Assignment 2

Reading:

By Thursday, September 13, read Chapter 1 (pages 1–18), Appendix A (pages 549–562), and Chapter 2 (pages 25–60) of *Sampling: Design and Analysis, second edition*; by Sharon L. Lohr.

For Thursday, September 20, read Sections 3.1–3.5 (pages 73–95) of Lohr.

Homework 2:

The following problems are due in class on Thursday, September 27. Homework can also be submitted to the course mailbox in Room 904 SSW by 5:00pm on Wednesday, October 3.

1. Let $N = 6$ and $n = 3$. For purposes of studying sampling distributions, assume that all population values are known.

$$y_1 = 98 \quad y_2 = 102 \quad y_3 = 133 \quad y_4 = 154 \quad y_5 = 175 \quad y_6 = 190$$

We are interested in $\bar{y}_U$, the population mean. Two sampling plans are proposed.

- Plan 1. Eight possible samples may be chosen.

Sample number	Sample, $\mathcal{S}$	$\Pr(\mathcal{S})$
1	$\{1, 3, 5\}$	1/8
2	$\{1, 3, 6\}$	1/8
3	$\{1, 4, 5\}$	1/8
4	$\{1, 4, 6\}$	1/8
5	$\{2, 3, 5\}$	1/8
6	$\{2, 3, 6\}$	1/8
7	$\{2, 4, 5\}$	1/8
8	$\{2, 4, 6\}$	1/8

- Plan 2. Three possible samples may be chosen.

Sample number	Sample, $\mathcal{S}$	$\Pr(\mathcal{S})$
1	$\{1, 3, 5\}$	1/4
2	$\{2, 3, 6\}$	1/2
3	$\{1, 4, 6\}$	1/4

(a) Let $\bar{y}$ be the mean of the sample values. For each sampling plan, find:

$$\text{(i) } E[\bar{y}]; \qquad \text{(ii) } V[\bar{y}]; \qquad \text{(iii) Bias}(\bar{y}); \qquad \text{(iv) MSE}(\bar{y}).$$

(b) Which sampling plan do you think is better? Why?

2. Consider an artificial situation in which we know the value of y_i for each of the $N = 8$ units in the whole population. The index set for the population is

$$\mathcal{U} = \{1, 2, 3, 4, 5, 6, 7, 8\}$$

and the values of y_i are

i	1	2	3	4	5	6	7	8
y_i	1	2	4	4	7	7	7	8

Consider the following sampling plan.

Sample number	Sample, $\mathcal{S}$	$\Pr(\mathcal{S})$
1	$\{1, 3, 5, 6\}$	$1/8$
2	$\{2, 3, 7, 8\}$	$1/4$
3	$\{1, 4, 6, 7\}$	$1/8$
4	$\{2, 4, 7, 8\}$	$3/8$
5	$\{4, 5, 6, 8\}$	$1/8$

(a) Find the probability of selection π_i for each unit i.

(b) What is the sampling distribution of $\hat{t} = 8\bar{y}$?

3. A university has 807 faculty members. For each faculty member, the number of refereed publications was recorded. A frequency table for number of refereed publications is given below for a simple random sample of 50 faculty members.

Refereed publications	0	1	2	3	4	5	6	7	8	9	10
Faculty members	28	4	3	4	4	2	1	0	2	1	1

(a) Plot the data using a histogram. Describe the shape of the data.

(b) Estimate the mean number of publications per faculty member, and give the standard error for your estimate.

(c) Do you think that $\bar{y}$ from part (b) will be approximately normally distributed? Why or why not?

(d) Estimate the proportion of faculty members with no publications, and give a 95% confidence interval.

4. The manager of a mail-in lottery game took a random sample of 1000 entries from the last few contests, and found that 175 of those came from the South.

(a) Find a 95% confidence interval for the percentage of *all* entries that come from the South.

(b) According to the *Statistical Abstract of the United States*, 30.9% of the U.S. population lives in states that the lottery manager considered to be in the South. Is there evidence from your interval in part (a) that the percentage of entries from the South differs from the percentage of persons living in the South?

5. Suppose we wish to take a simple random sample of the 580 children served by a family medical practice, to estimate the proportion who are overdue for a vaccination. What sample size would be necessary to estimate the proportion with 95% confidence and margin of error 0.08?