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

Experiment design and validity
January 15, 2019

Recap: 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.

Recap: 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?
Recap: 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 (Share)
   a. Review the final draft for missing/duplicate information.
   b. Goal: agree on drafted terminology.

3 Groups: Validity, Variables/Measures, Design

Today

- **Discussion**: Should computer scientists experiment more?
- Experiment design and validity: **Terminology (continued)**
  - Goal: Collaboratively develop a set of definitions and a common understanding of important terms
  - TPS: Think, Pair, Share -> 3 groups (expert panels)
- **Discussion**: Class project

Should computer scientists experiment more?

Paper discussion:
1. Is computer science an experimental science?
2. What can we learn from the Knight-and-Leveson experiment?
3. Traditional scientific method isn't applicable.
4. The current level of experimentation is good enough (1998 vs. 2018).
5. Experiments cost too much.
6. Demonstrations will suffice (proof of concept is good enough).
7. There is too much noise in the way (the easy way out).
8. Progress will slow.
10. You'll never get it published.
11. Feature comparison is good enough (comparison on paper or verbally).
12. Trust your intuition.
13. Trust the experts.
14. Flawed experiments (unrealistic assumptions etc.).
15. Competing theories (RISC vs. CISC, OO vs. functional programming).