# Questions?

- Project #3
  - Survey question samples at the end of the last lecture
  - Due next Thursday
  - Presentation starting next Tuesday
    - *10 min presentation followed by 3 min for questions*
- Grad Projects
  - A little bit over one month left
- Midterm
  - Working memory versus long term memory
  - Having several master is different from having several market segments
  - Direct manipulation is not good for repeated actions
    - *Rename all the files in a directory*

# 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*

# Quantitative approach outcome

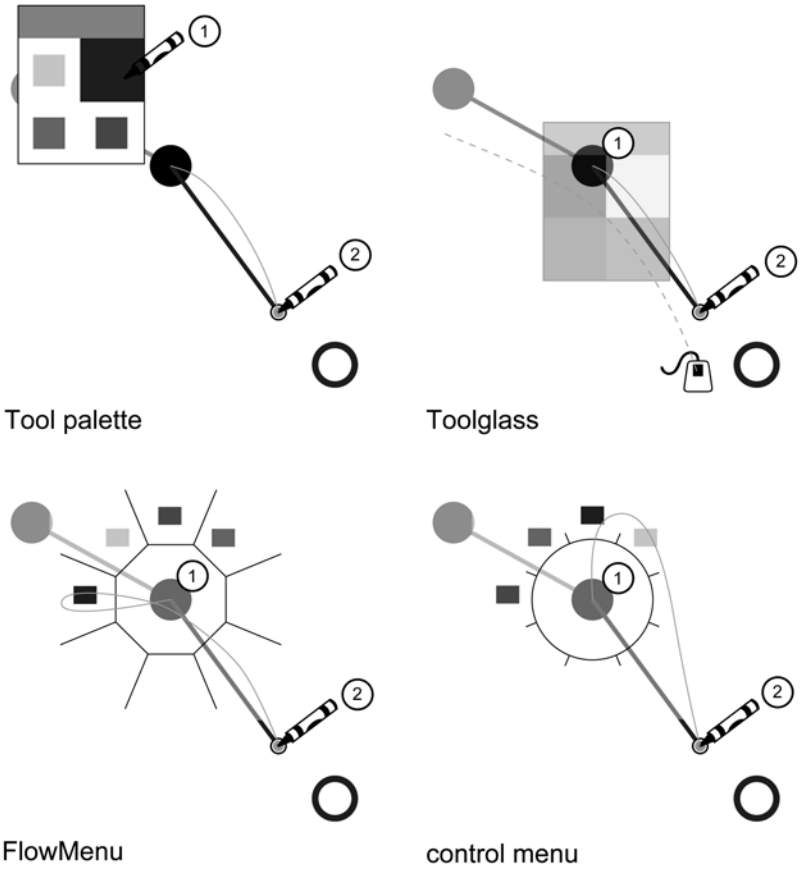
- Low level effects
  - Patterns of use
  - Menu selection method A faster than method B
- Pros and cons
  - Objective measurements
    - *Good internal validity*
  - Real world implications sometime difficult to foresee
  - Effects might be dwarfed in real world settings
    - *3.05s versus 3.00s?*

# 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 4 command mechanisms
  - Used in geometric drawing tasks



# State a lucid, testable hypothesis

"Because of the time it takes to reach for the Tool palette, Tool palette will be the slowest condition."

# 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…

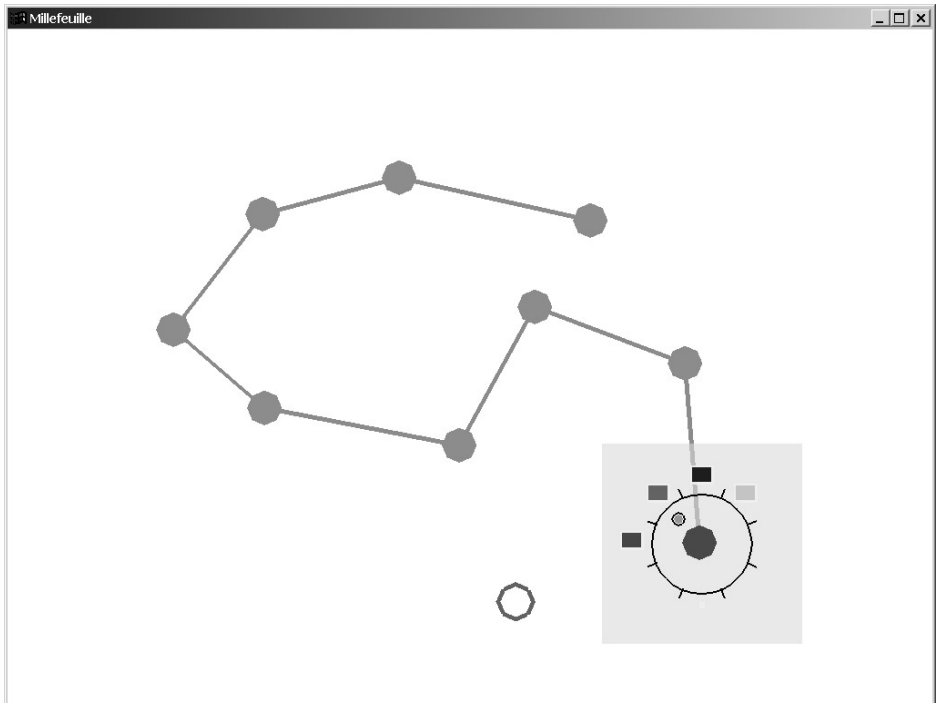
# Design the experimental protocol

- Between or within subjects?
  - Between subjects: each subject run one condition
    - *Need more subjects but more powerful*
  - Within subjects: each subject run several conditions
    - *Need less subjects but less powerful*
  - 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:
  - "Connect the colored dots" task
    - *Similar to area selection and shape creation*
  - Using a pen and a puck in an indirect setting



# Chose the user population

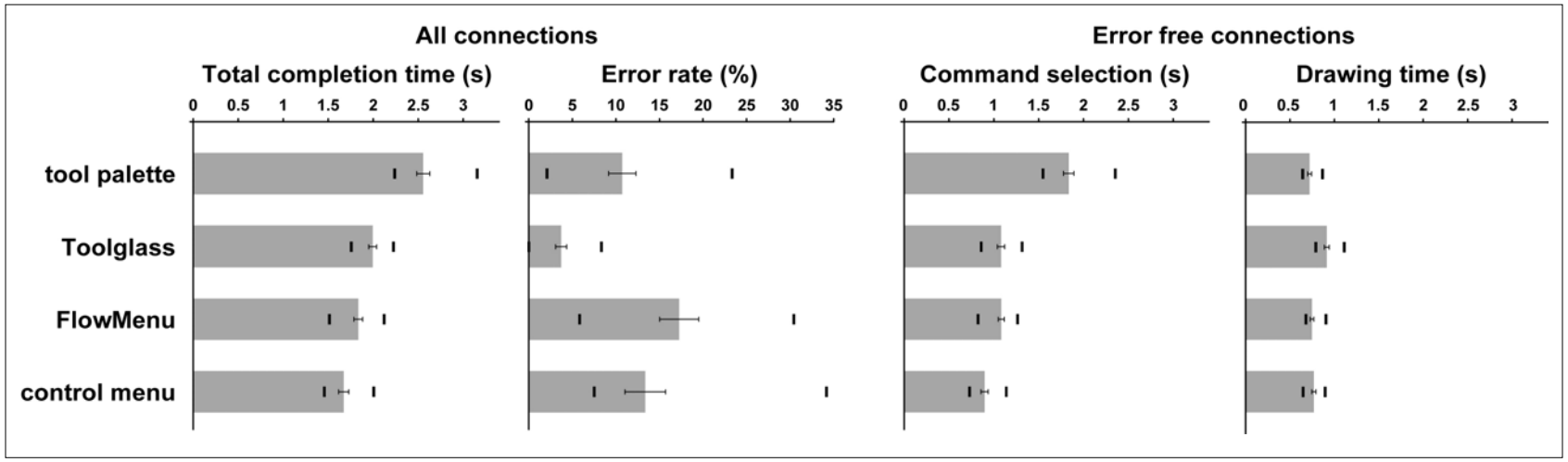
- Pick a well balanced sample
  - Novices, experts, average
  - Age group
  - Sex…
- Population group may be one of the independent variable

- Running example:
  - Non-color blind, right handed adults (male and female)

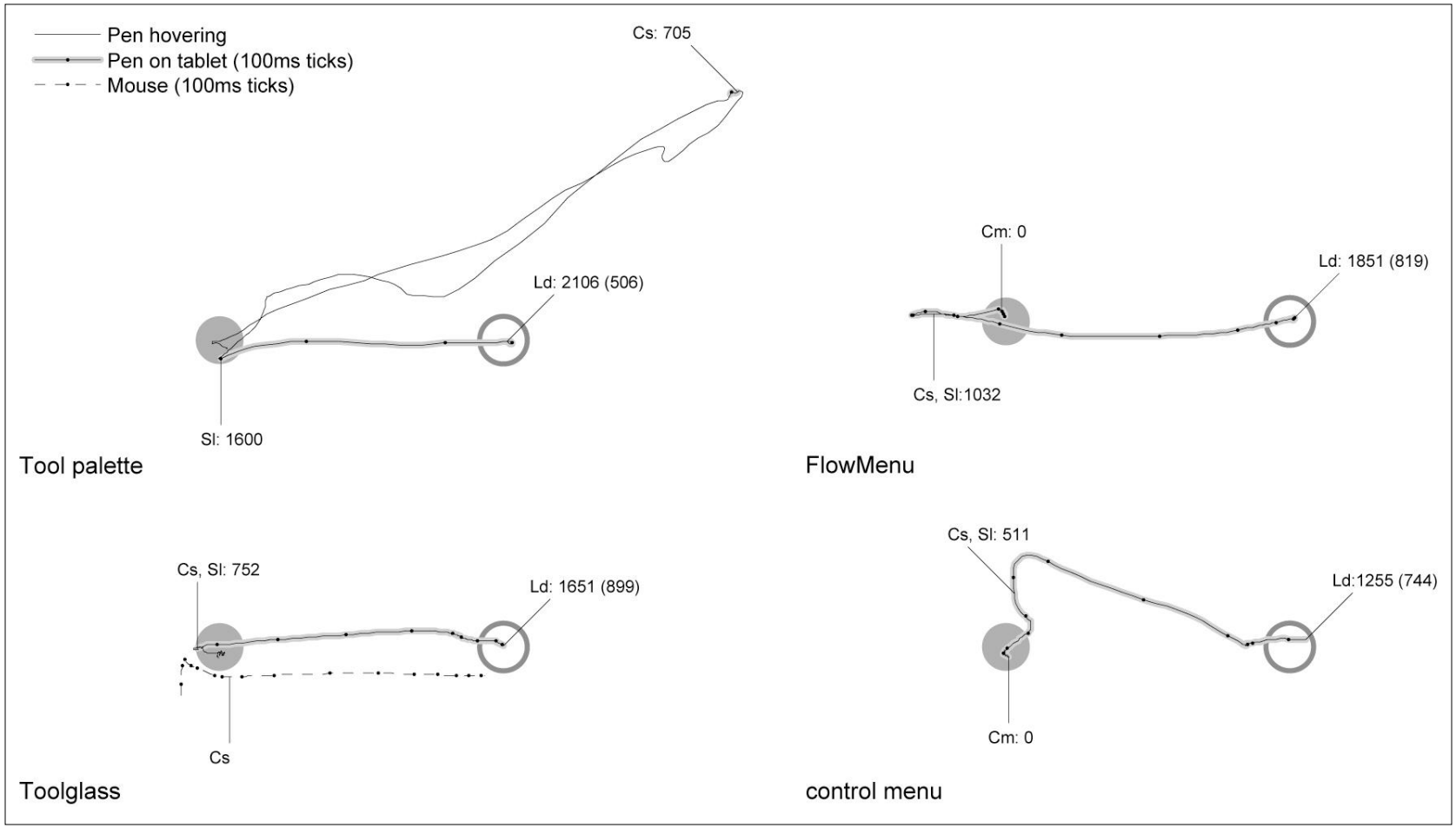
# Run the experiment

- 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 of different distributions?
- Probability that our claims are correct
  - Statistical significance:

"The hypothesis that using a pen in direct mode is faster is accepted ($p < .05$)" means that there is a higher than 95% chance the hypothesis is true

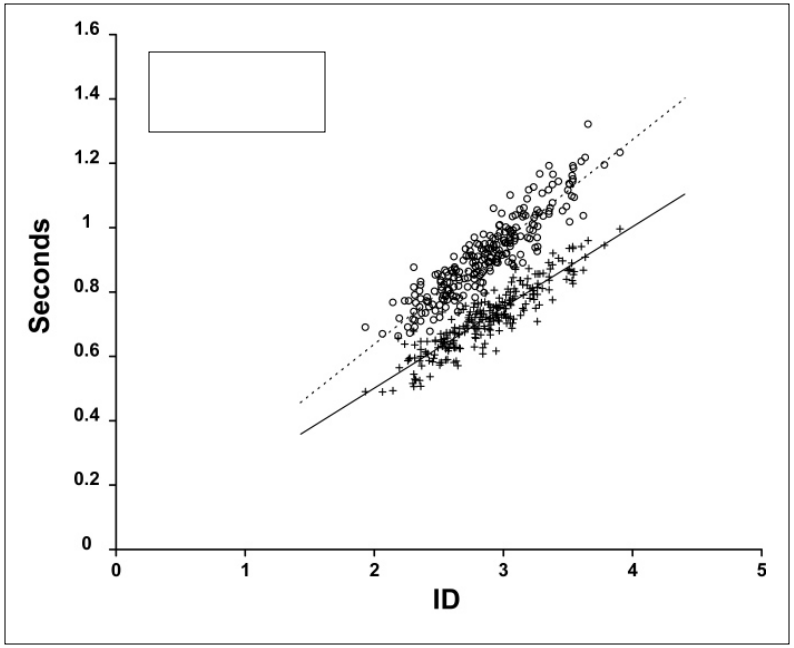
  - Typical level are .05 and .01 level

# 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
  - 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!

# 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

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