DCC / ICEx / UFMG

# Experiment Planning

Eduardo Figueiredo

http://www.dcc.ufmg.br/~figueiredo

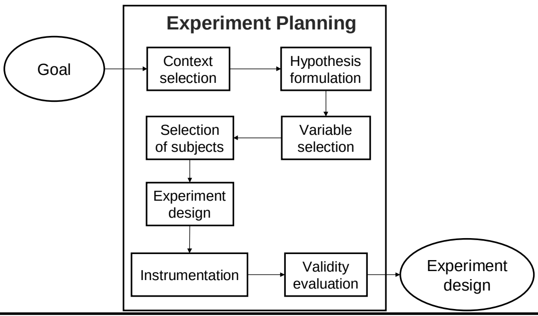
## Scoping and Planning

- Scoping determines the foundations for the experiment
  - Why the experiment is conducted?
- Planning prepares the experiment
  - How the experiment is conducted?

## Planning Steps

- The planning phase can be divided into seven steps
  - Context selection
  - Hypothesis formulation
  - Variable selection
  - Selection of subjects
  - Experiment design
  - Instrumentation
  - Validity evaluation

## Planning Phase Overview




# Context Selection

## Context Selection

- The context of the experiment can be characterized by four dimensions
  - Offline vs. online
  - Student vs. professional
  - Toy vs. real problems
  - Specific vs. general
- Cheaper: offline with students using a toy problem in a specific domain

## Hypothesis Formulation

## Hypothesis Formulation

- Hypothesis formulation is the basis for the statistical analysis of an experiment
- If the hypothesis can be rejected, the conclusions can be drawn
- We aim to reject the hypothesis
  - A hypothesis is stated formally
  - Data are collected

## Null and Alternative Hypotheses

- The experiment is formalized into two hypotheses
  - Null hypothesis ($H_0$)
  - Alternative hypothesis ($H_1$)
- An alternative hypothesis is in favor of the null hypothesis rejection

## Null Hypothesis

- A null hypothesis ($H_0$) states that there are no real underlying trends or patterns in the experiment setting
  - Therefore, the only reason for differences are coincidental
- The experimenter wants to reject this hypothesis

## Types of Errors

- Hypothesis testing involves two main types of risks (or types of errors)
  - Type-I-error: the risk of rejecting a true hypothesis
  - Type-II-error: the risk of do not rejecting a false hypothesis

## Variable Selection

## Variable Selection

- The independent variables we can control and change in the experiment
  - The choice of independent variables also includes choosing the measurement scales
- The effect of treatment is measured in the dependent variable
  - There is usually only one dependent variable
  - The dependent variable is derived from the hypothesis

## Selection of Subjects

## Selection of Subjects

- The selection of subjects is related to the generalization of the results
  - In order to generalize, the selection must be representative for the population
- The selection (or sampling) can be of two types
  - A probability sample
  - A non-probability sample

## Types of Sampling

- Probability Sampling: the probability of selecting each subject is known
  - Example: subjects are selected from a list of the population at random
- Non-Probability Sampling: the probability of selecting each subject is unknown
  - Example: the nearest and most convenient persons are selected as subjects

## Experiment Design

## Experiment Design

- To get the most out of the experiment, it must be carefully planned and designed
  - Design is closely related to interpretation
- An experiment consists of a series of tests of the treatments
  - A design describes how the tests are organized and run

## General Design Principles

- A proper design allows replication
- The general design principles are
  - Randomization
  - Blocking
  - Balancing

## Randomization

- All statistical methods require observations from independent random variables
- Randomization applies to the allocation of objects, subjects, and the order of the tests
  - It is also used to select subjects that are representative of the population

## Blocking

- Blocking is used to systematically eliminate the undesired effect in the comparison among treatments
  - Within one block (group), the undesired effect is the same
- Example:
  - Subjects with different experience in OOP
  - They are grouped into two groups: one of people with and one without experience

## Balancing

- We have a balanced design if each treatment has equal number of subjects
  - That is, the same number of subjects in each group (block)

## Standard Design Types

- Some of the most frequently used experiment designs are
  - One factor with two treatments
  - One factor with more than two treatments
  - Two factors with two treatments
  - More than two factors, each factor with two treatments

## One Factor, Two Treatments

- In this design, we want to compare the two treatments against each other
- Example
  - Factor: Heuristic for Bad Smell Detection
  - Treatments: Concern-Sensitive Heuristics and Traditional Heuristics
- Assigning treatments to subjects
  - Completely Randomized Design
  - Paired Comparison Design

## Completely Randomized Design

- We assign subjects randomly to each treatment
- Each subject uses only one treatment
- If we have the same number of subjects per treatment, the design is balanced

## Paired Comparison Design

- It is also called crossover design
  - We assign each subject to both treatment
  - The order is assigned randomly to each subject
  - If we have the same number of subjects starting with the first and the second treatments, the design is balanced
- Warning! Subjects may gain information from the first to the second treatment

## Example: One Factor, Two Treatments

**Completely Randomized Design**

| | Treatments | |
|---|---|---|
| **Subjects** | **Treat 1** | **Treat 2** |
| 1 | X | |
| 2 | | X |
| 3 | | X |
| 4 | X | |
| 5 | | X |
| 6 | X | |

**Paired Comparison Design**

| | Treatments | |
|---|---|---|
| **Subjects** | **Treat 1** | **Treat 2** |
| 1 | 2 | 1 |
| 2 | 1 | 2 |
| 3 | 2 | 1 |
| 4 | 2 | 1 |
| 5 | 1 | 2 |
| 6 | 1 | 2 |

## One Factor, More Treatments

- In this design type, comparison is often performed on the treatment means
- Example
  - Factor: Modeling Tool
  - Treatments: ArgoUML, StarUML, LucidChart
- Assigning treatments to subjects
  - Completely randomized design
  - Randomized complete block design

## Completely Randomized Design

| | Treatments | | |
|---|---|---|---|
| **Subjects** | **Treat 1** | **Treat 2** | **Treat 3** |
| 1 | X | | |
| 2 | | X | |
| 3 | | | X |
| 4 | X | | |
| 5 | | | X |
| 6 | | X | |

## Randomized Complete Block

| | Treatments | | |
|---|---|---|---|
| **Subjects** | **Treat 1** | **Treat 2** | **Treat 3** |
| 1 | 1 | 3 | 2 |
| 2 | 3 | 1 | 2 |
| 3 | 2 | 3 | 1 |
| 4 | 2 | 1 | 3 |
| 5 | 3 | 2 | 1 |
| 6 | 1 | 2 | 3 |

## Two Factors

- The experiment gets more complex when two factors are analyzed
- It requires three hypotheses
  - One hypothesis for the effect of the first factor
  - One hypothesis for the effect of the second factor
  - One hypothesis for the interaction between the two factors

## Examples of Two Factors

- Factor 1: Measurement Tool
  - Metrics Plugin
  - Google AnalytiX
- Factor 2: Analyzed Code
  - JHotDraw
  - ArgoUML
- Assigning treatments to subjects
  - 2 * 2 factorial design
  - Two-stage nested design

## 2 * 2 Factorial Design

| | | Factor A | |
|---|---|---|---|
| | | Treatment A1 | Treatment A2 |
| **Factor B** | Treatment B1 | Subjects 4, 6 | Subjects 1, 7 |
| | Treatment B2 | Subjects 2, 3 | Subjects 5, 8 |

## Two-Stage Nested Design

| Factor A | | | |
|---|---|---|---|
| Treatment A1 | | Treatment A2 | |
| **Factor B** | | **Factor B** | |
| Treatment B1' | Treatment B2' | Treatment B1" | Treatment B2" |
| Subjects 1, 3 | Subjects 6, 2 | Subjects 7, 8 | Subjects 4, 5 |

## More than Two Factors

- It is also called factorial design
- The effect on the dependent variable considers both
  - Each factor separately
  - The interaction between the factors
- Interactions can be between two or more factors

## Assigning Treatments

- $2^k$ factorial design (general case)
  - k is the number of factors
- $2^2$ factorial design (k = 2)
  - The same of 2 * 2 factorial design
  - Two factors with two treatments
- $2^3$ factorial design (k = 3)
  - Three factors, each with two treatments

## Example of $2^3$ Factorial Design

| Factor A | Factor B | Factor C | Subjects |
|---|---|---|---|
| A1 | B1 | C1 | 2, 3 |
| A2 | B1 | C1 | 1, 13 |
| A1 | B2 | C1 | 5, 6 |
| A2 | B2 | C1 | 10, 16 |
| A1 | B1 | C2 | 7, 15 |
| A2 | B1 | C2 | 8, 11 |
| A1 | B2 | C2 | 4, 9 |
| A2 | B2 | C2 | 12, 14 |

# Instrumentation

## Instrumentation

- The goal of instrumentation is to provide means for performing and monitoring the experiment
- The instruments for an experiment are of three types
  - Objects
  - Guidelines
  - Measurements

## Experiment Objects

- When planning an experiment, it is important to choose appropriate objects
- Example
  - In an experiment for bad smell detection, the number of bad smells must be known
  - You can either introduce bad smells or use objects with known bad smells

## Guidelines

- Guidelines are needed to guide the participants in the experiment
  - Participants may also need training in the methods or tools to be used
- Guidelines include, for instance, process descriptions and checklists

## Measurement

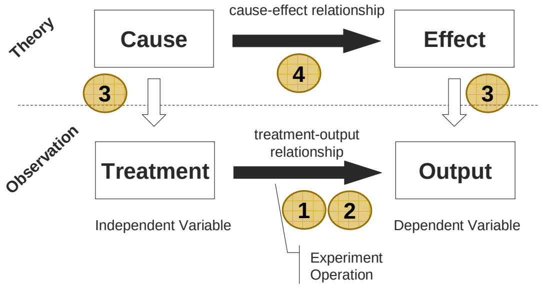
- Measurements in an experiment are conducted by data collection
  - Data is usually collected via manual forms and interviews
  - Automated data collection tends to be more reliable than manual collection

# Validity Evaluation

## Validity Evaluation

- It is important to consider the experiment validity already in the planning phase
- The results must be valid for the population of interest
  - If the population is one organization, results have to be valid for this organization
- The four main types of validity threats
  - Conclusion, Internal, Construct, External

## Experiment Principles




## Four Categories of Threats

1. Conclusion Validity: concerned with the relationship between treatment and outcome
2. Internal Validity: there is a causal relationship between treatment and outcome
3. Construct Validity: concerned with the relationship between theory and observation
4. External Validity: there is a causal relationship between cause and effect

## Bibliography


- C. Wohlin et al. **Experimentation in Software Engineering**, Springer. 2012.
  - Chapter 8 – Planning