Introduction to the experimental method

Why do we need experiments in computer science ?

- Computers and software are technological artifacts, designed and built by humans. They are not natural phenomena, so we should know how they work.

- Computer use Boolean algebra, computer science is based on theory, use abstract models, the hardware and software are the result of engineering process, etc, etc...

Why do we need experiments?

(how do you know whether your program is correct?)

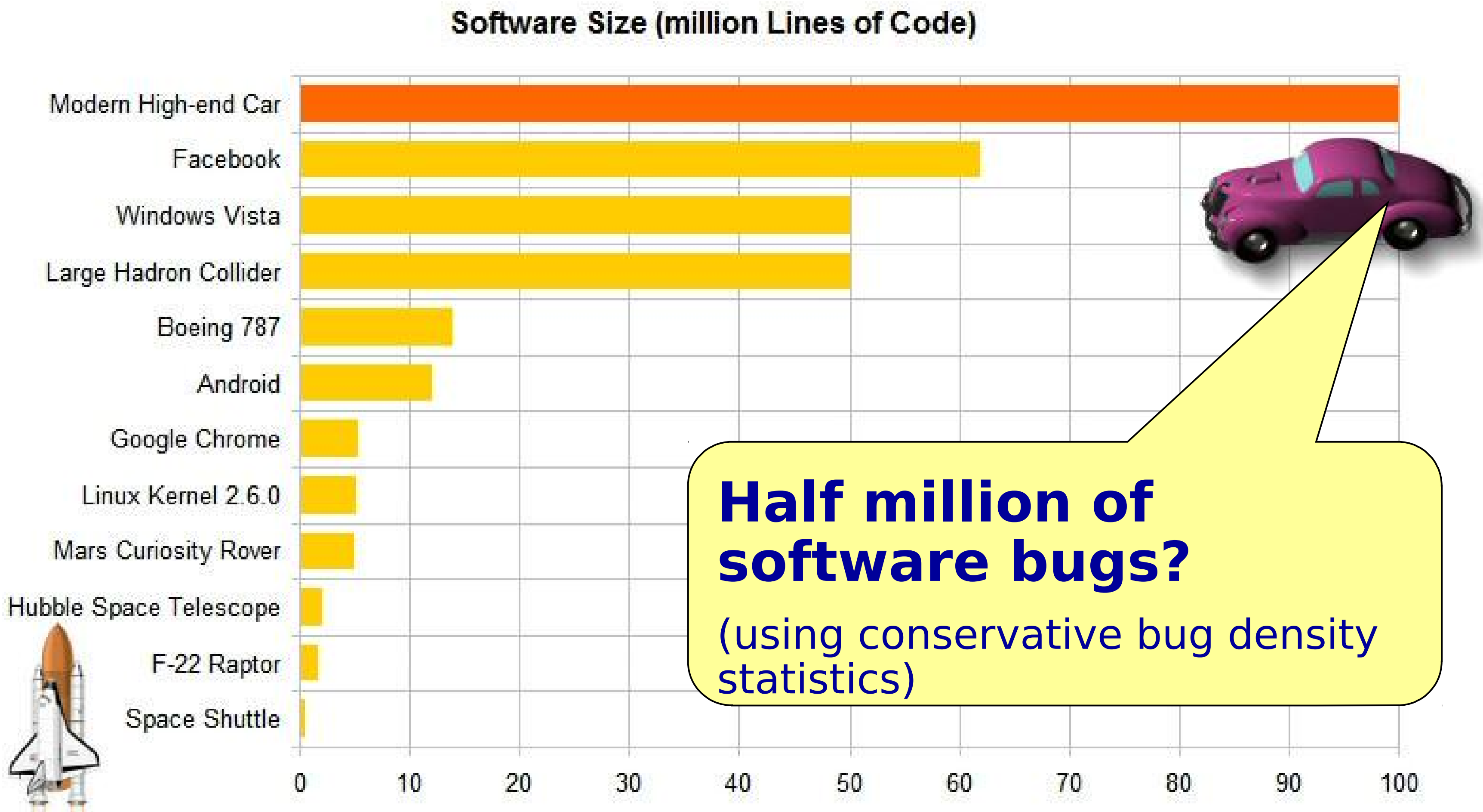
Why do we need experiments in computer science?

- Merge sort has O(n log n) time complexity in the worst case and quick sort has $O(n^2)$ time complexity worst case, but O(n log n) in the average case.

- A web server (Intel i5-7200U, 64 GB RAM) needs to sort an input of at most 10 000 integers for each query (PageRank).

- Which algorithm am I going to use? How can time complexity help me in this particular context?

Why do we need experiments in computer science?



From Rich Rogers, https://twitter.com/richrogersiot/status/958112741218111489

Why do we need experiments?

Researchers:

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

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
- Measure/evaluate features
- Etc, etc…

"Experimentation as the feedback step in the engineering loop"

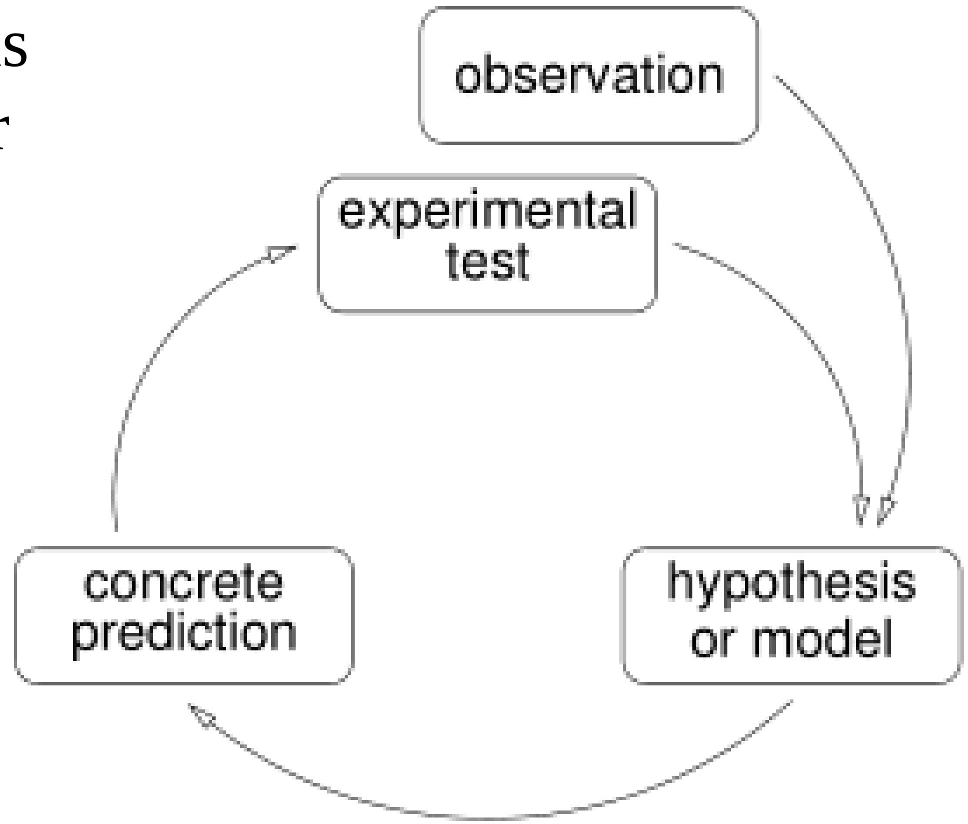
Why do we need experiments?

Business engineers, entrepreneurs, managers:

- Survey to confirm product market potential or to select product features (product-oriented, marketing-oriented)
- Select the best internal organization strategies for companies (production-oriented, management-oriented)

Scientific method (oversimplified)

The Scientific Method is an empirical method for the construction of knowledge



(adapted from [1])

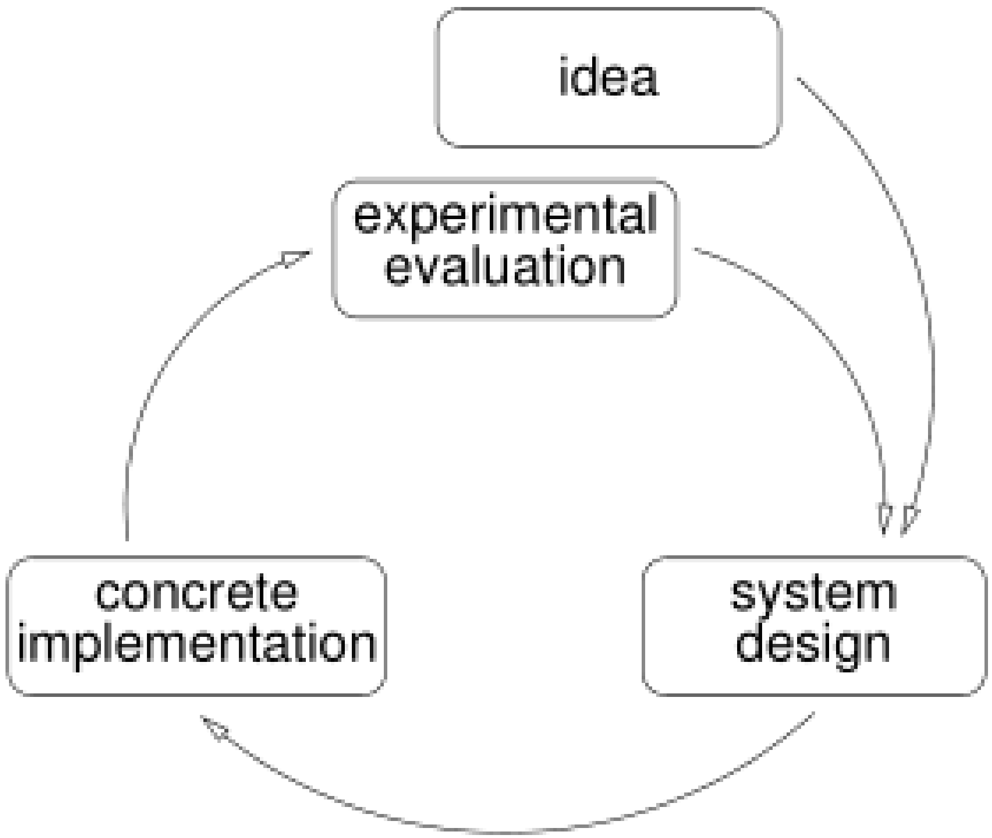
Scientific method (oversimplified)

- Use previous knowledge and observations to gain insight about a phenomenon
- Formulate hypothesis
- Construct a model of the phenomenon
- Use model (hypothesis) to predict outcomes
- Test hypothesis by experimenting
- Analyze outcome of experiment
- Go back to the beginning, refine, …

Scientific method (oversimplified)

- Hypothesis: Loud music affects the number of bugs in code over time
- Prediction: The number of bugs increases over time if sound is above 85 dB
- Experimental test: Measure the number of bugs obtained with several programming tasks over time by several programmers hearing heavy metal above and below 85 dB.
- Results: the number of bugs increases over time if sound is above 85 dB, but there is a large variance. Maybe there is other factor that affects the number of bugs? Code size ? Type of programming tasks? Type of music ?

Experiments in system/product design



(adapted from [1])

Experiments in system/product design

- Only 45% satisfaction with the current version of *Inforestudante* for course enrollment. Possible to improve the satisfaction by using less buttons in the UI. But this change adds cost.
- With the new UI, it is possible to reach 60% satisfaction or more
- Several students test the new UI prototype and fill a questionnaire to express their satisfaction.
- 58% satisfaction achieved. Does this contradict the hypothesis?

Typical computer science scenario

- A particular task needs to be solved by a software system
- This task is currently solved by an existing system (a baseline)
- You propose a new, in your opinion, better system
- You argue why your proposed system is better than the baseline
- You support your arguments by providing evidence that your system indeed beats the baseline

Key properties

- **Relevance**

Are the goals of the experiment and the expected results important for the progress?

- **Representativeness**

Is the experiment realist 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 other to reproduce the experiment?

- **Results analyses and generalization**

Is there enough information that allows to have sound conclusions and to generalize?

- **Cost**

Is the cost of the experiments compatible with the expected benefits?

Many types of experiments...

- Laboratory experiments
- Pilot studies
- Case studies
- Benchmarks
- Surveys

Many types of experiments...

- Laboratory experiments: Experimental investigation of a testable hypothesis, in which conditions are set up to isolate the variables of interest ("independent variables") and test how they affect certain measurable outcomes (the "dependent variables")
- Pilot studies
- Case studies
- Benchmarks
- Surveys

Many types of experiments...

- Laboratory experiments
- Pilot studies: Small experimental investigation to determine the feasibility (cost, duration) of performing a large-scale investigation
- Case studies
- Benchmarks
- Surveys

Many types of experiments...

- Laboratory experiments
- Pilot studies
- Case studies: A technique for detailed exploratory investigations of a subject of study that attempts to understand and explain phenomena or test theories
- Benchmarks
- Surveys

Many types of experiments...

- Laboratory experiments
- Pilot studies
- Case studies
- Benchmarks: A test or set of tests used to compare alternative tools or techniques. A benchmark comprises a motivating comparison, a task sample, and a set of measures (e.g., performance)
- Surveys

Many types of experiments...

- Laboratory experiments
- Pilot studies
- Case studies
- Benchmarks
- Surveys: A comprehensive system for collecting information to describe, compare or explain knowledge, attitudes and behavior over large populations

Design of experiments (a first look)

- Design of experiments (or experimental design) is the process of systematically defining and planning experiments in such a way that the data obtained can be analyzed to draw valid and objective conclusions
- **The goal is to design and perform valid experiments that allow good technical decisions: characterize and optimize process/product decisions**
- Basic idea:
  - Introduce controlled changes to input variables in order to to study their effect on an observable variable (or variables)
  - Get the maximum amount of information on cause and effect relationships with the minimum effort

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

Design of the experiment

6. Run experiments and collect the data/measurements — Measurements
7. Perform data analysis — Analysis
8. Draw conclusions (often go back to the beginning and reformulate the problem statement or test a different hypothesis) — Conclusions

Design of experiments (a first look)

Laboratory experiments, controlled experiments

1. Problem statement (or research question)
2. Identify var
3. Generate hy
4. Define the
5. Develop to
6. Run experi
7. Perform da
8. Draw conclusions (often go back to the beginning and 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?

Design of experiments (a first look)

Laboratory experiments, controlled experiments

1. Problem statement (or research question)
2. Identify var
3. Generate hy
4. Define the
5. Develop to
6. Run experi
7. Perform da
8. Draw conclusions (often go back to the beginning and 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

Design of experiments (a first look)

Laboratory experiments, controlled experiments

1. Problem statement (or research question)
2. Identify variables
3. Generate
4. Define th
5. Develop
6. Run expe
7. Perform
8. Draw co
   reformulat

- **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 continuous (e.g., time, size in bytes) or discrete (type of system, type of algorithm, etc.)

Design of experiments (a first look)

Laboratory experiments, controlled experiments

1. Problem statement (or research question)
2. Identify variables
3. Generate
4. Define th
5. Develop
6. Run expe
7. Perform
8. Draw co
   reformulat

- **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 dependent variable.

- **Full factorial**

  Change two or more factors simultaneously. It allows the study of possible interactions among factors

Design of experiments (a first look)

Laboratory exp

1. Problem state
2. Identify var
3. Generate
4. Define th
5. Develop
6. Run expe
7. Perform
8. Draw co
reformulat

Allow the study of possible interactions among factors

Terminology:

- **Repetition:** Reduce the experimental error
- **Randomization:** Minimize potential uncontrollable biases in the experiments by randomly assigning levels to 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).

Design of experiments (a first look)

Laboratory experimen

1. Problem statement (o
2. Identify variables
3. Generate hypothesis
4. Define the experiment
5. Develop tools and pro
6. Run experiments and
7. Perform data analysis
8. Draw conclusions (ofte
   reformulate the problem st

- **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)

Design of experiments (a first look)

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

Laboratory exp

1. Problem state
2. Identify variab
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
8. Draw conclusions (often go back to the beginning and reformulate the problem statement or test a different hypothesis)

Design of experiments (a first look)

Laborat

1. Probl
2. Identi
3. Gene
4. Defin
5. Develop tools and procedures for the experiment
6. Run experiments and collect the data/measurements
7. Perform data analysis
8. Draw conclusions (often go back to the beginning and reformulate the problem statement or test a different hypothesis)

- 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

Design of experiments (a first look)

Laborat

1. Probl
2. Ident
3. Gene
4. Defi
5. Deve

- **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,…)

6. Run experiments and collect the data/measurements
7. Perform data analysis
8. Draw conclusions (often go back to the beginning and reformulate the problem statement or test a different hypothesis)

Design of experiments (a first look)

Laborat

1. Probl
2. Ident
3. Gene
4. Defi
5. Deve
6. Run
7. Perform data analysis
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

Readings

- Dror G. Feitelson, *Experimental Computer Science: The Need for a Cultural Change*
- One question in the exam is related to this article.
- Suggestion: write a summary of the article