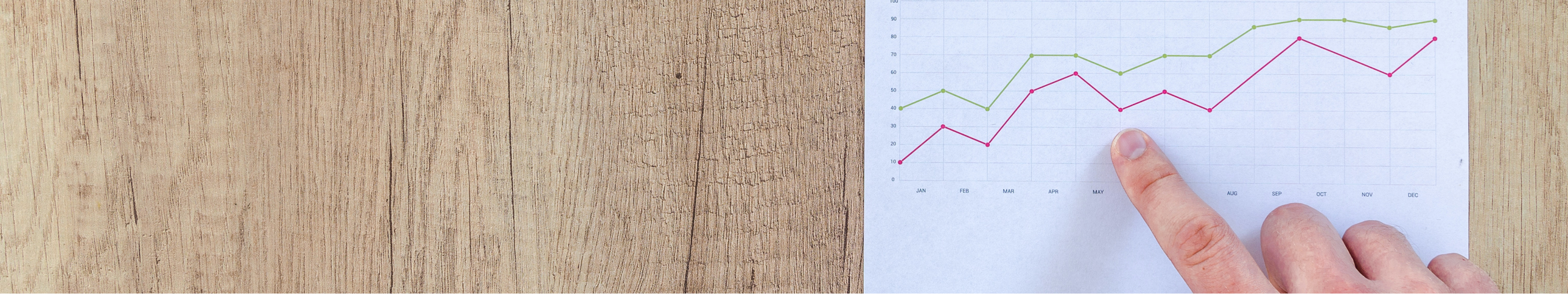


Current Topics in Media Computing and HCI

Quantitative and Qualitative Analyses

Krishna Subramanian, M.Sc.
Media Computing Group
RWTH Aachen University

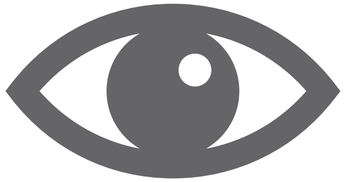
Summer Semester 2020

https://hci.rwth-aachen.de/cthci

RWTH AACHEN UNIVERSITY

Recap: Three Approaches to HCI Research

Empirical

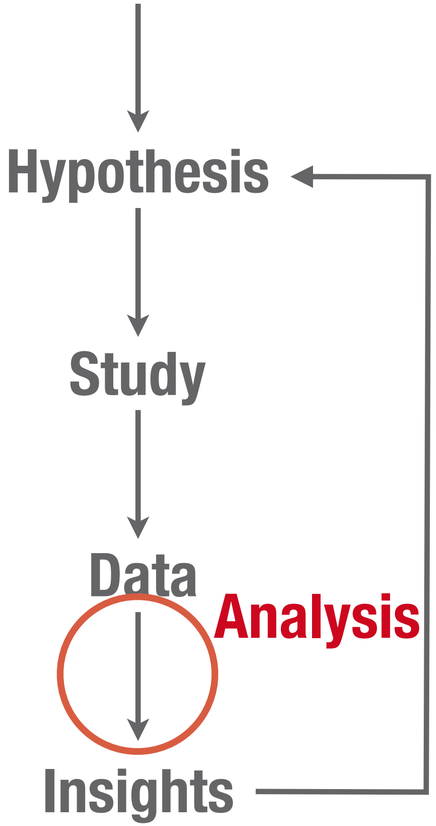


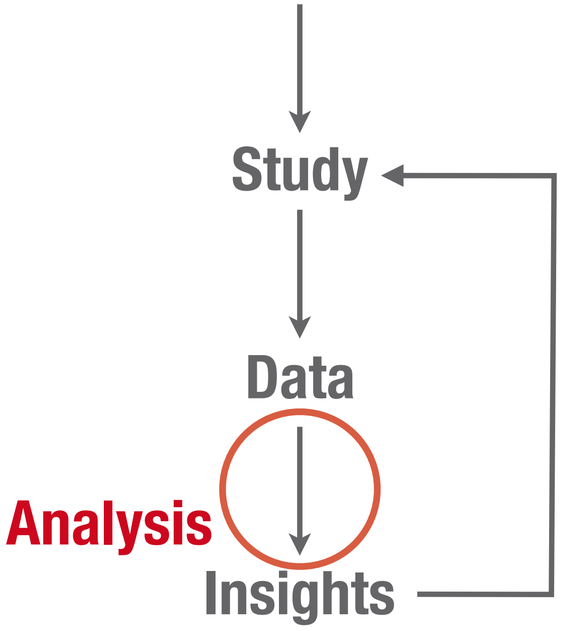
Ethnography

Engineering & Design

Recap: Empirical Research and Ethnography

Empirical Research





Ethnography

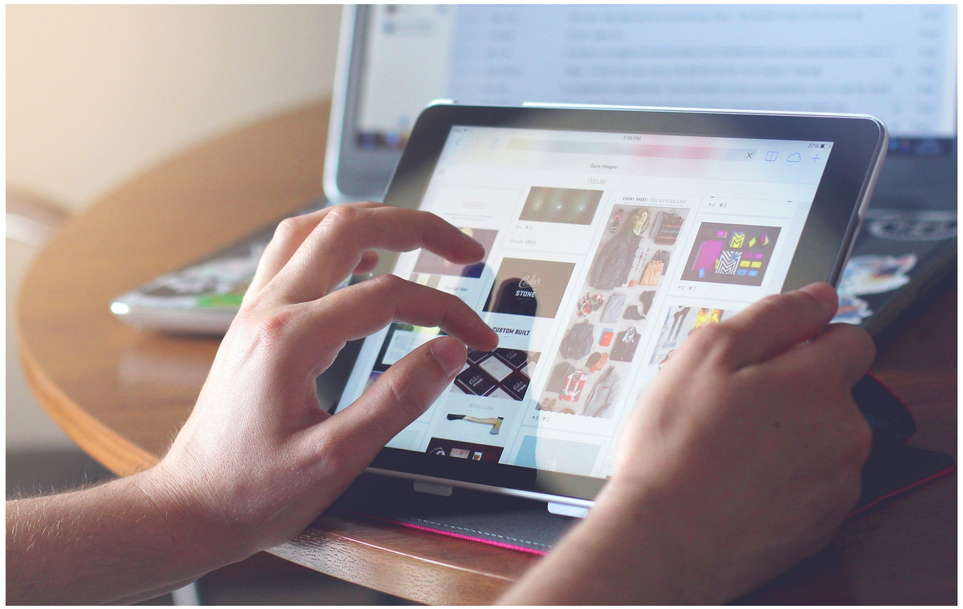
Quantitative vs. Qualitative Analyses

Quantitative	Qualitative
Use numbers to present a research finding	Use text, videos, or pictures to present a research finding
Used to confirm theories and assumptions—mostly in empirical research	Used to understand people and processes—mostly in ethnography
Data collection through lab experiments and surveys	Data collection through interviews, observations, and diary studies
Data analysis through significance testing, regression models, Bayesian analysis, etc.	Data analysis through grounded-theory, affinity diagramming, etc.

In reality, you often mix *aspects* of quantitative and qualitative analyses.

Quantitative Analysis

Example Empirical Research: Comparing Input Methods for Typing



Fingers

Stylus

Steps in Empirical Research

1. Formulate hypothesis

2. Design experiment by identifying the dependent and independent variables while limiting extraneous variables

3. Recruit participants

4. Run the experiment to collect experimental data

5. Perform quantitative analysis on experimental data to accept or reject hypothesis

1. **Formulate hypothesis**
2. Design experiment, pick dependent & independent variables, and limit extraneous variables
3. Recruit subjects
4. Run experiment (to collect data which you will analyze)
5. Perform statistical analysis on the collected data to accept or reject hypothesis

- Null hypothesis (H_0): The typing speed when using fingers is not different from the typing speed when using a stylus.
- Alternative hypothesis (H_1): The typing speed when using fingers is different from the typing speed when using a stylus.

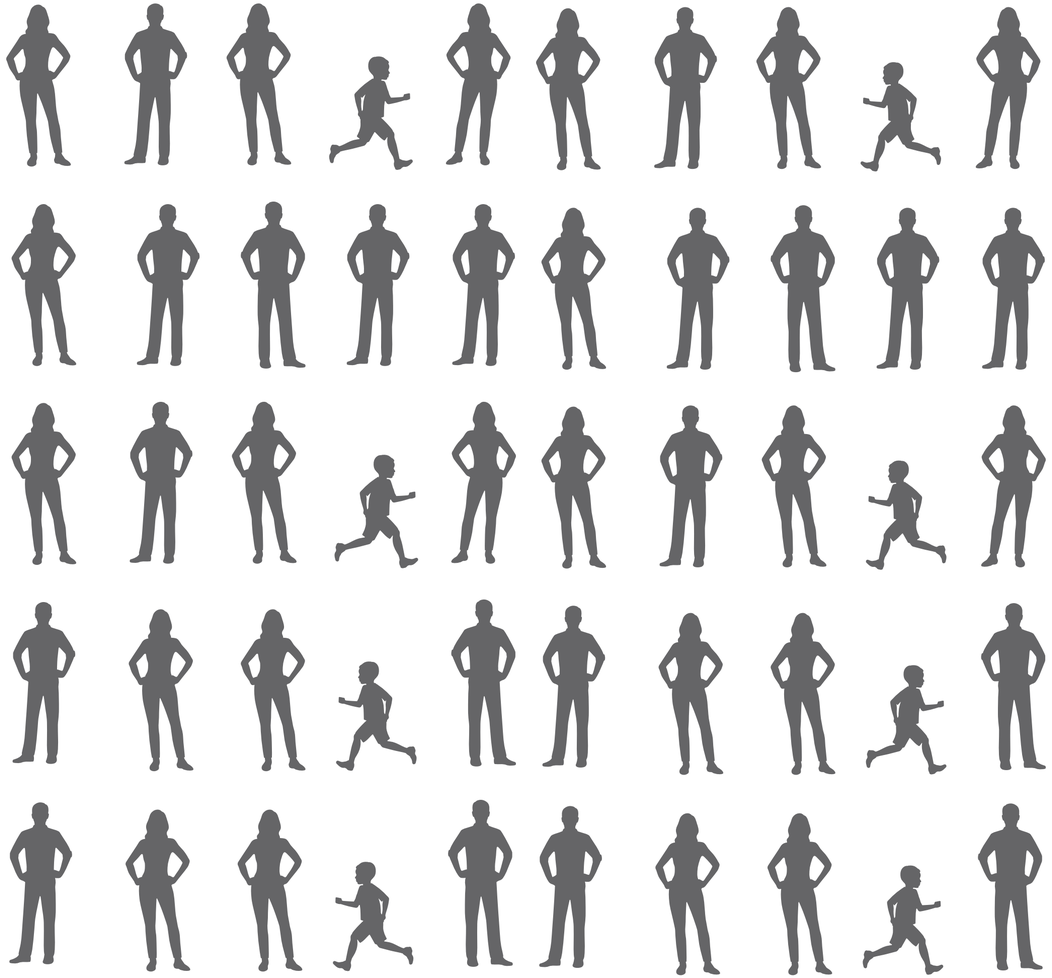
1. Formulate hypothesis
2. **Design experiment, pick dependent & independent variables, and limit extraneous variables**
3. Recruit subjects
4. Run experiment (to collect data which you will analyze)
5. Perform statistical analysis on the collected data to accept or reject hypothesis

- Experimental design: Between-subjects design
- Variables
  - Independent variable (IV): Input method with levels *fingers* and *stylus*
  - Dependent variable (DV): Task completion time (in seconds)
- Control other variables (user experience, model of the smartphone/tablet, etc.)

1. Formulate hypothesis
2. Design experiment, pick dependent & independent variables, and limit extraneous variables
3. **Recruit subjects**
4. Run experiment (to collect data which you will analyze)
5. Perform statistical analysis on the collected data to accept or reject hypothesis

- Select a representative sample

Sample vs. Population



Sample vs. Population

1. Formulate hypothesis
2. Design experiment, pick dependent & independent variables, and limit extraneous variables
3. Recruit subjects
4. **Run experiment (to collect data which you will analyze)**
5. Perform statistical analysis on the collected data to accept or reject hypothesis

1. Formulate hypothesis
2. Design experiment, pick dependent & independent variables, and limit extraneous variables
3. Recruit subjects
4. Run experiment (to collect data which you will analyze)
5. **Perform statistical analysis on the collected data to accept or reject hypothesis**

Significance Testing

- Difference in means between sampled distributions can be due to

  => an actual difference in the populations

Significance Testing

- Difference in means between sampled distributions can be due to

  => an actual difference in the populations

  => no actual difference, but difference in means is due to a sampling error

Significance Testing

- Difference in means between sampled distributions can be due to

  => an actual difference in the populations

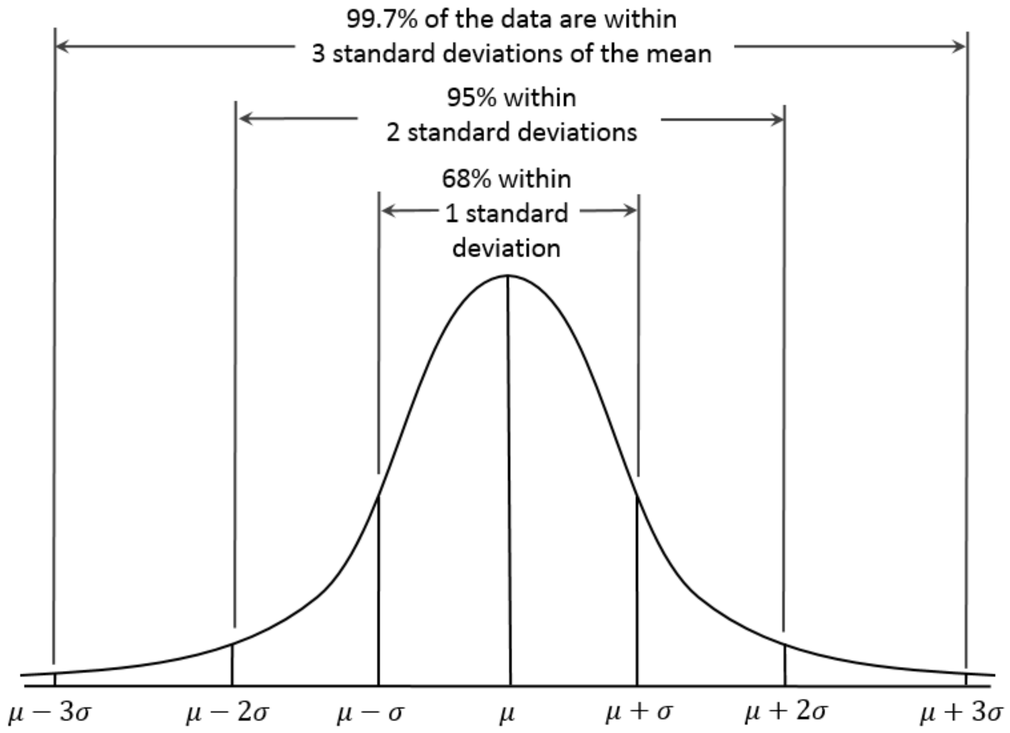
  => no actual difference, but difference in means is due to a sampling error

- Significance tests can tell these two apart

Significance Testing

- Assume H_0 to be true (i.e., no difference at the population level)
- Conduct the experiment and collect data
- Fit a statistical model to the data (e.g., *t*-distribution, F-distribution)
- Compute ***p*-value,** which is defined as:
  - **“The chances of obtaining the experimental data we’ve collected assuming the null hypothesis is true”**

How p-Values Work



- The **68–95–99.7 rule,** a characteristic of Gaussian distributions

p-Value

- *p*-value gives us confidence in accepting or rejecting the null-hypothesis (i.e., no difference between distributions)
  - $p = 0$ => There is no chance that the null hypothesis is true, which means that the alternate hypothesis is true (there is a difference between distributions)
  - $p = 1$ => The means of the samples' distributions are the same
- Remember: *"If the p-value is low, the null-hypothesis has to go!"*
- In HCI, use a cut-off of 0.05
  - **$p \leq 0.05$** => **reject** H_0 (and accept H_1)
  - **$p > 0.05$** => **accept** H_0

In-Class Exercise: p-value

- Which of the following statements are correct?

A. There is a 3% probability that school students watch TV more than college students

B. There is a 3% probability that school students watch TV in a different amount than college students

C. Assuming that school students watch TV in different amount than college students, there is a 3% probability that this result occurs

D. Assuming that school students and college students watch TV in the same amount, there is a 3% probability that this result occurs

In-Class Exercise: p-value

- Which of the following statements are correct?

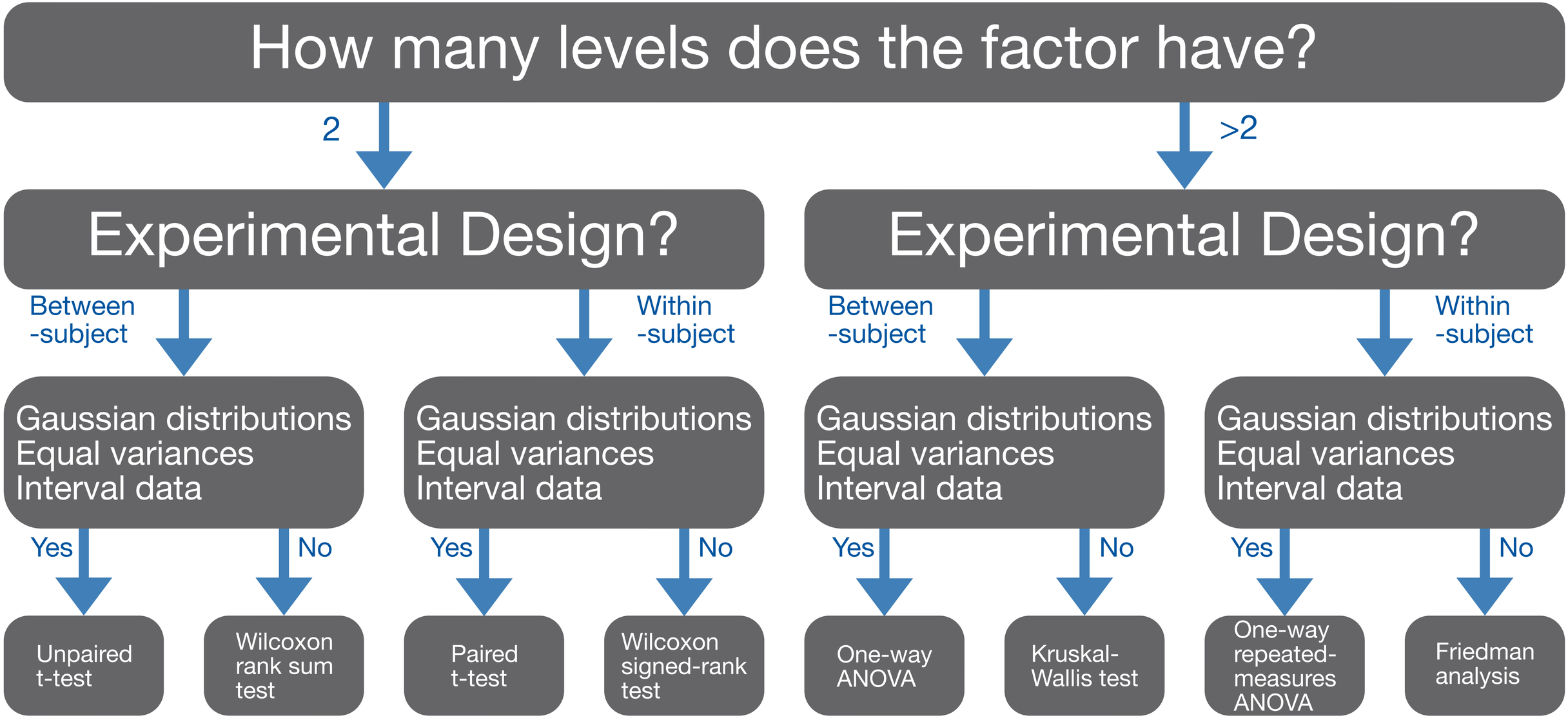
A. There is a 3% probability that school students watch TV more than college students

B. There is a 3% probability that school students watch TV in a different amount than college students

C. Assuming that school students watch TV in different amount than college students, there is a 3% probability that this result occurs

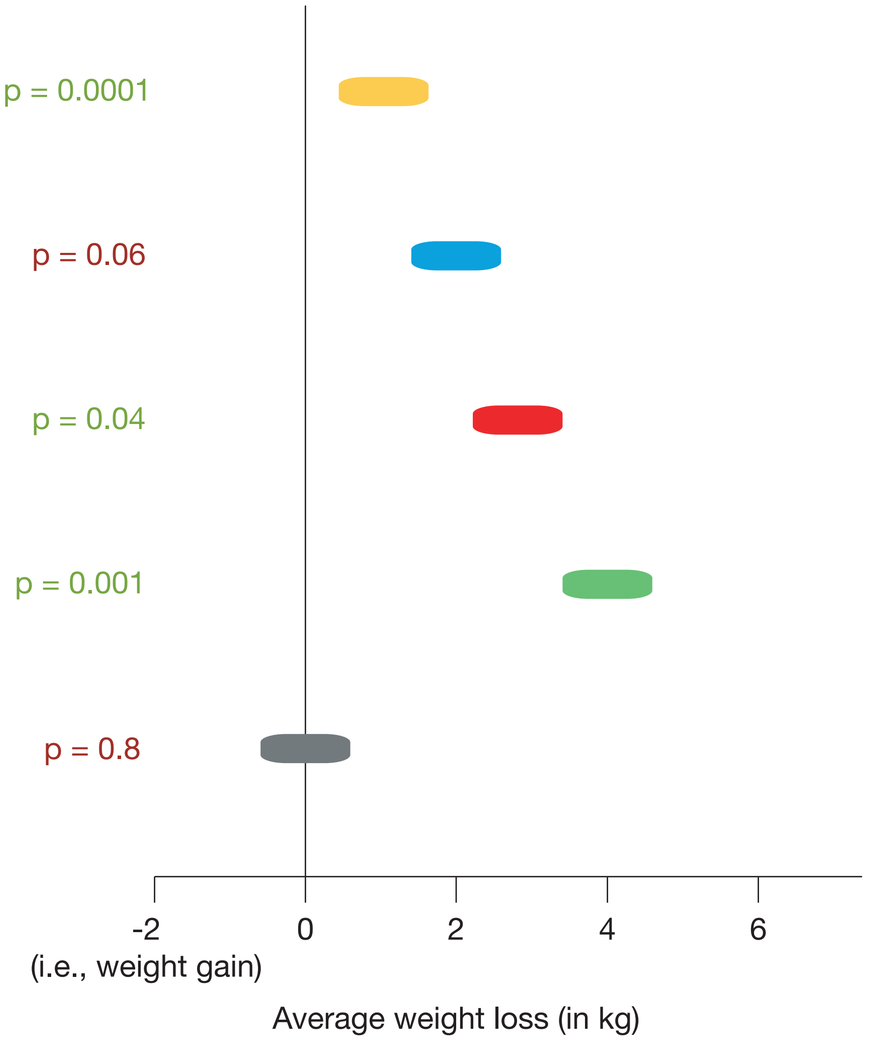
D. Assuming that school students and college students watch TV in the same amount, there is a 3% probability that this result occurs

Selecting the Correct Statistical Test



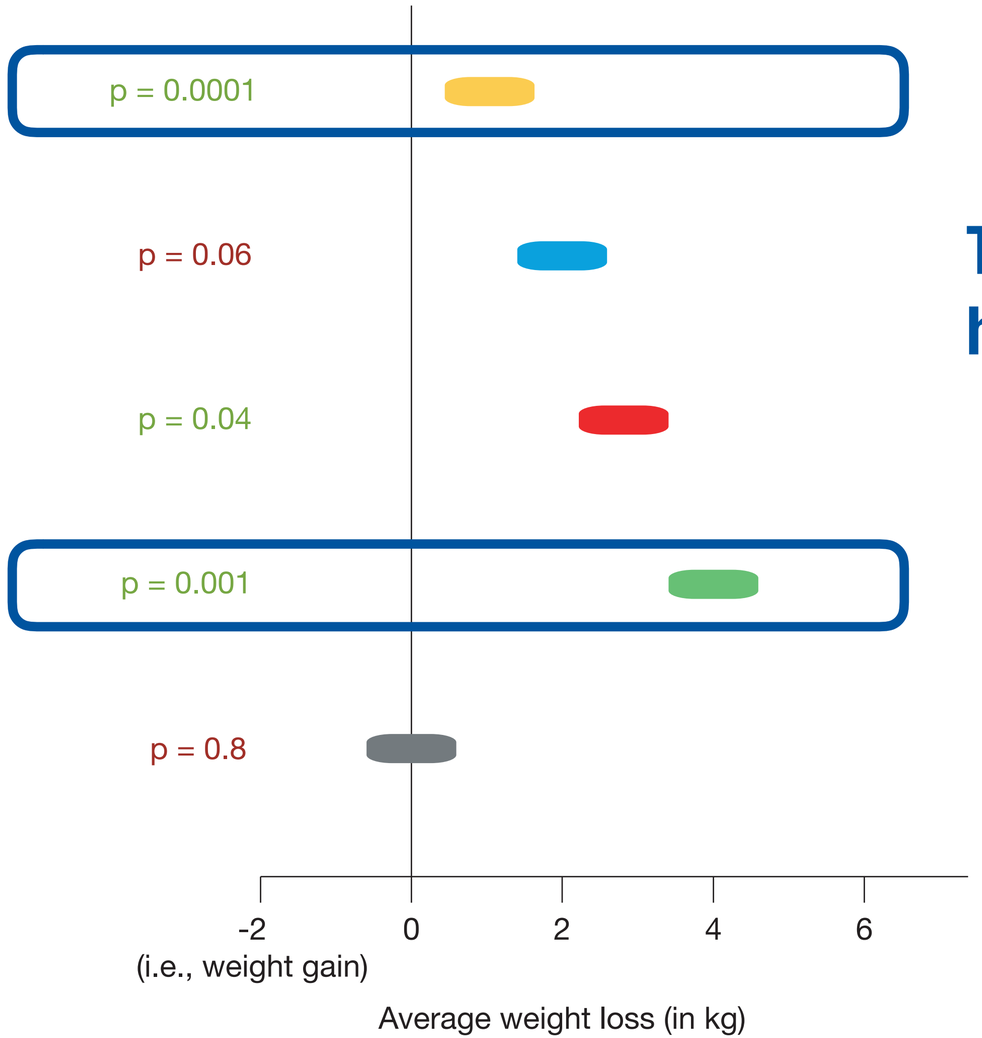
Statistically Significant = Practically Significant?

Scenario: Weight Loss via Pills



Adapted from Ziliak and Mccloskey, 2009

Scenario: Weight Loss via Pills



Takeaway: p-values do not help with interpretation!

Adapted from Ziliak and Mccloskey, 2009

Effect Size

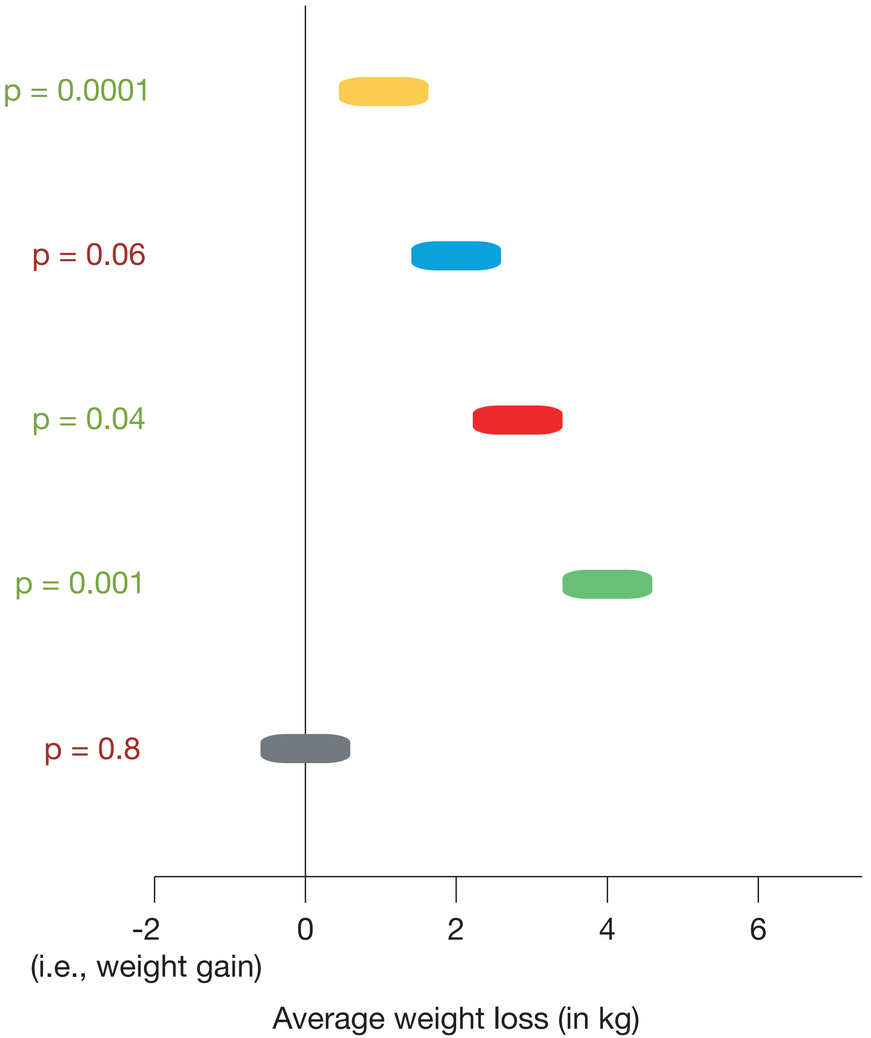
- *p*-value tells us the chances of the sample distributions coming from the same population distribution
  - **But:** Statistically significant ($p < 0.05$) != practically significant
- For practical significance, we use effect sizes
- Effect size: Measure of how big the difference between distributions are

Effect Size: Examples

- Difference between two means
  - E.g., Stylus is 40s slower than Touch
  - In original unit, intuitive
- Percentage and ratio
  - E.g., Stylus is twice slower than Touch
  - Emphasize the magnitude of effect
- Difference between means has a measurement unit (e.g., seconds, points, etc.) and therefore requires domain knowledge

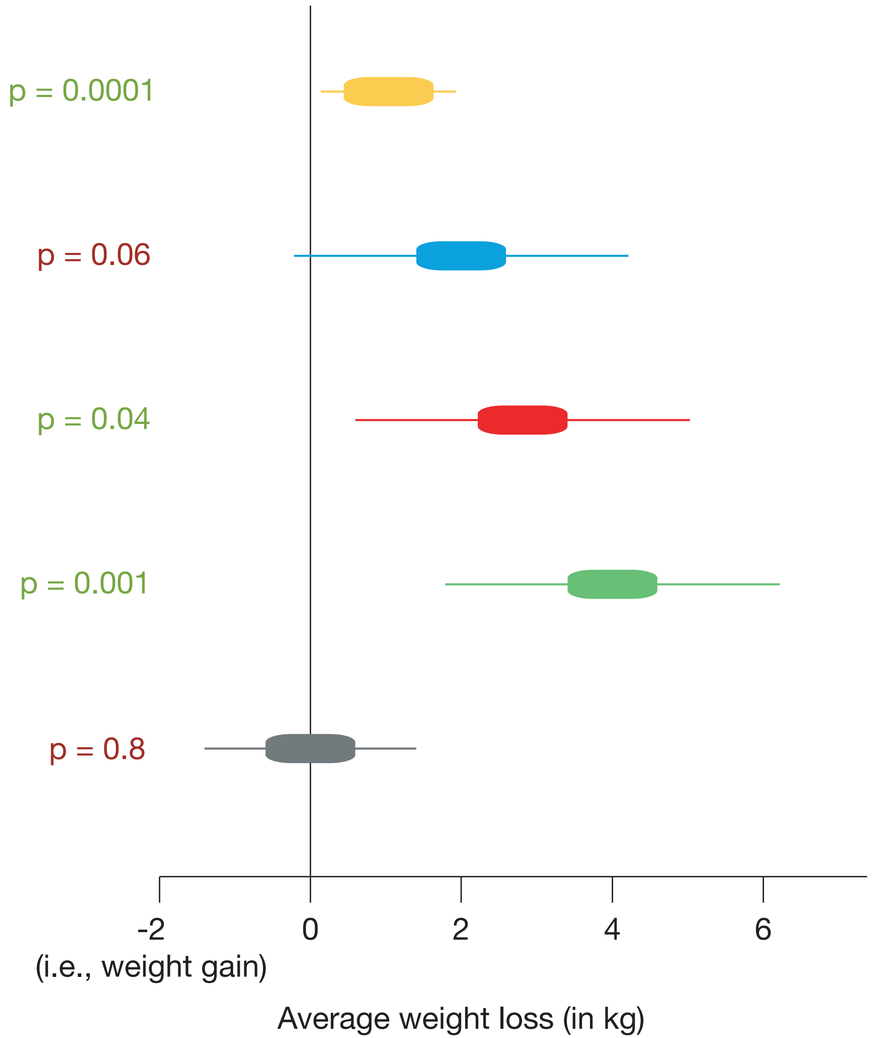
Confidence Intervals

Scenario: Weight Loss via Pills



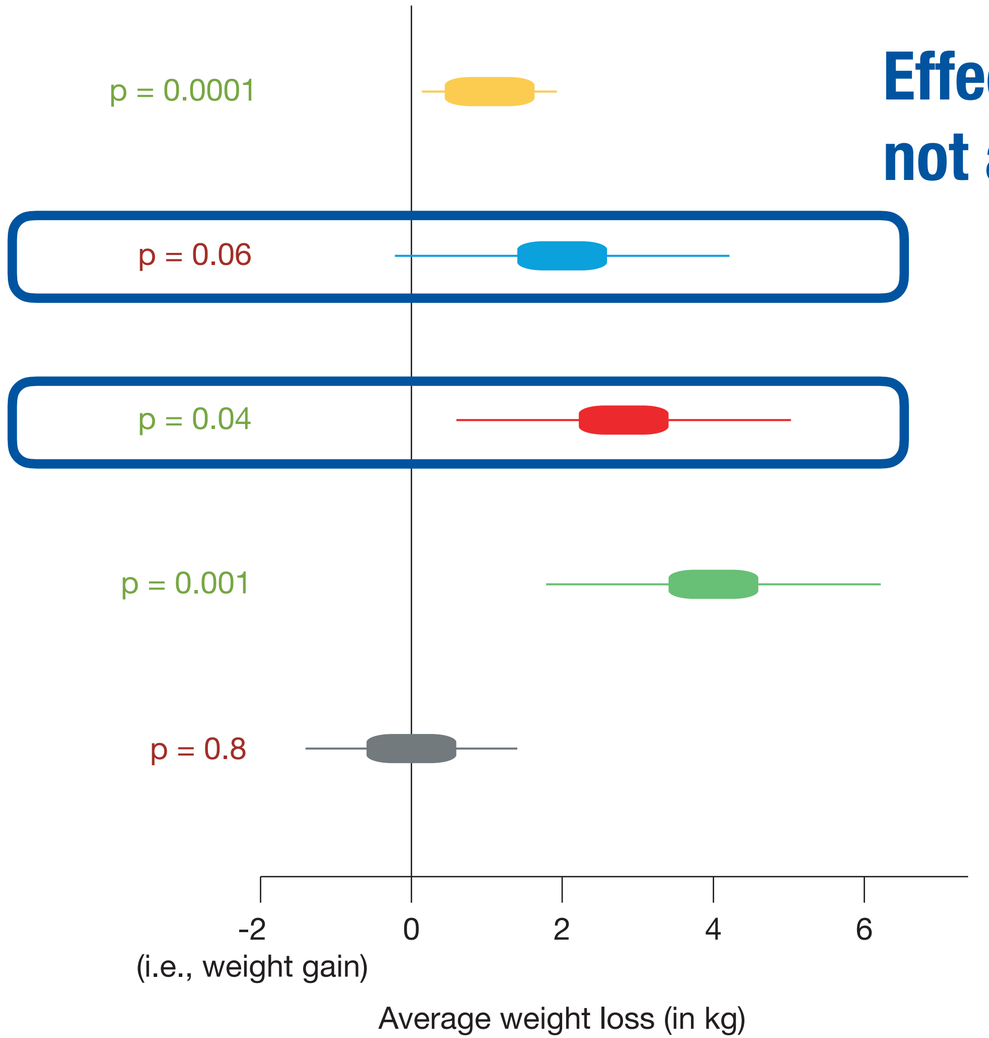
Adapted from Ziliak and Mccloskey, 2009

Scenario: Weight Loss via Pills



Adapted from Ziliak and Mccloskey, 2009

Scenario: Weight Loss via Pills



Effect sizes by themselves are not adequate for interpretation!

Adapted from Ziliak and Mccloskey, 2009

95% Confidence Interval

- An interval estimate (i.e., a range) of the population mean
- In an infinite number of experiments, 95% of the time, the 95% CIs will contain the population mean
- 95% is a convention, might vary across domains
  (e.g., medicine, psychology have different conventions)

Guidelines for Reporting

- Report effect sizes and 95% confidence intervals in addition to descriptive statistics (mean, sd) and results of the significance test

- **Be transparent:** Include the experimental data, analysis script, and your rationale for analysis

Required Reading

- (Cumming and Finch, American Psychologist 2005) Inference by Eye: Confidence Intervals and How to Read Pictures of Data

- (Delmas et al., 2005) Using Assessment Items To Study Students' Difficulty Reading and Interpreting Graphical Representations of Distributions

- An exercise sheet on interpreting graphs (named "In-Class Exercise 1 — Interpreting Graphs.pdf") will be uploaded to RWTHmoodle.

  Note: This will not be graded; it is only to help you get some practice.

Useful Resources

- Statistical Methods for HCI Research by Koji Yatani, U. of Tokyo
    - Link: **http://yatani.jp/teaching/doku.php?id=hcistats:start**
- Practical Statistics for HCI by Jacob O. Wobbrock, U. of Washington
    - Uses SPSS and JMP (trial version available for free download)
    - Link: **http://depts.washington.edu/aimgroup/proj/ps4hci/**
- How 95% CI's Work: **http://www.latrobe.edu.au/psychology/research/research-areas/cognitive-and-developmental-psychology/esci/understanding-the-new-statistics**
    - Chapters 1-4, CIJumping tab

Useful Resources

- How to compute 95% CI: **http://www.stat.yale.edu/Courses/1997-98/101/confint.htm**
- How to report statistics in thesis/research papers: **http://my.ilstu.edu/~jhkahn/apastats.html** (APA style)
- The New Statistics: Cumming, G. (2013). The New Statistics. Psychological Science, 25(1), 7–29. **http://doi.org/10.1177/0956797613504966**
- (Alternative approach) Bayesian analysis: Kruschke, J. (2014). Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan. Academic Press.

Summary

- We need statistical analysis to establish causal relationship between our dependent and independent variables
- Significance testing can be used to accept or reject null hypothesis
- Effect size quantifies the effect of independent variable on the dependent variable
- 95% confidence intervals help deal with uncertainty in data

Qualitative Analysis

When to Use Qualitative Analysis?

- To answer “how” and “why” questions, e.g.,
  - How do people use ticket vending machines?
  - Why do elderly people find it difficult to use technology?
- Can be useful to identify problems, a useful first step before building artifacts
- Can also be used to evaluate an artifact
  - But: You cannot make generalizable claims about any hypothesis; use quantitative analysis for this

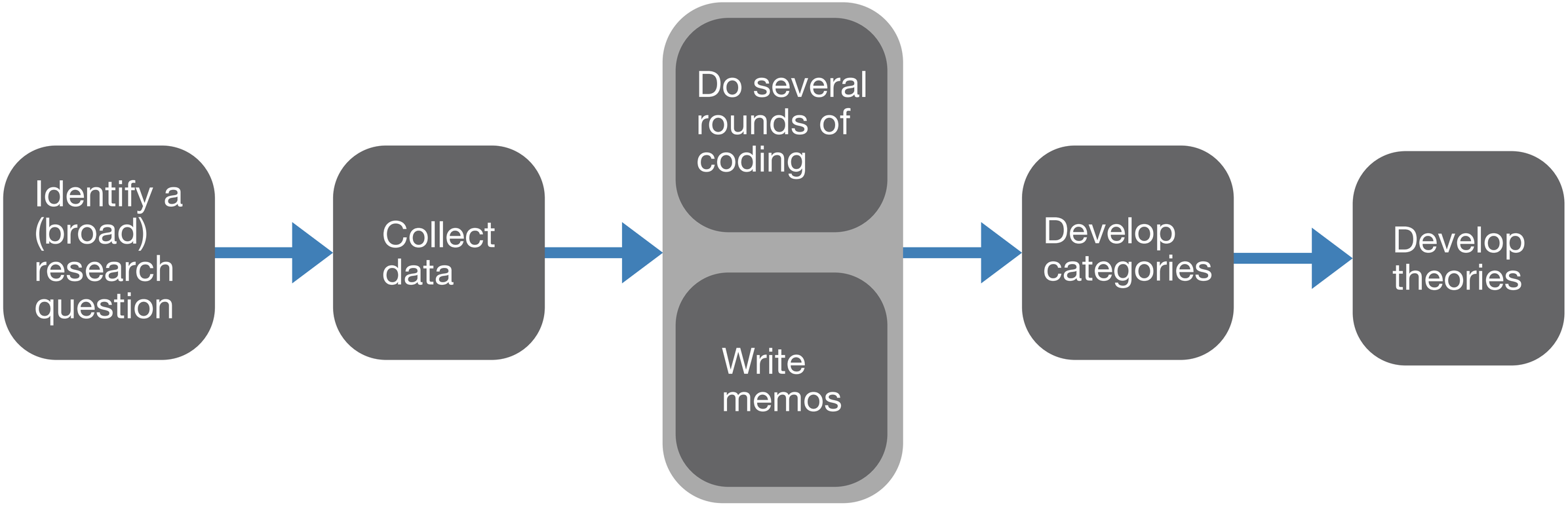
Data Collection Techniques

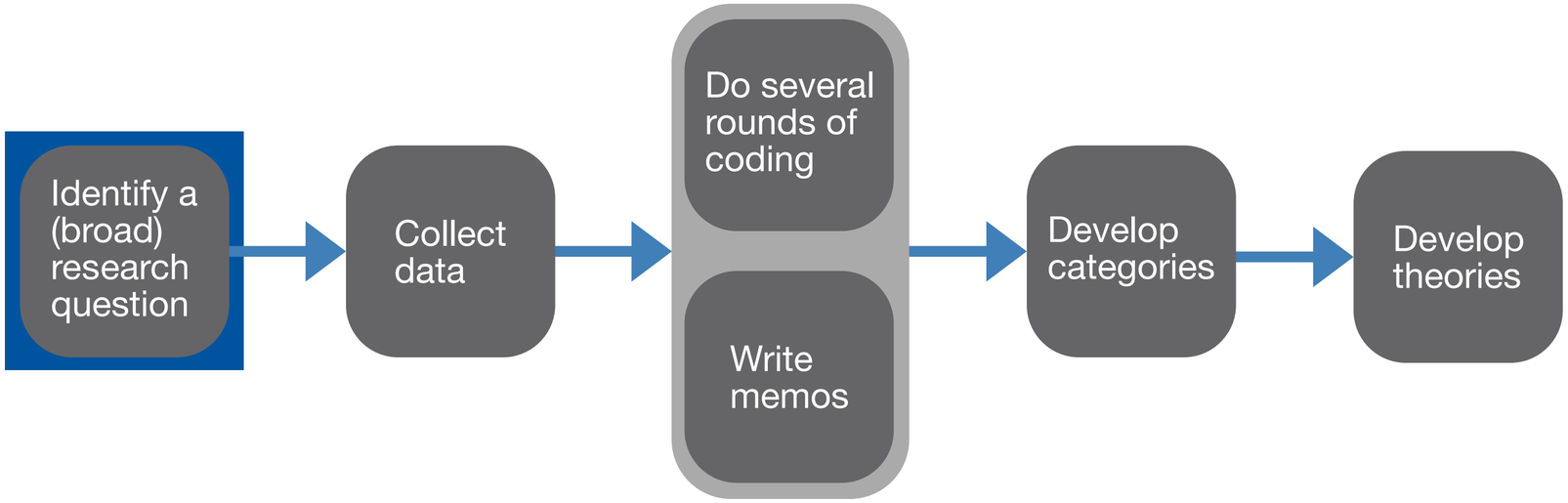
- Interviews
- Diary studies
- Observation
- Focus groups
- …

Evaluation Techniques

- Grounded-theory
- Affinity diagramming/contextual design

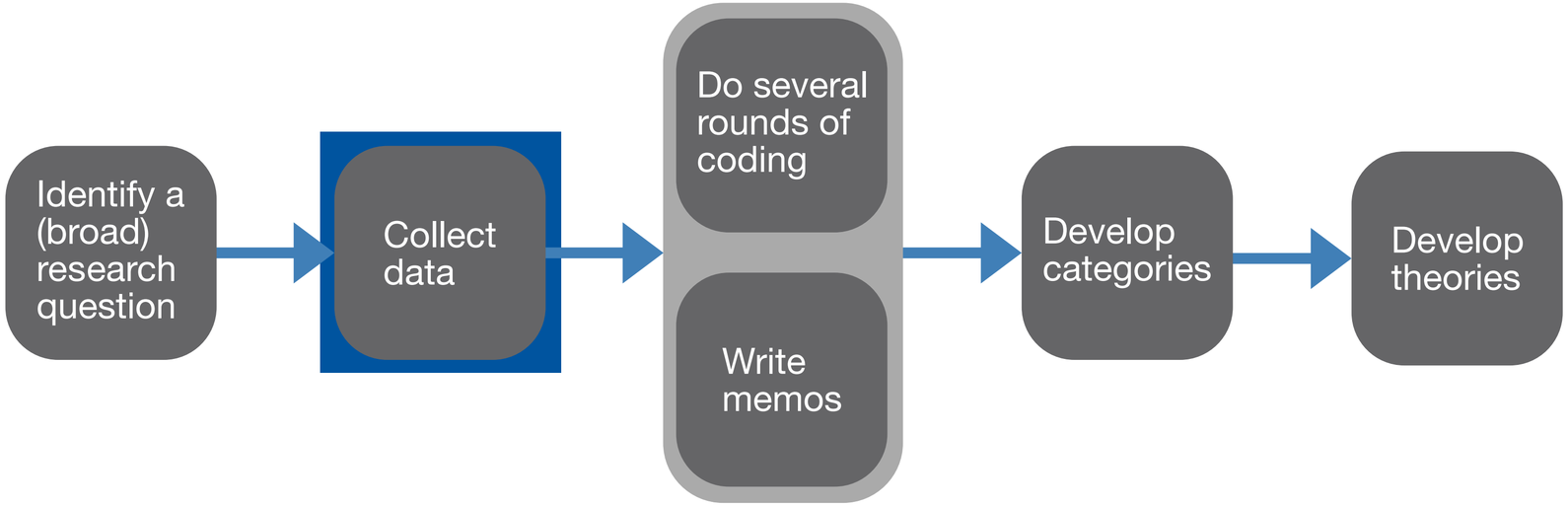
Overview of Grounded-Theory Method



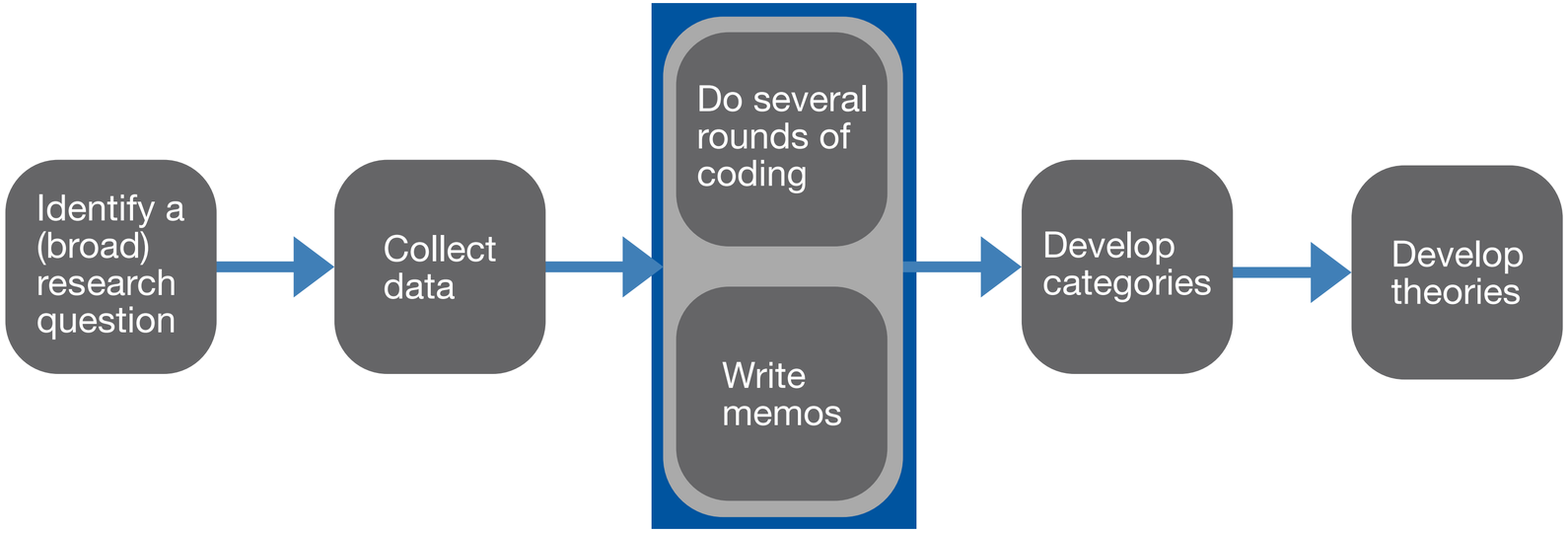


- “How do foreigners use the Deutsche Bahn ticket vending machine?”

needpix.com



- Observation and follow-up interviews
- Questions could be based on what you observed
  - E.g., “I noticed that you took some time to figure out the payment method—could you talk about what happened?”



Passenger:ExpressTrain

UseCase:OnlyCash

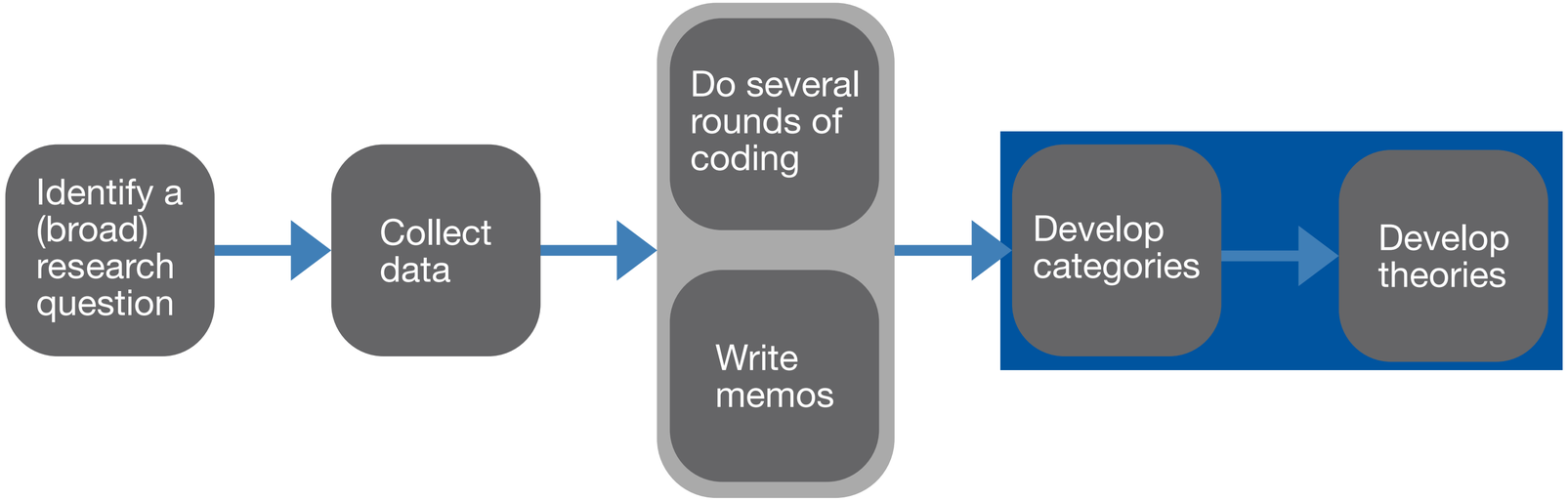
"I had to take an ICE train this morning. I had only 5 minutes left to buy a ticket because I only had cash to pay for the ticket. There was a bit of a queue so I had to wait for a couple of minutes before I can buy the ticket. The process was overall OK, but I had difficulty selecting the destination—I first selected Bochum instead of Bonn."

Problem:Waiting

Problem:CitySelection

Example Memo

Many participants have problems selecting the destination. Is this because of the interface? May be I can check if the participants who are not so rushed also make this mistake. Also, does this occur for specific cities? These are things I want to investigate.



- Combine codes to develop categories
- Use categories and memos to develop theories
  - E.g., “Foreigners face problems even before using the interface which impacts their experience more than usability issues”

I. Problems

Cat. 1: Problems before purchase

I.I. Queues before purchase

I.II. Finding the least used machines

I.III. Determining which type of ticket to buy

Cat. 2: Interface problems

I.IV. Incorrect selection of cities

I.V. Unable to pay with the correct payment method

I.VI. Changing language

Validity, Reliability, and Transparency

- To ensure **validity**
  - Perform **triangulation,** i.e., use more than one source of data to confirm a finding
  - Use a **representative sample,** e.g., make sure you have people of different countries, ethnicities, gender, age group, and experience with technology
- To ensure **reliability,** compute **inter-coder agreement** after coming up with a coding scheme
- Guidelines for **transparent** qual. research: Aguinis, Herman, and Angelo M. Solarino. *"Transparency and Replicability in Qualitative Research: The Case of Interviews With Elite Informants." Strategic Management Journal* 40.8 (2019): 1291-1315. (optional read)

Further Reading

- Leung, Lawrence. "Validity, Reliability, and Generalizability in Qualitative research." Journal of Family Medicine and Primary Care 4.3 (2015): 324.

- (For Grounded-theory analysis) Saldaña, Johnny. The Coding Manual for Qualitative Researchers. SAGE, 2015.

- (For affinity diagramming/contextual design) Beyer, Hugh, and Karen Holtzblatt. Contextual Design: Design for Life. M. Kaufmann, 2017.

Summary

- Use qualitative analysis to answer “how” and “why” questions
- Often a useful first step to better understand problems or to identify problems
- Can also be a useful tool for evaluating an artifact
- Requires qualitative coding, inductive reasoning, and rigorous comparisons of data