# MODULE SECTION: ESTIMATION

## Discussion Questions (Set A).

**Responses to questions**

1. Explain what "standard deviation" is and why it may be helpful to estimate it. Give an example of when it is of practical use in representing a data set.
   * Standard deviation is a measure of variability of a variable in our data.
   * It is useful to compare how much the data has been dispersed around the mean. For example, when you have two groups of students with the same mean score, the standard deviation will help you check the dispersion of the individual scores around the mean within the two groups.
2. Explain what "standard error" is and what its role is in the estimation process.
   * Standard error (s.e) tells us how precise we are with our estimate i.e the level of precision on our estimate. It is calculated as the standard deviation divided by the square root of the sample size, n. It decreases with increase in the sample size. The higher the standard error, the less precise our estimate is. It is commonly represented in the form of confidence intervals which are +/- 2s.e from the estimate. The confidence intervals and standard error does not tell us much about the variability in our data but about the variability in our estimate.
3. Find out what "bias" is in estimation, what are possible sources of bias, and how to reduce bias. You can search the internet!
   * Bias is the systematic error that causes a researcher to consistently overestimate or underestimate the true value of a population parameter.
   * Sources of bias include: selection bias (due to sampling methodology used), measurement bias (e.g. due to a not so well designed data collection or measurement tool) , non-response bias (e.g a certain group in a population are less likely to respond to a survey), omitted variable bias (e.g. when there is a confounding factor in the survey that is overlooked but has implications on the survey results)
   * How to reduce bias: use of probability-based sampling, improving the sampling methodology, standardising data collection methods, piloting data collection tools, doing thorough research to understand potential confounding variables before designing data collection tool, control for confounding variables…
4. Why do we expect that a probability-based sampling process minimises biases in the estimation of a characteristic of a population?
   * Probability-based sampling (or random sampling) is a sampling method where every individual in the population has a known, non-zero chance of being selected. It is expected to minimize biases in estimating a population characteristic because it ensures that every member of the population has a known, non-zero chance of being selected. This helps to provide (i) equal representation of groups of interest from the population (ii) reduce selection bias (iii) minimisation of confounding variables.
5. Explain to a person who has not taken a statistics course the difference between "being precise" and "being accurate" in estimating a population characteristic.
   * Being precise is being consistently or repeatedly getting more or less the same estimate of a variable from your sample, while being accurate means getting a more representative (true) estimate of the population characteristic from your sample. Precision is mostly dependant on the sample size, while being accurate depends on the sample size, sampling methodology, quality of measuring device, quality of data collection and data entry procedures. Between the two, in research, it is better to be imprecise but accurate rather than precise but inaccurate.
6. Explain the difference between “precision” and “variability”.
   * Variability is the underlying differences within our data, and it is independent of the sampe size i.e. as sample size increases, the variability will approximately remain the same. Precision is the uncertainty around our estimate based on the sampling methodology. It increases with increase in sample size.