Collecting data – for evaluations or experiments

Five P's:

Performance – response time, errors

Preference – preference in a comparative task

Perception – free-form opinions 1. quality (good / bad) 2. ideas for change of the design

Process – actions performed in doing a task

Product – artefact created by the participant

Data Collection Methods

- Questionnaire
- Interview
- Answers on a task sheet
- Observational notes
- Automatic collection
  - answers
  - time taken
  - action log (internal system logs, video/audio of users, screen recordings)
- Artefact storage

data	questionnaire	interview	task sheet	observation	automatic collection	artefact storage
Performance			x	x	x	x
Preference	x	x		x		
Perception	x	x		x		
Process				x	x	x
Product				x		x

record	questionnaire	interview	task sheet	observation	automatic collection	artefact storage
evaluator	design	x		x	x	
user	x		x			x

questionnaire	interview	task sheet	observation	automatic collection	artefact storage
easy to administer	not always easy to conduct	easy to administer	obvious thing to do	easy to collect	(maybe) easy to collect
difficult to design	difficult to design	based on task	difficult to gather systematic data	tempting to collect more than is needed	being unambiguous is crucial
typically quick	takes time	takes as long as the task itself	takes as long as the task itself	takes as long as the task itself	takes as long as the task itself
difficult to get detailed qualitative answers	can elicit detailed qualitative answers	quantitative responses to direct task-based questions	can be qualitative and/or quantitative	quantitative only; time consuming to analyse	qualitative product data
easy to get large numbers of participants	difficult to get participants	availability of participants will depend recruitment to do the task	availability of participants will depend recruitment to do the task	availability of participants will depend recruitment to do the task	availability of participants will depend recruitment to do the task

Terminology

subjective/ objective

quantitative/ qualitative

preference/ performance

open/ closed questions

formative/ summative

eg. objective qualitative data

evaluation/ experiment

left not always match with right

'Evaluation' *vs* 'Experiment'

- qualitative *vs* quantitative
- formative *vs* summative

eg. qualitative formative data

- small numbers *vs* large numbers
- visualisation tool *vs* visualisation method/idea

evaluation experiment

- large task *vs* small task
- lengthy *vs* brief

exp eva (eg.)

Experiments

- **Hypothesis-driven**, or driven by a **Research Question**
- In IV, seeking verification that a proposed tool/ method is 'good'

- The experiment is defined by
  - the independent variable (what is being tested)
  - conditions (different values of the independent variable)
  - control variables (stay the same)
  - random variables (that you don't care about)
  - dependent variables (the data you collect)
  - tasks (what the participants are asked to do)

Experimental factors are things that can change during an experiment, e.g.

- the data used in a visualisation
- number of menu options available
- the colour palette used
- the tasks being performed by the participants
- the participants themselves
- temperature of the room
- the computer used

Some of these you want to change, some you want to keep the same, some you have no control over, some you don't care about

Variables (or 'factors')

Independent variables

Factors you want to change, because they are related to your RQ

Defined by a set of 'conditions'

Control variables (not really 'variables')

Factors you need to make sure do not change

Random variables

Factors whose values you allow to change randomly

Confounding factors

Factors that changed during the experiment (although you wish they hadn't)

Dependent variables

The data that you collect

We would like to show that:

*a change in the values of the **independent variable** will result in a change in the values of the **dependent variable***

And we want to be sure that:

*any change in the value of the **dependent variable** is a result of the change in the value of the **independent variable** (rather than a result of change in the value of any other factors)*

Confounding factors

- Factors that changed during the experiment…
- … but only discovered after completion of the experiment

- These variables may have affected the value of the dependent variable…
- … and may have done so in way that means we cannot be sure that it was change in the value of the independent variable that affected the value of the dependent variable…
- … since it may have been the change in the confounding factor instead

Control variables

- Kept constant throughout the experiment
- If constant, these variables cannot affect the value of the dependent variable…
- … thus ensuring that we can be sure that it is the value of the independent variable that affects the value of the dependent variable

Qualitative Data Analysis

Analysing data from, e.g.:

Interviews

Focus groups

Open-ended question responses in a questionnaire

Transcribe any audio/video data

Familiarise yourself with *all* the data

Inductive coding

Find the themes in the data, in a structured and rigorous way

Quantitative data analysis overview

Task1	4	6	7	8	4	3		5.33	53.33%
Task2	56	23	98	3	10	67	55	44.57	44.57%
Task3	7	8		8	8	8	7	7.67	76.67%
Task4	10	12	15	14	13	16	20	14.29	71.43%
Task5	115	128	109	98	133	87	101	110.14	73.43%

63.89%

Qualitative data analysis overview

Task1								summary for T1
Task2								summary for T2
Task3								summary for T3
Task4								summary for T4
Task5								summary for T5

overall summary

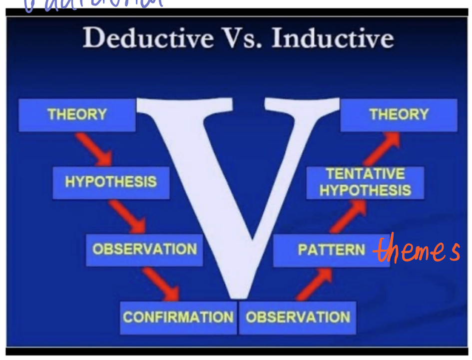
Qualitative Data Analysis: Inductive Approaches

Thematic analysis

One form: grounded theory
- No preconceived theories that bias
- Open mind to what may be in data
- 'Theory' is what eventually 'emerges' from the process

Verify coding with multiple coders at different stages:
1. Open coding
2. Axial coding
3. Selective coding

traditional



Stage 1: Open coding

Identify distinct pieces of information
- Assign an open code
- In-vivo coding: use participants' own words to define codes

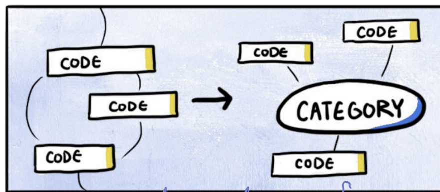
Size/scope of pieces are determined by the coder's interpretive process

Participant: I think teens use drugs as a release from their parents [rebellious act]. Well, I don't know. I can only talk for myself. For me, it was an experience [experience] [in vivo code]. You hear a lot about drugs [drug talk]. You hear they are bad for you ["negative connotation" to the drug talk]. There is a lot of them around [available supply]. You just get into them because they're accessible [easy access] and because it's kind of a new thing [novel experience]. It's cool! You know, it's something that is bad for you, taboo, a "no" [negative connection]. Everyone is against it [adult negative stance]. If you are a teenager, the first thing you are going to do is try them [challenge the adult negative stance].

Stage 2: Axial coding

Organise open codes into sets of concepts/categories

Think about the relationships between concepts



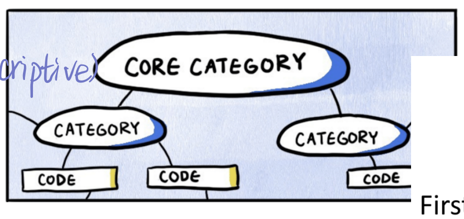
Concepts and categories don't have to all be at the same level of specificity, or need the same number of codes assigned to them
- Be driven by what is in the data, not what you think of as 'even' or ideas of symmetry/balance…

more general and specific when grouping together.
Something do happen more often than others, so we'll get more codes around that kind of topic / category

Stage 3: Selective coding

Combining concepts and categories, to form larger concept framework (more descriptive)

Re-code original transcripts using this new framework



Reporting Results from Qualitative Data Analysis

First, set out the coding framework

Then if, for example, a thematic analysis uncovers 5 main themes
- 5 main sections in your report, each explaining one theme
- Use participant quotes as evidence, perhaps also discussion of how often or in what contexts instances of the theme happened

Then, discuss overall or higher-level findings
- Put this in the context of related work, e.g. reinforce/constrast other studies, consider different factors that might influence the data and the participants
- If appropriate, discuss implications for future system design requirements

Number of participants

Formative evaluations:

Jakob Nielsen (2000), "*Why You Only Need to Test with 5 Users*"

"Elaborate usability tests are a waste of resources. The best results come from testing no more than 5 users and running as many small tests as you can afford."

Summative evaluations:

Probably more than this, e.g. 10, to be sure of the validity and generalisability of the 'proof of worth' outcome

Experiments:

For statistical analysis, even more; more than 30 is an oft-quoted minimum

Key principles of evaluations & experiments

Participant time is your most precious resource

Pilot: try out your evaluation/experimental methodology in advance to make sure it will produce the data needed

- phrasing of tasks never let participant confused
- adequacy of training or tutoring, before starting tasks (a problem of evaluation/experiment
- expected duration of experiment session (too long/short?) not of system)

Don't start until you are sure that the data you will collect will be useful to you; create 'dummy' data to see if it makes any sense

Participants should never be confused as to what to do

Get as much as you can out of the participant in each session

A Word of Warning

Quantitative analysis is NOT better than qualitative analysis

Too often people are biased towards 'clean' quant models, but they forget the subjectivity/variability/errors of parameters, metrics, models, etc.

It's common in qualitative analysis to get multiple people to code the same data independently, and then compare. It's rare to get multiple people to do quant analysis on the same data, independently, and then compare.

Each can be done well or badly, but neither is better than the other

Rigour of good measurement and modelling is similar in level to the rigour of good coding, concept formation, and theory building

quantitative modeling can just be as rigorous as qualitative modeling

Ideally, do both: Use quant to learn *what* people do, and qual to learn *why*

more deeper . what is good