Designing and Analyzing HCI Experiments

Thomas George
College of IST
Penn State University
thomasg@ist.psu.edu
Jan 15, 2008
What do you do in HCI research?

- Develop the “right” research questions (in context)
- Use the “right” data collection methods
- Use the “right” data collection tools
- Using the “right” data analysis
- Arrive at interesting insights about the context and the human computer interaction/work
In this talk..

- Look into data collection methods, specifically on controlled experiments
  - Taking validity into account (ecological, internal, external)
  - Bad (mediocre) designs
  - Good designs
  - Quasi-controlled experiments / Exploratory controlled experiments

- How to analyze data in different contexts
  - I will NOT talk about common known methods – anova, regression etc
  - Will talk about higher order Manova’s, profiling, time-series, path-models, non-parametric models, probability models etc
  - Focus on when these methods are useful and why
Controlled Experiments

• Look into a data collection methods, specifically on controlled experiments
  – Taking into validity into account (ecological, internal, external)
  – Common mistakes
  – When do controlled experiments work
  – Quasi-controlled experiments / Exploratory controlled experiments

• How to analyze data in different contexts
  – I will NOT talk about common known methods – anova, regression etc
  – Will talk about higher order Manova’s, profiling, time-series, path-models, non-parametric models, probability models etc
  – Focus on when these methods are useful and why
The “Scientific Method”

- Scientific method
  - Perspectives differ among scientists (physicists vs. anthropologists; possibly even between psychologists and anthropologists)
  - Epistemological and philosophical differences
  - Read more about “Science Wars” (*Social Text Affair*)
    http://physics.nyu.edu/~as2/

- In this talk I will focus on controlled experiments
  - Not necessarily looking for causal relationships
  - Exploratory studies
Experimental Validity

- Validity: best available approximation to truth or falsity of propositions (Cook & Campell, 1957)
- Internal Validity: best approximate truth about inferences regarding causal relationships (i.e. did the treatment really cause the effect?)
- External Validity: generalizability of the results
- Statistical Conclusion Validity: related more to statistical power and significance of tests
Internal Validity

- **History**: e.g. events between a pre-test and post-test
- **Maturation**: relevant only in long term studies (e.g. growing older, getting more proficient with software)
- **Effects of testing**
- **Instrumentation**: due to experimenters (e.g. judgment scores)
- **Selection biases**
- **Experiment mortality**: loss of respondents
External Validity

• Interaction effects: e.g. effect of a pre-test on the results
• Interaction effects of selection bias
• Effects of multiple treatments: when multiple treatments are used for the same group (e.g. in a multivariate repeated measures study)
Experimental Designs (1)

- One shot case study [X O]
  - Single group, studied once
  - Some agent is presumed to cause some change
  - Almost no control
  - Inferences are based on what would have been if the design was not used

- Such studies require
  - Detailed data collection
  - Chances of errors on inferences; significant internal invalidity
Experimental Designs (2)

• One group Pretest – Posttest design \([O_1 \times O_2]\)
  – Single group, studied once
  – Group observed and pre-test is administered
  – Some treatment is introduced
  – Get posttest scores based on the treatment
  – Very common in HCI literature

• Such studies can have
  – Chances of errors on inferences due to internal invalidity (only control is on the selection bias)
Experimental Designs (3)

• Static group comparison \([X \ O_1 \ O_2]\)
  - Two groups, studied once
  - One group gets a treatment, second does not
  - Get scores/measurement, compare differences (typical ANOVA design)
  - Very common in HCI literature

• Such studies can have
  - Selection bias, violation of normality assumptions
Summarizing…

• What can we say about experimental designs 1, 2 and 3
  – *Easy to set up and run*
  – *Are really pre-experiments!*
  – *Almost no Internal validity*
  – *NO external validity*
Experimental Designs (4)

• Pretest – Posttest Control group \( [R O_1 X O_2 \ R O_1 \ O_2] \)
  – Two groups (expt and control); random assignments
  – Control groups for checking treatment effects
  – Test for pre and post test effects on treatment X
  – Test pre and post test effects with no treatment

• Such studies are
  – Very robust in internal validity
  – Most good experimental hci papers use this
Experimental Designs (5)

- 4 group design \([R \ O_1 \times \ O_2 \ (e1)]
  \[R \ O_3 \ O_4 \ (c1)
  \ R \ X \ O_5 \ (e2)
  \ R \ O_6 \ (c2)\]
  - Robust design; increasing external validity
  - Test for \(O_2 > O_4; O_2 > O_1; O_5 > O_6; O_5 > O_3\)

- Such studies are
  - Rare (or absent) in HCI literature
  - Difficult to run (cost and time), but fairly easy to set up
Exploratory Studies

- To establish a baseline for further experimentation
- Establish requirements for design
- To measure variables such as performance, efficiency etc
- Task analysis, time-motion studies
Analysis of Data

• Comparison (are the groups different?)
• Correlation (are the groups similar?)
• Causality (does x cause y?)
  – Direction of causality
• Effects, patterns
Comparison of groups (1/2)

• Are two groups different (can be applied to experimental designs 2 - 5)
  – Univariate vs. Multivariate
  – Normality assumptions
  – Sample characteristics

• What can you say about the results?
  – Group A is different than B when treatment X is applied
  – Anova, Manova, t-tests etc
Comparison of groups (1/2)

• Are two groups different (can be applied to experimental designs 2 - 5)
  – Univariate vs. Multivariate
  – Normality assumptions
  – Sample characteristics

• What can you say about the results?
  – Group A is different than B when treatment X is applied
  – Anova, Manova, t-tests etc
What more can you do with comparison data?

• What they don’t teach you in methods classes:
  – Analyzing the shape of comparison (of means) graphs for
    • Growth/decay curves (identifying what variables increase (or decrease) for what groups)
    • Identifying profiles (e.g. male vs. female; high school vs. graduates)
  – Overcoming correlation effects
    • Most researchers tend to overlook correlation effects among variables and avoid data transformation

• Why they don’t teach you these methods:
  – No statistical software does these directly (the likes of Minitab, SPSS etc)
  – Need to write your own code SAS or more efficiently in R/S-plus
Causality

• Correlation DOES NOT imply causality
• Usually associated with a probability (you can never be 100 % certain)
• Assumes a time dependent relation
• Cause and effect directions can be interchanged

A --> B (A causes B)
Analysis methods to establish causality (general)

- Simplest: chi square (A and B have an odds ratio); the likelihood of A is occurring is x times as of B
- Regression: A is caused by a combination of several variables
- Using Categorical Analysis:
  - Log-linear models (using marginal probabilities)
  - Logistic regression (using log odds ratios)
  - Poisson regression (for rare events)
Analysis methods for causality

• Graphical models using directed graphs from transition probabilities
  – Transitions
  – Graph properties (in degrees, out degrees, centrality)
  – Several Java based tools available (e.g. Jung)
• Probability Models
  – Developing Bayes, Markov based models based on the tpm
  – Simulation as an effective mechanism for understanding “population” parameters
    • MCMC (if based on human behavioral models)
    • Social simulation (for group/social behavior) - Swarm based; Axelrod models etc
Analysis methods for causality

• Time based models
  – time series analysis (helps generate general properties)
  – Lag sequential analysis
Some useful links..

- Research methods for social science
  http://www.socialresearchmethods.net/