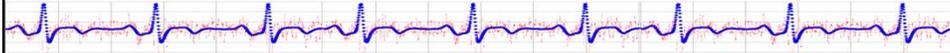


Empirical Research Methods in Information Science

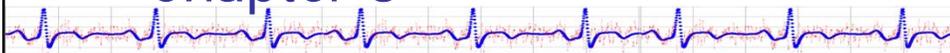
IS 4800 / CS6350



Lecture 8 Miscellaneous Measures

1

Chapter 8



Using Nonexperimental Research (?)

5

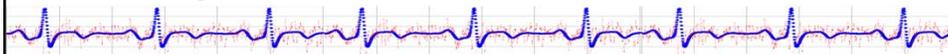
Observational
Research



Nonexperimental
Research

6

Example: Handheld ECAs



- Research Question:
 - Do people exhibit the same nonverbal conversational behavior when talking to a 2" tall character than when talking to another person face-to-face?
- Exercise:
 - Design the study

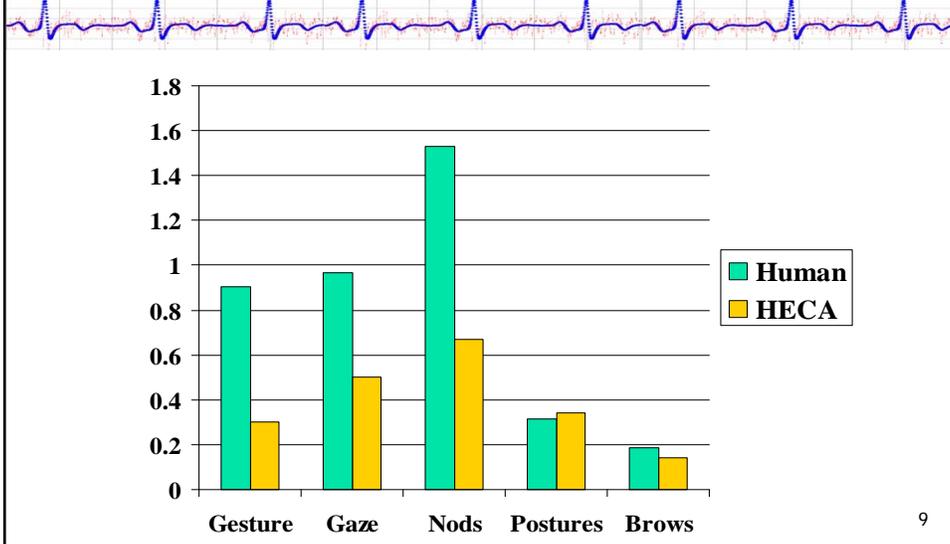


7

Example: Handheld ECAs



Results

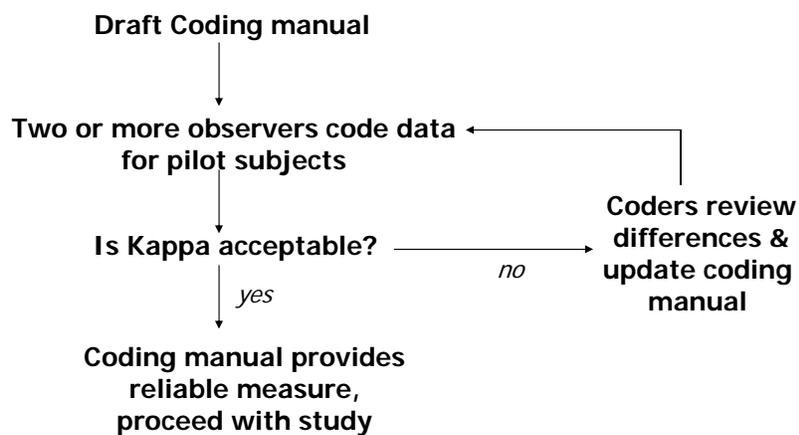


Defining Behavioral Categories

- Only need enough detail to provide a reliable measure.
 - What is reliability?
 - How to measure it?
- e.g. do people in the student center get more rude 10 minutes before class times?

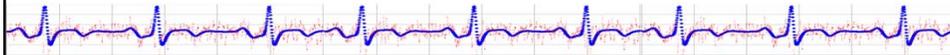


Defining a behavioral protocol



11

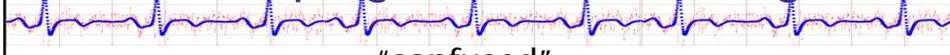
Developing Behavioral Categories



- Categories must be operationally defined
- Behavioral categories must be clearly defined to avoid ambiguity
 - “flailing arms around”
vs.
■ “moved arms from below to above waist and back more than 3 times per minute”

12

Developing Behavioral Categories



“confused”

vs.

(“clicked mouse at least 5 times on inappropriate menu”

OR

“gazed at interface with mouth open AND no mouse clicks or keyboard presses for 5 minutes”)

AND

“furrowed brows”

13

Coding Manual

- You should write your behavior identification rules down so that you could give them to someone else to follow reliably.
- You should also write down the sampling and coding methods you will use, as well as your recording instrument (e.g., paper form).

14

Quantifying Behavior in Observational Research

- *Frequency Method*
 - Record the frequency with which a behavior occurs within a time period
- *Duration Method*
 - Record how long a behavior lasts
- *Intervals Method*
 - Divide the observation period into several discrete time intervals (e.g., ten 2-minute intervals), and record whether a behavior occurs within each interval

16

Example: Code Posture Shifts



17

Example

■ Posture shifts

■ Body part

- Upper body
- Lower body
- Both

■ Type

- Shift
- Return

■ Energy level

- 0-100%

- Hand gestures and other communicative behavior does not count – nor their effects.

StartTime	EndTime	BodyPart	Type	Energy
00:00:03	00:00:04	Upper	Return	50%
...

18

Posture Shifts

Duration, Frequency, or Interval Measures?

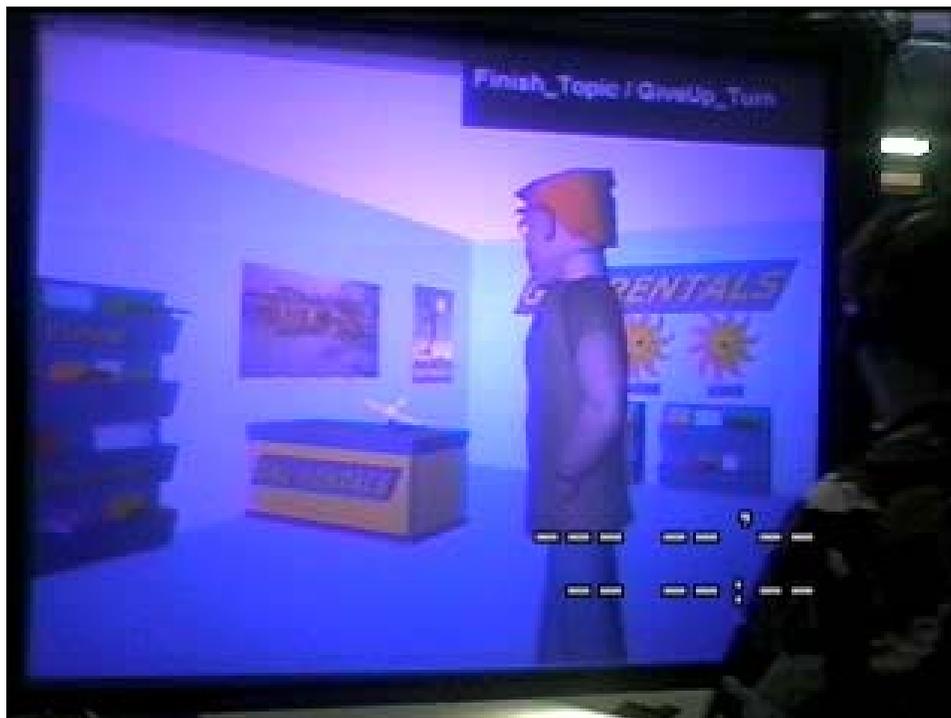


Posture shifts with respect to discourse segment

	Monologues (0.06/s)			Dialogues (0.07/s)		
	ps/s	ps/int	energy	ps/s	ps/int	energy
Inter-dseg	<u>0.340</u>	0.837	0.832	<u>0.332</u>	0.533	0.844
intra-dseg	<u>0.039</u>		0.701	<u>0.053</u>		0.723

Lecture 1 - Introduction

19



Tools for Coding: ANVIL

The screenshot displays the ANVIL software interface, which is used for coding gestures in video. It consists of several windows:

- Top Left:** A console window showing the following text:

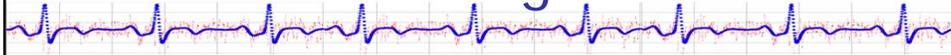
```
DEBUG MODE IS ON
open file tq1-kara-anvil
closing annotation
read anvil file: anvil.Annotation@34263a
create player for: video.quicktime
FrameRate = 0.0
Movie duration: 02:27:28
No. of frames: 3681
open player (first)
creating annotation window
```
- Top Center:** A video player window showing a scene with two men sitting on a couch. A vertical volume slider is on the left, and playback controls are at the bottom.
- Top Right:** A window for a selected track, "gesture.metaphoric". It shows the start and end times (00:18:23 to 00:18:47) and a list of attributes: "phase: stroke", "location-height: chest", "location-side: outer-right", and "handedness: right". A comment field contains "copy gesture -> elaboration".
- Bottom:** A timeline window showing a sequence of frames (15 to 24). The timeline is divided into rows for different coding categories: "tr", "ling", "posture" (with sub-categories "head", "upper", "lower"), "gesture" (with sub-categories "emblem", "iconic", "metaphoric"), and "scholastic". Colored blocks indicate the duration of each coded event across the frames.

Behavior Coding Exercise



23

Behavior Coding Exercise

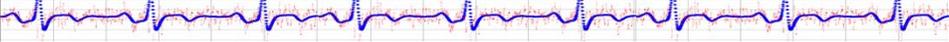


- Does playing with Sam cause children to play together more?

24



Coping With Complexity in Observational Research



- *Recording*
 - Use a recording device to make a record of behavior for later review
- *Time Sampling*
 - Scan subjects for a specific period (e.g., 30 seconds), and then record your observations during the next period
- *Individual Sampling*
 - Select a subject and observe behavior for a given period (e.g., 30 seconds), and then shift to another subject and repeat observations

26

Coping With Complexity in Observational Research



- *Event Sampling*
 - Select one behavior for observation and record all instances of that behavior
 - It is best if one behavior can be specified as more important than others

27

Coping With Complexity in Observational Research

- Ecological momentary assessment
- Intelligent/Context Aware EMA
- Smart rooms

- What kind of sampling is this?



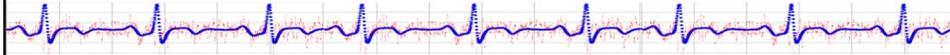
28

Smart Rooms – e.g. PlaceLab



29

Issues in Observational Research

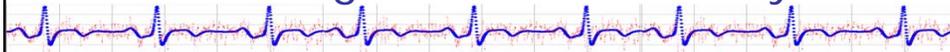


- IRB issues with video/audio recording?

- Behavior vs. Function/Intent

30

Evaluating Interrater Reliability

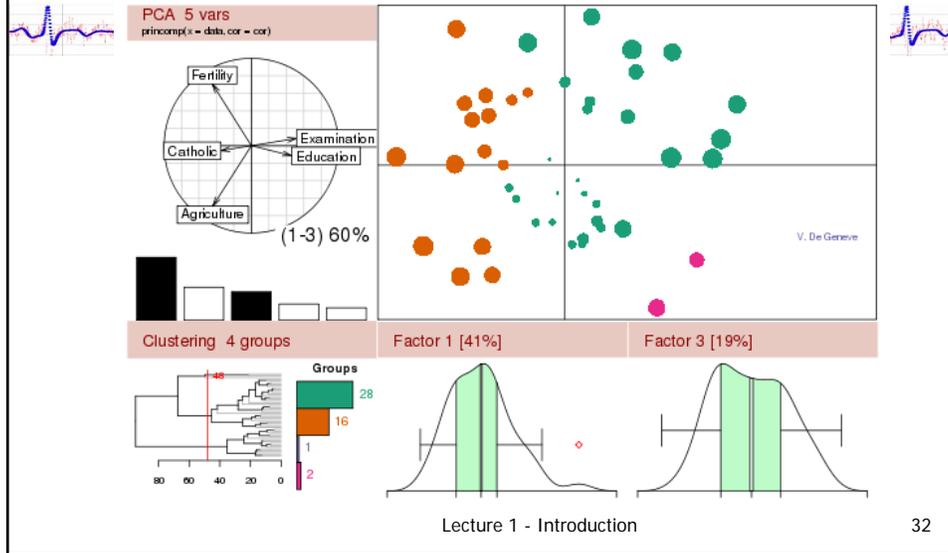


- You must establish reliability of observations from multiple observers (*interrater reliability*)
- Most common/acceptable method for evaluating interrater reliability
 - *Cohen's Kappa*
 - Allows you to determine if agreement observed is due to chance
 - Kappa of 0.70 or more indicates acceptable interrater reliability

31

The R Project for Statistical Computing

Interrater Reliability



Example R data setup

Time	Judge1	Judge2
1	together	together
2	apart	apart
3	together	together
4	apart	together
5	apart	apart
6	together	together
7	together	together

Kappa

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

> require(psych) #every session

> wkappa(table(data$Judge1,data$Judge2))

$kappa [1] 0.6

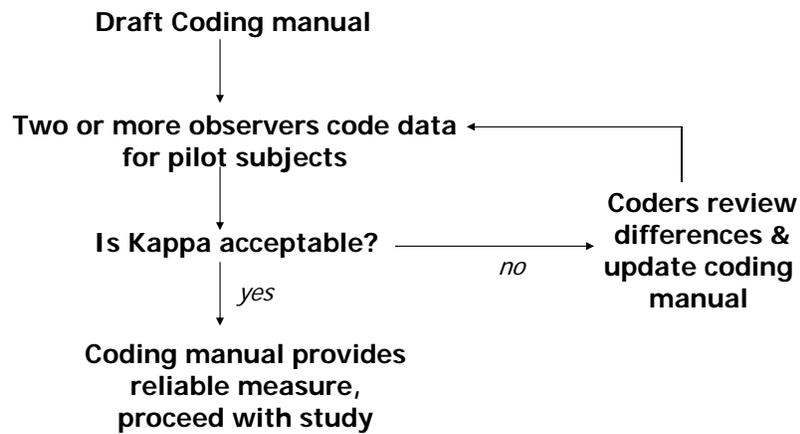
#accounts for distance of each discrepancy
$weighted.kappa [1] 0.2
```

Other statistics for inter-rater reliability

- Pearson correlation coefficient
 - Problematic

- Intraclass correlation coefficient
 - If numeric (interval, ratio) judgments
 - See 'icc' function in 'irr' R package.

Refining a behavioral protocol



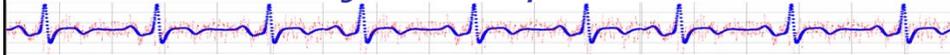
37

Exercise: Evaluate Enjoyment



38

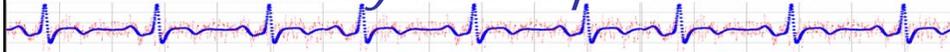
Approaches to Data Collection *Necessarily non-experimental?*



- *Naturalistic Observation*
 - Unobtrusive observations of subjects' naturally occurring behavior are made
- *Ethnography*
 - The researcher becomes immersed in the behavioral or social system being studied. May be conducted as a participant or non-participant observation study
- *Sociometry*
 - You identify and measure interpersonal relationships within a group

39

Approaches to Data Collection *Necessarily non-experimental?*



- *Case History*
 - You observe and report on a single case
- *Content Analysis*
 - You analyze spoken or written records for the occurrence of specific categories of events (e.g., a word or phrase)

40

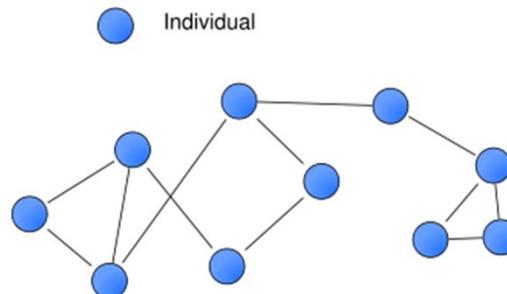
Approaches to Data Collection *Necessarily non-experimental?*

- *Archival Research*
 - You use existing records (e.g., police records) as your source of data
- *Meta-Analysis*
 - Compute overall statistics based on a number of previously-published studies.

41

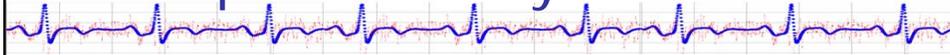
Sociometry on Steroids: Social Network Analysis

- Analyze structures
- Measures
 - Betweenness – how often a node is an intermediary
 - Centrality – number of ties to others
 - Many more...



42

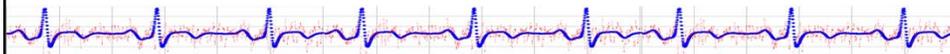
Sequential Analysis



- B&A say recording sequences of behavior may yield more information than individual events.
 - e.g. interruption followed by grimace followed by rolling eyes

43

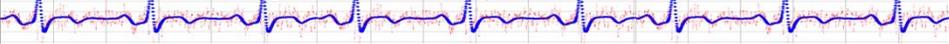
Content Analysis: Defining Characteristics



- Used to analyze a written or spoken record for occurrence of specific behaviors or events
- Archival sources often used as sources for data
- Response categories must be clearly defined
- A method for quantifying behavior must be defined

44

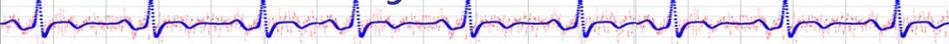
Example Study



- The CEO of Global Enterprises, Inc. is very worried about the low morale in the company, as evidenced by the amount of flame email she receives. She considers sending every office on a “ropes” course, but to do this would cost the company \$10M. She asks you to do a study to tell how well her scheme might actually work in reducing her flame mail.

45

Meta-Analyses



- Compare/Integrate “all” studies that have investigated a given phenomena
 - E.g., use of a particular medication for a particular disease
- Common in the literature (esp. medical)
- Very methodical
 - Search for articles
 - Eligibility criteria
 - Statistical analyses

49

Meta-Analysis

- New terms(?)
 - Level of Significance
 - Effect Size
 - Type I & II errors

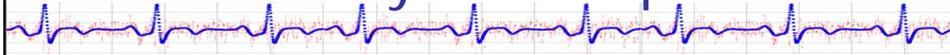
50

Meta-Analyses

- Effect Size
 - Measure of how much difference exists between treatment groups in an experiment
 - How to assess as common metric?
 - E.g., compare effect of large monitors on productivity
 - Study 1 measures widgets per day
 - Study 2 measures subjective assessment of managers
 - How to integrate across studies?

51

Meta-analysis example



CHI 2007 Proceedings • Faces & Bodies in Interaction

April 28-May 3, 2007 • San Jose, CA, USA

A Meta-Analysis of the Impact of the Inclusion and Realism of Human-Like Faces on User Experiences in Interfaces

Nick Yee, Jeremy N. Bailenson, Kathryn Rickertsen

Department of Communication
Stanford University, Stanford, CA
{nyee, bailenson, kathryn}@stanford.edu

52

METHOD

Selection of Studies



The studies considered for inclusion in this analysis were culled from bibliographic indexes related to the fields of psychology, computer-mediated communication (CMC), and virtual reality. These included Expanded Academic ASAP, Google Scholar, Google keyword, PsycInfo, PsycArticles Fulltext Search, InterDok, ProQuest, and SearchPlus. In this initial pass, articles that appeared to report an experimental study of anthropomorphism, embodied agents, or agent realism were collected and reviewed. Sources were only considered if they were published in a peer-reviewed journal or in published conference proceedings. This ensured a basic level of

53

The literature review yielded 106 studies. Several selection criteria were then applied. First, an article was included only if it was an experimental study that manipulated the variables of interest and contained clear reports of quantitative data relating to the outcome of different conditions. Thus, purely qualitative studies involving open-ended self-reports or observational user studies without quantitative coding schemes or dependent variables were removed.



54

Of these 25 studies, the average year of publication was 2001.96 (SD = 2.29) with a median of 2002. The average sample size within each study was 45.40 (SD = 35.55). With regard to study location, 13 were conducted in the US or Canada, 9 were performed in Europe, and the remaining 3 were conducted in Asia. And finally, with regard to equipment used, 17 were conducted on desktop equipment, 6 were conducted using immersive virtual reality, and the remaining 2 were conducted on a large projected screen.



55

Effect Size Calculations



To generate the necessary effect size tabulations in order to test our hypotheses, we tabulated several possible effect sizes for each paper depending on the available conditions. First, we tabulated the results of performance data separately from the results of subjective data. Performance data might include time to task completion, accuracy measures, or similar behavioral measures. Subjective data, on the other hand, was any measure that was based on self-report or survey data. Second, we tabulated effect sizes based on two kinds of comparisons between conditions. We

56

RESULTS

Formal Meta-Analyses



The results of the effect size and significance value aggregation are listed in Appendix A for each individual study and the overall values. The overall effect sizes of the four comparison conditions ranged from -.04 to .14. While three of the four comparison conditions were highly significant at p levels of less than .05, the comparison of high-low realism using performance measures was not significant, with $p = .14$.

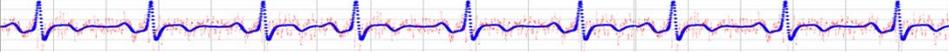
57

APPENDIX A – EFFECT SIZES AND SIGNIFICANCE VALUES OF STUDIES INCLUDED

	Performance		Subjective		N
	Face vs. No Face	High vs. Low Realism	Face vs. No Face	High vs. Low Realism	
Okonkwo & Vassileva, 2001 [41]		$r = 0, z = 0.24$		$r = 0.03, z = 0.84$	12
Moundridou, Virvou 2002 [37]	$r = 0.1, z = 0.39$		$r = 0.48, z = 4$		48
Hongpaisanwiwat & Lewis, 2003 [23]	$r = 0, z = -0.02$	$r = 0.07, z = 0.45$			50
Burgoon, Bengtsson, Bonito, Ramirez, & Dunbar, 1999 [11]	$r = 0.03, z = 0.2$	$r = -0.03, z = -0.17$	$r = 0, z = -0.04$	$r = 0.12, z = 0.8$	50
Bailenson, Beall, & Blasovich, 2002 [2]			$r = 0.51, z = 1.92$	$r = 0.16, z = 0.46$	30
Burgoon, Bonito, Bengtsson, Cederberg, Lundeberg,

In our meta-analysis, we had also separated: 1) studies that compared interacting with an agent that had no facial representation versus an agent that had a facial representation (i.e., the yes-no comparisons), and 2) studies that compared interacting with faces of low realism versus faces of high realism (i.e., the high-low comparison). A comparison of these two groups of effect sizes revealed that the effect sizes from yes-no comparisons ($n = 25, r = .16$) were significantly larger than those from the high-low comparison ($n = 18, r = .07$), $z = 2.43, p = .02$.

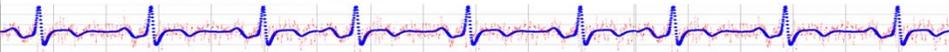
Homework



- **Read B&A Ch 9 and paper.**
- **Do Homework 18 – Observational Measure**
 - Your manager thinks everyone in the department is slacking off. She asks you to do an objective study of how hard people are working at their computers.
 - Define observational measures to categorize a computer worker's behavior as "intensely working", "casually working", or "not working". Think about how to encode a worker's behavior during 15 minutes of observation (hint: intervals method). Write a brief coding manual so that someone else could follow your directions.
 - Pick a public area in which people are using computers and observe two people for 15 minutes each. If you cannot surreptitiously observe users, be sure to obtain verbal consent first. Code their behavior.

60

Homework



- **Individual:** turn in your coding manual, data, and any descriptive statistics that make sense.
 - **Team of 2 students (preferred):** both code observations for the same subjects at the same time. In addition to the above, compute and report Kappa. **If your Kappa is below 0.7** discuss the nature of your discrepancies and how you would update your coding manual to be more reliable.
-
- Due in 1 week!

61