Software Engineering Experimentation

Experimental Terms

Jeff Offutt

http://www.ise.gmu.edu/~offutt/

Do We Know This Already?

- We saw the basics of experimental design in 5th grade
  - Not many details
  - Many of us were not paying attention
- College students in science learn this in their freshman year
  - Physics, chemistry, biology, …

A major failing of CS is that we do not teach this material

Descriptive Research

Used to discover trends and tendencies

- *Observational studies*: systematic measurement of behavior
  - interrater reliability: degree to which independent observers agree on their coding of data
- *Archival studies*: examine records of past events and behaviors
- *Surveys*: asking questions about attitudes, beliefs, and behavior

Scientific Experiment

Used to understand effects

- Changes a set of variables to elicit a response
- Imposes a *treatment* on a group of *objects* or *subjects*
  - Treatment defines a way to change variables

Correlational Research

Used to establish associations between variables

- *Correlation coefficient*: statistical measure of the strength and direction of association between two variables (varies between –1.0 and +1.0)
  - *Positive correlation*: As one variable increases the other also increases
  - ***N**egative correlation*: As one variable increases the other decreases

Basic Experimental Terms

- *Hypothesis*: A testable prediction about the conditions under which an event will occur
- *Theory*: An organized set of principles used to explain observed phenomena
- *Operational Definition*: A specific way in which a variable is measured or manipulated (*treatment*)

Variables

- *Independent variable* : manipulated by the researcher to determine if it causes a change in the dependent variable
  - Also called *factor*
- *Dependent variable* : measured by the researcher to determine if it is affected by the IV
- *Confounding variables* : alternative explanations for the results
- *Measured variable* : If the dependent variable cannot be directly measured, we measure a related variable to approximate

Validity

- *Internal validity* : degree to which there is certainty that the IV caused the effects on the DV
- *External validity* : degree to which the results from a study can be generalized to other situations and people
- *Conclusion validity* : degree to which conclusions relationships in the data are reasonable
- *Construct validity* : degree to which inferences can be made from the specific objects in your study to the theoretical constructs on which those objects were based

Experimental Bias

- *Bias*: A flaw in the experimental design or conduct that can change the dependent variable
  - This is often due to an inadvertent introduction of a confounding variable
- *Bias (psychology)*: A flaw introduced by an experimenter whose expectations about the outcome of the experiment can be subtly communicated to the participants in the experiment
  - Often happens when experimenters are also subjects

Example

Data flow testing finds more faults than branch testing

- Independent Variables: Data flow, branch testing
- Dependent Variable: Faults found
- Confounding Variables: tool support, characteristics of subjects, specific values chosen, knowledge of testers, …
  - Effects the internal validity
- Bias: If I invented data flow, I expect it to do better

Correlation and Causality

- *Correlated*: Two things always happen at the same time
  - Brake lights and car slowing down
- *Causality*: Understanding what causes something to happen
  - Brake light causes the car to slow down
- If A and B are correlated:
  - A causes B
  - B causes A
  - C causes A and B
  - Pressing brake activates brake light AND slows car down

Confusing Correlation and Causality

- In "*the old days*", we believed that being cold caused us to get colds
- Colds are caused by viruses, not temperature
- Viruses breed very well in warm, damp, low-oxygen, carbon-dioxide rich environments
- When the weather turns cold, we often close up our houses and turn up the heat … creating …
- In Virginia, we have a secondary cold season in July-August … when the weather turns hot and humid …

Cognitive Dissonance

- We feel uncomfortable when new data or a new model contradicts a previously held model
- Revising our mental model to accommodate new data is hard
  - We resist the new idea

Correlation and Prediction

- Correlation: if A happens, then B happens
  - Brake lights and car slowing down
- Causality: if A happens, then it causes B to happen
  - Pressing brake slows the car down
- *Predictability*: if A happens, I can predict that B will happen

We do not need to show causality to have predictability

Experimental Design

- Choosing variables, subjects, objects, process and analysis method
- *Pilot study*: small-scale experiment used to design the full experiment
  - Identify potential confounding variables
  - Refine experimental design

Avoiding Bias in Experimental Design

- *Control*: Ensuring the confounding variables do not influence the results
  - I want to measure whether maintenance programmers understand programs better by studying statecharts or reading comments
  - Comments already existed in the program, statecharts generated by experimenter
  - Statecharts were of much higher quality
  - Programmers understood statecharts better …
- Must control for differences in quality

Placebos in Experimental Design

- Patients expect a treatment to work, and sometimes respond to a treatment that has no effect – like a sugar pill
- Programmers in an experiment will do better because they think they should – they concentrate better or work harder
  - Give some programmers a tool that actually does not work
- *Double-blind*: Neither the programmers or the experimenter knows who has the placebo

Hawthorne Effect

- When people know they are being studied, they will behave differently
- How can we study people without letting them know they are being studied?

Avoiding Errors of Judgment

- *Randomization*: Objects are assigned randomly to experimental groups
  - *Randomized block design*: Divide subjects into homogeneous blocks, then randomly assign from each block
    - Programmers: undergraduate students, MS students, PhD students, professional
- *Replication*: perform the experiment again, with different subjects, experimenters, or experiment design
  - Most reviewers will not accept replicated experiments