Measurement and descriptive statistics

4 types of measurement

- *Nominal measurement* involves using numbers as labels
- *Ordinal measurement* is when the categories are ordered on some continuum
- *Interval measurement* are like ordinal but also provide information about the magnitude of the difference between items
- *Ratio measurement* have all the properties of interval but also provides ratios of the items measured (e.g. it is twice as expensive with ratio of 2:1)

Qualitative and quantitative

- Variables measured on an ordinal, interval, or ratio level are *quantitative*
- Variables measured on a nominal level are *qualitative*

Discrete and continuous variables

- *Discrete* variables have only a finite number of values that can occur between any 2 points (e.g. no of family members)
- *Continuous* variables can theoretically assume an infinite number of values between any 2 points (e.g. reaction time)

Populations and samples

- *Parametric tests* are based on a population
- *Statistics* are based on a sample (from a population)
- Always want to make sure you have a representative sample of the population

Descriptive and inferential statistics

- *Descriptive statistics* involves describing a group of scores in a clear and precise manner (e.g. getting the mean/SD)
- *Inferential statistics* takes measurements on a sample then uses them to infer something about the population

The concept of probability

- p = number of observations favouring A / total number of possible observations

- A frequency distribution is a compilation of all the score values in a set of scores and the number of times each occurs (presented in frequency tables)

Measures of central tendency and variability

Measures of central tendency

- *Measures of central* tendency are referred to as descriptive statistics
- Central tendencies measure an 'average' of scores and also are measures of variability

Measures of central tendency for quantitative variables

- *Mode* = most common score
- *Median* = middle score
- *Mean* = average score
- *Range* = highest minus lowest score
- *Interquartile range* = divide into 4 quartiles then minus highest from lowest
- *Sum of squares* (SS) = sum of the squared deviations from the mean (never negative)
- *Variance* (S^2) = sum of squares / observations (N) $s^2 = \frac{SS}{N-1}$
- *Variance from sample estimate of population value*
- *Standard deviation (*S) = square root of the average squared deviation from the mean
- The sum of squares, variance, and standard deviation will always be greater than or equal to 0 (can never be negative)

Measures of central tendency for qualitative variables

- *Skewness* = is the tendency for scores to cluster on 1 side of the mean (pos or neg skewed)
- *Kurtosis* = how flat the peak of a distribution is, and how long/flat the tails are

Normal distribution

Standard scores

- *Standard scores* are used to compare your position in a set of scores
- The z-score is a standard score in a normal distribution (formula)
- A *standard score* (z-score) represents the number of standard deviation a score falls above or below the mean
- A standard score of 0 is equal to the mean
- The SD and variance of a set of standard scores is always equal to 1.00

Characteristics of a normal distribution

- Mean is always 0 and SD is always 1
- All normal distributions are symmetrical and have a 'bell curve'
- The mode, median, and mean all have the same value

Chi-square test

When to use

- Used when both variables in the relationship we are interested in are nominal level
- both variables are measured on the same individuals
- We compare the observed value to the expected value (what we would expect if the null hypothesis was true)
- Our expected value (E) is basically the amount we would "expect" there to be in each cells if the null hypothesis was true (so there is no relationship between the two variables)
- Can be used to test independence and/or goodness of fit
    - *Independence* means variables A and B are not related

Assumptions of the chi-square test

- The observations are independently and randomly selected
- The expected value for each cell is nonzero (preferable if > 5)
- Nominal data

Calculations

- $$E = \left(\frac{CMF}{N}\right)(RMF)$$

- CMF is the amount of columns
- RMF is the amount of rows
- We calculate the expected value for each cell
- Then we compare the Observed value for that cell (how many times that particular combination of the two variables occurred) with the expected value for that cell (how many times we would expect that particular combination of the two variables to occur).
- The table you put the values in is called the *contingency table*
- χ^2 is the *chi-square statistic* and reflects the overall difference between the observed and the expected
- H0: … are unrelated in the population

  H1: … are related in the population
- Df = (r-1) (c-1)

More information