

So You Want to be a Scientist?

What is an Experiment, Anyway?

Computer Science & Research?!

- Many students view CS as an applied science, not as a “pure” science
 - Writing programs
 - Building Networks
- There does exist a large theoretical component!

Research Process (Refresher)

- Form an Idea
- Review the Literature
- Define the Problem
- Develop a Solution/Hypothesis
- Evaluate the Solution/Hypothesis
- Report the Results & Analysis

We Do Experiments?

- Depends on Nature of Problem/Solution
- Evaluated via
 - Experimentation
 - Simulation
 - Hypothesis/Proof
- Do you need data?
 - Systems areas frequently use experimentation/simulation
 - Theory of Computation, Algorithm Design/Analysis frequently use proofs

Experimentation-Simulation: What's the Diff?

- Depends on who you talk to
- Experiment:
 - Conducted under controlled condition
 - Measurements record *actual* characteristics
 - Example: Benchmark new hardware
- Simulation
 - Conducted under controlled condition
 - Measurements record *simulated* characteristics
 - Example: Study new network routing algorithm
- When is one or the other best?

So I Need Numbers & Charts

- Run a few benchmarks & plot the results...
- NOT SO FAST
- You must follow Scientific Method to be taken seriously

Scientific Method

Principles and procedures for the systematic pursuit of knowledge involving the recognition and formulation of a problem, the collection of data through observation and experiment, and the formulation and testing of hypotheses. Merriam-Webster

- Key Points
 - Systematic
 - Data Collection
 - Hypothesis Testing

Before Collection Data

- Plan out your experiment/simulation!
 - What are you studying?
 - What are your independent/dependent variables?
 - What must you measure
 - What controls will you use?
 - What data? Where will it come from?
 - Do you need human interaction?

More Planning

- If studying a system, what is the execution environment?
- What analysis is required to prove hypothesis?
 - Statistical tests?
 - Hypothesis Testing (F, T, Chi-square)
 - Correlation
 - Analysis of Variance (ANOVA)
 - Proofs only?

Input Data

- Must take time to pick good representative of data/input
- Programs?
 - All sizes (small to large)
 - From computation bound to wait-full
 - Compute over little to massive data
- Other characteristics
 - User Sessions
 - Network Traffic

Computation Environment

- Machine Configuration
 - Available Memory
 - CPU
 - Software version
- Working with “real world” or modeled environment
 - Cover range of
 - Communication nodes (few to many)
 - Communication interconnect?
 - Computation capability?

The “Old” Way

- Some Computer Scientists “did not do” experimentation
- Data analysis was conducted even less frequently
- This is changing!
- Quality conferences & Journals *require* solid empirical evaluation (where appropriate)
- If you have not had a statistics course, think about taking one

Just Before You Start

- Write up your Experiment Plan
 - Be detailed!
 - Answer all the previous questions
 - Motivate why you are measuring something
 - Explain relationships between variables & behaviors
 - Expected values?

Final Note

- Document everything!
- Keep a research log/journal
 - Documents when you had ideas
 - Reminds you what those ideas are/were
 - Helps you sort out your ideas