

## Questions on Sampling Technique-II Lab Report

1. A village has 6 orchards containing 150, 50, 100, 200, 160, and 40 trees, respectively. Select a sample of size 3 orchards with replacement and with probability proportion to the number of trees.
2. A village has 10 holdings consisting of 50, 30, 45, 25, 40, 26, 24, 35, 28, and 27 fields. Select a sample of four holdings using Cumulative totals of PPSWR.
3. A village has 6 orchards containing 150, 50, 100, 200, 160, and 40 trees, respectively. Select a sample of size 3 orchards Using Lahiri's Method of PPSWR
4. A village has 10 holdings consisting of 50, 30, 45, 25, 40, 26, 24, 35, 28, and 27 fields. Select a sample of four holdings using Lahiri's method of PPSWR.
5. A sample survey was conducted to study the yield of wheat in Haryana. A sample of 20 farms from a total of 100 was taken, with probability proportional to the area under wheat crop, with replacement method. The total area under wheat crop (X) was 484.5 hectares. The area under crop (X) and yield (Y) were noted in hectares and quintals per hectare, respectively. The sample selected by the cumulative method was

X	5.2	5.9	3.9	4.2	4.7	4.8	4.9	6.8	4.7	5.7	5.2	5.2	4.9	4.0	1.3	7.4	7.4	4.8	6.2	6.2
Y	28	29	30	22	24	25	28	37	26	32	25	38	31	16	6	61	61	29	47	47

The sample selected by Lahiri's method was

X	4.8	4.1	1.3	5.2	6.6	6.0	2.0	6.3	5.2	4.2	4.8	5.9	5.8	5.8	5.1	4.7	5.6	5.2	4.0	4.6
Y	22	19	6	25	54	43	4	40	28	29	22	39	39	44	30	27	34	31	18	31

- (i) Estimate the average yield per farm along with its standard errors for these two samples. Also, estimate the population total with its standard errors. Also construct the 95% CI.
  - (ii) Estimate the gain in efficiency due to PPS sampling compared to SRSWR.
  - (iii) Make a comment on your findings.
6. At an experimental station, there were 100 fields sown with wheat. Each field was divided into 16 plots of equal size (1/16th hectare). Out of 100 fields, 10 were selected by SRSWOR. From each selected field, 4 plots were chosen by SRSWOR. The yields in kg/plot are given below:

Selected field	Plots			
	1	2	3	4
1	4.32	4.84	3.96	4.04
2	4.16	4.36	3.50	5.00
3	3.06	4.24	4.76	3.12
4	4.00	4.84	4.32	3.72
5	4.12	4.68	3.46	4.02
6	4.08	3.96	3.42	3.08
7	5.16	4.24	4.96	3.84
8	4.40	4.72	4.04	3.98
9	4.20	4.66	3.64	5.00
10	4.28	4.36	3.00	3.52

- (i) Estimate the wheat yield per hectare for the experimental station along with its standard error.
- (ii) How can an estimate obtain from an SRS of 40 plots be compared with the estimate obtained above, in (i)?
- (iii) Obtain optimum n and m under cost function  $100 = 4n + nm$ .
- (iv) Make a comment on your findings.
7. For study of feeding and rearing practices of sheep and yield of wool in the Rajasthan State, during the year 1980-81, two stage sampling design with tehsils as first stage units and villages in the tehsil as second stage units was adopted. The data given below are the stationary sheep population in the selected villages in each of 4 tehsils selected from 12 tehsils of the Ajmer Division, as counted in the survey along with the number of villages in the tehsil.

Selected Tegisil	No. of villages in the tehsil ( $M_i$ )	Stationary sheep population in the selected villages
Behrar	102	266, 890, 311, 46, 174, 31, 17, 186, 224, 31, 102, 46, 31, 109, 275, 128, 125, 267, 153, 152, 84, 21, 52, 10, 0, 48, 94, 123, 87, 89, 109, 0, 310, 3
Bairath	105	129, 57, 64, 11, 163, 77, 278, 50, 26, 127, 252, 194, 350, 0, 572, 149, 275, 114, 387, 53, 34, 150, 224, 185, 157, 244, 466, 203, 354, 816, 242, 140, 66, 590, 747, 147
Ajmer	200	247, 622, 225, 278, 181, 132, 659, 403, 281, 236, 595, 265, 431, 190, 348, 232, 88, 1165, 831, 120, 987, 938, 197, 614, 187, 896, 330, 485, 60, 60, 1051, 651, 552, 968, 987
Bansur	88	347, 362, 34, 11, 133, 36, 34, 61, 249, 170, 112, 42, 161, 75, 68, 0, 247, 186, 473, 0, 143, 198, 65, 0, 308, 122, 345, 0, 223, 302, 219, 120, 199, 35, 0, 0

Estimate the mean stationary sheep population in the Ajmer division during the year 1980-81, together with its standard error when  $\bar{M} = 124$ . [Hints;  $\bar{y} = \frac{1}{n\bar{M}} \sum_{i=1}^n M_i \bar{y}_i$ ;  $v(\bar{y}) = \left(\frac{1}{n} - \frac{1}{N}\right) S_b^2 + \frac{1}{Nn} \sum_{i=1}^n W_i^2 \left(\frac{1}{m_i} - \frac{1}{M_i}\right) s_{w_i}^2$ ;  $s_b^2 = \frac{1}{n-1} \sum_{i=1}^n (W_i \bar{y}_i - \bar{y})^2$ ;  $s_{w_i}^2 = \frac{1}{m_i-1} \sum_{j=1}^{m_i} (y_{ij} - \bar{y}_i)^2$ ;  $W_i = \frac{M_i}{\bar{M}}$  ]

8. A list of 23 former agricultural districts of Bangladesh with areas in thousand acres of land (x) is given in the accompanying table together with district-wise production of rice (in '000 metric tons) in 1998-99 (y).

District	X	Y	District	X	Y
Banderban	61	48	Tangail	561	483
Chittagong	1079	994	Barisal	133	662
Khagrachari	30	26	Jessore	1452	1352
Comilla	1519	1313	Khulna	1134	853
Noakhali	1036	779	Kushtia	567	479
Rangamati	48	40	Patuakahli	1027	543
Sylhet	2309	1512	Bogra	1169	1093
Dhaka	936	859	Dinajpur	1573	1069
Faridpur	1018	577	Pabna	738	660
Jamalpur	811	723	Rajshahi	1799	1753
Kidhorgonj	1341	1121	Rangpur	2243	1873
Mymensingh	1715	928	Total	24299	19740

- Draw a sample 5 districts without/with replacement using (i) simple random sampling method and (ii) PPS method;
  - Estimate the average and the total production of rice per district for both the samples;
  - Compute the 95% confidence interval in each case.
9. A nurseryman wants to estimate the average height of seedlings in a large field that is divided into 50 plots that vary slightly in size. He believes that the heights are fairly constant throughout each plot but may vary considerably from plot to plot. Therefore, he decides to sample 10% of trees within each of 10 randomly selected plots using a two-stage cluster sample. The accompanying table displays the data.

Plot	$M_i$	$m_i$	$y_{ij}$
1	52	5	12, 11, 12, 10, 13
2	56	6	10, 9, 7, 9, 8, 10
3	60	6	6, 5, 7, 5, 6, 4
4	46	5	7, 8, 7, 7, 6
5	49	5	10, 11, 13, 12, 12
6	51	5	14, 15, 13, 12, 13
7	50	5	6, 7, 6, 8, 7
8	61	6	9, 10, 8, 9, 9, 10
9	60	6	7, 10, 8, 9, 9, 10
10	45	6	12, 11, 12, 13, 12, 12

- Estimate the average height of seedlings in the field;
- compute a 95% confidence interval for the average;
- Make a comment on your findings.

**Instructor:** Dr. Md. Maniruzzaman, Associate Professor, Statistics Discipline, Khulna University

10. If the farm size is taken as  $X_i$ , and area under wheat as  $Y_i$ , the data from a census of all farms in a district divided into 2 strata are given below:

Stratum	$N_i$	$\bar{X}_i$	$\bar{Y}_i$	$S_{y_i}^2$
1	1580	82.56	19.40	312
2	430	244.85	51.63	922
Total	2010	117.28	26.30	620

Assuming that the cost of measuring the wheat, area is 10 times that for farm size and the total amount available for the survey is Rs 100, draw a random sample of 100 farms by double sampling with optimum allocations. Make a comment on your findings.

11. In a survey conducted by an agency in Saraibal block of Santhal Parganas in Bihar (India) during the Rabi season of 1967-68, 499 fields growing wheat were selected at random and the yield of wheat (kg/ha) was recorded by eye estimation. The data showing the distribution of the number of fields falling in each of the 6 strata are given below. A sub-sample of the fields was selected from the fields falling in each of 6 strata and crop-cutting experiments were conducted. The yield rates observed are given in column (4) for each of the selected fields.

Strata	No. of Eye Estimate	No. of Crop-cutting experiments	Yield (kg/ha)
(1)	(2)	(3)	(4)
101-200	154	40	200,208,152,224,104,168,160,152,247,178, 84,360,340,380,340,340,136,180,420,172, 216,128,136,124,114,139,160,112,104,140, 192,184,124,81,82,70,380,440,400,400
201-300	189	46	104, 152, 148, 256, 280, 260, 320, 288, 288, 140, 144, 124, 496, 450, 492, 256, 256, 252, 332, 384, 276, 330, 344, 243, 296, 292, 314, 424, 360, 416, 200, 192, 195, 326, 248, 322, 104, 144, 61, 38, 112, 137, 144, 206, 496, 535
301-400	91	22	280, 192, 192, 280, 200, 304, 440, 320, 192, 448, 448, 345, 326, 420, 212, 326, 163, 326, 325, 496, 304, 243
401-500	40	15	288,156,280,136,384, 472, 345, 333, 300, 326, 324, 720, 672, 568, 520
501-600	13	6	428, 368, 506, 824, 624, 768
Above	12	2	344, 712

- Obtain an estimate of the average yield of wheat in kg/ha for the block and also the standard error of the estimate.
- Assuming the cost function  $150n + 0.05n'$  and the optimum allocation to be used for allocation of  $n$  to various strata, estimate the values of  $n$  and  $n'$  so that the variance of the double sampling estimate is minimum. Also estimate the variance of the estimate with these values of  $n$  and  $n'$ .

**Instructor:** Dr. Md. Maniruzzaman, Associate Professor, Statistics Discipline, Khulna University

(iii) Make a comment on your findings.

12. The following data, relating to the yield of paddy, have been taken from crop cutting surveys conducted in certain parts of Uttar Pradesh (India). In all, 40 random cuts of size 1/100th of an acre were taken and the yield of paddy observed on the day of harvest was noted. On a sub-sample of 20 cuts out of 40, the dry yield was also noted. The results are given below:

SN of Cut	Harvest yield (in kg)	Dry yield (in kg)	SN of cut	Harvest yield (in kg)
1	16.8	15.2	21	8.7
2	12.7	11.8	22	11.6
3	18.8	17.5	23	11.5
4	13.9	12.5	24	14.4
5	11.3	10.4	25	17.8
6	10.9	10.1	26	8.4
7	12.5	11.2	27	8.7
8	17.4	15.8	28	14.6
9	14.1	13	29	12.1
10	11.9	10.8	30	7.9
11	13.4	12.3	31	8.9
12	13.5	12.4	32	11.1
13	8.3	7.6	33	13
14	13.7	12.5	34	10.5
15	14.6	13.5	35	14.2
16	14.5	13.5	36	12.7
17	17.1	16.2	37	11.9
18	14.5	13.5	38	15.5
19	11.4	10.3	39	17.1
20	14	12.8	40	10.9

- (i) Estimate the average yield per acre of dry paddy, along with its sampling error, by utilizing whatever information is available both on harvest as well as on dry yield.
- (ii) What would have been loss in precision, had we neglected the additional information on harvest yield from the sub-sample of only 20 cuts for which dry yield is available.
- (iii) If the cost of obtaining dry weight from the cut is  $1\frac{1}{2}$  times that of obtaining the harvest weight of the cut, determine the optimum proportion of cuts for which dry weight is to be observed.
- (iv) Make a comment on your findings.

### **Instructions for Students to Submit the Assignment**

#### **1. File Preparation:**

- Complete your analysis using **R Markdown (.Rmd)**.
- Ensure that all code, outputs, tables, and graphs are clearly visible in the final report.
- Save the output file in **PDF, HTML, or Word (.docx)** format.

## 2. File Naming Convention:

- Use the following format for your file name:
  - **YourName\_ID\_STAT\_3105\_PUST.Rmd** and
  - **YourName\_ID\_STAT\_3105\_PUST.pdf/html/docx**(e.g., *Maniruzzaman\_10024730\_STAT\_3105\_PUST.pdf*)

## 3. Submission Requirements:

- Submit both the **R Markdown file** (.Rmd) and the **output file** (PDF/HTML/Word) within the deadline (**November 16, 2025**).
- Late submissions may not be accepted unless prior approval is granted.

## 4. Content Guidelines:

- Include appropriate **titles, section/subsection headings, and make appropriate comments** in your *R Markdown* file.
- Ensure that all analyses and interpretations are presented neatly and clearly.
- Provide brief explanations for each result and interpretation of model performance.

## 5. Academic Integrity:

- **Each student must complete the assignment independently.**
- Plagiarism or code copying will result in penalties.