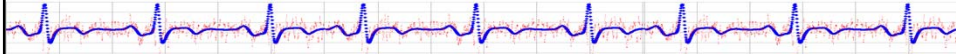


Empirical Research Methods in Information Science

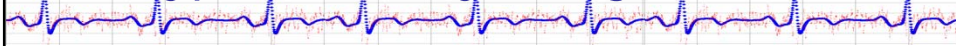
IS 4800 / CS 6350



Lecture 24 Some Advanced Designs

1

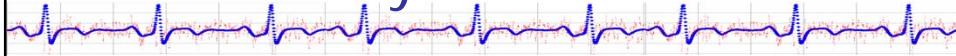
Types of Study Designs



- Qualitative
 - Ethnography
- Quantitative
 - Descriptive
 - Correlational
 - Demonstrative
 - Experimental
 - Between-subjects
 - Single factor, two-level
 - Single factor, N-level (for $N > 2$)
 - M factor, N-level, full factorial (for $M, N \geq 2$)
 - Within-subjects
 - Single factor, two-level

3

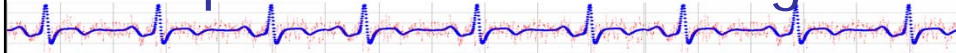
What you need to remember from today...



- Concepts / Terms
 - Repeated-measures design
 - Mixed design
 - Covariate
 - Quasi-experimental design
 - Pretest-Posttest design
 - Development design
 - Cross-sectional design
 - Longitudinal design
 - Single subject design

4

Repeated Measures Design

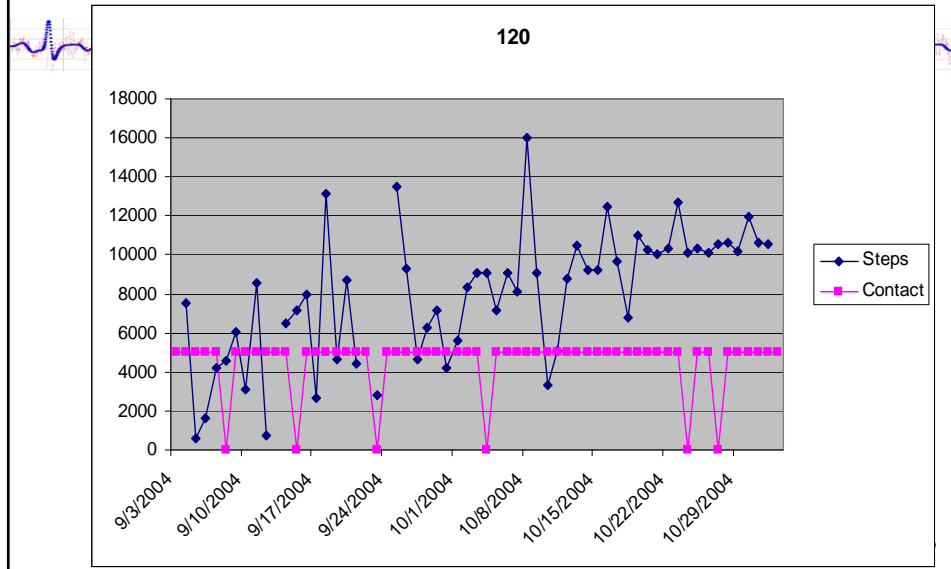


- Taking the same measure multiple times from the same subject.

- For two measurements = ?
 - 2 treatment within-subjects design.

5

Repeated Measures Design



Repeated Measures Design

- One-way ANOVA?
 - No!
- Repeated-measures ANOVA or more advanced methods.

Mixed Design

- Some between-subjects factors and some within-subjects factors

8

Mixed Design aka "Split plot" design

Within-Subjects
Levels

	Between-Subjects Level 1		Between Subjects Level 2	
	A	B	A	B
S₁	→	→	S₆	→
S₂	→	→	S₇	→
S₃	→	→	S₈	→
S₄	→	→	S₉	→
S₅	→	→	S₁₀	→
S₆	→	→	S₁₁	→

9

Mixed Design Example

- Boysen & Keller, 1980, "Measuring computer program comprehension"
- Displayed a PL expression and asked subjects to evaluate whether last line was true or false, e.g.
 $X := 6$
 $X > 4$
- Outcome measure (DV): reaction time
- Between-Ss factor:
 - months experience
- Repeated measures (within-Ss factor):
 - Expression ($::=$, IF, CASE, etc.)
 - Value of X (digits 0-9)

10

Mixed Design Example

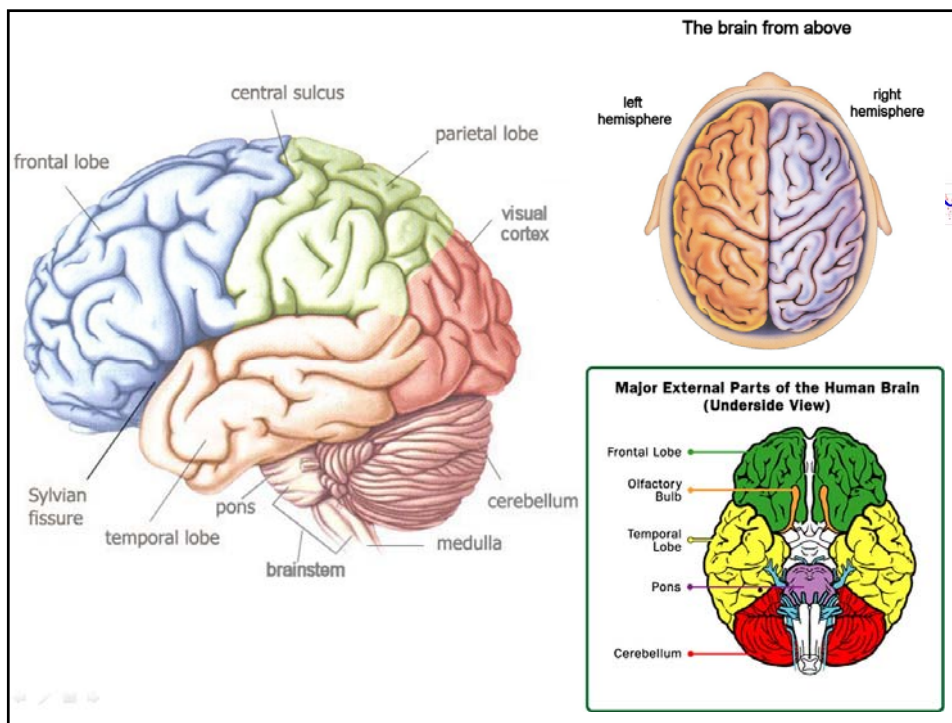


11

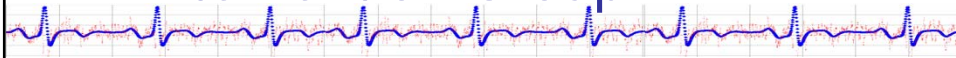
Design a Study to Evaluate the Effectiveness of Pinky..

- Compare studying text & image, to this plus video.
- Hypothesize recall ability decays rapidly after learning exposure (e.g., over 10 minutes), and differs by teaching method.
- Design the study...

12

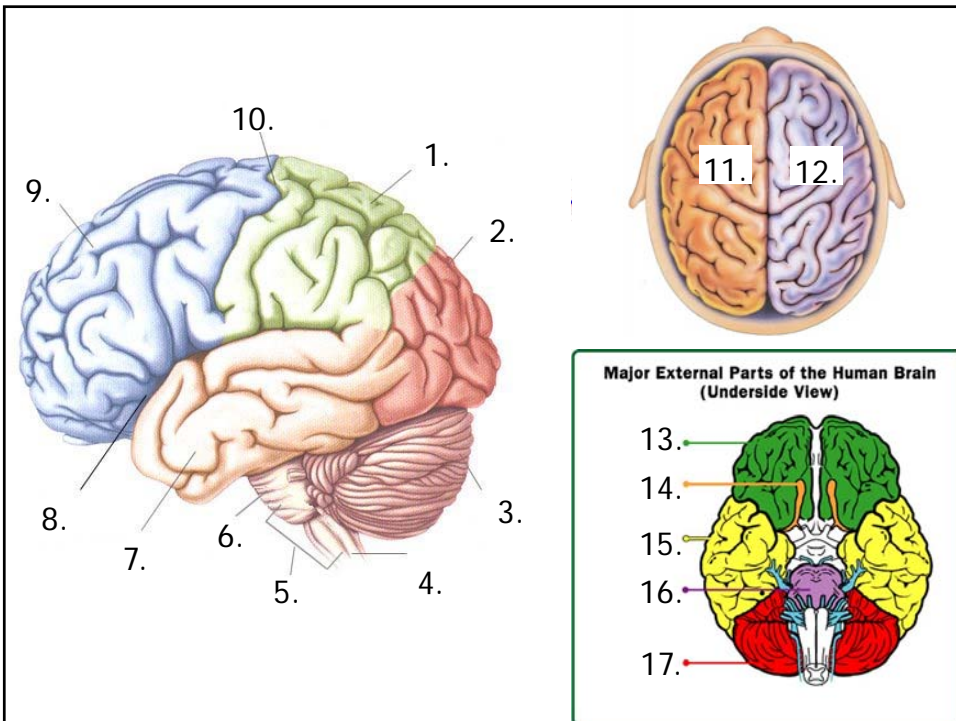


Intervention Group

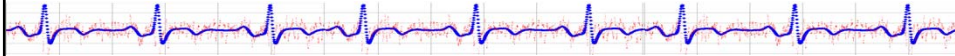


1. Show Textbook page
 1. "Please try to learn the labels on the diagrams. You may not take any notes or use your laptop or talk to anyone. You are welcome to move closer to the screen if you need to."
 2. "I will give you 3 minutes to study them, then I will show you a short educational video on the topic."
 3. Give subjects 3 minutes to study.
 4. "Now I will show you the video."
 5. Play video, wait 30 seconds.
 6. "I'm going to play it one final time."
 7. Play video again.
 8. Close video and pps file.
2. Immediately hand out Test Time 1 sheets, then display **TEST** Give subjects one minute to complete (in silence, no notes or laptop, no peeking at others). Collect.

14

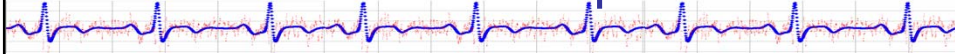


Test



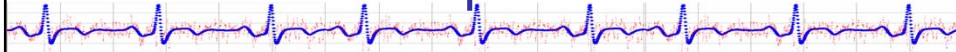
ID _____	Time 1	9. _____
1. _____		10. _____
2. _____		11. _____
3. _____		12. _____
4. _____		13. _____
5. _____		14. _____
6. _____		15. _____
7. _____		16. _____
8. _____		17. _____

Intervention Group



3. Ask one of the VJs to show some YouTube videos (unrelated to anatomy). Stop the VJ after exactly 5 minutes. (Score number correct for each student during the wait.)
4. Hand out Test Time 2 sheets, then display **TEST**. Give subjects one minute to complete (as above). Collect.
5. Ask the second VJ to show some YouTube videos (unrelated to anatomy). Stop the VJ after exactly 5 minutes. (Score number correct for each student during the wait.)
6. Hand out Test Time 3 sheets, then display **TEST**. Give subjects one minute to complete (as above). Collect. Score number correct for each student.

Control Group

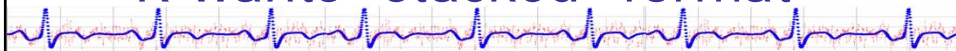


1. Show 1-TEACH.pps (on the desktop).
 1. "Please try to learn the labels on the diagrams. You may not take any notes or use your laptop or talk to anyone. You are welcome to move closer to the screen if you need to."
 2. "I will give you 6 minutes to study them."
 3. Give subjects 6 minutes to study.
 4. Close pps file.
2. Immediately hand out Test Time 1 sheets, then display **TEST**. Give subjects one minute to complete (in silence, no notes or laptop, no peeking at others). Collect.
3. Ask one of the VJs to show some YouTube videos (unrelated to anatomy). Stop the VJ after exactly 5 minutes. (Score number correct for each student during the wait.)
4. Etc (as in Intervention group)

18

Data

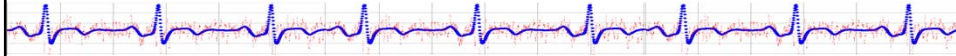
R wants "stacked" format



ID	group	time	score
100	1	1	10
101	1	1	4
102	1	1	8
103	1	1	10
201	1	1	4
202	1	1	10
203	1	1	3
204	1	1	10
205	1	1	2
300	0	1	11
301	0	1	5
302	0	1	9
303	0	1	11
401	0	1	6
402	0	1	12

19

R



```
> library(ez)
> d$group<-factor(d$group)
> d$time<-factor(d$time)
> d$ID<-factor(d$ID)

> res<-ezANOVA(data=d, dv=.(score),
wid=.(ID), between=.(group),
within=.(time), detailed=TRUE)
```

20

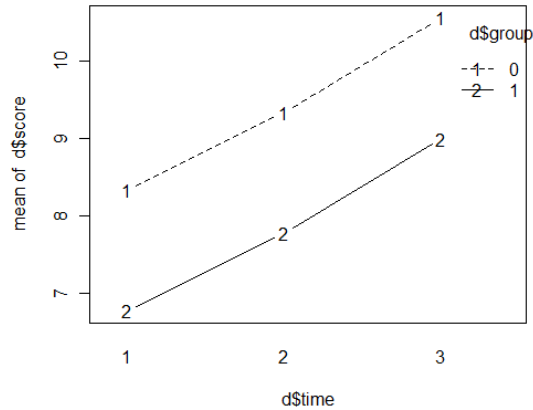
R

```
$ANOVA
      Effect DFn DFd      SSn      SSd      F      p p<.05
1 (Intercept)  1  16 4.021407e+03 640.59259 1.004422e+02 2.663801e-08 *
2      group   1  16 3.266667e+01 640.59259 8.159112e-01 3.797770e-01
3      time   2  32 4.459259e+01 28.74074 2.482474e+01 3.098450e-07 *
4 group:time   2  32 7.691394e-30 28.74074 4.281807e-30 1.000000e+00
```

21

R

```
interaction.plot(x.factor=d$time, trace.factor=d$group,  
response=d$score, fun=mean, type="b")
```



22

Nested Design.

Objective: Generality across Bet-Ss Levels

Between-Subjects Level 1		Between Subjects Level 2	
A	B	C	D
S ₁	→	S ₆	→
S ₂	→	S ₇	→
S ₃	→	S ₈	→
S ₄	→	S ₉	→
S ₅	→	S ₁₀	→
S ₆	→	S ₁₁	→

23

Example Nested Design *Related – Between/Between*

- Pfleeger '95, "Experimental design and analysis in software engineering, part 4: choosing an experimental design"
- Suppose you want to test the effectiveness of two design methods (e.g., scrum and micromanaged waterfall) on quality of design, with and without SE tool support.

25

Example: "Crossed Design"

- Assign projects to cells – look at every treatment combination.

Crossed		Design Method	
		Method A	Method B
Tool Usage	Not used	Projects 1, 2 and 3	Projects 7, 8 and 9
	Used	Projects 4, 5 and 6	Projects 10, 11 and 12

- However, what if tools for Method A are different for Method B? Some combinations don't make sense.

26

Example Nested Design

Design Method			
Method A		Method B	
Tool Usage		Tool Usage	
Not used	Used	Not used	Used
Projs. 1,2,3	Projs. 4,5,6	Projs. 7,8,9	Projs. 10,11,12

- Kind of study?
 - between-Ss factors only
- How to analyze?
 - Use one-way ANOVA instead of 2x2 ANOVA to analyze.
- Tools per Method are different.

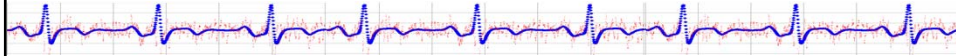
27

Combining Experimental and Correlational Designs

- Including a *covariate* in an experimental design
 - A *covariate* is a correlational variable (e.g., self-esteem) in an experimental design
 - Usually continuous
 - Can test to see if significant effect on DV
 - More commonly, "Subtract out" the influence of the covariate to reduce error variance
 - Makes your design more sensitive to the effects of the independent variable
- AKA "potentially confounding variable"
 - Often demographics, such as age, etc.

28

Most common use of covariates



- Correct for imbalances in randomization
 - When significant differences between groups at baseline (e.g., in demographics)
- Correct for known factors that have very significant effect on DV
 - Check quality of models with various combinations of covariates

29

Example Baseline Measures Table

Variable		All N=263	Control N=131	Intervention N=132
Sex (Female)	N (%)	161 (61.2)	72 (55.0)	89 (67.4)
Age at enrollment	Mean (SD)	71.3 (5.4)	70.8 (5.2)	71.7 (5.6)
Race	Black, N(%)	165 (62.7)	86 (65.6)	79 (59.8)
	White, N(%)	75 (28.5)	35 (26.7)	40 (30.3)
	Other, N(%)	23 (8.7)	10 (7.6)	13 (9.8)
Hispanic or Latino	N (%)	20 (7.6)	10 (7.6)	10 (7.6)
Education	<HS, N(%)	56 (21.3)	30 (22.9)	26 (19.7)
	HS, N(%)	79 (30.0)	34 (26.0)	45 (34.1)
	>HS, N(%)	128 (48.7)	67 (51.1)	61 (46.2)

30

Quasi-Independent Variable

- Including a *quasi-independent variable* in an experimental design
 - A *quasi-independent variable* is a correlational variable (e.g., gender) that looks like an experimental variable
 - Resulting design looks like a factorial experimental design
 - The quasi-independent variable must not be interpreted as causing changes in the dependent variable

31

Example

CHI 2000 • 1-6 APRIL 2000

Papers

The Effects of Animated Characters on Anxiety, Task Performance, and Evaluations of User Interfaces

Raoul Rickenberg and Byron Reeves



Figure 1: Examples of the Monitoring Character (Top Row) and Idle Character (Bottom Row).

32

Locus of control

Now imagine two different people being monitored while they work on a complex task. The first person believes that she controls her own destiny and that other people have little to do with whether she fails or succeeds. The second person is convinced, however, that he is at the mercy of

35

EXPERIMENTAL METHODS

Subjects. Eighty-four people participated in the experiment (60% male and 40% female). An additional 20 people were used to pretest stimulus materials. All subjects were either undergraduate or graduate students recruited at Stanford University. All were experienced computer users (i.e., they knew how to word-process and manage a UNIX email account).

Experimental Design. The experiment was a between-subjects, full-factorial two-by-three design. The two factors were (1) the subjects' locus of control and (2) the monitoring activity of an animated character.

36

Locus of Control Pretest. Rotter's [27] Locus of Control Scale was used to determine the internal versus external orientations of potential subjects. This instrument consists of 23 forced-choice items that each present a pair of statements. In each pair, one statement expresses an internal viewpoint and the other an external viewpoint.

Scores on this scale can range from 0, indicating that no external statements are endorsed, to 23, indicating that all external statements are endorsed. The mean score on the pretest was 13.27 ($SD = 3.8$). Only the 42 subjects that scored lowest (internal) and highest (external) on the pretest were selected to participate. A two-tailed t -test indicated that the scores on the Locus of Control scale for these two groups was significantly different ($t(82) = 18.47, p < .001$).

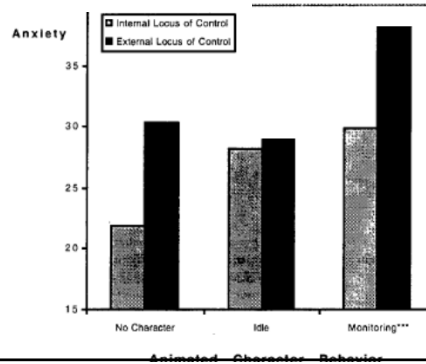
37

RESULTS

Full factorial ANOVAs were performed on all measures. A summary of these ANOVAs appears in Table 1. The planned tests of all hypotheses are discussed in detail below, as are results pertaining to the relationship of animated characters and locus of control to the evaluation of the websites.

42

Figure 2 shows the results for anxiety. One-tailed, *a priori* contrasts showed that people were more anxious when an idle character was present than when no character was present ($t(72) = 1.4, p = .08$). Also, people were more anxious if an animated character appeared to monitor them than if no character was present ($t(72) = 4.6, p < .001$). And people were more anxious if an animated character monitored them than if an idle character was present ($t(72) = 4.6, p < .001$). Users with an external locus of control were also more anxious when monitored by an animated character than were users with an internal locus of control ($t(72) = 4.6, p < .001$).



43

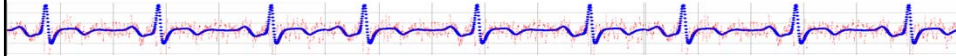
DISCUSSION

The perception of being monitored by an animated character has the same effects on Anxiety and Performance as being monitored by a human, either electronically or in person. When a character watches, users are more likely to feel anxious about their work and to perform less well. This anxiety is most pronounced among users who think that other people control their success.

At the most general level, these results suggests that decisions concerning the use of animated characters should address the details of execution and social presentation. It is not sufficient—for celebration or condemnation—to focus on whether or not an animated character is present. Rather, the ultimate evaluation is similar to those for real people—it depends on what the character does, what it says, and how it presents itself. The effects of animated characters are not

44

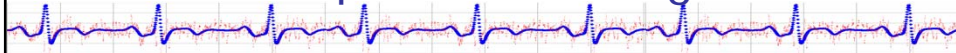
Example Quasi-Independent Variable



- You want to compare your new game (A) to your competitor's game (B) on satisfaction.
- You hypothesize that players who already own B game may prefer B, whereas those who have not already purchased B will like your game more in a head-to-head comparison.

45

Quasi-Experimental Designs



- Similar to an experiment, but lacking randomization
- May allow you to control the assignment to the treatment condition, but using some method other than random assignment (eg. an eligibility score)
- May have no control over assignment
- *Example: when the only or primary IV is quasi-experimental.*
 - *You can't randomize to gender or personality trait.*

46

Quasi-Experimental Designs

- Similar to an experiment, but lacking randomization
- *Example: Time Series Design (no control)*
 - Essentially a within-subjects design without control and without counterbalancing
 - IV = pre or post an event you cannot manipulate
 - E.g., enactment of a new law, hurricane, etc.
 - Make several observations of behavior before and after introducing your independent variable
 - t-test for dependent means if 2 measures (compare after-before), else repeated measures ANOVA.
- When would you use this?

47

Example: Depression following Catastrophe

Pre-test



Post-test

48

Example: Acceptance & Usability studies

- Especially if small N
 - Budget, resource constraints
 - Difficult to recruit population

49

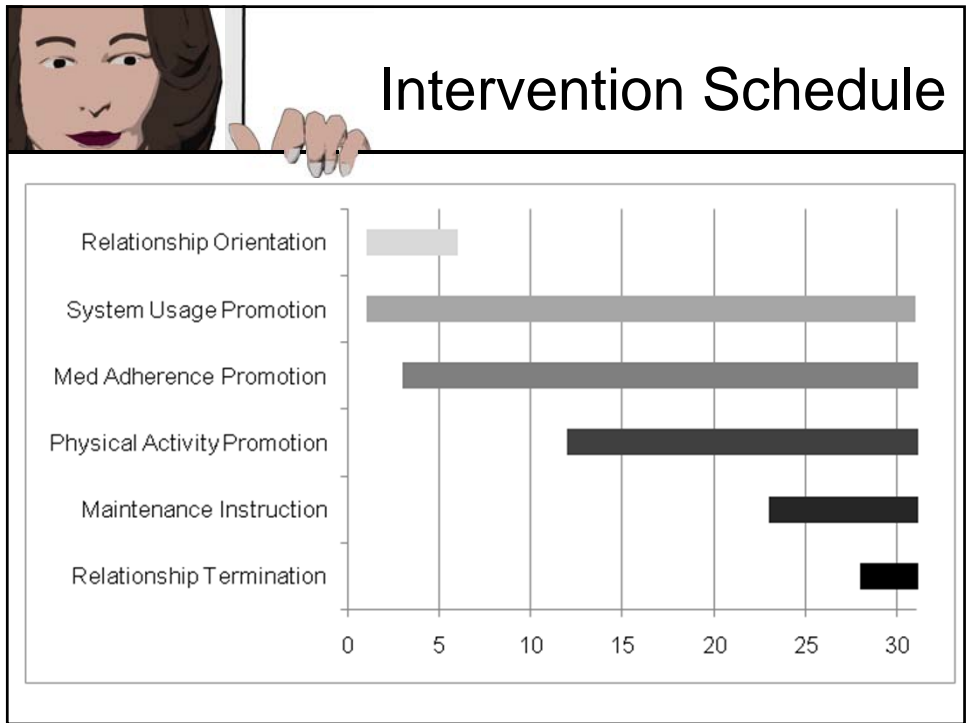
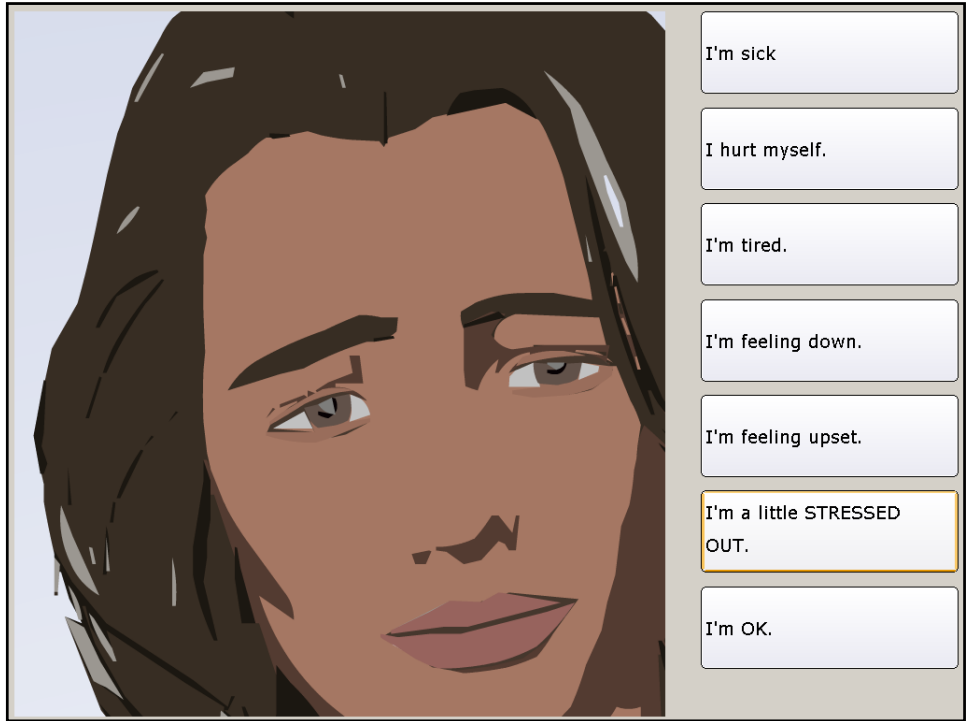
Antipsychotic Medication Adherence for Young Adults with Schizophrenia


Funded by Eli Lilly Pharmaceuticals – In collaboration with University of Pittsburgh School of Nursing

- 30 day, pre-post design study
- Pilot study: 20 subjects
- Intervening on three behaviors in parallel:
 - System use
 - Medication adherence
 - Physical activity

Maintaining Reality: Relational Agents for Antipsychotic Medication Adherence
 Interacting with Computers, 2010, 22(4)
 276-288

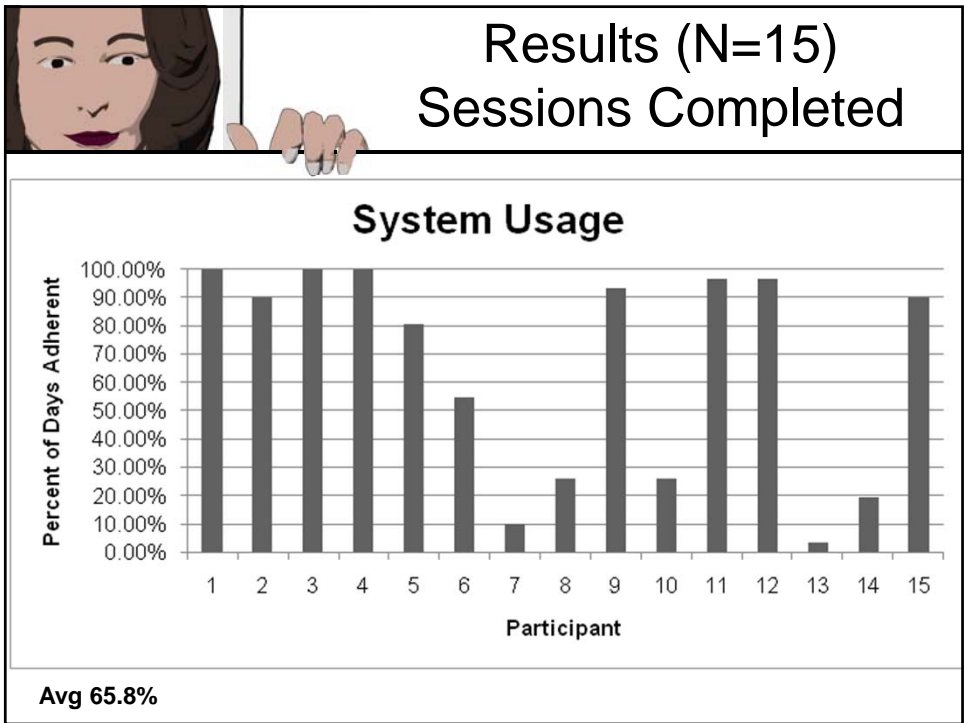


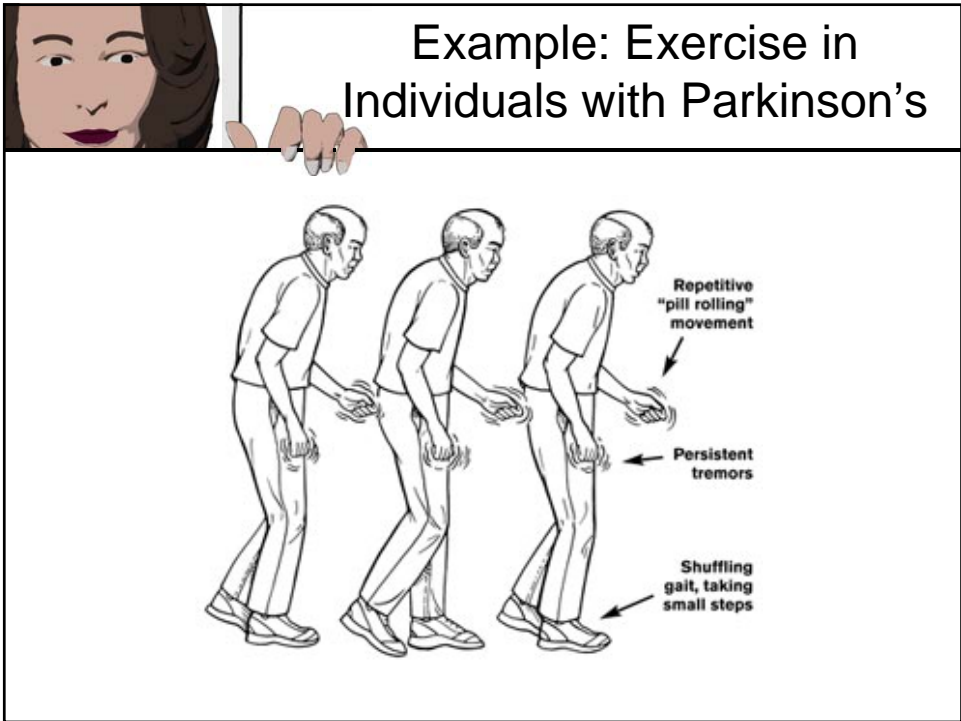
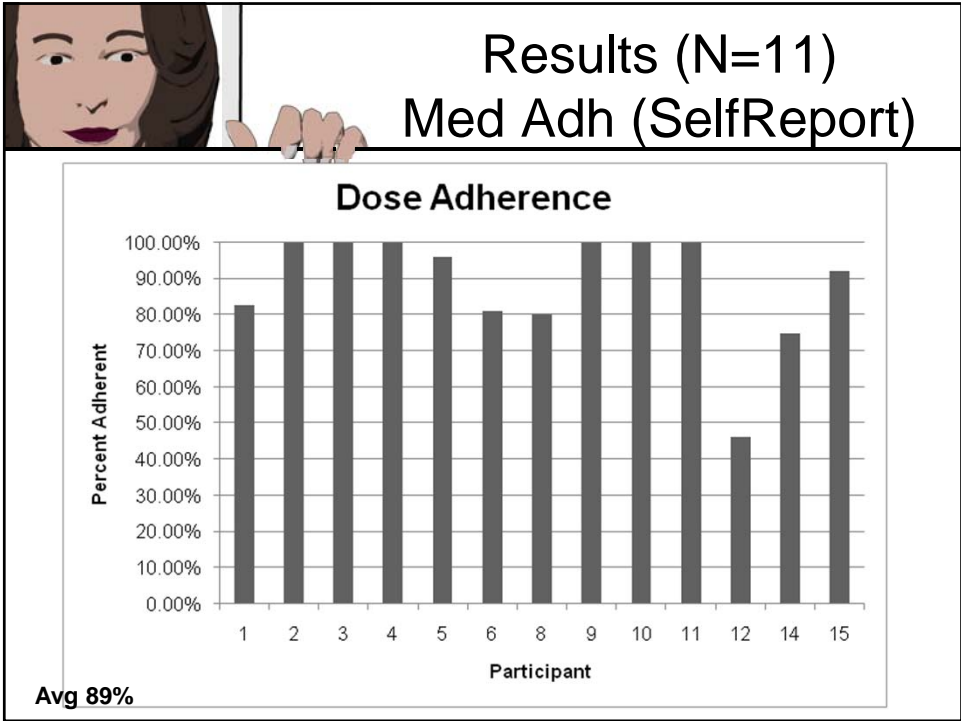




Study Participants

- DSM IVR criteria for Schizophrenia
- 19-58 years old
- On any antipsychotic medication
 - Prolixin, Zyprexa, Abilify, Respirdal, Haldol, Clozaril, Seroquel
- Have had two or more episodes of non-adherence in the 72 hours prior to recruitment
- 20 Ss enrolled.
 - 13 African American F, 3 African American M, 4 Caucasian M
- Self-report data from 15 Ss.
 - 4 w/d before start of study, 1 data file corrupted







Exercise Promotion in Patients with Parkinson's

- Acceptance, Usability & Feasibility study
- 30-day “Elderwalk” intervention
- N=20, 55% female, age 65.6
- 100% retention
- Satisfaction 5.6/7
- Use: 25.4 / 30 days
- Walking adherence 85%
- Significant improvements in gait speed and 6-minute walk test



Question

- What is another name for a quasi-experimental design (no control) with only post-test measures?

The Pretest-Posttest Design

- Pretest administered before and after exposure to experimental treatment.
- Control group used (absent treatment)
- Unlike quasi-experimental designs, this is a true experimental design
- Used to assess the impact of some change on performance

Group 1: Pretest-Treatment-Posttest

Group 2: Pretest Posttest

60

Analysis: a Mixed Design With Pretest-Posttest as the Within-Subjects Factor

Group 1: Pretest-Treatment-Posttest

Group 2: Pretest Posttest

- What is the possible concern with this design?
=> Pretest may affect subject

61

Various solutions to this...

The Solomon Four-Group Design

- Variation on the pretest-posttest design
- Allows you to evaluate the impact of a pretest on posttest performance
- Adds two groups to the basic pretest-posttest design
 - A treatment-posttest group
 - A posttest only group

62

The Solomon Four-Group Design

Group 1:	Pretest-Treatment-Posttest
Group 2:	Pretest Posttest
Group 3:	Treatment-Posttest
Group 4:	Posttest

63

Developmental Designs

- Associate Time/Age with Measure
 - Satisfaction vs. years of using a software product
 - Productivity vs. seniority
- Example: You hypothesize customer satisfaction with your product increases over time. How would you test?

64

Developmental Designs

- *The Cross-Sectional Design*
 - Participants from different age groups are run through a study at the same time
 - Creating "cohort" groups based on participants' ages
 - Allows you to collect developmental data in a short period of time
 - May not be appropriate for studies using widely ranging age groups
 - *Generation effects* may be a problem
 - *E.g., "Flynn effect" – IQ increases between generations*

65

Developmental Designs

- *The Longitudinal Design*
 - A single group of participants is measured several times over some period of time (e.g., months or years)
 - Avoids the generation effect that may plague a cross-sectional study
 - May still have a *cross-generational problem*
 - Results from a longitudinal study on one generation may not generalize to another
 - Problems with the longitudinal design
 - Subject burden, attrition, mortality
 - Multiple observation effects

66

Theory of Rapport

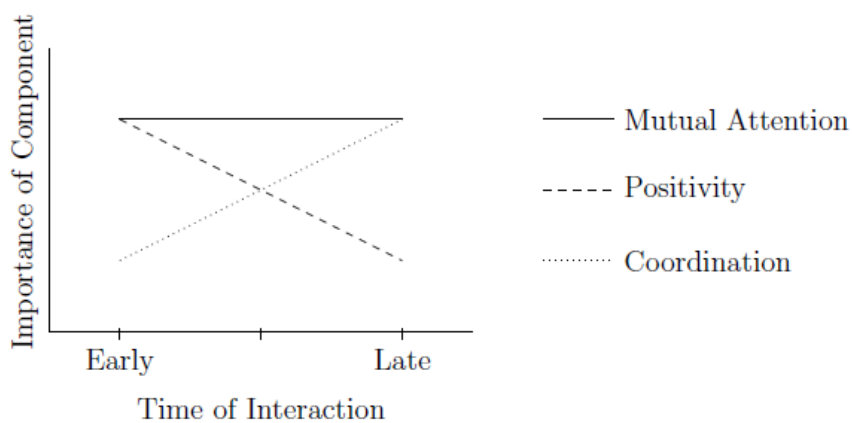


Figure 2.1: Relative importance of different components of rapport (from Tickle-Degnen and Rosenthal, 1990 [162])

Example development study:

Cassell et al, Coordination in Conversation and Rapport, 2007

- Compared direction-giving dialogues between friends and between strangers
- Design.
 - “Manipulated” long-term rapport, visibility, and subsequent route in a 2x2x3 design.
 - Operationalized long-term rapport as a binary, between-subjects variable, with conditions Friends (self-reported as friends for at least one year) and Strangers.

68

Example development study:

Cassell et al, Coordination in Conversation and Rapport, 2007

- Findings: Strangers used more explicit acknowledgments than friends when giving or receiving information

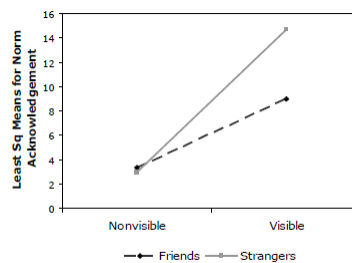


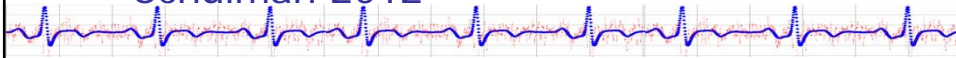
Figure 3: Receiver Acknowledgment by condition

- Kind of study? Issues with design?

69

Criticism

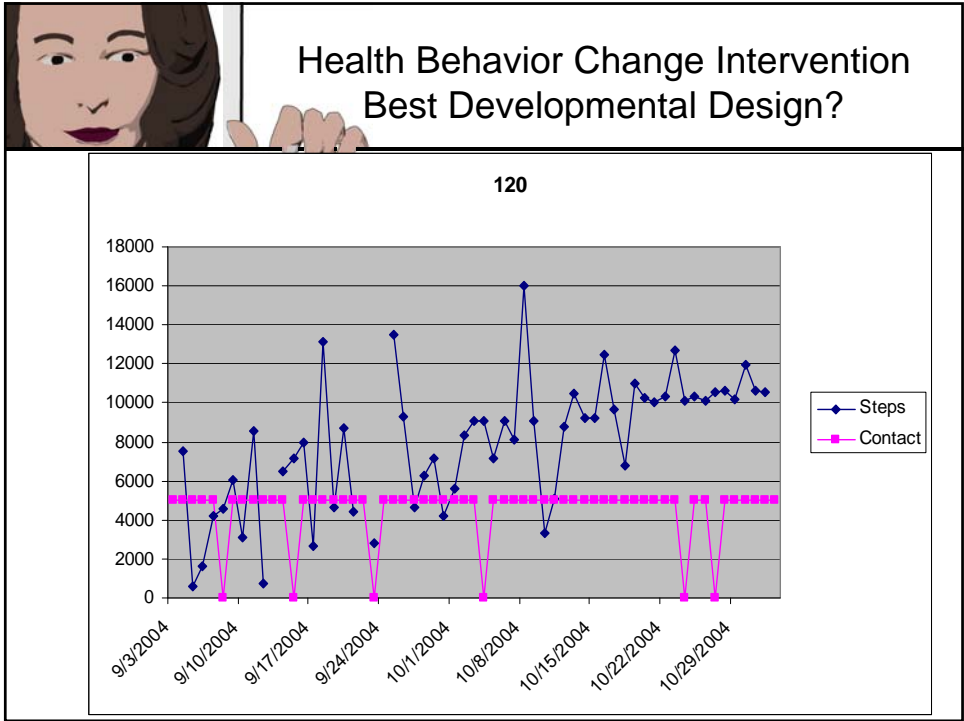
Schulman 2012



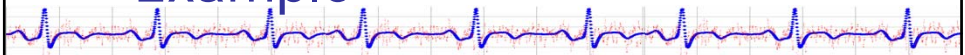
- Cassell et al.'s study, which is cross-sectional, cannot easily distinguish differences between dyads from changes that occur over time
- Their findings...could indicate either that
 - behavioral differences before meeting indicates that a dyad is more or less likely to become friends, or
 - the behaviors of interest change in the hypothesized direction as a function of number of interactions, or
 - the behaviors change as a function of liking/friendship.

70





Developmental Design Example



- You want to assess the correlation between satisfaction with your website and the length of time (months) a user has been visiting.
 - Cross sectional?
 - Longitudinal?

Chapter 12

Using Single-Subject Designs

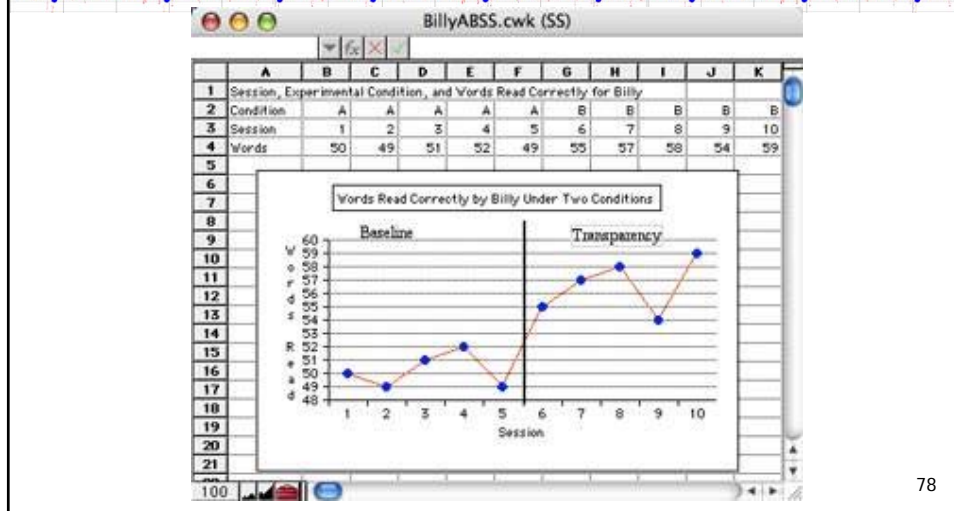
75

Single Subject Designs

- Baseline Design
 - Repeat
 - Change IV
 - Sample DV until stable
- Dynamic Design
 - Continuously vary IV & measure DV response
- Discrete Trials Design
 - Repeat
 - Give randomly assigned IV
 - Measure DV

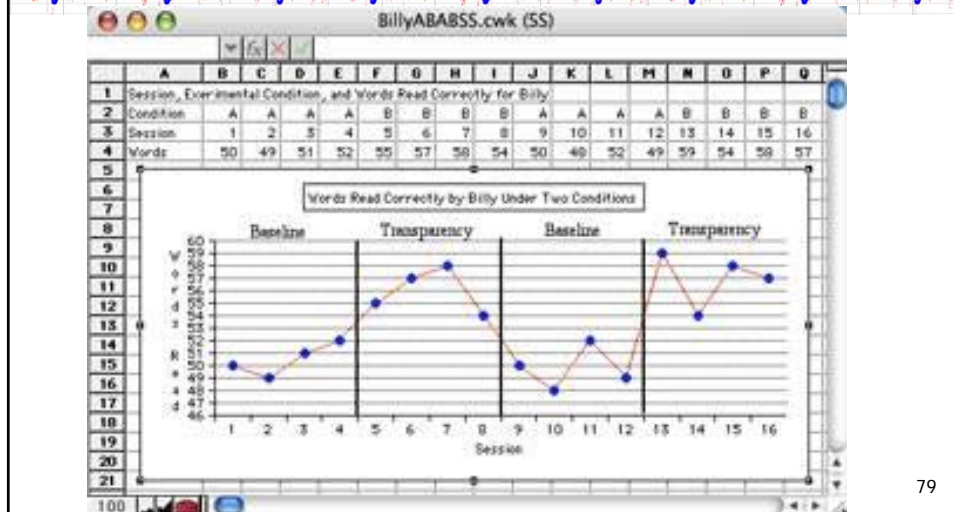
77

Single Subject Baseline Design A-B



78

Single Subject Baseline Design A-B-A-B



79

Example Discrete Trials Design

Longitudinal Affective Computing - Virtual Agents that Respond to User Mood, Ring et al, 2012

- N=21 participants interacted with a virtual exercise coach for 2 months (696 samples)
- Each day, their mood was assessed via self-report.
- The coach would randomly say on of:
 - **Favor:** I was wondering if you'd mind doing me a favor and take a walk before our next session.
 - **Request:** Would you take a walk before our next session.
- Walking was assessed via pedometer.

80

Virtual Coach

Relational Agent Karen

Yes.

No, not much.

I might want to exercise.

But I'm not sure I want to exercise.

If I did exercise, it might be better.

Northeastern
Relational Agents Group

81

Example Discrete Trials Design

Longitudinal Affective Computing - Virtual Agents that Respond to User Mood, Ring et al, 2012

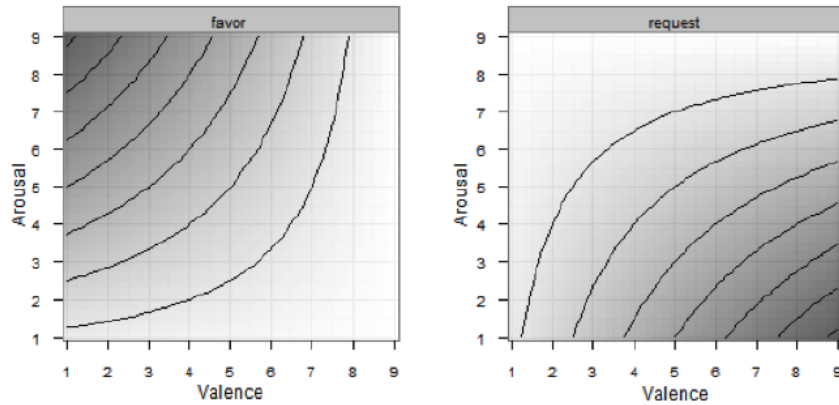
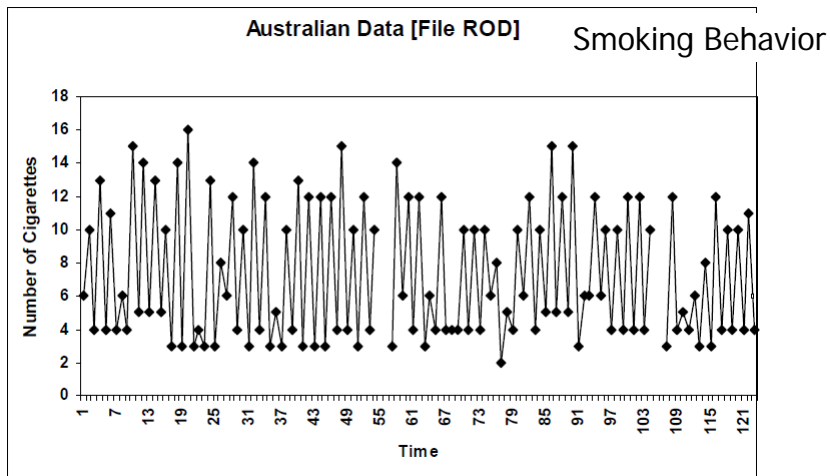


Fig. 1: Change in the number of steps walked based on mood and dialogue manipulation. Darker areas represent where each dialogue had the most positive effect on step count.

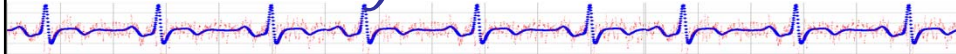
Time Series Analysis

Velicer & Fava, 2003



Why is this science?

Generality?

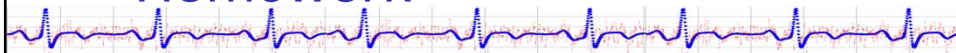


1. Large number of observations
 1. Average out small error variations
2. Rigid control of extraneous variables
3. Focus on powerful effects

4. Still show inter-subject generality
 1. But usually small N (3-6)

84

Homework



- T3
- Regression (Aron Ch 12; [Gilbert](#)).
- Factor Analysis (review B&A Ch 9; [Buys & Brown](#))

98