Experimentation and Case Studies in Software Engineering

By

Kristian Sandahl
Why do we need empirical studies?

Software Engineering has great variation in:

- Scale
- Domain
- Tools
- Infrastructure
- Human resources
- Organization
- Locality
- Technique
- Quality
What is an experiment?

Units (subjects) ➔ Treatment ➔ Comparison

Control group ➔ No treatment ➔
Types of experiments

- **Randomized experiment**: Units receiving the treatment are selected by random
- **Quasi-experiment**: Units are not selected randomly
- **Controlled experiment**: Comparison between treatments (Sjøberg et al, 2005)
- **Correlation study**: Observes relationships with variables (empirical evaluation)
- **Replication**: Repeating the study
- **Differentiated replication**: Replication with variation of essential conditions
# Variables

<table>
<thead>
<tr>
<th>Background variables</th>
<th>Controlled variables</th>
<th>Independent variables</th>
<th>Dependent variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Time of day</td>
<td>Method used</td>
<td>No of errors done</td>
</tr>
<tr>
<td>Sex</td>
<td>Temperature</td>
<td>Tool used</td>
<td>Time to complete task</td>
</tr>
<tr>
<td>Education</td>
<td>Available resources</td>
<td>Size of task</td>
<td>Judgement of quality</td>
</tr>
<tr>
<td>Experience</td>
<td>...</td>
<td>Group size</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

| Different          | Same                  | Manipulated           | Assumed to change as effect manipulation of independent variables |
| Not changeable     | Observed              | Observed              | as effect manipulation of independent variables |

Different Not changeable
Validity threats

• **Internal validity**: Are differences in dependent variables really due to changes of independent variables?

• **Conclusion validity**: Are our measurement and analysis methods appropriate?

• **Construct validity**: Are we measuring the phenomena we intend to do?

• **External validity**: To what population can we generalise our results?
Comparing means

Under certain conditions: Student’s t-test

Significance level: nominally 5%
Comparing distributions

Comparing severity ratings

<table>
<thead>
<tr>
<th>Severity</th>
<th>Tester 1</th>
<th>Tester 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catastrophic</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Severe</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>Moderate</td>
<td>53</td>
<td>27</td>
</tr>
<tr>
<td>Minor</td>
<td>105</td>
<td>58</td>
</tr>
</tbody>
</table>

- Are the testers’ methods the same?
- Under certain conditions: use the Chi-square test
- For 2x2 contingency tables other methods apply, for instance Cohen’s Kappa
The box plot

- Maximum value
- 75th percentile
- 50th percentile (median)
- Mean (optional)
- 25th percentile
- Minimum value
Comparing variance
### Linear regression

#### ANOVA

<table>
<thead>
<tr>
<th>Model</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>120,250</td>
<td>2</td>
<td>60,125</td>
<td>25,880</td>
<td>.002*</td>
</tr>
<tr>
<td>Residual</td>
<td>11,625</td>
<td>5</td>
<td>2,325</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>131,875</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*a. Predictors: (Constant), course, diagram*  
*b. Dependent Variable: time*

#### Coefficients

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
<th>t</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>Std. Error</td>
<td>Beta</td>
<td>Beta</td>
</tr>
<tr>
<td>1</td>
<td>(Constant)</td>
<td>14,375</td>
<td>.934</td>
<td></td>
</tr>
<tr>
<td></td>
<td>diagram</td>
<td>.250</td>
<td>1.078</td>
<td>.031</td>
</tr>
<tr>
<td></td>
<td>course</td>
<td>-7,750</td>
<td>1.078</td>
<td>-.954</td>
</tr>
</tbody>
</table>

*a. Dependent Variable: time*
Case Studies...

• ... are investigating contemporary phenomena in their context

• ... can be:
  – Exploratory
  – Descriptive
  – Explanatory
  – Improving
### Related methods

Table 1  Overview of research methodology characteristics

<table>
<thead>
<tr>
<th>Methodology</th>
<th>Primary objective</th>
<th>Primary data</th>
<th>Design</th>
</tr>
</thead>
<tbody>
<tr>
<td>Survey</td>
<td>Descriptive</td>
<td>Quantitative</td>
<td>Fixed</td>
</tr>
<tr>
<td>Case study</td>
<td>Exploratory</td>
<td>Qualitative</td>
<td>Flexible</td>
</tr>
<tr>
<td>Experiment</td>
<td>Explanatory</td>
<td>Quantitative</td>
<td>Fixed</td>
</tr>
<tr>
<td>Action research</td>
<td>Improving</td>
<td>Qualitative</td>
<td>Flexible</td>
</tr>
</tbody>
</table>

(Runeson & Höst 2009)
Triangulation

Increases precision and thrust-worthiness

Types:

• Data (source)
• Observer
• Methodological
• Theory

(Picture: Wikimedia commons)
Some more definitions

- **Objective**: overall statement of expectations
- **Research questions**: refinement of objective
- **Hypothesis**: supposed explanation
- **Case**: the **object** of the study
- **Subjects**: those providing information
- **Qualitative data**: Descriptions, text
- **Quantitative data**: Numbers, measurements
- **Case study protocol**: Procedures of study
- **Protocol**: Data obtained by observation, interview etc.
Case study research process

1. Case study design
2. Preparation for data collection
3. Collecting evidence
4. Analysis of collected data
5. Reporting

(Runeson & Höst 2009)
## Interviews

(Runeson & Höst 2009)

<table>
<thead>
<tr>
<th></th>
<th>Unstructured</th>
<th>Semi-structured</th>
<th>Fully structured</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Typical focus</strong></td>
<td>How individuals qualitatively experience the phenomenon</td>
<td>How individuals qualitatively and quantitatively experience the phenomenon</td>
<td>Researcher seeks to find relations between constructs</td>
</tr>
<tr>
<td><strong>Interview questions</strong></td>
<td>Interview guide with areas to focus on</td>
<td>Mix of open and closed questions</td>
<td>Closed questions</td>
</tr>
<tr>
<td><strong>Objective</strong></td>
<td>Exploratory</td>
<td>Descriptive and explanatory</td>
<td>Descriptive and explanatory</td>
</tr>
</tbody>
</table>
Interviews

Process:
• Start
• Q & A
• Summary teach-back
• Thank you!
• What’s next

Kinds:
• Structured
• Unstructured

Tips
• Be 2 interviewers – shift roles
• Plan the interview
• Don’t stick to the plan – use feelings
• Let the customer talk
• Prepare ice-breakers
• Probe thinking
• Look for body language
• Think of human bias
• Why do you get the answers you get?
• Record for memory and repeatability!
Coding in protocol analysis

- Input: some type of media
- Output: annotated media
- Open coding – exploration and classification
- Axial coding – connections between categories
- Factual coding – use a more concrete pre-understanding, events, actions, setting...
- Interpretative coding – use more abstract pre-understanding, conditions, perspectives...
- Pattern or frequency analysis
Discussion

• Why are only 2 % of papers dealing with controlled experiments?
• Is case-study research appropriate for Software Engineering?
• Why are the disposition of articles so diverse?
• What would happen if we had more controlled experiment, well reported case-studies and common ways of reporting?
What happens if the line is broken?