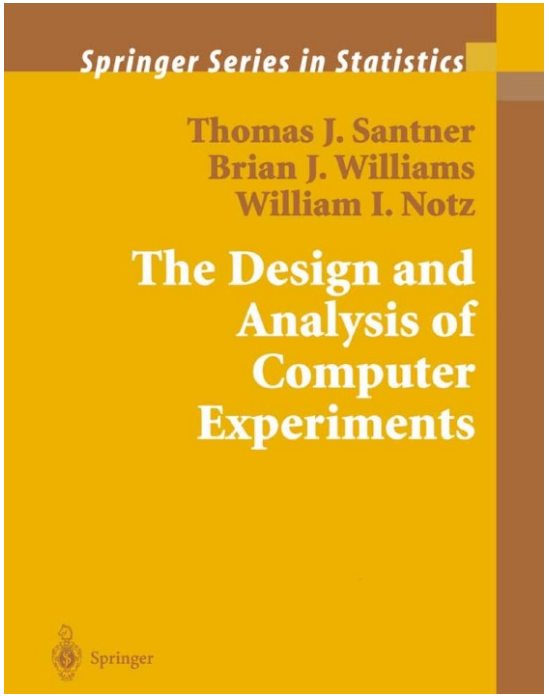
CIC0203 - COMPUTAÇÃO EXPERIMENTAL - TA (2021.1 - 35T23)
Abordagens à Computação Experimental

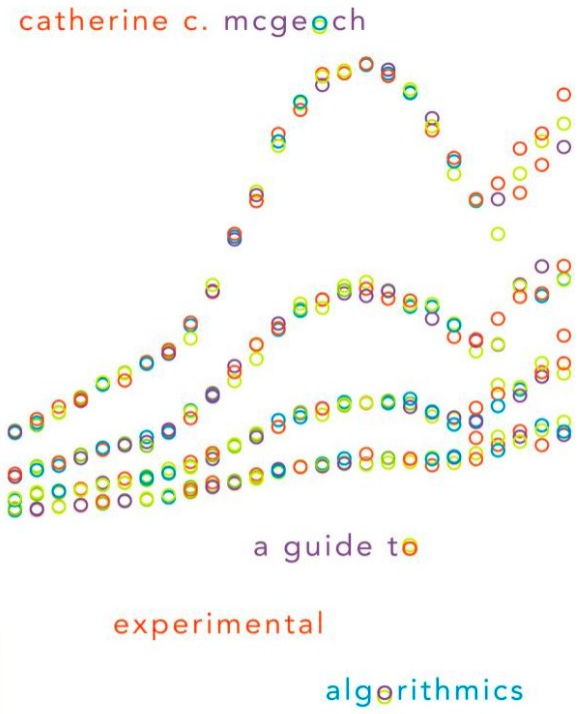
Prof. Jorge H C Fernandes
Março de 2022

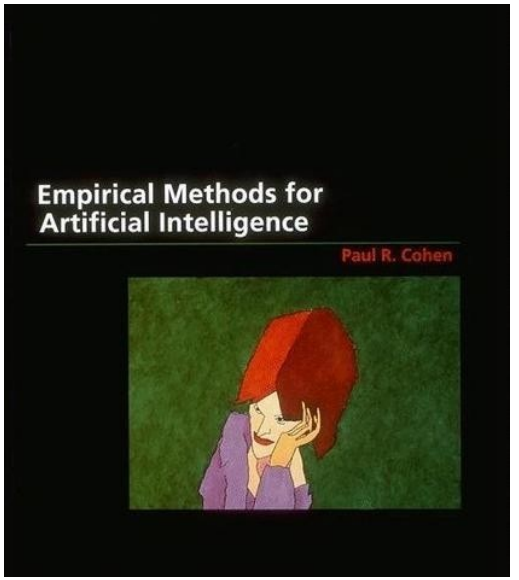
Abordagens à Computação Experimental

- Experimentos Computacionais
- Algorítmica Experimental ou Engenharia de Algoritmos Empírica
- Computação Empírica

Abordagens à Computação Experimental







Santner et al., 2003

McGeogh, 2007

Cohen, 1995

https://web.archive.org/web/20191105212926/https://www.cs.amherst.edu/alglab/labs.html

Experimentos Computacionais, segundo Santner, 2003, p. vii

- Many **physical processes** are difficult or even impossible to study by conventional experimental methods. As computing power has increased, it has become possible to **model some of these processes by sophisticated computer code**. In such cases, **the code can serve as a proxy for the physical (químico, biológico, ecológico, social) process**. As in a physical experiment, **one can vary the inputs to the code and observe how the process output is affected**. **Such studies are called computer experiments** and are becoming increasingly popular surrogates for and adjuncts to physical experiments.
- **Much of the methodology for designing, modeling, and analyzing computer experiments can only be found in research papers**.
- Our goal in writing this book is to make these methods accessible to a more general audience. To accomplish this, we have tried to keep the mathematics at a level that will be understandable to readers with Masters level training in Statistics. This has been a challenging task.

Experimentos Computacionais versus Experimentos Físicos, segundo Santner, 2003, p. 1-17

Experimental Algorithmics

Algoritmica Experimental ou Engenharia de Algoritmos Empírica

https://en.wikipedia.org/wiki/Empirical_algorithmics

- In computer science, empirical algorithmics (or experimental algorithmics) is the practice of using empirical methods to study the behavior of algorithms.
- The practice combines algorithm development and experimentation: algorithms are not just designed, but also implemented and tested in a variety of situations. In this process, an initial design of an algorithm is analyzed so that the algorithm may be developed in a stepwise manner.

Algoritmica Experimental ou Empírica ou Engenharia de Algoritmos Visão Geral

https://en.wikipedia.org/wiki/Empirical_algorithmics

- Methods from empirical algorithmics complement theoretical methods for the analysis of algorithms. Through the principled application of empirical methods, particularly from statistics, it is often possible to obtain insights into the behavior of algorithms such as high-performance heuristic algorithms for hard combinatorial problems that are (currently) inaccessible to theoretical analysis. Empirical methods can also be used to achieve substantial improvements in algorithmic efficiency.
- American computer scientist Catherine McGeoch identifies two main branches of empirical algorithmics:
  - the first (known as empirical analysis) deals with the analysis and characterization of the behavior of algorithms, and
  - the second (known as algorithm design or algorithm engineering) is focused on empirical methods for improving the performance of algorithms.
- The former often relies on techniques and tools from statistics, while the latter is based on approaches from statistics, machine learning and optimization.
- Dynamic analysis tools, typically performance profilers, are commonly used when applying empirical methods for the selection and refinement of algorithms of various types for use in various contexts.

Experimental Algorithmics
Why Experiment? (McGeoch, p. 3)

- Experimental algorithmics represents a third approach that treats algorithms as laboratory subjects, emphasizing control of parameters, isolation of key components, model building, and statistical analysis. This is distinct from the purely empirical approach, which studies performance in “natural settings,” in a manner akin to field experiments in the natural sciences.
- Instead, experimental algorithmics combines the tools of the empiricist – code and measurement – with the abstraction-based approach of the theoretician. Insights from laboratory experiments can be more precise and realistic than pure theory provides, but also more general than field experiments can produce.
- This approach complements but does not replace the other two approaches to understanding algorithm performance. It holds promise for bridging the long-standing communication gap between theory and practice, by providing a common ground for theoreticians and practitioners to exchange insights and discoveries about algorithm and program performance.

Experimental Algorithmics
algorithm engineering (McGeoch, p. 3)

- Similarly, the experimental approach has made important contributions to prob-lems in algorithm design and engineering. The term algorithm engineering has been coined to describe a systematic process for transforming abstract algorithms into production-quality software, with an emphasis on building fast, robust, and easy-to-use code.
- Algorithm design, which focuses on implementation and tuning strategies for specific algorithms and data structures (see Chapter 4 for details), is just one part of the larger algorithm engineering process, which is also concerned with requirements specification, interfaces, scalability, correctness, and so forth.

Escalas de instanciação de programas, segundo Mcgeoch (2007, p. 6)

- We make no qualitative distinction here between “algorithms” and “programs.” Rather, we consider algorithms and programs to represent two points on a scale of instantiation, according to how much specificity is in their descriptions. Here are some more recognizable points on this scale:
  - At the most abstract end are **metaheuristics and algorithm paradigms**, which describe generic algorithmic structures that are not tied to particular problem domains. For example, Dijkstra’s algorithm is a member of the greedy paradigm, and tabu search is a metaheuristic that can be applied to many problems.
  - The **algorithm** is an abstract description of a process for solving an abstract problem. At this level we might see Dijkstra’s algorithm written in pseudocode. The pseudocode description may be more or less instantiated according to how much detail is given about data structure implementation.
  - The **source program** is a version of the algorithm implemented in a particular high-level language. Specificity is introduced by language and coding style, but the source code remains platform-independent. Here we might see Dijkstra’s algorithm implemented in C++ using an STL priority queue.
  - The **object code** is the result of compiling a source program. This version of the algorithm is written in machine code and specific to a family of architectures.
  - The **process** is a program actively running on a particular machine [N.E: using an assembly of object codes and operating system services, APIs, or system calls] at a particular moment in time. Performance at this level may be affected by properties such as system load, the size and shape of the memory hierarchy, and process scheduler policy.
- Interesting algorithmic experiments can take place **at any point on the instantiation scale**. We make a conceptual distinction between the experimental subject, which is instantiated somewhere on the scale, and the test program, which is implemented to study the performance of the subject.

Cohen
Métodos empíricos em Inteligência Artificial

Pesquisa e Evidência Empírica

https://en.wikipedia.org/wiki/Empirical_research

- Empirical research is research using empirical evidence.
- It is also a way of gaining knowledge by means of **direct** and **indirect observation** or **experience**.
- Empiricism values some research more than other kinds.
- Empirical evidence (the record of one's direct observations or experiences) can be analyzed quantitatively or qualitatively.
- **Quantifying** the evidence or **making sense of it in qualitative form**, a researcher can answer empirical questions, which should be clearly defined and answerable with the evidence collected (usually called data).
- **Research design** varies by field and by the question being investigated. Many researchers combine qualitative and quantitative forms of analysis to better answer questions which cannot be studied in laboratory settings, particularly in the social sciences and in education.

Métodos empíricos

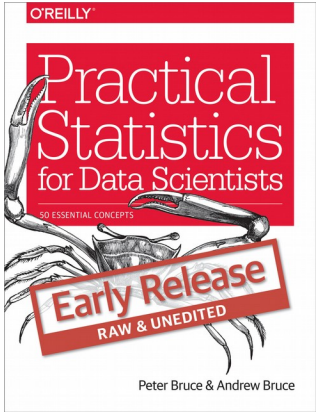
Empirical Research, cf. Cohen (1995, p. 1)

- Empirical methods cluster loosely into
  - exploratory techniques for visualization, summarization, exploration, and modeling;
  - and confirmatory procedures for testing hypotheses and predictions.
- In short,
  - empirical = exploratory + experimental.

Uso de Métodos Estatísticos

Empirical Research, cf. Cohen (1995, p. 1)

- Because empirical studies usually observe nondeterministic phenomena, which exhibit variability, both exploratory and experimental methods are based in statistics.
- Exploratory statistical methods are called, collectively, **exploratory data analysis**;
- whereas methods for **confirmatory experiments** go by the name statistical hypothesis testing.

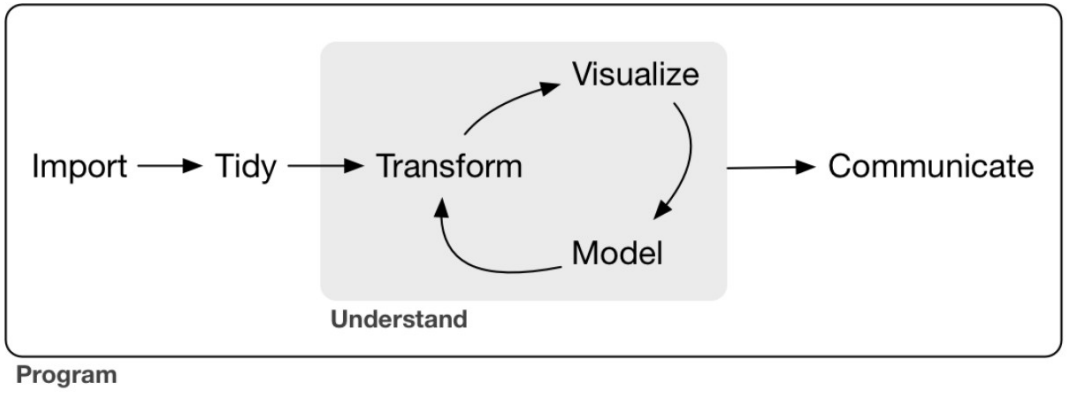


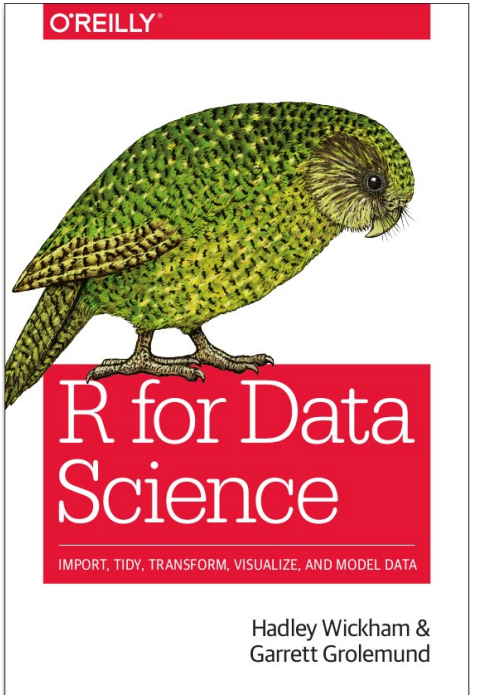
Dificuldades e Facilidades da Abordagem Confirmatória

Empirical Research, cf. Cohen (1995, p. 2)

- Experiments and hypothesis testing answer yes-or-no questions. You can reduce all your research interests to a series of yes-or-no questions, but you'll find it slow work.
- A more efficient approach is to flip back and forth between exploratory and experimental work, engaging in the latter only when the former produces a question you really want to answer.

Análise Exploratória de Dados =~ Data Science





Empiricismo e Empírico, conforme American Heritage Dictionary

Empirical Research, cf. Cohen (1995, p. 2)

- Empirical
  - (1) **Relying upon or derived from observation or experiment: empirical methods.**
  - (2) Relying solely on practical experience and without regard for system or theory
- Empiricism
  - (1) The view that experience, esp. of the senses, is the only source of knowledge.
  - (2) **The employment of empirical methods, as in science**.

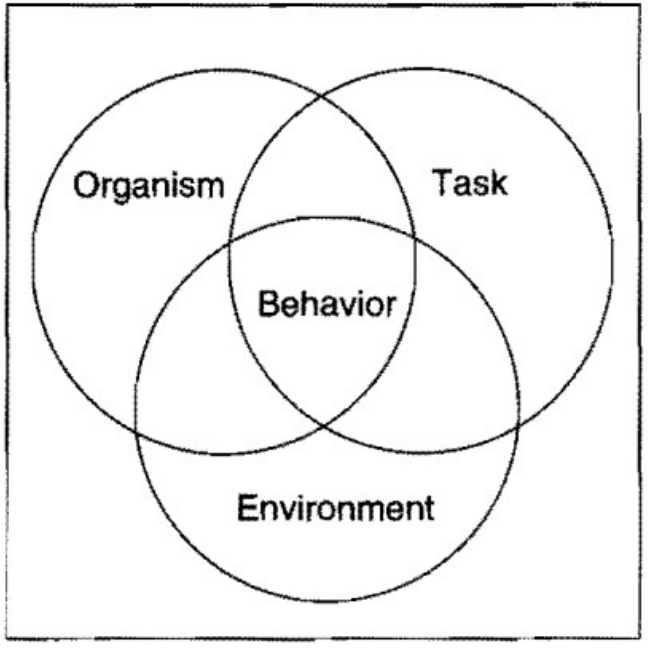
Programas de computador enquanto sujeitos de investigação

Empirical Research, cf. Cohen (1995, p. 2)

- Our subject is empirical methods for studying ...AI... programs, methods that involve running programs and recording their behaviors.
- Unlike other scientists, who study chemical reactions, processes in cells, bridges under stress, animals in mazes, and so on, we study computer programs that perform tasks in environments.

Espaço de variabilidade dos experimentos e três perguntas básicas

Empirical Research, cf. Cohen (1995, p. 3-4)



- How will a change in the agent's structure affect its behavior given a task and an environment?
- How will a change in an agent's task affect its behavior in a particular environment
- How will a change in an agent's environment affect its behavior on a particular task?

Figure 1.1 How the structure, task, and environment of an organism influence its behavior.

Discussão: Estudantes devem exemplificar três perguntas básicas, para um caso

- How will a change in the agent's structure affect its behavior given a task and an environment?
- How will a change in an agent's task affect its behavior in a particular environment
- How will a change in an agent's environment affect its behavior on a particular task?

Capacidade de generalização científica em função da melhoria do entendimento

Empirical Research, cf. Cohen (1995, p. 5)

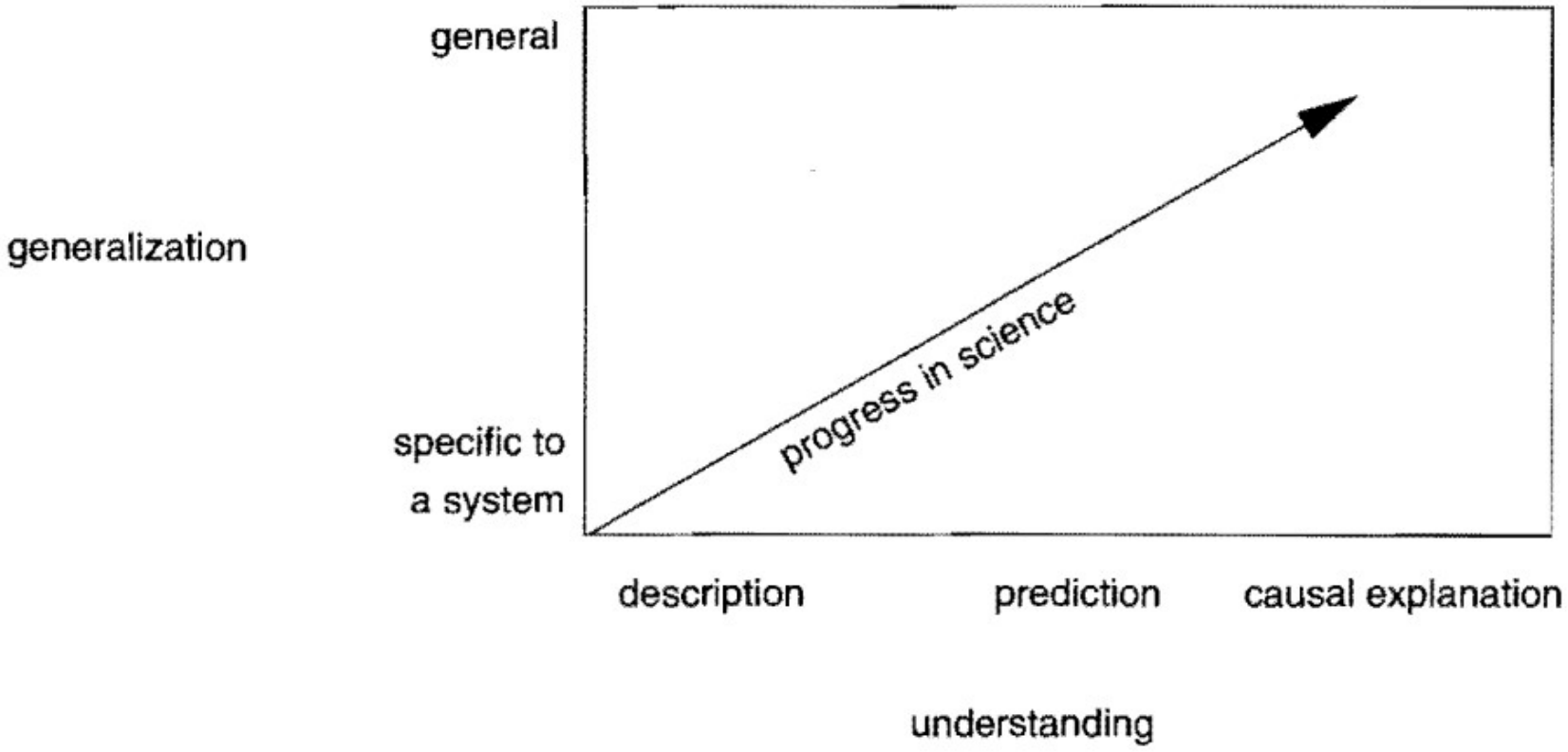


Figure 1.2 Generalization and understanding define a space of basic research questions.

Capacidade de generalização científica em função da melhoria do entendimento

Empirical Research, cf. Cohen (1995, p. 5)

- Early in a project we ask **short questions that often have lengthy descriptions** of behavior as answers;
- **later, the questions are lengthy** because they refer to predictive and causal models, **but the answers are short**. often just "yes" or "no,"
- This shift in balance characterizes progress from exploratory studies to experimental studies.

Abordagem Empírica de Cohen

Empirical Research, cf. Cohen (1995, p. 6)

The empirical generalization strategy, …, goes like this:

1. build **a program that exhibits a behavior of interest** while performing **particular tasks** in **particular environments**;

2. identify specific **features** of the program, its **tasks** and **environments** that influence the target behavior;

3. develop and **test a causal model** of how these features influence the target behavior;

4. once the **model makes accurate predictions**, generalize the features so that other programs, tasks, and environments are encompassed by the causal model:

5. **test whether the general model predicts accurately the behavior** of this larger set of programs, tasks, and environments.

Quatro tipos de estudos empíricos, segundo Cohen

Empirical Research, cf. Cohen (1995, p. 7)

- Exploratory studies
  - **yield causal hypotheses** that are tested in observation or manipulation experiments. To this end, exploratory studies usually collect lots of data, analyzing it in many ways to find regularities.
- Assessment studies
  - **establish baselines and ranges**, and other assessments of the behaviors of a system or its environment.
- Observation experiments
  - **disclose effects of factors on measured variables** by observing associations between levels of the factors and values of the variables. These are also called natural and quasi-experimental experiments.
- Manipulation experiments
  - **test hypotheses about causal influences of factors by manipulating them and noting effects**, if any, on one or more measured variables.

Estudos empíricos: Exploratory studies, segundo Cohen

Empirical Research, cf. Cohen (1995, p. 7)

- Exploratory studies **yield causal hypotheses** that are tested in observation or manipulation experiments. To this end, exploratory studies usually collect lots of data, analyzing it in many ways to find regularities.

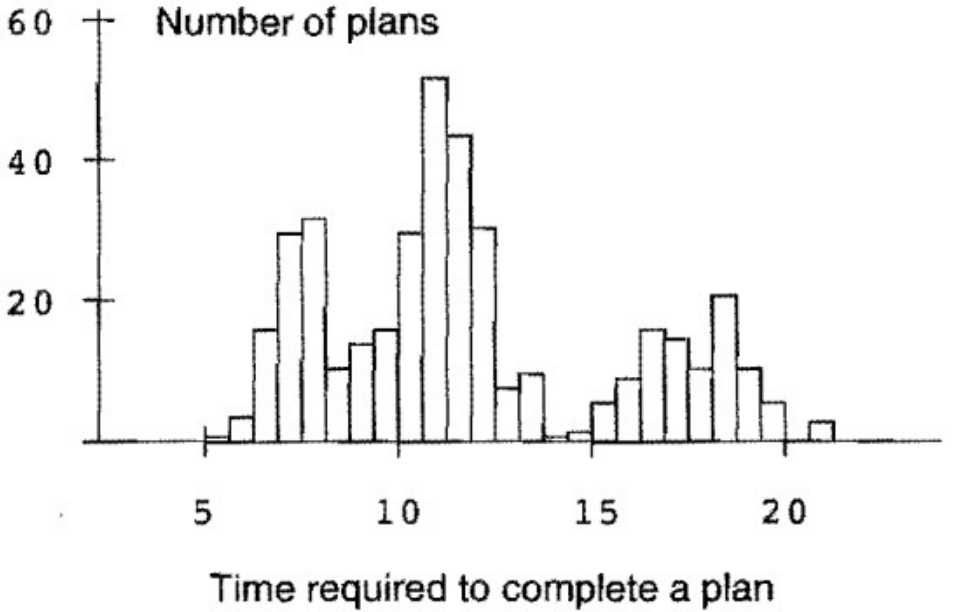


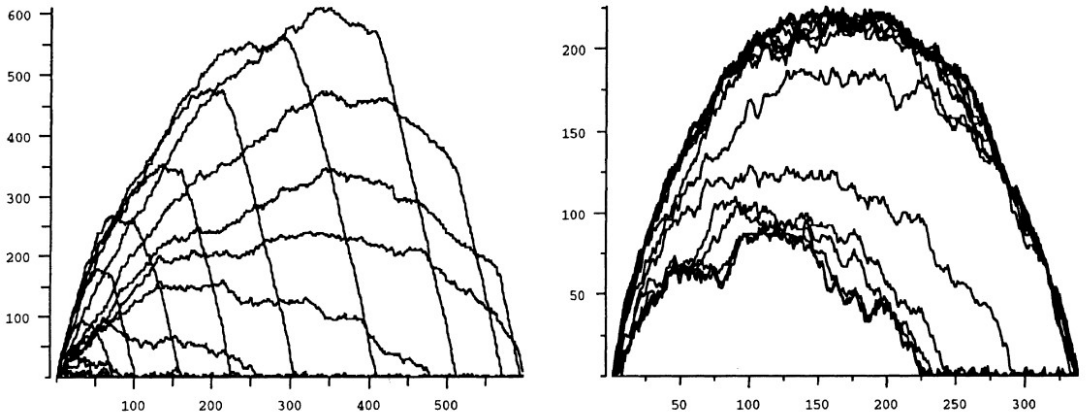
Figure 1.3 A frequency distribution.

What does the distribution in figure 1.3 suggest to you?

Estudos empíricos: Assessment studies, segundo Cohen

Empirical Research, cf. Cohen (1995, p. 7)

- Assessment studies **establish baselines and ranges**, and other assessments of the behaviors of a system or its environment.



(e) KOSO with $p = 16$ PEs (f) KOSO* with $p = 16$ PEs

Fig. 4. Queue sizes over the course of a trial.

Fonte: BARROWS, Miranda E.; GREGORY, Dawn E.; GAO, Lixin; et al. An empirical study of dynamic scheduling on rings of processors. Parallel Computing, v. 25, n. 9, p. 1063–1079, 1999.

Quais são os limites de desempenho dessas arquiteturas?

Estudos empíricos: Observation experiments, segundo Cohen

Empirical Research, cf. Cohen (1995, p. 7)

- Observation experiments **disclose effects of factors on measured variables** by observing associations between levels of the factors and values of the variables.

Estudos empíricos: Manipulation experiments, segundo Cohen

Empirical Research, cf. Cohen (1995, p. 7)

- Manipulation experiments
  - **test hypotheses about causal influences of factors by manipulating them and noting effects**, if any, on one or more measured variables.

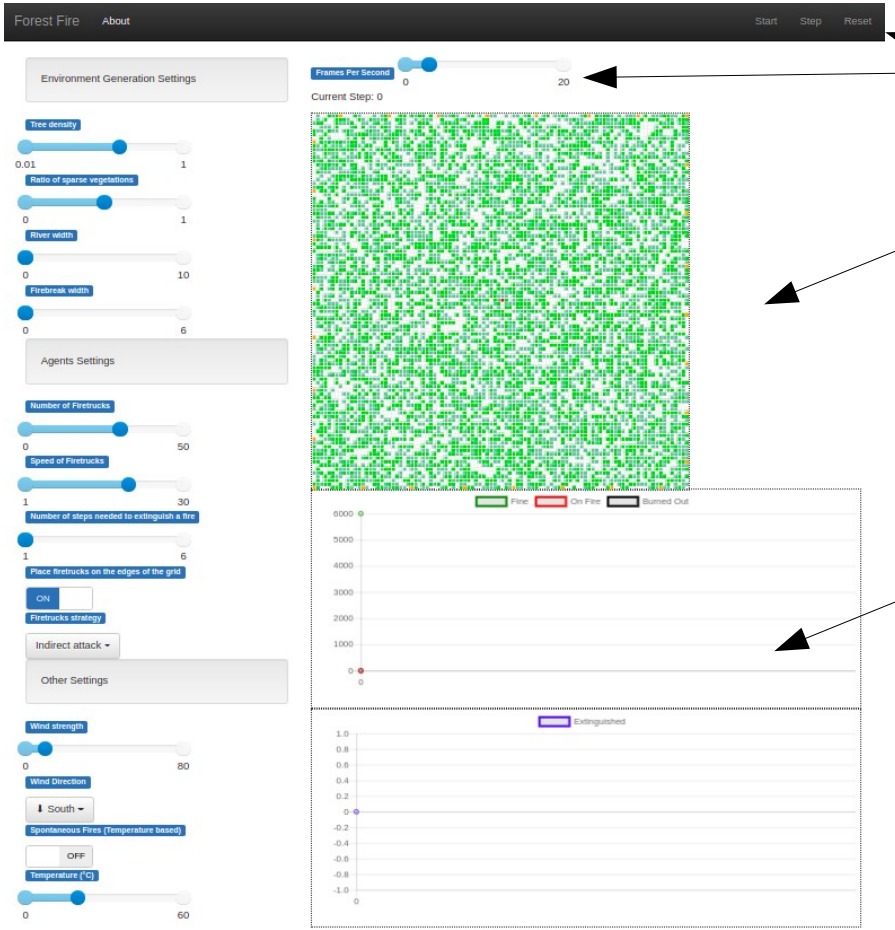
Fire-suppression-abm
https://github.com/hildobby/fire-suppression-abm)

Fire-suppression-abm (https://github.com/hildobby/fire-suppression-abm)
Uma instância (simulação) sob potencial experimentação

Parâmetros Ambientais

Parâmetros do agente

Parâmetros da Tarefa



Controle da simulação

Comportamento da Simulação

Sumário de Algumas Mensurações

Fire-suppression-abm (https://github.com/hildobby/fire-suppression-abm) README (1)

- Agent-based Modeling: Wildfire prevention simulation using agent-based modeling written in Python 3+
- This project uses [Mesa](https://github.com/projectmesa/mesa), an Apache2 licensed agent-based modeling (or ABM) framework in Python.
- This repository allows users to simulate a wildfire in a randomly generated forest environment. The random generation is based on user settable settings from the visualisation tool provided by mesa. This simulation is created in a 2D grid of size 100x100. The firefighting agents will attempt to extinguish said fire using a strategy chosen by the user, the firefighter's success depends on the settings used in the simulation. The program also allows for sensitivity analysis with a built-in script.

Fire-suppression-abm (https://github.com/hildobby/fire-suppression-abm) README (2)

- Agent-based Modeling: Wildfire prevention simulation using agent-based modeling written in Python 3+
- This project uses [Mesa](https://github.com/projectmesa/mesa), an Apache2 licensed agent-based modeling (or ABM) framework in Python.
- This repository allows users to simulate a wildfire in a randomly generated forest environment. The random generation is based on user settable settings from the visualisation tool provided by mesa. This simulation is created in a 2D grid of size 100x100. The firefighting agents will attempt to extinguish said fire using a strategy chosen by the user, the firefighter's success depends on the settings used in the simulation. The program also allows for sensitivity analysis with a built-in script.
- Features
  - User settable settings for environment generation such as wind direction, wind strength, rivers, rain, # of firefighting agents and other parameters.
  - Multiple firefighting strategies ('Go to the closest fire', 'Go to the biggest fire', 'Random movements', 'Parallel attack' and 'Indirect attack')
  - Sensitivity analysis (One-factor-a-time OFAT) of the environmental settings

Fire-suppression-abm (https://github.com/hildobby/fire-suppression-abm) README (3)

- **Cloning the repository**
  - git clone https://github.com/hildobby/fire-suppression-abm.git
- To download all the packages using `pip`, navigate to the repository's local directory and run the following:
  - pip install -r "requirements.txt"
- To run the simulation with the GUI in python, run the following fromt your cloned repository's local directory:
  - python3 src/server.py
- Running the Sensitivity Analysis
  - python3 src/sensitivity_analysis/ofat_sa.py
- where the built-in BatchRunner of mesa is used. More precisely, the child class BatchRunnerMP is used which allows for parallel computing. One needs to determine manually which parameters to feed to the mesa build in BatchRunner such as the wind strength, the bounds and the number of cores to use.
- What remains to be worked on
  - Add tests and use codecov
  - Adding new types of agents
  - Implementing new fire fighting methods
  - Make agents able to change their method depending on the circumstances

Fire-suppression-abm (https://github.com/hildobby/fire-suppression-abm) README (4)

- Alexandridis, A., Vakalis, D., Siettos, C. I., & Bafas, G. V. (2008). **A cellular automata model for forest fire spread prediction: The case of the wildfire that swept through spetses island in 1990**. Applied Mathematics and Computation, 204 (1), 191–201
- Brooks, M. L., D'antonio, C. M., Richardson, D. M., Grace, J. B., Keeley, J. E., DiTomaso, J. M.,... Pyke, D. (2004). **Effects of invasive alien plants on fire regimes.** BioScience, 54 (7), 677–688
- Cheney, N. P., Gould, J. S., McCaw, W. L., & Anderson, W. R. (2012). **Predicting fire behaviour in dry eucalypt forest in southern australia.** Forest Ecology and Management, 280, 120-131. Retrieved from http://www.sciencedirect.com/science/article/pii/S0378112712003325 doi:https://doi.org/10.1016/j.foreco.2012.06.012
- DeBano, L. F., Neary, D. G., & Ffolliott, P. F. (1998). **Fire effects on ecosystems.** John Wiley & Sons. Grimm, V., Berger, U., DeAngelis, D. L., Polhill, J. G., Giske, J., & Railsback,S. F. (2010).**The odd protocol:A review and first update**. Ecological Modelling, 221 (23), 2760-2768. Retrieved from http://www.sciencedirect.com/science/article/pii/S030438001000414X doi: https://doi.org/10.1016/j.ecolmodel.2010.08.019
- Hansen, R. (2012, 05). **Estimating the amount of water required to extinguish wildfires under different conditions and in various fuel types.** International Journal of Wildland Fire, 21, 525-536. doi: 10.1071/WF11022 CO
- Hu, X., & Sun, Y. (2007). **Agent-based modeling and simulation of wildland fire suppression.** In 2007 winter simulation conference (pp. 1275-1283).
- Lee, Y.-H., Fried, J., Albers, H., & Haight, R. (2012, 11). **Deploying initial attack resources for wildfire suppression: Spatial coordination, budget constraints, and capacity constraints.** Canadian Journal of Forest Research, 43, 56-65. doi: 10.1139/cjfr-2011-0433
- McKinney, W. (2012). **Python for data analysis: Data wrangling with pandas, numpy, and ipython**." O'Reilly Media, Inc.".
- Millman, K. J., & Aivazis, M. (2011). **Python for scientists and engineers.** Computing in Science & Engineering, 13 (2), 9–12.
- Nolan, R. H., Boer, M. M., Collins, L., Resco de Dios, V., Clarke, H., Jenkins, M.,... Bradstock, R. A. (2020). **Causes and consequences of eastern australia's 2019-20 season of mega-fires.** Global change biology.
- Russoa, L., Vakalisb, D., & Siettos, C. (2013). **Simulating the wildfire in rhodes in 2008 with a cellular automata model.** CHEMICAL ENGINEERING, 35.
- Shinneman, D. J., Germino, M. J., Pilliod, D. S., Aldridge, C. L., Vaillant, N. M., & Coates, P. S (2019). **The ecological uncertainty of wildfire fuel breaks: examples from the sagebrush steppe**. Frontiers in Ecology and the Environment, 17 (5), 279–288.
- Williams, F. (1978). **Mechanisms of fire spread**. In Symposium (international) on combustion (Vol. 16, pp. 1281–1294)

Fire-suppression-abm (https://github.com/hildobby/fire-suppression-abm) Descrição dos parâmetros do simulador

- About Forest Fire
  - Create a forest fire model with fire fighting agents that try to extinguish the fire
- Args:
  - height, width: The size of the grid to model;
  - density: What fraction of grid cells have a tree in them;
  - temperature: influences the number of spontaneous fires;
  - truck_strategy: the tactic that firetrucks will adhere to;
  - river_width: in the case that a river is present, what is its width;
  - break_width: in the case that a pre-made fire break is present, what is its width;
  - random_fires: boolean indicating whether spontaneous fires are present;
  - vision: the distance fire fighting agents can look around them;
  - truck_max_speed: the max speed with which firetruck agents can move around (grid cells/step);
  - wind_strength: the speed with which the wind moves;
  - wind_dir: string specifying the direction of the fire;
  - sparse_ratio: the fraction of the vegetation that is sparse instead of dense;
  - steps_to_extinguishment: number of steps it takes for firetrucks to extinguish a burning cell;
  - placed_on_edges: indicates whether the firetrucks are placed randomly over the grid, or equispaced on the rim

Fire-suppression-abm

Exercício 1

- Baixe e execute o fire-supression-lab, conforme as instruções
  - https://github.com/hildobby/fire-suppression-abm
- Leia atentamente as descrições dos parâmetros da simulação, e execute pelo menos dez simulações que exploram o funcionamento do simulador, para ganhar intuição de como o simulador funciona.
- Faça uma proposta de agrupamento dos parâmetros apresentados no “about” do simulador, justificando a classificação de cada parâmetro conforme os conceitos apresentados por Cohen (1999, p.3-4).
- Descreva, em até 500 palavras, a sua experiência com o uso do simulador, e como você acredita que o simulador fire-supression-lab possa ser usado em uma experimentação computacional, usando necessariamente os conceitos de:
  - **Procedimento**
  - **Hipótese**
  - **Causa e efeito**
  - **Demonstração de resultados**
  - **Manipulação de fatores**
  - **Procedimentos repetíveis**
  - **Análise lógica dos resultados**
- Registre no fórum de discussões no Moodle: 1 – sua proposta de agrupamento; 2 – Possibilidade de uso do simulador
- Faça uma crítica construtiva, sugerindo aprimoramentos à proposta de agrupamento de um colega, à maneira como ele descreveu as possibilidades de uso do simulador.

Cohen, Capítulo 2

Análise Exploratória de Dados

Empirical Research, cf. Cohen (1995, p. 11-65)

Motivação: Análise Exploratória de Dados

Empirical Research, cf. Cohen (1995, p. 11)

- A crude but useful distinction can be made between confirmatory experiments and exploratory studies.
- When experiments are reported in journals and proceedings, one gets the impression that the authors started with a clear question and breezed through the experiment to answer it.
- In reality, however, confirmatory experiments are the product of a sometimes lengthy process in which vague questions are refined to precise ones and pilot experiments that produce murky data are refined to produce crisp results.
- **Exploratory studies, then, are the informal prelude to experiments**. Exploratory studies are like working in a test kitchen whereas experiments are like recipes in cook books.

Análise Exploratória de Dados: Os Dados

Empirical Research, cf. Cohen (1995, p. 12)

- The act that defines a project as empirical [artificial intelligence/computing] is **running a program and collecting data**.
- The data might be:
  - execution traces,
  - the percentage of correct answers,
  - the time to complete a task,
  - the trees generated by a search process,
  - the number of times a process is interrupted,
  - the frequency with which a piece of knowledge is used,
  - or the difference between the program's answer and a human judge's answer.

Análise Exploratória de Dados: Quando os Dados são Coletados?

Empirical Research, cf. Cohen (1995, p. 12)

- Data are sometimes collected **while the program is running** (e.g., execution traces) and sometimes are collected **after the program has finished** (e.g., the percentage of correct answers).
- In most experiments a **program will be run more than once, and something will be slightly different on each trial**.
- For example, the program might be given ten slightly different problems to solve.

Análise Exploratória de Dados: Amostras e Medidas

Empirical Research, cf. Cohen (1995, p. 12)

- A sample is a collection of measurements on individuals [N.E: instância]: hair color for a sample of five colleagues, or quality ratings of a sample of 20 imported beers, **or the times required by a program [N.E: instância] to solve some problems**.
- It is common to make many measurements of each member [N.E: instância] of a sample.
- For an **individual run of a program [N.E: instância]** you might measure
  - run time,
  - whether the answer was correct,
  - how many times particular pieces of knowledge were used,
  - and so on.

Análise Exploratória de Dados: Variáveis e Representações (Abstrações)

Empirical Research, cf. Cohen (1995, p. 13)

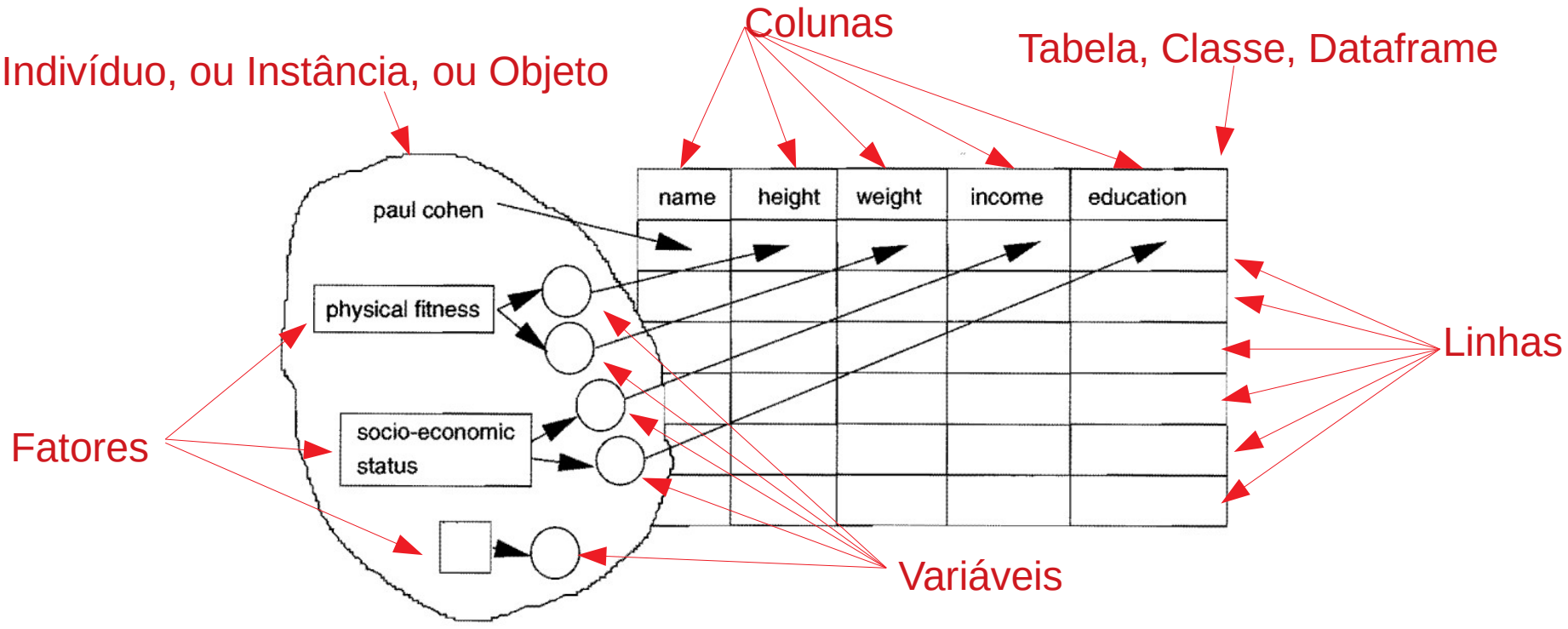


Figure 2.1 Viewing variables as functions that generate representations.

Factors are abstract characteristics of individuals

Variables, represented by circles, produce measurements of factors.

columns are variable names and rows represent individuals

Fire-suppression-abm

Exercício 1

1) Faça uma análise exploratória de um dos arquivos CSV disponibilizados no diretório src/sensitivity_analysis/data, usando Cohen (1999, p. 13-15)

2) Descreva, textualmente, o que significa o arquivo analisado que você escolheu no passo 1

3) Faça um script R que apresenta os dados de forma tabular, usando os termos distribuição univariada, distribuição multivariada, partição, como expostos por Cohen (1999, p. 13-15)

4) Faça pelo menos uma plotagem de distribuição de frequência de um dos dados, e apresente sua interpretação desses dados.

5) Descreva, após sua análise, pelo menos um achado de interesse para análise experimental do simulador

6) Apresente seus resultados no Fórum 2

CIC0203 - COMPUTAÇÃO EXPERIMENTAL - TA (2021.1 - 35T23)
Abordagens à Computação Experimental

Prof. Jorge H C Fernandes
Março de 2022