

MÔN HỌC

THỐNG KÊ ỨNG DỤNG (KC107)

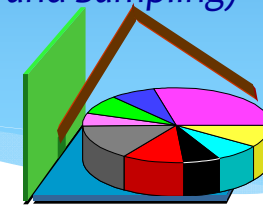
GIÁO VIÊN PHỤ TRÁCH

ĐẶNG THỂ GIA

Bộ môn Kỹ Thuật Xây Dựng
Khoa Công Nghệ, Trường Đại Học Cần Thơ

Chương 5:

THU THẬP DỮ LIỆU & CHỌN MẪU
(Data Collection and Sampling)



Nội dung chương

- 1. Giới thiệu**
- 2. Nguồn dữ liệu** (Sources of data)
- 3. Chọn mẫu** (Sampling)
- 4. Phương pháp & Kế hoạch chọn mẫu**
(Sampling Methods & Plans)
- 5. Các lỗi trong chọn mẫu** (Errors in Sampling)

GIỚI THIỆU
INTRODUCTION

Thu thập dữ liệu

- Là quá trình gom góp và đo lường thông tin về các biến số quan tâm theo một cách làm được thiết lập có hệ thống, cho phép người ta trả lời các câu hỏi nghiên cứu đã được nêu, các kiểm định các giả thuyết, và đánh giá kết quả.
- Việc thu thập dữ liệu của nghiên cứu là phổ biến trong tất cả các lĩnh vực nghiên cứu bao gồm: khoa học tự nhiên và xã hội, nhân văn, kinh doanh, quản lý, kỹ thuật,...
- Dù nhiều phương pháp khác nhau tùy theo chuyên ngành, nhưng sự nhấn mạnh về đảm bảo thu thập **chính xác** và **trung thực** là giá trị chung.

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

Mẫu ngẫu nhiên - Định nghĩa toán học

Định nghĩa 4.1. Mẫu ngẫu nhiên là 1 dãy n biến ngẫu nhiên (X_1, \dots, X_n) từ phân phối $F(x, \theta)$, n được gọi là kích thước mẫu. Các giá trị của mẫu được ký hiệu bằng chữ thường x_1, \dots, x_n . Về mặt hình học, một mẫu (X_1, \dots, X_n) được xem như 1 điểm trong không gian R^n chiều. Không gian R^n được gọi là không gian mẫu.

Chú ý: Thông thường người ta hay xét mẫu ngẫu nhiên (X_1, \dots, X_n) là dãy các biến ngẫu nhiên độc lập có cùng phân phối.

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

NGUỒN DỮ LIỆU

SOURCES OF DATA

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

Nguồn dữ liệu (Sources of Data)

- Độ tin cậy (reliability) và độ chính xác (accuracy) của dữ liệu ảnh hưởng đến tính hợp lệ (validity) của kết quả phân tích thống kê.
- Độ tin cậy và độ chính xác của dữ liệu phụ thuộc vào phương pháp thu thập dữ liệu.
- Ba trong số các nguồn dữ liệu thống kê phổ biến nhất là:
 - Dữ liệu đã xuất bản (dữ liệu thứ cấp)
 - Nghiên cứu quan sát
 - Nghiên cứu thực nghiệm

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

Dữ liệu đã xuất bản (Published Data)

- Đây thường là nguồn dữ liệu ưa thích do chi phí thấp và tiện lợi.
- Dữ liệu được xuất bản được tìm thấy dưới dạng tài liệu in, băng, đĩa và trên Internet.
- Dữ liệu được xuất bản bởi tổ chức.

Dặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

Nghiên cứu quan sát & thực nghiệm (Observational and experimental studies)

- Khi dữ liệu thứ cấp không có sẵn hoặc nhất thiết phải có dữ liệu sơ cấp.
- **Nghiên cứu quan sát** là phép đo thực hiện trên một biến được quan sát và ghi lại mà không kiểm soát bất kỳ yếu tố nào có thể ảnh hưởng đến giá trị của chúng.
- **Nghiên cứu thực nghiệm** là phép đo thực hiện trên một biến được quan sát và ghi lại, đồng thời kiểm soát các yếu tố có thể ảnh hưởng đến các giá trị của chúng.

Dặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

Khảo sát (Surveys)

- Khảo sát thu thập
- Khảo sát có thể
 - Phỏng vấn cá nhân
 - Phỏng vấn qua điện thoại
 - Phiếu trắc nghiệm

Bảng câu hỏi chất lượng cần được thiết kế tốt

- Bảng câu hỏi càng ngắn càng tốt
- Ngắn, đơn giản, từ ngữ đơn giản
- Bắt đầu với các câu hỏi nhân khẩu học để giúp người trả lời khởi đầu thoải mái
- Sử dụng các câu hỏi có hai phương án trả lời (dichotomous questions) và câu hỏi nhiều lựa chọn
- Sử dụng các câu hỏi mở một cách thận trọng
- Tránh sử dụng các câu hỏi hàng đầu
- Kiểm tra trước bảng câu hỏi
- Khi soạn bảng câu hỏi, hãy suy nghĩ về cách bạn dự định sử dụng các dữ liệu thu thập được

Dặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

CHỌN MẪU

SAMPLING

Dặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

Mẫu ngẫu nhiên

- Giả sử ta cần nghiên cứu một tính chất (thông số) nào đó của một quần thể.
- Trong thực tế số phần tử của đám đông quá lớn hoặc vì lý do nào đó không thể khảo sát được toàn bộ quần thể.
- Tuy nhiên chúng ta vẫn muốn có một kết luận chính xác về tính chất (thông số) của của các cá thể/phần tử trong quần thể.
- Khi đó, ta thường chọn ra một tập hợp các phần tử từ đám đông đó đại diện cho đám đông. Tập hợp này được gọi là **mẫu**.

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

Chọn mẫu (Sampling)

- Động lực để tiến hành một thủ tục lấy mẫu:
 - Chi phí
 - Quy mô tổng thể
 - Khả năng phá hoại bản tính của quá trình lấy mẫu
- Kích thước mẫu và mục tiêu tổng thể phải giống nhau.

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

KẾ HOẠCH CHỌN MẪU

SAMPLING PLANS

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

Phương pháp chọn mẫu ngẫu nhiên

1. Mẫu ngẫu nhiên đơn giản
 - Mẫu ngẫu nhiên hoàn lại
 - Mẫu ngẫu nhiên không hoàn lại
2. Mẫu ngẫu nhiên phân tầng
3. Mẫu ngẫu nhiên hệ thống
4. Mẫu ngẫu nhiên cơ học
5. Mẫu “điển hình”
6. Mẫu ngẫu nhiên nhiều nhóm

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

Mẫu ngẫu nhiên đơn giản (Simple random sampling)

- Là một tập hợp con của các cá thể/phần tử (một mẫu) được lựa chọn từ một tập hợp lớn hơn (một quần thể). Mẫu ngẫu nhiên đơn là một kỹ thuật khảo sát không thiên vị.
- Mỗi cá thể/phần tử được chọn ngẫu nhiên và hoàn toàn tình cờ, như vậy mỗi cá thể đều có cùng xác suất được chọn ở các giai đoạn trong quá trình lấy mẫu, và mỗi tập hợp con của k cá thể này có cùng xác suất được chọn mẫu như bất kỳ tập con của k cá thể khác.

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

Mẫu hoàn lại

a) *Mẫu ngẫu nhiên hoàn lại.* Từ tập hợp tổng quát (tập gồm các đối tượng cùng loại) gồm N phần tử ta chọn ngẫu nhiên 1 phần tử, khảo sát và ghi lại kết quả X_1 ; Sau đó trả phần tử đó vào tập tổng quát, rồi chọn ngẫu nhiên phần tử thứ hai khảo sát và ghi lại kết quả X_2 . Sau đó ta trả lại phần tử này vào tập tổng quát và tiếp tục chọn ngẫu nhiên phần tử thứ ba v.v.. Lập lại như thế đến n lần. Ta nhận được dãy kết quả (X_1, \dots, X_n) . Mẫu ngẫu nhiên này được gọi là mẫu ngẫu nhiên hoàn lại. Các X_i được chọn với xác suất như nhau và bằng $\frac{1}{N}$.

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

Mẫu không hoàn lại

b) *Mẫu ngẫu nhiên không hoàn lại.* Từ tập tổng quát gồm N phần tử, ta chọn ngẫu nhiên 1 phần tử, khảo sát và ghi lại kết quả X_1 . X_1 nhận được với xác suất $\frac{1}{N}$. Bỏ phần tử ấy ra ngoài, ta lại chọn ngẫu nhiên phần tử thứ hai từ tập còn lại, khảo sát và ghi kết quả X_2 ; X_2 nhận được với xác suất $\frac{1}{N-1}$. Sau đó ta bỏ phần tử đó ra ngoài và tiếp tục chọn ngẫu nhiên phần tử thứ 3 từ tập còn lại và cứ tiếp tục như thế đến lần thứ n . Ta nhận được mẫu (X_1, X_2, \dots, X_n) với xác suất tương ứng $\frac{1}{N}, \frac{1}{N-1}, \dots, \frac{1}{N-n+1}$. Mẫu ngẫu nhiên này được gọi là mẫu ngẫu nhiên không hoàn lại. Hai loại mẫu a) và mẫu b) được mang một tên chung là *mẫu ngẫu nhiên đơn giản*.

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

Cách thực hiện

- Đối với phương pháp này trước tiên người nghiên cứu cần lập danh sách các đơn vị của tổng thể chung theo một trật tự nào đó ví dụ như lập theo tên, theo quy mô, hoặc theo địa chỉ,... Sau đó đánh STT vào trong danh sách.
- Dùng các phương pháp ngẫu nhiên như rút thăm, dùng bảng số ngẫu nhiên, dùng hàm random của máy tính để chọn ra từng đơn vị trong tổng thể chung vào mẫu.
- Phương pháp này thường vận dụng khi các đơn vị của tổng thể chung nằm ở vị trí địa lý gần nhau, các đơn vị đồng đều nhau về đặc điểm. Phương pháp này thông thường được áp dụng trong quá trình kiểm tra chất lượng sản phẩm trong dây chuyền sản xuất hàng loạt.

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

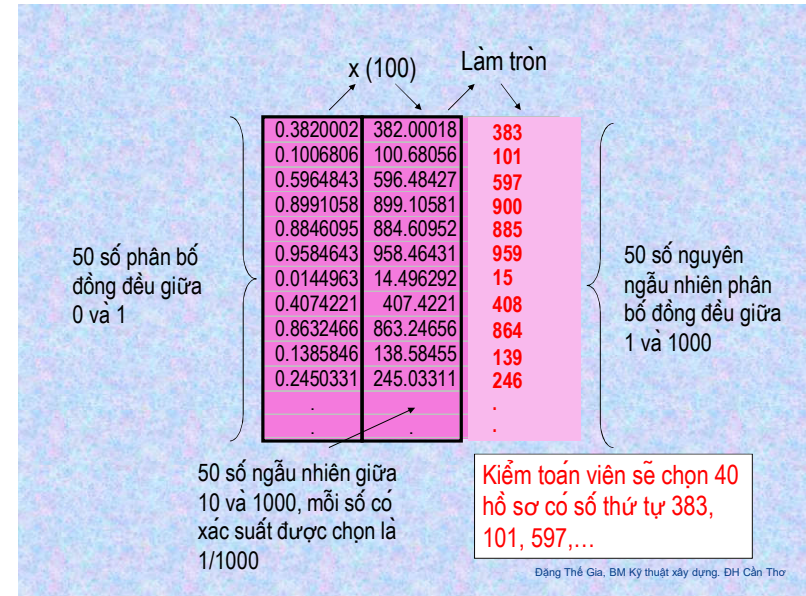
Ví dụ

- Một kiểm toán viên thuế thu nhập của chính phủ chịu trách nhiệm về 1000 tờ khai thuế.
- Kiểm toán viên sẽ lựa chọn ngẫu nhiên 40 mức lợi nhuận để kiểm toán.
- Sử dụng hàm ngẫu nhiên của Excel để chọn tờ khai.

Bài giải

- Lấy ngẫu nhiên 50 số trong khoảng từ 1 đến 1000. Thực ra chỉ cần 40 số cho 40 tờ khai, nhưng ta chọn dư để phòng trường hợp có sự trùng lặp từ các số được tạo ra.

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ



Mẫu ngẫu nhiên phân tầng (Stratified Random Sampling)

- Trước tiên phân chia tổng thể thành các tổ theo 1 tiêu thức hay nhiều tiêu thức có liên quan đến mục đích nghiên cứu (như phân tổ các DN theo vùng, theo khu vực, theo loại hình, theo quy mô,...).
- Sau đó trong từng tổ, dùng cách chọn mẫu ngẫu nhiên đơn giản hay chọn mẫu hệ thống để chọn ra các đơn vị của mẫu.

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

- Phương pháp lấy mẫu này chia quần thể thành các nhóm loại trừ nhau (tầng - strata), sau đó lấy mẫu ngẫu nhiên đơn giản từ mỗi tầng.
- Với thủ tục này, chúng tôi có thể thu thập thông tin về:
 - Toàn bộ tổng thể
 - Mỗi tầng
 - Mối quan hệ giữa các tầng



Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

Mẫu ngẫu nhiên phân tầng (Stratified Random Sampling)

- Đối với chọn mẫu phân tầng, số đơn vị chọn ra ở mỗi tổ có thể tuân theo tỷ lệ số đơn vị tổ đó chiếm trong tổng thể, hoặc có thể không tuân theo tỷ lệ.
- Ví dụ: Một toà soạn báo muốn tiến hành nghiên cứu trên một mẫu 1000 doanh nghiệp trên cả nước về sự quan tâm của họ đối với tờ báo nhằm tiếp thị việc đưa thông tin quảng cáo trên báo. Toà soạn có thể căn cứ vào các tiêu thức: vùng địa lý (miền Bắc, miền Trung, miền Nam) ; hình thức sở hữu (quốc doanh, ngoài quốc doanh, công ty 100% vốn nước ngoài,...) để quyết định cơ cấu của mẫu nghiên cứu.

Dặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

Có một số cách để lấy mẫu phân tầng. Chẳng hạn, giữ nguyên tỷ lệ của mỗi tầng trong tổng thể.

Một mẫu có kích thước 1,000 phần tử cần được lấy

Tầng	Thu nhập	Tỷ lệ tổng thể	Kích thước tầng
1	< \$15,000	25%	250
2	15,000-29,999	40%	400
3	30,000-50,000	30%	300
4	over \$50,000	5%	50

Tổng 1,000

Dặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

Mẫu ngẫu nhiên cơ học

c) *Mẫu được chọn theo phương pháp cơ học.* Trường hợp tập tổng quát có vô hạn phần tử làm theo cách a) và b) rất khó khăn, ta phải tiến hành bằng phương pháp sau : Dùng máy để phân chia ngẫu nhiên tập tổng quát thành các tập nhỏ, rồi từ những tập nhỏ đó ta chọn ra một số phần tử đại diện rồi hợp nhất lại ta được một mẫu chung. Ngoài ra ta có thể dùng bảng số ngẫu nhiên để chọn ; tức là ta có thể đánh số thứ tự tất cả các phần tử của tập tổng quát (thường các số này đã được ghi từ trước). Ta chọn ngẫu nhiên 1 số trong bảng số ngẫu nhiên, sau đó chọn những phần tử nào của tập tổng quát có số thứ tự trùng với những số ta vừa lấy ra để khảo sát. Phương pháp chọn mẫu này được gọi là phương pháp cơ học.

Dặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

Mẫu ngẫu nhiên “điển hình”

d) *Mẫu “điển hình”.* Mẫu “điển hình” là mẫu mà phần tử của nó không chọn từ toàn bộ tập tổng quát mà từ bộ phận “điển hình” của nó. Tất nhiên từ bộ phận “điển hình” ta chọn mẫu theo cách a) hoặc cách b).

Dặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

Cách thực hiện

- Trước tiên lập danh sách tổng thể chung theo từng khối (như làng, xã, phường, lượng sản phẩm sản xuất trong một khoảng thời gian...). Sau đó, ta chọn ngẫu nhiên một số khối và điều tra tất cả các đơn vị trong khối đã chọn.
- Thường dùng phương pháp này khi không có sẵn danh sách đầy đủ của các đơn vị trong tổng thể cần nghiên cứu.
- Ví dụ: Tổng thể chung là sinh viên của một trường đại học. Khi đó ta sẽ lập danh sách các lớp chứ không lập danh sách sinh viên, sau đó chọn ra các lớp để điều tra.

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

Mẫu ngẫu nhiên hệ thống

- Khi một mô hình hệ thống được giới thiệu vào lấy mẫu ngẫu nhiên, nó được gọi là "hệ thống (ngẫu nhiên) lấy mẫu".
- Một ví dụ là nếu các sinh viên trong trường đã gắn số liền với tên tuổi của họ, ví dụ 0001-1000. Chọn một điểm khởi đầu ngẫu nhiên, ví dụ như 0533, sau đó chọn tất cả các tên thứ 10 sau đó để lấy cho mẫu. Với kích thước mẫu là 100, bắt đầu lại với 0003 sau khi đạt 0993.

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

Cách thực hiện

- Trước tiên lập danh sách các đơn vị của tổng thể chung theo một trật tự quy ước nào đó, sau đó đánh số thứ tự các đơn vị trong danh sách.
- Đầu tiên chọn ngẫu nhiên 1 đơn vị trong danh sách; sau đó cứ cách đều k đơn vị lại chọn ra 1 đơn vị vào mẫu,... cứ như thế cho đến khi chọn đủ số đơn vị của mẫu.
- Ví dụ : Dựa vào danh sách bầu cử tại 1 thành phố, ta có danh sách theo thứ tự vần của tên chủ hộ, bao gồm 240.000 hộ. Ta muốn chọn ra một mẫu có 2.000 hộ. Vậy khoảng cách chọn là : $k = 240.000/2.000 = 120$, có nghĩa là cứ cách 120 hộ thì ta chọn một hộ vào mẫu.

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

Mẫu ngẫu nhiên nhiều cụm (Cluster sampling)

- Phương pháp này thường áp dụng đối với tổng thể chung có quy mô quá lớn và địa bàn (geographical) nghiên cứu quá rộng. Việc chọn mẫu phải trải qua nhiều giai đoạn (nhiều cấp).
- Trước tiên phân chia tổng thể chung thành các đơn vị cấp I, rồi chọn các đơn vị mẫu cấp I. Tiếp đến phân chia mỗi đơn vị mẫu cấp I thành các đơn vị cấp II, rồi chọn các đơn vị mẫu cấp II...
- Trong mỗi cấp có thể áp dụng các cách chọn mẫu ngẫu nhiên đơn giản, chọn mẫu hệ thống, chọn mẫu phân tầng,... để chọn ra các đơn vị mẫu.

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

Mẫu ngẫu nhiên nhiều cụm (Cluster sampling)

- Ví dụ: Muốn chọn ngẫu nhiên 50 hộ từ một thành phố có 10 khu phố, mỗi khu phố có 50 hộ. Cách tiến hành như sau : Trước tiên đánh số thứ tự các khu phố từ 1 đến 10, chọn ngẫu nhiên trong đó 5 khu phố. Đánh số thứ tự các hộ trong từng khu phố được chọn. Chọn ngẫu nhiên ra 10 hộ trong mỗi khu phố ta sẽ có đủ mẫu cần thiết.
- Lấy mẫu ngẫu nhiên nhiều cụm có thể làm tăng lỗi lấy mẫu vì có thể có sự tương tự giữa các phần tử trong cùng cụm.

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

Phương pháp chọn mẫu phi ngẫu nhiên

1. Mẫu phi ngẫu nhiên thuận tiện
2. Mẫu phi ngẫu nhiên phán đoán
3. Mẫu phi ngẫu nhiên định ngạch

Đặng Thế Gia, BM Kỹ thuật
xây dựng, ĐH Cần Thơ

Mẫu phi ngẫu nhiên thuận tiện

- Lấy mẫu dựa trên sự thuận lợi hay dựa trên tính dễ tiếp cận của đối tượng, ở những nơi mà nhân viên điều tra có nhiều khả năng gặp được đối tượng.
- Chẳng hạn, có thể gặp bất cứ người nào ở trung tâm thương mại, đường phố, cửa hàng,... để xin thực hiện cuộc phỏng vấn. Nếu người được phỏng vấn không đồng ý thì họ chuyển sang đối tượng khác.
- Lấy mẫu thuận tiện thường được dùng trong nghiên cứu khám phá, để xác định ý nghĩa thực tiễn của vấn đề nghiên cứu; hoặc để kiểm tra trước bảng câu hỏi nhằm hoàn chỉnh bảng; hoặc khi muốn ước lượng sơ bộ về vấn đề đang quan tâm mà không muốn mất nhiều thời gian và chi phí.

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

Mẫu phi ngẫu nhiên phán đoán

- Là phương pháp mà phỏng vấn viên là người tự đưa ra phán đoán về đối tượng cần chọn vào mẫu.
- Tính đại diện của mẫu phụ thuộc nhiều vào kinh nghiệm và sự hiểu biết của người tổ chức việc điều tra và cả người đi thu thập dữ liệu.
- Chẳng hạn, nhân viên phỏng vấn được yêu cầu đến các trung tâm thương mại chọn các phụ nữ ăn mặc sang trọng để phỏng vấn. Như vậy không có tiêu chuẩn cụ thể "thế nào là sang trọng" mà hoàn toàn dựa vào phán đoán để chọn ra người cần phỏng vấn.

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

Mẫu phi ngẫu nhiên định ngạch

- Đối với phương pháp chọn mẫu này, trước tiên ta tiến hành phân tổ tổng thể theo một tiêu thức nào đó mà ta đang quan tâm, cũng giống như chọn mẫu ngẫu nhiên phân tầng. Tuy nhiên sau đó ta lại dùng phương pháp chọn mẫu thuận tiện hay chọn mẫu phán đoán để chọn các đơn vị trong từng tổ để tiến hành điều tra.
- Sự phân bổ số đơn vị cần điều tra cho từng tổ được chia hoàn toàn theo kinh nghiệm chủ quan của người nghiên cứu.

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

Mẫu phi ngẫu nhiên định ngạch

- Chẳng hạn nhà nghiên cứu yêu cầu các phỏng vấn viên đi phỏng vấn 800 người có tuổi trên 18 tại 1 thành phố.
- Nếu áp dụng phương pháp chọn mẫu định ngạch, ta có thể phân tổ theo giới tính và tuổi như sau: chọn 400 người (200 nam và 200 nữ) có tuổi từ 18 đến 40, chọn 400 người (200 nam và 200 nữ) có tuổi từ 40 trở lên.
- Sau đó nhân viên điều tra có thể chọn những người gần nhà hay thuận lợi cho việc điều tra của họ để dễ nhanh chóng hoàn thành công việc.

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

Tính chất

- Chọn mẫu phi ngẫu nhiên (hay chọn mẫu phi xác suất) là phương pháp chọn mẫu mà các phần tử trong tổng thể không có khả năng ngang nhau để được chọn vào mẫu.
- Việc chọn mẫu phi ngẫu nhiên hoàn toàn phụ thuộc vào kinh nghiệm và sự hiểu biết về tổng thể của người nghiên cứu nên kết quả điều tra thường mang tính chủ quan của người nghiên cứu.
- Mặt khác, ta không thể tính được sai số do chọn mẫu, do đó không thể áp dụng phương pháp ước lượng thống kê để suy rộng kết quả trên mẫu cho tổng thể.

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

LỖI CHỌN MẪU

(ERRORS INVOLVED IN SAMPLING)

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

Các loại lỗi trong lấy mẫu

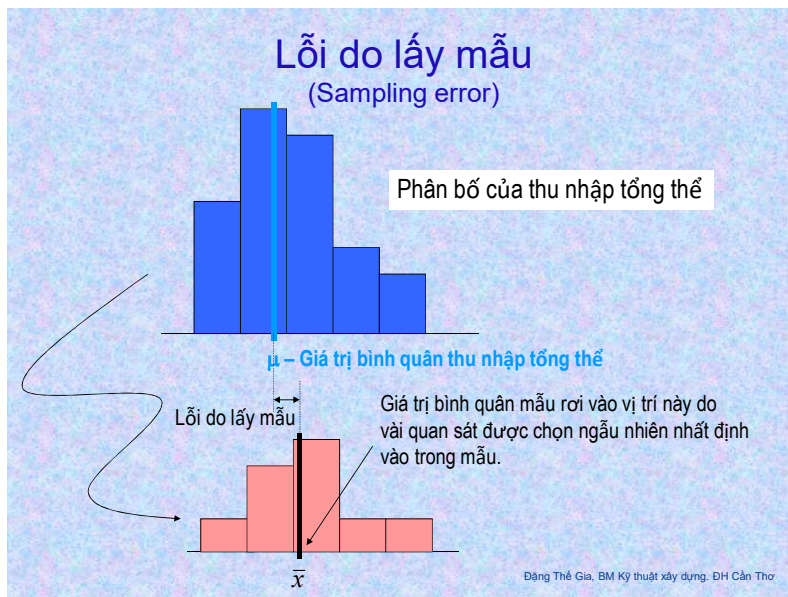
- Có 2 loại lỗi có thể xảy ra khi thực hiện lấy mẫu.
 - **Lỗi do lấy mẫu** (Sampling error): lỗi lấy mẫu xảy ra khi các đặc trưng thống kê của một quần thể được tính toán từ một tập con, hoặc mẫu, của tổng thể đó.
 - **Lỗi không do lấy mẫu** (Non-sampling error): Là một thuật ngữ chỉ tất cả các sai lệch so với giá trị thực mà không phải là một chức năng của mẫu được chọn, bao gồm các lỗi hệ thống khác nhau và bất kỳ lỗi ngẫu nhiên nào không phải do lấy mẫu.
- Các lỗi không do lấy mẫu khó định lượng hơn nhiều so với lỗi do lấy mẫu.

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

Lỗi do lấy mẫu (Sampling error)

- Lỗi do lấy mẫu liên quan đến sự khác biệt giữa mẫu và tổng thể, bởi vì những quan sát nhất định nào đó xảy ra để được chọn.
- Lỗi do lấy mẫu dự kiến sẽ xảy ra khi đưa ra một tuyên bố về tổng thể dựa trên mẫu lấy.

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ



Lỗi không do lấy mẫu (Non-sampling error)

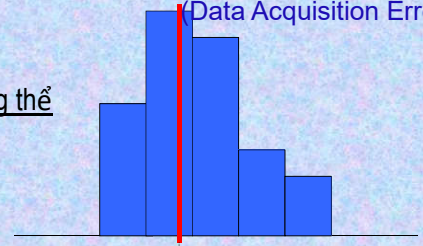
- Lỗi không do lấy mẫu xảy ra do những sai lầm được thực hiện trong suốt quá trình thu thập dữ liệu.
- Tăng kích cỡ mẫu sẽ không làm giảm loại lỗi này.
- Có ba loại lỗi không do lấy mẫu:
 - * Lỗi trong thu thập dữ liệu (Errors in data acquisition)
 - * Lỗi không đáp ứng (Non-response errors)
 - * Lựa chọn thiên vị (Selection bias): Thiên vị trong lấy mẫu là một nguồn có thể gây ra lỗi lấy mẫu, trong đó mẫu được chọn theo cách mà một số cá thể dường như không thuộc về mẫu so với những cá thể khác. Nó dẫn đến các lỗi lấy mẫu là làm tăng hoặc giảm tính trội nào đó.

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

Lỗi trong thu thập dữ liệu

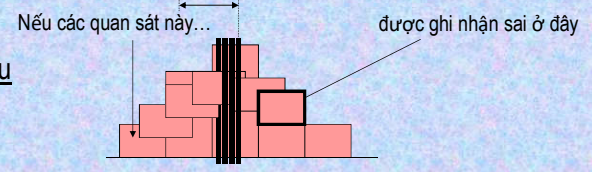
(Data Acquisition Error)

Tổng thể



Lỗi do lấy mẫu +
Lỗi do thu thập số liệu

Mẫu



Nếu các quan sát này...

được ghi nhận sai ở đây

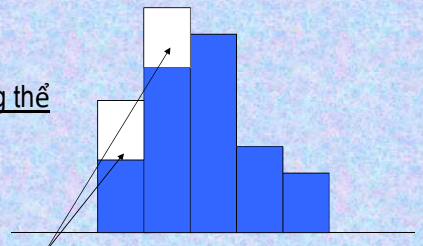
Khi đó bình quân mẫu bị ảnh hưởng

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

Lỗi do không đáp ứng

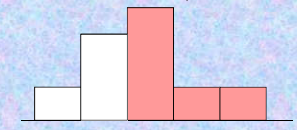
(Non-response errors)

Tổng thể



Không đáp ứng ở đây... có thể dẫn đến kết quả thiên vị ở đây

Mẫu

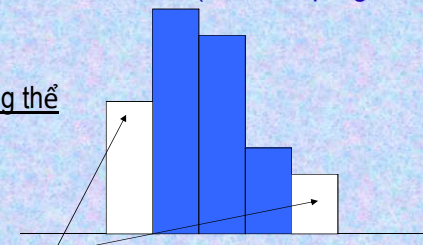


Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ

Lỗi không do lấy mẫu

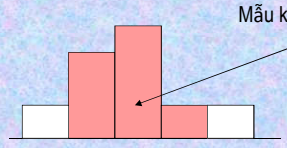
(Non-sampling error)

Tổng thể



Khi vài bộ phận của tổng thể không được chọn...

Mẫu



Mẫu không thể đại diện cho tổng thể

Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ



Đặng Thế Gia, BM Kỹ thuật xây dựng, ĐH Cần Thơ