Statistics
Chapter Three

- **Central Tendency**: Indexes of central tendancy provide a single number to characterize a distribution.
  - → Measures of central tendencies come from the centre of the distribution of data values, indicating what is "typical", and where data values tend to cluster
  - → Popularly called an "average"
  - → Three alternative indexes:
    - → Mode → Median → Mean

- **The mode**: the score value with the highest frequency
  - → Advantages
    - → Can be used w/ data measured on any measurement level (including nominal)
    - → Easy to compute
    - → Reflects an actual value in the distribution
  - → Disadvantages
    - → Ignores most info in the distribution
    - → Tends to be unstable (varies sample to sample)
    - → Some distributions may not have a mode

- **The Median**: the score that divides the distribution into 2 equal halves - 50% above 50% below
  - → Advantages
    - → Not influenced by outliers
    - → Good index of what is "typical" when distribution is skewed
    - → Easy to compute
    - → Appropriate when data are ordinal
  - → Disadvantages
    - → Only an index of position, doesn't take actual data values into account
    - → Not necessarily an actual data value → more difficult to understand

Chapter Three

- The mean: the arithmic average, data values are summed and divided by # of values > N
  - → Usu. preferred for interval & ratio level data
  - → Equation: $M = \Sigma X \div N$
    - → M: Sample mean also $\bar{X}$
    - → X: actual data values
    - → N: # of people
  - → Advantages
    - → The balance point in the distribution
    - → Doesn't ignore any information
    - → Most stable index of central tendancy
  - → Disadvantages
    - → Sensitive to outliers
    - → Value is not often an actual data value
    - → Gives a distorted view of what is "typical" when data are skewed

- In a normal distribution, all three indexes coincide, exact same

- In a skewed distribution, the mean is pulled in the direction of the skew
  - → Positive Skew mean > median & mode
  - → Negative Skew mean < median & mode

- Variability: how spread out or dispersed data values in a distribution are
  - → 2 distributions w the same mean could have very different dispersion
  - → High variability > heterogeneous
  - → low variability > homogeneous

Chapter Three

- Indexes of variability
→ Range → Interquartile Range → Standard deviation
→ Variance

- Range: the difference btwn the highest & lowest value in the distribution
→ Advantages
→ Easy to compute & understand
→ Disadvantages
→ Depends on only 2 scores & influenced by sample size
→ Sensitive to outliers
→ Tends to be unstable

- Interquartile Range: Based on quartiles > Q3 - Q1, where the middle 50% of scores lie
→ lower quartile Q1: Below which 25% of score lie
→ Q2 > median
→ Q3 : Below which 75% of scores lie

- Standard Deviation: An index that conveys how much, on average, scores in a distribution vary
→ $x = X - M$
→ x = deviation score → X = Original Score → M = mean
→ Sum always equals zero ∴ deviation scores must be squared, added, then $\sqrt{}$
→ $SD = \sqrt{\Sigma x^2 \div (N-1)}$

→ Indicates the average amount of deviation of scores from the mean

Chapter Three

- Standard deviation provides valuble information when the distribution is normal
  - → There are approximately three SD's above & below the mean in a normal distribution
  - → In a normal distribution a fixed percentage of cases lie within certain distances from the mean

- Variance: not used descriptively, not easily interpreted because it is not in units of original data, it's in units squared
  - → Variance = SD^2

- Relative Standing
  - → Central tendancy and variability indexes describe a distribution
  - → There are also descriptive statistics to describe individual scores > their relative standing in a distribution
    - → Percentile ranks
    - → Standard Scores

- A percentile is one hundredth of a distribution, a quartile divides a distribution into quarters

- Percentile Rank: the location of a given score in a distribution of scores > What percentage of cases fall at or below that value

- Standard score (aka z score): a score expressed in standard deviation units > how far a score is in relative distance from the mean
  - → $Z = (X - M) \div SD$
    - → Mean subtracted from original score, divided by SD