Empirical investigation

- Software engineering investigation
- Investigation principles
- Investigation techniques
- Formal experiments: Planning
- Formal experiments: Principles
- Formal experiments: Types
- Formal experiments: Selection
- Guidelines for empirical research

Empirical SE

- Fill the gap between research and practice by:
  - **Developing methods** for studying SE practice
  - **Building a body of knowledge** of SE practice
  - **Validating research** before deployment in industrial settings

SE Investigation

- What is software engineering investigation?
  – Applying “scientific” principles and techniques **to investigate** properties of software and software related tools and techniques.
- Why talking about software engineering investigation?
  – Because the standard of empirical software engineering research is quite poor.

SE Investigation: Examples

- **Experiment to confirm rules-of-thumb**
  - Should the LOC in a module be less than 200?
  - Should the number of branches in any functional decomposition be less than 7?
- **Experiment to explore relationships**
  - How does the project team experience with the application affect the quality of code?
  - How does the requirements quality affect the productivity of the designer?
- **Experiment to initiate novel practices?**
  - Would it be better to start OO design with UML?
  - Would the use of SRE improve software quality?

SE Investigation: Why?

- To improve (process and/or product)
- To evaluate (process and/or product)
- To prove a theory or hypothesis
- To disprove a theory or hypothesis
- To understand (a scenario, a situation)
- To compare (entities, properties, etc.)

SE Investigation: What?

- Person’s performance
- Tool’s performance
- Person’s perception
- Tool’s usability
- Document’s understandability
- Program’s complexity

etc.

SE Investigation: How?

- Hypothesis/question generation
- Data collection
- Data evaluation
- Data interpretation
- Feedback into iterative process

SE Investigation: Characteristics

- Data sources come from industrial settings
  - This may include people, program code, etc.

  - Surveys
  - Case-studies
  - Experiments

Investigation Principles

- 4 main principles of investigation
  1. **Stating the hypothesis**: What should be investigated?
  2. **Selecting investigation technique**: conducting surveys, case studies, formal experiments
  3. **Maintaining control over variables**: dependent and independent variables
  4. **Making meaningful investigation**: verification of theories, evaluating accuracy of models, validating measurement results.

SE Investigation Techniques

- Three ways of investigate:
- **Formal experiment**: a controlled investigation of an activity, by identifying, manipulating and **documenting key factors of that activity.**
- **Case study**: document an activity by identifying **key factors (inputs, constraints & resource)** that may affect the outcomes of that activity.
- **Survey**: a demonstration study of a situation to try to **document relationships and outcomes**.

Case-study or Experiment?

- How to decide whether conduct an experiment or perform a case-study?

Factor	Experiment	Case-study
Retrospective \ Review	Yes	No
Level of control	High	low
Difficulty of control	Low	High
Level of replication	High	Low
Cost of replication	Low	high
Can generalize?	Yes (may be)	no

Hypothesis

- Deciding what to investigate.
- The goal for the research can be expressed as a hypothesis in quantifiable terms that is to be tested.
- The test result (the collected data) will confirm or refute the hypothesis.
- Eg: Can Software Reliability Engineering (SRE) help us to achieve an overall improvements in software development practice in our company?

Hypothesis

- Eg:
- Can integrated development and testing tools improve our productivity?
- Does cleanroom software development product better quality software than using the conventional development methods?

Control

- What variables may affect truth of a hypothesis? How do they affect it?
- Variable:
  - Independent (values are set by experiment or initial conditions)
  - Dependent (values are affected by change of other variable
  - Eg: effect of “programming language” on “quality” of resulting code.

Control

- Ignoring other variables that may affect the values of a dependent variable.
- How to identify the dependent and independent variable?

Formal Experiments: Planning

- Conception
  - Defining the goal of investigation
- Design
  - Generating quantifiable hypotheses to be tested
  - Defining experimental objects or units
  - Identifying experimental subject
  - Identifying the response variables

Formal Experiments: Planning

- Preparation
  - Getting ready to start, e.g., purchasing tools, hardware, training personnel, etc.
- Execution
- Review and analysis
  - Review the results for soundness and validity
- Dissemination & decision making
  - Documenting conclusions

Formal Experiments: Principles

- Replication
  - Experiment under identical conditions should be repeatable.
  - Confounded results should be avoided.
- Randomization
  - The experimental **trials must be organized** in a way that the effects of uncontrolled variables are minimized

Formal Experiments: Principles

- Local control
  - **Blocking**: allocating experimental units to blocks or groups so the units within a block are relatively homogeneous. The blocks are designed so that the experimental design captures the anticipated variation in the blocks **by grouping like varieties**, so that the variation does not contribute to the experimental error.
  - **Balancing**: is the blocking and assigning of treatments so that an equal number of subjects is assigned to each treatment. Balancing is desirable because it simplifies the statistical analysis.
  - **Correlation** – Linear and non-linear correlation

Formal Experiments: Types

Factorial design:

- Crossing (each level of each factor appears with each level of the other factor)
- Nesting (each level of one occurs entirely in conjunction with one level of another)
- May reduce the number of cases to be tested.

Formal Experiments: Types

- Advantage of factorial design
  - Resources can be used more efficiently
  - Coverage (completeness) of the target variables range of variation.
  - Implicit replication

  Disadvantages of factorial design
  - Higher costs of preparation, administration and analysis
  - Number of combinations will grow rapidly