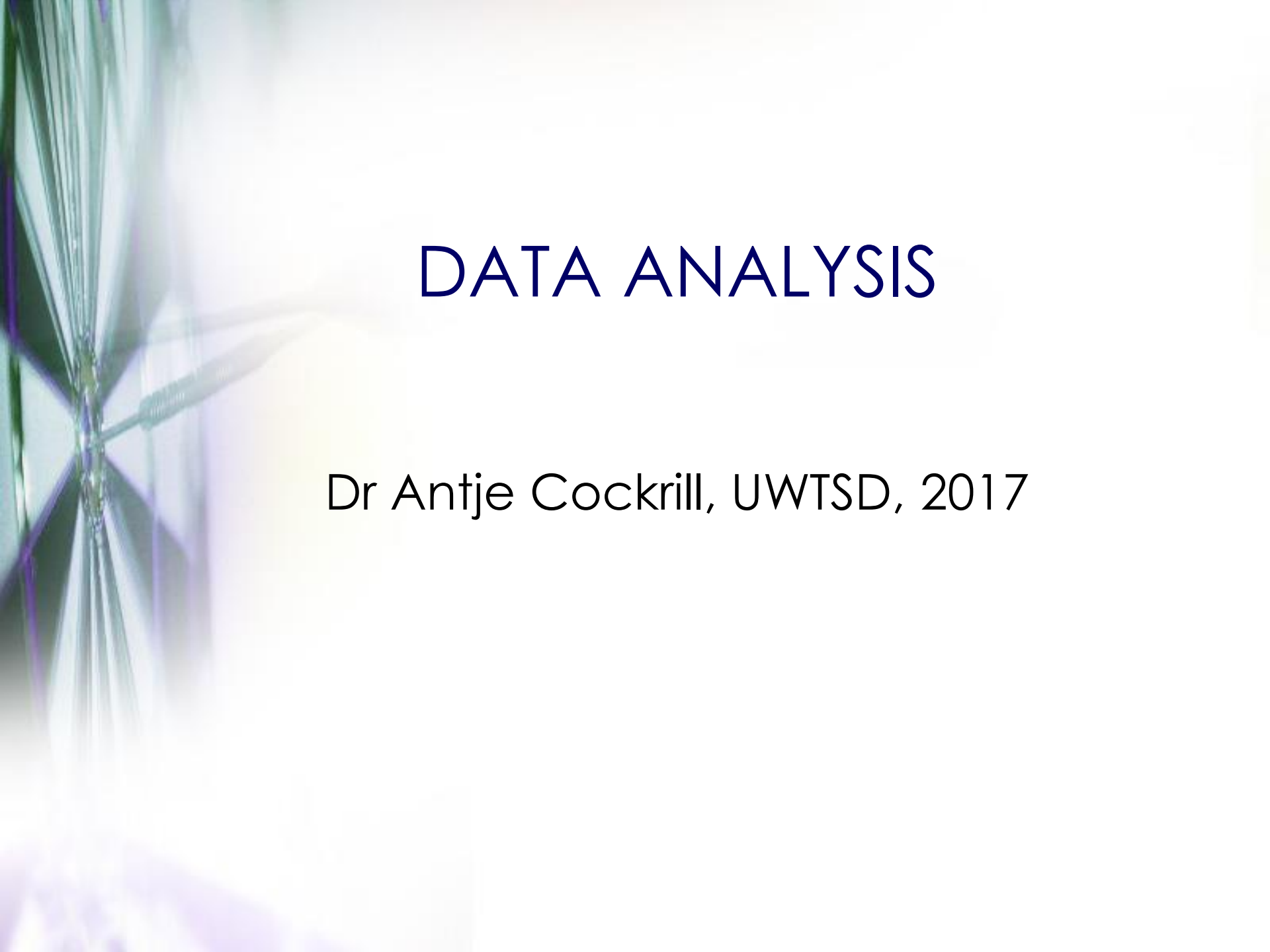




DATA ANALYSIS

Dr Antje Cockrill, UWTSD, 2017

At the end of this session you should be able to

- Know how to approach qualitative analysis
- Know where to start basic statistical analysis
- Interpret a p value
- Understand reliability and validity

Qualitative Analysis

- Generation of data which aim to explore, describe and contextualise experience (Henwood and Pidgeon, 2013)
- Good at answering 'How?' and 'What?' questions
- Based on methods which involve appreciation of complexity and detail (Mason, 2006)



Stage 1. Make Your Data 'Analysis Friendly'

- Turn data into text using word processing package e.g. transcripts, field notes/memos/research diary, documents
- Ensure you can trace where excerpts of text are from (e.g. carefully labelled with identifier for participant/focus group, dated, backed up)

Stage 2. Decide On Tools To Help You With Analysis

- By using Word (or similar) and sorting/cut and paste
- By using specialist software like NVIVO (the university has a license)



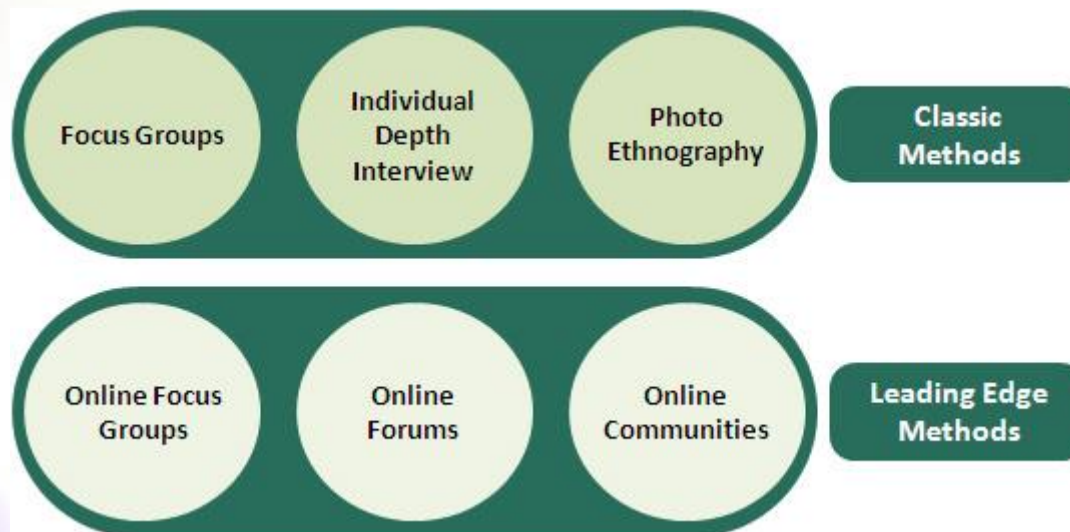
Stage 3. Deal With Demographic Data

- Most qualitative studies include some quantitative data (eg age, gender)
- Think about how best to display this data (eg table, bar chart, pie chart)
- Interval/ratio data (eg age):
mean/median, range, standard deviation
- Nominal data (eg gender, grade,):
mode, frequencies

Stage 4. Carry Out Your Qualitative Analysis

Content analysis

- Thematic analysis
- 'Methodology specific analyses' eg discourse analysis, grounded theory, narrative analysis



Using Matrix Analysis (based on questions)

OBJECTIVE 1	QUESTION 1	QUESTION 2	QUESTION 3	QUESTION 4
Interview 1	Comments & key points made	Comments & key points made	Comments & key points made	Comments & key points made
Interview 2 And so on	Comments & key points made	Comments & key points made	Comments & key points made	Comments & key points made
	Summary of critical and most common points made	Summary of critical and most common points made	Summary of critical and most common points made	Summary of critical and most common points made

the data

Using Matrix Analysis (based on themes)

Qualitative data

OBJECTIVE 1	Theme 1	Theme 2	Theme 3	Theme 4
Interview 1	<i>Comments & key points made</i>	<i>Comments & key points made</i>	<i>Comments & key points made</i>	<i>Comments & key points made</i>
Interview 2 And so on	<i>Comments & key points made</i>	<i>Comments & key points made</i>	<i>Comments & key points made</i>	<i>Comments & key points made</i>
	Summary of critical and most common points made	Summary of critical and most common points made	Summary of critical and most common points made	Summary of critical and most common points made

Common Difficulties/ Problems ...

- Can't find in books 'How to do it'
- Underestimate how long it takes to turn tape-recorded data into text
- Time to analyse properly
- Completed analysis
 - i) seems like common sense
 - ii) feels like it is invented



... And Some Solutions!!

- Clarify what you understand by your qualitative research approach (using references), and state clearly what you did
- Allow plenty of time for all parts of research process (eg 1:8 for transcribing of interviews)
- Don't spend all your time transcribing to detriment of analysis. Recognise this requires focused time, and work through difficulties!
- Think about how you can show that you have done your research rigorously and carefully



Quantitative Analysis



Coding terms

- Before you can analyse numerical data, it needs to be coded.
- Code: numbers or letters assigned to an observation of a variable
- In quantitative analysis, you try to turn as much information into data as possible
- This means you have to develop a coding frame



Example of a coding Frame

Coding frame

✖ Q1 Male = 1, Female = 2

✖ Q2 Age in years

✖ Q3 Temporary = 1, Operator = 2,
Office = 3, Management = 4, Other
= 5

✖ Q4 Salary in £

A coding frame contains your coding information.

1. Variable names
2. Values for each response option
3. Any recoding information

Considerations in Choosing Statistical Analysis

- Level of measurement of data collected (nominal, ordinal, interval, ratio)
- Goals of data analysis:
 - Descriptive statistics: procedures that assist in organizing, summarizing and interpreting data
 - Inferential statistics: Procedures that allow one to make generalizations from a sample to the larger population
- Audience – For whom is the analysis intended?

Measures of Variability

Range: "Difference between the highest and lowest score" $(28-15) = 13$

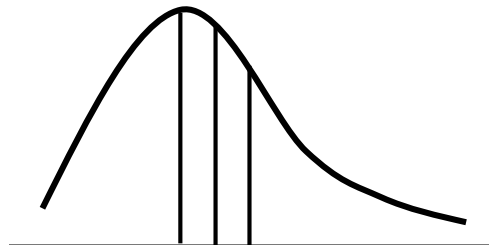
Standard Deviation: "Average distance of the scores from the mean" $SD=4$

**Symmetrical
Distribution**



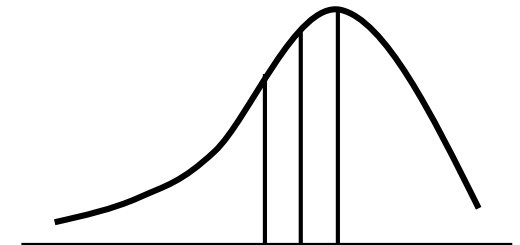
(mean, median, mode)

Positive Skew



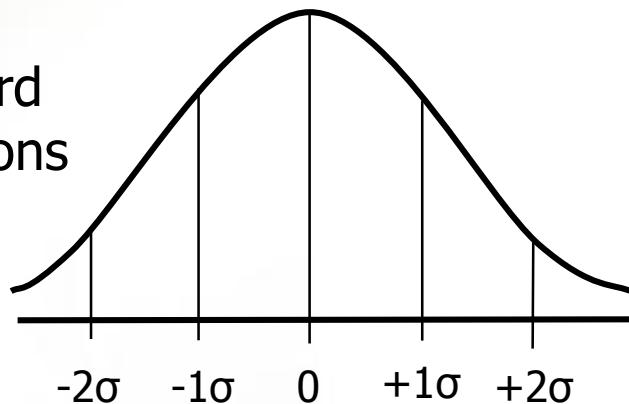
(mode, median, mean)

Negative Skew



(mean, median, mode)

**Standard
deviations**



$$\sigma = \sqrt{\frac{\sum (x - M)^2}{N}}$$

x = individual scores

M = mean

N = number of scores in group

Types of Statistical Tests: Brief e.g.s

When running a multivariate tests comparing the means between groups and/or variables

- We assume
 - random sampling
 - the groups are homogeneous
 - distribution is normal
 - samples are large enough to represent population (>30)
 - DV Data: represented on an **interval or ratio** scale

These are Parametric tests!

Types of Tests

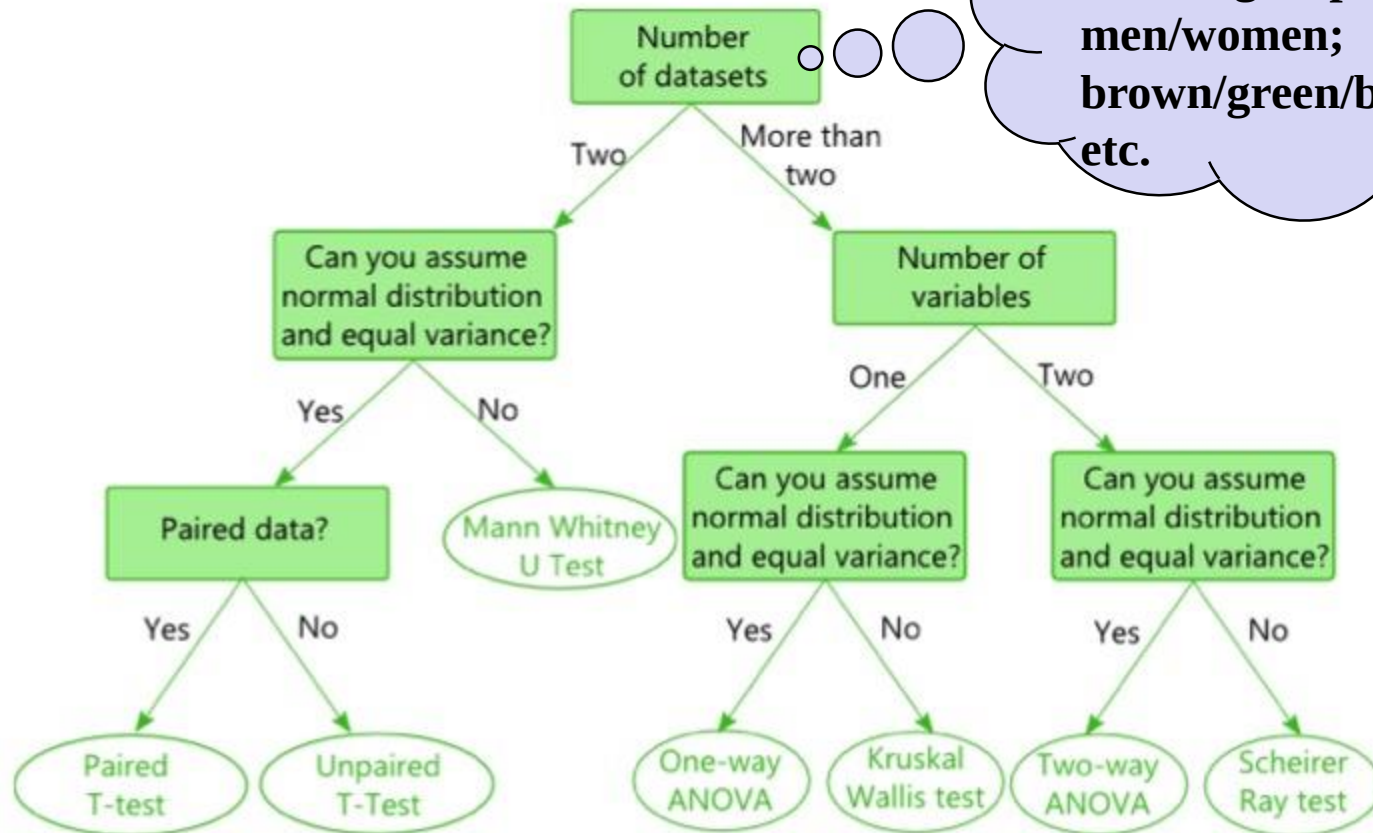
When the assumptions are violated:

- Subjects were not randomly sampled
- If variables are:
 - Ordinal (ranked)
 - Nominal (categorized: brands of organic foods, types of rock, levels of education, learning styles)
 - The scores are greatly skewed or we have no knowledge of the distribution

Non-Parametric Test!

If in doubt – use non parametric tests

Non parametric and parametric tests



‘Dataset’ refers here to groups e.g. men/women; brown/green/black etc.

<http://bitesizebio.com/wp-content/uploads/2014/01/Slide15.jpg>

Chi Square

Used when assumptions are violated e.g. data are nominal (both IV and DV), or you simply don't know

- Comparing frequencies of distributions occurring in different categories or groups
- Tests whether group distributions are different
 - *Customer preference of 3 brands of nuts*
- determines the association between two variables by counting the frequencies of distribution
 - *Gender relative to study preference (alone or in group)*

p-value interpretation

- The interpretation of a p-value is a little tricky
- We can usually say that a value is statistically significant if it is below 0.05 and it is highly significant if it is below 0.001
- The p-value represents the chance that we would see a difference as big as we saw (or bigger) if there were really nothing happening other than chance variability. A .05 p value means that we would expect to get the same results 95 times out of a hundred.



Correlation

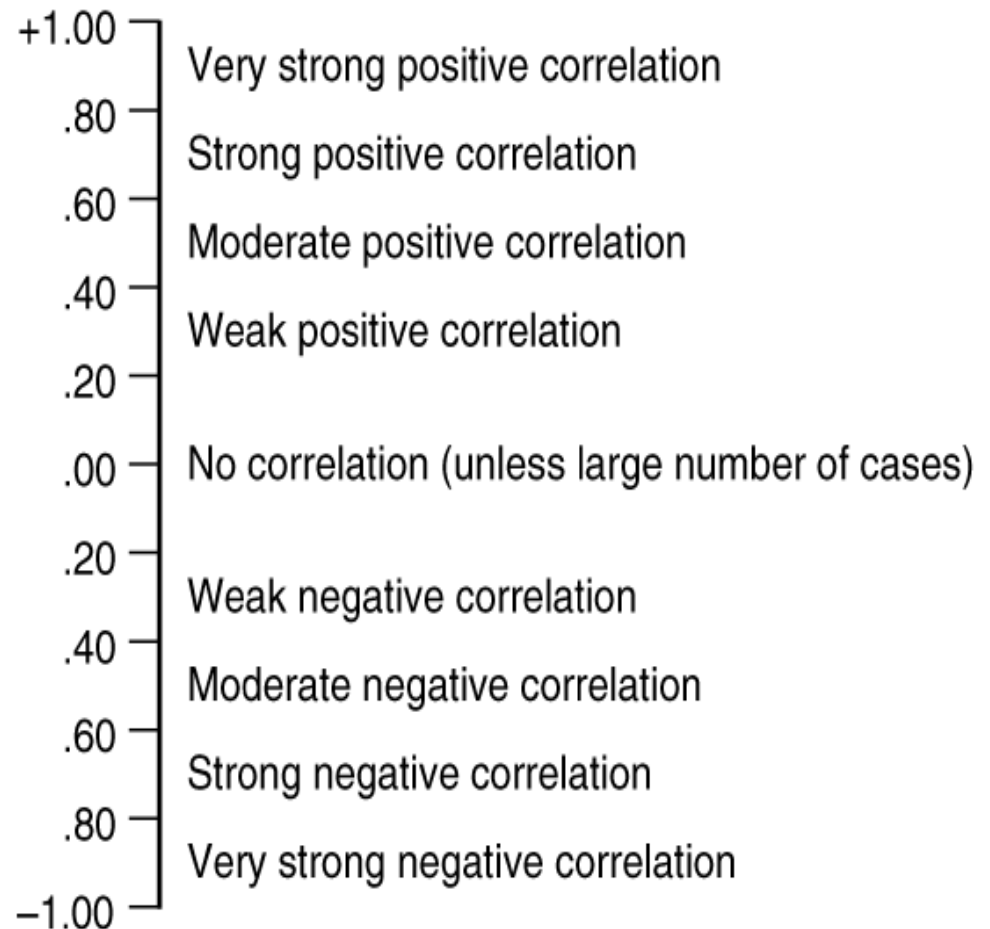
Correlation: Measure of **relationship** between two or more quantitative variables

Correlation Coefficient:

A number between -1 and $+1$ Which indicates the direction and strength of the relationship.

Pearson Product-
Moment
Correlation

$$r = .45$$



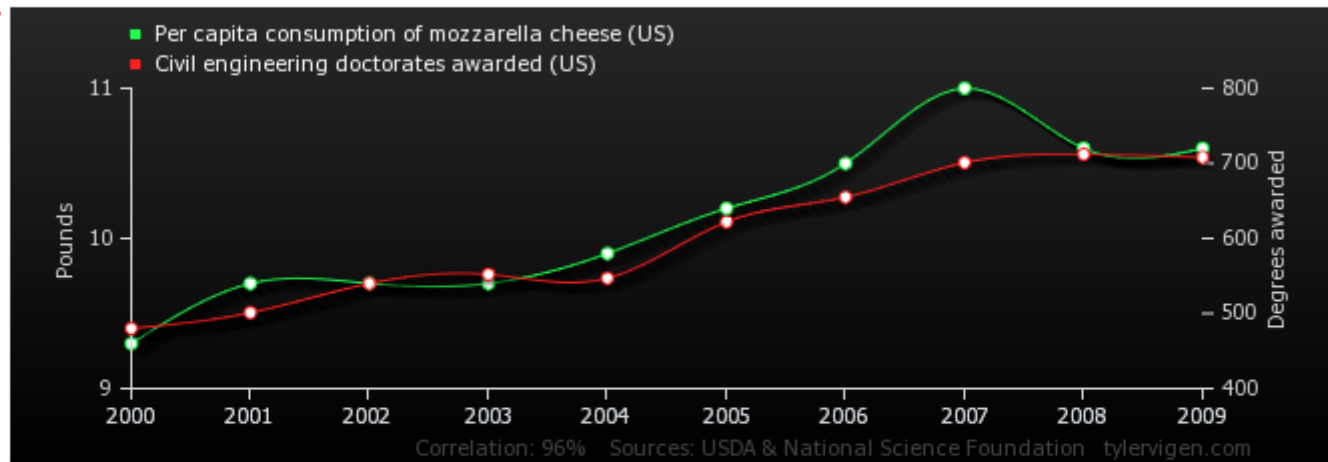
Hypothesis: Eating mozzarella cheese improves the likelihood of getting a civil engineering doctorate

Reliability?
Validity?

Per capita consumption of mozzarella cheese (US)

correlates with

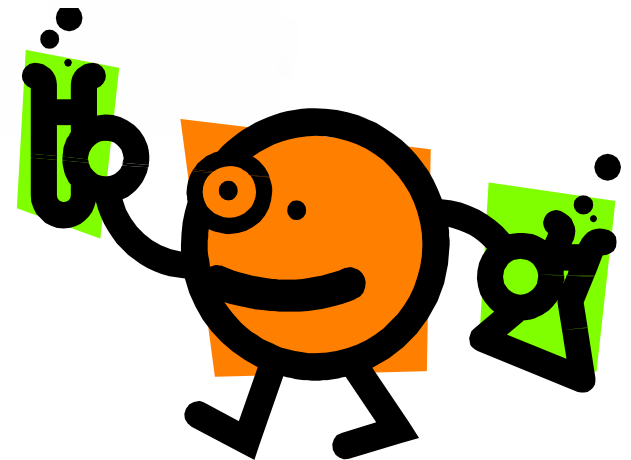
Civil engineering doctorates awarded (US)



	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Per capita consumption of mozzarella cheese (US) Pounds (USDA)	9.3	9.7	9.7	9.7	9.9	10.2	10.5	11	10.6	10.6
Civil engineering doctorates awarded (US) Degrees awarded (National Science Foundation)	480	501	540	552	547	622	655	701	712	708
Correlation: 0.958648										

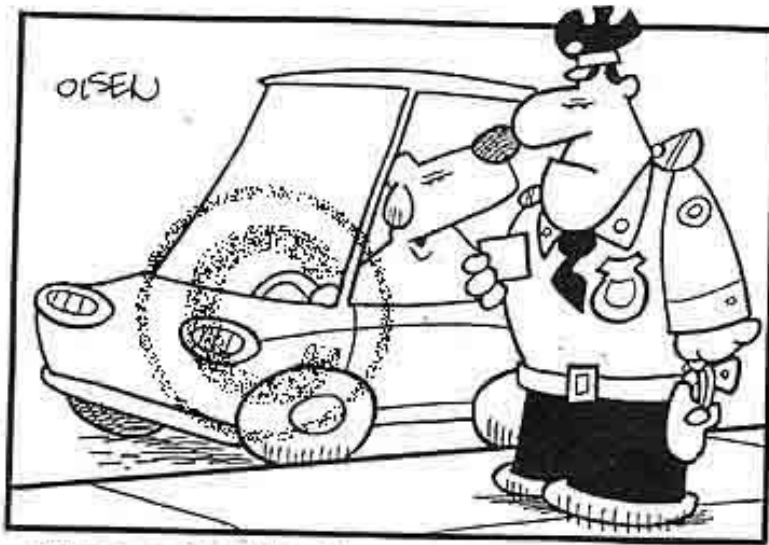
Reliability

- Results should be consistent
- Any significant results must be more than a one-off finding and be inherently repeatable.
- The results of an experiment should be the same when re-tested in the same conditions.
- If they are not consistent it indicates random error



Validity

The measuring device must measure what it's supposed to measure.



"Of course it's valid, officer...don't forget, you have to multiply dogs years by seven !"

Internal validity

The extent to which the observed relationship reflects the actual relationship between the variables in a study.

External validity

The extent to which the results generalise to other participants and other situations

Validity in data collection means that your findings truly represent the phenomenon you are claiming to measure

http://linguistics.byu.edu/faculty/henrichsen/ResearchMethods/RM_2_18.html

Research should be:



Reliable
Not Valid



Low Validity
Low Reliability

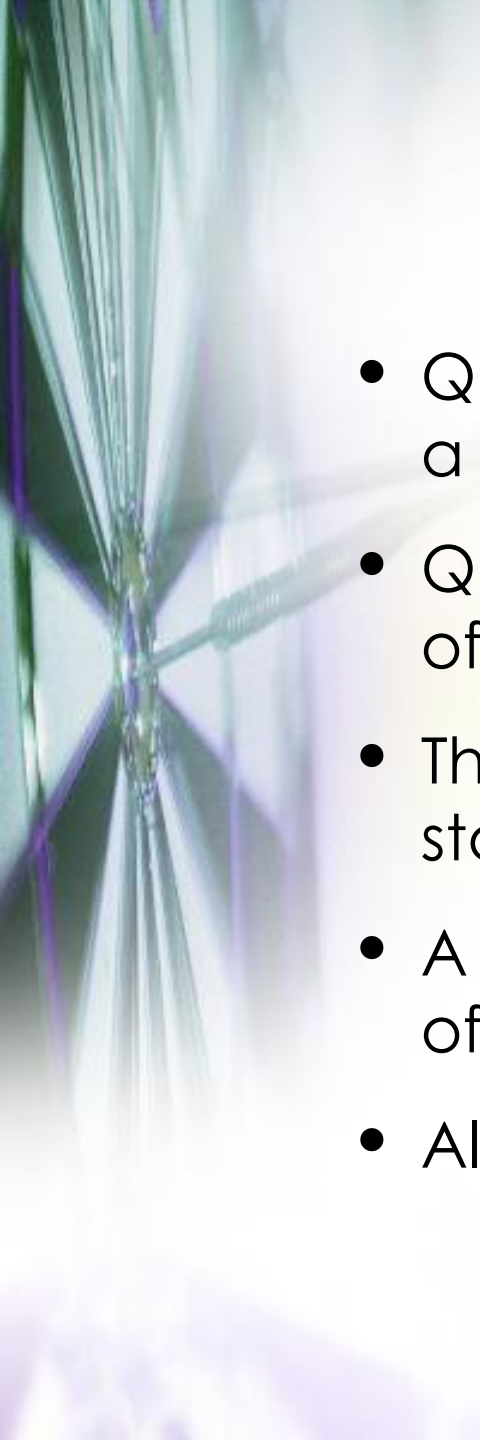


Not Reliable
Not Valid



Both Reliable
and Valid

by Experiment-Resources.com



Summary

- Qualitative research can be analysed using a thematic or a question based approach.
- Quantitative analysis depends on the type of data collected
- There are parametric and non-parametric statistical tests
- A p-value reflects the statistical significance of a result
- All research has to be reliable and valid