# MODULE SECTION: ESTIMATION

## Discussion Questions (Set B)

**Responses to questions**

In groups discuss the following questions:

1. You have a population of farmers where 75% are smallholders, and 25% have large farms. You need to estimate the mean farm income from the population and decide to take a sample of 50 small and 50 large holdings. To do so, you calculate the mean for the income of the 100 farms by adding all the incomes and dividing by 100. Do you see any problem with this estimate?
   * Yes, there is a problem with the estimate, because the sample is not a good representative of the population. The estimate is likely to be overestimated than the true value of the population income due to the biased representation of large holdings reducing the accuracy of the estimate.
2. You decide to increase the sample size to 300 small and 300 large holdings. What is the consequence of this in your estimation of income? Does this solve your problem?
   * By increasing the sample size without adjusting the sampling frame to account to the proportion of small and large holdings in the population, the estimate is likely to be remain an overestimate of the true value of the population income. The precision of the estimate will increase due to the increase in sample size but the estimate will remain biased hence inaccurate.
3. How can you solve this problem before the data is collected?
   * Before data collection, the problem can be solved by: adjusting the sampling methodology to account for the proportions of the small and large holdings in the population when calculating the sample size from each sub-group of farmers. For example, use stratified sampling methods where the population is divided into two strata and you sample each strata depending on its size in the population.
4. How can you solve this problem after the data is collected?
   * After data collection; the selection bias can be solved by applying weights to the the incomes from the two sub-groups and calculate the weighted mean. (A weight can be calculated as population proportion of sub-group 1/sample proportion of sub-group 1.
5. What is a clustered sample?
   * Cluster sampling is a probability-based sampling method where the population is divided into smaller, naturally occurring groups, called clusters, and a random selection of clusters is made. Instead of sampling individuals directly from the entire population, all individuals within the selected clusters (or a random subset of them) are included in the sample.
6. What is the effect on the precision of an estimate that comes from a clustered sample?
   * The precision of an estimate from a clustered sample is generally lower compared to other sampling methods (like simple random sampling) with the same sample size. This is because individuals within a cluster tend to be more similar to each other (i.e., intra-cluster homogeneity) than to individuals in other clusters. As a result, the diversity of the sample decreases, leading to an increase in sampling error.