

Questions?

- Project #4
 - Consent form
- HW #7 due next week
- Exam: 17th of Dec, 1:30pm-3:30pm
 - Closed books

Quantitative Evaluation

- Gather (performance) measurements
- Methods
 - User events collection
 - *Mouse clicks, keys pressed,...*
 - *Data collected during system use*
 - Google, Amazon
 - Controlled experiments
 - *Set forth a testable hypothesis*
 - *Manipulate one or more independent variable*
 - *Observe effect on one or more dependent variable*
 - *Can be reproduced by others*

Controlled experiment

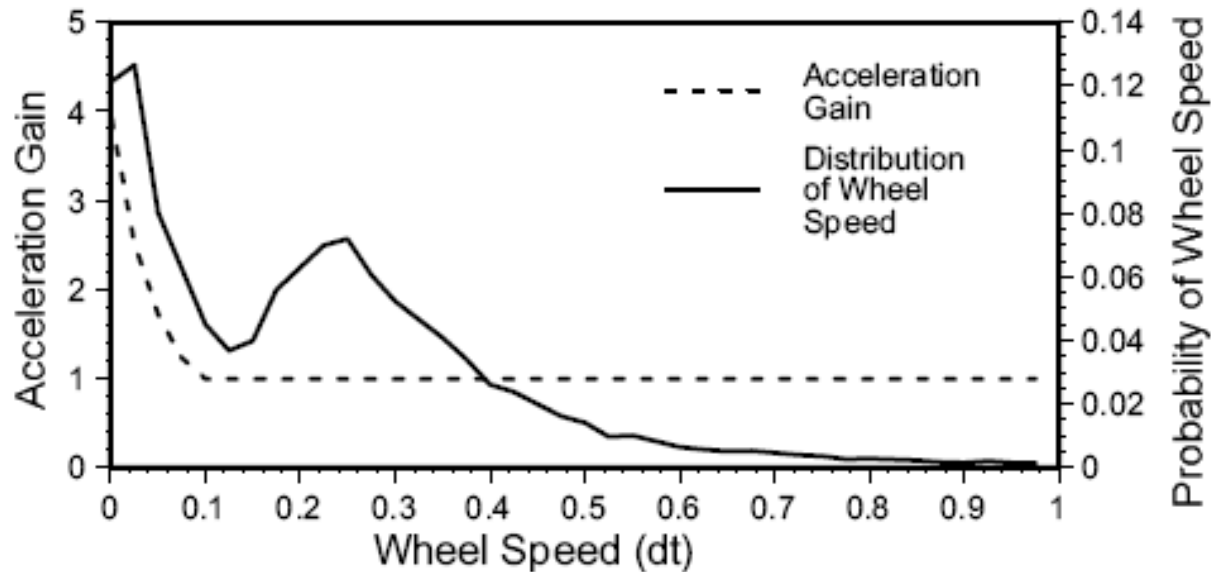
- State a lucid, testable hypothesis
- Identify independent and dependent variables
- Design the experimental protocol
- Choose the user population
- Apply for human subjects protocol review
- Run a couple of pilots
- Run the experiment
- Run statistical analysis
- Draw conclusions

Running example

- Compare Scrolling Techniques [Hinckley et al. '02]
 - ScrollPoint
 - Standard Wheel
 - Accelerated Wheel (2 methods)

State a lucid, testable hypothesis

“With a proper acceleration function, a scroll-wheel based system can be faster than a ScrollPoint.”



Choose the variables

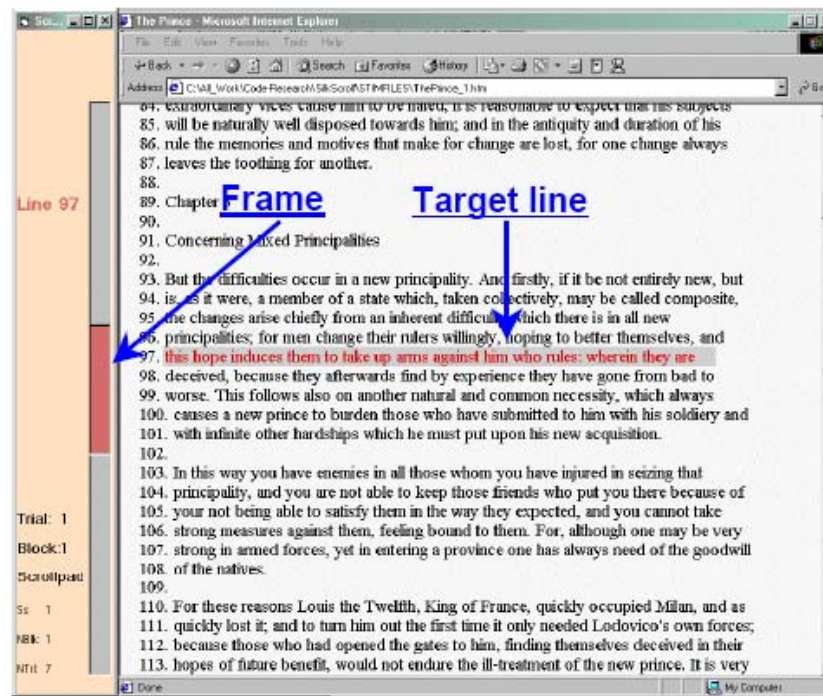
- Manipulate one or more *independent* variable
 - Method
 - Device type...
- Observe effect on one or more *dependent* variable
 - Time to completion
 - Accuracy
 - Error rate...
- Running example
 - Independent variable: method
 - Dependent variable: speed, error rate, user satisfaction...

Design the experimental protocol

- Between or within subjects?
 - Between subjects: each subject run one condition
 - *Need more subjects*
 - *Difference between subjects might introduce a bias*
 - Within subjects: each subject run several conditions
 - *Need less subjects but possible problem with skill transfer*
 - Very important for the statistical analysis phase
- Which task?
 - Must reflect the hypothesis
 - Must avoid bias
 - *Instructions, ordering...*
 - *In doubt, always favor the null hypothesis*

Design the experimental protocol

- Running Example:
 - Navigating in a document
 - *Using a simplified navigation task*
 - Use Fitts' law as the experimental framework



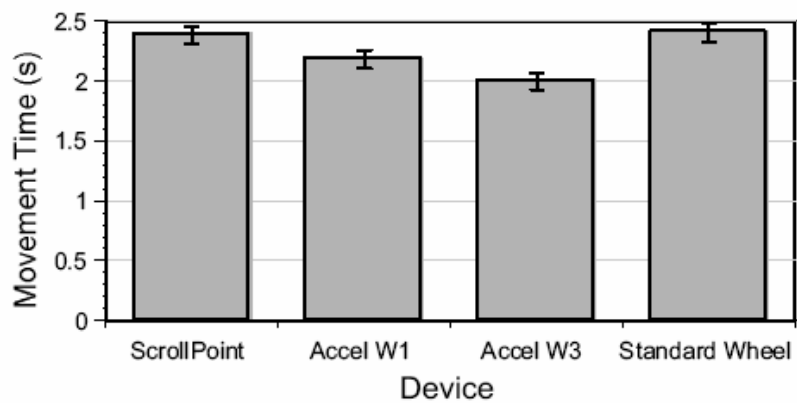
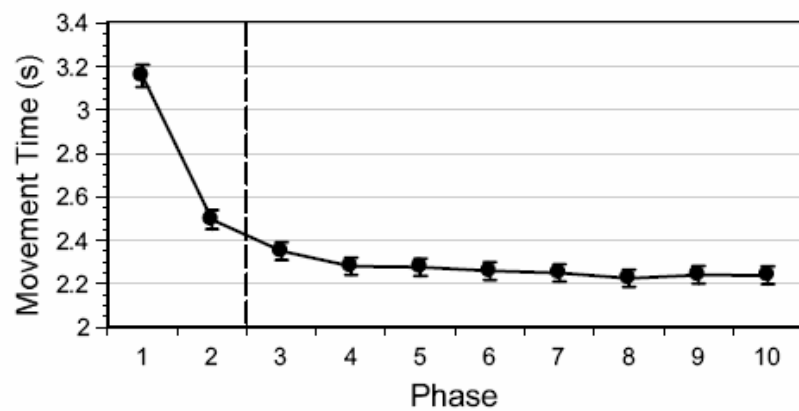
Chose the user population

- Pick a well balanced sample
 - Novices, experts, average
 - Age group
 - Sex...
- Population group may be one of the independent variable
- Running example
 - Used a wide range of age

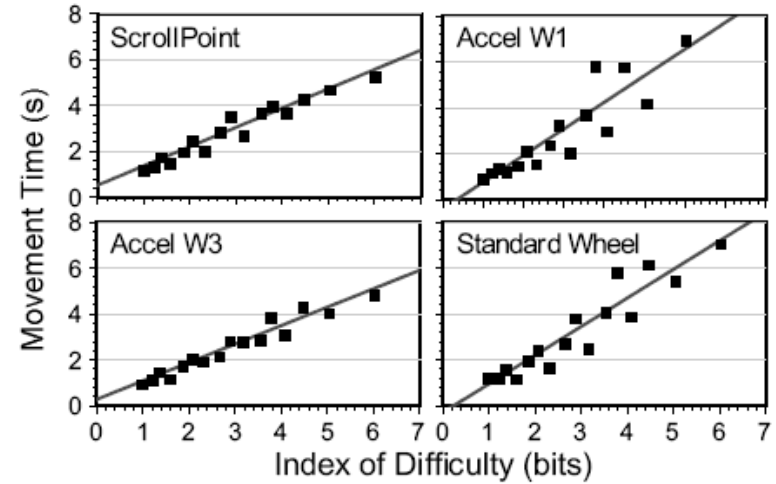
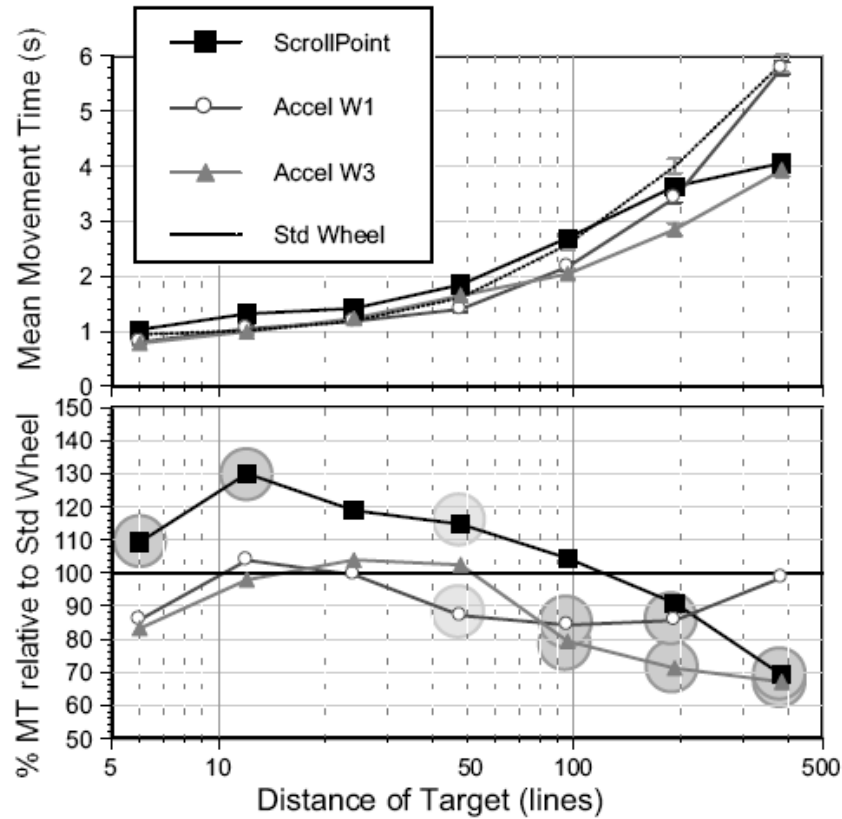
Run the experiment

- Always run pilots first!
 - There are always unexpected problem!
 - When the experiment has started you cannot pick and choose
- Use a check-list so that all subjects follow the same steps
- Don't forget the consent form!
- Don't forget to debrief each subjects

Running example result I



Running example result II



Run statistical analysis

- Properties of our population
 - Mean, variance...
- How different data sets relate to each other
 - Are we sampling from similar or different distributions?
- Probability that our claims are correct
 - Statistical significance:
 - “The hypothesis that technique X is faster is accepted ($p < .05$)” means that there is a higher than 95% chance the hypothesis is true
 - Typical level are .05 and .01 level

Statistical tools I

- T-test
 - Compare the mean of 2 populations
 - *Null hypothesis: no difference between means*
 - Assumptions
 - *Samples are normally distributed*
 - Very robust in practice
 - *Population variances are equal*
 - Reasonably robust for differing variances
 - *Individual observations in samples are independent*
 - Very important

Statistical tools II

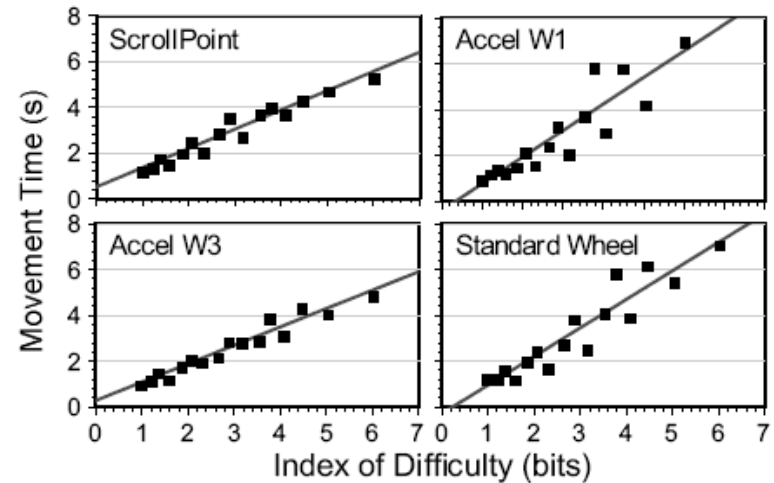
- Correlation

- Measure the extent to which 2 concepts are related
- Caveats
 - *Correlation does not imply cause and effect (hidden variable)*
 - Ice cream consumption and drowning
 - *Need a large enough group*

- Regression

- Calculate the “best fit”

| | R | R ² | Slope | Intercept (s) | IP (bps) |
|-------------|------|----------------|-------|---------------|----------|
| ScrollPoint | 0.97 | 0.94 | 0.84 | 0.42 | 1.19 |
| Accel W1 | 0.90 | 0.81 | 1.16 | -0.51 | 0.86 |
| Accel W3 | 0.97 | 0.95 | 0.80 | 0.18 | 1.25 |
| Wheel Std | 0.94 | 0.88 | 1.25 | -0.42 | 0.80 |



Statistical tool III

- ANOVA
 - Single factor analysis of variance
 - *Compare three or more means*
 - Analysis of variance
 - *Compare relationship between many factor*
 - Beginners type at the same speed on all keyboards,
 - Touch-typist type fastest on the qwerty
- Running example
 - Accept the hypothesis
- Your protocol influence the kind of test you can use
 - In doubt consult with a statistician before starting the experiment!

Draw conclusions

- Running example
 - What is the scope of the finding?
 - *Does the experiment reflect real use?*
 - *Are there other parameters at play?*

Quantitative approach outcome

- Low level effects
 - Patterns of use
 - Scrolling method A faster than method B
- Pros and cons
 - Objective measurements
 - *Good internal validity*
 - Real world implications sometime difficult to foresee
 - Effects might be dwarfed in real world settings
 - *3.05s versus 3.00s?*