Experimentation

Professional Communications
Designing Experiments

• The use of experimentation to verify hypotheses is one of the central elements of science

• In computing, experiments are usually used to confirm hypotheses about algorithms or human factors
  – For example, we usually want to know if a system can complete a specified task, and if it can do so with reasonable use of resources
Designing Experiments

• Tests should be fair, rather than constructed to support hypotheses

• Choose an appropriate baseline
  – My sorting algorithm is faster than bubblesort
  – My sorting algorithm is faster than quicksort

• Choose an appropriate dataset
  – Choosing parameters to suit data, or choosing data to suit parameters, invalidates the research
Designing Experiments

• Care is particularly needed when checking the outcome of a negative or failed experiment
  – What happened? Why did it happen?
  – A failure is likely to expose a problem with the assumptions, with the methodology, or with the hypothesis

• A sanity check is always necessary
  – Do the results make sense?
Designing Experiments

• Results must be conclusive and obvious
  – Results should not be subject to interpretation
  – Remember cold fusion?

• Don’t discard anomalies unless you are certain that they are irrelevant
  – They may represent problems you haven’t considered
  – As the graph shows, the algorithm was much slower on two of the data sets. We are still investigating this behavior.
Describing Experiments

• The purpose of experimentation is to take measurements that can be used as evidence in support of the hypothesis

• What should you measure?
  – Consider this: the sole reason for computer programming is to build tools that collect, process, and produce information (data)

• Measurements are intended to be a consequence of some underlying phenomenon that is described by a theory or hypothesis
Describing Experiments

• Measurements should be simple
  – Measure everything that is necessary and only what is necessary

• Measurements can be quantitative
  – Number, duration, volume, etc.

• Measurements can be qualitative
  – The outcome was achieved, features observed, etc.
Describing Experiments

• Experiments should be reproducible
  – Experiments should be described in enough detail so that another researcher could reproduce your results exactly
  – If the results cannot be reproduced by an independent researcher, then the results cannot be trusted
Describing Experiments

• Your interpretation and understanding of the results is as important as the results themselves

• Analyze the results and explain their significance
  – Select typical results, explain why they are typical, theorize about anomalies, explain why the results confirm or disprove the hypothesis
  – Make the results interesting
Describing Experiments

• Not all results need to be included
  – Not all results are directly relevant to the hypothesis
  – Researchers must decide which results to report, even though the reported results should be a fair reflection of the experiment’s outcome
  – If a test fails on some data sets and succeeds on others, it is unethical to conceal the failures
Hypothesis Testing

• A hypothesis test compares the distributions of two sets of observations
• The means of the two sets of observations might be different, but are they *significantly* different?
The Role of Intuition

• A cautious researcher should consider whether assumptions are statistically reasonable
  – If a robot successfully traverses a room once, avoiding all obstacles (as it was designed to do), does that imply that it will be able to do so again?
  – Consider the role that randomness plays in any experimental paradigm