Introduction to Biostatistics
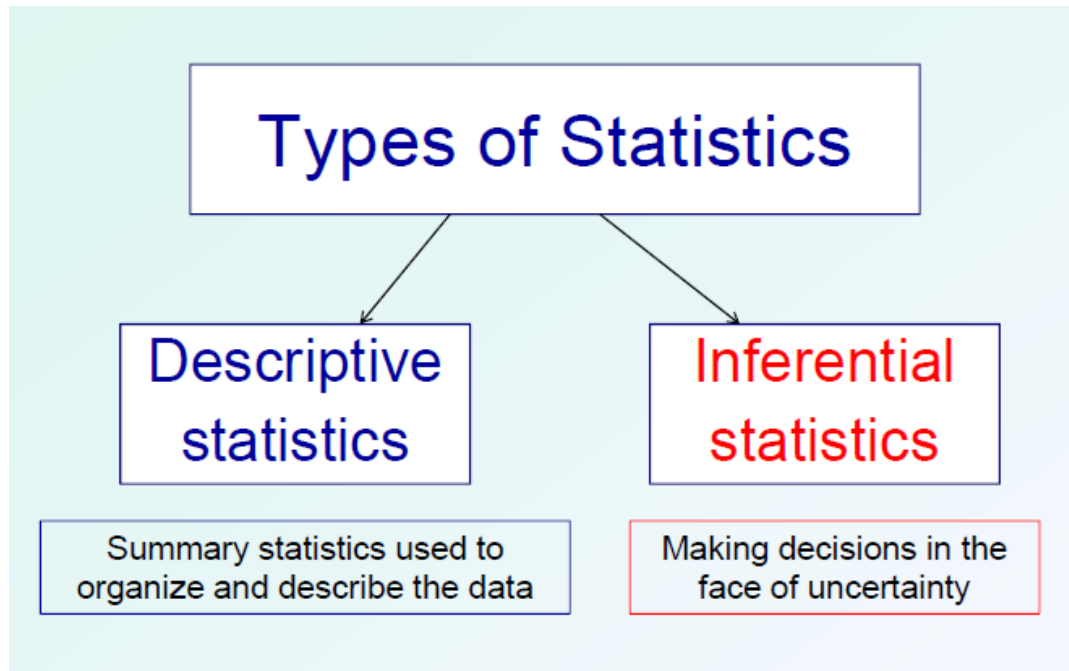
Kashif Kamran

Statistics

- "Statistics may be defined as a body of methods for making wise decisions in the face of uncertainty."

- Statistics is not just an extension of mathematic.

Applicable to

- Clinical research
- Basic science and laboratory research
- Epidemiological research

Descriptive statistics uses the data to provide descriptions of the population, either through numerical calculations or graphs or tables.

Inferential statistics makes inferences and predictions about a population based on a sample of data taken from the population in question.

Types of Data

Variable – anything that varies within a set of data

- Mortality rates
- Survival time
- LDL cholesterol
- Surgery type
- Biopsy stage
- Compliance
- Marital status
- Age
- Weight
- Smoking status
- Adverse drug reaction
- Energy intake
- Parity
- Drug dose

Types of Data

Important in deciding which analysis
methods will be appropriate

Categorical (qualitative) variables
- Sex, ethnicity, smoker/non-smoker, blood type

Numerical (quantitative) variables are measured
- Age, weight, parity, triglycerides, tumor size

Sample and population

- Populations are rarely studied because of logistical, financial and other considerations

- Researchers must rely on study samples

- Many types of sampling design • Most common is simple random sampling
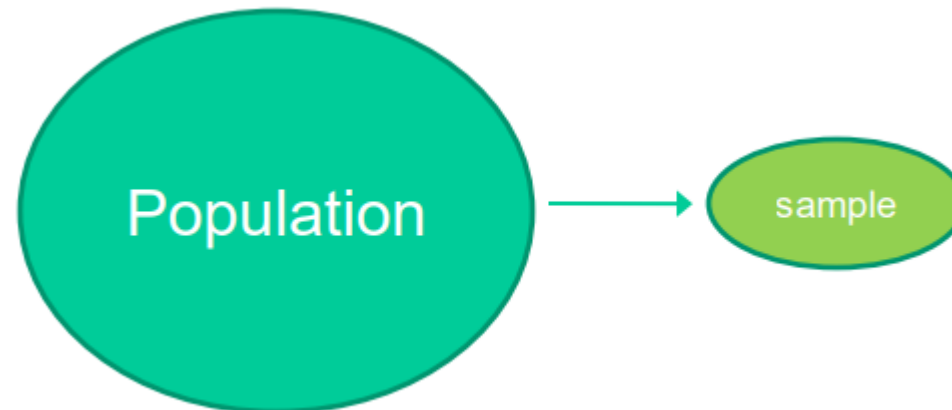
Random sampling

- Suppose that we want to estimate the mean birthweights of Malay male live births in Singapore

- Due to logistical constraints, we decide to take a random sample of 100 Malay live births at the National University Hospital in a given year
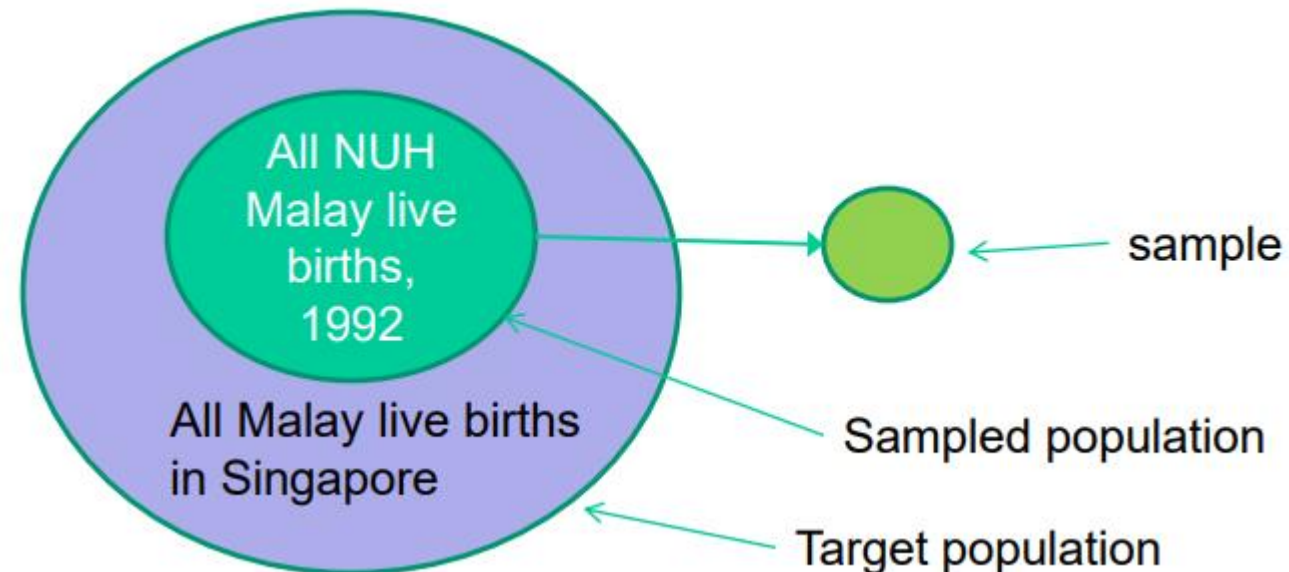
Standard deviation vs. standard error

• Standard deviation (SD) tells us variability among individuals • • Standard error

$$S.E._{\bar{X}} = \frac{s}{\sqrt{n}}$$

standard deviation of the population. Where s is sample standard deviation and n is the sample size

• Standard error (S.E.) tells us
• variability of sample means

$$\sigma = \sqrt{\frac{\sum(x_i - \mu)^2}{N}}$$

σ = population standard deviation

N = the size of the population

x_i = each value from the population

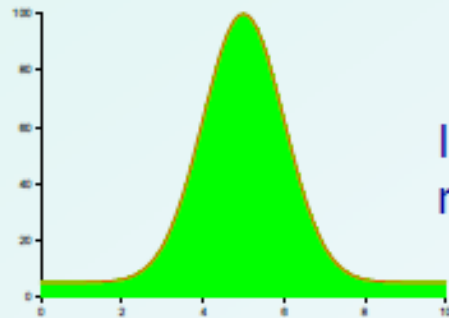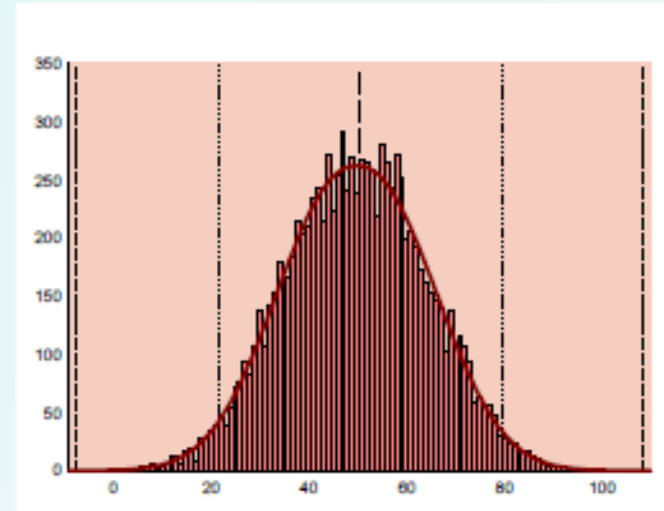μ = the population mean

Descriptive statistics

- Measures of Central Tendency

- Measures of Dispersion

Measures of Central Tendency*

*or Measures of Location

- Mean
- Median
- *Geometric mean*
- Mode

In a <u>symmetric</u> distribution, the median, mode and mean will have the same value.