

CHƯƠNG 4. BIẾN SỐ VÀ KỸ THUẬT THU THẬP SỐ LIỆU

4.1. Khái niệm biến số trong nghiên cứu y học

Trong thống kê y học, **biến số (Variable)** là **đặc điểm hoặc thuộc tính có thể thay đổi giữa các cá thể, nhóm, hoặc theo thời gian**, có thể đo lường, phân loại hoặc ghi nhận được [1].

Ví dụ:

- Huyết áp tâm thu (mmHg)
- Giới tính (nam/nữ)
- Tình trạng thiếu máu (có/không)
- Tuổi (năm)

Một **nghiên cứu khoa học y học** luôn xoay quanh việc xác định, đo lường và phân tích các biến số để kiểm định giả thuyết hoặc mô tả hiện tượng.

4.2. Phân loại biến số

Có nhiều cách phân loại biến số. Cách phổ biến nhất là dựa theo **bản chất dữ liệu** (định tính/định lượng) và **vai trò trong mô hình nghiên cứu** (độc lập/phụ thuộc).

4.2.1. Phân loại theo bản chất dữ liệu

Bảng 12. Phân loại biến số theo bản chất dữ liệu

Loại biến	Đặc điểm	Ví dụ	Cách mã hóa trong SPSS
Biến định tính (Qualitative)	Không thể đo bằng số, chỉ phân loại nhóm	Giới tính (Nam/Nữ), nghề nghiệp, nhóm máu	1 = Nam, 2 = Nữ
Biến định lượng (Quantitative)	Đo được bằng số, có thể tính trung bình, độ lệch chuẩn	Cân nặng (kg), huyết áp (mmHg), tuổi (năm)	Ghi giá trị thực
Biến thứ bậc (Ordinal)	Có thứ tự, khoảng cách không đều	Mức độ đau (nhẹ, vừa, nặng), học lực (Yếu → Giỏi)	1, 2, 3, 4
Biến nhị phân (Binary)	Chỉ có 2 giá trị (có/không, đúng/sai)	Thiếu máu (Có/Không), Tiêm vắc xin (Có/Không)	0 = Không, 1 = Có

4.2.2. Phân loại theo vai trò trong nghiên cứu

Bảng 13. Phân loại biến theo vai trò

Vai trò	Khái niệm	Ví dụ
Biến độc lập (Independent Variable)	Yếu tố có thể ảnh hưởng đến kết quả	Tuổi, giới, chế độ ăn
Biến phụ thuộc (Dependent Variable)	Kết quả hoặc hiện tượng chịu ảnh hưởng	Tình trạng thiếu máu, huyết áp
Biến nhiễu (Confounding Variable)	Yếu tố ngoài ý muốn ảnh hưởng đến mối liên hệ	Thu nhập, trình độ học vấn
Biến trung gian (Mediating Variable)	Trung gian giữa nguyên nhân và kết quả	Hành vi ăn uống giữa thu nhập và béo phì

4.3. Thang đo biến số (Scale of Measurement)

Thang đo là **mức độ đo lường biến số**, quyết định loại phép thống kê được sử dụng trong SPSS. Có bốn thang đo chính theo Stevens (1946) [2]:

Bảng 14. Bốn loại thang đo biến số

Thang đo	Đặc điểm	Phép toán có thể thực hiện	Ví dụ
Danh nghĩa (Nominal)	Chỉ phân loại, không thứ tự	Đếm, tần suất, tỷ lệ	Giới tính, nhóm máu
Thứ bậc (Ordinal)	Có thứ tự nhưng khoảng cách không đều	Xếp hạng, trung vị	Mức độ đau (1–3)
Khoảng (Interval)	Có thứ tự, khoảng cách đều, không có gốc 0 tuyệt đối	Cộng, trừ, trung bình	Nhiệt độ °C
Tỷ lệ (Ratio)	Có thứ tự, khoảng cách đều, có gốc 0 tuyệt đối	Mọi phép toán	Cân nặng (kg), chiều cao (cm)

Lưu ý:

Khi phân tích bằng SPSS, việc chọn đúng loại thang đo rất quan trọng.

Ví dụ: thang đo “Ordinal” thì nên dùng **Kiểm định Mann–Whitney** thay vì t-test.

4.4. Xây dựng và mã hóa biến số trong SPSS

4.4.1. Định nghĩa biến

Trong **SPSS**, mỗi biến được định nghĩa trong “Variable View” với các trường thông tin sau:

Bảng 15. Các thành phần định nghĩa biến trong SPSS

Trường	Ý nghĩa	Ví dụ
Name	Tên biến (viết liền, không dấu)	Gioitinh, Tuoi
Type	Kiểu dữ liệu	Numeric, String
Width / Decimals	Độ rộng và số chữ số thập phân	8 / 0
Label	Nhãn mô tả biến	“Giới tính của đối tượng”
Values	Mã hóa giá trị	1 = Nam, 2 = Nữ
Missing	Mã trị số bị thiếu	99
Measure	Thang đo	Nominal / Ordinal / Scale

4.4.2. Mã hóa biến

Ví dụ 4.1:

Biến “Giới tính” được mã hóa như sau:

Nhãn Giá trị mã hóa

Nam 1

Nữ 2

Khi nhập dữ liệu, ta chỉ cần gõ “1” hoặc “2”. SPSS sẽ hiển thị nhãn tương ứng khi chọn chế độ “Value Labels”.

4.5. Kỹ thuật thu thập số liệu trong nghiên cứu y học

Thu thập số liệu là giai đoạn trọng yếu, ảnh hưởng trực tiếp đến **độ chính xác và độ tin cậy** của kết quả nghiên cứu.

Tùy loại nghiên cứu (mô tả, phân tích, can thiệp), ta có các kỹ thuật khác nhau.

4.5.1. Các phương pháp thu thập số liệu

Bảng 16. Các phương pháp thu thập số liệu phổ biến

Phương pháp	Đặc điểm	Ưu điểm	Hạn chế
Phỏng vấn trực tiếp	Người nghiên cứu hỏi – ghi câu trả lời	Kiểm soát tốt, tỷ lệ trả lời cao	Dễ thiên lệch người phỏng vấn
Bảng câu hỏi tự điền (Survey)	Đối tượng tự ghi câu trả lời	Tiết kiệm thời gian	Khó kiểm soát sai sót
Quan sát (Observation)	Ghi nhận hành vi, triệu chứng	Thích hợp lâm sàng	Chủ quan, tồn thời gian
Khai thác hồ sơ bệnh án	Dữ liệu sẵn có	Tiết kiệm, ít tốn công	Dễ thiếu thông tin

Phương pháp	Đặc điểm	Ưu điểm	Hạn chế
Đo lường sinh học (Biochemical measurement)	Thu thập thông tin khách quan	Chính xác cao	Chi phí cao, yêu cầu kỹ thuật

4.5.2. Công cụ thu thập số liệu

- **Bảng câu hỏi (Questionnaire):**
 - Gồm các phần: thông tin chung, yếu tố nguy cơ, kết quả sức khỏe.
 - Câu hỏi nên rõ ràng, tránh gợi ý, hạn chế câu hỏi nhạy cảm.
- **Phiếu thu thập số liệu lâm sàng (CRF – Case Record Form):**
 - Dành cho nghiên cứu bệnh viện hoặc thử nghiệm lâm sàng.
- **Thiết bị đo lường:** cân, máy đo huyết áp, máy đo Hb, phần mềm theo dõi.

4.6. Thiết kế bảng câu hỏi

4.6.1. Cấu trúc bảng câu hỏi

1. **Phần mở đầu:** Giới thiệu mục đích, cam kết bảo mật, hướng dẫn điền.
2. **Phần nội dung:**
 - Thông tin nhân khẩu học (tuổi, giới, nghề, học vấn).
 - Câu hỏi chuyên đề (yếu tố nguy cơ, hành vi, triệu chứng).
3. **Phần kết thúc:** Lời cảm ơn và xác nhận đồng ý tham gia.

Bảng 17. Trích bảng câu hỏi nghiên cứu thiếu máu ở phụ nữ mang thai

Câu hỏi	Dạng câu	Ghi chú mã hóa
Tuổi (năm)	Mở	Numeric
Trình độ học vấn	Lựa chọn	1 = Tiểu học, 2 = THCS, 3 = THPT, 4 = Cao đẳng/ĐH
Số lần mang thai	Mở	Numeric
Có sử dụng viên sắt không? Có/Không	0 = Không, 1 = Có	

4.6.2. Nguyên tắc viết câu hỏi

- Câu hỏi **ngắn gọn, rõ ràng, không định hướng;**
- Sử dụng **ngôn ngữ phù hợp với người trả lời;**
- Tránh dùng từ chuyên môn khó hiểu;
- Với câu hỏi định lượng, cần xác định **đơn vị đo lường** (năm, kg, mmHg...).

4.7. Kiểm tra và làm sạch dữ liệu

Trước khi nhập vào SPSS, dữ liệu cần được kiểm tra bằng tay hoặc bằng phần mềm (EpiData, Excel) để đảm bảo:

1. Không thiếu giá trị ở biến chính;
2. Không có giá trị ngoài phạm vi;
3. Mã hóa đúng quy ước;

4. Tên biến thống nhất với bảng định nghĩa.

Trong SPSS 27, có thể làm sạch dữ liệu bằng:

- **Descriptives → Frequencies:** xem giá trị tối đa/tối thiểu, phát hiện ngoại lệ.
- **Data → Select Cases:** lọc dữ liệu hợp lệ.
- **Transform → Recode into Same/Different Variables:** chuẩn hóa mã trị số.

4.8. Đảm bảo độ tin cậy và giá trị của công cụ

4.8.1. Độ tin cậy (Reliability)

Thể hiện **mức độ ổn định và nhất quán** của công cụ đo.

Đánh giá bằng:

- **Cronbach's Alpha:** đo độ nhất quán nội bộ ($\alpha \geq 0,7$ là chấp nhận được).
- **Kiểm tra lặp lại (Test-retest):** dùng hai lần đo cách nhau thời gian ngắn.

4.8.2. Giá trị (Validity)

Phản ánh **mức độ công cụ đo lường được đúng khái niệm cần đo**.

Các loại giá trị:

- **Giá trị nội dung (Content validity):** được chuyên gia xác nhận.
- **Giá trị cấu trúc (Construct validity):** kiểm định bằng phân tích nhân tố (EFA).
- **Giá trị dự báo (Predictive validity):** dự đoán được kết quả thực tế.

4.9. Các khái niệm thống kê cơ bản và ý nghĩa trong y học

4.9.1. Trung bình (Mean) – Giá trị đại diện

Khái niệm

- Trung bình ((\bar{X})) là **giá trị trung tâm** đại diện cho toàn bộ dữ liệu, giúp mô tả “xu hướng chung” của quần thể hoặc mẫu.
- Nó **nhạy cảm với giá trị ngoại lai** (outlier): một vài giá trị quá cao hoặc thấp có thể làm thay đổi trung bình đáng kể.

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

Ví dụ y học

Giả sử ta đo huyết áp tâm thu (mmHg) của 10 bệnh nhân:

Bệnh nhân Huyết áp (mmHg)

1	118
2	122
3	115
4	120
5	200
6	124

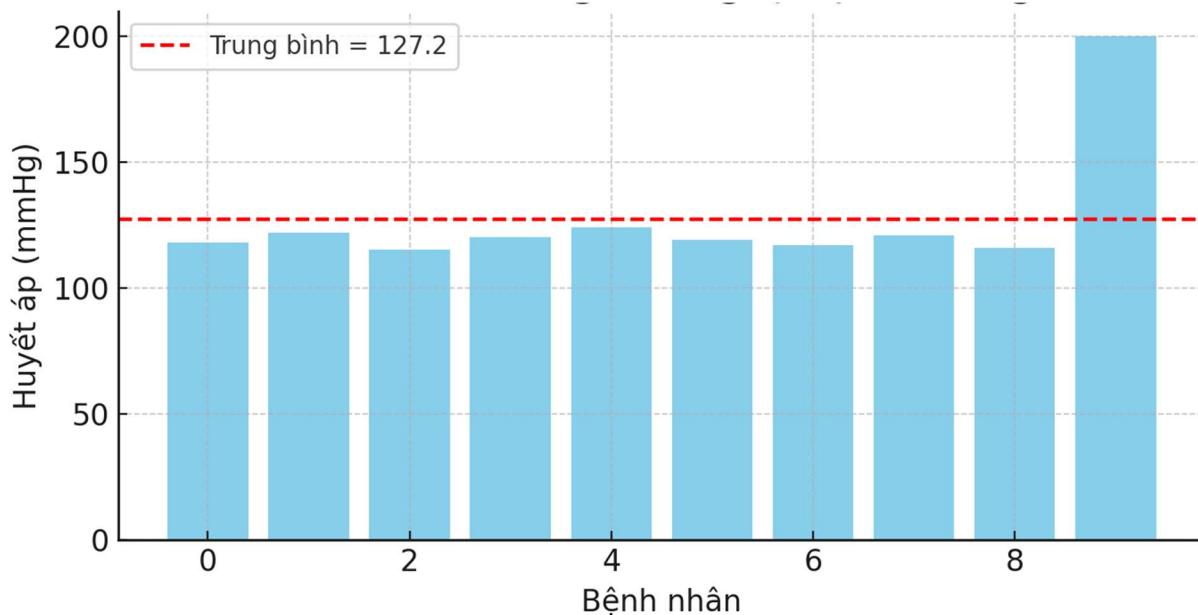
Bệnh nhân Huyết áp (mmHg)

7	119
8	117
9	121
10	116

$$\bar{X} = \frac{118 + 122 + 115 + 120 + 200 + 124 + 119 + 117 + 121 + 116}{10} = 137.2$$

Trung bình = 137.2 mmHg, nhưng có **một giá trị 200** khiến trung bình “đội” lên.

→ Nên dùng **trung vị (median)** khi có ngoại lệ.



Hình 1. Biểu đồ minh họa ảnh hưởng của giá trị ngoại lai đến trung bình
(Biểu đồ cột: các cột 115–124 tương đối đều, riêng 200 vượt hẳn lên — thể hiện trung bình bị kéo lệch.)

4.9.2. Độ lệch chuẩn (Standard Deviation – SD)

Khái niệm

- SD đo **mức độ phân tán dữ liệu quanh trung bình**.
- Dữ liệu càng đồng nhất, SD càng nhỏ.

$$SD = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}}$$

Ví dụ

Hb (g/dL) của 5 sản phẩm: 10.5, 10.7, 10.8, 10.6, 10.9.

$$\bar{X} = 10.7, \quad SD = 0.15$$

→ Các giá trị gần nhau, SD nhỏ → nhóm đồng nhất.

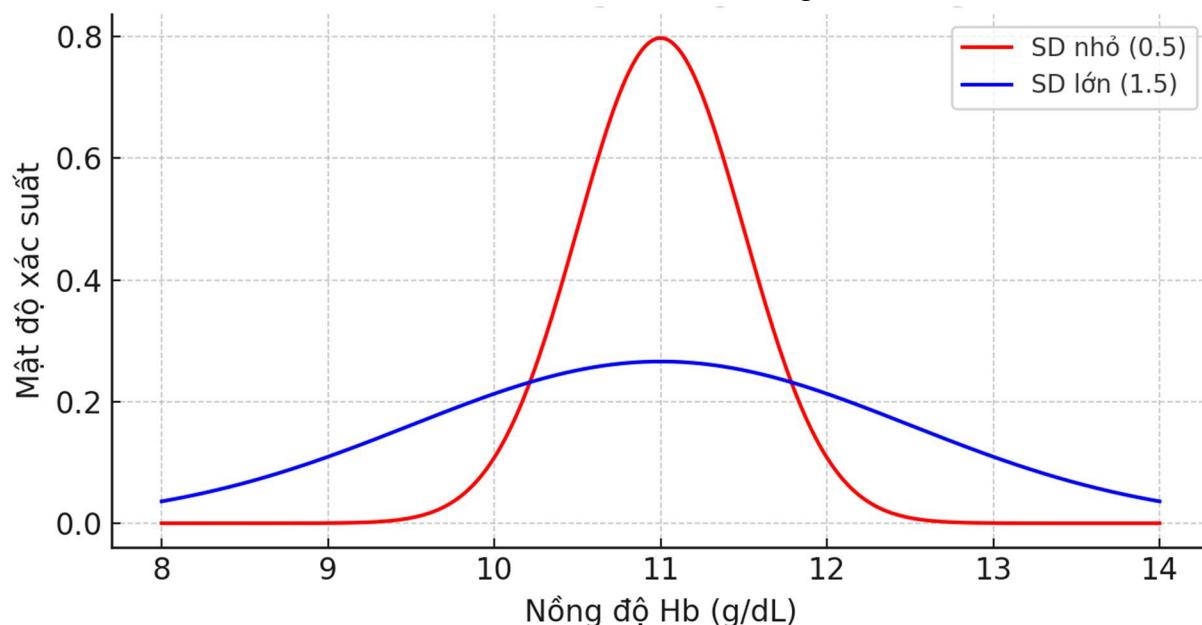
Nếu dữ liệu là 9.0, 11.5, 13.0, 10.0, 11.0:

$$\bar{X} = 10.9, \quad SD = 1.5$$

→ SD lớn → nhóm biến động nhiều.

Ý nghĩa

- SD nhỏ → dữ liệu ổn định, dễ khái quát.
- SD lớn → cần cỡ mẫu lớn hơn để đạt độ chính xác tương tự.



Hình 2. Biểu đồ so sánh hai nhóm có cùng trung bình nhưng SD khác nhau
(Đường cong màu đỏ hẹp – SD nhỏ; đường xanh rộng – SD lớn. Cả hai có trung bình 11.0 g/dL.)

4.9.3. Sai số chuẩn (Standard Error – SE)

Khái niệm

- SE đo mức độ sai lệch giữa trung bình mẫu và trung bình quần thể thật.
- Càng nhiều mẫu → SE càng nhỏ → trung bình càng “ ổn định.”

$$SE = \frac{SD}{\sqrt{n}}$$

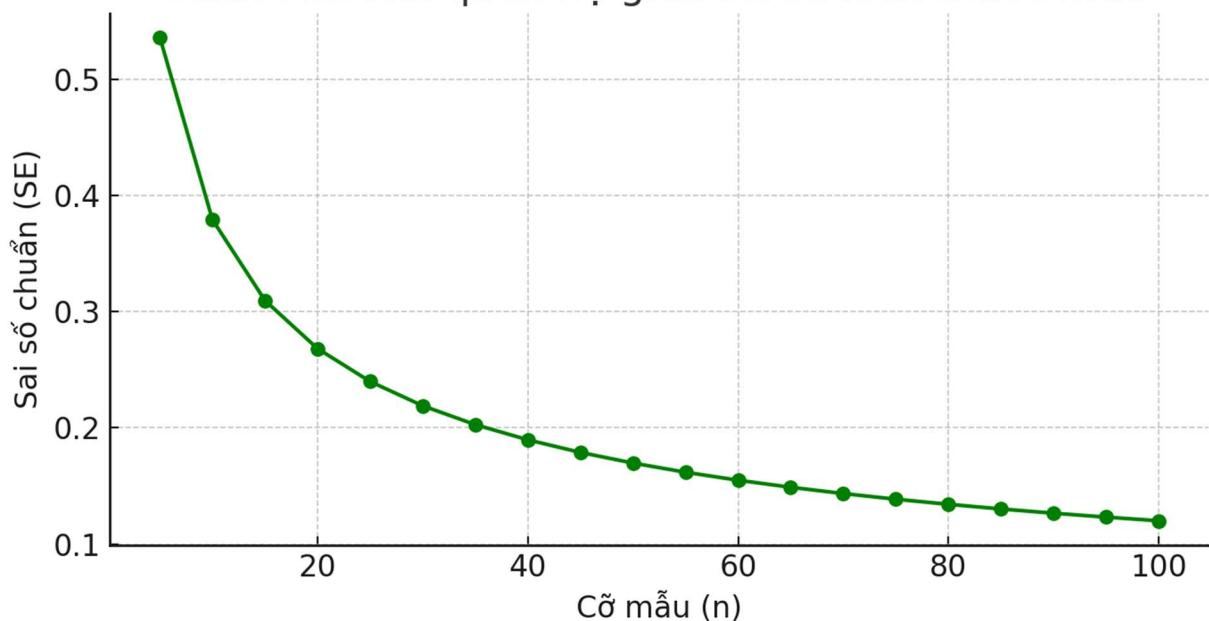
Ví dụ

Với SD = 1.2, n = 100 →

$$SE = \frac{1.2}{\sqrt{100}} = 0.12$$

Nếu chỉ có n = 25 → SE = 0.24 → độ tin cậy giảm đi một nửa.

Hình 5.3. Mối quan hệ giữa SE và kích thước mẫu



Hình 3. Mối quan hệ giữa SE và kích thước mẫu
 (Biểu đồ đường: SE giảm dần theo \sqrt{n} — đường cong đi xuống nhanh rồi dẹt ra.)

4.9.4. Khoảng tin cậy (Confidence Interval – CI)

Khái niệm

- CI là **khoảng ước lượng** trong đó giá trị thật của quần thể có khả năng nằm trong đó.
- Thường dùng mức **tin cậy 95%**.

Công thức

Nếu quần thể lớn và SD biết:

$$CI = \bar{X} \pm Z_{1-\alpha/2} \times SE$$

Nếu mẫu nhỏ hoặc SD không biết:

$$CI = \bar{X} \pm t_{1-\alpha/2, df=n-1} \times SE$$

Mức tin cậy Giá trị tối hạn Ký hiệu

90 %	1.645	(Z_{0.95})
95 %	1.960	(Z_{0.975})
99 %	2.576	(Z_{0.995})

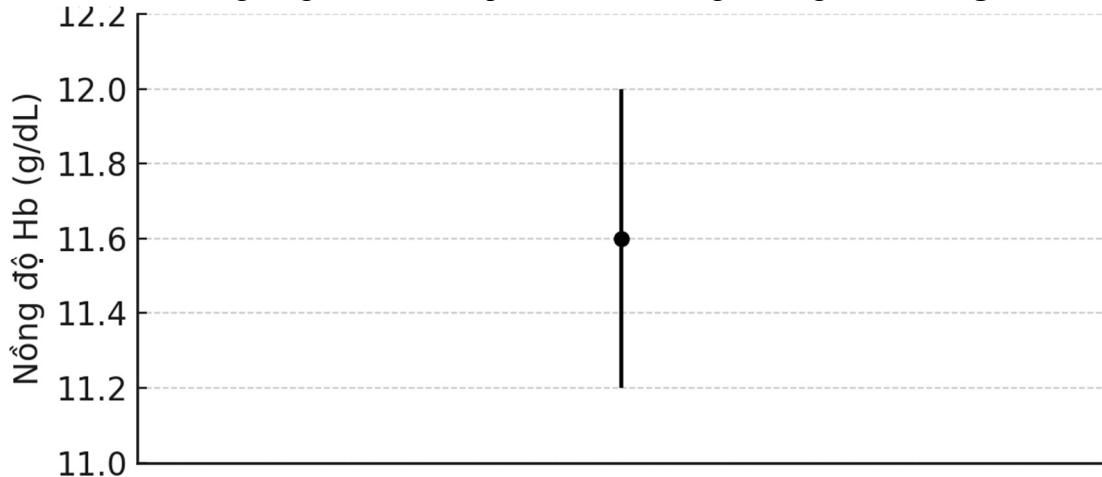
Ví dụ y học

Hb trung bình của 25 thai phụ là 11.6 ± 1.0 (SD).

$$SE = \frac{1.0}{\sqrt{25}} = 0.2$$

$$CI_{95\%} = 11.6 \pm 1.96 \times 0.2 = (11.2; 12.0)$$

→ Có 95% tin tưởng rằng Hb thật của quần thể nằm trong khoảng **11.2–12.0 g/dL**.



Hình 4. Biểu đồ minh họa khoảng tin cậy (CI)

(Điểm trung bình ở giữa, hai gạch ngang 11.2–12.0 tạo khoảng CI 95%).

4.9.5. Mức ý nghĩa (Significance Level – α)

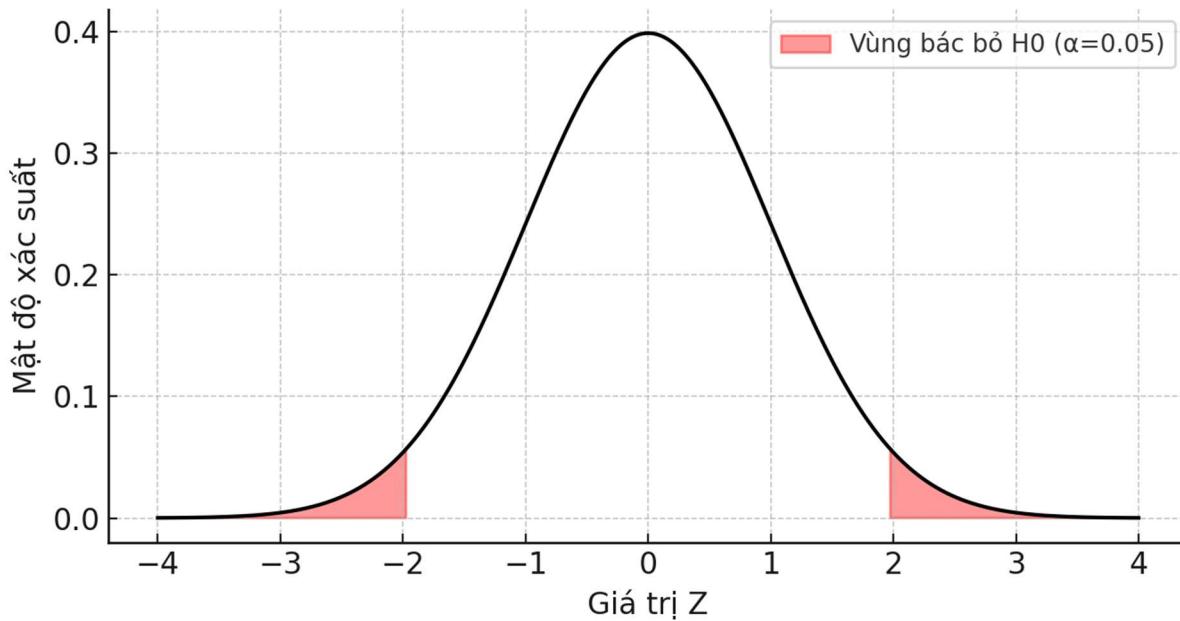
Khái niệm

- α là xác suất **chấp nhận sai lầm loại I** – tức là bác bỏ giả thuyết gốc (H_0) khi H_0 đúng.
- Thường chọn $\alpha = 0.05 \rightarrow$ cho phép 5% rủi ro sai lầm.

$$1 - \alpha = \text{Mức độ tin cậy (Confidence level)}$$

Ví dụ

Nếu $\alpha = 0.05 \rightarrow$ ta chấp nhận rằng **5 lần trong 100 lần kiểm định**, có thể kết luận sai rằng “có khác biệt” khi thực tế không có.



Hình 5. Minh họa phân phối chuẩn và vùng bác bỏ H_0
 (Vùng màu đỏ hai bên chiếm 5% tổng diện tích → tương ứng $\alpha = 0.05$.)

4.9.6. Giá trị p (p-value)

Khái niệm

- p là xác suất quan sát được kết quả bằng hoặc cực đoan hơn nếu H_0 đúng.
- p càng nhỏ \rightarrow bằng chứng chống lại H_0 càng mạnh.

p-value **Điễn giải**

$p < 0.05$ Có ý nghĩa thống kê

$p < 0.01$ Rất có ý nghĩa

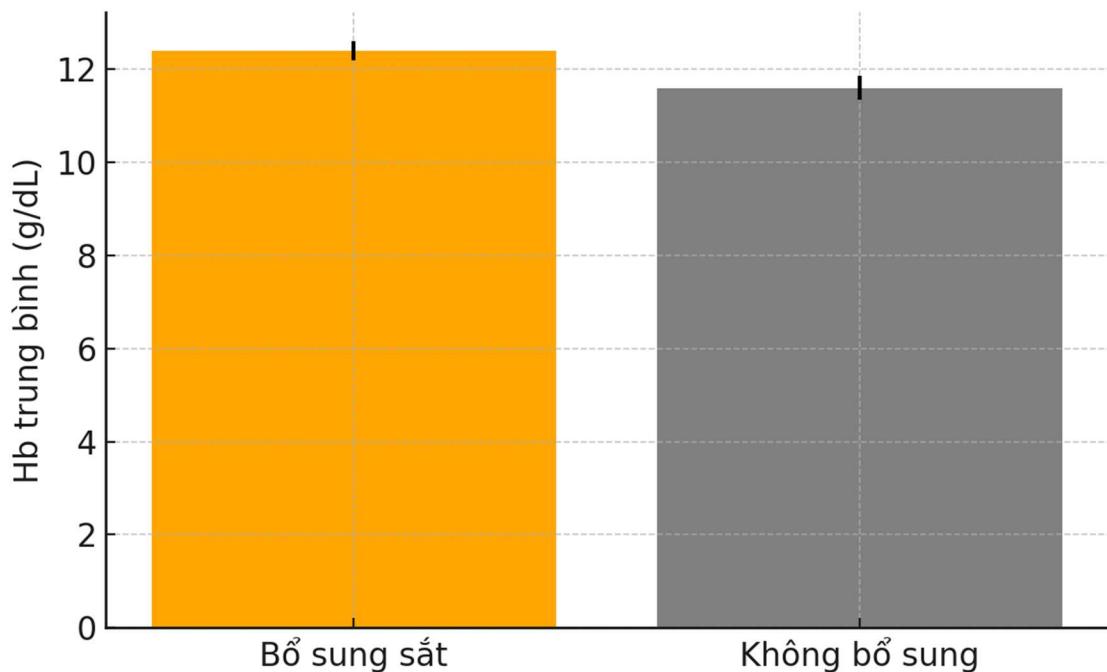
$p < 0.001$ Cực kỳ có ý nghĩa

$p \geq 0.05$ Không có ý nghĩa

6.2. Ví dụ

Nghiên cứu kiểm định Hb giữa nhóm dùng và không dùng viên sắt:

- Hb trung bình nhóm sắt = 12.4 ± 1.1
- Hb nhóm không sắt = 11.6 ± 1.2
 $\rightarrow p = 0.002$.
 \rightarrow Vì $p < 0.05$, ta kết luận **có sự khác biệt có ý nghĩa thống kê**.



Hình 6. Biểu đồ so sánh hai nhóm theo giá trị trung bình và CI

Hai cột: cột nhóm sắt cao hơn, có đường CI không chồng lên nhau → thể hiện khác biệt đáng kể.

4.9.7. Tác động mẫu (Effect Size)

Khái niệm

- Effect size đo **mức độ mạnh của mối quan hệ hoặc sự khác biệt** – cho biết “ý nghĩa lâm sàng.”
- Không phụ thuộc vào cỡ mẫu.

Công thức Cohen's d

$$d = \frac{M_1 - M_2}{SD_{pooled}}$$

$$SD_{pooled} = \sqrt{\frac{SD_1^2 + SD_2^2}{2}}$$

Điễn giải

d Mức độ tác động	Ý nghĩa y học
0.2 Nhỏ	Khác biệt nhẹ, ít ý nghĩa
0.5 Trung bình	Ảnh hưởng vừa phải
0.8 Lớn	Ảnh hưởng rõ, có ý nghĩa lâm sàng

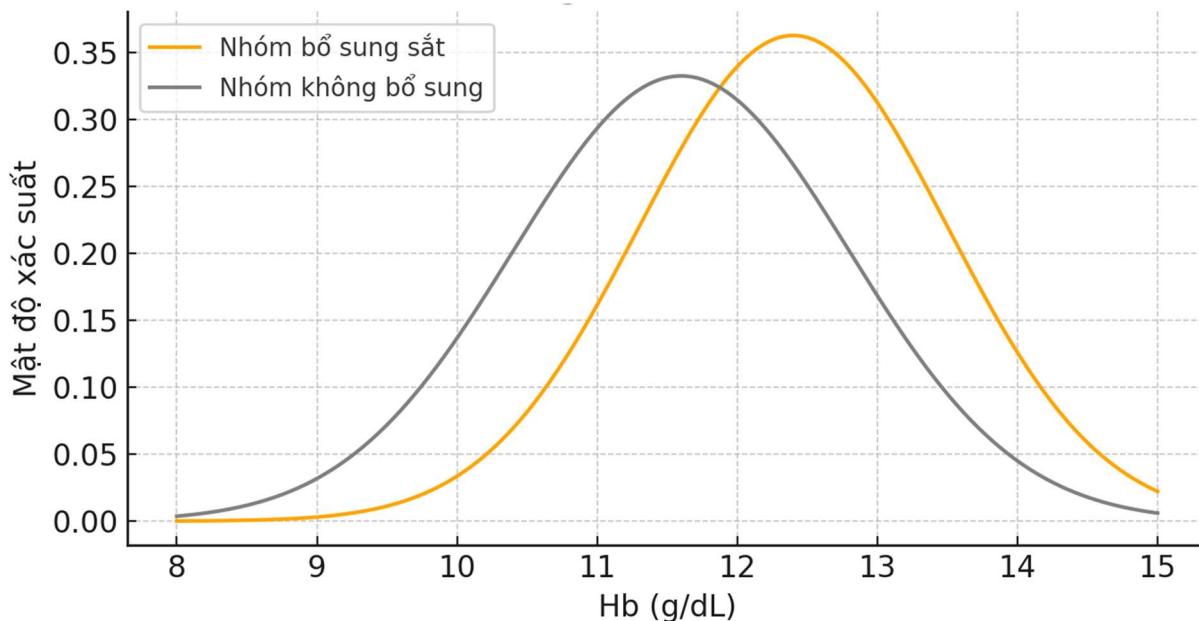
7.4. Ví dụ

Hb trung bình nhóm bổ sung sắt = 12.4 ± 1.1

Hb nhóm không sắt = 11.6 ± 1.2

$$SD_{pooled} = 1.15, \quad d = \frac{12.4 - 11.6}{1.15} = 0.7$$

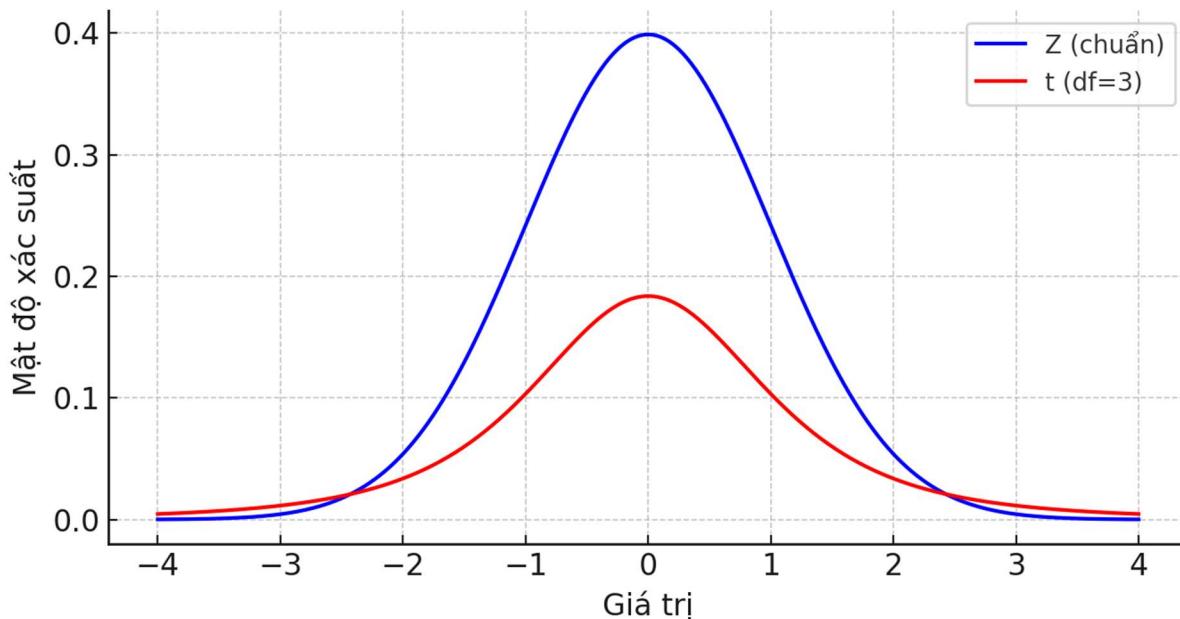
→ Effect size trung bình-lớn → **hiệu quả bổ sung sắt đáng kể.**



Hình 7. Biểu đồ mô phỏng “khoảng chòng lấp” giữa hai phân phối Hb
(Hai đường cong: nhóm sắt dịch sang phải rõ rệt, chòng lấp ít hơn → effect size lớn.)

4.9.8. Phân phối Student's t và Z

Trường hợp	Phân phối sử dụng	Ký hiệu	Ghi chú
Mẫu lớn ($n \geq 30$)	Chuẩn (Normal)	$(Z_{1-\alpha/2})$	Có SD quần thể
Mẫu nhỏ ($n < 30$)	Student's t	$(t_{\{1 - \alpha/2, df = n - 1\}})$	Chưa biết SD quần thể



Hình 8. So sánh phân phối Z và t

Đường cong t “bẹt hơn” Z → phản ánh độ bất định lớn hơn khi mẫu nhỏ. Khi n tăng, t → Z.

4.9.9. Kết nối các tham số với nhau

Tham số	Công thức chính	Ý nghĩa	Ảnh hưởng thực hành
SD	$\left(\sqrt{\frac{\sum(X - \bar{X})^2}{n - 1}} \right)$	Biến động dữ liệu	Cần kiểm tra trước khi phân tích
SE	(SD/\sqrt{n})	Sai số của trung bình mẫu	Giảm khi tăng n
CI	$(\bar{X} \pm Z \times SE)$	Khoảng ước lượng tham số thật	CI hẹp → tin cậy cao
α	Mức ý nghĩa	Rủi ro chấp nhận sai lầm loại I	Thường chọn 0.05
p	Xác suất kết quả nếu H_0 đúng	Bằng chứng chống H_0	$p < \alpha \rightarrow$ có ý nghĩa thống kê
d	$((M_1 - M_2)/SD_{\text{pooled}})$	Cường độ ảnh hưởng	Giúp đánh giá ý nghĩa lâm sàng

4.10. Tổng kết

- Biến số là nền tảng của mọi phân tích thống kê.
- Cần xác định rõ loại biến, thang đo và vai trò của từng biến trong nghiên cứu.
- Việc thu thập số liệu phải được chuẩn hóa, đảm bảo độ tin cậy và giá trị.
- SPSS 27 hỗ trợ mạnh mẽ cho giai đoạn **mã hóa, làm sạch và kiểm tra dữ liệu** trước khi tiến hành phân tích thống kê.

Trong y học, **một kết quả chỉ có giá trị khi hiểu rõ bản chất thống kê** phía sau:

- SD, SE và CI cho biết **độ tin cậy và biến động**;
- p và α cho biết **ý nghĩa thống kê**;
- Effect size cho biết **ý nghĩa lâm sàng thực tế**.

Tài liệu tham khảo

- [1] Lưu Ngọc Hoạt, *Nghiên cứu khoa học trong y học*, NXB Y học, Hà Nội, 2014.
- [2] Stevens S. S., “On the Theory of Scales of Measurement,” *Science*, vol. 103, pp. 677–680, 1946.
- [3] WHO, *Health Research Methodology: A Guide for Training in Research Methods*, 3rd ed., Geneva, 2014.
- [4] Polit D. F., Beck C. T., *Nursing Research: Generating and Assessing Evidence for Nursing Practice*, 10th ed., Lippincott Williams & Wilkins, 2017.
- [5] Field A., *Discovering Statistics Using IBM SPSS Statistics*, 5th ed., SAGE Publications, 2020.

CHƯƠNG 5. TỔNG HỢP, PHÂN TÍCH VÀ TRÌNH BÀY SỐ LIỆU

5.1. Khái niệm và vai trò của phân tích số liệu

Phân tích số liệu (Data Analysis) là **quá trình sử dụng các công cụ thống kê để mô tả, so sánh, và rút ra kết luận từ dữ liệu nghiên cứu**.

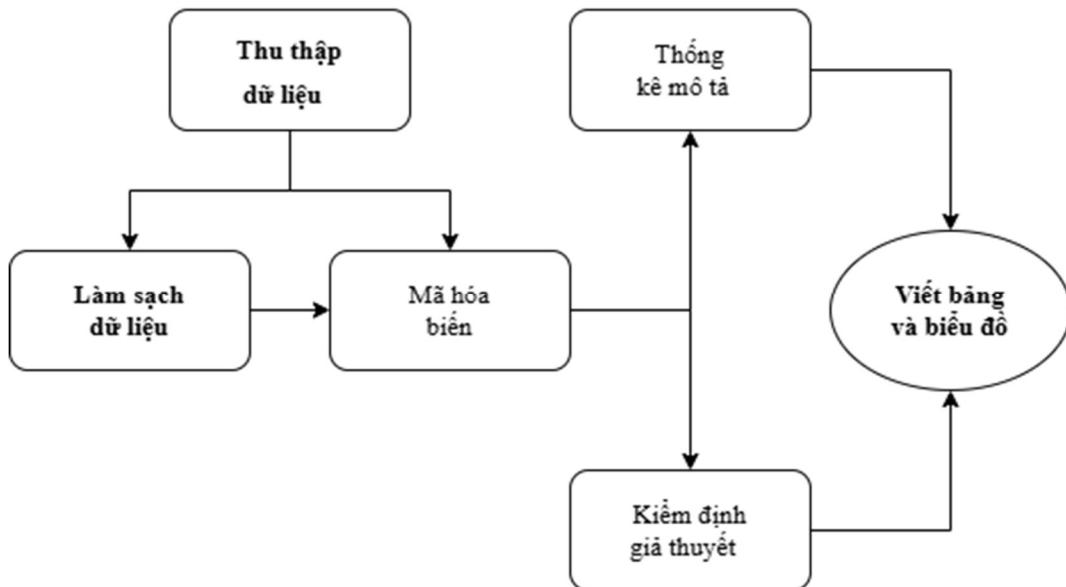
Mục đích chính là biến dữ liệu thô thành **thông tin có ý nghĩa khoa học** giúp trả lời các câu hỏi nghiên cứu đã đặt ra.

Trong nghiên cứu y học, phân tích số liệu giúp:

- Kiểm tra giả thuyết nghiên cứu;
- Xác định mối quan hệ giữa các biến;
- Tính toán các chỉ số y học (tỷ lệ, trung bình, độ lệch chuẩn...);
- Đưa ra các kết luận dựa trên bằng chứng thống kê.

Theo Field (2020), “thống kê y học không chỉ là việc tính toán, mà là cách chúng ta kể câu chuyện của dữ liệu bằng các con số và biểu đồ có ý nghĩa” [1].

5.2. Quy trình xử lý và phân tích dữ liệu



Hình 9. Quy trình xử lý và phân tích dữ liệu

5.2.1. Làm sạch và kiểm tra dữ liệu

Trước khi phân tích, dữ liệu phải được kiểm tra để loại bỏ lỗi nhập liệu hoặc giá trị bất thường (outlier).

Trong SPSS 27, thực hiện qua:

- Analyze → Descriptive Statistics → Frequencies: xem tần suất, phát hiện giá trị sai.
- Graphs → Boxplot: phát hiện ngoại lệ.
- Data → Select Cases: lọc dữ liệu hợp lệ.

5.2.2. Mã hóa biến

Biến định tính được mã hóa bằng số (0,1,2...), đảm bảo tương thích với SPSS.

Ví dụ:

Giới tính: 1 = Nam, 2 = Nữ.

Thiếu máu: 0 = Không, 1 = Có.

5.3. Thông kê mô tả

5.3.1. Khái niệm

Thông kê mô tả (Descriptive Statistics) nhằm tóm tắt và mô tả đặc điểm của tập dữ liệu bằng các chỉ số trung tâm và phân tán, hoặc bằng bảng, biểu đồ.

Tùy vào loại biến (định tính – định lượng) mà ta chọn chỉ số mô tả thích hợp.

5.3.2. Các chỉ số thống kê cơ bản

Loại biến	Chỉ số trung tâm	Chỉ số phân tán	Ví dụ
Định tính	Tần suất, tỷ lệ (%)	–	Tỷ lệ bệnh nhân nam: 55%
Định lượng phân phối chuẩn	Trung bình (Mean)	Độ lệch chuẩn (SD)	Huyết áp TB = 120 ± 10 mmHg
Định lượng không chuẩn	Trung vị (Median)	Khoảng từ phân vị (IQR)	Hb = 11,2 (IQR: 10,3–12,4) g/dL

5.3.3. Cách thực hiện trong SPSS

- Vào menu **Analyze → Descriptive Statistics → Descriptives / Frequencies**
- Chọn các biến cần mô tả (tuổi, Hb, BMI...)
- Chọn “Mean, Std. Deviation, Minimum, Maximum”
- SPSS sẽ xuất ra bảng kết quả như sau:

Biến	N	Trung bình	SD	Min	Max
Tuổi (năm)	350	28.4	6.2	18	45
Hb (g/dL)	350	11.6	1.3	8.5	14.8

5.4. Trình bày dữ liệu bằng bảng

5.4.1. Nguyên tắc chung

- Bảng phải có **số thứ tự, tiêu đề ngắn gọn, đơn vị rõ ràng**.
- Không nên có quá 10 hàng hoặc 6 cột để đảm bảo dễ đọc.
- Mọi số liệu nên được **làm tròn** đến 1 hoặc 2 chữ số thập phân.
- Dưới bảng phải có **chú thích (Ghi chú / Source)** nếu cần.

5.4.2. Cấu trúc bảng chuẩn trong nghiên cứu y học

Đặc điểm Số lượng (n) Tỷ lệ (%)

Giới tính

– Nam 190 54.3

– Nữ 160 45.7

Nhóm tuổi

– < 25 tuổi 85 24.3

– 25–34 tuổi 200 57.1

– ≥ 35 tuổi 65 18.6

Tổng 350 100.0

Ghi chú: N = 350 đối tượng khảo sát tại Bệnh viện Đa khoa Tiền Giang (2024).

5.5. Trình bày dữ liệu bằng biểu đồ

Biểu đồ giúp minh họa trực quan mối quan hệ giữa các biến và xu hướng dữ liệu.

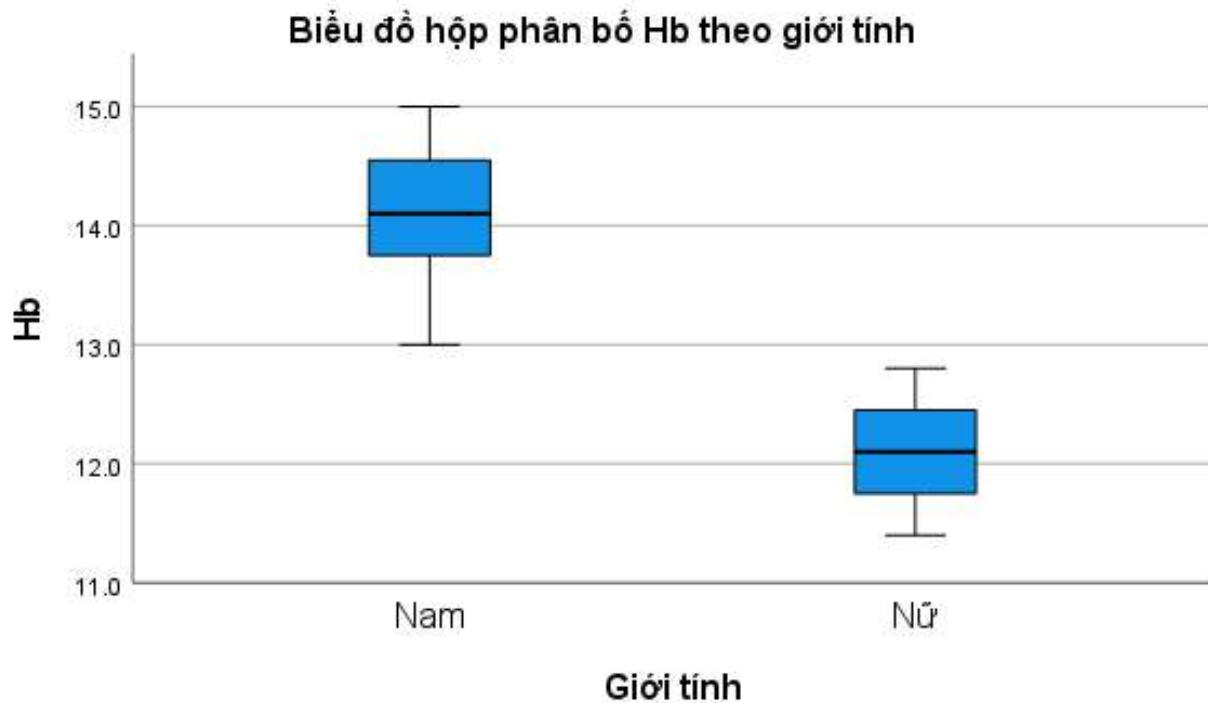
Một số loại biểu đồ thường dùng trong thống kê y học:

Bảng 18. Các loại biểu đồ thường dùng

Loại biểu đồ	Mục đích	Ví dụ
Biểu đồ cột (Bar chart)	So sánh tần suất giữa các nhóm	Tỷ lệ thiếu máu theo nhóm tuổi
Biểu đồ tròn (Pie chart)	Thể hiện tỷ lệ phần trăm	Phân bố giới tính trong mẫu
Biểu đồ hộp (Boxplot)	Mô tả phân bố, phát hiện ngoại lệ	Phân bố Hb theo giới tính
Histogram	Kiểm tra phân phối chuẩn của biến định lượng	Phân bố tuổi trong nghiên cứu
Biểu đồ đường (Line chart)	Thể hiện xu hướng theo thời gian	Tỷ lệ nhiễm bệnh qua các năm

5.5.1. Thực hiện trong SPSS

- **Bar chart:** Graphs → Chart Builder → chọn “Bar”.
- **Pie chart:** Graphs → Chart Builder → chọn “Pie”.
- **Boxplot:** Graphs → Boxplot → chọn biến định lượng (Y) và nhóm (X).
- **Histogram:** Graphs → Legacy Dialogs → Histogram → chọn biến cần vẽ.



Hình 10. Ví dụ biểu đồ hộp (Boxplot) trong SPSS – Phân bố Hb theo giới tính

5.6. Kiểm định giả thuyết thống kê (Statistical Hypothesis Testing)

5.6.1. Khái niệm

Kiểm định giả thuyết thống kê là **phương pháp xác định xem bằng chứng trong mẫu có đủ mạnh để bác bỏ giả thuyết không (H_0) hay không** [2].

5.6.2. Các bước kiểm định

1. **Đặt giả thuyết H_0 và H_1**
 - H_0 : Không có sự khác biệt / không có mối liên quan.
 - H_1 : Có sự khác biệt / có mối liên quan.
2. **Chọn mức ý nghĩa (α)** – thường 0,05.
3. **Chọn phép kiểm định phù hợp** tùy theo loại biến và phân phối dữ liệu.
4. **Phân tích bằng SPSS**, xem giá trị p .
5. **Kết luận:**
 - Nếu $p < 0,05 \rightarrow$ Bác bỏ H_0 (có khác biệt có ý nghĩa).
 - Nếu $p \geq 0,05 \rightarrow$ Chấp nhận H_0 (không khác biệt có ý nghĩa).

Bảng 19. Lựa chọn phép kiểm định theo loại biến

Mục tiêu	Loại biến	Phép kiểm định	Ghi chú
So sánh trung bình giữa 2 nhóm	Định lượng, phân phối chuẩn	t-test độc lập	Ví dụ: tuổi trung bình giữa 2 giới
So sánh trung bình > 2 nhóm	Định lượng, phân phối chuẩn	ANOVA	Kiểm tra tuổi giữa 3 nhóm nghề
So sánh tỷ lệ giữa 2 nhóm	Định tính	Chi-square / Fisher	So sánh tỷ lệ thiếu máu giữa nam – nữ
Mối tương quan giữa 2 biến định lượng	Định lượng	Pearson / Spearman	Hb và tuổi thai

5.7. Trình bày kết quả phân tích trong báo cáo nghiên cứu

Khi viết báo cáo, các kết quả thống kê cần được trình bày **ngắn gọn, đủ thông tin và dễ hiểu**. Cấu trúc phổ biến:

1. **Bảng mô tả đặc điểm chung** của mẫu (tuổi, giới, nghề nghiệp).
2. **Bảng hoặc biểu đồ mô tả biến chính** (Hb, BMI, bệnh, hành vi).
3. **Kết quả kiểm định thống kê** – nêu giá trị p , thống kê kiểm định (t , F , χ^2), và diễn giải.

Ví dụ 5.1 – Trình bày kết quả trong báo cáo

Tuổi trung bình của mẫu nghiên cứu là $28,4 \pm 6,2$ tuổi.

Tỷ lệ thiếu máu ở phụ nữ mang thai là 32,5%.

Có mối liên quan có ý nghĩa thống kê giữa thiếu máu và trình độ học vấn ($\chi^2 = 5,87$; $p = 0,015$).

Tuy nhiên, không ghi nhận sự khác biệt đáng kể theo nhóm tuổi ($p > 0,05$).

5.8. Các lỗi thường gặp khi phân tích số liệu

Bảng 20. Các lỗi thường gặp khi phân tích số liệu

Lỗi thường gặp	Nguyên nhân	Cách khắc phục
Sử dụng sai phép kiểm định	Không xác định đúng loại biến và thang đo	Xem lại Bảng 5.5 hoặc dùng “SPSS Decision Tree”
Quên kiểm tra phân phối chuẩn	Áp dụng t-test/ANOVA sai	Thực hiện Shapiro–Wilk trước
Không làm sạch dữ liệu	Có giá trị ngoại lai hoặc thiếu	Kiểm tra bằng Frequencies và Boxplot
Làm tròn sai số	Làm sai lệch ý nghĩa thống kê	Làm tròn 1–2 chữ số thập phân
Không nêu rõ giá trị p	Viết “không khác biệt” mà không có số liệu cụ thể	Luôn báo cáo p-value cụ thể (ví dụ: $p = 0,032$)

5.9. Kết luận chương

- Phân tích số liệu là bước trung tâm của nghiên cứu y học, chuyển dữ liệu thô thành tri thức.
- SPSS 27 hỗ trợ toàn diện từ mô tả, kiểm định, đến biểu đồ hóa dữ liệu.
- Kết quả cần được trình bày theo chuẩn khoa học: bảng, biểu đồ, và diễn giải p-value rõ ràng.
- Việc lựa chọn đúng phép kiểm định và trình bày mạch lạc sẽ nâng cao chất lượng báo cáo và giá trị học thuật của nghiên cứu.

Tài liệu tham khảo

- [1] Field A., *Discovering Statistics Using IBM SPSS Statistics*, 5th ed., SAGE Publications, 2020.
- [2] Bland M., *An Introduction to Medical Statistics*, 4th ed., Oxford University Press, 2015.
- [3] WHO, *Health Research Methodology: A Guide for Training in Research Methods*, 3rd ed., Geneva, 2014.
- [4] Lưu Ngọc Hoạt, *Nghiên cứu khoa học trong y học*, NXB Y học, Hà Nội, 2014.
- [5] Polit D. F., Beck C. T., *Nursing Research: Generating and Assessing Evidence for Nursing Practice*, 10th ed., Lippincott Williams & Wilkins, 2017.

CHƯƠNG 6. VIẾT VÀ TRÌNH BÀY BÁO CÁO NGHIÊN CỨU

6.1. Khái niệm và ý nghĩa của báo cáo nghiên cứu

Báo cáo nghiên cứu khoa học (Research Report) là tài liệu mô tả toàn bộ quá trình thực hiện một nghiên cứu, bao gồm cơ sở lý luận, phương pháp, kết quả và kết luận.

Trong y học, báo cáo nghiên cứu là **hình thức chuyển hóa dữ liệu khoa học thành tri thức y học có thể ứng dụng vào thực tiễn** [1].

Vì việc viết báo cáo không chỉ là tổng hợp kết quả, mà còn thể hiện:

- Tư duy khoa học và khả năng trình bày của người nghiên cứu;
- Sự minh bạch, khách quan của dữ liệu;
- Bằng chứng phục vụ việc ra quyết định y tế, điều trị và đào tạo.

6.2. Các loại báo cáo nghiên cứu trong y học

Bảng 21. Phân loại các loại báo cáo nghiên cứu phổ biến

Loại báo cáo	Mục đích	Đặc điểm
Báo cáo đề tài khoa học (Research Project Report)	Tổng kết toàn bộ quá trình nghiên cứu	Dài, có phụ lục và bảng biểu chi tiết
Báo cáo tóm tắt (Abstract Report)	Tóm tắt cho hội nghị, hội thảo	250–300 từ, trình bày ngắn gọn
Bài báo khoa học (Scientific Paper)	Công bố kết quả trên tạp chí	Theo cấu trúc IMRaD (Introduction–Methods–Results–Discussion)
Báo cáo kỹ thuật (Technical Report)	Trình bày quy trình, sản phẩm nghiên cứu	Thường dùng trong nghiên cứu ứng dụng, thử nghiệm lâm sàng
Luận văn / Luận án	Báo cáo nghiên cứu chuyên sâu ở bậc học cao	Có hội đồng phản biện, kiểm tra chặt chẽ về hình thức và nội dung

6.3. Cấu trúc cơ bản của một báo cáo nghiên cứu y học

Một báo cáo nghiên cứu y học hoàn chỉnh thường có cấu trúc **IMRaD** (Introduction – Methods – Results – Discussion), được chấp nhận bởi hầu hết các tạp chí quốc tế [2].

Bảng 22. Cấu trúc và nội dung cơ bản của báo cáo nghiên cứu

Phần	Nội dung chính	Lưu ý
1. Trang bìa và thông tin chung	Tên đề tài, tác giả, cơ quan, thời gian, đơn vị tài trợ (nếu có)	Ngắn gọn, đúng chuẩn hành chính

Phần	Nội dung chính	Lưu ý
2. Tóm tắt (Abstract)	Mục tiêu, phương pháp, kết quả chính, kết luận (250–300 từ)	Viết cuối cùng, dù đặt đầu báo cáo
3. Giới thiệu (Introduction)	Lý do nghiên cứu, tổng quan, mục tiêu, giả thuyết	Dẫn chứng bằng trích dẫn chuẩn IEEE
4. Phương pháp (Methods)	Thiết kế, đối tượng, cỡ mẫu, biến, công cụ, phân tích thống kê	Rõ ràng, chi tiết, có thể tái lập
5. Kết quả (Results)	Các bảng, biểu đồ, phân tích thống kê	Điển giải theo logic mục tiêu
6. Bàn luận (Discussion)	So sánh với nghiên cứu khác, giải thích kết quả, hạn chế	Không lặp lại kết quả, tránh suy diễn
7. Kết luận và kiến nghị (Conclusion, Recommendation)	Tóm tắt phát hiện chính, khuyến nghị ứng dụng	Ngắn gọn, bám sát mục tiêu
8. Tài liệu tham khảo	Theo chuẩn IEEE hoặc Vancouver	Sắp xếp theo thứ tự trích dẫn
9. Phụ lục (nếu có)	Bảng câu hỏi, bộ số liệu, hình ảnh minh chứng	Không bắt buộc, tùy yêu cầu

6.4. Viết các phần của báo cáo nghiên cứu

6.4.1. Phần Giới thiệu (Introduction)

Mục đích: Giải thích lý do tại sao nghiên cứu được tiến hành.

Cần có các yếu tố:

- Vấn đề sức khỏe hiện tại** (có dẫn chứng số liệu, ví dụ từ WHO hoặc Bộ Y tế);
- Khoảng trống tri thức** (chưa có nghiên cứu hoặc còn mâu thuẫn);
- Mục tiêu nghiên cứu** (tổng quát và cụ thể).

Ví dụ:

“Thiếu máu là một vấn đề sức khỏe cộng đồng quan trọng, đặc biệt ở phụ nữ mang thai. Theo WHO (2021), tỷ lệ thiếu máu toàn cầu ở nhóm này là 36,5%. Tuy nhiên, tại Việt Nam, số liệu cập nhật còn hạn chế ở cấp tỉnh. Vì vậy, nghiên cứu này được thực hiện nhằm xác định tỷ lệ và các yếu tố liên quan đến thiếu máu ở phụ nữ mang thai tại Tiền Giang năm 2024.”

6.4.2. Phần Phương pháp (Methods)

Phần này quyết định **độ tin cậy và khả năng tái lập** của nghiên cứu.

Cấu trúc gợi ý:

- Thiết kế nghiên cứu:** (mô tả, phân tích, can thiệp...)
- Đối tượng nghiên cứu:** tiêu chuẩn chọn và loại trừ;
- Thời gian và địa điểm nghiên cứu;**
- Cỡ mẫu và phương pháp chọn mẫu;**
- Biến số và công cụ thu thập dữ liệu;**

6. **Phân tích thống kê:** phần mềm sử dụng (SPSS 27), phép kiểm định, mức ý nghĩa ($p < 0,05$);
7. **Đạo đức nghiên cứu:** được hội đồng đạo đức thông qua.

Ví dụ tóm tắt:

“Nghiên cứu mô tả cắt ngang được thực hiện từ 01/2024–06/2024 tại Bệnh viện Đa khoa Tiền Giang. Cỡ mẫu 350 phụ nữ mang thai được chọn ngẫu nhiên. Dữ liệu được nhập và phân tích bằng SPSS 27, sử dụng thống kê mô tả, kiểm định Chi-square và hồi quy logistic.”

6.4.3. Phần Kết quả (Results)

Nguyên tắc:

- Trình bày **theo đúng trình tự mục tiêu nghiên cứu;**
- Mỗi mục tiêu nên có **1 bảng hoặc 1 biểu đồ chính;**
- Không lặp lại số liệu giữa bảng và biểu đồ;
- Chỉ trình bày **số liệu đã phân tích**, không liệt kê thô.

Ví dụ 6.1. Trích bảng kết quả từ SPSS

Bảng 23. Kết quả kiểm định Chi-Square cho hiện tượng thiếu máu theo độ tuổi

Yếu tố	Thiếu máu (n=350)	Có (%)	Không (%)	p-value
Nhóm tuổi <25	25	29.4	70.6	0.071
25–34	65	33.8	66.2	
≥35	24	36.9	63.1	
Tổng	114	32.5	67.5	

Ghi chú: kiểm định Chi-square; $p < 0,05$ có ý nghĩa thống kê.

6.4.4. Phần Bàn luận (Discussion)

Đây là phần **thể hiện năng lực khoa học và khả năng phân tích của tác giả.**

Cần tuân thủ 4 nguyên tắc:

1. **Trả lời mục tiêu nghiên cứu:** So sánh kết quả với giả thuyết ban đầu;
2. **So sánh với các nghiên cứu khác:** Dẫn chứng từ trong nước và quốc tế;
3. **Giải thích nguyên nhân khác biệt:** Yếu tố dịch tễ, phương pháp, đối tượng;
4. **Trình bày hạn chế nghiên cứu:** Thiếu mẫu, sai lệch, giới hạn thời gian.

Ví dụ:

“Tỷ lệ thiếu máu trong nghiên cứu này (32,5%) tương tự kết quả của Nguyễn Đức Hinh (2018) tại Hà Nội (33,1%) [3], nhưng thấp hơn so với báo cáo của WHO tại khu vực Đông Nam Á (40,0%) [4]. Nguyên nhân có thể do chương trình bổ sung viên sắt cho phụ nữ mang thai tại Tiền Giang được triển khai hiệu quả hơn. Tuy nhiên, nghiên cứu này chỉ mang tính mô tả, chưa phân tích nguyên nhân – quả.”

6.4.5. Phần Kết luận và Kiến nghị

Phần **kết luận** tóm tắt ngắn gọn phát hiện chính, tránh lặp lại số liệu.

Phần **kiến nghị** đưa ra hướng ứng dụng thực tiễn hoặc gợi ý nghiên cứu tiếp theo.

Ví dụ:

“Tỷ lệ thiếu máu ở phụ nữ mang thai tại Tiền Giang là 32,5%. Các yếu tố liên quan có ý nghĩa thống kê gồm trình độ học vấn và sử dụng viên sắt. Cần tăng cường truyền thông và theo dõi định kỳ hemoglobin trong thai kỳ.”

6.5. Trình bày và diễn giải kết quả định tính trong báo cáo nghiên cứu

6.5.1. Cách trình bày kết quả định tính

Khác với định lượng, dữ liệu định tính được trình bày bằng **chủ đề (themes)** kèm **trích dẫn nguyên văn (quotes)** từ người tham gia.

Nên trình bày ngắn gọn, mạch lạc, xen kẽ giữa mô tả và trích dẫn.

Ví dụ 6.5.1:

Chủ đề 1: Rào cản tuân thủ điều trị

Nhiều bệnh nhân cho biết khó duy trì lịch uống thuốc đều đặn do công việc và trách nhiệm gia đình:

“Tôi hay quên uống thuốc vì phải chăm con nhỏ và đi làm ca đêm.” (Nữ, 32 tuổi)

Một số khác gấp tác dụng phụ khiến họ ngại tiếp tục điều trị:

“Uống thuốc xong tôi chóng mặt và buồn nôn, nên bỏ vài hôm.” (Nam, 45 tuổi)

Lưu ý:

- Dấu ngoặc kép để nguyên phát biểu gốc, không sửa ngữ pháp;
- Giữ ẩn danh người tham gia (chỉ ghi giới tính, tuổi, hoặc vai trò).

6.5.2. Trích dẫn và mã hóa chủ đề

Có thể trình bày bằng bảng mã hóa tóm tắt:

Chủ đề	Mã con (Subcodes)	Trích dẫn tiêu biểu
Rào cản tuân thủ	Quên thuốc, tác dụng phụ, bận rộn	“Tôi hay quên uống thuốc vì phải chăm con nhỏ.”
Động lực điều trị	Hỗ trợ từ gia đình, nhân viên y tế	“Khi bác sĩ nhắc nhở, tôi thấy yên tâm hơn.”

6.5.3. Diễn giải kết quả định tính trong phần Bàn luận

Trong phần thảo luận, kết quả định tính nên được **so sánh hoặc bổ sung cho dữ liệu định lượng**.

Ví dụ:

Phân tích định lượng cho thấy tỷ lệ tuân thủ điều trị chỉ đạt 62%.

Kết quả định tính bổ sung lý do chính là “thiếu hỗ trợ từ gia đình” và “tác dụng phụ thuốc”.

Điều này phù hợp với kết luận của Polit & Beck (2021), rằng sự hỗ trợ xã hội là yếu tố trung gian quan trọng trong quản lý bệnh mạn tính.

6.5.4. Kết hợp định tính – định lượng trong báo cáo hỗn hợp

- Trình bày kết quả định lượng (bảng, biểu đồ) trước, sau đó là phần định tính giải thích;
- Hoặc gộp song song bằng **mô hình hội tụ (Convergent Model)**;
- Nhấn mạnh rằng **dữ liệu định tính giúp “giải nghĩa con số”**, không chỉ minh họa.

6.6. Hình thức trình bày báo cáo

6.6.1. Quy định chung về trình bày

Thành phần	Quy định định dạng	
Khổ giấy	A4, lề trên 2,5 cm, lề trái 3,0 cm, lề phải 2,0 cm, lề dưới 2,5 cm	
Font chữ	Times New Roman, cỡ 13 hoặc 14, giãn dòng 1,5	
Dánh số trang	Ở giữa hoặc góc phải, bắt đầu từ phần Giới thiệu	
Dánh số bảng và hình	Theo chương, ví dụ: Bảng 6.1, Hình 6.1	
Trích dẫn	Chuẩn IEEE, đặt trong ngoặc vuông [n]	
Tài liệu tham khảo	Cuối chương hoặc cuối báo cáo, sắp theo thứ tự trích dẫn	
6.6.2. Trình bày kết quả thống kê trong văn bản	Một số quy tắc khi viết kết quả thống kê trong phần kết quả hoặc bàn luận:	
Nội dung	Cách trình bày đúng	Cách sai thường gặp
p-value	“(p = 0.032)”	“p < 0.05 là có ý nghĩa”
Trung bình ± SD	“Hb = 11.6 ± 1.3 g/dL”	“Hb trung bình là 11.6 và SD 1.3”
Tỷ lệ phần trăm	“32.5%” (không làm tròn 33%)	“Khoảng 33%”
So sánh nhóm	“Không khác biệt có ý nghĩa thống kê (p > 0.05)”	“Không có khác biệt” (thiếu giá trị p)

6.7. Đạo đức trong viết báo cáo nghiên cứu

Viết báo cáo y học phải tuân thủ **chuẩn mực đạo đức học thuật**, cụ thể:

1. **Trung thực và minh bạch dữ liệu:** không làm sai lệch, không “bóp méo” kết quả;
2. **Tôn trọng quyền tác giả:** trích dẫn đầy đủ mọi nguồn tài liệu;
3. **Không đạo văn:** tỷ lệ trùng lặp < 20% theo Turnitin;
4. **Bảo mật thông tin người tham gia nghiên cứu;**
5. **Khai báo xung đột lợi ích và nguồn tài trợ** nếu có.

6.8. Một số lỗi thường gặp trong viết báo cáo

Lỗi thường gặp	Biểu hiện	Cách khắc phục
Thiếu mối liên hệ giữa mục tiêu và kết quả	Mục tiêu A nhưng kết quả không đề cập	Rà soát lại mục tiêu, viết kết quả theo từng mục tiêu
Lặp lại số liệu	Bảng và biểu đồ chia sẻ cùng thông tin	Chỉ chọn 1 hình thức thể hiện

Lỗi thường gặp	Biểu hiện	Cách khắc phục
Dẫn chứng không chuẩn	Trích dẫn không đúng chuẩn IEEE	Sử dụng Zotero / Mendeley để quản lý tài liệu
Không có phần đạo đức nghiên cứu	Thiếu xác nhận Hội đồng đạo đức	Luôn thêm “Nghiên cứu được Hội đồng đạo đức thông qua số .../QĐ”
Viết kết luận lan man	Lặp lại toàn bộ kết quả	Chỉ nêu phát hiện chính và khuyến nghị

6.9. Kết luận chương

- Viết và trình bày báo cáo nghiên cứu là kỹ năng quan trọng trong đào tạo y học.
- Cấu trúc IMRaD là chuẩn mực quốc tế, giúp báo cáo rõ ràng và logic.
- Trình bày kết quả bằng bảng, biểu đồ và diễn giải đúng ngữ pháp thống kê là yêu cầu bắt buộc.
- Tuân thủ đạo đức và trích dẫn đúng chuẩn giúp nâng cao giá trị học thuật và tính minh bạch của nghiên cứu.

Tài liệu tham khảo

- [1] Lưu Ngọc Hoạt, *Nghiên cứu khoa học trong y học*, NXB Y học, Hà Nội, 2014.
- [2] International Committee of Medical Journal Editors (ICMJE), *Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals*, 2023.
- [3] Nguyễn Đức Hinh và cộng sự, *Thống kê y học và phương pháp nghiên cứu khoa học y học*, NXB Y học, 2018.
- [4] WHO, *Health Research Methodology: A Guide for Training in Research Methods*, 3rd ed., Geneva, 2014.
- [5] Field A., *Discovering Statistics Using IBM SPSS Statistics*, 5th ed., SAGE Publications, 2020.