Quasi-Experimental Design

Ken Koedinger

Based on Trochim Chpt 10

Preliminaries

- Why are you interested in education research?
- Readings
  - Tues: Clark et al
  - Thur: Rittle-Johnson & Koedinger
- Assignment 2 overview
  - Do a Rational Cognitive Task Analysis in a research domain to identify critical, hard-to-learn “knowledge components”
  - Tell us, in a sentence, what is the educational subject matter of a domain in which you might do this assignment

Quasi-Experimental Designs

- 10-1 Nonequivalent-Groups Design
- 10-2 The Regression-Discontinuity Design
- 10-3 Other Quasi-Experimental Designs
  - 10-3a The Proxy Pretest Design
  - 10-3b The Separate Pre-Post Samples Design
  - 10-3c The Double-Pretest Design
  - 10-3d The Switching-Replications Design
  - 10-3e The Nonequivalent Dependent Variables (NEDV) Design
  - 10-3f The Regression Point Displacement (RPD) Design

Nonequivalent Groups Design (NEGD)

N	O	X	O
N	O		O

N = nonequivalently assigned group (not done by random assignment)

O = observation or measure taken at one time for one group

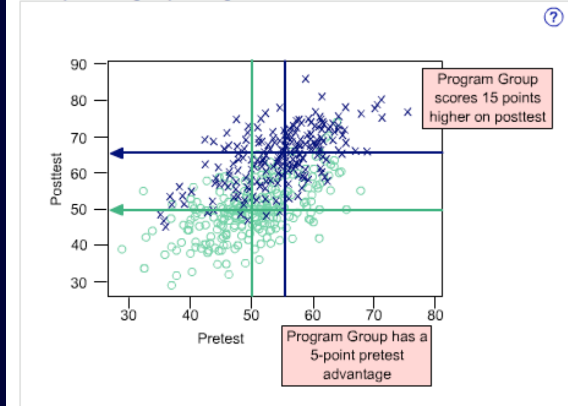
X = a treatment or program administered to one group

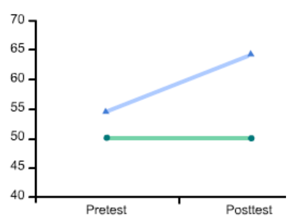
What kind of study?

- Study of Cognitive Tutor Algebra course
  - Treatment:
    - A number of teachers at 3 different schools already been using course during early development
  - Control
    - We asked other teachers at the same schools to give our post-test
  - We used prior test-scores & grades as a pre-test
- Is this a NEGD?
- Or an Archival Proxy Pre-Test design?

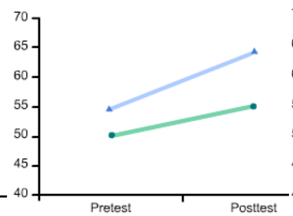
Pre-post scatterplot ("bivariate distribution") of treatment & control students

Figure 10.2. Bivariate distribution for a hypothetical example of a nonequivalent-groups design

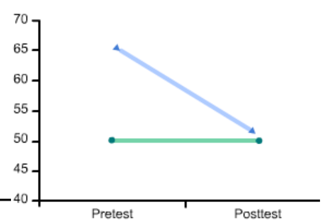




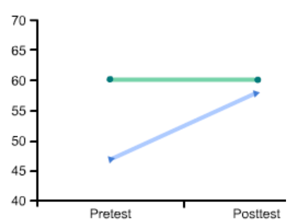
Sel maturation	0
Sel history	+
Sel regression	-
Sel instrumentation	+
Sel testing	+
Sel mortality	+



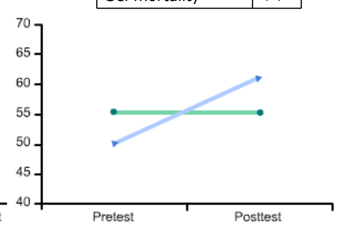
Sel maturation	+
Sel history	+
Sel regression	-



Sel maturation	-
Sel history	+
Sel regression	+
Sel mortality	++



Sel maturation	0
Sel regression	+



Sel maturation	-
Sel regression	-
Sel history …	+

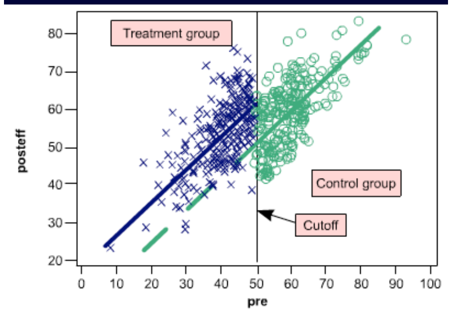
Quasi-Experimental Designs

- 10-1 Nonequivalent-Groups Design
- 10-2 The Regression-Discontinuity Design
- 10-3 Other Quasi-Experimental Designs
  - 10-3a The Proxy Pretest Design
  - 10-3b The Separate Pre-Post Samples Design
  - 10-3c The Double-Pretest Design
  - 10-3d The Switching-Replications Design
  - 10-3e The Nonequivalent Dependent Variables (NEDV) Design
  - 10-3f The Regression Point Displacement (RPD) Design

Regression-Discontinuity Design

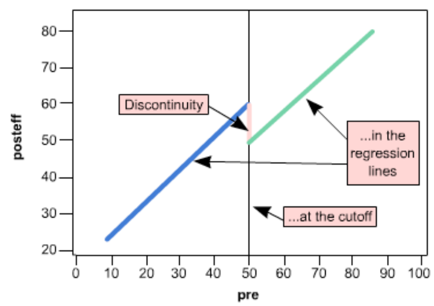
- A pretest-posttest, program-comparison group quasi-experimental design in which a cutoff criterion on the preprogram measure is the method of assignment to group
- C = groups assigned by means of a cutoff score

C	O	X	O
C	O		O



Regression-Discontinuity design with ten-point treatment effect.

If there is a treatment effect, there will be a...



How the Regression-Discontinuity Design got its name.

Example RD design

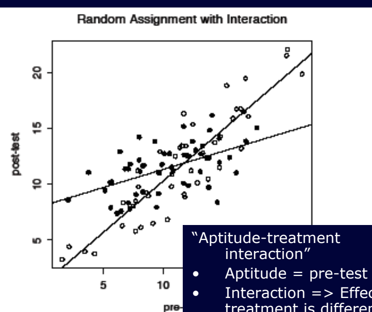
- Does assigning community college students to a remedial math sequence improve their college completion?
- Cut-off: Math placement test at college entrance
- What's a positive result?
- Those just below the cut-off do somewhat better than those just above
- What's a null result?

A Problem for RD

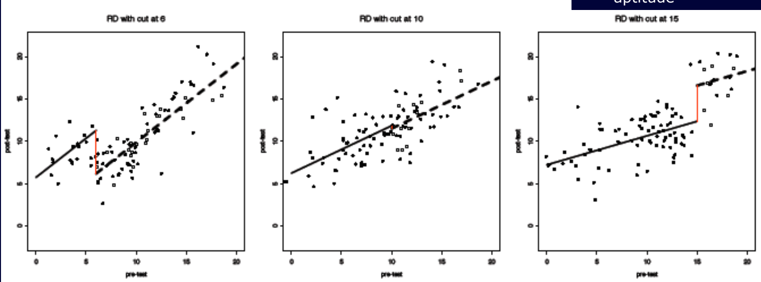
In this example, the poor performing students benefit from the intervention while the high performing students are adversely affected.

This is a bad case for a Regression Discontinuity design, because now, the *treatment effect* is also dependent on the cut point.

Notice that with the cut at 15, we actually see a negative treatment effect, even though the program benefitted some students.



"Aptitude-treatment interaction"
- Aptitude = pre-test level
- Interaction => Effect of treatment is different for students with low aptitude (pre-test) than for for high aptitude



Questions about other Quasi-Experimental Designs?

- 10-1 Nonequivalent-Groups Design
- 10-2 The Regression-Discontinuity Design
- 10-3 Other Quasi-Experimental Designs
  - 10-3a The Proxy Pretest Design
  - 10-3b The Separate Pre-Post Samples Design
  - 10-3c The Double-Pretest Design
  - 10-3d The Switching-Replications Design
  - 10-3e The Nonequivalent Dependent Variables (NEDV) Design
  - 10-3f The Regression Point Displacement (RPD) Design

Nonequivalent dependent variables (NEDV) design

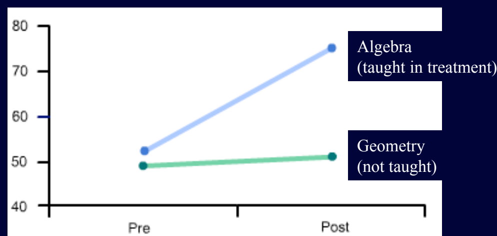
N $\frac{O_1}{O_2}$ X $\frac{O_1}{O_2}$

- A single-group pre-post quasi-experimental design with two outcome measures, where only one measure is theoretically predicted to be affected by the treatment and the other is not

Questions:

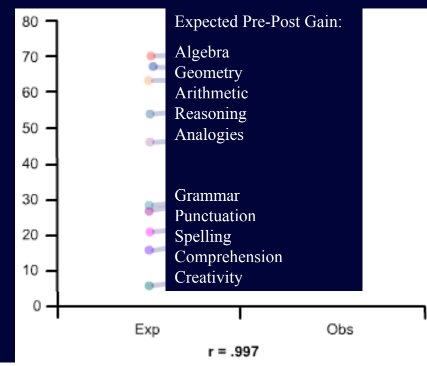
What reduces history, maturation threats?

What can you conclude?



Pattern-Matching NEDV Design

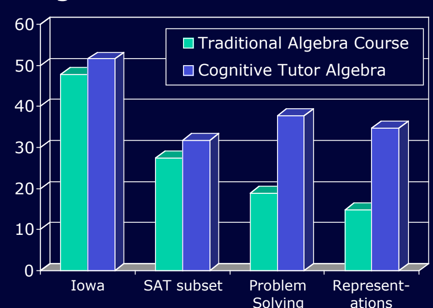
- Again: Algebra taught as treatment
- This time many other tests & a prediction of the order



Cognitive Psychology Goes to School

Cognitive Tutor Algebra course

- Based on ACT-R theory & computational models of student problem solving
- Wide-spread use & evaluation
  - 500,000 students use daily at 2600 schools
  - 8 of 10 full-year field studies demonstrate significantly better student learning



Koedinger, Anderson, Hadley, & Mark (1997). Intelligent tutoring goes to school in the big city.

Carnegie Learning™
THE COGNITIVE TUTOR® COMPANY

Minimizing Threats to Validity

Can you think of examples of these in studies you like, don't like, have been involved?

- By argument
- By measurement or observation
- By design
- By analysis
- By preventive action

Criteria fo Good Design

- Theory-grounded
- Situational
  - Example: Two cut-off design with randomization in between
- Feasible
- Redundant
  - Ex: Adding nonequivalent dependent variables (or “pattern matching” to an existing design

Summary

- Constraints aside, experimental designs are best
- Practicality, ethnics, stage of research may lead you to a quasi-experimental design
- Tips
  - *Know the threats & know your options*
  - Use theory … to help address threats
  - *Don’t over-pitch your results!!*
- Pattern matching … lead in to Cognitive Task Analysis
  - What kinds of learner changes can/should you expect from your planned treatment?

Extras

Example of a Regression-Discontinuity Study Proposal

Research Questions, Hypotheses and Design

The research team has designed an evidence-based evaluation of the effects of the project. The effects for which we designed this effort were driven by two fundamental research questions:

1. What is the effect on middle school students’ math and science CSAP test results of their teachers’ participation in the RM-MSMSP professional learning programs?
2. With what level of complexity and fidelity do teachers trained in the RM-MSMSP professional learning programs implement what they have learned in those programs?

To address the first evaluation question above, a regression-discontinuity design will be implemented. To address the second question a standardized interview protocol called the Levels of Use protocol, drawn from over three decades of theoretical research through the Concerns Based Adoption Model (CBAM) will be used.

<u>Question 1: Regression-Discontinuity Design</u>

The regression-discontinuity (RD) design is a relative newcomer in quasi-experimental design literature, first explicated in detail by Trochim (1984). RD is basically a two-group, pretest posttest design, but its uniqueness stems from how participants are assigned to the two groups. Trochim (2001) states:

> In RD designs, participants are assigned to program or comparison groups solely on the basis of a cutoff score on a pre-program measure….This cutoff criterion implies a major advantage of RD designs: they are appropriate *when you want to target a program or treatment to those who most need or deserve it*. (p. 221, emphasis added)

RD designs are considered in the family of pretest-posttest control group designs, and when well implemented, are considered equal to randomized experiments in their ability to support causal inferences by controlling threats to internal validity. This particular research design appears well-suited to evaluating the effects of the RM-MSMSP professional learning program because the PL program specifically targets middle school teachers most in need of training.

Figures 1 and 2 on the following page give a conceptual idea of how this design will function in assessing the effects of this PL intervention.

Suppose the scatterplot below reflects the distribution of average CSAP 6th grade elementary school scores (horizontal axis – "pre") plotted against average CSAP 7th or 8th grade (depending on the level of the teacher) for all the middle level math teachers in the RM-MSMSP II. Thus, each data point in the scatterplot below represents a teacher, and the value of the data point reflects the bivariate value associated with the 6th grade CSAP scaled score average of each teacher's students on the CSAP back when they were in elementary school with the classroom 7th grade average while they were students in that teacher's middle school math class. Thus, this scatterplot in Figure 1 reflects how this distribution might look with no PL intervention.

Assume as well, that a cutoff point is established by the RM-MSMSP project that identifies those teachers whose students' 7th grade CSAP scores are in the lowest half of the distribution (see cutoff point in Figure 1). These would be the teachers who will be recruited and served by the PL activities of the project.

Figure 1: Pre-RM-MSMSP Professional Development Scatterplot

Now look at the scatterplot on the following page. The visual display of "discontinuity" at the cutoff point reflects similar post-treatment measurement of pre-PL and post-PL CSAP scores and assumes an effective PL "treatment" as well. Visually, you can see that the "treated" teachers' average students' CSAP scores are elevated to a level roughly equal to those of the "untreated" control group. The analysis of this "discontinuity" is completed using regression analysis and modeling, and if all of the assumptions underlying the RD design are satisfied, the analysis can produce an accurate and interpretable estimate of PL effect.

Figure 2: Post-RM-MSMSP Professional Development Scatterplot (assumes effective treatment)

<u>Assumptions Underlying the Use of this RD Design for the RM-MSMSP Project</u>

Several key assumptions must be satisfied in order for this research design to provide a scientifically valid estimate of program effect. These assumptions interact with each other such that it is impossible to know the answer to them in advance; the data must be collected, analyzed, and modeled iteratively over the life of the project to know just how well it will work.

1. There must be sufficient numbers of teachers (each individual is a data point in the scatterplots above) served by the project to generate satisfactory variability in the outcome measures and sufficient power to detect PL program effect.
2. The selection of "high need" teachers to receive PL treatment must be aligned with their "pre-postnull" placement below the cutoff point and not for political or convenience reasons. While some teachers whose "pre-postnull" values fall above the cutoff point can be served by the project, the data from these teachers cannot be included in either the treatment or control groups in the analysis.
3. Variability in scoring on the outcome measure (CSAP math overall score) must be highly sensitive to the PL services in the treatment. Put another way, if a "high need" teacher learns the PL content well and implements it with fidelity (see assumption 4), the CSAP scores of her students should rise noticeably.
4. The teachers must implement the treatment package with some measure of fidelity as they teach their classes after the PL treatment.

Probabilistic equivalence

- Two groups are "probabilistically equivalent" if when repeatedly measured, one is not consistently better than the other
  - Will such groups obtain the exact same average score on a particular measure?
  - Can a statistical difference between probabilistically equivalent groups be found?

Figure 9.3. Probabilistic equivalence does not mean that two randomly selected groups will obtain the exact same average score