What is “Ph.D.”?

http://gizmodo.com/5613794/what-exactly-a-doctorate
“THE” Method in Physics

- Form hypothesis about a natural phenomenon
- Design experiment
- Collect data
- Compare data to hypothesis
- Accept or reject hypothesis
  - Publish (in Nature)
  - Get someone else to repeat experiment (replication)

Method in Medicine

- Form hypothesis about a treatment
- Select experimental and control groups that are comparable except for the treatment
- Collect data
- Commit statistics on the data
- Treatment ⇒ difference (statistically significant)
Research Methods in CI

- Rough classification of the methods
  - Survey
  - Formal
  - Experimental
  - Build
  - Process
  - Model
  - ...

Survey Method

- What is known?
- How is one the same/different from other?
- What questions are still open?

- Need to provide new perspective, taxonomy, classification, interpretation, understanding,
  ...

5

6
Formal Method

- Often used in theoretical computing
  - To prove facts about algorithms (e.g., complexity, P vs. NP) and systems (e.g., quality, verification)
- Often use:
  - Mathematical proof
  - Algorithm design/analysis
  - Complexity/automata theory

Experimental Method

- Broadly used in CI to evaluate new solutions
- Often in two phases:
  - Exploratory phase: taking measurements that will help identify what are the questions that should be asked about the system under evaluation
  - Evaluation phase will attempt to answer these questions
- A well-designed experiment will start with a list of the questions that the experiment is expected to answer
Experimental Method (cont)

- Needs good record keeping in experiments
- Repeatability
  - Careful design
    - Questions, variables, measures
  - Reporting results
    - Careful in using aggregates (e.g., AVG)

Experimental Reproducibility

The goal of producing reproducible results is to ensure your research is robust, verifiable, and valuable for future research. The purpose is that experimental results will be useful when they have been tested and generated by objective third parties.

The Review Process

The committee consists of reviewers, who are accomplished in the field. They are expected to continue the review process even if the submission process is not submitted to another major publication. The committee takes the decision to submit or not the following criteria:

- Reproducible: the experiments reproduced by the committee support the statistical results reported in the paper.
- Statistical: the experiments are made available for the committee and they have been tested by the committee in the future.
- If the paper does not meet the criteria, the committee may ask for further evidence or reject the contribution.

Eg, In 2005, it was reported that “average (via mean) household income in UK fell by 0.2%...”
- Distribution was not Gaussian
- In reality, “average (via median) household income increased by 0.5%”
**Build Method**

- Building an artifact (e.g., physical or S/W system) to demonstrate that it is possible
  - E.g., feasibility testing
- To be considered research
  - The construction of the artifact must be *new* or
  - It must include *new* features that have not been demonstrated before in other artifacts
- E.g., the early web browser

---

**Process Method**

- Used to understand the processes used to accomplish tasks in CI
- Useful in the study of activities that involve “humans”
  - E.g., Software Engineering, HCI, and games
- The study of processes may also be used to understand cognition in AI
- Check prof. Poole’s materials on this method for more details
Model Method

- How to define an abstract model for a real system
- Model allows to better understand a system
- Model allows to perform experiments that could not be performed in the system itself because of cost or accessibility
- Often used together with other 4 methods
  - Eg, experiments based on a model → simulations
  - Eg, formal description of the model to verify the functionality or correctness of a system → model checking

Model Method (cont)

- Q: how to build more powerful web server?

- Use the model method to "model" the workload (the form and pattern of requests) to a web server using particular distributions
- These distributions are used to express:
  - time between request arrivals
  - the sizes of documents retrieved
  - the probabilistic relationship between successive item requests
## Examples of Research Questions

**Existence:**
- Does X exist?

**Description & Classification**
- What is X like?
- What are its properties?
- How can it be categorized?
- How can we measure it?
- What are its components?

**Descriptive-Comparative**
- How does X differ from Y?

**Frequency and Distribution**
- How often does X occur?
- What is an average amount of X?

**Descriptive-Process**
- How does X normally work?
- By what process does X happen?
- What are the steps as X evolves?

**Relationship**
- Are X and Y related?
- Do occurrences of X correlate with occurrences of Y?

**Causality**
- What causes X?
- What effect does X have on Y?
- Does X cause Y?
- Does X prevent Y?

**Causality-Comparative**
- Does X cause more Y than does Z?
- Is X better at preventing Y than is Z?
- Does X cause more Y than does Z under one condition but not others?

**Design**
- What is an effective way to achieve X?
- How can we improve X?

---

## Reference

- About Computing Science Research Methodology, José Nelson Amaral

- Research Method in Computer Science, Serge Demeyer