**9 Quantitative Data Analysis:**

Population - group of people of interest to you

Sample - a number of the population that you have collected data (sample is representative of the population with a large enough sample to draw conclusions from - response rates as high as possible)

Representative - validity of the sample towards the accuracy of the results

Descriptive statistics: describe findings from the sample but don’t attempt to make inferences from the data - univariate (1 variable) - states a fact

Bivariate analysis - 2 variables and tests their correlation with each other, tests to see if variation of one variable coincides with the other - context statement

Inferential statistics” describing the larger population from what you collected from a smaller sample of people - uses statistical tests and significance tests to identify trends in the larger population

Associations - bar graphs for comparisons

Strength and direction between two variables

r= correlation coefficient, -1,0,1

Use scatter plot = if one goes up so does the other

Correlation coefficient = assigning a numerical value to the association shown in the scatter plot

Nearer to zero the coefficient, the less correlation between the variables

Draw a line of best fit - steepness of line shows strength of correlation, direction indicates negative or positive correlation

Just because lines go the same way doesn’t mean that one caused the other

Coefficients say nothing about which variables causes the other to change

Prove causality - statistical association, temporal precedence (cause than effect), no alternate explanations

Inferential: Significance levels: how likely it is a result is observed through chance, common level 95%

Attempt to prove or disprove the null hypothesis and only accept the hypothesis if we can be 95% sure that there is a genuine association

Null hypothesis - no association between two variables

**How to analyze Qualitative Data:**

Read your data again and again - comes down to reading it in close detail and familiarizing yourself with the data

Like writing a book report, summarize the plot, describe the characters and their actions/motivations, themes

1st read - basic details, later readings - subtle details

Requires depth of the data

Unstructured data, random nobody is trying to create patterns nor follow a particular structure

The transcripts result in 100s of words that require time to interrogate the data

Use the data to answer your research questions

Read, compare, interpret,

1. Read - what's going on in the data, do you understand the data, no read again in a different way

Write up

Content analysis -what people are literally saying

Thematic analysis - themes of topics

Discourse analysis - how people say things, what words or language they use

Not just looking for what people say but for what they don’t say, what aren’t they not talking about, subtleties, subtext

Structured using quotes

Coding - interesting sections, line by line

2. Comparison - what’s different about the data, people, periods of time, categories, events, same or difference, why

3. Interpretation - what does the data mean, look at all the previous steps, identify the story in the data

Involves uncovering patterns in data and involves uncovering meaning and multiple perspectives

Tests or formulates more hypothesis, does the data support our hypothesis

Qualitative analysis software data - can keep track of your codes and themes but doesn’t do any analysis

Excel, word, paper to write on transcripts

**Beginners Guide to Coding Qualitative Data:**

Coding - used to analyze data by identifying themes or codes that appear in our qualitative data and then assigning intersections of data to those codes

They symbol labels of units of meaning

Purpose is to break up very long and detailed qualitative data like interviews, focus groups, and documents into common themes

Read across different sources to compare what people are saying on one particular issue or topic

It’s like putting things into categories - its organizational and easy to find things later and used to see variation

Easily referenceable through categorization

Codes: milk, porridge, children, family structure, social breakdown, time

What is put into the codes are quotes or certain sentences that relate to those categories

Takes time

Pithy labels identifying what is of interest in the data

Themes are a recurring pattern across the data set

Different types of coding:

Grounded theory (emergent coding) - start with a blank sheet, no preconceptions, open to having the data speak to you, define codes and themes that is suggested by the data

Framework analysis (structured coding) - framework with list of codes and themes before hand, match data with the codes already identified

Mixed Method of coding - flexible combination, they have some kind of idea of the things they want to find in the data, but are open to new things, something surprising from the data they’ll create new codes or themes in response

Descriptive coding: coding what is being literally being discussed

Thematic coding - create themes and codes

Line by line coding - assign each line a code to summarize

Interpretative phenomenological analysis - see how participants experience and make meaning of things that are happening in their daily lives

In-vivo coding - use participants own words to summarize/code words

Discourse analysis - how people express themselves

Coding is nearly always an iterative cyclical process, go back and do it again and again

Reading it through multiple times, different levels, different types of coding, coding your coding

Putting your coding into different groups and different structures

Requires further interpretation when complete - need to address what the data says

Coding is not required - takes very long, very complex, low level base level analysis

Just make it make sense in your head

**Statistics Made Easy:**

When looking at sample data we see two things: differences between groups and relationships between variables

These two things lead into the question whether or not those associations/relationships are real

By making a table that visually categorizes various combinations of variables we can establish very specific differences between groups and specific relationships between variables

Statistical tests - proportions, chi-squared, t-test, ANOVA, correlation - used to interpret results

Take a random sample of the population and measure the intended variables, collect additional information (gender and age group in sample) to arrange data in a spreadsheet with the various attributes in columns known as variables which are the object of your inquiry

Variables are either:

Categorical or numeric

Categorical = groups, categories

Summary to understand data - represent data by counting number of observations in table or bar graph

Numerical = numbers, number lines

Summary to understand data - spread of distribution of data, range, standard deviation, median, mean, visualize using box plot visual representation of range, histogram gives shape of the data

Combination of categorical and numerical

If what we are seeing in the sample data has implications for what we think about the wider population, are what we seeing statistically important

One categorical = bar graph, 1 sample proportion test

Two categorical = multiple bar graphs, chi squared

One numeric = box plot, t-test

One numeric and one categorical = box plot, t-test or ANOVA

Two numeric = scatterplot, correlation test

Before interrogating data, define question, define alpha value/null hypothesis, and then analyze data

There is always a possibility that the sample shows a no, the null hypothesis is wrong and needs rejecting

Alpha Value = at what value do we decide it is too small and must reject the hypothesis

If p is less than the alpha value we can reject the null hypothesis

Null hypothesis = no change

Correlation coefficient = -1 - 1