CSE 599F
Research Methods and Data Analysis in Software Systems Research
Winter 2019

Experiment design and validity

January 10, 2019
Recap: the scientific method

Question → Observations → Hypothesis → Predictions

Broader theory
Recap: the scientific method

- Question
- Observations
- Hypothesis
- Experiment
- Predictions

Broader theory

Data collection and analysis!

Repeatable

Falsifiable
Today

- Discussion: Is computer science science?
- Experiment design and validity: Terminology
  - **Goal**: Collaboratively develop a set of definitions and a common understanding of important terms
  - **TPS**: Think, Pair, Share -> 3 groups (expert panels)
Is computer science science?

Paper discussion:

- CS = science, engineering, and mathematics.
- “CS is a grab bag of tenuously related areas thrown together”
- “CS is not a science, and its ultimate significance has little to do with computers”
- “Computing is not a science because it studies man-made objects”
- “Most scientific fields have saturated”
- “Science will never again yield revelations as monumental as the theory of evolution, general relativity, quantum mechanics, …”
- “Has computer science already made all the big discoveries it’s going to? Is incremental progress all that remains?”

- CS constantly forms new relationships with other fields => new fields.
- Overclaiming (empty promises) hurts the credibility of CS.
- Is the scientific method applicable to CS?
Is computer science science?

Paper discussion:

- CS = science, engineering, and mathematics.

"Latour defines science-in-the making as the processes by which scientific facts are proposed, argued, and accepted. A new proposition is argued and studied in publications, conferences, letters, email correspondence, discussions, debates, practice, and repeated experiments. It becomes a “fact” only after it wins many allies among scientists and others using it. To win allies, a proposition must be independently verified by multiple observations and there must be no counterexamples.

*Latour sees science-in-the making as a messy, political, human process, fraught with emotion and occasional polemics.*
Is computer science science?

Paper discussion:

- CS = science, engineering, and mathematics.
- "CS is a grab bag of tenuously related areas thrown together"
- "CS is not a science, and its ultimate significance has little to do with computers"
- "Computing is not a science because it studies man-made objects"
- "Most scientific fields have saturated"
- "Science will never again yield revelations as monumental as the theory of evolution, general relativity, quantum mechanics, …"
- "Has computer science already made all the big discoveries it’s going to? Is incremental progress all that remains?"
- CS constantly forms new relationships with other fields => new fields.
- Overclaiming (empty promises) hurts the credibility of CS.
- Is the scientific method applicable to CS?
Today

- Discussion: Is computer science science?
- **Experiment design and validity: Terminology**
  - **Goal**: Collaboratively develop a set of definitions and a common understanding of important terms
  - **TPS**: Think, Pair, Share -> 3 groups (expert panels)
Timeline and goals

1. **Group assignment (Think)**

2. **Breakout session (Pair)**
   a. Each group identifies relevant concepts and gives an explanation+example.
   b. **Goal:** identify and document relevant concepts related to the group’s focus.

3. **Regroup (Share)**
   a. Reshuffle groups and each group member explains her/his findings.
   b. **Goal:** convey identified concepts and comment on clarity.

4. **Class discussion (next time)**
   a. Review the final draft for missing/duplicate information.
   b. **Goal:** agree on drafted terminology.

3 Groups: Validity, Variables/Measures, Design
Group topics

Group 1: Validity
  ○ What are limitations of an experiment?
  ○ What could possibly invalidate an experiment?
  ○ What measures could improve validity?

Group 2: Variables/Measures
  ○ What types of variables exist?
  ○ What variables can we control vs. not?
  ○ What relationships between variables can we observe?

Group 3: Design
  ○ What are conceptual differences between experiments?
  ○ What subjects are studied and how?
  ○ What is a good data layout for what experiment design?
Experiments: a systematic approach
Two example experiments

Inputs

Controllable characteristics

pro'gram'mer
An organism that converts caffeine into code

Uncontrollable characteristics

Output

[Image of a coffee mug with text: I've cut back to just one cup of coffee a day!]

[Image of a code matrix]
Two example experiments

High-level research question:
Does coffee consumption improve programmer productivity and code quality?

Experiment 1:
- I program on project 1 for 20 weeks on Mondays with coffee.
- I program on project 2 for 20 weeks on Fridays without coffee.
- Code quality: number of defects encountered in each project.
- Productivity: number of lines of code written.
- Coffee consumption: dollars spent on coffee (Monday receipts).

Experiment 2:
- 20 participants, randomly assigned to two groups of 10.
- Each participant gets the same coding assignment.
- Code quality: number of defects encountered in each project.
- Productivity: number of lines of code written.
- Coffee consumption: Participants in group 1 get a free 64oz coffee.