Empirical Research Methods in Information Science

IS 4800 / CS6350

Lecture 7

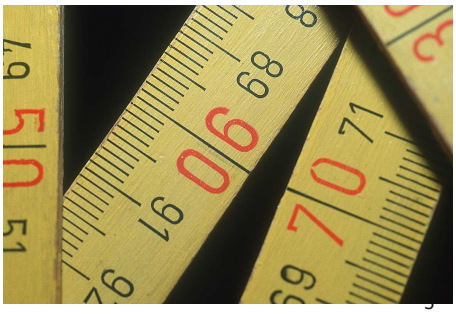
Even more Measures

Quantitative Usability Tests

Ethnography Homework Status?

- Goal: idea to make the library more efficient and friendly
  - Pick a location and spend an hour people watching with a notebook and pencil.
  - Identify an activity you find interesting.
  - Watch several people do it.
  - Interview one or two about it.

Making Systematic Observations



Questionnaire Validation - Summary

- Reliability
  - Test-retest
  - Internal consistency
- Validity
  - Face
  - Content
  - Criterion-related
    - Concurrent
    - Predictive
  - Construct
    - Convergent
    - Discriminant

4

Scales of measurement
aka Levels of measurement

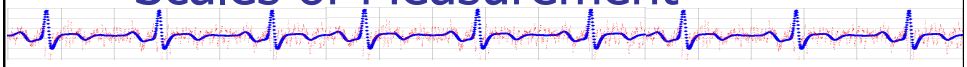
5

Scales of Measurement

- *Nominal Scale*
  - Lowest scale of measurement involving variables whose values differ by category (e.g., male/female)
  - Values of variables have different names, but no ordering of values is implied
- *Ordinal Scale*
  - Higher scale of measurement than nominal scale
  - Different values of a variable can be ranked according to quantity (e.g., high, moderate, or low self-esteem)

6

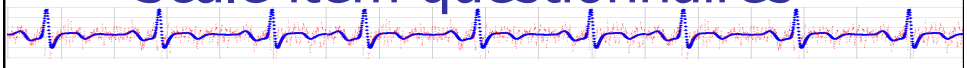
Scales of Measurement



- *Interval Scale*
  - Scale of measurement on which the spacing between values is known (e.g., IQ)
  - No meaningful zero point
- *Ratio Scale*
  - Similar to interval scale, but with a true zero point (e.g., number of lever presses)

7

Scale item questionnaires



- Treat a single item as ordinal
- Treat a composite questionnaire (with at least six items) as interval

8

Concerns with Measures

- Sensitivity
  - Is a dependent measure sensitive enough to detect the change you are interested in?
  - An insensitive measure will not detect subtle behaviors
- Range Effects
  - Occur when a dependent measure has an upper or lower limit
    - *Ceiling effect:* When a dependent measure has an upper limit
    - *Floor effect:* When a dependent measure has a lower limit.

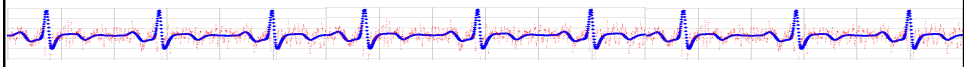
9

Chapter 13

Describing Data

10

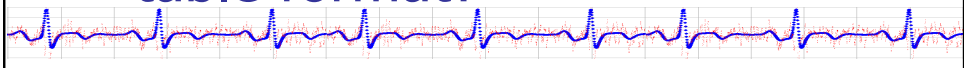
Doing Exploratory Data Analysis



- Use *EXPLORATORY DATA ANALYSIS* (EDA) to search for patterns in your data
- Before conducting any inferential statistic, use EDA to ensure that your data meet the requirements and assumptions of the test you are planning to use (e.g., normally distributed)
- More on data prep later…

11

Stacked vs. Unstacked data table format?



- Unstacked = 1 row per subject
- Stacked = 1 row per observation

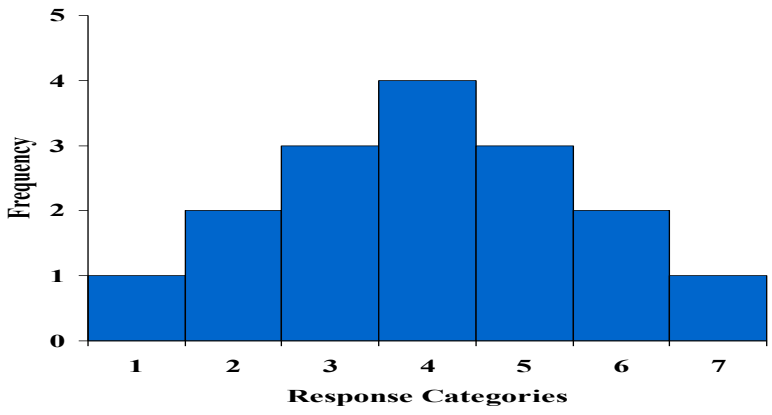
12

The Frequency Distribution

- Represents a set of mutually exclusive categories into which actual values are classified
- Can take the form of a table or a graph
- Graphically, a frequency distribution is shown on a *histogram*
  - A bar graph in which the bars touch
  - The y-axis represents a frequency count of the number of observations falling into a category
  - Categories represented on the x-axis

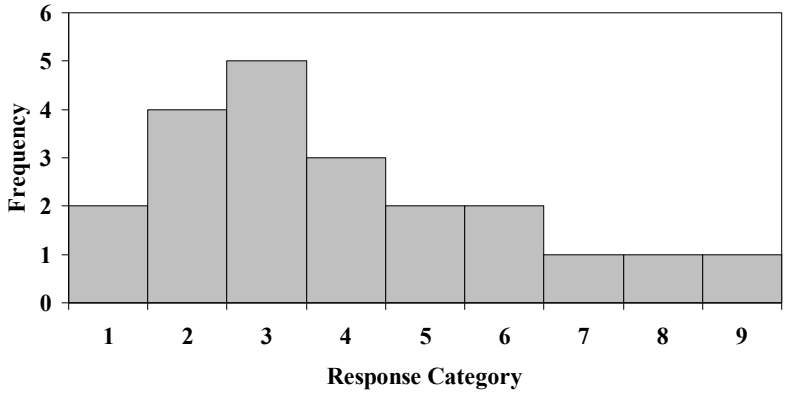
13

Histogram Showing a Normal Distribution



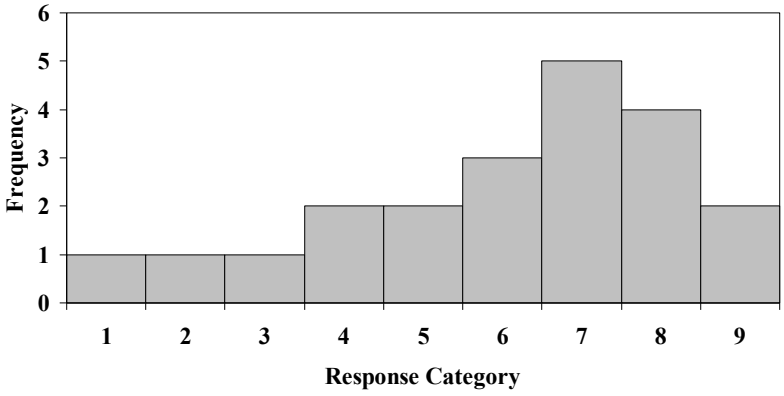
14

Histogram Showing a Positive Skew



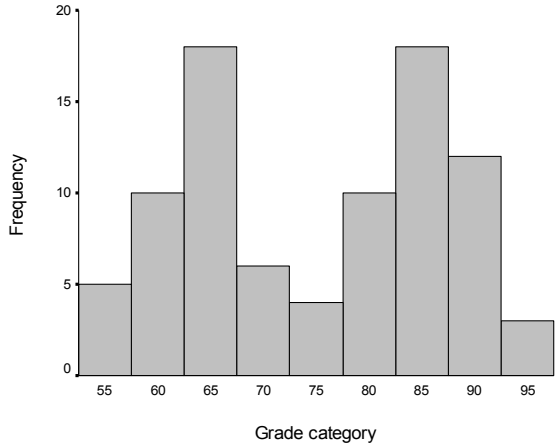
15

Histogram Showing a Negative Skew



16

A Bimodal Distribution



17

Descriptive Statistics

- Statistic = a number used to describe some feature of a group of measurements
- Tim's rule: For every measure you must have
  - Exactly one statistic describing a measure of center
  - Zero or one statistic describing a measure of spread

18

Measures of Center

- Mean
- Median
- Mode

- Whazzit?
- When to use?

19

Measures of Center: Characteristics and Applications

- *Mode*
  - Most frequent score in a distribution
  - Simplest measure of center
  - Scores other than the most frequent not considered
  - Limited application and value
- *Median*
  - Central score in an ordered distribution
  - More information taken into account than with the mode
  - Relatively insensitive to outliers
  - Prefer when data is skewed
  - Used primarily when the mean cannot be used

20

Measures of Center: Characteristics and Applications

- *Mean*
  - Average of all scores in a distribution
  - Value dependent on each score in a distribution
  - Most widely used and informative measure of center

21

Measures of Center: Applications

- *Mode*
  - Used if data are measured along a nominal scale
- *Median*
  - Used if data are measured along an ordinal scale
  - Used if interval or ratio data do not meet requirements for using the mean (skewed but unimodal), or if significant outliers

22

Measures of Center: Applications

- *Mean*
  - Used if data are measured along an interval or ratio scale
  - Most sensitive measure of center
  - Used if scores are normally distributed

23

Measures of Spread

- Std Deviation
- Inter-quartile range
- Range

- Whazzit?
- When to use?

24

Measures of Spread: Characteristics

- *Range*
  - Subtract the lowest from the highest score in a distribution of scores
  - Simplest and least informative measure of spread
  - Scores between extremes are not taken into account
  - Very sensitive to extreme scores
- *Interquartile Range*
  - Less sensitive than the range to extreme scores
  - Used when you want a simple, rough estimate of spread

25

Measures of Spread: Characteristics

- *Variance*
  - Average squared deviation of scores from the mean
  - Divide sum of squared deviations by N or N-1?
- *Standard Deviation*
  - Square root of the variance
  - Most widely used measure of spread

Sample Variance

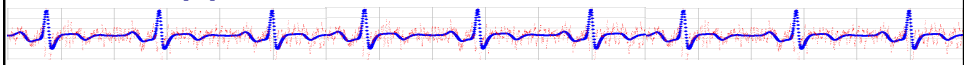
$$s^2 = \frac{\sum(x - \bar{x})^2}{n - 1}$$

Sample Standard Deviation

$$s = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}}$$

26

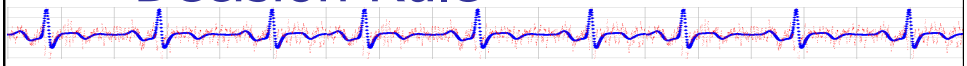
Measures of Spread: Applications



- The range and standard deviation are sensitive to extreme scores ("outliers")
- When your distribution of scores is skewed, the standard deviation does not provide a good index of spread

27

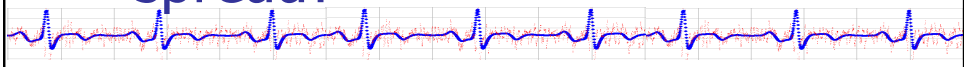
Measures of Spread: Decision Rule

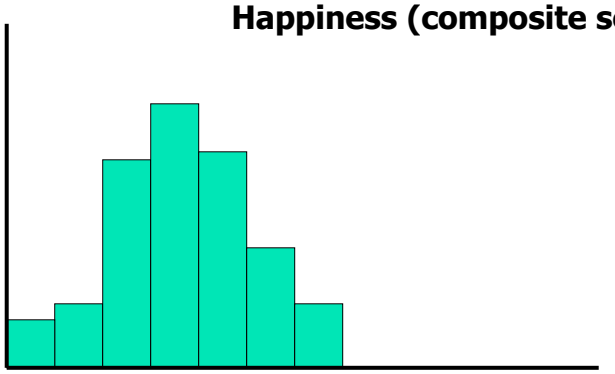


- Nominal, Ordinal => no measure of spread
- Interval, Ratio & Normal & no outliers => SD
- Else: IQR

28

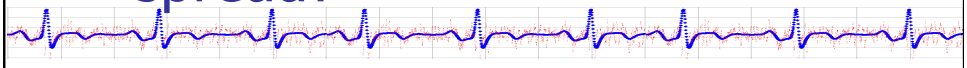
Which measures of center and spread?

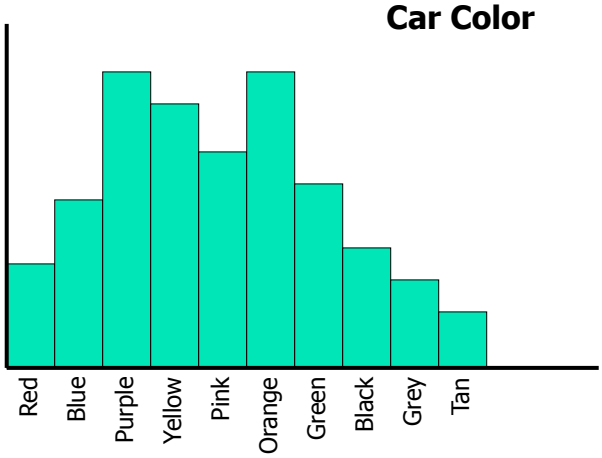




29

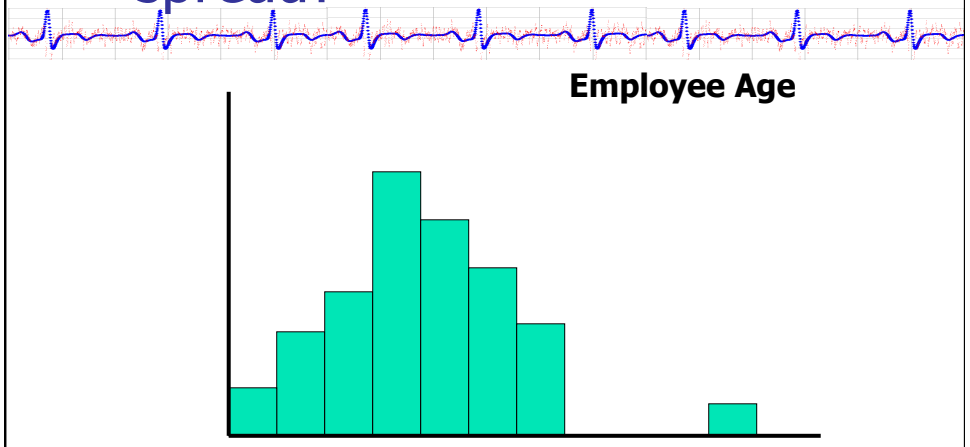
Which measures of center and spread?





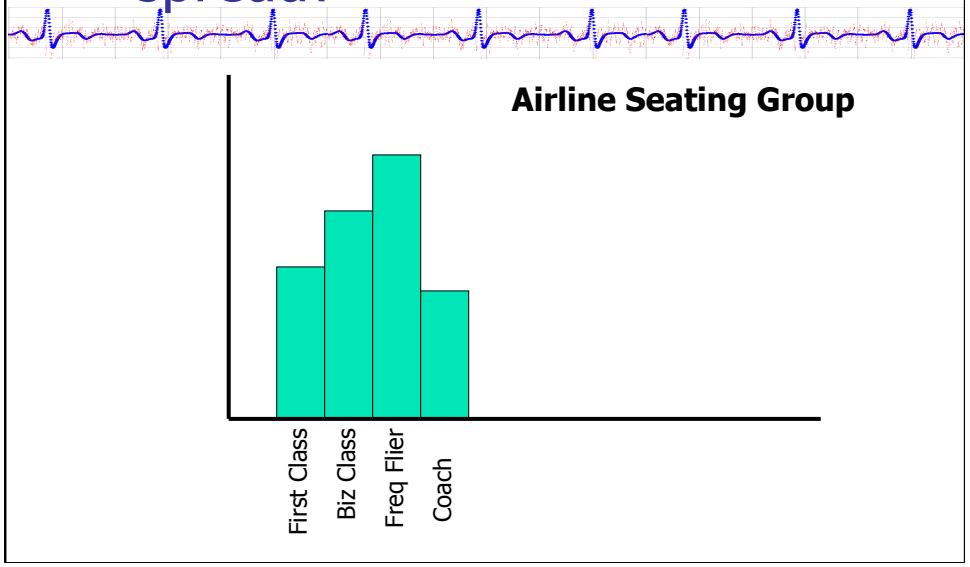
30

Which measures of center and spread?

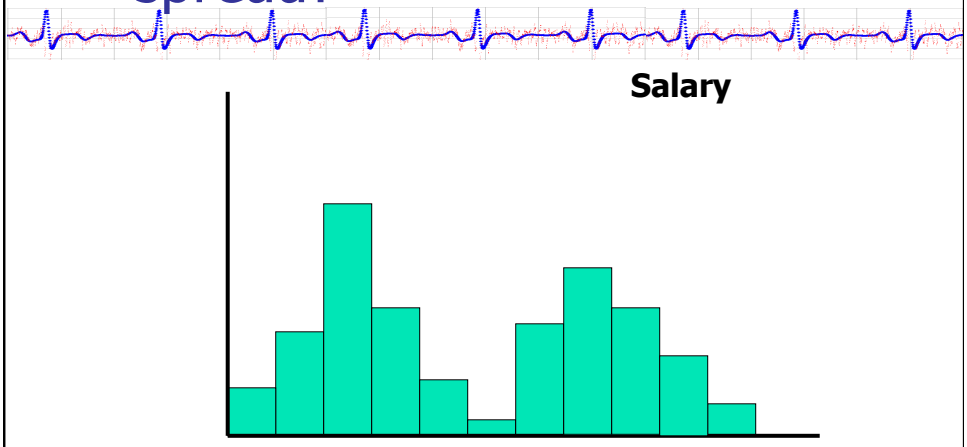


31

Which measures of center and spread?

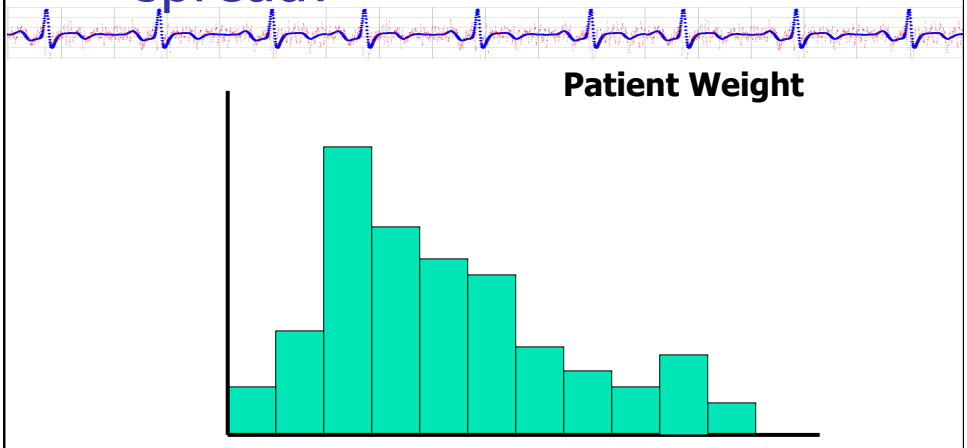


Which measures of center and spread?



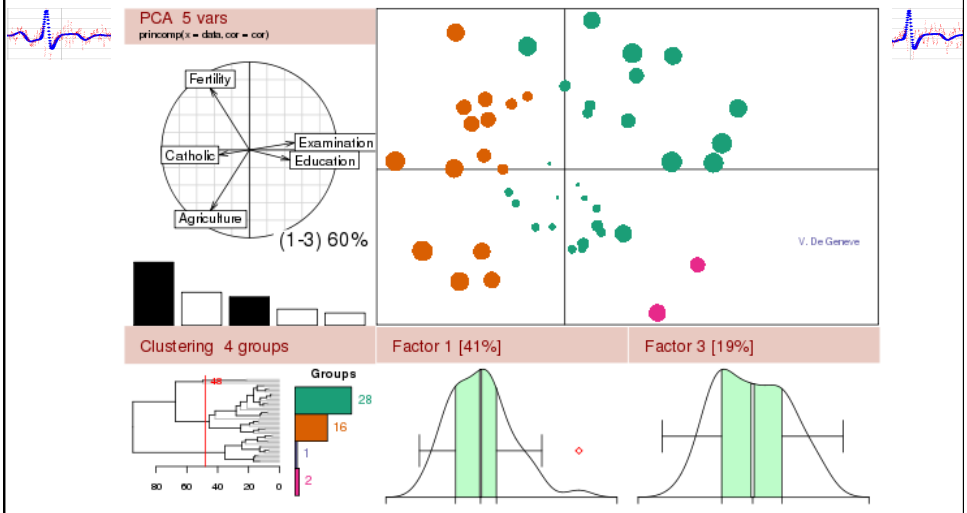
33

Which measures of center and spread?



34

The R Project for Statistical Computing



Descriptive Statistics in R

- Basic descriptives for numeric values

```
summary(eruptions)
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
1.600   2.163   4.000   3.488   4.454   5.100
```

Descriptive Statistics in R for numeric values

- `mean(data)`
- `median(data)`

- `var(data)`
- `sd(data)`
- `IQR(data)  //inter-quartile range`

Frequency Tables for Mode or some Histograms

- `table(vectordata) //1D array of freq counts`

- `prop.table(table(data))`
  `//relative frequencies`

- `round(100*prop.table(table(data)),1)`
  `//by pct`

Descriptive Statistics in R

- Basic descriptives

```
summary(eruptions)
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
1.600   2.163   4.000   3.488   4.454   5.100
```

- Histogram (for numeric values)

```
hist(eruptions)
```

- Histogram (for non-numeric values)

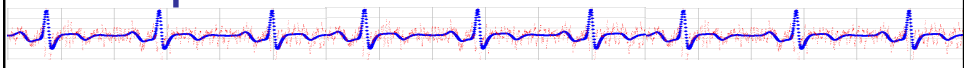
```
barplot(table(favoritecolor))
```

Descriptive Statistics Reports

- Every report for every homework or project must have a "Descriptive Statistics" section
- For every measure (subsection)
  - Explicitly state the level of measurement (nominal, etc.)
  - Show the histogram
  - For interval, ratio explicitly state your conclusion regarding normality
  - Present descriptive statistics
    - 1 measure of center
    - 0 or 1 measure of spread
  - Do not just copy R output
- Include all raw data in appendix

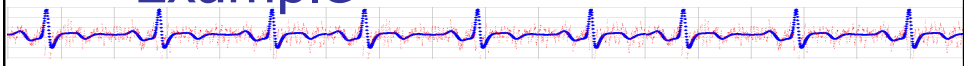
40

Example descriptive stats report



41

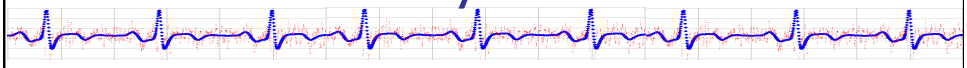
Example



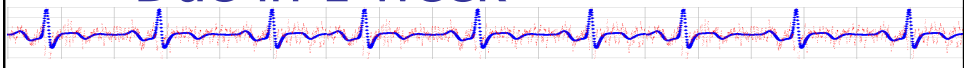
- You collect some data on time to find a table in Curry… what do you do next?

42

Usability Evaluation



Homework I3
Due in 1 week



- Conduct a small usability study
  - Descriptive
  - Quantitative
  - At least Two tasks
  - At least Two measures
  - At least Three subjects

Homework I3
Due in 1 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. 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.

45

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

46

What is usability?

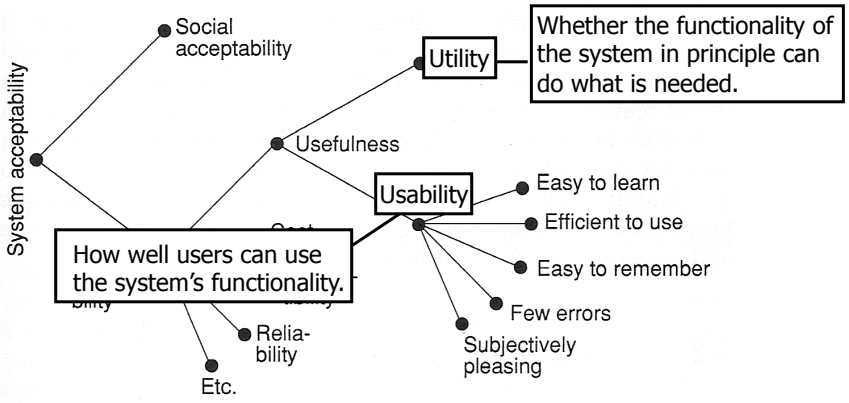


Figure 1 *A model of the attributes of system acceptability.*

From Nielsen, Usability Engineering

47

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."

48

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

49

Usability Measures Examples of each?

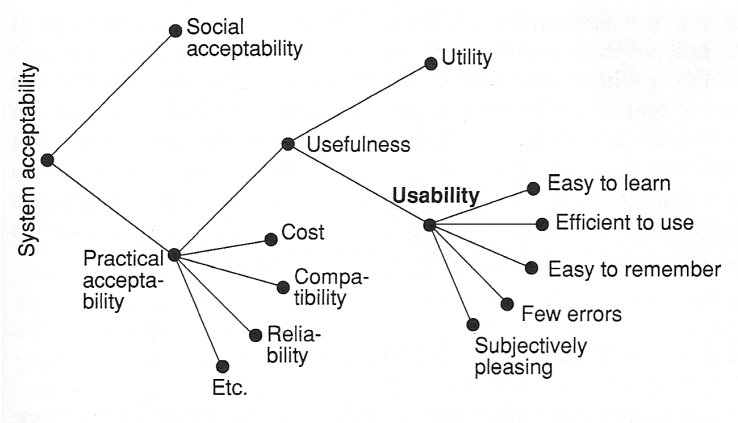


Figure 1 *A model of the attributes of system acceptability.*

50

Performance Measures

- Time to complete task
- %tasks successfully completed
- Number of errors
- Time spent recovering from errors
- Number of commands used (steps or types)
- %users who say they prefer the system over competitor
- etc. (pg 194 Nielsen has more)

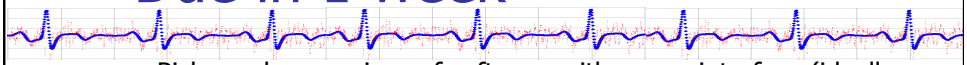
51

Papers

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

52

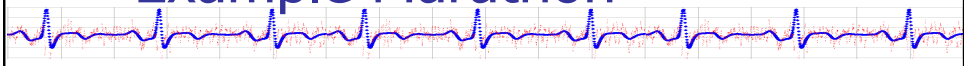
Homework I3
Due in 1 week



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. 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.

53

Example Marathon



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

54

Example

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

55

Example

- Fast food restaurant
- Customer ordering time
- Customer satisfaction

56

Example

- Student Diet
- Student BMI

Example

- Sign displaying wait times per restaurant vs. no sign
- Student happiness (single 7-pt scale item)

58