Research Methods A - Exam Revision

Table of Contents

Week 1

Population vs Sample

Population: the entire set of individuals, or events, of interest in a particular study

- E.g. refers to every office worker in Australia.
- To make the study feasible we draw a random sample from this population, which we hope is representative.

Sample: set of individuals selected from a population

- Representative sample: A sample that shares the key characteristics of the population from which it has been taken
- e.g., 100 workers from each state

Parameter vs. Statistic

Parameter: a value that describes a key characteristic of the population

Statistic: a value that describes a key characteristic of the sample

- Statistics are used to estimate values that exist in the population.

Types of Variables

Discrete: Contains only a small number of values

- E.g. handedness (right/left/ambi), favorite season (summer/autumn/ winter/ spring)
- Often referred to as Categorical data

Continuous: Contains many different values

- Weight (40kg- 140kg), age (0-100 years)
- Measured along a continuum
- Often referred to as Measurement data

Independent variable (IV): The variable that we manipulate

- The IV is controlled by the researcher
- For example, group membership – the researcher assigns participants to either high anxiety or low anxiety groups
- Researcher randomly allocates people to either a treatment group or a control group.

- The IV 'causes' differences/changes in the dependent variable

Dependent variable: the variable that we measure

- The actual measured data
- The variable that is observed for differences/changes
- For example, levels of depression in control vs. treatment groups
- The DV is influenced by the IV

Measurement Scales

Nominal: Nominal scales are simply categories with different names

- There is no underlying scale and no ordering
- Example: Religion is usually measured on a nominal scale
- Christian, Muslim, Jewish, Other. (1, 2, 3, 4)
- Can't say that one category is larger/higher/more than another
- They are just labels for the different categories.

Ordinal: Ordinal scales are categories with different names AND are organised into an ordered sequence

- Example: Degree of illness – None, Mild, Moderate, Severe.
- This allows us to determine the direction of the difference
- We can say that severe is greater than moderate; and moderate is greater than mild
- But distances between the categories is unknown

Interval: Equal distances between points on the scale.

- Many more points than on an ordinal scale (continuous data)
- Temperature (F, C) is measured on an interval scale
- Distance between 10 – 20 degrees is exactly the same as the distance between 50 – 60 degrees.
- But, no true zero point.
- 0 degrees does not mean an absence of temperature.

Ratio: Equal distances between points on the scale and a true zero point.

- Time, length, age.
- Such variables have true zero points.

Type of Scale	Scale Qualities	Example
Nominal	Categories None	Gender Hair colour Religion
Ordinal	Ordered categories Magnitude	Small, medium, large None, moderate, severe
Interval	Magnitude Equal intervals	Temperature
Ratio	Magnitude Equal intervals True zero point	Age Time Weight & Height..

Type of Statistics

Descriptive statistics: organizing and summarizing data

- Frequency tables and graphs: a way of organising and simplifying the data we've collected
- They provide two key pieces of information:
- 1. The set of scores or range of categories that people could have either obtained or fallen into on the variable on interest
- 2. A record of frequency, or number of individuals, who obtained each score or fell in each category

Inferential statistics: interpreting data and drawing conclusions

- Inferential statistics are generally used after we have described the data and are interested in answering specific research questions.

Types of distributions

Bimodal distributions
When a frequency distribution has two peaks it is termed bimodal.

Negatively skewed distributions
When more people score at the high end of a scale, the distribution is said to be negatively skewed.

Positively skewed distributions
When more people score at the low end of a scale, the distribution is said to be positively skewed.

Kurtosis
Kurtosis refers to how 'flat' or 'peaked' the distribution appears.

Week 2

The Central Tendency

- Statistics that represent the 'centre' of a distribution.

Measures of central tendency

The mode

- The mode is simply the most common score in a data set. It is the most frequently occurring value and thus represented by the highest bar in a histogram.

The median

- There are many definitions of the median.

- Refers to it as 'the score that corresponds to the point at or below which 50% of the scores fall when the data are arranged in numerical order'
- Thus, when we find the median score, half of the scores will be higher and half will be lower.
- A simple rule of thumb formula for finding the median location in a set of scores is (N + 1)/2, where N = the number of scores.
- Therefore, if we have 9 numbers, the location of the median will be (9 + 1)/2 = 5. This means that the median is the fifth number in an ordered series of 9 scores (e.g. 1, 2, 2, 3, 3, 4, 5, 5, 5).

The mean

- By far the most common and popular measure of central tendency is the mean.
- The mean is simply the average score in a data set and is calculated as the sum (Σ) of the scores (X) divided by the number of scores (N):
- To find the mean of the 9 scores listed above: ΣX = 30 30/9 = 3.33

Measures of variability

Range and interquartile range-

- Range statistics report the minimum and maximum score in a data set.
- Using the same 9 scores, the range would be 1–5.
- Clearly, the range is influenced by extreme scores.
- The interquartile range overcomes the problem of extreme values by removing the upper 25% and lower 25% of the distribution.
- The range of the middle 50% of scores is then reported.
- For example, by removing 50% of the data we could be losing important information.
- The decision to trim the sample should be based on a consideration of the influence of extreme scores in the data.

Average deviation and mean absolute deviation

- If we want to assess the degree to which scores vary from the mean, why not calculate the mean and then compute how much each score deviates from this average value
- We could average these deviations to provide a single measure of variability (average deviation)
- The mean absolute deviation (m.a.d.) overcomes this problem by using the absolute values of the deviations.
- The m.a.d. is the sum of the absolute deviations divided by N to provide an average deviation.

The variance (s2)

- The variance is a fundamental concept which is calculated in a number of statistical techniques used in psychological research.
- This statistic uses squared deviations to deal with the problem of positive/negative deviations.
- A negative value, once squared, becomes positive.
- So we calculate the sum of the squared deviations from the mean (or sums of squares, SS).
- The variance cannot be directly interpreted because it is in squared form.

$$s^2 = \frac{\sum(X - \overline{X})^2}{N - 1}$$

The standard deviation (s)

- The standard deviation is simply the square root of the variance. It provides a measure of the average deviation from the mean and is the most commonly reported statistic of variability.
- Once the s2 has been calculated, it can be converted into the SD by simply taking the square root of s2
- If the Standard deviation (s)= 2.16
- This means that on average, scores in this sample deviated from the mean by 2.16.

The calculation of the variance and standard deviation using the definitional approach is based on a four step procedure:

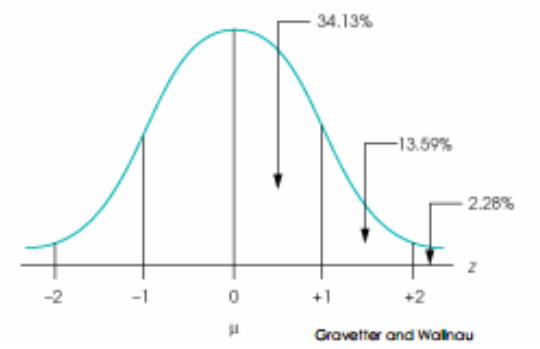
1. The deviations from the mean must be calculated for each score, i.e. the mean is subtracted from each score.
2. Each deviation must be squared, i.e. 2.
3. All of the squared mean deviations must be summed, i.e. ∑2 to obtain the 'sum of squares'.
4. Lastly, the sum of squares must be divided by N–1, i.e. , which provides us with the variance (s2). If we take the square root of this variance we will obtain the standard deviation (s).

The Normal Distribution

- Assumption is that dependent variables are normally distributed in the population
- If we could sample the entire population, scores would fall into a normal distribution
- This assumption allows us to answer research questions
- Mean = 0, SD = 1
- Convert your distribution of scores into a standard normal distribution by converting each score into z- score:

$$z = \frac{X - \overline{X}}{s}$$

- A z-score of +1 shows that a score is 1 standard deviation above the mean
- A z-score of -1 shows that a score is 1 standard deviation below the mean
- A z-score of 0 shows that a score fell right on the mean
- When a behaviour or attribute is normally distributed, then the same proportion of people in the population will always fall within particular ranges on the distribution
- Knowing this, we can use zscores to tell us the probability of someone obtaining a given score

- 

- Z-scores are therefore 'standard scores'
- Very useful in comparing scores from different measurements
- If a distribution of scores is normally distributed then you can use it to answer questions about probability.

Example:

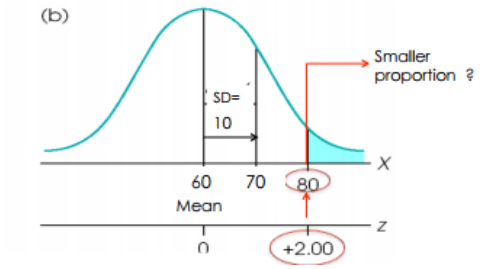
- A Psychology class sits an exam. The scores in the class are normally distributed with the mean being 60 and the SD being 10.
- What is the probability of selecting someone from this class with a score greater than 80?
- Step 1: identify the exact position of 80 in the distribution of scores by computing a z score

$$z = \frac{X - \overline{X}}{s}$$

$$z = \frac{80 - 60}{10}$$

$$z = +2$$

- The score of 80 has a z-score of +2.
- Step 2: We use this information to determine what proportion of scores in the distribution fall above and below the value of interest.



- Consult z-table