

Phương pháp nghiên cứu trong khoa học liên ngành - Các phương pháp định lượng

Nguyễn Bích Ngọc

Khoa các khoa học liên ngành, ĐHQGHN

Thông tin lớp học

- Lý thuyết: 26/03 (T), 27/03 (T)
- Thực hành: 29/03 (S/C)
 - Excel
 - JASP

Mục tiêu lớp học

- Giới thiệu khái niệm cơ bản của các phương pháp định lượng và thống kê
- Xác định vấn đề và định hướng phương pháp sử dụng (tên phương pháp)

Tài liệu tham khảo

- Cẩm nang nghiên cứu khoa học: từ ý tưởng đến công bố – Nguyễn Văn Tuấn (2nd edition, 2020)
- Từng bước nhập môn nghiên cứu khoa học xã hội – Phạm Hiệp & cộng sự (2022)
- Fundamentals of data visualization – Claus O. Wilke (<https://clauswilke.com/dataviz/index.html>)
- Applied statistics with R – David Dalpiaz (<https://book.stat420.org/>)
- The Scientist's Guide to Writing: How to Write More Easily and Effectively throughout Your Scientific Career – Stephen B. Heard (2nd 2022)
- Understanding research methods – Coursera (<https://www.coursera.org/learn/research-methods/home/info>)

Nội dung

- Giới thiệu chung
- Dữ liệu và nguồn dữ liệu
- Phân tích dữ liệu thăm dò
- Phân tích dữ liệu khẳng định
- Các phương pháp nâng cao

Giới thiệu chung

Khoa học?

Khoa học?

Science

[Article](#) [Talk](#)

From Wikipedia, the free encyclopedia

For a topical guide, see [Outline of science](#). For other uses, see [Science \(disambiguation\)](#).

Science is a rigorous, systematic endeavor that builds and organizes knowledge in the form of testable explanations and predictions about the universe.^{[1][2]} Modern science is typically divided into three major

Khoa học?

Science

[Article](#) [Talk](#)

From Wikipedia, the free encyclopedia

For a topical guide, see [Outline of science](#). For other uses, see [Science \(disambiguation\)](#).

Science is a rigorous, systematic endeavor that builds and organizes knowledge in the form of testable explanations and predictions about the universe.^{[1][2]} Modern science is typically divided into three major

Nghiên cứu khoa học?

Nghiên cứu khoa học?

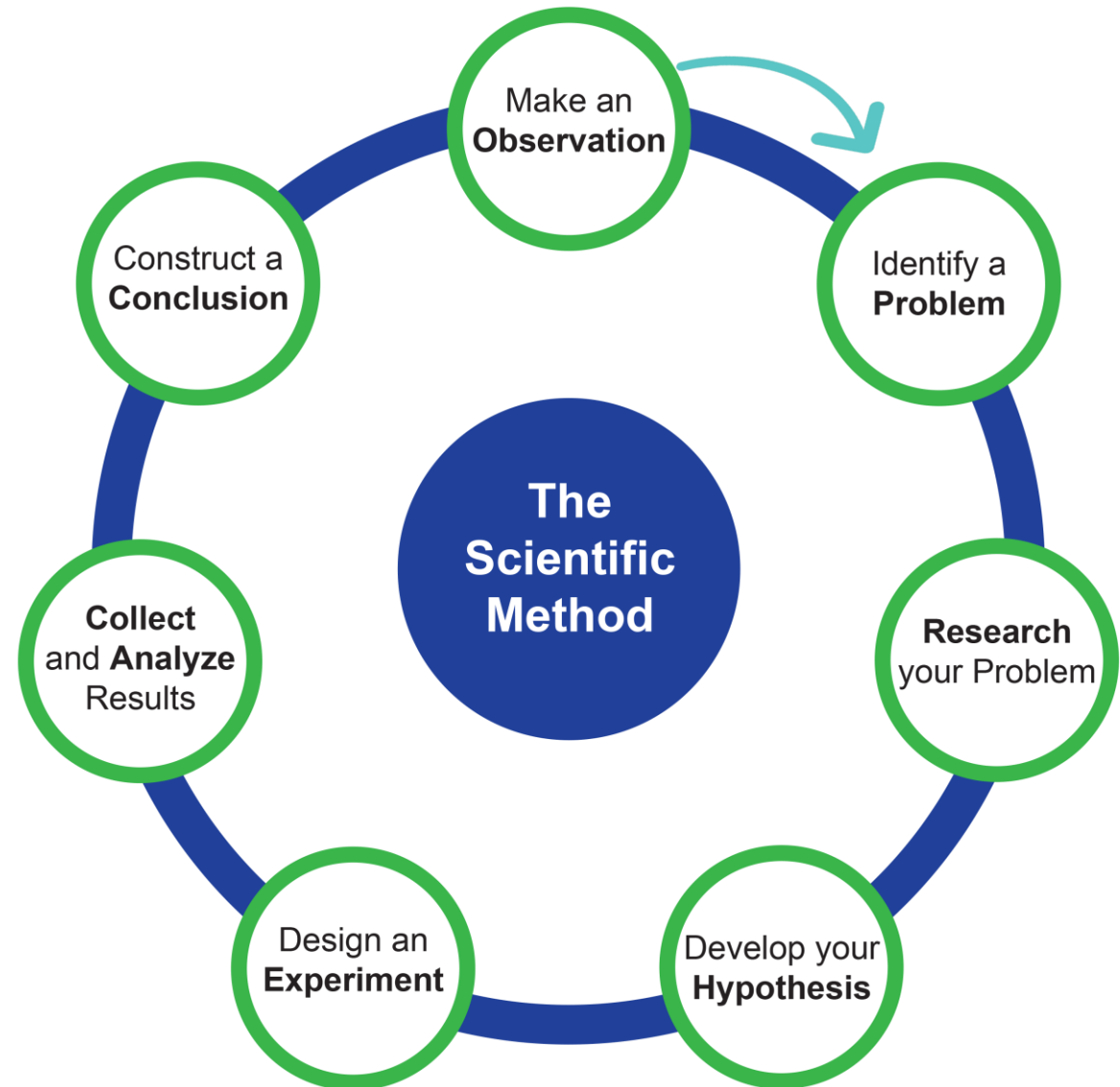
Research is **systematic inquiry** that helps to **make sense of the world** and that helps to make sensible the debates and interpretations that we have of issues of **contemporary significance**.

Professor Sandra Halperin

<https://www.coursera.org/learn/research-methods/home/info>

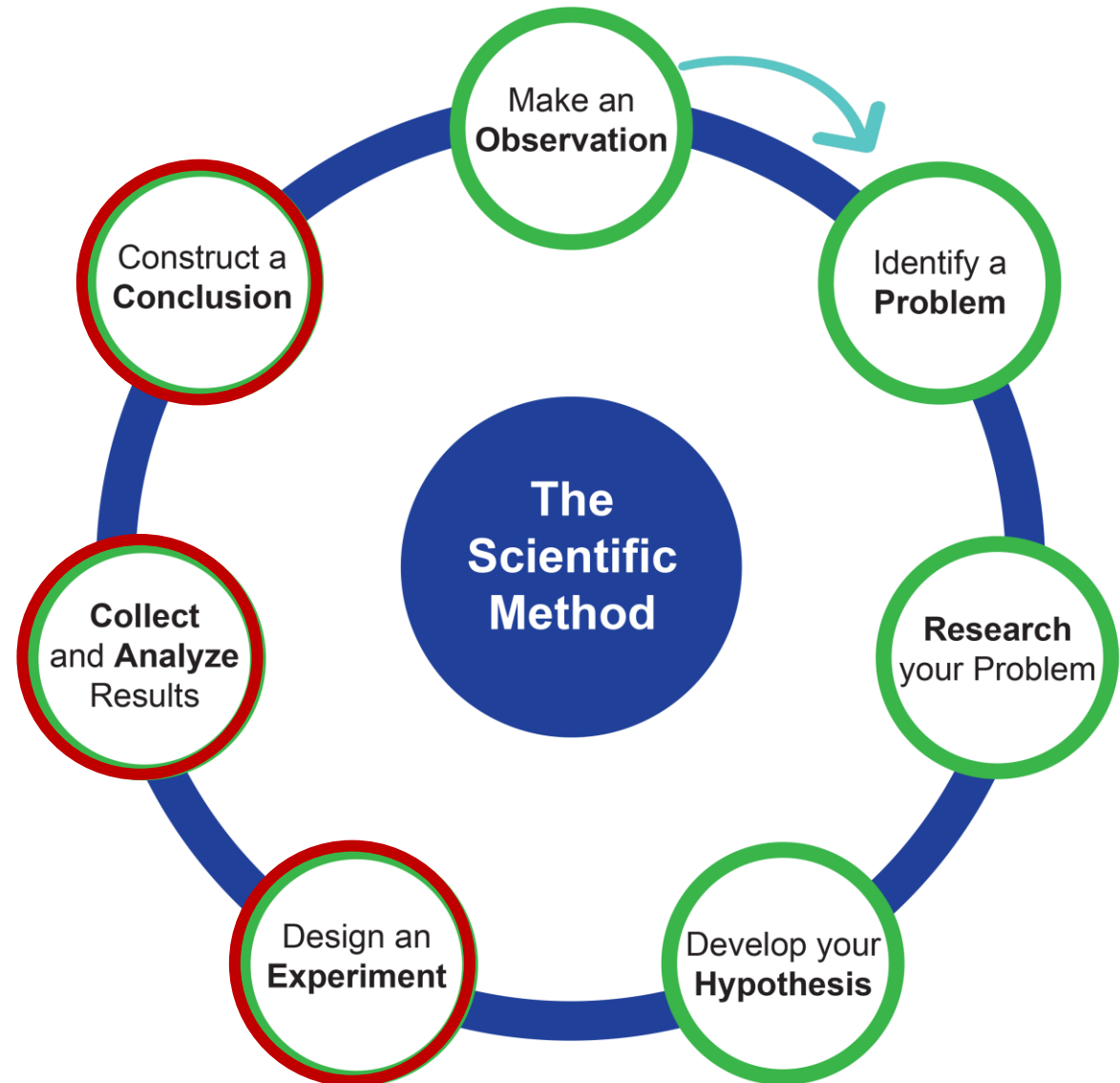
Quá trình nghiên cứu

Quá trình nghiên cứu



Source: <https://www.arfreethinkers.org/>

Phương pháp nghiên cứu?



Source: <https://www.arfreethinkers.org/>

Dữ liệu và nguồn dữ liệu

Dữ liệu

- Dữ liệu là gì? (Dữ liệu vs. thông tin?)

Dữ liệu

- Dữ liệu là gì? (Dữ liệu vs. thông tin?)
- Dữ liệu có cấu trúc và dữ liệu phi cấu trúc


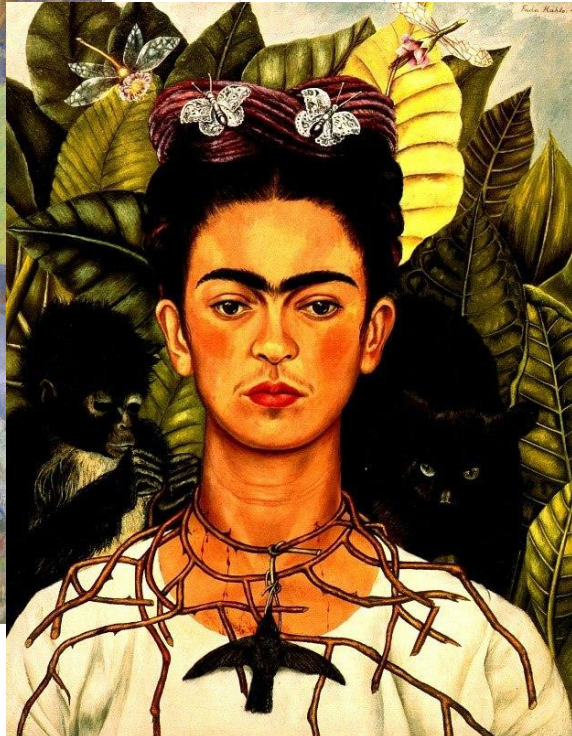
Dữ liệu

- Dữ liệu là gì? (Dữ liệu vs. thông tin?)
- Dữ liệu có cấu trúc và dữ liệu phi cấu trúc

	name	gender	bplace_country	birthyear	deathyear	occupation	genre
1	Camille Pissarro	M	U.S. Virgin Islands	1830	1903	PAINTER	Impressionism,Post-Impressionism
2	Francisco Goya	M	Spain	1746	1828	PAINTER	Romanticism
3	Piet Mondrian	M	Netherlands	1872	1944	PAINTER	Neoplasticism
4	Eugène Delacroix	M	France	1798	1863	PAINTER	Romanticism
5	Henri de Toulouse-Lautrec	M	France	1864	1901	PAINTER	Post-Impressionism
6	Caravaggio	M	Italy	1571	1610	PAINTER	Baroque
7	Henri Rousseau	M	France	1844	1910	PAINTER	Primitivism
8	Peter Paul Rubens	M	Germany	1577	1640	PAINTER	Baroque
9	Frida Kahlo	F	Mexico	1907	1954	PAINTER	Primitivism,Surrealism
10	Paul Gauguin	M	France	1848	1903	PAINTER	Symbolism,Post-Impressionism
11	Paul Cézanne	M	France	1839	1906	PAINTER	Post-Impressionism

Dữ liệu

- Dữ liệu là gì? (Dữ liệu vs. thông tin?)
- Dữ liệu có cấu trúc và dữ liệu phi cấu trúc

	name	gender	bplace_country	birthyear	deathyear	occupation	genre
1	Camille Pissarro	M	U.S. Virgin Islands	1830	1903	PAINTER	Impressionism,Post-Impressionism
2	Francisco Goya	M	Spain	1746	1828	PAINTER	Romanticism
3	Piet Mondrian	M	Netherlands	1872	1944	PAINTER	Neoplasticism
4					1863	PAINTER	Romanticism
5					1901	PAINTER	Post-Impressionism
6					1610	PAINTER	Baroque
7					1910	PAINTER	Primitivism
8					1640	PAINTER	Baroque
9							
10							ssionism
11							

Dữ liệu

Phi cấu trúc
=> cấu trúc?

Act 1, Scene 1

[Enter Sampson and Gregory, two high-ranking servants of the Capulet household, carrying swords and shields. Gregory is making fun of Sampson, who sees himself as a fearsome fighter]

Sampson

Gregory, on my word, we'll not carry coals.

Gregory

No, for then we should be **colliers**.
coal workers

Sampson

I mean, **an** we be **in choler** we'll draw.
if angered (our swords)

Gregory

Ay, while you live, draw your neck out of collar.

Sampson

I strike quickly, being **moved**.
provoked

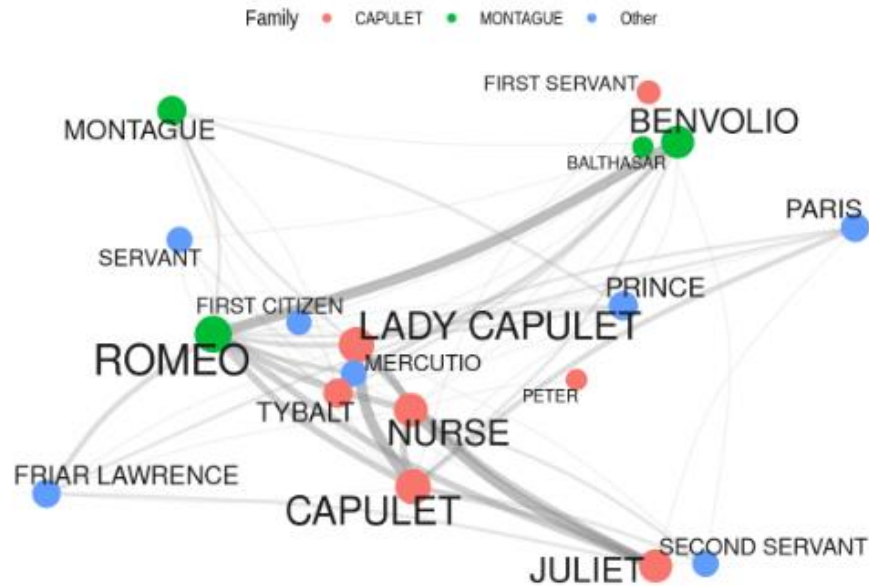
1

2

3

4

5



actscene	person	contrib	occurrences
ACT I_SCENE I	BENVOLIO	24	7
ACT I_SCENE I	CAPULET	2	9
ACT I_SCENE I	FIRST CITIZEN	1	2
ACT I_SCENE I	LADY CAPULET	1	10
ACT I_SCENE I	MONTAGUE	6	3
ACT I_SCENE I	PRINCE	1	3
ACT I_SCENE I	ROMEO	16	14
ACT I_SCENE I	TYBALT	2	3
ACT I_SCENE II	BENVOLIO	5	7
ACT I_SCENE II	CAPULET	3	9
ACT I_SCENE II	PARIS	2	5
ACT I_SCENE II	ROMEO	11	14
ACT I_SCENE II	SERVANT	8	3
ACT I_SCENE III	JULIET	5	11
ACT I_SCENE III	LADY CAPULET	11	10



Persona	BALTHASAR	BENVOLIO	CAPULET	FIRST CITIZEN	FIRST SERVANT
BALTHASAR	0	0	1	0	0
BENVOLIO	0	0	3	2	1
CAPULET	1	3	0	1	2
FIRST CITIZEN	0	2	1	0	0
FIRST SERVANT	0	1	2	0	0

Dữ liệu

- Dữ liệu là gì? (Dữ liệu vs. thông tin?)
- Dữ liệu có cấu trúc và dữ liệu phi cấu trúc
- Nguồn sơ cấp vs thứ cấp



Sơ cấp

Thực nghiệm

Khảo sát/Bảng hỏi

Đo đạc ngoài thực địa



Thứ cấp

Cơ sở dữ liệu mở

Báo cáo kỹ thuật của chính phủ

Báo cáo nội bộ

Bài tập

- Dữ liệu trong nghiên cứu của bạn?
- Có cấu trúc, phi cấu trúc?
- Nguồn sơ cấp, thứ cấp?

Dữ liệu

Biến

	Gender	Age.Range	Year	Nationality	Q1	Q2	Q3	Q4	Q5
1	Female	20 - 21 years old	Year 4	Thai	7	5	7	7	7
2	Female	20 - 21 years old	Year 4	Thai	6	5	7	5	6
3	Female	20 - 21 years old	Year 4	Thai	7	7	7	7	7
4	Female	20 - 21 years old	Year 4	Thai	7	2	7	6	7
5	Female	22 - 23 years old	Year 4	Thai	6	6	7	7	7
6	Male	20 - 21 years old	Year 3	Thai	5	4	4	4	4
7	Male	20 - 21 years old	Year 3	Thai	6	4	5	7	6
8	Female	20 - 21 years old	Year 3	Thai	7	4	7	6	7
9	Female	20 - 21 years old	Year 3	Thai	7	5	7	7	7
10	Male	20 - 21 years old	Year 3	Thai	5	5	5	7	6
11	Female	20 - 21 years old	Year 3	Thai	7	5	7	7	7

Đối tượng
quan sát/Mẫu

Biến

- Các loại biến

- Định danh (nominal)

- Thứ bậc (ordinal)

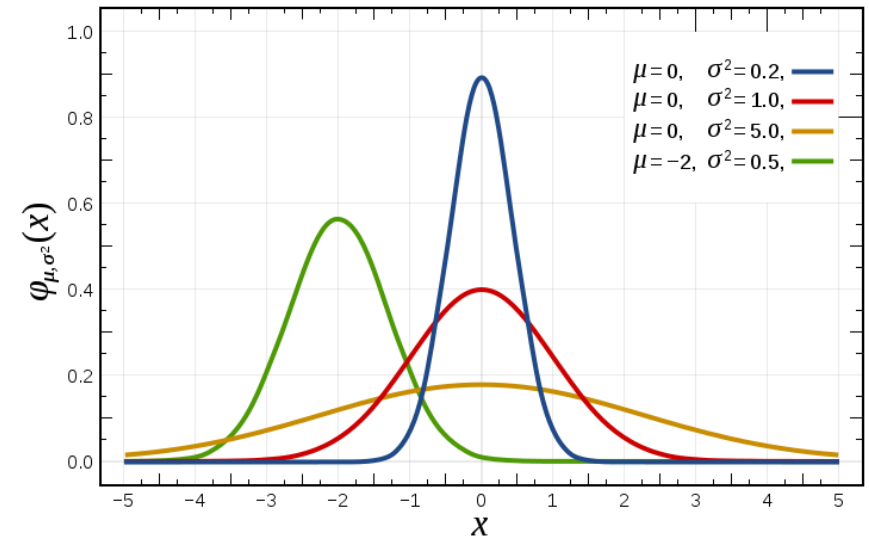
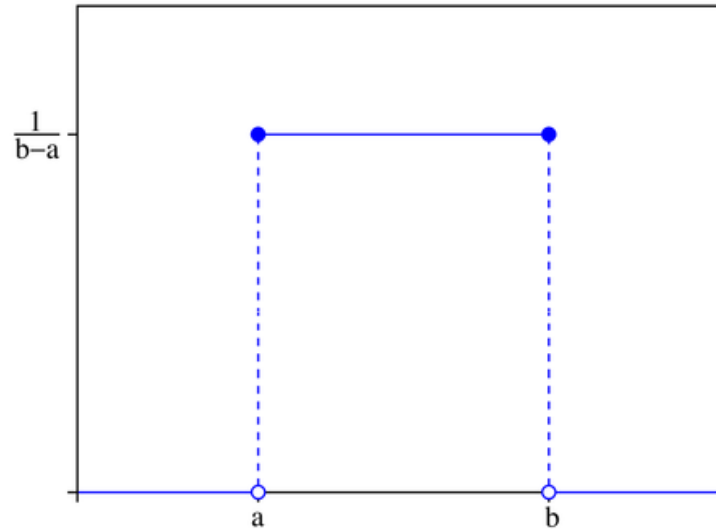
- Liên tục/Định lượng (continuous variables)



Biến rời rạc/định tính
(discrete variables)

Biến định danh – nominal variables

- Ví dụ:
 - Giới tính
 - Tôn giáo
 - Quốc tịch
- Phân phối
 - Đồng nhất
 - Nhị phân
 - Đa thức

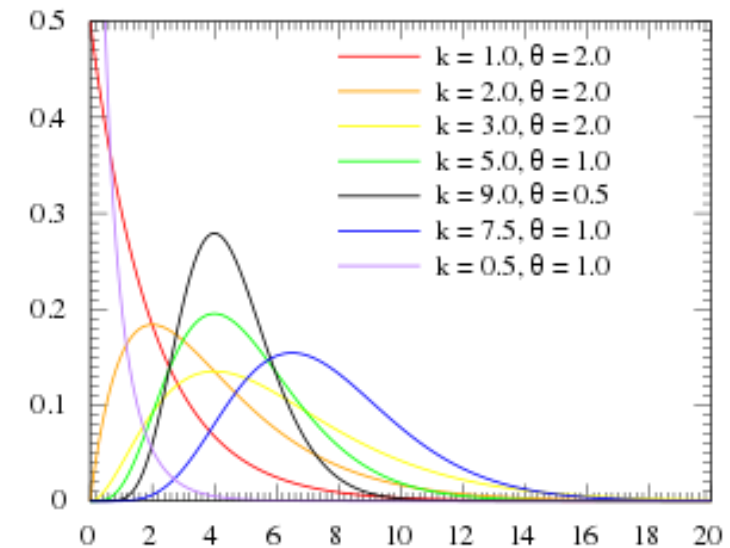
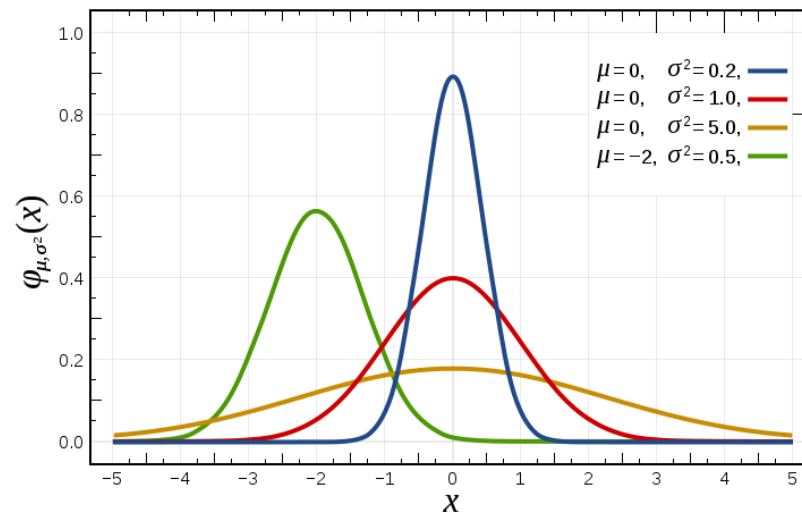


Biến thứ bậc – ordinal variables

- Ví dụ:
 - Thang Likert: hoàn toàn không đồng ý – hoàn toàn đồng ý
 - Trình độ học vấn: THCS, THPT, trung cấp, đại học, sau đại học
 - Điều kiện kinh tế xã hội: thấp, trung bình, cao
 - Đánh giá/chấm điểm: 1 – 5 ★
- Đặc điểm
 - Có tính thứ bậc tự nhiên
 - Không thể khẳng định khoảng cách bằng nhau giữa các giá trị

Biến liên tục/định lượng – continuous variables

- Ví dụ:
 - Tuổi đời (người/di tích)
 - Diện tích
 - Dân số
- Phân phối
 - Đồng nhất
 - Chuẩn (chuẩn tắc)
 - Gamma
 - ...



Bài tập

name	gender	bplace_country	birthyear	deathyear	occupatio	genre	paintings	l	non_en_p	hpi	hpi_cat
Camille Pissarro	M	U.S. Virgin Islands	1830	1903	PAINTER	Impressionism,Post-Impressionism	91	67	171897	74.6474	1
Francisco Goya	M	Spain	1746	1828	PAINTER	Romanticism	291	144	719332	84.0449	2
Piet Mondrian	M	Netherlands	1872	1944	PAINTER	Neoplasticism	84	66	495337	79.8805	2
Eugène Delacroix	M	France	1798	1863	PAINTER	Romanticism	31	83	351122	79.6869	2
Henri de Toulouse-Lautrec	M	France	1864	1901	PAINTER	Post-Impressionism	81	77	436156	77.0527	2
Caravaggio	M	Italy	1571	1610	PAINTER	Baroque	55	87	938325	84.2831	2
Henri Rousseau	M	France	1844	1910	PAINTER	Primitivism	70	62	188919	74.2238	1
Peter Paul Rubens	M	Germany	1577	1640	PAINTER	Baroque	141	130	500191	85.3987	3
Frida Kahlo	F	Mexico	1907	1954	PAINTER	Primitivism,Surrealism	120	141	4417902	85.919	3
Paul Gauguin	M	France	1848	1903	PAINTER	Symbolism,Post-Impressionism	311	102	823724	82.8685	2
Paul Cézanne	M	France	1839	1906	PAINTER	Post-Impressionism	47	103	583083	82.0348	2
Andrei Rublev	M	Russia	1360	1428	PAINTER	Byzantine Art	99	60	145561	71.3574	1
Alfred Sisley	M	France	1839	1899	PAINTER	Impressionism	259	62	85215	73.058	1
Paul Klee	M	Switzerland	1879	1940	PAINTER	Expressionism,Abstractionism,Surrealis	188	67	388492	76.607	2
Diego Rivera	M	Mexico	1886	1957	PAINTER	Social Realism,Muralism	70	64	784272	73.4297	1
Henri Matisse	M	France	1869	1954	PAINTER	Impressionism,Post-Impressionism	186	135	558341	81.9253	2
Pablo Picasso	M	Spain	1881	1973	PAINTER	Cubism	439	184	3535833	90.2854	3
Edvard Munch	M	Norway	1863	1944	PAINTER	Symbolism,Expressionism	67	88	721936	83.5982	2
Gustave Courbet	M	France	1819	1877	PAINTER	Realism	59	62	334576	77.3955	2
Hieronymus Bosch	M	Netherlands	1450	1516	PAINTER	Northern Renaissance	137	76	560732	80.8084	2
Gustav Klimt	M	Austria	1862	1918	PAINTER	Symbolism,Art Nouveau	117	82	897167	81.589	2
Claude Monet	M	France	1840	1926	PAINTER	Impressionism	73	115	1226028	85.3814	3
El Greco	M	Greece	1541	1614	PAINTER	Mannerism	87	95	392464	81.2606	2
Raphael	M	Italy	1483	1520	PAINTER	High Renaissance	109	151	1003727	89.2779	3
Marc Chagall	M	Belarus	1887	1985	PAINTER	Primitivism	239	74	556003	78.8858	2
Rembrandt	M	Netherlands	1606	1669	PAINTER	Baroque	262	161	1085677	90.137	3

Bài tập

	Gender	Age.Range	Year	Nationality	Q1	Q2	Q3	Q4	Q5
1	Female	20 - 21 years old	Year 4	Thai	7	5	7	7	7
2	Female	20 - 21 years old	Year 4	Thai	6	5	7	5	6
3	Female	20 - 21 years old	Year 4	Thai	7	7	7	7	7
4	Female	20 - 21 years old	Year 4	Thai	7	2	7	6	7
5	Female	22 - 23 years old	Year 4	Thai	6	6	7	7	7
6	Male	20 - 21 years old	Year 3	Thai	5	4	4	4	4
7	Male	20 - 21 years old	Year 3	Thai	6	4	5	7	6
8	Female	20 - 21 years old	Year 3	Thai	7	4	7	6	7
9	Female	20 - 21 years old	Year 3	Thai	7	5	7	7	7
10	Male	20 - 21 years old	Year 3	Thai	5	5	5	7	6
11	Female	20 - 21 years old	Year 3	Thai	7	5	7	7	7

Mẫu, quần thể, cỡ mẫu

- Quần thể? Mẫu?

- Tính đại diện?

https://youtu.be/rxv_sB-wOkY

- Cỡ mẫu?

https://nckh.huph.edu.vn/sites/nckh.huph.edu.vn/files/Ph%C6%B0%C6%A1ng%20ph%C3%A1p%20ch%E1%BB%8Dn%20m%E1%BA%ABu%20v%C3%A0%20t%C3%ADnh%20t%C3%A1n%20c%E1%BB%A1%20m%E1%BA%ABu_revised%20l%E1%BA%A7n%201_5.8.2020_0.pdf

Phương pháp lấy mẫu

- Mẫu ngẫu nhiên (Probability/Random sample)
 - Mẫu ngẫu nhiên đơn giản (Simple random sample)
 - Mẫu ngẫu nhiên hệ thống (Systematic sample)
 - Mẫu ngẫu nhiên phân loại (Stratified sample)
 - Mẫu ngẫu nhiên cụm (Cluster sample)
- Mẫu không ngẫu nhiên (Nonprobability sample)
 - Mẫu thuận tiện (Convenience sample)
 - Mẫu hạn ngạch (Quota sample)
 - Mẫu có mục đích (Judgement (or purposive) sample)
 - Mẫu bóng tuyết (Snowball sample)

Sử dụng
trọng số để
khắc phục
tính không
đại diện

Bảng hỏi

Table 1
Description of the characteristics in the dataset.

Column	Data label	Explanation
Column A	Student Status	Degree student; Exchange student
Column B	Institution	Prince of Songkla University
Column C	Faculty	Faculty of Hospitality and Tourism; College of Computing; Faculty of International Studies
Column D	Gender	Male; Female; I do not wish to say; Other
Column E	Age Range	18–19 years old; 20–21 years old; 22–23 years old; 24 years or above
Column F	Year	Year 1; Year 2; Year 3; Year 4
Column G	Nationality	Thai; Foreign
Column H	Probe	Have you heard about environmental sustainability before? [Answer options: (Yes) and (No)].

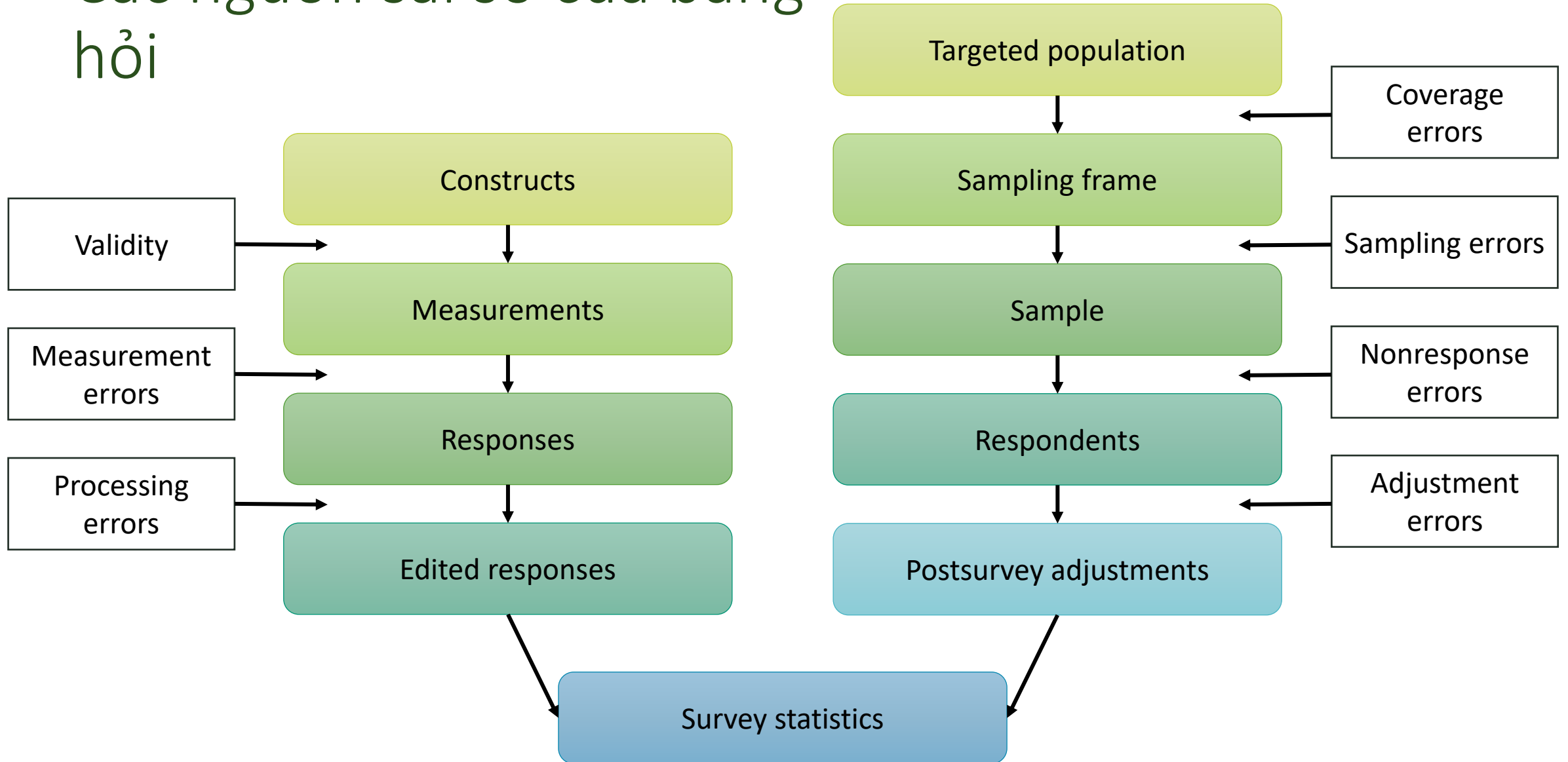
Table 2
Questionnaire organized by their respective factor.

Column	Data label	Explanation
<i>Attitude</i>		
Column I	Question 1	In my opinion, it is important to protect the environment.
Column J	Question 2	I actively practice environmental sustainability at home (e.g., energy conservation, recycling).
Column K	Question 3	Everyone is responsible for caring for the environment
Column L	Question 4	I am concerned about the long-term future of the environment.
Column M	Question 5	In my opinion, it is important to conserve natural resources.
Column N	Question 6	I think that environmental sustainability is a waste of time and effort.
Column O	Question 7	I am a passionate advocate of environmental sustainability.
<i>Perceived behavioral control</i>		
Column P	Question 8	It is easy for me to perform environmentally sustainable activities (e.g., energy conservation, recycling).
Column Q	Question 9	I have control over my actions to support the environment.
Column R	Question 10	It is my decision whether or not to perform environmentally sustainable activities.
Column S	Question 11	I have the ability to carry out environmentally sustainable activities.
Column T	Question 12	I have control over performing environmentally sustainable activities.

Thiết kế bảng hỏi

- Phạm trù (Construct) cần quan tâm
 - Là gì?
 - **Làm sao để đo?**
- Thiết kế bảng hỏi cần chú ý
 - Cách dùng từ
 - Tránh việc sử dụng chỉ một câu hỏi để đo lường cho 1 phạm trù
 - Tâm lý người hỏi và người trả lời
 - **Luôn luôn thử nghiệm trước** bộ câu hỏi

Các nguồn sai số của bảng hỏi

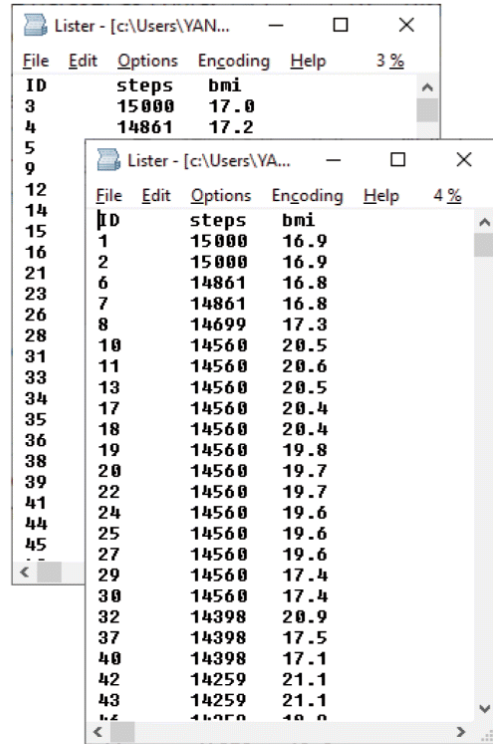


Phân tích dữ liệu thăm dò

Tìm hiểu dữ liệu/Biểu diễn dữ liệu

- Là bước không thể bỏ qua
- Giúp phát hiện những vấn đề trong dữ liệu
- Giúp có hình dung chung về dữ liệu và các mối tương quan giữa các dữ liệu
- Phát triển giả thuyết, và lý thuyết mới

a



The screenshot shows a Notepad window titled 'Lister - [c:\Users\YAN...]' with a menu bar (File, Edit, Options, Encoding, Help) and a status bar (3 %). The text content is as follows:

ID	steps	bmi
3	15000	17.0
4	14861	17.2

Below this, there is a second Notepad window titled 'Lister - [c:\Users\YA...]' with a menu bar (File, Edit, Options, Encoding, Help) and a status bar (4 %). It contains a larger table:

ID	steps	bmi
1	15000	16.9
2	15000	16.9
6	14861	16.8
7	14861	16.8
8	14699	17.3
10	14560	20.5
11	14560	20.6
13	14560	20.5
17	14560	20.4
18	14560	20.4
19	14560	19.8
20	14560	19.7
22	14560	19.7
24	14560	19.6
25	14560	19.6
27	14560	19.6
29	14560	17.4
30	14560	17.4
32	14398	20.9
37	14398	17.5
40	14398	17.1
42	14259	21.1
43	14259	21.1
44	14259	21.1

Yanai, I., Lercher, M. A hypothesis is a liability. *Genome Biol* **21**, 231 (2020). <https://doi.org/10.1186/s13059-020-02133-w>

a

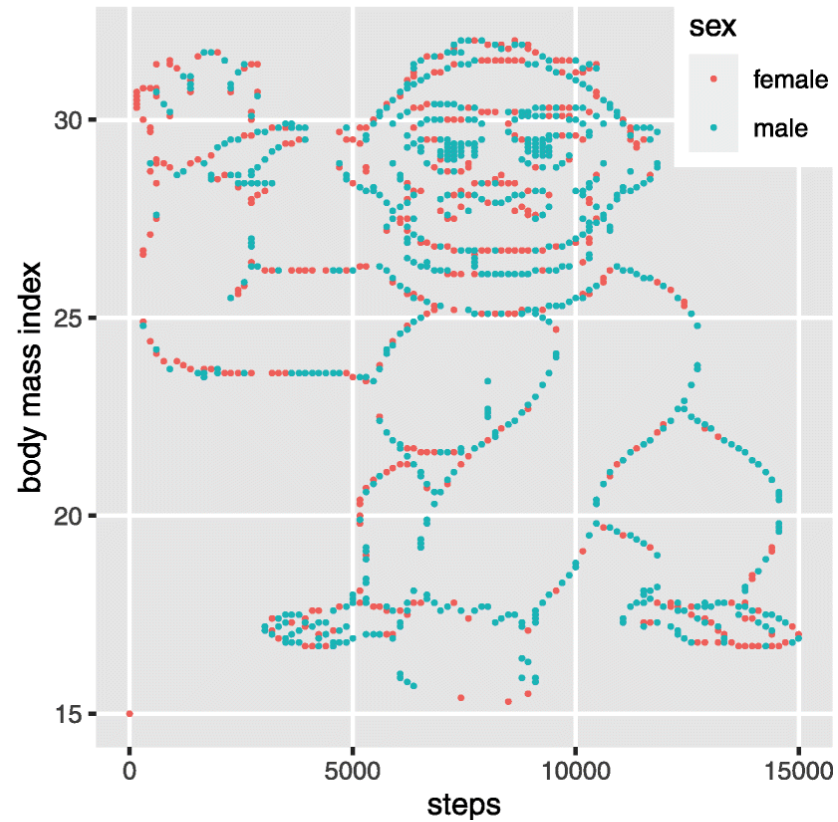
List - [c:\Users\YAN...]

ID	steps	bmi
3	15000	17.0
4	14861	17.2

List - [c:\Users\YA...]

ID	steps	bmi
1	15000	16.9
2	15000	16.9
6	14861	16.8
7	14861	16.8
8	14699	17.3
10	14560	20.5
11	14560	20.6
13	14560	20.5
17	14560	20.4
18	14560	20.4
19	14560	19.8
20	14560	19.7
22	14560	19.7
24	14560	19.6
25	14560	19.6
27	14560	19.6
29	14560	17.4
30	14560	17.4
32	14398	20.9
37	14398	17.5
40	14398	17.1
42	14259	21.1
43	14259	21.1
44	14259	21.1
45	14259	21.1

b



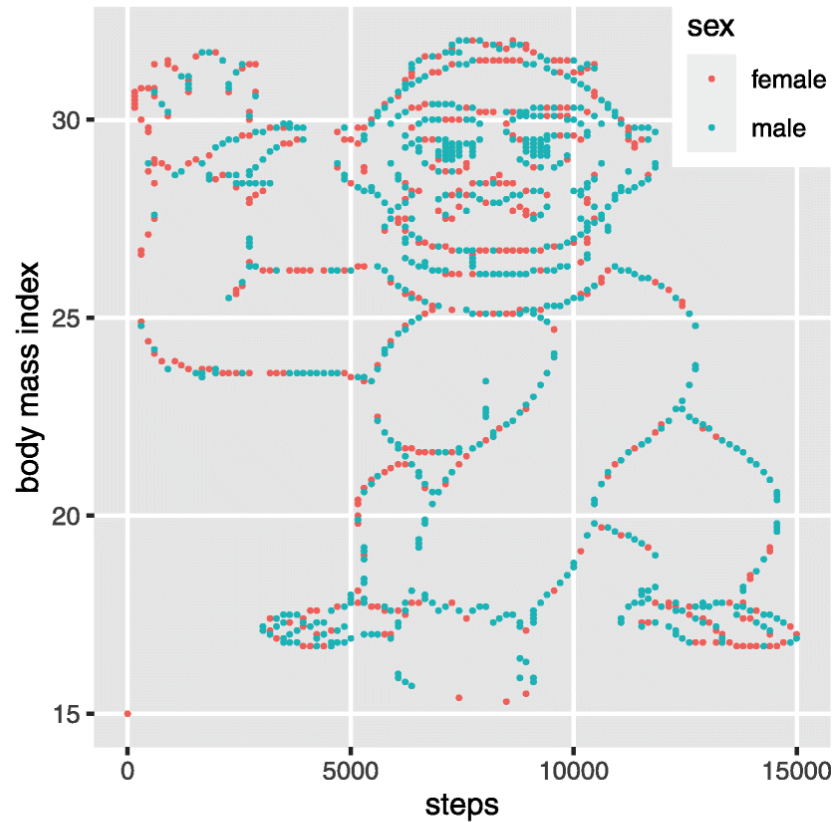
Yanai, I., Lercher, M. A hypothesis is a liability. *Genome Biol* **21**, 231 (2020). <https://doi.org/10.1186/s13059-020-02133-w>

a

Figure a shows two screenshots of a spreadsheet application (likely Excel) displaying data for individuals (ID, steps, bmi) across different rows. The top screenshot shows rows 3 and 4, and the bottom screenshot shows rows 1 through 45.

ID	steps	bmi
3	15000	17.0
4	14861	17.2
1	15000	16.9
2	15000	16.9
6	14861	16.8
7	14861	16.8
8	14699	17.3
10	14560	20.5
11	14560	20.6
13	14560	20.5
17	14560	20.4
18	14560	20.4
19	14560	19.8
20	14560	19.7
22	14560	19.7
24	14560	19.6
25	14560	19.6
27	14560	19.6
29	14560	17.4
30	14560	17.4
32	14398	20.9
37	14398	17.5
40	14398	17.1
42	14259	21.1
43	14259	21.1
44	14259	21.1
45	14259	21.1

b

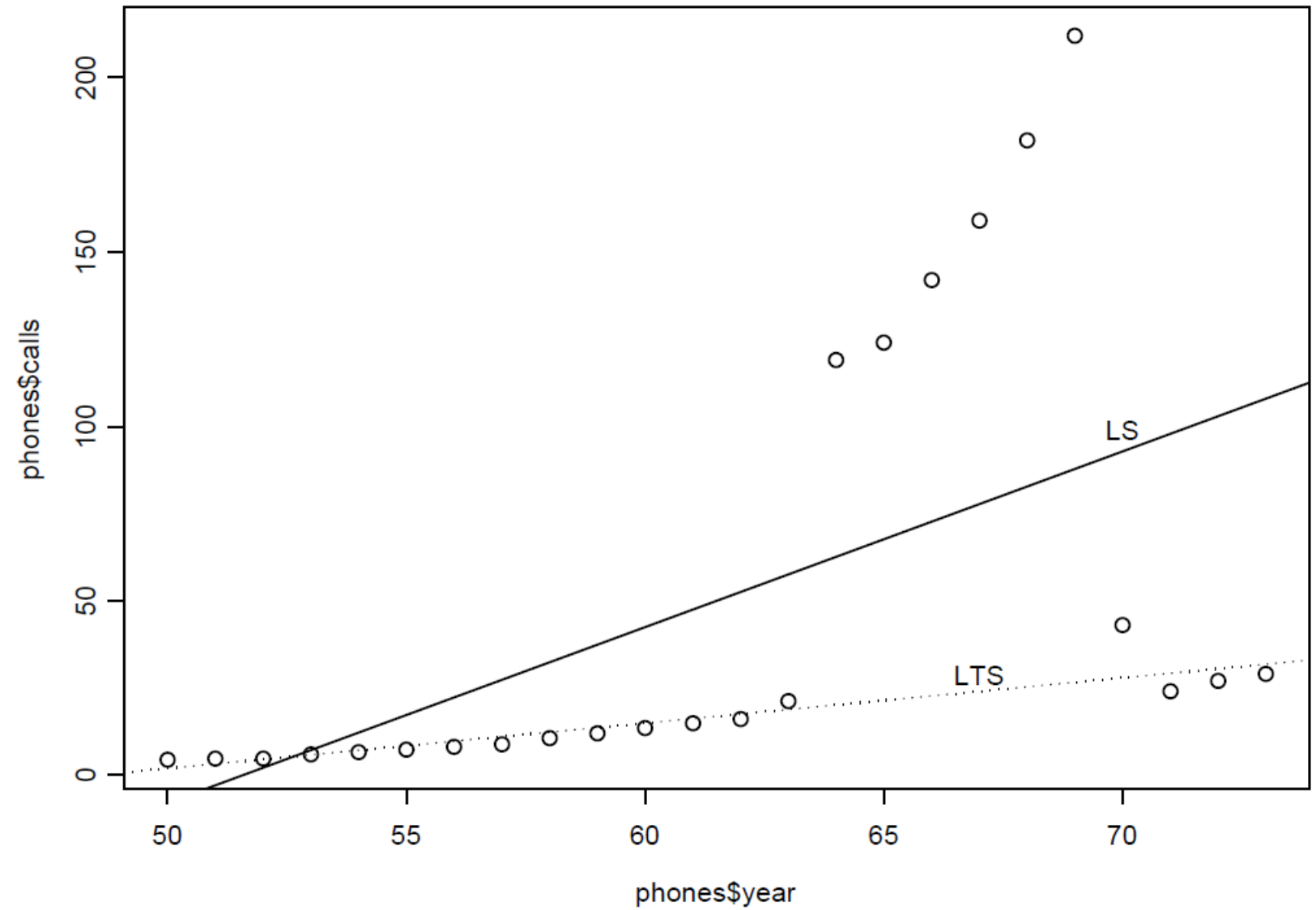


c

	Gorilla <u>not</u> discovered	Gorilla discovered
Hypothesis-focused	14	5
Hypothesis-free	5	9

Yanai, I., Lercher, M. A hypothesis is a liability. *Genome Biol* **21**, 231 (2020). <https://doi.org/10.1186/s13059-020-02133-w>

- Dữ liệu điện thoại
- Cuộc gọi (triệu) ra nước ngoài từ Bỉ từ 1950-1973.



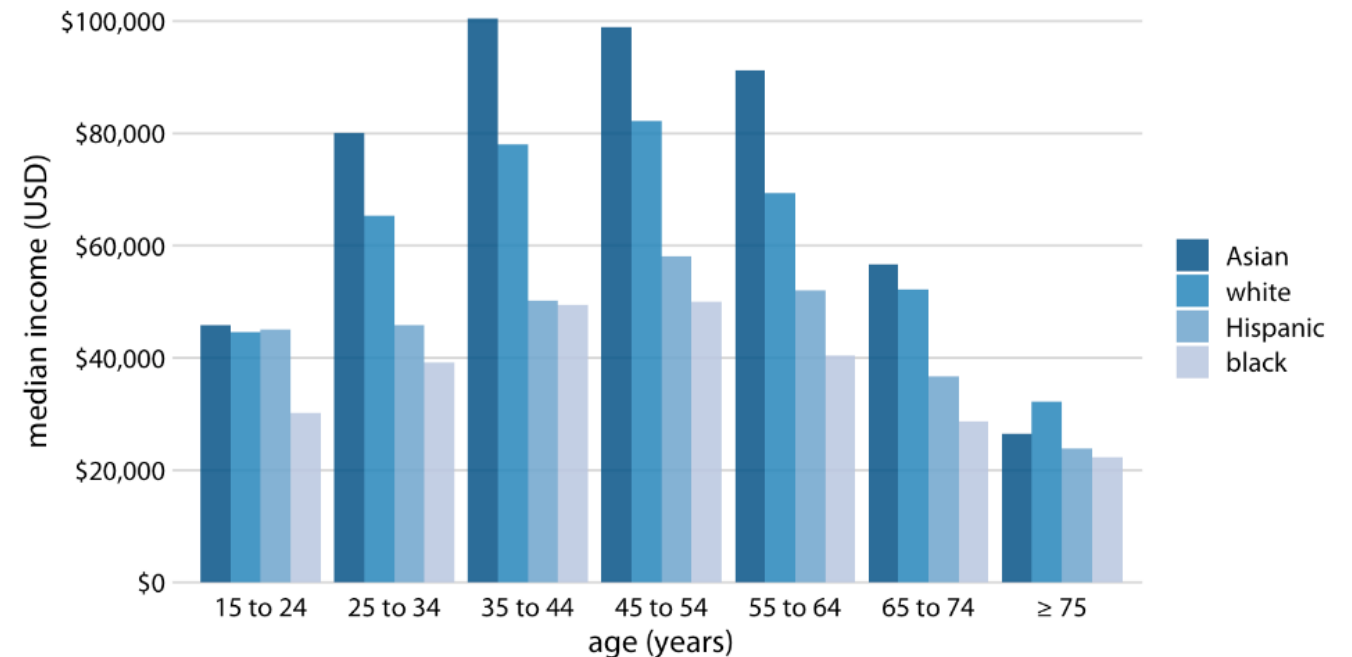
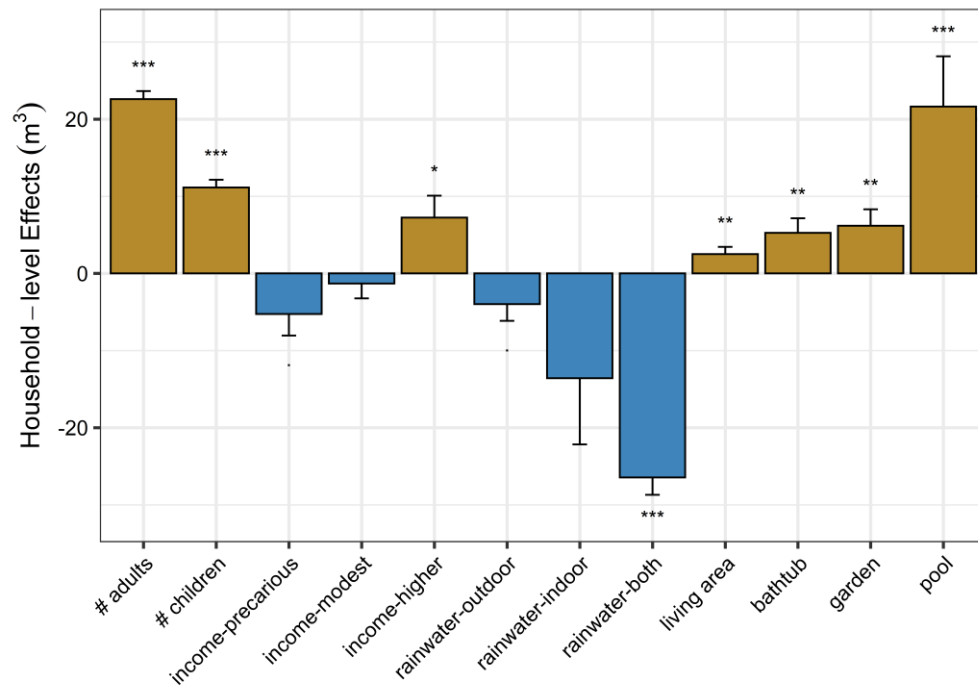
Đồ thị

- Rõ ràng
- Chính xác
- Hiệu quả
- Tối đa thông tin, tối thiểu mực in

https://www.ted.com/talks/hans_rosling_the_best_stats_you_ve_ever_seen

Một số loại đồ thị thông thường

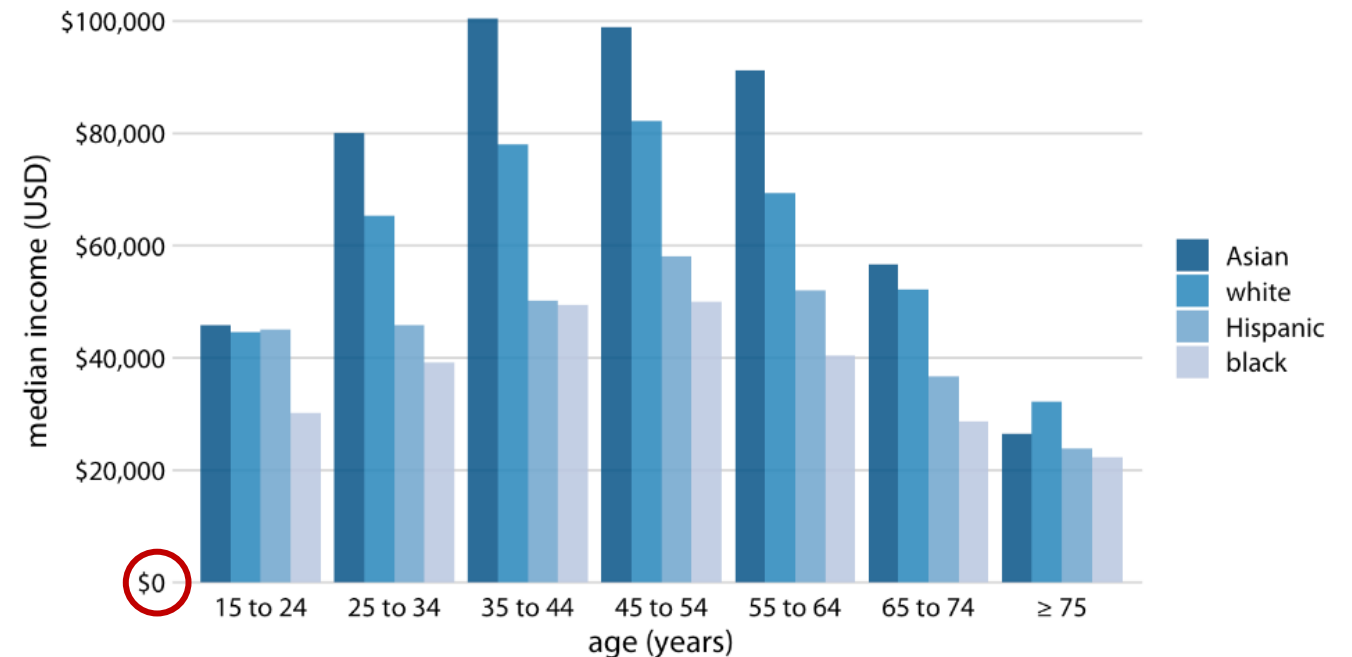
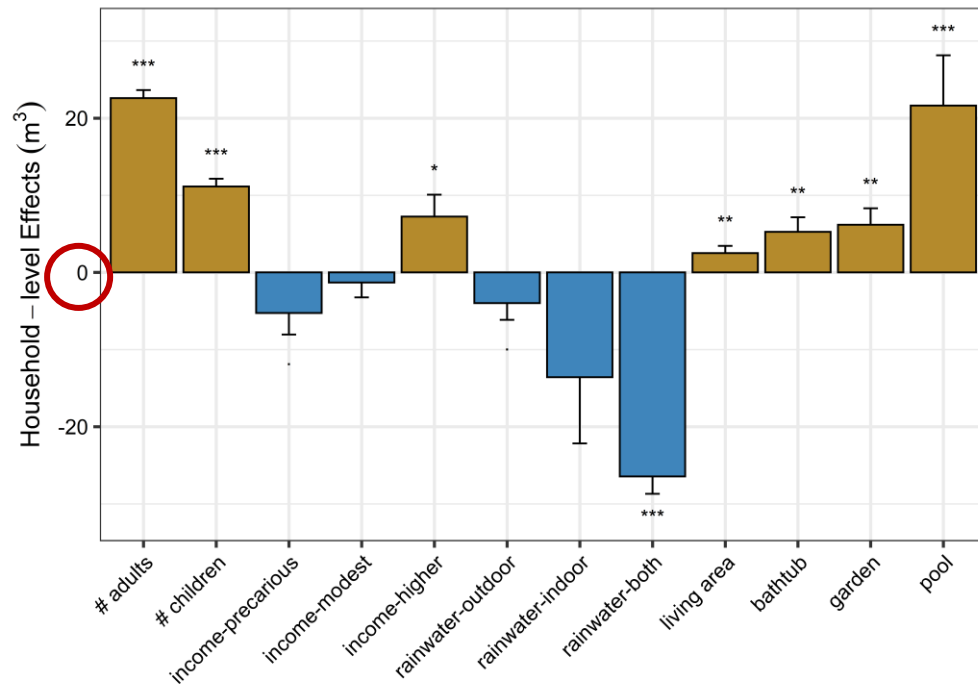
Đồ thị cột – bar charts



Wilke (2018)

Một số loại đồ thị thông thường

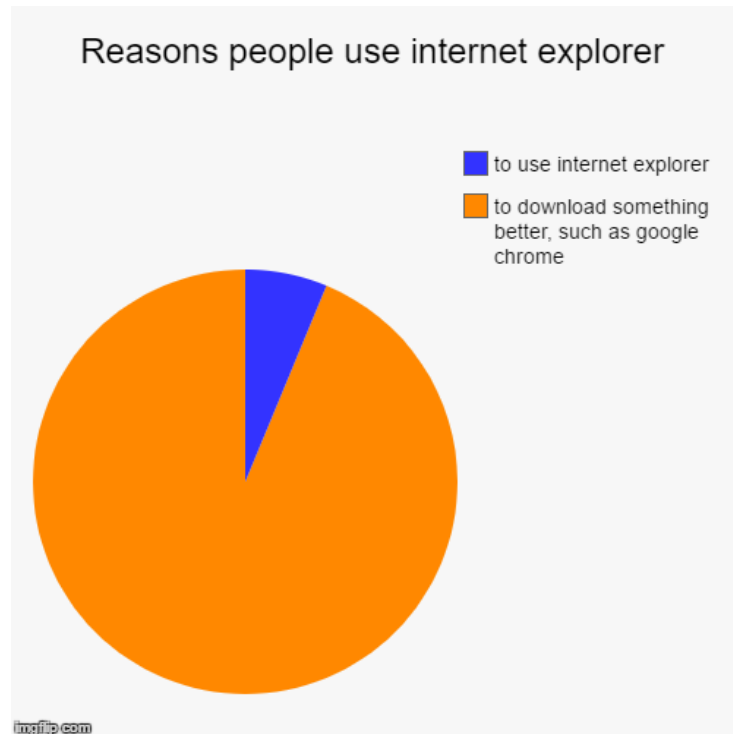
Đồ thị cột – bar charts



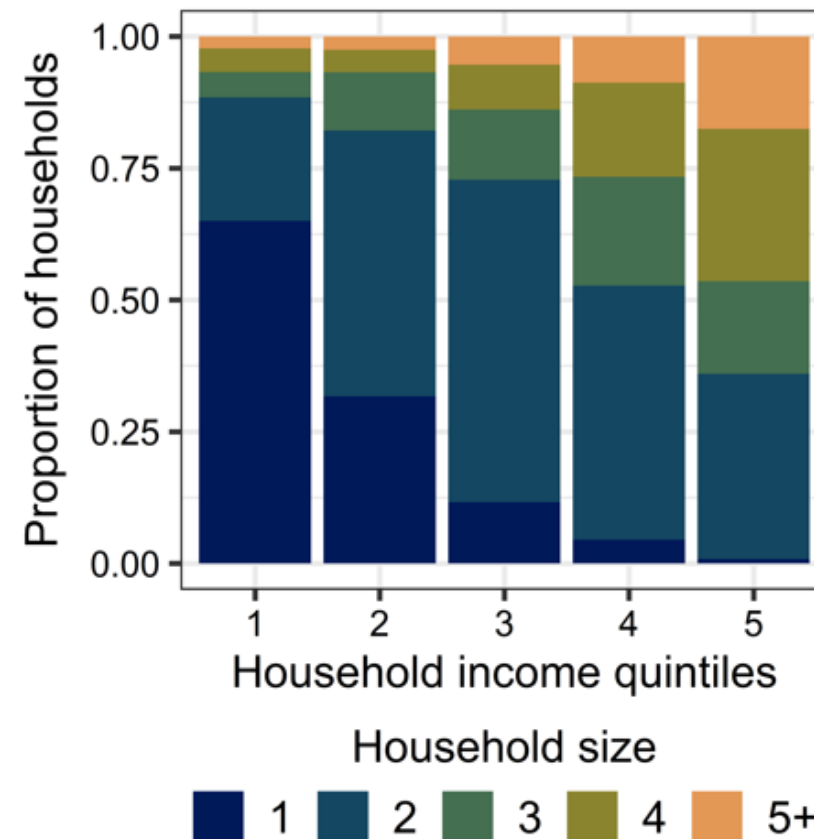
Wilke (2018)

Một số loại đồ thị thông thường

Đồ thị quạt – pie chart

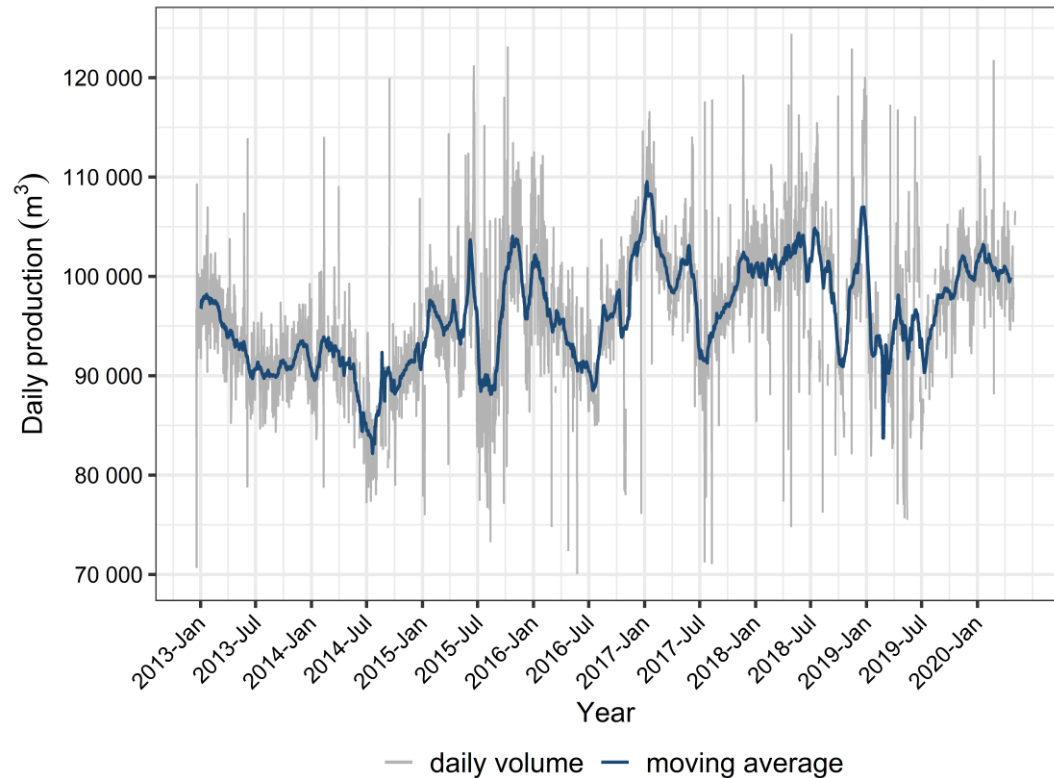
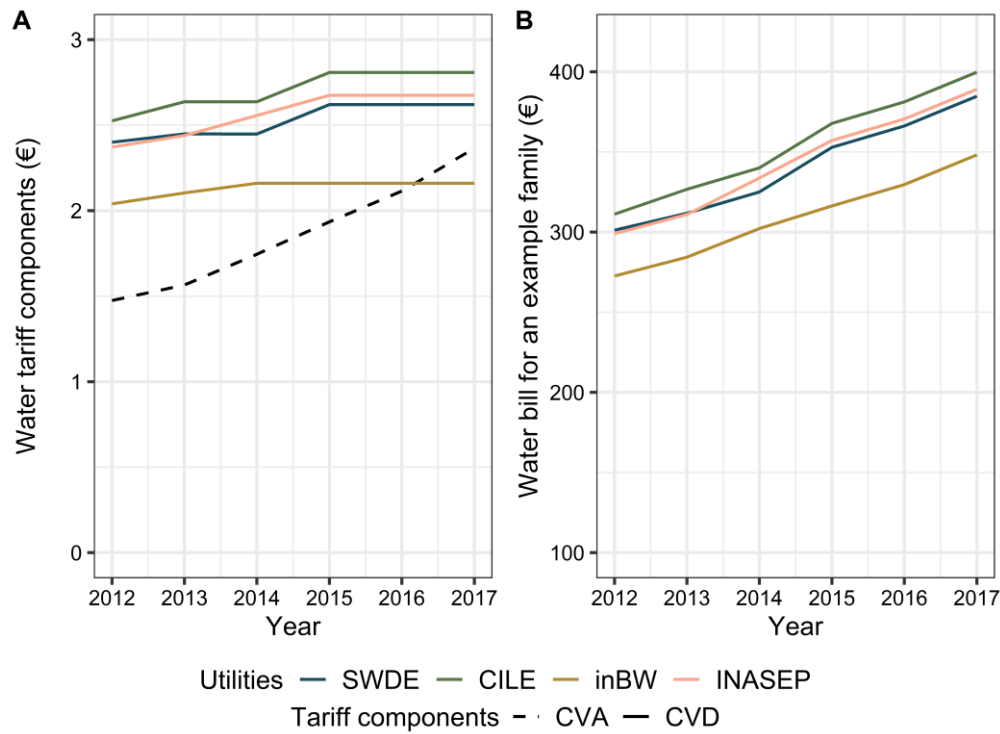


Đồ thị cột xếp chồng – Stacked bar charts



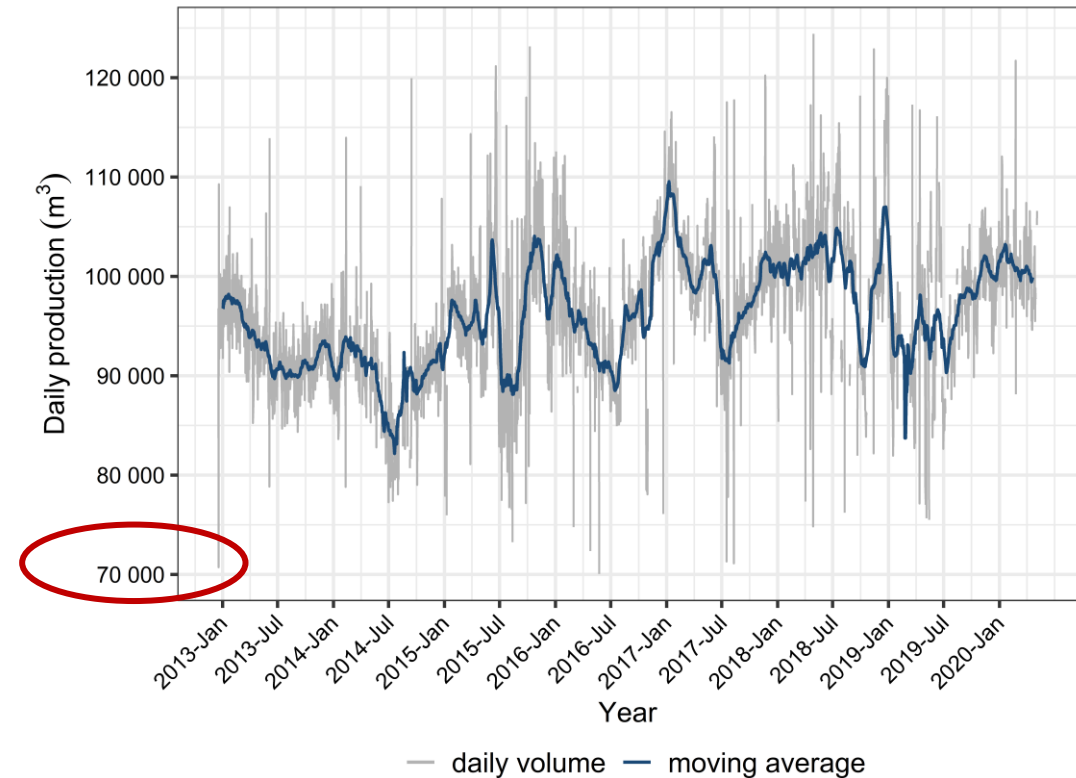
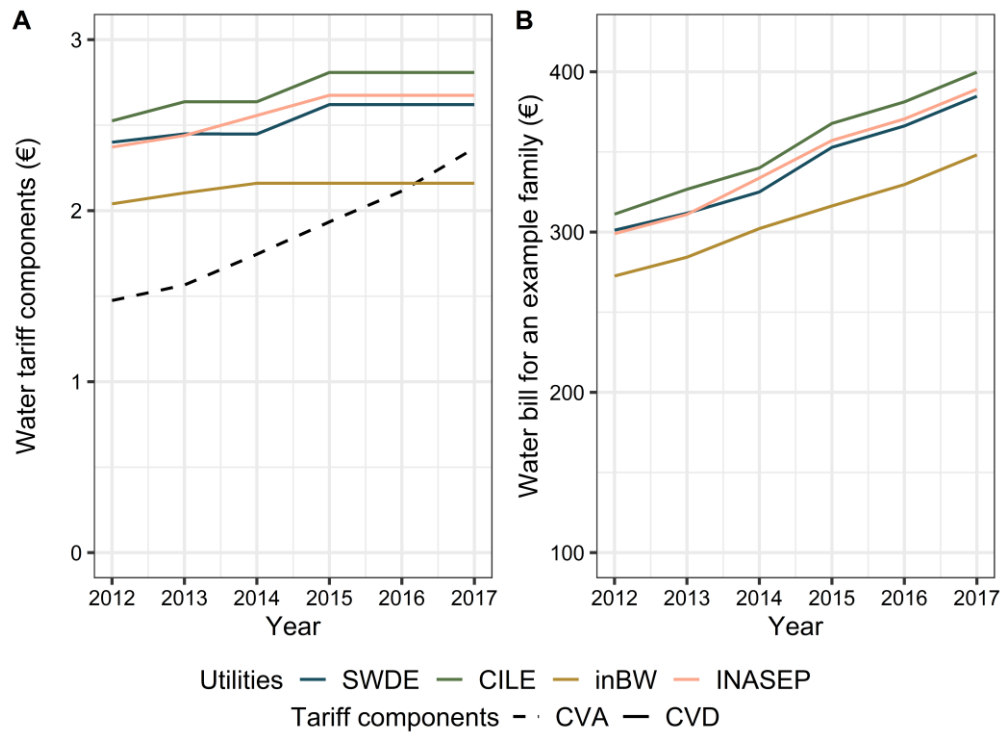
Một số loại đồ thị thông thường

Đồ thị đường – line charts



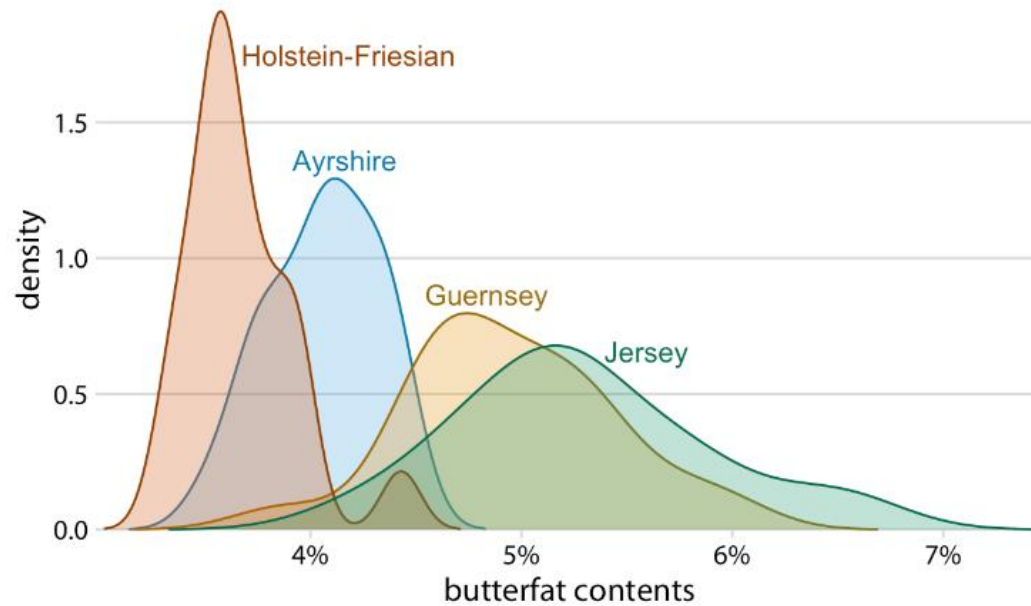
Một số loại đồ thị thông thường

Đồ thị đường – line charts

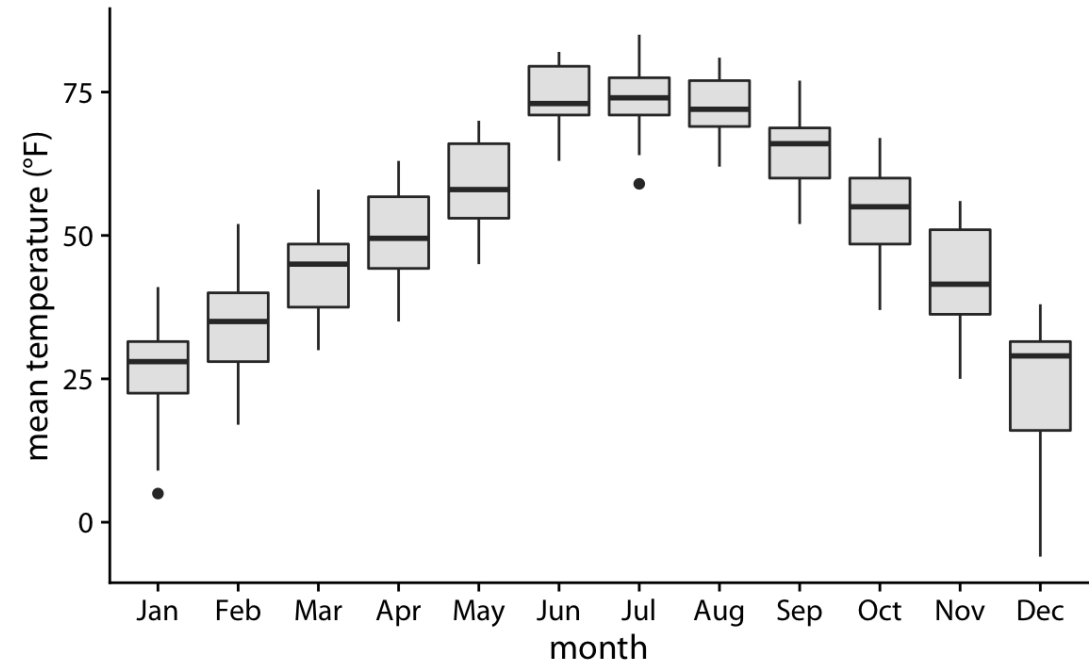


Một số loại đồ thị thông thường

Biểu đồ tần suất - Histogram



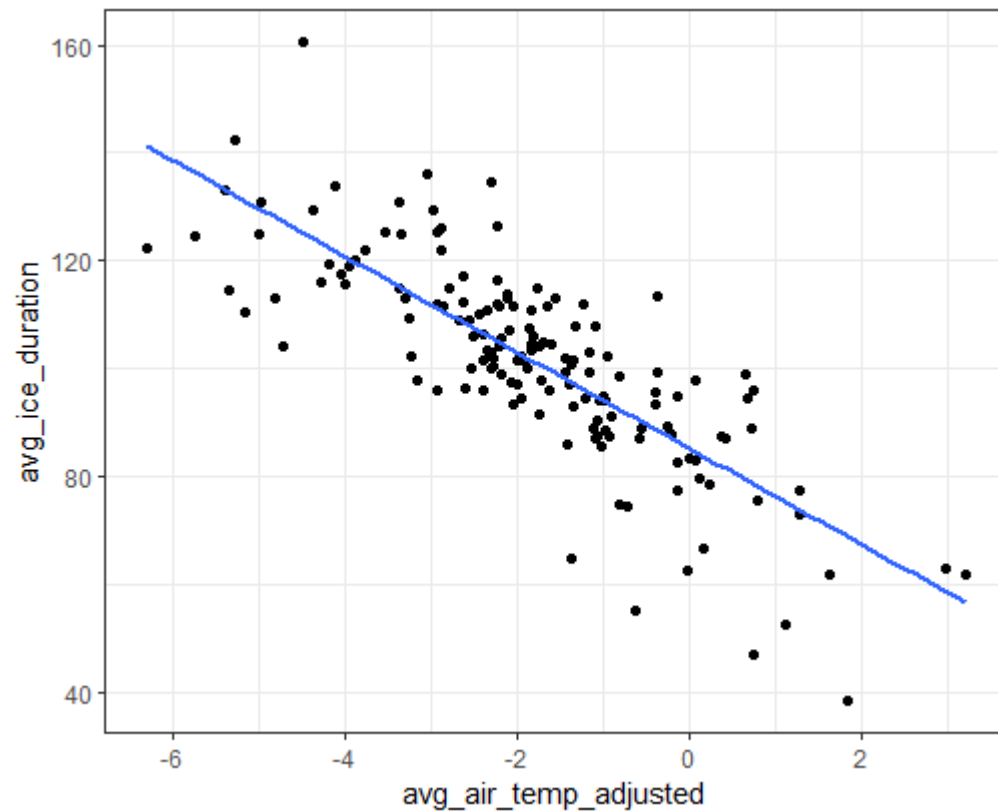
Đồ thị hộp – Box plots



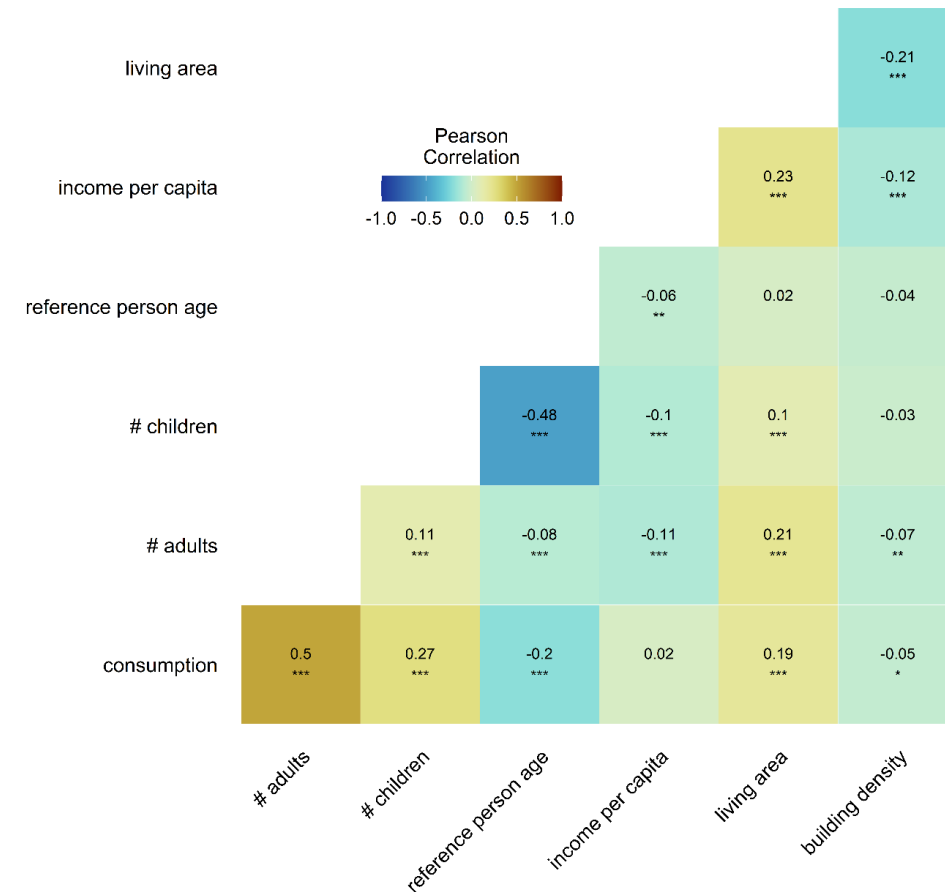
Source Wilke (2018)

Một số loại đồ thị thông thường

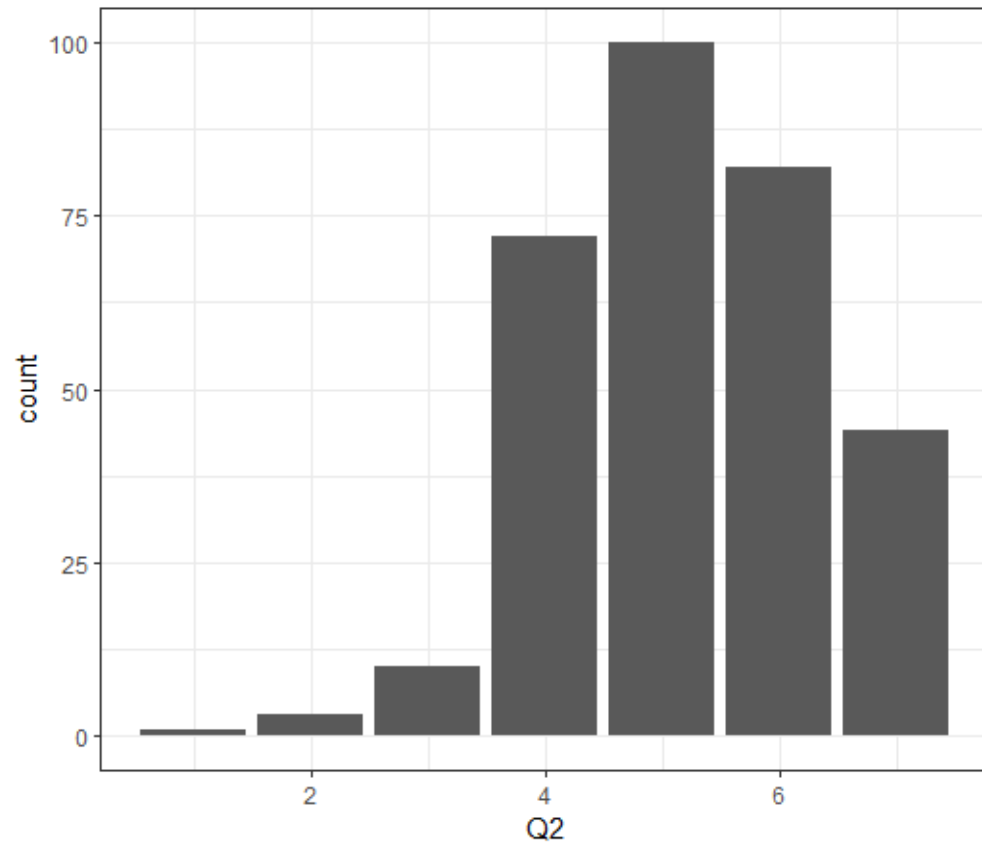
Biểu đồ phân tán - Scatter plots



Bản đồ nhiệt - heatmap



Thống kê mô tả - biến rời rạc



Question 2

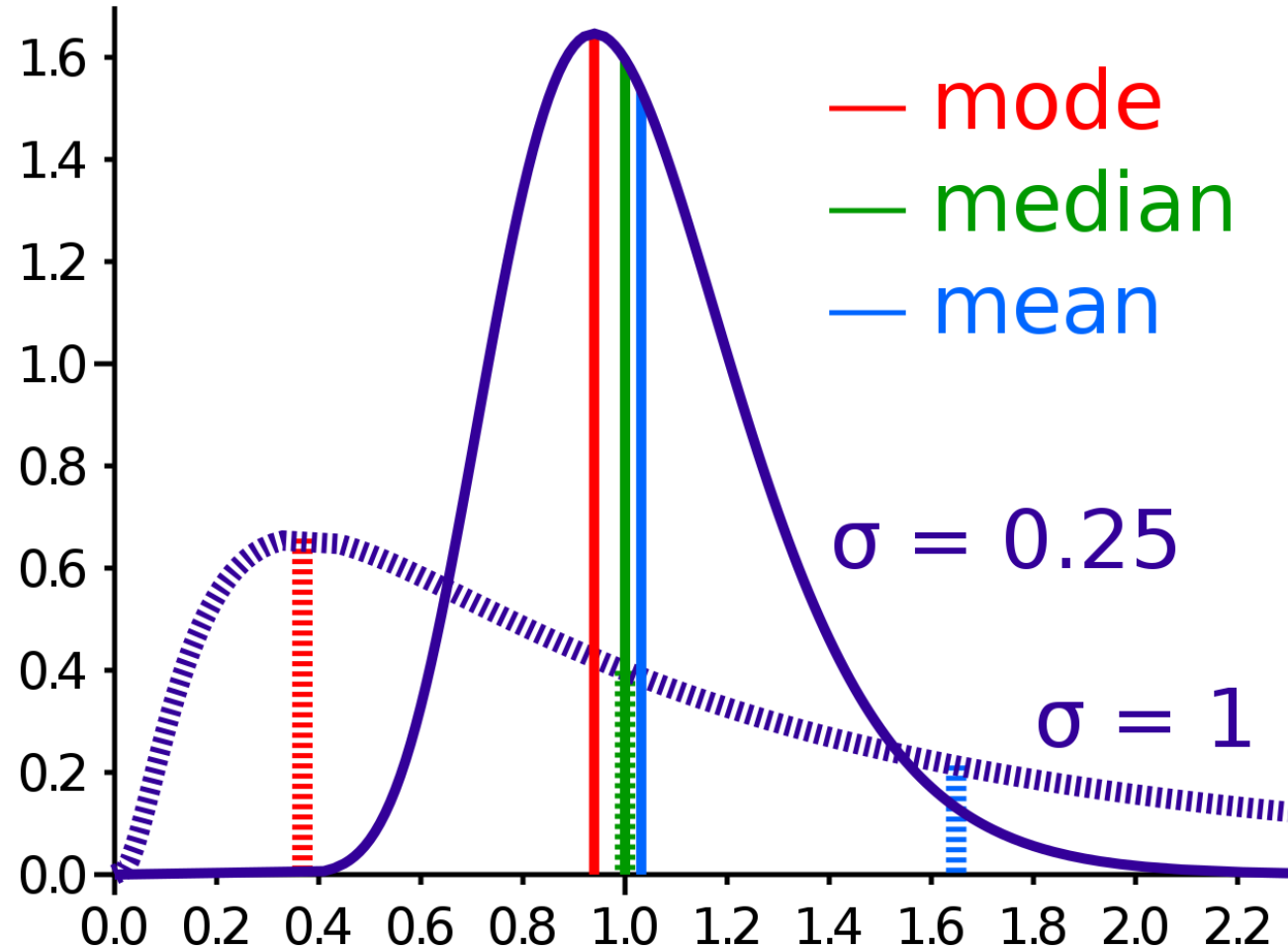
I actively practice environmental sustainability at home (e.g., energy conservation, recycling).

	Q2	count	percent
1	1	1	0.3205128
2	2	3	0.9615385
3	3	10	3.2051282
4	4	72	23.0769231
5	5	100	32.0512821
6	6	82	26.2820513
7	7	44	14.1025641

Thống kê mô tả - Biến liên tục

Đo độ tập trung

Trung bình, trung vị, mode

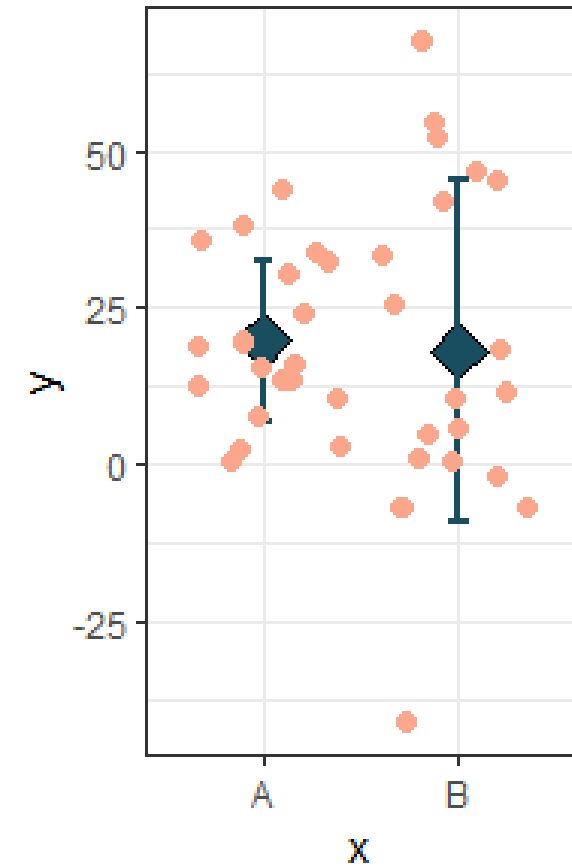
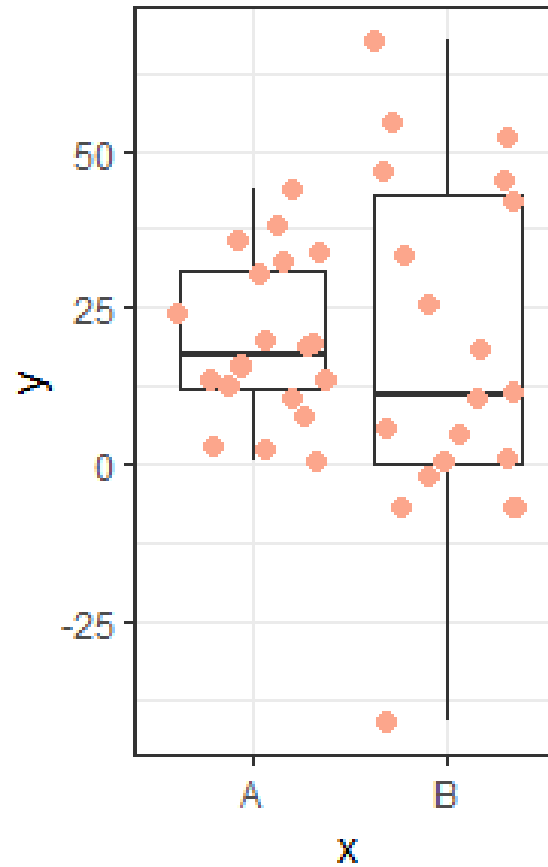


[This Photo](#) by Unknown Author is licensed under [CC BY-SA](#)

Thống kê mô tả - Biến liên tục

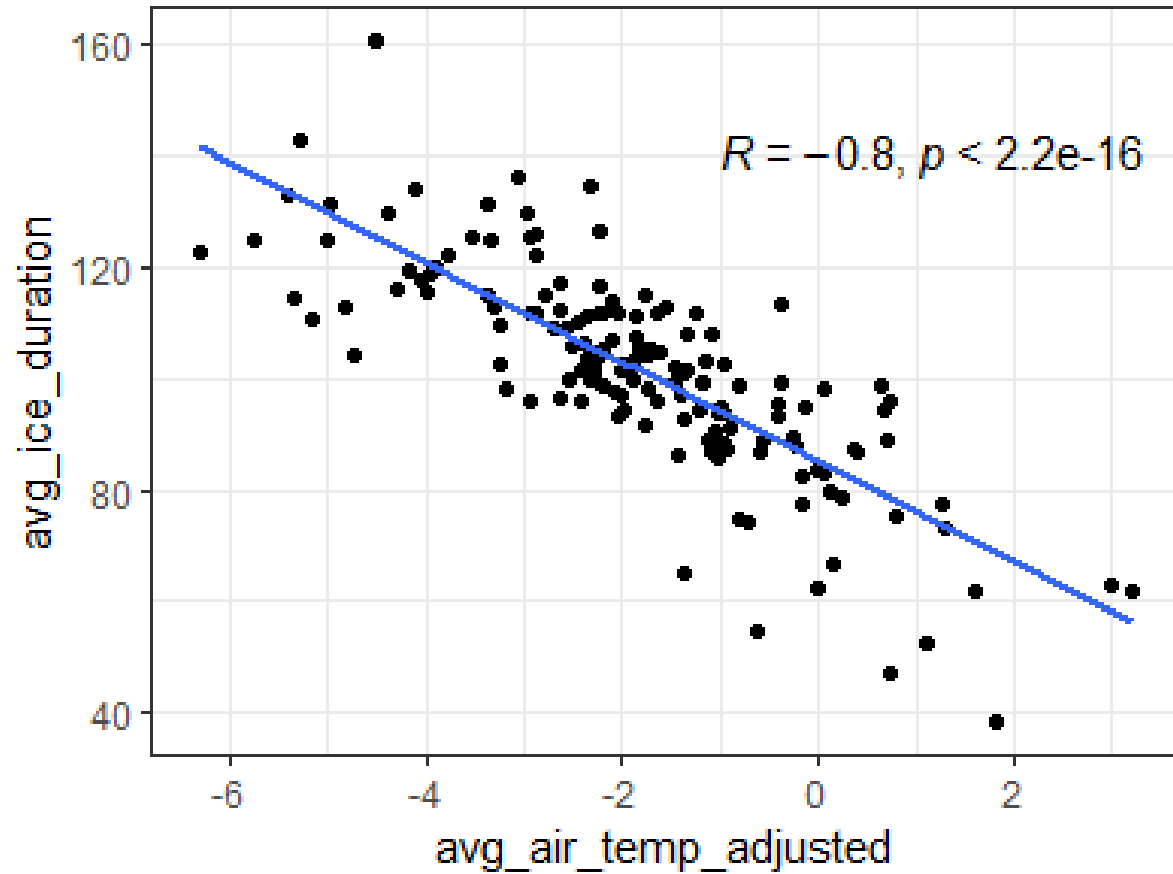
Đo độ phân tán

- Phương sai (Variance)
- Độ lệch chuẩn (Standard deviation)
- Phân vị (Quantile)
- Điểm tứ phân vị (Quartile)
- Giá trị tối thiểu/tối đa (min/max)

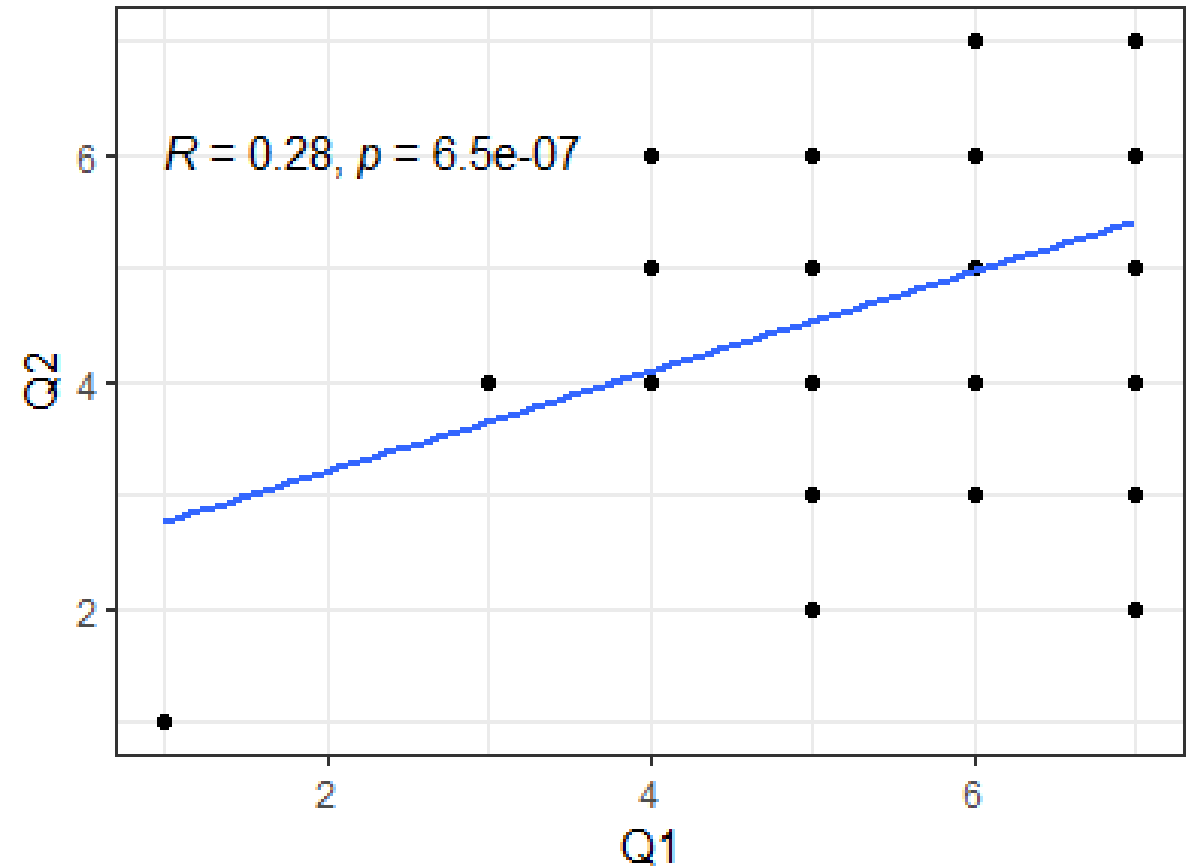


Thống kê mô tả - tương quan

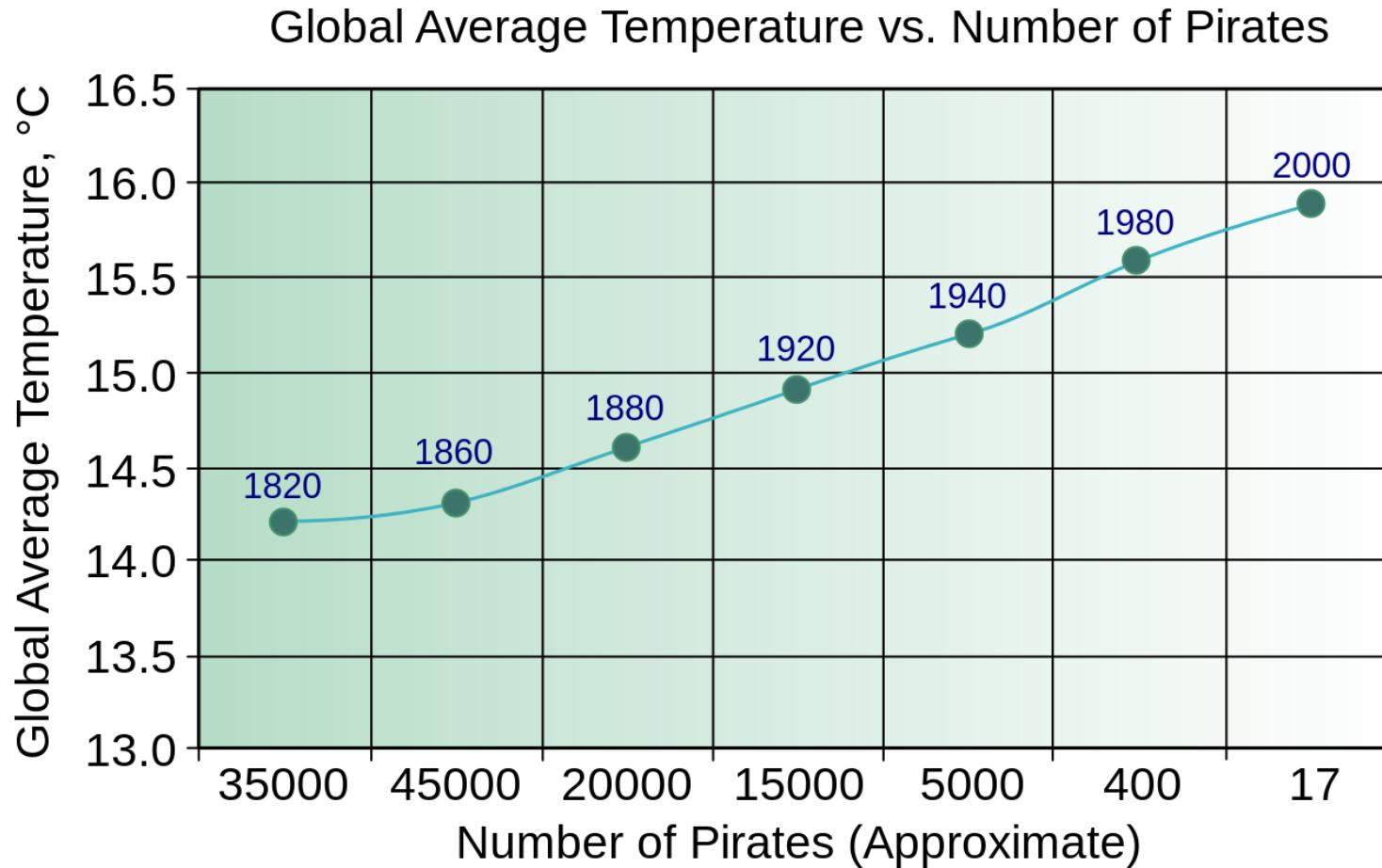
Pearson



Spearman



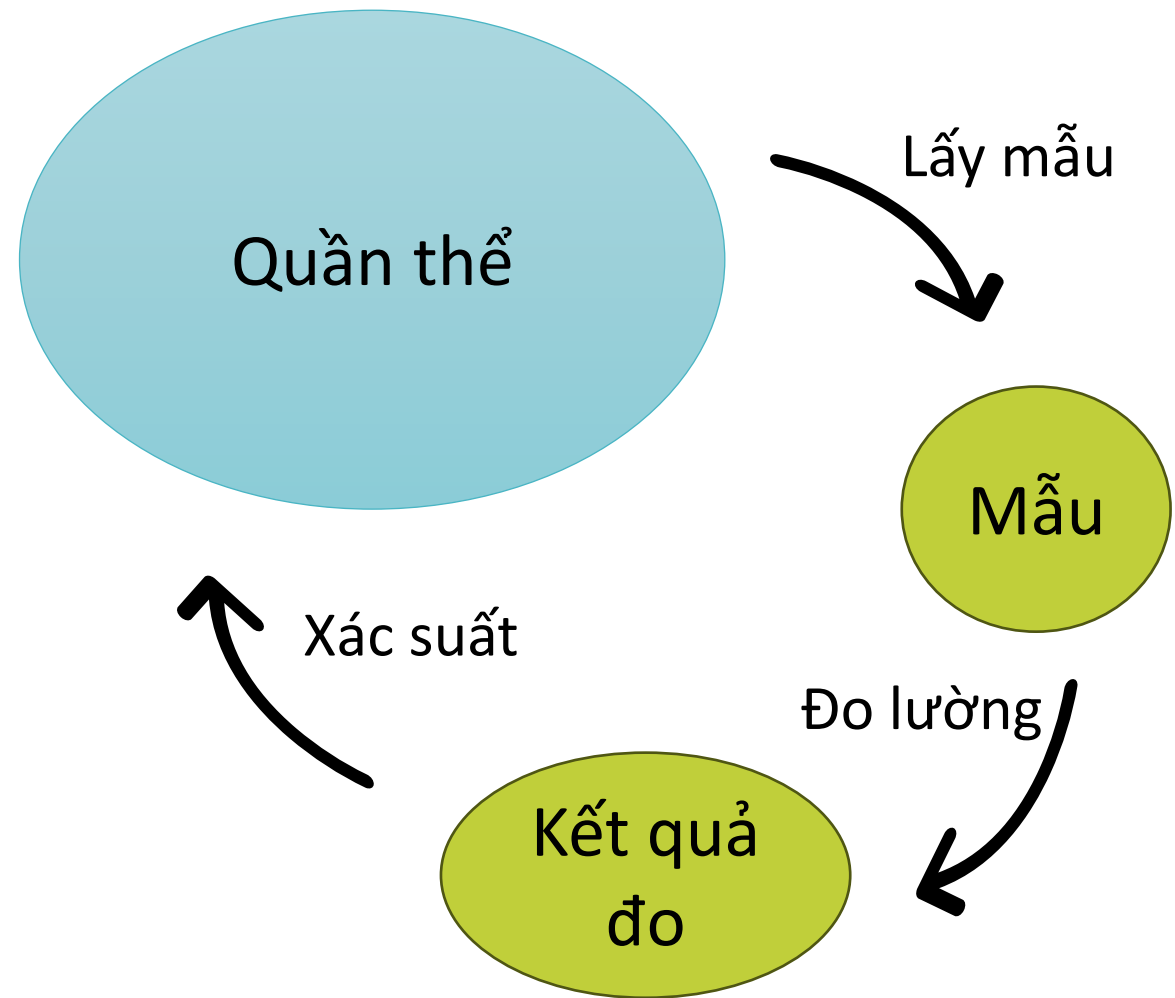
Tương quan và quan hệ nhân quả



Phân tích dữ liệu khẳng định

Thống kê suy luận

- Dùng thông tin về mẫu để suy luận về quần thể
- Kiểm định các giả thuyết (hypothesis) nghiên cứu
- Đưa ra kết luận về mối quan hệ giữa các biến



Kiểm định thống kê (Hypothesis testing)

- Giả thuyết trống/không (Null hypothesis)
- Giả thuyết thay thế (Alternative hypothesis)

$$H_0: \mu = 0$$

$$H_a: \mu \neq 0$$

- Giả thuyết thú vị với nhà nghiên cứu luôn là giả thuyết thay thế
- Giả thuyết được kiểm định luôn là giả thuyết không/trống

Bài tập

- Mức độ tham gia của cộng đồng trong việc bảo tồn các di sản vật thể ảnh hưởng tích cực đến cảm giác về nơi chốn và bản sắc của cư dân địa phương
- Việc thêm các bài giảng về di sản phi vật thể vào trong trường học giúp tăng cường đối thoại và sự tôn trọng đối với các dân tộc khác nhau
- Việc đô thị hóa quá nhanh dẫn đến gia tăng chênh lệch giàu nghèo tại các đô thị

Giá trị p

- Xác suất để thu được kết quả tương tự hoặc cực đoan hơn khi giả thiết rằng giả thuyết trống là đúng
- Giá trị $p < 0.05$: có ý nghĩa về mặt thống kê

	H_0 đúng	H_0 sai
Bác bỏ H_0	Lỗi loại I	✓
Không bác bỏ H_0	✓	Lỗi loại II

- Ý nghĩa về mặt thống kê vs. ý nghĩa thực tế

Kiểm định t (t-test)

- So sánh 2 giá trị trung bình

$$H_0: \mu_1 = \mu_2$$

$$H_a: \mu_1 \neq \mu_2$$

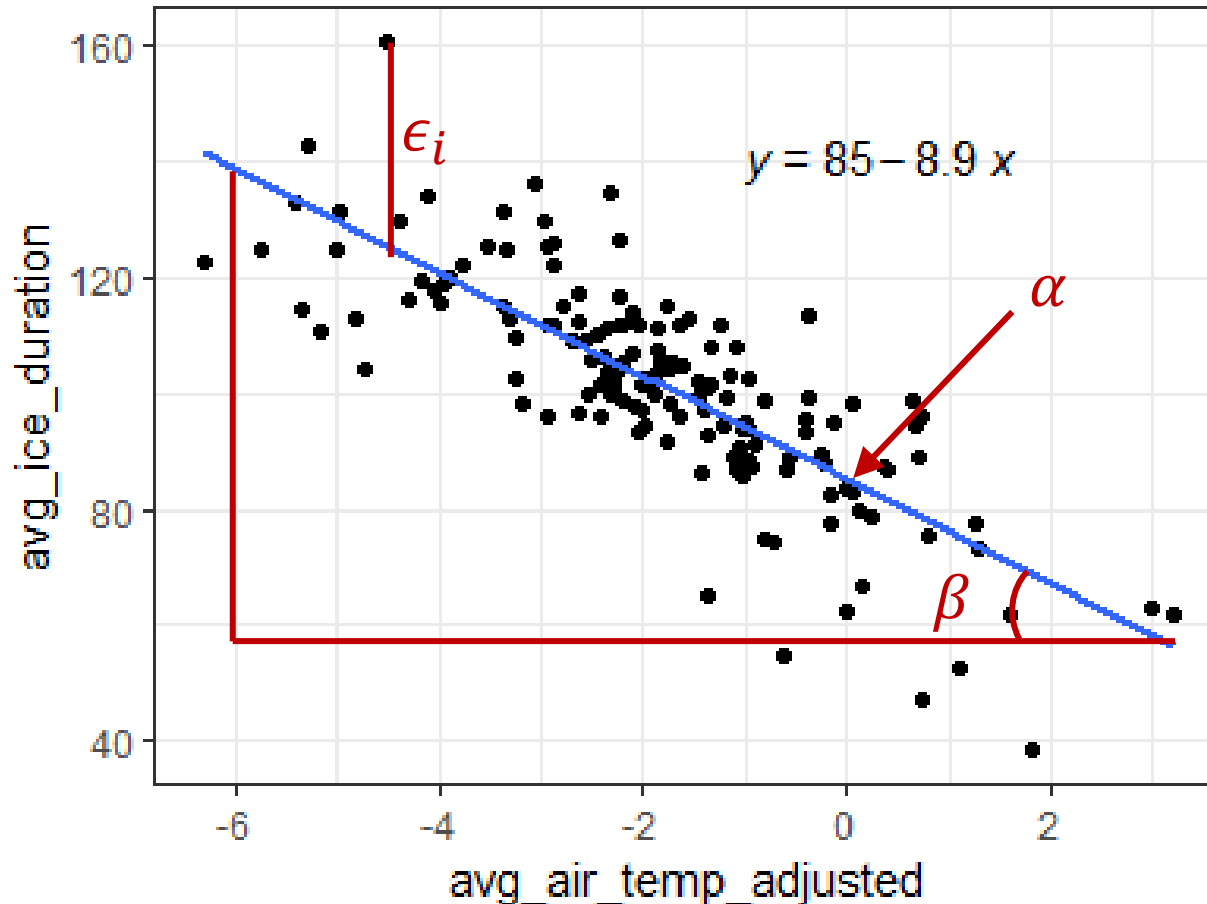
- Giả định
 - Liên tục/định lượng
 - Độc lập
 - Phân phối chuẩn
 - Độ phân tán tương đương
- Trường hợp đặc biệt: Kiểm định t ghép cặp

Giới thiệu các kỹ thuật phân tích định lượng

Một số kỹ thuật định lượng

- Hồi quy tuyến tính đơn giản (Simple linear regression)
- Hồi quy tuyến tính suy rộng (Generalized linear regression)
- Mô hình đa cấp/Mô hình ảnh hưởng hỗn hợp (Multilevel model/Mixed effects model)
- Phân tích nhân tố (Factor analysis)
- Phân tích thành phần chính (PCA - Principal component analysis)
- Mô hình phương trình cấu trúc (SEM – Structural equation model)
- Thống kê không gian (Spatial statistics)
- ...

Hồi quy tuyến tính đơn giản



$$Y_i = \alpha + \beta X_i + \epsilon_i$$
$$\epsilon_i \sim N(0, \sigma^2)$$

- Y biến phụ thuộc (dependent/out come variable)
- X biến độc lập (independent/explanatory variable, predictor)
- ϵ phần dư, sai số ngẫu nhiên (residuals, random errors)

Hồi quy tuyến tính suy rộng (Generalized linear regression)

$$Y_i = \alpha + \beta_1 X_{i1} + \beta_1 X_{i1}^2 + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} + \epsilon_i$$
$$\epsilon_i \sim N(0, \sigma^2)$$

Hồi quy tuyến tính bội

Y : liên tục/định lượng

X : liên tục/định lượng hoặc rời rạc

Khi Y là biến rời rạc?

Hồi quy tuyến tính suy rộng (Generalized linear regression)

$$Y_i = \alpha + \beta_1 X_{i1} + \beta_1 X_{i1}^2 + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} + \epsilon_i$$
$$\epsilon_i \sim N(0, \sigma^2)$$

Hồi quy tuyến tính bội

Y : liên tục/định lượng

X : liên tục/định lượng hoặc rời rạc

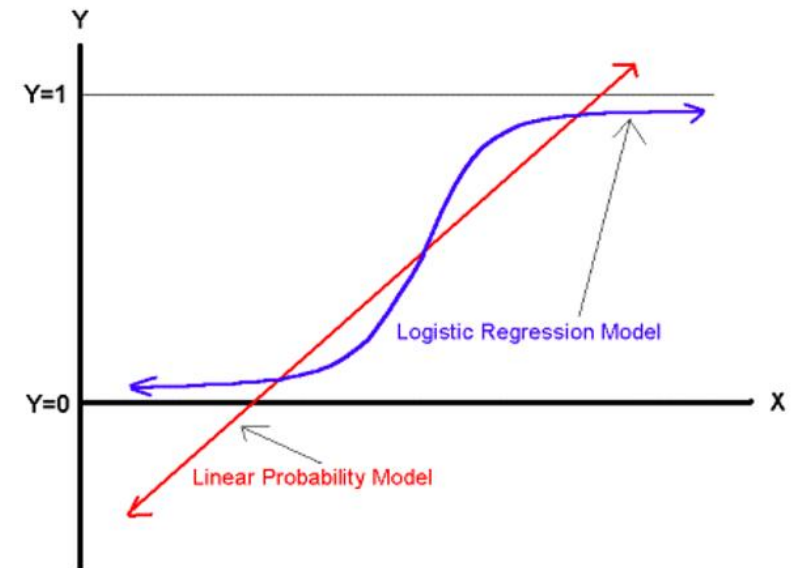
Khi Y là biến rời rạc?

Nhị phân (Yes/No): Hồi quy Logistic (Logistic regression)

Định danh: Multinomial logistic regression

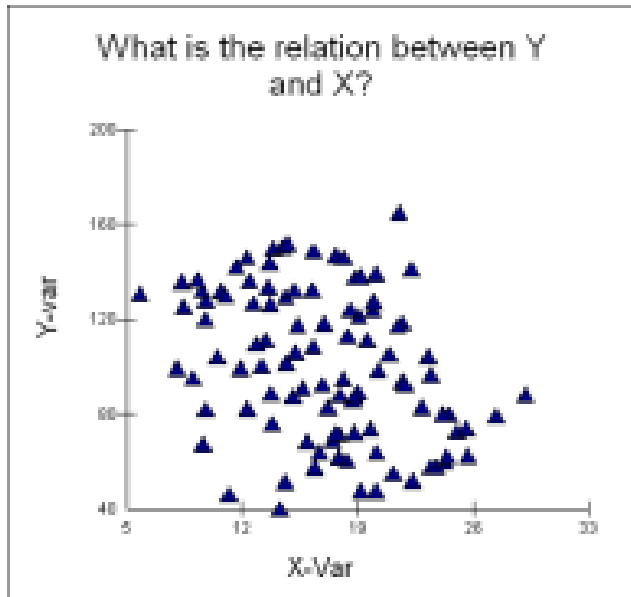
Thứ bậc: Cumulative logistic regression

Biến đếm: Poisson regression



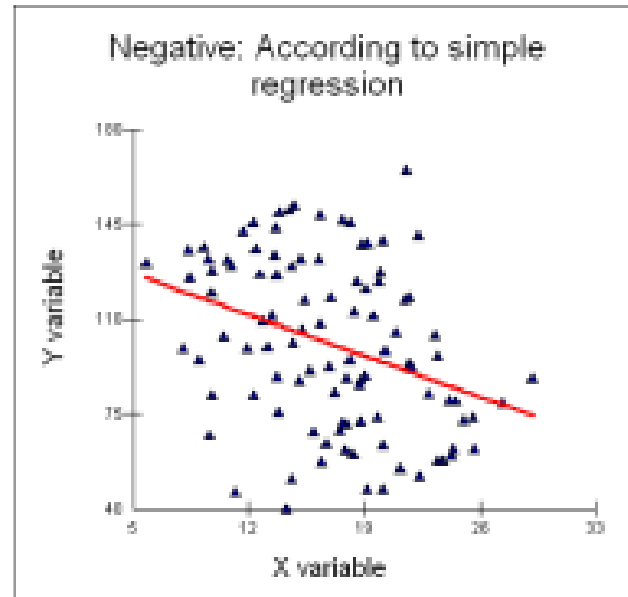
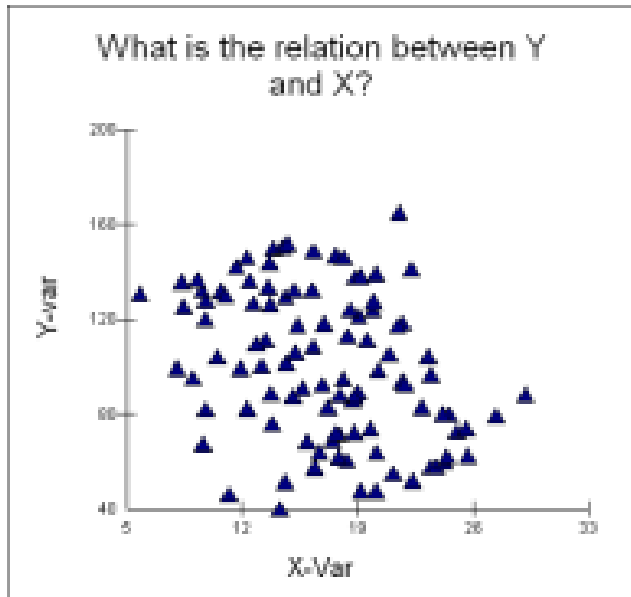
Mô hình đa cấp/Mô hình ảnh hưởng hỗn hợp

Multilevel model/Mixed effect model



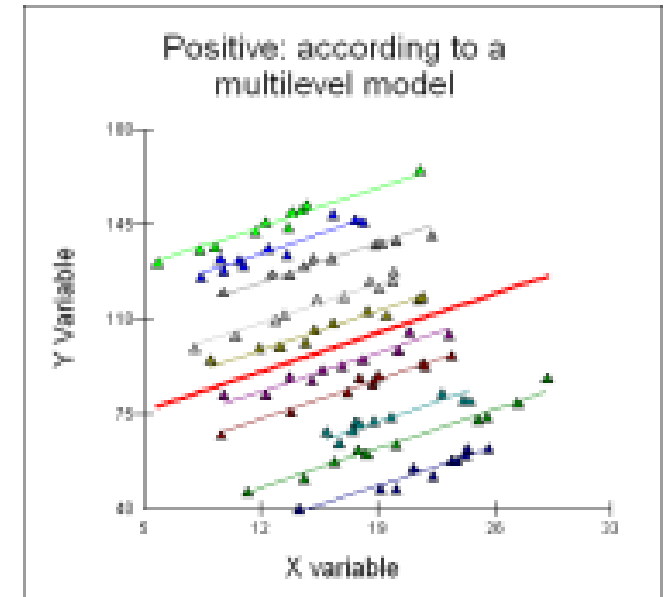
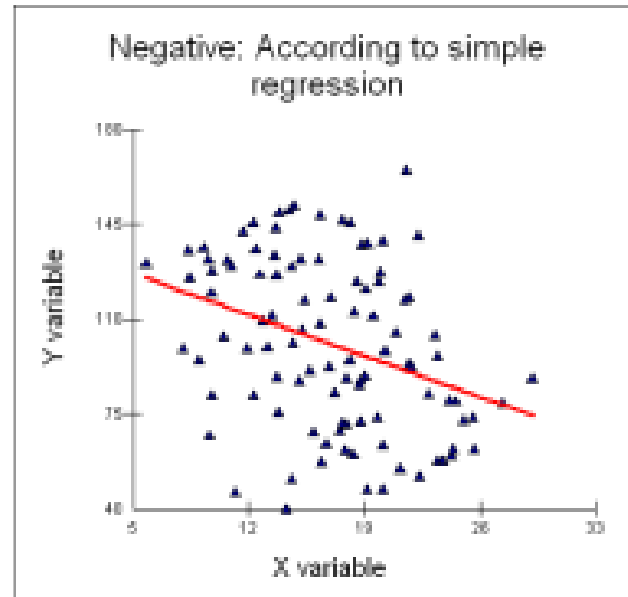
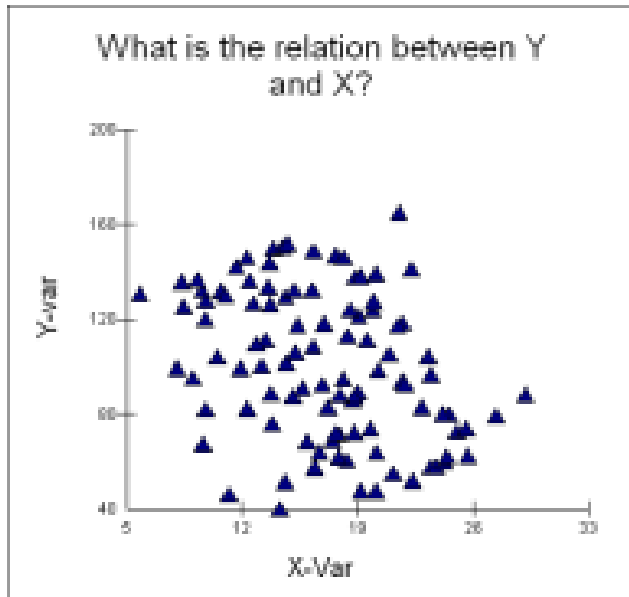
Mô hình đa cấp/Mô hình ảnh hưởng hỗn hợp

Multilevel model/Mixed effect model



Mô hình đa cấp/Mô hình ảnh hưởng hỗn hợp

Multilevel model/Mixed effect model

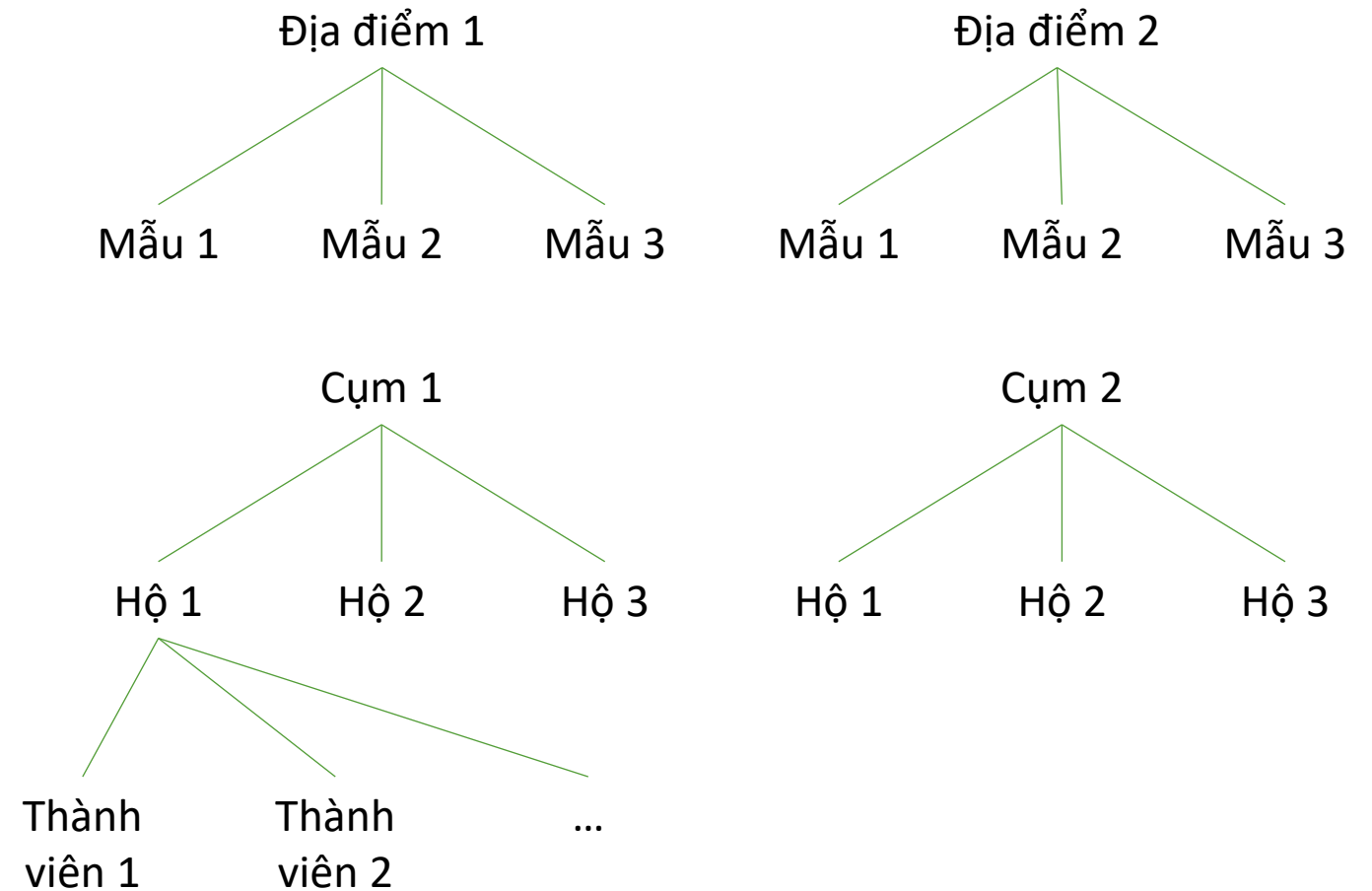


Mô hình đa cấp/Mô hình ảnh hưởng hỗn hợp

Multilevel model/Mixed effect model

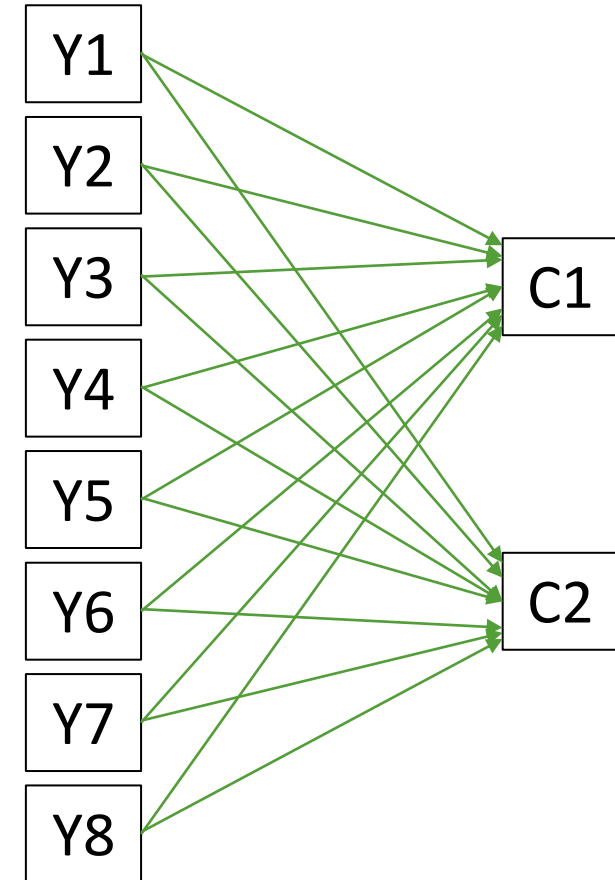


Image by [Chelsea Parlett-Pelleriti](#)



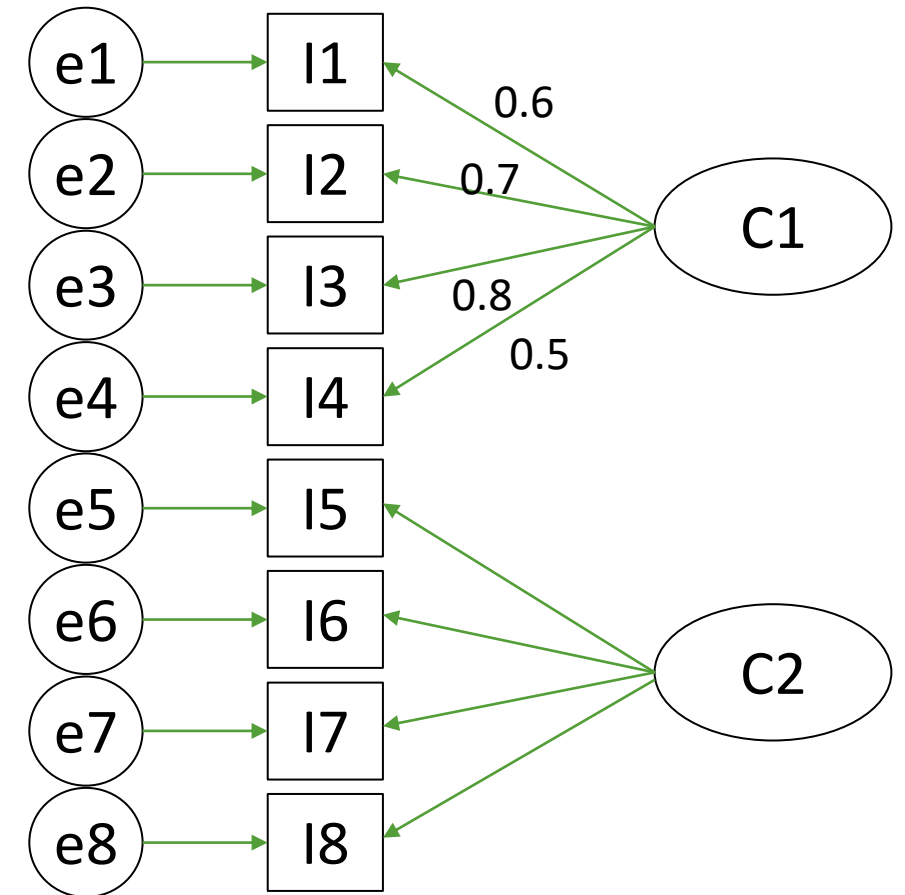
Phân tích thành phần chính (PCA - Principal component analysis)

- Phương pháp giảm chiều dữ liệu
- Không phân biệt biến độc lập hay phụ thuộc
- Phương pháp khảo sát (không phải phương pháp suy luận)
- Bước trước cho hồi quy tuyến tính để giảm đa cộng tuyến (multicollinearity)



Phân tích nhân tố (Factor analysis)

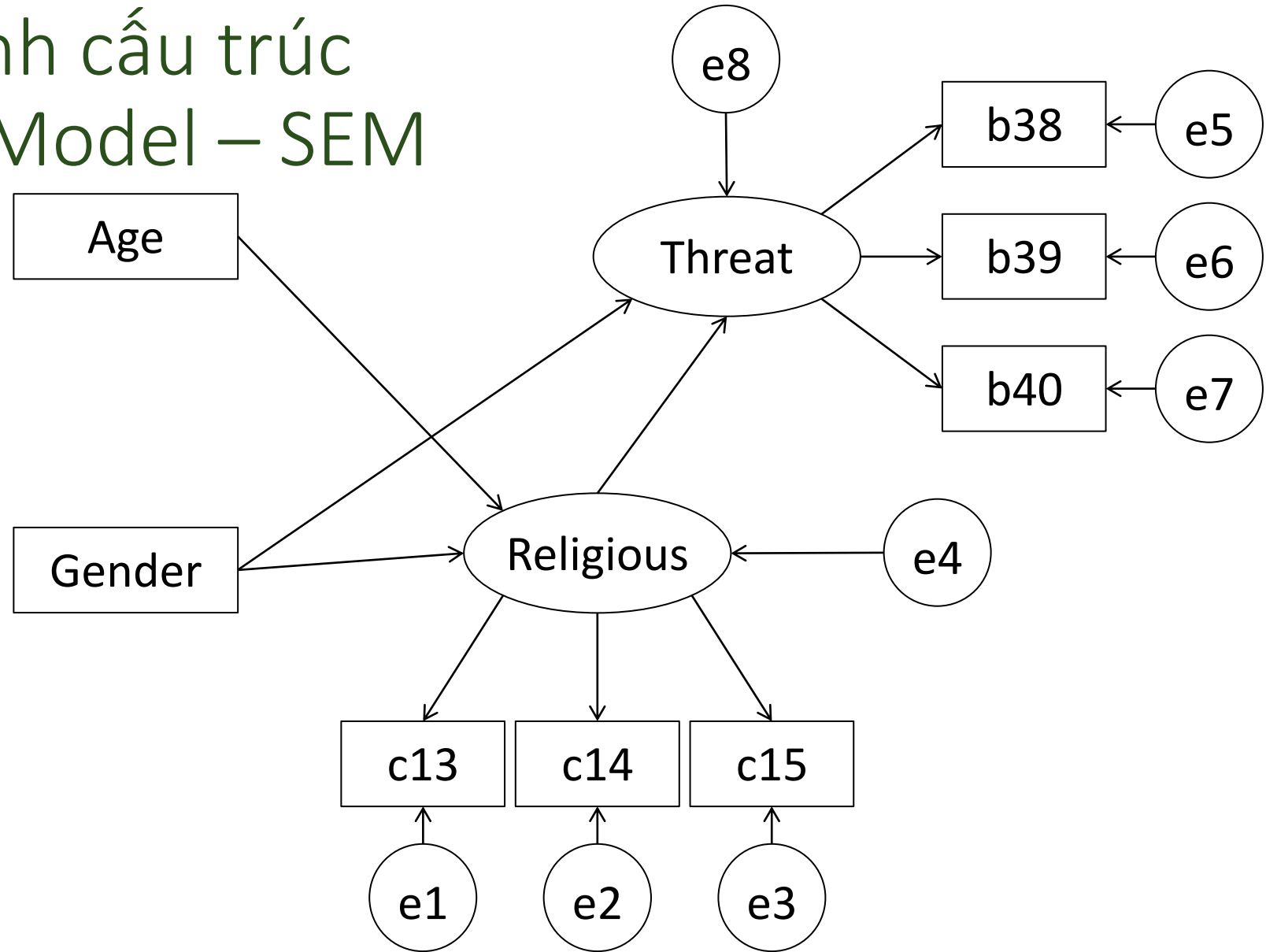
- Thường áp dụng cho dữ liệu bảng hỏi
- Đo phạm trù tiềm ẩn (Latent construct)
- Phân tích nhân tố khám phá/khẳng định (Exploratory/Confirmatory Factor Analysis)



Mô hình phương trình cấu trúc

Structural Equation Model – SEM

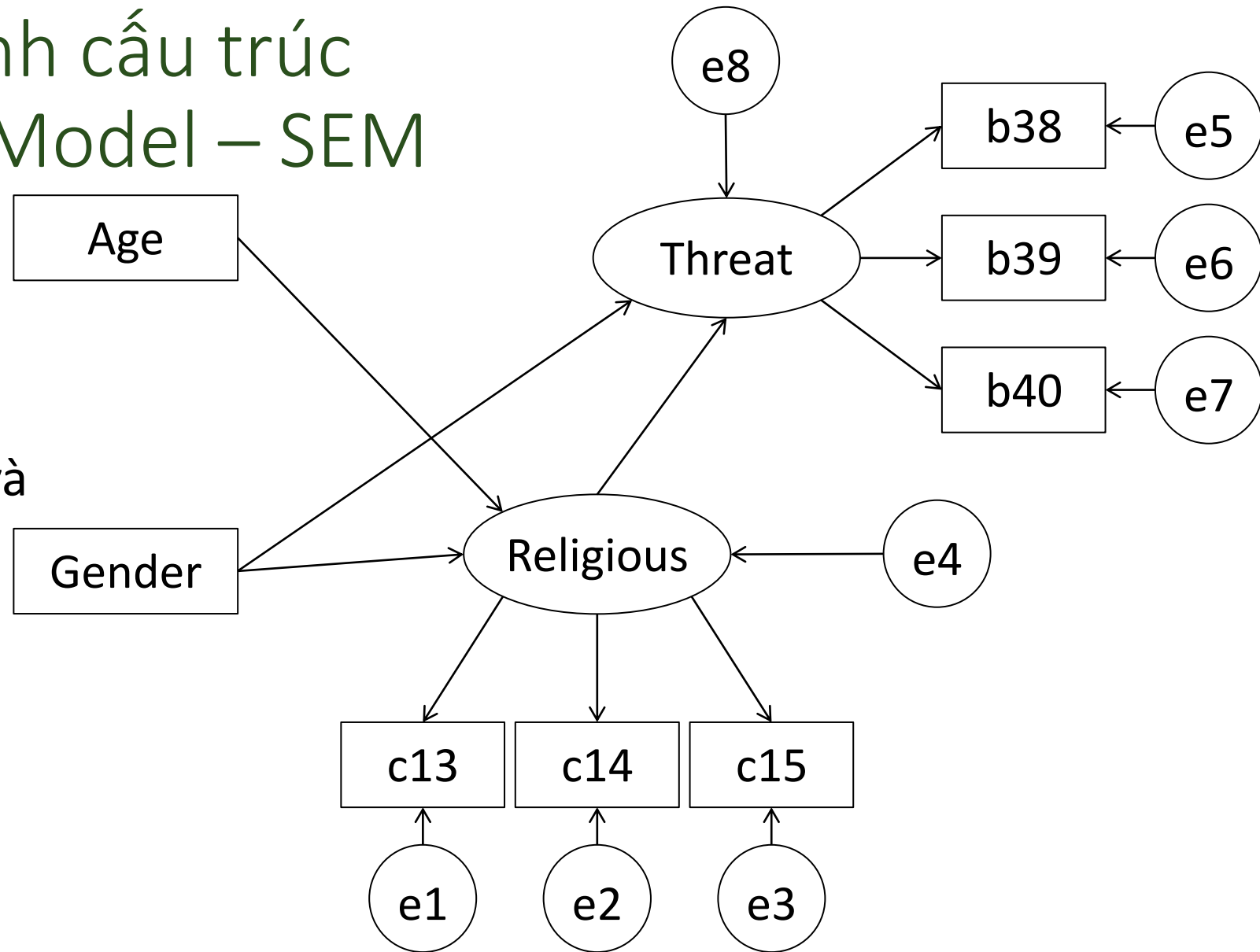
- Tiềm ẩn (latent) và biểu hiện (manifest)



Mô hình phương trình cấu trúc

Structural Equation Model – SEM

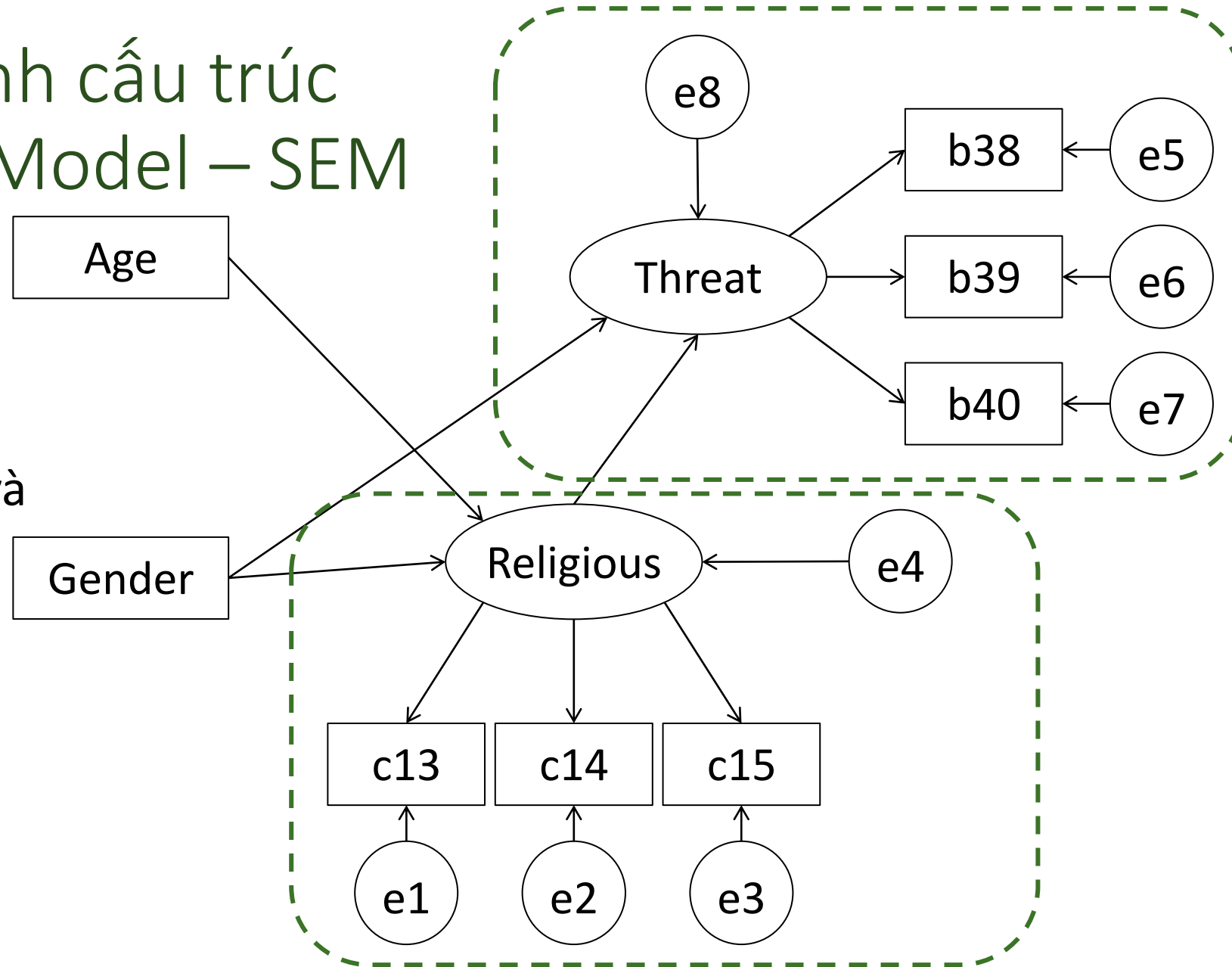
- Tiềm ẩn (latent) và biểu hiện (manifest)
- Nội sinh (endogenous) và ngoại sinh (exogenous)



Mô hình phương trình cấu trúc

Structural Equation Model – SEM

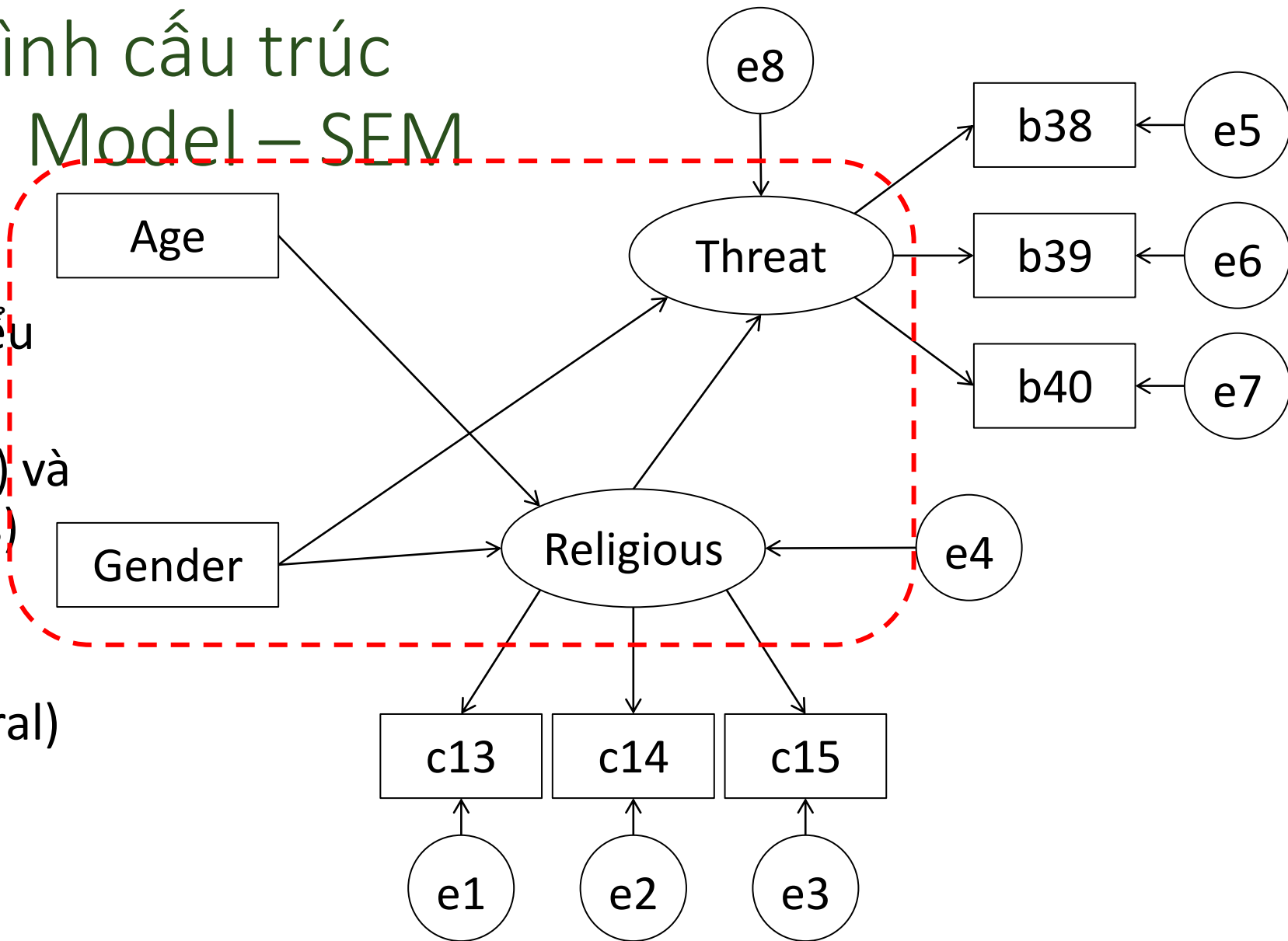
- Tiềm ẩn (latent) và biểu hiện (manifest)
- Nội sinh (endogenous) và ngoại sinh (exogenous)
- Mô hình đo lường (measurement)



Mô hình phương trình cấu trúc

Structural Equation Model – SEM

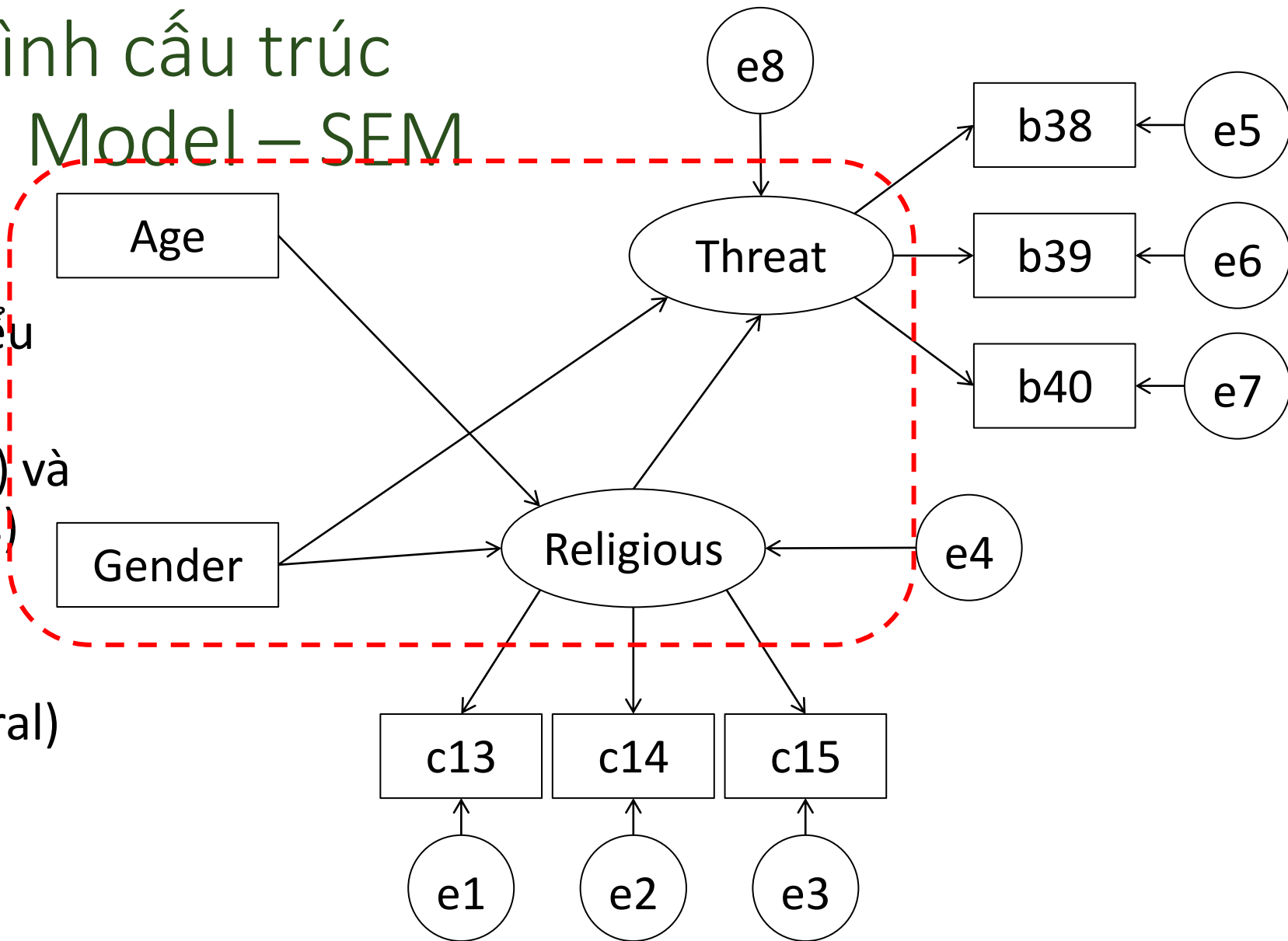
- Tiềm ẩn (latent) và biểu hiện (manifest)
- Nội sinh (endogenous) và ngoại sinh (exogenous)
- Mô hình đo lường (measurement) và mô hình cấu trúc (structural)



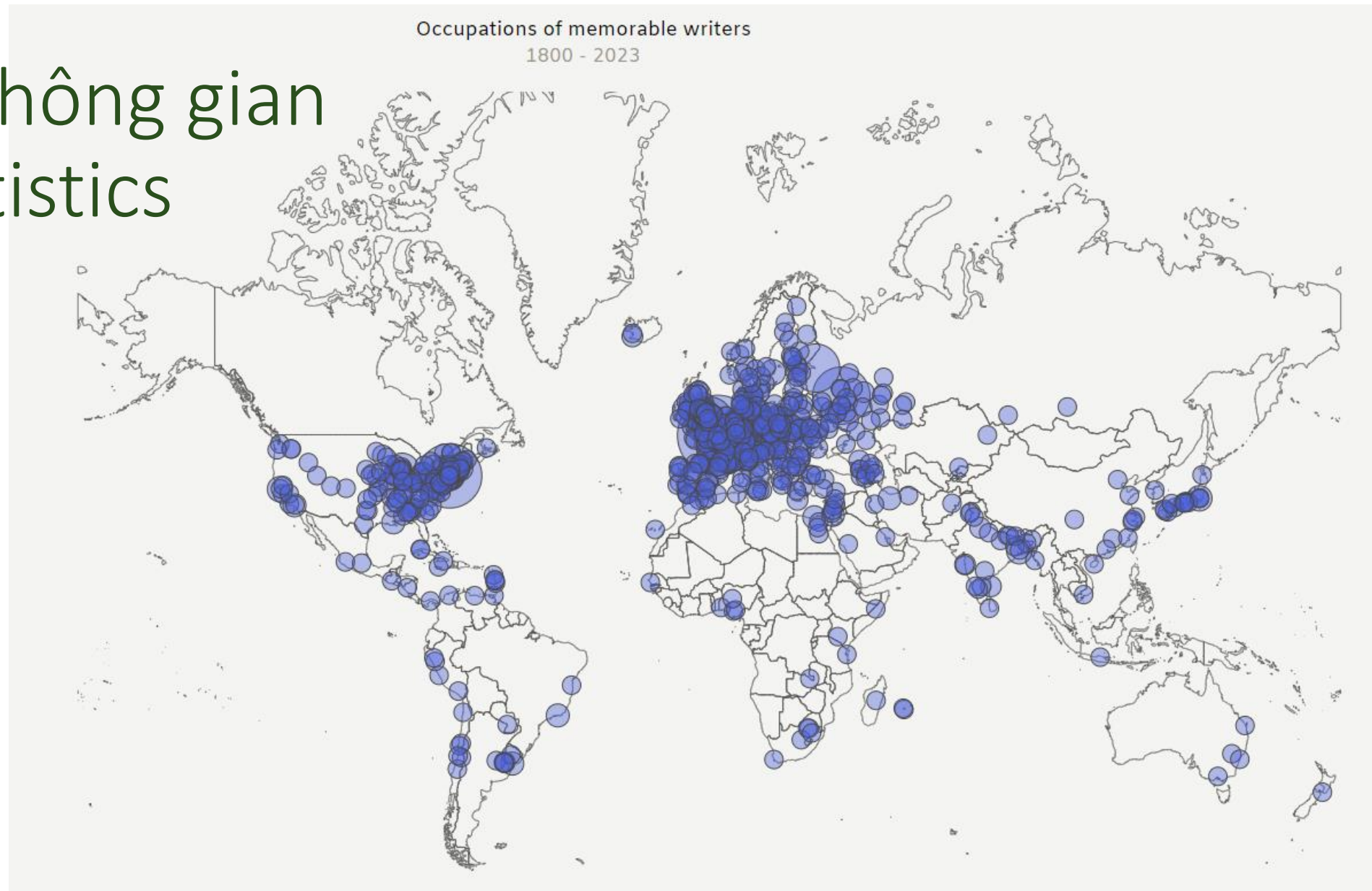
Mô hình phương trình cấu trúc

Structural Equation Model – SEM

- Tiềm ẩn (latent) và biểu hiện (manifest)
- Nội sinh (endogenous) và ngoại sinh (exogenous)
- Mô hình đo lường (measurement) và mô hình cấu trúc (structural)
- Tác động trực tiếp và gián tiếp (Direct vs indirect effects)



Thống kê không gian Spatial Statistics



Bài tập

- Với dữ liệu dự kiến của bạn phương pháp phân tích phù hợp là gì?