Human-Computer Interaction IS4300

I6 – Swing Layout Managers *due now*

- **You have two choices for requirements**:

1) try to duplicate the functionality of an existing applet; or,

2) create your own (ideally project-related) applet with the following minimum requirements:

  - A JFrame and a modal JDialog.
  - A JTabbedPane and JScrollPane.
  - Nested JPanels including the following layout managers: GridLayout, FlowLayout, BorderLayout
  - Some interaction widgets (JButton, etc.) on every JPanel and tab.
  - Reasonable behavior when the JFrame is resized.

- You may not use GridBagLayout or absolute layout anywhere in the project.

2

T5b – Paper Prototyping
due next class

- Recruit 3-5 users who are as close as possible to your target demographic.
- Be sure to record demographic information (age, gender, education, occupation, etc.) for your report.
- **Testing Users** When you run your prototype on a user, you should do the following things:
  - Obtain verbal consent for participation.
  - Brief the user.
  - Present one task.
  - Watch the user do the task. Take notes of your observations.
  - Repeat with the other tasks.
  - Interview users, take any measures you think are important.

3

T5b – Paper Prototyping
due next class

- **What to Post**
- Prototype photos. Digital photos of the pieces of your prototype. Try to show the prototype in an interesting state, not just a blank window.
- Briefing. The briefing you gave to users.
- Scenario tasks. The tasks you gave to users, exactly as you wrote them on the cards.
- Demographics of your test users, and description of the test scenario (time, place, equipment, etc).
- Observations. Usability problems you discovered from the testing, and possible solutions. Describe what users did. You must test at least 3 users.
- Results from interviews & other measures.

4

Analysis and interpretation of user observation evaluation data

Sample Usability Report Template

Usability.gov

Short - Informal

7

Usability.gov
Your guide for developing usable and useful Web sites

Short Usability Test Report for [Site]

Date of Report:	[Month Day, Year]
Date of Test:	[Month Day, Year]
Location of Test:	[City, State]
Prepared for:	[Name]
Phone Number:	[XXX-XXX-XXXX]
Email:	[name@address.gov]
Prepared by:	[Name]
Phone Number:	[XXX-XXX-XXXX]
Email:	[name@address.gov]

Executive summary

NOTE: This section describes the main goal and rationale of the study. Briefly describe the scenarios that participants completed, how the sessions were conducted, and how many participants took part in the study. This section should also discuss overall trends, such as whether or not participants were able to complete all the tasks. Data should be reported as both a number of completed scenarios as well as a percentage. Is there is a reason why tasks were completed or not? Be sure to give an overall impression (theme) about what the reader will encounter in the report.

Methodology

Who we tested

[Eight] participants, having the following characteristics, evaluated [product name].

NOTE: Add or delete main categories as needed. Refer to screener for main demographic information.

Audience Type

User Profile 1	2
User Profile 2	4
User Profile 3	2
TOTAL (participants)	**8**

Computer Usage

0 to 10 hrs. wk.	2
11 to 25 hrs. wk.	4
26+ hrs. wk.	2
TOTAL (participants)	**8**

Age

18-25	2
26-39	2
40-59	2
60-74	2
TOTAL (participants)	**8**

Gender

Women	4
Men	4
TOTAL (participants)	**8**

9

What participants did

NOTE: State about how long participants met with the study facilitator and how many tasks they completed. Describe anything else that participants did, such as filling out questionnaires.

What data we collected

NOTE: Describe the data collected, including paths selected, task completion rates, and verbal feedback. Describe any other data that were collected, such as time on task or satisfaction ratings.

Major findings and recommendations

- **List major issues** – Use the bullet format to enable quick scanning.
- **Identify solutions** – Spot issues and trends via user testing and then make recommendations.

NOTE: Readers should be able to use this section to get a good grasp of what the issues are and what possible solutions exist. It is not a list of every single problem, but an overview of the major stumbling blocks identified during testing.

10

Detailed findings and recommendations

Scenario 1 - Exercise/Sports; Fitness

You've heard that many people get injured due to bike accidents. You want to know if wearing a bike helmet will reduce your chance of injury. Where would you look for this information?

Number of participants	X
Percent successful	XX%

Sample Findings	Sample Recommendations
X participants completed the task with ease (score of "2") by finding the [target]. X participants needed prompting or had significant difficult completing the task (score of "1") X participants did not complete the task (score of "0"). X participants found the [target] via [path].	State each separate recommendation in the form of an action item. List as many as needed to resolve the issue, keeping in mind how the recommendations affect the site or application as a whole.

- ***You should also describe the problems (what did you observe?) How many users had this problem?***
- ***Severity: critical, important, cosmetic***
- ***Great to augment these with images or screenshots highlighting the problems***

11

Exit Questions/User Impressions

NOTE: It may be useful to put responses in a table, especially if you want to list all participant responses rather than an overall summary. Following are two examples.

Sample 1. Summary of user impressions

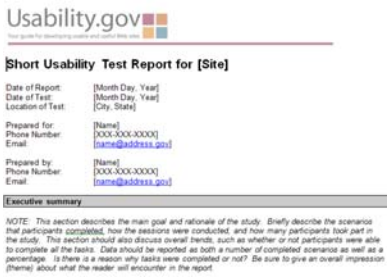
Sample Questions	Sample Responses
What is your overall impression to [site]?	Participants liked being able to find health info.
What is your impression of the search capability?	Search worked well, efficiently. Seemed to have the most relevant information at the top.
What did you like best about the site?	1) Resources 2) Articles 3) Health tips
What did you like least about the site?	1) No way to share info 2) Everything in PDF 3) Not enough pictures (images, charts, graphs)
Is there anything that you feel is missing on this site?	1) "Email a friend" 2) FAQs
If you were to describe this site to a colleague in a sentence or two, what would you say?	It's a useful site with a lot of good information.
Do you have any other final comments or questions?	Needs more pictures and color.

Sample 2. User impressions by participant

Participant No.	Like best?	Like least?	Improvements
1	Search	Registration	Add more color
2	A-Z index	Pop-ups	Clickable charts

12

Problem



Usability.gov

Short Usability Test Report for [Site]

Date of Report: [Month Day, Year]
Date of Test: [Month Day, Year]
Location of Test: [City, State]

Prepared for: [Name]
Phone Number: [XXX-XXX-XXXX]
Email: [name@address.gov]

Prepared by: [Name]
Phone Number: [XXX-XXX-XXXX]
Email: [name@address.gov]

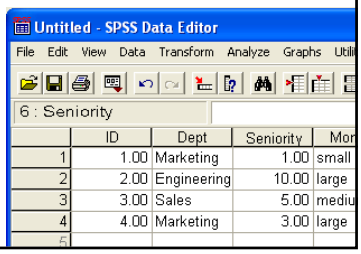
Executive summary

NOTE: This section describes the main goal and rationale of the study. Briefly describe the scenarios that participants completed, how the sessions were conducted, and how many participants took part in the study. This section should also discuss overall trends, such as whether or not participants were able to complete all the tasks. Data should be reported as both a number of completed scenarios as well as a percentage. Is there is a reason why tasks were completed or not? Be sure to give an overall impression (theme) about what the reader will encounter in the report.

Example: Paper Prototyping usability test.

Coding data

- Transcribe all interviews
- Code (data entry) all measures
  - Questionnaires
  - Metrics (times, errors)
  - Check for errors, missing data
  - Unstacked format typical
  - Excel ok for very simple analyses, recommend SPSS for more complex (and R if you are a hacker)
- Do asap, by people in the room



6 : Seniority

	ID	Dept	Seniority	Mor
1	1.00	Marketing	1.00	small
2	2.00	Engineering	10.00	large
3	3.00	Sales	5.00	mediu
4	4.00	Marketing	3.00	large
5				

Summarizing Data

- Qualitative
  - Problem analysis
  - Text analysis
- Quantitative
  - Descriptive statistics

15

Problem ("usability defect") analysis Example: Optometrist website

- U1: Could not find SEARCH function. Failed to complete.
- U2: Spent long time finding contents of cart. Completed.
- U3: Spent long time finding SEARCH function. Completed.
- U4: No problems.
- U5: Could not find SEARCH function. Failed to complete.
- U6: Did not like colors on checkout page.

16

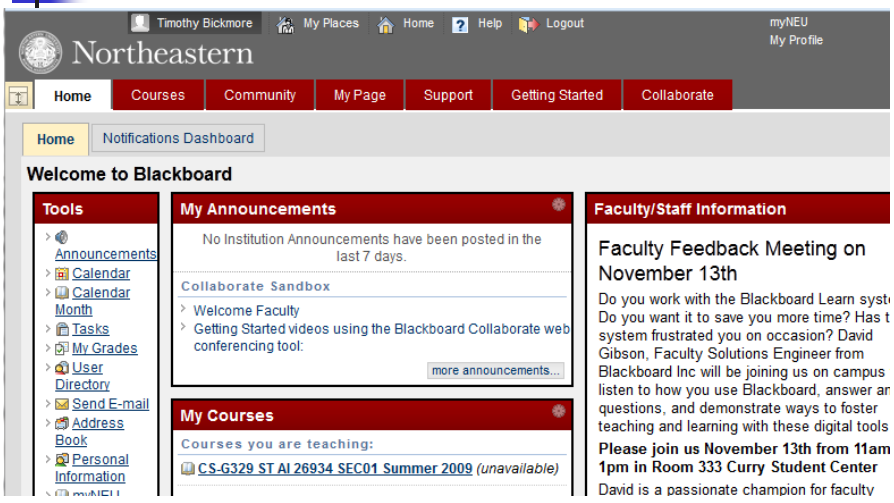
Summarizing Qualitative Data

Use with Interview & Think Aloud data

More on 11/1 with Barbara

17

Example Analytic Induction



Analytic Induction (Znaniecki)

Nonexperimental, Qualitative analogue to scientific method

1. Phenomenon tentatively defined
2. Hypothesis is developed
3. A single instance is considered to determine if hypothesis is confirmed
4. If hypothesis fails, then phenomenon or hypothesis is redefined
5. Additional cases are examined and, if the new hypothesis is repeatedly confirmed, some degree of certainty results
6. Each negative case requires that the hypothesis be reformulated until there are no exceptions

19

Summarizing Quantitative Data

Kinds of Measures

Primary source:

Bordens & Abbott, *Research Design and Methods*

21

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)

22

Scales of Measurement

- *Interval Scale*
  - Scale of measurement on which the spacing between values is known (e.g., rating a book on a scale ranging from 0 to 10)
  - No true zero point
- *Ratio Scale*
  - Similar to interval scale, but with a true zero point (e.g., number of lever presses, height)

23

What kind is it?

- Age
- Gender
- Job Category (Engineer, Manager…)
- Efficiency (time to complete)
- School Year (Freshman…)
- Temperature (Celsius)
- Think aloud quotes / themes
- Monitor Size
- Weather (Rain, Snow, …)
- Debrief quotes / themes
- Productivity (wpd)
- Owns Pet (or not)

24

Descriptive Statistics Practically speaking

- You will decide on statistical methods depending on whether your measures are
  - Nominal, or
  - Numeric (Interval, Ratio)
- Ordinal values can sometimes be treated as numeric, sometimes as non-numeric.
- Text (structured qualitative data) should be subjected to other analyses first.

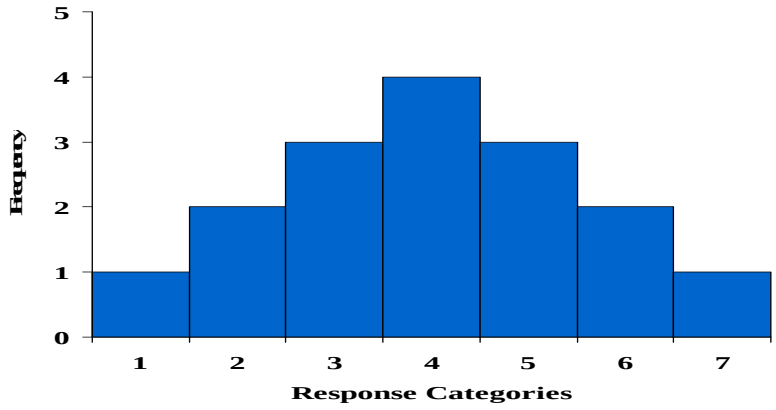
26

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 on 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

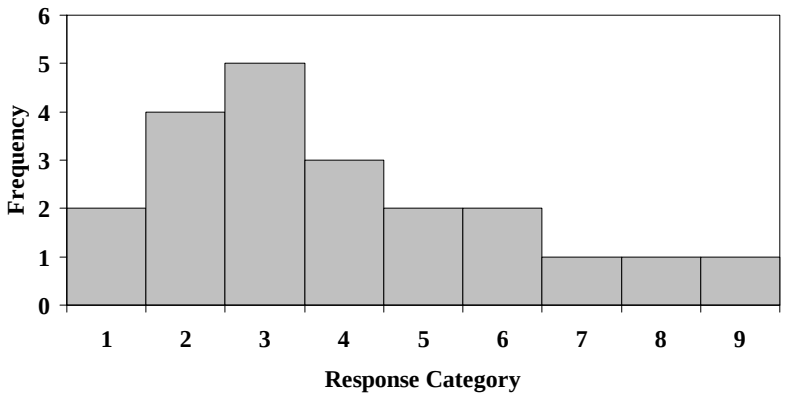
27

Histogram Showing a Normal Distribution



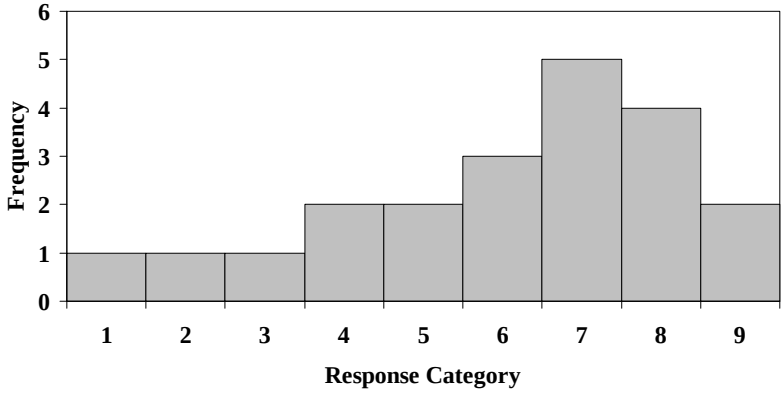
28

Histogram Showing a Positive Skew



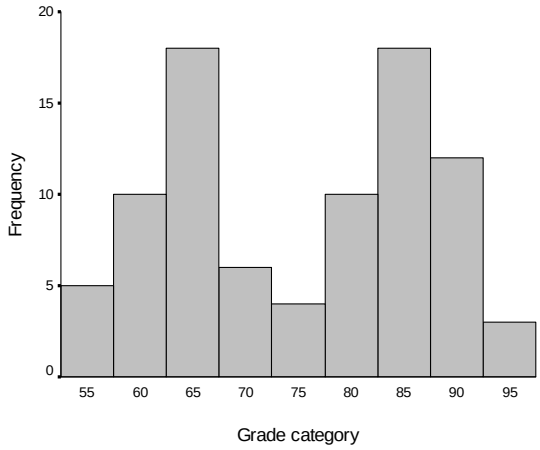
29

Histogram Showing a Negative Skew



30

A Bimodal Distribution



31

Measures of Center

- Mean
- Median
- Mode

32

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

33

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 data do not meet requirements for using the mean (skewed but unimodal), or if significant outliers

35

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

36

Measures of Spread

- Std Deviation
- Inter-quartile range
- Range

37

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

38

Measures of Spread: Characteristics

- *Variance*
  - Average squared deviation of scores from the mean
- *Standard Deviation*
  - Square root of the variance
  - Most widely used measure of spread

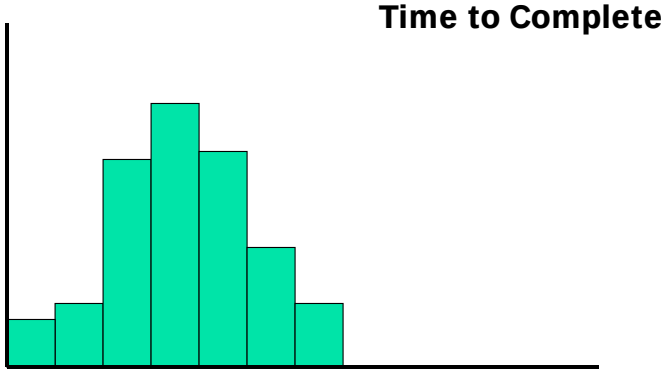
39

Measures of Spread: Applications

- The range and standard deviation are sensitive to extreme scores
  - In such cases the interquartile range is best
- When your distribution of scores is skewed, the standard deviation does not provide a good index of spread
  - use the interquartile range

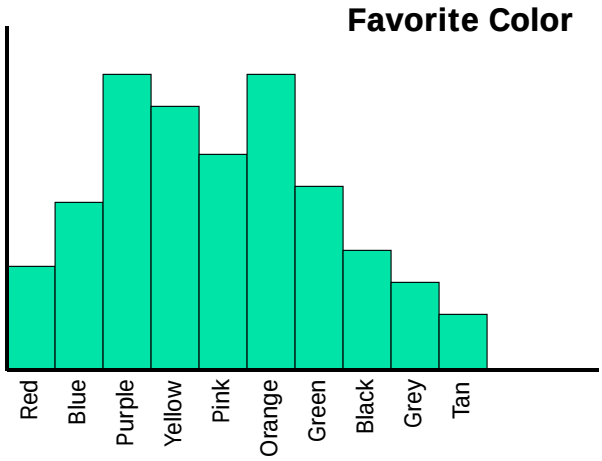
40

Which measures of center and spread?



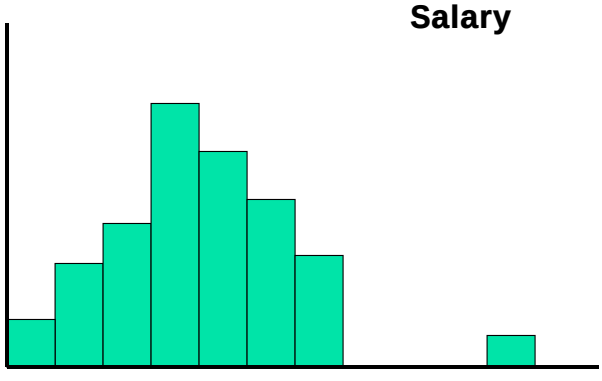
41

Which measures of center and spread?



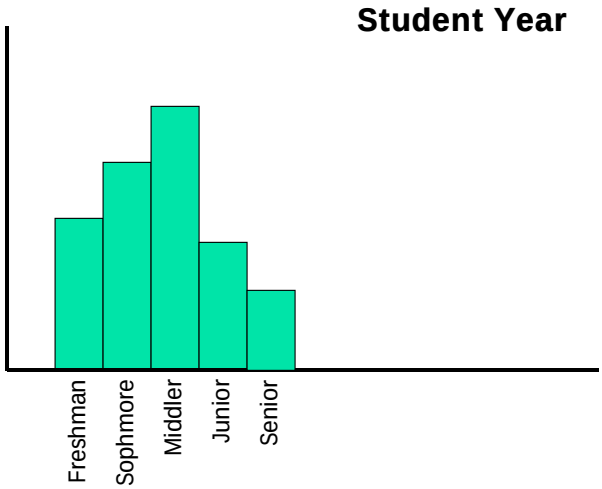
42

Which measures of center and spread?



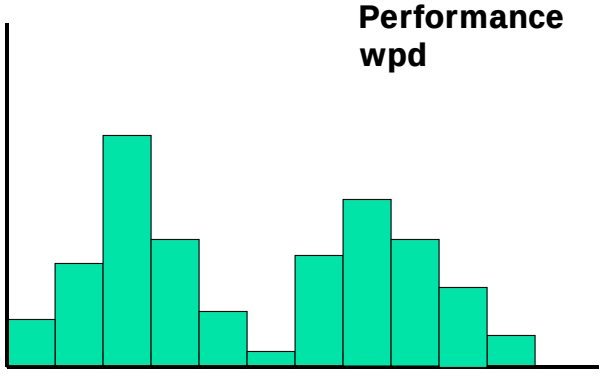
43

Which measures of center and spread?



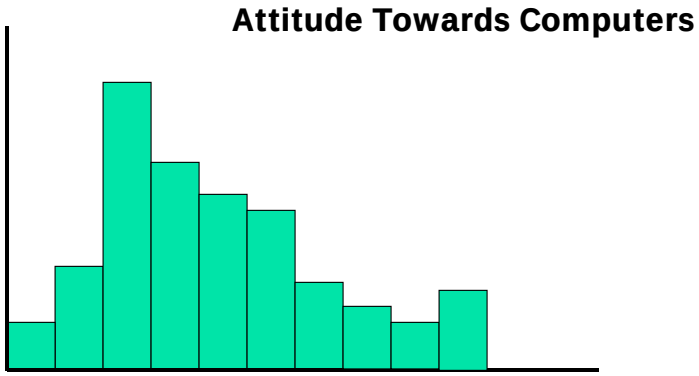
44

Which measures of center and spread?



45

Which measures of center and spread?



46

Recent controversy over analysis of scale measures

- Historically, have been treated as interval if they appear normal (i.e., with mean, stdev, and t-test)
- Some statisticians say NEVER. They are ordinal measures – must use median, no meaningful range measures, and non-parametric inferential statistics (e.g., Mann-Whitney)
- See
  - "Stats: We're Doing It Wrong" on ACM.ORG

47

Inferential Analyses for Quantitative "Metric" studies

Users performed the set of standardized tasks in a significantly shorter time using interface FOO compared to interface BAR, $t(27)=3.4$, $p<.05$

48

Samples & Populations

- Population = everyone you care about
  - E.g., all of your primary stakeholders, all of your customers, all gamers in the US, etc
- Sample = everyone in your study

- Usually $|Sample| << |Population|$
- Inferential statistics let us make claims about the Population based on data from one or more Samples.
- If you could experiment on everyone in the population you would not need inferential statistics.

49

Typical case

- You are trying to demonstrate there is a difference between two metrics
  - E.g., performance with interface FOO vs. performance with interface BAR

50

Type of Errors in Inferential Statistics

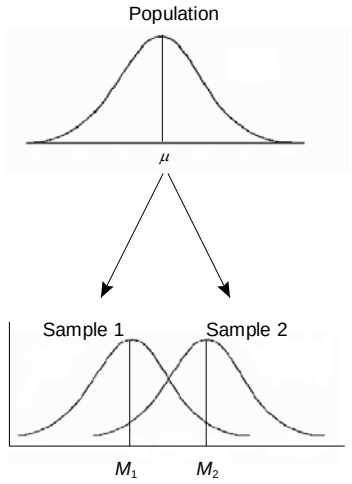
Research Hypothesis: There is a difference
(e.g., FOO better than BAR)

	"The Truth"	
	No diff	**Diff**
Conclude diff	**Type I Error**	**Correct Decision**
Conclude no diff	**Correct Decision**	**Type II Error**

'p' = Probability of Type I Error
The likelihood the difference observed is not real.

51

Relationship Between Population and Samples When a Treatment Had No Effect



'p' = Likelihood of this happening.

52

Inferential Analyses

- Correlational
- Demonstrative
- Experimental
  - Between-subjects
    - Single factor, two-level
    - Single factor, N-level (for N>2)
    - Two factor, N-level (for N>=2)
  - Within-subjects
    - Single factor, two-level

53

t-test for independent means

- Two samples, interval or ratio
- No other information about comparison distribution
- Assumptions:
  - Sample randomly selected from population.
  - The sampling distribution of means is normal
  - Variances of the two populations (whether they are the same or different) are the same.

54

Excel T.TEST, returns 'p'

Syntax

```
T.TEST(array1,array2,tails,type)
```

The T.TEST function syntax has the following arguments:

- **Array1** Required. The first data set.
- **Array2** Required. The second data set.
- **Tails** Required. Specifies the number of distribution tails. If tails = 1, T.TEST uses the one-tailed distribution. If tails = 2, T.TEST uses the two-tailed distribution.
- **Type** Required. The kind of t-Test to perform.

Parameters

If type equals	This test is performed
1	Paired
2	Two-sample equal variance (homoscedastic)
3	Two-sample unequal variance (heteroscedastic)

55

t-test

- If assumptions are followed, T.TEST returns ‘p’
  - Likelihood of differences observed being due to chance, or error
  - Probability of Type I error
- If p<threshold (conventionally 0.05), we say there is a significant difference
- If p>=threshold, what can we conclude?

56

Reporting results

- Significant results, scientific articles
  t(df)=*tscore*, p<*sig*
  e.g., t(38)=4.72, p<.05
- Non-significant results
  e.g., t(38)=4.72, n.s.
- Informal usability reports:
  - t-test for independent means indicated that performance with FOO was significantly better than performance with BAR, p<.05
  - t-test for independent means for performance with FOO vs. BAR failed to yield significant results.

57

Example t-test

User	UI	Time
1	FOO	5.1
2	FOO	3.5
3	FOO	4.2
4	FOO	1.7
5	FOO	4.9
6	FOO	6.4
7	BAR	2.1
8	BAR	4.1
9	BAR	1.1
10	BAR	2.8
11	BAR	3.2
12	BAR	1.4

59

Another note about "Power"

- For small, informal, qualitative, debugging usability tests
  - 5 users gets 80% of "usability defects"
- For quantitative usability experiments
  - Should do a "Power Analysis"
    - See online "Power Analysis Calculator"
    - Parameters: anticipated α, β (or power=1- β), effect size, tails
    - A "pilot study" with 12 users/condition will let you estimate effect size

60

Type of Errors in Inferential Statistics

Research Hypothesis: There is a difference
(e.g., FOO better than BAR)

	"The Truth"	
	No diff	**Diff**
Conclude diff	**Type I Error**	**Correct Decision**
Conclude no diff	**Correct Decision**	**Type II Error**

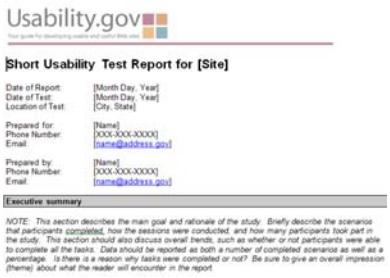
61

Final note about inferential statistics

- You likely won't need them for small usability studies
- We'll cover in more detail on 10/25 and 11/5

62

Problem


Usability.gov

Short Usability Test Report for [Site]

Date of Report: [Month Day, Year]
Date of Test: [Month Day, Year]
Location of Test: [City, State]

Prepared for: [Name]
Phone Number: [XXX-XXX-XXXX]
Email: [name@address.gov]

Prepared by: [Name]
Phone Number: [XXX-XXX-XXXX]
Email: [name@address.gov]

Executive summary

NOTE: This section describes the main goal and rationale of the study. Briefly describe the scenarios that participants completed, how the sessions were conducted, and how many participants took part in the study. This section should also discuss overall trends, such as whether or not participants were able to complete all the tasks. Data should be reported as both a number of completed scenarios as well as a percentage. Is there a reason why tasks were completed or not? Be sure to give an overall impression (theme) about what the reader will encounter in the report.

Example: Paper Prototyping usability test.

67

To Do

- Read
  - Advanced Evaluation (Stone Ch 27).
  - Case study (Gould).
- Finish by next class
  - T5b – Conduct and writeup User Tests of paper prototype.

68