

Nhập môn Đánh giá Tác động Chính sách (Impact Evaluation of Public Policy)

Lê Việt Phú
Chương trình Giảng dạy Kinh tế Fulbright

Ngày 10 tháng 5 năm 2016

Table of contents

1. Thế nào là đánh giá tác động chính sách
2. Tổng quan khung lý thuyết đánh giá tác động chính sách
3. Mô hình hồi quy đánh giá tác động chính sách
4. Thực hành đánh giá tác động chính sách với dữ liệu điều tra ngẫu nhiên

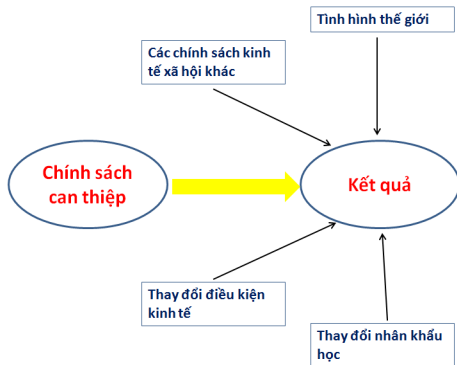
1. Thế nào là đánh giá tác động chính sách

- ▶ Các chính sách kinh tế xã hội cần phải đánh giá hiệu quả và điều chỉnh nếu cần thiết. Ví dụ:
 - ▶ Chính sách thuế bảo vệ môi trường (!) đối với xăng dầu có làm giảm lượng người sử dụng phương tiện cá nhân hay không? Nếu có thì giảm bao nhiêu phần trăm (quãng đường, số lượng xe cộ mua mới)?
 - ▶ Chính sách bảo hiểm (nông nghiệp, y tế) tại sao lại có số lượng người tham gia ít? tác động của các chương trình này như thế nào đối với hoạt động sản xuất nông nghiệp và tỷ lệ đi khám chữa bệnh của người dân?
 - ▶ Các chương trình tín dụng ưu đãi đối với hộ nghèo tại các vùng miền núi và dân tộc thiểu số có giúp cải thiện sinh kế của người dân hay không?

Các hình thức đánh giá tác động chính sách

- ▶ Đánh giá có thể được thực hiện trước khi có chương trình (đánh giá tiên nghiệm), hoặc sau khi có chương trình (hậu nghiệm hoặc hồi cứu).
- ▶ Đánh giá tiên nghiệm dựa trên các kịch bản và dự báo thay đổi (dòng tiền, chi phí, lợi ích). Ví dụ báo cáo khả thi của dự án, hay phân tích CBA.
- ▶ Đánh giá hậu nghiệm thực hiện trong quá trình vận hành dự án, hay sau khi dự án đã hoàn thành. Đánh giá hậu nghiệm dựa trên dữ liệu thu thập được trong quá trình thực hiện dự án, và dữ liệu trước khi có dự án.

2. Mục đích của đánh giá tác động chính sách



- ▶ Xác lập được quan hệ nhân quả giữa chính sách can thiệp và kết quả. Quan hệ nhân quả được xác lập khi chúng ta có thể kết luận chính sách X gây ra tác động Y.

Khi nào thì thiết lập được quan hệ nhân quả trong đánh giá tác động chính sách?

- ▶ Những gia đình thuộc diện đói nghèo nếu được tham gia các chương trình tín dụng xóa đói giảm nghèo, chương trình dạy nghề *có thể* sẽ có cuộc sống tốt hơn so với chính họ nếu như họ không tham gia.
- ▶ Người nào hút thuốc lá nhiều sẽ có khả năng bị ung thư cao hơn so với trường hợp họ không hút thuốc.
- ▶ Lao động trong các doanh nghiệp vừa và nhỏ được tiếp cận vốn ưu đãi của chính phủ không có thu nhập cao hơn so với khi doanh nghiệp đó không được vay vốn ưu đãi.

Bản chất của đánh giá tác động chính sách

- ▶ Là so sánh giữa kết quả đã thực hiện dưới tác động của chính sách với kết quả *có thể đã xảy ra* trong điều kiện không có chính sách.
- ▶ Kết quả có thể đã xảy ra gọi là **phản thực (phản chứng) - counterfactual**.
- ▶ Không bao giờ quan sát được phản chứng. Một gia đình chỉ có thể được hưởng lợi từ một chính sách nào đó, tại một thời điểm nhất định, hoặc không. Không thể có số liệu đối với cả hai trường hợp có và không có chính sách đối với một quan sát trong một thời điểm nhất định.

→ **Rất khó xác lập được quan hệ nhân quả do yêu cầu phải biết điều gì sẽ xảy ra nếu không có chính sách. Điều này đòi hỏi phải có các thiết kế nghiên cứu (research design) hợp lý. Các mô hình hồi quy thông thường chỉ xác định được quan hệ tương quan mà không xác định được quan hệ nhân quả vững chắc.**

Tương quan hay nhân quả (Correlation vs Causation)

- ▶ Người trong độ tuổi lao động thường có thu nhập tăng cùng với độ tuổi. Đây là quan hệ tương quan hay quan hệ nhân quả?
- ▶ Số liệu cho thấy nam giới có tỷ lệ gây tai nạn xe cộ nhiều hơn phụ nữ. Quan hệ giữa giới tính và hành vi lái xe nguy hiểm là tương quan hay nhân quả?
- ▶ Trẻ em tiếp xúc với các trò chơi bạo lực thường có xu hướng ứng xử bạo lực đối với các bạn đồng môn. Đây là quan hệ tương quan hay quan hệ nhân quả?

Lưu ý đối với đánh giá tác động chính sách

- A Muốn xác lập quan hệ nguyên nhân - kết quả, nguyên nhân (causes) phải thay đổi chi phối được dưới tác động của chính sách ("no causality without manipulation"). Ví dụ chỗ ở, đi học ở đâu, có tham gia vay mượn hay không ảnh hưởng như thế nào đến sinh kế.

- B Thuộc tính (attributes) chỉ có thể có quan hệ tương quan đến kết quả. Ví dụ giới tính, độ tuổi, dân tộc không thể bị chi phối hay thay đổi bởi chính sách can thiệp, do đó không thể thiết lập được quan hệ nhân quả giữa thuộc tính với kết quả.

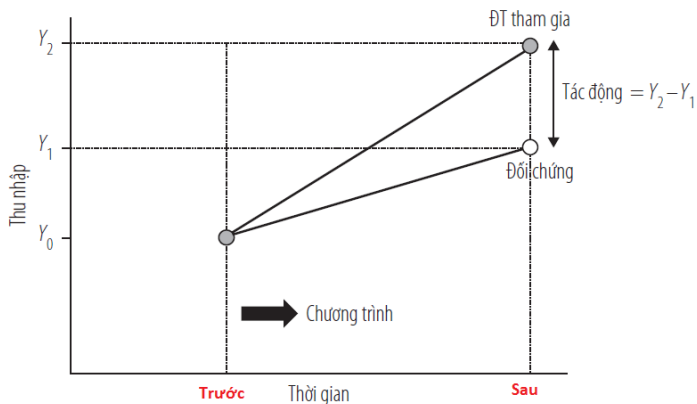
- ▶ Lý do: chỉ có thể xây dựng được phản chứng đối với [A]. Không tồn tại phản chứng với [B].

2. Tổng quan khung lý thuyết đánh giá tác động chính sách

Đánh giá tác động chính sách xoay quanh việc xây dựng phản chứng - còn gọi là nhóm đối chứng. Tùy vào điều kiện áp dụng và khả năng tài chính có các phương pháp khác nhau.

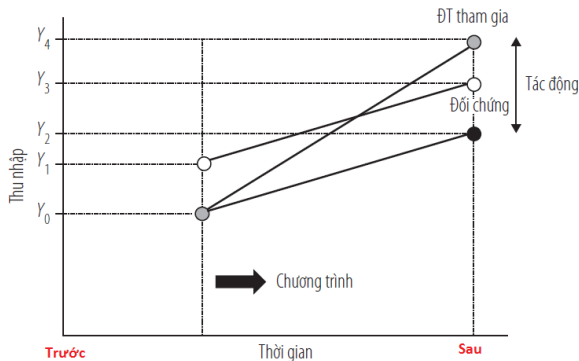
- ▶ Thiết kế **mẫu ngẫu nhiên** (randomization) là tiêu chuẩn vàng với kết quả vững chắc nhất.
- ▶ Với dữ liệu bán thực nghiệm (còn gọi là dữ liệu thử nghiệm tự nhiên, quasi- hoặc là natural- experiments), có thể sử dụng **phương pháp ghép cặp** (còn gọi là phương pháp điểm xu hướng - propensity score matching, PSM). Với dữ liệu điều tra dạng bảng, có thể sử dụng hồi quy **khác biệt trong khác biệt** (diff-in-diff regressions, DiD) để đánh giá tác động chính sách. Tuy nhiên các phương pháp này đều yêu cầu thỏa mãn một số giả định tương đối mạnh.
- ▶ Một phương pháp ít phổ biến hơn và khó áp dụng hơn là sử dụng **biến công cụ hoặc hồi quy gián đoạn** (instrumental variables (IV), regression discontinuity (RD)). Tuy nhiên phương pháp này yêu cầu một số giả định mạnh.

Trường hợp chuẩn - Thiết kế mẫu ngẫu nhiên



- ▶ Nếu nhóm đối chứng hoàn toàn tương thích với nhóm hưởng lợi thì khác biệt về kết quả giữa 2 nhóm sau khi thực hiện chính sách là tác động của chính sách can thiệp.

Đối với dữ liệu không ngẫu nhiên



- ▶ Phương pháp PSM xây dựng nhóm đối chứng bằng các ghép cặp các quan sát có chung các thuộc tính quan sát được.
- ▶ Phương pháp DiD xây dựng nhóm đối chứng bằng cách so sánh cùng một nhóm đối tượng trước và sau khi có chính sách.

3. Mô hình hồi quy đánh giá tác động chính sách

Giả sử chúng ta muốn ước lượng tác động của việc tham gia chính sách tín dụng hộ gia đình lên thu nhập của hộ bằng một mô hình đơn giản sau:

$$Y_i = \alpha X_i + \beta * T_i + \varepsilon_i \quad (1)$$

trong đó:

- ▶ Y_i là thu nhập của hộ.
- ▶ X_i là các đặc tính nhân khẩu học. Ví dụ như tuổi, giới tính, trình độ học vấn của chủ hộ.
- ▶ T là biến chính sách, nhận giá trị 0 và 1 đối với gia đình không và có tham gia.

$$T = \begin{cases} 1 & \text{Tham gia} \\ 0 & \text{Không tham gia} \end{cases}$$

- ▶ ε_i là phần dư.
- ▶ Tham số cần ước lượng β là tác động của việc tham gia chính sách đến thu nhập của hộ gia đình.

Ôn tập lý thuyết hồi quy tuyến tính cổ điển - CLRM

- ▶ Các điều kiện để ước lượng của β trong mô hình (1) là BLUE.
 - ▶ $E[\varepsilon_i] = 0$
 - ▶ $E[\varepsilon_i * T_i] = 0$
 - ▶ $V[\varepsilon_i] = \sigma^2$
 - ▶ $Cov[\varepsilon_i, \varepsilon_j] = 0$

- ▶ Các đặc tính của ước lượng BLUE:
 - ▶ Giá trị tới hạn ước lượng được của β bằng với giá trị chuẩn.
 - ▶ Sai số ước lượng được của β là nhỏ nhất.
 - ▶ Xác suất ước lượng β hội tụ về giá trị chuẩn.

Việc tham gia chính sách có được coi là ngẫu nhiên không?

- ▶ Chính sách luôn có mục tiêu cụ thể, ví dụ hướng vào đối tượng ưu tiên thay vì cho toàn bộ dân số (purposive placement).
- ▶ Tự lựa chọn mẫu (self selection or selection bias): những hộ thực sự cần thiết tham gia chưa chắc đã là những hộ được tham gia chính sách, hoặc ngược lại, do những nguyên nhân không quan sát được.
- ▶ Ngoài ra, thiếu biến quan trọng trong mô hình. Ví dụ mức độ nỗ lực làm việc, hay yếu tố may mắn, ảnh hưởng đến khả năng thành công của chính sách thì không thể quan sát được.

Hậu quả của việc ước lượng mô hình (1) nếu việc tham gia chính sách là không ngẫu nhiên

$$\text{plim } \hat{\beta} = \beta + \frac{\text{Cov}(\varepsilon * T)}{\text{Var}(T)}$$

- ▶ Khi này ước lượng tác động của chính sách sẽ bị chệch và không nhất quán.
- ▶ Hướng chệch (lên hay xuống) phụ thuộc vào tương quan giữa phần dư với biến chính sách. Nếu chỉ hộ giàu được tham gia chính sách (ε_i lớn khi $T = 1$) thì ước lượng tác động chính sách sẽ bị chệch lên. Khi này kết luận chính sách có tác động tích cực bị phóng đại so với thực tế.

Các thuật ngữ sử dụng trong đánh giá tác động chính sách

Ký hiệu sử dụng đối với hộ gia đình i :

- ▶ T_i : biến chính sách, có tham gia (1) hay không (0).
- ▶ Y_i : biến kết quả (ví dụ thu nhập).
- ▶ $Y_i|T_i$: biến kết quả với điều kiện T.

Thuật ngữ:

- ▶ Treatment: chính sách can thiệp, tham gia, bị ảnh hưởng, hưởng lợi.
- ▶ Treatment group: nhóm xử lý, nhóm hưởng lợi, nhóm tham gia, nhóm bị ảnh hưởng bởi chính sách.
- ▶ Control: kiểm soát, đối chứng.
- ▶ Control group: nhóm kiểm soát, nhóm đối chứng, nhóm không tham gia.

Công thức tính tác động can thiệp trung bình

Tác động can thiệp trung bình (average treatment effect-ATE) cho toàn bộ dân số là:

$$D = \underbrace{\mathbf{E}(Y_i | T = 1)}_{\text{Expected income with participation}} - \underbrace{\mathbf{E}(Y_i | T = 0)}_{\text{Expected income without participation}} \quad (2)$$

$Y_i | T = 1$ là thu nhập với điều kiện hộ gia đình i tham gia vào chương trình. $Y_i | T = 0$ là thu nhập với điều kiện hộ gia đình i không tham gia vào chương trình.

- ▶ Ví dụ lấy thu nhập trung bình của nhóm tham gia chương trình 135 so với nhóm không tham gia. Nếu nhóm tham gia chủ yếu là những người nghèo ít cơ hội nghề nghiệp \Rightarrow sử dụng D để nhận định tác động của chính sách thì kết quả có thể bị giảm thiểu so với thực tế.
- ▶ Do vấn đề tự chọn mẫu (selection bias), nhóm tham gia và nhóm không tham gia có thể có những khác biệt mang tính hệ thống, do đó D có thể không phản ánh chính xác tác động của chính sách can thiệp.

Định nghĩa phản chứng

Chúng ta quan tâm đến tác động của chính sách đối với những người tham gia. Đó là sự khác biệt giữa thu nhập của hộ tham gia so với trường hợp *chính họ không tham gia* \Rightarrow Cần phải định nghĩa **phản chứng - counterfactual**.

- ▶ $Y_i^0|T = 1$: Thu nhập của hộ gia đình i nếu như họ không tham gia chính sách, nhưng trên thực tế là có tham gia. Phản chứng $Y_i^0|T = 1$ không quan sát được trên thực tế.
- ▶ $Y_i^1|T = 1$: Thu nhập thực tế của hộ gia đình i sau khi họ tham gia chính sách.

$\Rightarrow (Y_i^1|T = 1) - (Y_i^0|T = 1)$ là tác động của chính sách đối với những người tham gia (treatment effect on the treated-TOT).

Tác động can thiệp trung bình và vấn đề lựa chọn mẫu

Biến đổi công thức (2):

$$D = \mathbf{E}(Y_i|T = 1) - \mathbf{E}(Y_i^0|T = 1) + \mathbf{E}(Y_i^0|T = 1) - \mathbf{E}(Y_i|T = 0)$$

- ▶ $\mathbf{E}(Y_i|T = 1) - \mathbf{E}(Y_i^0|T = 1)$ chính là tác động của chính sách đối với những hộ tham gia. **Tác động này được gọi là tác động trung bình với người tham gia (average treatment effect on the treated-ATT, hoặc ATOT), là mục tiêu nghiên cứu của việc đánh giá tác động chính sách.**
- ▶ $\mathbf{E}(Y_i^0|T = 1) - \mathbf{E}(Y_i|T = 0)$ là tác động của lựa chọn mẫu (selection bias) lên tác động trung bình D. Đó là sự khác biệt giữa thu nhập của những hộ nếu như họ không tham gia, nhưng trên thực tế là có tham gia ($Y_i^0|T = 1$), so với thu nhập của những hộ không tham gia ($Y_i|T = 0$). $Y_i^0|T = 1$ không quan sát được nên không thể ước lượng D một cách chính xác tuyệt đối.

Tác động can thiệp trung bình và vấn đề lựa chọn mẫu

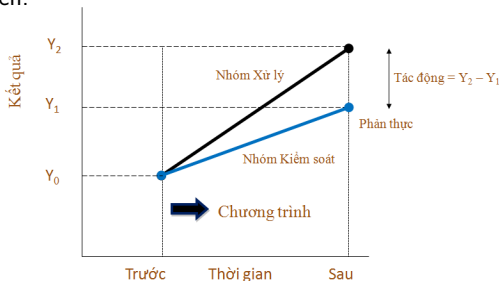
$$D = ATT + Bias \quad (3)$$

Mục tiêu của các phương pháp đánh giá tác động là loại bỏ tác động của lựa chọn mẫu, hoặc tìm phương pháp để xử lý nó.

- ▶ Thiết kế mẫu ngẫu nhiên để đảm bảo nhóm đối chứng tương đồng với nhóm hưởng lợi $\Rightarrow Bias = 0 \Rightarrow D = ATT$. Khi này chỉ cần so sánh kết quả giữa 2 nhóm hưởng lợi và không hưởng lợi là biết tác động của việc tham gia chương trình.
- ▶ Thiết kế phương pháp đánh giá sử dụng DiD, PSM, IV, RD đối với dữ liệu bán thực nghiệm.
- ▶ Lưu ý là đối với nhà hoạch định chính sách chúng ta quan tâm đến ATT hơn là ATE.

Phương pháp thiết kế mẫu ngẫu nhiên

- ▶ Các tên khác: randomization, randomized experiments, randomized control trials (RCT).
- ▶ Việc lấy mẫu ngẫu nhiên yêu cầu phải đảm bảo nhóm đối chứng và nhóm hưởng lợi tương đồng về mọi mặt \Rightarrow Khác biệt về kết quả là do khác biệt về chính sách.



- ▶ RCT thường tốn kém, khó thực hiện. Ví dụ thử nghiệm thuốc mới trên cơ thể bệnh nhân - drug trials.
- ▶ Một trường hợp không cần thiết kế mẫu ngẫu nhiên là sử dụng số liệu của cặp sinh đôi (còn gọi là ghép cặp hoàn hảo, exact matching): Các cặp sinh đôi có cùng yếu tố về năng lực hay các đặc tính sinh học nhưng bị ảnh hưởng bởi các điều kiện kinh tế, xã hội, gia đình khác nhau.

4. Thực hành đánh giá tác động chính sách với dữ liệu điều tra ngẫu nhiên

- ▶ Sử dụng bộ dữ liệu của World Bank (hh_98.dta)
- ▶ Stata dofile randomization.do

Mô tả dữ liệu

.describe

Contains data from C:\Users\x200\Desktop\Quantitative Methods II\hh_98.dta

obs: 1,129
vars: 24 1 Apr 2009 12:04
size: 115,158

variable name	storage type	display format	value label	variable label
nh	double	%7.0f		HH ID
year	float	%9.0g		Year of observation
villid	double	%9.0g		Village ID
thanaid	double	%9.0g		Thana ID
agehead	float	%3.0f		Age of HH head: years
sexhead	float	%2.0f		Gender of HH head: 1=M, 0=F
educhead	float	%2.0f		Education of HH head: years
famsize	float	%9.2f		HH size
hhland	float	%9.0g		HH land: decimals
hhasset	float	%9.0g		HH total asset: Tk.
expfd	float	%9.0g		HH per capita food expenditure: Tk/year
expnfd	float	%9.0g		HH per capita nonfood expenditure: Tk/year
exptot	float	%9.0g		HH per capita total expenditure: Tk/year
dmmfd	byte	%8.0g		HH has male microcredit participant: 1=Y, 0=N
dfmfd	byte	%8.0g		HH has female microcredit participant: 1=Y, 0=N
weight	float	%9.0g		HH sampling weight
vaccess	float	%9.0g		Village is accessible by road all year: 1=Y, 0=N
pcirr	float	%9.0g		Proportion of village land irrigated
rice	float	%9.3f		Village price of rice: Tk./kg
wheat	float	%9.3f		Village price of wheat: Tk./kg
milk	float	%9.3f		Village price of milk: Tk./liter
potato	float	%9.3f		Village price of potato: Tk./kg
egg	float	%9.3f		Village price of egg: Tk./4 counts
oil	float	%9.3f		Village price of edible oil: Tk./kg

Mô tả dữ liệu

.sum

Variable	Obs	Mean	Std. Dev.	Min	Max
nh	1129	169970.4	87505.85	11054	323103
year	1129	1	0	1	1
villid	1129	2.178919	.9471206	1	4
thanaid	1129	16.77236	8.752229	1	32
agehead	1129	46.0124	12.67865	18	95
sexhead	1129	.9078831	.2893191	0	1
educhead	1129	2.317095	3.47617	0	16
famsize	1129	5.300266	2.20517	1	18
hhland	1129	76.83225	204.0166	.2	4208
hhasset	1129	155576.4	849719.9	1	2.42e+07
expfd	1129	3660.191	1558.638	945.3242	15270.67
expnfd	1129	1813.078	3316.891	89.54749	43411.15
exptot	1129	5473.268	4140.221	1193.328	47981.01
dmmfd	1129	.1948627	.3962705	0	1
dfmfd	1129	.5270151	.4994909	0	1
weight	1129	.9973043	.8185954	.125	5
vaccess	1129	.8352524	.371117	0	1
pcirr	1129	.5603686	.3320238	0	.99
rice	1129	10.28298	1.566328	6.879596	15.4509
wheat	1129	7.466872	.8467278	5.413453	9.473542
milk	1129	10.89583	3.381805	6.766816	20.30045
potato	1129	6.958305	1.059905	4.655296	10.45534
egg	1129	1.953326	.3740207	1.353363	2.706726
oil	1129	39.40337	4.008882	23.00717	50.75112

Tạo biến log và mã số của làng

```
gen lexptot=ln(1+exptot)
gen lnland=ln(1+hhland/100)
gen vill=thanaid*10+villid
```

- ▶ Có thể phân tích bằng dữ liệu gốc hoặc lấy log.
- ▶ Chương trình được thực hiện ngẫu nhiên ở cấp độ làng (village). Mỗi làng có một mã định danh riêng.
- ▶ Cấu trúc dữ liệu: có tổng cộng 1,129 quan sát ở cấp độ hộ gia đình, nằm trong 32 thana (xã), mỗi xã có tối đa 4 làng.

Tạo biến chính sách và tìm hiểu cấu trúc dữ liệu

```
egen progvillm=max(dmmfd), by(vill)
```

```
egen progvillf=max(dfmfd), by(vill)
```

- ▶ Chương trình được chọn ngẫu nhiên ở cấp độ làng. Ở mỗi làng có hộ gia đình có và không tham gia. Mỗi hộ có thể có thành viên nam hoặc nữ hoặc cả hai cùng tham gia. Do đó chúng ta tạo ra 2 biến chính sách ở cấp độ làng (progvillm và progvillf) để phản ánh làng đó có thành viên nam hoặc nữ tham gia chương trình hay không.
- ▶ Biến dmmfd = 1 nếu hộ có thành viên là nam tham gia chương trình tín dụng, = 0 nếu không có. Tương tự, dfmfd cho thành viên nữ.

thanaid	villid	vill	dmmfd	dfmfd	progvillm	progvillf
1	1	11	1	1	1	1
1	1	11	0	1	1	1
1	1	11	0	1	1	1
1	1	11	0	1	1	1
1	2	12	0	0	0	1
1	2	12	0	0	0	1
1	2	12	0	1	0	1
1	2	12	0	0	0	1
1	2	12	0	1	0	1
1	2	12	0	1	0	1

Kiểm tra tính hợp lệ của việc chọn mẫu ngẫu nhiên

Các đặc tính của làng có và không tham gia vào chương trình là giống nhau. Sử dụng thống kê mô tả và T-test để so sánh 2 mẫu.

```
sort progvillm
```

```
by progvillm: sum
```

```
ttest agehead, by(progvillm)
```

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
0	447	46.34899	.5998129	12.68147	45.17018	47.5278
1	682	45.79179	.4855902	12.68125	44.83835	46.74522
combined	1129	46.0124	.377334	12.67865	45.27204	46.75276
diff		.5572044	.7717312		-.9569872	2.071396

diff = mean(0) - mean(1)

t = 0.7220

Ho: diff = 0

degrees of freedom = 1127

Ha: diff < 0

Ha: diff != 0

Ha: diff > 0

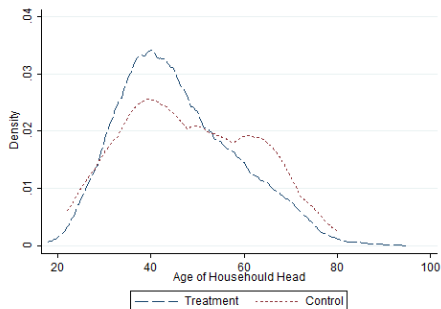
Pr(T < t) = 0.7648

Pr(|T| > |t|) = 0.4704

Pr(T > t) = 0.2352

Sử dụng đồ thị phân phối để so sánh việc chọn mẫu

```
twoway ///  
(kdensity agehead if progvillf==1,  
lpattern(longdash)) ///  
(kdensity agehead if progvillf==0,  
lpattern(shortdash)), ///  
legend(label(1 "Treatment") label(2 "Control")) ///  
bgcolor(white) graphregion(color(white))  
ytitle("Density") xtitle("Age of Household Head")
```



Tính tác động can thiệp trung bình - ATE

So sánh chỉ tiêu trung bình của hộ tại làng có và không tham gia chương trình. Ví dụ đối với làng có phụ nữ tham gia:

`ttest lexptot, by(progwillf)`

$$ATE = \mathbf{E}(Y_i|T = 1) - \mathbf{E}(Y_i|T = 0)$$

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
0	67	8.328525	.0644093	.5272125	8.199927	8.457122
1	1062	8.458371	.0157201	.5122923	8.427525	8.489217
combined	1129	8.450665	.0152934	.5138679	8.420659	8.480672
diff		-.1298466	.0646421		-.2566789	-.0030142

diff = mean(0) - mean(1)

t = -2.0087

Ho: diff = 0

degrees of freedom = 1127

Ha: diff < 0

Ha: diff != 0

Ha: diff > 0

Pr(T < t) = 0.0224

Pr(|T| > |t|) = 0.0448

Pr(T > t) = 0.9776

Sử dụng hồi quy để tính ATE

```
reg lexptot progwillf
```

Source	SS	df	MS	
Model	1.06259118	1	1.06259118	Number of obs = 1129
Residual	296.797338	1127	.263351676	F(1, 1127) = 4.03
				Prob > F = 0.0448
				R-squared = 0.0036
				Adj R-squared = 0.0027
Total	297.85993	1128	.264060221	Root MSE = .51318

lexptot	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
progwillf	.1298466	.0646421	2.01	0.045	.0030142	.2566789
_cons	8.328525	.0626947	132.84	0.000	8.205513	8.451536

Sử dụng hồi quy mở rộng để tính ATE

Kiểm soát các biến có thể có ảnh hưởng đến chi tiêu:

```
reg lexptot progvillf sexhead agehead educhead lnland  
vaccess pcirr rice wheat milk oil egg [pw=weight]
```

```
(sum of wgt is 1.1260e+03)
```

```
Linear regression
```

```
Number of obs = 1129  
F( 12, 1116) = 20.16  
Prob > F = 0.0000  
R-squared = 0.2450  
Root MSE = .46179
```

lexptot	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
progvillf	-.0455621	.1046759	-0.44	0.663	-.2509458	.1598217
sexhead	-.0373236	.0643335	-0.58	0.562	-.1635519	.0889047
agehead	.0030636	.0012859	2.38	0.017	.0005405	.0055867
educhead	.0486414	.0057184	8.51	0.000	.0374214	.0598614
lnland	.1912535	.0389079	4.92	0.000	.1149127	.2675943
vaccess	-.0358233	.0498939	-0.72	0.473	-.1337197	.0620731
pcirr	.1189407	.0608352	1.96	0.051	-.0004236	.238305
rice	.0069748	.0110718	0.63	0.529	-.0147491	.0286987
wheat	-.029278	.0196866	-1.49	0.137	-.0679049	.009349
milk	.0141328	.0072647	1.95	0.052	-.0001211	.0283867
oil	.0083345	.0038694	2.15	0.031	.0007424	.0159265
egg	.1115221	.0612063	1.82	0.069	-.0085702	.2316145
_cons	7.609248	.2642438	28.80	0.000	7.090777	8.127718

Tác động lên hộ tham gia chương trình

- ▶ Chúng ta đã xem tác động của việc làng được chọn tham gia chương trình hay không một cách ngẫu nhiên ảnh hưởng như thế nào đến chi tiêu của hộ (placement impact).
- ▶ Việc hộ trong làng có tham gia hay không (participation impact) cũng cần được nghiên cứu. Lý do? Việc tham gia của hộ là không ngẫu nhiên. Có thể hộ nào thật nghèo, hay có tín chấp, mới được vay nợ.

```
ttest lexptot, by(dmmfd)
```

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
0	534	8.447977	.023202	.5361619	8.402398	8.493555
1	595	8.453079	.0202292	.4934441	8.413349	8.492808
combined	1129	8.450665	.0152934	.5138679	8.420659	8.480672
diff		-.005102	.0306448		-.0652292	.0550253

```
diff = mean(0) - mean(1)
```

```
t = -0.1665
```

```
Ho: diff = 0
```

```
degrees of freedom = 1127
```

```
Ha: diff < 0
```

```
Ha: diff != 0
```

```
Ha: diff > 0
```

```
Pr(T < t) = 0.4339
```

```
Pr(|T| > |t|) = 0.8678
```

```
Pr(T > t) = 0.5661
```

Tác động của việc tham gia chương trình - mở rộng (2)

```
reg lexptot dfmfd sexhead agehead educhead lnland  
vaccess pcirr rice wheat milk oil egg [pw=weight]
```

```
(sum of wgt is 1.1260e+03)
```

```
Linear regression
```

```
Number of obs = 1129  
F( 12, 1116) = 19.72  
Prob > F = 0.0000  
R-squared = 0.2478  
Root MSE = .46093
```

lexptot	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
dfmfd	.0654911	.0348852	1.88	0.061	-.0029569	.133939
sexhead	-.0331386	.0647884	-0.51	0.609	-.1602593	.0939822
agehead	.0031133	.001314	2.37	0.018	.000535	.0056915
educhead	.0493265	.0060583	8.14	0.000	.0374395	.0612134
lnland	.2058408	.0421675	4.88	0.000	.1231043	.2885774
vaccess	-.0295222	.0501813	-0.59	0.556	-.1279825	.0689381
pcirr	.1080647	.0610146	1.77	0.077	-.0116515	.2277809
rice	.0057045	.0112967	0.50	0.614	-.0164607	.0278696
wheat	-.0295285	.0195434	-1.51	0.131	-.0678744	.0088174
milk	.0136748	.0073334	1.86	0.062	-.0007139	.0280636
oil	.0079069	.0038484	2.05	0.040	.000356	.0154579
egg	.1129842	.0612986	1.84	0.066	-.0072893	.2332577
_cons	7.560953	.278078	27.19	0.000	7.015339	8.106568

Tác động của việc tham gia chương trình - mở rộng (3)

```
reg lexptot dfmfd progvillf sexhead agehead educhead  
lnland vaccess pcirr rice wheat milk oil egg  
[pw=weight]
```

lexptot	Robust		t	P> t	[95% Conf. Interval]	
	Coef.	Std. Err.				
dfmfd	.0737423	.0359919	2.05	0.041	.0031228	.1443618
progvillf	-.0747142	.107158	-0.70	0.486	-.2849682	.1355397
sexhead	-.0377076	.0641847	-0.59	0.557	-.1636439	.0882288
agehead	.0030077	.0012831	2.34	0.019	.0004901	.0055254
educhead	.0499607	.0057753	8.65	0.000	.038629	.0612924
lnland	.2040906	.040482	5.04	0.000	.1246611	.2835201
vaccess	-.0348664	.0494669	-0.70	0.481	-.1319252	.0621924
pcirr	.1071558	.0609133	1.76	0.079	-.0123617	.2266734
rice	.0053896	.011106	0.49	0.628	-.0164013	.0271806
wheat	-.028722	.0196859	-1.46	0.145	-.0673476	.0099036
milk	.0137693	.0072876	1.89	0.059	-.0005297	.0280683
oil	.0077801	.0038339	2.03	0.043	.0002576	.0153025
egg	.1137676	.0614016	1.85	0.064	-.0067082	.2342433
_cons	7.64048	.2627948	29.07	0.000	7.124852	8.156108

Tác động của việc tham gia chương trình - mở rộng (4)

- ▶ Tác động của chương trình đối với những làng có tham gia chương trình:

```
reg lexptot dmfmd sexhead agehead educhead lnland  
vaccess pcirr rice wheat milk oil egg if  
progvillf==1 [pw=weight]
```

- ▶ Sử dụng cluster để kiểm soát sự tự tương quan giữa các quan sát trong một làng:

```
reg lexptot dmfmd progvillf sexhead agehead  
educhead lnland vaccess pcirr rice wheat milk oil  
egg [pw=weight], cluster(vill)
```