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

Course introduction
January 08, 2019

Today

- Brief introduction
- Your expectations
- Course overview
- My expectations
- Logistics
- The scientific method
- Some “classic” examples

My background

My research areas
- Software testing and verification
- Software debugging
- Software security
- Empirical software engineering
- Mining software repositories
My background

My research areas
● Software testing and verification
● Software debugging and repair
● Software security
● Empirical software engineering
● Data science / Applied ML

The Role of Software/Systems Engineering

Experimental infrastructure is software, too!
Example: Design space exploration

Infrastructure
Parallel executions of all possible configurations

|   | 1   | 2   | 3   | 4   | ...
|---|-----|-----|-----|-----|-----
| 1 | 0.34| 0.52| 0.21| 0.81| ... |
| 2 | 0.81| 0.32| 0.53| 0.52| ...
| 3 | 0.32| 0.22| 0.21| 0.81| ...
| 4 | 0.81| 0.32| 0.53| 0.22| ...

● 150 configurations, 1000+ benchmarks
● 1-85 hours per execution
● 200,000+ CPU hours (~23 CPU years)
### Your background and expectations

**Introduction and a very brief survey**

- **Field**: What is your area of research?
- **Stage**: What stage of the (PhD) program are you at?
- **Top-2 expectations**: What do you expect from this course?

---

### Course overview: the big picture

**Computer science as science**
- Reasoning about experimental designs, studies, and threats to validity.
- Correlation, causality, and confounding.
- Statistical vs. practical significance (effect size).
- Simpson’s paradox and censored data.

**The technical part and R tutorial**
- Hypothesis testing, confidence intervals, and bootstrapping.
- Parametric vs. non-parametric statistics.
- Modeling and parameter estimation.
- Data visualization and reporting.

**Research project**
- Related to empirical research/data analysis -- ideally in your field.
- You can team up or work alone.
Course overview: the big picture

- **Computer science as science** 2 paper reviews + 1 presentation
  - Reasoning about experimental designs, studies, and threats to validity.
  - Correlation, causality, and confounding.
  - Statistical vs. practical significance (effect size).
  - Simpson’s paradox and censored data.

- **The technical part and R tutorial** 2 in-class exercises
  - Hypothesis testing, confidence intervals, and bootstrapping.
  - Parametric vs. non-parametric statistics.
  - Modeling and parameter estimation.
  - Data visualization and reporting.

- **Research project** Entire quarter
  - Related to empirical research/data analysis -- ideally in your field.
  - You can team up or work alone.

Questions?

Course overview: grading

- 40% Class project
- 20% In-class exercises (2 sessions)
- 20% Paper reviews (2 papers)
- 10% Paper presentation (1 paper)
- 10% Participation

Questions?

Course overview: the even bigger picture

This course
- is experimental!
- is feedback-driven -- your input matters!
- is not a comprehensive summary of statistical methods.
- is not a replacement for various statistics courses.

Other (UW) resources
- INFO 270: Calling Bullshit: Data reasoning in a digital world
  [https://callingbullshit.org](https://callingbullshit.org)
- Practical Statistics for HCI
  [https://depts.washington.edu/madlab/proj/ps4hci/](https://depts.washington.edu/madlab/proj/ps4hci/)
- Statistical Analysis and Reporting in R

My expectations
- Reading and reviewing research papers.
- Presenting a research paper.
- Conducting a quarter-long data analysis project.
- Working with the R programming language.
- Some programming experience.
- Have fun!
Logistics

- CSE 305, Tu/Th, 10:30am – 11:50am.
- Lectures, presentations, and lab session.
- Course material, schedule, etc. on website: https://homes.cs.washington.edu/~rjust/courses/2019Winter/CSE599
- Submission of assignments via Canvas: https://canvas.uw.edu
- Discussions Piazza https://piazza.com

First paper readings and project proposal

- **01/10**: Paper reading: *Is computer science science?*
- **01/15**: Paper reading: *Should computer scientists experiment more?*
- **01/15**: informal project proposal.

Today

- Brief introduction
- Your expectations
- Course overview
- My expectations
- Logistics
- **The scientific method**
- Some “classic” examples

The scientific method

[Question]
The scientific method

Observations ➔ Question

Hypothesis ➔ Experiment ➔ Predictions

Seems quite simple. What’s important?

Broader theory

Falsifiable

Broader theory
The scientific method

- Question
- Observations
- Hypothesis
- Experiment
- Predictions
- Data collection and analysis
- Falsifiable
- Broader theory

My favorite quotes

Collaborators, students, reviewers:

- These results are bad and cannot be true.
- I trust my intuition, no need to run experiments.
- These results are entirely expected.
- I have computed all the data; which statistical test should I use to show that my results are significant?
- Most papers are wrong or later obsolete, so who cares?
- I don’t understand these intervals, can you give a p value?

"If you torture the data long enough, it will confess." [Ronald Harry Coase]
My favorite quotes

Collaborators, students, reviewers:

- These results are bad and cannot be true.
- I trust my intuition, no need to run experiments.
- These results are entirely expected.
- I have computed all the data; which statistical test should I use to show that my results are significant?
- Most papers are wrong or later obsolete, so who cares?
- I don’t understand these intervals, can you give a p value?

Avoid confirmation bias; always try to prove yourself wrong (regardless of whether the results are good or bad).

"Statistical significance is the least interesting thing about the results"
[Sullivan and Fein: Using effect size -- or why the p value is not enough]
### A classic example: correlation vs. causation

**Observation:**
Murder rates and ice cream sales are strongly positively correlated.

**Conclusion:**
Eating ice cream is really dangerous.

**Possible explanations?**
- Resurrected zombies primarily feed off ice cream
- Excessive ice cream consumption makes others jealous

*The weather is a non-controlled confound!*

### Another example: errors and baselines

Neural correlates of interspecies perspective taking in the post-mortem Atlantic Salmon: An argument for multiple comparisons correction

Yet another example: study design

Goal:
Studying the relationship between time spent on studying Java and success rate in completing coding assignment.

Methodology:
- 75 participants are randomly selected in front of CSE.
- Each participant is given a high-level overview of the study.
- Each participant decides on how long to study before attempting to solve any coding assignment.
- Each participant solves as many coding assignments as possible in one hour (after studying).

Yet another example: conclusions

Conclusion: Spending more time on learning Java makes you a worse Java programmer.

Yet another example: what’s wrong?

Goal:
- Comes up with a testable hypothesis about the data.
- 2 methodology questions.

Yet another example: Simpson’s paradox

This phenomenon is called: Simpson’s paradox.