Empirical Study in SE [1]

Overview

Software Engineering

- IEEE : « Application of a systematic, disciplined, <u>quantifiable</u> approach to development, operation and maintenance of a software »

=> How to assess the benefits of your research

Quantitative / Qualitative

- Quantitative: quantifying a relationship or comparing two or more groups. Identify a cause-effect.
- Qualitative: studying objects in their natural setting. Interpret a phenomenon based on explanation.

Strategies

- Survey: Investigation performed in retrospect. Analysing a representative sample with the intent to generalise the conclusion.
- Case study: Monitor a living project or an activity. Aim at identify a specific attribute or a relationship.
- Experiment: Done in a laboratory. Different treatments performed at random. Measure the effect of variables.

Strategies vs Quant. & Qual

Strategy	Quantitative vs Qualitative
Survey	Both
Case Study	Both
Experiment	Quantitative

Measurement

- Measurement is a mapping from the empirical world to the formal, relational world.
- The input and the output must be measured to control the study and to see the effects
- Metrics: (1) The field of software engineering measurement, or (2) the entity which is measured

Scale

- Nominal: maps attributes into a name or a symbol (a class)
  - Ex: Red, Blue, Yellow
- Ordinal: ranks attributes
  - Ex: low, middle, high
- Interval: ranks attributes and gives them a distance
  - Ex: number of stars for hotel
- Ratio: ranks and distance + zero and ratio
  - Ex: Real

Qualitative (Nominal, Ordinal)

Quantitative (Interval, Ratio)

Goal / Question / Metric paradigm

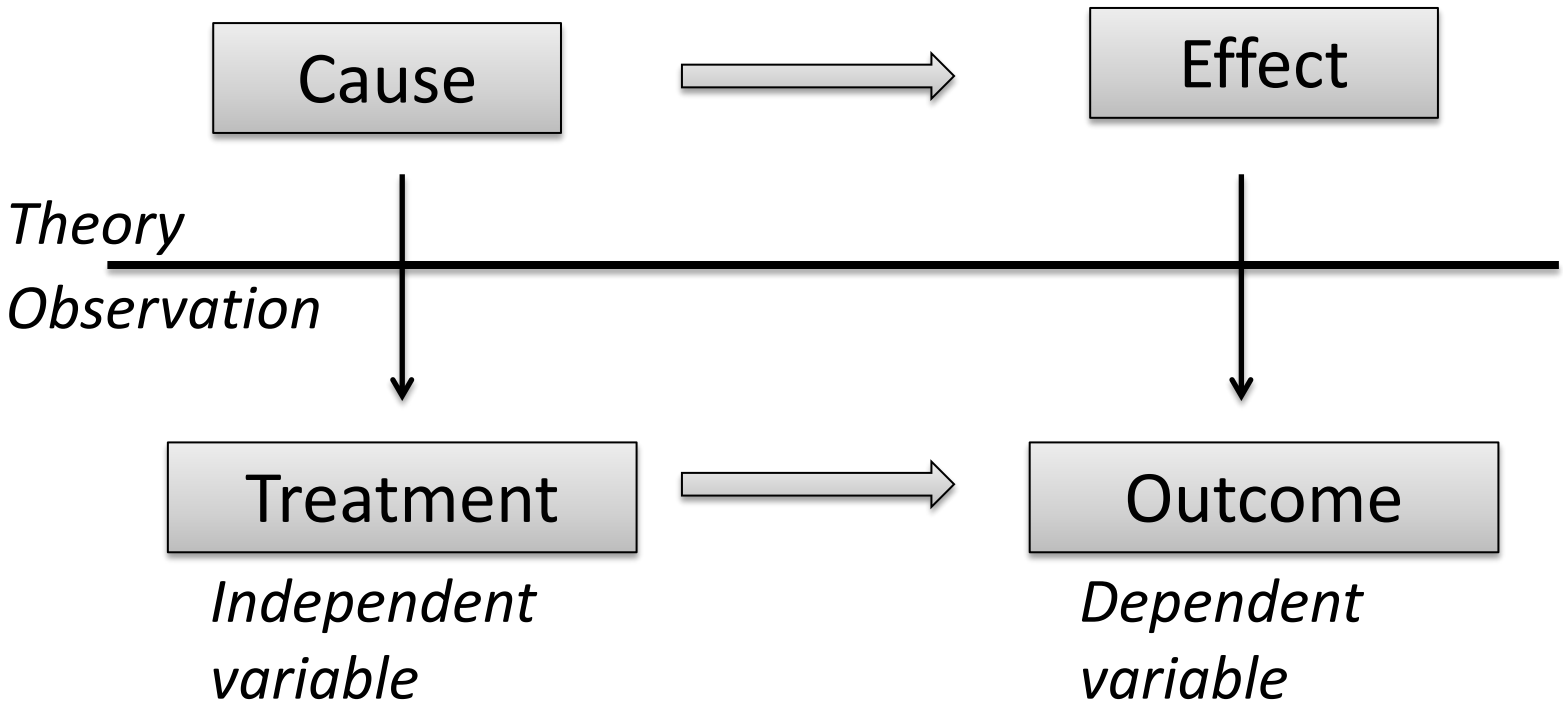
1. Conceptual level (Goal)
   – Define the goal of the empirical study (objects)
2. Operational level (Question)
   – Set the questions that characterize the assessment of the goal.
3. Quantitative level (Metric)
   – Identify the data and the measure needed to answer the question.

Exercise

- Define your empirical study:
  - Goal
  - Question
  - Data

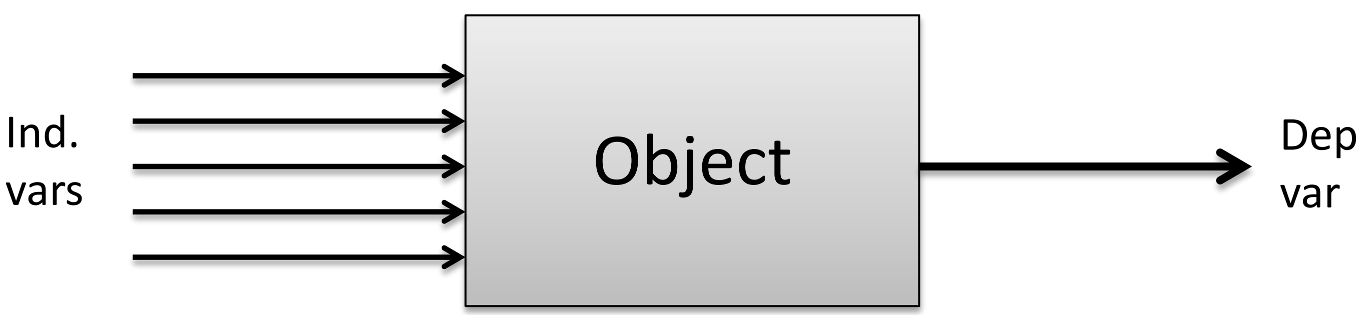
=> Need of a process to design a study

Experiment Principles



Concepts (1/2)

- **Dependent variable (response variable)**: the studied variable
- **Independent variables**: variables that can be either controlled or manipulated.
- **Factors**: independent variables that will change (manipulated)
- **Treatment**: Value of a factor



Concepts (2/2)

- Object: Object of the study (product, process, resource)
- Subjects: people that realise the treatment (if needed)

Process

Idea

1. Definition
2. Planning
3. Operation
4. Analysis & Interpretation
5. Package

$\Rightarrow$Conclusion

Definition: Goal Template

- Analyze <Object(s) of study>
- For the purpose of <Purpose>
- With respect to their <Quality focus>
- From the point of vies of the <Perspective>
- In the context of <Context>

Definition: Examples

Object	Purpose	Quality Focus	Perspective	Context
Product	Characterize	Effectiveness	Developer	Student
Process	Monitor	Cost	Modifier	Project
Metric	Predict	Reliability	Maintainer	Large software
Theory	Control	Maintenability	Project Manager	OSS
	Change	Portability	Customer	Senior
			User	
			Researcher	

Planning: 7 steps

1. Context selection
2. Hypothesis formulation
3. Variables selection
4. Selection of subjects
5. Experiment design
6. Instrumentation
7. Validity evaluation

$\Rightarrow$Experiment design

Planning: 1- context selection

Trade-off between cost and generalisation

- Student vs Professionals
- Toy vs real software
- Off-line vs on-line

Planning: 2 - Hypothesis

- Null hypothesis (H_0): No effect.
  - So we expect to reject it !
- Alternative hypothesis (H_a, H_1, etc.): Effect.
  - We expect to accept it!
- Type-I-Error: Reject H_0 although it is true (Conclude that there is an effect although there is no).
  - P(Type-I-Error)=P(reject H_0 | H_0 is true)
- Type-II-Error: Not reject H_0 although it is false (Conclude that there is no effect although there is one).
  - P(Type-II-Error)=P(not reject H_0 | H_0 is false)

Planning: 3 – Variables Selection

- Independent Variables:
  - No advice or guideline.
  - Requires domain knowledge.
  - Include the choice of the scale.
- Dependent Variables:
  - Support for the hypothesis.
  - Have to be easily measurable (!metrics aggregation).

Planning: 4 – Selection of Subjects

- Large impact on the generalisation
- Randomized
  - Pure random
  - Quota
  - Blocked

Planning 5 – Experiment Design

- One factor with two treatments
- One factor with more than two treatments
- Two factors with two treatments
- More than two factors each with two treatments

Planning 6 - Instrumentation

- Objects
- Guidelines
- Measurement

Planning 7 – Validity evaluation

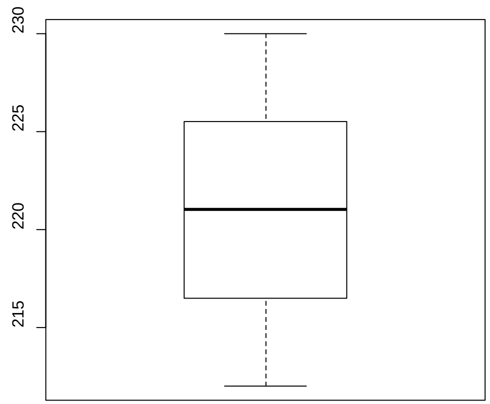
- Conclusion validity
  - Low statistical power, …
- Internal validity
  - History, maturity, …
- Construct validity
  - Mono-method bias, …
- External validity
  - Selection, …

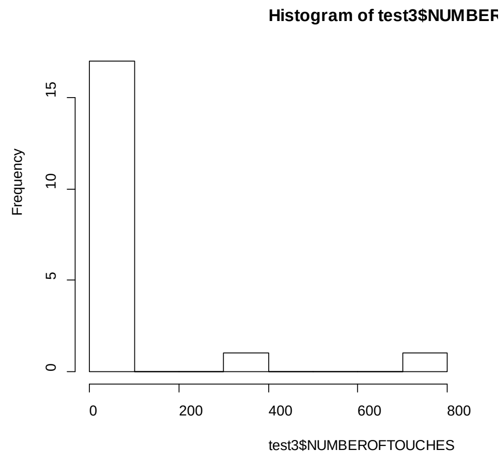
Analysis

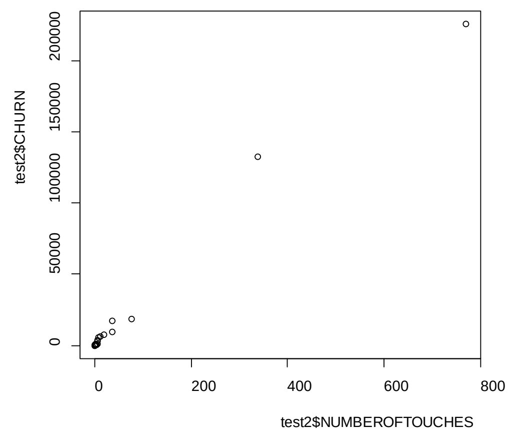
- Descriptive Statitics
- Data set reduction
- Hypothesis testing

Analysis: Descriptive Statistics

scale	Central tendency	Dispersion
Nominal	Mode	Frequency
Ordinal	Median	Interval of variation
Interval	Mean	Standard deviation, variance
Ration	Geometric mean	Coefficient of variation

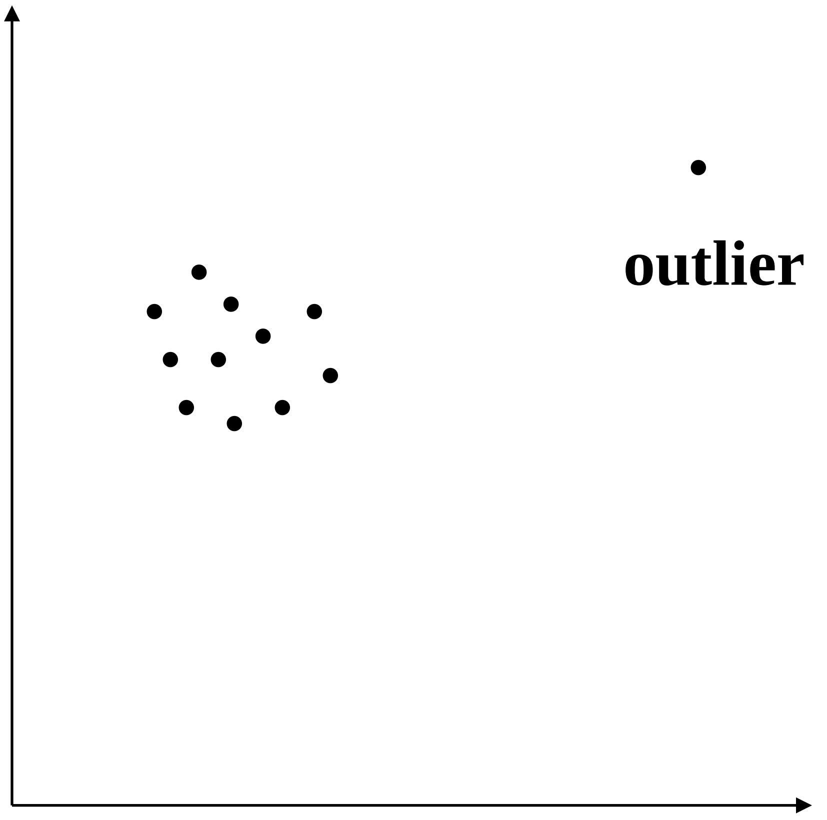






Analysis: Outliers

- Remove data that clearly does not fit the sample



Scatter plots

Analysis: Hypothesis Testing 1/3

- Rewrite the null hypothesis according to the measure
  - Ex: one factor, two treatments => same probability p(A)=p(B)=1/2
- Test Error-I : p(Reject Ho | Ho is true)
  - p(Reject | p(A)=p(B)=1/2)
  - If we have 15 tests. P(4 or less) = 0.059, P(5 or less) = 0.1509. If we obtain 4 or less A then, this is not due to hazard. If we obtain 5 or less, then it may be due to hazard.
  - 0, 1, 2, 3 or 4 => reject
  - 5, 6 … => cannot reject

Analysis: Hypothesis Testing 2/3

Design	Parametric	Non-param
One factor, one treat		Chi-2, Binomial test
One factor, two treatments, rand design	t-test, F-test	Mann-Whitney, Chi-2
One factor, two treatments, paired	Paired t-test	Wilcoxon, Sign test
One factor, more that two treatments	ANOVA	Kruskal-Wallis, Chi-2
More than one factor	ANOVA	

Example

- Goal: Show that our tool improves efficiency for developers to evolve their software
- H_0 : time needed to realize an evolution is the same with or without our tool (one factor, two treatment)
  - t-test to compare the mean (check normal law)
- Two groups (with and without our tool)
- Object: realize one evolution on an existing product

Conclusion

- Empirical Study as a validation
- **Quantitative** / Qualitative
- Design Process
  - Design, Plan, Analysis

Ref

[1] “Experimentation in Software Engineering, An introduction”. C. Wohlin, P. Runeson, M. Höst, M.C. Ohlsson, B. Regnell, A. Wesslén. Kluwer Academic Publishers.