



PHƯƠNG PHÁP NGHIÊN CỨU

Lý thuyết & thực tiễn

Tháng 8/2016

Powering
Informed
Decisions

Nội dung chương trình

Bài 1

- ☐ Lập kế hoạch nghiên cứu
- ☐ Giới thiệu lý thuyết về mẫu nghiên cứu & chọn mẫu nghiên cứu trong thực tiễn (phần 1)

Bài 2

- ☐ Giới thiệu lý thuyết về mẫu nghiên cứu & chọn mẫu nghiên cứu trong thực tiễn (phần 2)
- ☐ Thiết kế bảng câu hỏi khảo sát & thu thập thông tin

Bài 3

- ☐ Tiến hành khảo sát thu thập dữ liệu từ thực địa

Bài 4

- ☐ Phân tích kết quả bằng phần mềm SPSS



Bài 1

- Lập kế hoạch nghiên cứu
- Lý thuyết về mẫu nghiên cứu
- Chọn mẫu nghiên cứu trong thực tiễn

Powering
Informed
Decisions

1.1 LẬP KẾ HOẠCH NGHIÊN CỨU

- ☐ Nghiên cứu là gì
- ☐ Các loại hình nghiên cứu
- ☐ Tại sao phải nghiên cứu
- ☐ Một số loại hình nghiên cứu
- ☐ Những câu hỏi căn bản trong nghiên cứu
- ☐ Các bước thực hiện một nghiên cứu
- ☐ Đề xuất nghiên cứu

Nghiên cứu là gì?

What is research?

- ❑ Hoạt động tìm kiếm kiến thức, thường là bằng 1 phương pháp khoa học, nhằm mục đích:
 - khám giá hoặc thiết lập những cái mới mẻ
 - giải quyết những vấn đề mới hoặc vấn đề hiện tại
 - chứng minh những ý tưởng mới
 - phát triển những lý thuyết mới hoặc kiểm chứng những lý thuyết/kiến thức đã có
- ❑ Quá trình thu thập và phân tích thông tin (dữ liệu) một cách có hệ thống nhằm nâng cao sự hiểu biết về hiện tượng mà chúng ta quan tâm.

Nghiên cứu thị trường là gì?

What is market research?

- ❑ Quá trình thu thập và phân tích thông tin (*dữ liệu*) về những khách hàng mà bạn muốn tiếp cận (*hay còn gọi là thị trường mục tiêu*). Nghiên cứu thị trường có thể giúp bạn hình thành 1 kế hoạch kinh doanh, tung ra 1 sản phẩm hay dịch vụ mới, tinh chỉnh các sản phẩm/dịch vụ hiện hữu, mở rộng thâm nhập vào các thị trường mới, phát triển 1 chiến dịch quảng cáo, ấn định giá cả, và/hoặc chọn lựa 1 địa điểm kinh doanh.
- ❑ Việc áp dụng một phương pháp khoa học nhằm tìm kiếm sự thật về các hiện tượng của thị trường. Các hoạt động bao gồm:
 - xác định cơ hội và các vấn đề của thị trường
 - hình thành và đánh giá những ý tưởng về thị trường
 - theo dõi hoạt động và đánh giá hiệu quả cũng như tìm hiểu quá trình tiếp thị

Nghiên cứu khán giả là gì?

What is audience research?

- ❑ Là phương pháp tìm hiểu một cách có hệ thống và chính xác về khán giả (*thính giả, độc giả, người dùng Internet v.v.*)
- ❑ Nghiên cứu khán giả có thể thực hiện 2 nhiệm vụ chính:
 - 1) đo lường (*ước tính*) qui mô khán giả, và
 - 2) khám phá những thói quen và sở thích của khán giả.
- ❑ Ngành phát thanh & truyền hình có nhu cầu đặc biệt về nghiên cứu khán giả bởi đây là 2 lĩnh vực duy nhất không thể đếm khán giả của mình một cách chính xác.
- ❑ Ngày nay, nghiên cứu khán giả không chỉ giới hạn ở lĩnh vực phát thanh và truyền hình, mà đã được áp dụng ở các ngành truyền thông khác như báo/tạp chí, và các phương tiện mới nổi như Internet và truyền thông di động.

Các loại hình nghiên cứu

Types of research

- ❑ Nhiều hình thức phân loại khác nhau
- ❑ Một số phân loại thường thấy
 - Nghiên cứu ứng dụng & nghiên cứu cơ bản
Applied/practical/empirical versus basic/fundamental research
 - Nghiên cứu định lượng & nghiên cứu định tính
Quantitative versus qualitative research
 - Nghiên cứu mô tả & nghiên cứu tương quan
Descriptive versus correlational
 - Nghiên cứu syndicated & nghiên cứu theo yêu cầu
Syndicated versus customized research

Nghiên cứu ứng dụng

Applied/practical/empirical research

- ❑ Còn gọi là *ngiên cứu hành động* (*action research*)
- ❑ Cung cấp những thông tin thực tế giúp cho việc hình thành quyết định trên cơ sở mô tả một hiện tượng nào đó được quan tâm hoặc làm sáng tỏ các kết quả của một quá trình hành động cụ thể.
- ❑ Chủ yếu quan tâm đến một vấn đề hay một nhu cầu cụ thể trước mắt và ít khi nào đặt tham vọng đưa ra những lý giải lâu bền cho các hiện tượng xảy ra quanh chúng ta. Nhưng đôi khi cũng cung cấp những hiểu biết thấu đáo hữu ích cho các nỗ lực xây dựng lý thuyết.
- ❑ Chiếm ưu thế trong các loại hình nghiên cứu như nghiên cứu thị trường, nghiên cứu khán giả trong ngành truyền thông.
- ❑ Nghiên cứu phương pháp luận (*Methodological research*) – Liên quan đến các cơ sở lý luận làm nền tảng cho việc xây dựng, phát triển và đề xuất các phương pháp nghiên cứu.

Nghiên cứu ứng dụng

Applied/practical/empirical research

❑ Một số ví dụ

- Nghiên cứu nâng cao năng suất sản xuất nông nghiệp
- Nghiên cứu chữa hoặc điều trị một loại bệnh cụ thể
- Nghiên cứu nâng cao hiệu quả sử dụng năng lượng ở gia đình, văn phòng, hoặc các phương tiện vận tải
- Nghiên cứu phương pháp đo lường mức sống của hộ gia đình
- Nghiên cứu phương pháp đo lường khán giả truyền hình



Nghiên cứu cơ bản

Basic/fundamental/theoretical research

- ❑ Mở mang sự hiểu biết của con người chứ không phải là sáng tạo hay phát minh.
- ❑ Kiểm định những lý giải được khái quát hóa thành lý thuyết về cơ chế vận hành của các hiện tượng xung quanh chúng ta.
- ❑ Thường thấy phổ biến hơn trong lĩnh vực hàn lâm (*academic*).
- ❑ Kết quả nghiên cứu không nhất thiết phải có giá trị thương mại.
- ❑ Ranh giới giữa nghiên cứu ứng dụng & nghiên cứu cơ bản nhiều khi chỉ là tương đối và đôi khi rất khó có thể phân định rõ ràng.
- ❑ Một nghiên cứu cũng có thể phục vụ cho cả 2 mục đích nghiên cứu ứng dụng & lý thuyết.
- ❑ Một số ví dụ:
 - *Nghiên cứu sự hình thành của vũ trụ*
 - *Nghiên cứu cấu tạo của protons, neutrons, electrons*
 - *Nghiên cứu mã gen di truyền của loài ruồi dấm*

Nghiên cứu định lượng

Quantitative research

- ❑ Chủ yếu dựa vào điều tra khảo sát, thí nghiệm và phân tích nội dung để thu thập những thông tin có thể lượng hóa hay có thể biểu diễn thông qua những con số định lượng.
- ❑ Đánh giá/đo lường thái độ, quan điểm, hành vi v.v. của đối tượng nghiên cứu và khái quát hóa (suy rộng) kết quả nghiên cứu.
- ❑ Sử dụng các mô hình, lý thuyết và các giả thuyết toán/thống kê có liên quan đến hiện tượng nghiên cứu để phân tích kết quả khảo sát, thí nghiệm, quan sát v.v..
- ❑ Thường sử dụng phương pháp thu thập dữ liệu như:
 - Phỏng vấn trực diện (*Face-to-face interviews*)
 - Phỏng vấn qua điện thoại (*Telephone interviews*)
 - Nghiên cứu theo thời gian (*Longitudinal studies*)
 - Phỏng vấn ngẫu nhiên qua website hoặc thu thập ý kiến trực tuyến (*Website interceptors/online polls*)
 - Quan sát một cách có hệ thống (*Systematic observations*)

Nghiên cứu định tính

Qualitative research

- ❑ Đưa ra những tóm tắt, kết luận không phải bằng những con số, ví dụ những ghi chép mô tả tại hiện trường hay những nhận xét rút ra từ các cuộc phỏng vấn.
- ❑ Cho phép người nghiên cứu tìm hiểu sâu hơn về một chủ đề nào đó, và kết quả nghiên cứu thường không thể suy rộng (*khái quát hóa*).
- ❑ Nhằm trả lời cho câu hỏi “**Tại sao**” chứ không phải “**Như thế nào**”.
- ❑ Dựa trên các phân tích thông tin phi cấu trúc (*unstructured information*), như các bản ghi chép và diễn giải kết quả phỏng vấn, các câu trả lời cho các câu hỏi dạng mở (*open ended*), thư từ, nhật ký, phiếu góp ý, phim ảnh v.v.
- ❑ Không chỉ dựa vào thống kê hay những con số
- ❑ Sử dụng để tìm hiểu thấu đáo về thái độ, hành vi, hệ thống giá trị, sự quan tâm, động cơ, cảm hứng, văn hóa hoặc lối sống của con người.
- ❑ Thường sử dụng các công cụ nghiên cứu như phỏng vấn, khảo sát & quan sát.

Nghiên cứu định lượng & định tính

Tóm tắt những sự khác biệt

Nghiên cứu định tính	Nghiên cứu định lượng
Mục đích là mô tả đầy đủ và chi tiết	Mục đích là phân loại các đặc điểm, đếm số lượng, và xây dựng các mô hình thống kê nhằm giải thích hiện tượng được quan sát/nghiên cứu.
Người nghiên cứu có thể chỉ biết trước rất ít về những gì mình tìm sẽ hiểu, nghiên cứu.	Người nghiên cứu đã biết khá rõ ràng từ trước về những gì mình sẽ tìm hiểu, nghiên cứu.
Được khuyến nghị sử dụng trong những giai đoạn đầu của các dự án nghiên cứu.	Được khuyến nghị sử dụng trong những giai đoạn sau của các dự án nghiên cứu.
Thiết kế theo kiểu nghiên cứu mở.	Mọi khía cạnh nghiên cứu đều phải được thiết kế cẩn thận trước khi tiến hành thu thập dữ liệu.
Người nghiên cứu chính là công cụ thu thập dữ liệu.	Người nghiên cứu sử dụng các công cụ như bảng câu hỏi hay thiết bị để thu thập dữ liệu.
Dữ liệu dưới dạng từ ngữ, hình ảnh, hiện vật.	Dữ liệu dưới dạng số học và số thống kê.
Nghiên cứu có tính chủ quan/tính cá nhân cao hơn	Nghiên cứu có tính khách quan cao hơn
Việc diễn giải các sự kiện là quan trọng, ví dụ sử dụng cách quan sát đối tượng tham gia nghiên cứu, phỏng vấn chuyên sâu, v.v.	Nỗ lực đo lường và phân tích một cách chính xác những khái niệm nghiên cứu, ví dụ sử dụng các cuộc khảo sát, bảng câu hỏi, v.v.
Dữ liệu định tính “phong phú” hơn, tốn nhiều thời gian, và ít có khả năng suy rộng (<i>khái quát hóa</i>) kết quả nghiên cứu.	Dữ liệu mang tính hiệu quả cao hơn, có khả năng dùng để kiểm định các giả thuyết, nhưng có thể bỏ qua các chi tiết mang tính hoàn cảnh.
Người nghiên cứu có xu hướng tự hòa nhập một cách chủ quan vào trong vấn đề nghiên cứu.	Người nghiên cứu có xu hướng duy trì tính khách quan đối với vấn đề nghiên cứu.

Nghiên cứu syndicated (1)

Syndicated research

- ❑ Là sản phẩm được chuẩn hóa để cung cấp đồng thời cho nhiều người đăng ký sử dụng (subscribers). Ví dụ các báo cáo về khán giả truyền hình cung cấp cho tất cả mọi người sử dụng ở một thị trường nào đó.
- ❑ Một số loại hình sản phẩm syndicated trong ngành truyền thông:
 - Đo lường khán giả truyền hình, thính giả nghe đài, độc giả báo chí, Internet v.v.
 - Theo dõi, ghi nhận và thống kê nội dung truyền thông như chương trình phát thanh, truyền hình, các mẫu quảng cáo (*creative*), vị trí sắp đặt quảng cáo (*spot*), nội dung quảng và chi phí quảng cáo.
 - Nghiên cứu mức độ & thói quen tiêu dùng các sản phẩm FMCG.
 - Kiểm chứng (*audit*) phương pháp thực hiện các báo cáo về khảo sát, đo lường khán giả, báo cáo số lượng phát hành ấn phẩm báo chí (*circulation*) do các chủ thể phát hành công bố.

Nghiên cứu syndicated (2)

Syndicated research

- ❑ Một số ưu điểm so với các loại hình nghiên cứu khác:
 - Chi phí nghiên cứu được chia sẻ bởi nhiều người cùng đăng ký sử dụng (*subscribers*), do đó chi phí tính trên mỗi người sử dụng khá thấp.
 - Cùng một kết quả nghiên cứu giống nhau được cung cấp cho nhiều người sử dụng khác nhau, do đó nhà cung cấp sản phẩm syndicated phải khách quan trong nghiên cứu. Các báo cáo kết quả nghiên cứu bản thân chúng đã mang tính chất bán công khai (*semi-public*), và điều đó khiến cho bất kỳ ai cũng khó có thể làm sai lệch kết quả báo cáo bởi nhiều người đăng ký sử dụng đều có thể kiểm chứng thông qua dữ liệu syndicated.
 - Tuy chưa phải là hoàn hảo, nhưng các số liệu syndicated, chẳng hạn như báo cáo về số lượng khán giả, thường trở thành các con số chính thức được dùng làm cơ sở để giao dịch kinh doanh.

Nghiên cứu theo yêu cầu (1)

Customized research

- ❑ Là nghiên cứu được thiết kế nhằm đáp ứng nhu cầu của bên tài trợ, vì thế kết quả nghiên cứu ít khi được công bố ra bên ngoài.
- ❑ Các nghiên cứu có thể được ủy thác bởi các chuyên gia như các nhà tư vấn sắp đặt chương trình và tin tức, hoặc có thể được thực hiện bởi một bộ phận nghiên cứu trong nội bộ.
- ❑ Kết quả nghiên cứu dạng này có thể có giá trị đối với bên tài trợ (và sử dụng), nhưng ngoài ra thì giá trị của chúng là khá hạn chế.
- ❑ Nhà tài trợ thường không muốn chia sẻ kết quả nghiên cứu với bất kỳ ai bên ngoài tổ chức của họ, và nếu có thì những thông tin được cung cấp thường bị coi là rất đáng ngờ.
- ❑ Các phương pháp thực hiện khó kiểm chứng được, và người ta thường cho rằng nhà tài trợ ít nhiều đều có động cơ vụ lợi trong việc quảng bá kết quả nghiên cứu (ví dụ họ chỉ công bố những kết quả có lợi và che dấu những kết quả bất lợi cho nhà tài trợ).

Nghiên cứu theo yêu cầu (2)

Customized research

- Phần lớn nghiên cứu ở các trường cao đẳng & đại học (Mỹ) là nghiên cứu theo yêu cầu, và thường được coi là *nghiên cứu gốc* hay *nghiên cứu sơ cấp* (*original or primary research*).
- Đôi khi các đặc tính của nghiên cứu syndicated và nghiên cứu theo yêu cầu được kết hợp trong một nghiên cứu hỗn hợp (*hybrid study*). Các nhà cung cấp sản phẩm syndicated thường có nhiều kho dữ liệu thô (*raw data*), và trong thời đại của máy vi tính như hiện nay thì việc khai thác các kho dữ liệu đó để cho ra những báo cáo theo yêu cầu của khách hàng là quá đơn giản. Vì những nghiên cứu như vậy được khai thác từ những dữ liệu sẵn có nên chúng được gọi là những *phân tích thứ cấp* (*secondary analyses*).

Nghiên cứu theo yêu cầu (3)

Customized research

- ❑ Ví dụ về nghiên cứu (*phân tích*) hỗn hợp của Kantar Media Vietnam:
 - Báo cáo khán giả truyền hình và báo cáo adex
 - Báo cáo cho Goldsun Focus Media
 - Báo cáo về Internet trên cơ sở kết hợp MHS và Yahoo Netindex
- ❑ Ưu điểm của nghiên cứu (*phân tích*) hỗn hợp
 - Làm tăng doanh thu của các nhà cung cấp sản phẩm syndicated với chi phí thấp hơn nhiều so với các nghiên cứu sơ cấp (*primary research*).
 - Vì các phân tích được thực hiện trên cơ sở khai thác dữ liệu syndicated, do đó các con số mang tính chính thống và khách quan hơn so với các kết quả nghiên cứu sơ cấp (*primary research*).
- ❑ Tuy nhiên khi thực hiện các phân tích thứ cấp cần phải chú ý đến các vấn đề như chọn mẫu và cơ sở mẫu (*rất quan trọng*).

Những câu hỏi thường đặt ra trong nghiên cứu

Ai?

Cái gì?

Như thế nào?

Who?

What?

Khi nào?

How?

Ở đâu?

When?

Tại sao?

Where?

Why?

Các bước thực hiện một nghiên cứu

- ❑ Mỗi dự án nghiên cứu đều khác nhau và có tính chất đặc thù riêng. Nhưng các bước thực hiện và hoạt động nghiên cứu đều có những điểm tương đồng để hình thành một quy trình nghiên cứu chung.
- ❑ Quy trình nghiên cứu này:
 - có thể tóm lược thành một chuỗi **10 bước** cần phải thực hiện.
 - mô tả việc thiết kế và thực hiện một dự án nghiên cứu nói chung.
 - có tính tương đối và người nghiên cứu có thể không cần phải tuân theo trật tự của 10 bước một cách chính xác như trình bày ở đây.

Các bước thực hiện một nghiên cứu

1. Xác định/nhận dạng hiện tượng/vấn đề cần nghiên cứu
2. Thiết lập mục tiêu nghiên cứu
3. Thiết kế nghiên cứu
4. Xác định nhu cầu về thông tin và nguồn thu thập thông tin
5. Chọn lựa phương pháp thu thập dữ liệu
6. Thiết kế công cụ thu thập dữ liệu
7. Thử kế mẫu & chọn mẫu
8. Thu thập dữ liệu
9. Nhập liệu, xử lý & phân tích dữ liệu
10. Chuẩn bị và trình bày báo cáo kết quả nghiên cứu

Các bước thực hiện một nghiên cứu

Bước 1: Nhận dạng/xác định vấn đề cần nghiên cứu (1)

- ❑ Là bước đầu tiên và quan trọng nhất trong nghiên cứu.

“Một vấn đề được xác định đúng coi như đã giải quyết được một nửa”

- ❑ Bản chất của vấn đề sẽ quyết định loại nghiên cứu nào cần thực hiện.
- ❑ Các dấu hiệu (hiện tượng) như suy giảm doanh số, lợi nhuận, thị phần hay mức độ trung thành của khách hàng **KHÔNG PHẢI LÀ VẤN ĐỀ**.

Một vấn đề cần nghiên cứu phải được xác định một cách đúng đắn và chính xác, nếu không sẽ rất khó có thể thiết kế được một nghiên cứu tốt.

- ❑ Ví dụ trong lĩnh vực tiếp thị, các vấn đề có thể liên quan đến sự khó khăn hay liên quan đến cơ hội. Đối với cả 2 loại vấn đề này, điều kiện tiên quyết để xác định vấn đề là phải nhận dạng và mổ xẻ vấn đề đó.
- ❑ Hãy phân tích hiện trạng. Nó sẽ cung cấp động lực cơ bản và đòn bẩy cho các nghiên cứu tiếp theo.

Bước 1: Nhận dạng/xác định vấn đề cần nghiên cứu (2)

- ❑ Tìm câu trả lời đúng cho câu hỏi: *Chính xác là khách hàng (công ty, tổ chức v.v.) bản thân người nghiên cứu muốn (hay cần) biết điều gì?*
- ❑ Một câu hỏi căn bản cần làm rõ là: *“Làm sao biết được là có vấn đề?”*
- ❑ Các vấn đề có thể trở nên rõ ràng từ những hoàn cảnh/dấu hiệu sau:
 - *đi chệch hướng so với kế hoạch kinh doanh;*
 - *ghi chép thống kê và báo cáo của doanh nghiệp;*
 - *những sự than phiền hay bất bình của khách hàng;*
 - *đối thoại với các nhân viên của doanh nghiệp;*
 - *quan sát thấy những hành vi hoặc điều kiện không thỏa đáng ở trong doanh nghiệp;*
 - *thành công của các đối thủ của doanh nghiệp;*
 - *Các tài liệu được công bố nói về các vấn đề như các thay đổi trên thị trường hay xu thế biến đổi môi trường, các qui định mới của chính phủ hay những thay đổi được nhìn thấy trước trong nền kinh tế v.v.*

Bước 1: Nhận dạng/xác định vấn đề cần nghiên cứu (3)

- ❑ Một khi các dấu hiệu của vấn đề được phát hiện thì:
 - *Hãy tiến hành tìm kiếm một số dữ kiện ban đầu để xác định bản chất của vấn đề thực sự.*
 - *Nói chuyện với những người khác (nhất là các chuyên gia) về vấn đề và tìm kiếm những tài liệu ban đầu về chủ đề đó.*
- ❑ Trong giai đoạn đầu, người ta có thể chỉ nhận thức được vấn đề một cách **khái quát và chung chung**. Điều này có thể làm cho chương trình nghiên cứu không được thiết kế một cách chu toàn.
- ❑ Người nghiên cứu cũng như nhà quản lý (*hay đối tác đặt hàng nghiên cứu*) cần phải làm việc với nhau để cụ thể hóa vấn đề cần nghiên cứu thành một mệnh đề chính xác và rõ ràng (*không lẫn lộn với vấn đề khác*).

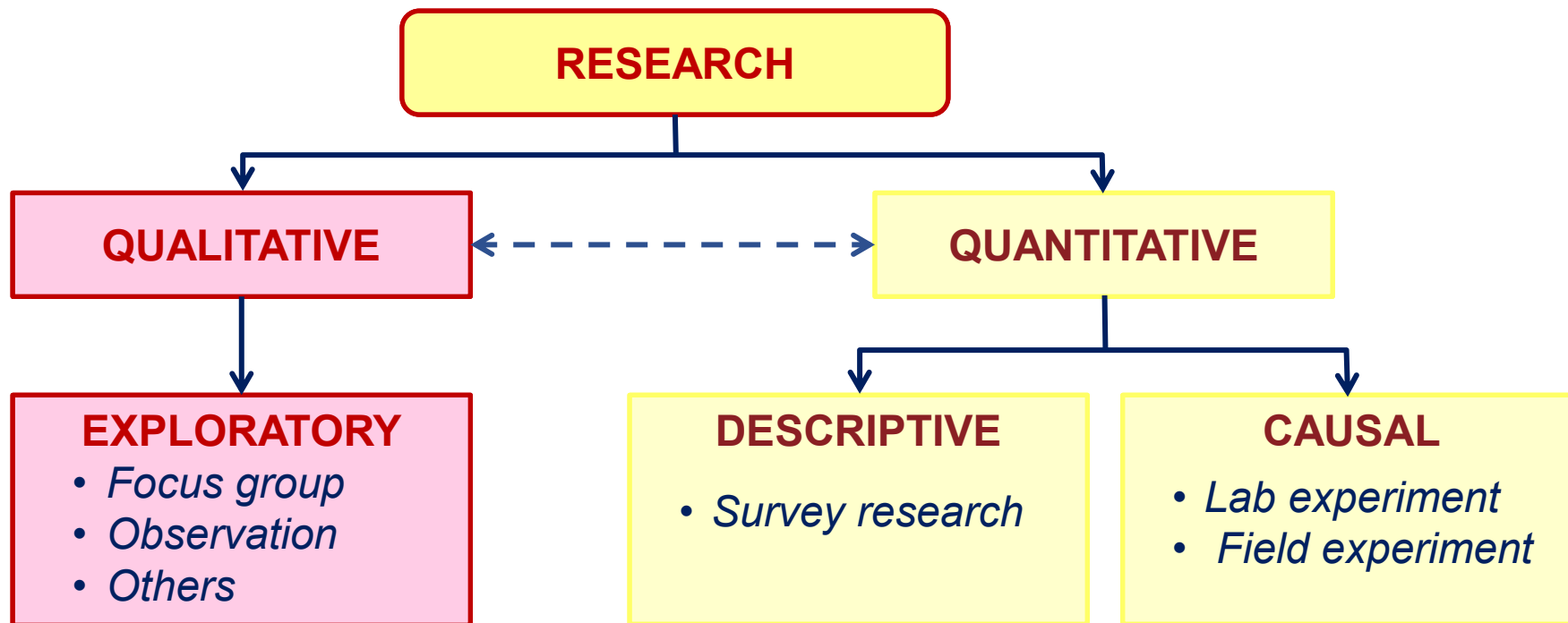
Bước 2: Thiết lập mục tiêu nghiên cứu

“Nếu bạn không biết mình cần tìm cái gì thì bạn sẽ không tìm thấy cái đó”

- ☐ Mục tiêu nghiên cứu có liên hệ đến, và được ấn định bởi, *việc xác định vấn đề*. Để thiết lập mục tiêu nghiên cứu, người nghiên cứu phải trả lời những câu hỏi sau đây:
 - 1) *Dự án nghiên cứu cần cung cấp những thông tin cụ thể nào?*
 - 2) *Nếu dự án cần cung cấp nhiều hơn 1 loại thông tin, thông tin nào là quan trọng nhất?*
 - 3) *Cái gì cần được ưu tiên?*
- ☐ Phát triển các giả thiết có thể sẽ rất hữu ích khi chi tiết hóa các mục tiêu nghiên cứu.
- ☐ Các mục tiêu nghiên cứu phải được thiết lập sao cho sau khi đạt được những mục tiêu đó thì chúng phải cung cấp được những thông tin cần thiết để giải quyết vấn đề.

Bước 3: Thiết kế nghiên cứu

- ❑ Bước tiếp theo nhằm phát triển một kế hoạch nghiên cứu để đạt được mục tiêu đã đề ra.
- ❑ Có nhiều loại thiết kế nghiên cứu và việc lựa chọn thiết kế nào cho phù hợp tùy thuộc vào mục đích nghiên cứu.



Bước 4: Xác định thông tin/dữ liệu cần thu thập (1)

- ❑ Sau khi xác định vấn đề, người nghiên cứu phải quyết định loại thông tin nào sẽ đáp ứng tốt nhất các mục tiêu nghiên cứu đã đề ra.

- ❖ **Thông tin/dữ liệu sơ cấp (hay còn gọi là thông tin cấp 1 - primary information)**

Thông tin sơ cấp là những tài liệu gốc (ví dụ kết quả của cuộc khảo sát được thiết kế riêng cho 1 dự án), và chưa được phân tích (hay sử dụng) bởi bất cứ ai ngoài người tạo ra (hay thu thập) thông tin đó (ví dụ chủ dự án nghiên cứu). Các nguồn thông tin sơ cấp thường gặp là nhật ký, thư, bài phỏng vấn, bài diễn văn, truyện, thơ ca, hay ghi chép từ những sự kiện, hiện trường, v.v. Ví dụ gói dữ liệu từ kết quả khảo sát MHS là nguồn thông tin sơ cấp cho các báo cáo phân tích về truyền thông của Kantar Media VN mà chưa từng công bố trước đó.

- ❖ **Thông tin/dữ liệu thứ cấp (hay còn gọi là thông tin cấp 2 - secondary info)**

Thông tin thứ cấp là những tài liệu xuất phát từ nguồn thông tin sơ cấp đã được xử lý, phân tích và diễn giải (ví dụ để phục vụ cho một dự án đã tạo ra nguồn thông tin đó). Ví dụ các báo cáo về thói quen sử dụng Internet (khai thác từ dữ liệu MHS) có thể trở thành nguồn thông tin thứ cấp cho những báo cáo phân tích khác.

- ❖ **Cả 2 loại thông tin/dữ liệu đều có tầm quan trọng như nhau đối với các nghiên cứu**

Bước 4: Xác định thông tin/dữ liệu cần thu thập (2)

- ❑ **Thông tin/dữ liệu sơ cấp** có thể cung cấp cho chúng ta dữ liệu gốc (*raw data*) và giúp chúng ta phân tích nhiều thứ mà thông tin thứ cấp không thể giúp được. Trong nhiều trường hợp, chỉ có thông tin sơ cấp mới đáp ứng được nhu cầu phân tích đã đề ra, và đó cũng chính là lý do tại sao chúng ta phải tiến hành các cuộc khảo sát riêng cho từng dự án.
- ❑ **Thông tin (dữ liệu) thứ cấp** có thể cung cấp (thêm) cho chúng ta sự hiểu biết về vấn đề đang được nghiên cứu và trong nhiều trường hợp có thể đáp ứng được nhu cầu về thông tin cho 1 dự án nào đó mà không cần phải làm điều tra khảo sát riêng. Trong những trường hợp như vậy, việc tìm hiểu và nắm bắt các nguồn thông tin thứ cấp có thể giúp chúng ta tiết kiệm được rất nhiều thời gian và chi phí.

Bước 5: Thiết kế phương pháp thu thập thông tin yêu cầu

- ❑ Thông tin có thể được thu thập bằng nhiều cách thức khác nhau:
 - *Thư gửi qua đường bưu điện, máy fax*
 - *Phỏng vấn qua điện thoại*
 - *Khảo sát qua Internet*
 - *Phỏng vấn cá nhân*
 - *Sử dụng panel nghiên cứu mẫu (sample panel), bao gồm những người đồng ý cung cấp điều đặn thông tin về thói quen mua sắm, tiêu dung, sử dụng các phương tiện truyền thông, v.v.*

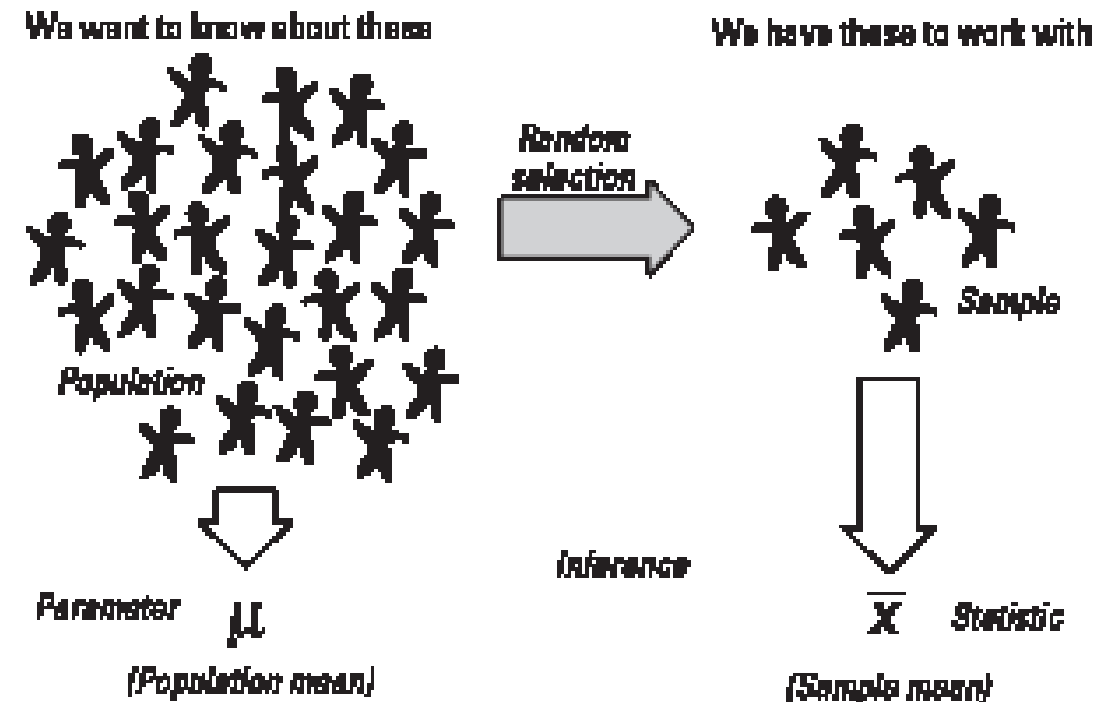
Bước 6: Thiết kế bảng câu hỏi (*questionnaire*)

- ❑ Một trách nhiệm quan trọng của người nghiên cứu là phải thiết kế được công cụ thu thập dữ liệu – ví dụ Bảng câu hỏi - sao cho dễ hiểu và dễ áp dụng cho các đối tượng được nghiên cứu (*đáp viên*).
- Người nghiên cứu phải xác định những tiêu chí để chọn lựa các đối tượng cần nghiên cứu. Ví dụ những người trong độ tuổi 15-54 tuổi và thường xem TV ít nhất 4 ngày/tuần trở lên.
- Thiết kế mẫu khảo sát phải tạo ra một *tập hợp mẫu bao gồm các đối tượng khảo sát phù hợp với mục tiêu nghiên cứu đã đặt ra*.
- Người nghiên cứu có thể có nhiều thiết kế mẫu khảo sát khác nhau để đạt được cùng một mục tiêu của dự án. Điều quan trọng là phải xác định được thiết kế nào là phù hợp nhất (*hay hiệu quả nhất*) đối với từng hoàn cảnh cụ thể của mỗi dự án nghiên cứu.

Các bước thực hiện một nghiên cứu

Bước 7: Thiết kế mẫu khảo sát/nghiên cứu & chọn mẫu

- Số lượng mẫu yêu cầu
- Tổng thể nghiên cứu
- Khung chọn mẫu
- Địa bàn chọn mẫu
- Đơn vị chọn mẫu
- Đối tượng nghiên cứu
- Phương pháp chọn mẫu



Bước 8: Thực hiện thu thập dữ liệu

- ☐ Người nghiên cứu phải quản lý, giám sát và theo dõi quá trình thu thập dữ liệu đúng cách (*theo đúng tiêu chuẩn nếu có*).
 - *Quality Control Standards for Interviewing (QCSI)*
- ☐ Nếu sử dụng phương pháp phỏng vấn, người nghiên cứu phải huấn luyện phỏng vấn viên (PVV) và xây dựng các quy trình kiểm soát chất lượng phỏng vấn.
- ☐ Nếu sử dụng phương pháp thu thập thông tin do đáp viên tự điền Bảng câu hỏi thì có thể bỏ qua các quy trình kiểm soát chất lượng phỏng vấn nêu trên. Nhưng người nghiên cứu phải chuẩn bị tài liệu hướng dẫn kỹ càng để đáp viên tham khảo khi cần.

Bước 9: Nhập liệu, xử lý & phân tích dữ liệu khảo sát

- ❑ Dữ liệu khảo sát “thô” (*‘raw’ research data*) cần phải được:
 - làm sạch (*mã hóa thông tin, bổ sung, chỉnh sửa các sai sót có thể phát sinh từ khâu phỏng vấn đến khâu nhập liệu, áp dụng các bước kiểm tra logic v.v.*);
 - xử lý các bước trung gian (*ví dụ tính tuổi từ thông tin về ngày, tháng & năm sinh*);
 - tổng hợp thành biểu bảng (*tabulation*); và
 - mổ xẻ, phân tích để tìm kiếm kết quả và diễn giải ý nghĩa.
- ❑ Có thể dùng phương pháp thủ công hoặc máy tính.
- ❑ Kế hoạch phân tích phải bám sát mục tiêu nghiên cứu.
- ❑ Sự tương quan cũng như các mối liên hệ giữa các biến số phải được nhận dạng và thảo luận theo hướng giải quyết các vấn đề cụ thể đã được đặt ra.

Các bước thực hiện một nghiên cứu

Bước 10: Báo cáo kết quả nghiên cứu

- ❑ Sản phẩm cuối cùng của các dự án nghiên cứu là những báo cáo do người nghiên cứu thực hiện.
- ❑ Người nghiên cứu cũng thường phải trình bày miệng (oral presentation) cho các cấp quản lý hoặc khách hàng hay bên tài trợ cho dự án.
- ❑ Các báo cáo kết quả nghiên cứu thường bao gồm các nội dung sau:
 - *Giới thiệu về dự án*
 - *Thiết kế dự án và quy trình thực hiện dự án*
 - *Tóm tắt những kết quả nghiên cứu chính yếu*
 - *Phân tích chi tiết kết quả nghiên cứu*
 - *Kết luận và khuyến nghị*
 - *Phần phụ lục bao gồm các tài liệu cần thiết có liên quan đến dự án, kể cả Bảng câu hỏi mẫu, bảng biểu số liệu và đồ thị, hình ảnh minh họa.*
- ❑ Báo cáo phải phù hợp với đối tượng tiếp nhận (*báo cáo dành cho ai?*) và yêu cầu hay sự quan tâm của họ. *Nói 1 cách khác, hãy báo cáo những gì khách hàng yêu cầu chứ không nên báo cáo những gì ta có, trừ phi mục đích của chúng ta là giới thiệu sản phẩm hay những điều mới mẻ.*

Đề xuất nghiên cứu (1)

Research proposal

- ☐ Là kế hoạch thể hiện chi tiết các bước thực hiện một dự án nghiên cứu được đề xuất.
- ☐ Phản ánh sự hiểu biết của người nghiên cứu về vấn đề được quan tâm và khả năng thực hiện nghiên cứu.
- ☐ Nếu việc nghiên cứu sẽ được thực hiện thông qua một cơ quan nghiên cứu (*research agency*), đề xuất nghiên cứu đóng vai trò như một tiêu chí lựa chọn quan trọng.
- ☐ Khi được chấp thuận, đề xuất nghiên cứu sẽ trở thành cơ sở để xác lập hợp đồng hay thỏa thuận giữa cơ quan nghiên cứu và khách hàng, và đóng vai trò như một biên bản ghi nhận những gì đã được thỏa thuận thực hiện.

Đề xuất nghiên cứu (2)

Research proposal

- ❖ Không có khuôn mẫu cố định hay khuôn mẫu chuẩn nào cho một đề xuất nghiên cứu cả, bởi nó phụ thuộc vào tính chất đặc thù của từng dự án nghiên cứu cụ thể.
- ❖ Tuy nhiên, phần lớn các đề xuất nghiên cứu đều có những nội dung sau:

1. *Giới thiệu*
2. *Phân tích & xác định vấn đề (cần quan tâm nghiên cứu, giải quyết v.v.)*
3. *Xác lập các mục tiêu nghiên cứu*
4. *Các chi tiết về kế hoạch thực hiện nghiên cứu được đề xuất*
5. *Tiến độ thực hiện*
6. *Dự trù kinh phí*
7. *Đội ngũ nghiên cứu*

Lý thuyết về mẫu & chọn mẫu trong thực tiễn

- ❑ Tổng thể & mẫu (*population & sample*)
- ❑ Tại sao phải chọn mẫu - Lý thuyết & thực tiễn
- ❑ Điều tra toàn diện/tổng điều tra & điều tra mẫu (*censuses & sample surveys*)
- ❑ Tổng thể nghiên cứu & khung chọn mẫu (*target population & sampling frames*)
- ❑ Các dạng mẫu nghiên cứu/khảo sát (*types of samples*)

Tổng thể & mẫu (1)

Population & sample

- ❑ **Tổng thể nghiên cứu** là tập hợp tất cả những đối tượng mà chúng ta muốn nghiên cứu tìm hiểu.
- ❑ **Tổng thể** là khái niệm tương đối và được xác định bởi:
 - ❖ Đặc điểm của đối tượng cụ thể được quan tâm nghiên cứu:
 - Dân số 15-54 tuổi
 - Phụ nữ có gia đình
 - Nhân viên văn phòng
 - Những người có sử dụng Internet
 - v.v.
 - ❖ Không gian (địa điểm) được quan tâm nghiên cứu:
 - Việt Nam
 - TP. HCM
 - Khu vực nội thành
 - v.v.

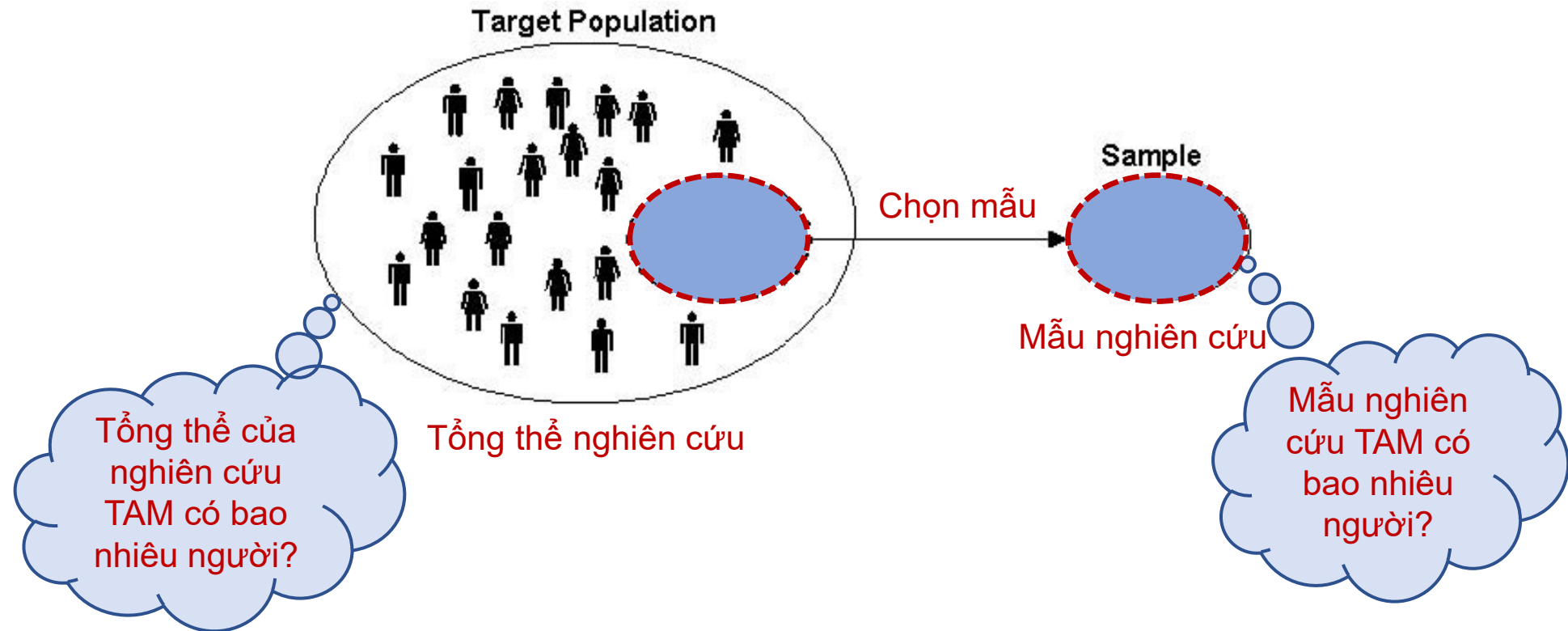


Tổng thể của
nghiên cứu khảo
sát MHS của
KMV là gì?

Tổng thể & mẫu (2)

Population & sample

- ❑ Mẫu là một tập hợp các đối tượng nghiên cứu được chọn lựa (*theo một cách thức nào đó*) từ tổng thể mà chúng ta quan tâm nghiên cứu tìm hiểu.
- ❑ Số lượng cá thể trong một tập hợp thường nhỏ hơn rất nhiều so với số lượng cá thể thuộc tổng thể mà chúng ta quan tâm nghiên cứu.



Tổng thể & mẫu (3)

Population & sample

Tại sao phải nghiên cứu mẫu?

- ❑ Các lý thuyết thống kê chứng minh rằng có thể nghiên cứu các đặc tính, thái độ, hành vi v.v. của các cá thể trong tập hợp mẫu và suy rộng kết quả nghiên cứu mẫu cho tất cả các cá thể trong tổng thể với mức độ chính xác và độ tin cậy xác định.
- ❑ Việc nghiên cứu (*khảo sát*) toàn bộ các cá thể trong 1 tổng thể là không thực tế, không khả thi hoặc không cần thiết vì một hay nhiều trong số những lý do chính yếu sau đây:
 - 1) **Chi phí:** Rõ ràng là chi phí để thu thập thông tin từ một số lượng ít cá thể (*mẫu*) sẽ thấp hơn nhiều so với toàn bộ cá thể trong 1 tổng thể.
 - 2) **Tính chính xác/độ tin cậy:** Nghiên cứu mẫu đem lại khả năng kiểm soát chất lượng và đảm bảo độ chính xác/độ tin cậy cho kết quả nghiên cứu cao hơn nhiều so với việc nghiên cứu một tổng thể (*thường có số lượng cá thể lớn hơn nhiều lần so với mẫu*).

Các nhà nghiên cứu quan điểm xã hội ở Mỹ thường có thể rút ra kết luận chính xác cho toàn bộ dân số Mỹ thông qua kết quả phỏng vấn chỉ 1.000 người.

[Denis List]

Tổng thể & mẫu (4)

Population & sample

Tại sao phải nghiên cứu mẫu? (tiếp theo)

- 3) **Chiều sâu:** Trong cùng điều kiện về chi phí, thời gian, nguồn nhân lực v.v., việc nghiên cứu mẫu tạo điều kiện nghiên cứu sâu sắc hơn về những vấn đề được quan tâm.
- 4) **Tốc độ/thời gian:** Việc thu thập thông tin hay quan sát các đối tượng nghiên cứu thuộc tập hợp mẫu sẽ dễ dàng và nhanh chóng hơn nhiều so với công việc được tiến hành trên toàn bộ các cá thể trong một tổng thể. Đây cũng là yếu tố sống còn mà trong đó yếu tố tốc độ là quyết định tất cả, như các cuộc khảo sát thăm dò kết quả bầu cử ở Mỹ.
- 5) **Tính khả thi:** Có những dự án nghiên cứu đòi hỏi phải có đội ngũ chuyên gia giỏi và những trang thiết bị chuyên dùng vốn chỉ có được với số lượng hạn chế. Trong những trường hợp như vậy thì việc khảo sát toàn bộ các cá thể là không thực tế hoặc không khả thi, và khảo sát mẫu là phương pháp thích hợp nhất để đạt được mục tiêu nghiên cứu.

Chúng ta muốn nghiên cứu phụ nữ 15-49 tuổi. Tổng số phụ nữ này là 1.298.503 người. Nếu mỗi PVV có thể phỏng vấn 5 phụ nữ 1 ngày, và chúng ta có 100 PVV. Phải mất bao nhiêu lâu để phỏng vấn hết số phụ nữ đó?

Tổng điều tra và điều tra mẫu

Censuses & sample surveys

- ❖ **Tổng điều tra:** Là hoạt động thu thập thông tin trên toàn bộ (hoặc gần như toàn bộ) các cá thể trong 1 tổng thể. Một số ví dụ:
 - Tổng điều tra dân số Việt Nam 1/4/2009 (*Population census*)
 - Tổng điều tra nhà ở Việt Nam 1/4/2009 (*Housing census*)
 - Tổng kiểm kê tài sản của 1 doanh nghiệp
- ❖ **Điều tra mẫu:** Là hoạt động thu thập thông tin trên một số lượng nhất định những cá thể (tức là mẫu khảo sát) được chọn lựa từ tổng các cá thể được quan tâm nghiên cứu, tìm hiểu. Một số ví dụ:
 - Điều tra mức sống HGĐ Việt Nam 2010 (*Vietnam Household Living Standard Survey*)
 - Điều tra lao động và việc làm (*Labor & Employment Survey*)
 - Điều tra nhân khẩu học giữa kỳ (*Intercensal Demographic Survey*)

Bạn biết đến những cuộc điều tra khảo sát nào ở Việt Nam?

Tổng điều tra và điều tra mẫu

Những ưu điểm/nhược điểm

Tổng điều tra

- ❖ Dữ liệu thu thập được có thể phân tích cho cả những nhóm đối tượng nhỏ nhất hay những cấp hành chính nhỏ nhất.
- ❖ Dữ liệu có thể sử dụng làm nền tảng cho việc dự báo dân số và hoạch định kinh tế-xã hội dài hạn.
- ❖ Dữ liệu có thể làm cơ sở để thiết kế các cuộc điều tra mẫu với nhiều mục đích khác nhau.
- ❖ Không có những rủi ro từ việc chọn mẫu.
- ❖ Dữ liệu thu thập được không có sai số do bản chất của việc chọn mẫu gây nên (sampling error).

Điều tra mẫu

- ❖ Chi phí (tiền bạc, nhân lực) thấp hơn.
- ❖ Thời gian triển khai thực hiện nhanh hơn.
- ❖ Có thể tiến hành những nghiên cứu phức tạp và sâu sắc hơn.
- ❖ Một số nghiên cứu thông qua nghiên cứu mẫu.
- ❖ Tạo điều kiện kiểm soát chất lượng quá trình thu thập dữ liệu tốt hơn.
- ❖ Tạo điều kiện giảm thiểu những sai sót không thuộc về bản chất của việc chọn mẫu (non-sampling error).
- ❖ Cung cấp kết quả nghiên cứu nhanh chóng & kịp thời.

Khung chọn mẫu

Sampling frames

❑ Khung chọn mẫu có thể là một trong 2 thứ sau:

- 1) *Danh sách của tất cả các cá thể trong một tổng thể*
- 2) *Phương pháp chọn lựa bất kỳ cá thể nào trong tổng thể*

❑ Nếu tổng thể là dân số ở một khu vực địa lý hay 1 đơn vị hành chính cụ thể thì khung chọn mẫu thường là:

- *Danh sách cử tri*
- *Danh sách theo hộ khẩu thường trú & tạm trú*
- *Danh bạ điện thoại*
- *Danh sách khách hàng sử dụng các dịch vụ công ích như điện & nước sinh hoạt.*

❑ Khung chọn mẫu phải là danh sách đầy đủ nhất, chính xác nhất và được cập nhật mới nhất.

❑ Mỗi nước khác nhau có thể có những căn cứ khác nhau dùng để xác lập khung chọn mẫu. Có nơi dùng danh sách HGD, có nơi sử dụng danh sách cá thể (hay nhân khẩu).

Xác định tổng thể, khung chọn mẫu & mẫu



Các dạng mẫu khảo sát

Types of samples

Mẫu nghiên cứu/khảo sát có thể được chọn lựa theo một số cách thức khác nhau:

- ❑ Mẫu xác suất hay mẫu ngẫu nhiên (*probability or random samples*)
- ❑ Mẫu phi xác suất (*non-probability samples*)
 - Mẫu hạn ngạch (*quota samples*)
 - Mẫu mục tiêu (*purposive samples*)
 - Mẫu tình nguyện (*volunteer samples*)
 - Mẫu đa dạng tối đa (*maximum variation samples*)
 - Mẫu tiện lợi (*convenience samples*)
 - Mẫu lan tỏa (*snowball samples*)

MẪU XÁC SUẤT

Probability samples

- ❑ Còn gọi là mẫu ngẫu nhiên (*random samples*);
- ❑ Mỗi phần tử trong tổng thể đều có xác suất (*hay cơ hội*) được chọn như nhau;
- ❑ Mẫu được hình thành từ quá trình chọn lựa mang tính ngẫu nhiên, vì vậy mẫu mang tính khách quan và không bị thiên lệch do vô tình hay cố ý;
- ❑ Cho kết quả khảo sát chính xác nhất & đáng tin cậy nhất;
- ❑ Giúp ước lượng chính xác mức độ sai biệt giữa kết quả khảo sát mẫu so với kết quả đo lường tổng thể (*census*);
- ❑ Là nền tảng để áp dụng các phương pháp đo lường, ước lượng và kiểm định thống kê.

CÁC PHƯƠNG PHÁP CHỌN MẪU XÁC SUẤT

Probability sampling methods

- 1) Chọn mẫu ngẫu nhiên đơn giản (*Simple Random Sampling –SRS*)
- 2) Chọn mẫu ngẫu nhiên có hệ thống (*Systematic Random Sampling*)
- 3) Chọn mẫu ngẫu nhiên phân tầng (*Stratified Random Sampling*)
- 4) Chọn mẫu ngẫu nhiên theo cụm (*Cluster Random Sampling*)
- 5) Chọn mẫu nhiều giai đoạn (*Multi-Stage Sampling*)

CÁC PHƯƠNG PHÁP CHỌN MẪU XÁC SUẤT

Probability sampling methods



1. Chọn mẫu ngẫu nhiên đơn giản

- ❑ Mục tiêu: Chọn lựa sao cho mọi phần tử trong tổng thể đều có cơ hội được chọn làm mẫu như nhau.
- ❑ Phương pháp nào đảm bảo mục tiêu này?
 - Sử dụng Bảng số ngẫu nhiên (*Table of random numbers*)
 - Sử dụng máy tính để tạo ra các con số ngẫu nhiên
 - Sử dụng dụng cụ cơ học để chọn mẫu
- ❑ Điều kiện thực hiện: Phải thiết lập được khung chọn mẫu (*sampling frame*), bao gồm tất cả các phần tử của tổng thể cần nghiên cứu.

73735	45963	78134	63873
02965	58303	90708	20025
98859	23851	27965	62394
33666	62570	64775	78428
81666	26440	20422	05720

15838	47174	76866	14330
89793	34378	08730	56522
78155	22466	81978	57323
16381	66207	11698	99314
75002	80827	53867	37797

99982	27601	62686	44711
84543	87442	50033	14021
77757	54043	46176	42391
80871	32792	87989	72248
30500	28220	12444	71840

CÁC PHƯƠNG PHÁP CHỌN MẪU XÁC SUẤT

Probability sampling methods

1. Chọn mẫu ngẫu nhiên đơn giản

□ Ưu điểm

- Đơn giản, dễ thực hiện và dễ giải thích về tính khách quan/công bằng
- Dễ sử dụng các lý thuyết/phép tính thống kê để suy rộng kết quả khảo sát mẫu

□ Hạn chế

- Trong nhiều trường hợp, việc tạo lập khung chọn mẫu là không khả thi (*về mặt chi phí tiền bạc, thời gian, nhân lực*)
- Không phải là phương pháp hiệu quả nhất về phương diện thống kê
- Không phải là phương pháp luôn luôn tạo ra mẫu có tính đại diện cao nhất cho tổng thể cần nghiên cứu (*ví dụ do bản chất chọn lựa ngẫu nhiên, nếu số mẫu không đủ lớn thì trong mẫu có thể thiếu đại diện của một hay nhiều nhóm phần tử nào đó*).

CÁC PHƯƠNG PHÁP CHỌN MẪU XÁC SUẤT

Probability sampling methods

2. Chọn mẫu ngẫu nhiên có hệ thống

- ❑ Mục tiêu: Chọn lựa sao cho mọi phần tử trong tổng thể đều có cơ hội được chọn làm mẫu như nhau.
- ❑ Phương pháp nào đảm bảo mục tiêu này?
 - Sử dụng bước nhảy có hệ thống (theo một khoảng cách k cố định);
 - Chọn 1 số ngẫu nhiên từ 1 đến k để làm điểm bắt đầu trong danh sách;
- ❑ Điều kiện thực hiện:
 - Phải thiết lập được khung chọn mẫu (*sampling frame*) và đánh số tất cả các phần tử trong danh sách;
 - Danh sách trong khung chọn mẫu phải được sắp xếp (trộn) ngẫu nhiên;
 - Tính khoảng cách bước nhảy k dựa trên số mẫu cần chọn (*sample size so với tổng thể*);

1	26	51	76
2	27	52	77
3	28	53	78
4	29	54	79
5	30	55	80
6	31	56	81
7	32	57	82
8	33	58	83
9	34	59	84
10	35	60	85
11	36	61	86
12	37	62	87
13	38	63	88
14	39	64	89
15	40	65	90
16	41	66	91
17	42	67	92
18	43	68	93
19	44	69	94
20	45	70	95
21	46	71	96
22	47	72	97
23	48	73	98
24	49	74	99
25	50	75	100

CÁC PHƯƠNG PHÁP CHỌN MẪU XÁC SUẤT

Probability sampling methods

2. Chọn mẫu ngẫu nhiên có hệ thống

❑ Ưu điểm

- Đơn giản, trong nhiều trường hợp dễ thực hiện hơn cả phương pháp SRS, và có thể đạt được tính đại diện cao hơn.
- Dễ sử dụng các lý thuyết/phép tính thống kê để suy rộng kết quả khảo sát mẫu

❑ Hạn chế

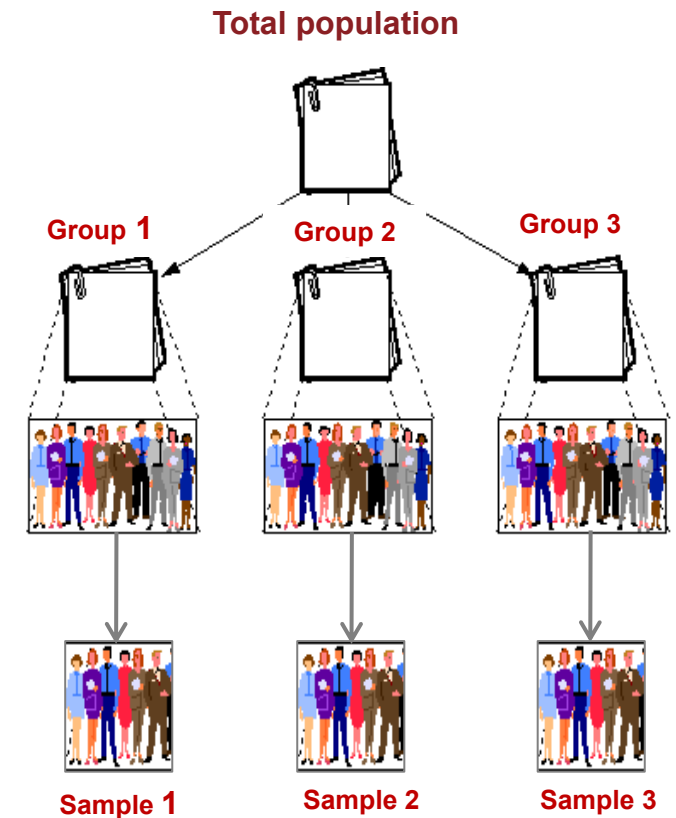
- Trong nhiều trường hợp, việc tạo lập khung chọn mẫu là không khả thi (*về mặt chi phí tiền bạc, thời gian, nhân lực*)
- Không phải là phương pháp hiệu quả nhất về phương diện thống kê

CÁC PHƯƠNG PHÁP CHỌN MẪU XÁC SUẤT

Probability sampling methods

3. Chọn mẫu ngẫu nhiên phân tầng

- ❑ Đôi khi còn gọi là phương pháp chọn mẫu tỷ lệ thuận (*proportional*) hay chọn mẫu ngẫu nhiên theo hạn ngạch (*quota random sampling*).
- ❑ Mục tiêu: Phân chia tổng thể thành các nhóm (tầng) không trùng lấp, sau đó chọn ngẫu nhiên một số mẫu nhất định trong từng nhóm (tầng).
- ❑ Phương pháp chọn mẫu trong từng nhóm (tầng):
- ❑ Chọn ngẫu nhiên với số mẫu tỷ lệ thuận với tổng số phần tử của từng nhóm/tầng (*proportionate stratified sampling*);
- ❑ Chọn ngẫu nhiên với số mẫu không tỷ lệ thuận với tổng số phần tử của từng nhóm/tầng (*disproportionate stratified sampling*);



CÁC PHƯƠNG PHÁP CHỌN MẪU XÁC SUẤT

Probability sampling methods

3. Chọn mẫu ngẫu nhiên phân tầng

❑ Ưu điểm

- Mẫu không những có thể đại diện cho tổng thể nói chung, mà có thể đại diện cho các nhóm phần tử nhỏ hơn trong tổng thể, nhất là các nhóm có số lượng phần tử tương đối ít (*xác suất rơi vào mẫu rất nhỏ*).
- Có thể sử dụng tỷ lệ chọn mẫu khác nhau trong từng nhóm/tầng để cho phép có được số mẫu tối thiểu cần thiết cho các phân tích.
- Kết quả khảo sát thường có mức độ chính xác cao hơn so với phương pháp SRS do mẫu có tính đại diện cao hơn.

❑ Hạn chế

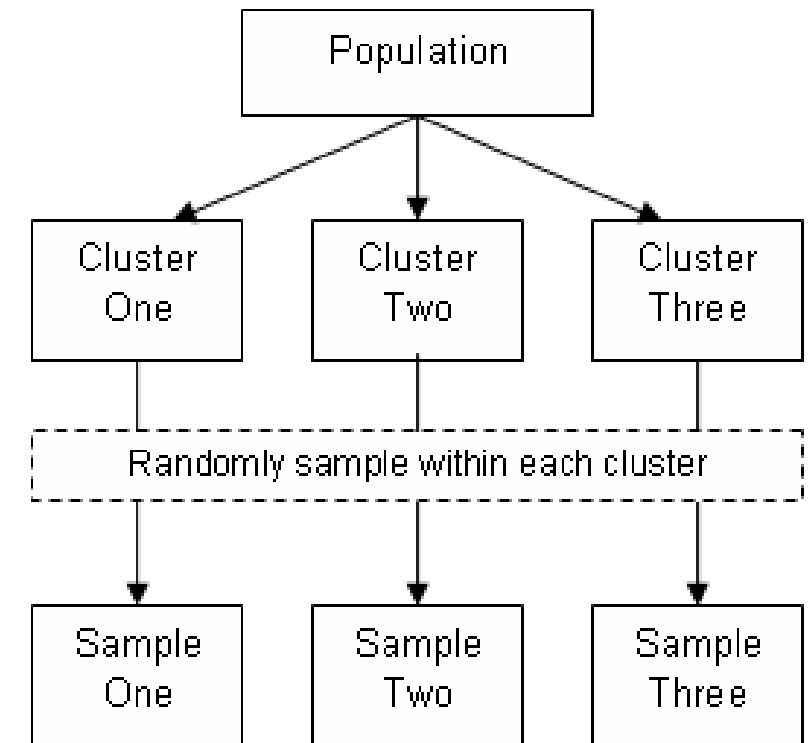
- Đòi hỏi phải có những thông tin cần thiết làm cơ sở cho việc phân nhóm (tầng);
- Nếu việc phân nhóm (tầng) không chính xác thì hiệu quả đạt được có thể không được như mong muốn.

CÁC PHƯƠNG PHÁP CHỌN MẪU XÁC SUẤT

Probability sampling methods

4. Chọn mẫu ngẫu nhiên theo cụm (khu vực)

- ❑ Khắc phục những nhược điểm về **tính khả thi** của phương pháp SRS.
- ❑ Phân chia tổng thể thành các cụm (*dân cư*) nhỏ (thường dựa vào ranh giới hành chính hoặc địa lý;
- ❑ Chọn mẫu một số cụm nhất định từ tổng số cụm trong toàn bộ tổng thể;
- ❑ Khảo sát tất cả các phần tử trong các cụm được chọn mẫu hoặc tiếp tục chọn mẫu ngẫu nhiên một số phần tử nhất định trong cụm được chọn để khảo sát.



CÁC PHƯƠNG PHÁP CHỌN MẪU XÁC SUẤT

Probability sampling methods

4. Chọn mẫu ngẫu nhiên theo cụm (khu vực)

❑ Ưu điểm

- Mang tính hiệu quả kinh tế cao;
- Thích hợp với những hoàn cảnh bị bó buộc về điều kiện đi lại, thời gian, nhân lực & chi phí đầu tư cho dự án khảo sát.

❑ Hạn chế

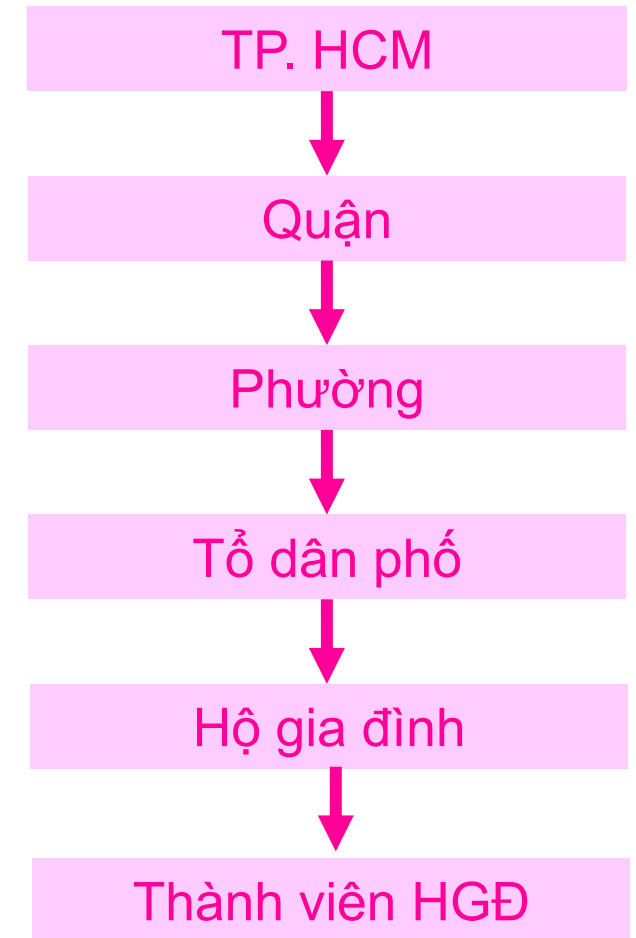
- Việc áp dụng các tính toán thống kê cho kết quả khảo sát mẫu phức tạp hơn so với trường hợp dùng phương pháp SRS;
- Để đạt được cùng mức độ chính xác & độ tin cậy như phương pháp SRS, phương pháp này đòi hỏi số mẫu phải lớn hơn.

CÁC PHƯƠNG PHÁP CHỌN MẪU XÁC SUẤT

Probability sampling methods

5. Chọn mẫu ngẫu nhiên nhiều giai đoạn

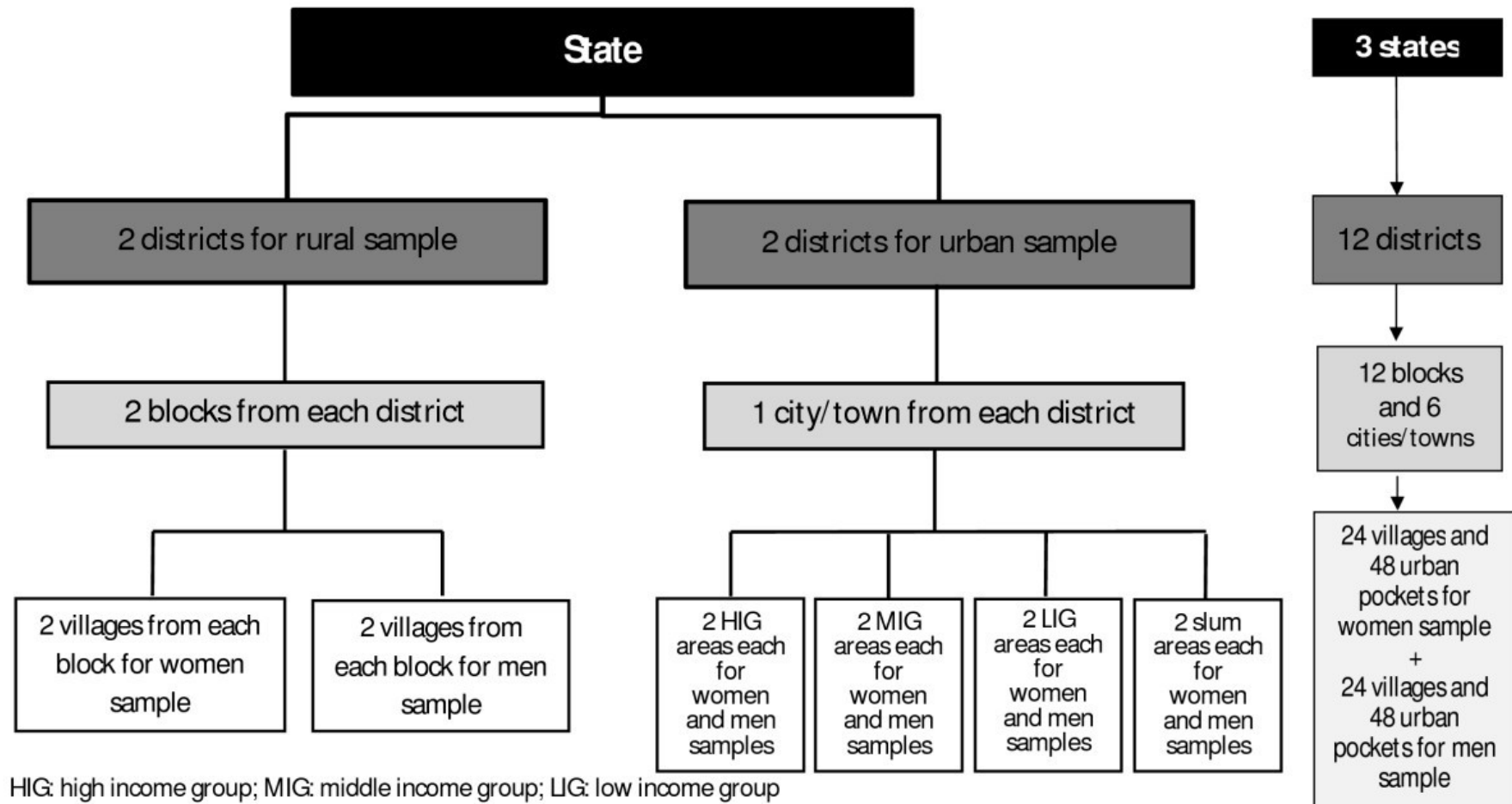
- ❑ Các điều kiện trong thực tiễn thường rất phức tạp, đòi hỏi phải có sự kết hợp nhiều phương pháp chọn mẫu ngẫu nhiên khác nhau để vừa đáp ứng yêu cầu đề ra, vừa đạt được hiệu quả về mặt kinh tế.
- ❑ Việc chọn đối tượng khảo sát có thể thực hiện qua nhiều giai đoạn.
- ❑ Mỗi giai đoạn có thể áp dụng một phương pháp chọn mẫu phù hợp với hoàn cảnh cụ thể & mục đích nghiên cứu đã đề ra.
- ❑ Được sử dụng rộng rãi trong thực tiễn.
- ❑ Ứng dụng cho nhiều loại khảo sát nghiên cứu kinh tế-xã hội khác nhau.



CÁC PHƯƠNG PHÁP CHỌN MẪU XÁC SUẤT

Probability sampling methods

5. Chọn mẫu nhiều giai đoạn



CÁC PHƯƠNG PHÁP CHỌN MẪU XÁC SUẤT

Probability sampling methods

Tìm hiểu thêm về chọn mẫu phân tầng

3 vấn đề then chốt cần giải quyết

- ❖ **Cơ sở phân tầng** – Những đặc điểm nào nên được sử dụng để phân chia tổng thể thành các tầng (nhóm)?
 - Các đặc điểm có ảnh hưởng quan trọng đến kết quả nghiên cứu/đo lường như giới tính, độ tuổi, học vấn, thu nhập v.v.
 - Mức độ khác biệt giữa các cá thể trong từng tầng là tối thiểu
 - Mức độ khác biệt giữa các tầng là tối đa
- ❖ **Số tầng** – Nên phân thành bao nhiêu tầng (*nhóm*)?
 - Càng nhiều tầng càng có cơ hội nâng cao mức độ đồng nhất trong từng tầng
 - Thông tin sẵn có để làm cơ sở cho việc phân tầng
 - Chi phí đầu tư và hiệu quả kinh tế
- ❖ **Cơ sở mẫu trong từng tầng** – Ở mỗi tầng (*nhóm*) cần chọn bao nhiêu mẫu?
 - Phân bổ mẫu tỷ lệ thuận với qui mô của từng tầng (*proportional allocation*)
 - Phân bổ mẫu tối ưu (*optimum allocation*)
 - Mục đích/yêu cầu nghiên cứu
 - Đặc điểm của mỗi tầng

CÁC PHƯƠNG PHÁP CHỌN MẪU XÁC SUẤT

Probability sampling methods

So sánh chọn mẫu phân tầng và chọn mẫu theo cụm

Chọn mẫu phân tầng

- ❖ Cơ sở phân tầng là (những) đặc điểm có ảnh hưởng đến kết quả nghiên cứu/đo lường như các đặc điểm nhân khẩu học hay KT-XH.
- ❖ Các phần tử trong mỗi tầng mang tính đồng nhất cao.
- ❖ Qui mô (dân số) của mỗi tầng có thể rất khác nhau.
- ❖ Tổng thể nên được chia thành nhiều tầng để nâng cao tính đồng nhất trong từng tầng & tính khác biệt giữa các tầng.
- ❖ Quá trình chọn mẫu có thể được tiến hành qua nhiều giai đoạn (*stage*).
- ❖ Ưu điểm chính là khả năng nâng cao mức độ chính xác cho kết quả khảo sát/đo lường.

Chọn mẫu theo cụm

- ❖ Cơ sở phân cụm là khu vực/vùng địa lý hay ranh giới hành chính. Ví dụ các vùng KTXH, các tỉnh/TP, huyện/quận, xã/phường v.v.
- ❖ Các phần tử trong mỗi cụm không nhất thiết phải có tính đồng nhất cao.
- ❖ Qui mô (dân số) của mỗi cụm nên tương đương nhau thì tốt hơn.
- ❖ Tổng thể nên được chia thành nhiều cụm có qui mô nhỏ thì tốt hơn là được phân chia thành ít cụm nhưng mỗi cụm lại có qui mô lớn.
- ❖ Chọn mẫu 1, 2 hay nhiều giai đoạn (*one-stage, two-stage or multi-stage*).
- ❖ Ưu điểm chính là tính khả thi cao về mặt kinh tế, nhưng cơ sở mẫu thường phải lớn hơn.

MẪU PHI XÁC SUẤT

Non-probability samples

- ❑ Không áp dụng qui trình chọn mẫu ngẫu nhiên.
- ❑ Không nhất thiết không có tính đại diện cho tổng thể.
- ❑ Không thể dựa vào lý thuyết xác suất đơn thuần để đánh giá tính chính xác và mức độ tin cậy của kết quả khảo sát.
- ❑ Được áp dụng trong những hoàn cảnh/điều kiện nhất định, thường là trong nghiên cứu khoa học xã hội ứng dụng (*applied social research*).
 - ❖ *Mẫu hạn ngạch (quota samples)*
 - ❖ *Mẫu mục tiêu (purposive samples)*
 - ❖ *Mẫu tình nguyện (volunteer samples)*
 - ❖ *Mẫu đa dạng tối đa (maximum variation samples)*
 - ❖ *Mẫu tiện lợi (convenience samples)*
 - ❖ *Mẫu lan tỏa (mẫu liên đới/mẫu theo mối quan hệ/mẫu dây chuyền) (snowball samples)*

MẪU PHI XÁC SUẤT

Non-probability samples

1. Mẫu hạn ngạch (*Quota samples*)

❑ Một dạng của Mẫu phân tầng

- Quota được xác lập theo tỷ lệ (*proportional quota sampling*). Ví dụ mẫu MHS.
- Quota được xác lập không theo tỷ lệ (*non-proportional quota sampling*). Ví dụ mẫu Netindex.

❑ Quy trình chọn mẫu:

- Có áp dụng nguyên tắc chọn mẫu ngẫu nhiên (nhiều hoặc ít, chặt chẽ hoặc không chặt chẽ).
- Hoàn toàn không áp dụng nguyên tắc chọn mẫu ngẫu nhiên. Ví dụ PVV có thể hoàn toàn quyết định chọn đối tượng PV theo các tiêu chí định trước.

❑ Ưu điểm, nhược điểm và tính ứng dụng cho đến nay vẫn còn là vấn đề tranh cãi (*ví dụ giữa các nhà thống kê và các nhà nghiên cứu thị trường*).

MẪU PHI XÁC SUẤT

Non-probability samples

1. Mẫu hạn ngạch (*Quota samples*)

❑ Những lý lẽ không ủng hộ

- Không thể ước lượng sai số chọn mẫu (*sampling errors*).
- Việc không áp dụng (*triệt để*) nguyên tắc chọn mẫu ngẫu nhiên có thể dẫn đến sự quyết định tùy tiện của PVV, và mẫu không mang tính đại diện. Chẳng hạn những người được chọn cho nhóm quota 45-49 tuổi có thể toàn là những người ở 1 độ tuổi nào đó chứ không rải đều ở cả 5 độ tuổi. Hoặc PVV có thể ưu tiên chọn những người dễ tiếp cận để phỏng vấn (*ví dụ người nội trợ hay có mặt ở nhà*).
- Việc kiểm soát quota theo các đặc điểm KT-XH có thể lệ thuộc quá nhiều vào phán xét chủ quan của PVV.
- Khó khăn hơn trong việc kiểm soát công tác thực địa (*fieldwork*). Chẳng hạn PVV có thể làm sai lệch thông tin về đáp viên để thỏa mãn các tiêu chí kiểm soát hạn ngạch (*quota*). Điều đó làm cho kết quả khảo sát/đo lường có thể bị sai lệch hoặc không đáng tin cậy.

MẪU PHI XÁC SUẤT

Non-probability samples

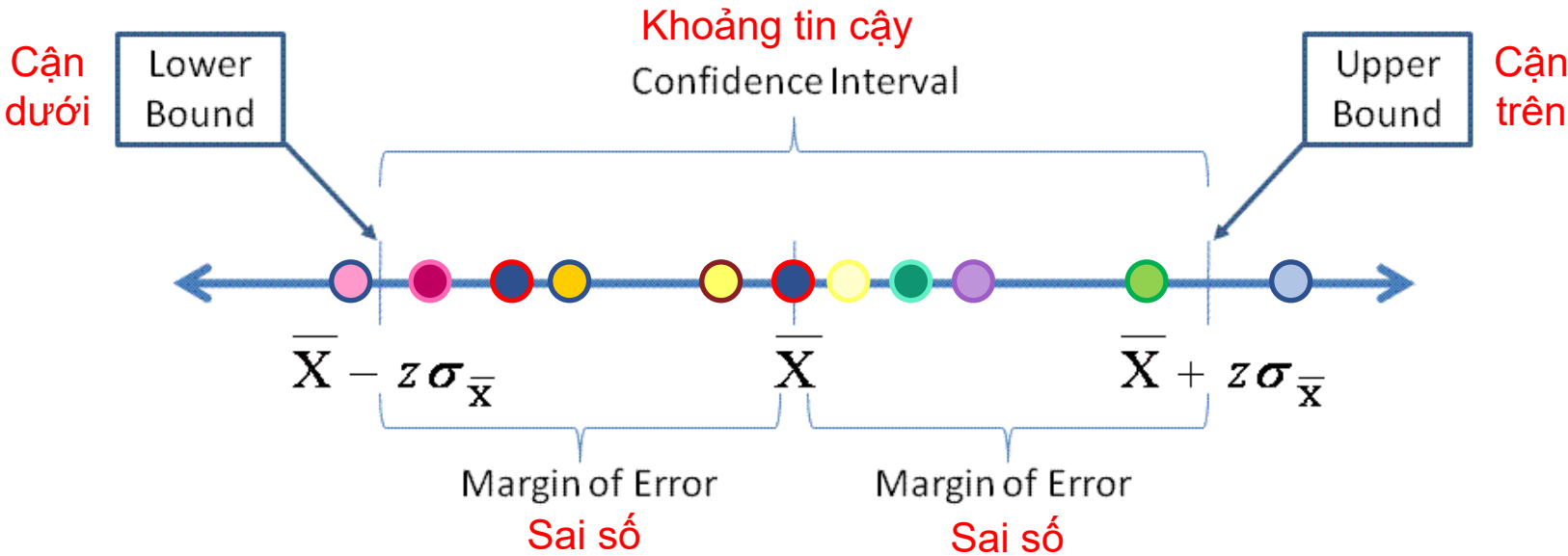
1. Mẫu hạn ngạch (*Quota samples*)

❑ Những lý lẽ ủng hộ

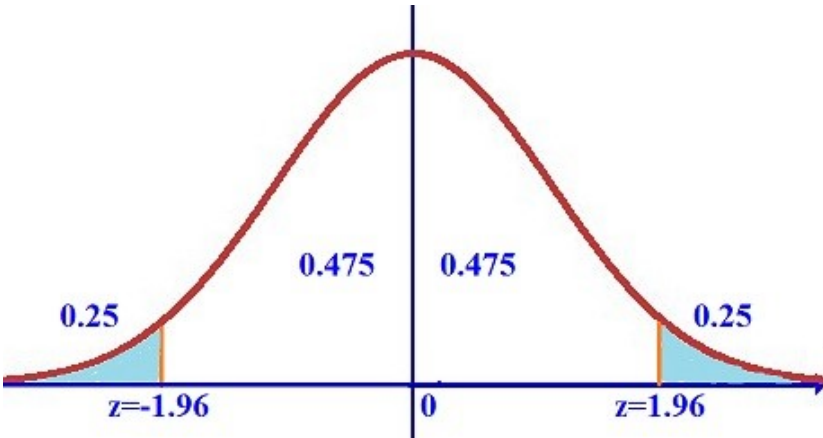
- Chi phí thấp hơn. Chi phí cho 1 cuộc khảo sát mẫu quota có thể chỉ bằng $\frac{1}{2}$ hay thậm chí $\frac{1}{3}$ chi phí của 1 cuộc khảo sát ngẫu nhiên. Tuy nhiên cần lưu ý rằng việc tiết kiệm chi phí phải đánh đổi với sự suy giảm về mức độ chính xác và độ in cậy của kết quả khảo sát.
- Quản lý & điều hành dễ dàng hơn. Yêu cầu về nhân lực thấp hơn có thể là giảm đáng kể áp lực về mặt quản lý con người và các công tác đảm bảo chất lượng khác.
- Nếu thời gian là yếu tố then chốt thì chọn mẫu quota có thể là lựa chọn duy nhất. Ví dụ để thăm dò công luận hay phản ứng của công chúng về 1 sự kiện nào đó trong 1 thời gian rất ngắn.
- Mẫu quota không phụ thuộc vào việc thiết lập khung chọn mẫu. Trong nhiều trường hợp, việc thiết lập khung chọn mẫu có thể không khả thi về chi phí đầu tư, nguồn nhân lực và thời gian.

KHOẢNG TIN CẬY, MỨC TIN CẬY & SAI SỐ MẪU

Confidence Interval, Confidence Level & Sampling Error



Mức tin cậy	Giá trị Z
Confidence level	Z value
90%	1.65
95%	1.96
99%	2.58
99,9%	3.291



KHOẢNG TIN CẬY, MỨC TIN CẬY & SAI SỐ MẪU

Confidence Interval, Confidence Level & Sampling Error

Những phát biểu sau đây có ý nghĩa như thế nào đối với bạn?

Kết quả đo lường khán giả truyền hình cho thấy tỷ lệ dân số 4+ xem tivi hàng ngày ước tính nằm trong khoảng 80%-90%

Chúng ta *chắc chắn* 95% rằng tỷ lệ dân số 4+ xem tivi hàng ngày sẽ nằm trong khoảng 80%-90%

Kết quả đo lường khán giả truyền hình cho thấy tỷ lệ dân số 4+ xem tivi hàng ngày được ước tính khoảng 85%, với sai số (*khảo sát mẫu*) là $\pm 5\%$

1 SỐ ĐẠI LƯỢNG THỐNG KÊ MÔ TẢ

Descriptive Statistics

- ☐ Giá trị nhỏ nhất (*Minimum – Min*)
- ☐ Giá trị lớn nhất (*Maximum – Max*)
- ☐ Giá trị trung bình (*Mean, Average*)
- ☐ Giá trị trung vị (*Median*)
- ☐ Giá trị Mode (*Mode*)
- ☐ Phương sai (*Variance*)
- ☐ Độ lệch chuẩn (*Standard Deviation – Stdev*)

THỐNG KÊ MÔ TẢ

Descriptive Statistics

- ❑ **Giá trị trung bình của tổng thể**
Population Mean/Average

$$\mu = \frac{\sum_{i=1}^N X_i}{N} = \frac{110}{20} = 5.5$$

- ❑ **Phương sai của tổng thể**
Population Variance

$$\sigma^2 = \frac{\sum_{i=1}^N (X_i - \mu)^2}{N} = \frac{415}{20} = 20.75$$

- ❑ **Độ lệch chuẩn của tổng thể**
Population Standard Deviation

$$\sigma = \sqrt{\sigma^2} = \sqrt{20.75} = 4.555$$

Phân phối thu nhập (X_i) của 20 cá thể

i	X_i	$X_i - \mu$	$(X_i - \mu)^2$
1	3	-2.5	6.25
2	8	2.5	6.25
3	5	-0.5	0.25
4	4	-1.5	2.25
5	3	-2.5	6.25
6	0	-5.5	30.25
7	2	-3.5	12.25
8	2	-3.5	12.25
9	1	-4.5	20.25
10	4	-1.5	2.25
11	10	4.5	20.25
12	8	2.5	6.25
13	7	1.5	2.25
14	12	6.5	42.25
15	20	14.5	210.25
16	5	-0.5	0.25
17	1	-4.5	20.25
18	7	1.5	2.25
19	6	0.5	0.25
20	2	-3.5	12.25
Σ	110	0	415

THỐNG KÊ MÔ TẢ

Descriptive Statistics

- ❑ **Giá trị trung bình của mẫu**
Sample Mean/Average

$$\bar{X} = \frac{\sum_1^N X_i}{N} = \frac{110}{20} = 5.5$$

- ❑ **Phương sai của mẫu**
Sample Variance

$$S^2 = \frac{\sum_1^n (X_i - \bar{X})^2}{n-1} = \frac{415}{19} = 21.84$$

- ❑ **Độ lệch chuẩn của mẫu**
Sample Standard Deviation

$$S = \sqrt{S^2} = \sqrt{21.84} = 4.674$$

Phân phối thu nhập (X_i) của 20 cá thể

i	X_i	$X_i - \mu$	$(X_i - \mu)^2$
1	3	-2.5	6.25
2	8	2.5	6.25
3	5	-0.5	0.25
4	4	-1.5	2.25
5	3	-2.5	6.25
6	0	-5.5	30.25
7	2	-3.5	12.25
8	2	-3.5	12.25
9	1	-4.5	20.25
10	4	-1.5	2.25
11	10	4.5	20.25
12	8	2.5	6.25
13	7	1.5	2.25
14	12	6.5	42.25
15	20	14.5	210.25
16	5	-0.5	0.25
17	1	-4.5	20.25
18	7	1.5	2.25
19	6	0.5	0.25
20	2	-3.5	12.25
Σ	110	0	415

PHÂN PHỐI CHUẨN

Normal Distribution

Hàm phân phối mật độ xác suất $f(x)$

Probability Density Function

❖ Tính chất:

- Đường cong hình chuông
- Đối xứng quanh giá trị trung bình μ
- Chỉ có 1 giá trị Mode = μ

❖ 2 tham số quyết định:

- 1) Giá trị trung bình μ
- 2) Độ lệch chuẩn σ

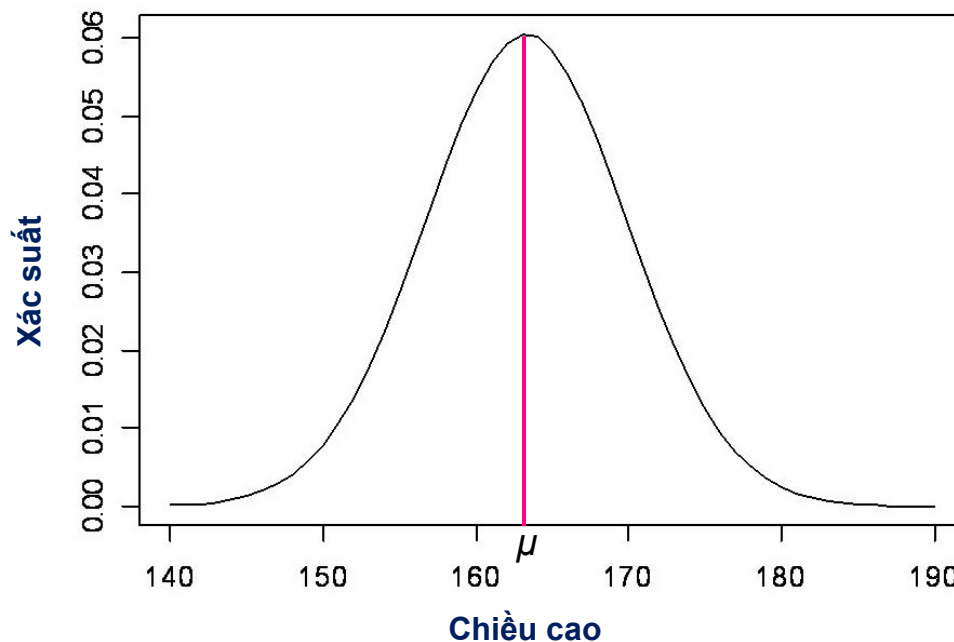
❖ Ví dụ: Có bao nhiêu % đàn ông VN cao 160 cm?

$$P(X = 160) = \frac{1}{6.6\sqrt{2 \times 3.1416}} \exp \left[-\frac{(160 - 163.3)^2}{2 \times (6.6)^2} \right]$$
$$= 0.0533 = 5.3\%$$

❖ Cách diễn đạt khác: nếu chọn ngẫu nhiên 1 người đàn ông VN, cơ hội để gặp được người cao 160 cm là **5.3%**

$$P(X = x \mid \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left[-\frac{(x - \mu)^2}{2\sigma^2} \right]$$

Xác suất phân bố đàn ông Việt Nam theo chiều cao



Trung bình $\mu = 163.3 \text{ cm}$

Độ lệch chuẩn $\sigma = 6.6 \text{ cm}$

PHÂN PHỐI CHUẨN

Normal Distribution

3 ứng dụng quan trọng

- 1) Nhiều hiện tượng diễn ra trong tự nhiên & xã hội có dạng **phân phối chuẩn** hoặc **gần giống dạng phân phối chuẩn**. Chúng ta có thể tìm thấy rất nhiều ví dụ ở những lĩnh vực khác nhau.
- 2) Các phân phối dựa trên kết quả của nhiều quan sát (nhiều cá thể được quan sát hoặc nhiều lần quan sát) sẽ tiến tới dạng **phân phối chuẩn** khi số lần quan sát n đủ lớn (**$n > 30$ thường được coi là đủ lớn**)
- 3) Phân phối xác suất (xảy ra hiện tượng nào đó) với n lớn là phân phối chuẩn. Định luật giới hạn trung tâm (*Central Limit Theorem*) phát biểu rằng nếu một tổng thể có giá trị trung bình là μ và độ lệch chuẩn là σ , thì một tập hợp mẫu với số lượng mẫu n được chọn đủ lớn ($n > 30$) sẽ cho kết quả quan sát (độc lập nhau) có dạng phân phối chuẩn với cùng giá trị trung bình là μ và độ lệch chuẩn là $\frac{\sigma}{\sqrt{n}}$.

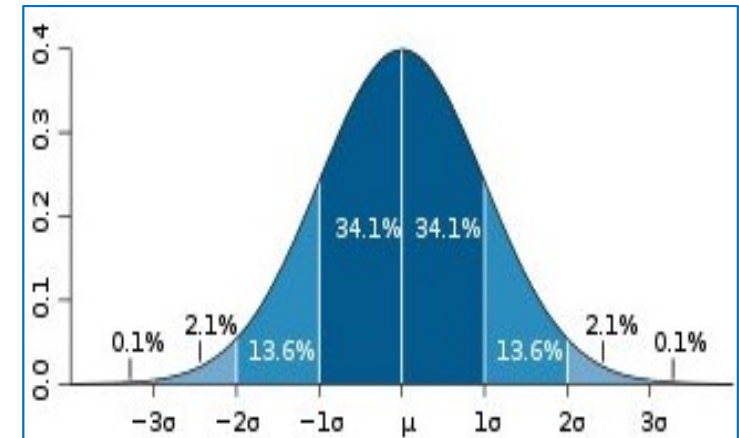
PHÂN PHỐI CHUẨN

Normal Distribution

Ứng dụng thực tiễn

- ❖ Có bao nhiêu % đàn ông VN có chiều cao 185 cm?
- ❖ Có bao nhiêu % người lao động VN có mức thu nhập từ 5–10 triệu đồng?
- ❖ Có bao nhiêu % dân số VN có trình độ từ đại học trở lên (hoặc dưới đại học)?
- ❖ Có thể ước đoán mức độ chính xác của kết quả khảo sát mẫu về TV rating?
- ❖ Kết quả của 1 cuộc khảo sát mẫu cho thấy chiều cao trung bình của đàn ông VN là 165-170 cm. Nếu làm 1 cuộc khảo sát tương tự khác, cơ hội để có cùng kết quả là bao nhiêu %?

Tỷ lệ cá thể có giá trị nằm trong 1 khoảng cách nhất định tính từ giá trị trung bình của tổng thể



Quy tắc đã kiểm chứng (Empirical rule)

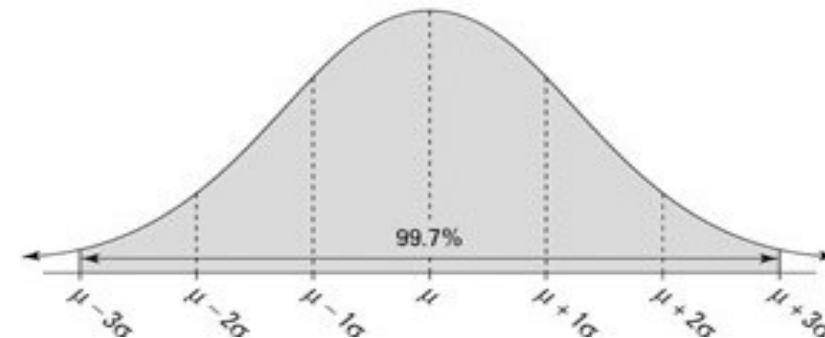
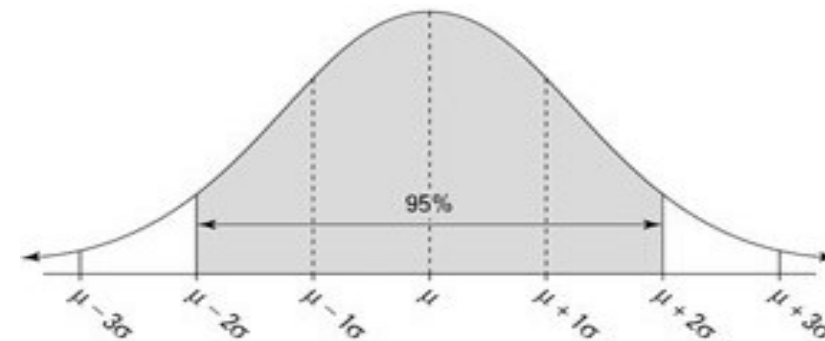
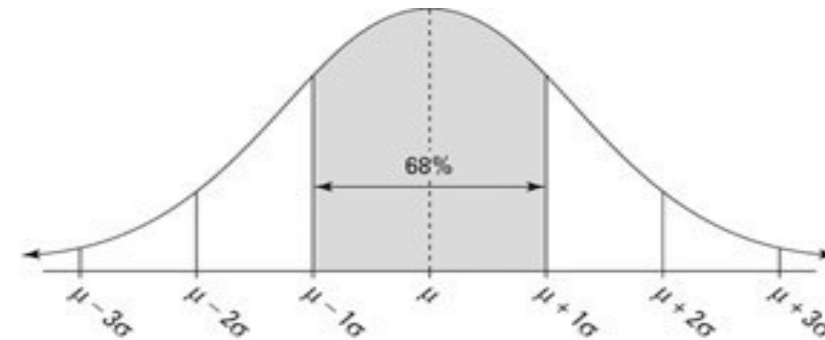
- ❖ Khoảng **68%** tổng số cá thể trong tổng thể có giá trị (ví dụ mức thu nhập) nằm trong khoảng $\mu \pm 1\sigma$
- ❖ Khoảng **95%** tổng số cá thể trong tổng thể có giá trị (ví dụ chỉ số IQ) nằm trong khoảng $\mu \pm 2\sigma$
- ❖ Khoảng **99.7%** tổng số cá thể trong tổng thể có giá trị (ví dụ số năm đi học) nằm trong khoảng $\mu \pm 3\sigma$

PHÂN PHỐI CHUẨN

Qui tắc đã kiểm chứng

Empirical rule

- ❖ Khoảng **68%** tổng số cá thể trong tổng thể có giá trị (ví dụ mức thu nhập) nằm trong khoảng $\mu \pm 1\sigma$
- ❖ Khoảng **95%** tổng số cá thể trong tổng thể có giá trị (ví dụ chỉ số IQ) nằm trong khoảng $\mu \pm 2\sigma$
- ❖ Khoảng **99.7%** tổng số cá thể trong tổng thể có giá trị (ví dụ số năm đi học) nằm trong khoảng $\mu \pm 3\sigma$



PHÂN PHỐI CHUẨN CHUẨN HÓA

Standardized Normal Distribution

Hàm phân phối xác suất $f(z)$

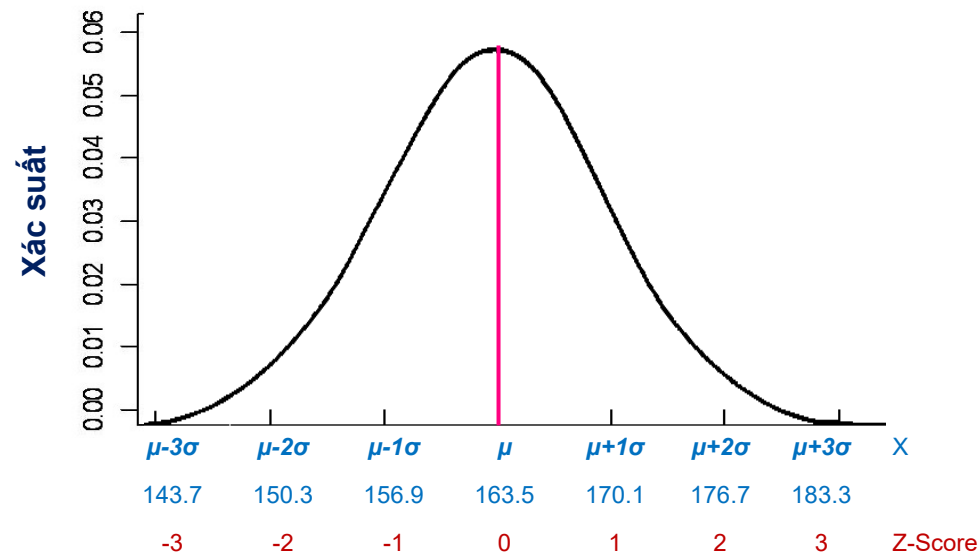
- ❖ Biến đổi giá trị X thành chỉ số Z (Z-Score)

$$Z = \frac{x - \mu}{\sigma}$$

- ❖ Chỉ số Z độc lập v lượng (ví dụ không có đơn vị là cm hay VN đồng)
- ❖ Giá trị trung bình $\mu = 0$
- ❖ Độ lệch chuẩn $\sigma = 1$
- ❖ Z chính là độ lệch chuẩn giữa từng cá thể và giá trị trung bình của tổng thể.
- ❖ $Z > 0$ có nghĩa là cá thể có giá trị lớn hơn giá trị trung bình của tổng thể
- ❖ $Z < 0$ có nghĩa là cá thể có giá trị nhỏ hơn giá trị trung bình của tổng thể

$f(z)$

Xác suất phân bố đàn ông Việt Nam theo chiều cao
[Với X đã được biến đổi sang chỉ số Z (Z-Score)]



PHÂN PHỐI CHUẨN CHUẨN HÓA

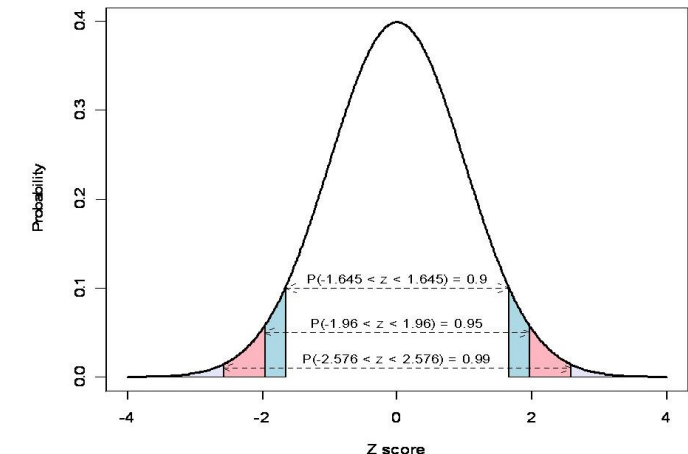
Standardized Normal Distribution

Ứng dụng thực tiễn

- ❖ Có thể tính được xác suất tương ứng với 1 giá trị Z cho trước mà không cần phải biết giá trị μ và σ .
- ❖ Sử dụng Bảng phân phối xác suất tích lũy chuẩn hóa (*Table of Standardized Normal Cumulative Distribution*)
- ❖ Sử dụng các hàm thống kê trong *Excel* để tính xác suất tương ứng với giá trị X hoặc Z cho trước:
 - Phân phối chuẩn: ***NORMDIST***(*x,mean,standard_dev,cumulative*)
 - Phân phối chuẩn chuẩn hóa: ***NORMSDIST***(*z*)
- ❖ Sử dụng các hàm thống kê trong *Excel* để tính giá trị X hoặc Z tương ứng với xác suất cho trước:
 - Phân phối chuẩn: ***NORMINV***(*probability,mean,standard_dev*)
 - Phân phối chuẩn hóa: ***NORMSINV***(*probability*)

Table of Standardized Normal Cumulative Distribution

Z	.00	.01	.02	.03
0.0	.0000	.0040	.0080	.0120
0.1	.0398	.0438	.0478	.0517
0.2	.0793	.0832	.0871	.0910
0.3	.1179	.1217	.1255	.1293
0.4	.1554	.1591	.1628	.1664
0.5	.1915	.1950	.1985	.2019
0.6	.2257	.2291	.2324	.2357
0.7	.2580	.2612	.2642	.2673
0.8	.2881	.2910	.2939	.2967
0.9	.3159	.3186	.3212	.3238
1.0	.3413	.3438	.3461	.3485
1.1	.3643	.3665	.3686	.3708

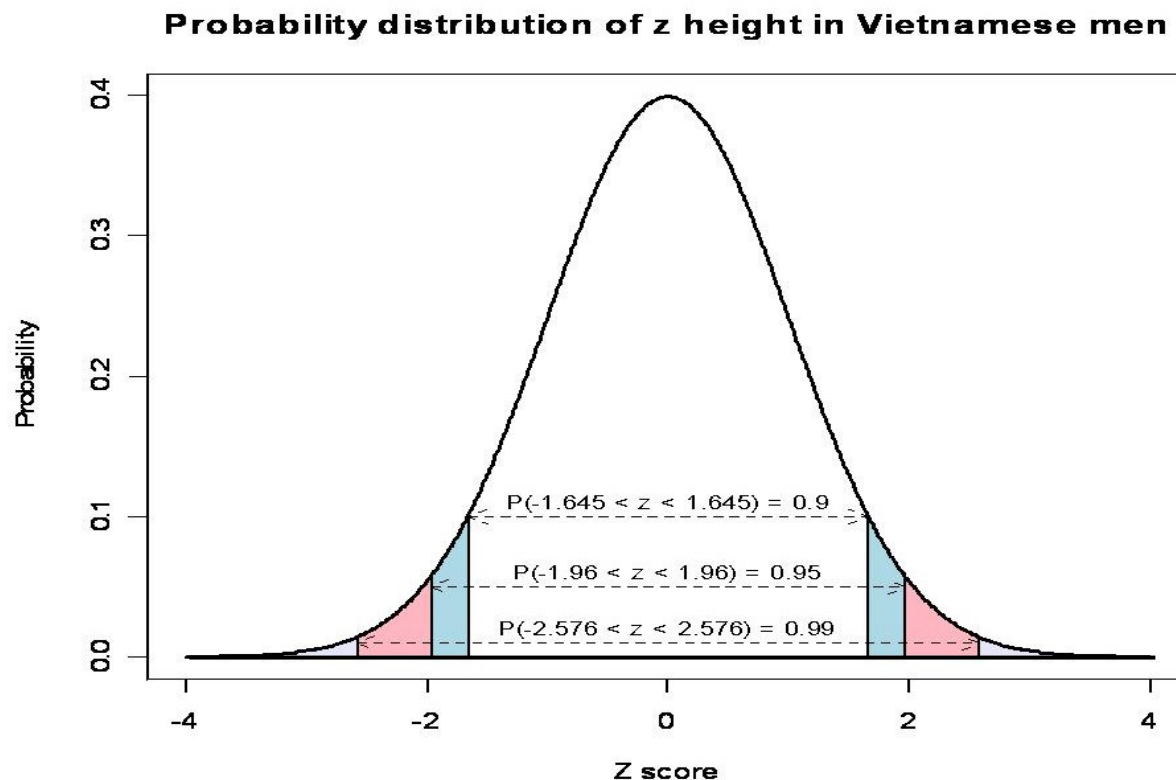


PHÂN PHỐI CHUẨN HÓA

Standardized Normal Distribution

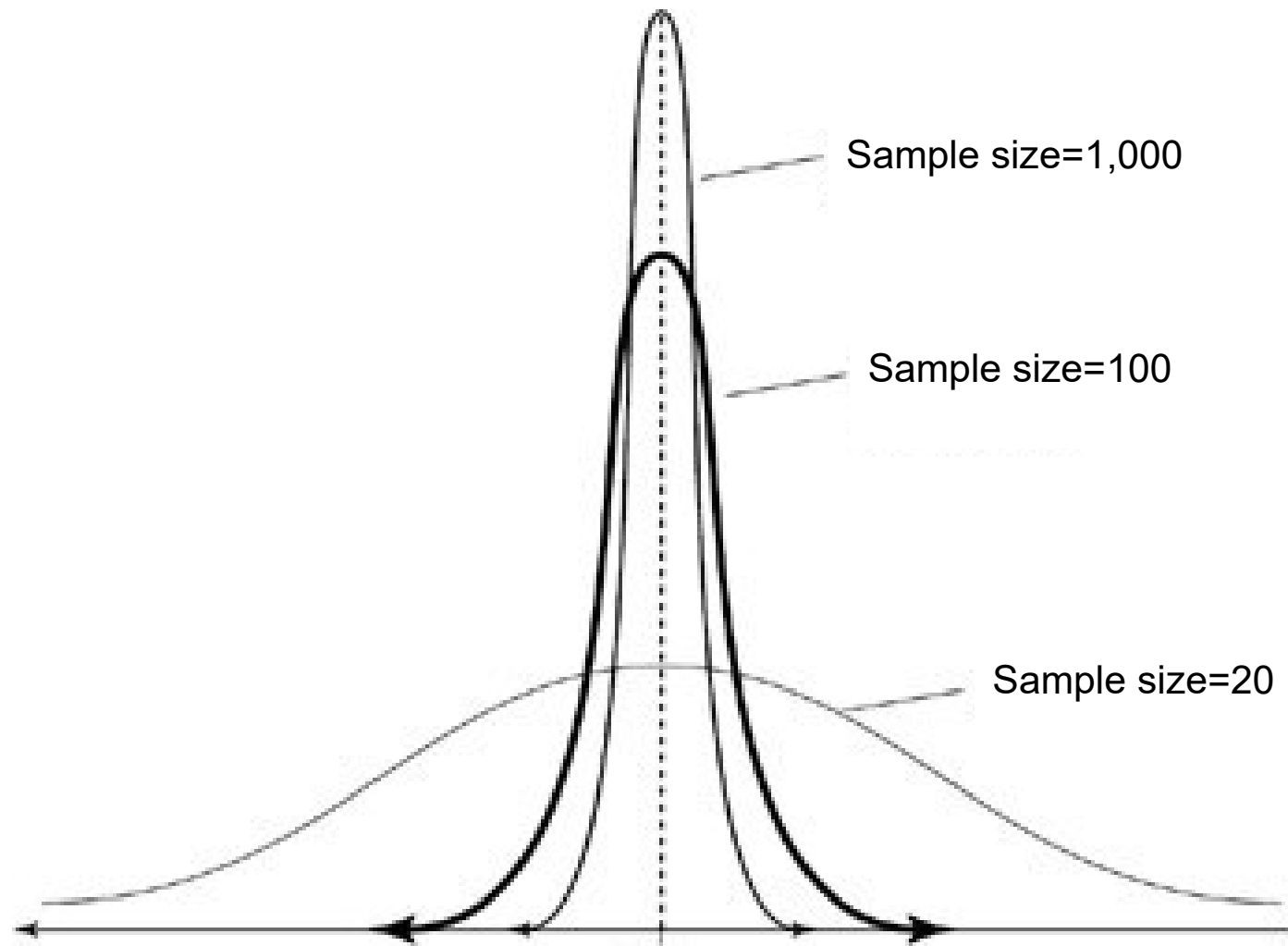
Quy tắc đã kiểm chứng (*Empirical rule*)

- ❖ Khoảng **90%** tổng số cá thể trong tổng thể có giá trị Z nằm trong khoảng **-1.65 đến 1.65**
- ❖ Khoảng **95%** tổng số cá thể có giá trị Z nằm trong khoảng **-1.96 đến 1.96**
- ❖ Khoảng **99.9%** tổng số cá thể có giá trị Z nằm trong khoảng **-3.291 đến 3.291**



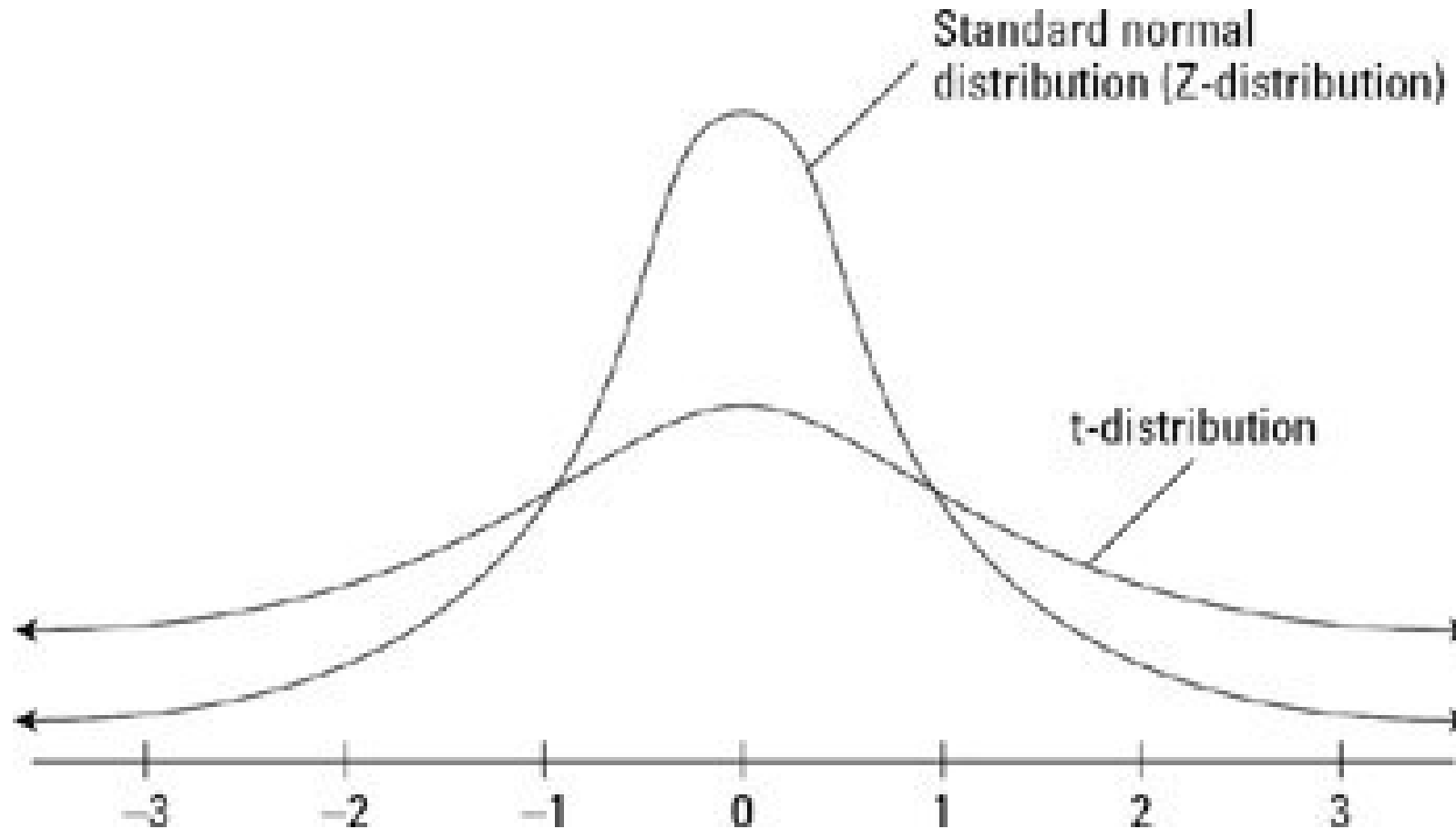
PHÂN PHỐI CHUẨN & PHÂN PHỐI Student t

Normal Distribution & t -Distribution



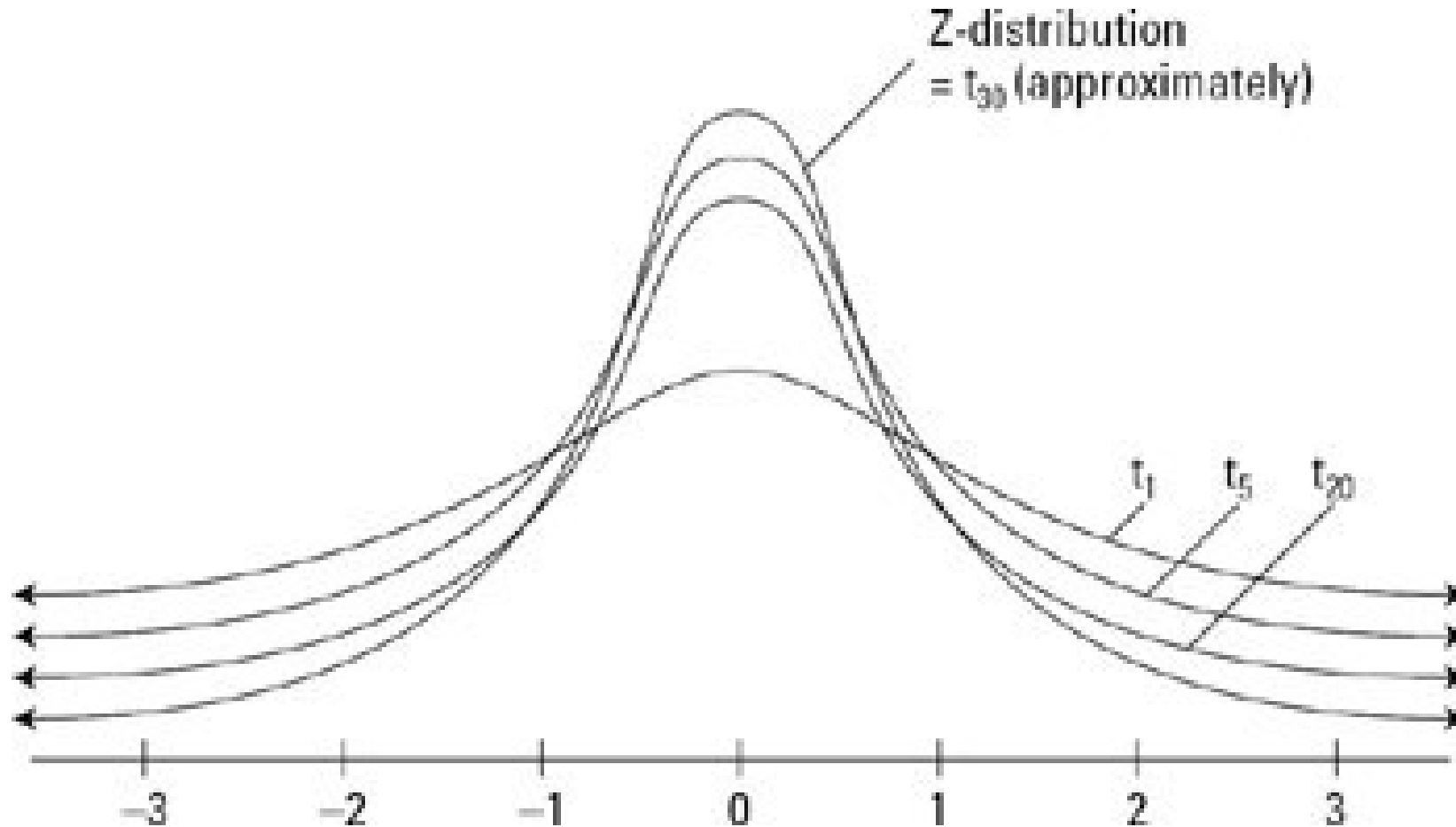
PHÂN PHỐI CHUẨN & PHÂN PHỐI Student t

Normal Distribution & t -Distribution



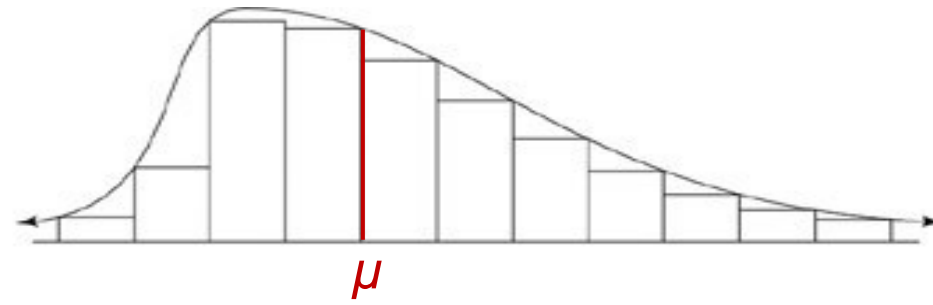
PHÂN PHỐI CHUẨN & PHÂN PHỐI Student t

Normal Distribution & t -Distribution

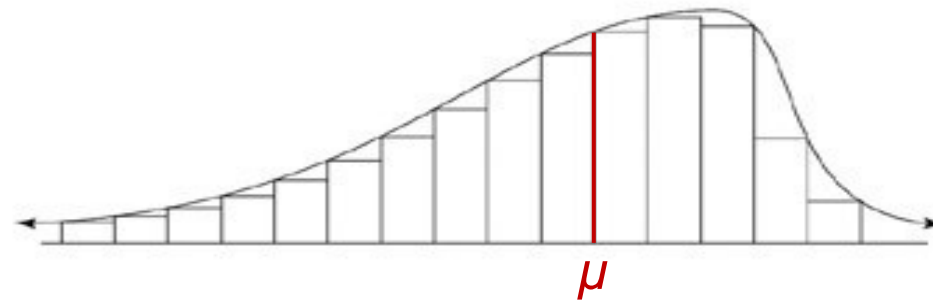


PHÂN PHỐI "LỆCH CHUẨN"

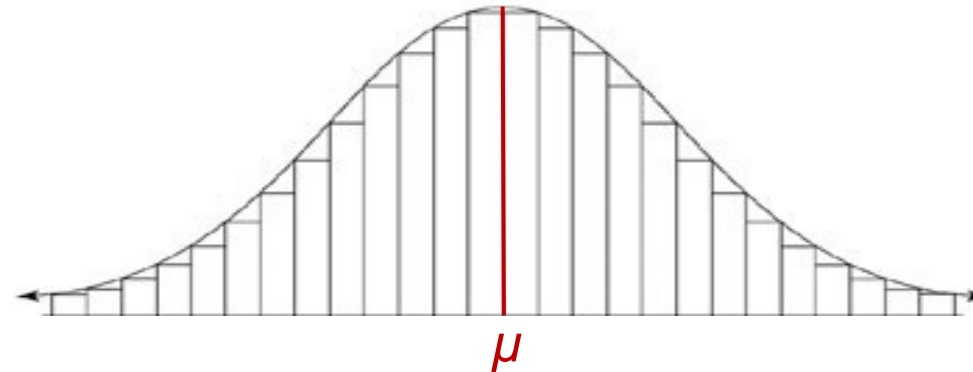
Nghiêng về bên phải
Skewed to the right



Nghiêng về bên trái
Skewed to the left



Đối xứng (phân phối chuẩn)
Symetric



ĐỘ LỆCH CHUẨN & SAI SỐ CHUẨN

Standard Deviation & Standard Error

- ❑ Độ lệch chuẩn của một tổng thể (σ) đo lường *mức độ khác biệt giữa các cá thể trong tổng thể đó.*
- ❑ Độ lệch chuẩn của một tập hợp mẫu (s) đo lường *mức độ khác biệt giữa các cá thể trong tập hợp mẫu đó.*
- ❑ Sai số chuẩn (*Standard Error* , viết tắt là *SE*) đo *lượng sự khác biệt giữa các kết quả đo lường được ước lượng từ N lần chọn mẫu khác nhau từ cùng 1 tổng thể.*
- ❑ Ví dụ:
 - Lần chọn mẫu thứ nhất cho kết quả đo lường là X_1 (X_1 có thể là thu nhập bình quân hay tỷ lệ người xem tivi bình quân)
 - Lần chọn mẫu thứ 2 cho kết quả đo lường là X_2
 - ...
 - Lần chọn mẫu thứ N cho kết quả đo lường là X_n
- ❖ Độ lệch chuẩn của dãy số $X_1, X_2, ..., X_n$ được gọi là Sai số chuẩn (*SE*)

ĐỘ LỆCH CHUẨN & SAI SỐ CHUẨN

Standard Deviation & Standard Error

❑ Sai số chuẩn có thể xác định được từ một tập hợp mẫu duy nhất:

❖ Áp dụng cho trường hợp tính giá trị trung bình (mean)

$$SE = \frac{\text{Độ lệch chuẩn của tập hợp mẫu}}{\sqrt{\text{Cơ số mẫu } (n)}} = \frac{s}{\sqrt{n}}$$

❖ Áp dụng cho trường hợp tính tỷ lệ (proportion)

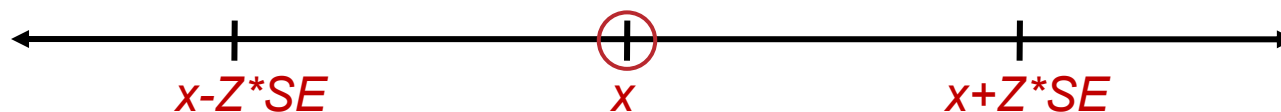
$$SE = \frac{\text{Độ lệch chuẩn của tập hợp mẫu}}{\sqrt{\text{Cơ số mẫu } (n)}} = \sqrt{\frac{p(1-p)}{n}}$$

❑ Nếu coi giá trị trung bình của nhiều lần khảo sát mẫu ($\bar{X}$) là ước lượng tốt nhất cho tham số của tổng thể (*thu nhập bình quân hay tỷ lệ xem tivi bình quân*), thì độ lệch bình quân của mỗi lần khảo sát mẫu so với $\bar{X}$ chính là sai số sinh ra do chúng ta sử dụng tập hợp mẫu để đo lường thay vì đo lường trên toàn bộ các cá thể của tổng thể (*census*). Sai số đó cũng chính là Sai số chuẩn (SE).

SAI SỐ CHUẨN & SAI SỐ CHỌN MẪU

Standard Error & Sampling Error or Margin of Error

- ❑ Nếu chúng ta chỉ chọn mẫu 1 lần duy nhất và kết quả đo lường cho 1 đại lượng nào đó (ví dụ mức thu nhập hay tỷ lệ xem tivi bình quân) là x , thì giá trị này trong thực tế (x) có thể nằm trong khoảng $x-Z*SE$ và $x+Z*SE$.



- ❑ $Z*SE$ chính là sai số xuất phát từ việc chúng ta *sử dụng 1 tập hợp mẫu* để ước lượng cho giá trị trong thực *tế thay vì sử dụng toàn bộ các cá thể trong tổng thể* mà chúng ta quan tâm nghiên cứu.
- ❑ $Z*SE$ được gọi là Sai số chọn mẫu (*Sampling error* hay *Margin of error*)

$$\text{Sampling Error} = Z^* \text{ Sai số chuẩn} = Z*SE$$

- ❑ Trị số Z có thể xác định tương ứng với xác suất để x nằm trong khoảng $x-Z*SE$ và $x+Z*SE$ (ví dụ 90%, 95%, 97% v.v.).
- ❑ Xác suất này cũng chính là mức tin cậy của kết quả khảo sát mẫu.

SAI SỐ CHUẨN & SAI SỐ CHỌN MẪU

Standard Error & Sampling Error or Margin of Error

- ❑ Mối quan hệ giữa Mức tin cậy (*Confidence level*) và trị số Z (*Z value*)

Confidence level	Z value
90%	1.65
95%	1.96
99%	2.58
99,9%	3.291

- ❑ Ở mức tin cậy = 95%, Z = 1.96
- ❑ Sai số chuẩn và Sai số chọn mẫu được xác định bằng công thức:

$$\text{Sai số chọn mẫu} = Z * SE = 1.96 * SE = 1.96 * \frac{s}{\sqrt{n}} = 1.96 * \sqrt{\frac{p(1-p)}{n}}$$

- ❑ Có thể nói rằng Z đại diện cho mức tin cậy của kết quả khảo sát mẫu

CÁC YẾU TỐ ẢNH HƯỞNG ĐẾN SAI SỐ CHỌN MẪU

❑ Sai số chọn mẫu (*Error*) = $Z * SE = Z * \frac{s}{\sqrt{n}} = Z * \sqrt{\frac{p(1-p)}{n}}$

❑ Các yếu tố ảnh hưởng:

❖ Mức tin cậy của kết quả khảo sát mẫu (Z)

- Khoảng tin cậy càng lớn thì Sai số mẫu càng lớn & ngược lại (?)

❖ Mức độ khác biệt (xét về đại lượng được quan tâm nghiên cứu) giữa các cá thể trong 1 tổng thể, hay nói cách khác là mức độ biến thiên (*variability*) của tổng thể được nghiên cứu, thể hiện qua giá trị của *Độ lệch chuẩn* (s hoặc $\sqrt{p(1-p)}$)

- Mức độ biến thiên càng lớn, Sai số mẫu càng lớn & ngược lại

❖ Kích thước của mẫu (cơ số mẫu) nghiên cứu (n)

- Cơ số mẫu càng lớn thì Sai số mẫu càng nhỏ & ngược lại

❑ Sai số mẫu (*Sampling error*) là lớn nhất khi $p = 1-p = 0.5$ hay 50% (Bạn có tin không?)

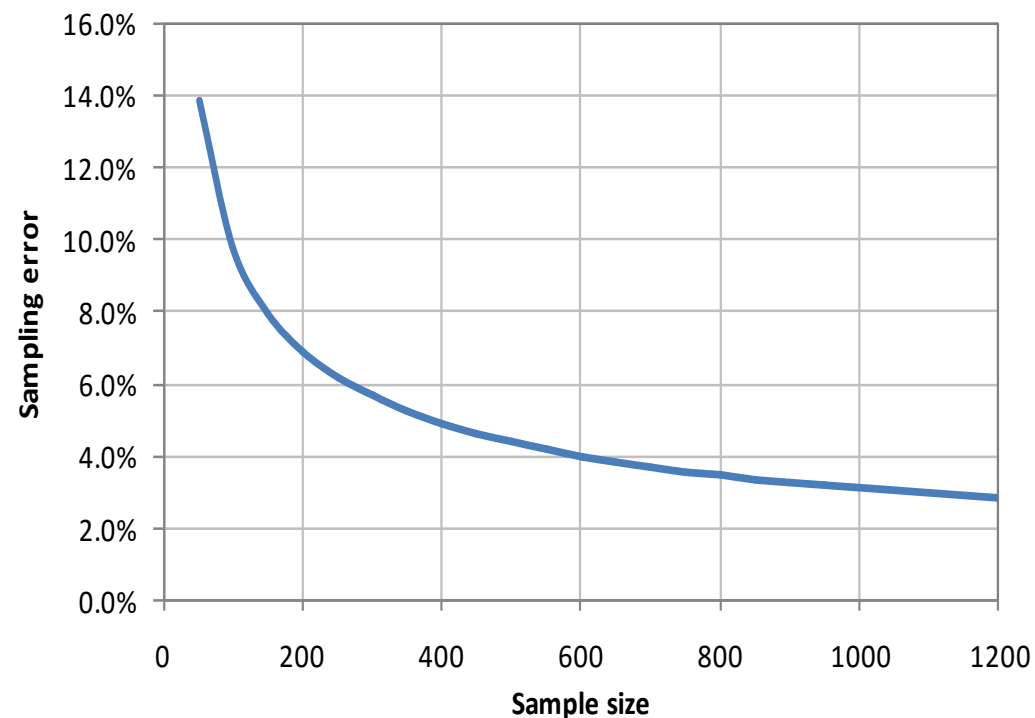
XÁC ĐỊNH SAI SỐ CHỌN MẪU TỐI ĐA KHI BIẾT n

- ❑ Sai số mẫu (Error) là lớn nhất khi $p = 1-p = 0.5$ hay 50%

$$Error = Z \sqrt{\frac{0.5 * 0.5}{n}} = Z \sqrt{\frac{0.25}{n}}$$

- ❑ Với mức tin cậy = 95%; $Z = 1.96$; $p = 0.5$

Sample size	Error	Error%
50	0.139	13.9%
100	0.098	9.8%
150	0.080	8.0%
200	0.069	6.9%
400	0.049	4.9%
500	0.044	4.4%
800	0.035	3.5%
1000	0.031	3.1%
5000	0.014	1.4%
10000	0.010	1.0%



XÁC ĐỊNH CƠ SỐ MẪU THEO YÊU CẦU

Required Sample Size

❑ Cơ số mẫu $n = Z^2 \frac{S^2}{Error^2} = Z^2 \frac{p(1-p)}{Error^2}$

- ❖ Khi biết s (hoặc p), n phụ thuộc vào mức tin cậy và sai số mẫu yêu cầu
- ❖ Khi không biết s :
 - Sử dụng giá trị s từ những nghiên cứu khác
 - Ước lượng giá trị s theo kinh nghiệm
- ❖ Khi không biết p :
 - Tính cơ số mẫu n cho trường hợp Sai số mẫu ($Error$) là tối đa (khi $p = 1-p = 0.5$ hay $p=50\%$)

$$n = Z^2 \frac{0.5 * 0.5}{Error^2} = Z^2 \frac{0.25}{Error^2}$$

XÁC ĐỊNH CƠ SỐ MẪU THEO YÊU CẦU (2)

Required Sample Size

- ❑ Tính cơ số mẫu với giả thiết Sai số mẫu lớn nhất không vượt quá ngưỡng giới hạn cho trước
- ❑ Với mức tin cậy = 95% và $p = 0.5$ (hay $p = 50\%$)

Error	Error%	Sample size
0.10	10.0%	96
0.09	9.0%	119
0.08	8.0%	150
0.07	7.0%	196
0.06	6.0%	267
0.05	5.0%	384
0.04	4.0%	600
0.03	3.0%	1067
0.02	2.0%	2401
0.01	1.0%	9604

CƠ SỞ MẪU & SAI SỐ CHỌN MẪU

