

## INTRODUCTION TO STATISTICS

### ❖ The role of Statistics

- *Statistics* is the science of collecting, organizing, analyzing and interpreting data.
- Our desire is to understand what is going on in a statistical population, by understanding data obtained through research. (*Research* is a systematic approach to finding answers to questions raised). Research is undertaken:
  - a) To solve a currently existing problem
  - b) To contribute to the general body of knowledge in a field of interest
- The data obtained through research helps us make decisions applied to the above reasons.

### **Steps of a statistical Investigation**

- *Ask a research question / identify a research problem* – e.g. looking at the effect of a particular fertilizer on yield. This can also be stated as an experimental hypothesis using the “if..., then..” statement. This is the outcome expected if the scientist’s speculations are true e.g. “if fertilizer X is used then yield Y will be...”.
- *Collect data* – This is done through either an observational study/ survey or an experimental study. An experimental study involves manipulation of data or imposing some treatment in order to record a response. E.g. changing the composition of a diet to see the effect.
- *Organize data* – usually data collected is not ordered and therefore has to be put in a format that can help us understand it. This can be in ascending or descending order, use of graphs, tables or frequency distributions etc. Organizing data also helps us identify unusual observations called outliers.
- *Analyze data* – This process gives us a sense of what the data is all about, how it is distributed and it’s variability. We use parameters like the mean, mode, median, range, variance and standard deviation to understand the differences between observations and whether these differences are large enough to indicate a real effect. Additionally, statistical tests are done to determine significance.
- *Interpret data and draw conclusions* – This process determines how the data analyzed answers the research question. Conclusions are usually based around a hypothesis called a null hypothesis. Research conclusions are particular to a study and may not be mirrored in other studies.

### ❖ Populations and Samples

- A statistical *population* is the entire group of elements about which we want information. However, it is difficult to observe all of the population and therefore a sample is used. A *sample* is a subset of a population that is randomly chosen and used to answer questions about the population.
- A measure of a characteristic of a population is called a *Parameter*. The 3 main parameters of a population are mean ( $\mu$ ), variance ( $\sigma^2$ ) and covariance ( $\sigma_{xy}$ ).
- A measure of a characteristic of a sample is called a *statistic*. Sample statistics are used to estimate population parameters since population parameters are generally unknown. The 3 main statistics are mean ( $\bar{x}$ ), variance ( $s^2$ ) and covariance ( $s_{xy}$ ).

### ***Sampling in a population***

We consider the following when sampling in a population:

- a) State the objective of your study – objectives should be **SMART** (simple, measurable, attainable, realistic, time bound). The objective is in line with what the research problem is.
- b) Specify population of interest – identify which population will best answer the questions at hand.
- c) Specify variables of interest and how you will measure them – a variable is any characteristic of interest that can be measured. The variable should also be related to the objective.
- d) Specify sampling units – a sampling unit is an object that has been sampled from a statistical population e.g. a student or a group of students. Sampling units can be equivalent to experimental units which are individual entities on which you take measurements. Each experimental unit is subjected to one of several experimental treatments.
- e) Sample size – this is the number of sampling units in your experiment or survey.
- f) Specify sampling frame – this is a list of units in a population that you want to study.
- g) Specify sampling method – this is dependent on whether you are carrying out probability sampling or non-probability sampling. *Probability sampling* entails each member of the population having an equal chance of being selected. *Non-probability sampling* entails each member not having equal chance of being selected. Probability sampling has the following methods; Random, Systematic, Stratified and Cluster sampling while non-probability sampling involves Quota sampling.

