Current Topics in Media Computing and HCI

Krishna Subramanian
Media Computing Group
RWTH Aachen University

Summer term 2016

http://hci.rwth-aachen.de/cthci

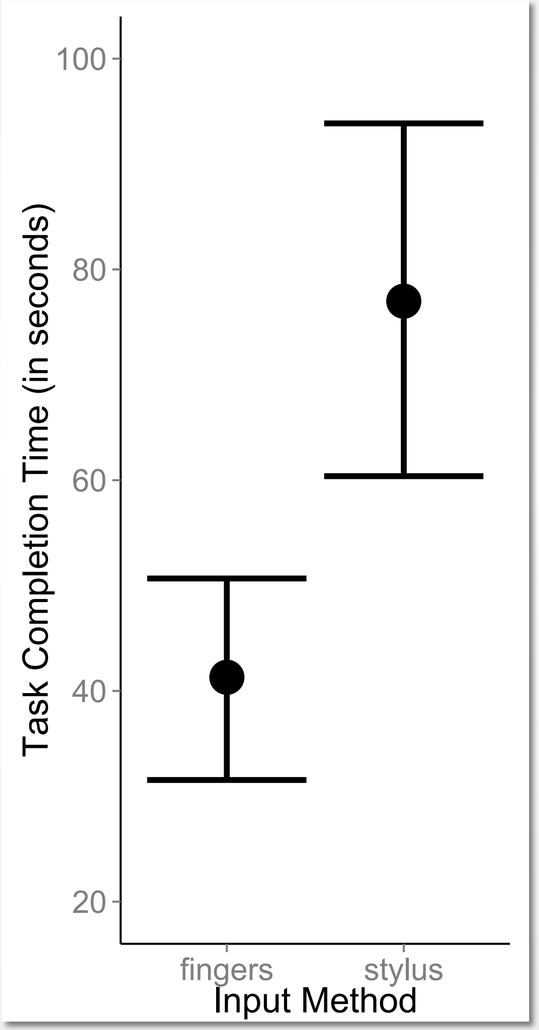
Last Tuesday in Current Topics…

- Types of variables?
- Types of validity?
- Basic experimental designs?

Understanding Statistics in HCI Research

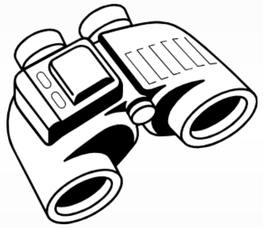
Result of Statistical Analysis

- The input method (fingers, stylus) had a significant effect on the task completion time, $t(20) = 4.03$, $p < .001$.
- Finger (M = 42.03 s; 95% CI [31.78, 52.22]) is faster than Stylus (M = 76.21 s; 95% CI [59.40, 93.02]). Difference between the means is 34.18 s.



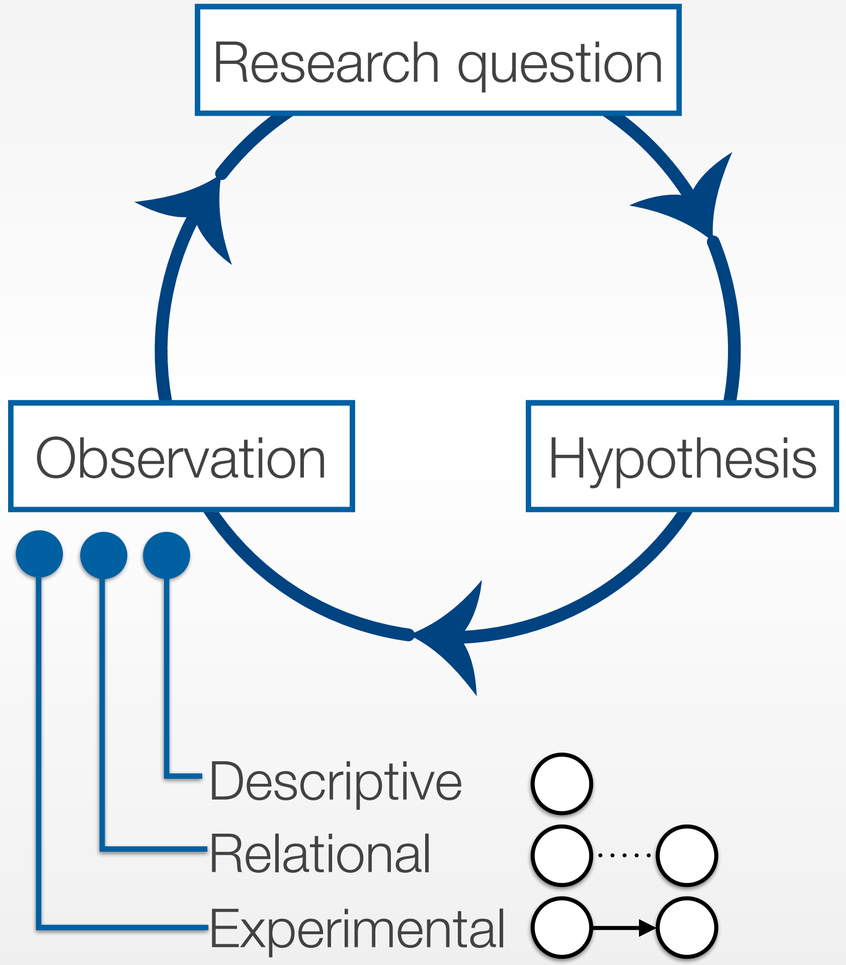
Way Back in Current Topics…

Empirical science



Ethnography

Engineering and design



Scenario: Comparing Input Methods for Typing

Fingers

Stylus

Steps in Experimental Research

1. Formulate hypothesis
2. Design experiment, pick dependent & independent variables, and limit extraneous variables
3. Recruit subjects
4. Run experiment (to collect data which you will analyze)
5. Perform statistical analysis on the collected data to accept or reject hypothesis

1. Formulate hypothesis
2. Design experiment, pick dependent & independent variables, and limit extraneous variables
3. Recruit subjects
4. Run experiment (to collect data which you will analyze)
5. Perform statistical analysis on the collected data to accept or reject hypothesis

- Null hypothesis (H_0): The typing speed when using fingers is not different from the typing speed when using a stylus
- Alternative hypothesis (H_1): The typing speed when using fingers is different from the typing speed when using a stylus

1. Formulate hypothesis
2. Design experiment, pick dependent & independent variables, and limit extraneous variables
3. Recruit subjects
4. Run experiment (to collect data which you will analyze)
5. Perform statistical analysis on the collected data to accept or reject hypothesis

- Experimental design: **Between-subjects design**
- Variables
  - Independent variable (IV): Input method with levels *fingers* and *stylus*
  - Dependent variable (DV): Task completion time (in seconds)
- Control other variables (user experience, model of the smartphone, etc.)

1. Formulate hypothesis
2. Design experiment, pick dependent & independent variables, and limit extraneous variables
3. Recruit subjects
4. Run experiment (to collect data which you will analyze)
5. Perform statistical analysis on the collected data to accept or reject hypothesis

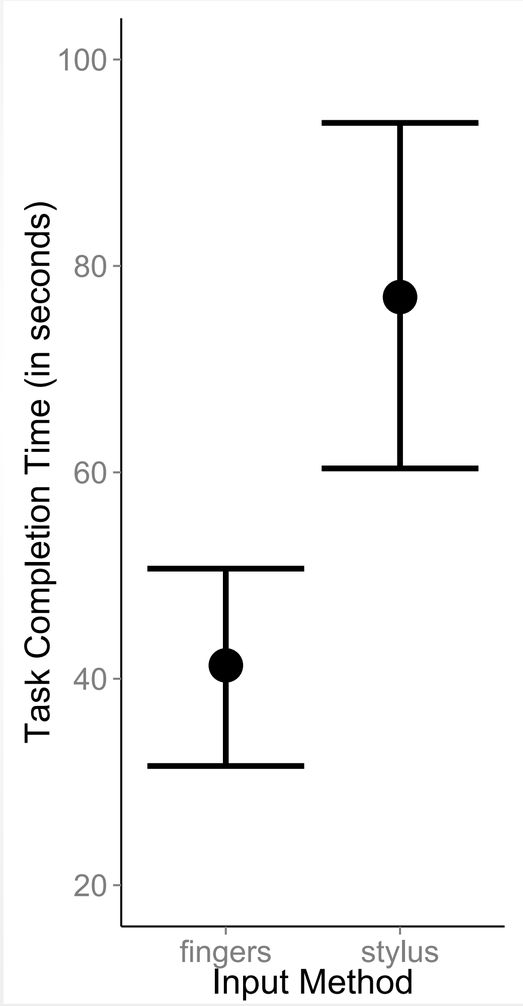
Raw Data

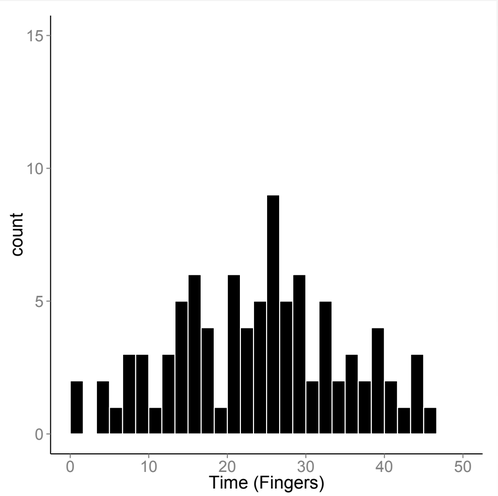
participant_ID, input_type, typing_speed
1, Fingers, 70
2, stylus, 90
3, Fingers, 50
4, stylus, 60
5, Fingers, 90
6, stylus, 85
…

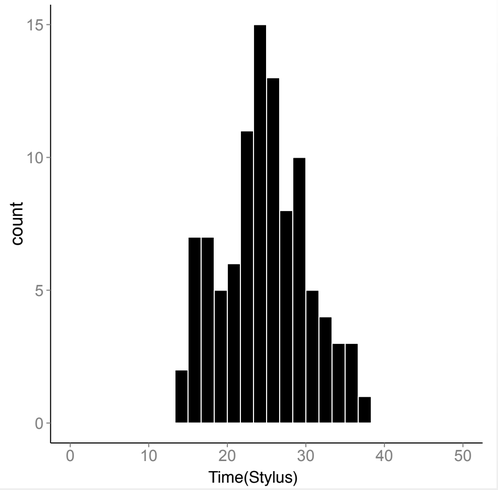
Hard to analyze!

Descriptive Statistics & Visualizations

- Measures of central tendency
  - Mean, median, and mode
- Measures of spread
  - Variance and standard deviation

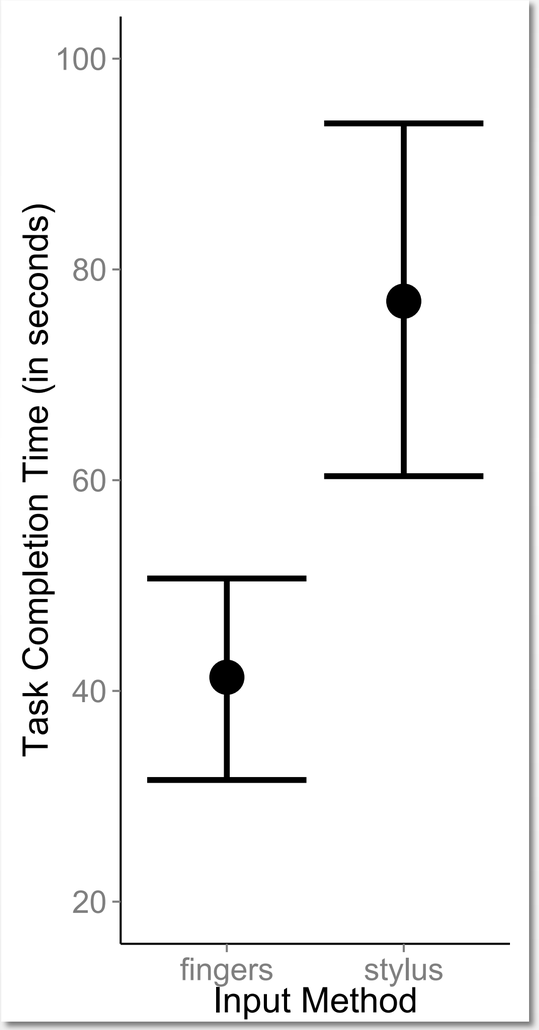






Result of Statistical Analysis

- The input method (fingers, stylus) had a significant effect on the task completion time, $t(20) = 4.03$, $p < .001$.
- Finger (**M = 42.03 s**; 95% CI [31.78, 52.22]) is faster than Stylus (**M = 76.21 s**; 95% CI [59.40, 93.02]). Difference between the means is 34.18 s.



Descriptive Statistics & Visualizations

+ Get a summary of data
+ Detect patterns in data

— Findings valid only for sample, not for the population

Statistical Significance Testing

Statistical Significance Testing

Is there a difference between the distributions at the population level?

Null Hypothesis Significance Testing (NHST)

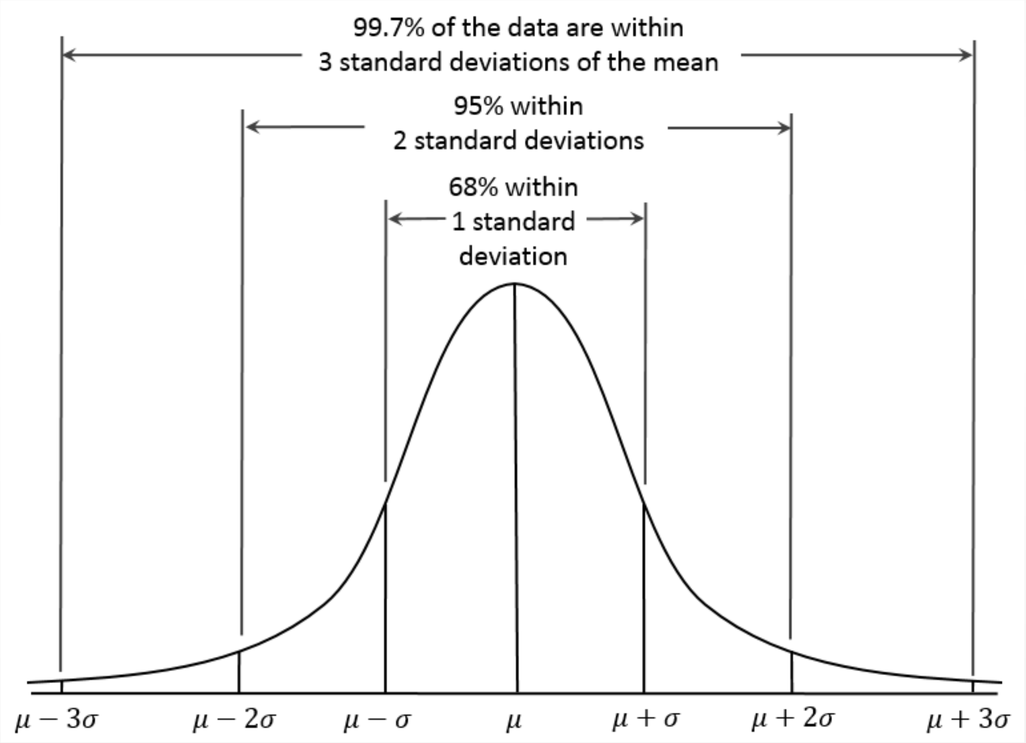
- Commonly used method for significance testing
- Difference in means between sampled distributions

  => difference in the populations (significant difference)

  => no difference in populations, difference is due to random chance (sampling)
- Purpose of NHST: To tell these two differences apart

Normal Distributions

- Characteristic “bell-shape” of the distribution
- Central Limit Theorem
  - “Distribution of a large number of independent, identically distributed variables will be approximately normal…”

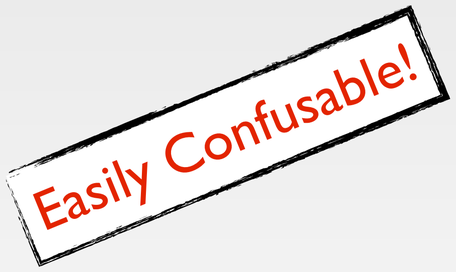


Null Hypothesis Significance Testing (NHST)

- Assume H_0 to be true (no difference)
- Conduct the experiment and collect data
- Fit a statistical model to the data (assuming H_0 is true)
- Compute *p*-value
  - “Chances of obtaining the experimental data we’ve collected assuming the null hypothesis is true”

Null Hypothesis Significance Testing (NHST)

- *De facto* cutoff level of $p = 0.05$ for statistical significance
  - **$p \leq 0.05$** => **reject** H_0 (and accept H_1)
  - **$p > 0.05$** => **accept** H_0



In-class Exercise: *p* value

- Which of the following statements are correct?

A. There is a 3% probability that school students watch more TV than college students

Incorrect: p-value cannot claim alternative hypothesis

B. There is a 3% probability that school students watch TV in a different amount than college students

Incorrect: p-value cannot claim alternative hypothesis

C. Assuming that school students watch TV in different amount than college students, there is a 3% probability that we obtained our data

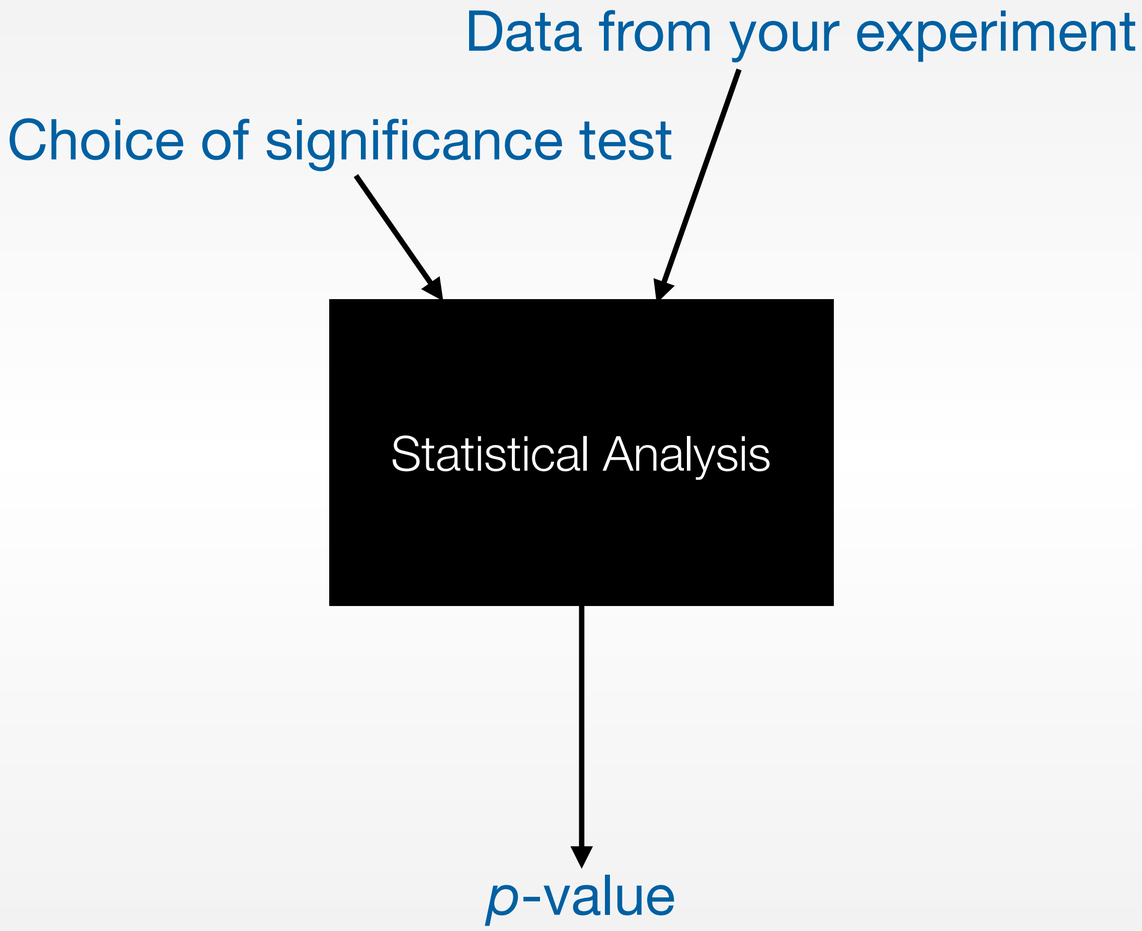
Incorrect: p-value doesn't tell you the direction of difference

D. Assuming that school students and college students watch TV in the same amount, there is a 3% probability that we obtained our data

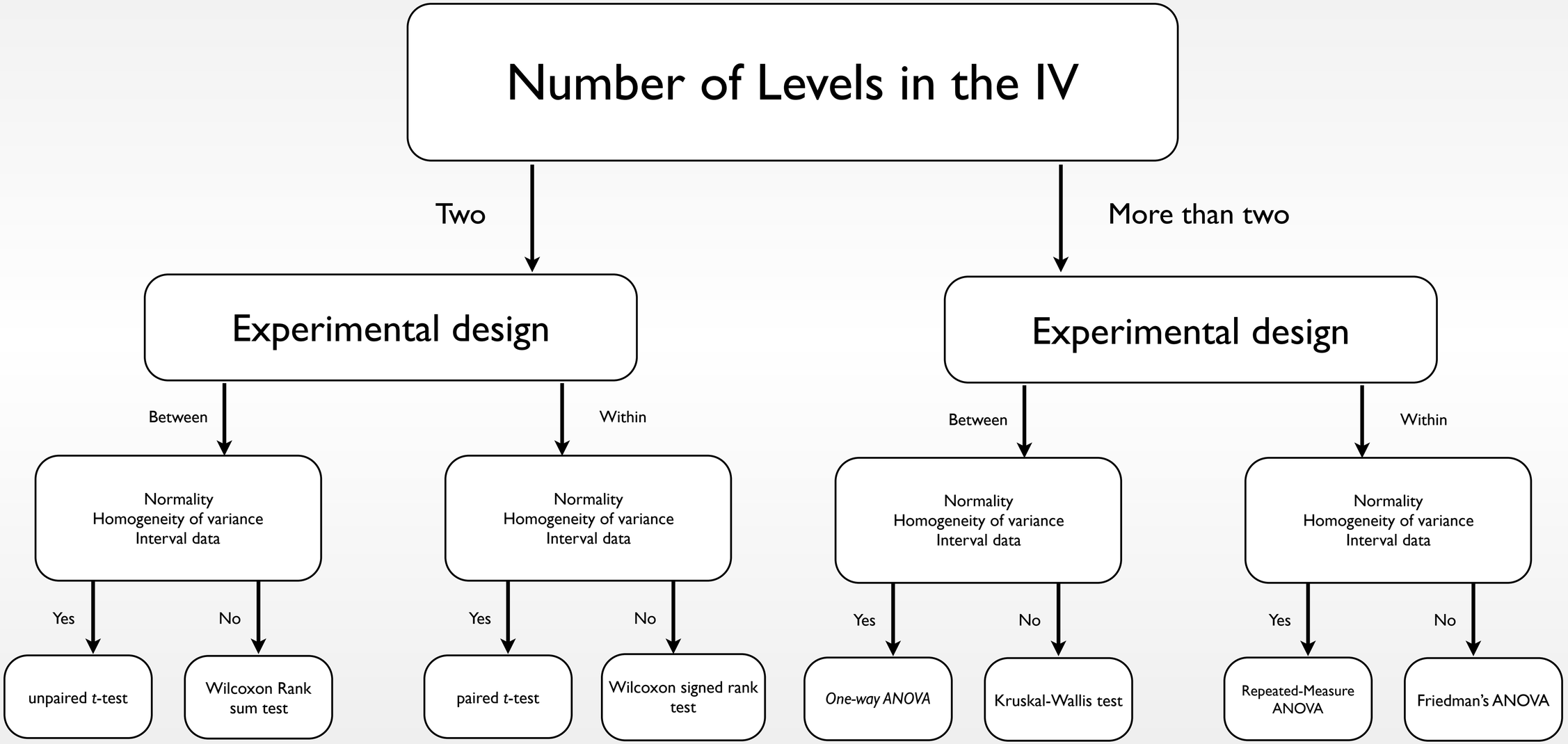
Correct!

A Few Words on NHST

- **Test statistic:** A measure of how well our data fits a statistical model (e.g., *t*-distribution, *F*-distribution, etc.)
- *p*-value is computed from test statistic
- *p*-value is sensitive to sample sizes

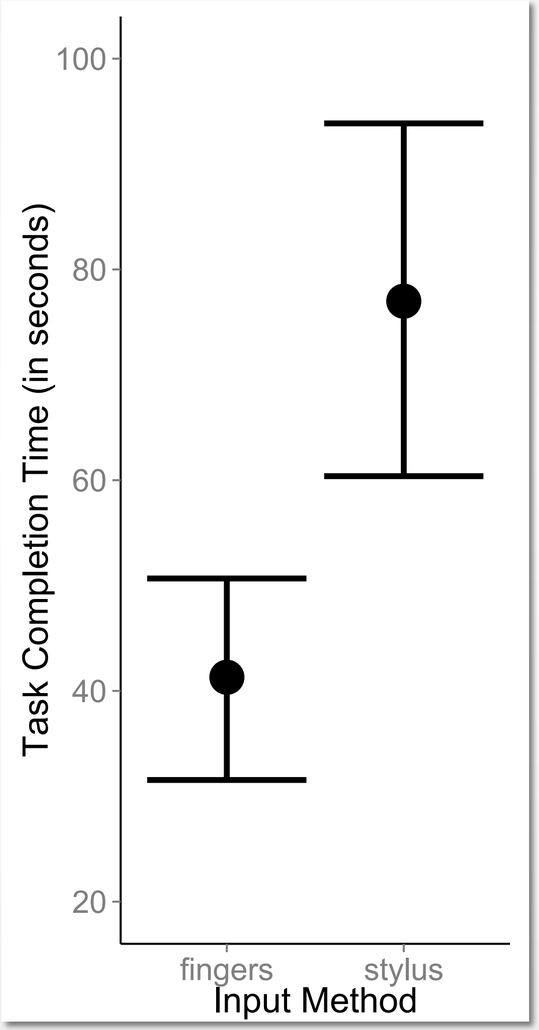

Data from your experiment
Choice of significance test
Statistical Analysis
p-value

Do the Authors Use the Correct



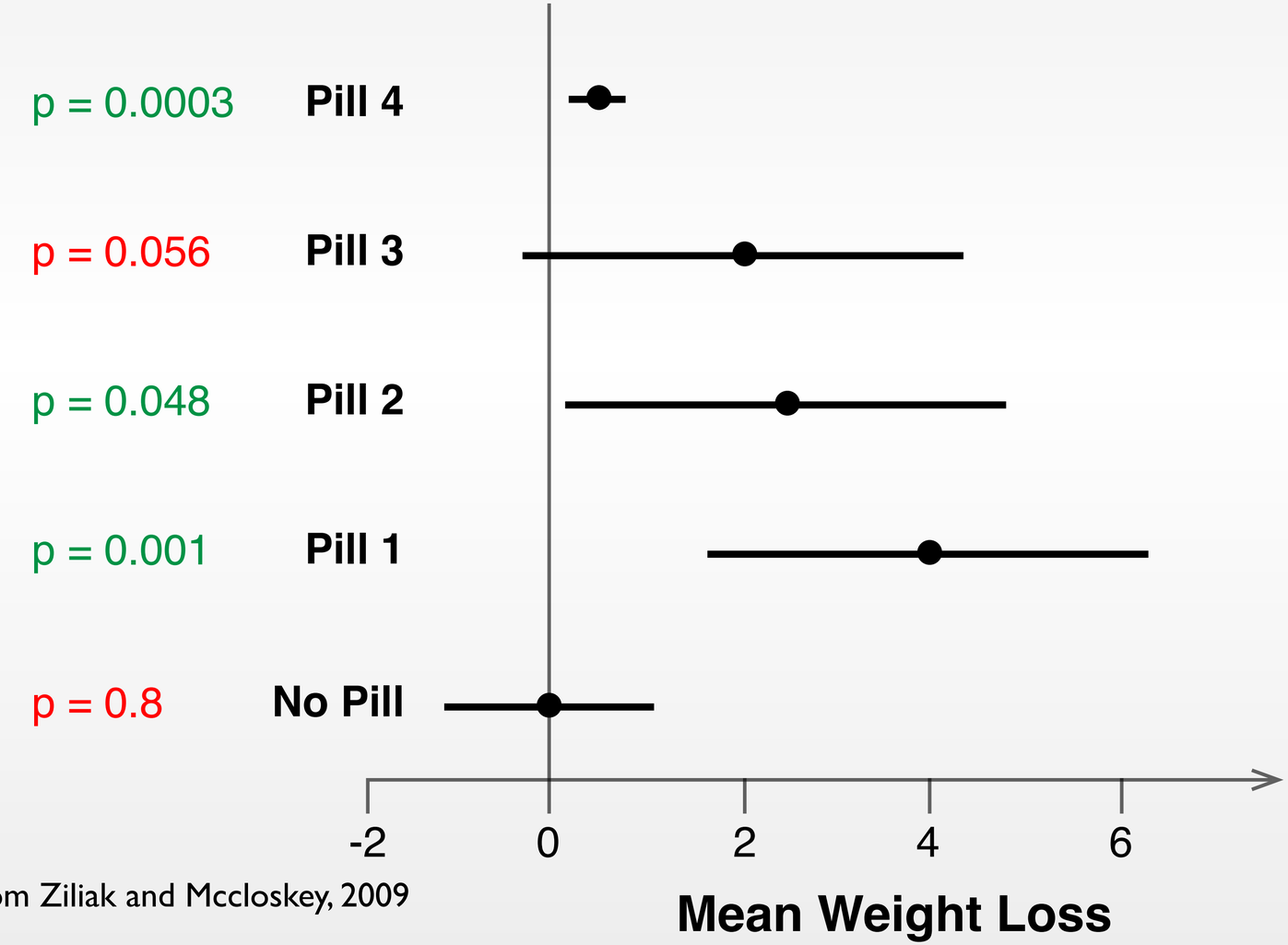
Result of Statistical Analysis

- The input method (fingers, stylus) had a significant effect on the task completion time, ***t*(20) = 4.03, *p* < .001**.
- Finger (*M* = 42.03 s; 95% CI [31.78, 52.22]) is faster than Stylus (*M* = 76.21 s; 95% CI [59.40, 93.02]). Difference between the means is 34.18 s.



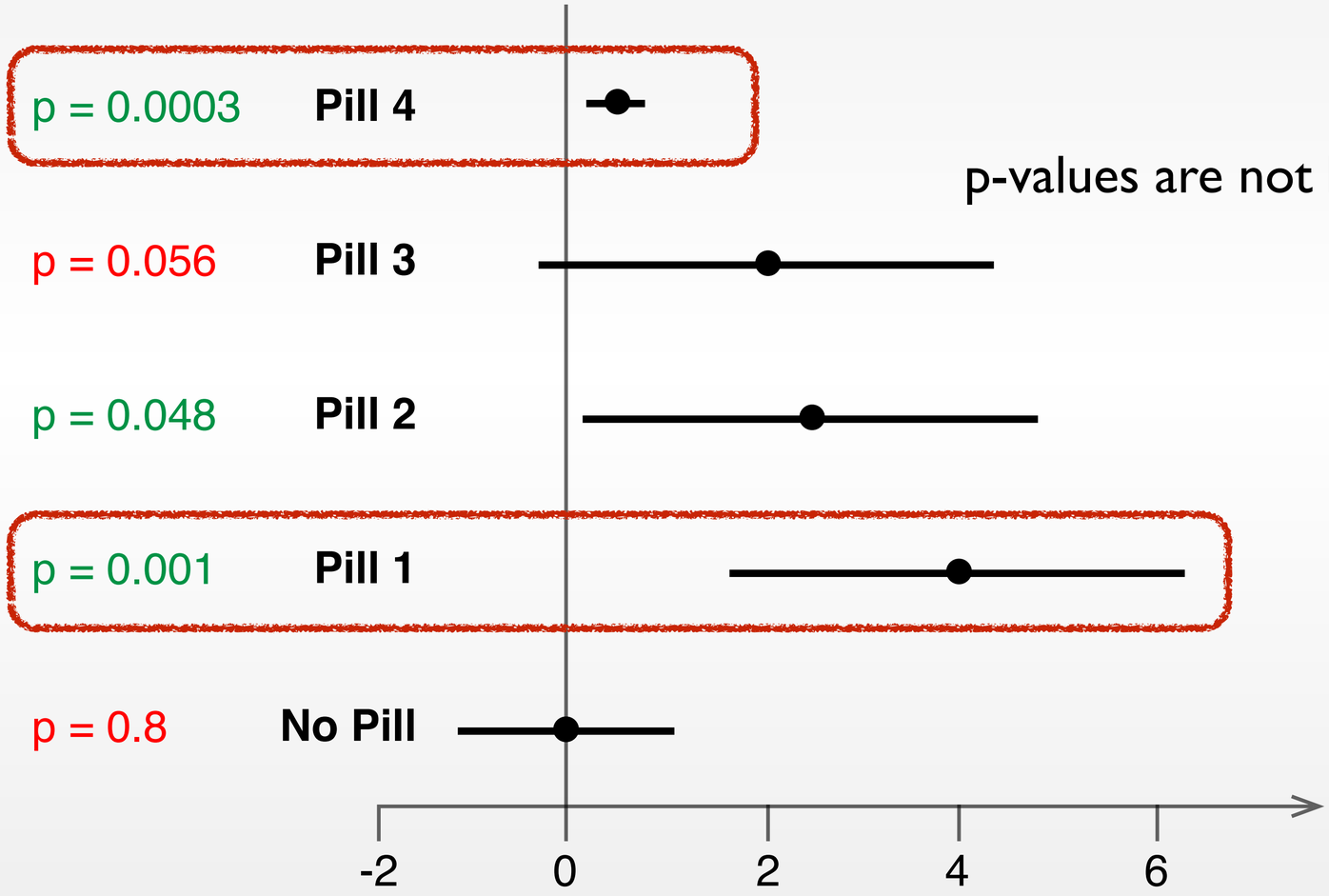
Interpreting Uncertainty in Data

How Uncertainty Influences our Interpretation



Adopted from Ziliak and Mccloskey, 2009

How Uncertainty Influences our Interpretation



p-values are not helpful!

Adopted from Ziliak and Mccloskey, 2009

Effect Size

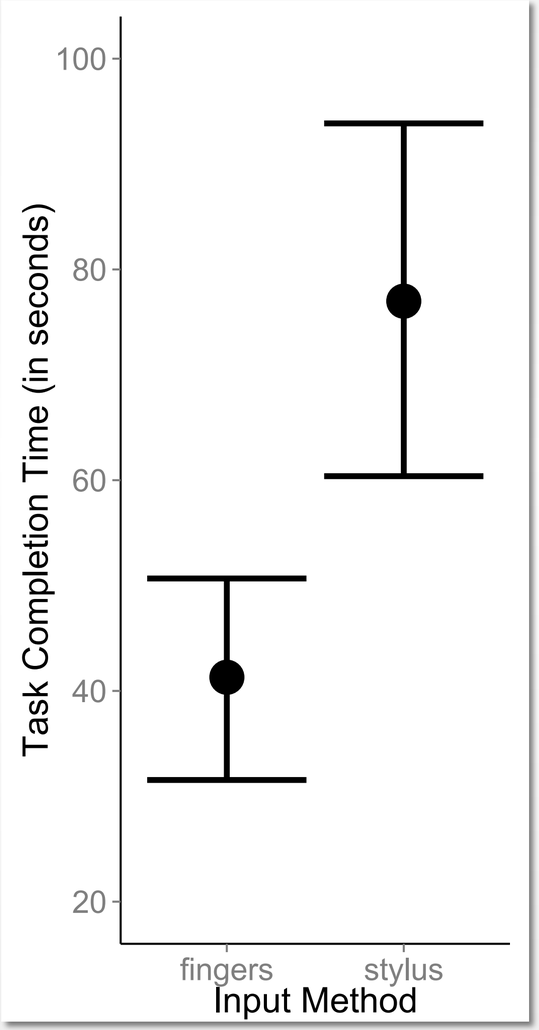
- *p*-value: Is there a difference between distributions at the population level?
  - **But:** Statistically significant ($p < 0.05$) != practically significant
- Need a measure of how the big the difference is (= effect size)

Effect Size: Examples

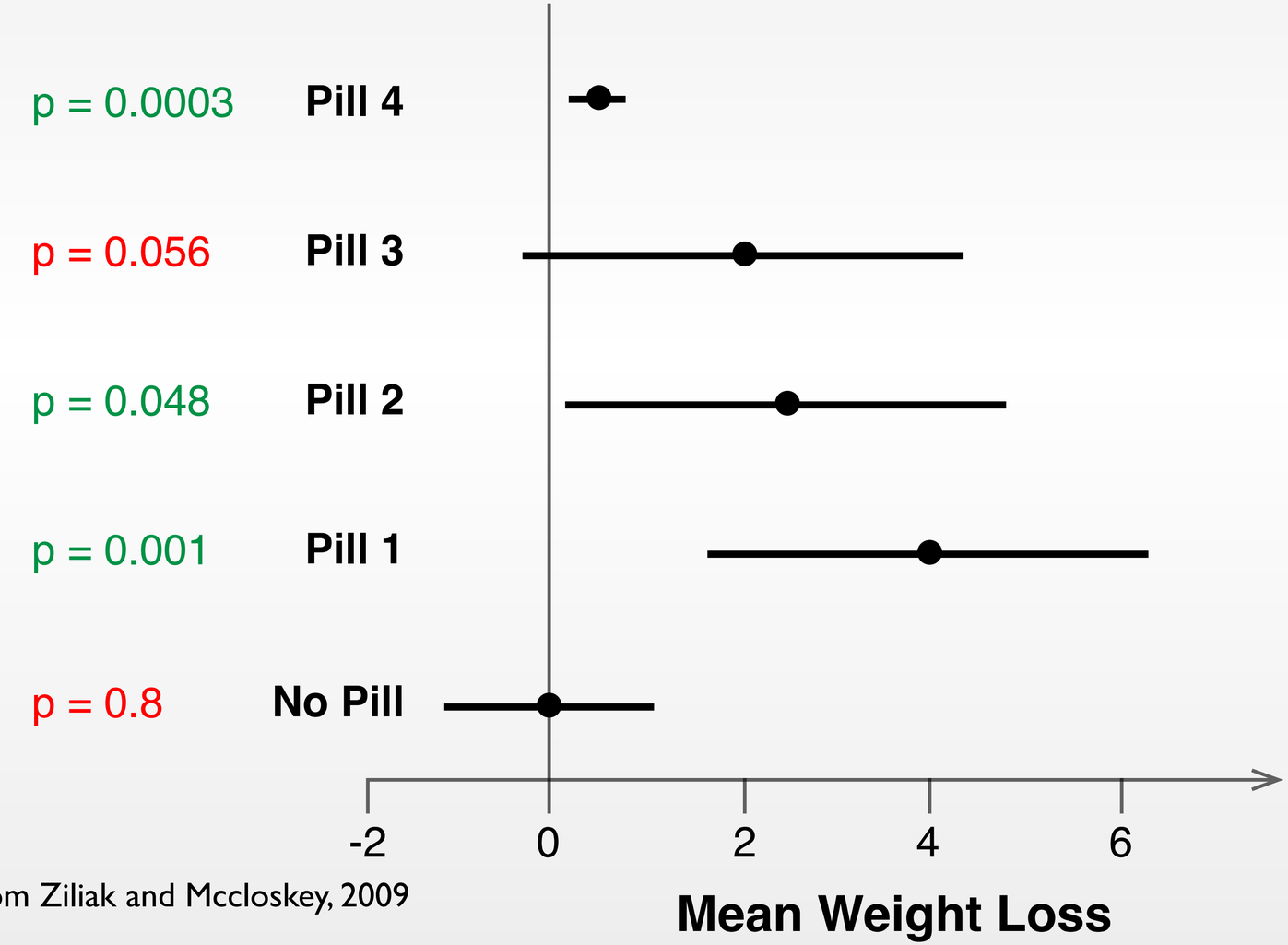
- Difference between two means
  - E.g., Stylus is 40s slower than Touch
  - In original unit, intuitive
- Percentage and ratio
  - E.g., Stylus is twice slower than Touch
  - Emphasize the magnitude of effect
- Difference between means has a measurement unit (e.g., seconds, points, etc.) and therefore requires domain knowledge

Result of Statistical Analysis

- The input method (fingers, stylus) had a significant effect on the task completion time, $t(20) = 4.03$, $p < .001$.
- Finger (M = 42.03 s; 95% CI [31.78, 52.22]) is faster than Stylus (M = 76.21 s; 95% CI [59.40, 93.02]). **Difference between the means is 34.18 s**.



How Uncertainty Influences our Interpretation



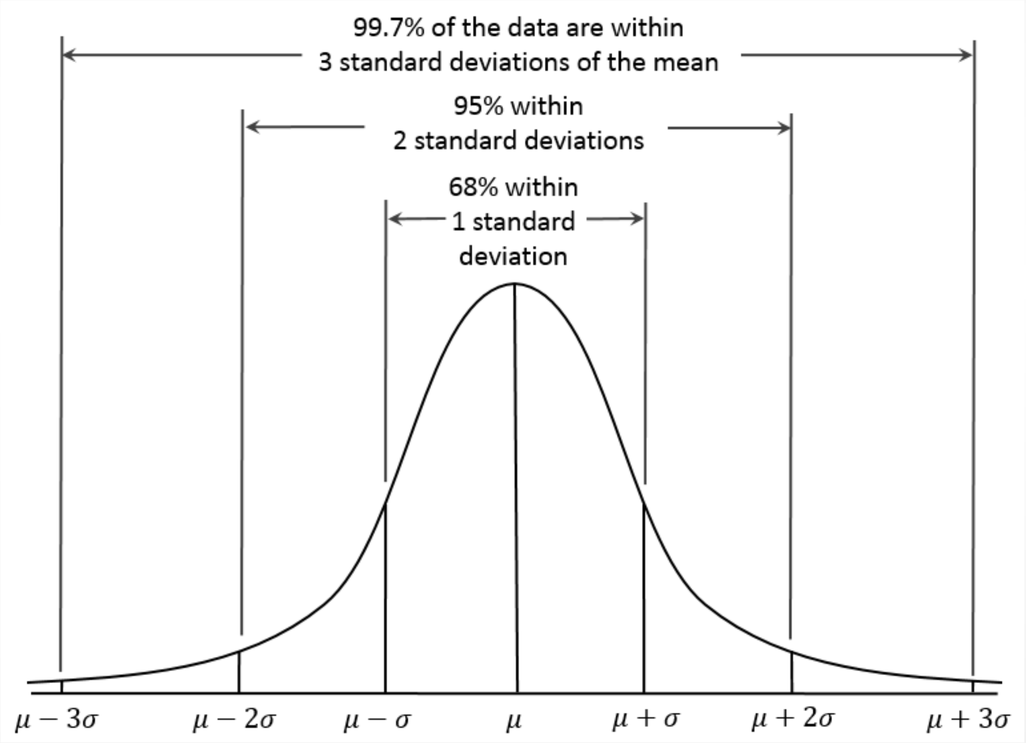
Adopted from Ziliak and Mccloskey, 2009

How Uncertainty Influences our Interpretation

Adopted from Ziliak and Mccloskey, 2009

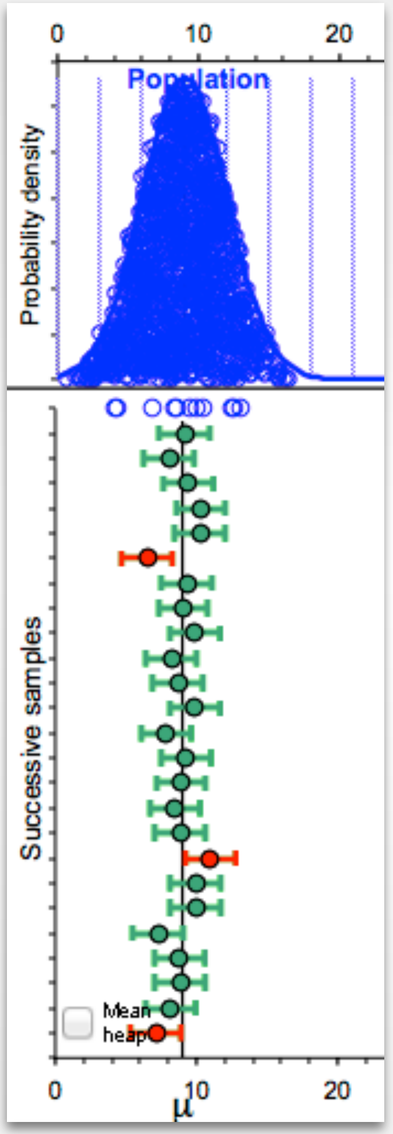
Normal Distributions

- Characteristic “bell-shape” of the distribution
- Central Limit Theorem
  - “Distribution of a large number of independent, identically distributed variables will be approximately normal…”



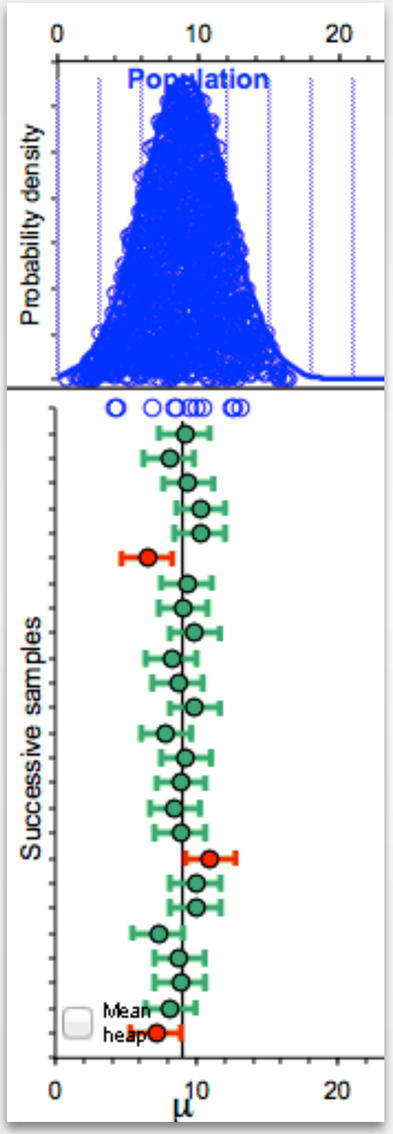
95% Confidence Interval

- An interval estimate (i.e., a range) of the population mean
- In an infinite number of experiments, 95% of the time, the 95% CIs will contain the population mean
- 95% is a convention



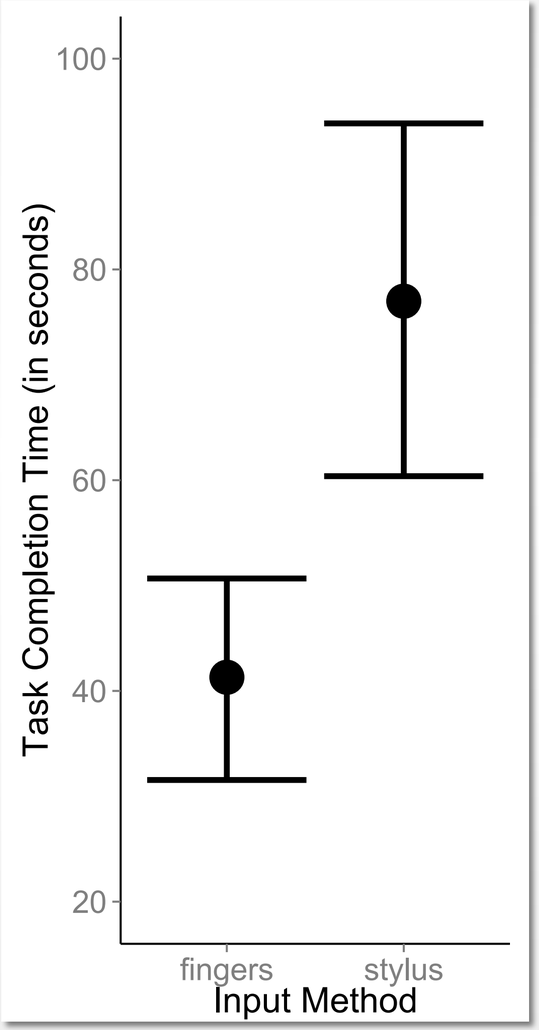
95% Confidence Interval

- Report both mean and confidence interval
  - E.g., M = 39.96 95% CI [25.30, 54.62]



Result of Statistical Analysis

- The input method (fingers, stylus) had a significant effect on the task completion time, $t(20) = 4.03$, $p < .001$.
- Finger (*M* = 42.03 s; **95% CI [31.78, 52.22]**) is faster than Stylus (*M* = 76.21 s; **95% CI [59.40, 93.02]**). Difference between the means is 34.18 s.



Required Reading

- (Cumming and Finch, American Psychologist 2005) *Inference by Eye: Confidence Intervals and How to Read Pictures of Data*
- (Delmas et al., 2005) *Using Assessment Items To Study Students' Difficulty Reading and Interpreting Graphical Representations of Distributions*
  - An exercise sheet on interpreting graphs will be uploaded to L2P (not graded).

Recommended Reading

- Statistical Methods for HCI Research by Koji Yatani, U. of Tokyo
  - Link: http://yatani.jp/teaching/doku.php?id=hcistats:start
- Practical Statistics for HCI by Jacob O. Wobbrock, U. of Washington
  - Uses SPSS and JMP (trial version available for free download)
  - Link: http://depts.washington.edu/aimgroup/proj/ps4hci/
- In-class demo of CI jumping: http://www.latrobe.edu.au/psychology/research/research-areas/cognitive-and-developmental-psychology/esci/understanding-the-new-statistics
  - Chapters 1-4, CIJumping tab

Summary

- We need statistical analysis to establish causal relationship between our IV and DV
- Raw data is hard to analyze
- Descriptive statistics (central tendency, spread) summarize data, but one can't make statements about the population
- NHST can be used to accept or reject null hypothesis
- Effect size quantifies the effect of IV on DV
- Confidence intervals help deal with uncertainty in data