

CHƯƠNG 7. TỔNG QUAN VỀ PHẦN MỀM SPSS 27 – ĐỊNH NGHĨA VÀ TẠO KHUNG BIẾN

7.1. Giới thiệu phần mềm SPSS 27

7.1.1. Tổng quan

SPSS (Statistical Package for the Social Sciences) là phần mềm thống kê được phát triển bởi IBM, hiện được dùng phổ biến trong các nghiên cứu y học, tâm lý học, điều dưỡng và khoa học xã hội [1].

Phiên bản 27 ra mắt năm 2020, hỗ trợ các tính năng nổi bật:

- Giao diện thân thiện, dễ thao tác;
- Tích hợp phân tích thống kê mô tả, kiểm định, hồi quy, và phân tích nhân tố;
- Hỗ trợ xử lý dữ liệu lớn, nhập từ Excel, CSV, EpiData, REDCap;
- Xuất báo cáo tự động sang Word, Excel hoặc PDF.

Lý do chọn SPSS trong nghiên cứu y học:

1. Giao diện trực quan, phù hợp với người không chuyên lập trình;
2. Tích hợp nhiều công cụ thống kê y học (t-test, ANOVA, Chi-square, Logistic Regression...);
3. Có khả năng xử lý và trực quan hóa dữ liệu nhanh chóng;
4. Dễ dàng xuất bảng kết quả cho luận văn hoặc bài báo khoa học.

7.2. Giao diện làm việc của SPSS 27

Khi mở SPSS, xuất hiện 3 cửa sổ chính:

Cửa sổ	Chức năng	Tương tự trong Excel
Data View	Nhập dữ liệu (mỗi dòng = 1 đối tượng, mỗi cột = 1 biến)	Sheet dữ liệu
Variable View	Định nghĩa đặc tính cho từng biến (tên, nhãn, giá trị, loại đo)	Bảng cấu hình biến
Output Viewer	Hiển thị kết quả phân tích, bảng và biểu đồ	Trang kết quả

Hình 7.1. Sơ đồ giao diện SPSS 27

Menu chính: File – Edit – Analyze – Graphs...	
Data View Variable View	
Các cột: biến Các dòng: đối tượng nghiên cứu	

7.3. Khung dữ liệu (Data Frame) trong SPSS

Trong SPSS, mỗi **file dữ liệu (.sav)** gồm hai phần:

1. **Variable View:** định nghĩa biến;
2. **Data View:** chứa dữ liệu của các biến đó.

Một **khung dữ liệu chuẩn** phải đảm bảo:

- Mỗi hàng = 1 đơn vị phân tích (bệnh nhân, mẫu, người tham gia);
- Mỗi cột = 1 biến (tuổi, giới, Hb...);
- Không có ô trống trong biến chính;
- Tên biến không trùng lặp, không có dấu cách.

Ví dụ 7.1. Khung dữ liệu nghiên cứu thiếu máu (rút gọn)

ID Tuoi Gioitinh Hb Thieumau

001	22	2	10.2	1
002	27	1	13.4	0
003	34	2	11.0	0

7.4. Định nghĩa biến trong SPSS

Mỗi biến cần được khai báo cụ thể trong **Variable View** với các trường thông tin sau:

Bảng 7.1. Cấu trúc định nghĩa biến trong Variable View

Trường	Mô tả	Ví dụ
Name	Tên biến (không dấu, không cách)	Tuoi, Hb, Thieumau
Type	Loại dữ liệu	Numeric / String
Width	Số ký tự tối đa	8
Decimals	Số chữ số thập phân	1 hoặc 2
Label	Mô tả đầy đủ biến	“Tuổi của phụ nữ mang thai (năm)”
Values	Mã hóa giá trị (cho biến định tính)	1 = Nam, 2 = Nữ
Missing	Ký hiệu giá trị thiếu (nếu có)	99
Columns	Độ rộng hiển thị	8–10
Align	Căn lề hiển thị	Right
Measure	Thang đo (Nominal, Ordinal, Scale)	Scale
Role	Vai trò (Input, Target, Both)	Input

7.5. Tạo biến và nhập dữ liệu

7.5.1. Các bước thực hiện

1. **Bước 1:** Mở SPSS → *File* → *New* → *Data*.
2. **Bước 2:** Chuyển sang tab *Variable View*.

3. **Bước 3:** Định nghĩa từng biến theo Bảng 7.1.
4. **Bước 4:** Chuyển sang *Data View*, nhập dữ liệu cho từng đối tượng.
5. **Bước 5:** Lưu file dạng *.sav* (ví dụ: Thieumau_TienGiang2024.sav).

7.5.2. Mã hóa biến định tính

Ví dụ 7.2. Mã hóa biến giới tính và nghề nghiệp

Biến	Giá trị	Mã hóa
Gioitinh	Nam / Nữ	1 = Nam, 2 = Nữ

Nghenghiệp Nội trợ, công nhân, nhân viên, khác 1, 2, 3, 4

Sau khi nhập, bật “Value Labels” để hiển thị giá trị nhãn thay vì mã số.

7.6. Nhập dữ liệu từ Excel hoặc EpiData

7.6.1. Từ Excel

1. Lưu tệp Excel dưới dạng *.xlsx* hoặc *.csv*.
2. Trong SPSS: *File* → *Open* → *Data* → chọn file Excel.
3. Chọn “Read variable names from the first row”.
4. Kiểm tra và hiệu chỉnh thang đo cho từng biến.

7.6.2. Từ EpiData

1. Trong EpiData, chọn *Export* → *SPSS (.sav)**.
2. SPSS tự động nhận định nghĩa biến và nhãn giá trị.
3. Mở file trong SPSS, kiểm tra lại cột “Measure” để xác định Nominal/Scale.

7.7. Kiểm tra dữ liệu sau khi nhập

Trước khi phân tích, cần kiểm tra dữ liệu để tránh sai sót.

7.7.1. Kiểm tra giá trị hợp lệ

- Vào **Analyze** → **Descriptive Statistics** → **Frequencies**.
- Chọn các biến định tính (Giới tính, Nghề nghiệp...).
- Chọn “Display frequency tables” → Kiểm tra tần suất.

Nếu có giá trị ngoài phạm vi (ví dụ: “3” trong biến giới tính), cần sửa lại hoặc gán “Missing”.

7.7.2. Kiểm tra phân phối biến định lượng

- Vào **Graphs** → **Legacy Dialogs** → **Histogram**.
- Chọn biến định lượng (Hb, Tuổi...).
- Nếu đường biểu diễn gần chuẩn, có thể sử dụng t-test, ANOVA.
- Nếu lệch nhiều, chọn kiểm định phi tham số (Mann–Whitney, Kruskal–Wallis).

Hình 7.2. Ví dụ biểu đồ Histogram trong SPSS 27 – Biến Hemoglobin

(Miêu tả: Biểu đồ thể hiện Hb trung bình $11,6 \pm 1,3$ g/dL, phân phối lệch nhẹ sang trái.)

7.8. Lưu và quản lý dữ liệu

7.8.1. Đặt tên file dữ liệu

- Tên file nên ngắn gọn, có năm và địa điểm:
Ví dụ: Hypertension_HCMC_2024.sav
- Không dùng ký tự đặc biệt (*, /, %, ...).

7.8.2. Lưu dữ liệu an toàn

- Lưu định kỳ 2 phiên bản: “raw_data.sav” (dữ liệu gốc) và “clean_data.sav” (dữ liệu đã xử lý).
- Lưu trữ kèm mã biến (Variable_Codebook.docx) để tra cứu nhanh.
- Nếu chia sẻ dữ liệu, cần **ẩn thông tin cá nhân** theo quy định đạo đức nghiên cứu.

7.9. Một số lỗi thường gặp trong bước định nghĩa biến

Lỗi	Nguyên nhân	Cách khắc phục
SPSS báo lỗi “Invalid variable name”	Tên biến có dấu hoặc khoảng trắng	Dùng tiếng Anh, không dấu (vd: Tuổi, BMI)
Không hiển thị nhãn giá trị	Chưa gán “Value Labels”	Khai báo lại trong Variable View
Biến định tính bị hiểu thành “Scale”	SPSS tự động gán sai	Chỉnh lại “Measure” → “Nominal”
Dữ liệu bị lệch dòng	File Excel có tiêu đề trống	Kiểm tra hàng đầu tiên của Excel
Không lưu được file	Chưa chọn định dạng .sav	Chọn <i>Save as type: SPSS Statistics (.sav)*</i>

7.10. Kết luận chương

- SPSS 27 là công cụ thống kê mạnh mẽ và thân thiện trong nghiên cứu y học.
- Hiểu rõ cách **định nghĩa và mã hóa biến** giúp tránh sai sót trong phân tích.
- **Variable View** là phần quan trọng nhất khi tạo khung dữ liệu.
- Cần kiểm tra dữ liệu ngay sau khi nhập để đảm bảo tính chính xác và tin cậy.
- Các bước kế tiếp (Chương 8–11) sẽ hướng dẫn chi tiết **phân tích thống kê mô tả, kiểm định giả thuyết, và trình bày kết quả trên SPSS**.

Tài liệu tham khảo

- [1] Field A., *Discovering Statistics Using IBM SPSS Statistics*, 5th ed., SAGE Publications, 2020.
- [2] Lưu Ngọc Hoạt, *Nghiên cứu khoa học trong y học*, NXB Y học, Hà Nội, 2014.
- [3] IBM, *IBM SPSS Statistics 27 Documentation*, IBM Corporation, 2020.
- [4] WHO, *Health Research Methodology: A Guide for Training in Research Methods*, 3rd ed., Geneva, 2014.

[5] Polit D. F. & Beck C. T., *Nursing Research: Generating and Assessing Evidence for Nursing Practice*, 10th ed., Lippincott Williams & Wilkins, 2017.

CHƯƠNG 8. NHẬP VÀ LÀM SẠCH DỮ LIỆU TRONG SPSS 27

8.1. Mục tiêu chương

Sau khi hoàn thành chương này, người học có thể:

1. Nhập dữ liệu từ các nguồn khác nhau (Excel, EpiData, CSV...) vào SPSS 27;
2. Phát hiện và xử lý lỗi nhập liệu, giá trị bị thiếu (missing values), và giá trị ngoại lai (outliers);
3. Chuẩn hóa, tái mã hóa, và tạo biến mới;
4. Xuất bộ dữ liệu đã làm sạch để phân tích thống kê.

8.2. Quy trình nhập và làm sạch dữ liệu

Làm sạch dữ liệu là **bước trung gian giữa thu thập và phân tích**, đảm bảo kết quả thống kê chính xác.

Hình 8.1. Quy trình nhập và xử lý dữ liệu

Thu thập dữ liệu → Nhập dữ liệu → Kiểm tra lỗi → Xử lý thiếu dữ liệu →

Phát hiện ngoại lệ → Chuẩn hóa biến → Xuất dữ liệu sạch (clean data)

Theo WHO (2014), giai đoạn làm sạch dữ liệu chiếm **30–40% thời gian của nghiên cứu y học**, nhưng lại **quyết định hơn 70% độ tin cậy** của kết quả thống kê [1].

8.3. Nhập dữ liệu vào SPSS

8.3.1. Nhập dữ liệu trực tiếp

1. Mở SPSS → *File* → *New* → *Data*
2. Trong **Variable View**, định nghĩa các biến (tên, kiểu, nhãn, mã hóa).
3. Chuyển sang **Data View**, nhập dữ liệu hàng theo hàng.
4. Lưu file dưới dạng *.sav* để đảm bảo định dạng.

Lưu ý:

- Mỗi dòng là **một đối tượng** (bệnh nhân, mẫu, quan sát).
- Mỗi cột là **một biến** (tuổi, giới tính, Hb...).

- Không nhập công thức, chỉ nhập giá trị thô.

8.3.2. Nhập dữ liệu từ Excel

Đây là cách thông dụng nhất khi dữ liệu được thu thập bằng Google Sheets, Excel hoặc EpiData.

Các bước thực hiện:

1. Chuẩn bị dữ liệu trong Excel:
 - Hàng đầu tiên là tên biến (không có dấu, không cách).
 - Mỗi dòng = 1 đối tượng.
 - Không để trống dòng hoặc cột.
2. Trong SPSS:
 - Chọn *File* → *Open* → *Data*.
 - Chọn “Files of type: Excel (*.xlsx)”.
 - Đánh dấu “Read variable names from the first row of data”.
 - Kiểm tra dữ liệu sau khi nhập.

Ví dụ:

File Thieumau_TienGiang2024.xlsx có 5 biến: ID, Tuổi, Gioitinh, Hb, Thieumau.

Sau khi mở trong SPSS, tự động tạo cấu trúc dữ liệu như Bảng 8.1.

Bảng 8.1. Dữ liệu sau khi nhập từ Excel

ID Tuổi Gioitinh Hb Thieumau

001 22 2 10.2 1

002 27 1 13.4 0

003 34 2 11.0 0

8.3.3. Nhập dữ liệu từ EpiData hoặc CSV

Từ EpiData:

1. Trong EpiData: *File* → *Export Data* → *SPSS (.sav)*.
2. SPSS sẽ tự động nhận các định nghĩa biến, nhãn và thang đo.

Từ CSV:

1. *File* → *Read Text Data* → *chọn file.csv*.
2. Chọn dấu phân cách (thường là “,”).
3. Đặt định dạng cột (Numeric hoặc String).

8.4. Kiểm tra dữ liệu sau khi nhập

8.4.1. Kiểm tra tần suất và mã hóa biến định tính

Vào menu:

Analyze → **Descriptive Statistics** → **Frequencies**

Chọn biến cần kiểm tra (Giới tính, Nghề nghiệp, Tình trạng thiếu máu, v.v.).

Kết quả hiển thị tần suất từng giá trị — nếu có mã “3” trong biến giới tính (chỉ có 1=Nam, 2=Nữ), cần sửa lại hoặc gán là *missing*.

Bảng 8.2. Kiểm tra tần suất biến giới tính

Giá trị	Tần suất	Tỷ lệ (%)
---------	----------	-----------

1 (Nam)	190	54.3
---------	-----	------

2 (Nữ)	160	45.7
--------	-----	------

3 (Lỗi nhập)	1	0.3
--------------	---	-----

→ Cần kiểm tra và sửa dòng có mã “3”.

8.4.2. Kiểm tra giá trị tối đa – tối thiểu của biến định lượng

Analyze → Descriptive Statistics → Descriptives

Chọn các biến như Tuổi, Hb, BMI.

Ví dụ:

Biến	N	Min	Max	Mean	SD
------	---	-----	-----	------	----

Tuoi	350	17	68	28.4	6.2
------	-----	----	----	------	-----

Hb	350	6.0	18.0	11.6	1.3
----	-----	-----	------	------	-----

Nếu giá trị Hb = 60 hoặc tuổi = 160 xuất hiện → lỗi nhập liệu, cần loại bỏ hoặc sửa.

8.5. Xử lý dữ liệu bị thiếu (Missing Data)

8.5.1. Phát hiện dữ liệu thiếu

Trong SPSS, giá trị thiếu hiển thị bằng dấu chấm (.) trong ô dữ liệu.

Thống kê nhanh bằng:

Analyze → Descriptive Statistics → Frequencies → Check “Display missing values.”

Bảng 8.3. Tỷ lệ dữ liệu thiếu theo biến

Biến	Số giá trị thiếu	Tỷ lệ (%)
------	------------------	-----------

Tuoi	0	0.0
------	---	-----

Hb	2	0.6
----	---	-----

Nghenghiiep	15	4.3
-------------	----	-----

Nếu tỷ lệ dữ liệu thiếu <5%, có thể loại bỏ. Nếu >10%, nên xem xét thay thế (imputation).

8.5.2. Xử lý dữ liệu thiếu

1. **Loại bỏ hàng có giá trị thiếu (Listwise Deletion):**

Dùng khi dữ liệu thiếu ít, không ảnh hưởng kết quả.

2. **Thay thế bằng giá trị trung bình hoặc trung vị:**

○ **Transform → Replace Missing Values → Series Mean.**

3. **Thay thế nâng cao:**

Dùng mô hình hồi quy hoặc multiple imputation (cần kiến thức thống kê nâng cao).

8.6. Phát hiện và xử lý giá trị ngoại lai (Outlier)

8.6.1. Phát hiện bằng Boxplot

Graphs → Boxplot → Simple → Define → Chọn biến cần xem.

Giá trị nằm ngoài $1.5 \times \text{IQR}$ được coi là ngoại lệ.

Hình 8.1. Ví dụ Boxplot biến Hemoglobin (Hb)

(Hình minh họa: Một điểm cực trị 18 g/dL được đánh dấu, nghi ngờ sai sót nhập liệu.)

8.6.2. Kiểm tra bằng Descriptive Statistics

Analyze → Descriptives → Save standardized values as variables (Z-scores).

Các giá trị có $|Z| > 3 \rightarrow$ nghi ngờ ngoại lệ.

Xem trong Data View biến ZHb: nếu ZHb > 3 hoặc $< -3 \rightarrow$ kiểm tra lại hồ sơ gốc.

8.7. Chuẩn hóa và tái mã hóa biến

8.7.1. Tái mã hóa (Recode Variables)

- Vào **Transform → Recode into Same Variables** (hoặc Different).
- Dùng khi muốn gộp nhóm, thay đổi thang đo hoặc mã giá trị.

Ví dụ: Gộp tuổi thành 3 nhóm:

Tuổi (năm) Mã mới Nhóm tuổi

<25	1	Trẻ
25–34	2	Trung niên
≥35	3	Lớn tuổi

Thao tác:

Transform → Recode into Different Variables → Old and New Values

Old: 0–24 = 1, 25–34 = 2, 35–99 = 3 → Output variable: Nhomtuoi

8.7.2. Tạo biến mới (Compute Variable)

Transform → Compute Variable

Công thức ví dụ:

$$BMI = \frac{\text{Cân nặng} (\text{Chiều ao})^2}{10000}$$

Nhập trong SPSS:

Target Variable: BMI

Numeric Expression: Cannang / ((Chieuca/100)**2)

8.8. Kiểm tra dữ liệu sau làm sạch

Khi hoàn tất quá trình làm sạch, cần kiểm tra lại:

1. Không còn giá trị ngoài phạm vi;
2. Không còn mã sai hoặc thiếu biến;
3. Các biến định nghĩa đúng loại (Nominal / Scale / Ordinal);
4. Kết quả mô tả hợp lý (Mean, SD, Min, Max phù hợp lâm sàng).

Xuất dữ liệu sạch:

File → Save As → Tên file: “Cleaned_Data_Study2024.sav”

8.9. Ghi chú quản lý dữ liệu (Data Documentation)

Mỗi bộ dữ liệu nghiên cứu y học cần có:

1. **Data Codebook:** mô tả từng biến, giá trị mã hóa, thang đo;
2. **Logbook:** nhật ký các lần chỉnh sửa dữ liệu;
3. **File dự phòng:** bản sao dữ liệu gốc chưa chỉnh sửa.

Ví dụ 8.1. Mẫu trích Data Codebook

Tên biến	Mô tả	Thang đo	Ghi chú
ID	Mã đối tượng	Nominal	Duy nhất
Gioitinh	Giới tính (1=Nam, 2=Nữ)	Nominal	Không trống
Hb	Nồng độ Hemoglobin (g/dL)	Scale	7–18 g/dL
Thieumau	Tình trạng thiếu máu (0=Không, 1=Có)	Nominal	Theo Hb<11

8.10. Tóm tắt thao tác làm sạch dữ liệu trong SPSS 27

Mục tiêu	Lệnh SPSS tương ứng
Kiểm tra giá trị ngoài phạm vi	Analyze → Frequencies
Phát hiện giá trị bất thường	Graphs → Boxplot
Tái mã hóa biến	Transform → Recode
Tạo biến mới	Transform → Compute Variable
Kiểm tra dữ liệu thiếu	Analyze → Descriptives
Thay thế dữ liệu thiếu	Transform → Replace Missing Values
Lưu dữ liệu sạch	File → Save As

8.11. Kết luận chương

- Làm sạch dữ liệu là bước **thiết yếu để đảm bảo độ chính xác và độ tin cậy** trong phân tích thống kê y học.
- SPSS 27 cung cấp đầy đủ công cụ giúp **kiểm tra, mã hóa, phát hiện và xử lý dữ liệu lỗi**.
- Một bộ dữ liệu sạch giúp kết quả thống kê không sai lệch và đảm bảo có thể tái lập.
- Sinh viên và nhà nghiên cứu nên lưu cả dữ liệu gốc và dữ liệu đã xử lý, kèm tài liệu “Codebook” để thuận tiện khi kiểm chứng sau này.

Tài liệu tham khảo

- [1] WHO, *Health Research Methodology: A Guide for Training in Research Methods*, 3rd ed., Geneva, 2014.
- [2] Field A., *Discovering Statistics Using IBM SPSS Statistics*, 5th ed., SAGE Publications, 2020.
- [3] IBM, *IBM SPSS Statistics 27 Documentation*, IBM Corporation, 2020.
- [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 9. TÍNH TOÁN VÀ XỬ LÝ SỐ LIỆU THỐNG KÊ TRONG SPSS 27

9.1. Mục tiêu chương

Sau khi hoàn thành chương này, người học có thể:

1. Thực hiện các phép **tính toán, mã hóa và chuyển đổi biến** trong SPSS;
2. **Lọc và chọn dữ liệu (Select Cases)** theo điều kiện;
3. **Tạo nhóm phân tích (Split File)** để so sánh các nhóm đối tượng;
4. Xuất dữ liệu thống kê đã xử lý để sử dụng trong các phân tích tiếp theo.

9.2. Tổng quan về xử lý dữ liệu trong SPSS

Sau khi làm sạch dữ liệu, bước tiếp theo là **xử lý biến (Data Transformation)** — giai đoạn biến dữ liệu thô thành dữ liệu sẵn sàng phân tích.

Các thao tác chính bao gồm:

- **Compute Variable** – Tạo biến mới từ phép tính;
- **Recode Variable** – Thay đổi giá trị hoặc nhóm biến;
- **Select Cases** – Lọc dữ liệu theo điều kiện;
- **Split File** – Chia dữ liệu thành các nhóm phân tích;
- **Aggregate** – Tính tổng, trung bình, hoặc tỷ lệ theo nhóm.

9.3. Tính toán và tạo biến mới (Compute Variable)

9.3.1. Mục đích

Tạo biến mới dựa trên công thức hoặc mối quan hệ giữa các biến hiện có.

Ví dụ: tính **chỉ số BMI, tổng điểm**, hoặc **chỉ số trung bình cộng**.

9.3.2. Thao tác thực hiện

1. Vào **Transform → Compute Variable**
2. Ô *Target Variable*: nhập tên biến mới (không dấu, không cách)
3. Ô *Numeric Expression*: nhập công thức
4. Bấm **OK** → SPSS tạo biến mới ở cột cuối cùng.

Ví dụ 9.1. Tính chỉ số BMI

$$BMI = \frac{\text{Cân nặng (kg)}}{(\text{Chiều cao (m)})^2}$$

Trong SPSS:

- Target Variable: BMI

- Numeric Expression: Cannang / ((Chieucac/100)**2)

Kết quả: biến BMI hiển thị ở Data View với giá trị tương ứng từng đối tượng.

9.4. Tái mã hóa biến (Recode Variable)

9.4.1. Khi nào cần tái mã hóa?

- Khi muốn **gộp nhóm** (tuổi → nhóm tuổi);
- Khi cần **chuyển đổi giá trị** (1=Nam, 2=Nữ → 0/1);
- Khi cần **tạo biến phân loại nhị phân** từ biến định lượng.

9.4.2. Thao tác thực hiện

Vào **Transform → Recode into Different Variables**

1. Chọn biến cần tái mã hóa;
2. Nhập tên biến mới (ví dụ: Nhomtuoi);
3. Chọn *Old and New Values*;
4. Đặt quy tắc:
 - 0–24 = 1,
 - 25–34 = 2,
 - 35–99 = 3.
5. Nhấn *Continue* → *OK*.

Bảng 9.1. Ví dụ kết quả tái mã hóa nhóm tuổi

Tuoi Nhomtuoi Diễn giải

22	1	Dưới 25 tuổi
27	2	25–34 tuổi
36	3	≥35 tuổi

9.4.3. Tái mã hóa giá trị nhị phân

Ví dụ: Biến Thieumau ban đầu là 0 = Không, 1 = Có, 2 = Không rõ.

Cần loại 2 và chỉ giữ 0,1:

Old value: 0 → 0; 1 → 1; 2 → System-missing.

9.5. Gộp nhóm hoặc tạo biến phân loại (Visual Binning)

Visual Binning cho phép tạo nhóm tự động dựa trên phân bố dữ liệu.

Thao tác:

1. *Transform → Visual Binning*
2. Chọn biến (vd: Hb)
3. SPSS gợi ý nhóm theo phân vị (tertile/quartile).
4. Đặt tên nhóm:
 - <11.0 = “Thiếu máu”
 - ≥11.0 = “Không thiếu máu”

Hình 9.1. Minh họa cửa sổ Visual Binning trong SPSS

(Hình mô tả: thanh phân phối giá trị Hb, đường cắt tại 11.0 g/dL).

9.6. Lọc dữ liệu theo điều kiện (Select Cases)

9.6.1. Mục đích

Lọc các đối tượng phù hợp để phân tích theo nhóm, loại bỏ dữ liệu không cần thiết.

9.6.2. Thao tác

1. Vào **Data** → **Select Cases**
2. Chọn “If condition is satisfied” → *If...*
3. Nhập điều kiện logic (ví dụ: Gioitinh = 1 hoặc Tuổi \geq 25).
4. Chọn “Filter out unselected cases”.
5. Dữ liệu không thỏa điều kiện sẽ bị gạch chéo (không xóa).

Ví dụ 9.2:

Phân tích riêng nhóm phụ nữ mang thai \geq 25 tuổi:

Condition: Tuổi \geq 25 AND Gioitinh = 2

9.7. Phân nhóm dữ liệu để so sánh (Split File)

9.7.1. Mục đích

Khi muốn SPSS **phân tích song song các nhóm** (nam/nữ, vùng miền, nghề nghiệp...) mà không cần tạo nhiều file.

9.7.2. Thao tác

1. Vào **Data** → **Split File**
2. Chọn “Compare groups” hoặc “Organize output by groups”
3. Chọn biến phân nhóm (ví dụ: Gioitinh)
4. Chạy lệnh phân tích → SPSS tự động tách kết quả theo nhóm.

Ví dụ 9.3.

Tính trung bình Hb cho nam và nữ riêng biệt bằng Split File.

Giới tính N Hb trung bình (g/dL) SD

Nam	190 13.0	1.1
Nữ	160 11.2	1.3

9.8. Kết hợp dữ liệu (Merge Files)

Trong nghiên cứu lớn, dữ liệu thường chia thành nhiều bảng (ví dụ: thông tin hành chính và xét nghiệm).

SPSS cho phép kết hợp dễ dàng.

9.8.1. Ghép theo cột (Add Variables)

Data → Merge Files → Add Variables

- Chọn file thứ hai, xác định biến chung (ID).

- SPSS ghép các biến theo hàng ID.

9.8.2. Ghép theo hàng (Add Cases)

Data → Merge Files → Add Cases

- Dùng khi hai file có cùng cấu trúc biến nhưng khác đối tượng.

Lưu ý: Trước khi gộp, cần đảm bảo:

- Tên biến và loại dữ liệu trùng khớp;
- Không trùng ID;
- Cùng đơn vị đo.

9.9. Tạo biến tổng hợp (Aggregate)

9.9.1. Mục đích

Tính **trung bình, tổng, hoặc tỷ lệ** theo nhóm (ví dụ: trung bình Hb theo xã).

9.9.2. Thao tác

1. **Data → Aggregate**
2. “Break variable”: chọn biến nhóm (vd: Xaphuong)
3. “Aggregated variables”: chọn biến cần tính trung bình (Hb)
4. Chọn “Function: Mean”
5. Output file: Hb_Xaphuong.sav

Ví dụ 9.4. Kết quả tổng hợp trung bình Hb theo xã

Xã/Phường N Hb trung bình (g/dL)

Long Hưng 60 11.2

Mỹ Phước 80 11.7

Tân Phú 70 12.0

9.10. Gán điều kiện và lọc dữ liệu ngẫu nhiên

9.10.1. Lọc ngẫu nhiên

Data → Select Cases → Random sample of cases → “Approximately” → Nhập tỷ lệ (ví dụ: 0.3 để chọn 30%).

9.10.2. Gán điều kiện (Compute with IF)

Trong cửa sổ Compute Variable, có thể đặt điều kiện bằng “If...”.

Ví dụ:

Tạo biến “TiemSat” = 1 nếu Hb < 11, ngược lại 0.

Target Variable: TiemSat

Numeric Expression: 1

If: Hb < 11

9.11. Sử dụng cú pháp (Syntax) để tự động hóa

SPSS cho phép ghi lại thao tác thành **cú pháp lệnh (Syntax)**, giúp tái lập kết quả và chia sẻ quy trình.

9.11.1. Ghi lại cú pháp

- Mỗi lần thực hiện thao tác → chọn *Paste* thay vì *OK*.
- Cửa sổ Syntax hiển thị lệnh SPSS (.sps).
- Lưu lại bằng *File* → *Save As* → “*Analysis_Hb.sps*”.

9.11.2. Chạy cú pháp

Chọn đoạn mã → Run → Selection.

Ví dụ 9.5. Lệnh cú pháp tạo biến BMI

COMPUTE BMI = Cannang / ((Chieuca/100)**2).

EXECUTE.

9.12. Tổng hợp các lệnh xử lý dữ liệu thông dụng trong SPSS 27

Bảng 9.2. Các lệnh thường dùng

Mục tiêu	Lệnh	Ghi chú
Tạo biến mới	Transform → Compute Variable	Tính BMI, điểm trung bình...
Tái mã hóa biến	Transform → Recode	Gộp nhóm, mã hóa mới
Lọc dữ liệu	Data → Select Cases	Lọc theo giới tính, tuổi
Chia nhóm phân tích	Data → Split File	So sánh nhóm
Tổng hợp dữ liệu	Data → Aggregate	Trung bình theo nhóm
Gộp file	Data → Merge Files	Kết hợp dữ liệu
Tính tổng/trung bình	Analyze → Descriptive Statistics	Thống kê mô tả
Lưu file đã xử lý	File → Save As	“Cleaned_processed.sav”

9.13. Kết luận chương

- Chương này cung cấp các thao tác **tính toán, mã hóa, gộp, và lọc dữ liệu** – nền tảng cho mọi phân tích thống kê.
- SPSS 27 cho phép thực hiện hầu hết các bước xử lý bằng thao tác trực quan hoặc cú pháp tự động.
- Việc lưu lại cú pháp giúp đảm bảo **tính tái lập (reproducibility)** – tiêu chuẩn vàng trong nghiên cứu y học hiện đại.
- Sau khi xử lý hoàn tất, dữ liệu sẵn sàng để **phân tích mô tả (Chương 10)** và **kiểm định giả thuyết (Chương 11)**.

Tài liệu tham khảo

[1] Field A., *Discovering Statistics Using IBM SPSS Statistics*, 5th ed., SAGE Publications, 2020.

- [2] IBM, *IBM SPSS Statistics 27 Documentation*, IBM Corporation, 2020.
- [3] Lưu Ngọc Hoạt, *Nghiên cứu khoa học trong y học*, NXB Y học, Hà Nội, 2014.
- [4] WHO, *Health Research Methodology: A Guide for Training in Research Methods*, 3rd ed., Geneva, 2014.
- [5] Polit D. F. & Beck C. T., *Nursing Research: Generating and Assessing Evidence for Nursing Practice*, 10th ed., Lippincott Williams & Wilkins, 2017.

CHƯƠNG 10. THỐNG KÊ MÔ TẢ VÀ VẼ BIỂU ĐỒ TRONG SPSS 27

10.1. Mục tiêu chương

Sau khi học xong chương này, sinh viên có thể:

1. Thực hiện các thống kê mô tả trong SPSS 27 cho cả biến định tính và định lượng;
2. Biết cách chọn chỉ số thống kê phù hợp theo loại thang đo;
3. Biết cách tạo, tùy chỉnh và xuất biểu đồ (cột, tròn, histogram, boxplot);
4. Viết kết quả mô tả đúng quy tắc báo cáo y học (theo chuẩn APA hoặc Vancouver).

10.2. Khái niệm và ý nghĩa của thống kê mô tả

Thống kê mô tả (Descriptive Statistics) là bước đầu tiên trong phân tích dữ liệu, dùng để **tóm tắt, mô tả và trình bày các đặc điểm của mẫu nghiên cứu**.

Theo Bland (2015) [1], thống kê mô tả giúp:

- Tạo cái nhìn tổng quan về phân bố dữ liệu;
- Phát hiện lỗi hoặc giá trị ngoại lệ;
- Hỗ trợ lựa chọn phép kiểm định thích hợp (tham số hoặc phi tham số);
- Diễn giải ý nghĩa lâm sàng trước khi suy luận thống kê.

10.3. Chọn loại thống kê mô tả phù hợp

Tùy theo **loại biến số và thang đo**, ta chọn các chỉ số mô tả khác nhau.

Bảng 10.1. Lựa chọn chỉ số mô tả theo loại biến

Loại biến	Thang đo	Chỉ số trung tâm	Chỉ số phân tán	Ví dụ
Định tính danh nghĩa	Nominal	Tần số, tỷ lệ (%)	—	Giới tính, nghề nghiệp
Định tính thứ bậc	Ordinal	Trung vị (Median), Tần suất	Khoảng tứ phân vị (IQR)	Mức độ đau: nhẹ, vừa, nặng
Định lượng (phân phối chuẩn)	Scale (chuẩn)	Trung bình (Mean)	Độ lệch chuẩn (SD)	Huyết áp, Hb
Định lượng (không chuẩn)	Scale (không chuẩn)	Trung vị (Median)	IQR	BMI, CRP, AST

10.4. Thống kê mô tả biến định lượng

10.4.1. Cách thực hiện trong SPSS

1. Vào **Analyze** → **Descriptive Statistics** → **Descriptives**;
2. Chọn các biến định lượng (ví dụ: Tuổi, Hb, BMI);
3. Bấm **Options** → **chọn Mean, Std. Deviation, Minimum, Maximum**;
4. Nhấn **OK** để xem kết quả.

10.4.2. Ví dụ

Bảng 10.2. Kết quả mô tả biến định lượng

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
BMI (kg/m ²)	350	21.8	2.5	16.3	29.1

10.4.3. Diễn giải

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

Nồng độ Hb trung bình là $11,6 \pm 1,3$ g/dL, nằm trong giới hạn sinh lý bình thường.

Giá trị nhỏ nhất là 8,5 và cao nhất là 14,8 g/dL.

10.5. Thống kê mô tả biến định tính

10.5.1. Cách thực hiện

1. Vào **Analyze** → **Descriptive Statistics** → **Frequencies**;
2. Chọn các biến định tính (Giới tính, Nghề nghiệp, Thiếu máu...);
3. Đánh dấu “Display frequency tables”;
4. Chọn “Charts → Bar Chart” nếu muốn kèm biểu đồ.

10.5.2. Ví dụ

Bảng 10.3. Phân bố đặc điểm định tính

Đặc điểm	Số lượng (n)	Tỷ lệ (%)
Giới tính		
– Nam	190	54.3
– Nữ	160	45.7
Tình trạng thiếu máu		
– Có	114	32.5
– Không	236	67.5

10.5.3. Diễn giải

Trong tổng số 350 đối tượng nghiên cứu, tỷ lệ nữ chiếm 45,7%.

Có 32,5% phụ nữ mang thai bị thiếu máu, cho thấy vấn đề dinh dưỡng vẫn đáng quan tâm tại cộng đồng.

10.6. Kiểm tra phân phối của biến định lượng

10.6.1. Histogram

Thao tác:

Graphs → Legacy Dialogs → Histogram → Chọn biến (ví dụ: Hb) → OK

SPSS tạo biểu đồ tần suất kèm đường cong chuẩn.

Hình 10.1. Phân phối nồng độ Hemoglobin (Hb)

(Hình minh họa: biểu đồ histogram có dạng đối xứng, trung bình ~11.6 g/dL, SD 1.3.)

10.6.2. Kiểm định phân phối chuẩn

Analyze → Descriptive Statistics → Explore → Plots → chọn “Normality plots with tests”.

Kết quả gồm:

- Kiểm định **Shapiro–Wilk** ($n < 50$) hoặc **Kolmogorov–Smirnov** ($n \geq 50$);
- Nếu $p > 0.05$ → phân phối chuẩn;
- Nếu $p < 0.05$ → không chuẩn → dùng phép kiểm định phi tham số.

10.7. Các loại biểu đồ thường dùng trong y học

Bảng 10.4. Loại biểu đồ và mục đích sử dụng

Loại biểu đồ	Mục đích	Loại biến	Ví dụ
Biểu đồ cột (Bar Chart)	So sánh tần suất giữa các nhóm	Định tính	Tỷ lệ thiếu máu theo nhóm tuổi
Biểu đồ tròn (Pie Chart)	Thể hiện tỷ lệ phần trăm	Định tính	Phân bố giới tính
Histogram	Kiểm tra phân phối chuẩn	Định lượng	Phân bố BMI
Biểu đồ hộp (Boxplot)	So sánh phân bố, phát hiện ngoại lệ	Định lượng	Hb giữa các nhóm nghề
Biểu đồ đường (Line Chart)	Xu hướng theo thời gian	Liên tục	Tỷ lệ bệnh qua năm

10.8. Thực hành tạo biểu đồ trong SPSS 27

10.8.1. Biểu đồ cột (Bar Chart)

Thao tác:

Graphs → Chart Builder → Bar → chọn “Simple Bar”

X-axis: biến định tính (Giới tính)

Y-axis: biến định lượng (Hb trung bình)

Element Properties → chọn “Display error bars” (95% CI).

Hình 10.2. Biểu đồ cột: Trung bình Hb theo giới tính

(Hình mô tả: Nam 13.0 g/dL, Nữ 11.2 g/dL.)

10.8.2. Biểu đồ tròn (Pie Chart)

Thao tác:

Graphs → Chart Builder → Pie → chọn biến định tính (Thiếu máu).

Chọn “Display % in Slices”.

Hình 10.3. Biểu đồ tròn: Tình trạng thiếu máu

Loại	Tỷ lệ (%)
Có	32.5
Không	67.5

10.8.3. Biểu đồ hộp (Boxplot)

Thao tác:

Graphs → Boxplot → Define Simple → Y: Hb, X: Nhóm tuổi.

Hình 10.4. Biểu đồ hộp: Hb theo nhóm tuổi

(Mô tả: nhóm ≥ 35 tuổi có Hb trung bình thấp hơn rõ rệt, xuất hiện một ngoại lệ Hb=8.5 g/dL.)

10.9. Tùy chỉnh và xuất biểu đồ

10.9.1. Tùy chỉnh giao diện biểu đồ

- Nhấp đôi vào biểu đồ để mở **Chart Editor**;
- Có thể chỉnh màu, font chữ, tiêu đề, trục X/Y;
- Dùng **Elements** → **Show Data Labels** để hiển thị giá trị phần trăm hoặc trung bình.

10.9.2. Xuất biểu đồ

File → Export → Chọn định dạng (Word, Excel, PDF, hoặc PNG).

Khuyến nghị: dùng **.PNG (300 dpi)** để chèn vào bài báo hoặc luận văn.

10.10. Trình bày kết quả thống kê mô tả theo chuẩn y học

10.10.1. Quy tắc trình bày

1. Viết rõ **N**, **trung bình \pm SD**, hoặc **Median (IQR)**;
2. Đối với biến định tính, nêu rõ **n (%)**;
3. Không viết số 0 thập phân không cần thiết ($p = 0.032$, không phải 0.0320);
4. Dùng dấu chấm thay cho dấu phẩy trong số thập phân (11.6 thay vì 11,6).

10.10.2. Ví dụ minh họa

“Tuổi trung bình của đối tượng nghiên cứu là 28.4 ± 6.2 tuổi (từ 18 đến 45).

Tỷ lệ thiếu máu là 32.5% (114/350), trong đó nhóm phụ nữ ≥ 35 tuổi chiếm tỷ lệ cao nhất (36.9%).

Nồng độ Hb trung bình ở nam là 13.0 ± 1.1 g/dL, cao hơn đáng kể so với nữ (11.2 ± 1.3 g/dL).”

10.11. Tổng hợp thao tác thống kê mô tả và biểu đồ

Bảng 10.5. Tóm tắt các thao tác trong SPSS

Mục tiêu	Lệnh SPSS	Kết quả
Thống kê mô tả biến định lượng	Analyze → Descriptive Statistics → Descriptives	Mean, SD, Min, Max
Thống kê mô tả biến định tính	Analyze → Descriptive Statistics → Frequencies	n, %
Kiểm tra phân phối	Analyze → Explore → Plots	Histogram, p-value Shapiro–Wilk
Biểu đồ cột	Graphs → Chart Builder → Bar	So sánh nhóm
Biểu đồ tròn	Graphs → Chart Builder → Pie	Tỷ lệ phần trăm
Biểu đồ hộp	Graphs → Boxplot	So sánh phân bố
Lưu biểu đồ	File → Export	.PNG / .DOCX

10.12. Kết luận chương

- Thống kê mô tả là bước nền tảng để hiểu rõ đặc điểm mẫu nghiên cứu trước khi tiến hành kiểm định.
- SPSS 27 cung cấp giao diện thân thiện, giúp mô tả dữ liệu nhanh chóng và tạo biểu đồ chuyên nghiệp.
- Biết lựa chọn chỉ số phù hợp và trình bày kết quả theo chuẩn quốc tế giúp báo cáo khoa học có giá trị và dễ công bố.
- Kỹ năng mô tả và trực quan hóa dữ liệu là yêu cầu bắt buộc đối với sinh viên và nhà nghiên cứu y học hiện đại.

Tài liệu tham khảo

- [1] Bland M., *An Introduction to Medical Statistics*, 4th ed., Oxford University Press, 2015.
- [2] Field A., *Discovering Statistics Using IBM SPSS Statistics*, 5th ed., SAGE Publications, 2020.
- [3] IBM, *IBM SPSS Statistics 27 Documentation*, IBM Corporation, 2020.
- [4] Polit D. F. & Beck C. T., *Nursing Research: Generating and Assessing Evidence for Nursing Practice*, 10th ed., Lippincott Williams & Wilkins, 2017.
- [5] Lưu Ngọc Hoạt, *Nghiên cứu khoa học trong y học*, NXB Y học, Hà Nội, 2014.

CHƯƠNG 11. KIỂM ĐỊNH GIẢ THUYẾT THỐNG KÊ TRONG SPSS 27

11.1. Mục tiêu chương

Sau khi học xong chương này, người học có thể:

1. Hiểu được bản chất của kiểm định giả thuyết thống kê trong nghiên cứu y học;
2. Lựa chọn và thực hiện đúng phép kiểm định phù hợp với loại biến và mục tiêu nghiên cứu;
3. Diễn giải kết quả kiểm định: giá trị p , khoảng tin cậy (CI), odds ratio (OR), risk ratio (RR);
4. Trình bày kết quả thống kê theo chuẩn quốc tế (APA hoặc Vancouver).

11.2. Khái niệm và nguyên tắc kiểm định giả thuyết

11.2.1. Giả thuyết thống kê

Trong phân tích suy luận, ta bắt đầu với hai giả thuyết:

- **Giả thuyết không (H_0):** Không có sự khác biệt hoặc mối liên hệ giữa các nhóm.
- **Giả thuyết đối (H_1):** Có sự khác biệt hoặc mối liên hệ giữa các nhóm.

Ví dụ:

H_0 : Không có khác biệt về nồng độ Hb giữa nam và nữ.

H_1 : Có khác biệt về nồng độ Hb giữa nam và nữ.

11.2.2. Mức ý nghĩa (α)

- Thường chọn $\alpha = 0,05$ (nghĩa là chấp nhận 5% khả năng sai lầm loại I).
- Nếu $p < 0,05 \rightarrow$ bác bỏ H_0 (có sự khác biệt có ý nghĩa thống kê).

11.2.3. Quy trình kiểm định giả thuyết

1. Xác định giả thuyết H_0 và H_1 ;
2. Chọn phép kiểm định thích hợp;
3. Thực hiện kiểm định bằng SPSS;
4. Đọc giá trị p và so sánh với α ;
5. Kết luận và diễn giải ý nghĩa.

Hình 11.1. Sơ đồ quy trình kiểm định giả thuyết

Xác định vấn đề \rightarrow Đặt H_0 & $H_1 \rightarrow$ Chọn phép kiểm định \rightarrow Phân tích SPSS \rightarrow So sánh $p \rightarrow$ Kết luận

11.3. Phân loại kiểm định thống kê

Bảng 11.1. Phân loại các phép kiểm định phổ biến trong nghiên cứu y học

Mục tiêu	Loại biến	Phân phối	Phép kiểm định	Ví dụ
So sánh trung bình giữa 2 nhóm độc lập	Định lượng	Chuẩn	Independent t-test	Hb giữa nam và nữ
So sánh trung bình giữa 2 nhóm liên quan	Định lượng	Chuẩn	Paired t-test	Hb trước và sau điều trị
So sánh trung bình ≥ 3 nhóm	Định lượng	Chuẩn	One-way ANOVA	Tuổi trung bình giữa 3 nghề
So sánh ≥ 2 nhóm không chuẩn	Định lượng	Không chuẩn	Kruskal–Wallis, Mann–Whitney	BMI giữa nhóm tuổi
So sánh tỷ lệ giữa 2 nhóm	Định tính	–	Chi-square hoặc Fisher's exact	Tỷ lệ thiếu máu giữa nam/nữ
So sánh tỷ lệ ≥ 3 nhóm	Định tính	–	Chi-square	Tỷ lệ bệnh theo trình độ học vấn
Mối tương quan giữa 2 biến định lượng	Định lượng	Chuẩn	Pearson correlation	Tuổi & Hb
Mối tương quan không chuẩn	Định lượng	Không chuẩn	Spearman correlation	BMI & tuổi thai
Dự đoán biến phụ thuộc định tính	Định tính (2 giá trị)	–	Logistic Regression	Yếu tố liên quan đến thiếu máu

11.4. Kiểm định cho biến định lượng

11.4.1. So sánh trung bình giữa 2 nhóm độc lập – Independent t-test

Ví dụ: So sánh Hb trung bình giữa nam và nữ.

Thao tác SPSS:

Analyze → Compare Means → Independent-Samples T Test

Test Variable(s): Hb

Grouping Variable: Gioitinh

Define Groups: 1 = Nam, 2 = Nữ → OK.

Kết quả mẫu (Bảng 11.2):

Giới tính	N	Mean (g/dL)	SD	t	df	p
Nam	190	13.0	1.1	12.56	348	<0.001
Nữ	160	11.2	1.3			

Kết luận:

Có sự khác biệt có ý nghĩa thống kê về nồng độ Hb giữa nam và nữ ($p < 0.001$).

11.4.2. So sánh trung bình trước và sau can thiệp – Paired t-test

Ví dụ: So sánh Hb trước và sau bổ sung sắt 3 tháng.

Thao tác:

Analyze → Compare Means → Paired-Samples T Test

Pair: Hb_Truoc WITH Hb_Sau → OK.

Kết quả mẫu:

Biến	Mean	SD	t	df	p
Hb_Sau – Hb_Truoc	+1.2	0.8	9.45	99	<0.001

Kết luận:

Sau bổ sung sắt, Hb trung bình tăng đáng kể ($p < 0.001$).

11.4.3. So sánh ≥ 3 nhóm – One-way ANOVA

Ví dụ: So sánh Hb trung bình theo 3 nhóm tuổi (<25, 25–34, ≥ 35).

Thao tác:

Analyze → Compare Means → One-Way ANOVA

Dependent List: Hb

Factor: Nhomtuoi → OK.

Nếu $p < 0.05$ → thực hiện Post Hoc (Bonferroni hoặc Tukey) để xác định nhóm khác biệt.

Kết quả mẫu:

Nhóm tuổi	N	Mean	SD
<25	85	11.9	1.1
25–34	200	11.6	1.2
≥ 35	65	10.8	1.4

$p = 0.032$

Kết luận:

Hb trung bình khác biệt có ý nghĩa giữa các nhóm tuổi ($p = 0.032$), nhóm ≥ 35 có Hb thấp nhất.

11.4.4. Dữ liệu không chuẩn – Mann–Whitney U test và Kruskal–Wallis

Mann–Whitney: dùng cho 2 nhóm độc lập không chuẩn.

Kruskal–Wallis: dùng cho ≥ 3 nhóm không chuẩn.

Thao tác:

Analyze → Nonparametric Tests → Independent Samples → chọn Test Type.

Ví dụ: So sánh BMI (không chuẩn) giữa 3 nhóm tuổi.

Kruskal–Wallis test → $p = 0.021$ → kết luận có khác biệt.

11.5. Kiểm định cho biến định tính

11.5.1. Kiểm định tỷ lệ – Chi-square (χ^2 test)

Ví dụ: So sánh tỷ lệ thiếu máu giữa nam và nữ.

Thao tác:

Analyze → Descriptive Statistics → Crosstabs

Row: Gioitinh; Column: Thieumau

Click *Statistics* → *Chi-square*; *Cells* → *Row percentages* → OK.

Bảng 11.3. Kết quả mẫu Chi-square

Giới tính	Có thiếu máu n (%)	Không thiếu máu n (%)
Nam	30 (15.8)	160 (84.2)
Nữ	84 (52.5)	76 (47.5)

Chi-square = 45.6; p < 0.001

Kết luận:

Tỷ lệ thiếu máu cao hơn đáng kể ở nữ so với nam (p < 0.001).

11.5.2. Kiểm định Fisher's Exact

Dùng khi tần suất mong đợi trong bảng 2×2 nhỏ hơn 5 (mẫu nhỏ).

Thao tác:

Trong Crosstabs, chọn *Exact* → *Fisher's Exact Test*.

11.5.3. Kiểm định tỷ lệ một mẫu (One-sample proportion test)

Không có trực tiếp trong SPSS, nhưng có thể dùng cú pháp:

NPAR TESTS /CHISQUARE=Thieumau(0,1) /EXPECTED=0.3.

→ Kiểm định xem tỷ lệ thiếu máu khác 30% (theo WHO) hay không.

11.6. Phân tích tương quan

11.6.1. Tương quan Pearson (cho biến định lượng chuẩn)

Ví dụ: Mỗi liên hệ giữa tuổi và Hb.

Analyze → Correlate → Bivariate → chọn “Pearson” → OK.

Bảng 11.4. Kết quả mẫu

Biến	Hệ số r	p
Tuổi – Hb	-0.28	0.002

Kết luận:

Có tương quan nghịch nhẹ giữa tuổi và Hb (r = -0.28, p = 0.002).

11.6.2. Tương quan Spearman (cho dữ liệu không chuẩn hoặc thứ bậc)

Analyze → Correlate → Bivariate → chọn “Spearman”.

11.7. Phân tích hồi quy Logistic (nâng cao cơ bản)

Dùng khi **biến phụ thuộc là định tính nhị phân** (ví dụ: có/không thiếu máu).

Mô hình:

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 +$$

Thao tác SPSS:

Analyze → Regression → Binary Logistic

Dependent: Thieumau (0/1)

Covariates: Tuoi, BMI, Gioitinh, Nghe.

Kết quả mẫu (Bảng 11.5):

Biến	B	Wald	p	OR (Exp(B))	95% CI
Tuổi ≥35	0.68	5.23	0.022	1.97	1.10–3.54
Nữ	1.24	11.61	0.001	3.46	1.67–7.18
BMI	–0.15	4.05	0.044	0.86	0.74–0.99

Diễn giải:

Phụ nữ có nguy cơ thiếu máu cao gấp 3,46 lần nam (OR = 3.46; 95% CI: 1.67–7.18; p = 0.001).

Tuổi ≥35 tăng gần gấp đôi nguy cơ (OR = 1.97; p = 0.022).

BMI cao là yếu tố bảo vệ (p = 0.044).

11.8. Diễn giải p-value, CI, OR, RR trong y học

Ký hiệu	Ý nghĩa	Cách hiểu thông dụng
p-value	Xác suất sai lầm khi bác bỏ H_0	$p < 0.05$: có ý nghĩa thống kê
CI (Confidence Interval)	Khoảng tin cậy (thường 95%)	Khoảng mà giá trị thật có khả năng nằm trong đó
OR (Odds Ratio)	Tỷ số chênh – xác suất xảy ra bệnh giữa nhóm phơi nhiễm và không phơi nhiễm	OR = 2 → nhóm phơi nhiễm có khả năng bị bệnh gấp 2 lần
RR (Relative Risk)	Nguy cơ tương đối	RR = 1 → không khác biệt; RR > 1 → tăng nguy cơ; RR < 1 → yếu tố bảo vệ

Lưu ý:

- OR và RR chỉ có ý nghĩa khi CI không chứa 1 và $p < 0.05$.
- Nếu CI rộng → cỡ mẫu chưa đủ hoặc biến dao động cao.

11.9. Báo cáo kết quả kiểm định trong nghiên cứu y học

Ví dụ cách viết đúng:

Nồng độ Hb trung bình của nam (13.0 ± 1.1 g/dL) cao hơn đáng kể so với nữ (11.2 ± 1.3 g/dL, $t = 12.56$, $df = 348$, $p < 0.001$).

Có mối tương quan nghịch giữa tuổi và Hb ($r = -0.28$, $p = 0.002$).

Phụ nữ có nguy cơ thiếu máu cao gấp 3,46 lần so với nam (OR = 3.46, 95% CI: 1.67–7.18, $p = 0.001$).

11.10. Lưu ý khi chọn phép kiểm định

1. **Kiểm tra phân phối trước** khi chọn phép kiểm định.
2. **Phân biệt nhóm độc lập – liên quan.**
3. Không dùng nhiều phép kiểm định cho cùng một dữ liệu (tránh *p-hacking*).
4. Luôn trình bày đầy đủ: t , df , χ^2 , p , OR , CI .

11.11. Kết luận chương

- Kiểm định giả thuyết là nền tảng để suy luận thống kê trong y học.
- SPSS 27 cung cấp công cụ hoàn chỉnh để thực hiện mọi phép kiểm định phổ biến.
- Người làm nghiên cứu cần hiểu **bản chất thống kê**, không chỉ “bấm nút” — biết **khi nào** và **tại sao** chọn phép kiểm định nào.
- Việc diễn giải đúng kết quả (p , CI , OR , RR) giúp báo cáo khoa học có giá trị thực chứng và ứng dụng lâm sàng cao.

Tài liệu tham khảo

- [1] Bland M., *An Introduction to Medical Statistics*, 4th ed., Oxford University Press, 2015.
- [2] Field A., *Discovering Statistics Using IBM SPSS Statistics*, 5th ed., SAGE Publications, 2020.
- [3] IBM, *IBM SPSS Statistics 27 Documentation*, IBM Corporation, 2020.
- [4] Lưu Ngọc Hoạt, *Nghiên cứu khoa học trong y học*, NXB Y học, Hà Nội, 2014.
- [5] Kirkwood B. R. & Sterne J. A. C., *Essential Medical Statistics*, 2nd ed., Wiley-Blackwell, 2003.

PHẦN PHỤ LỤC

(Appendices)

PHỤ LỤC A. CẤU TRÚC DỮ LIỆU MẪU NGHIÊN CỨU “THIẾU MÁU Ở PHỤ NỮ MANG THAI TẠI TIỀN GIANG – 2024”

1. Mô tả bộ dữ liệu (SPSS file: Thieumau_TienGiang2024.sav)