

Tutorial 3 (chapter 3 part1)

Simple random sampling

Example 3.1

Appendix C gives data related to the number of tractors in 30 serially numbered villages of Doraha development block in Punjab (India). Select (1) WR and (2) WOR simple random sample of 10 villages using direct approach method.

Appendix C Number of tractors, tube wells, and net irrigated area (in hectares) for 69 villages of Doraha development block of Punjab, India

Village No.	Village name	Tractors	Tube wells	Irrigated area
1	Ajnaud	20	102	281
2	Aracha	19	126	337
3	Afjullapur	5	31	77
4	Alampur	10	39	108
5	Bishanpur	12	85	191
6	Shahpur	21	130	208
7	Sirthala	44	222	698
8	Sultanpur	12	70	180
9	Sihora	20	219	458
10	Hole	10	115	288
11	Kotli Afgana	11	41	161
12	Katari	30	133	512
13	Kartarpur	15	55	123
14	Katana Sahib	3	50	100
15	Kaddon	36	249	675
16	Gobindpura	24	119	258
17	Gurditpura	17	95	143
18	Gidrho	15	100	267
19	Ghadani Kalan	76	551	1178
20	Ghadani Khurd	35	209	523
21	Ghalouti	46	380	583
22	Ghanrash	16	160	310
23	Chankoiyan Kalan	7	33	129
24	Chankoiyan Khurd	21	142	184
25	Chapran	7	64	91
26	Jaipura	22	130	240
27	Jahangir	5	27	80
28	Jandali	21	118	429
29	Jargarhi	38	193	583
30	Jarag	59	360	888

$$N = 30$$

$$\bar{Y} = \frac{\sum_{i=1}^N Y_i}{N} = 22.5667$$

$$\sigma^2 = \frac{\sum_{i=1}^N (Y_i - \bar{Y})^2}{N} = 273.3789$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{273.3789} = 16.5342$$

$$S^2 = \frac{\sum_{i=1}^N (Y_i - \bar{Y})^2}{N - 1} = 282.8057$$

Here village is the sampling unit. The villages in the population are already serially numbered which, otherwise, is the first step involved in the sample selection. Refer to Appendix B, and use first column by dropping the last two digits of each four-digit number. Then we see that the first random number thus formed is 11. Similarly, the subsequent random numbers are seen to be 5, 26, 11 ... , 3

(1) By selecting the first 10 random numbers from 1 to 30, without discarding Repetitions (**WR**), we obtain the serial numbers of villages in the sample. These are given below along with their variable values (number of tractors).

Village	11	5	26	11	11	24	12	22	9	3
Tractors y_i	11	12	22	11	11	21	30	16	20	5

One can see that 11th village has been selected twice in the with replacement simple random sample where repeated selection of units is permitted.

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n} = \frac{159}{10} = 15.9 \approx 16$$

$$s^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n - 1} = 53.8778$$

$$\text{Sampling variance: } V(\bar{y}) = \frac{\sigma^2}{n} = \frac{273.3789}{10} = 27.3379$$

$$\text{Standard error of the sample mean: } se(\bar{y}) = \sqrt{V(\bar{y})} = \sqrt{27.3379} = 5.2286$$

The confidence interval for the population mean \bar{Y} are given by

$$\bar{y} \pm 2 se(\bar{y})$$

$$15.9 \pm 2 * 5.2286$$

$$\bar{Y} \in [5.4428, 26.3572]$$

$$\text{Unbiased estimator of } V(\bar{y}) = v(\bar{y}) = \frac{s^2}{n} = \frac{53.8778}{10} = 5.38778$$

(2) In without replacement sample, any repetition (11th village in the present case) is omitted, and another random number is selected as its replacement. Next random numbers from 1 to 30 are 7 and 25. Thus the WOR simple random sample of 10 villages from the population under study is with the following serial numbers:

Village	11	5	26	24	12	22	9	3	7	25
Tractors y_i	11	12	22	21	30	16	20	5	44	7

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n} = \frac{188}{10} = 18.8$$

$$s^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n - 1} = 135.7333$$

$$V(\bar{y}) = \frac{N-n}{Nn} S^2 = \frac{30-10}{30*10} * 282.8057 = 18.8537$$

Standard error of the sample mean: $se(\bar{y}) = \sqrt{V(\bar{y})} = \sqrt{18.8537} = 4.3421$

The confidence interval for the population mean \bar{Y} are given by

$$\begin{aligned} & \bar{y} \pm 2se(\bar{y}) \\ & 18.8 \pm 2*4.3421 \\ & \bar{Y} \in [10.1158, 27.4842] \end{aligned}$$

$$\text{Unbiased estimator of } V(\bar{y}) = v(\bar{y}) = \frac{N-n}{Nn} s^2 = \frac{30-9}{30*9} * 135.7333 = 10.557$$

Example 3.2 (by.R):

The height (in cm) of 6 students of M.Sc., majoring in statistics, from Punjab Agricultural University, Ludhiana was recorded during 1985. The data, so obtained, are given below:

Table 3.1 Heights of M.Sc. students

Student	Name	Height
1	Sarjinder Singh	168
2	Gurmeet Singh	175
3	Varinder Kumar	185
4	Sukhjinder Singh	173
5	Devinder Kumar	171
6	Gulshan Kumar	172

1. Calculate (a) population mean \bar{Y} , and (b) population variance σ^2 .
2. Enumerate all possible SRS with replacement samples of size $n=2$. Obtain sampling distribution of mean, and hence show that
 - a. $E(\bar{y}) = \bar{Y}$
 - b. $V(\bar{y}) = \frac{\sigma^2}{n}$
 - c. $E(s^2) = \sigma^2$
 - d. $E[v(\bar{y})] = V(\bar{y})$

By R:

(1)

```

> Y <- c(168,175,185,173,171,172)
> (Ybar <- mean(Y))
[1] 174
> N <- 6
> (Sigmasquare <- var(Y)*((N-1)/N))
[1] 28.66667

```

(2)

```

> n <- 2
> (No.s<- N^n)      #Number_of_all_possible_samples
[1] 36
> install.packages("tidyverse")
> library(tidyverse)
> sample <- crossing (Var1=Y, Var2=Y)
> samples<- sample (Y, 2, replace= TRUE)
> view(sample)
> (ybars= apply(sample,1 , mean))
#for a matrix 1 indicates rows,2 indicates columns or rowMeans(sample)
[1] 168.0 169.5 170.0 170.5 171.5 176.5 169.5 171.0
[9] 171.5 172.0 173.0 178.0 170.0 171.5 172.0 172.5
[17] 173.5 178.5 170.5 172.0 172.5 173.0 174.0 179.0
[25] 171.5 173.0 173.5 174.0 175.0 180.0 176.5 178.0
[33] 178.5 179.0 180.0 185.0
> (unbiased_mean <- mean(ybars))
[1] 174
> Ybarsquare <- (ybars)^2
> (Variance_Of_ybar <- mean(Ybarsquare) -(unbiased_mean)^2)
[1] 14.33333
> (Sigmasquare/n)
[1] 14.33333
> install.packages("matrixStats")
> library(matrixStats)
> ssquare <- rowVars(as.matrix(sample))
> mean(ssquare)
[1] 28.66667

```

$$E(\bar{y}) = \bar{Y}$$

$$V(\bar{y}) = \frac{\sigma^2}{n}$$

$$E(s^2) = \sigma^2$$

Example 3.3:

From the data given in example 3.2, enumerate all the SRS without replacement samples of size $n=2$, and write down sampling distribution of mean. Using this distribution, show that:

- a. $E(\bar{y}) = \bar{Y}$
- b. $V(\bar{y}) = \frac{N-n}{Nn} S^2$
- c. $E(s^2) = S^2$
- d. $E[v(\bar{y})] = V(\bar{y})$

By R:

```
> Y <- c(168,175,185,173,171,172)
> n <- 2
> (No.s<- choose(N,n))
[1] 15
> samples <- combn(Y,n)
> samples
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9]
[1,]  168  168  168  168  168  175  175  175  175
[2,]  175  185  173  171  172  185  173  171  172
      [,10] [,11] [,12] [,13] [,14] [,15]
[1,]   185   185   185   173   173   171
[2,]   173   171   172   171   172   172
> (ybars= apply(samples,2, mean))
#for a matrix 1 indicates rows,2 indicates columns or colMeans(samples)
[1] 171.5 176.5 170.5 169.5 170.0 180.0 174.0 173.0
[9] 173.5 179.0 178.0 178.5 172.0 172.5 171.5
> (unbiased_mean <- mean(ybars))
[1] 174
```

$$E(\bar{y}) = \bar{Y}$$

```
> Ybarsquare <- (ybars)^2
> (Variance_Of_ybar <- mean(Ybarsquare) -(unbiased_mean)^2)
[1] 11.46667
> (Ssquare <- var(Y))
[1] 34.4
> (N-n)/(N*n) * Ssquare
[1] 11.46667
```

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

```
> install.packages("matrixStats")
> library(matrixStats)
> ssquare <- colVars(samples)
> mean(ssquare)
[1] 34.4
```

$$E(s^2) = S^2$$

Example 3.4:(using WR)

From the WR sample of 10 villages selected in example 3.1, estimate the average number of tractors per village in the block along with its standard error. Also, set up confidence interval for the population mean.

Village :	34	61	58	62	47	34	11	43	5	35
Tractors :	14	8	15	39	9	14	11	19	12	18

Average number of tractors per village in the block is:

$$\bar{y} = \frac{1}{10} \sum_{i=1}^{10} y_i = \frac{159}{10} = 15.9$$

To find the standard error, we first calculate sample mean square:

$$s^2 = \frac{1}{n-1} \left(\sum_{i=1}^n y_i^2 - n\bar{y}^2 \right)$$
$$= \frac{1}{9} (3233 - 10 * 15.9^2) = 78.32222$$

The estimate of the variance is given by:

$$v(\bar{y}) = \frac{s^2}{n} = \frac{78.3222}{10} = 7.83222$$

Estimate the standard error of the sample mean is:

$$se(\bar{y}) = \sqrt{v(\bar{y})} = \sqrt{7.83222} = 2.7986$$

The lower and upper limits of the confidence interval for \bar{Y} are given by:

$$\bar{y} \pm 2se(\bar{y})$$
$$15.9 \pm 2 * 2.7986$$
$$(10.3028, 21.497)$$