Empirical Methods

Empirical Methods

$t = a + b$

Research Landscape

- Quantitative = Positivist/post-positivist approach
  - Evaluate hypotheses via experimentation
- Qualitative = Constructivist approach
  - Build theory from data

Overview: Empirical Methods

- Wikipedia
  - Any research which bases its findings on observations as a test of reality
  - Accumulation of evidence results from planned research design
  - Academic rigor determines legitimacy
- Frequently refers to scientific-style experimentation
  - Many qualitative researchers also use this term

Positivism

- Describe only what we can measure/observe
  - No ability to have knowledge beyond that
- Example: psychology
  - Concentrate only on factors that influence behaviour
  - Do not consider what a person is thinking
- Assumption is that things are deterministic

Post-Positivism

- A recognition that the scientific method can only answer question in a certain way
- Often called critical realism
  - There exists objective reality, but we are limited in our ability to study it
  - I am often influenced by my physics background when I talk about this
    - Observation => disturbance
  - We can't test everyone and everything
    - We are just accumulating evidence.

Implications of Post-Positivism

- The idea that all theory is fallible and subject to revision
  - The goal of a scientist should be to disprove something they believe
- The idea of triangulation
  - Different measures and observations tell you different things, and you need to look across these measures to see what's really going on
- The idea that biases can creep into any observation that you make, either on your end or on the subject's end

Experimental Biases in the RW

- Hawthorne effect/John Henry effect
- Experimenter effect/Observer-expectancy effect
- Pygmalion effect
- Placebo effect
- Novelty effect

Hawthorne Effect

- Named after the Hawthorne Works factory in Chicago
- Original experiment asked whether lighting changes would improve productivity
  - Found that anything they did improved productivity, even changing the variable back to the original level.
  - Benefits stopped or studying stopped, the productivity increase went away
- Why?
  - Motivational effect of interest being shown in them
- Also, the flip side, the John Henry effect
  - Realization that you are in control group makes you work harder

Experimenter Effect

- A researcher's bias influences what they see
- Example from Wikipedia: music backmasking
  - Once the subliminal lyrics are pointed out, they become obvious
- Dowsing
  - Not more likely than chance
- The issue:
  - If you expect to see something, maybe something in that expectation leads you to see it
- Solved via double-blind studies

Pygmalion effect

- Self-fulfilling prophecy
- If you place greater expectation on people, then they tend to perform better
- Studied teachers and found that they can double the amount of student progress in a year if they believe students are capable
- If you think someone will excel at a task, then they may, because of your expectation

Placebo Effect

- Subject expectancy
  - If you think the treatment, condition, etc has some benefit, then it may
- Placebo-based anti-depressants, muscle relaxants, etc.
- In computing, an improved GUI, a better device, etc.
  - Steve Jobs: http://www.youtube.com/watch?v=8JZBLjxPBUU
  - Bill Buxton: http://www.youtube.com/watch?v=Arrus9CxUiA

Novelty Effect

- Typically with technology
- Performance improves when technology is instituted because people have increased interest in new technology
- Examples: Computer-Assisted instruction in secondary schools, computers in the classroom in general, etc.

What can you test?

- Three things?
  - Comparisons
  - Models
  - Exploratory analysis
- Reading was comparative

Concepts

- Randomization and control within an experiment
  - Random assignment of cases to comparison groups
  - Control of the implementation of a manipulated treatment variable
  - Measurement of the outcome with relevant, reliable instruments
- Internal validity
  - Did the experimental treatments make the difference in this case?
- Threats to validity
  - History threats (uncontrolled, extraneous events)
  - Instrumentation threats (failure to randomize interviewers/raters across comparison groups)
  - Selection threat (when groups are self-selected)

Themes

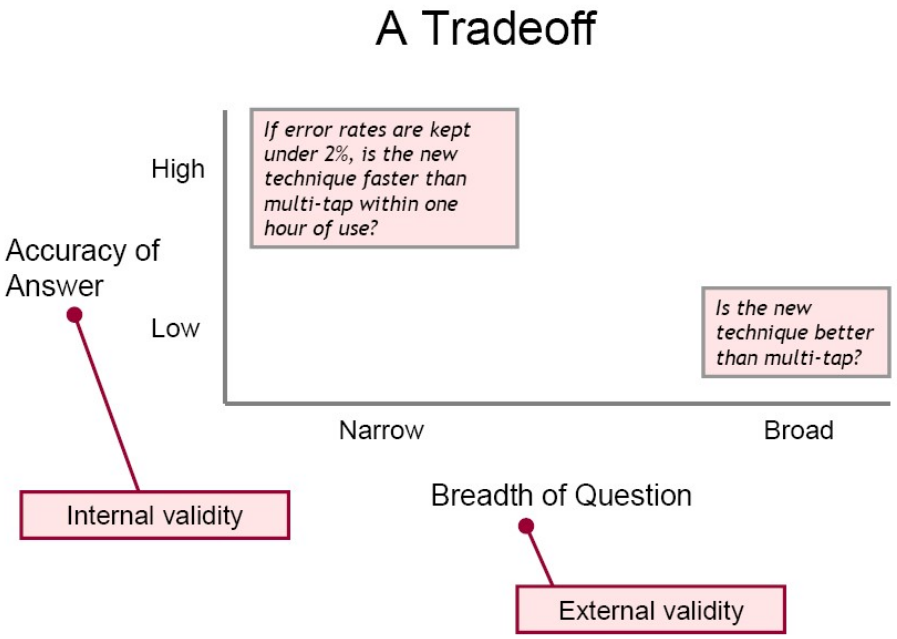
- HCI context
- Scott MacKenzie's tutorial
  - Observe and measure
  - Research questions
  - User studies – group participation
  - User studies – terminology
  - User studies – step by step summary
  - Parts of a research paper

Observations and Measures

- Observations
  - Manual (human observer)
    - Using log sheets, notebooks, questionnaires, etc.
  - Automatically
    - Sensors, software, etc.
- Measurements (numerical)
  - Nominal: Arbitrary assignment of value (1=male, 2=female
  - Ordinal: Rank (e.g. 1^{st}, 2^{nd}, 3^{rd}, etc.
  - Interval: Equal distance between values, but no absolute zero
  - Ratio: Absolute zero, so ratios are meaningful (e.g. 40 wpm is twice as fast as 20 wpm typing)
- Given measurements and observations, we:
  - Describe, compare, infer, relate, predict

Research Questions

- You have something to test (a new technique)
- Untestable questions:
  - Is the technique any good?
  - What are the technique's strengths and weaknesses?
  - Performance limits?
  - How much practice is needed to learn?
- Testable questions seem narrower
  - See example at right



Scott MacKenzie's course notes

Research Questions (2)

- Internal validity
  - Differences (in means) should be a result of experimental factors (e.g. what we are testing)
  - Variances in means result from differences in participants
  - Other variances are controlled or exist randomly
- External validity
  - Extent to which results can be generalized to broader context
  - Participants in your study are “representative”
  - Test conditions can be generalized to real world
- These two can work against each other
  - Problems with “Usable”
  - Noted by many with the readings

Research Questions (3)

- Given a testable question (e.g. a new technique is faster) and an experimental design with appropriate internal and external validity
- You collect data (measurements and observations)
- Questions:
  - Is there a difference
  - Is the difference large or small
  - **Is the difference statistically significant**
  - Does the difference matter

Significance Testing

- R. A. Fisher (1890-1962)
  - Considered designer of modern statistical testing
- Fisher's writings on Decision Theory versus Statistical Inference:
  - An important difference is that Decisions are final while the state of opinion derived from a test of significance is provisional, and capable, not only of confirmation but also of revision (p.100).
  - A test of significance ... is intended to aid the process of learning by observational experience. In what it has to teach each case is unique, though we may judge that our information needs supplementing by further observations of the same, or of a different kind (pp. 100-101).
- Implications?
  - What is the difference between statistical testing and qualitative research?

Testing

- Various tests
  - t- and z-tests for two groups
  - ANOVA and variants for multiple groups
  - Regression analysis for modeling
- Also
  - Binomial test for distributions
  - CHI-Square test for tabular values
- Great on-line resources:
  - http://www.statisticshell.com/
  - http://www.statisticshell.com/html/limbo.html
  - Jacob Wobbrock's tutorial

Research Design

- Participants
  - Formerly “subjects”
  - Use appropriate number (e.g. similar to what others have used)
- Independent variable
  - What you manipulate, and what levels of iv were tested (test conditions)
- Confounding variables
  - Variables that can cause variation
  - Practice, prior knowledge

Research Design (2)

- Within subjects versus between subjects
  - Within = repeated measures
  - Sometimes a choice:
    - Controls subject variances (easier stat significance), but can have interference
- Counterbalancing
  - Typing on qwerty versus numeric keyboard
    - Could learn phrases, some phrases could be easier, so vary order of devices
  - Latin square
  - http://www.yorku.ca/mack/RN-Counterbalancing.html

Reading Experimental Results

- Sometimes you need to read carefully to fully appreciate what data is saying