

CHƯƠNG 6: XỬ LÝ VÀ PHÂN TÍCH DỮ LIỆU

Sau khi học xong chương 6, người học có thể:

- Thực hiện được các bước cơ bản trong xử lý dữ liệu định lượng
- Thực hiện được các bước cơ bản trong xử lý dữ liệu định tính
- Lựa chọn được các cách phân tích số liệu phù hợp với mục tiêu nghiên cứu của mình
- Trình bày được kết quả phân tích dữ liệu ở dạng bảng

TS. Nguyễn Thị Kim Liên

6.1. XỬ LÝ VÀ PHÂN TÍCH DỮ LIỆU ĐỊNH LƯỢNG

6.1. XỬ LÝ VÀ PHÂN TÍCH DỮ LIỆU ĐỊNH LƯỢNG

6.1.1. Chuẩn bị dữ liệu

6.1.2. Phân tích dữ liệu

6.2. XỬ LÝ VÀ PHÂN TÍCH DỮ LIỆU ĐỊNH TÍNH

6.2.1. Chuẩn bị dữ liệu

6.2.2. Một số phương pháp phân tích dữ liệu định tính

6.3. TRÌNH BÀY KẾT QUẢ PHÂN TÍCH DỮ LIỆU

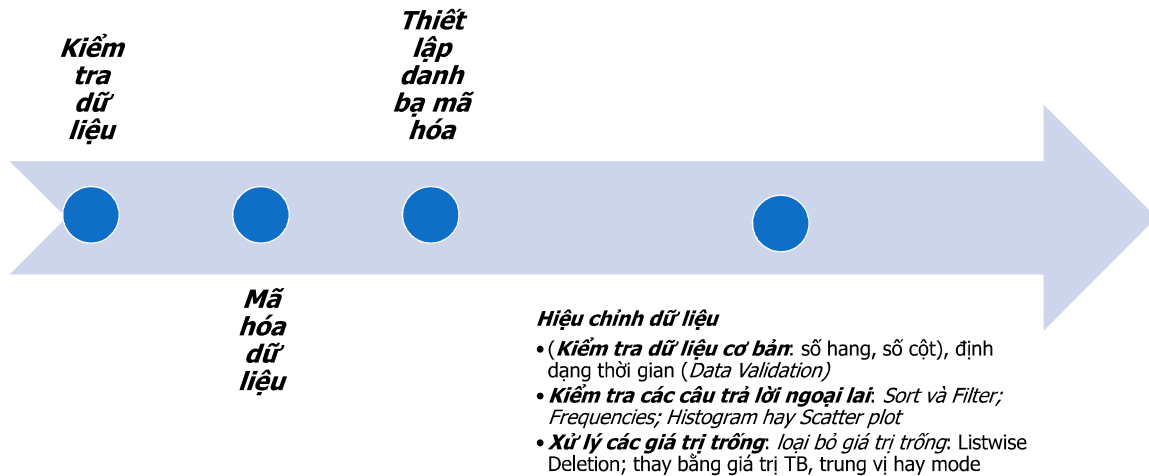
6.3.1. Từ ngữ

6.3.2. Bảng biểu

6.3.3. Biểu đồ, đồ thị

TS. Nguyễn Thị Kim Liên

6.1.1. Chuẩn bị dữ liệu



TS. Nguyễn Thị Kim Liên

1 Trong suốt năm học này tại trường Đại học Công nghiệp TP.HCM, bạn đã thực hiện các hoạt động dưới đây ở mức độ thường xuyên như thế nào? Đánh dấu câu trả lời của bạn vào ô thích hợp

	Không bao giờ	Thỉnh thoảng	Thường xuyên	Rất thường xuyên
	1	2	3	4
a Đặt câu hỏi hoặc tham gia thảo luận trên lớp	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b Thuyết trình trên lớp	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c Cố gắng hiểu và nắm vững các nội dung khó trong bài	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

- Mã hóa các nhóm 20 – 29 = 1; 30 – 39 = 2; 40 – 49 = 3; 50 – 59 = 4; trên 60 = 5.
- Lưu ý: Khi mã hóa cần tránh các trường hợp các giá trị bị chồng chéo với nhau.

TS. Nguyễn Thị Kim Liên

Ví dụ: Danh bạ mã hóa cho các mục hỏi

Số của cột	Số của mục hỏi	Tên của biến số (Name)	Nhãn biến số (Label)	Phương án trả lời	Giá trị (Values)
1	Số thứ tự của phiếu	ID	ID	Tương ứng với kích thước mẫu	Từ 1 đến số thứ tự lớn nhất của mẫu
2	1a	Tluan	Thảo luận trên lớp	Không thường xuyên Thỉnh thoảng Thường xuyên Rất thường xuyên	1 2 3 4
3	1b	Ttrinh	Thuyết trình trên lớp	Không thường xuyên Thỉnh thoảng Thường xuyên Rất thường xuyên	1 2 3 4

TS. Nguyễn Thị Kim Liên

6.1.2. Phân tích dữ liệu

Thống kê mô tả:

giá trị trung bình (mean),
 giá trị trung vị (median),
 giá trị xuất hiện thường xuyên nhất trong tập hợp (mode),
 tỷ lệ % (percentage),
 tần số xuất hiện (frequency),
 khoảng biến thiên (range),
 phương sai (variance),
 độ lệch chuẩn (standard deviation)

Thống kê suy luận:

khái quát hóa kết quả nghiên cứu từ mẫu cho toàn bộ dân số nghiên cứu và đưa ra các dự báo về dân số nghiên cứu.

Một số phép thống kê suy luận phổ biến bao gồm: tương quan, hồi quy, so sánh trung bình (ttest, ANOVA).

TS. Nguyễn Thị Kim Liên

Cách sử dụng thống kê suy luận kiểm định sự khác biệt

Mục tiêu nghiên cứu	Ví dụ	Thống kê
So sánh giá trị trung bình của tổng thể với một số cụ thể	Giả thuyết: Thu nhập TB/ tháng của các hộ trong địa bàn X là 5 triệu. So sánh thu nhập TB của các hộ gia đình với 5 triệu	One-sample T test
So sánh giá trị trung bình của 2 đám đông độc lập	Giả thuyết: Sinh viên nữ chăm học hơn sinh viên nam.	Independent samples T-test
So sánh từng cặp giá trị trung bình trước và sau khi có 1 tác động nào đó.	Giả thuyết: Điểm TB môn toán của lớp X được cải thiện sau khi GV sử dụng phương pháp giảng dạy mới	Paired samples T-test
So sánh giá trị trung bình của từ 3 đám đông trở lên (ảnh hưởng của BDL định tính lên BPT định lượng)	Giả thuyết: Có sự khác biệt trong mức độ nỗ lực học tập của sinh viên thuộc các nhóm ngành học khác nhau	One - way ANOVA

TS. Nguyễn Thị Kim Liên

Cách sử dụng thống kê suy luận kiểm định sự khác biệt

Mục tiêu nghiên cứu	Ví dụ	Thống kê
Kiểm tra mối tương quan tuyến tính giữa BDL và BPT, giữa nhiều biến với nhau	Giả thuyết: Trí thông minh có tương quan đồng biến với kết quả học tập	Correlation Bivariate
Kiểm định mối quan hệ giữa một biến độc lập và một biến phụ thuộc	Giả thuyết: Chế độ lương, thưởng có ảnh hưởng đến sự hài lòng của nhân viên	Simple Linear Regression
Kiểm định mối quan hệ giữa nhiều biến độc lập với một biến phụ thuộc	Giả thuyết: Chất lượng của SP, thái độ phục vụ, chính sách hậu mãi có ảnh hưởng đến sự hài lòng của KH.	Multiple Linear Regression
Kiểm định mối quan hệ giữa một BDL và một BPT	Giả thuyết: Chế độ lương, thưởng có ảnh hưởng đến sự hài lòng của NV	Simple Linear Regression

TS. Nguyễn Thị Kim Liên

6.2.1. Chuẩn bị dữ liệu trong phân tích định tính

Bước 1: Làm quen với dữ liệu

Bước 3: Thiết lập danh bạ mã hóa

Bước 2: Xem lại mục tiêu nghiên cứu

Bước 4: Xác định các chủ đề và các mối quan hệ giữa các khái niệm hay chủ đề

TS. Nguyễn Thị Kim Liên

6.2.2. Một số phương pháp phân tích dữ liệu định tính

- **Phân tích nội dung:** dùng để phân tích các thông tin ở dạng văn bản, âm thanh, hình ảnh, vật thể, thường được dùng để phân tích các câu trả lời phỏng vấn.
- **Phân tích tường thuật:** dùng để phân tích nội dung từ các nguồn phỏng vấn, quan sát từ thực địa, hoặc khảo sát. Nó tập trung vào tường thuật và diễn giải các câu chuyện và kinh nghiệm được các đối tượng khảo sát/ phỏng vấn chia sẻ để tìm ra câu trả lời cho câu hỏi nghiên cứu
- **Phân tích diễn ngôn:** phân tích sự tương tác giữa nhà nghiên cứu và các đối tượng được phỏng vấn; phân tích ngữ cảnh xã hội nơi diễn ra sự giao tiếp, xem xét môi trường sinh hoạt hàng ngày của người được phỏng vấn
- **Lý thuyết nền:** dữ liệu định tính để giải thích nguyên nhân xảy ra một hiện tượng nào đó xảy ra. Nghiên cứu một loạt các trường hợp tương tự nhau trong các ngữ cảnh khác nhau → rút ra các giải thích về mối quan hệ nhân quả

TS. Nguyễn Thị Kim Liên

6.3. TRÌNH BÀY KẾT QUẢ PHÂN TÍCH DỮ LIỆU

- Diễn giải kết quả phân tích sao cho người đọc có thể hiểu được ý nghĩa của các số liệu, mối quan hệ giữa chúng...
- Kết quả nghiên cứu cần phải được trình bày trung thực, chính xác, đầy đủ, rõ ràng và dễ hiểu.
- Kết quả nghiên cứu có thể được trình bày dưới nhiều dạng:
 - + từ ngữ,
 - + biểu thức toán học,
 - + mô hình,
 - + bảng biểu,
 - + biểu đồ,
 - + đồ thị ...

TS. Nguyễn Thị Kim Liên

6.3.1. Từ ngữ

- là phương tiện chính để diễn giải kết quả nghiên cứu
- cần chú ý đến các yếu tố sau: trình độ hiểu biết, mức độ quan tâm đến đề tài và các yêu cầu chặt chẽ về học thuật và khoa học của độc giả.
- Ngôn ngữ sử dụng trong báo cáo cần rõ ràng và trôi chảy, bố cục trình bày phải hấp dẫn và dễ nhìn

TS. Nguyễn Thị Kim Liên

6.3.2 Bảng biểu (cách diễn giải bảng)

Đưa ra các nhận định tổng quát về số liệu → phân tích chi tiết số liệu

Chỉ nên trình bày số liệu nổi bật (số cao nhất, thấp nhất ...) và chỉ ra sự tương quan giữa các số liệu.

CÁC ĐIỂM CHÍNH

- Nêu nội dung chính của bảng
- Nhận định tổng quát về số liệu trong bảng, chỉ ra các điểm chung của số liệu
- Nêu các số liệu nổi bật
- So sánh các số liệu
- Phân tích cụ thể một số dữ liệu tiêu biểu

TS. Nguyễn Thị Kim Liên

6.3.3. Biểu đồ, đồ thị

Thể hiện các mối tương quan của các đại lượng định lượng một cách trực quan

- Biểu đồ: biểu đồ dạng cột, hình quạt, hình tròn, biểu đồ phối hợp ...
- Đồ thị mô tả mối quan hệ phụ thuộc của một đại lượng đối với một hay nhiều đại lượng khác

Diễn giải biểu đồ, đồ thị

- nhận xét tổng quát về biểu đồ → phân tích chi tiết các số liệu
- tập trung chú ý vào những chi tiết nổi bật
- chú ý đến mối quan hệ giữa những chuỗi dữ liệu khác nhau được biểu diễn trong biểu đồ
- cần so sánh tương quan về độ lớn giữa các đại lượng trong biểu đồ

TS. Nguyễn Thị Kim Liên

CÂU HỎI VÀ BÀI TẬP

1. Mô tả quy trình chuẩn bị dữ liệu định lượng
2. Trình bày chi tiết các thao tác hiệu chỉnh dữ liệu định lượng
3. Mô tả quy trình chuẩn bị dữ liệu định tính
4. Trình bày một số phương pháp phân tích dữ liệu định tính
5. Mã hóa các câu hỏi, phương án trả lời cho phiếu khảo sát nhóm nghiên cứu của bạn đã thiết kế trong chương 4. Sau đó thiết lập danh bạ mã hóa cho các mục/câu hỏi và giá trị được gán cho các phương án trả lời
6. Mô tả quy trình chuẩn bị dữ liệu định lượng
7. Trình bày chi tiết các thao tác hiệu chỉnh dữ liệu định lượng
8. Mô tả quy trình chuẩn bị dữ liệu định tính
9. Trình bày một số phương pháp phân tích dữ liệu định tính
10. Mã hóa các câu hỏi, phương án trả lời cho phiếu khảo sát nhóm nghiên cứu của bạn đã thiết kế trong chương 4. Sau đó thiết lập danh bạ mã hóa cho các mục/câu hỏi và giá trị được gán cho các phương án trả lời
11. So sánh thống kê mô tả và thống kê suy luận

TS. Nguyễn Thị Kim Liên

CÂU HỎI VÀ BÀI TẬP

- 12.** Giả sử một nhà nghiên cứu quan tâm đến thực trạng sử dụng dịch vụ chăm sóc sức khỏe B của dân cư trên địa bàn X. Đọc các mục tiêu nghiên cứu dưới đây và cho biết nhà nghiên cứu có thể sử dụng các phép tính thống kê nào cho từng mục tiêu.
- a) Tìm hiểu mức độ sử dụng dịch vụ B của dân cư ở địa bàn X
 - b) So sánh mức độ sử dụng dịch vụ B của các nhóm dân cư trên địa bàn X. Các nhóm dân cư được phân chia theo (i) giới tính; (ii) nghề nghiệp; (iii) điều kiện kinh tế
 - c) Tìm hiểu mối quan hệ giữa các đặc điểm giới tính, độ tuổi, nghề nghiệp, điều kiện kinh tế đối với mức độ sử dụng dịch vụ B của dân cư trên địa bàn X.
 - d) Xác định mức độ ảnh hưởng của các yếu tố (i) thông tin về dịch vụ; (ii) chất lượng dịch vụ; (iii) giá cả của dịch vụ đối với mức độ sử dụng dịch vụ B của dân cư trên địa bàn X

TS. Nguyễn Thị Kim Liên

CÂU HỎI VÀ BÀI TẬP

13. Đọc và diễn giải các thông tin trong bảng dưới đây.

Bảng 1: Số lượng và tỷ lệ người dùng Internet tại 4 quốc gia Đông Nam Á năm 2017

Quốc gia	Dân số (thống kê năm 2017)	Người dùng Internet (tính đến 30/06/2017)	Tỷ lệ % dân số
Indonesia	263.500.146	132.700.000	50,4%
Việt Nam	95.414.640	64.000.000	67,1%
Philippines	103.796.832	57.607.242	55,5%
Thái Lan	68.297.547	57.000.000	83,5%

(Nguồn: Hiệu chỉnh từ *Các số liệu thống kê Internet Việt Nam*, Dammio.com, 2018)¹