Experimental Methods in Computer Science

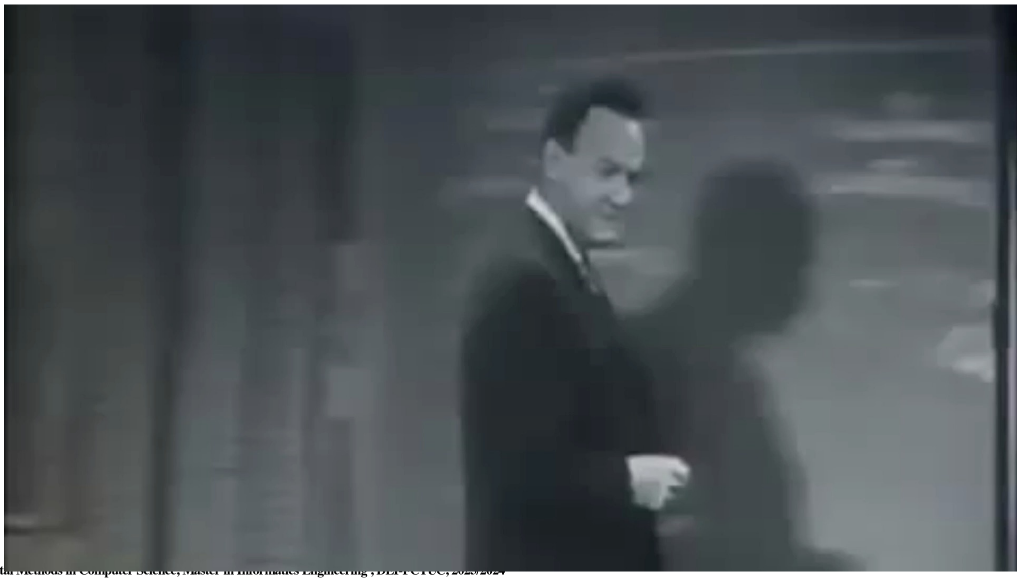
(*Metodologias Experimentais em Informática*)

Henrique Madeira

Master in Informatics Engineering
Departamento de Engenharia Informática
Faculdade de Ciências e Tecnologia da Universidade de Coimbra
2023/2024

1

Why do we need experiments?



2

Why do we need experiments?

Researchers:

- Collect evidence facts about the world (or system)
- Validate hypothesis
- Support the definition, validation, parameterization of models
- Validate models
- Confirm theories
- Etc, etc…

Henrique Madeira, DEI-FCTUC, 2018-2023

Experimental Methods in Computer Science, Master in Informatics Engineering , DEI-FCTUC, 2023/2024 3

3

Why do we need experiments?

Engineers (including informatics engineers):

- Tune up systems
- Compare and select among different project choices
- Verify that requirements or specifications are met
- Validate mechanisms and/or solutions
- Measure/evaluate features, e.g., to access efficiency of mechanisms.
- Assess the effectiveness of processes, e.g., software development processes
- Etc, etc…

"Experimentation as the feedback step in the engineering loop"

Henrique Madeira, DEI-FCTUC, 2018-2023

Experimental Methods in Computer Science, Master in Informatics Engineering , DEI-FCTUC, 2023/2024 4

4

Key properties

- **Relevance**
  Are the goals of the experiment and the expected results important for the progress (do they have impact in science, technology, market, etc.)?
- **Representativeness**
  Is the experiment realistic and representative of real-world scenarios?
- **Repeatability**
  Is it possible to repeat the experiment and achieve same or statistically similar results?
- **Reproducibility**
  Is there enough information to allow others to reproduce the experiment?
- **Results analysis and generalization**
  Is the analysis of results sound? Is the generalization of conclusions credible?
- **Cost**
  Is the cost of the experiments compatible with the expected benefits?

Many types of experiments...

- Controlled experiments
- Field studies
- Case studies
- Pilot studies
- Benchmarks
- Simulations
- Surveys
- Rational reconstructions
- Artifact/archive analysis
- Ethnographies
- Quasi-experiments

Design of experiments (a first look)

Laboratory experiments, controlled experiments

1. Problem statement (or research question)
2. Identify variables
3. Generate hypothesis
4. Define the experimental setup/scenario
5. Develop tools and procedures for the experiment
6. Run experiments and collect the data/measurements
7. Perform data analysis and test hypothesis
8. Draw conclusions (often go back to the beginning and reformulate the problem statement or test a different hypothesis)

Design of the experiment

Measurements

Analysis

Conclusions

Henrique Madeira, DEI-FCTUC, 2018-2023

7

Design of experiments (a first look)

Laboratory experiments, controlled experiments

1. Problem statement (or research question)
2. Identify va
3. Generate h
4. Define the
5. Develop to
6. Run experi
7. Perform da
8. Draw conc
reformulate the problem statement or test a different hypothesis)

A good (i.e., relevant) problem statement should be focused enough to allow the clear identification of the variables of the problem but, at the same time, should be sufficiently open to allow different hypothesis to answer the problem/question.

Possible generic formulation:
How does X affect Y under conditions Z?

Henrique Madeira, DEI-FCTUC, 2018-2023

8

Design of experiments (a first look)

Laboratory experiments, controlled experiments

1. Problem statement (or research question)
2. Identify va
3. Generate h
4. Define the
5. Develop to
6. Run experi
7. Perform da
8. Draw conc
   reformulate the problem statement or test a different hypothesis)

To formulate good problem statements:

- Must know the subject area: process, system, technique, product, product market, etc.
- Must be precise and clear
- Must be sure that the problem/question is relevant

Henrique Madeira, DEI-FCTUC, 2018-2023

Experimental Methods in Computer Science, Master in Informatics Engineering , DEI-FCTUC, 2023/2024 9

9

Design of experiments (a first look)

Laboratory experiments, controlled experiments

1. Problem statement (or research question)
2. Identify variables
3. Genera
4. Define
5. Develo
6. Run ex
7. Perforı
8. Draw c
   reformu

- **Dependent variable** (response variable)
  Measured output (e.g., response time, throughput, no. bugs, downtime, latency, error detection coverage, etc., etc.)
- **Independent variables** (factors)
  Input variables that can be changed in the experiment (e.g., memory size, clock rate, file size, channel bandwidth, etc., etc.)
- **Levels**
  Values taken by the variables. Can be (nearly) continuous (e.g., ~time, size in bytes) or discrete (type of system, type of algorithm, etc.)

Henrique Madeira, DEI-FCTUC, 2018-2023

Experimental Methods in Computer Science, Master in Informatics Engineering , DEI-FCTUC, 2023/2024 10

10

Design of experiments (a first look)

Laboratory experiments, controlled experiments

1. Problem statement (or research question)
2. Identify variables
3. Genera
4. Define
5. Develo
6. Run ex
7. Perfor
8. Draw c
   reformul

ne,

ne,

- **Change one factor at the time**
  Simple scenario. The analysis is simple, as it is easy to understand the effect of a given factor on the independent variable.
- **Full factorial**
  Change two or more factors simultaneously. Much more complex (to be seen in detail later on in the course). Has two important advantages:
  - More efficient experiments (save time and effort)
  - Allow the study of possible interactions among factors

Henrique Madeira, DEI-FCTUC, 2018-2023

Experimental Methods in Computer Science, Master in Informatics Engineering , DEI-FCTUC, 2023/2024 11

11

Design of experiments (a first look)

Laboratory experiments, controlled experiments

1. Problem statement (or research question)
2. Identify variables
3. Genera
4. Define
5. Develo
6. Run ex
7. Perfor
8. Draw c
   reformul

e,

,

Terminology:

- **Baseline** (often called **golden run**)
  Set of factor values (i.e., independent variables) that represent a baseline scenario for the experiments
- **Repetition of golden run**
  Used to estimate the experimental error (noise) in the system and identify small effects that may cause variations in the results.

Allow the study of possible interactions among factors

Henrique Madeira, DEI-FCTUC, 2018-2023

Experimental Methods in Computer Science, Master in Informatics Engineering , DEI-FCTUC, 2023/2024 12

12

Design of experiments (a first look)

Laboratory

1. Problem st
2. Identify v
3. Genera
4. Define
5. Develo
6. Run ex
7. Perfor
8. Draw c
   reformul

Terminology:

- **Randomization**
  Minimize potential uncontrollable biases in the experiments by randomly assigning factors to "average out" the effects of possible extraneous factors.
- **Blocking**
  The experiment is divided in homogeneous segments (blocks such as sets of machines, users, loads, etc.) to improve precision. The goal is to control the variability block to block.
- **Confounding variable**
  Extraneous variable that influences the relationship between the dependent and independent variables (i.e., correlates with both the dependent and independent variables).

Henrique Madeira, DEI-FCTUC, 2018-2023

13

Design of experiments (a first look)

Laboratory experim

1. Problem statement (o
2. Identify variables
3. Generate hypothesis
4. Define the experimen
5. Develop tools and pro
6. Run experiments and
7. Perform data analysis
8. Draw conclusions (oft
   reformulate the problem statement or test a different hypothesis)

- Hypothesis describe provisional relationships between factors (independent variables) and the response variable (dependent). It is a interim answer to the problem statement.
- Can be directional or non-directional
- May lead to a model allowing prediction of what is going to happen in future cases.
- Quite often (in computers) the goal of the experiments is to quantify the relationship (not just confirm that exists)

Henrique Madeira, DEI-FCTUC, 2018-2023

14

Design of experiments (a first look)

Laboratory ex

- Experiment complexity
- Experiment cost
- Availability of tools and frameworks that may help
- Degree of automation

1. Problem state
2. Identify varial
3. Generate hypo
4. Define the experimental setup/scenario
5. Develop tools and procedures for the experiment
6. Run experiments and collect the data/measurements
7. Perform data analysis and test hypothesis
8. Draw conclusions (often go back to the beginning and reformulate the problem statement or test a different hypothesis)

Henrique Madeira, DEI-FCTUC, 2018-2023

Experimental Methods in Computer Science, Master in Informatics Engineering , DEI-FCTUC, 2023/2024 15

15

Design of experiments (a first look)

Labora

- Continuous and/or discrete measurements
- Accuracy, precision, and resolution
- Basic measurements in computers…
  - Count
  - Duration
  - Size
  - Any value derived from the combination of basic measurements

1. Probl
2. Ident
3. Gene
4. Defir
5. Deve
6. Run experiments and collect the data/measurements
7. Perform data analysis and test hypothesis
8. Draw conclusions (often go back to the beginning and reformulate the problem statement or test a different hypothesis)

Henrique Madeira, DEI-FCTUC, 2018-2023

Experimental Methods in Computer Science, Master in Informatics Engineering , DEI-FCTUC, 2023/2024 16

16

Design of experiments (a first look)

Labor

1. Prob
2. Iden
3. Gen
4. Defi
5. Dev
6. Run experiments and collect the data/measurements
7. Perform data analysis and test hypothesis
8. Draw conclusions (often go back to the beginning and reformulate the problem statement or test a different hypothesis)

- Exploratory data analysis
- Statistical data analysis
  - Tables, charts, etc., average, standard deviation
  - Coping with measurement errors
  - Confidence intervals
  - Statistical comparison of alternatives
  - Tests to check if measured data fit know distributions (chi-square, K-S tests,…)

Henrique Madeira, DEI-FCTUC, 2018-2023

Experimental Methods in Computer Science, Master in Informatics Engineering , DEI-FCTUC, 2023/2024 17

17

Design of experiments (a first look)

Labor

1. Prob
2. Iden
3. Gen
4. Defi
5. Dev
6. Run
7. Perform data analysis and test hypothesis
8. Draw conclusions (often go back to the beginning and reformulate the problem statement or test a different hypothesis)

- The written **report** of the experiments is quite often the **single outcome** of months or years of work
- **Quality of writing** is essential. Some relevant attributes of the report:
  - Clear (in the goals, approach, setup, steps, analysis, discussion, conclusions)
  - Credible (in the data reported, conclusion, etc.)
  - Self-contained

Henrique Madeira, DEI-FCTUC, 2018-2023

Experimental Methods in Computer Science, Master in Informatics Engineering , DEI-FCTUC, 2023/2024 18

18

Example

Consider the following problem/question: Is the number of software bugs found in the tests of program units developed by programmers dependent on the average number of sleeping hours of the programmers?

Assume that you have the detailed specifications of a set of program units to be developed and consider that the program units include units of high, medium and low complexity. Additionally, you have comprehensive unit test suits to test each program unit.

In these circumstances, describe how you would organize an experiment to answer the proposed question (problem statement). Your answer should be as complete as possible, focusing on the experiment design steps (obviously, it does not make sense to speculate about the experiment results and conclusions), and indicate at least the following:

a) Dependent and independent variables.

b) Levels you would consider for the independent variables.

c) Hypothesis under evaluation.

d) Hypothesis testing technique you would use.

e) Brief description of the experimental setup, taking into account in your answers to the previous points and the fact that the experiment deals with people (the programmers).