

Đánh giá chính sách

Bài giảng 3: Thiết kế và thực hiện RCT

Edmund Malesky & Le Viet Phu

June 21, 2019

Duke University & Fulbright University

Những vấn đề đặc biệt

- Đo lường tác động
- Phân tầng và phân khối
- Tính toán cỡ mẫu
- Kiểm tra điều kiện cân bằng

Nhân tố cơ bản của RCT – Đo lường tác động

- Yêu cầu dữ liệu
 - Dữ liệu kết quả của nhóm tham gia và nhóm đối chứng
 - Dữ liệu tham chiếu (Baseline data) nếu có
- Tác động
 - Tác động can thiệp trung bình (Average Treatment Effect)

$$E[Y_1 - Y_0 | D = 1] \\ = \bar{y}_T - \bar{y}_C$$

- Thử nghiệm – Phản thực
- Trung bình tham gia – Trung bình đối chứng

THỨ TỰ NGẪU NHIÊN CỦA THIẾT KẾ THỬ NGHIỆM THEO GIAI ĐOẠN (PHASE-IN DESIGN)

Thử nghiệm theo giai đoạn: Tận dụng lợi thế của việc mở rộng chương trình

- Vấn đề đạo đức: Cuối cùng thì ai cũng được tham gia
- Thực tế: Phương pháp tiếp cận tự nhiên khi mở rộng quy mô của chương trình trong khi phải đối mặt với các hạn chế về nguồn lực
- Ngẫu nhiên hóa: Nhân tố nào quyết định trường học nào, chi nhánh nào... sẽ được tham gia chương trình vào năm nào?

Đặc tính của thiết kế ngẫu nhiên hóa theo giai đoạn

- **Phản thực:**

- Sau năm 1, những người/địa điểm bắt đầu tham gia chương trình ở năm 2, 3... sẽ được sử dụng làm nhóm đối chứng. Sau năm 2, những người bắt đầu tham gia chương trình ở năm 3, 4... sẽ được sử dụng làm nhóm đối chứng...

- **Yêu cầu dữ liệu:**

- Dữ liệu tham chiếu (tùy thuộc) và kết quả

- **Cân nhắc:**

- Theo thời gian, các nhóm đối chứng sẽ chuyển thành nhóm tham gia, do đó sẽ bị mất đi
- Tác động nhiều do nhóm những người chưa tham gia hiện nay dự kiến sẽ được tham gia trong tương lai

Thiết kế theo giai đoạn

Vòng 1

Tham gia: 1/3

Đối chứng: 2/3

Vòng 2

Tham gia: 2/3

Đối chứng:

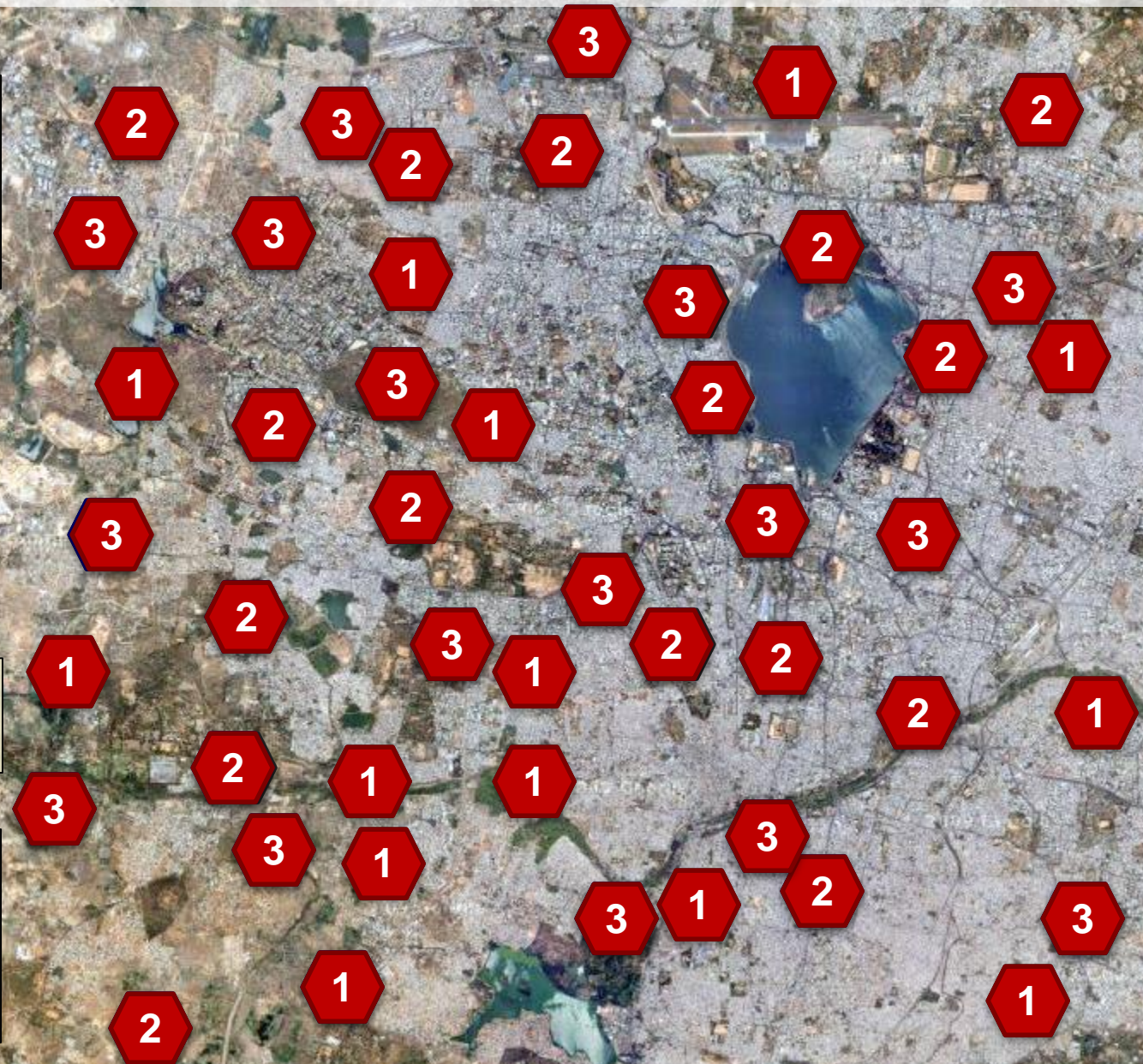
1/3

Kết thúc đánh giá thử nghiệm

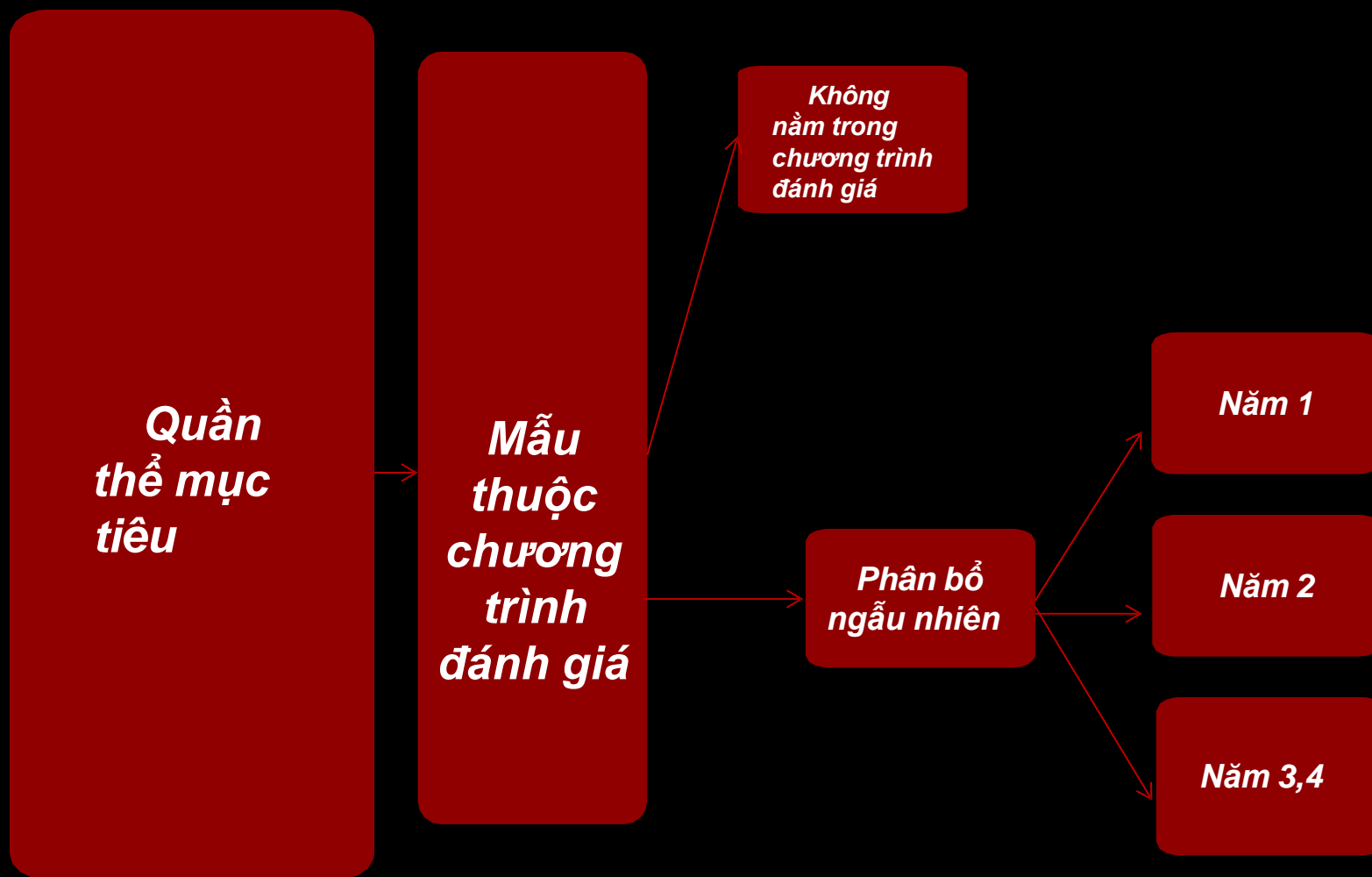
Vòng 3

Tham gia: 3/3

Đối chứng: 0



RCTs | Thiết kế theo giai đoạn



Thiết kế theo giai đoạn – Đo lường tác động

- Tác động

- Sau năm 1: Trung bình của nhóm tham gia (những người được tham gia năm 1) trừ đi trung bình của nhóm sẽ tham gia vào năm 2 & 3.

$$\begin{aligned} E[Y_1 - Y_0 | D = 1] \\ = \bar{y}_{Y1} - \bar{y}_{Y2\&3} \end{aligned}$$

- Sau năm 2: Trung bình của nhóm tham gia năm 1 & 2 trừ đi trung bình của nhóm sẽ tham gia năm 3 & 4.

$$\begin{aligned} E[Y_1 - Y_0 | D = 1] \\ = \bar{y}_{Y1\&2} - \bar{y}_{Y3\&4} \end{aligned}$$

Thiết kế theo giai đoạn: Ưu và nhược điểm

- Ưu điểm
 - Mọi người cuối cùng sẽ được tham gia nên có động lực để giữ liên hệ
- Một số vấn đề
 - Có thể làm phức tạp vấn đề ước lượng tác động dài hạn
 - Theo thời gian, có thể mất nhóm đối chứng
 - Yêu cầu phải cân trọng với các khung thời gian thực hiện theo giai đoạn
 - Liệu kỳ vọng được tham gia có làm thay đổi hành vi hiện tại không?
 - Có thể bị nhiễu do tác động dự báo của những người sẽ được tham gia trong tương lai.

THIẾT KẾ KHUYẾN KHÍCH

Khuyến khích

- *Phải làm gì khi bạn không thể sử dụng phân bố ngẫu nhiên?*
 - Đôi khi phân bố ngẫu nhiên là không thực tế hay phi đạo đức
 - Nhưng có rất nhiều chương trình có tỷ lệ chấp nhận tham gia dưới 100%
 - Có thể ngẫu nhiên hóa việc khuyến khích tham gia chương trình

Khuyến khích là gì?

- Một thứ gì đó làm cho đối tượng có xu hướng chấp nhận đăng ký chương trình đánh giá
- Khuyến khích không phải là việc được tham gia hay hưởng lợi từ chương trình
- Chúng ta ước lượng tác động đối với nhóm đối tượng nào?
- Nghĩ về “ai sẽ phản ứng đối với các khuyến khích?”

RCTs | Thiết kế khuyến khích

- Yêu cầu dữ liệu:
 - Tham chiếu (nên có) và kết quả đối với nhóm được nhận khuyến khích và nhóm không
- Cân nhắc:
 - Khuyến khích cần thiết phải được thiết kế sao cho tăng xác suất đăng ký chương trình
 - Tác động can thiệp trung bình có thể khác giữa nhóm đăng ký chương trình do được khuyến khích và quần thể hay dân số nói chung.

Thiết kế khuyến khích

 Khuyến khích

 Không khuyến khích

So sánh nhóm được khuyến khích với nhóm không được khuyến khích

Các nhóm này tương quan nhau

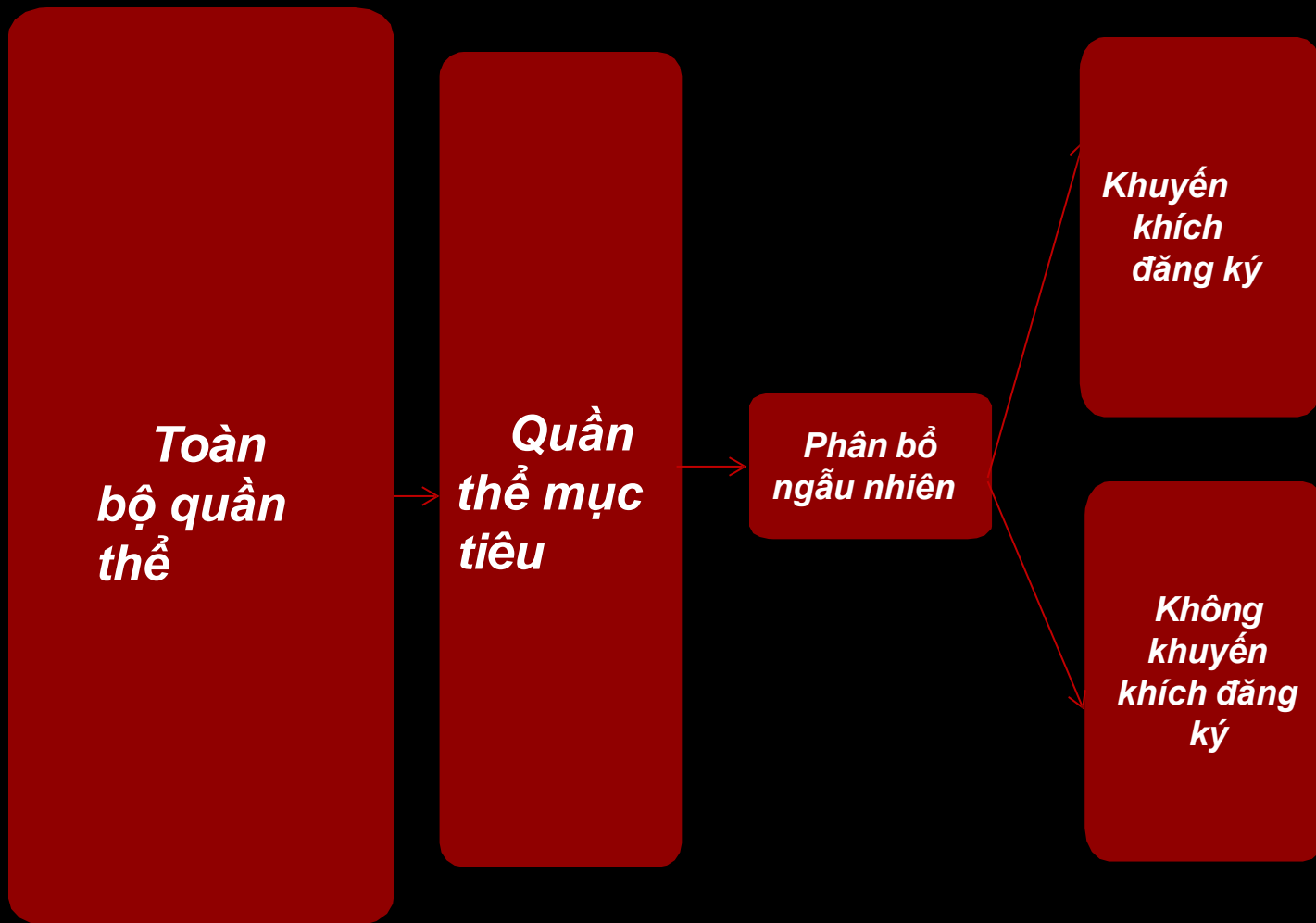
 Đăng ký tham gia

 Không tham gia

Không so sánh nhóm tham gia với nhóm không tham gia



RCTs | Thiết kế khuyến khích



Thiết kế khuyến khích – Đo lường tác động

- Tác động

- Trung bình nhóm tham gia (người nhận được khuyến khích) trừ đi trung bình nhóm không được khuyến khích.
 $E[Y_{-1} - Y_{-0} | D=1]$
- Chia cho phần trăm khác biệt về đăng ký chương trình.
- Quan trọng: Tác động này gọi là tác động can thiệp với nhóm muốn tham gia chương trình (intention to treat effect-ITE), không phải là tác động can thiệp trung bình (ATE).

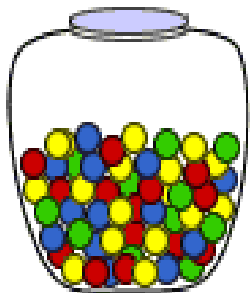
$$\begin{aligned} E[Y_{-1} - Y_{-0} | D = 1] \\ = \frac{y_E - \bar{y}_C}{e_E - e_C} \end{aligned}$$

Thiết kế khuyến khích

Ví dụ tác động

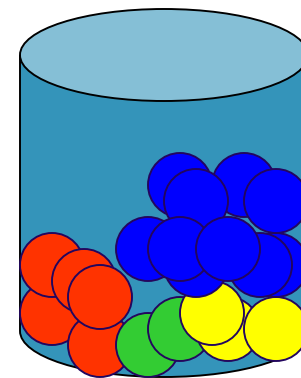
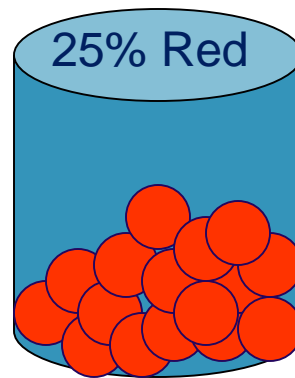
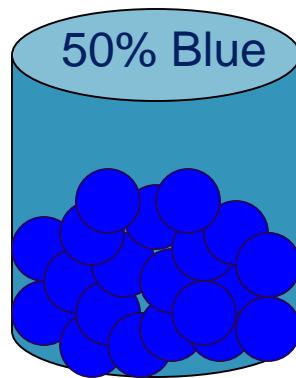
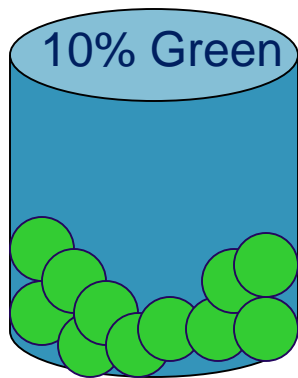
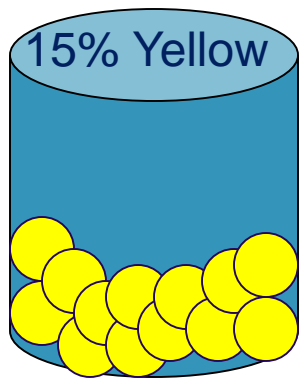
- Bạn mở một chương trình đào tạo nghề cho tất cả các đối tượng và phân bổ ngẫu nhiên ở một số khu vực tại đó các cá nhân nhận được khuyến khích để tham gia.
- Bạn phát hiện ra là ở khu vực nhận được khuyến khích tỷ lệ dân số đăng ký tham gia là 25% cao hơn so với khu vực khác. Sau 1 năm, thu nhập trung bình ở khu vực nhận được khuyến khích là \$100, so với khu vực khác là \$80.
- Tác động can thiệp ước lượng được là:
$$ITE = (\$100 - \$80) / .25 = \$25$$

Phân khoảng



Lấy mẫu ngẫu nhiên đơn giản: Sẽ tính được phần trăm số quả bóng màu nào đó (+/- 3%)

Nếu tôi muốn chắc chắn hơn, tôi sẽ phân tầng theo từng màu và lấy mẫu ngẫu nhiên trong từng màu đó



Phân tầng và phân khối

Tại sao bạn có thể không muốn thực hiện ngẫu nhiên hóa một lần duy nhất (ngẫu nhiên hóa đơn giản)?

Mường tượng rằng bạn có một biến số X liên tục quan sát được có tương quan với kết quả của việc tham gia chương trình.

- Tại sao lại sử dụng việc lấy mẫu ngẫu nhiên sao cho việc được tham gia chương trình hoàn toàn không phụ thuộc vào X ?

Bạn có thể phân tầng theo các giá trị của X để tạo ra xác suất việc tham gia chương trình không tương quan với biến X .

Điều gì nếu bạn có biến số X rời rạc quan sát được có tương quan với kết quả của việc tham gia chương trình, hoặc nếu bạn muốn phân tích tác động của việc tham gia chương trình theo các giá trị khác nhau của biến rời rạc này?

- Bạn có thể phân khối (**Block**) đối với biến số này để đảm bảo mỗi nhóm đối tượng trong toàn bộ mẫu có tỷ lệ tham gia chương trình như nhau và bằng với tỷ lệ tham gia của cả mẫu.

Phương sai kỳ vọng của một ước lượng được phân tầng hoặc phân khối không thể cao hơn phương sai kỳ vọng của ước lượng dựa trên ngẫu nhiên hóa một lần duy nhất.

Khi nào thì phân tầng

- Khi cỡ mẫu nhỏ, để giảm sai số thì có thể làm:
 - Phân tầng theo các biến có khả năng tác động lớn đến biến kết quả
 - Phân tầng theo các nhóm trong mẫu mà bạn thực sự quan tâm (nhất là khi bạn nghi ngờ tác động của chương trình có thể khác nhau)
 - Phân tầng quan trọng với mẫu dữ liệu có ít quan sát
- Cảnh báo 1: Có thể trở nên phức tạp nếu phân tầng dựa trên nhiều biến số
- Cảnh báo 2: Khi phân tầng càng nhiều thì kết quả lấy mẫu càng mất tính minh bạch

Phân tầng như thế nào

1. Cần có danh sách các quan sát trong khung dữ liệu cần ngẫu nhiên hóa.
2. Tạo một số ngẫu nhiên cho mỗi quan sát.
3. Xếp hạng theo tầng hay theo khối trước, sau đó theo số ngẫu nhiên.
4. Tung đồng xu để xác định quan sát đầu tiên trong bảng trong nhóm tham gia hay nhóm đối chứng
5. Sau đó thay đổi trạng thái tham gia đối với các quan sát tiếp theo. Cách này sẽ giả định tỷ lệ tham gia = 50%

Đối với nhiều tầng hoặc nhiều khối:

- Xếp hạng các tầng hay các khối, sau đó theo số ngẫu nhiên, và trình tự cũng tương tự như trên.

Lấy mẫu ngẫu nhiên phân tầng trong điều tra PCI

<u>Province</u>	<u>Joint Stock Companies</u>								<u>Sole Proprietorships</u>							
	<u>Manufacturing</u>		<u>Services</u>		<u>Construction</u>		<u>Agriculture</u>		<u>Manufacturing</u>		<u>Services</u>		<u>Construction</u>		<u>Agriculture</u>	
	<u>New</u>	<u>Old</u>	<u>New</u>	<u>Old</u>	<u>New</u>	<u>Old</u>	<u>New</u>	<u>Old</u>	<u>New</u>	<u>Old</u>	<u>New</u>	<u>Old</u>	<u>New</u>	<u>Old</u>	<u>New</u>	<u>Old</u>
An Giang	46	2	18	1	58	1	25	1	236	6	115	73	750	165	34	3
Bac Can	89	0	13	1	31	0	30	0	68	3	60	0	60	1	29	0
Bac Giang	128	0	107	2	160	3	41	1	6	2	22	1	110	4	5	0
Bac Lieu	26	1	10	1	23	3	9	2	59	6	53	53	385	45	23	11
Bac Ninh	172	2	191	2	198	3	24	1	27	3	177	10	140	5	7	1
Ben Tre	19	0	18	0	31	5	8	0	142	5	109	24	706	295	181	362
Binh Dinh	44	1	46	6	156	17	37	6	134	15	120	14	503	74	74	31
Binh Duong	161	0	234	3	292	0	33	0	88	1	527	63	1227	117	27	3
Binh Phuoc	60	0	34	0	51	1	61	1	45	4	88	7	652	82	117	1
Binh Thuan	61	1	30	1	112	2	63	1	75	2	131	17	539	35	118	9
BR-VT	242	4	78	2	287	9	52	2	96	2	156	11	803	66	104	9
Ca Mau	53	0	14	0	52	2	10	0	123	8	113	48	1115	140	23	4
Can Tho	232	0	84	3	231	2	26	1	89	4	268	38	837	41	21	0
Cao Bang	28	0	9	1	25	3	29	1	127	7	12	0	56	2	17	2
Da Nang	285	5	169	10	1239	43	36	1	179	13	113	14	728	71	23	1
Dak Lak	102	2	36	1	105	6	52	0	101	7	65	3	727	88	48	1
Dak non	43	0	9	0	33	0	29	1	29	0	52	0	223	13	42	0

Hãy xem xét khung dữ liệu mẫu của dự án này

Tại sao bạn không nhất thiết phải phân tầng hay phân khối?

- Bruhn & McKenzie chứng minh cho thấy cấu trúc thiết kế nghiên cứu phải được thể hiện trong việc xử lý sai số trong phương trình ước lượng.
- Ví dụ, nếu bạn muốn phân khối theo các giá trị rời rạc thì bạn cần đưa thêm tác động cố định của các giá trị đó trong phương trình ước lượng. Làm vậy sẽ mất đi số bậc tự do (DOF). Điều này có đáng làm không?
 - Trả lời: cần thiết phải phân khối nếu bạn nghi ngờ các đặc tính của biến số đó có tác động mạnh mẽ lên biến kết quả.
 - Nếu không, phân khối sẽ làm mất bậc tự do và giảm hiệu lực kiểm định thống kê.
- Sự khác biệt giữa ngẫu nhiên hóa có phân khối hoặc phân tầng và ngẫu nhiên hóa đơn giản nói chung không đáng kể khi số quan sát > 300 đơn vị.



SỨC MẠNH KIỂM ĐỊNH THỐNG KÊ - STATISTICAL POWER

Sức mạnh kiểm định thống kê

- Độ mạnh của một kiểm định là khả năng (xác suất) bác bỏ một giả thuyết thông kê (giả thuyết không – Null) khi giả thuyết thay thế là đúng.
- Khi nghiên cứu so sánh 2 nhóm, độ mạnh của kiểm định là xác suất bác bỏ giả thuyết không là hai nhóm có cùng một giá trị trung bình (bằng trung bình của quần thể), do đó kết luận là không có sự khác biệt giữa giá trị trung bình của hai quần thể, khi mà trên thực tế có sự khác biệt ở một mức độ nhất định.
- Do đó nó là xác suất kết luận đúng là hai nhóm khác biệt nhau.

Sức mạnh kiểm định thống kê

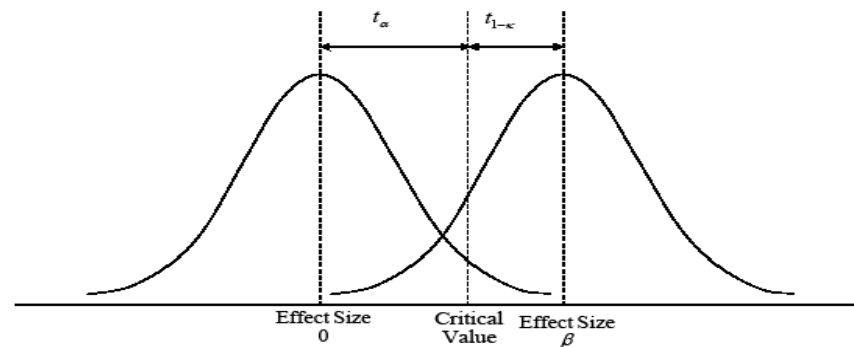
Test result

	“Reject Null,” Find an effect!	“Fail to Reject Null” Conclude no effect.
Truth: There is an effect	Great! (κ)	“Type II Error” (low power $1 - \kappa$)
Truth: There is NO effect	“Type I Error” (test size α)	Great! ($1 - \alpha$)

- Probability of Type I error is “size” of test, α (typically 0.05)
- Probability of Type II error is $(1 - \kappa)$, where κ is “power” of the test

Sức mạnh và mức ý nghĩa:

Figure 1



Đồ thị trái là phân phối của beta mũ theo giả thuyết không là nó bằng 0.

Đồ thị phải là phân phối của beta mũ nếu độ lớn thực tế là beta.

Mức ý nghĩa được nhận diện là vùng đuôi bên phải của phân phối bên trái.

Sức mạnh thống kê là vùng đuôi bên trái của phân phối bên phải.

(source: Duflo & Kremer 'Toolkit')

Sức mạnh và mức ý nghĩa:

Cần bao nhiêu quan sát thì đủ?

- Không có câu trả lời dễ dàng.
- Ngay cả công thức tính độ mạnh thống kê đơn giản nhất cũng yêu cầu bạn phải biết tác động can thiệp kỳ vọng, ETE, và phương sai của kết quả: σ ,
- Và tỷ lệ tham gia chương trình p .
- Từ đó, bạn cần chọn mức độ mạnh (xác suất bạn có thể bác bỏ giả thuyết thì bạn cần, do đó tránh được sai lầm loại II), thông thường $\kappa = .8$ and (one-tailed). $t_{1-\kappa} = 0.84$
- Sau đó, lựa chọn mức ý nghĩa (xác suất bạn bác bỏ sai khi đáng lẽ phải chấp nhận, do đó mắc sai lầm loại I), thông thường $\alpha = .05$ and $t_{1-\kappa} = 1.96$ (two-tailed).

Sau đó bạn có thể tính được cỡ mẫu tối thiểu dựa trên hàm số của độ mạnh mong muốn.

Cỡ mẫu tối thiểu

$$N > \frac{\sigma^2}{\left(\frac{ETE}{\left(t_{1-\kappa} + t_{\alpha/2} \right) \sqrt{\frac{1}{p(1-p)}} \right)^2}$$

Hơn nữa,

Bạn có thể chấp nhận một cỡ mẫu nhỏ nếu:

- Kỳ vọng tác động can thiệp là lớn
- Kết quả có độ dao động nhỏ
- Nhóm tham gia và đối chứng là các nhóm tương đồng về số quan sát ($p=.5$)
- Sẵn lòng chấp nhận mức ý nghĩa thấp và độ mạnh thống kê thấp.

Các mức độ ngẫu nhiên hóa

- Nên ngẫu nhiên hóa ở mức độ nào?
 - Vấn đề:
 - Càng nhiều nhóm thì càng cần nhiều quan sát để đạt được một mức nhất định.
 - Nếu vấn đề chệch do lan tỏa có thể xảy ra nghiêm trọng, thì có thể ngẫu nhiên hóa ở mức độ cao để ước lượng được các tác động đó.
 - Ngẫu nhiên hóa ở cấp độ nhóm có thể dễ thực hiện hơn.
 - Ngẫu nhiên hóa ở cấp độ cá nhân có thể gây ra bất mãn lớn đối với tổ chức thực hiện.

Thiết kế can thiệp theo nhóm (Clustered Treatment Designs):

Thông thường ngẫu nhiên hóa được áp dụng ở cấp độ cao hơn cấp độ chi tiết của dữ liệu.

- Ví dụ:
 - Ngẫu nhiên hóa ở cấp độ trường học hoặc làng bản đối với các chương trình áp dụng ở cấp độ sinh viên hay học viên
 - Các kiểm chứng ở cấp độ thị trường hoặc thành phố về tác động của các thông điệp chính trị áp dụng cho người bầu cử
 - Các thay đổi ở cấp độ bệnh viện trong các nghiên cứu y học sử dụng trên bệnh nhân

Ảnh hưởng của “tác động thiết kế” lên sức mạnh thống kê của kiểm định tương tự như việc điều chỉnh nhóm (cluster) khi ước lượng mô hình hồi quy.

Rốt cục: sức mạnh của kiểm định có liên quan chặt chẽ với số quan sát mà bạn sẽ thực hiện ngẫu nhiên hóa hơn là tổng số quan sát trong một nghiên cứu.

Thiết kế can thiệp theo nhóm

Nhận xét về sự khác biệt giữa “tác động tối thiểu có thể phát hiện được” – tác động thực nhỏ nhất mà một thử nghiệm có thể phát hiện được với xác suất cao.

Không có thiết kế theo nhóm:

$$MDE > (t_{1-\kappa} + t_{\alpha/2}) \sqrt{\frac{1}{p(1-p)}} \sqrt{\frac{\sigma^2}{N}}$$

Với thiết kế theo nhóm:

$$MDE > (t_{1-\kappa} + t_{\alpha/2}) \sqrt{\frac{1}{p(1-p)J}} \sigma \sqrt{\rho + \frac{1-\rho}{n}}$$

(J là số nhóm cùng kích cỡ, n là tương quan nội nhóm, và ρ là số quan sát trong một nhóm.)

Tính độ mạnh thống kê trên thực tế:

- Sử dụng phần mềm!

Có nhiều phần mềm cho phép tính toán trên Internet:

– EGAP

- https://egap.shinyapps.io/Power_Calculator/

– STATA

- power

– ‘Optimal Design’

- http://sitemaker.umich.edu/group-based/optimal_design_software

– ‘G*Power’

- <http://www.psych.uni-duesseldorf.de/aap/projects/gpower/>

Nhiều chương trình sử dụng khái niệm về tham số thống kê trong khoa học y tế chứ không phải khoa học xã hội, do đó có thể gây rắc rối khi sử dụng.

Cần thiết phải sử dụng công thức tính độ mạnh thống kê với chức năng cho phép tính thiết kế can thiệp theo nhóm nếu đơn vị nghiên cứu trong bài nghiên cứu không cùng với đơn vị can thiệp.

Thực tế: Bạn thường xuyên phải đối mặt với hạn chế về cỡ mẫu do những khó khăn



ToonClips.com

#1693

service@toonclips.com

CÂN BẰNG THỐNG KÊ

Các kiểm định cân bằng sau khi ngẫu nhiên hóa

Các nhà nghiên cứu thường viết các vòng lặp đệ quy để thực hiện việc ngẫu nhiên hóa nhiều lần, kiểm tra cân bằng theo các tiêu chí khác nhau, và lặp lại quá trình cho đến khi điều kiện cân bằng theo các tiêu chí cho trước được đảm bảo.

Có các tranh cãi liên quan đến hoạt động này.

Tất nhiên, hoạt động này cung cấp một bảng thống kê kiểm định t của các kết quả tham chiếu. Sử dụng như một tiêu chí phân tầng đa chiều. Tuy nhiên:

- Kiểm định T về sự khác biệt dựa trên sự so sánh đơn thuần giữa các giá trị trung bình không còn đúng nữa, và
- Không dễ để chỉnh sửa cấu trúc thiết kế nghiên cứu khi ước lượng các tác động can thiệp (Bruhn & McKenzie, 2008).

Ví dụ bảng thống kê cân bằng

TABLE 2. Summary Statistics

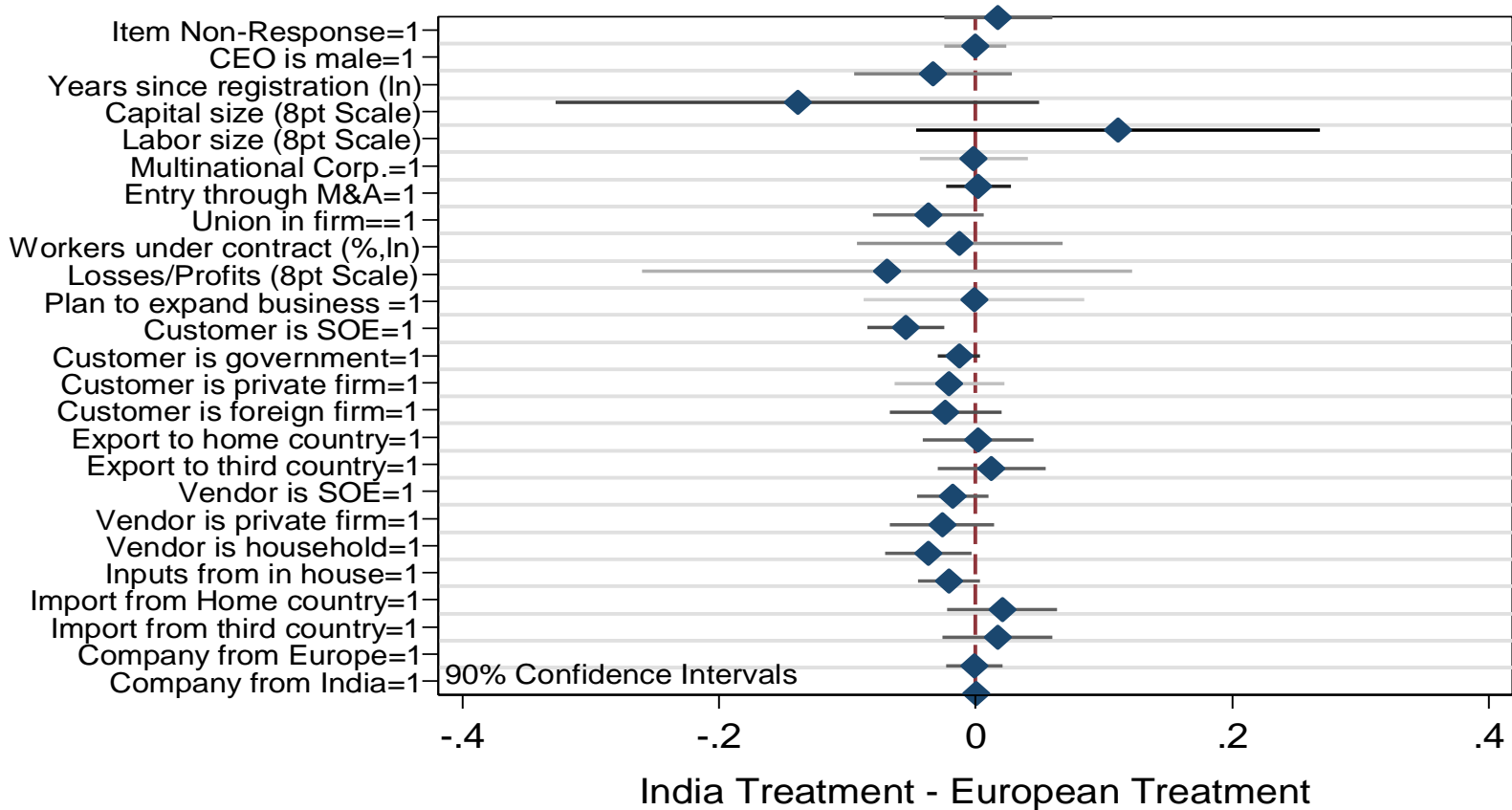
	(1)	(2)	(3)	(4)		(1)	(2)	(3)	(4)
	Mean in Meeting Group	Difference between Plebiscite and Meeting Group	<i>p</i> Value	Num Obs		Mean in Meeting Group	Difference between Plebiscite and Meeting Group	<i>p</i> Value	Num Obs
<i>Village characteristics</i>					<i>Village government characteristics</i>				
Village population (1,000 inhabitants)	2.401 [2.726]	-0.295 (0.598)	0.625	49	Village head age	45.935 [8.370]	2.368 (3.059)	0.443	47
Agricultural wage (1,000 Rupiah)	21.023 [5.892]	-1.061 (1.443)	0.466	43	Village head years of education	11.645 [2.026]	-1.409 (0.788)	0.081*	47
Percent village roads that are asphalt	0.305 [0.269]	-0.042 (0.062)	0.507	49	Number of village head candidates in last village head election	2.207 [1.013]	0.304 (0.383)	0.432	44
Number of hamlets per village	4.813 [1.839]	-0.633 (0.423)	0.142	49	More than one candidate in last village head election	0.724 [0.455]	0.089 (0.116)	0.449	44
Number of churches and mosques per village	2.438 [1.933]	-0.220 (0.563)	0.698	49	Share of population that voted in last village head election	0.888 [0.100]	-0.004 (0.031)	0.910	43
Distance to subdistrict capital (km)	5.766 [6.509]	3.548 (2.173)	0.109	49	Village head's margin of victory in last election (if challenger)	0.263 [0.262]	-0.011 (0.069)	0.870	33
Village ethnic fragmentation	0.268 [0.250]	-0.075 (0.056)	0.190	49	Number of village government executive branch members	8.516 [2.850]	-0.616 (0.703)	0.386	47
Village religious fragmentation	0.106 [0.137]	0.011 (0.051)	0.827	49	Share of hamlets represented in village executive branch	0.853 [0.240]	0.043 (0.056)	0.442	47
					Number of people in village parliament	7.750 [3.627]	-0.976 (0.832)	0.249	36
<i>Survey respondent characteristics</i>									
Survey respondent predicted log per capita expenditure	11.505 [0.279]	0.034 (0.066)	0.602	224	Share of hamlets represented in village parliament	0.843 [0.202]	0.054 (0.056)	0.339	36
Survey respondent years education	8.925 [3.088]	-0.519 (0.616)	0.404	244	Number of village parliament meetings in last year	5.714 [4.689]	-1.853 (0.878)	0.041**	44
Survey respondent is female	0.431 [0.497]	0.025 (0.023)	0.292	245	Village parliament district system (1 = district, 0 = at large)	0.241 [0.435]	0.081 (0.148)	0.587	45
Survey respondent age	41.700 [12.021]	1.896 (1.701)	0.271	245	Number of previous KDP projects	1.875 [0.976]	-0.239 (0.318)	0.455	49
Survey respondent is farmer	0.594 [0.493]	-0.052 (0.084)	0.541	245					

Notes: Column (1) presents the mean of the listed variable in the meeting villages, with standard deviations in brackets. Column (2) presents the difference between election and meeting villages, estimated with wave fixed effects, with robust standard errors in parentheses clustered at the village level. Column (3) shows the *p* value from a test of the null hypothesis that the listed variable is not different between elections and meeting villages. Column (4) shows the number of observations of the listed variable.

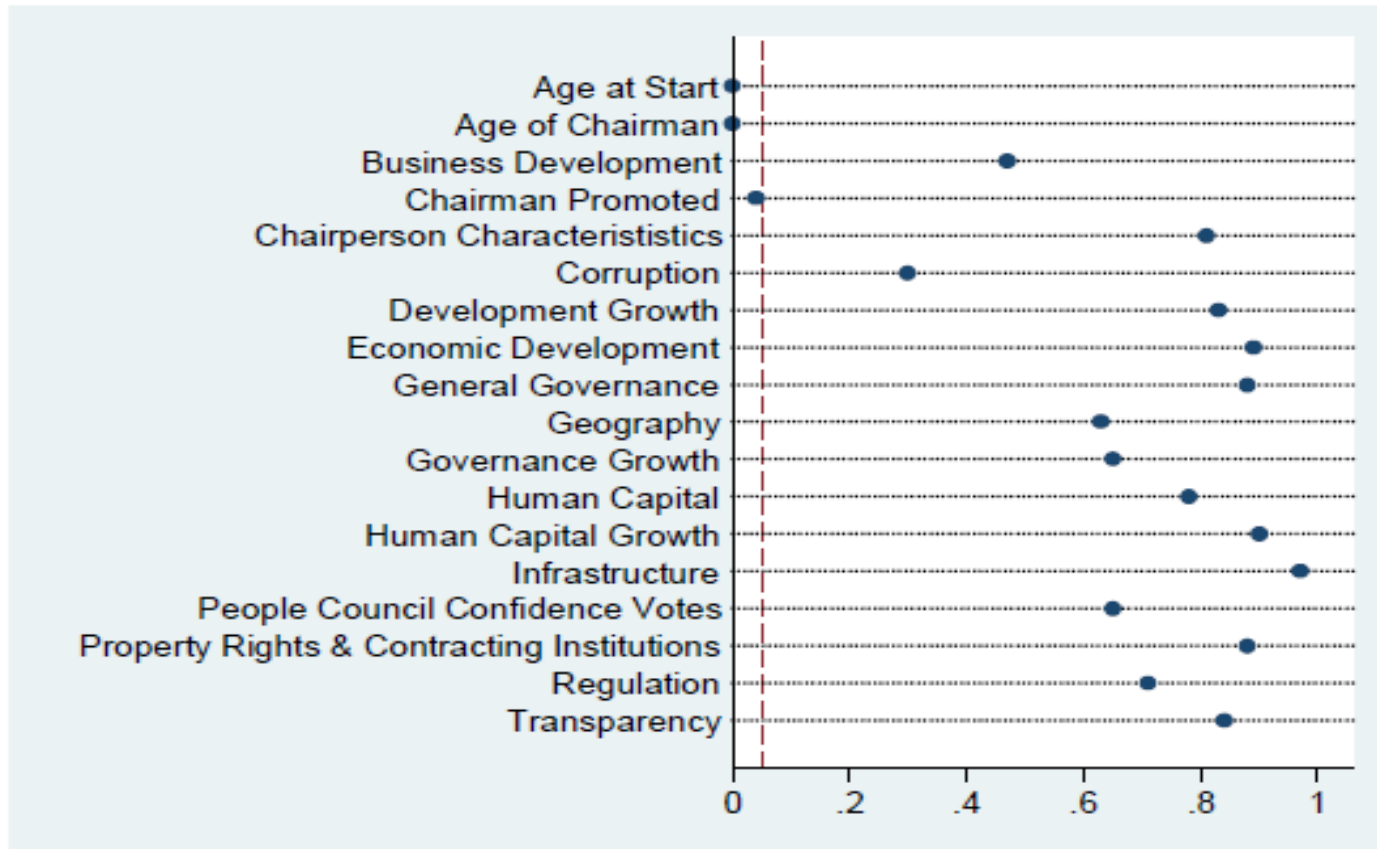
* significant at 10%; ** significant at 5%; *** significant at 1%.

Cân bằng với khoảng tin cậy

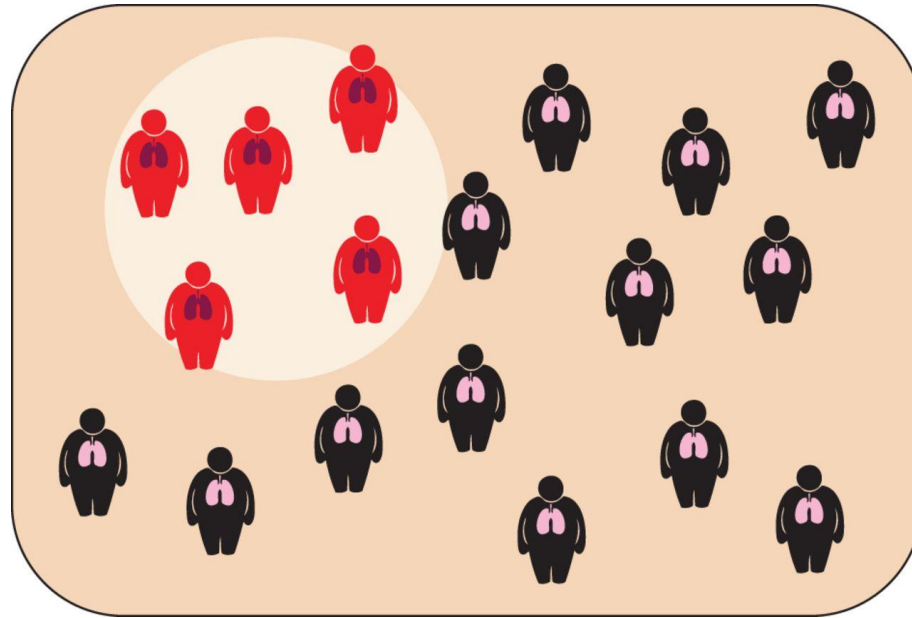
Figure 2: Survey Attrition & Balance of Confounders



Cân bằng với giá trị p-value



Note: Blue dots represent p-values from MANOVA analyses of grouped variables. The y-axis supplies the title of each grouping. A full list of indicators under each title can be found in Online Appendix B. Dashed line represents $p=.05$ from the MANOVA analysis. For dots below that number, we reject the null hypothesis that the treatment and control are different on that set of criteria.



PHÂN TÍCH NHÓM MẪU (SUB-GROUP ANALYSIS)/TÁC ĐỘNG CAN THIỆP KHÁC BIỆT (HETEROGENOUS TREATMENTS)

Tác động can thiệp khác biệt là gì?

- Một can thiệp bất kỳ có thể ảnh hưởng đến đối tượng được thử nghiệm theo nhiều cách khác nhau:
 - Tác động lớn đến đối tượng nào?
 - Tác động nhỏ với đối tượng nào?
 - Với đối tượng nào thì can thiệp tạo ra ích lợi hay thiệt hại?
- Nghiên cứu các câu hỏi đó giúp thiết lập lý thuyết về điều kiện thực thi sao cho can thiệp có hiệu quả nhất hay kém hiệu quả nhất.
- Nó cũng giúp thiết lập các hình thức thiết kế và triển khai chính sách để tối đa hóa tính hiệu quả.

Tác động can thiệp trung bình có điều kiện - Conditional Average Treatment Effects (CATEs)

- CATE là tác động can thiệp trung bình đối với nhóm mẫu, khi nhóm mẫu được xác định bởi các đặc tính của chủ thể (ví dụ tác động can thiệp trung bình ATE đối với nhóm phụ nữ), hoặc đặc tính của bối cảnh thử nghiệm (ví dụ tác động can thiệp trung bình ATE tại một địa điểm cố định trong một thử nghiệm ở nhiều địa điểm)

Sử dụng ảnh hưởng tương tác

- Nhà nghiên cứu cũng có thể ước lượng tác động can thiệp trong mối tương quan với các biến giải thích khác (treatment-by-covariate interaction effects), hoặc sự khác biệt giữa hai tác động CATE khi các nhân tố kiểm soát cho phép phân tách mẫu thành các nhóm mẫu (và các nhân tố kiểm soát này không được chủ định thử nghiệm)
- Tham số δ là ảnh hưởng tương tác, và được giải thích là sự khác biệt giữa ATE của nhóm hưởng lợi (X) trong các nhóm mẫu Z và ATE của chương trình dạy nghề trong các nhóm mẫu không phải là Z
- *Nếu Z không được phân bố ngẫu nhiên, không có tính chất nhân quả, chỉ có tính chất định tính*

$$Y_i = \alpha + \beta Z_i + \gamma X_i + \delta Z_i X_i + \varepsilon_i$$



VẤN ĐỀ THỰC TẾ CUỐI CÙNG

Cái gì dễ ngẫu nhiên hóa nhất?

1. Thông tin:

- Chương trình đào tạo
- Tuyên truyền các thông điệp chính trị
- Tuyên truyền các thông điệp về chất lượng ứng cử viên, mức độ tham nhũng
- Gửi thư cho phép thay đổi cách tiếp cận
- Khuyến khích tham gia.
 - Vấn đề đối với tất cả các nội dung trên là chúng có thể chỉ là nhân tố phụ trội đối với những nội dung mà bạn thực sự quan tâm
 - Điều này dẫn đến có hàng loạt nghiên cứu về loại hình ngẫu nhiên hóa nào có thể thực hiện được, thay vì hỏi tại sao chúng ta lại muốn thực hiện các nghiên cứu đó.

2. Phân cấp, can thiệp ở cấp độ cá nhân:

- Dẫn đến việc đánh giá nhiều câu hỏi trọng tâm về chính sách khó khăn.
- Không dễ đối với hệ thống bầu cử, chính sách quốc gia, tác động ở cấp độ đại diện, và các đồng thuận quốc tế.
- Dễ thực hiện hơn với các chương trình kêu gọi cử tri, thông điệp, tái phân bố địa bàn, và kiểm toán.

Những vấn đề thực tiễn trong thiết kế thử nghiệm

1. Bạn có trực tiếp kiểm soát được vấn đề thực hiện?
 - Nếu vậy, bạn có thể có thiết kế nghiên cứu khá tham vọng.
 - Nếu không, bạn cần phải hết sức thực tế về những mục đích chiến lược của cơ quan thực hiện. Cần giữ mức độ giản đơn:
 - Cơ quan thực hiện có giao nhân viên điều tra tại hiện trường trách nhiệm đảm bảo tính xác đáng của thiết kế nghiên cứu? Nếu không, bạn phải làm.
2. Chương trình can thiệp có quá trình lựa chọn phức hợp không?
 - Nếu có, bạn phải thiết kế việc đánh giá theo quá trình này.
 - Hoặc lựa chọn trước và ước lượng TET, hoặc ước lượng ITE.
 - Nếu mức độ tham gia thấp, bạn cần chọn trước mẫu có tỷ lệ tham gia cao để ước lượng ITE.
3. Có ràng buộc tự nhiên nào đối với việc thực hiện chương trình không?
 - Nếu có, sử dụng nó để nhận diện:
 - Phương pháp đăng ký vượt mức (Oversubscription)
 - Nếu việc thực hiện được tiến hành dần dần, bạn có thể nghiên cứu ứng dụng ngẫu nhiên hóa thứ tự thực hiện.

Ngẫu nhiên hóa không hoàn hảo

- Tác động can thiệp trung bình nội tại - Local Average Treatment Effect (LATE)
- Tuân thủ một phần - Partial Compliance
 - Cố gắng chọn thiết kế với mức độ tuân thủ cao nhất
- Ngoại tác
 - Tác động lan tỏa xảy ra đối với nội bộ nhóm và giữa các nhóm.
 - Nếu xảy ra ngoại tác lan tỏa, cần thiết kế chương trình để xử lý (Miguel and Kremer).
- Rơi rớt mẫu - Attrition
 - Rơi rớt ngẫu nhiên chỉ ảnh hưởng đến sai số chuẩn của ước lượng.
 - Rơi rớt có hệ thống sẽ làm sai lệch kết quả.
 - Cần thiết phải theo dõi và giám sát vấn đề rơi rớt mẫu.