VIỆN NGHIÊN CỨU Y XÃ HỘI HỌC

Chọn mẫu trong nghiên cứu

Nguyễn Trương Nam

Copyright – Bản quyền thuộc về tác giả và thongke.info. Khi sử dụng một phần hoặc toàn bộ bài giảng đề nghị mọi người trích dẫn: tên tác giả và thongke.info. Ví dụ: Nguyễn A – Thongke.info.

Chọn mẫu

- Chọn mẫu xác suất (probability sampling)
  - Mỗi đối tượng nghiên cứu được chọn vào trong mẫu nghiên cứu với một xác suất (khác 0)
  - Thường sử dụng danh sách mẫu hoặc khung mẫu (sampling fames) để chọn mẫu
  - Các quy trình chọn mẫu thường được xác định rõ ràng chặt chẽ
  - Các chỉ số (estimates) có thể gần xấp xỉ với chỉ số thật trong quần thể ('true' population values, ví dụ tỷ lệ hoặc giá trị trung bình)

Chọn mẫu

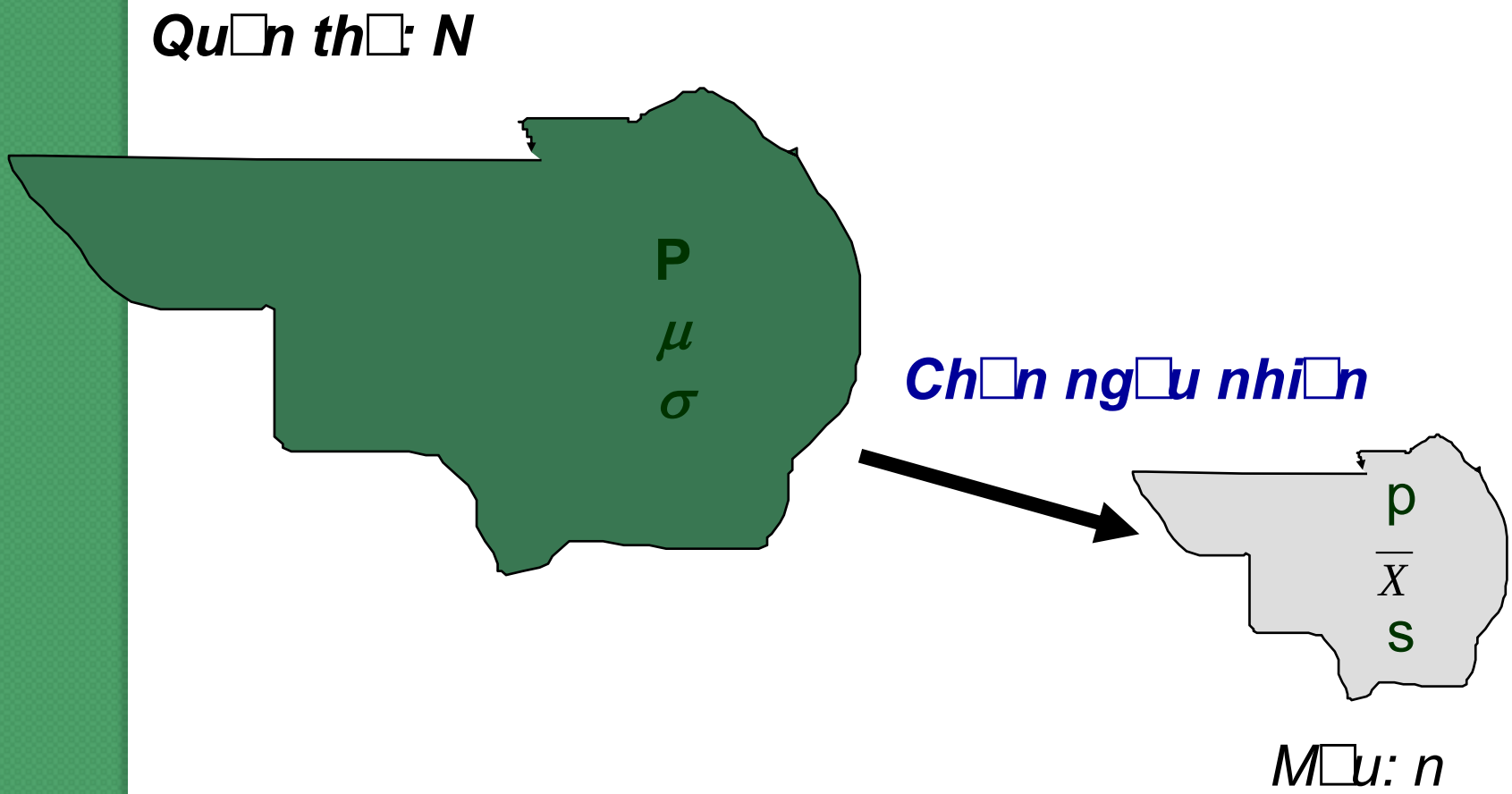
- Chọn mẫu không xác suất (non- probability sampling)
  - Chọn mẫu không dựa trên các nguyên lý thống kê học
  - Mẫu không đại diện cho quần thể
  - Quả bóng tuyết (Snowball - network sampling)
  - Chọn mẫu dây chuyền có kiểm soát (RDS)
  - Chọn mẫu chủ đích (purposive sampling, quota sampling?)
  - Chọn mẫu thuận tiện.

Các phương pháp chọn mẫu xác suất

- Chọn mẫu ngẫu nhiên đơn (simple random)
- Chọn mẫu ngẫu nhiên hệ thống (systematic sampling)
- Chọn mẫu phân tầng (stratification)
- Chọn mẫu chùm (cluster sampling) – *conventional cluster and time-location cluster*
- Chọn mẫu nhiều giai đoạn (multi stages)
- Chọn mẫu tỷ lệ với cỡ dân số (probability Propotional to Size – PPS)

Chọn mẫu ngẫu nhiên đơn (không thay thế)

- Mỗi thành viên của quần thể được đánh số (không trùng lặp)
- Bảng số ngẫu nhiên hoặc một phương pháp rút thăm được sử dụng để chọn từng đối tượng một cho tới khi đủ mẫu
- Ứng dụng với các nghiên cứu có cỡ mẫu nhỏ.
- Sử dụng bảng số ngẫu nhiên
- Rút thăm
- Máy tính


Qu□n th□: N
P
μ
σ
Ch□n ng□u nhi□n
p
X̄
s
M□u: n

Phương pháp

- Lập danh sách tất cả cá thể trong quần thể định chọn mẫu. Đánh số thứ tự các cá thể.
- Xác định cỡ quần thể N
- Xác định cỡ mẫu n
- Rút thăm từng đối tượng cho tới khi đủ mẫu
- Dùng bảng số ngẫu nhiên chọn từng đối tượng cho tới khi đủ cỡ mẫu
- Máy tính

Sử dụng bảng số ngẫu nhiên

- Đánh số các cá thể trong quần thể mẫu
- Xác định cỡ quần thể N
- Xác định cỡ mẫu n
- Xác định điểm bắt đầu trên bảng số ngẫu nhiên bằng cách nhắm mắt và ngẫu nhiên chỉ ngón tay vào một điểm trên bảng số
- Lựa chọn hướng: ví dụ từ trên xuống và từ phải sang trái
- Lấy số được chọn thứ nhất là số có số chữ số X cuối giữa 0 và N (nếu N là số 2 chữ số, X = 2, nếu N là số hàng trăm X = 3)
- Không chọn lại số đã được chọn
- Cứ như vậy theo hướng đã định đọc đủ số mẫu n
- Nếu hết bảng vẫn chưa đủ số n, chọn điểm khởi đầu khác, chọn hướng khác.

Ví dụ/thực hành

- Bằng rút thăm
- Bằng bảng số ngẫu nhiên
- Random number table.pdf
- Bằng máy tính (Stata, OpenEpi, EpiCalc 2000)

- Thực hành: Chọn ngẫu nhiên 40 học sinh từ một trường tiểu học có tổng số 95 học sinh.

ISMS
VIỆN NGHIÊN CỨU Y - XÃ HỘI HỌC

OpenEpi, EpiCalc2000

OpenEpi

Random Number Generator		
Range	Lowest value	1
	Highest value	99
How many would you like?		35
Format into how many columns?		5
Omit text from output?		No

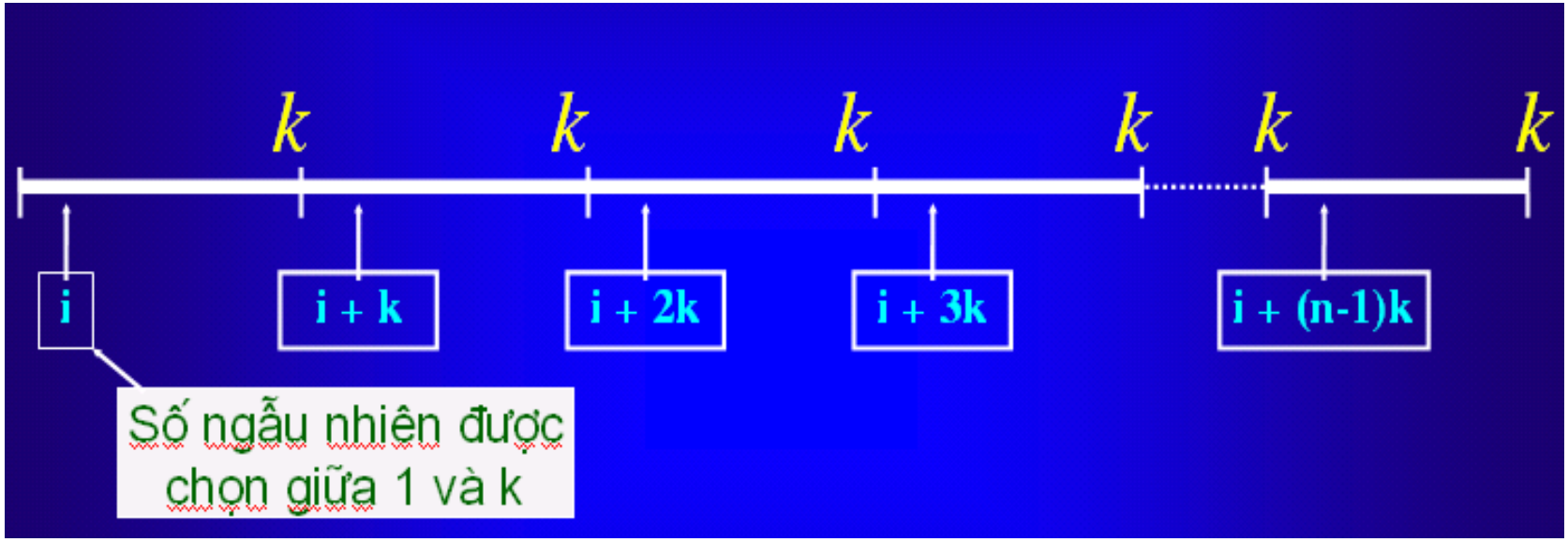
EpiCalc2000

ISMS
VIỆN NGHIÊN CỨU Y - XÃ HỘI HỌC

Chọn mẫu ngẫu nhiên hệ thống

- Xác định và đánh số thứ tự đơn vị mẫu (khung mẫu).
- Xác định cỡ quần thể N
- Xác định cỡ mẫu n
- Tính khoảng cách mẫu k (k=N/n)
- Chọn đơn vị mẫu đầu tiên (i) nằm giữa 1 và k bằng phương pháp ngẫu nhiên (sử dụng bảng số ngẫu nhiên hoặc rút thăm).
- Chọn các đơn vị mẫu tiếp theo bằng cách cộng k với đơn vị mẫu đầu tiên, tiếp tục cho đến khi đủ số mẫu: i + 1k; i+2k; i+3k…i+(n-1)k.

Chọn mẫu ngẫu nhiên hệ thống



Ví dụ/thực hành: Chọn 400 HỘ GIA ĐÌNH (HGĐ) từ 40 TỔ DÂN PHỐ (TDP).

- 10 HGĐ sẽ được chọn từ mỗi TDP theo phương pháp chọn mẫu ngẫu nhiên.
- Liệt kê danh sách danh sách tất cả các HGĐ có trong TDP.
- Khoảng cách mẫu k=tổng số hộ gia đình/10.
- X là số ngẫu nhiên giữa 1 và k, được chọn thông qua bảng số ngẫu nhiên.
- Hộ thứ nhất là số ngẫu nhiên, các hộ tiếp theo là (x+k); (x+2k)…; (x+9k).

Chọn mẫu ngẫu nhiên hệ thống

□u □i□m:

- Mẫu được phân bổ dàn đều trong khung mẫu (nêu các đơn vị mẫu được đánh số một cách chính xác)
- Nhanh và dễ áp dụng.
- Đơn giản trong điều kiện thực địa.

H□n ch□:

- Đơn vị mẫu không xếp ngẫu nhiên, thiếu đại diện.

Chọn mẫu phân tầng

- Được lựa chọn khi nghiên cứu muốn đảm bảo tính đại diện của mẫu cho từng nhóm quần thể, ví dụ các nhóm tuổi, giới
- Quần thể mẫu được chia thành các tầng
- Các tầng có chung đặc điểm (ví dụ, vùng miền, giới tính, nhóm tuổi…).
- Mẫu được chọn riêng biệt cho từng tầng (phương pháp ngẫu nhiên đơn giản hay ngẫu nhiên hệ thống).
- Chỉ số cho từng tầng được tính toán riêng biệt cùng với chỉ số toàn mẫu
- Vì trong các tầng, các cá thể giống nhau với một số đặc điểm, Chỉ số tính toán trong từng tầng ước tính chính xác hơn so với chỉ số được tính nếu như toàn mẫu được chọn ngẫu nhiên đơn giản.

Chọn mẫu phân tầng

- Cỡ mẫu tại từng tầng được chọn dựa trên tỷ lệ của cỡ dân số tại tầng đó với quần thể (proportionate samples).
  - Ví dụ mẫu 2 tầng: thành thị (60% quần thể) và nông thôn (40%). Nếu cỡ mẫu 5000 thì tầng thành thị chọn 3000 và tầng nông thôn chọn 2000.
- Cỡ mẫu được chọn tương đương tại các tầng (equal-size samples/disproportionate samples)

Ví dụ/thực hành: Chọn mẫu trong điều tra ban đầu A&T.

- Giai đoạn 1:
  - 40 xã tại các huyện thuộc 4 tỉnh đã được lựa chọn. Số lượng các huyện, xã được lựa chọn ở mỗi tỉnh dựa trên cỡ dân số của tỉnh, huyện đó.
  - Các xã được lựa chọn từ các huyện dựa trên các tiêu chí: tương tự về dân số, về tình trạng kinh tế-xã hội.
- Giai đoạn 2: chọn mẫu phân tầng: Chọn 4000 bà mẹ có con <5 tuổi.
  - Tại từng tỉnh danh sách tất cả các trẻ tại các xã đã được lựa chọn sẽ được lập theo 2 nhóm can thiệp và nhóm chứng và theo 3 nhóm tuổi: 0-5.9T; 6-23.9T; và 24-59.9T.
  - Số lượng trẻ từ tầng tuổi < 6 tháng và 6-24 tháng = ½ số trẻ từ tầng 24-60 tháng
  - Tại mỗi tầng Sử dụng phương pháp chọn mẫu ngẫu nhiên hệ thống để chọn các bà mẹ của mỗi nhóm tuổi

ISMS
VIỆN NGHIÊN CỨU Y - XÃ HỘI HỌC

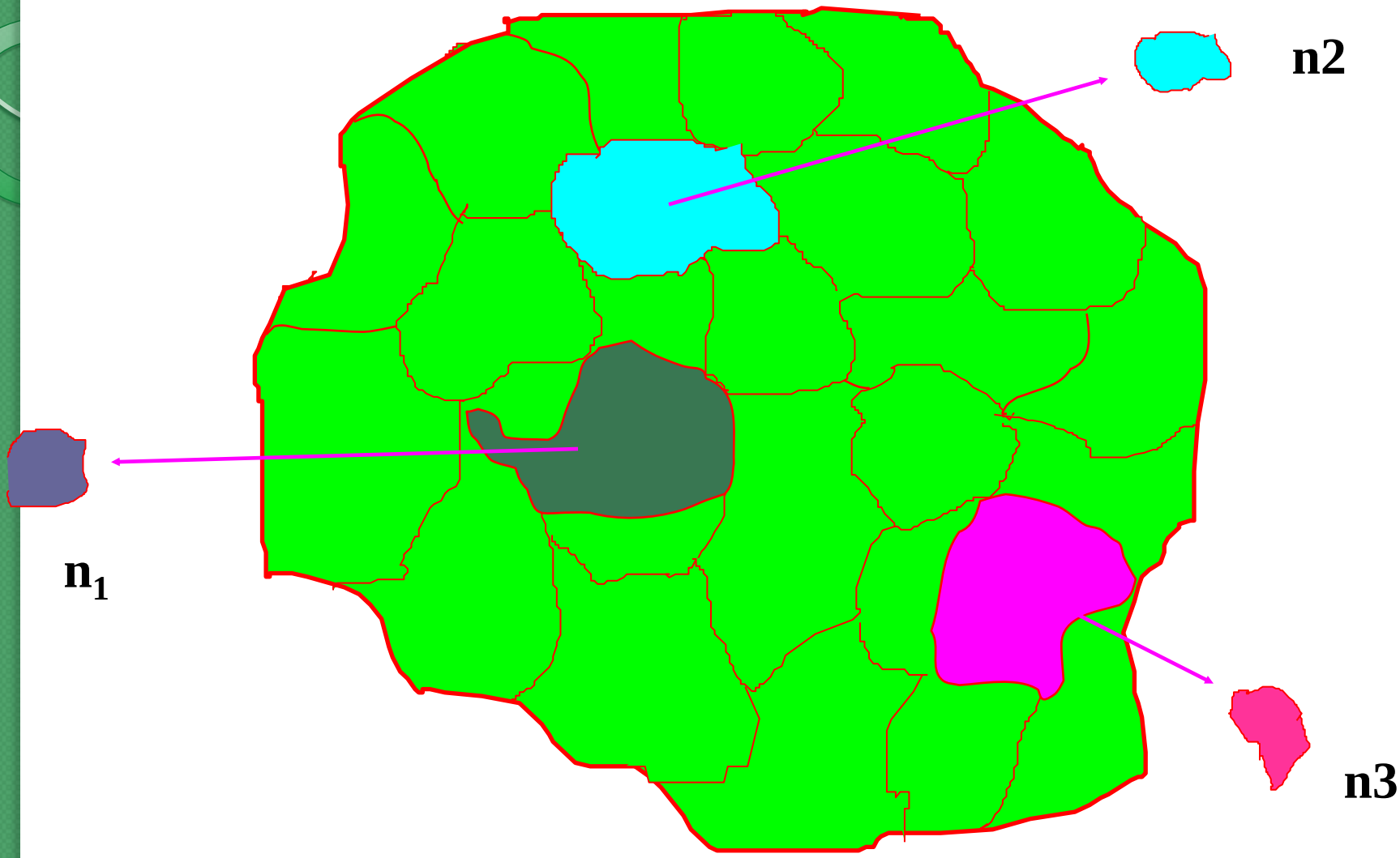
Ví dụ/thực hành: Chọn mẫu trong điều tra ban đầu A&T.

Province	<6 months		6-23.9 months		24-59.9 months	
	Intervention (franchise)	Control	Intervention (franchise)	Control	Intervention (franchise)	Control
Thanh Hoa	175	175	175	175	350	350
Thai Nguyen	100	100	100	100	200	200
Vinh Long	100	100	100	100	200	200
Quang Ngai	125	125	125	125	250	250

ISMS
VIỆN NGHIÊN CỨU Y - XÃ HỘI HỌC

Chọn mẫu cụm/chùm

- Xác định cụm/chùm (theo địa lý: tỉnh, huyện, xã; theo tổ chức: phòng khám, bệnh viện)
- Lập danh sách cụm/chùm
- Chọn chùm theo phương pháp ngẫu nhiên đơn hoặc ngẫu nhiên hệ thống
- Chọn các cá thể tại mỗi cụm/chùm bằng cách:
  - Lựa chọn tất cả các đơn vị mẫu trong các cụm/chùm vào nghiên cứu (chùm 1 bậc).
  - Lập danh sách tại mỗi cụm/chùm, chọn cá thể bằng phương pháp ngẫu nhiên đơn hoặc hệ thống (chùm 2 bậc).


n2
n1
n3

Ví dụ

- Để chọn một mẫu ngẫu nhiên hộ gia đình tại một tỉnh, cách đơn giản nhất là chọn từ danh sách toàn bộ số hộ gia đình trong tỉnh đó.
- Phương pháp này có thể không thực hiện được vì danh sách đó không có sẵn và việc có một danh sách hoàn chỉnh tất cả các hộ gia đình là khó thực hiện.
- Danh sách các phường xã được lập
- 30 phường xã được chọn.
- Tại mỗi phường xã được chọn, danh sách hộ gia đình được lập.
- 10% số hộ gia đình được chọn từ danh sách mỗi xã phường.

Chọn mẫu cụm/chùm

Ưu □i□m:

- Có thể áp dụng trong điều tra có phạm vi rộng, phân tán, không có được danh sách các đơn vị nghiên cứu.
- Khung mẫu đơn giản (danh sách các chùm) dễ lập.
- Điều tra dễ và nhanh vì đối tượng nghiên cứu được nhóm lại.
- Nâng cao chất lượng của giám sát và đảm bảo chất lượng số liệu.
- Tiết kiệm kinh phí, thời gian.

Chọn mẫu cụm/chùm

H□n ch□:

- Tính chính xác và tính đại diện thấp (hệ số thiết kế).
- Cần số cụm/chùm lớn. Thường số chùm >30.

Chọn mẫu chùm – thời gian-địa điểm

- Thông thường đối với nhóm dân cư cố định chùm được định nghĩa là không gian địa lý.
- Tuy nhiên với nhóm dân cư di biến động, chùm được định nghĩa thời gian-địa điểm.
- Ví dụ: Với nhóm lái xe tải đường dài – tại một điểm dừng chân với các khung giờ khác nhau số lượng lái xe dừng chân khác nhau. Chùm sẽ được định nghĩa = khoảng thời gian tại địa điểm đó: 8-10h sáng/điểm A, 10-12h/điểm A, 8-10h sáng/điểm B, 10-12h/điểm B…
- Khung mẫu được lập bằng phương pháp mapping – bản đồ. Chùm thời gian-địa điểm được chọn. Tiếp theo sẽ chọn các lái xe tại các chùm được chọn.

ISMS
VIỆN NGHIÊN CỨU Y - XÃ HỘI HỌC

Chọn mẫu nhiều giai đoạn

- Phương pháp được sử dụng nhiều trong các nghiên cứu lớn, phạm vi địa lý rộng.
- 2 giai đoạn
  - Giai đoạn 1: quần thể được chia thành các cụm/chùm, mẫu cụm/chùm được chọn.
  - Giai đoạn 2: các cá thể được chọn từ các cụm/chùm (được chọn từ giai đoạn 1).
- Nhiều giai đoạn
  - Giai đoạn 1: quần thể được chia thành các cụm/chùm, mẫu cụm/chùm được chọn.
  - Giai đoạn 2: Các cụm/chùm chọn từ giai đoạn 1 được chia thành các cụm/chùm nhỏ hơn, mẫu cụm/chùm bậc 2 được chọn.
  - Giai đoạn 3: : các cá thể được chọn từ các cụm/chùm (được chọn từ giai đoạn 2)
- Tại các giai đoạn các phương pháp ngẫu nhiên đơn giản, hệ thống, hay phân tầng được sử dụng.

Ví dụ chọn mẫu nhiều giai đoạn

- Giai đoạn 1: Chọn tỉnh
- Giai đoạn 2: Chọn huyện
- Giai đoạn 3: Chọn xã
- Giai đoạn 4: Lập danh sách mẫu tại các xã để chọn hộ gia đình.

Chọn mẫu tỷ lệ với cỡ dân số PPS

- Là một phương pháp chọn mẫu nhiều giai đoạn.
- Xác suất được chọn vào mẫu của các cụm/chùm được tỷ lệ với cỡ dân số của cụm/chùm đó.
- PPS rất hữu ích khi cỡ dân số của các cụm/chùm khác biệt nhau nhiều.
- Kết hợp với việc chọn số mẫu tương đương nhau tại các cụm/chùm, PPS đảm bảo các cá thể mẫu được chọn vào mẫu với cùng xác xuất.
- PPS được sử dụng rất nhiều trong các điều tra nghiên cứu hành vi, trong điều tra hộ gia đình.

Phương pháp

1. Chuẩn bị danh sách đơn vị mẫu đầu tiên với dân số tương ứng cho mỗi đơn vị mẫu.
2. Bắt đầu từ phần đầu danh sách, tính dân số lũy tích và ghi lại vào cột bên cạnh cột về dân số tương ứng cho mỗi đơn vị mẫu.
3. Tính khoảng cách mẫu (SI) bằng cách chia tổng số dân số lũy tích (M) cho tổng số đơn vị mẫu cần chọn (a). Do đó SI=M/a.
4. Chọn một số ngẫu nhiên (RS) giữa 1 và khoảng cách mẫu (SI). So sánh số ngẫu nhiên này và với số dân lũy tích. Số nào gần nhất với RS sẽ được chọn làm đơn vị mẫu đầu tiên.
5. Các đơn vị mẫu tiếp theo sẽ được chọn theo công thức: RS + SI, RS + 2SI, RS + 3SI,…RS + (a-1) SI.

Chú ý: Trong việc lựa chọn các đơn vị mẫu, điều rất quan trọg là các số thập phân trong khoảng cách mẫu được giữ lại. Nguyên tắc là khi phần thập phân của các số lấy mẫu là nhỏ hơn .5, cụm có số thấp hơn sẽ được chọn và khi phần thập phân của số lấy mẫu bằng .5 hoặc lớn hơn cụm có số lớn hơn sẽ được chọn. **(FHI-BSS Guide)**

Ví dụ: Chọn 40 cụm (TỔ DÂN PHỐ) theo phương pháp PPS.

- Yêu cầu: chọn 40 cụm tổ dân phố tại 5 xã triển khai dự án – hai huyện Trảng Bàng, Gò Dầu – Tây Ninh
- Liệt kê danh sách các đơn vị mẫu là TỔ DÂN PHỐ (các cụm) tại 5 xã can thiệp. Bản danh sách gồm tên các tổ, tổng số dân trong mỗi tổ, dân số lũy tích-số lượng này thu được bằng cách cộng dồn số dân của mỗi tổ với dân số của tất cả các nhóm trước đó trong danh sách.
- Khoảng cách mẫu (k) được tính bằng cách chia tổng số dân số lũy tích cho 40.
- Một số ngẫu nhiên (x) giữa nằm trong khoảng giữa 1 và khoảng cách mẫu (k) sẽ được chọn. So sánh số này với số dân số luỹ tích, tổ dân phố nào có số dân số luỹ tích gần sát nhất với số ngẫu nhiên sẽ là đơn vị mẫu thứ nhất.
- TỔ DÂN PHỐ được chọn tiếp theo được chọn bởi cộng khoảng cách mẫu với số ngẫu nhiên (x+k), các tổ tiếp theo theo công thức sau: (x+2k), (x+3k), …x+19k.

Điều tra ban đầu dự án Phòng chống HIV sau xây dựng đường cao tốc HCMC – Phnompenh: thành tố điều tra hộ gia đình (thực hiện bởi ISMS - 2010)

ISMS
VIỆN NGHIÊN CỨU Y - XÃ HỘI HỌC

Phương pháp lập bản đồ trong chọn mẫu

- Phương pháp lập bản đồ phục vụ việc tạo khung mẫu.
- Ứng dụng trong các nghiên cứu với nhóm di biến động.
- Ứng dụng trong các nghiên cứu định tính.
- *Chi tiết các phương pháp trong bài trình bày riêng.*

Áp dụng các phương pháp chọn mẫu trong thực tiễn nghiên cứu

- Khảo sát, điều tra (large scale): chọn mẫu nhiều giai đoạn, kết hợp các phương pháp, ví dụ PPS sau đó ngẫu nhiên hệ thống.
- Khảo sát nhỏ: ngẫu nhiên đơn giản, ngẫu nhiên hệ thống.
- Thử nghiệm lâm sàng: ngẫu nhiên đơn giản, ngẫu nhiên hệ thống.