Quantitative Research Methods

Quasi-Experimental Design

Nikolaj Thallmayer (01608292)

Quasi-Experimental design differs from true experimental design in its non-random group assignment

"True" experimental designs

Random assignment

- Gold standard for quantitative research
- Truly random assignment can be challenging or ethically concerning

Quasi-experimental designs

Non-random assignment

- Often easier to implement
- Challenges with internal validity

Focus of presentation

Today's agenda

- Nonequivalent Groups Design
- Regression-Discontinuity Design
- Other Quasi-Experimental Designs

Today's agenda

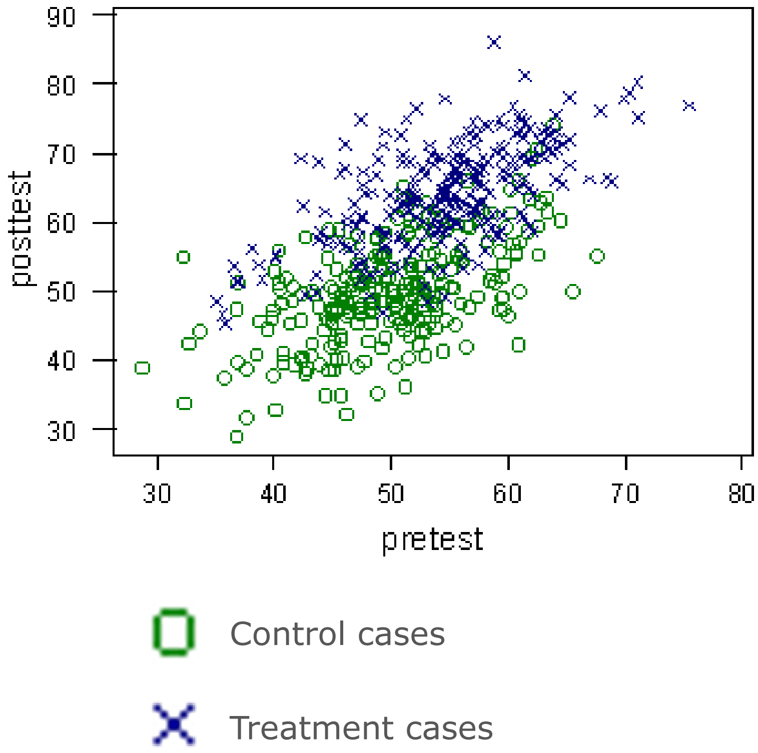
- Nonequivalent Groups Design
- Regression-Discontinuity Design
- Other Quasi-Experimental Designs

Nonequivalent groups design uses already existing groups for treatment and control

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

- Non-equivalent groups design (NEGD) is structured like **pretest-posttest randomized experiment without random group assignment**
- **Treatment and control group are formed based on previously existing groups**, i.e., in a school using two comparable classrooms or using two similar communities
- Even if similar groups are selected, they are unlikely as similar as if chosen at random -> therefore **"nonequivalent groups"**
- Prior differences in groups raises **internal validity threat of selection**, meaning that differences of groups may affect outcome of study

Bivariate distribution reveals issue of selection threat to internal validity

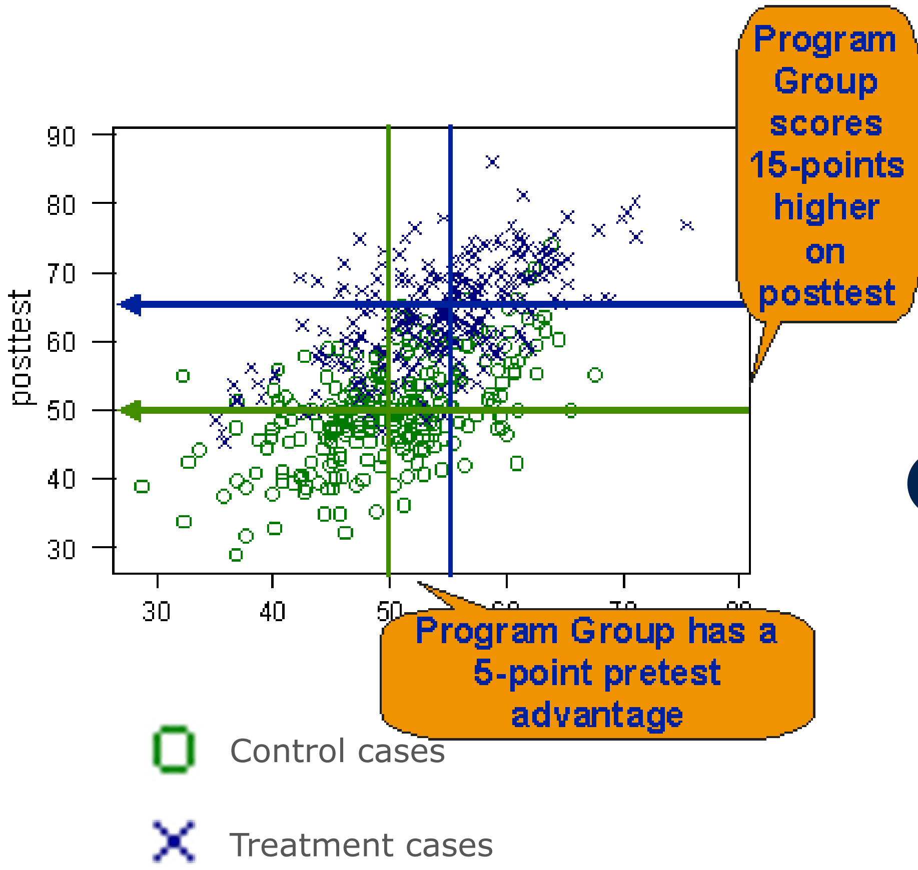


Looking at hypothetical bivariate distribution of NEGD reveals two insights:

1. **Treatment cases clearly score higher on posttest** than control cases (y-axis)
2. On average, **treatment cases performed slightly better on pretest already** (x-axis)

Hard to determine, whether difference is partially or fully caused by treatment or if initial advantage led to better outcomes (selection threat)

Bivariate distribution reveals issue of selection threat to internal validity



Looking at hypothetical bivariate distribution of NEGD reveals two insights:

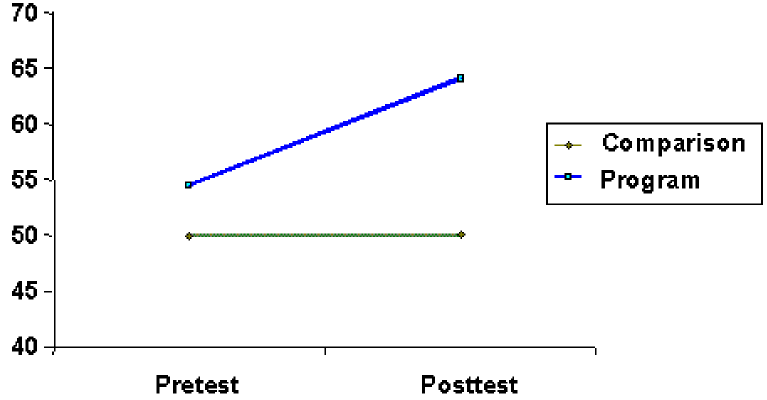
1. **Treatment cases clearly score higher on posttest** than control cases (y-axis)
2. On average, **treatment cases performed slightly better on pretest already** (x-axis)

Hard to determine, whether difference is partially or fully caused by treatment or if initial advantage led to better outcomes (selection threat)

Most important types of selection threats to internal validity

Selection-maturation threat	Implies that groups are maturing at different rates, which creates illusion of program effect
Selection-history threat	Occurrence of event that only affected one group or that only occurred for one group
Selection-regression	Implies that observed differences stem from sample mean of group regressing to population mean over time

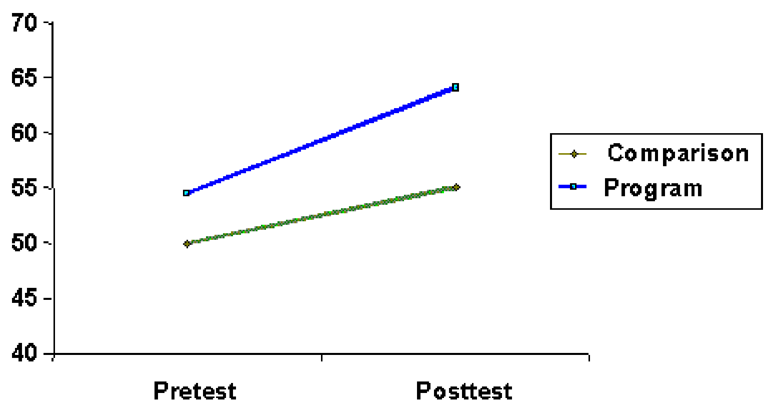
Exemplary graph #1: No clear evidence for effectiveness can be inferred



Different representation of previous graph – what possible threats apply?

- **Selection-maturation threat**
  - Unlikely, as no maturation at all can be observed in comparison group
- **Selection-history threat**
  - Plausible for case due to differences in development
- **Selection-regression**
  - Unlikely, as upwards trend in program group implies they were below population mean – if regression to the mean was cause we would see it in comparison group too

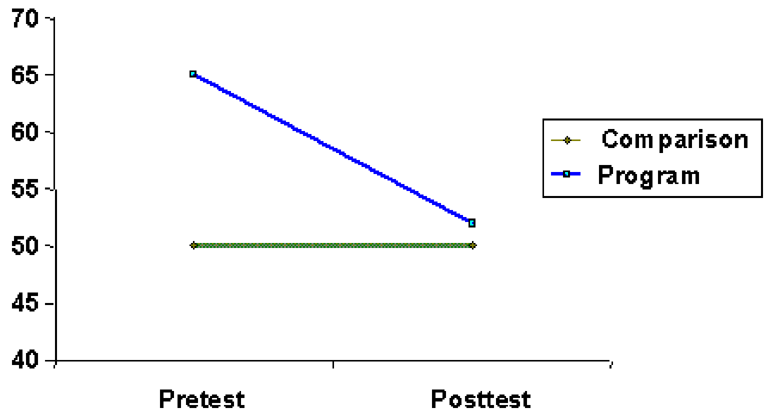
Exemplary graph #2: No clear evidence for effectiveness can be inferred

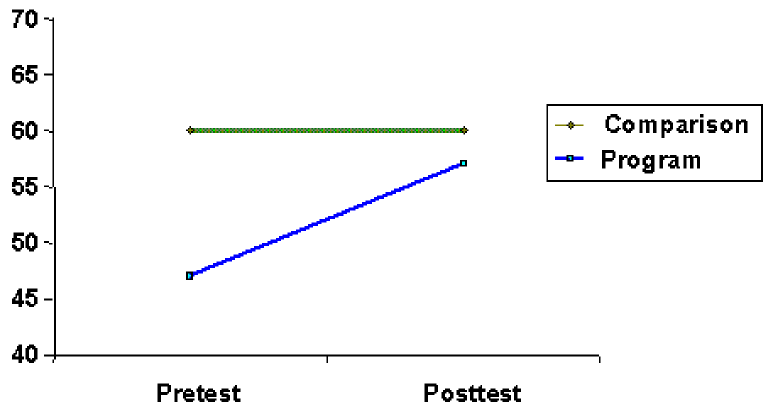


What possible threats apply?

- **Selection-maturation threat**
  - Likely, as previous differences might have arisen from different maturation rates, which further increased
- **Selection-history threat**
  - Also likely, if groups react differently to some event, which causes different developments
- **Selection-regression**
  - Unlikely, as upwards trend in program group implies they were below population mean – if regression to the mean was cause we would see it in comparison even more strongly

Exemplary graph #3 & #4: No clear evidence for effectiveness can be inferred

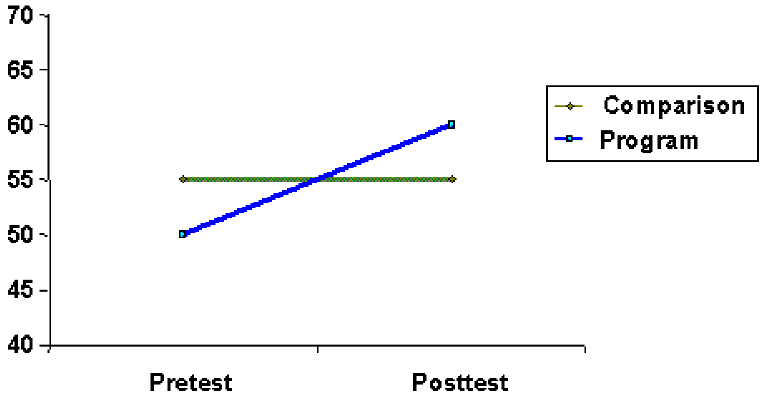




What possible threats apply?

- **Selection-maturation threat**
  - Unlikely, as maturation in program group with no maturation in comparison group seems hard to explain
- **Selection-regression**
  - Very likely -> program group was very high/low on pre-test and simply regressing to population mean, whereas comparison group was already at population mean levels

Exemplary graph #5: Cross-over pattern implies strongest evidence for effectiveness



What possible threats apply?

- **Selection-maturation threat**
  - Unlikely, as one would need to argue that program group matured from below average to above average with no maturation in comparison group
- **Selection-history threat**
  - Unlikely, as program group started out worse off and improved beyond comp
- **Selection-regression**
  - Unlikely, as program group would need to approach comparison group, but not cross over, while no movement in comparison group

Cross-over pattern implies strong evidence for treatment effect, however one shouldn't structure research intentionally in such way (treating disadvantaged group and hoping it even outperforms comparison group)

Today's agenda

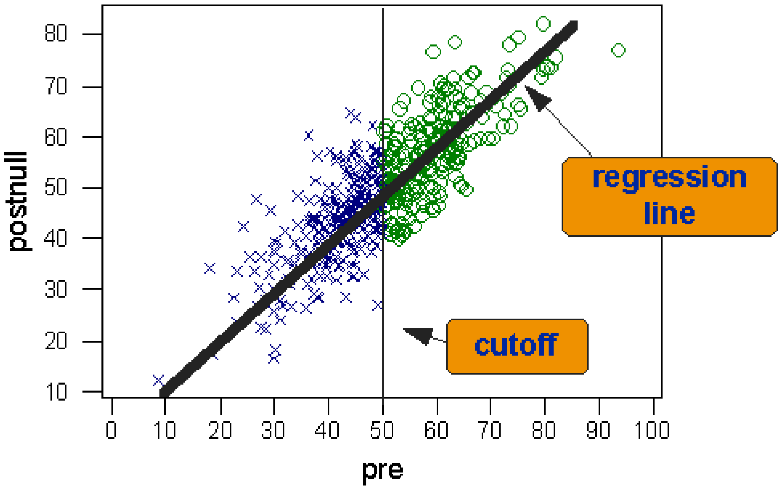
- Nonequivalent Groups Design
- Regression-Discontinuity Design
- Other Quasi-Experimental Designs

Regression-Discontinuity Design (RD)

- **Regression-Discontinuity Design (RD design)** can refer to several design variations - simplest traditional form **represents pretest-posttest program-comparison group strategy**
- **Assignment to groups is based on a cutoff score** on a pre-program measure
- Biggest advantage is that it allows to **assign treatment to those who need it most**
- **Internal validity comparable to randomized experiments**, however statistical power lower by a factor of ~2.75[1]
- **RD design not yet frequently implemented**, as its rather novel (first introduced in mid 1970s), not flexible (single quantitative measure determines group assignment), and **counterintuitive as it maximizes group differences** instead of trying to have similar groups

1. E.g., 100 participants required for significance in randomized experiment, then 275 needed for RD

Bivariate distribution without any intervention shows continuous distribution

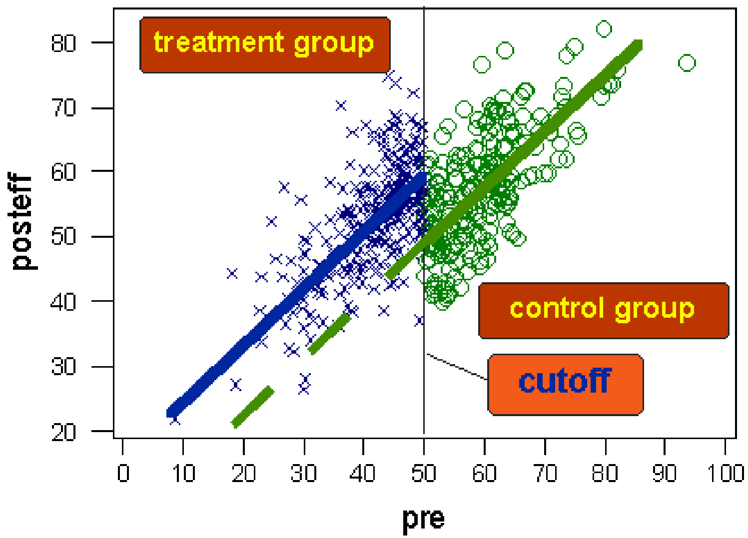


Control cases

Treatment cases

- Example of cases where composite health score (100= healthy, 0 = not healthy) is used as pre-test and post-test
- Chart shows bivariate distribution of cases **without any program intervention**
- On average, being healthy on pretest means being healthy on posttest
- Smooth regression line can be put on data points

Bivariate distribution with intervention shows discontinuity in regression line

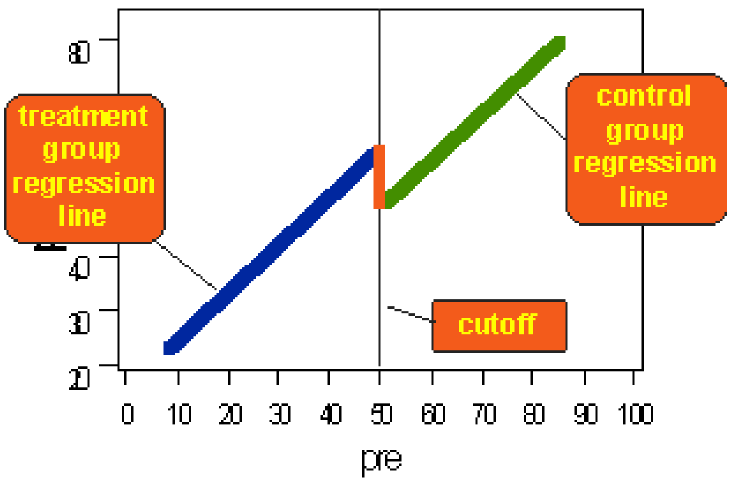


Control cases

Treatment cases

- Hypothetical distribution, assuming that treatment has a constant positive effect
- Green dotted line implies expected regression line if treatment had no effect
- Under presence of a treatment effect, there will be disruption (or discontinuation) in regression line -> therefore Regression Discontinuity Design

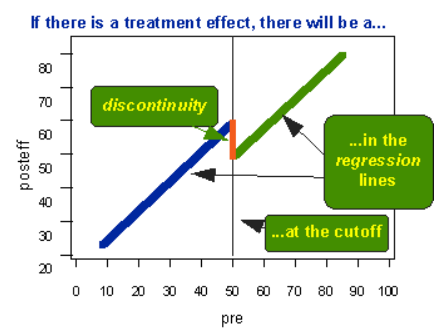
Bivariate distribution with intervention shows discontinuity in regression line



Control cases

Treatment cases

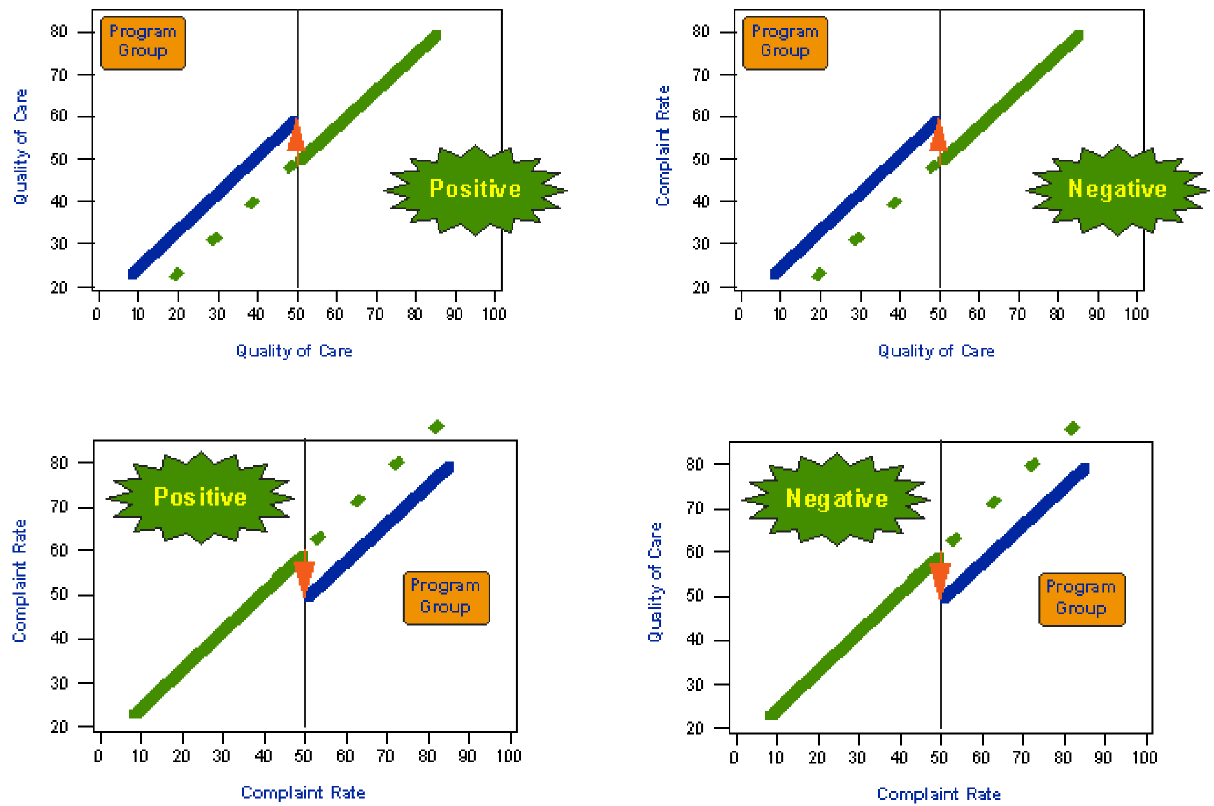
- Hypothetical distribution, assuming that treatment has a constant positive effect
- Green dotted line implies expected regression line if treatment had no effect
- Under presence of a treatment effect, there will be disruption (or discontinuation) in regression line -> therefore Regression Discontinuity Design



Interpretation of results depends on nature of assignment and outcome variable

Example:

- Hospital staff to receive treatment to improve quality of care
- Two measures available: Quality of care (judged by supervisor) and complaint rate (number of complaints relative to patients)

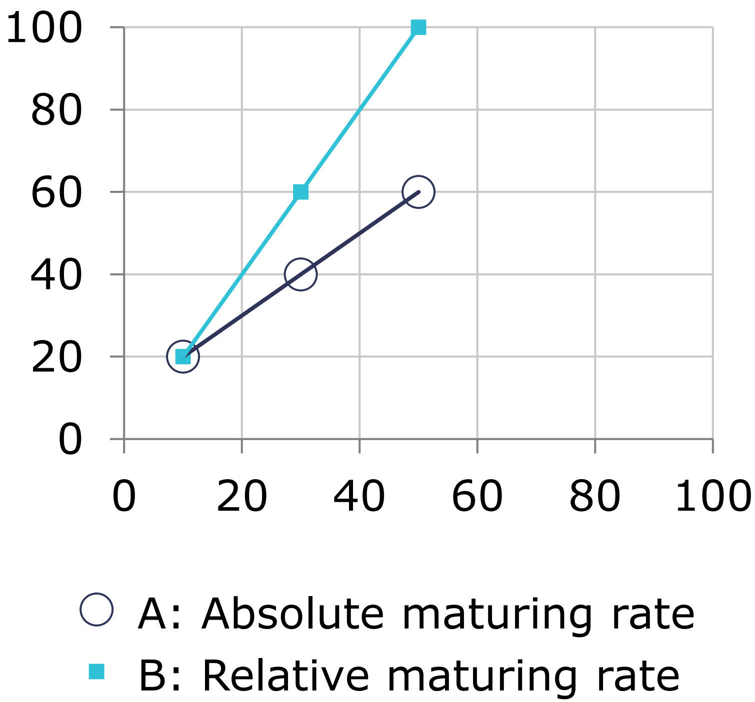


RD design requires a continuous pre-program variable and a defined cutoff value

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

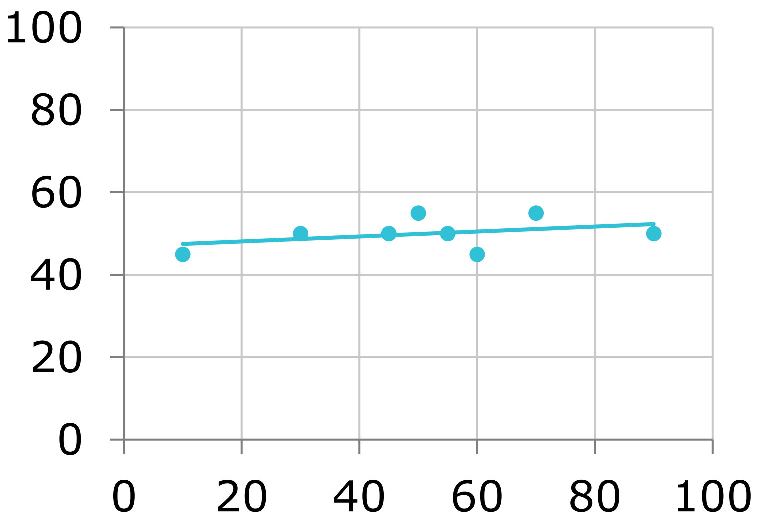
- RD design requires a continuous quantitative pre-program measure (which can but doesn't have to be different from the pretest measure)
- All persons on one side of cutoff are assigned to treatment, all persons on other side are assigned to control
- Selection of cutoff
  - **Based on project resources**: I.e., if 25 people can be treated, cutoff can be set such that 25 fall into treatment and rest into control
  - **Based on expert views**: I.e., if expert believe health score of 50 or lower indicates treatment, then 50 can be set as cutoff
- In other designs, we assume or provide evidence that treatment and control group are equivalent and that differences can be attributed to program
- In RD design we instead assume that in absence of program pre-post-relationship is equivalent for both groups – this assumes the following:
  - No spurious discontinuity that coincides with cutoff point
  - Correct modelling of the pre-post-relationship
  - **Introduces possibility of selection threat to internal validity**

Selection maturation threat to internal validity



- Selection maturity threat implies that different groups could have different rates of maturing
- I.e., group A shows a constant absolute maturing rate of +10 from pretest to posttest
- I.e., group B shows constant relative maturing rate of x2, which leads to differing absolute changes from pre to post, depending on the pre-level
- Both scenarios can be captured by a regression line, so lang that differences in maturing rates do not coincide with the cutoff point

Selection regression threat to internal validity



- Regression to the mean would imply that sampled cases approach the population mean over time
- The expected regression to the mean from pretest to posttest score is exactly what is modelled by the regression line itself
- As we do not expect a difference in regression to the mean between any two groups, there should not be a discontinuity in the regression line without intervention

RD design shows strong internal validity with some practical caveats to be considered

- RD design by itself is not susceptible to selection threats and shows in theory the same internal validity as a randomized experiment
- There is some caveats that can practically hinder internal validity:
  - **Participants can manipulate their pre-program measure**, making it not random – i.e., if using students' grades as pre-program measure and 50% as the cutoff, teachers might be giving students slightly below 50% a mercy pass lifting them above 50% and thus changing their group assignment
  - **There is another treatment coinciding with the cutoff of the actual treatment** – i.e., studying alcohol's effects on mental health using legal drinking age as cutoff coincides with legal gambling age, potentially contaminating the results
  - **Correct modelling of the pretest-posttest relationship** – incorrect modelling of non-linear relationships could be mistaken as discontinuity

Today's agenda

- Nonequivalent Groups Design
- Regression-Discontinuity Design
- Other Quasi-Experimental Designs

Proxy Pretest Design

- In general same setup as normal pretest-posttest design
- However, no pretest conducted, instead proxy measure used instead of pretest:
  - **Recollection proxy pretest**: Participants own assessment, what they believe their pretest score might have been
  - **Archived proxy pretest**: Using measure from before treatment that is readily available and was collected independent of study
- Proxy pretest design should not be actively pursued, but is rather a backup if study has begun and no proper pretest was conducted

	Participants	Pretest	Treatment	Posttest
Treatment	Group 1	-	X	Test
Comparison	Group 2	-	-	Test

Separate Pre-Post Sample Design

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

- Two non-equivalent groups, where pretest and posttest are conducted on different subgroups each
- This setup can occur, if you can not be sure to track the same participants within each group at pretest and posttest
- I.e., trying to improve customer satisfaction: one organization is getting the treatment, another one is serving as control; customers that are surveyed during pretest will most likely not be the same as during posttest as different customers will have issues
- Variation of experiment design includes random subsampling within group – this doesn't change issues of design however

	Participants	Pretest	Treatment	Posttest
Treatment	Group 1a	Test	-	-
	Group 1b	-	X	Test
Comparison	Group 2a	Test	-	-
	Group 2b	-		Test

Double Pretest Design

- Similar to classical Nonequivalent Groups Design, however two pretests are conducted
- This eliminates possible selection maturation biases, as differences in maturation could be observed between the two pretests
- Often referred to as "dry run" quasi-experimental design, as it simulates the null case
- Design is very strong in internal validity

	Participants	Pretest 1	Pretest 2	Treatment	Posttest
Treatment	Group 1	Test	Test	X	Test
Comparison	Group 2	Test	Test	-	Test

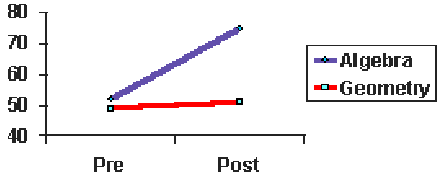
Switching Replications Design

- Design involves two separate non-random groups, where both are treated consecutively with posttests in between
- Strong in internal validity, and strong in external validity due to two separate implementations of treatment
- Ethically strong, as all participants receive treatment at some point

	Participants	Pretest 1	Treatment 1	Posttest 1	Treatment 2	Posttest 2
Treatment	Group 1	Test	X	Test	-	Test
Comparison	Group 2	Test	-	Test	X	Test

Nonequivalent Dependent Variables (NEDV) Design

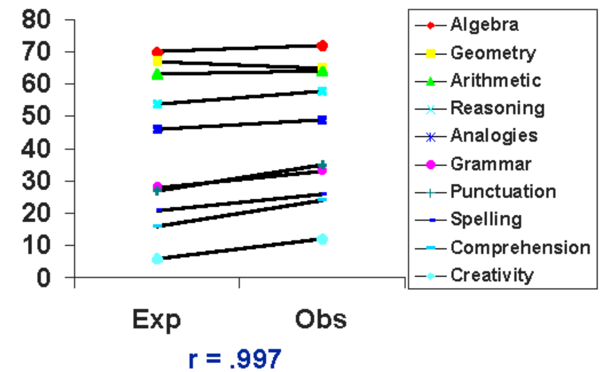
- Only treatment group (no comparison group) which is evaluated based on two different variables in pretest and posttest
- Treatment is supposed to affect one variable, but not the other variables
- Other variables serve as "control", to capture any selection threats (i.e., history, maturation, etc.)
- Key is that control variables are similar enough, so that selection threats realize identical to target variable, but not so similar that treatment affects them
- E.g., program to improve algebra score, which is controlled with geometry scores
- Overall design is rather weak in internal validity



	Participants	Pretest	Treatment	Posttest
Treatment	Group 1	Test 1 Test 2	X	Test 1 Test 2
Comparison	-	-	-	-

Pattern Matching NEDV Design

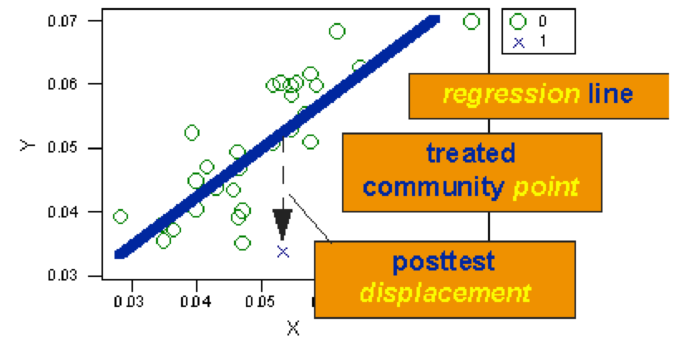
- Similar to previous design, however many more test metrics are being used
- Ex ante the researcher specifies what the expected impact of the treatment is on each variable
- Matching patterns between expectation of observation is strong evidence for effect, as only other event with exactly same impact as expected could cause results
- The more variables the better internal validity, but also harder to find expected patterns in data



Treatment	Group 1	Test 1 Test 2 Test [...]	X	Test 1 Test 2 Test [...]
Comparison	-	-	-	-
	Participants	Pretest	Treatment	Posttest

Regression Point Displacement (RPD) Design

- Useful when implementation of treatment is expensive and thus only a single unit can be treated
- Design is enhanced by leveraging a larger number of comparison units
- I.e., one community receives AIDS education program; before and after HIV rates for all communities in the state are captured -> posttest displacement of treated community is observed
- Useful, when many control case units are available and routine measurement are conducted



	Participants	Pretest	Treatment	Posttest
Treatment	Unit 1	Test	X	Test
Comparison	Unit 2 […] Unit X	Test	-	Test

Today's agenda

- ✓ Nonequivalent Groups Design
- ✓ Regression-Discontinuity Design
- ✓ Other Quasi-Experimental Designs

Summary

Nonequivalent Groups Design

- Uses **pre-existing groups** for control and comparison
- Relatively **easy to implement** and allows for some discretion over group selection
- Multiple **threats to internal validity**

Regression-Discontinuity Design

- **Group assignment based on cutoff value** of pre-program metric
- Theoretically **strong internal validity**, however lower statistical power than random experiments
- Practically a few **challenges to ensure internal validity**
- **Ethically strong** as it allows treatment of those who need it most

Other Quasi-Experimental Designs

- **Switching Replications Design**: Strong internal validity, ethically strong, however double effort for treatment
- **Pattern Matching NEDV Design**: Theoretically strong internal validity, but difficult to practically achieve
- **Regression Point Displacement Design**: Relatively strong internal validity, useful when treatment is expensive (i.e. community research)
- **Double Pretest Design**: Strong internal validity
- Proxy Pretest Design, Separate Pre-Post Sample Design,, Nonequivalent Dependent Variables Design: Not desirable but sometimes useful as least bad option

Now some quiz!

https://create.kahoot.it/details/511c171c-520d-4787-a9bd-8e3636de4b5d