Fundamental Concepts of Quantitative Impact Assessment: Module 4

Aneta Bonikowska, PhD
Social Analysis and Modelling Division

March 19, 2024

Module 4

Common approaches to QIA

Module 4 contents

- Self-assessment
- Difference estimators
- Discontinuity estimators
- Instrumental variables
- Matching
- Regression

Example:

- A regional development agency (RDA) funds the acquisition of specific machinery and equipment
- The RDA is interested in the sales of the firms. Other outcomes of interest could be the number of employees or the propensity to export. It could also be the *growth* rather than *levels*.
- The RDA wants to know whether the recipients of the program (A) performed better than their counterparts (B)

Self-assessment

- Program participants are asked how the program impacted them
- “How much higher are your sales this year as a result of the program?”

Self-assessment

- Pros:
  - Always available as long as a survey can be funded
    - More rigorous approaches are not always possible
  - Easiest to understand

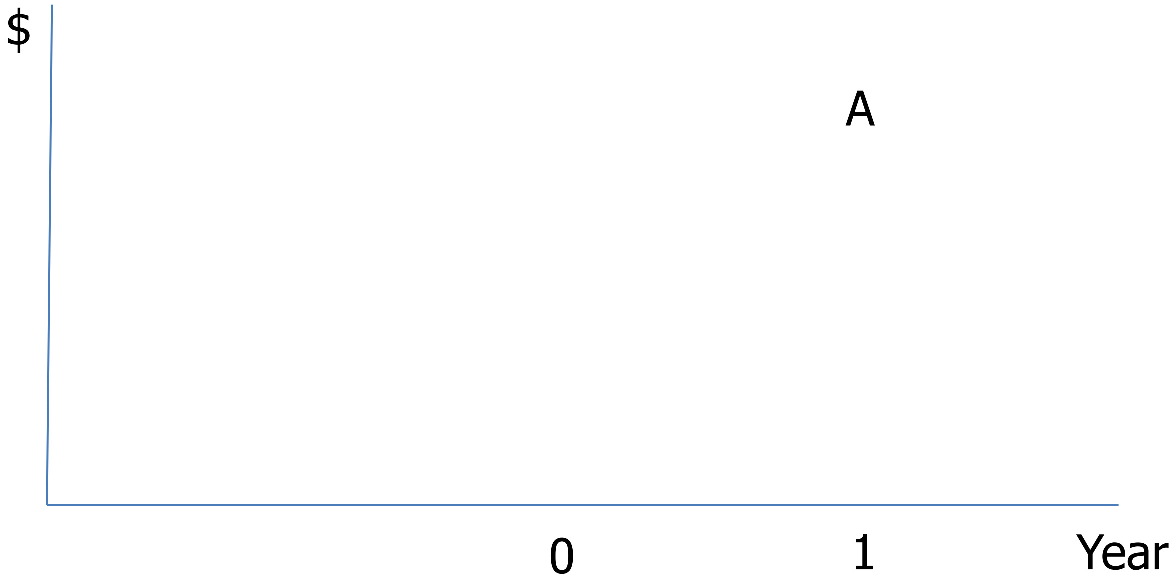
- Cons:
  - Participants may not know how their sales would have evolved <u>had they not taken the program</u>
  - Placebo effects are well documented in various settings

Difference estimators

- Method #1: No differences
- Simply look at the sales of program participants after taking the program

Difference estimators

- Method #1: No differences



One year after the program was implemented, participating firms had $A worth of sales on average.

$A is the actual outcome of participating firms. Outcome, not impact.

No comparison → cannot tell if the program had an impact.

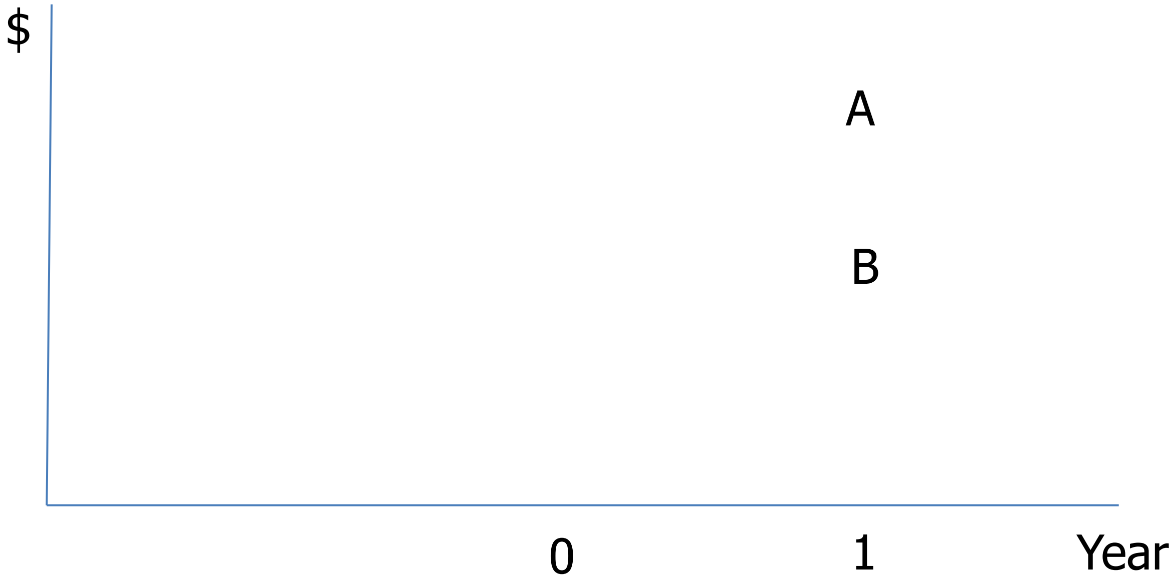
Difference estimators

- Method #2: First difference after program implementation

- Compare the sales of program participants and (somewhat) similar non-participants after the program was launched

Difference estimators

- Method #2: First difference after program implementation



Non-participant firms had $B worth of sales on average

Estimated impact = A - B

What if the sales of program participants fell, rather than increase, after implementation?

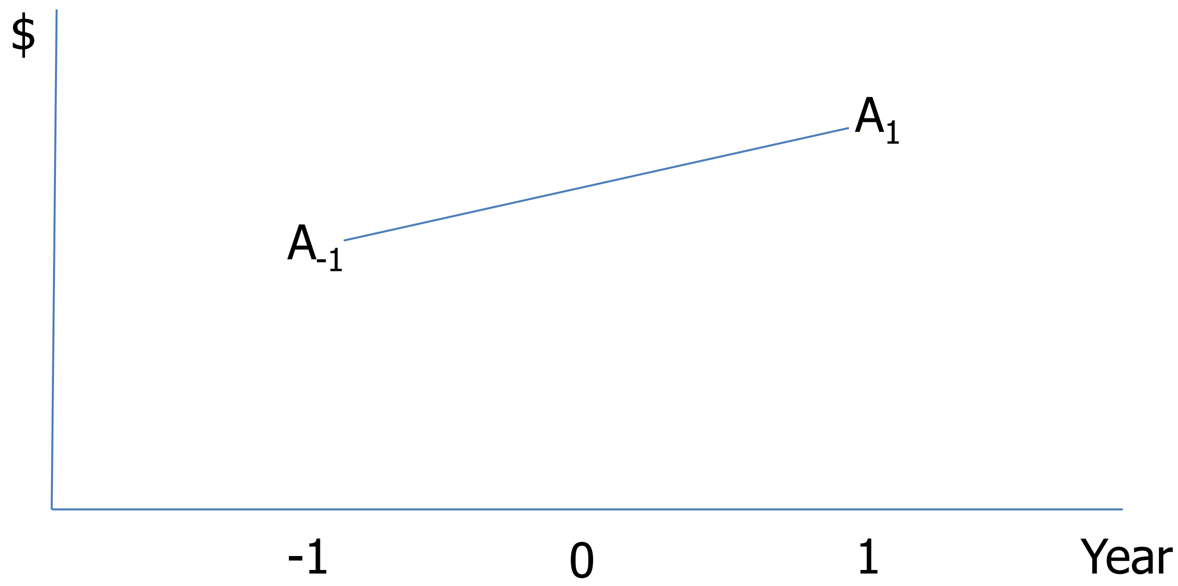
Difference estimators

- Method #3: First difference before and after program implementation (pre/post)

- Compare sales of program participants before and after the program was launched

Difference estimators

- Method #3: First difference before and after program implementation



Estimated impact = $A_1 - A_{-1}$

Looks like program worked, but this could be because of improvements in the economy which could also affect the sales of non-participants.

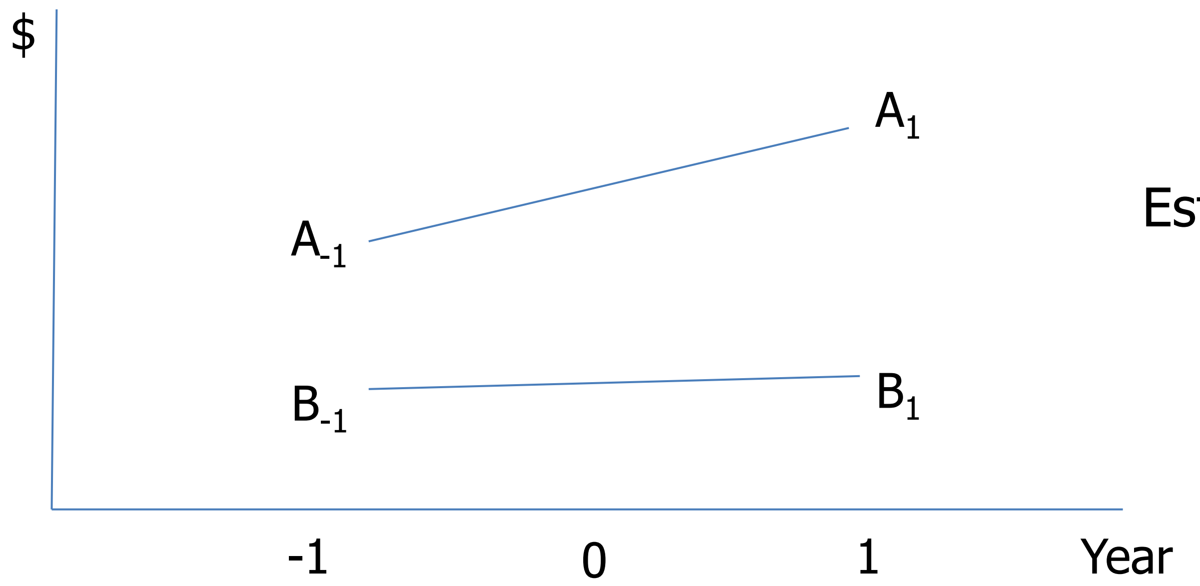
Difference estimators

- Method #4: Difference-in-differences

- Compare the change in sales (before and after the program was launched) of program participants and non-participants

Difference estimators

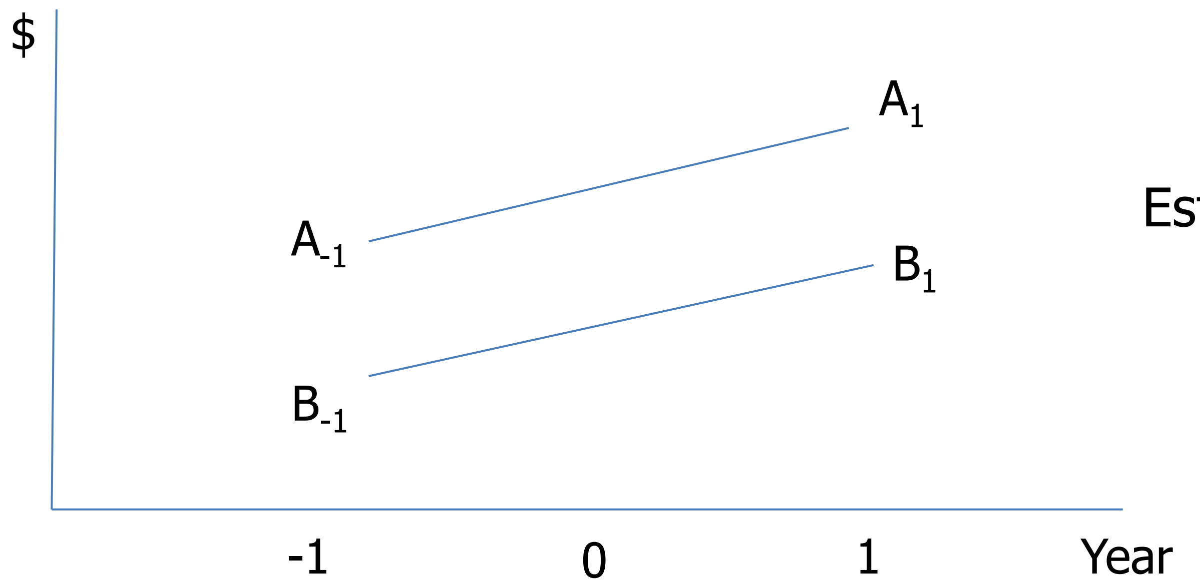
- Method #4: Difference-in-differences



Estimated impact = $(A_{\mathbf{1}} - A_{\mathbf{-1}}) - (B_{\mathbf{1}} - B_{\mathbf{-1}})$

Difference estimators

- Method #4: Difference-in-differences



Estimated impact = $(A_1 - A_{-1}) - (B_1 - B_{-1})$

Difference estimators

- Method #4: Difference-in-differences
- Starting point for most serious quantitative impact assessments
- However, there are many reasons why it may still not be enough

Difference estimators

- Problem #1 with difference-in-differences:

Confounding factors

- – Other factors may affect program group in particular
- – Usually a problem with provincial or other group analysis (e.g. a new provincial policy came into effect at the same time as the program was being implemented)
- – Can also be a problem when tracking individual outcomes (e.g. non-participants may seek alternative treatment)

Difference estimators

- Problem #1 with difference-in-differences:
  - Solution to confounding factors problem is to perform an environmental scan to ensure there are none

Difference estimators

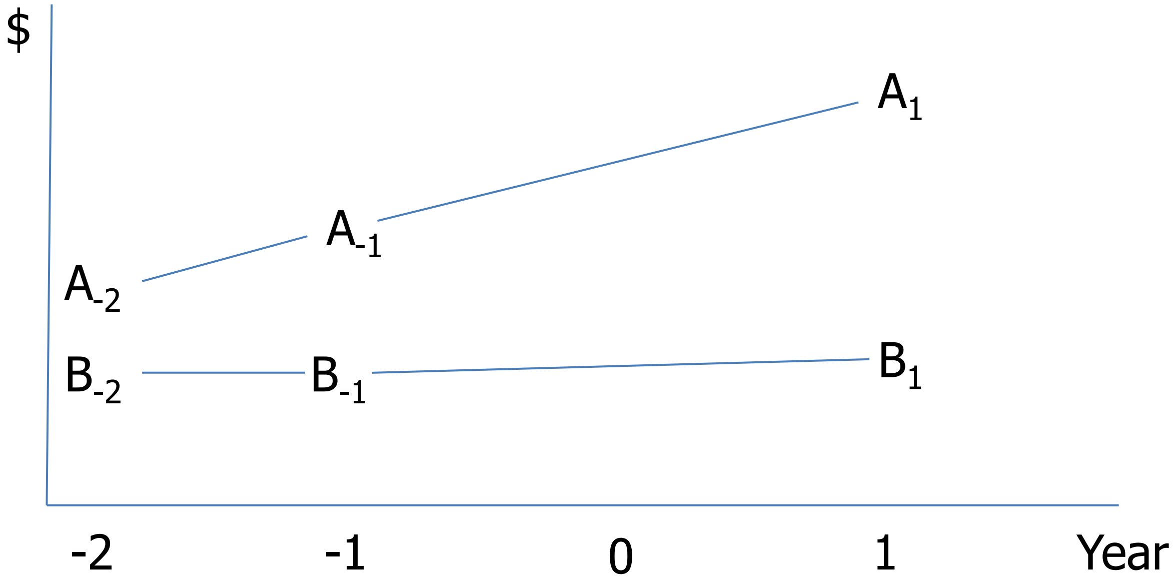
- Problem #2 with difference-in-differences:

  Common trends assumption

  – Program and comparison groups may have been on different paths before program implementation

Difference estimators

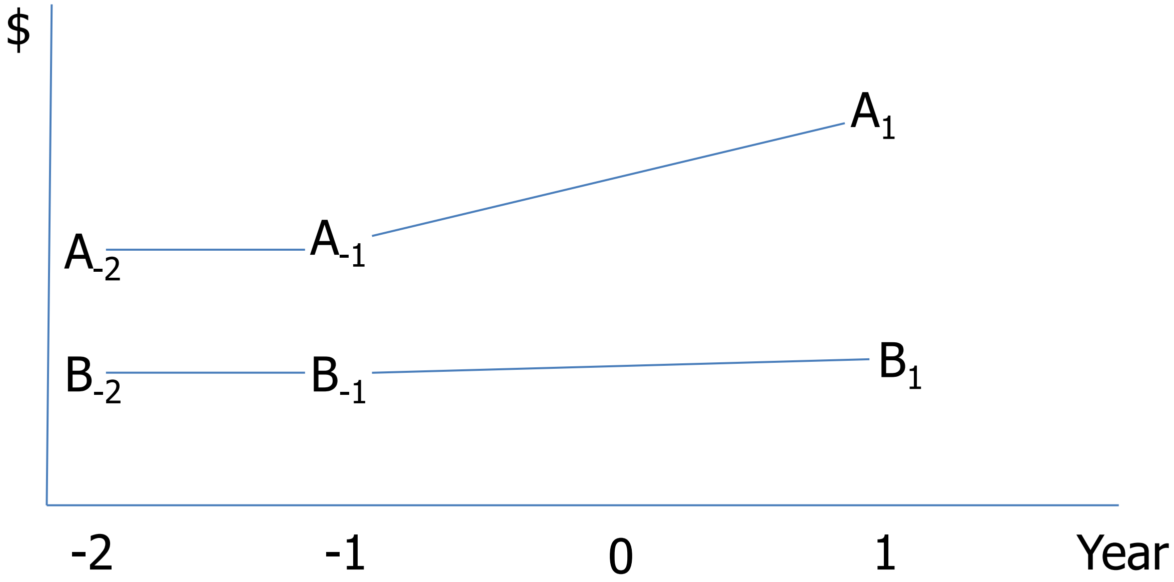
- Did the program work?



It doesn't look like it since the trends did not change after program implementation

Difference estimators

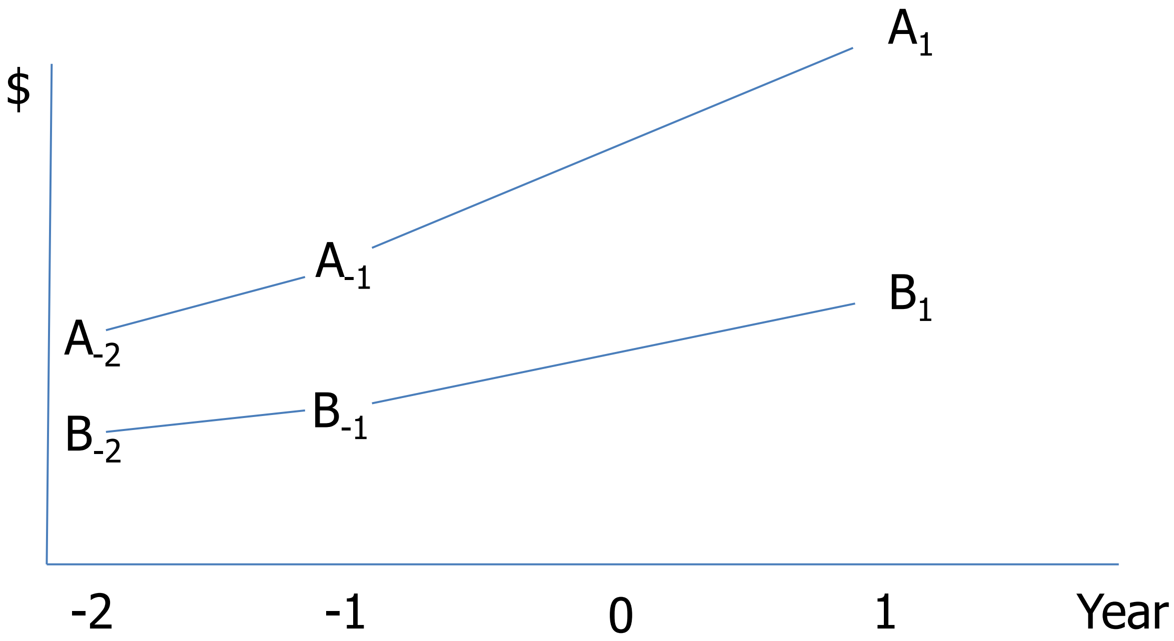
- Did the program work?



It appears so since the trends were similar before program implementation

Difference estimators

- Did the program work?



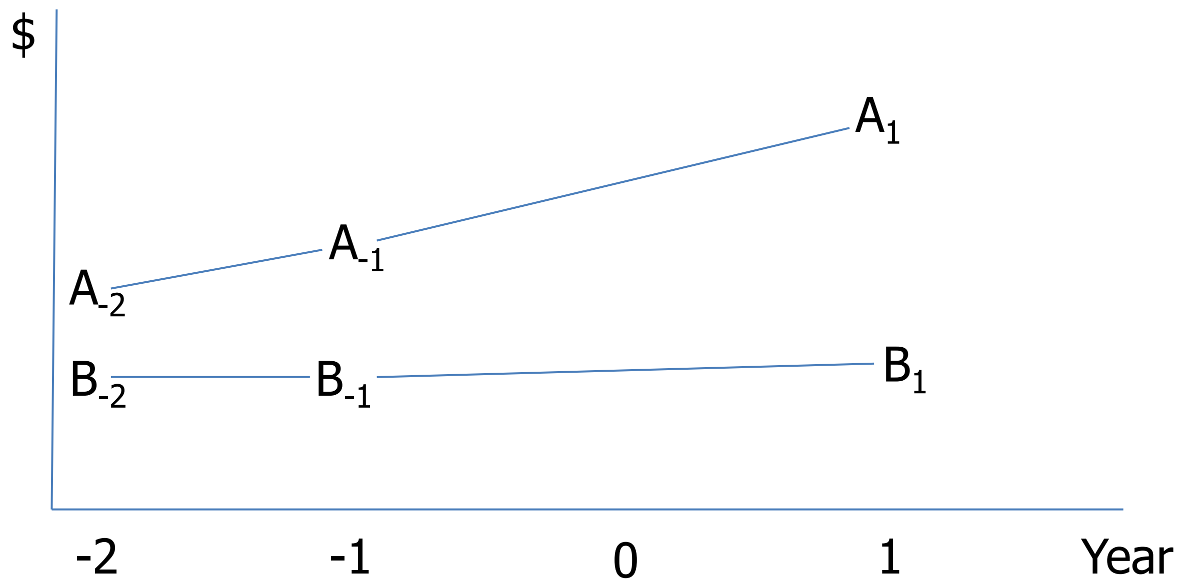
It appears so since sales grew faster for A

- after the program implementation compared to before implementation *and*
- compared to B's growth after program implementation

Difference estimators

- Problem #2 with difference-in-differences:
  - Solution to common trends assumption is to not make this assumption
  - Instead, take pre-existing trends into consideration

Difference estimators



Estimated program impact: compare differences in trends before and after program implementation

Trend after: $(A_1 – A_{-1}) – (B_1 – B_{-1})$

Trend before: $(A_{-1} – A_{-2}) – (B_{-1} – B_{-2})$

Difference in these two numbers is our estimated program impact

Difference estimators

- Problem #3 with difference-in-differences:

Selection on treatment

  - Program group may have opted to take program because they knew that it would benefit them
  - Comparison group may have chosen to not take program because they knew it would <u>not</u> benefit them
  - Major problem in QIA

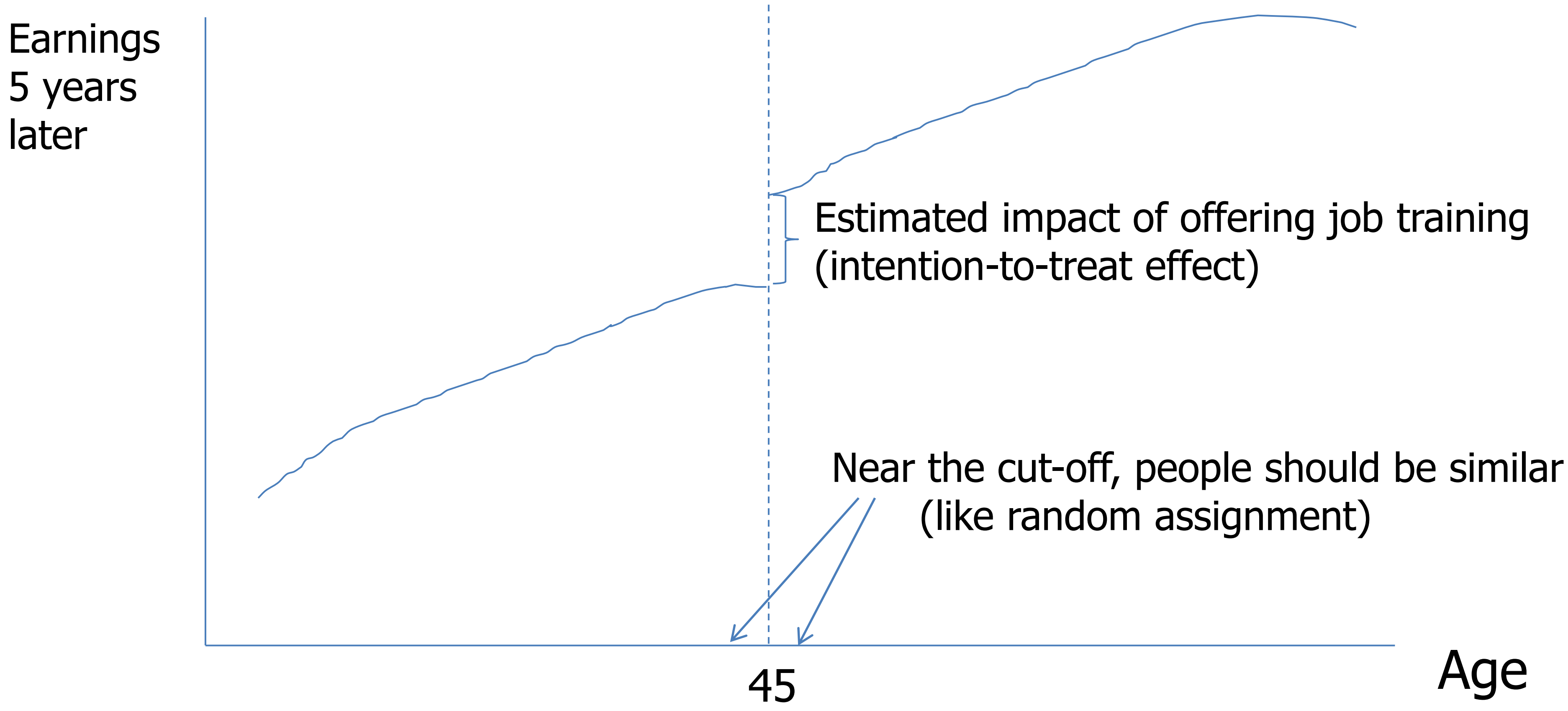
Difference estimators

- Problem #3 with difference-in-differences:
  - Solutions for selection on treatment, in descending order of effectiveness:
    - Eliminate choice (e.g. random assignment)
    - Find situations in the real world where choice is removed (discontinuities and instrumental variables)
    - Ensure program and comparison groups are as similar as possible through mathematical adjustments (matching and regression)

Discontinuity estimators

- Example: A one-time job training program is offered to ‘older’ workers, defined as 45+ on a specified day (one day less removes eligibility)

Discontinuity estimators



Discontinuity estimators

- A **fuzzy discontinuity** is one where the probability of receiving the treatment is affected by the discontinuity (like the job training program above)
  - Like intention-to-treat

- A **sharp discontinuity** is one where the treatment will be determined by the discontinuity

Discontinuity estimators

- Limitations of using discontinuities in eligibility:
  - Rare (eligibility rules often more gradual)
  - Need very large sample very close to cut-off
  - Manipulation effects - people can alter (game) or lie about their eligibility criteria to get treatment (selection on treatment)
  - Overlapping discontinuities (confounding factors)
  - LATE (effects only apply near the discontinuity)

Instrumental variable estimators

- Example: Job training program again, but instead of being based on age, the program is geared towards lone-parents with at least three children
- Cannot apply a discontinuity estimator since two and three kids are very different choices likely made by very different people, i.e. people with two kids could be quite different from people with three kids and thus not be a suitable control group for people with three kids
- Instead, can we think of a situation in which there is an element of luck in the number of children and therefore eligibility for the program?

Instrumental variable estimators

- Many parents aim for two children, preferably one of each sex
- Two instances in which parents may end up with three, even if they hoped for only two, of interest to researchers[1]:
    1. A multiple birth on the second birth
    2. Two children of the same sex on the first two births (girl-girl or boy-boy)
- We refer to 1. and 2. as instrumental variables
- To a large extent, luck determines multiple births and the sex of the children
- Luck is like random assignment (i.e. it takes away choice, which is a good thing for QIA)

[1] Source: Frenette, Marc. (2011). How does the stork delegate work? Childbearing and the gender division of paid and unpaid labour. *Journal of Population Economics*. 24. 895-910. 10.1007/s00148-010-0307-y.

Instrumental variable estimators

Multiple birth on second birth, example

BIRTH	GROUP 1	GROUP 2
	Number of kids born	
1st	1	1
2nd	1	2

- Group 2 becomes eligible for the program by luck
- Having twins on the second birth is a potential candidate for an instrumental variable for program eligibility

Instrumental variable estimators

2 children of same sex on first 2 births, example

BIRTH	GROUP 1	GROUP 2
	Sex of child	
1st	Boy	Boy
2nd	Girl	Boy
3rd	N/A	Girl

- Parents with 2 children of the same sex on the first 2 births are empirically more likely to have more than 2 children in total!
- 2 children of the same sex on the first 2 births is (generally) due to luck
- 2 children of the same sex on the first 2 births is a 'predictor' of having more than 2 kids in total
- Therefore, having 2 children of the same sex on the first 2 births is a potential candidate for an instrumental variable for program eligibility

Instrumental variable estimators

- The actual mechanics of instrumental variable estimators is quite complex and best studied in an advanced methods class
- Note that you will need data on both program eligibility (in this example the number of children) AND the chosen instrumental variable(s)
- A credible instrumental variable has two characteristics:
  - Strength
  - Validity

Instrumental variable estimators

Strength:

- It should be highly correlation with the eligibility criteria (having a multiple birth on the second birth, or having two same sex children on the first two births, must be highly correlated with program eligibility)
  - Easy to verify with the data

Instrumental variable estimators

Validity:

- It must be related to the outcome (earnings in our example) only through its impact on program eligibility.

  - Note: Anything resembling luck is always best
  - In general, very difficult to establish other than on conceptual grounds

Instrumental variable estimators

- Limitations of using instrumental variables:
  - Rare (both conceptually and in terms of availability in datasets)
  - Validity is hard to establish
  - Weak instruments pose their own problems
  - LATE (effects only apply to those who would change their training decision as a result of the instrumental variable value, but it says nothing about the impact of the training program on those who would either take training or not regardless of the instrumental variable value)

Matching estimators

- A mathematical tool for making program and comparison groups more similar
- Can be used on its own in a QIA, or in conjunction with other approaches we have seen to improve them
- They can even improve randomized controlled experiments since treatment and control groups are likely not identical

Matching estimators

- We are given the following data on job training program participants and non-participants:

Job training?	Outcome (earnings)	Age
Yes	100,000	55
Yes	50,000	35
Yes	30,000	30
No	40,000	36
No	25,000	34
No	25,000	29
No	10,000	21

Average program group outcome = (100,000+50,000+30,000)/3 = 60,000

Average comparison group outcome = (40,000+25,000+25,000 +10,000)/4 = 25,000

Estimated program impact = 60,000 – 25,000 = 35,000

Matching estimators

- On the surface, it appears that the program improved earnings by $35,000 per person
- But the program participants and non-participants are very different
- Specifically, participants are older on average
- Let’s compare apples to apples by finding an appropriate counterfactual for each individual

Matching estimators

- Other than the outcome and the treatment status, we only know the age of the individuals
- Let’s find the best counterfactual for everyone based on age proximity
- Furthermore, let’s ensure that only individuals close in age are compared (say, no more than five years apart)

Matching estimators

- We are given the following data on job training program participants and non-participants:

Job training?	Outcome (earnings)	Age	Counterfactual outcome	Program impact
Yes	100,000	55	--	
Yes	50,000	35	(40,000+25,000)/2=32,500	17,500
Yes	30,000	30	25,000	5,000
No	40,000	36	50,000	10,000
No	25,000	34	50,000	25,000
No	25,000	29	30,000	5,000
No	10,000	21	--	

Estimated program impact = (17,500+5,000+10,000 +25,000+5,000)/5 = 12,500

Matching estimators

- This example highlighted two other aspects of QIA:
  - The impact may be different for different people (heterogeneous impacts)
  - We can estimate an impact for those who did not take the program, by looking at outcomes of those who did (i.e. the counterfactual to not taking the program is… taking the program)

Matching estimators

- We can take several matches (e.g. 10 closest counterfactuals)
- Danger in accepting too many matches – some are not that similar
- Trying different criteria to see how results change is a good approach (robustness testing)

Matching estimators

- When matching estimators were developed, they were used in the context of randomized controlled experiments
- Researchers tended to project the qualities of experiments onto matching estimators (they got blended together)
- In reality, matching estimators are only as good as the data we have to match with
- There may be unobserved factors that matter ('matching on observables' issue)

Matching estimators

- Different types of matching estimators:
  - Nearest neighbour matching (what we just saw)
  - Exact matching (impose that matches be identical)
  - Propensity score matching (popular, but more technical)
  - Coarsened exact matching (more recent)

Regressions

- Regressions are very similar to matching estimators in their purpose:
  - Both approaches make the program and comparison groups more similar through mathematical techniques
  - Matching estimators do this directly by comparing members of both groups who are similar
  - But what if the samples are too small? It may not be possible to compare similar individuals
  - Regressions also make program and comparison groups similar, but by predicting what would happen if both groups had the same characteristics

Regressions

- Back to our job training example
- We have data on earnings and age for the program and comparison groups
- Program group is much older and since older workers generally have more experience, they generally earn more

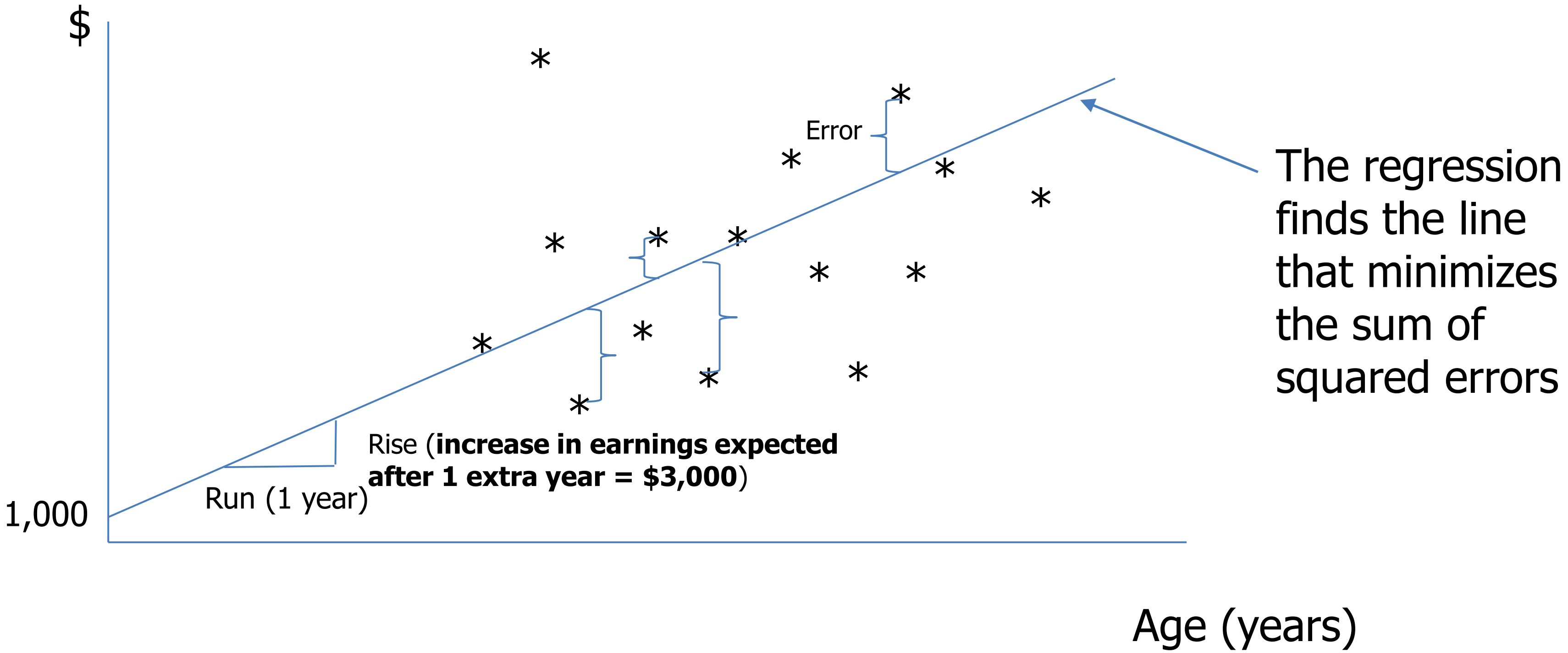
Regressions

- In words, this is what a regression does:
  - Estimates the correlation between earnings and age (say one additional year is associated with $3,000 more in earnings, on average)
  - Adjusts the gap in earnings between program and comparison group by assuming that both are the same age
  - E.g. If the actual gap in earnings is $15,000, and the program group is two years older, then the adjusted gap will be

    $15,000 – 2 years * $3,000 = $9,000
  - You can adjust for differences in many factors between those who took the training (program) and those who did not, provided you have data on those factors

Regressions

- Graphically, this is what a regression does:



Regressions

- Mathematically, this is what a regression does (and what you might see as a consumer):

Simple model of earnings:

How much do earnings rise with 1 extra year of age, on average

$$\text{Earnings}_i \quad = \quad \alpha \ + \ \beta \ \text{age}_i \ + \ e_i$$

$$\text{Earnings} \quad = \quad 1{,}000 + 3{,}000 \times \text{age}$$

Estimated coefficients allow you to predict **mean earnings** at different ages

Regressions

- Mathematically, this is what a regression does (and what you might see as a consumer):

= average earnings of those who took the program
 – average earnings of those who did not
 take the program

$$\text{Earnings}_i = \alpha + \beta\, \text{Iprogram}_i + e_i$$

$$\text{Earnings}_i = \alpha + \beta_1\, \text{Iprogram}_i + \beta_2\, \text{age}_i + e_i$$

= average earnings of those who took the program
 – average earnings of those who did not take the
 program, HOLDING AGE CONSTANT

Regressions

- Like matching estimators, regressions can be used on their own in a QIA, or in conjunction with other approaches we have seen to improve them
- With very large samples, matching estimators are potentially better since they allow for more direct matches
- With smaller samples, regressions are more feasible, but they require a lot of skill to implement and interpret them correctly
- Note though that using regression analysis alone with observational data (i.e. not data obtained through randomised assignment) is not sufficient to estimate a causal impact

Main points to remember from course

- Quandary: *can improved outcomes be attributed to program participation or to individual/firm characteristics of those who chose to participate, or other factors altogether?*
- Improved outcomes may be <u>correlated</u> with program participation, but not necessarily <u>caused</u> by program participation
- Correlation ≠ Causation
- To estimate the causal impact of a program, outcomes of program participants must be put in context – must find a credible counterfactual outcome
- A randomized control trial is the best way to do this because…

Main points to remember from course

- In real life, random assignment is rare. Most policies/programs will not be tested in an experimental framework. Why does this complicate things?
  - People/firms *choose* to participate in programs
  - Can lead to self-selection bias
  - Those who choose to participate may differ systematically from those who choose not to participate in terms of characteristics that might influence their outcomes
  - Other factors (e.g. economic conditions, government policies, etc.) may change at the same time as a program is being implemented and affect outcomes of program participants
- In the absence of random assignment, we must look towards other (less credible) approaches that try to mimic random assignment (e.g. difference-in-differences, discontinuity, instrumental variables, matching)

Thank You!

Aneta.Bonikowska@statcan.gc.ca

Stay connected!

StatsCAN app

Eh Sayers podcast

StatsCAN Plus

The Daily

Website

Surveys and statistical programs

Data service centres

My StatCan

Questions? Contact us: infostats@statcan.gc.ca

Statistics Canada—
Your National Statistical Agency

Delivering insight through data for a better Canada