

## Improving on quantitative and qualitative analysis

(first and second level analysis e.g. describing and then identifying patterns in the data)

### 1. Statistics Basics (Adapted from Rumsey, 2016)

When working with statistics note the different types of data (fig 1.1)

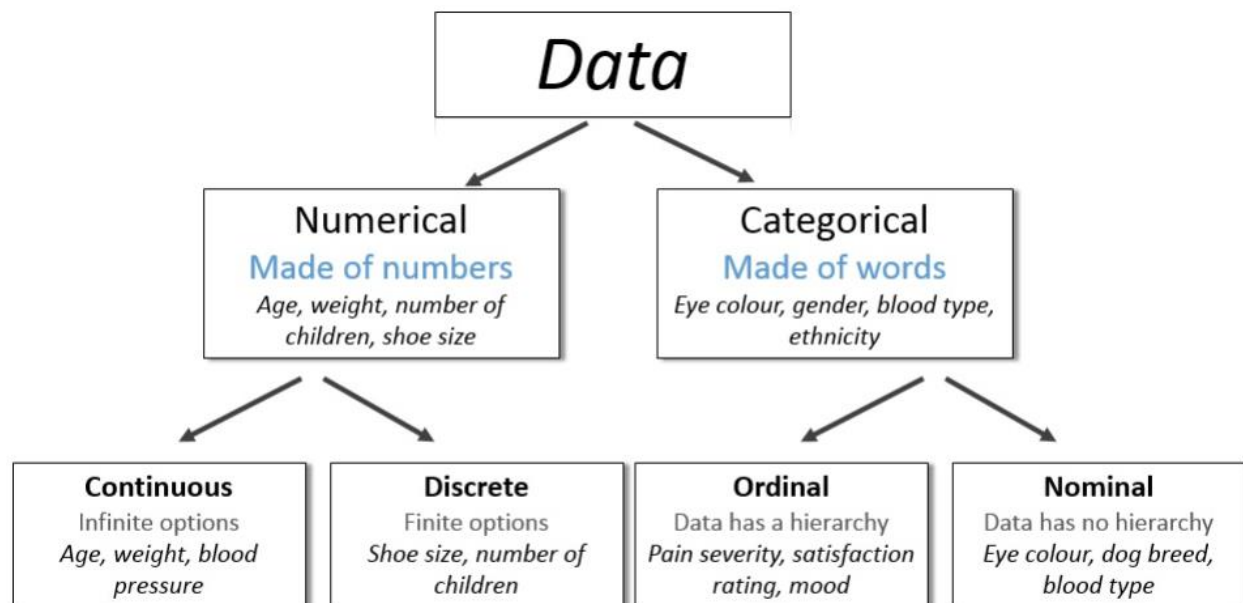


Fig 1.1 Types of data

#### 1.1 Data types explained

**A. numerical** (discrete and continuous) - these data have meaning such as a measurement, such as a person's height

Either **discrete** – countable/finite e.g. dice side, no. of eggs laid

OR **continuous** – not countable/infinite – usually a measurement such as temperature, length, time

**B. categorical** - represent characteristics such as a person's gender, marital status, hometown. Can take on numerical values (such as "1" indicating male and "2" indicating female), but those numbers don't have mathematical meaning. You couldn't add them together, for example. (Other names for categorical data are *qualitative data*, or *Yes/No data*.)

**Can be nominal** –unranked categories e.g. eye colour, gender

**OR ordinal** - mixes numerical and categorical data. The data fall into categories, but the numbers placed on the categories have meaning e.g. rating something on a scale from 0 (lowest) to 4 (highest) stars gives ordinal data. Ordinal data are often treated as categorical, where the groups are ordered when graphs and charts are made. However, unlike categorical data, the numbers do have mathematical meaning. For example, if you survey 100 people and ask them to rate a restaurant on a scale from 0 to 4, taking the average of the 100 responses will have meaning. This would not be the case with categorical data

How to remember:

*Nominal data involves naming or identifying data; because the word "nominal" shares a Latin root with the word "name". Ordinal data involves placing information into an order, and "ordinal" and "order"*

Some basics on different types of graphical data to use:

<http://www.learnhigher.ac.uk/learning-at-university/numeracy-maths-and-statistics/using-graphical-data/>

## 2. Qualitative Research

### 2.1 Doing Qual Research

A good overview of the process is supplied in fig 2.1.

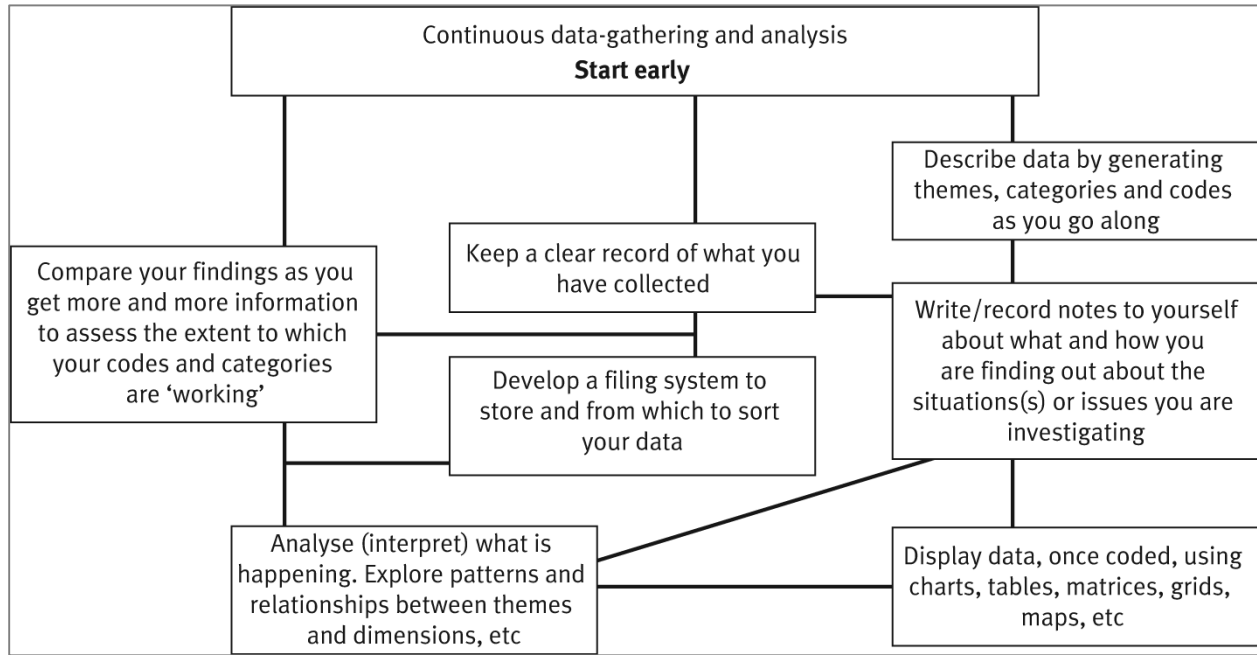


Fig. 2.1 Doing qualitative data analysis (Anderson, 2013, p236)

### 2.2 Categorising and Coding Data

*'Coding is the process of identifying themes, ideas and categories from within your data set so that you can make comparisons and look for regularities and different characteristics of the situation you are investigating'* (Anderson, 2013, p.240). It is an **ITERATIVE process**: a process for arriving at a decision or a desired result by repeating rounds of analysis or a cycle of operations (Business Dictionary, 2018).

'A category is a grouping of behaviors, descriptions or particular events that have shared characteristics' (Anderson, 2013, p.240). The codes you devise may result from a range of different features of your project or derived from your research aims and objectives, or seen as a result of a review of the literature.

Coded analysis involves taking data groups and organising the results into groups. Each one of these groups is then given a "code" that reflects what makes these results similar. From there, a researcher can begin to understand why there may be more consensus in one code versus another. As just one example of an approach, Seidel (1998) adopts three steps to the process.

1. Collecting, or gathering data
2. Noticing, or dividing them into different codes
3. Thinking, or analysing the data based on these codes

Remember codes are '*nothing more than labels you attach to chunks of text that represent a unit of meaning*' (Anderson, 2013, p.241).

### 2.3 Four levels/types of code

- Descriptive/manifest codes
  - First-level codes: identify them for further analysis later
  - Case codes
    - organisation codes
    - sector codes
    - codes for individuals
- Analytical codes
  - Second-level codes
  - Themes, topics, concepts, ideas
- Axial/relationship codes
  - To show relationships or connections within or between categories or sub-categories.

### 2.4 Questions to follow the coding process

Once the data has been coded key questions to consider (Anderson, 2013, p243):

- Are there themes, trends or patterns from groups of people or organisations?
- Are there similarities and differences between the different parts of the data?
- Are there interrelationships between different parts of the data?
- What interesting issues emerge from the data?
- Do the findings suggest that additional data might be usefully be collected?
- Are there deviations from the main patterns or trends that might be identified?

## 2.5 Formulations of conclusions

Formulations of conclusions based on the analysis of qualitative data are shown in table 2.1.

| Factor                 | Example                                                                                                                         |
|------------------------|---------------------------------------------------------------------------------------------------------------------------------|
| Counting               | <i>Less than half of the interviews perceived....</i>                                                                           |
| Patterning             | <i>Recurrent patterns in the analysis of the organisational documents were...</i>                                               |
| Clustering             | <i>Responses from focus groups in location B reported particular difficulties with...whereas those closer to...suggested...</i> |
| Factoring              | <i>Key factors underpinning the perceived effectiveness of the appraisal scheme were...</i>                                     |
| Variables              | <i>The analysis suggests that the practice of...occurs when...but it is unlikely to take place when...</i>                      |
| Casual Networks        | <i>The following model indicates the relationships between the six different factors...</i>                                     |
| Relationship to theory | <i>The incidence of...that the analysis has identified may best be understood within the...theory relating to...</i>            |

**Table 2.1: Factors in drawing conclusions (Robson2011, Neuman 2011, Yin 2009) cited in Anderson (2013 p.256)**

It is important to clarify and justify the analysis processes adopted in your methodology section. The description of the main themes that have been identified and how it has contributed to your understanding of the research question should be housed in the data presentation and analysis chapter.

### 3. Quantitative Analysis

Using quantitative data you can **describe** a current situation in terms of:

- **Frequency:** how many people over the age of 50? How many times were certain behaviours manifested?
- **Central tendency** – what is the average salary of employees?
- **Dispersion:** how wide is the difference between the lowest and highest rates of absence between different departments?

**You can compare this information from other sources & describe trends. It can also be used to explain different phenomena..**

#### 3.1 First level data analysis (descriptive analysis)

Describing the frequencies of different questionnaire responses is helpful in making sense of the data. However it is good practice to also establish some measures by which you compare responses with different variables in a more concise way. How this is approached depends on the type of data you have gathered.

If you have nominal (e.g. how many times?) or ordinal data (e.g. are you male or female, what dept. you work in) – counting the frequencies and describing these as proportions is possible and allows for means of comparison

If you have interval scale data (Likert scales – asking for a rating b/w strongly agree or disagree) it is possible to calculate measures of **central tendency (mean, mode & median)** to assist comparisons between different variables.

Mode – most frequently occurring

Median – middle point range

Mean – average value

Another worthwhile characteristic to explore is **dispersion of the data variables** (also called variability, scatter, or spread). Popular measures of dispersion include the range, interquartile range and standard deviation. The following link provides easy to follow worked examples of how to calculate these:

<https://www2.le.ac.uk/offices/ld/resources/numerical-data/variability>

These may be used to produce normal distribution curve diagrams (if applicable to your data).

### 3.2 Second level analysis – identifying patterns, making inferences through use of statistical tests

Once you have looked at the distribution of your data and perhaps conducted some descriptive statistics to find out the mean, median, or mode, it is timely to make some **inferences** (a conclusion reached on the basis of evidence and reasoning) about the data.

Inferential statistics are the set of statistical tests we use to make inferences about data. These statistical tests allow us to make inferences because they can tell us if the **pattern** we are observing is real or just due to chance.

#### 3.2.1 How do you know what kind of test to use?

One of the first steps in inferential analysis is to answer the question what does the distribution of data look like? The type of test you choose will be guided by the distribution of the data. Distributions fall into two categories, normal and non-normal. You should always check the distribution of your data before beginning inferential analysis.

**Normal Distribution:** This type of distribution looks like the Bell Curve. The graph (fig 3.1) below is an example of what a normal distribution looks like.

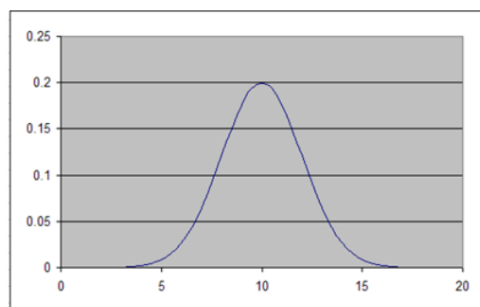


**Fig 3.1** example of normal distribution

If you look at your distribution, you could draw a curve over the distribution that would fit your data. If your distribution looks like the one in the image below (or close to it) your distribution is normal. This type of distribution shows us that the majority of the data is clustered around one number or value. Usually if the data is normal, we choose from statistical tests called parametric tests.

**Normal distributed data tests: parametric**

**Non-normal distributed data tests: non-parametric**

***Warning – point of confusion – some test can be used across the two types depending on the data and their use***

A more detailed easy to use process chart to help your decision follows in fig 3.2:

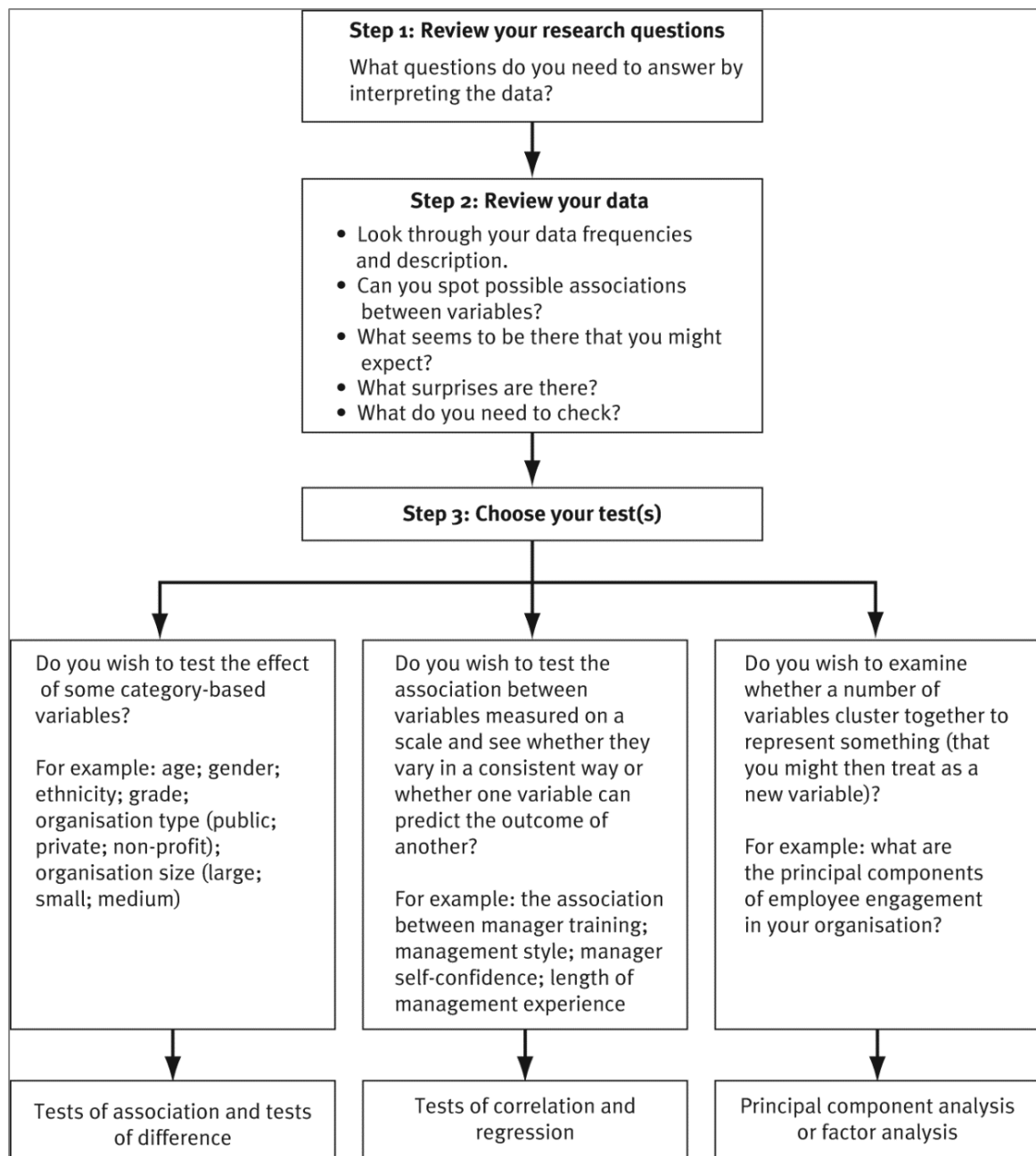


Fig 3.2 Steps and options for quantitative data analysis (Anderson, 2013, p321)

In summary the type of test depends on what you are trying to achieve as well as what type of data or measurement scales you used in your questionnaires AND whether the data generated was normally distributed or not. Therefore if you are clear about your research questions, and the types of data you have collected, you are well positioned to identify the most appropriate analysis tools. Table 3.1 provides an indication of the main options for parametric and non-parametric data.

| TEST & PURPOSE                                                                                                                   | TYPES of DATA                                                                                          | What are you trying to establish?                                                                                                                                                                                                                                                                                                                       |
|----------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Test of association – cross tabulation                                                                                           | All types – useful for nominal (category) data                                                         | How significant is the association between 2 variables?                                                                                                                                                                                                                                                                                                 |
| Test of association – scattergram (plotting a graph where each variable represented on each axis)                                | Interval or ratio scale data only                                                                      | How strong is the association?<br>How significant is the association?                                                                                                                                                                                                                                                                                   |
| Assessment of significance of association (chi-square) also known as $\chi^2$ test or <i>Pearson's</i> chi-squared test.         | All types                                                                                              | Used to determine whether there is a significant difference between the expected frequencies and the observed frequencies in one or more categories.<br><br>Does the number of individuals or objects that fall in each category differ significantly from the number you would expect?                                                                 |
| <u>Test of difference</u><br><br>t-test<br><br><br>or Mann-Whitney U test                                                        | Never for nominal data                                                                                 | t-test: <i>Is the observed difference larger than what you would expect from random fluctuations alone?</i><br><br>u-test: used to compare two sample means that come from the same population, and used to test whether two sample means are equal or not                                                                                              |
| <u>Test of correlation</u><br><br>Spearman's rho (for non-parametric data)<br><br>or Pearson's correlation (for parametric data) | Never for nominal data<br><br>(both calculate a correlation coefficient that ranges between -1 and =1) | Measure the strength and direction of association between different variables<br><br>Can be distorted by outliers & the possibility of incorrectly being used to infer a causal relationship between the variable                                                                                                                                       |
| Factor analysis                                                                                                                  | Never for nominal data                                                                                 | To assess whether there are key variables in the data or whether some variables can be grouped or clustered together to form a coherent composite variable<br><br>What is the validity of questionnaire items (taken together)? Do the questions in your questionnaire actually relate in a coherent way to the concept that you are trying to measure? |

Table 3.1 Analysis Options and Choices

### 3.3 Online help: Which Statistics Test Should I Use?

Try out: [https://www.socscistatistics.com/tests/what\\_stats\\_test\\_wizard.aspx](https://www.socscistatistics.com/tests/what_stats_test_wizard.aspx)

This wizard will ask you a few questions, and then based on your answers, will recommend a statistics test.

Takes you through the first level analysis e.g. measure of spread

<https://www.skillsyouneed.com/num/simple-statistical-analysis.html>

Takes you through second level analysis e.g. looking for patterns (making use of tests)

<https://www.skillsyouneed.com/num/statistics-identifying-patterns.html>