SCI1020

Lecture 1 – Introduction

Basic Terminology

- Unit → The objects described by a set of data
- Variable → Any characteristic of an individual; can take different values
- Random Variable → Any random unit will have a random value
- Categorical Variable → Places an individual into groups/categories
- Quantitative → Numerical values; continuous scale
- Ordinal → Numbered order is evident
- Discrete → Numbers are counts of a characteristic value
- Explanatory → Fixed value
- Response → Variable that responds to explanatory values
- Distribution → What values a variable takes and how often it takes these values
- Histogram → Frequency distribution of continuous quantitative data

Lecture 2 – Graphical Distributions (Chapter 1)

Examining Quantitative Continuous Data

- Location: Around what value are the date located?
  - Magnitude
  - Central tendency
- Spread: What is the variability among the data values?
  - Range
  - Most values are within what limits?
- Shape: What is the distribution of the data?
  - Overall pattern of data
  - Deviations from overall patterns
  - Symmetric or skewed; single peak or multimodal
  - Outliers

Histograms

- Frequency distribution of continuous quantitative data
  - Divide the possible values into class intervals
  - Count how many observations fall in each interval
  - Draw chart representing this distribution (x-axis is continuous)
  - Direction of the 'tail' is the direction of which it is skewed

Shape of Histograms

- Symmetrical: Usually biological measurements; bell shaped
- Skewed: Data trails off/picks up and forms a 'tail'; distinguish the "vital few" from the "trivial many"
- Bi-modal: When sub-groups have different values to each other but similar values within the sub-group

Lecture 3 – Numerical Distributions (Chapter 2)

Centre of a Distribution

- Mean: Arithmetic average of all the data values; $\frac{\sum xi}{n}$
- Median: The 'middle' value; not skewed by outliers; $\frac{n+1}{2}$

Spread of a Distribution

- Standard Deviation: Variability (on average) that the individual data values are from the mean; $\sigma = \sqrt{\frac{\sum(x-\mu)^2}{n}}$
  - In excel: STDEV.S(_:_)
- Quartiles: The 25% and 75% position in the ordered list of data; inter-quartile range is the middle 50% of values (IQR=Q3-Q1)

Five Number Summary

- Provides a quick overview by dividing the data values into quarters
- (Min, first quartile Q1, median, third quartile Q3, max)
  1. Sort data from minimum to maximum values
  2. Determine the median
  3. Determine first quartile (middle of lower half of values)
  4. Determine third quartile (middle of upper half of values