# Empirical Research Methods in Information Science

IS 4800
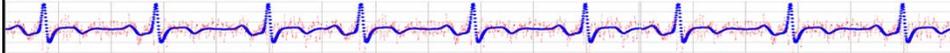
Lecture 9

Survey Design
Composite Measure Design
Survey Administration
Sampling
Preview of Hypothesis Testing

1

Quiz
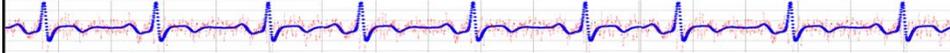
- 10 minutes
- Closed book, closed notes

2

1

Homework Review

- Usability/Performance Measures
- Behavioral Measures

Chapter 8

Using Survey Research
Part I – Questionnaire Design

Terminology Soup

- Questionnaire = Self-Report Measure = Instrument
- Field Survey vs. Lab Instrument/Questionnaire
- Composite Measure ~ Index ~ Scale
- Item

5

Overview – Design of Self-Report Questionnaires

- Constructing Questionnaires
- Types of Questionnaire Items
- Composite measures
- Example
- Sampling & Administration

6

Overview of Questionnaire Construction
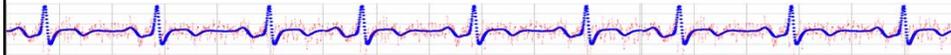
Note: Most of the heuristics on questionnaire design in the text are most appropriate for field surveys.

7

Parts of a Questionnaire
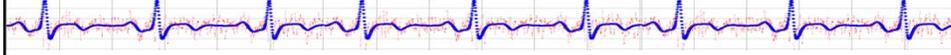
- In any study you normally want to collect demographics – usually done through questionnaire
- Single items
- Composite items

8

4

Sample Questionnaire

Participant ID_____ Date _____

Single item
Single item

Composite measure

Single item

Demographics

9

Questionnaire Construction

- Items can be optional. Flow often depicted verbally and/or pictorially.

14. Have you ever participated in the Model Cities program?

[ ] Yes

[ ] No

If Yes: When did you last attend attend a meeting?

_____

10

Questionnaire Construction
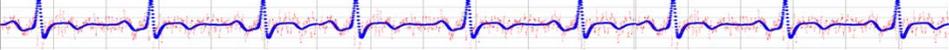
- Many heuristics for ordering questions, length of surveys, etc. For example:
  - Put interesting questions first
  - Demonstrate relevance to what you've told participants
  - Group questions in to coherent groups

Questionnaire Construction
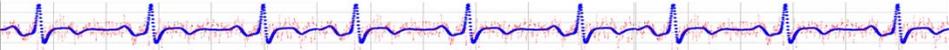
- Additional heuristics
  - Organize questions into a coherent, visually pleasing format
  - Do not present demographic items first
  - Place sensitive or objectionable items after less sensitive/objectionable items
  - Establish a logical *navigational path*

Types of Questionnaire Items

- *Open-Ended*
  - Respondents are asked to answer a question in their own words
- *Restricted* (close-ended)
  - Respondents are given a list of alternatives and check the desired alternative
- *Partially Open-Ended*
  - An "Other" alternative is added to a restricted item, allowing the respondent to write in an alternative

13

Types of Questionnaire Items

- *Rating Scale*
  - Respondents circle a number on a scale (e.g., 0 to 10) or check a point on a line that best reflects their opinions
  - Two factors need to be considered
    - Number of points on the scale
    - How to label ("anchor") the scale (e.g., endpoints only or each point)

14

Types of Questionnaire Items

- A *Likert Scale* is a scale used to assess attitudes
  - Respondents indicate the degree of agreement or disagreement to a series of statements
  - I am happy.
    Disagree 1 2 3 4 5 6 7 Agree

- A *Semantic Differential Scale* allows participate to provide a rating within a bipolar space
  - How are you feeling right now?
    Sad 1 2 3 4 5 6 7 Happy

15

Visual Analog Scale

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

No pain | Mild, annoying pain | Nagging, uncomfortable, troublesome pain | Distressing, miserable pain | Intense, dreadful, horrible pain | Worst possible, unbearable, excrutiating pain

16

8
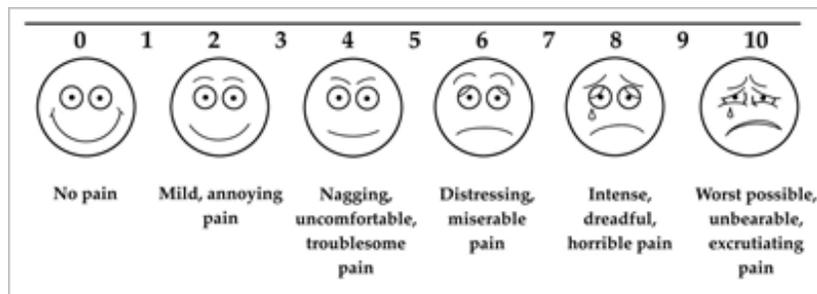
Writing Good Items
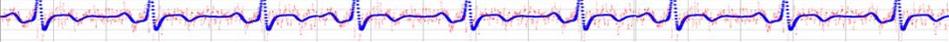
- Use simple words
- Avoid vague questions
- Don't ask for too much information in one question
- Avoid "check all that apply" items
- Avoid questions that ask for more than one thing
- Soften impact of sensitive questions
- Avoid negative statements (usually)

17

Two Most Important Rules in Designing Questionnaires?
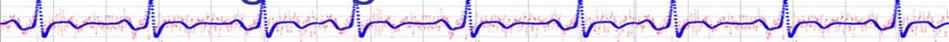
1. Use an existing validated questionnaire if you can find one.
2. If you must develop your own questionnaire, **pilot test** it!

18

Composite Measures
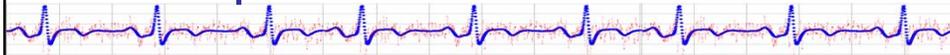
Example 'Composite Scale Questionnaire' UCLA Loneliness Scale (excerpt)

1. I feel in tune with the people around me.
NEVER   RARELY   SOMETIMES   ALWAYS

2. I lack companionship.
NEVER   RARELY   SOMETIMES   ALWAYS

3. There is no one I can turn to.
NEVER   RARELY   SOMETIMES   ALWAYS

4. I do not feel alone.
NEVER   RARELY   SOMETIMES   ALWAYS

5. I feel part of a group of friends.
NEVER   RARELY   SOMETIMES   ALWAYS

Example Composite Measure
Working Alliance Inventory (5 of 36 Qs)

I feel uncomfortable with the advisor.
disagree   agree
completely   •   •   •   •   •   •   •   completely

The advisor and I understand each other.
disagree   agree
completely   •   •   •   •   •   •   •   completely

I believe the advisor likes me.
disagree   agree
completely   •   •   •   •   •   •   •   completely

I believe the advisor is genuinely concerned about my welfare.
disagree   agree
completely   •   •   •   •   •   •   •   completely

The advisor and I respect each other.
disagree   agree
completely   •   •   •   •   •   •   •   completely

21

'Scoring' a Composite Measure

- Generally:
  - Negate negative items
    - Score' = (max score + 1) − Score
  - Sum scores
- Can normalize by averaging
- Weight items equally unless you have a compelling reason to do otherwise
- Missing data:
  - "impute the average" by excluding unanswered items from the average

22

Composite measures:
Why ask the same question 10 ways?

- It is seldom possible to arrive at a single question that adequately represents a complex variable.
  - Any single item is likely to misrepresent some respondents (e.g., church-going)
- A single item may not provide enough variation for your purposes.
- Single items give crude assessments; several items may give a more comprehensive and accurate assessment.

23

# Terminology: Factors, Subscales & Constructs

- Construct
  - a psychological entity that you are interested in measuring (e.g., loneliness, working alliance)
- Factor
  - A construct may have more than one parts or dimensions or aspects, referred to as "factors" that may be independently assessed by your questionnaire.
- Subscale
  - A part of your questionnaire that assesses one factor.
  - Usually: score subscales separately, in addition to aggregate

- Factors can be informed by theory, or emerge from data analysis ("exploratory factor analysis" 4/9)

24

1. (B) I feel uncomfortable with George
2. (T) George and I agree about the things I will need to do to help improve my level of physical activity.
3. (G) I am worried about the outcome of my sessions with George.
4. (T) What I am doing in my discussions with George gives me new ways of looking at physical activity.
5. (B) George and I understand each other.
6. (G) George perceives accurately what my goals are.
7. (B) I find what I am doing with George confusing.
8. (B) I believe George likes me.
9. (G) I wish George and I could clarify the purpose of our sessions.
10. (G) I disagree with George about what I ought to get out of my discussions with him.
11. (T) I believe the time George and I are spending together is not spent efficiently.
12. (G) George does not understand what I am trying to accomplish.
13. (T) I am clear on what my responsibilities are with respect to physical activity.
14. (G) My physical activity goals are important to me.
15. (G) I find what George and I are doing are unrelated to my concerns.
16. (T) I feel that the things I do with George will help me to accomplish the changes that I want.
17. (B) I believe George is genuinely concerned about my welfare.
18. (T) I am clear as to what George wants me to do in our discussions.

19. George and I respect each other.
20. I feel that George is not totally honest about his feelings toward me.
21. I am confident in George's ability to help me.
22. George and I are working towards mutually agreed upon goals.
23. I feel that George appreciates me.
24. We agree on what is important for me to work on.
25. As a result of my discussions with George I am clearer as to how I might be able to change.
26. George and I trust one another.
27. George and I have different ideas on what my problems are.
28. My relationship with George is very important to me.
29. I have the feeling that if I say or do the wrong things, George will stop working with me.
30. George and I collaborate on setting goals for us to work on.
31. I am frustrated by the things I do with George.
32. We have established a good understanding of the kind of changes that would be good for me.
33. The things that George is asking me to do don't make sense.
34. I don't know what to expect as the result of my discussions with George.
35. I believe the way we are working with my problem is correct.
36. I feel George cares about me even when I do things that he does not approve of.

Designing a Composite Measure

Literature Review
Previous measures, theoretical concepts

↓

Brainstorm on Factors ←

↓

Brainstorm on Items ←

↓

Preliminary /Validity Reliability testing

↓

Factor analysis

↓

Reliability testing

↓

Validity testing
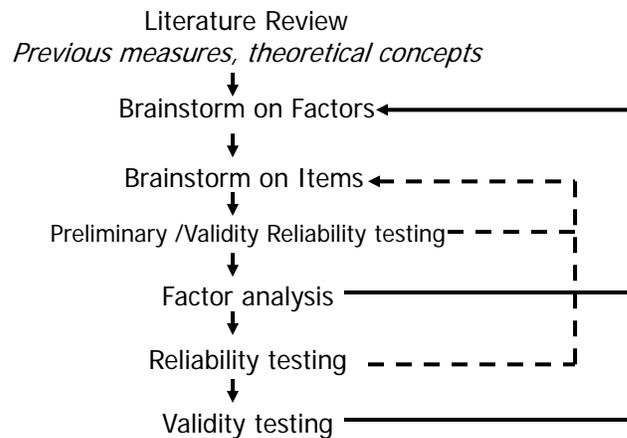
27

Composite Measures

- Indexes (aka "scales") provide an ordinal ranking of respondents with respect to a concept of interest (e.g., liking of computers)
- Usually assessed through a series of related questions.
- Psychological concepts
  - Most have no real meaning, no ultimate definition
  - ad hoc summaries of experience and observations

28

Psychological Concepts aka "Constructs"

- Concepts are general codifications of experience and observations.
    - Observe differences in social standing -> concept of social status.
    - Observe differences in religious commitment -> concept of religiosity
- Most psychological concepts have no real meaning, no ultimate definitions
- Concepts are ad hoc summaries of experience and observations

29

Operationalization

- The process of specifying empirical observations that are indicators of the concept of interest
- Begin by enumerating all the subdimensions ("factors") of the concept
    - Review previous research
    - Use commonsense

30

Example: religiosity

- Subdimensions/indicators/factors
    - Ritual involvement
        - E.g., going to church
    - Ideological involvement
        - Acceptance of religious beliefs
    - Intellectual involvement
        - Extent of knowledge about religion
    - Experiential involvement
        - Range of religious experiences
    - Consequential involvement
        - Extent to which religion guides social decisions
- (there are many others)

31

Example

- "NU Husky Fanatic"

1. What are some factors?
2. What are some items per factor?

32

16

Discriminant indicators

- Also think about related measures which should <u>not</u> be indicators of your construct
- In particular if you will be measuring another related variable, make sure none of your indicators include any attributes of it.
- Example
    - Want to study the relationship between religiosity and attitudes towards war => including a measure of adherence to "peace on earth" doctrine is not a good idea.

33

Picking items for a Composite

- Face validity
- Unidimensionality
    - All items measure same concept
- Should provide variance in responses
    - Don't pick items that classify <u>everyone</u> one way.
    - If you are interested in a binary classification (e.g., liberal vs. conservative), each item should split respondents roughly in half
- Negate up to half of the items to avoid response bias.

34

Picking items: bivariate analysis

- Every pair of items should be related, but not too strongly
  - Scoring high on item A should increase likelihood of scoring high on item B
  - But, if two items are perfectly correlated (e.g. one logically implies the other), then one can be dropped.

35

Scoring a Composite Measure

- Average the item scores
- Weight items equally unless you have a compelling reason to do otherwise
- Missing data
  - Omit dataset
  - Impute average/intermediate score
  - "Last value forward" for repeated measures
  - Many other strategies

36

Validating a Composite Measure
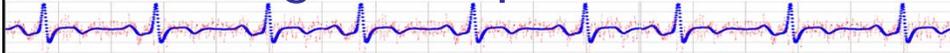
What is a validated measure?
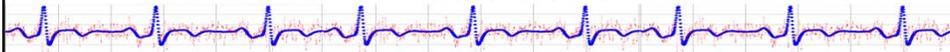
- Has reliability
- Has validity

- For psychological measures, these are collectively referred to as a measure's "psychometrics".

Measure Reliability

- A reliable measure produces similar results when repeated measurements are made under identical conditions
- Reliability can be established in several ways
    - *Test-retest reliability:* Administer the same test twice
    - *Parallel-forms reliability:* Alternate forms of the same test used
    - *Split-half reliability:* Parallel forms are included on one test and later separated for comparison

39

Reliability

- For surveys, this also encompasses *internal consistency:*
    - Do all of the questions address the same underlying construct of interest?
    - That is, do scores covary?
    - A standard measure is Cronbach's alpha
        - 0 = no correlation
        - 1 = scores always covary in the same way
        - 0.7 used as conventional threshold

40

Measure Validity

- A valid measure measures what you intend it to measure
- Most important when using psychological tests (e.g., IQ test)
- Validity can be established in a variety of ways
  - *Face validity:* Assessment of adequacy of content. Least powerful method
  - *Content validity:* How adequately does a test sample behavior it is intended to measure?
    - Does each item relate to the concept?
    - Do the items collectively cover the concept?

41

Measure Validity

- *Criterion-related validity:* How adequately does a test score match some criterion score? Takes two forms
  - Concurrent validity: Does test score correlate highly with score from a measure with known validity?
  - Predictive validity: Does test predict behavior known to be associated with the behavior being measured?
- *Construct validity:* Do the results of a test correlate with what is theoretically known about the construct being evaluated?
  - Convergent validity (subtype): measures of constructs that *should* be related to each other are
  - Discriminant validity (subtype): measures of constructs that *should not* be related are not

42

Example

- Assume we have good evidence for this model of the world..

  MonitorSize ⟶ **Productivity**

  Seniority
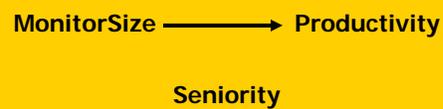
- We now propose a new measure for **Productivity**
  - What would be evidence for convergent validity?
  - What would be evidence for discriminant validity?

43

Validation - Summary

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

44

Overall Process to Develop a Composite Measure

- Identify factors
- Identify items
- Face, content validity for each item
- Check Response Variance for each item
- Bi-variate analysis
- Test reliability
- Test validity

45

Developing a New Measure

- Say you decide you need a new survey measure, "attitude towards large computer monitors" (ATLCM)
  - I like big monitors.
  - Big monitors make me nervous.
  - I prefer small monitors, even if they cost more.
  - *7-pt Likert scales*

- How would you validate this measure?

46

The R Project for Statistical Computing
Reliability of a Questionnaire
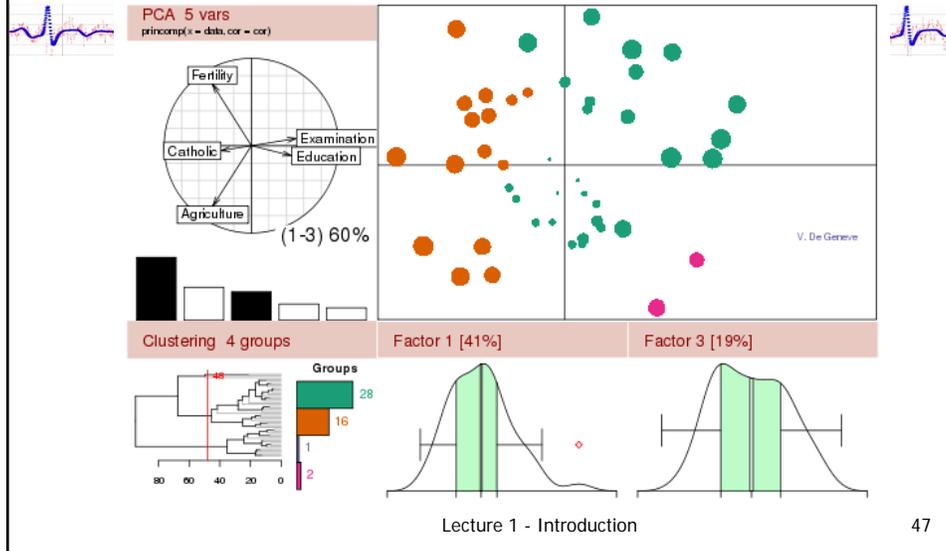
Cronbach's alpha

```
> install.packages("psych") #one time

> require(psych)  #every session
> alpha(data2)     #each column of frame = 1 item
Reliability analysis
Call: alpha(x = data2)
raw_alpha    std.alpha G6(smc) average_r mean sd
0.11         0.72 0.79 0.39 98 38

… #lots of diagnostic info follows…
```

24

Increasing the Reliability
of a Composite Questionnaire

- Increase the number of items on your questionnaire
- Standardize the conditions under which the test is administered (e.g., timing procedures, lighting, ventilation, instructions)
- Make sure you score your questionnaire carefully, eliminating scoring errors
- Check to be sure the items on your questionnaire are clearly written and appropriate for those who will complete your questionnaire

49

Exercise

- Break into Teams
- Design a Survey to assess...*X*

- Present your survey to the class – be prepared to discuss:
  - Underlying factors considered
  - Bivariate analysis
  - Justification for individual items & format
  - How you would assess reliability
  - How you would assess validity

50

Questionnaire Administration & Results Analysis
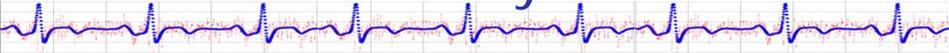
51

Administering Your Questionnaire
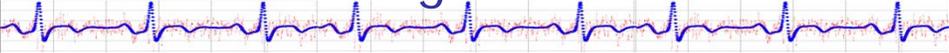
- *MAIL SURVEY*
  - A questionnaire is mailed directly to participants
  - Mail surveys are very convenient
  - Nonresponse bias is a serious problem resulting in an unrepresentative sample
- *INTERNET SURVEY*
  - Survey distributed via e-mail or on a Web site
  - Large samples can be acquired quickly
  - Biased samples are possible because of uneven computer ownership across demographic groups
  - *Check out surveymonkey.com*

52

Administering Your Questionnaire
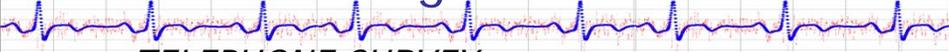
- *TELEPHONE SURVEY*
    - Participants are contacted by telephone and asked questions directly
    - Questions must be asked carefully
    - The plethora of "junk calls" may make participants suspicious
- *GROUP ADMINISTRATION*
    - A questionnaire is distributed to a group of participants at once (e.g., a class)
    - Completed by participants at the same time
    - Ensuring anonymity may be a problem

53

Administering Your Questionnaire
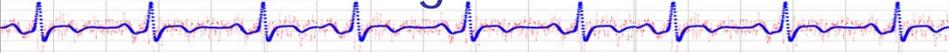
- *INTERVIEW*
    - Participants are asked questions in a face-to-face structured or unstructured format
    - Characteristics or behavior of the interviewer may affect the participants' responses

54

Mechanical Turk

55

What is a Turk ?

- Amazon mechanical turk is a Crowdsourcing tool developed by Amazon.
- Used to have people perform small tasks for micropayments.
- Developed by Peter Cohen at Amazon for it's internal use. [To find duplicates among webpages in 2005]
- Amazon recommends paying $6/hour. But, people have reported some pay as low as $1/hour
- "Sunday Times" reports Amazon Turk as "digital sweatshop"

Mechanical Turk

Turk: Demographics

Ross, et al, CHI 2010, Who are the Crowdworkers? Shifting Demographics in Mechanical Turk


58

29

Turk: Demographics

59

Turk: Demographics
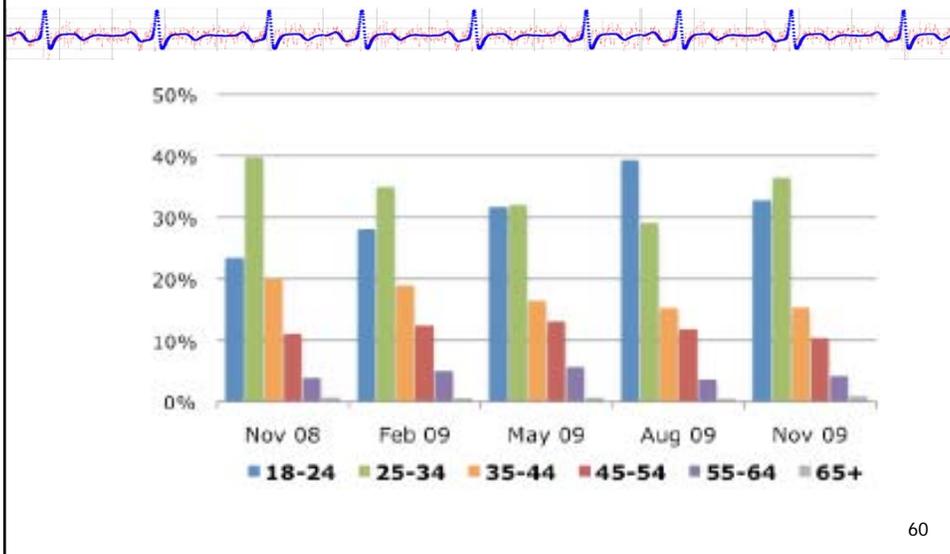
60

Administering Your Questionnaire
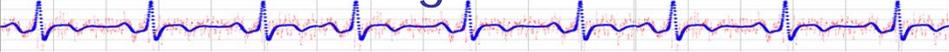
- In general
  - Personal techniques (interview, phone) provide higher response rates, but are more expensive and may suffer from bias problems.

61

Sampling
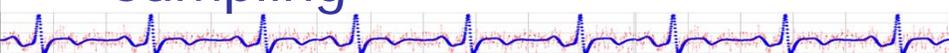
62

Sampling

- Sometimes you really can measure the entire population (e.g., workgroup, company), but this is rare...

- "Convenience sample"
  - Cases are selected only on the basis of feasibility or ease of data collection.

63

Acquiring A Survey Sample

- You should obtain a *representative sample*
  - The sample closely matches the characteristics of the population
- A *biased sample* occurs when your sample characteristics don't match population characteristics
  - Biased samples often produce misleading or inaccurate results
  - Usually stem from inadequate sampling procedures

64

Sampling Techniques
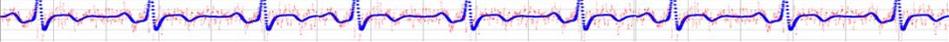
- *Simple Random Sampling*
  - Randomly select a sample from the population
  - *Random digit dialing* is a variant used with telephone surveys
  - Reduces systematic bias, but does not guarantee a representative sample
    - Some segments of the population may be over- or underrepresented

65

Sampling Techniques
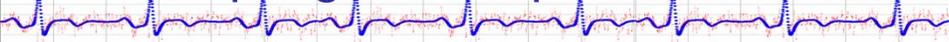
- *Systematic Sampling*
  - Every k^{th} element is sampled after a randomly selected starting point
    - Sample every fifth name in the telephone book after a random page and starting point selected, for example
  - Empirically equivalent to random sampling (usually)
    - May still result in a non-representative sample
  - Easier than random sampling

66

Sampling Techniques

- *Stratified Sampling*
  - Used to obtain a representative sample
  - Population is divided into (demographic) strata
    - Focus also on variables that are related to other variables of interest in your study (e.g., relationship between age and computer literacy)
  - **A random sample of a fixed size is drawn from each stratum**
  - May still lead to over- or underrepresentation of certain segments of the population
- *Proportionate Sampling*
  - Same as stratified sampling except that the proportions of different groups in the population are reflected in the samples from the strata

67

Sampling Example:

- You want to conduct a survey of job satisfaction of all employees but can only afford to contact 100 of them.
- Personnel breakdown:
  - 50% Engineering
  - 25% Sales & Marketing
  - 15% Admin
  - 10% Management
- Examples of
  - Stratified sampling?
  - Proportionate sampling?

68

Sampling Techniques

- *Cluster Sampling*
    - Used when populations are very large
    - The unit of sampling is a group (e.g., a class in a school) rather than individuals
    - Groups are randomly sampled from the population (e.g., ten classes from a particular school)

69

Sampling Techniques

- *Multistage Sampling*
    - Variant of cluster sampling
    - First, identify large clusters (e.g., school districts) and randomly sample from that population
    - Second, sample individuals from randomly selected clusters
    - Can be used along with stratified sampling to ensure a representative sample

    - Note: Multilevel, hierarchical statistical analysis can tease apart differences due to individual vs. cluster

70

Sampling

- Most statistics assume a random sample.

71

Sample size

- In all empirical research, you should motivate your *sample size*

- B&A Ch 9 provide formula for estimating sample size for binomial descriptive studies.
  - For binomial (two category) measures
  - Based on
    - Amount of acceptable *sampling error*
    - Expected magnitude of population proportions

72

Sample Size

- You should try to select an *economic sample*
  - Includes enough respondents to ensure a valid survey and no more
- Two factors are taken into account when determining necessary sample size
  - Amount of acceptable *sampling error*
  - Expected magnitude of population proportions
- Formula to calculate sample size using the above parameters

73

Homework

- Do Homework 19 – Designing a Composite Self-Report Measure
  - Design a new composite self-report measure to assess a person's degree of addiction to Facebook. Assume it only has one factor, but use at least five scale items. Incorporate information from at least one literature reference. Assess the face and content validity of your measure and work through a bivariate anlaysis of your items.
  - Decide on one method for assessing validity (besides face & content) for your measure that you can also assess in a self-report questionnaire (should be a numeric measure).
  - Add any other items to your questionnaire that you feel are important.
  - Email your questionnaire to the class (is4800-all@ccs.neu.edu). (You are also obligated to reply to any questionnaires mailed to you within 48 hrs.)
  - Compute the reliability (internal consistency) of your measure using R Alpha. Compute descriptive statistics for your measure and any other items you may have included on the questionnaire. Assess the validity of your measure (you can do this qualitatively, e.g., using scatterplots).
  - Document and submit all of the above.
  - You may work individually or in teams of two. Due in one week.
 - Extra credit: implement questionnaire on surveymonkey.com

74

Aronson Reading Preview: Hypothesis Testing
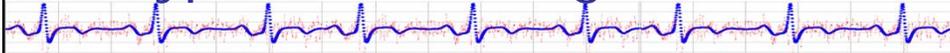
75

Lecture
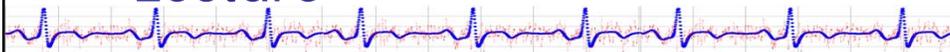
- How would you do hypothesis testing?
- Kinds of errors?
- Which are worse?

76

A typical scenario

- Between-subject design
- Let every subject try both Wizziword & Word
- Measure performance
- Research Hypothesis:
  - Wizziword is better

77

What if ...

- You can test every subject in your population, AND
- There is no measurement error?
- →
- Compute descriptives for two treatments
- If Wizziword perf > Word perf conclude H1 is supported
- No uncertainty, No 'p' value

78

Basic Process of Hypothesis Testing
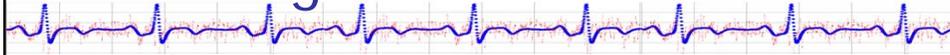
- H1: Research Hypothesis:
  - Population 1 is different than Population 2
- H0: Null Hypothesis:
  - No difference between Pop 1 and Pop 2
  - *The difference is "null"*
- Compute *p(observed difference/H0)*
  - 'p' = probability observed difference is due to random variation
- If p<*threshold* then reject H0 => accept H1
  - p typically set to 0.05 for most work
  - p is called the "level of significance"

79

Other ways of thinking about this...
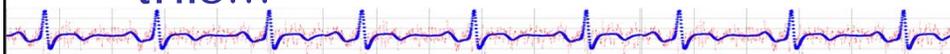
- "Innocent until proven guilty."

- How surprising would this result be if there really were no difference?

- Why do things this way???

80

The grand plan

- X^2 tests
  - For nominal measures
  - Can apply to a single measure
- Correlation tests
  - For two numeric measures
- t-test for independent means
  - For categorical IV, numeric DV

81

Assignment: customer satisfaction with your bathroom games division

- Marketing is thinking about rolling out a new game...
- Want to measure customer satisfaction with one of your product lines (a console to let users play games while seated in the bathroom)
- Want the components of satisfaction ...

82