

Seminar

# Phân tích theo dõi dọc

---

Thực hành sử dụng Stata

Trung tâm Nghiên cứu Khoa học Sức Khỏe  
Khương Quỳnh Long  
Hà Nội, 08/05/2021

# Nội dung

---

## Phần 1: Dữ liệu theo dõi dọc

- Định nghĩa
- Ví dụ
- Cấu trúc dữ liệu
- Các kiểm định và mô hình hồi quy thường sử dụng

## Phần 2: Mô hình ảnh hưởng hỗn hợp (Mixed-effects models)

- Đặc điểm
- Mô hình random intercept và random effect
- Nghiên cứu theo dõi dọc (không nhóm chứng)
- Nghiên cứu bán thực nghiệm (Quasi-experimental study)
- Thực hành với Stata

## Phần 3: Một số phương pháp hiệu chỉnh cho sự mất cân bằng giữa các nhóm so sánh trong nghiên cứu bán thực nghiệm (đọc thêm)

- Mô hình đa biến
- Propensity score matching (PSM)
- Propensity score inverse-probability weighting (IWP)

# Phần 1: Dữ liệu theo dõi dọc

# Định nghĩa

---

- Dữ liệu theo dõi dọc (longitudinal data, panel data) là dữ liệu được thu thập thông qua đo lường lặp lại trên cùng một đối tượng trong khung thời gian nhất định
- Mục đích đo lường sự thay đổi qua thời gian (can thiệp)

**Table 1.1** Blood lead levels ( $\mu\text{g/dL}$ ) at baseline, week 1, week 4, and week 6 for 10 randomly selected children from the TLC trial.

ID	Group <sup>a</sup>	Baseline	Week 1	Week 4	Week 6
79	P	30.8	26.9	25.8	23.8
8	S	26.5	14.8	19.5	21.0
44	S	25.8	23.0	19.1	23.2
11	P	24.7	24.5	22.0	22.5
69	S	20.4	2.8	3.2	9.4
29	S	20.4	5.4	4.5	11.9
46	P	28.6	20.8	19.2	18.4
13	P	33.7	31.6	28.5	25.1
74	P	19.7	14.9	15.3	14.7
53	P	31.1	31.2	29.2	30.1

<sup>a</sup> P = placebo; S = succimer.

# Phân biệt

---

- ✓ Longitudinal data: Đo lường lặp lại trên cùng 1 đối tượng
- ✓ Dữ liệu cắt ngang (Cross-sectional data): Đo lường 1 lần cho 1 đối tượng
- ✓ Dữ liệu cắt ngang lặp lại (Repeated cross-sectional data): Đo lường lặp lại trên đối tượng khác nhau


# Đặc điểm

---

- Thuật ngữ
  - ✓ Individuals / subjects
  - ✓ Occasions / times
  - ✓ “balanced” / “unbalanced”
- Vấn đề số liệu trống (missing data)
- Các đo lường lặp lại trên cùng một đối tượng → các quan sát không độc lập

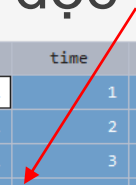
# Cấu trúc dữ liệu

- Dạng ngang (wide)



	id	trt	value1	value2	value3	value4
1	1	1	3.5	4.5	7.5	7.5
2	2	1	6.5	5.5	8.5	8.5
3	3	1	3.5	4.5	7.5	9.5
4	4	1	3.5	3.5	6.5	8.5
5	5	2	1	2	5	10
6	6	2	2	3	6	10
7	7	2	2	4	5	9
8	8	2	2	3	6	11

- Dạng dọc (long)



	id	time	trt	value
1	1	1	1	3.5
2	1	2	1	4.5
3	1	3	1	7.5
4	1	4	1	7.5
5	2	1	1	6.5
6	2	2	1	5.5
7	2	3	1	8.5
8	2	4	1	8.5
9	3	1	1	3.5
10	3	2	1	4.5
11	3	3	1	7.5
12	3	4	1	9.5
13	4	1	1	3.5
14	4	2	1	3.5
15	4	3	1	6.5
16	4	4	1	8.5



# Cấu trúc dữ liệu

---

- Chuyển từ wide → long

**reshape long** value, i (id) j (time)

- Chuyển long → wide

**reshape wide** value, i (id) j (time)

## Bài tập:

- ✓ Dữ liệu “MC\_CochranQ.dta”
- ✓ Chuyển sang dạng long sau đó chuyển lại dạng wide

# Phân tích đơn biến

---

	Đo lường lặp lại 2 lần	Đo lường lặp lại >2 lần
Định lượng, phân bố chuẩn	t ghép cặp	ANOVA lặp lại
Định lượng, không phân bố chuẩn	Wilcoxon signed rank	Friedman
Nhị giá	McNemar	Cochran's Q

# t ghép cặp

	<b>PaO<sub>2</sub></b>			<b>PaCO<sub>2</sub></b>		
	Trước	Sau	Hiệu số	Trước	Sau	Hiệu số
<b>1</b>	70	82	12	49	45	4
<b>2</b>	59	66	7	68	54	14
<b>3</b>	53	65	12	65	60	5
<b>4</b>	54	62	8	57	60	-3
<b>5</b>	44	74	30	76	59	17
<b>6</b>	58	77	19	62	54	8
<b>7</b>	64	68	4	49	47	2
<b>8</b>	43	59	16	53	50	3
<b>Trung bình</b>	55.6	69.1	13.5	59.9	53.6	6.3
<b>Độ lệch chuẩn</b>	9.2	7.9	8.2	9.6	5.9	6.5

## t ghép cặp

- Mở số liệu

pao2.dta

- Thực hiện ghép cặp

**ttest** {biếntrước} = {biếnsau}

Ví dụ: `ttest PaO2_tr = PaO2_s`

```
. ttest Pa02_tr = Pa02_s
```

Paired t test

Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
Pa02_tr	8	55.625	3.267467	9.241792	47.89867	63.35133
Pa02_s	8	69.125	2.780272	7.863796	62.5507	75.6993
diff	8	-13.5	2.915476	8.246211	-20.39401	-6.605995

mean(diff) = mean(Pa02\_tr - Pa02\_s) t = -4.6305  
Ho: mean(diff) = 0 degrees of freedom = 7

Ha: mean(diff) < 0 Ha: mean(diff) != 0 Ha: mean(diff) > 0  
Pr(T < t) = 0.0012 Pr(|T| > |t|) = 0.0024 Pr(T > t) = 0.9988

# Wilcoxon signed rank

**signrank** {biếntước} = {biếnsau}

Ví dụ: `signrank PaO2_tr = PaO2_s`

```
. signrank PaO2_tr = PaO2_s
Wilcoxon signed-rank test
```

sign	obs	sum ranks	expected
positive	0	0	18
negative	8	36	18
zero	0	0	0
all	8	36	36

```
unadjusted variance      51.00
adjustment for ties      -0.13
adjustment for zeros      0.00
-----
adjusted variance        50.88

Ho: PaO2_tr = PaO2_s
      z = -2.524
Prob > |z| = 0.0116
Exact Prob = 0.0078
```

# Anova lặp lại

- Số liệu phải ở dạng dọc
- Mở số liệu “time\_score.dta”

	person	time	score
1	1	1	30
2	1	2	28
3	1	3	16
4	1	4	34
5	2	1	14
6	2	2	18
7	2	3	10
8	2	4	22
9	3	1	24
10	3	2	20
11	3	3	18
12	3	4	30
13	4	1	38
14	4	2	34
15	4	3	20
16	4	4	44
17	5	1	26
18	5	2	28
19	5	3	14
20	5	4	30

# Anova lặp lại

`anova score person time, repeated(time)`

```
. anova score person time, repeated(time)
```

```
Number of obs =      20    R-squared      = 0.9244  
Root MSE      = 3.06594    Adj R-squared = 0.8803
```

Source	Partial SS	df	MS	F	Prob>F
Model	1379	7	197	20.96	0.0000
person	680.8	4	170.2	18.11	0.0001
time	698.2	3	232.73333	24.76	0.0000
Residual	112.8	12	9.4		
Total	1491.8	19	78.515789		

# Friedman

---

- Sử dụng gói “emh”

```
ssc install emh
```

```
emh score time, strata(person) anova transformation(rank)
```

```
. emh score time, strata(person) anova transformation(rank)
```

```
Extended Mantel-Haenszel (Cochran-Mantel-Haenszel) Stratified Test of Association
```

```
ANOVA (Row Mean Scores) Statistic:
```

```
Q (3) = 13.5600, P = 0.0036
```

```
Transformation: Ranks
```



# McNemar

Sử dụng số liệu “MC\_CochranQ.dta”

mcc value1 value3

```
. mcc value1 value3
```

Cases	Controls		Total
	Exposed	Unexposed	
Exposed	6	2	8
Unexposed	0	7	7
Total	6	9	15

McNemar's chi2(1) = 2.00 Prob > chi2 = 0.1573  
Exact McNemar significance probability = 0.5000

	ID	value1	value2	value3
1	1	1	1	1
2	2	1	1	1
3	3	1	0	0
4	4	1	0	0
5	5	0	1	0
6	6	0	0	0
7	7	0	0	0
8	8	0	0	0
9	9	1	0	1
10	10	0	1	0
11	11	1	1	1
12	12	0	0	0
13	13	1	1	1
14	14	1	0	1
15	15	0	1	0

# Cochran's Q

---

- Sử dụng gói “cochran”

```
ssc install cochran
```

```
cochran value1 value2 value3
```

```
. cochran value1 value2 value3
```

```
Test for equality of proportions of nonzero  
outcomes in matched samples (Cochran's Q):
```

```
Number of obs      =      15  
Cochran's chi2(2)   =   .8571429  
Prob > chi2         =   0.6514
```

# Hồi quy

---

- Một số mô hình hồi quy thường dùng
  - ✓ Phương trình ước tính tổng quát – Generalized Estimating Equations (GEE)
  - ✓ Mô hình ảnh hưởng hỗn hợp – Mixed-effects models

## **Phần 2: Mô hình ảnh hưởng hỗn hợp (Mixed-effects models)**

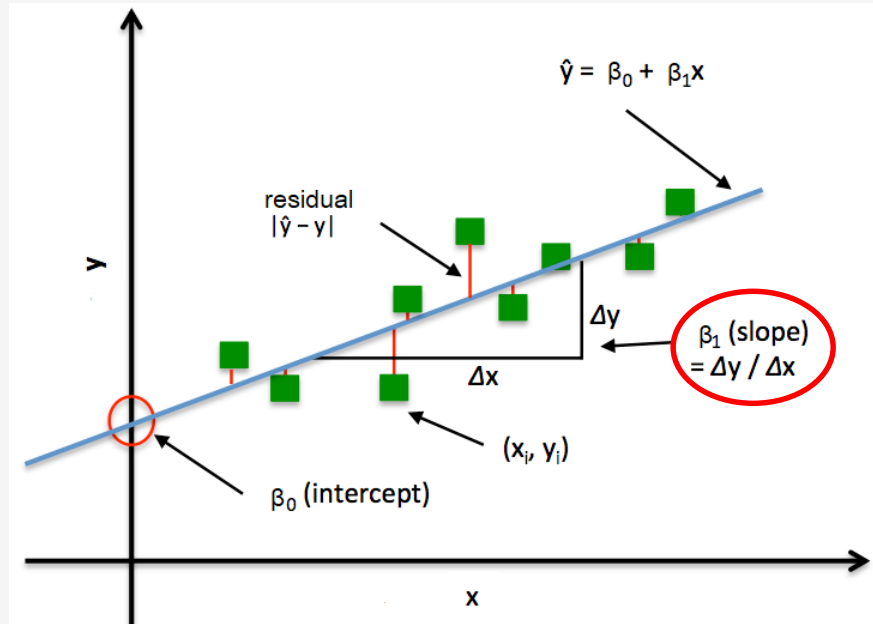
# Mô hình hồi quy tuyến tính

---

- Hồi quy tuyến tính cho phép mô hình hóa mối quan hệ tuyến tính giữa biến phụ thuộc định lượng với một hay nhiều biến độc lập
- Sử dụng phương pháp tổng bình phương nhỏ nhất (Ordinary Least Squares – OLS)

$$y_i = \beta_0 + \beta_1 x_i + e_i \quad [1]$$

# Mô hình hồi quy tuyến tính



# Mô hình hồi quy tuyến tính

---

- Giả định:  $e \sim N(0, \sigma^2)$
- ✓ Phần dư có phân phối chuẩn với trung bình bằng 0
- ✓ Phần dư có phương sai bất biến  $\sigma^2$
- ✓ Phần dư của từng cá nhân  $i$  là độc lập với nhau
- Vấn đề trong dữ liệu theo dõi dọc
  - ✓ Các đo lường lặp lại trên cùng 1 đối tượng  $\rightarrow$  phần dư không độc lập

# Mô hình ảnh hưởng hỗn hợp

---

- Tên gọi khác
  - ✓ Mô hình đa bậc (multi-level models)
  - ✓ Mô hình thứ bậc (hierarchical models)
  - ✓ Random effect model, random coefficients model
- Dùng để chỉ các nhóm mô hình
  - ✓ Linear mixed-effects model
  - ✓ Generalized mixed-effects models
  - ✓ ...



# Mô hình ảnh hưởng hỗn hợp

---

- Giải quyết vấn đề các giá trị quan sát tương quan với nhau
- Bao gồm 2 phần:
  - ✓ Ảnh hưởng cố định (fixed effect)
  - ✓ Ảnh hưởng ngẫu nhiên (random effect)

# Mô hình ảnh hưởng hỗn hợp

---

- Ảnh hưởng cố định hàm ý rằng ảnh hưởng là như nhau đối với các cá nhân có cùng đặc điểm
- Ảnh hưởng ngẫu nhiên cho phép mô hình hóa sự dao động về đo lường giữa các cá nhân

Ví dụ, thử nghiệm lâm sàng đánh giá hiệu quả của thuốc X

- ✓ Ảnh hưởng cố định là tác dụng của thuốc X trên những bệnh nhân điều trị thuốc X là như nhau.
- ✓ Ảnh hưởng ngẫu nhiên có thể hiểu là cơ địa của bệnh nhân, do đó ảnh hưởng của thuốc X lên các bệnh nhân này là dao động và không giống nhau

# Mô hình random intercept

---

- Mô hình chỉ bao gồm phần cố định (hồi quy tuyến tính đơn bậc)

$$y_i = \beta_0 + \beta_1 x_i + e_i \quad [1]$$

- Chỉ có 1 điểm chặn  $\beta_0$  cố định cho mọi giá trị  $y_i$
- Đa phần các trường hợp có giá trị khác nhau tại thời điểm ban đầu  $\rightarrow$  giá trị đo lường tại các thời điểm sau phụ thuộc giá trị ban đầu
- Thêm phần dao động cho phần điểm chặn ( $\beta_0$ )  $\rightarrow$  mô hình random intercept

# Mô hình random intercept

---

- Khi đó công thức [1] ( $y_i = \beta_0 + \beta_1 x_i + e_i$ ) được cộng thêm phần random intercept (kí hiệu là  $\mu_0$ ) và trở thành:

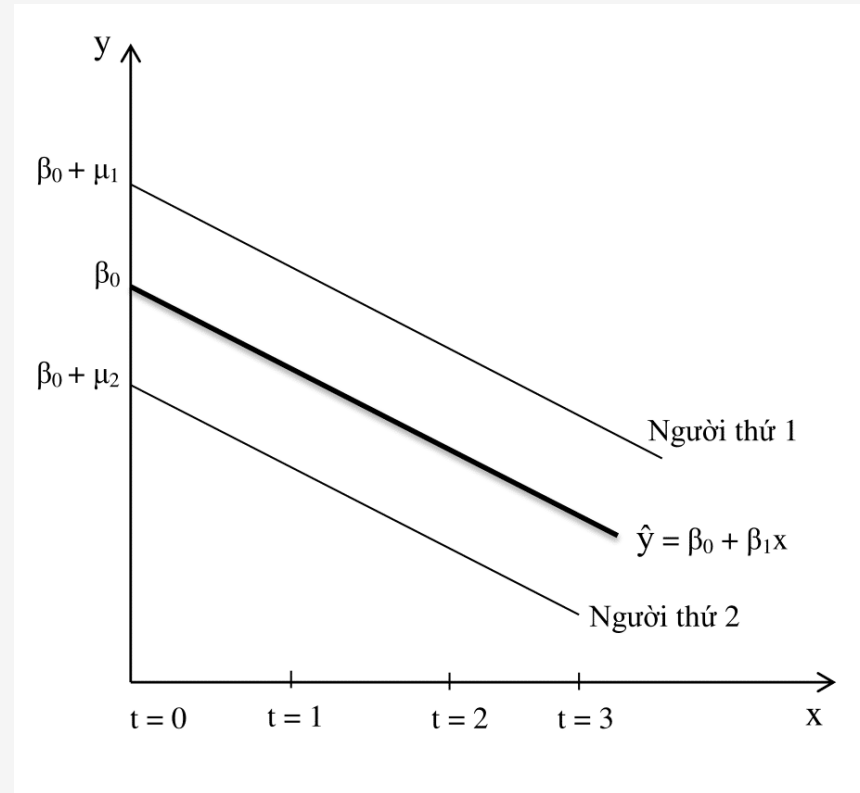
$$y_{ij} = (\beta_0 + \mu_{0j}) + \beta_1 x_{ij} + e_{ij} \quad [2]$$

- Trong đó  $i$  là lần đo thứ  $i$  của người thứ  $j$ . Phương trình [2] cũng có thể viết lại theo phần cố định và phần ngẫu nhiên

$$y_{ij} = (\beta_0 + \beta_1 x_{ij}) + (\mu_{0j} + e_{ij}) \quad [3]$$

Với  $\mu_{0j} \sim N(0, \sigma_{\mu_0}^2)$  và  $e_{ij} \sim N(0, \sigma_e^2)$

# Mô hình random intercept



# Mô hình random intercept

---

- Một nghiên cứu tiến hành nhằm đánh giá hiệu quả giảm đường huyết của thuốc X, các bệnh nhân được đo đường huyết ban đầu tại thời điểm  $t = 0$ . Sau đó là 3 tháng, 6 tháng và 9 tháng
- Giả định mỗi liên hệ giữa giảm đường huyết theo thời gian là mỗi liên hệ **tuyến tính** và **tốc độ giảm đường huyết là cố định** (fixed slope – nội dung này sẽ được đề cập trong mục tiếp theo)
- Đường huyết tại các thời điểm đo lường (tháng thứ 3, 6 và 9) sẽ khác nhau giữa các cá nhân phụ thuộc vào phần random intercept (ví dụ  $\mu_1$  và  $\mu_2$  ở hình minh họa trên)

# Mô hình random slope

---

- Phần độ dốc (slope) hay hệ số phương trình ( $\beta_1$ ) thể hiện mối liên hệ giữa biến phụ thuộc  $y$  và biến độc lập  $x$
- Ở mô hình chỉ bao gồm cố định: mối liên hệ giữa  $y$  và  $x$  là như nhau cho mọi cá nhân (trung bình)
- Trên thực tế: tác động (slope) là khác nhau giữa các cá nhân
- Thêm tham số thể hiện sự dao động này  $\rightarrow$  mô hình random slope

# Mô hình random slope

---

$$y_{ij} = (\beta_0 + \mu_{0j}) + (\beta_1 + \mu_{1j}) x_{ij} + e_{ij} \quad [4]$$

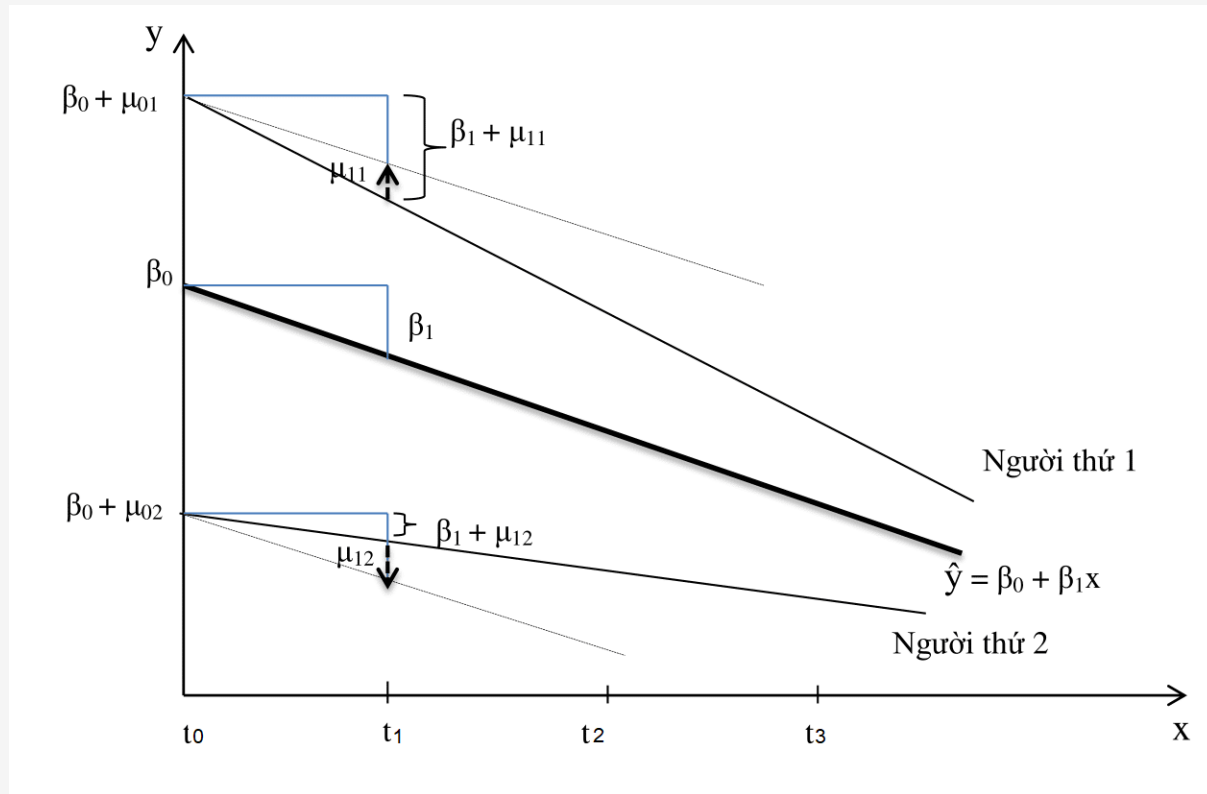
- Trong đó  $i$  là lần đo thứ  $i$  của người thứ  $j$ . Phương trình [4] cũng có thể viết lại theo phần cố định và phần ngẫu nhiên

$$y_{ij} = (\beta_0 + \beta_1 x_{ij}) + (\mu_{1j} x_{ij} + \mu_{0j} + e_{ij}) \quad [5]$$

Với:  $\mu_{0j} \sim N(0, \sigma_{\mu_0}^2)$ ,  $\mu_{1j} \sim N(0, \sigma_{\mu_1}^2)$  và  $e_{ij} \sim N(0, \sigma_e^2)$



# Mô hình random slope



# Mô hình random slope

---

- Trở lại ví dụ thuốc điều trị đái tháo đường, mức độ giảm đường huyết là phần độ dốc
- Mô hình random slope còn quan tâm đến mối liên hệ giữa hai phần random intercept và random slope, còn gọi là hiệp biến (covariance), hiệp biến có thể nhận ba loại giá trị: âm, dương và bằng 0
- Dương: điểm chặn lớn và có độ dốc cũng lớn
- Âm: điểm chặn lớn và có độ dốc nhỏ
- Bằng 0: mức độ giảm đường huyết không có mối liên hệ nào với mức độ đường huyết ban đầu

# Ví dụ

---

- bộ dữ liệu “*Isoproterenol.dta*”
- Nghiên cứu đánh giá tác động của isoproterenol, một chất chủ vận  $\beta_2$ -adrenergic, đến lưu lượng máu cẳng tay trên 22 người đàn ông khỏe mạnh. Trong đó có 9 người da đen và 13 người da trắng. Lưu lượng máu cẳng tay (ml/min/dl) của mỗi người được đo lường tại thời điểm ban đầu và tại những thời điểm tăng liều isoproterenol (các liều 10, 20, 60, 150, 300 và 400 (ng/min))
- Câu hỏi nghiên cứu
  - ✓ Liều isoproterenol có ảnh hưởng tới lưu lượng máu ở cẳng tay (fbf) hay không?
  - ✓ Mức độ đáp ứng với isoproterenol có khác nhau giữa hai nhóm màu da (da đen và da trắng) hay không?

# Ví dụ

	id	race	fbf0	fbf10	fbf20	fbf60	fbf150	fbf300	fbf400
1.	1	0	1	1.4	6.4	19.1	25	24.6	28
2.	2	0	2.1	2.8	8.3	15.7	21.9	21.7	30.1
3.	3	0	1.1	2.2	5.7	8.2	9.3	12.5	21.6
4.	4	0	2.44	2.9	4.6	13.2	17.3	17.6	19.4
5.	5	0	2.9	3.5	5.7	11.5	14.9	19.7	19.3
6.	6	0	4.1	3.7	5.8	19.8	17.7	20.8	30.3
7.	7	0	1.24	1.2	3.3	5.3	5.4	10.1	10.6
8.	8	0	3.1	.	.	15.45	.	.	31.3
9.	9	0	5.8	8.8	13.2	33.3	38.5	39.8	43.3
10.	10	0	3.9	6.6	9.5	20.2	21.5	30.1	29.6
11.	11	0	1.91	1.7	6.3	9.9	12.6	12.7	15.4
12.	12	0	2	2.3	4	8.4	8.3	12.8	16.7
13.	13	0	3.7	3.9	4.7	10.5	14.6	20	21.7
14.	14	1	2.46	2.7	2.54	3.95	4.16	5.1	4.16
15.	15	1	2	1.8	4.22	5.76	7.08	10.92	7.08
16.	16	1	2.26	3	2.99	4.07	3.74	4.58	3.74
17.	17	1	1.8	2.9	3.41	4.84	7.05	7.48	7.05
18.	18	1	3.13	4	5.33	7.31	8.81	11.09	8.81
19.	19	1	1.36	2.7	3.05	4	4.1	6.95	4.1
20.	20	1	2.82	2.6	2.63	10.03	9.6	12.65	9.6
21.	21	1	1.7	1.6	1.73	2.96	4.17	6.04	4.17
22.	22	1	2.1	1.9	3	4.8	7.4	16.7	21.2

# Ví dụ

---

- Dữ liệu đang ở dạng rộng/ngang (wide form) bao gồm 9 cột và 22 hàng
- id Mã số người tham gia nghiên cứu, gồm 22 người
- race Màu da (0 = da trắng, 1 = da đen)
- fbf0 Lưu lượng máu cẳng tay (ml/min/dl) tại liều isoproterenol = 0 ng/min
- fbf10 Lưu lượng máu cẳng tay (ml/min/dl) tại liều isoproterenol = 10 ng/min
- .....
- fbf400 Lưu lượng máu cẳng tay (ml/min/dl) tại liều isoproterenol = 400 ng/min

# Các bước

---

- Mô tả, thăm dò số liệu, vẽ biểu đồ
- Xây dựng mô hình random intercept
- Kiểm tra và thêm random slope
- Kiểm tra và thêm các ảnh hưởng khác vào mô hình như biến gây nhiễu, ảnh hưởng tương tác...

# Mô tả, thăm dò số liệu

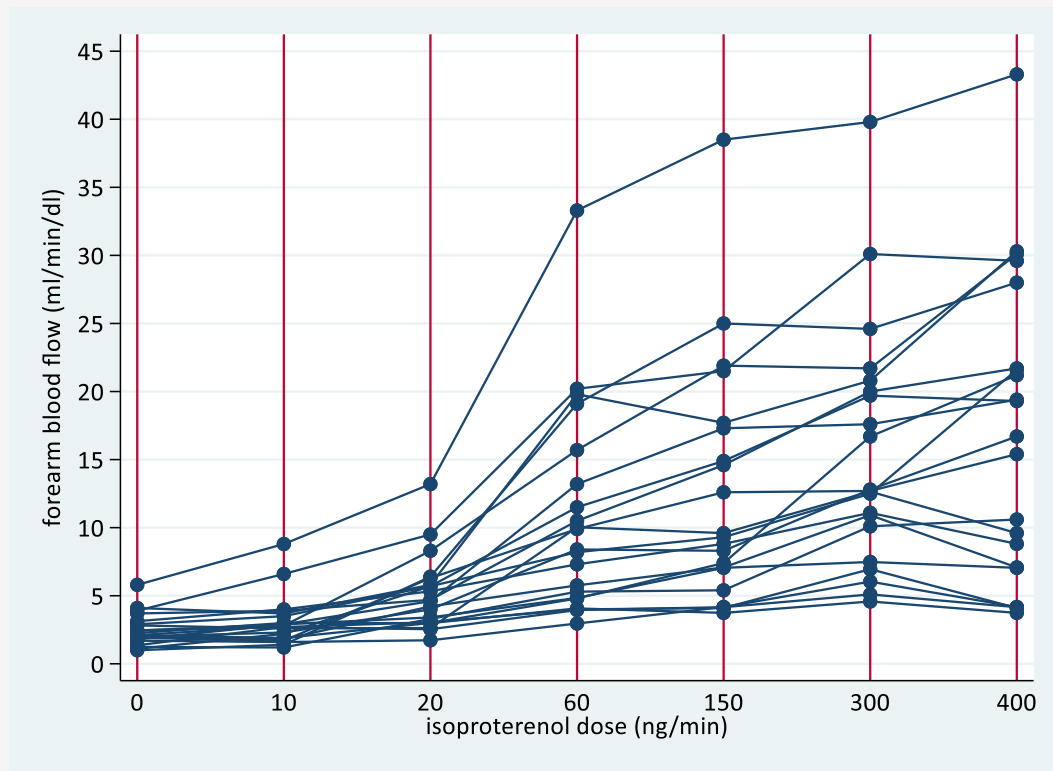
---

- Cài đặt gói “parplot” để vẽ biểu đồ

```
ssc install parplot
```

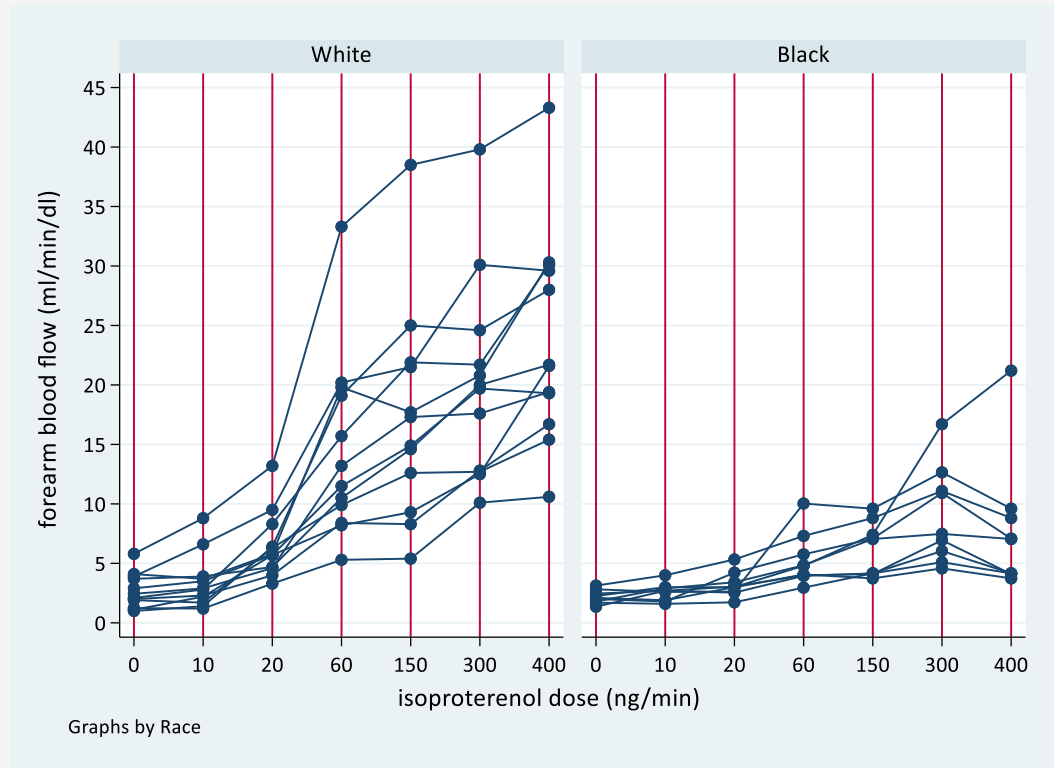
```
parplot fbf0-fbf400, transform(raw) xlabel(1 "0" 2 "10" 3 "20" 4  
"60" 5 "150" 6 "300" 7 "400") ylabel(0(5)45, angle(horizontal))  
yttitle("forearm blood flow (ml/min/dl)") xtitle("isoproterenol  
dose (ng/min)")
```

```
parplot fbf0-fbf400, transform(raw) xlabel(1 "0" 2 "10" 3 "20" 4  
"60" 5 "150" 6 "300" 7 "400") ylabel(0(5)45, angle(horizontal))  
yttitle("forearm blood flow (ml/min/dl)") xtitle("isoproterenol  
dose (ng/min)") by(race)
```





# Theo màu da



# Chuyển đổi cấu trúc số liệu

- Dữ liệu đang ở dạng ngang → chuyển dạng dọc  
`reshape long fbf, i(id) j(dose)`

	id	dose	race	fbf
1.	1	0	0	1
2.	1	10	0	1.4
3.	1	20	0	6.4
4.	1	60	0	19.1
5.	1	150	0	25
6.	1	300	0	24.6
7.	1	400	0	28
8.	2	0	0	2.1
9.	2	10	0	2.8
10.	2	20	0	8.3
11.	2	60	0	15.7
12.	2	150	0	21.9
13.	2	300	0	21.7
14.	2	400	0	30.1

# Mô hình linear mixed-effect models trong Stata

---

- Cấu trúc lệnh

```
mixed {biếnphụthuộc} {biếnđộc lập} || {random_intercept} :  
{random_slope}, [options]
```

- Ví dụ “isoproterenol”

- ✓ Random intercept là biến gì?
- ✓ Random slope là biến gì?

# Random intercept

```
mixed fbf dose || id:
```

$$y_{ij} = (\beta_0 + \mu_{0j}) + \beta_1 x_{ij} + e_{ij}$$


Mixed-effects ML regression  
Group variable: id

Number of obs = 150  
Number of groups = 22

Obs per group:  
min = 3  
avg = 6.8  
max = 7

Log likelihood = -476.95216  
Wald chi2(1) = 163.42  
Prob > chi2 = 0.0000

**Phần cố định (fixed)**

fbf	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
dose	.0352574	.0027581	12.78	0.000	.0298517	.0406631
_cons	4.966772	1.24597	3.99	0.000	2.524717	7.408828

**Phần ngẫu nhiên (random)**

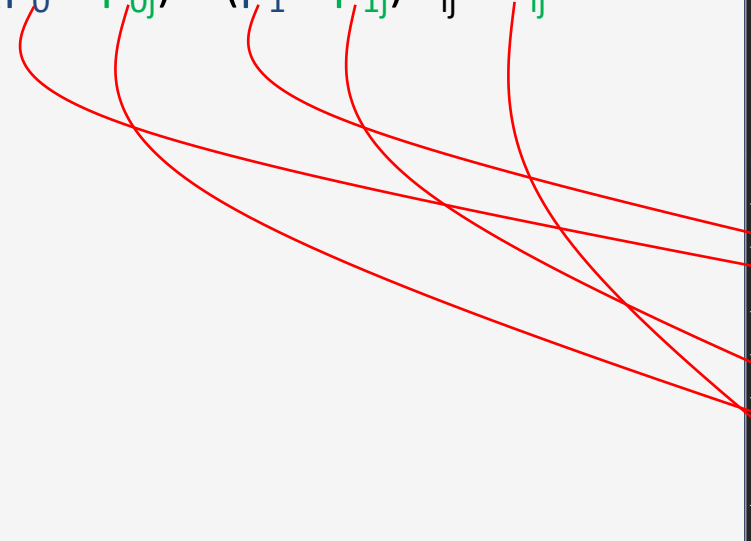
Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
id: Identity				
var(_cons)	27.38189	9.390553	13.98124	53.62668
var(Residual)	24.73093	3.091446	19.35697	31.59682

LR test vs. linear model: chibar2(01) = 64.59      Prob >= chibar2 = 0.0000

# Random slope

- Độ dốc sự thay đổi fbf là khác nhau qua liều lượng khác nhau

`mixed fbf dose || id: dose, cov(unstructured)`

$$y_{ij} = (\beta_0 + \mu_{0j}) + (\beta_1 + \mu_{1j}) x_{ij} + e_{ij}$$


```
Computing standard errors:
Mixed-effects ML regression      Number of obs   =    150
Group variable: id              Number of groups =    22

                                Obs per group:
                                min =     3
                                avg =    6.8
                                max =     7

                                Wald chi2(1)   =    50.84
                                Prob > chi2    =    0.0000

Log likelihood = -437.38499
```

fbf	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
dose	.0356267	.0049967	7.13	0.000	.0258334	.0454199
_cons	4.929402	.6681074	7.38	0.000	3.619935	6.238868

Random-effects Parameters		Estimate	Std. Err.	[95% Conf. Interval]	
id: Unstructured					
	var(dose)	.0004634	.0001642	.0002314	.0009279
	var(_cons)	6.388516	2.791472	2.713102	15.04298
	cov(dose,_cons)	.0544077	.	.	.
	var(Residual)	12.67931	1.585114	9.92389	16.19979

LR test vs. linear model: chi2(3) = 143.72      Prob > chi2 = 0.0000

# Câu hỏi 1

---

- Liệu isoproterenol có ảnh hưởng tới lưu lượng máu ở cẳng tay hay không?
  - ✓ liều lượng isoproterenol có ảnh hưởng thuận tới lưu lượng máu cẳng tay (cứ tăng liều isoproterenol lên 1 ng/min thì trung bình lưu lượng máu cẳng tay tăng lên 0,036 ml/min/dl)
  - ✓ Những người có lưu lượng máu cẳng tay cao hơn có xu hướng đáp ứng mạnh hơn với liều isoproterenol

## Câu hỏi 2

---

- Mức độ đáp ứng với isoproterenol có khác nhau giữa hai nhóm màu da (da đen và da trắng) hay không?
  - Hay nhóm màu da có thay đổi ảnh hưởng của isoproterenol lên fbf hay không?
  - Nhóm màu da có phải là effect modifier hay không?
- Interaction giữa liều isoproterenol và màu da

---

- Sử dụng ##

```
mixed fbf race## c.dose || id: dose, cov(unstructured)
```



- Interaction giữa isoproterenol và màu da = -0.03 ( $p < 0.001$ )
- Mức độ đáp ứng với isoproterenol có khác nhau giữa hai nhóm màu da. Trong đó những người da đen có mức độ đáp ứng với isoproterenol thấp hơn những người da trắng

Log likelihood = -429.13311					Wald chi2(3) = 128.50	Prob > chi2 = 0.0000
fbf	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
race						
Black	-3.01543	1.203554	-2.51	0.012	-5.374352	-.6565081
dose	.0497674	.004543	10.95	0.000	.0408634	.0586715
race#c.dose						
Black	-.0345474	.0070824	-4.88	0.000	-.0484286	-.0206661
_cons	6.161752	.7749015	7.95	0.000	4.642974	7.680531
Random-effects Parameters		Estimate	Std. Err.	[95% Conf. Interval]		
id: Unstructured						
	var(dose)	.0001824	.000077	.0000798	.0004171	
	var(_cons)	4.342912	2.109244	1.676385	11.25093	
	cov(dose,_cons)	.0281484	.0093923	.0097398	.0465569	
	var(Residual)	12.51898	1.564474	9.799323	15.99345	
LR test vs. linear model: chi2(3) = 86.84					Prob > chi2 = 0.0000	

# Nghiên cứu bán thực nghiệm (Quasi-experimental study)

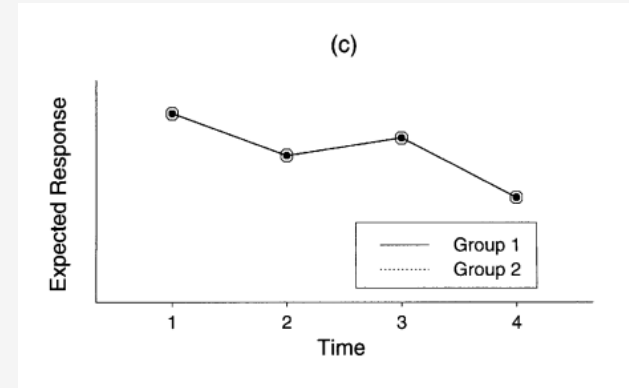
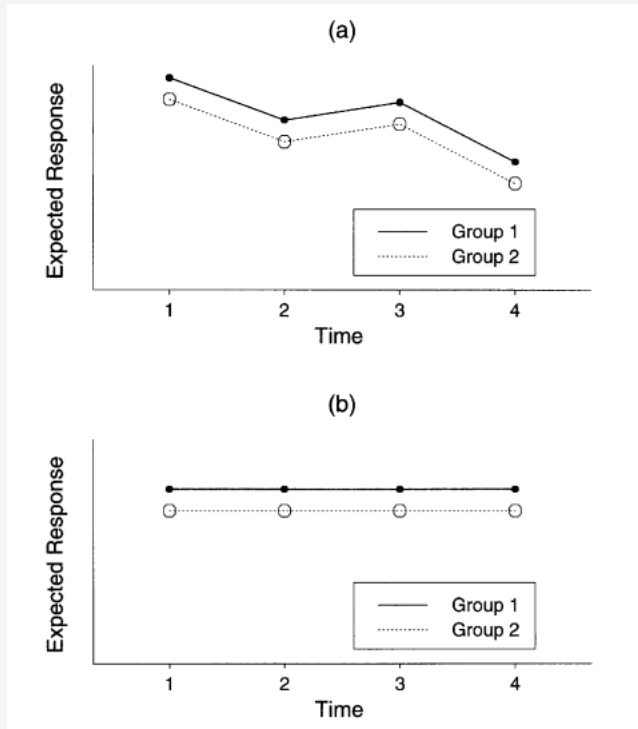
---

- Một dạng của nghiên cứu thuần tập
- Đánh giá hiệu quả của can thiệp (có nhóm chứng)
- Các đối tượng nghiên cứu không được phân bổ ngẫu nhiên

# Các giả thuyết trong phân tích

---

- Time effect: Giá trị đo lường khác nhau qua thời gian hay không?
- Group effect: Giá trị đo lường khác biệt giữa các nhóm so sánh hay không?
- **Treatment effect**: Sự thay đổi qua thời gian của các giá trị đo lường có khác biệt giữa các nhóm so sánh hay không? (*thời gian \* nhóm so sánh*)



## Ví dụ 2

---

- Dữ liệu “BP\_management.dta”
  - ✓ Nghiên cứu bán thực nghiệm “Communities for Healthy Hearts” do trường YTCC thực hiện
  - ✓ Đối tượng nghiên cứu bệnh nhân tăng huyết áp
  - ✓ Đánh giá hiệu của chương trình can thiệp cộng đồng
  - ✓ Biến đầu ra: sử dụng công cụ hỗ trợ kiểm soát huyết áp (1/0)
  - ✓ (Dữ liệu thực hành đã được loại bỏ một số lần đo và mẫu khảo sát so với dữ liệu đã công bố)

## Ví dụ 2

---

**id**: mã cá nhân

**BP\_tools**: biến outcome, sử dụng công cụ (có/không)

**trt**: nhóm can thiệp (can thiệp/ chứng)

**round**: đo lường lặp lại (dữ liệu thực hành gồm 4 vòng)

Các biến số đặc điểm cá nhân: **nhomtuoi gender edu job marriage smoke nonalc**

# Câu hỏi nghiên cứu

---

- Can thiệp cộng đồng có hiệu quả trong việc tăng cường sử dụng công cụ hỗ trợ kiểm soát huyết áp theo thời gian trên bệnh nhân tăng huyết áp hay không?  
(*treatment effect*)

- 
- Gợi ý

## Sử dụng mô hình logistic mixed-effect model

```
melogit {biếnphụthuộc} {biếnđộc lập} || {random_intercept} :  
{random_slope}, or [options]
```

Random intercept là biến số gì?

Random slope là biến số gì?



**Phần 3: Một số phương pháp hiệu chỉnh cho sự mất cân bằng giữa các nhóm so sánh trong nghiên cứu bán thực nghiệm (đọc thêm)**

# Vấn đề

---

- Do đối tượng nghiên cứu trong nghiên cứu bán thực nghiệm không được phân bổ ngẫu nhiên → sự khác biệt về đặc điểm ở baseline → cần được hiệu chỉnh

```
. table1 if round ==0, by(trt) vars(nhomtuoi cat\ gender cat\ edu cat\ job cat\ marriage cat\ smoke cat\ nonalc cat\) format(%9.1) one
```

Factor	Control	Intervention	p-value
N	1274	1265	
Age group			<0.001
40-49 year	409 (32.1%)	293 (23.2%)	
50-59 year	419 (32.9%)	487 (38.5%)	
60-69 year	446 (35.0%)	485 (38.3%)	
Gender			0.71
Male	627 (49.2%)	632 (50.0%)	
Female	647 (50.8%)	633 (50.0%)	
Education			<0.001
Illiteracy	45 (3.5%)	17 (1.3%)	
Primary school and below	420 (33.0%)	271 (21.4%)	
Secondary school completed	430 (33.8%)	424 (33.5%)	
High school completed	268 (21.0%)	368 (29.1%)	
College/University and above	110 (8.6%)	175 (13.8%)	
Unknown/refused	1 (0.1%)	10 (0.8%)	
Occupation			<0.001
Business owner	59 (4.6%)	88 (7.0%)	
Retired	128 (10.0%)	211 (16.7%)	
Housewife/ househusband	377 (29.6%)	328 (25.9%)	
Industrial labourer	80 (6.3%)	113 (8.9%)	
Freelancer	312 (24.5%)	298 (23.6%)	
Other	318 (25.0%)	227 (17.9%)	
Marital status			0.32
Currently married	1060 (83.2%)	1071 (84.7%)	
Other	214 (16.8%)	194 (15.3%)	
Smoking status			0.007
Non-smoker	880 (69.1%)	879 (69.5%)	
Formal smoker	79 (6.2%)	115 (9.1%)	
Current smoker	315 (24.7%)	271 (21.4%)	
Ever consume alcohol			0.002
Never	562 (44.1%)	483 (38.2%)	
Ever	712 (55.9%)	782 (61.8%)	

# Một số phương pháp hiệu chỉnh cân bằng giữa các nhóm so sánh

---

- Mô hình đa biến
- Propensity score matching (PSM)
- Propensity score inverse-probability weighting (IWP)

# Tài liệu tham khảo

---

