Questions?

- HW#7 due in 1 week
- Project #4 due this coming Thursday

- Away until Monday (included)

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
- Identify confounding 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
Quantitative Analysis of Scrolling Technique
Hinckley et al., CHI 2002

- Comparing several scrolling technique
  - Standard (isotonic) wheel
  - Isometric wheel
  - Accelerated isotonic wheel

State a lucid, testable hypothesis

- Different hypothesis have different strength [Vincente 98]
  - Point prediction
    - *Gravity is an attractive force and* $G = 6.67\ 10^{-11}\ Nm^2kg^{-2}$
  - Interval prediction
    - *Gravity is an attractive force* $6\ 10^{-11}\ Nm^2kg^{-2} < G < 7\ 10^{-11}\ Nm^2kg^{-2}$
  - Ordinal prediction
    - *Gravity is an attractive force*
  - Categorical prediction
    - *Gravity is a non-zero force*

- Running example

State a lucid, testable hypothesis

- Different hypothesis have different strength [Vincente 98]
  - Point prediction
    - *Gravity is an attractive force and* $G = 6.67\ 10^{-11}\ Nm^2kg^{-2}$
  - Interval prediction
    - *Gravity is an attractive force* $6\ 10^{-11}\ Nm^2kg^{-2} < G < 7\ 10^{-11}\ Nm^2kg^{-2}$
  - Ordinal prediction
    - *Gravity is an attractive force*
  - Categorical prediction
    - *Gravity is a non-zero force*

- Running example
  - Fitt's law can help us designing better scrolling techniques
  - Acceleration will help
    - *Bi-modal distribution of wheel speed*

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…
- Identify possible confounding variables
  - Previous experience, training, order effect…

- Running example

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…
- Identify possible confounding variables
  - Previous experience, training, order effect…

- Running example
  - Independent variables: technique, Distance, tolerance
  - Dependent variables: speed, error rate, user satisfaction…
  - Confounding variables: skill, age, learning effect

Measuring dependant variables

- Variable and measurement result are not the same!
  - Reliability
    - *For a given state of the variable*
      the same measurement method provide the same results
  - Convergent validity
    - *For a given state of the variable*
      different measurement methods provide the same results
  - Discriminant validity
    - *For different state of the variable,*
      a given method will provide different results

Design the experimental protocol (I)

- Between or within subjects?
  - Between subjects: each subject run one condition
    - *Need more subjects*
    - *Less "powerful" for detecting differences*
    - *No learning effects*
  - Within subjects: each subject run several conditions
    - *Need less subjects*
    - *More "powerful" for detecting differences*
    - *Learning effects*

<u>Your protocol influence the kind of test you can use</u>

In doubt consult with a statistician before starting the experiment!

- Running example

Design the experimental protocol (II)

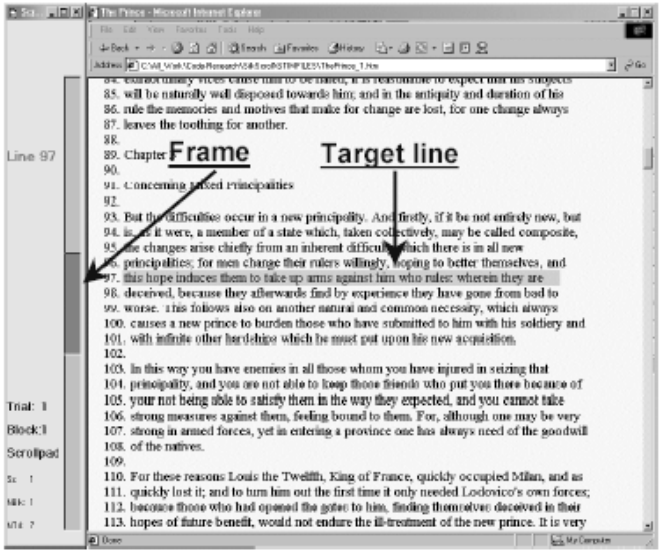
- Which task?
  - Must reflect the hypothesis
  - Must avoid bias
    - *Instructions, ordering…*
    - *In doubt, always favor the null hypothesis*

- Running example

Design the experimental protocol (II)

- Which task?
  - Must reflect the hypothesis
  - Must avoid bias
    - *Instructions, ordering…*
    - *In doubt, always favor the null hypothesis*

- Running example
  - Scrolling between two areas on the screen
  - Using 4 different scrolling technique



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:

Chose the user population

- Pick a well balanced sample
  - Novices, experts, average
  - Age group
  - Sex…
- Population group may be one of the independent variable

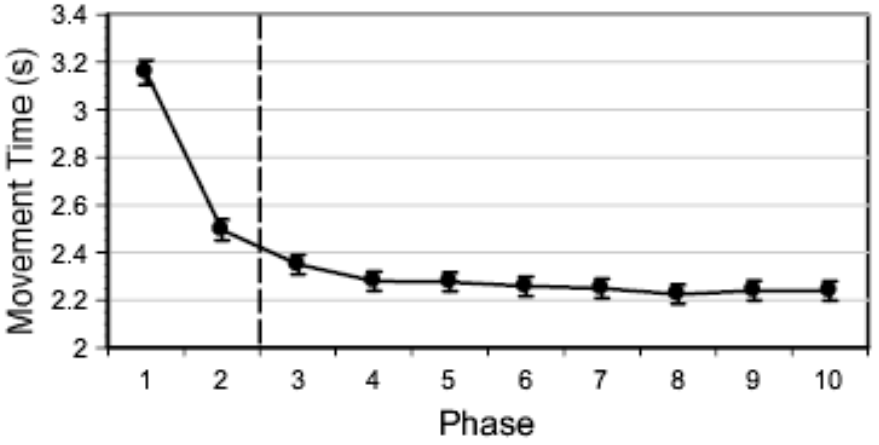
- Running example:
  - 15 females, 12 males
  - Non-color blind, normal vision, right handed, no prior experiences

Run the experiment

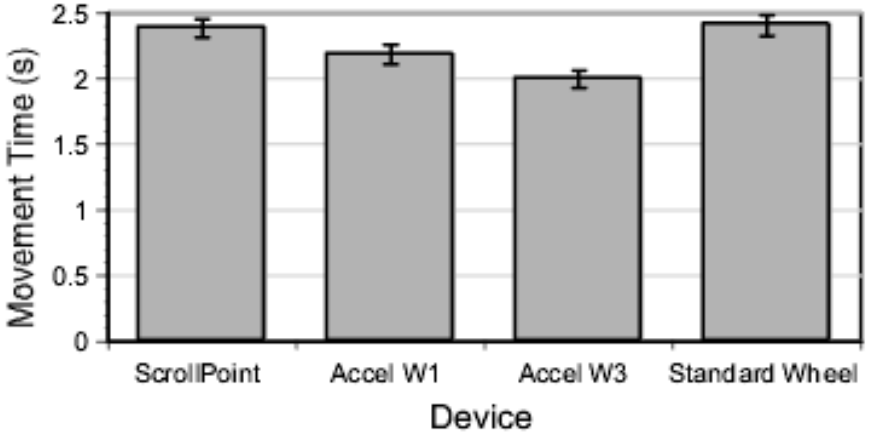
- Always run pilots first!
  - There are always unexpected problem!
  - When the experiment has started you cannot pick and choose
- Use a check-list so that all subjects follow the same steps
- Don’t forget the consent form!
- Don’t forget to debrief each subjects

Running example results

- Learning effects



- Main effects



Run statistical analysis

- Properties of our population
  - Mean, variance…
- How different data sets relate to each other
  - Are we sampling from similar of different distributions?
- Probability that our claims are correct
  - Statistical significance:

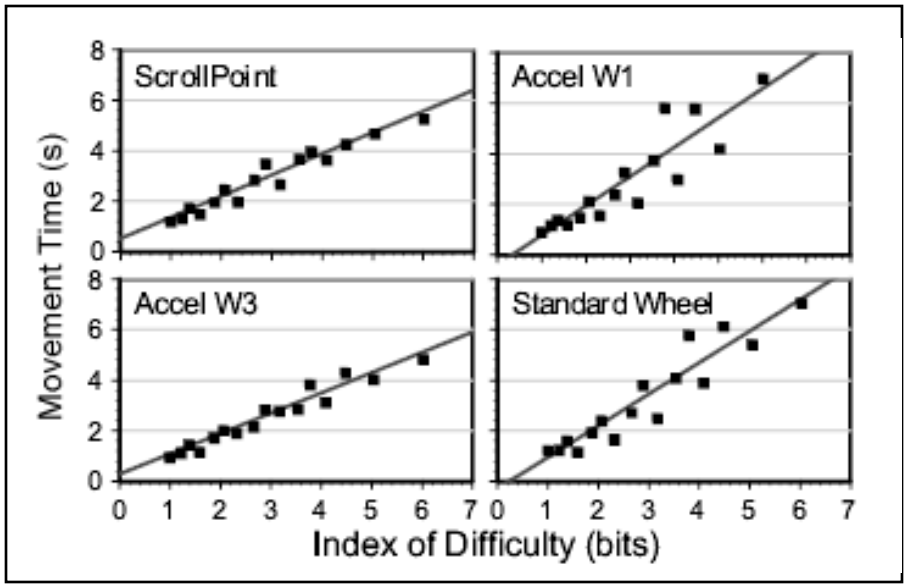
    "The hypothesis that accelerated scrolling is faster is accepted ($p < .05$)" means that there is a higher than 95% chance the hypothesis is true
  - Typical level are .05 and .01 level

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"



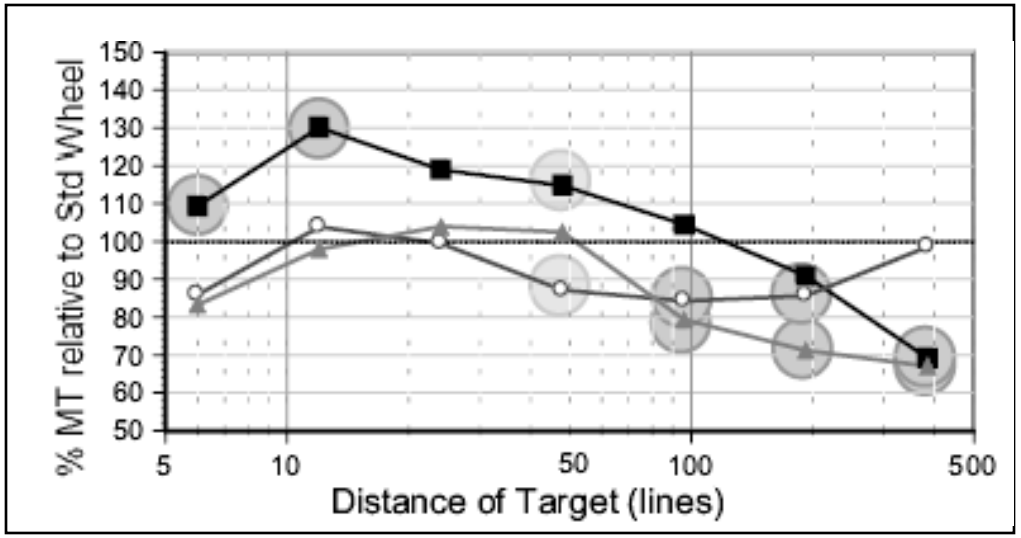
Statistical tool III

- ANOVA (analysis of variance)
  - Compare relationship between factors
    - *Main effect*
      - Describes overall behavior
    - *Interaction*
      - Describes how 2 or more variables interact
- Running example

Statistical tool III

- ANOVA (analysis of variance)
  - Compare relationship between factors
    - *Main effect*
      - Describes overall behavior
    - *Interaction*
      - Describes how 2 or more variables interact

- Running example
  - Device × Width × Distance



Statistical significance

- Statistical significance
  - Comparing to the null hypothesis: “There is no effect”
  - Type I errors are the most disruptive

Researcher’s Decision	Actual Situation: Null Hypothesis is True	Actual Situation: Null Hypothesis is False
Accept the null hypothesis	Correct decision	**Type II error**
Reject the null hypothesis	**Type I error**	Correct decision

- Design significance?
  - 3.00s versus 3.05s?

Draw conclusions

- Be critical about your results
  - Are there other possible explanations?
- Draw the scope of your results
  - How can they be applied in practice?
  - Could they be applied in other contexts?
- Running example