SIMPLE RANDOM SAMPLING

1. (a) To estimate the average income in a certain area, a simple random sample of size n is chosen without replacement from a the population of size N.

    i. Show that $\overline{y}$ is an unbiased estimate of the true population mean $\overline{Y}$.

    ii. Derive the variance of the estimator.

    iii. Obtain a 95% confidence interval for the $\overline{Y}$.

   (a) To estimate the total disposable income in an area, a simple random sample of size n is chosen without replacement from a the population of size N.

    i. Write down an appropriate unbiased estimator for the total income Y.

    ii. Show that it is unbiased.

    iii. Derive the variance of the estimator.

    iv. Obtain a 95% confidence interval for the Y.

2. (a) To estimate the proportion unemployed in a certain area, a simple random sample of size n is chosen without replacement from a the population of size N. The number of unemployed is found to be r.

    i. Show that $\frac{r}{n}$ is an unbiased estimate of the true proportion unemployed.

    ii. Derive the variance of the estimator.

    iii. Obtain a 95% confidence interval for the true proportion unemployed in the area.

SAMPLE SIZE

1. Verify that in simple random sampling from a finite population, the sample size necessary to estimate a population average is given by

$$n = \frac{n_0}{1 + n_0/N}$$

   where n_0 is the sample size indicated for given confidence and error levels without inclusion of the finite population correction factor.

2. A simple random sample is to be chosen from a population of size $N = 1,000$ to estimate the average income. Calculate the sample size necessary to ensure that the estimator is within 10 euro of the true value. From previous studies it is known that the population variance $S^2 = 81$.

3. Verify that in simple random sampling from a finite population, the sample size necessary to estimate a particular attribute is given by

$$n = \frac{n_0}{1 + n_0/N}$$

   where n_0 is the sample size indicated for given confidence and error levels without inclusion of the finite population correction factor.

4. A simple random sample is to be chosen from a population of size $N = 3,000$ to estimate the percentage unemployed. Calculate the sample size necessary to ensure that the estimator is within 2% of the true proportion unemployed with 95% confidence.

5. Sometimes a survey co-ordinator will decide to take a '5% sample'. Discuss why such a relative formulation of sample size is not adequate.

STRATIFICATION

1. Define *proportionate stratification* and *optimum allocation* as used when assigning the number of elements to be allocated to each stratum in a stratified random sample design. Discuss the advantages and disadvantages of each method. Explain when each method should be used.

2. Prove that
$$V(\overline{y}_{srs}) \geq V(\overline{y}_{prop})$$
where $\overline{y}_{srs}$ is the mean of a simple random sample and $\overline{y}_{prop}$ is the mean of a proportionate stratified random sample.

3. A stratified random sample of total size n is to be drawn from a finite population of size N consisting of k strata of sizes $N_1, N_2, \ldots, N_k$. The overall fixed cost of the survey is estimated at $\$C_0$, and the cost per unit in stratum is $\$c_h$. An independent random sample of size n_h $(1 \leq h \leq k)$ is to be selected from each stratum. Show that the variance of the sample mean $\sum_{h=1}^{k} W_h \sum_{i=1}^{n_h} y_{hi}/n_h$ is minimised provided $n_h \propto W_h S_h/\sqrt{c_h}$ where S_h is the standard deviation of the elements in the h^{th} stratum and $W_h = N_h/N$.

4. A study is being planned to estimate employee expenses in a particular firm. There are four departments in the firm and estimates of the standard deviations for each department available from previous studies are given as follows:

Stratum	Size	S
1	1500	3.7
2	1500	1.9
3	1500	2.8
4	1500	4.3

The budget for the study allows a sample of size 160.

(a) Allocate the sample to the four departments using proportionate allocation and using optimum allocation.

(b) Calculate the variance of the sample mean for the proportionate stratified design and for the optimum allocation.

5. It is proposed to take a stratified random sample from a finite population which has been divided into three strata. It is expected that the field cost is of the form $\$C_0 + \sum_{h=1}^{3} n_h c_h$. Estimates of the relevant quantities for the three strata are given below:

Stratum	W_h	S_h	c_h
1	0.4	10	4
2	0.3	15	6
3	0.3	20	2

(a) How should a sample of 200 be allocated among the three strata in order to minimise the variance of the mean?

(b) Calculate the variance with this allocation.

CLUSTER SAMPLING

1. Describe briefly the essential features of cluster sampling, indicating clearly the main advantages and disadvantages of the sampling method. Give examples of some cluster sampling designs used in practice.

2. For one-stage cluster sampling with m clusters from M equal-sizes clusters,

   (a) show that

   $$\overline{y}_{cl} = \frac{1}{m} \sum_{i=1}^{m} \overline{y}_i$$

   is an unbiased estimator of the population mean, where $\overline{y}_i$ is the mean of the units falling into the ith sampled cluster.

   (b) Write down the variance of $\overline{y}_i$. Prove that the variance of the mean with cluster sampling is less than that with simple random sampling .

3. A population consists of the following 12 units:

i	1	2	3	4	5	6	7	8	9	10	11	12
Y_i	1	1	2	1	4	3	3	5	3	6	9	10

   This population is divided into 6 clusters each containing 2 units by the following two schemes:

   Scheme 1 Units by cluster: (1,2), (3,4), (6, 7), (5, 9), (8,10), (11,12).

   Scheme 2 Units by cluster: (1,7), (2,8), (3, 9), (4, 10), (5,11), (6,12).

   Scheme 3 Units by cluster: (1,12), (2,11), (4, 10), (3, 8), (5,6), (7,9).

   A sample of size four is selected by choosing two clusters at random from the six for each of the three schemes. Calculate the variance of the estimator in each case, and compare it with the variance obtained with simple random sampling for a sample of size 4 four. What does this example tell you about the design of cluster sampling?

4. The following table gives the sizes of the six DEDs (District Electoral Divisions) in the Blackrock area of Dublin:

DED	1	2	3	4	5	6
Population	1,974	2,104	2,682	2,677	1,711	2,676

In a two-stage cluster sample design, it is desired to select, at the first stage, three DEDs with probability proportional to size.

(a) Show carefully how this is done, and select the units.

(b) At the second stage a sample of size 150 will be selected. How many elements should be selected from each selected DED to maintain epsem sampling?