Warmup Exercise

• Think about your current or upcoming research

• Write down a hypothesis you would like to test in that project

• Turn to your neighbor and share

• Briefly discuss whether the hypothesis is “good”

• Brief group discussion on exercise
Let’s actually begin at the beginning...

• A good research project starts with a good hypothesis

• It is not uncommon for the hypothesis to “morph” as investigators perform research
JELLY BEANS CAUSE ACNE!

SCIENTISTS! INVESTIGATE!

BUT WE'RE PLAYING MINECRAFT!

...FINE.

WE FOUND NO LINK BETWEEN JELLY BEANS AND ACNE (P > 0.05).

THAT SETTLES THAT.

I HEAR IT'S ONLY A CERTAIN COLOR THAT CAUSES IT.

SCIENTISTS!

BUT MINECRAFT!
We found no link between purple jelly beans and acne ($P > 0.05$).

We found no link between brown jelly beans and acne ($P > 0.05$).

We found no link between pink jelly beans and acne ($P > 0.05$).

We found no link between blue jelly beans and acne ($P > 0.05$).

We found no link between teal jelly beans and acne ($P > 0.05$).

We found no link between salmon jelly beans and acne ($P > 0.05$).

We found no link between red jelly beans and acne ($P > 0.05$).

We found no link between turquoise jelly beans and acne ($P > 0.05$).

We found no link between magenta jelly beans and acne ($P > 0.05$).

We found no link between yellow jelly beans and acne ($P > 0.05$).

We found no link between grey jelly beans and acne ($P > 0.05$).

We found no link between tan jelly beans and acne ($P > 0.05$).

We found no link between cyan jelly beans and acne ($P > 0.05$).

We found no link between green jelly beans and acne ($P > 0.05$).

We found a link between lime jelly beans and acne ($P < 0.05$).

We found no link between malv jelly beans and acne ($P > 0.05$).

We found no link between beige jelly beans and acne ($P > 0.05$).

We found no link between lilac jelly beans and acne ($P > 0.05$).

We found no link between black jelly beans and acne ($P > 0.05$).

We found no link between peach jelly beans and acne ($P > 0.05$).

We found no link between orange jelly beans and acne ($P > 0.05$).
News

Green Jelly Beans Linked to Acne!

95% Confidence

Only 5% Chance of Coincidence!

Scientists...
That’s Not What We Want!

• It’s understandable how this happens

• Here’s a pattern that can happen:
  • Think of a new “thing” you want to try (e.g. an attack, a defense, …)
  • Do the “thing” on some code you have easily accessible
  • Notice at the end of the test “Hey - that thing seemed to work well!”
  • Use that test as a “baseline”
  • Do the new “thing” on another piece of code
  • Write up the whole thing for a conference

What are some problems with this approach?
Problems

• Introduces bias

• “Fishing”
  • Violates assumptions of most statistical tests
  • Increases family-wise error rate
  • Could lead to spurious conclusion

• Might not have all the required info/data

• Lack proper study setup
Why Does This Happen?

• Find something interesting and unexpected that was not the original study goal

• Pressure to get papers out

• Fear of publishing bad results

• Lack of time for more planning
So, Let’s Not Let This Happen

• Start with refining your hypothesis

• What makes a good hypothesis?
A Good Hypothesis Is... 

- Interesting
- Appropriately Scoped
- Repeatable
- Answerable
- Measurable
Hypotheses

• **Null Hypothesis**
  • Hypothesis that there is no difference between the treatments/groups

• **Alternate Hypothesis**
  • Hypothesis that there is a difference between the treatments/groups
  • One-tailed or Two-tailed
Example:

**Goal**

The goal is to help security experts to focus effort on OSS code changes likely to contain vulnerabilities by identifying attributes that indicate the presence of those changes.
Example:
Hypothesis

$H_{10}$ - The amount of code churn in files/patchsets that contain vulnerabilities is the same as in other files/patchsets

$H_{1A}$ - The amount of code churn in files/patchsets that contain vulnerabilities is larger than in other files/patchsets
Example:
Hypothesis

$H_{20}$ – A code author’s experience (measured by contributions) has no impact on the likelihood of inserting vulnerabilities

$H_{2A}$ – A code author’s experience (measured by contributions) reduces the likelihood of inserting vulnerabilities
Look at Your Hypothesis

• Is there a larger problem you are trying to address? (Ex. a general system characteristic)

• Is there a particular type/category of thing that you are interested in? (Ex. A particular type of attack)

• What is your main variable that you are adjusting? (Ex. Which tool is used)

• What population are you interested in studying? (Ex. Students, professionals, …)

• What other questions can you think of?
Goal-Question-Metric

• Think “top-down” rather than “bottom-up”

• “What do I really want to know” vs. “What data can I collect/analyze”?

• Helps focus on the underlying scientific question
Goal-Question-Metric

Conceptual Level
• The overarching goal/hypothesis

Operational Level
• Questions to help determine whether your hypothesis is supported

Quantitative Level
• The data you need to answer the questions to test the hypothesis
Goal-Question-Metric

Goal/Hypothesis <-> Tested Hypothesis

Questions <-> Answer

Metrics <-> Measurement

Data Collection
H1A - The amount of code churn in files/patchsets that contain vulnerabilities is larger than in other files/patchsets

How much churn is in each patchset?

Which patchsets contain vulnerabilities?

# lines changed/added/removed in a file

Presence of a vulnerability
Examine Your Hypothesis

- What is the overarching goal?
- What questions should you try to answer?
- What metrics do you need to answer the questions?
- Are there specific data items you need?
- How might you collect them?
Examine Your Hypothesis

• Individual Activity
  • Rework your hypothesis
  • Is it Interesting, Answerable, Repeatable, Measurable, and Appropriately Scoped?
  • Write some questions and related metrics

• Group Activity
  • In groups of 3-4
  • Share your hypothesis, questions, and metrics
  • Other members provide feedback
Acknowledgements

• NC State SoSL Lablet

• NSF DEERS (Designing Empirical Education Research Studies) project
  • NSF-1525028, 1525173, 1525373
  • Mark Sherriff
  • Sarah Heckman