Experimental Methods II

The evaluation problem and the perfect experiment

Step 1: Identify the target population

Step 2: Create clones of the population
The evaluation problem and the perfect experiment

Step 3: Give money to the target group, and nothing to the clones
Step 4: The target group and the clones spend their money
The evaluation problem and the perfect experiment

Step 5: Measure consumption in both groups and compare

<table>
<thead>
<tr>
<th>Group 1</th>
<th>Group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>Average</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>0.75</td>
</tr>
</tbody>
</table>

Treatment effect = 1.25
How to randomize?

Have to wait a few years still for the cloning technology to reach humans

Till then we will use alternative methods...
Experimental Methods II

Agenda

Part I

How to randomize?

- Randomization methods
- Level of randomization
- Internal and external validity

Part II

Quasi-experimental methods

- Differences-in-differences
- Regression Discontinuity Design
- Matching
- Other methods
Experimental Methods II

Agenda

Part I

How to randomize?

- Randomization methods
- Level of randomization
- Internal and external validity

Part II

Quasi-experimental methods
How to randomize?

**Basic Setup of a randomized evaluation**

- **Target Population**
- **Random Sampling**
- **Evaluation Sample**
- **Random Assignment**
- **Treatment group**
- **Control group**
- **Baseline Survey**
- **Endline Survey**

**Key Concepts**

- **Balance Check**
- **Implement Program**
- **Measure Impact**
- **Internal Validity**
- **External Validity**
How to randomize?

External and Internal validity

External validity:
An evaluation is externally valid if the evaluation sample accurately represents the population of eligible units.
The results are then generalizable to the population of eligible units.

Internal validity:
An evaluation is internally valid if it uses a valid comparison group (or the comparison group represents the true counterfactual.

For evaluation results to be credible, the evaluation needs to be both internally and externally valid!
How to randomize?

Target population and level of randomization

Target population can be the whole country or districts.

Randomly select villages as treatment or control villages.

Select beneficiaries within villages to survey.
How to randomize?

Target population and level of randomization

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>1457</td>
<td>1442</td>
</tr>
</tbody>
</table>

Income per person, per month, Rs.

- Treatment: 1457
- Comparison: 1442
How to randomize?

Which level of randomization?

Individual

Cluster (group)
How to randomize?

Which level of randomization?

Individual
Example: Job training program

Cluster (group)
Example: Schools
How to randomize?

Which level of randomization?

Unit of randomization should be at least at the level of the unit of program intervention:

Example: IF the program targets *classes* ->

- Randomization may be at class, school, community level
- But not at the pupil level

Program evaluation costs can also help to decide the unit of randomization!
How to randomize?

Check that randomized groups are indeed similar

Balancing on observable characteristics = Balance in unobservable characteristics
How to randomize?

With a large enough sample size, randomization produces two groups that have similar characteristics on average.

Question: How large is “sufficiently large”?
A. 3 T + 3C
B. 20 T + 10 C
C. 30 T + 30 C
D. 200 T + 200 C
E. 1500 T + 1500 C
How to randomize?

**External and Internal validity**

**Evaluation:** Researchers in Nepal interested in raising student attendance in five districts randomized a cash transfer program conditional on student attendance in a randomly sampled set of villages in those districts.

**Districts:** Kailali, Kathmandu, Taplejung, Baglung and Rukum

**Result:** One year after the program started, student attendance was on average 20% higher in schools in the villages receiving the transfer.

**Newspaper headline:** A new study finds that cash transfer program keeps children in school.
How to randomize?

**RCTs excellent for testing alternative program designs**

- **Target Population**
- **Random Sampling**
- **Evaluation Sample**
- **Random Assignment**
- **Treatment group 1**
- **Treatment group 2**
- **Treatment group 3**
- **Control group**
- **Baseline Survey**
- **Endline Survey**
- **Implement Program**
- **Measure Impact**
- **Internal Validity**
- **Balance Check**
- **External Validity**
- **Not in evaluation**
How to randomize?

Testing multiple program features: Incentives and support for schools to improve promotion of students in Malawi

Problems: High repetition and drop-out, especially for girls
Core treatment: All schools receive base grant and guidelines to improve promotion
Variable treatment: Schools receive training and/or performance-based incentives

<table>
<thead>
<tr>
<th>Group A</th>
<th>Group B</th>
<th>Group C</th>
<th>Group D</th>
</tr>
</thead>
<tbody>
<tr>
<td>25 Zones 200 Schools</td>
<td>25 Zones 200 Schools</td>
<td>25 Zones 200 Schools</td>
<td>25 Zones 200 Schools</td>
</tr>
<tr>
<td>Base Grant</td>
<td>Base Grant</td>
<td>Base Grant</td>
<td>Base Grant</td>
</tr>
<tr>
<td>Guidelines</td>
<td>Guidelines</td>
<td>Guidelines</td>
<td>Guidelines</td>
</tr>
<tr>
<td>Training</td>
<td>Incentive</td>
<td>Training</td>
<td>Incentive</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Which program do you think will produce the greatest impact?
How to randomize?

Cross cutting program features

- Target Population
- Random Sampling
- Evaluation Sample
- Baseline Survey
- Random Assignment
- Treatment 1 (No, No, Control)
- Treatment 2 (Yes, No, Treatment 1)
- Treatment 3 (Yes, Yes, Treatment 2)
- Endline Survey

- External Validity
- Internal Validity
- Balance Check
- Measure Impact
How to randomize?

RCTs can analyze cross-cutting program features: Incentives and support for schools to improve promotion of students in Malawi

<table>
<thead>
<tr>
<th>Training</th>
<th>Performance based incentives</th>
</tr>
</thead>
<tbody>
<tr>
<td>No, No (Group A)</td>
<td>No, Yes (Group C)</td>
</tr>
<tr>
<td>Yes, No (Group B)</td>
<td>Yes, Yes (Group D)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Compare</th>
<th>Learn</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A vs. Group B</td>
<td>= Impact of Training</td>
</tr>
<tr>
<td>Group A vs. Group C</td>
<td>= Impact of Incentives</td>
</tr>
<tr>
<td>Group A vs. Group D</td>
<td>= Joint impact of Training and Incentives</td>
</tr>
</tbody>
</table>
Part I

How to randomize?

Part II

Quasi-experimental methods

- Differences-in-differences
- Regression Discontinuity Design
- Matching
- Other methods
Quasi-experimental methods

What if we can't randomize?

• Programs that **have to** target particular populations
  • Example: Pension programs that target population over a certain age

• Means-tested anti-poverty programs
  • Example: provide benefit to **all households** below a certain income threshold

• Sometimes, it may be logistically impossible to randomize

This is when we turn to quasi-experimental methods!
Quasi-experimental methods

Difference-in-differences (DD)

Goal
Improve girls’ learning outcomes in eight disadvantaged districts in Malawi

Target
Eight districts with high girls’ dropout

Intervention
Improvements in male-to-female teacher ratio in Standards 6-8

Question
Where will we find a counterfactual?
Quasi-experimental methods

The DD solution: Find and use similar non-program districts

Girls’ learning outcomes

Not enrolled (C)
Enrolled (T)

Similar trends before the program

Impact=0.11

A=0.60
B=0.74
C=0.78
D=0.81

Time

t=0
t=1

Question: How do we check that program and non-program districts are really comparable?
Quasi-experimental methods

**Regression Discontinuity Design (RDD)**

**Goal**
Improve promotion in lower grades

**Intervention**
Additional teachers for schools with PTR above 50

**Target**
Schools with lower grade PTRs above 50 get extra teachers
Schools with lower grade PTRs below 50 get no extra teachers

**Question**
Where will we find a counterfactual? We want ALL high-PTR schools to enroll
Quasi-experimental methods

The RDD Solution: Use those who are right above the cut-off as control group

- Compare outcomes for schools just **above and below** the cut-off point
- Schools just above the cut-off point are very similar to those just below – valid comparison

Question: How do we check that program and non-program households are indeed similar?
Quasi-experimental methods

**Regression Discontinuity Design**

Baseline or before intervention

![Graph showing distribution of outcomes before intervention.](image1)

- Eligible
- Not eligible

Endline or after intervention

![Graph showing distribution of outcomes after intervention.](image2)

- Impact
Quasi-experimental methods

Matching

Goal
Evaluate if a low-cost online course supported by teaching assistants increases student learning as much as regular university classes

Intervention
Weekly online classes show video lectures by professors with study sessions led by teaching assistants

Evaluation Method
A group of students volunteers to test the program

Question
Where will we find a counterfactual?
Quasi-experimental methods

**Matching**

**Intuition:**

- The group that enrolled may be different from the group of students that did not volunteer
- However, some individuals are similar
- So, we can match up the treatment students with similar students who did not enroll
Quasi-experimental methods

Matching:

Baseline
Treatment group:
100 Kepler students

Comparison group:
200 students from 4 other universities – match 100 from this pool

Matching:

Baseline
Gender: Male
Age: 19
Poverty level: 3
High school GPA: 3.4
National exam score: 82
CLA+ score: 45
Grit score: 4.5
Expected salary: 700,000

Comparison group:
Gender: Male
Age: 20
Poverty level: 3
High school GPA: 3.3
National exam score: 85
CLA+ score: 43
Grit score: 4.6
Expected salary: 500,000

Gender: Male
Age: 19
Poverty level: 3
High school GPA: 3.5
National exam score: 79
CLA+ score: 46
Grit score: 4.2
Expected salary: 900,000
Quasi-experimental methods

Matching

Basic approach:
- Stratify the data into cells defined by each particular value of $X$.
- Within each cell (i.e., conditioning on $X$) compute the difference between the average outcomes of the treated and the controls.
- Average these differences with respect to the distribution of $X$ in the population of treated units (i.e., a weighted average).

Challenges:
- Finding all the relevant data for good match-ups (e.g., Grit scores) of treated and control students within each cell.
- Can never control for unobservable differences.
Other methods

- **Pre-post**: Measure how program participants improved or changed over time.

- **Simple difference**: Measure difference between program participants and non-participants after the program has been completed.

- **Multivariate regressions**: Program participants are compared to non-participants and other factors that might explain differences are “controlled for”.

- **Instrumental variables**: Program participation can be predicted by an incidental (almost random) factor, or an “instrumental variable” which uncorrelated with the outcome other than the fact that it predicts participation.
### Impact of Progresa on Consumption (Y)

<table>
<thead>
<tr>
<th>Case</th>
<th>Impact (Y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1: Before &amp; After</td>
<td>34.28**</td>
</tr>
<tr>
<td>Case 2: Randomized Assignment</td>
<td>29.75**</td>
</tr>
<tr>
<td>Case 3: Discontinuity Design</td>
<td>30.58**</td>
</tr>
<tr>
<td>Case 4: Differences in Differences</td>
<td>25.53**</td>
</tr>
</tbody>
</table>

**Note:** If the effect is statistically significant at the 1% significance level, we label the estimated impact with 2 stars (**). If significant at 10% level, we label impact with +.
Quasi-experimental methods

Randomized vs. quasi-experimental methods: the downsides

• Quasi-experimental methods do not always control for unobservable differences
  • Matching estimates

• Quasi-experimental methods produce estimates for a specific population
  • Regression Discontinuity

• Non-experimental methods require much larger sample sizes
  • Matching estimates

• Data analysis can be more complicated!

Overall, randomized experiments are much cheaper and more credible than quasi-experimental methods!
Two things to remember

To estimate impact, we need to estimate the counterfactual.

- What would have happened to the beneficiaries in the absence of the program?
- For this, we need a comparison or control group.

Choose the best evaluation method that is feasible in the program’s operational context.
Thank you!