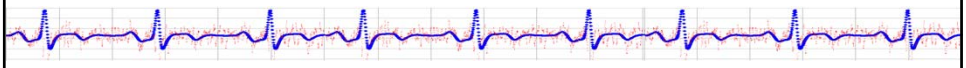
Empirical Research Methods in Information Science

IS 4800 / CS 6350

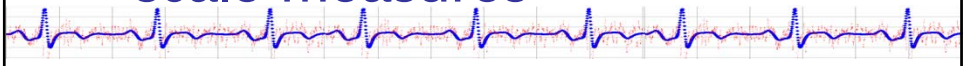


Lecture 7

Information System Measures

1

Rule of thumb for self-report scale measures



- Single item = ordinal
- Composite (~8 or more items) = interval

3

Construct validity

- Your company develops a new “Wild Thing” game that marketing says will make players feel like kids again. They ask you to prove it.
- You develop a “Kidlike” self-report measure with 12 items.
- How do you establish
  - Reliability?
  - Convergent construct validity (wrt existing Irresponsibility and Happiness scales)?
  - Discriminant construct validity (wrt age)?

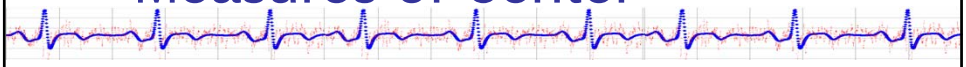
6

Measure example

- *The Processes of Change Questionnaire for Exercise* (Marcus, Rossi, et al., 1992) will be used. Individuals are asked to recall the past month and rate the frequency of occurrence of each of the 39 items on a five-point Likert scale (1=Never to 5=Repeatedly). The internal consistency coefficients for adults range from .62 to .89 (Marcus, Rossi et al, 1992). Discriminant validity was demonstrated by correlating the 10 process subscales with seven demographic variables (sex, age, race, education, income, body mass index and smoking status). None of the absolute values of the correlation coefficients exceeded .25 and the absolute median value was .07 (Marcus, Rossi, et al., 1992). Coefficient Alphas ranged from .67 to .86 ($\underline{M}$ = .78, Nigg et al., 1999).

7

Review: Measures of Center

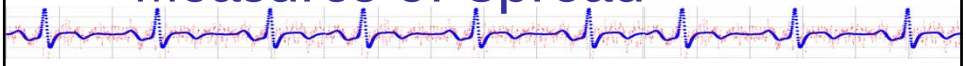


- Mean
- Median
- Mode

- Whazzit?
- When to use?

9

Measures of Spread

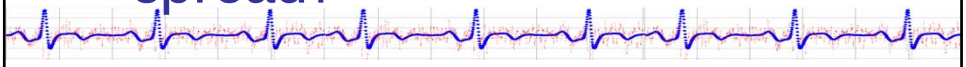


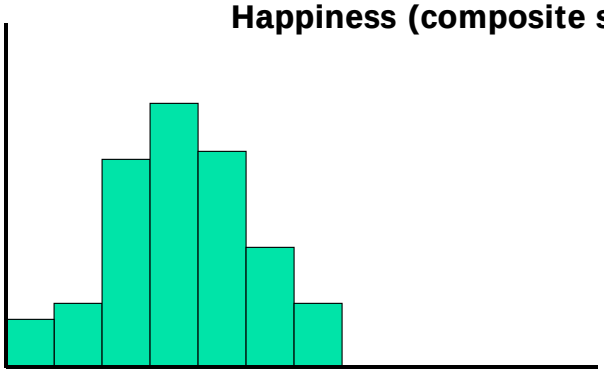
- Std Deviation
- Inter-quartile range
- Range

- Whazzit?
- When to use?

10

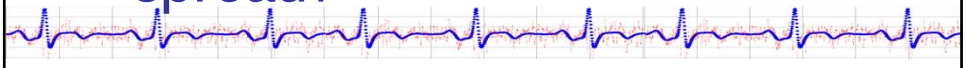
Which measures of center and spread?

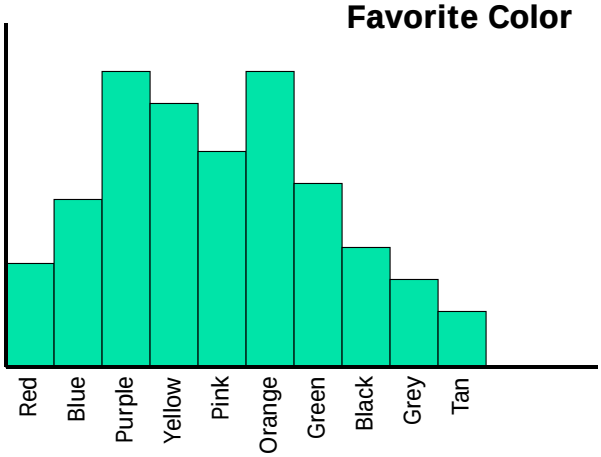




11

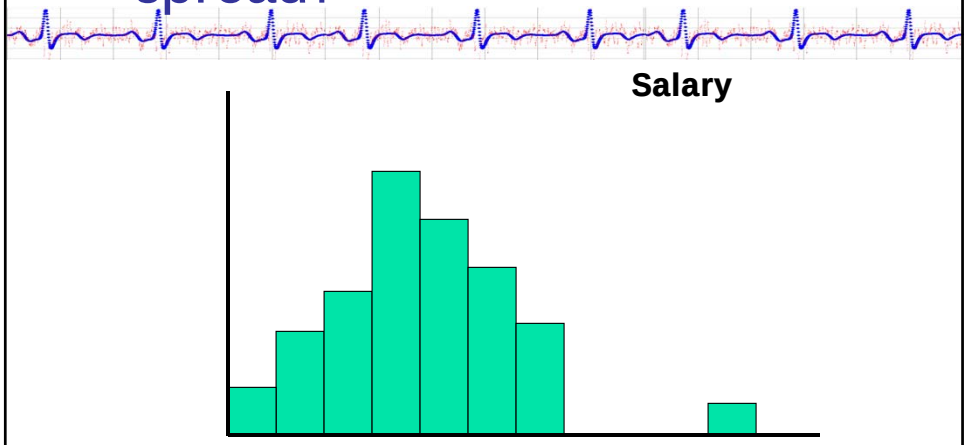
Which measures of center and spread?





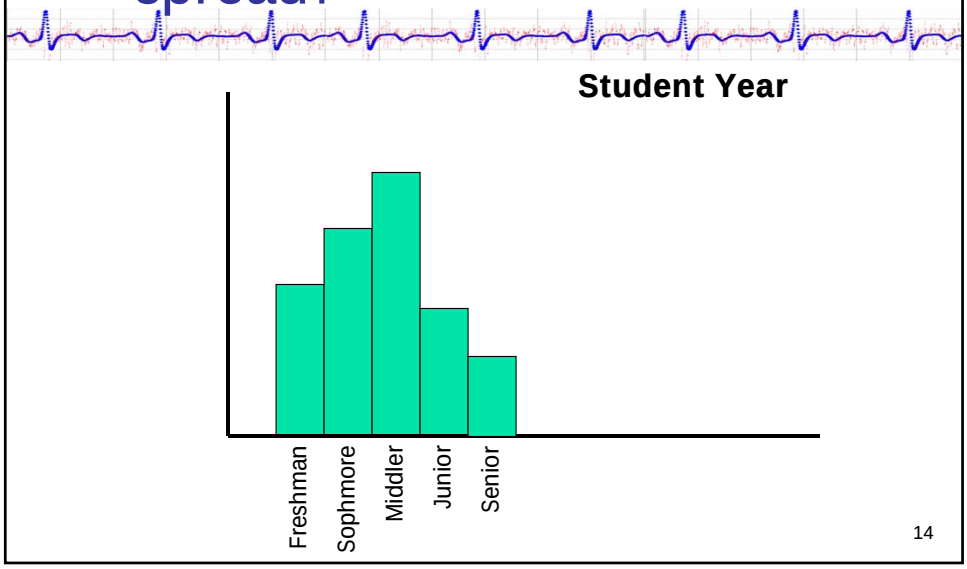
12

Which measures of center and spread?



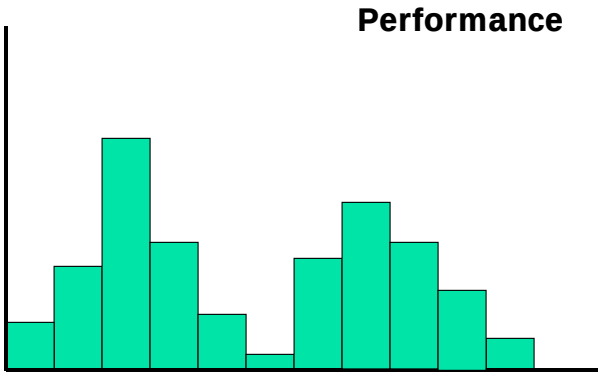
13

Which measures of center and spread?



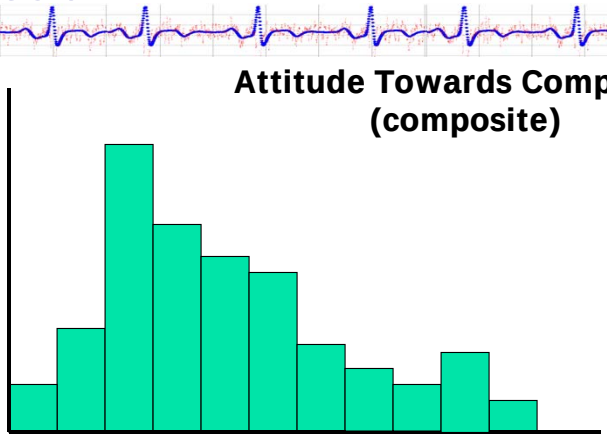
14

Which measures of center and spread?



15

Which measures of center and spread?



16

Aron & Aron Nomenclature

$$M = \frac{\sum X}{N}$$

$$SS = \sum (X - M)^2$$

$$SD^2 = \frac{SS}{N}$$

18

Z-scores

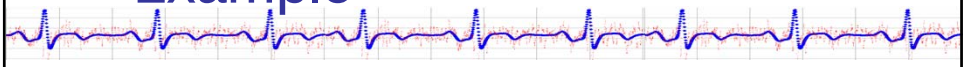
- Measures that have been normalized to make comparisons easier.

$$Z = \frac{X - M}{SD}$$

- Z-scores descriptives
  - Mean?
  - SD?
  - Variance?

19

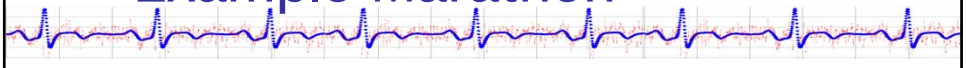
Example



- You collect some data on time to find a table in Curry... what do you do next?

20

Example Marathon



- For each example:
  - What are the types of measures?
  - Describe a possible study
    - Descriptive study
    - Demonstration
    - Correlational study
    - Experimental study

21

Example

- Number of items a person is holding
- Time to get through doorway

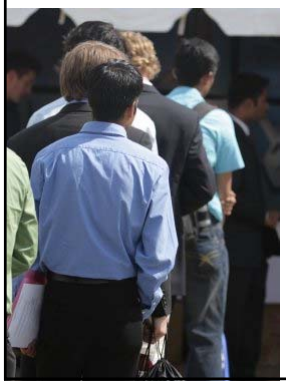
22

Example

- Sign displaying wait times per restaurant vs. no sign
- Time to decide on a restaurant

23

Example



- Time of day relative to class schedule
  - During class time
  - Between class time
- Number of people in line to use a computer

24

Example

- Student age
- Diet

Example

- Fast food restaurant cashier education (highest degree earned)
- Customer ordering time
- Customer satisfaction

26

Example

- Manipulation: InfoCommons seat availability display vs. As-is
- Student happiness (1-7 composite scale)

27

Example

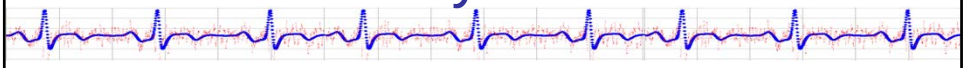
- Money willing to spend on lunch
- Choice of restaurant in food court

Homework I6

- Do Homework I6
  - Enter the following physician order entry dataset into Excel, and import into R.
  - Provide frequency tables, histograms, and descriptive statistics as appropriate.
  - Tabulate counts of JobCategory by Gender. Create a scatter plot of EHREntryTime vs. YearsComputerExperience. Provide boxplots of Accuracy by gender.
  - Turn in: the results of your analyses with narrative text describing the results.
  - Extra credit: turn in the R program for computing all of the above directly from the imported data frame as specified.

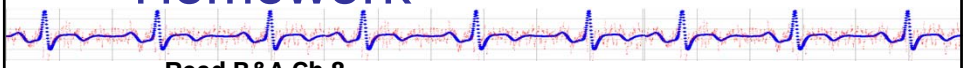
30

Usability Evaluation



31

Homework



- **Read B&A Ch 8**
- **Do Homework I7 – Performance Test (due in one week)**
  1. Pick an obscure piece of software with a user interface (ideally one you may have created for a class). Define two simple tasks using the software (something you can describe in 1-3 sentences and take less than 3 minutes to do) and write them down on two pieces of paper. Select two measures from pages 194-195 of the Nielsen reading that you think may be relevant to the software.
  2. Ask three (or more) classmates or friends to help you with a user study. Make sure they have not used the software before. Obtain verbal consent (as discussed earlier). Provide a brief description of the software (but not how to use it). Then, give each participant each task and watch them attempt to complete it. Do not provide any help. Collect your measures.
  3. Submit a brief writeup of your test plan, descriptive statistics of your data, and any design recommendations resulting from your tests.

32

Terminology from Nielsen Chapter

- Formative vs. Summative evaluation
- Between-Subject vs. Within-Subject evaluation
- Wizard-of-Oz test
- Pilot test
- Thinking aloud study
- Constructive interaction study
- Retrospective testing
- Coaching method

- Example confound: 24x80 alpha vs. 1024x1024 pixel display

33

What is usability?

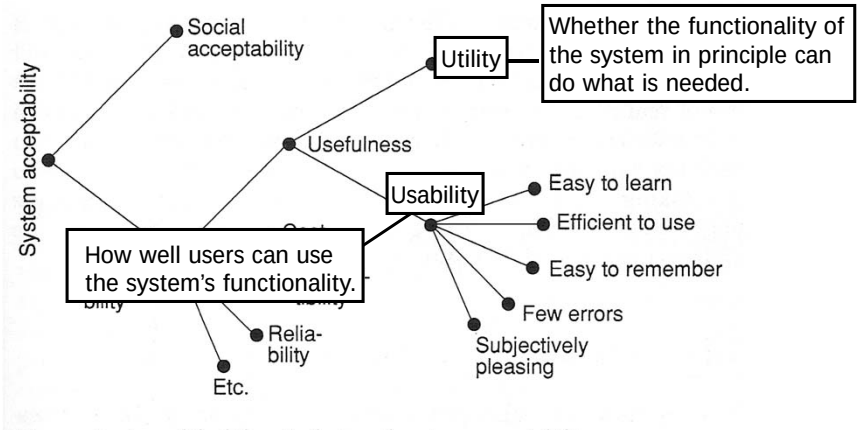


Figure 1 *A model of the attributes of system acceptability.*

From Nielsen, Usability Engineering

34

Usability Evaluation methods

- Analytic/Expert (performed by experts)
  - Inspection methods
  - Model-based evaluation
- Empirical/Usability/User (performed by representative users)
  - Qualitative methods (interviews, questionnaires)
  - Observation in the field
  - Controlled demonstrations or experiments (same environment & task with N alternative designs)

35

Usability Testing

- Usually in the “laboratory”
  - Demonstration
  - Experiment
- Could also do “in the field”
  - Descriptive
  - Demonstration
  - Experiment

36

Kinds of Measures

- Quantiative: Performance measurement
  - Time to complete
  - Error rate
  - etc.
- Qualitative: Thinking aloud
  - aka “verbal protocol analysis”
  - Several variants discussed in Nielsen

37

Test Artifacts

- Final system
- Prototype system
- “Paper prototype”

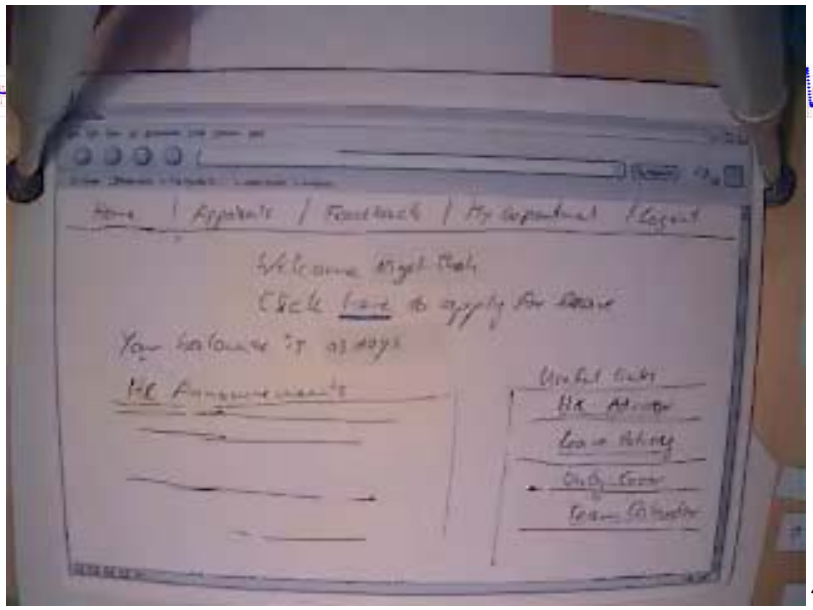
38

Paper Prototyping

- Perform user testing with a paper mock up of your interface
  - One person “plays computer” updating the interface whenever the user interacts with it

39

Example: leave submission system



40

Test Tasks

- Representative as possible
- Reasonable coverage
- Relatively short
- Precise
- Provided to subject in writing

- E.g., “Send an email to Mike Johnson asking him to schedule a meeting today at noon.”

41

Conducting study

- Prepare everything
- Consent subject
- Brief subject on overall purpose of system (best in writing)
- Give subject each task (in writing)
- Experimenter should have no interaction with subject during test
  - May be several observers
- Debrief subject

42

Homework

- Conduct a (small) usability study
  - Descriptive
  - Quantitative
  - Two tasks
  - Two measures
  - Three (or more) subjects

43

Usability Measures
Examples of each?

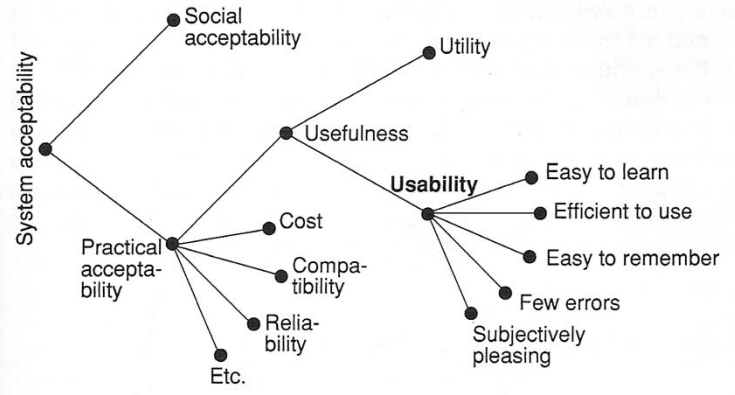


Figure 1 *A model of the attributes of system acceptability.*

44

Papers

- Sensing techniques for mobile interaction
- Usability Study of Physicians' Interaction…

46

Homework

- **Read B&A Ch 8**
- **Do Homework I7 – Performance Test (due in one week)**
  1. Pick an obscure piece of software with a user interface (ideally one you may have created for a class). Define two simple tasks using the software (something you can describe in 1-3 sentences and take less than 3 minutes to do) and write them down on two pieces of paper. Select two measures from pages 194-195 of the Nielsen reading that you think may be relevant to the software.
  2. Ask three (or more) classmates or friends to help you with a user study. Make sure they have not used the software before. Obtain verbal consent (as discussed earlier). Provide a brief description of the software (but not how to use it). Then, give each participant each task and watch them attempt to complete it. Do not provide any help. Collect your measures.
  3. Submit a brief writeup of your test plan, descriptive statistics of your data, and any design recommendations resulting from your tests.

47