Introduction to HCI

Experiments I

Prof. Narges Mahyar
UMass Amherst

nmahyar@cs.umass.edu
Courses, projects, papers, and more:
http://groups.cs.umass.edu/nmahyar/

Today

- Focus group [10 min]
- Readings [5 min]
- Experiments 1 lecture [30 min]
- In class activity [30 min]

Learning goals

- What is the experimental method?
- What is an experimental hypothesis?
- How do I plan an experiment?
- Why are statistics used?
- Within- & between-subject comparisons: how do they differ?
- Much of this material is well covered in today's readings:

  Hochheiser, H., Feng, J. H., & Lazar, J. (2017).
  - Experimental research. **Chapter 2**.
  - Experimental design. **Chapter 3**.

Acknowledgement: Some of the material in this lecture is based on material prepared for similar courses by Saul Greenberg (University of Calgary)

Material I assume you already know and will not be covered in lecture

- Types of variables
- Samples & populations
- Normal distribution
- Variance and standard deviation

Controlled experiments

The traditional scientific method

- Clear convincing result on specific issues
- In HCI
  - Insights into cognitive process, human performance limitations, ...
  - Allows comparison of systems, fine-tuning of details ...

Strives for

- Lucid and testable hypothesis
- Quantitative measurement
- Measure of confidence in results obtained
- Replicability of experiment
- Control of variables and conditions
- Removal of experimenter bias

Desired outcome of a controlled experiment

Statistical inference of an event or situation's probability:

"design A is better <in some specific sense>
than design B"

Or, design A meets a target:

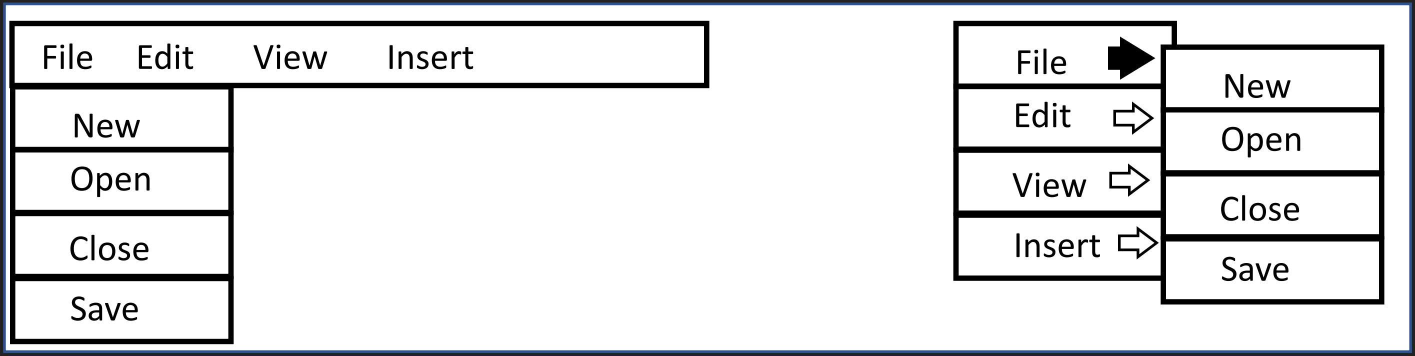
"90% of incoming students who have web experience can complete course registration within 30 minutes"

Steps in the experimental method

Step 1: begin with a testable hypothesis

Example 1:

- H_0: there is no difference in user performance (time and error rate) when selecting a single item from a pop-up or a pull down menu
- H_1: selecting from a pop-up menu will be faster and less error prone than selecting from a pull down menu



General: hypothesis testing

Hypothesis = **prediction** of the outcome of an experiment.

- Framed in terms of **independent** and **dependent** variables:
  - A variation in the independent variable will cause a difference in the dependent variable.
- Aim of the experiment: prove this prediction
  - **By**: *disproving* the “null hypothesis”
  - **Never** by: *proving* the “alternate hypothesis”

H_0: experimental conditions **have *no* effect** on performance (to some degree of **significance**) → null hypothesis

H_1: experimental conditions **have an effect** on performance (to some degree of **significance**) → alternate hypothesis

Step 2: explicitly state the independent variables

Independent variables

- **things you control/manipulate (independent of how a subject behaves)** to produce different conditions for comparison

- E.g. age and time

Step 3: carefully choose the dependent variables

Dependent variables

- Things that are measured
- Expectation that they depend on the subject’s behaviour / reaction to the independent variable (but unaffected by other factors)

E.g. hight

Activity

- Determine independent and dependent variables in these scenarios:

- You want to run a study to determine how long a student sleeps affects test scores.

- You want to know whether caffeine affects your appetite

step 4: consider possible nuisance variables & determine mitigation approach

"Systematic errors" in reading

- Undesired variations in experiment conditions which cannot be eliminated, but which may affect dependent variable
  - Critical to know about them

- Experiment design & analysis must generally accommodate them:
  - Treat as an additional experiment independent variable (if they can be controlled)
  - Randomization (if they cannot be controlled)

- Common nuisance variable: subject (individual differences)

Step 5: design the task to be performed

Tasks must:

Be externally valid

- External validity = do the results generalize?
- … Will they be an accurate predictor of how well users can perform tasks as they would in real life?
-
- Exercise the designs, bringing out any differences in their support for the task
- E.g., If a design supports website navigation, test task should not require subject to work within a single page

Be feasible - supported by the design/prototype, and executable within experiment time scale

Step 6: design experiment protocol

- Steps for executing experiment are prepared well ahead of time
- Includes unbiased instructions + instruments (questionnaire, interview script, observation sheet)
- Double-blind experiments, ...

Step 7: make formal experiment design explicit

Simplest: 2-sample (2-condition) experiment

- Based on comparison of **two sample means**:
  - Performance data from using design A & design B
    - E.g., New design & status quo design
    - E.g., 2 new designs
- Or, comparison of **one sample mean with a constant**:
  - Performance data from using design A, compared to performance requirement
    - Determine whether single new design meets key design requirement

Step 7: make formal experiment design explicit

More complex: factorial design

In menu experiment:

- 2 menu types (pop-up, pull down)
- X 5 menu lengths (3, 6, 9, 12, 15)
- X 2 levels of expertise (novice, expert)

Within/between subject comparisons

within-subject design:

subjects exposed to multiple treatment conditions

→ primary comparison internal to each subject

- allows control over subject variable
- greater statistical power, fewer subjects required
- not always possible (exposure to one condition might "contaminate" subject for another condition; or session too long)

Within/between subject comparisons

Between-subject design:

Subjects only exposed to one condition

- → Primary comparison is from subject to subject
- Less statistical power, more subjects required
- Why? Because greater variability due to more individual differences

Split-plot design (also called mixed factorial design)

Step 8: carefully select/recruit and assign subjects to groups

Subject pool: similar issues as for informal and field studies

- Match expected user population as closely as possible
- Age, physical attributes, level of education
- General experience with systems similar to those being tested
- Experience and knowledge of task domain

Sample size: more critical in experiments than other studies

- Going for “statistical significance”
- Should be large enough to be “representative” of population
- Guidelines exist based on statistical methods used & required significance of results
- Pragmatic concerns may dictate actual numbers
- “10” is often a good place to start

Step 8: carefully select/recruit and assign subjects to groups

- If there is too much variability in the data collected, you will not be able to achieve statistical significance
- You can reduce variability by controlling subject variability
- How?
  - Recognize classes and make them an independent variable
    - E.g., Older users vs. Younger users
    - E.g., Superstars versus poor performers
  - Use reasonable number of subjects and random assignment

Step 9: apply statistical methods to data analysis

Examples: t-tests, ANOVA, correlation, regression

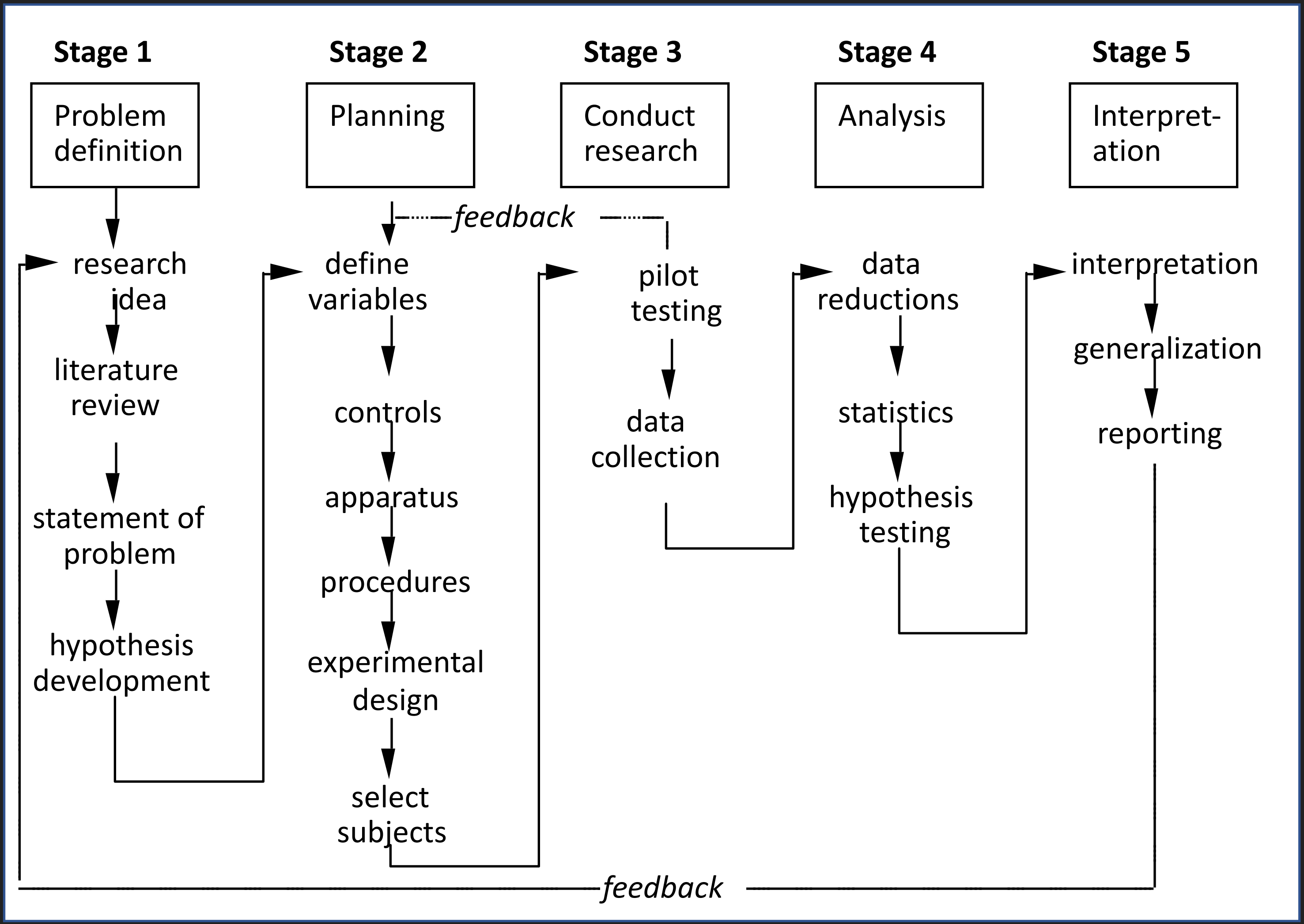
Confidence limits: the confidence that your conclusion is correct

- "The hypothesis that mouse experience makes no difference is rejected at the .05 level" (i.e., Null hypothesis rejected)
- This means:
  - A 95% chance that your finding is correct
  - A 5% chance you are wrong

Step 10: interpret your results

- What you believe the results mean, and their implications
- Yes, there can be a subjective component to quantitative analysis

The experiment planning flowchart



Summary of steps

- Step 1: begin with a testable hypothesis
- Step 2: explicitly state the independent variables
- Step 3: carefully choose the dependent variables
- step 4: consider possible nuisance variables & determine mitigation approach
- Step 5: design the task to be performed
- Step 6: design experiment protocol
- Step 7: make formal experiment design explicit
- Step 8: carefully select/recruit and assign subjects to groups
- Step 9: apply statistical methods to data analysis
- Step 10: interpret your results

To summarize:
how a controlled experiment works

1. Formulate an **alternate** and a **null** hypothesis:
   - H_1: experimental conditions have an effect on performance
   - H_0: experimental conditions have no effect on performance
2. Through **experimental task**, try to demonstrate that the null hypothesis is false (reject it),
   - For a particular level of significance
3. If successful, we can a**ccept** the alternate hypothesis,
   - And state the probability p that we are wrong (the null hypothesis is true after all) → this is result's confidence level
   - E.g., Selection speed is significantly faster in menus of length 5 than of length 10 ($p<.05$)

→ 5% chance we've made a mistake, 95% confident