CS 160: User Interface Design

Data Analysis

03/07/12

Berkeley

UNIVERSITY OF CALIFORNIA

Touch Projector:
Mobile Interaction through Video

Sebastian Boring¹, Dominikus Baur¹, Andreas Butz¹,
Sean Gustafson², Patrick Baudisch²

¹University of Munich
²Hasso Plattner Institute
Plan Through Midterm

Today 3/7:
Statistics & Analyzing Study Data

Monday 3/12:
Midterm Review
Due: Lo-fi test with three users

Wednesday 3/14:
In-class Midterm

Midterm on 3/14

In class. 75 minutes
Closed book & notes
Review on Monday 3/12

If you are registered with the DSP office and have special needs, we need to see your letter by today at 1pm to make accommodations.
Designing Controlled Experiments

Variables

Independent variables
- Cursor type (bubble, normal, area?)
- Target Distance
- Target Width (Effective vs. Actual?)

Dependent variables
- Movement Time
- Error Rate
- User Satisfaction

Control variables
- Color scheme, input device, screen size

Random variables
- Location, environment,
- Attributes of subjects
  - Age, gender, handedness, …

Conducting studies online vs. in person strongly influences which variables are controlled and which are random.
Choosing Subjects

Pick balanced sample reflecting intended user population
Novices, experts
Age group
Sex

Example
12 non-colorblind right-handed adults (male & female)

Population group can also be an IV or a controlled variable
What is the disadvantage of making population a controlled var?

Between Subjects Design

Wilma and Betty use one interface    Dino and Fred use the other
Within Subjects Design

Everyone uses both interfaces

Between vs. Within Subjects

**Between subjects**
Each participant uses one condition
+/- Participants cannot compare conditions
+ Can collect more data for a given condition
- Need more participants

**Within subjects**
All participants try all conditions
+ Compare one person across conditions to isolate effects of individual diffs
+ Requires fewer participants
- Fatigue effects
- Bias due to ordering/learning effects
Within Subjects: Ordering Effects

In within-subjects designs ordering of conditions is a variable that can confound results
Why?

Turn it into a random variable
Randomize order of conditions across subjects
Counterbalancing (ensure all orderings are covered)
Latin square (partial counterbalancing)
...

Run the Experiment

Always pilot it first!
Reveals unexpected problems
Can’t change experiment design after starting it

Always follow same steps – use a checklist

Get consent from subjects

Debrief subjects afterwards
For Wed: Bubble Cursor Experiment

Managing Study Participants
Run the Experiment

Always pilot it first!
Reveals unexpected problems
Can’t change experiment design after starting it

Always follow same steps – use a checklist

Get consent from subjects

Debrief subjects afterwards

The Participants’ Standpoint

Testing is a distressing experience
Pressure to perform
Feeling of inadequacy
Looking like a fool in front of your peers, your boss, …

(from “Paper Prototyping” by Snyder)
The Three Belmont Principles

**Respect for Persons**
Have a meaningful consent process: give information, and let prospective subjects freely chose to participate

**Beneficience**
Minimize the risk of harm to subjects, maximize potential benefits

**Justice**
Use fair procedures to select subjects (balance burdens & benefits)

To ensure adherence to principles, most schools require Institutional Review Board approval of research involving human subjects.

Ethics: Stanford Prison Experiment

1971 Experiment by Phil Zimbardo at Stanford
24 Participants – half prisoners, half guards ($15 a day)
Basement of Stanford Psychology bldg turned into mock prison
Guards given batons, military style uniform, mirror glasses,…
Prisoners wore smocks (no underwear), thong sandals, pantyhose caps

Experiment quickly got out of hand
Prisoners suffered and accepted sadistic treatment
Prison became unsanitary/in hospitable
Prisoner riot put down with use of fire extinguishers
Guards volunteered to work extra hours

Zimbardo terminated experiment early
Grad student Christina Maslach objected to experiment
Important to check protocol with ethics review boards

[from Wikipedia]
http://www.youtube.com/watch?v=rmwSC5fS40w
Ethics: Stanford Prison Exp. Video

http://www.youtube.com/watch?v=Z0jYx8nwjFQ

Ethics (more recently)

“In 2001, a faculty member [...] designed a study to see how restaurants would respond to complaints [...] As part of the project, the researcher sent letters to restaurants falsely claiming that he and/or his wife had suffered food poisoning that ruined their anniversary celebration. The letters [...] stated that the only intent was to convey to the owner what had occurred "in anticipation that you will respond accordingly." Restaurant owners were understandably upset and some employees lost their jobs before it was revealed that the letter was a hoax.“

CITI Human Subject Training Material
Beneficience: Example

**MERL DiamondTouch:**
User capacitively coupled to table through seating pad.
No danger for normal users, but possibly increased risk for participants with pacemakers.

Inform subjects in consent!


Privacy and Confidentiality

**Privacy:** having control over the extent, timing, and circumstances of sharing oneself with others.

**Confidentiality:** the treatment of information that an individual has disclosed with the expectation that it will not be divulged.

Examples where privacy could be violated or confidentiality may be breached in HCI studies?
Treating Subjects With Respect

**Follow human subject protocols**
- Individual test results will be kept confidential
- Users can stop the test at any time
- Users are aware (and understand) the monitoring technique(s)
- Their performance will not have implications on their life
- Records will be made anonymous

**Use standard informed consent form**
- Especially for quantitative tests
- Be aware of legal requirements

Conducting the Experiment

**Before the experiment**
- Have them read and sign the consent form
- Explain the goal of the experiment in a way accessible to users
- Be careful about the demand characteristic
  (Participants biased towards experimenter’s hypothesis)
- Answer questions

**During the experiment**
- Stay neutral
- Never indicate displeasure with users performance

**After the experiment**
- Debrief users (Inform users about the goal of the experiment)
- Answer any questions they have
Managing Subjects

Don’t waste users’ time
Use pilot tests to debug experiments, questionnaires, etc…
Have everything ready before users show up

Make users comfortable
Keep a relaxed atmosphere
Allow for breaks
Pace tasks correctly
Stop the test if it becomes too unpleasant

If you want to learn more…

Online human subjects certification courses:
E.g., http://phrp.nihtraining.com/users/login.php

The Belmont Report: Ethical Principles and Guidelines for the protection of human subjects of research
1979 Government report that describes the basic ethical principles that should underly the conduct of research involving human subjects
http://ohsr.od.nih.gov/guidelines/belmont.html
Data Analysis

Bubble Cursor Online Experiment
UC Berkeley CS160

In this short experiment, you will be asked to click on a sequence of targets on
screen. You will do this both with a regular mouse cursor and with a different type of
cursor that dynamically expands in size to always select the closest target. This
Bubble Cursor was introduced by Tovi Grossman and Ravin Balakrishnan at CHI 2005
[1].

This experiment does not work in Google Chrome. Firefox is preferred. Safari
might work.

Warm-Up
First, get familiar with the two tasks. Here is the normal cursor. Your job is to click
on the red targets - quickly but accurately. Click on a few of them.
Start by counting

1700 trials total

normal: mean time 966.3 ms, mean errors 1.286
bubble: mean time 758.8 ms, mean errors 0.279
Start by counting

21 users completed condition normal, size 10
mean time: 1129.37 ms, mean errors: 0.08
median time: 1055 ms, median errors: 0

21 users completed condition normal, size 30
mean time: 803.24 ms, mean errors: 0.05
median time: 766 ms, median errors: 0

21 users completed condition bubble, size 10
mean time: 796.22 ms, mean errors: 0.01
median time: 751 ms, median errors: 0

22 users completed condition bubble, size 30
mean time: 723.04 ms, mean errors: 0.02
median time: 701 ms, median errors: 0

Descriptive Statistics

Continuous data:

Central tendency
mean, median, mode

Dispersion
Range (max-min)
Standard deviation

Shape of distribution
Skew, Kurtosis

Categorical data:
Frequency distributions
What's missing from this bar chart?
Descriptive Statistics for Error

Exploratory Data Analysis (EDA):
Look at your data from different perspectives to get better intuition for it.
Show the raw data!
Use different visualizations: Histograms, scatterplots, box plots, …
Median vs. Mean

For normal distribution, median = mean

Most data is fitted to normal distribution so median and mean may differ

When data has outliers median more robust

Many data sets gathered online are strongly skewed (they exhibit power law distributions - “long tails”)

Outliers pull the mean to the right/left (mean of true power law distribution = \(\infty\))
Power Law Distributions

\[ p(x) \propto x^{-\alpha} \]

Figure 5.1: The distribution of photographers contributing photos of the 2005 Coney Island Mermaid Parade.

From C. Shirky, Here Comes Everybody
Cleaning Data

Don’t discard data just because it doesn’t fit your expectation! Maybe your assumptions were wrong.

In online experiments, discarding extreme outliers can make sense if you believe they reflect users not following normal task protocol (e.g., multitasking in a reaction-time study)

Effect Sizes: Time

**Normal vs. Bubble cursor at target size 10:**
1129ms vs. 796ms: Bubble cursor 30% faster

**Normal vs. Bubble cursor at target size 30:**
803ms vs. 723ms: Bubble cursor 10% faster

**Target size for normal cursor:**
1129ms vs 803ms: Larger targets 29% faster

**Target size for Bubble cursor:**
796ms vs. 723ms: Larger targets 9% faster
Effect Sizes: Time

<table>
<thead>
<tr>
<th>Factor</th>
<th>Group</th>
<th>Sample size</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>CURSOR x SIZE</td>
<td>bubble x 10</td>
<td>420</td>
<td>796.21905</td>
</tr>
<tr>
<td>CURSOR x SIZE</td>
<td>bubble x 30</td>
<td>440</td>
<td>723.03864</td>
</tr>
<tr>
<td>CURSOR x SIZE</td>
<td>normal x 10</td>
<td>420</td>
<td>1,129.36667</td>
</tr>
<tr>
<td>CURSOR x SIZE</td>
<td>normal x 30</td>
<td>420</td>
<td>803.2381</td>
</tr>
<tr>
<td>CURSOR</td>
<td>bubble</td>
<td>860</td>
<td>758.77791</td>
</tr>
<tr>
<td>CURSOR</td>
<td>normal</td>
<td>840</td>
<td>966.30238</td>
</tr>
<tr>
<td>SIZE</td>
<td>10</td>
<td>840</td>
<td>962.79286</td>
</tr>
<tr>
<td>SIZE</td>
<td>30</td>
<td>860</td>
<td>762.20581</td>
</tr>
</tbody>
</table>

Effect Sizes: Error

Normal vs. Bubble cursor, target size 10:
1.67 vs. 0.24 Errors per 20 trials: 85% fewer errors! (6.95x)

Normal vs. Bubble cursor, target size 30:
0.90 vs. 0.02 Errors per 20 trials: 98% fewer errors!
Effect Sizes: Error

### Descriptive Statistics

<table>
<thead>
<tr>
<th>Factor</th>
<th>Group</th>
<th>Sample size</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>CURSOR x SIZE</td>
<td>NORMAL x 10</td>
<td>21</td>
<td>1.66667</td>
</tr>
<tr>
<td>CURSOR x SIZE</td>
<td>NORMAL x 30</td>
<td>21</td>
<td>0.90476</td>
</tr>
<tr>
<td>CURSOR x SIZE</td>
<td>BUBBLE x 10</td>
<td>21</td>
<td>0.2381</td>
</tr>
<tr>
<td>CURSOR x SIZE</td>
<td>BUBBLE x 30</td>
<td>22</td>
<td>0.31818</td>
</tr>
<tr>
<td>CURSOR</td>
<td>NORMAL</td>
<td>42</td>
<td>1.28571</td>
</tr>
<tr>
<td>CURSOR</td>
<td>BUBBLE</td>
<td>43</td>
<td>0.27907</td>
</tr>
<tr>
<td>SIZE</td>
<td>10</td>
<td>42</td>
<td>0.95238</td>
</tr>
<tr>
<td>SIZE</td>
<td>30</td>
<td>43</td>
<td>0.60465</td>
</tr>
</tbody>
</table>

Example of Interactions

**Group problem solving**

Independent variable: Leadership

[from Martin 04]
Example of Interactions

Group problem solving
Independent variable: Leadership
Independent variable: Group size

Problem Solving Time

![Graph showing the relationship between group size and problem solving time with and without leadership.](from Martin 04)

Example of Interactions

Group problem solving
Change in time due to leadership is same regardless of group size

Problem Solving Time

![Graph showing the relationship between group size and problem solving time with and without leadership.](from Martin 04)
Example of Interactions

**Group problem solving**
Change in time due to leadership is same regardless of group size
Change in time due to group size is same regardless of leadership
Independent variables do not interact

\[ \Delta t \text{ due to group size is same whether or not there is a leader} \]

Example of Interactions

**Multiple IVs effect DV non-additively**
Change in time due to leadership differs with changes in group size
Independent variables do interact

\[ \text{[from Martin 04]} \]
Population versus Sample

Are the Results Meaningful?

**Hypothesis testing**
- Hypothesis: Manipulation of IV effects DV in some way
- Null hypothesis: Manipulation of IV has no effect on DV
- Null hypothesis assumed true unless statistics allow us to reject it

**Statistical significance (p value)**
- Likelihood that results are due to chance variation
- \( p < 0.05 \) usually considered significant (Sometimes \( p < 0.01 \))
- Means that < 5% chance that null hypothesis is true

**Statistical tests**
- T-test (1 factor, 2 levels)
- Correlation
- ANOVA (1 factor, > 2 levels, multiple factors)
- MANOVA ( > 1 dependent variable)