Week One: Introduction

Population:
- The entire collection of events in which you are interested
- All individuals that you are interested in studying

Sample:
- A subset of individuals selected from the entire population

Parameter:
- Characteristic of a *population*

Statistic:
- Characteristic of the *sample*

Two main types:
1. Descriptive statistics – simply used to *describe* the data
    - Summarise the data
    - Averages and range of scores etc.
    - Makes the data more 'manageable'
    - Can't infer anything from these statistics
2. Inferential statistics – used when we want to answer research questions
    - Allows us to make generalisations from the sample to the population
    - Eg. can infer the behaviour of all office workers based on the data we collect from the sample of 100 office workers

Variables
- A variable is anything that can take on different values
- Data is obtained by measuring variables
- *Independent variable (IV)* – controlled by the researcher (the 'cause')
- *Dependent variable (DV)* – the actual measured date (the 'effect')

Two main types:
- *Discrete variable* only has a limited number of values
    - Categorical
    - Eg. gender (male or female), age (young or old)
- *Continuous variable* can take on many different values
    - Measured on a continuum
    - Eg. age, IQ score, height, reaction time

Data Types:
- The type of data we collect influences the type of statistical approach we use
- When we are dealing with *continuous variables* we have *measurement data*
    - Usually summarised using means
- When we are dealing with *discrete variables* we have *categorical data*
    - Usually summarised using percentages

Measurement Scales:
1. Nominal Scale
   - Simple categories with different names
   - No underlying scale and no ordering
   - Eg. what's your favourite colour? Red, blue, green, yellow
2. Ordinal Scale
   - Categories with different names
   - Simple categories in an ordered sequence
   - Allows us to determine the direction of the difference
   - Differences between the categories is unknown
   - Eg. degree of illness – none, mild, moderate, severe
3. Interval Scale
   - Equal distances between points on the scale
   - Generally continuous variables
   - Eg. temperature
   - No true zero point – eg. zero degrees does not mean an absence of temperature
4. Ratio Scale
   - Equal distances between points on the scale
   - Has a true zero point
   - Eg. questionnaire measure of anxiety (0-100)
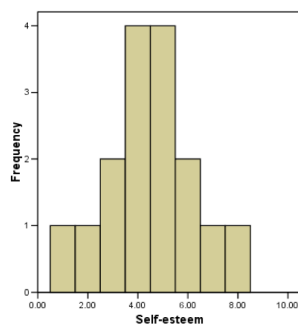
Data Description and Exploration:

Organisation of data – summarise and simplify information.
This can be done through:
- Frequency distributions

Self-esteem score	Frequency
0	0
1	1
2	1
3	2
4	4
5	4
6	2
7	1
8	1
9	0
10	0

- Histograms

- Stem-and-leaf displays

Raw data	Stem	Leaf
	0	000235679
0 0 0 2 3 5 6 7 9	1	011279
10 11 11 12 17 19	2	022336677777
20 22 22 23 23 26	3	0012225
26 27 27 27 27 27	4	29
30 30 31 32 32 32	5	07
35 42 49 50 57 62	6	28
68 71 72 72 79 80	7	1229
	8	0

Types of Distributions:
1. *Normal* distribution – most people score in the middle of the scale; score become less frequent in the extremes
2. *Bimodal* distribution – at least two peaks/modes (bimodal or multimodal)
3. *Negatively skewed* distribution – more high scores
4. *Positively skewed* distribution – more low scores
5. *Kurtosis* distributions – kurtosis refers to how flat or peaked the distribution appears
   o *Leptokurtic* – the distribution is characterised by a high peak in the centre of the scale. Unlike the normal distribution there's no people in the extremities
   o *Platykurtic* – the distribution is flatter, less scores in the centre

Week Two: Central Tendency, Variability & Z-Scores

Measures of Central Tendency:
- Any statistics that represent the centre of a distribution
- Scores that represent the data
- There are three measures:
  1. *Mode* – most common
     $$\bar{x} = \frac{\sum X}{N}$$
  2. *Median* – mid-point
  3. *Mean* - average
- If there are two modes:
  o Adjacent – average of the two values
  o Nonadjacent – bimodal distribution (report both modes)
- Median and mode ARE NOT influenced by extreme scores
- The mean IS influenced by extreme scores and therefore is not a good measure of central tendency in this case

Measures of Variability:
- How representative is the measure of central tendency?
- How well does the mean characterise the data we have?
- Measures of variability tell us the degree to which scores vary from the average value

Calculating Variability:

The average deviation
1. Calculate mean
2. Compute how much each score deviates from the mean
3. Average the deviations