Probability Methods in Engineering

Prepared by
Dr. Safdar Nawaz Khan Marwat
DCSE, UET Peshawar

Lecture 1

Resource Material

- Course Book
  - A. Leon-Garcia, "Probability and Random Processes for Electrical Engineering", 3rd Edition, Pearson Prentice Hall, 2008
- Reference Books
  - D. Bertsekas and J. N. Tsitsiklis, "Introduction to Probability", 2nd Edition, Athena Scientific, 2008
  - Hossein Pishro-Nik, "Introduction to Probability, Statistics, and Random Processes", Kappa Research, 2014

Course Group

- Class code
  - cbefhkk
- Class link
  - https://classroom.google.com/c/NTk0MDE1ODA1NjMw?cjc=cbefhkk
  - Teaching method
  - Combination of slides and white board
  - Interaction about concepts encouraged
  - Interruption to ask questions during lectures allowed

Tentative Grading Criteria

- Exams
  - Final exam: **50%**
  - Mid-term exam: **25%**
- Sessional
  - Attendance: **10%**
  - Assignments: **7.5%**
  - Quizzes: **7.5%**
- All lectures interrelated
  - Each lecture provides base for next lecture
  - Missing any lecture would result in problems in understanding subsequent lectures

- No mobile phone usage during class

Course Outline

- Statistics
  - Descriptive Statistics and Inferential Statistics
  - Applications
  - Sample and Population
  - ….
- Introduction to Probability
  - Axioms
  - Probabilities using Counting methods
  - Conditional Probability
  - Law on total Probability
  - Bayes' Rule
  - …
- Random Variables (RVs)
  - Cumulative Distribution Function (CDF)
  - Probability Density Function (PDF)
  - Mean and variance
- Random Processes
  - Stationary processes
  - Integrals
  - Power spectral density
  - …

How is probability used in everyday life?

- Weather forecasting
- Sports Strategies
- Insurance
- Calculation of batting average in cricket.
- How likely one can win a lottery ticket.
- Playing cards
- Voting strategy in politics
- Rolling a dice.
- Pulling black socks from a drawer of white socks.
- Online shopping
- Online games

Course Significance

- Basis for numerous advanced technologies
  - Wave propagation
  - Wireless communication
  - Communication theory
  - Information theory
  - Pattern recognition
  - Radar and sonar signal processing
  - Network design and optimization

Course Significance (cont.)

- Disadvantages of weak probability concepts
  - No scope in research fields
  - Poor analytical skills
  - Fear of interview questions
  - Inability to conceptualize techniques
  - No major role possible in engineering problem solving
  - Minimum contribution towards nation building
  - Incapability to carry out feasibility studies for mega projects

Weekly Course Outline

Week	Contents
Week 1	Statistics, Descriptive Statistics and Inferential Statistics, Applications of Statistics, Sample and Population
Week 2	Introduction to Mathematical Models, Deterministic Models, Probabilistic Models Basic Concepts of Probability, Axioms of Probability
Week 3	Computing Probabilities using Counting Methods, Conditional Probability
Week 4	Law on Total Probability, Bayes' Rule
Week 5	Independence of Events, Sequential Experiments
Week 6	Binomial Probability Law, Geometric Probability Law
Week 7	Sequences of Dependent Experiments, Random Variables, Notation of a Random Variable
Week 8	Types of Random Variable, Probability Mass Function
	Midterm Examination

Weekly Course Outline

Week	Contents
Week 9	Discrete Random Variables
Week 10	Expected Value, Variance, Standard Deviation
Week 11	Functions of a Random Variable, Expected Value of Function of Random Variables
Week 12	Entropy, Continuous Random Variables
Week 13	CDF, PDF, Memoryless Property
Week 14	Multiple Random Variables, Joint CDF and PDF, Conditional CDF and PDF, CCDF
Week 15	Python codes for Generation of Random variables
Week 16	Course Revision
	Final Term Examination

MAPPING OF CLOS WITH PLOS:

CLO #	Course Learning Outcomes (CLOs)	Level of Learning (Bloom's Taxonomy)	Program Learning Outcomes (PLOs)
1	Use essential concepts of probability and apply analytical methods for solving engineering problems.	Cog-3 (Application)	PLO1 (Engineering Knowledge)
2	Use the concepts of random variables and solve mathematical problems related to stochastic systems.	Cog-3 (Application)	PLO3 (Design/Development of Solutions)
3	Apply mathematical skills and demonstrate the use of software tools for implementation of probabilistic models.	Cog-3 (Application)	PLO5 (Modern Tool Usage)

Introduction to Statistics

- History of Statistics
- An Overview of Statistics
- Data Classification
- Statistics Examples in Real Life

Introduction to Statistics

History of Statistics

- The word "Statistics" was first used by a German Scholar Goti Fried Achenwall about the middle of the 18th Century

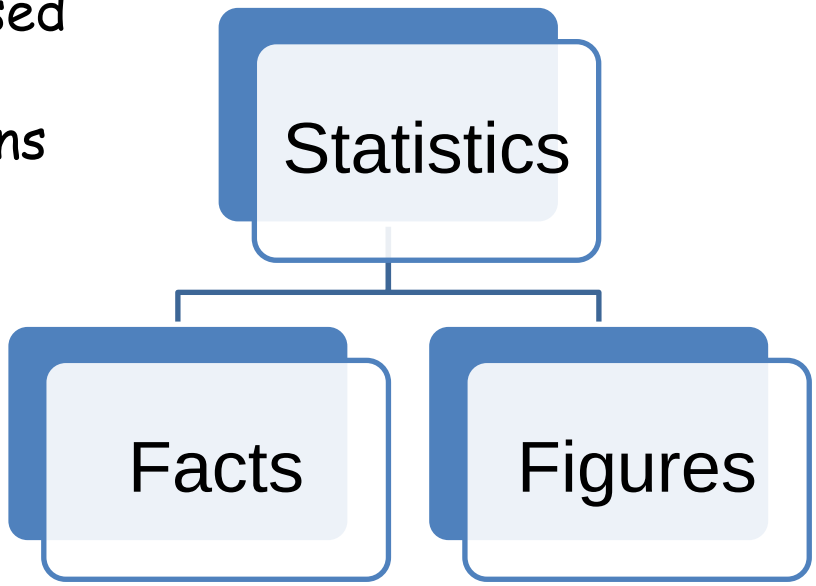
- The science of the statecraft concerning the collection and use of the data by the state.

- The word "Statistics" is derived from the "Latin" word "Status"
- The word "Statistics" is derived from the "Italian" word "Statista"
- The word "Statistics" is derived from the "German" word "Statistak"

A man equipped with the knowledge of statistics is a better ruler a better policy maker and a better administrator.

Introduction to Statistics

Statistics is a group of methods used to collect, analyze, present, and interpret data and to make decisions



Data are the facts and figures that are collected, analyzed, and summarized for presentation and interpretation.

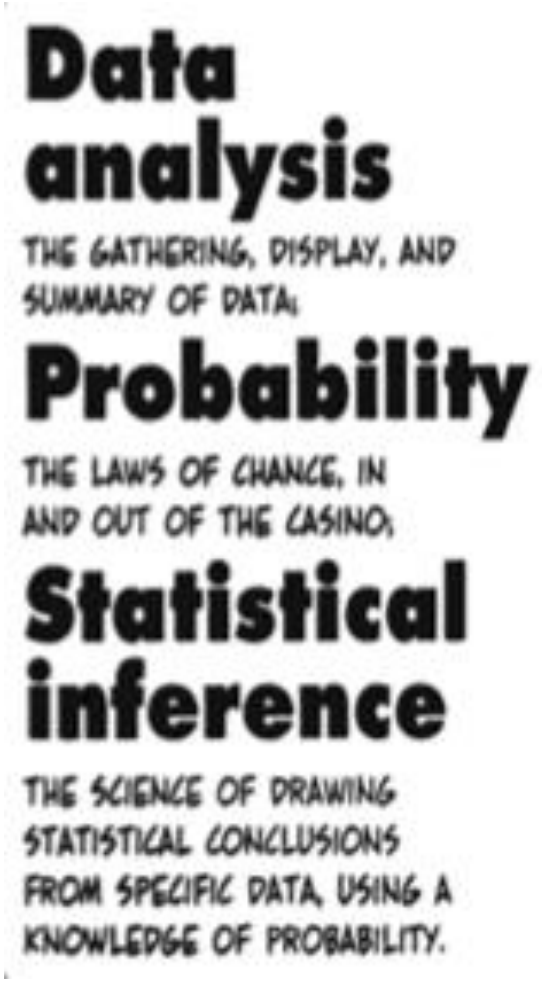
Figures are visual presentation of results, including graphs, diagrams and pictorial charts etc.

Facts which are obtained from analyzing information expressed in numbers for example about the number of times something happens.

Introduction to Statistics

An Overview of Statistics

The science of collectiong, organizing, presenting, analyzing, and interpreting data to assist in making more effective decisions



What is statistics

- Science of data
- Data are numbers with context
- It can be broken down to three branches:
  - Data analysis
  - Probability
  - Statistical Inference

Introduction to Statistics

Data

Statistical data are usually obtained by counting or measuring items. Most data can be put into the following categories:

- Continuous
- Discrete
- Digital

- **Qualitative** - data are measurements that each fail into one of several categories. (hair color, ethnic groups and other attributes of the population)

- **Quantitative** - data are observations that are measured on a numerical scale (distance traveled to college, number of children in a family, etc.)

Introduction to Statistics

Decide whether the following data are qualitative, discrete quantitative or continuous quantitative.

- 1. Number of cars
- 2. Mass of an object
- 3. Distance of University from home
- 4. Day of the week
- 5. Color of cars
- 6. Pocket money
- 7. Favorite Football team
- 8. World ranking
- 9. Birth place
- 10. Age

Introduction to Statistics

Nowadays the word statistics refers to "Numerically facts systematically arranged". It is always used in plural sense

- Statistics of Literacy
- Statistics of Corona Patient
- Statistics of Petrol Prices
- Statistics of Road accident
- Statistics of Rain
- Statistics of Crimes
- Statistics of Birth
- Statistics of Education

Introduction to Statistics

Statistics Examples in Real Life

- Population Record
- Budgeting and Finance
- Record of Production Goods and Services
- Stock Market Data Analysis
- Weather Forecasting
- Medical Records
- Sales Tracking
- Health Care Departments
- Educational Data
- Natural Disaster Prediction
- Sports

Introduction to Statistics

- Research and Analysis
- Banking
- Business Statistics
- Data Science
- Transportation
- Cryptocurrency
- Predicting Diseases
- Research Interpretations and Conclusions

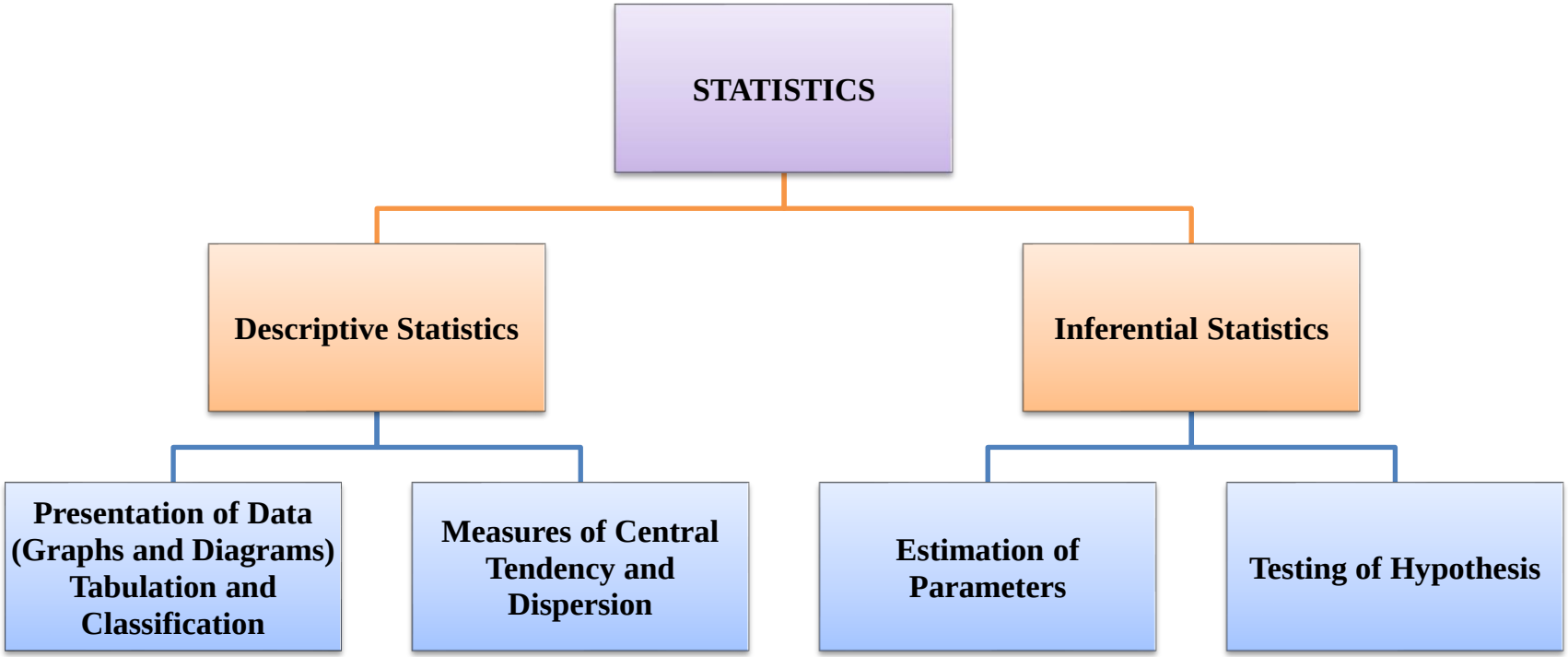
Introduction to Statistics

Types of Statistics

- **Descriptive Statistics:** deals with concepts and methods related with the summarization and description of the important aspects of numerical data. It consists of condensation of data, their graphical displays and computation of numerical quantities that can provide information about the centre and spreadness of observations of a data set.

- **Inferential Statistics:** deals with methods and procedures used for drawing inferences about the true but unknown characteristics of a population based on the sample data derived from the same population. Inferential statistics can be further classified into estimation of parameters and testing of hypothesis.

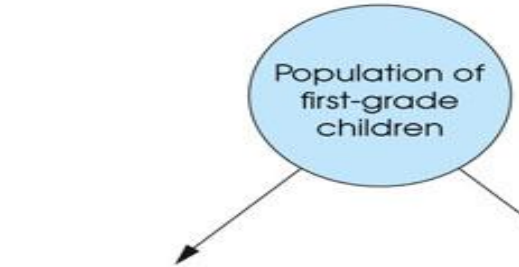
Introduction to Statistics



Introduction to Statistics

Step 1
Experiment:
Compare two teaching methods

Data
Test scores for the students in each sample



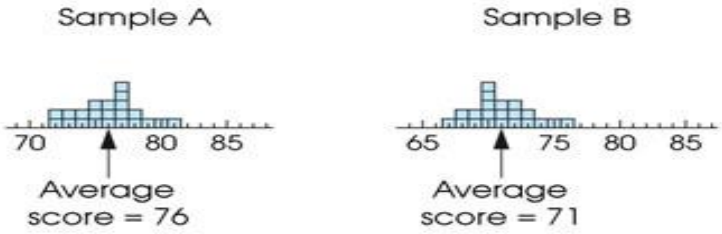
Sample A
Taught by Method A

73	75	72	79
76	77	75	77
72	75	76	78
80	74	76	78
73	77	74	81
77	77		

Sample B
Taught by Method B

68	70	73	71
67	72	70	71
75	68	70	71
72	74	69	72
76	73	70	70
69			

Step 2
Descriptive statistics:
Organize and simplify



Step 3
Inferential statistics:
Interpret results

The sample data show a 5-point difference between the two teaching methods. However, there are two ways to interpret the results:

1. There actually is no difference between the two teaching methods, and the sample difference is due to chance (sampling error).
2. There really is a difference between the two methods, and the sample data accurately reflect this difference.

The goal of inferential statistics is to help researchers decide between the two interpretations.

Introduction to Statistics

Population:

- An aggregate or totality having some common characteristics on interest is called population. It is also called *Universe*.
- For example, total number of students enrolled in im|Sciences-Peshawar, total number of markets in Peshawar city, total number of banks in Hayatabad, number of industries in the province, monthly/yearly sales of the stores in Peshawar district, etc.

Sample:

- A small representative part of population is called sample.
- For example, a small portion/part of students of im|Sciences-Peshawar will constitute a sample of students. Similarly, a randomly/purposively selected number of markets from a bulk of markets is called a sample of markets.

Introduction to Statistics

Parameter:

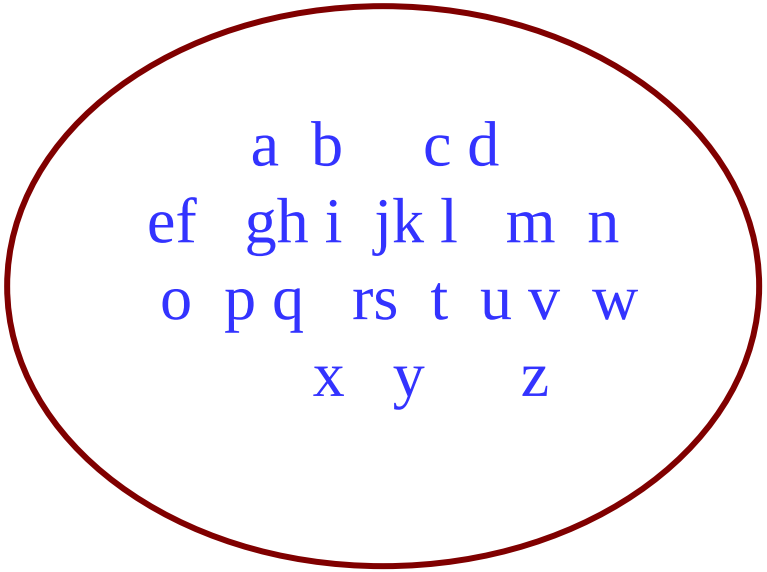
- An numerical quantity like mean, standard deviation etc computed/obtained from population data is known as parameter.
- For example, average monthly/yearly sale of all the stores located in district Peshawar etc. Parameters are generally used to specify the distribution of data.

Statistic:

- An numerical quantity like mean, standard deviation etc computed from sample data is called statistic.
- For example, the average GPA of the 50 students that are selected from a population of 300 students. Similarly, the average sale of 100 stores instead of 1000 stores etc are the examples of statistic(s).

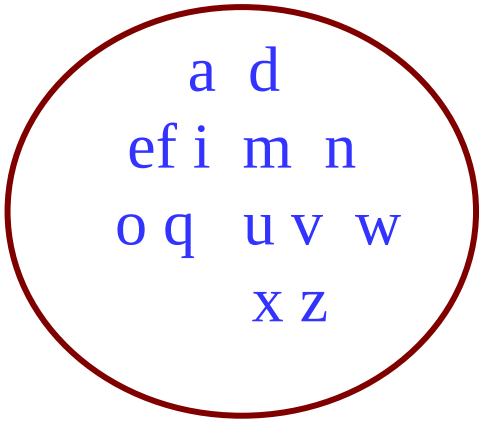
Population vs Sample

Population



Measures used to describe a population are called parameters

Sample



Measures computed from sample data are called statistics

Introduction to Statistics

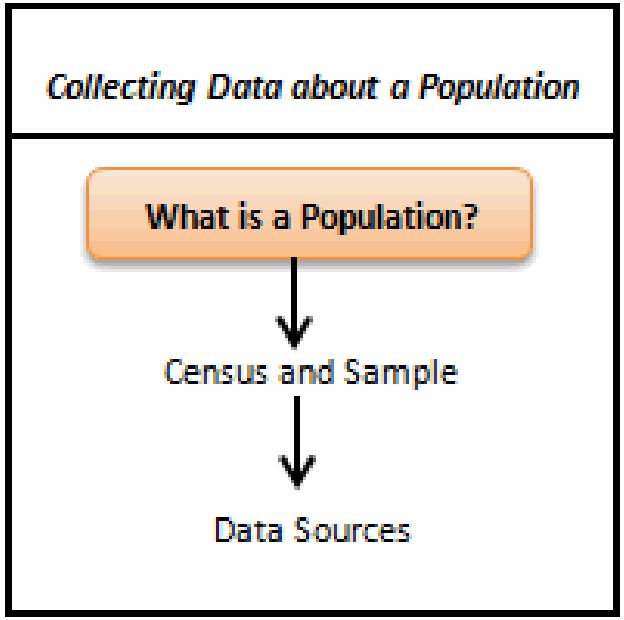
- A population is any complete group with at least one characteristic in common.

- Populations are not just people. Populations may consist of, but are not limited to, people, animals, businesses, buildings, motor vehicles, farms, objects or events.

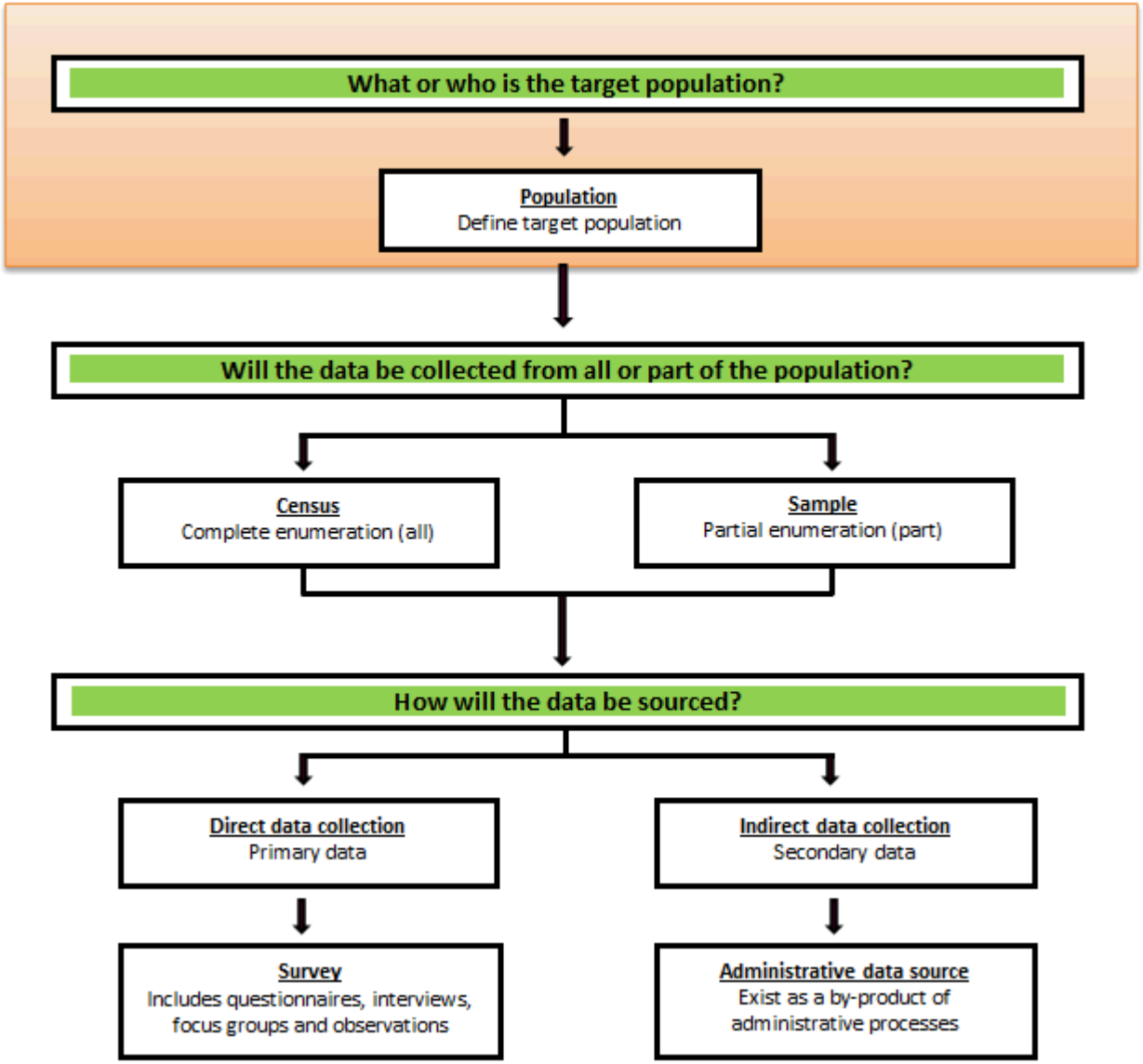
- **Why do you need to know who or what are in a population?** When looking at data, it is important to clearly identify the population being studied or referred to, so that you can understand *who* or *what* are included in the data. For example, if you were looking at some Australian farming data, you would need to understand whether the population the data refers to is all farms in Australia, just farms that grow crops, those that only have livestock, or some other type of farm.

Introduction to Statistics

- **When is a population identified?**
  The population needs to be clearly identified at the beginning of a study. The study should be based on a clear understanding of who or what is of interest, as well as the type of information required from that population.



Collecting Data about a Population Flowchart: What is a Population?



Introduction to Statistics

- A population may be studied using one of two approaches: taking a census, or selecting a sample.

- It is important to note that whether a census or a sample is used, both provide information that can be used to draw conclusions about the whole population.

- **A census is a study of every unit, everyone or everything, in a population.** It is known as a complete enumeration, which means a complete count.

- **A sample is a subset of units in a population, selected to represent all units in a population of interest.** It is a partial enumeration because it is a count from part of the population.
- Information from the sampled units is used to estimate the characteristics for the entire population of interest.

Introduction to Statistics

Pros of a CENSUS	Cons of a CENSUS
•Provides a true measure of the population (no sampling error)	•May be difficult to enumerate all units of the population within the available time
•Benchmark data may be obtained for future studies	•Higher costs, both in staff and monetary terms, than for a sample
•Detailed information about small sub-groups within the population is more likely to be available	•Generally takes longer to collect, process, and release data than from a sample

Introduction to Statistics

Pros of a SAMPLE	Cons of a SAMPLE
•Costs would generally be lower than for a census	•Data may not be representative of the total population, particularly where the sample size is small
•Results may be available in less time	•Often not suitable for producing benchmark data
•If good sampling techniques are used, the results can be very representative of the actual population	•As data are collected from a subset of units and inferences made about the whole population, the data are subject to 'sampling' error
	•Decreased number of units will reduce the detailed information available about sub-groups within a population

Introduction to Statistics

How are samples selected?

- A sample must be robust in its design and large enough to provide a reliable representation of the whole population. Aspects to be considered when designing a sample include the level of accuracy required, cost, and the timing. Sampling can be random or non-random.

- **In a random (or probability) sample each unit in the population has a chance of being selected, and this probability can be accurately determined.**

- Probability or random sampling includes, but is not limited to, simple random sampling, systematic sampling, and stratified sampling. Random sampling makes it possible to produce population estimates from the data obtained from the units included in the sample.

Introduction to Statistics

- ***Simple random sample:*** All members of the sample are chosen at random and have the same chance of being in the sample. A lottery draw is a good example of simple random sampling where the numbers are randomly generated from a defined range of numbers (i.e. 1 through to 45) with each number having an equal chance of being selected.

- ***Systematic random sample:*** The first member of the sample is chosen at random then the other members of the sample are taken at intervals (i.e. every 4th unit).

- ***Stratified random sample:*** Relevant subgroups from within the population are identified and random samples are selected from within each strata.

Introduction to Statistics

- **In a non-random (or non-probability) sample some units of the population have no chance of selection, the selection is non-random, or the probability of their selection can not be determined.**

- In this method the sampling error cannot be estimated, making it difficult to infer population estimates from the sample. Non-random sampling includes convenience sampling, purposive sampling, quota sampling, and volunteer sampling

- **Convenience sampling:** Units are chosen based on their ease of access;
- **Purposive sampling:** The sample is chosen based on what the researcher thinks is appropriate for the study
- **Quota sampling:** The researcher can select units as they choose, as long as they reach a defined quota;
- **Volunteer sampling:** participants volunteer to be a part of the survey (a common method used for internet based opinion surveys where there is no control over how many or who votes).