)-MIN(E41:E76) No. of classes

*(k)= 1+3.3221*LOG10(n) where "n" is sample size*

*k= 1+3.3221*LOG10(36)*

= 6.170192538

Class width(h)= R/k

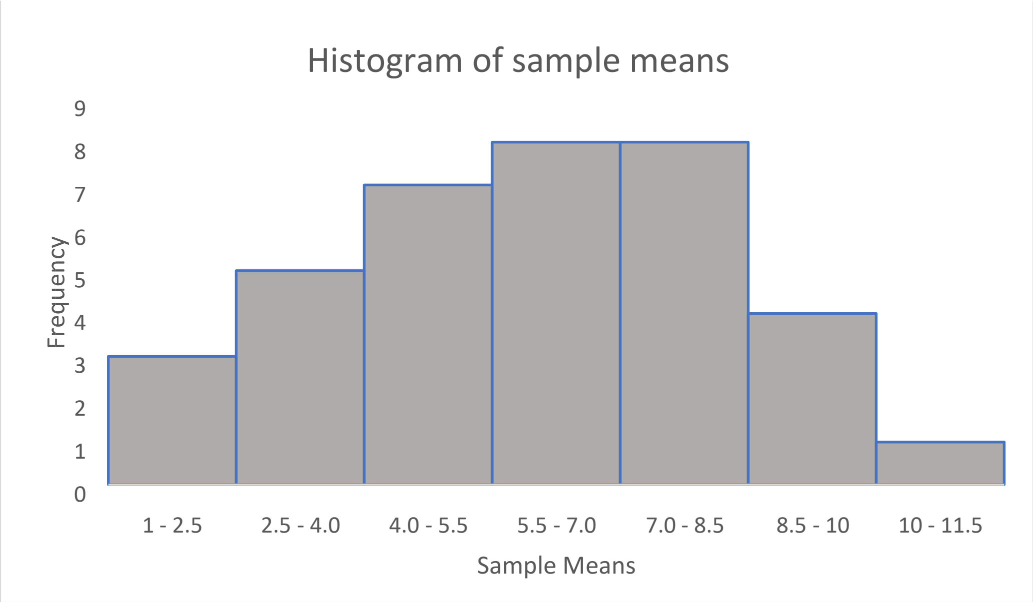
= C132/C133

= 1.620694968

From above information we construct the table below:

Sample Mean	*Bin Limit*	*Frequency*
1 - 2.5	*2.4*	*3*
2.5 - 4.0	*3.9*	*5*
4.0 - 5.5	*5.4*	*7*
5.5 - 7.0	*6.9*	*8*
7.0 - 8.5	*8.4*	*8*
8.5 - 10	*9.9*	*4*
10 - 11.5	*11.4*	*1*
	Total	*36*

Histogram represented by above table:



Distribution of sample mean is approximately normal.

Conclusion: *Shape of values of population has uniform distribution while the shape of the sample means in both SRSWOR and SRSWR sampling plans are approximately normally distributed.*

e) The population mean is calculated by using the formula:

Mean(μ)=AVERAGE(B5:B10)

=5.6667

Sampling without replacement(SRSWOR):

$E(\overline{X})$= *AVERAGE(E22:E36)*

=5.6667

Where the data range E22:E36 is sample mean of SRSWOR.

Sampling with replacement(SRSWR):

$E(\overline{X}r)$= *AVERAGE(E41:E76)*

=5.6667

Where the data range E41:E76 is the sample mean of SRSWR.

Hence it verifies that the population mean is equal to mean of the sample means i.e.,$E(\overline{x})= \mu$ *for both sampling plans. Hence, sample mean is an unbiased estimator of population mean(μ).*

f) *For SRSWOR:*

$$S.E.(\bar{X}) = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

LHS=SE=Standard deviation of $\bar{x}$

=√ (((5.5-5.67)^2+ (4.5-5.67)^2+……. + (5.5-5.67)^2)/15)

=2.1186

RHS= $\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$

=3.35/√2√ (6-2/6-1

=2.1186

Hence, verified.

For SRSWR:

To show: SE = σ/√n

LHS: SE = SD of $\overline{x}$

=√ (((8-0.567)^2+ (5.5-5.67)^2+…+ (7-0.567)^2) /36)

=2.368

RHS = σ/√n

= 3.35/√2

=2.368

Hence, it is verified.