PHƯƠNG PHÁP NGHIÊN CỨU

Lý thuyết & thực tiễn

Trương Sĩ Ánh
CRO, Kantar Media Vietnam
Tháng 5/2018

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 (giới thiệu chung)

Bài 2

- Lý thuyết về mẫu nghiên cứu &
  Chọn mẫu nghiên cứu trong thực tiễn (phần 2)
- Thiết kế bảng câu hỏi khảo sát

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 (1)

- Một dạng của Mẫu phân tầng
  - Quota được xác lập theo tỷ lệ (*proportional quota sampling*).
  - Quota được xác lập không theo theo tỷ lệ (*non-proportional quota sampling*).
- Qui 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 (2)

- 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 hạnh 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

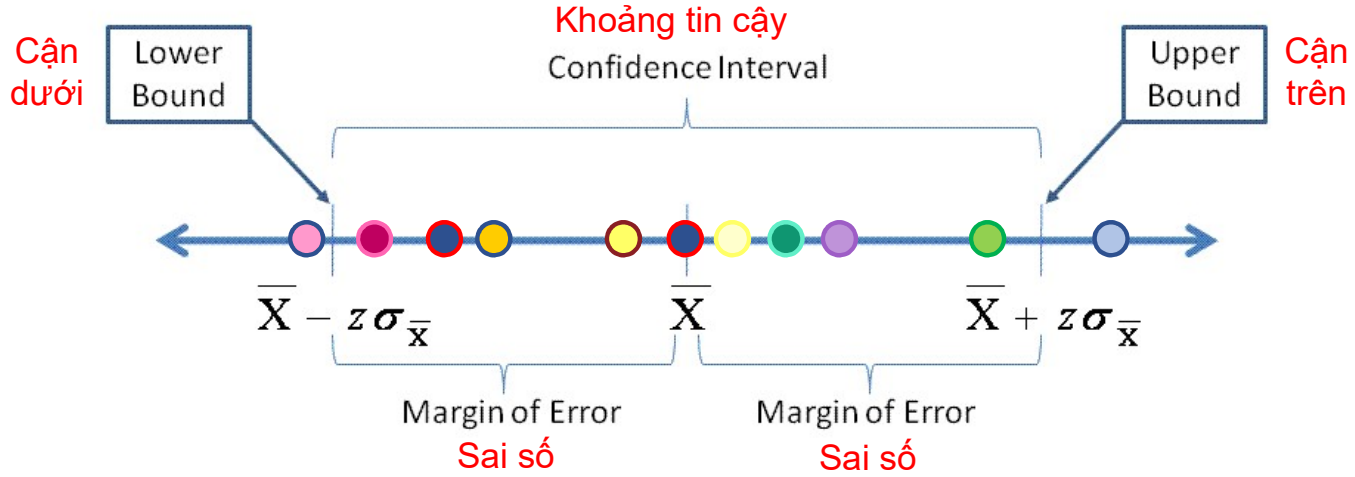
Non-probability samples

1. Mẫu hạn ngạch (3)

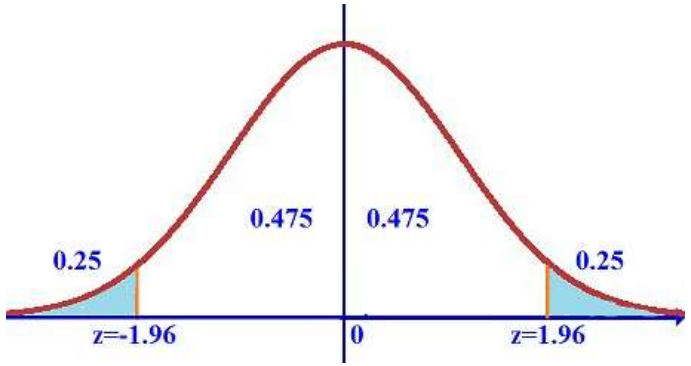
- Những lý lẽ ủng hộ
  - Chi phí thấp hơn. Chi phí cho 1 cuộc khảo sát mẫu quota có thể chỉ bằng 1/2 hay thậm chí 1/3 chi phí của 1 cuộc khảo sát ngẫu nhiên. Tuy nhiên cần lưu ý rằng việc tiết kiệm chi phí phải đánh đổi với sự suy giảm về mức độ chính xác và độ in cậy của kết quả khảo sát.
  - Quản lý & 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 tuổi trở lên (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 *+/- 5%*

MỘT 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 (1)

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

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

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

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

- **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 (tổng dân số)

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
Min	**0**	**-**	**-**
Max	**20**	**-**	**-**
Σ	**110**	**0**	**415**
Average	**5.5**	**0**	**20.75**
σ^2			**20.75**
σ			**4.555**

THỐNG KÊ MÔ TẢ

Descriptive Statistics (2)

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

$$\overline{X} = \frac{\sum_{1}^{N} X_i}{N} = \frac{110}{20} = 5.5$$

- **Phương sai của mẫu**
  Sample Variance

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

- **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 (tập hợp mẫu)

i	*Xi*	*Xi-μ*	*(Xi-μ)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
Min	**0**	**-**	**-**
Max	**20**	**-**	**-**
Σ	**110**	**0**	**415**
Average	**5.5**	**0**	**21.84**
S^2			**21.84**
S			**4.674**

PHÂN PHỐI CHUẨN
Normal Distribution

Hàm phân phối mật độ xác suất *f(x)*
Probability Density Function

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

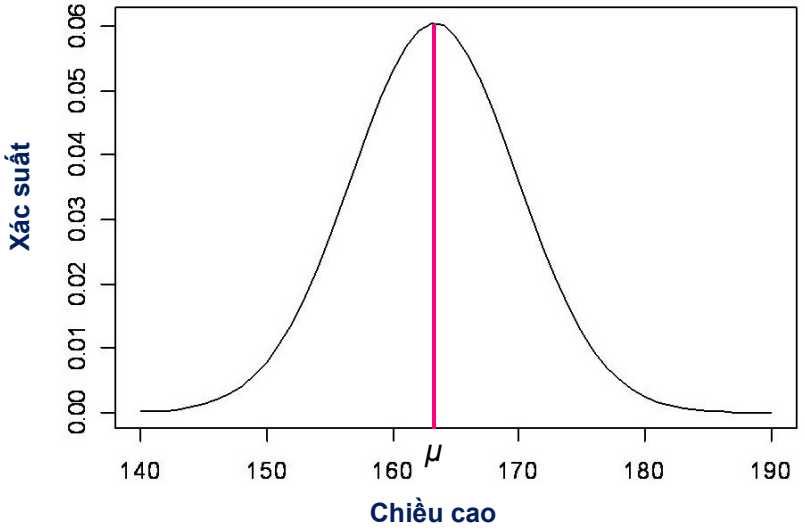
- Tính chất:
  - Đường cong hình chuông
  - Đối xứng quanh giá trị trung bình μ
  - Chỉ có 1 giá trị Mode = μ
- 2 tham số quyết định:
  1) Giá trị trung bình μ
  2) Độ lệch chuẩn σ
- Ví dụ: Có bao nhiêu % đàn ông VN cao 160 cm?

$$P(X = 160) = \frac{1}{6.6\sqrt{2x3.1416}} \exp\left[-\frac{(160-163.3)^2}{2 \times (6.6)^2}\right]$$

= 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%*

Xác suất phân bố đàn ông Việt Nam theo chiều cao



Trung bình $\mu = 163.3$ *cm*

Độ lệch chuẩn $\sigma = 6.6$ *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 $\frac{\sigma}{\sqrt{n}}$.

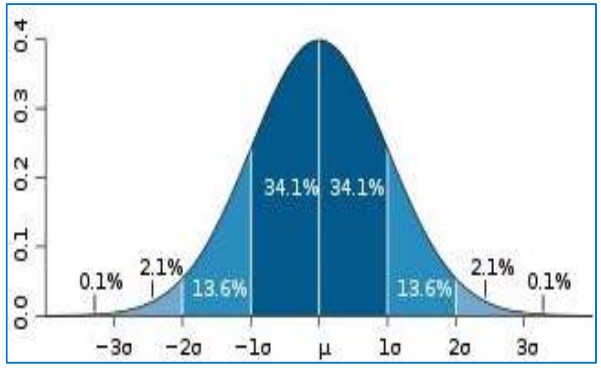
PHÂN PHỐI CHUẨN

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



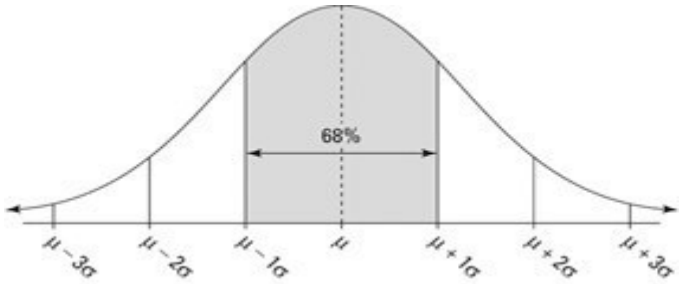
Qui tắc đã kiểm chứng (*Empirical rule*)

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

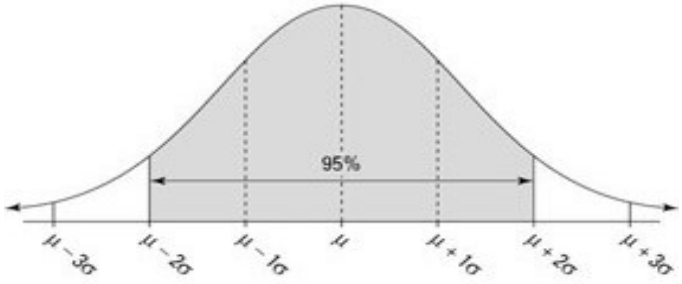
PHÂN PHỐI CHUẨN

Qui tắc đã kiểm chứng
Empirical rule

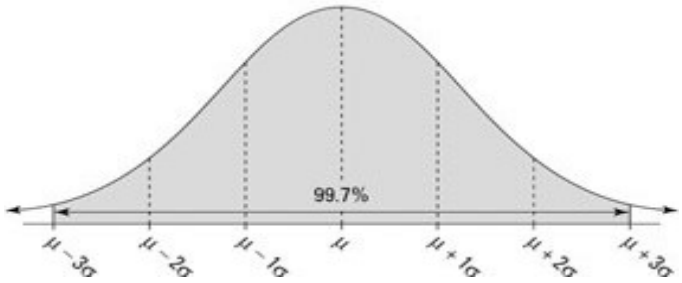
- Khoảng *68%* tổng số cá thể trong tổng thể có giá trị (*ví dụ mức thu nhập*) nằm trong khoảng *μ+/- 1σ*



- Khoảng *95%* tổng số cá thể trong tổng thể có giá trị (*ví dụ chỉ số IQ*) nằm trong khoảng *μ+/- 2σ*



- Khoảng *99.7%* tổng số cá thể trong tổng thể có giá trị (*ví dụ số năm đi học*) nằm trong khoảng *μ+/- 3σ*



PHÂN PHỐI CHUẨN CHUẨN HÓA

Standardized Normal Distribution

Hàm phân phối xác suất *f(z)*

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)$$

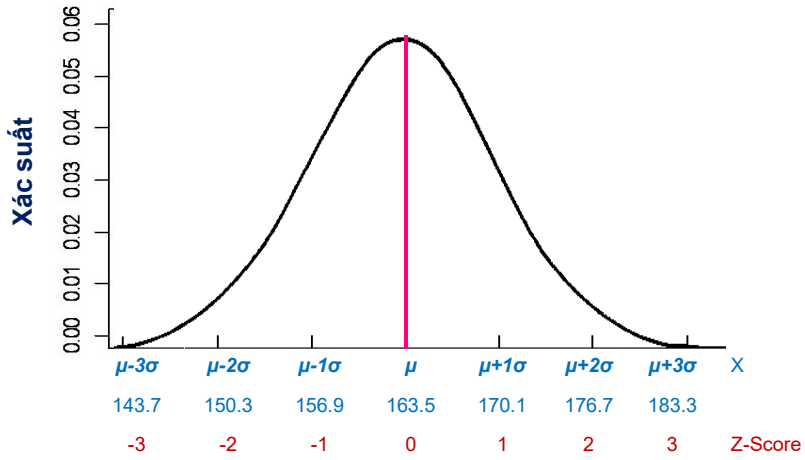
- Biến đổi giá trị X thành chỉ số Z (*Z-Score*)

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

- Chỉ số Z độc lập với đơn vị đo lường (*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

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

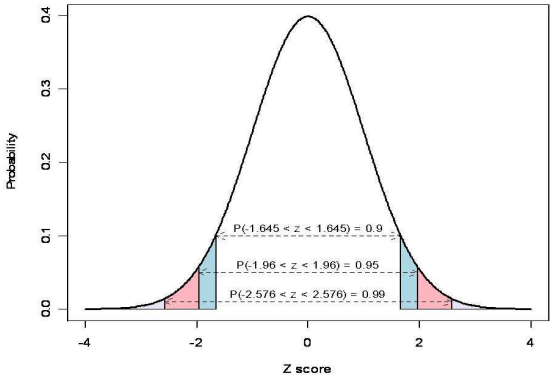
Standardized Normal Distribution

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

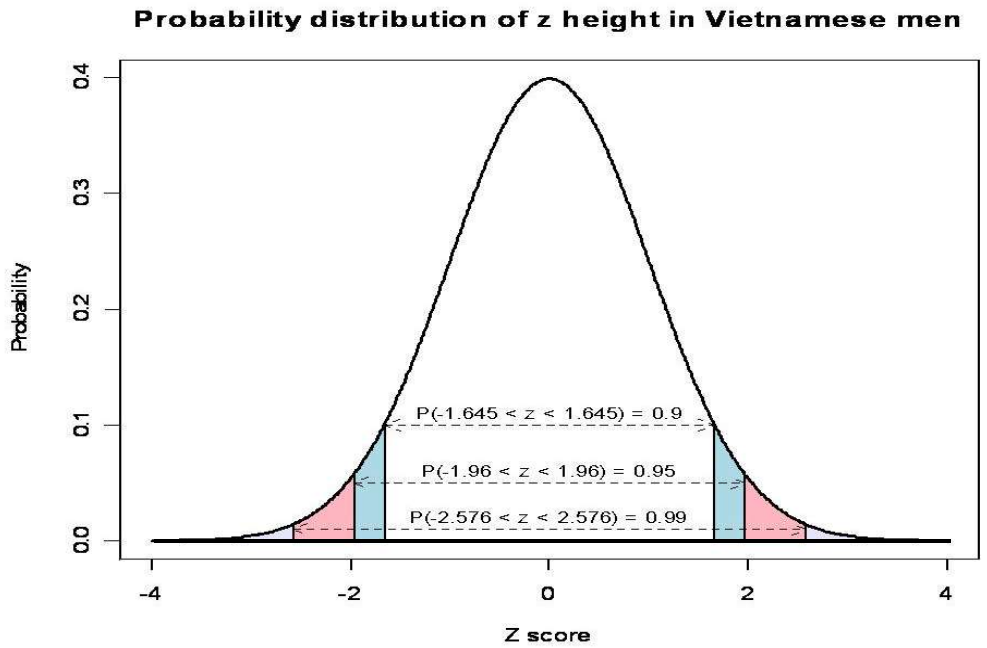
Z	.00	.01	.02	.03
0.0	.0000	.0040	.0080	.0120
0.1	.0398	.0438	.0478	.0517
0.2	.0793	.0832	.0871	.0910
0.3	.1179	.1217	.1255	.1293
0.4	.1554	.1591	.1628	.1664
0.5	.1915	.1950	.1985	.2019
0.6	.2257	.2291	.2324	.2357
0.7	.2580	.2612	.2642	.2673
0.8	.2881	.2910	.2939	.2967
0.9	.3159	.3186	.3212	.3238
1.0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



PHÂN PHỐI CHUẨN HÓA
Standardized Normal Distribution

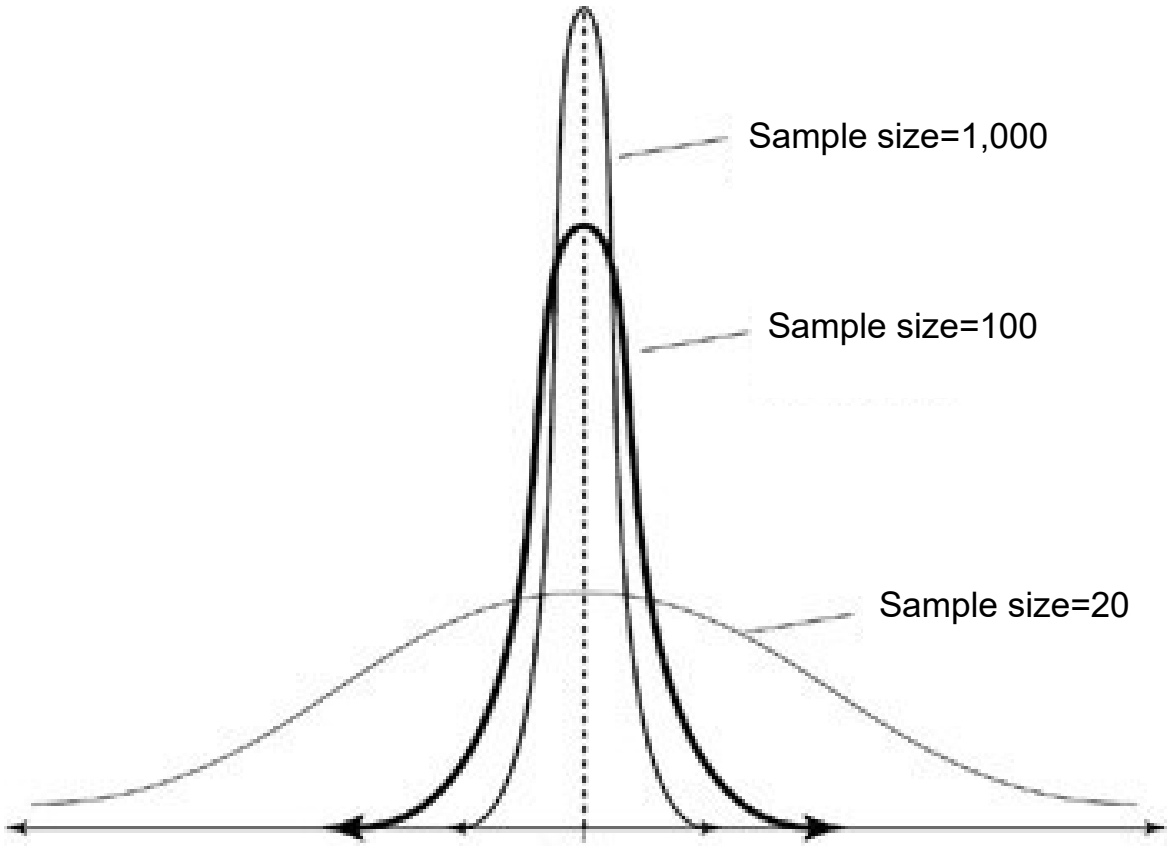
Qui tắc đã kiểm chứng (*Empirical rule*)

- 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*



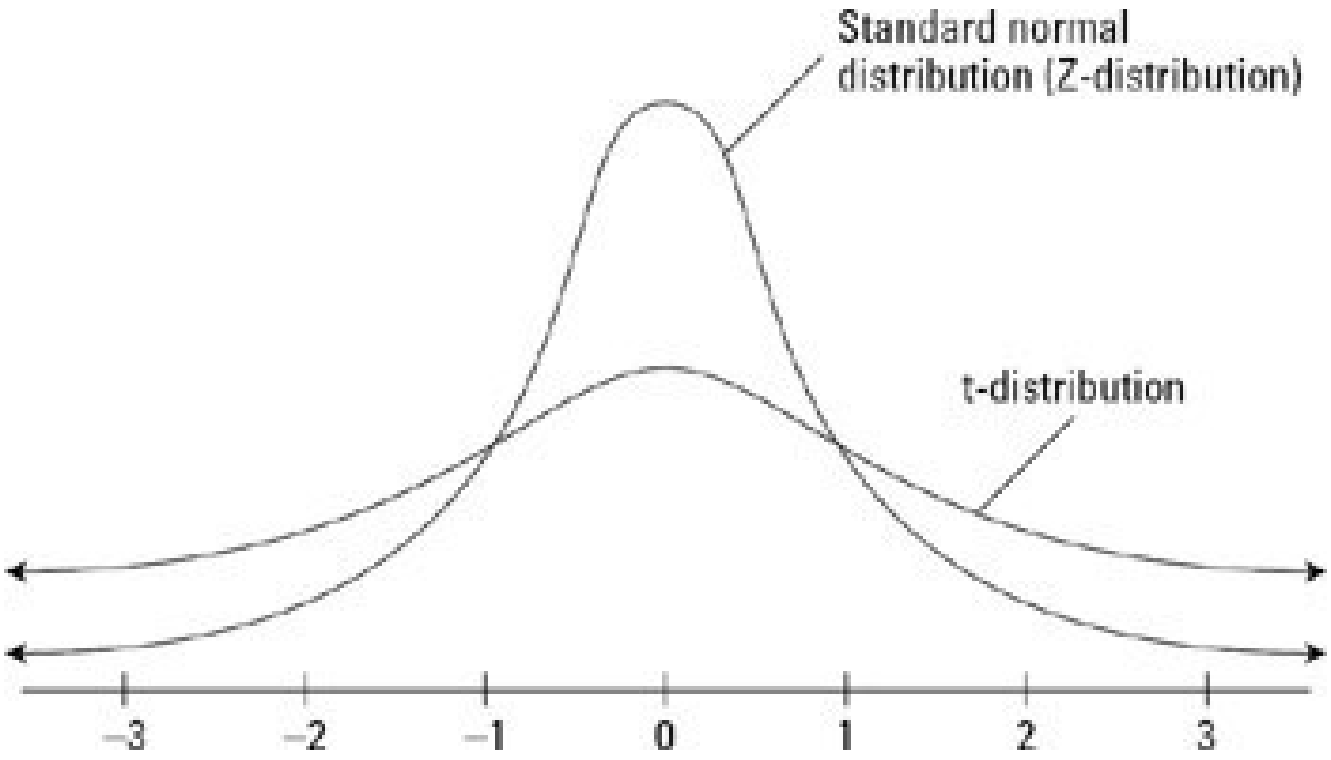
PHÂN PHỐI CHUẨN & PHÂN PHỐI Student *t*

Normal Distribution & t-Distribution



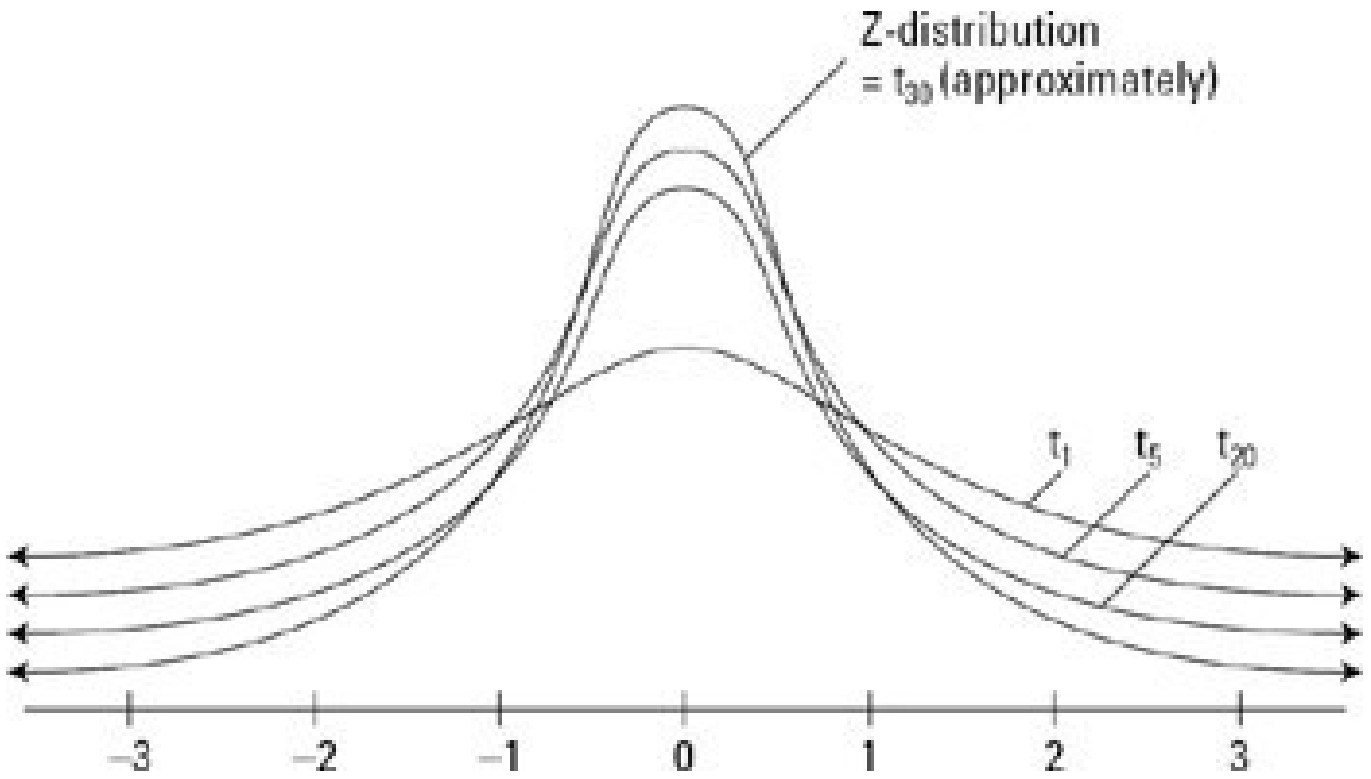
PHÂN PHỐI CHUẨN & PHÂN PHỐI Student *t*

Normal Distribution & t-Distribution



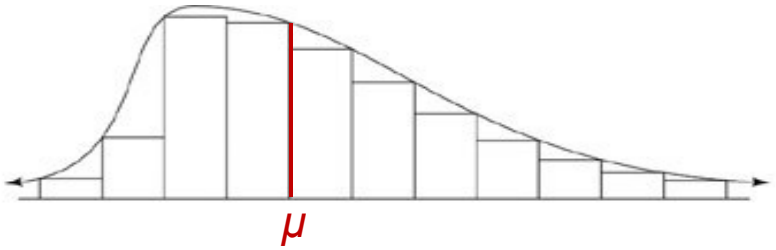
PHÂN PHỐI CHUẨN & PHÂN PHỐI Student *t*

Normal Distribution & t-Distribution

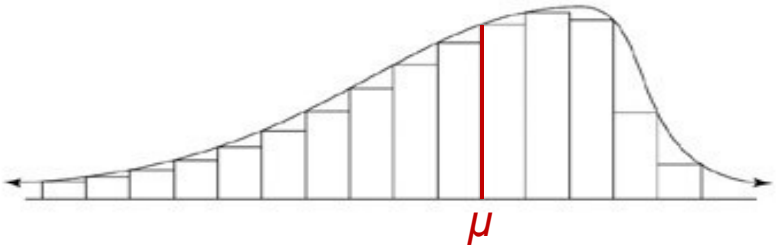


PHÂN PHỐI "LỆCH CHUẨN"

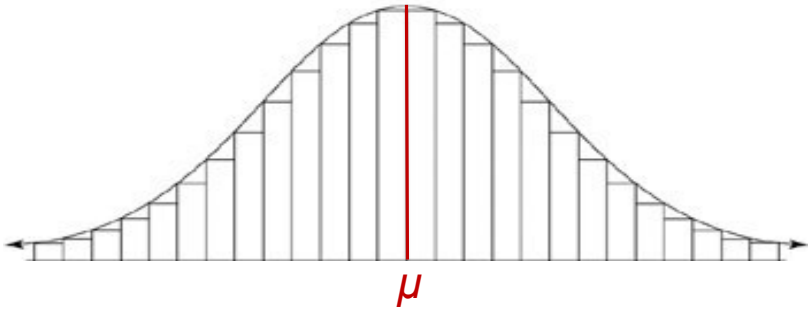
Nghiêng về bên phải
Skewed to the right



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 *lường 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)

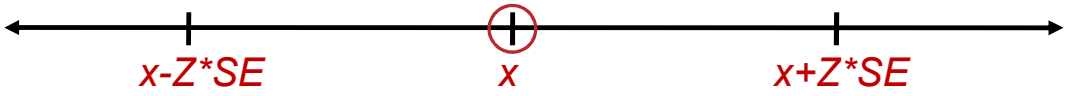
$$SE = \frac{\text{Độ lệch chuẩn của tập hợp mẫu}}{\sqrt{\text{Cơ số mẫu } (n)}} = \sqrt{\frac{p(1-p)}{n}}$$

- 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 tể 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*)

Sampling Error = Z* 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:

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

- 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 MẪU *(Sampling error)*

- Sai số 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)
    - Bạn nghĩ sao về nhận định "*Khoảng tin cậy càng lớn thì sai số mẫu càng lớn & ngược lại*" ?
  - Mức độ khác biệt giữa các cá thể trong 1 tổng thể (*xét về đại lượng được quan tâm nghiên cứu*), 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)*
    - Bạn nghĩ sao về nhận định "*Mức độ biến thiên càng lớn, sai số mẫu có xu hướng 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 là lớn nhất khi *p = 1-p = 0.5* hay *50%* (*Bạn có tin không?*)

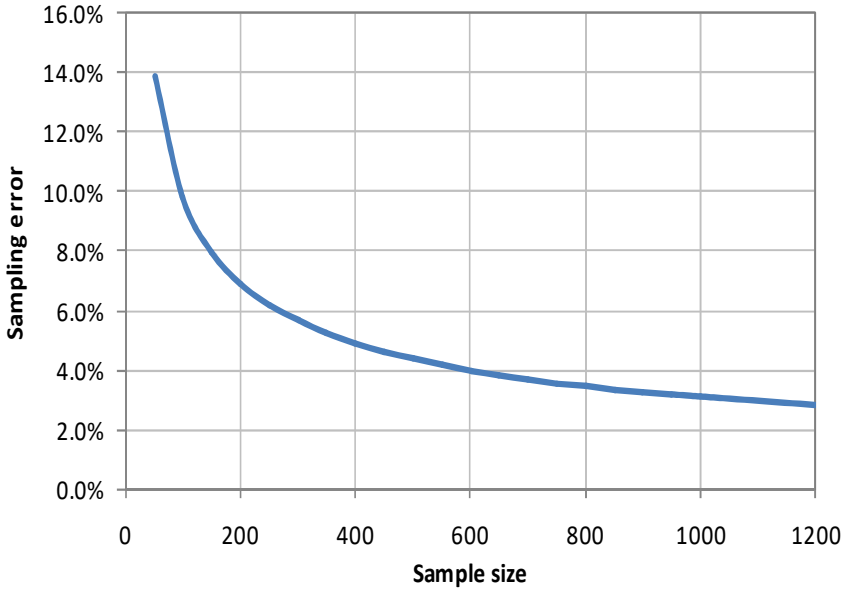
XÁC ĐỊNH SAI SỐ MẪU TỐI ĐA KHI BIẾT *n*

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

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

- 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 CỠ 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 $p = 1\text{-}p = 0.5$ 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)

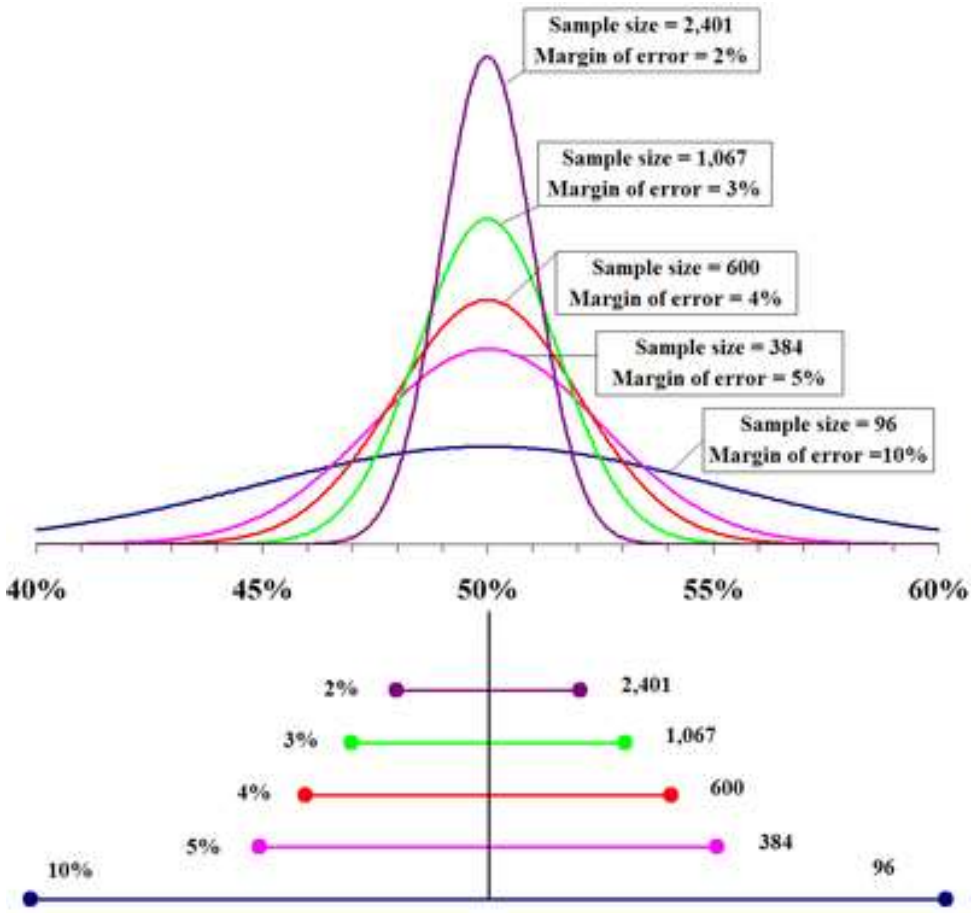
Required Sample Size

- 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%)*

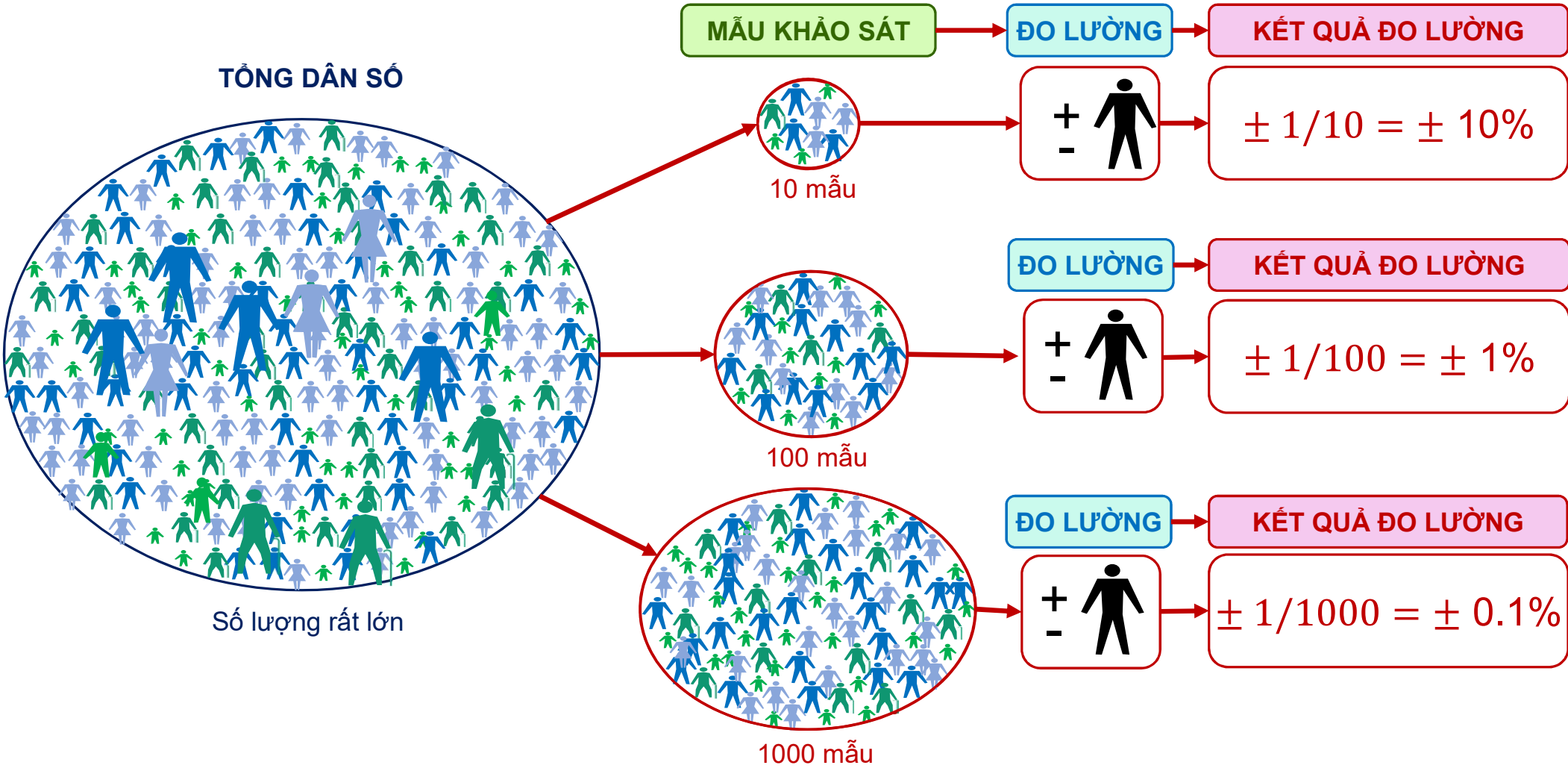
Error	Error%	Sample size
0.10	10.0%	96
0.09	9.0%	119
0.08	8.0%	150
0.07	7.0%	196
0.06	6.0%	267
0.05	5.0%	384
0.04	4.0%	600
0.03	3.0%	1067
0.02	2.0%	2401
0.01	1.0%	9604

CƠ SỐ MẪU & SAI SỐ CHỌN MẪU

Sampling error = Margin of error



Tổng thể - Mẫu - Kết quả nghiên cứu/đo lường



Bài 2

- Lý thuyết về mẫu nghiên cứu &
  Chọn mẫu nghiên cứu trong thực tiễn (phần 2)
- **Thiết kế bảng câu hỏi khảo sát**

Thiết kế Bảng câu hỏi khảo sát

Questionnaire Design

- Nguyên tắc thiết kế bảng câu hỏi (questionnaire)
- Những câu hỏi bạn cần đặt ra trước khi thiết kế câu hỏi
- Các dạng câu hỏi phổ biến
- Trình bày Bảng câu hỏi
- Thử nghiệm & đánh giá các câu hỏi khảo sát

Bảng câu hỏi là gì?

- Là công cụ để:
  - *Ghi nhận các câu trả lời cho những câu hỏi được đặt ra*
  - *Thu thập các dữ kiện thực tế (tập hợp thông tin hoặc các đại lượng đo lường)*
- Là một tập hợp các câu hỏi được sắp xếp theo 1 trình tự hợp lý
- BCH là 1 công cụ nghiên cứu bao gồm các câu hỏi và gợi ý hoặc hướng dẫn nhằm mục đích thu thập thông tin từ các đối tượng nghiên cứu (*đáp viên*)

Tại sao lại dùng Bảng câu hỏi

- Một BCH được thiết kế tốt sẽ:
  - *cung cấp thông tin chính xác và phù hợp cho những câu hỏi trong nghiên cứu*
  - *giảm thiểu tối đa các yếu tố làm sai lệch kết quả nghiên cứu*
  - *có cơ hội thành công cao hơn*
- Các câu hỏi càng đơn giản và càng tập trung vào chủ đề chính thì càng tốt.

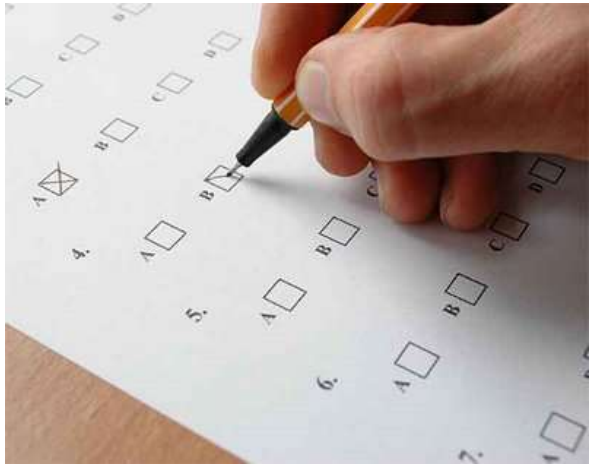
Ưu điểm & nhược điểm của BCH

- **Ưu điểm:**
  - *Có thể thu thập thông tin từ 1 số lượng lớn các đối tượng cần nghiên cứu một cách khá dễ dàng và có hiệu quả (về kinh tế, về thời gian, nhân lực v.v.)*
  - *Có thể cung cấp thông tin lượng hóa được (quantifiable)*
  - *Phân tích kết quả tương đối dễ dàng*
- **Nhược điểm:**
  - Chỉ có thể cung cấp:
    - *Các câu hỏi có số lượng phương án trả lời hạn chế (nên ít hơn 7)*
    - *Các câu hỏi có thể không phù hợp với thực tiễn cần nghiên cứu*
  - Các câu hỏi có thể bị hiểu nhầm hoặc diễn đạt sai (*có liên quan nhiều đến khâu thiết kế BCH*)
  - Cần phải có các câu hỏi đúng ngay từ đầu
    - *Một khi đã mất thông tin thì rất khó xoay xở để phân tích*

Các loại Bảng câu hỏi

- BCH phỏng vấn – được thiết kế dành cho phỏng vấn viên (interviewer)
  - *Phỏng vấn trực diện (face-to-face interview)*
  - *Phỏng vấn qua điện thoại*

- BCH tự điền - thiết kế dành cho đối tượng khảo sát (*đáp viên - respondent*) tự điền câu trả lời
  - *Phân phát trực tiếp*
  - *Gửi qua bưu điện*
  - *Gửi qua email hoặc điền trực tuyến qua Internet (online survey)*



Các loại Bảng câu hỏi

Bảng câu hỏi phỏng vấn

- **Ưu điểm:**
  - Có thể thu thập thông tin từ những người không biết đọc/viết
  - Phỏng vấn viên có thể giải thích trong trường hợp cần thiết
  - Phỏng vấn viên có thể sử dụng các công cụ hỗ trợ (*ví dụ showcard*) để giúp người được phỏng vấn nắm rõ vấn đề hơn.
  - Các câu hỏi có thể được trả lời nhanh hơn

- **Nhược điểm:**
  - Kết quả phỏng vấn phụ thuộc nhiều vào chất lượng của phỏng vấn viên, thậm chí có thể bị thiên lệch do việc giải thích/diễn gải câu hỏi không đúng hoặc không khách quan.
  - Chi phí đầu tư tốn kém hơn
  - Thường chỉ sử dụng những câu hỏi ngắn gọn (nhất là phỏng vấn qua điện thoại)
  - Khó có thể áp dụng để nghiên cứu những vấn đề mang tính nhạy cảm

Các loại Bảng câu hỏi

Bảng câu hỏi tự điền

- **Ưu điểm:**
  - Chi phí thấp và dễ triển khai
  - Đảm bảo tính riêng tư/tính bảo mật
  - Có thể điền câu trả lời khi thấy thuận tiện hay rảnh rỗi
  - Không bị tác động bởi người phỏng vấn
- **Nhược điểm:**
  - Tỷ lệ phản hồi thấp
  - Các câu hỏi có thể bị hiểu nhầm (*nếu không kèm theo hướng dẫn đầy đủ*)
  - Không có sự hướng dẫn/kiểm soát của phỏng vấn viên
  - Khó xác định được người điền BCH có phải là đối tượng cần nghiên cứu (*nhất là trong trường hợp cần tôn trọng quyền riêng tư*)
  - Thời gian khảo sát có thể kéo dài
  - Có thể gặp khó khăn trong việc suy rộng kết quả khảo sát

Các bước thiết kế bảng câu hỏi (1)

Lập kế hoạch nghiên cứu

- Xác lập mục tiêu nghiên cứu
  - *Tìm hiểu nguyên nhân vì sao tầng lớp trẻ và những người làm văn phòng ít đọc báo giấy hơn những đối tượng khác*
- Tìm hiểu về chủ đề (*có ý định/kế hoạch*) nghiên cứu
  - *Tổng quan tài liệu, tham vấn chuyên gia*
- Thiết lập giả thiết nghiên cứu ban đầu
  - *Tầng lớp trẻ và nhân viên văn phòng sử dụng Internet nhiều hơn những đối tượng khác*
- Xác định thông tin cần có để đáp ứng mục tiêu nghiên cứu
  - *Giới tính và độ tuổi; nghề nghiệp; mức độ và thói quen đọc báo giấy; mức độ và mục đích sử dụng Internet.*

Các bước thiết kế Bảng câu hỏi (2)

Xác định đối tượng nghiên cứu

- Hiểu biết về đối tượng khảo sát (đáp viên)
  - *Giới tính*
  - *Độ tuổi*
  - *Nghề nghiệp*
  - *Trình độ học vấn*
  - *Sắc tộc/ngôn ngữ*
  - *Những đặc điểm khác (nhất là đặc điểm mang tính nhạy cảm)*

Bảng câu hỏi phải được thiết kế sao cho *phù hợp với đối tượng khảo sát*

chứ không phải là *chọn đối tượng nào phù hợp với BCH để khảo sát*

Các bước thiết kế bảng câu hỏi (3)

Thiết kế câu hỏi

- Nội dung câu hỏi
  - *Mục tiêu nghiên cứu*
  - *Khái niệm hóa/cụ thể hóa vấn đề nghiên cứu. Ví dụ thông tin nào có thể dùng làm đại diện để đánh giá mức sống, thái độ, quan điểm, thói quen hành động v.v..*
- Dạng câu hỏi
- Hình thức trình bày
- Sử dụng mã số đại diện (*codes*) trong các câu hỏi *(trong hoàn cảnh thích hợp)*

Luôn làm thí điểm và điều chỉnh BCH trước khi triển khai chính thức

Thế nào là một BCH được thiết kế tốt

- Hình thức trình bày đẹp
- Ngắn gọn & đơn giản
- Phù hợp và có lôgic (*hợp lý*)
  - *Skip pattern*
  - *Double check*

❖ Tỷ lệ câu hỏi được trả lời cao

❖ Dễ dàng tổng hợp và phân tích dữ liệu

Những thông tin cần có trên BCH

- **Trang đầu tiên**
  - Tiêu đề của dự án nghiên cứu
  - Mã số nhận dạng BCH
  - Thông tin về địa bàn khảo sát
  - Thông tin về PVV
- **Các trang tiếp theo**
  - Số thứ tự của các câu hỏi
  - Số trang
  - Các chỉ dẫn cần thiết (nếu có) & cần in đậm

Trật tự câu hỏi trong BCH (1)

- **Xác định trật tự của các câu hỏi/hạng mục trong BCH**
  - Dễ → Khó
  - Khái quát → Cụ thể
  - Thực tế → Trừu tượng
  - Ít nhạy cảm → Nhạy cảm
- **Cân nhắc vị trí đặt các câu hỏi mang tính chất nhạy cảm**
  - Cần ý thức rõ tầm quan trọng & tác động của trật tự các câu hỏi trong BCH

Trật tự câu hỏi trong BCH (2)

- Phân nhóm các câu hỏi theo chủ đề/phương án chọn câu trả lời
- Các câu hỏi *mở đầu* nên:
  - đơn giản
  - có dạng câu hỏi đóng (*có sẵn các phương án trả lời để chọn*)
  - phù hợp với chủ đề chính
  - mang tính ôn hòa (*đáp viên không cảm thấy khó xử hay mất lòng*)
  - không hỏi thông tin mang tính cá nhân hay đặc điểm nhân khẩu học

Không nên để các câu hỏi quan trọng ở cuối cùng

Phần giới thiệu trong BCH

- Bạn (*phỏng vấn viên*) là ai/làm việc cho ai
- Mục đích của cuộc phỏng vấn và/hoặc mục đích nghiên cứu
- Lý do đáp viên được chọn để phỏng vấn (*chọn ngẫu nhiên, chọn từ danh bạ điện thoại v.v.*)
- Có thể liên lạc với bạn bằng cách nào & ở đâu
- Cam kết bảo mật thông tin
- Độ dài của cuộc phỏng vấn

Mọi đáp viên cần được hiểu rõ về *mục đích/ý nghĩa của dự án nghiên cứu*

Nội dung của các câu hỏi

- Tập trung vào vấn đề nghiên cứu
- Tránh đi lạc đề
- Tránh thu thập thông tin không cần thiết
- Thông tin nhân khẩu học/tâm lý (*demographics/psychographics*)
- Thông tin liên hệ *(nếu không phải là cuộc khảo sát ẩn danh [non-anonymized])*

Hình thức đặt câu hỏi (1)

- Điều chỉnh cho phù hợp với đối tượng khảo sát
  - *Những người có trình độ học vấn môn cao* ⟷ *đại chúng*
  - *Tầng lớp ưu tú* ⟷ *bình dân*
- Câu văn đơn giản, ngắn gọn & dễ hiểu
- Định nghĩa cụ thể cho những khái niệm mang tính tương đối
  - *Ví dụ: Thường xuyên/thỉnh thoảng; nhiều/ít*
- Luôn có phương án trả lời "*Không biết/không nhớ*" hoặc "*Khác*"

Hình thức đặt câu hỏi (2)

- Mỗi câu hỏi chỉ hỏi về 1 thông tin, không hỏi nước đôi
  - *Bạn có đọc báo hoặc tạp chí không?*

Có ☐
Không ☐

- Các phương án trả lời phải đảm bảo tính chất:
  - *Không trùng lắp:* ☐ *0-5 triệu* ☐ *5-10 triệu* ☐ *Trên 10 triệu*
  - *Không bỏ sót:* ☐ *1-2 lần/ngày* ☐ *5-10 lần ngày* ☐ *Trên 10 lần/ngày*

- Các phương án trả lời nên xếp theo chiều dọc thay vì chiều ngang

Có ☐
Không ☐

☐ *Có* ☐ *Không*

Hình thức đặt câu hỏi (3)

- Cần đặt câu hỏi càng rõ ràng, càng cụ thể càng tốt:
  - *Bạn có thường xuyên xem tivi không?*

Có ☐
Không ☐

  - *Bạn xem tivi thường xuyên như thế nào?*

Hàng ngày ☐
5-6 lần/tuần ☐
3-4 lần tuần ☐
1-2 lần/tuần ☐
Ít hơn 1 lần/tuần ☐
Không xem tivi ☐

Hình thức đặt câu hỏi (4)

2 dạng câu hỏi chính:

- Câu hỏi đóng (*close-ended question*)

Bạn có xem các trang quảng cáo trên báo không?

Có ☐
Không ☐

Bạn có sử dụng Internet không?

Có, dùng hàng ngày ☐
Có, nhưng không thường xuyên ☐
Không sử dụng ☐

- Câu hỏi mở (*open-ended question*)

Bạn có thể cho biết lý do không sử dụng Internet?

_____ ☐
_____ ☐
_____ ☐

Sử dụng câu hỏi đóng hay mở

Câu hỏi đóng

- **Ưu điểm**
  - *Đơn giản và nhanh chóng*
  - *Giảm bớt sự phụ thuộc vào khả năng đọc/viết*
  - *Dễ mã hóa, ghi nhận*
  - *Dễ so sánh câu trả lời giữa các đối tượng khảo sát với nhau*
  - *Dễ tổng hợp, phân tích & báo cáo kết quả*
- **Nhược điểm**
  - *Hạn chế phương án trả lời*
  - *Mất thông tin*

Nên thêm phương án "*Khác (ghi rõ) ……...*" (*kết hợp với câu hỏi mở*)

Sử dụng câu hỏi đóng hay mở (2)

Câu hỏi mở

- **Ưu điểm**
  - Không mang tính cưỡng ép trả lời
  - Cho phép tìm hiểu/khám phá vấn đề để xây dựng giả thuyết
    - *Nghiên cứu định tính; phỏng vấn nhóm,*
  - Sử dụng khi không lường trước được hết các phương án trả lời
  - Thích hợp để nghiên cứu/tìm hiểu kiến thức, thái độ, quan điểm/nhận thức v.v.
  - Có thể có những câu trả lời chi tiết và bất ngờ
- **Nhược điểm**
  - Có thể bị ảnh hưởng bởi phỏng vấn viên
    - *Khả năng nhận thức/hiểu câu trả lời & ghi chép câu trả lời*
  - Tốn kém thời gian
  - Khó khăn khi mã hóa, phân loại & tổng hợp thông tin
  - Khó phân tích
  - Khó so sánh các nhóm đối tượng khảo sát

Các dạng câu hỏi đóng (1)

- Câu hỏi đóng có câu trả lời đơn giản (*1 phương án trả lời*)
  - *Bạn sinh năm nào?*

Năm sinh | | | | |

  - *Bạn đi xem phim ở rạp lần gần đây nhất là khi nào (chọn 1 đáp án duy nhất)?*

Hôm qua	1
Trong 7 ngày qua	2
Trong 2 tuần qua	3
Trong 30 ngày qua	4
Lâu hơn 1 tháng qua	5
Chưa bao giờ	6

1 *Hôm qua*
2 *Trong 7 ngày qua*
3 *Trong 2 tuần qua*
4 *Trong 30 ngày qua*
5 *Lâu hơn 1 tháng qua*
6 *Chưa bao giờ*

1	*Hôm qua*
2	*Trong 7 ngày qua*
3	*Trong 2 tuần qua*
4	*Trong 30 ngày qua*
5	*Lâu hơn 1 tháng qua*
6	*Chưa bao giờ*

Các dạng câu hỏi đóng (2)

- Câu hỏi đóng có hơn 1 phương án trả lời
  - Trong vòng 7 ngày qua bạn đã đọc những tờ báo nào sau đây?

Báo Thanh Niên ☐
Báo Tuổi Trẻ ☐
Báo CA TP. HCM ☐
Báo Bóng Đá ☐

  - Bạn thường truy cập Internet bằng phương tiện nào sau đây?

Máy tính để bàn ☐
Máy tính xách tay ☐
Điện thoại DĐ ☐
Phương tiện khác (ghi rõ) ☐ _____

Các dạng câu hỏi đóng (3)

- Câu hỏi đóng có câu trả lời dạng thang đo (thang đánh giá)
  - *Xin cho biết mức độ yêu thích của bạn đối với các phương tiện truyền thông sau đây là như thế nào?*

	Rất thích	*Thích*	*Bình thường*	*Không thích*	*Rất không thích*
	1	2	3	4	5
Truyền hình					
Báo in					
Tạp chí in					
Radio					
Internet					
ĐTDĐ					

	Rất thích	*Thích*	*Bình thường*	*Không thích*	*Rất không thích*
Truyền hình	1	2	3	4	5
Báo in	1	2	3	4	5
Tạp chí in	1	2	3	4	5
Radio	1	2	3	4	5
Internet	1	2	3	4	5
ĐTDĐ	1	2	3	4	5

Các dạng câu hỏi đóng (4)

- Câu hỏi đóng có câu trả lời dạng thang đo/thang đánh giá bằng điểm số
  - *Bạn hãy sử dụng thang điểm từ 1 đến 10 để cho biết sở thích của bạn đối với các phương tiện truyền thông sau đây, với 1 = ít thích nhất và 10 = thích nhiều nhất*

	Thang điểm đánh giá mức độ yêu thích									
Truyền hình	1	2	3	4	5	6	7	8	**9**	10
Báo in	1	2	3	4	**5**	6	7	8	9	10
Tạp chí in	1	2	**3**	4	5	6	7	8	9	10
Radio	**1**	2	3	4	5	6	7	8	9	10
Internet	1	2	3	4	5	6	7	8	**9**	10
ĐTDĐ	1	2	3	4	5	6	7	8	9	**10**

Các dạng câu hỏi đóng (5)

- Câu hỏi đóng có câu trả lời dạng thang đo thái độ/quan điểm
  - *Bạn hãy cho biết quan điểm của mình về câu nói "Đã là đàn ông thì phải biết uống rượu"*

Hoàn toàn không đồng ý	*1*
Không đồng ý	*2*
không có ý kiến	*3*
Đồng ý	*4*
Hoàn toàn đồng ý	*5*

Các dạng câu hỏi đóng (6)

- Câu hỏi đóng có câu trả lời dạng thứ tự xếp hạng
  - *Bạn hãy sử dụng số thứ tự từ 1 đến 5 để cho biết tầm quan trọng của các yếu tố có ảnh hưởng đến quyết định mua sắm ĐTDĐ của bạn, với 1 = quan trọng nhất và 5 = ít quan trọng nhất.*

Giá	***5***
Chất lượng (độ bền)	***4***
Hình thức/mẫu mã	***3***
Tính năng của SP	***1***
Uy tín thương hiệu	***2***

Những sai lầm nên tránh (1)

- Tránh dùng những câu hỏi để hỏi về 2 thứ cùng 1 lúc
  - *Bạn đã từng sử dụng máy tính bảng và ĐTDĐ để truy cập Internet chưa?*
- Tránh những câu hỏi chung chung, không rõ ràng về khái niệm
  - *Bạn có biết nhiều về sản phẩm này không?*
  - *Bạn có hay uống nước ngọt không?*
- Tránh dùng từ chuyên môn, từ lóng, từ viết tắt trong câu hỏi
- Không chú ý đến sự khác biệt về từ ngữ giữa các địa phương (*ví dụ: máy điều hòa và máy lạnh*)
- Tránh dùng các phương án trả lời trùng lắp hoặc không bao trùm
  - *Năm nay bạn bao nhiêu tuổi?*

15-20	☐
20-25	☐
25-35	☐
35-55	☐

15-20	☐
25-35	☐
36-54	☐
55-65	☐

Những sai lầm nên tránh (2)

- Tránh dùng những câu hỏi quá dài
- Tránh sử dụng câu hỏi kiểu "*gợi ý*" câu trả lời
  - *Theo bạn thì ăn uống ngoài đường là không hợp vệ sinh?*
  - *Theo bạn thì quảng cáo trên tivi như vậy là quá nhiều phải không?*
- Tránh dùng câu hỏi dạng phủ định
  - *Theo bạn thì thái độ của nhân viên phục vụ như vậy là không đúng?*

Đúng ☐
Không đúng ☐ ✗

Có ☐
Không ☐ ✗

Kiểm định & đánh giá bảng câu hỏi

- Thử nghiệm BCH với những người tương tự như đối tượng nghiên cứu/khảo sát
  - Khả năng hiểu câu hỏi
  - Trật tự & lôgic của các câu hỏi
  - Mức độ sẵn lòng trả lời câu hỏi
  - Thời gian hoàn tất BCH
- Quan sát đáp viên trong quá trình phỏng vấn
- Thu thập ý kiến phản hồi từ phỏng vấn viên
- Đúc kết kinh nghiệm và chỉnh sửa BCH
- Thử nghiệm BCH lần 2 (*nếu thấy cần thiết*)
- Hoàn thiện BCH

Trình bày Bảng câu hỏi

- Rõ ràng & nhất quán
  - *Có đủ chỗ để ghi nhận câu trả lời*
- Font chữ đủ lớn để có thể đọc được
  - *Phù hợp với đối tượng khảo sát*
- Ngắt trang hợp lý
  - *Các phương án trả lời của 1 câu hỏi nên nằm trên cùng 1 trang*
- Trong trường hợp cần thiết, in màu sẽ tốt hơn
- Nên sử dụng các câu hỏi sàng lọc hay câu hỏi "*dẫn đường*" nếu cần thiết
- Nên có hướng dẫn hỏi/trả lời in ngay trên BCH để giúp PVV làm việc hiệu quả hơn
- Các phương án trả lời nên được mã hóa (*coded*) để giúp cho việc nhập liệu vào máy tính, tổng hợp & phân tích thông tin dễ dàng hơn.