# Techniques of analyzing qualitative data

Mainack Mondal

CS 60081
Autumn 2022

# Before we start… Study design

- RQ: is login interface 1 cause less number of password-enter errors than login interface 2?
- Between subjects
  - Each participant tests 1 version (either interface 1 OR interface 2)
  - You compare results from these groups
  - Groups should be mostly similar by demographic (need to verify)
- Within subjects
  - Every participant tests everything (both interface 1 and 2)
  - Think of a medicine trial – before and after
  - Very important to randomize order of testing

# Before we start… Study design

- RQ: Do old people make more password entering errors ?
- Factor 1: 5 age groups
- Factor 2: 3 type of services
- Total number of groups: 15
  - Divide participant in each of the 15 groups
  - Can discount a few groups
  - Full factorial design

# Qualitative Data Analysis

# Roadmap

- Qualitative Data Analysis
  - Selecting participants
  - Data analysis techniques
- Inter-rater agreement
- Quantitative data analysis

# Selecting participants

- How to select representative sample?
- External validity (Generalizability)
  - The extent to which information learned in a study can be generalized to the world at large
  - E.g.,
    - You created a end to end encrypted messaging tool and evaluate usability with members of this class
    - How generalizable are the results?
    - Actual users are more representative and hard to get
    - College students are less representative and easy to get
    - Convenience sample (good to start with and pilot)

# Control vs. external validity

- You need to control in your study
  - Only pick people who ever used banking in a password study
  - AND who never have seen the login screen you present
- More control → the effect you observe is *ONLY* because of your study
- More control → less generalizability
- Challenge: Balancing these two via sample selection

# Selecting participants

- Find people who represent your target population
  - Similar skills and abilities
  - Group you are most concerned about
  - Have similar limitations (e.g., refugees)
- Recruitment can be hard
  - Highly skilled
  - Vulnerable (children, students, people with mental disorder)
  - Rare ( CTOs, CEOs, actual police)

# Limitations

- Do include a limitation section in your report
  - How does the bias in your population might affect your result
  - One example: “One sample of very active banking users are doing more error in interface 1. So the generic less active population will do even worse”

# Roadmap

- Qualitative Data Analysis
  - Selecting participants
  - Data analysis techniques

- Inter-rater agreement
- Quantitative data analysis

# Type of qualitative data

- Text
  - Transcription of interviews and focus groups
  - Notes and memos
- Audio
  - Recording (need to “Transcribe”)
- Visual
  - Video
  - Photograph

# Qualitative data management: Ethics

- How do you store the data?
  - Need to do it securely (e.g., password protected computer )
  - Mention it in your IRB application
- Removal of names and identifying info
  - E.g., removal of email ids, phone numbers, Facebook profile link.

# Qualitative data analysis: Definitions

- Codes
  - Short hand notation for themes that you see in the data (e.g., “business” might be a code for “I walked down first the McDonalds and the KFC. Most of the small shops are getting closed.”)
- Coding
  - The act of linking themes/codes with passages of qualitative data
- Codebook
  - Lists of codes and definitions of codes

# Coding cycles

Now, Dr. Lucas-Smith Is a bit *cold,* maybe *too "professional, but* she's relatively fresh out of med school so her *knowledge is state-of-the-art.*
That's what *I like* about her: she was *able* to *clear up* two health problems of mine

**Data**

DECORUM

EXPERTISE

YOUNG DOCTOR

**First cycle:** Microscopic themes

DECORUM

EXPERTISE

YOUNG DOCTOR

**Second cycle:** condensing the vast array of initial analytic details into broader themes

# Coding cycles: First

- Grammatical Methods
  - Attribute Coding
  - Magnitude Coding
  - Simultaneous Coding
- Elemental Methods
  - Structural Coding
  - Descriptive Coding
  - In Vivo Coding
  - Process Coding
  - Initial Coding

# Coding cycles: Second

- Pattern Coding
- Focused Coding
- Axial Coding
- Theoretical Coding
- Elaborative Coding
- Longitudinal Coding

# Guidelines for creating codes

- similarity (things happen the same way)
- difference (they happen in predictably different ways)
- frequency (they happen often or seldom)
- sequence (they happen in a certain order)
- correspondence (they happen in relation to other activities or events)
- causation (one appears to cause another)

# Definitions of first cycle coding: Grammatical

- Attribute coding
  - essential information about the data and demographic characteristics of the participants
  - E.g., FB user, age 29, acted as primary hate speech receiver
- Magnitude coding
  - Describe their variable characteristics such as intensity or frequency,
  - E.g., Strongly (STR) Moderately (MOD) No opinions (NO)
- Simultaneous coding
  - Multiple codes for same text

# Definitions of first cycle coding: Elemental

- structural coding
  - Categorizes data, allowing access to data relevant to a particular analysis from a larger data set.
- descriptive coding
  - Create categories with a word or noun the basic topic of a passage of qualitative data.
- In Vivo Coding
  - Use a word or short phrase from the actual language found in the qualitative data record

# Definitions of first cycle coding: Elemental

- Process coding
  - Use only “-ing” word exclusively for labelling
- Initial Coding
  - break down qualitative data into discrete parts
  - closely examining them
  - Comparing them for similarities and differences and come up with codes
  - Create codebooks
  - Mark the text with codes

# Definitions of Second cycle coding (1)

- Pattern coding
  - grouping summaries into a smaller number of sets, themes, or constructs
- Focused coding
  - search for the most frequent or significant codes: categorize coded data based on thematic or conceptual similarity

# Definitions of Second cycle coding (2)

- Axial coding
  - describes a category's properties and dimensions and explores how the categories and subcategories relate to each other
  - Goal: come up with broader category
- Theoretical coding
  - discovering the central or core category that identifies the primary theme of the research

# Definitions of Second cycle coding (3)

- Elaborative coding
  - Borrow a previous study's codes, categories, and themes while a current and related study is underway
  - support or modify the researcher's observations developed in an earlier paper
- Longitudinal coding is the attribution of selected change processes to qualitative data collected and compared across time.

# Specific examples: Descriptive coding

- label the data to understand what is there
  - in a very general flexible way
  - summarize the data
  - Typically a word or short phrase, often a noun

"As I walked towards the campus, there is cheddis, a canteen for the students and then there is SBI and next to it Indian post service" --- BUSINESS

# Specific examples: in-vivo coding

- label the data using participant's own words
  - Go through the text and identify key phrases
  - The set of phrases – your codebook

"I hated college last semester, this semester is a bit better, I don't know why. I guess in this semester I stopped caring what people think of me " --- HATED COLLEGE, THIS SEMSTER IS BETTER, STOPPED CARING

# Specific examples: Process coding

- label the data using only “-ing” words
  - Actions of people, processes

“Well, that’s a problem. My department is pretty small and if you say one thing to one person, then they will tell others and soon everyone knows what’s going on ” --- SPREADING RUMORS, KNOWING WHAT YOU SAID

# Specific examples: Initial coding

- Open coding: each researcher reads through the text and marks passages with “codes” which are similar to labels

# Specific examples: Initial coding

- Open coding: each researcher reads through the text and marks passages with “codes” which are similar to labels

I decided that I wasn’t going
to change the password
because I have heard all the
reviews online about how
hard is it for people
generally if you forget your
new password

# Specific examples: Initial coding

- Open coding: each researcher reads through the text and marks passages with “codes” which are similar to labels

| | |
|---|---|
| I decided that I wasn’t going to change the password | Not changing |
| because I have heard all the reviews online about how | Recommendations |
| hard is it for people generally if you forget your new password | memorability |

# Initial coding: code book

- Awareness
- Deciding
- Preparation
- Changing password
  - Time, cost, resources, forgetting, misplacing
- Issues
- …

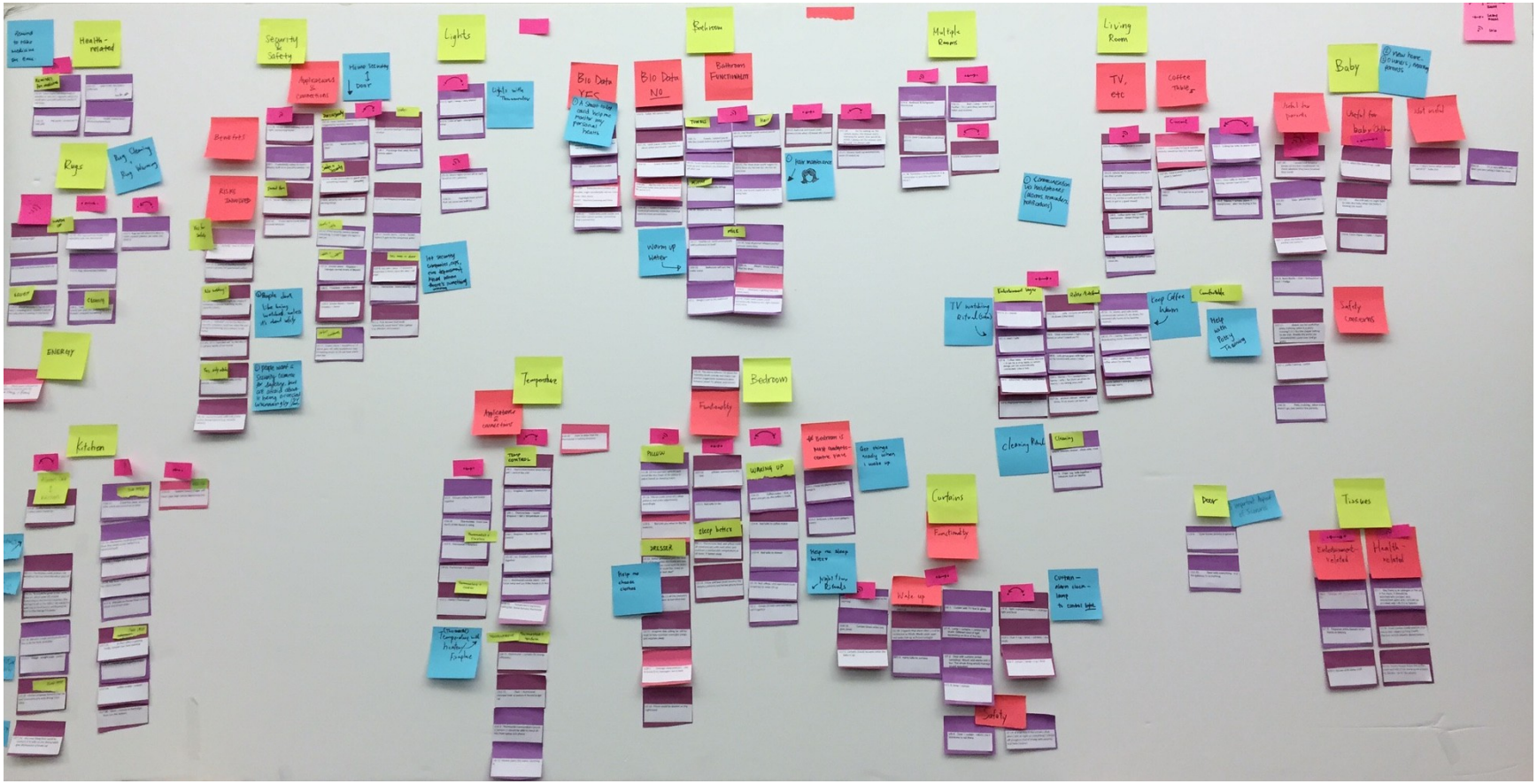
# Initial coding: Using the code book

| | |
|---|---|
| I decided I was not going to change the password | DECIDING: did not update |
| I have heard from reviews online | DECIDING: RESEARCH |
| how hard is it for people generally if you forget your new password | CHANGING PASSWORD: FORGETTING |

# Specific example: Affinity diagram

- Go through the data and identify “themes”
  - these themes become your outcomes.
  - Affinity diagrams are one of the easiest ways to do thematic analysis with a group or by yourself
- Pulls the main concepts of the data out
  - Easy for someone else to understand
  - Themes are grounded in the data with clear examples
  - Only works with a small amount of data
  - May require more than one person to improve validity

# Specific example: Affinity diagram



# Specific example: Affinity diagram

| Job Application |
| --- |
| Requirement |
| Waiting for offer |
| After acceptance |

| Finance |
| --- |
| Living costs |
| Banking |
| Tuition |

# Roadmap

- Qualitative Data Analysis
  - Selecting participants
  - Data analysis techniques

- Inter-rater agreement
- Quantitative data analysis

# Inter-rater reliability

# What Is Inter-Rater Reliability (IRR)?

- The extent to which two or more raters agree
- A fair measurement of label competency
- Addresses the uniformity of the labeling
  - You can label something as “unfair treatment”
  - I can label it “fair treatment”

# Set up

- For open coding you take 10% of data
  - Seat together with your partner to create codes
  - Now you code them separately
  - How to measure where both of you reached at same conclusion?

- You coding scheme should not be subjective

# Cohen's kappa

- Two raters classify each of N items into one of C categories
  - P0 is the observed agreement
  - PE is the expected agreement (when each rater behave randomly)
  - The kappa = (P0 – PE)/ (1- PE)
  - Max - min value?

# More on Cohen's kappa

| **Sentences** | **Label assigned by coder 1 (any of the C labels)** | **Label assigned by coder 2 (any of the C labels)** |
|---|---|---|
| Sentence 1 | X1 | X1 |
| Sentence 2 | X3 | X1 |
| … | | |
| … | | |
| … | | |
| Sentence n-2 | X10 | X5 |
| Sentence n-1 | X11 | X11 |
| Sentence n | X4 | X4 |

# More on Cohen's kappa

| **Sentences** | **Label assigned by coder 1 (any of the C labels)** | **Label assigned by coder 2 (any of the C labels)** |
|---|---|---|
| Sentence 1 | X1 | X1 |
| Sentence 2 | X3 | X1 |
| … | | |
| … | | |
| … | | |
| Sentence n-2 | X10 | X5 |
| Sentence n-1 | X11 | X11 |
| Sentence n | X4 | X4 |

# More on Cohen's kappa

- Lets take the two coder example– each of the code want to label N sentences with "Yes", "NO" labels
  - So there are C = 2 labels (Yes, NO)
  - Lets assume, total rows to label, N = a + b + c + d
  - First create the confusion matrix

| | Coder2_YES | Coder2_NO |
|---|---|---|
| Coder1_YES | a | b |
| Coder1_NO | c | d |

# More on Cohen's kappa

- Lets take the two coder example– each of the code want to label N sentences with "Yes", "NO" labels
  - So there are C = 2 labels (Yes, NO)
  - Lets assume, total rows to label, N = a + b + c + d
  - First create the confusion matrix

| | Coder2_YES | Coder2_NO |
|---|---|---|
| Coder1_YES | a | b |
| Coder1_NO | c | d |

  - P0 = proportion of agreement = (a + d)/(a+b+c+d)

# More on Cohen's kappa

- Lets take the two coder example– each of the code want to label N sentences with "Yes", "NO" labels
  - So there are C = 2 labels (Yes, NO)
  - Lets assume, total rows to label, N = a + b + c + d
  - First create the confusion matrix

| | Coder2_YES | Coder2_NO |
|---|---|---|
| Coder1_YES | a | b |
| Coder1_NO | c | d |

  - P0 = proportion of agreement = (a + d)/(a+b+c+d)
  - PE = Pr (both will say YES at random) + Pr (both will say NO at random) = $\frac{a+b}{a+b+c+d} * \frac{a+c}{a+b+c+d} + \frac{c+d}{a+b+c+d} * \frac{b+d}{a+b+c+d}$

# Example

| | Coder2_YES | Coder2_NO |
|---|---|---|
| Coder1_YES | 34 | 26 |
| Coder1_NO | 19 | 21 |

PO = (34 + 21) / (34+26+19+21) = 55/100= 0.55

PE = (34 + 26)/(100) * (34 + 19)/100 +

(19 + 21)/(100) * (26 + 21)/100 = 0.318 + 0.188 = 0.506

Kappa = (PO - PE) / (1 - PE) = (0.55 – 0.506) / (1 – 0.506) = 0.08

# Interpretation

| | | | |
|---|---|---|---|
| <.20 | Poor | | |
| .21-.40 | Fair | .61-.80 | Substantial |
| .41-.60 | Moderate | >.81 | Excellent |

# Other variations

- Scott's Pi
- Fleiss's Kappa  (multi-rater agreemnt)
- Krippendorff's alpha (multi-rater agreement, handles missing data)

# QUESTION

- The following Cohen's kappa (*k*) values strongly suggest that the instrument, the raters, the training protocol, or other aspects of the measurement situation need to be modified or there is an error in the kappa calculation (select all that apply):

*A.* $k = .69$

*B.* $k = .20$

*C.* $k = 3.2$

*D.* $k = .80$

# ANSWER

- Answer: B and C

- In general, kappa values under .60 may indicate need for modifications in the instrument, the raters, the training protocol, or other aspects of the measurement situation. A value of 3.2 is not within the range of an accurately calculated  kappa score.
- Bonus question: How to rectify it?