Exercise 2

#Systematic Sampling

Example: A company's desired to estimate employee savings. The savings ($10,000) of 40 employees are arranged in the following Table :

i	savings	i	savings	i	savings	i	savings
1	**23**	11	**14**	21	**14**	31	**15**
2	**30**	12	**17**	22	**22**	32	**14**
3	**19**	13	**40**	23	**24**	33	**15**
4	**38**	14	**17**	24	**14**	34	**29**
5	**23**	15	**40**	25	**17**	35	**10**
6	**38**	16	**23**	26	**12**	36	**10**
7	**19**	17	**20**	27	**30**	37	**16**
8	**18**	18	**20**	28	**14**	38	**22**
9	**43**	19	**24**	29	**20**	39	**16**
10	**23**	20	**14**	30	**14**	40	**12**

A- List all possible systematic samples of size 10, that can be drawn from this set of employees using systematic sampling. Also, obtain corresponding sample means and variance.

B- Compute variance of the systematic sample mean, and variance of the SRS mean. Also, compare the precision systematic sample and SRS.

Solution

N=40 , n=10, k=40/10=4 , $1 \leq r \leq 4$

so we can draw 4 groups of size 10. In actual sampling, one of the four samples will be selected randomly.

	Samples			
	1	2	3	4
	23	30	19	38
	23	38	19	18
	43	23	14	17
	40	17	40	23
	20	20	24	14
	14	22	24	14
	17	12	30	14
	20	14	15	14
	15	29	10	10
	16	22	16	12
Means	23.1	22.7	21.1	17.4
Variances	104.1	62.01	77.66	65.16

$sample\ Mean\ and\ Variance\ \ of\ h^{th}\ sample\ are$

$$\bar{y}_h = \frac{\sum_{i=1}^{n} y_i}{n}, \qquad s_h^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y}_h)^2}{n-1} \qquad , h = 1, .., k$$

variance of the systematic sample mean is

$$V(\bar{y}_{sys}) = \frac{1}{k}\sum_{h=1}^{k}(\bar{y}_h - \bar{Y})^2$$

$$\mu = \bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{N} = 21.075$$

$$V(\bar{y}_{sys}) = \frac{1}{4}[(23.1 - 21.075)^2 + (22.7 - 21.075)^2 + (21.1 - 21.075)^2 + (17.4 - 21.075)^2]$$

$$= 5.062$$

variance of the SRS is

$$V(\bar{y}_{srs}) = \frac{N-n}{N}\frac{S^2}{n}$$

$$S^2 = \frac{\sum_{i=1}^{n}(Y_i - \mu)^2}{N-1} = 76.48$$

$$V(\bar{y}_{srs}) = \frac{40-10}{40}\frac{76.48}{10} = 5.736$$

The difference in the variances of the mean $\bar{y}$ of a simple random sample and $\bar{y}_{sys}$ is

$$V(\bar{y}) - V(\bar{y}_{sys}) = 5.736 - 5.062 = 0.674$$

Therefore, the precision of the systematic mean is higher. The precision of systematic sampling is larger with a gain of $\frac{V(\bar{y}) - V(\bar{y}_{sys})}{V(\bar{y}_{sys})} 100\% = 13\%$

NOTE:
- Systematic sampling is precise when units within the same sample are heterogeneous and his imprecise when they are homogeneous.
- If the units in the population are arranged at random then systematic sampling is equivalent to SRS without replacement.

By use R program

```
#population(employee savings)
Y=c(23,30,19,38,23,38,19,18,43,23,14,17,40,17,40,23,20,20,24,14,14,22,24,14,1
7,12,30,14,20,14,15,14,15,29,10,10,16,22,16,12)
# compute population mean:
mean(Y)
```

```
[1] 21.075

# perform all 1-in-4 systematic samples
k=4 ;n=10 ; N=40
sys_samples=matrix(0,10,4)
sys_samples

      [,1] [,2] [,3] [,4]
 [1,]    0    0    0    0
 [2,]    0    0    0    0
 [3,]    0    0    0    0
 [4,]    0    0    0    0
 [5,]    0    0    0    0
 [6,]    0    0    0    0
 [7,]    0    0    0    0
 [8,]    0    0    0    0
 [9,]    0    0    0    0
[10,]    0    0    0    0

for (i in 1:4) sys_samples[,i]=Y[seq(i,40,k)]
sys_samples

      [,1] [,2] [,3] [,4]
 [1,]   23   30   19   38
 [2,]   23   38   19   18
 [3,]   43   23   14   17
 [4,]   40   17   40   23
 [5,]   20   20   24   14
 [6,]   14   22   24   14
 [7,]   17   12   30   14
 [8,]   20   14   15   14
 [9,]   15   29   10   10
[10,]   16   22   16   12
```

To compute all systematic sample mean y_{sys}

```
sys_mean=apply(sys_samples,2,mean)
#compare their mean with the population mean
mean(sys_mean)

[1] 21.075

mean(Y)

[1] 21.075
```

conclude: the systematic mean is an unbiased estimator of the population mean.

```
# compute the variance of the systematic sample mean
 var_sys_mean=1/k*sum((sys_mean-mean(Y))**2)
var_sys_mean
```

```
[1] 5.061875

# compute the variances of the systematic samples s_{i}²
sys_var=apply(sys_samples,2,var)
sys_var

[1] 104.10000  62.01111  77.65556  65.15556

# compute the variance of the SRS mean V(ybar)
S=sd(Y)
varSRS=((40-10)/(40*10))*(S^2)
varSRS

[1] 5.736106
```

compute the difference between the variance of the SRS mean and the variances of the systematic mean

```
varSRS-var_sys_mean

[1] 0.6742308
```

Conclude: the precision of the systematic mean is higher. The precision of systematic sampling is larger with a gain of 13%

Estimation of Proportion (SRS)

Example 1:

Punjab Agricultural University, Ludhiana, is interested in estimating the proportion P of teachers who consider semester system to be more suitable as compared to the trimester system of education. A **with replacement simple random sample** of n= 120 teachers is taken from a total of N=1200 teachers. The response is denoted by **0** if the teacher does not think the semester system suitable, and **1** if he/she does.

From the sample observations given below, Estimate the proportion and total number of teachers who consider semester system is suitable with 95% confidence interval for each.

Teacher	**1**	**2**	**3**	**4**	**5**	**6**	**….**	**119**	**120**	**total**
Response	**1**	**0**	**1**	**1**	**0**	**1**		**0**	**1**	**72**

Solution:

Total number with response 1 is $a = \sum_{i=1}^{N} Y_i = 72$, sample **WR**

The Estimate of proportion and total number of teachers who consider semester system to be more suitable are

$$p = \frac{a}{n} = \frac{72}{120} = 0.6 \ ; \ \ q = 1 - p = 1 - 0.6 = 0.4$$

$$A = Np = 1200(0.6) = 720$$

Estimate of the standard error of $\boldsymbol{p}$ will then be obtained as

$$v(p) = \widehat{V(p)} = \frac{pq}{n-1} = \frac{(0.6)(0.4)}{119} = \frac{0.24}{119} = 0.0022$$

$$se(p) = \sqrt{\frac{pq}{n-1}} = \sqrt{\frac{0.24}{119}} = 0.0449$$

The 95% confidence interval for **P** would be obtained following

$$p \mp Z_{1-\frac{\propto}{2}} se(p)$$

$$Z_{1-\frac{\alpha}{2}} = Z_{1-\frac{0.05}{2}} = Z_{0.975} = 1.96$$

$$0.6 \mp 1.96\,(0.0449)$$

$$[0.5119\,,0.688]$$

The proportion of teachers in the university favoring semester system is, therefore, likely to be in the closed interval [0.5119, 0.688].

or We are 95% confidence that the proportion of teachers in the university favoring semester system belong in the closed interval [0.5119, 0.688]

The 95% confidence interval for total number of teachers who consider semester system to be more suitable is

$$1200\,(0.5119) < A < 1200(\,0.688)$$
$$614 < A < 826$$

Example 2

From the data given in Example 1, suppose that the SRS **without replacement** samples.

- Estimate proportion P along with the standard error of your estimate.
- Estimate the proportion and total number of teachers who consider semester system is suitable.
- Calculate confidence interval for P and total number of teacher.

Solution:

$$p = \frac{a}{n} = \frac{72}{120} = 0.6\ ;\ \ q = 1 - p = 1 - 0.6 = 0.4$$
$$A = Np = 1200(0.6) = 720$$

Estimate of the standard error of $\boldsymbol{p}$ will then be obtained as

$$v(p) = \widehat{V(p)} = \frac{pq}{n-1}\frac{N-n}{N} = \frac{(0.6)(0.4)}{119}\frac{1200-120}{1200} = \frac{27}{14875} = 0.001815$$

$$se(p) = \sqrt{\frac{pq}{n-1}\frac{N-n}{N}} = \sqrt{\frac{27}{14875}} = 0.0426$$

The 95% confidence interval for **P** would be obtained following

$$p \mp Z_{1-\frac{\propto}{2}}\, se(p)$$

$$0.6 \mp 1.96\,(0.0426)$$

$$[0.5165\,,0.6835\,]$$

The 95% confidence interval for total number of teachers who consider semester system to be more suitable is

$$620 < A < 821$$

H.W

Example: A geneticist is interested in the proportion of African males who have a certain minor blood disorder. In **replacement simple random sample** of 100 African males, 24 are found to be afflicted.

(a) Compute a 99% confidence interval for the proportion and total number of African males who have this blood disorder.
(b) What can we assert with 99% confidence about the possible size of our error if we estimate the proportion of African males with this blood disorder to be 0.24?