Today

- The scientific method.
- Internal, external, and construct validity.
- Reasoning about two empirical studies.
- Paper discussion:
  Views on Internal and External Validity in Empirical Software Engineering
The scientific method

Observations → Hypothesis → Experiment → Predictions

Question

Falsifiable

Data collection and analysis!

Repeatable

Falsifiable

The scientific method

Observations → Hypothesis → Experiment → Predictions

Question

Repeatable
The scientific method: common mistake

Data collection → Observations → Hypothesis → Data Analysis → Predictions

"If you torture the data long enough, it will confess."
(Ronald Harry Coase)

Internal, external, and construct validity

**Internal validity**
How well does the experimental design isolate the effect/variables that it studies (i.e., control for confounds)?

**External validity**
How well does the experimental design generalize to the real world (i.e., other populations, situations, etc.)?

**Construct validity**
How well does the experimental design measure what it is supposed to measure? Does it use the right metrics and collect the right measurements?
### Internal validity: a classic example

**Internal validity**
How well does the experimental design isolate the effect/variables that it studies (i.e., control for confounds)?

**Classic example**
Murder rates and ice cream sales are highly positively correlated. Possible explanations?

- Possibilities:
  - Resurrected zombies primarily feed off ice cream
  - Excessive ice cream consumption makes others jealous

*Actually, the weather is a non-controlled confound!*

---

### Threats to validity: example experiment

**Research question:**
Does coffee consumption improve code quality?

**Methodology**
- I program on project 1 on Mondays with coffee.
- I program on project 2 on Fridays without coffee.
- Measure code quality in number of defects I encounter.
- Measure coffee consumption in dollars spent on coffee beans, as listed on my grocery-shopping receipt.
Threats to validity: example experiment

Research question:
Does coffee consumption improve code quality?

Methodology
- I program on project 1 on Mondays with coffee.
- I program on project 2 on Fridays without coffee.
- Measure code quality in number of defects I encounter.
- Measure coffee consumption in dollars spent on coffee beans, as listed on my grocery-shopping receipt.

What are threats to construct, internal, and external validity?

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 LGRT.
- 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).

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

Overall results

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

Something is fishy… Is there a dead salmon in here somewhere?
This phenomenon is called: **Simpson’s paradox.**