

Bài 5

MỘT SỐ ĐIỀU CẦN TRÁNH & TĂNG SỐ LƯỢNG QUAN SÁT



Vũ Thành Tự Anh

Trường Chính sách công và Quản lý Fulbright

MỘT SỐ ĐIỀU CẦN TRÁNH



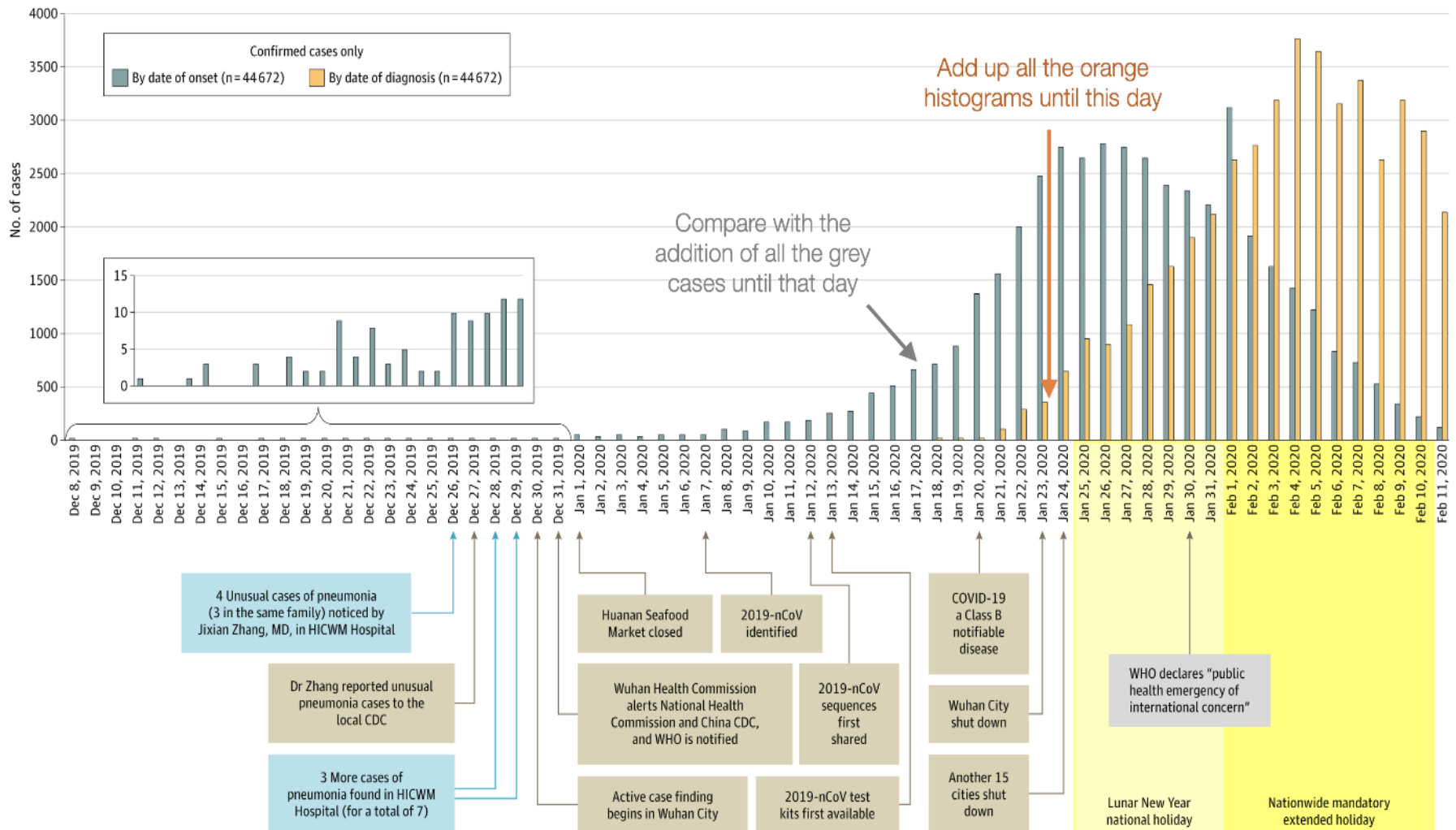
- ❧ Sai số đo lường
- ❧ Bỏ sót biến quan trọng
- ❧ Thừa biến không liên quan
- ❧ Biến nội sinh

Nhắc lại tiêu chí đánh giá suy luận mô tả và suy luận nhân quả

- ❧ Thiết kế nghiên cứu **xác định**
- ❧ **Nhất quán** nội tại
- ❧ **Không thiên lệch**: Trung tâm của khoảng giá trị gần với giá trị ước lượng đúng
- ❧ **Hiệu quả**: Thu hẹp khoảng giá trị ước lượng

1. Sai số đo lường: Khi ghi nhận khác thực tế

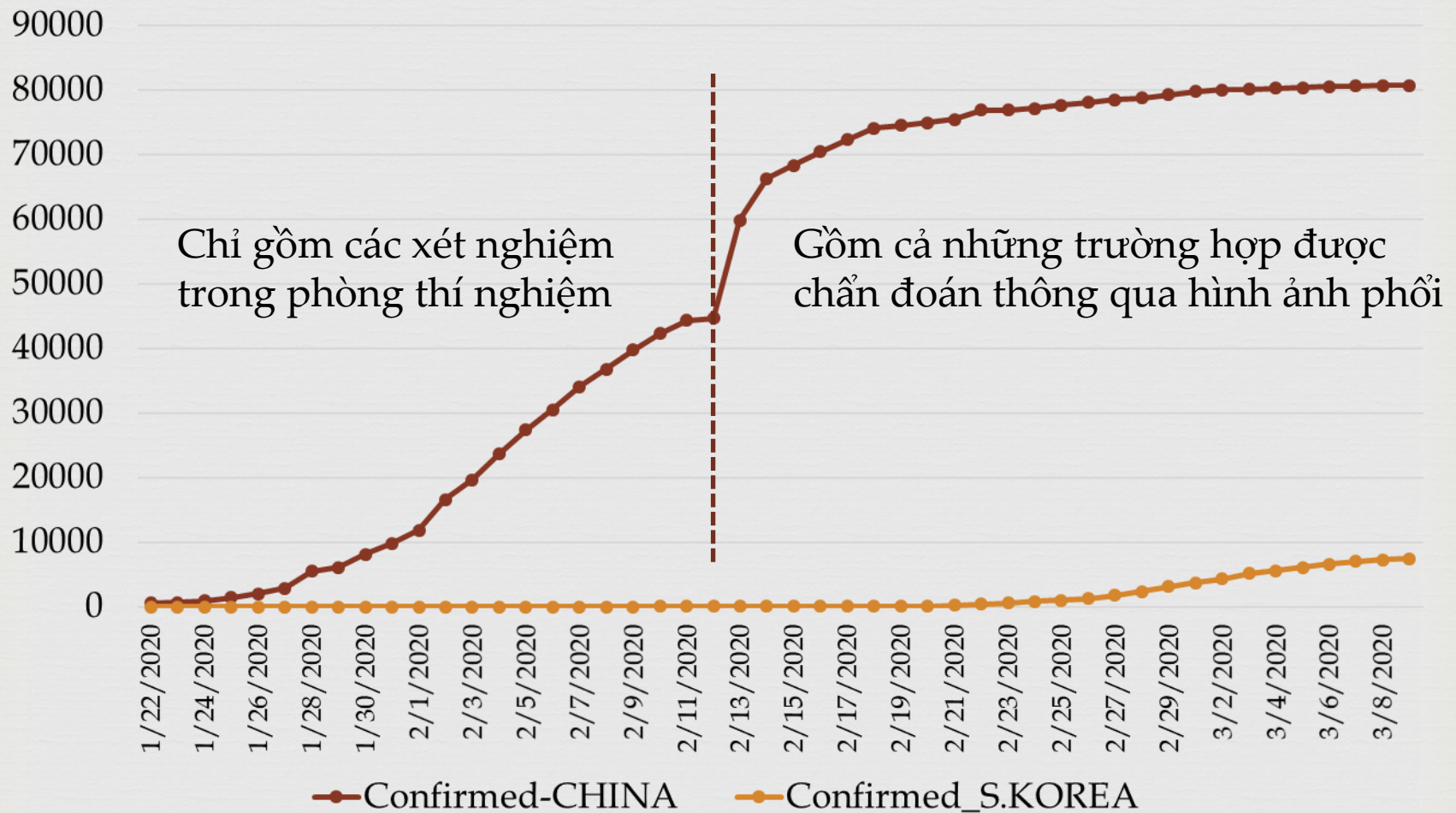
Chart 11: Timeline of Events in Hubei



Source: Tomas Pueyo analysis over chart and data from the [Journal of the American Medical Association](#)

1. Sai số đo lường

[Ví dụ: Số ca ghi nhận nhiễm COVID-19]



1. Sai số đo lường

[Các loại thang đo định lượng]

- ❧ Thang đo *biểu danh* (nominal): Sử dụng cho dữ liệu mà thuộc tính của chúng không ngầm định một sự hơn kém hay khác biệt về thứ bậc (VD giới tính, tôn giáo, sắc tộc). Số liệu có tính mã hóa, không dùng để tính toán đại số.
- ❧ Thang đo *thứ bậc* (ordinal): Thứ bậc phản ánh sự khác biệt, song khoảng cách giữa các bậc lại không có nhiều ý nghĩa (VD trình độ học vấn). Thang đo này không dùng trong các phép tính đại số
- ❧ Thang đo *khoảng* (interval): Dùng để đo các đại lượng liên tục, trong đó khoảng cách bằng nhau trên thang đo đại diện cho khoảng cách bằng nhau của thuộc tính (VD điểm thi). Phép chia đại số không có ý nghĩa.
- ❧ Thang đo *tỷ lệ* (ratio): Cũng dùng để đo các đại lượng liên tục, nhưng phép chia đại số lại có ý nghĩa (VD GDP)

Sai số đo lường

[Sự khác biệt giữa đo lường định tính và định lượng]

❧ **Định lượng:** Mọi thang đo đều bằng số

❧ **Định tính:**

❧ Thang đo không nhất thiết bằng số (VD chế độ chính trị, hình thức sở hữu, hay mức độ phân cấp giữa chính quyền trung ương và địa phương)

❧ Những thang đo gần đúng (proxy) không nhất thiết phản ánh đúng thuộc tính (VD số lượng thương vong không nhất thiết phản ánh tầm quan trọng của cuộc chiến tranh)

❧ Điều *then chốt* là nên sử dụng thang đo phù hợp nhất với mục tiêu nghiên cứu.

Tác động của sai số đo lường

[Sai số hệ thống sv. ngẫu nhiên]

	Sai số có tính hệ thống	Sai số có tính ngẫu nhiên
Biến phụ thuộc	• Làm thiên lệch suy luận mô tả • Không làm thiên lệch suy luận nhân quả	• Làm thiên lệch suy luận mô tả • Không làm thiên lệch suy luận nhân quả • Giảm hiệu quả ước lượng nhân quả
Biến độc lập	• Làm thiên lệch suy luận mô tả • Không làm thiên lệch suy luận nhân quả	• Làm thiên lệch suy luận mô tả • Làm suy yếu suy luận nhân quả • Giảm hiệu quả ước lượng nhân quả

2. Bỏ sót biến quan trọng



- ❧ **Bỏ sót biến** sẽ không ảnh hưởng đến kết quả ước lượng nếu:
 - ❧ Biến (bị sót) không tương quan với biến phụ thuộc
 - ❧ Biến bị sót tuy có tương quan với biến phụ thuộc song không tương quan với (các) biến giải thích khác trong mô hình. Khi ấy:
 - ❧ Ước lượng quan hệ giữa (các) biến giải thích và biến phụ thuộc không bị ảnh hưởng
 - ❧ Nhưng mức độ chính xác trong dự báo giá trị của biến phụ thuộc sẽ bị giảm.
 - ❧ Ví dụ [Xé rào ưu đãi FDI]: Khoảng cách đến thị trường chính và chỉ số PCI không có tương quan.
- ❧ Trong các trường hợp còn lại, bỏ sót biến (quan trọng) sẽ dẫn tới ước lượng thiên lệch.
 - ❧ Ví dụ: Khi biến bỏ sót tương quan với cả biến phụ thuộc lẫn biến độc lập khác trong mô hình.

Bỏ sót biến quan trọng



❧ Ví dụ bạn muốn ước lượng giá của xe hơi cũ:

$$\text{Giá} = \beta_1 + \beta_2 \cdot \text{Số dặm} + \beta_3 \cdot \text{Tuổi xe}$$

❧ Giả sử trong ước lượng của mình, bạn bỏ quên biến “tuổi xe”:

$$\text{Giá} = \beta_1 + \beta_2 \cdot \text{Số dặm}$$

❧ Ước lượng này có bị thiên lệch không?

❧ Có - vì tuổi xe tương quan với cả “giá” và “số dặm”

❧ Chiều hướng của thiên lệch này sẽ như thế nào?

❧ Nó sẽ làm cho β_2 nhận giá trị âm lớn hơn vì bây giờ β_2 phản ánh cả tác động (tiêu cực) của “tuổi xe” đến “giá”

Bỏ sót biến quan trọng

[Chiều hướng thiên lệch]

❧ Giả sử Y là biến phụ thuộc, A và B là 2 biến giải thích, và B là biến bị bỏ quên trong mô hình.

	B và A tương quan (+)	B và A tương quan (-)
B và Y tương quan (+)	Thiên lệch (+)	Thiên lệch (-)
B và Y tương quan (-)	Thiên lệch (-)	Thiên lệch (+)

Reading: Vietnam is the 94th happiest country in the world

Vietnam is the 94th happiest country in the world

By Ngoc Dinh March 21, 2019 | 11:34 am GMT+7



A young man in Hanoi celebrated after the Vietnamese football team won against

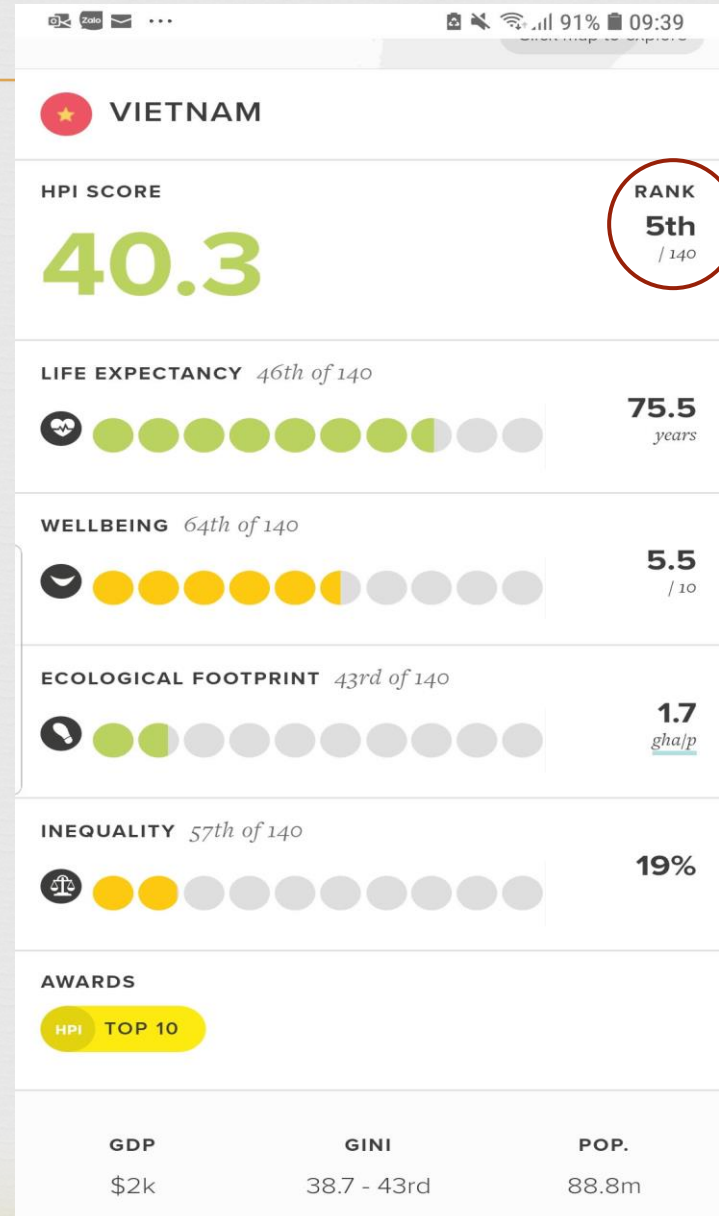
Reading: Vietnam climbs up happiness index ranking

Vietnam climbs up happiness index ranking

By Nguyen Quy March 24, 2020 | 03:25 pm GMT+7



A girl sits on her dad's back to watch fireworks on Lunar New Year's Eve on Nguyen Hue walking street in HCMC, January 24, 2020. Photo by VnExpress/Quynh Tran.



3. Thừa biến không liên quan



- ❧ Việc đưa thêm quá nhiều biến liên quan không phải là giải pháp cho nguy cơ bỏ sót biến quan trọng.
- ❧ Cơ sở của mô hình (cả định lượng và định tính) thường là những lý thuyết đã được kiểm chứng.
- ❧ Khi số lượng quan sát ít, đưa quá nhiều biến giải thích sẽ có nguy cơ biến mô hình trở thành “không xác định”.
- ❧ Đưa biến không liên quan vào mô hình tuy không làm thiên lệch ước lượng, song làm giảm hiệu quả của ước lượng, đặc biệt là khi biến này có tương quan với biến giải thích chính.

4. Nội sinh



- ❧ Vấn đề **nội sinh**: Giá trị của biến giải thích là hệ quả chứ không phải là nguyên nhân của biến phụ thuộc
- ❧ Ví dụ 1: “Matthew Effect” hay “nước chảy chỗ trũng”
 - ❧ Đồng tác giả trong 1 bài báo hay 1 dự án
- ❧ Ví dụ 2: “Self-selection” hay “đồng thanh tương ứng”
 - ❧ Học viên Fulbright
- ❧ Ví dụ 3: Chi KH-CN với năng lực KH-CN

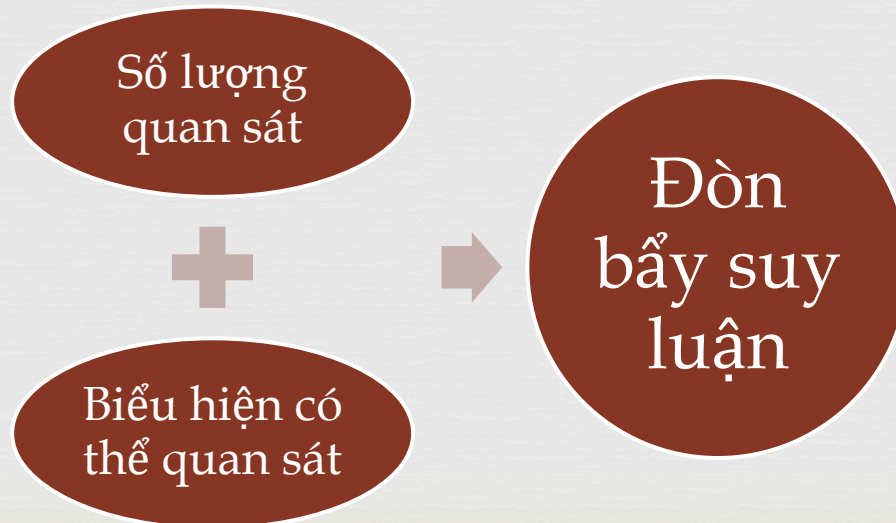
TĂNG SỐ LƯỢNG QUAN SÁT

- ❧ Tối đa hóa đòn bẫy suy luận
- ❧ Vấn đề của nghiên cứu chỉ dựa vào 1 quan sát
- ❧ Bao nhiêu quan sát là đủ?
- ❧ Làm thế nào để tăng số lượng quan sát?

Tối đa hóa đòn bẫy suy luận



- ❧ Một nhiệm vụ của thiết kế nghiên cứu là *tối đa hóa “đòn bẫy”* cho suy luận (mô tả và nhân quả).
- ❧ Để thực hiện mục tiêu này, biểu hiện có thể quan sát được *càng nhiều* và *càng mạnh* thì càng tốt.
- ❧ *Số lượng quan sát* càng nhiều thì khả năng tối đa hóa đòn bẫy càng cao.



Vấn đề với những nghiên cứu chỉ dựa trên 1 quan sát

- ❧ Không kiểm định được *giả thuyết thay thế*
- ❧ Rủi ro cao nếu gặp *sai số đo lường*
- ❧ *Thế giới xác suất* (ngẫu nhiên sv. tất định)

Làm thế nào để khắc phục các nhược điểm này?

- ❧ **Suy luận tương tự** (reasoning by analogy) đòi hỏi giả định quá mạnh – tất cả các biến số khác giữ nguyên
- ❧ **Suy luận so sánh**
- ❧ **Tăng số quan sát và tăng số biểu hiện** có thể quan sát được

Bao nhiêu quan sát là đủ?



❧ **Số lượng quan sát cần thiết** phụ thuộc vào:

- ❧ Nghiên cứu định tính hay định lượng
- ❧ Thiết kế nghiên cứu
- ❧ Suy luận nhân quả cần thực hiện
- ❧ Phạm vi và mức độ các nhân tố ngoài tầm kiểm soát

❧ **Kích thước mẫu cần thiết** phụ thuộc vào:

- ❧ Mức độ bất định của biến phụ thuộc
- ❧ Mức độ chắc chắn mong muốn của suy luận nhân quả
- ❧ Mức độ cộng tuyến giữa biến xử lý và các biến kiểm soát
- ❧ Mức độ biến thiên của biến giải thích chính

Làm thế nào để tăng số quan sát?



- ❧ Một quan sát = một đo lường của biến phụ thuộc cho một đơn vị phân tích (và của các biến giải thích tương ứng của đơn vị phân tích đó)
- ❧ Để tăng số quan sát, 2 câu hỏi đầu tiên cần trả lời là:
 - ❧ Đây là những biểu hiện có thể quan sát được của lý thuyết/giả thuyết?
 - ❧ Đây là những trường hợp trong đó ta có thể kiểm chứng được những biểu hiện có thể quan sát này?
- ❧ Cách tăng số lượng quan sát:
 - ❧ Tăng số đơn vị phân tích
 - ❧ Dùng thêm đo lường khác
 - ❧ Đơn vị phân tích mới & đo lường mới

Đo lường cũ, đơn vị phân tích mới



- ❧ Sử dụng cùng một lý thuyết (giả thuyết), sử dụng cùng một tập hợp các biến giải thích và phụ thuộc, nhưng áp dụng cho các trường hợp mới.
- ❧ **Xuyên không gian:** Ví dụ như
 - ❧ Các tỉnh thành của Việt Nam (sub-units)
 - ❧ Các quốc gia đang phát triển ...
- ❧ **Xuyên thời gian**
 - ❧ Các giai đoạn khác nhau (lưu ý tính nhất quán)
- ❧ Ví dụ: Bảo hộ khu vực DNNN khi hội nhập
 - ❧ Mức độ bảo hộ ở các ngành khác nhau
 - ❧ Mức độ bảo hộ trong các thể hệ hiệp định thương mại tự do khác nhau

Cùng đơn vị phân tích, đo lường mới



- ❧ Cách tiếp cận này tìm kiếm các tác động khác nhau của cùng một nguyên nhân.
- ❧ Tìm thêm biến phụ thuộc để kiểm định giả thuyết
- ❧ Ví dụ: Tác động của chính sách cải thiện môi trường kinh doanh đối với kết quả hoạt động kinh tế của các địa phương: Có thể đo lường kết quả bằng:
 - ❧ Tốc độ tăng trưởng GDP
 - ❧ GDP bình quân đầu người
 - ❧ Thu nhập bình quân đầu người
 - ❧ Chi tiêu bình quân đầu người
 - ❧ Tỷ lệ tăng lao động/dân số (cơ học) v.v.

Đơn vị phân tích mới, đo lường mới



- ❧ Ví dụ 1: Mối quan hệ giữa giá đền bù giải tỏa đất nông nghiệp và sự phản kháng của người dân ở VN
 - ❧ Đơn vị phân tích thấp hơn: tỉnh/thành phố | quận, huyện
 - ❧ Sự phản kháng có thể được đo bằng nhiều thước đo (số đơn khiếu kiện, khiếu kiện vượt cấp, biểu tình ...)
 - ❧ Sự phản kháng ngoài nguyên nhân giá đền bù thấp còn có thể do một số *nhân tố có tính địa phương*
- ❧ Ví dụ 2: Giải thích tình trạng bỏ học giữa các cấp của học sinh phổ thông ở Việt Nam
 - ❧ Đơn vị phân tích thấp hơn
 - ❧ Việc học sinh bỏ học có thể do *nhân tố có tính địa phương* (văn hóa, tập quán)