

PhD Seminar

The Scientific Method

Jeff Offutt

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

Measuring and Science

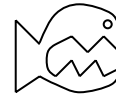
When you can measure what you are speaking about, and express it in numbers, you know something about it.

– Lord Kelvin, 1889

<http://zapatopi.net/kelvin/quotes.html>

What is a Scientific Test

- The Budweiser Test
 - Drinkers of another brand were given a “live” challenge
 - which beer is better?
 - Results?
 - 50% chose Budweiser!!!
 - Conclusion:
 - Budweiser is better !!!
- Hmmm ... something's fishy ...



Scientific Test

- Test: Live TV, lots of noise and confusion.
- Subjects wouldn't be able to tell any difference, so we should expect each beer to be chosen ...
- Half the time!
- There are three kinds of lies ...

Lies, Damn Lies, and Statistics



Berkshire Eagle, October 7, 1993 page A3 (An AP story from Boston)

Guns in the home found to increase risk of death

- People who keep guns at home nearly triple their chances of being murdered, usually by friends or relatives, but fail to protect themselves from intruders ...
- The article goes on to describe how the study was conducted, summarizes aspects of the population cross sections and conclusions of the study, and concludes with a refutation by a representative of the NRA:
- However, Paul Blackman, research coordinator at the National Rifle Association, criticized the study ...
 - "These people were highly susceptible to homicide," he said.
 - "We know that because they were killed."

© Jeff Offutt, 2004

5

Ways to Acquire Knowledge



1. Tenacity
2. Intuition
3. Authority
4. Rationalism
5. Empiricism
6. Science

© Jeff Offutt, 2004

6

1. Tenacity

Knowledge based on superstition or habit

- Examples:
 - “Good research can only be done by those in their 20s”
 - “OO design has too many subroutine calls and is too inefficient”
- Exposure: The more we see something, the more we like it
- Tenacity has
 - No guarantee of accuracy
 - No mechanism for error correction

2. Intuition

Guesswork: An approach that is not based on reasoning or inference

- No mechanism to separate accurate from inaccurate knowledge
- Can be valuable as a way to suggest hypotheses
- Can be very misleading

3. Authority

Accepted because it comes from a respected source

- Examples:
 - Religion
 - Totalitarian government
 - Rules our parents taught us
- No way to validate or question the knowledge
- Not the same as asking an expert – we can accept, reject, or challenge an expert
 - Teachers are experts, not authorities

4. Rationalism (Reasoning)

Acquisition of knowledge through reasoning

- Logical deduction
- Assume knowledge is correct if the correct reasoning process is used
- Middle ages relied almost exclusively on rationalism
- Important for theory and pure math
 - Theoretical physics ... experimental physics
- Easy to reach incorrect conclusions
- Use rationalism to arrive at a hypothesis, then test with the scientific method

5. Empiricism

Acquiring knowledge through experience

- “I have experienced it, therefore it is true”
- Experience is subjective and hard to control
- “I wrote 3 programs without designing and they worked – designs are worthless!”
 - Who wrote them?
 - What programs?
 - Was the the design present but just unwritten?
- Much of computer “science” is just empiricism

6. Science

Testing ideas empirically according to a specific testing procedure that is open to public inspection

- Based on reality
- Devoid of personal beliefs, perceptions, biases, attitudes, emotions
- Based on objectively observed evidence

Scientific Method

1. Identify a problem & form hypothesis
 - Problem must be clear, precise, measurable
 - Hypothesis must be testable and refutable
2. Design the experiment
 - The most creative part
3. Conduct the experiment
4. Perform hypothesis testing
 - Analyze data with appropriate statistics
5. Dissemination
 - Write legible papers and teach classes

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: specific way in which a variable is measured or manipulated

Variables and Validity

- Variables
 - Independent variable: manipulated by the researcher to determine if it causes a change in the dependent variable
 - Dependent variable: measured by the researcher to determine if it is affected by the IV
 - Confounding variables: alternative explanations for the results
- 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

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
 - Pressing brake slows the car 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

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

Statistical Tests and Software

- Experimental data based on programs cannot, with validity, be subjected to inferential statistical tests since the population is unknown
- An unknown population nullifies any statistical result that would be obtained, regardless of the number of programs
- Only descriptive statistics can be used
- That is, statistical hypothesis testing, at least in the statistical sense, is not accurate

Be Problem Solvers

- As Computer Science / Software Engineering majors, you have proven yourselves to be good problem solvers
- Much of life is about solving problems
- Education is not about skills, it is about knowledge
- Use your education knowledge to help you:
 - Think rationally
 - Question authority
 - Solve all of life's problems

A Suggested Categorization of Empirical Research Studies

IT 821, Fall 2001

1. Demonstration of Concept
2. Demonstration of Feasibility
3. Controlled Case Study
4. Practical Field Study
5. Empirical Analysis
6. Limited Lab Experiment
7. Realistic Experiment
8. Survey

Demonstration of Concept

An idea is tried. An implementation is built, a process is followed, or a technique is used. This should demonstrate that the concept can actually be used.

Summary: This idea can be applied.

Demonstration of Feasibility

An idea is evaluated in some way, either by trying it and analyzing the results or comparing the results with a traditional method. This does not necessarily control variables or attempt to generalize to a broad range of cases.

Summary: This idea can work and be competitive with other ideas.

Controlled Case Study

An idea is used in a limited situation but using realistic assumptions. Variables are not controlled, in fact, a bias might be intentionally included to indicate under what situations the idea can be useful. This does not necessarily generalize to a broad range of cases, but can give hope that the idea is useful.

Summary: The idea works, at least in some situations.

Practical Field Study

An idea is used in a practical, realistic situation to help produce a real product. This has all the characteristics of a case study, but is on a large scale.

Summary: The idea works in practice.

Empirical Analysis

Data is collected and analyzed, with care and attention to potential biases. There does not necessarily have to be an a priori hypothesis, and the data may or may not be based on realistic subjects.

Summary: To analyze current practice or theory and to generate new directions for empirical research.

Limited Lab Experiment

A hypothesis is generated, variables carefully identified and controlled, and data is collected in a scientific manner. The data is analyzed to evaluate the hypothesis. Detailed statistical analysis may or may not be necessary. The experiment may suffer from restrictions such as size of artifacts, number of subjects and artifacts, and technique used. The hypothesis may or may not be of broad applicability.

Summary: To give strong evidence that the hypothesis is true.

Realistic Experiment

Much like a Limited Lab Experiment, but using realistic artifacts. A hypothesis is generated, variables carefully identified and controlled, and data is collected in a scientific manner. The data is analyzed to evaluate the hypothesis, perhaps using detailed statistical analysis. The artifacts and subjects should be typical for the field of study -- e.g., professional programmers and actual software products. The hypothesis may or may not be of broad applicability.

Summary: To give convincing evidence that the hypothesis is true.

Survey

A "meta-empirical" paper. A paper about experimentation, about experimental papers, or that summarizes a group of experimental papers.

Summary: To educate about experimentation or to coalesce results from a number of related studies.

Software Metrics

- Why measure software?
 - Quality evaluation
 - Planning for change
 - Resource estimation
 - Improvements in software and process
 - Guide to design and implementation methods

Terms

- Factor: A quality that we wish to base a decision on but that cannot be measured directly
 - Maintainability, complexity, usability
- Metric: Something that can be measured
 - Height, age, size, coupling, time to learn

It is easy to get factor and metric confused. We use metrics to estimate factors. We often use the same term at different levels.

Measuring Metrics

- Measurement: A specific way to evaluate a metric. A measurement has units.
 - Number of inches/cm/feet, years, lines, ...
- Validation: A metric is validated if it tells us something useful ... that is, if it has been shown to accurately estimate a factor
 - Coupling has been shown to be correlated with maintainability and reliability

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