

## STAT150 LECTURE WEEK 1

- Statistics is the science of learning from data (collecting, presenting, analysing and interpreting data)

### **Objectives and scope of statistical studies**

#### Objectives

- Primary objective of obtaining information of **target population** using a **sample**
- Target population: relevant subjects of interest
- Sample: manageable subset making study feasible
- Sample used to answer questions about target population

#### Scope of statistical studies

- Study design - data analysis – interpretation of results – next study?

### **Study design, types of studies, variables**

#### Study design

- Formulate question (**who, what, why**)
- Specify target population (**who/what, where, when**)
- Determine measurements (**variables**)
- Define method of collection (**how, when, where**)

#### Populations and samples

- Sample should be non-biased representation of target population
- Observations should be independent

#### Selecting a sample

- **Representative sample** is unbiased and large populations giving accurate information used to make inferences (characteristics should represent those of the target population without bias)
- **Simple random sample** when each member has the same chance (names out of a hat) – ensuring sample is representative of the target population

#### Types of study

**Deductive (non- empirical)** – **no particular evidence (Sherlock Holmes)**

**inductive (empirical)** – **qualitative (unstructured)/ quantitative (structured)** –

**observational/experimental** **collecting data**

- **Observational** no intervention or treatment imposed by investigator
- **Experimental** investigator has some control over the determinant

#### The data we collect: variables

- When data is collected values are collected of variables
- Called **random variables** as values differ between subjects
- Variables take on 3 roles: **predictors** influences **outcomes**

### **Bias, sample size issues**

#### Bias

- Any **systematic error** resulting in incorrect parameter or association between variables
- Types: collection bias, measurement bias, response bias, confounding

#### Selection bias

- Systematic differences when selecting the subjects for the study