This course...

http://www.cs.montana.edu/courses/esof522/

Book References:
- Wholin et al. “Experimentation in Software Engineering”
- Mens and Demeyer, “Software Evolution”
- Juristo and Moreno “Basics of Software Engineering Experimentation”
- Yin “Case Study Research”

Note References:
- Byron Williams, Mississippi State University
- Technical University of Munich
- Steve Eastbrook, University of Toronto
Format

Lectures:
- 2 classes per week
- Mix of discussions, lectures, presentations

Readings:
- Assigned readings from the book, journals, conference papers, reference books.
- This is a major component of our weekly discussions
- You will be randomly asked to lead a discussion

Assessment
- Project (50%)
- Homework (20%)
- Discussions and class participation (10%)
- R Programming (20%)
A replication study
A new experiment validated against real world data
Perform a simulation (a model/hypothesis)
A case study

Places to look for papers:
- SQJ (Software Quality Journal),
- ESEM (Empirical Software Engineering Measurement)
Introduction

How do we know that claims made by software engineers are true?

Do computer scientists understand the scientific process?

How do we compare tools, processes, techniques?
Introduction

“One of the hallmarks of software becoming an engineering discipline is to be able to lay aside perceptions, bias, and market-speak to provide fair and impartial analysis and information.”

We achieve this through experimentation:

- Observation
- Look for quantitative relationships
- Look for causality, not correlation alone
- Acceptance or rejection of hypotheses
- Replication of studies
<table>
<thead>
<tr>
<th>Fallacy</th>
<th>Rebuttal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traditional scientific method isn’t applicable.</td>
<td>To understand the information process, computer scientists must observe phenomena and formulate and test explanations. This is the scientific method.</td>
</tr>
<tr>
<td>The current level of experimentation is good enough.</td>
<td>Relative to other sciences, the data show that computer scientists validate a smaller percentage of their claims.</td>
</tr>
<tr>
<td>Experiments cost too much.</td>
<td>Meaningful experiments can fit into small budgets; expensive experiments can be worth more than their cost.</td>
</tr>
<tr>
<td>Demonstrations will suffice.</td>
<td>Demos can provide incentives to study a question further. Too often, however, these demos merely illustrate a potential.</td>
</tr>
<tr>
<td>There’s too much noise in the way.</td>
<td>Fortunately, techniques can be used to simplify variables and answer questions.</td>
</tr>
<tr>
<td>Experimentation will slow progress.</td>
<td>Increasing the ratio of papers with meaningful validation has a good chance of actually accelerating progress.</td>
</tr>
<tr>
<td>Technology changes too fast.</td>
<td>If a question becomes irrelevant quickly, it is too narrowly defined and not worth spending a lot of effort on.</td>
</tr>
<tr>
<td>You will never get it published.</td>
<td>Smaller steps are still worth publishing because they improve our understanding and raise new questions.</td>
</tr>
</tbody>
</table>
Introduction

Research

- Finding good research questions
- Theory building
- Research Design
- Ethics
- Evidence and Measurement
- Sampling
- Peer Review Process

2004-5 Steve Eastbrook
Introduction

Two kinds of complementary analysis methods in empirical studies:

- Quantitative (and the focus of this class)
  - Establish relationships between variables
  - *Explanatory* studies
  - Collect numerical measurements
  - Use statistics to show significance

- Qualitative
  - Gain a holistic view of the problem
  - Analysis is done using “words” (sometimes nominal scales)
  - No formal procedure for getting objective conclusions
  - Closer to the subjects and thus are better for *exploratory* studies in the context of their setting
  - Better when it is hard to come up with a hypothesis
Empirical Studies: Types of Data

Types of Data

Quantitative
- Controlled Measures
  - Tend to be more objective
- Verification Oriented

Qualitative
- Naturalistic and Uncontrolled
  - Tend to be more subjective
- Discovery Oriented

Byron Williams, Mississippi State University
Introduction

Lab Experiment (In Vitro)

Real World (In Vivo)

Historical data collection

Real World (In Vivo)

Software Engineer

Researcher

Controlled environment

Case Study (quasi-experiment) by Beta adopters

Observational Case Study (survey)

Programmers
Introduction

Theory → World

Manipulation
Observation

Validation ??

2004-5 Steve Eastbrook
The Scientific Approach:

Observations accurately reported to others (Data plays a central role) +

Open exchange and competition among ideas (Science is adversarial) +

Search for discovery and verification of ideas (Scientists are not alone) +

Peer review of research

Provides an objective set of rules for gathering, evaluating, and reporting information

Byron Williams, Mississippi State University
Why Experiment?

There are no opinions in scientific research. Truth and thus validity must be validated against reality.

Established laws, theories can be used to predict behaviors on new artifacts that we build.

In Software Engineering there are no established relationships...

To date, the results of software construction cannot be predicted by the body of SE knowledge.
Theories

In science a theory is defined as:
- A set of statements that explain observations
- Correlations, causal relationships of variables
- The best explanation of all the available evidence.

A hypothesis is not a theory. It is a testable statement derived from a theory
- A conjecture, a model

We do not have theories in software engineering

2004-5 Steve Eastbrook
Theories

There are 3 reasoning modes for arriving at a hypothesis:

- Deduction
- Induction
- Abduction
Hypothesis Formation

REAL WORLD

DATA
(facts, phenomenon)

HYPOTHESES
(conjectures, models, theories ...)

THOUGHT

induction

abduction

deduction

abduction

induction

deduction

induction

...
Hypothesis Formation

In Software Engineering we are looking for relationships between variables that help predict:

- Process output, and
- Product output
- Causality
- Quality attributes of software artifacts
Relation to Other Fields: **Physics**

**Goal**
- Understand and predict the behavior of the physical universe

**Actors**
- **Theorists** (researchers) build models to explain the universe
  - Predict the results of events that can be measured
  - Models based on the theory about the essential variables and their interaction data from prior experiments
- **Experimentalists** (researchers) observer, measure, experiment to:
  - Test or disprove a hypothesis or theory
  - Explore a new domain

**Beginning**
- Early experimentalists only observed, did not manipulate the objects

**Cycle of Evolution**
- Progressed due to the interaction of the two types of researchers
- Regardless of where the cycle starts it follows the same pattern: Model, Experiment, Learn, Remodel, ...

**Current State**
- Modern physicists have learned to manipulate the physical universe, e.g. particle physics
Relation to Other Fields: Medicine

Goal
- Understand the human body and how it functions

Actors
- **Researcher** aims at understanding the workings of the human body to predict the effects of various procedures and drugs
- **Practitioner** applies knowledge by manipulating processes on the body for the purpose of curing it

Beginning
- Began as an art form
- Evolved as a field when it began observation and model building

Cycle of Evolution
- Knowledge built by feedback from practitioner to researcher
- Clear relationship between the two classes of actors
- Experimentation
  - From controlled experiments to case studies
  - Human variance causes problems in interpreting results
  - Data may be hard to acquire

Current State
- Knowledge of the human body continues to evolve
Replication

- Knowledge becomes science when it can be verified.
- Observations cannot be taken seriously unless repeated.
- Experimental results must be replicable by external independent agents.
- Scientists and engineers consider repeatability as a critical test to be passed by any new knowledge.
Replication

Even though it is impossible to replicate an environment and very difficult to even find similar conditions...

The manner in which we design experiments means that conclusions can be drawn about data.

Replication is used to confirm a hypothesis, otherwise we use variation.

Which variables can we change/alter?
Can we generalize from variations?
Replication

- Identical
- Differentiated
  - Variations:
    - Site
    - Experimenters or Subjects
    - Design chosen for the experiment
    - Instrumentation
    - Variables
Empirical Strategies

Survey

- Often performed in retrospect
- Cannot manipulate variables
- Primary method for gathering qualitative or quantitative data is through questionnaires and interviews
- Performed on samples of a population. The goal is to understand the population
- Results are analyzed to derive descriptive, explanatory or exploratory conclusions
Empirical Strategies

Systematic Literature Review

- Identification, analysis and interpretation of all available evidence related to a specific research question
- Rigorous undertaking that follows a strict protocol
- Rationale, research questions, search strategy for studies, selection criteria of studies, study quality assessments, data extraction strategy, dissemination, etc.
Empirical Strategies

Mapping Study

- Performed when the field of study is broader, or when not a lot of research done already.
- Follows the same process as a systematic literature review but has different criteria for inclusion of studies.
- Uses qualitative data to go beyond just descriptive statistics.
- More generic than systematic literature reviews.
- Broad in scope. They focus on a research area, rather than the more precise focus on a specific process or tools, etc.
Empirical Strategies

Case Studies
- An empirical inquiry that draws (possibly) on multiple sources of evidence to investigate a phenomenon in its context
- Typically aimed at tracking a specific attribute or establishing a relationship
- Results are analyzed to derive explanatory or exploratory conclusions
- Level of control is less than in an experiment
- Case studies are observational not controlled (like experiments)
- Easier to plan and more realistic than experiments but much harder to generalize
- Difficult to deal with confounding factors since we have little control
Empirical Strategies

Case Study Examples:

- Compare the results of using a new process/tool against a company baseline
- Compare a new process/tool against a sister project that is currently under use
- A new process/tool is applied to some components randomly, but not others (quasi-experiment)
- Study/observe the evolution of attributes (over time) in a code base (longitudinal)
Empirical Strategies

History
- Preferred when the researcher has no access or control
- Deal with the “dead” past, when no relevant persons are alive to respond, even retrospectively
- Researcher can only rely on documents, cultural and physical artifacts
- The difference with a case study is that the researcher cannot perform direct observation of events
Empirical Strategies

Ethnography
- Not a case study!
- Requires participant observation
- Requires long periods of time in the “field” and emphasize detailed observational evidence
- Large investment in field efforts
- In contrast, case studies can be done entirely from your office or library
Empirical Strategies

Experiments

- Done *in-vitro*
- An empirical inquiry that manipulates a variable
- Based on randomization, different treatments are applied by subjects while keeping other factors constant
- It is systematic
- High level of control of one or more variables
- Results are analyzed to derive explanatory conclusions
Empirical Strategies

Quasi-Experiments

- Done in-vitro
- Similar to an experiment, but the assignment of treatments by subjects cannot be based on randomization, but emerges from the characteristics of the subjects
- Results are analyzed to derive explanatory conclusions
## Major Strategy Comparisons

<table>
<thead>
<tr>
<th>Factor</th>
<th>Survey</th>
<th>Case Study</th>
<th>Experiment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Execution Control</td>
<td>NO</td>
<td>NO</td>
<td>YES</td>
</tr>
<tr>
<td>Measurement Control</td>
<td>NO</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Investigation Cost</td>
<td>Low</td>
<td>Medium</td>
<td>High</td>
</tr>
<tr>
<td>Ease of Replication</td>
<td>High</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Type</td>
<td>Descriptive</td>
<td>Exploratory, Explanatory (in context)</td>
<td>Explanatory</td>
</tr>
</tbody>
</table>

Wohlin et al., Experimentation in Software Engineering, 2012
Yin, Case Study Research.