

TRƯỜNG ĐẠI HỌC PHAN THIẾT
KHOA CƠ BẢN

THỐNG KÊ ỨNG DỤNG



Lưu hành nội bộ
- 2021 -

CHƯƠNG 1

NGUYÊN LÝ THỐNG KÊ VÀ CÁC KHÁI NIỆM CƠ BẢN

1. Thống kê là gì?

Thống kê là ngành khoa học về phương pháp thu thập, tổng hợp, trình bày số liệu và xử lý số liệu để đưa ra các phân tích, dự đoán và ra quyết định.

Thống kê được chia thành hai lĩnh vực:

- **Thống kê mô tả:** tập trung vào việc thu thập, tóm tắt và trình bày bộ dữ liệu. Các dữ liệu sau khi thu thập xong, phải được tóm tắt và trình bày dưới dạng một số mẫu có ý nghĩa và có khả năng mô tả các đặc trưng chung của hiện tượng nghiên cứu.
- **Thống kê suy diễn:** sử dụng dữ liệu của mẫu đã được tóm tắt và trình bày để phân tích, giải thích chúng và rút ra các kết luận về một tổng thể.

Sơ đồ của quá trình thống kê là:

Quan sát và đo lường → Thu thập dữ liệu → Mô tả, phân tích → Dự đoán, quyết định.

Thống kê được ứng dụng trong nhiều lĩnh vực khác nhau. Với phương pháp thống kê, chúng ta có thể hiểu được các hiện tượng, các sự kiện xã hội, ... thường là phức tạp và không ngừng biến thiên, đồng thời giúp chúng ta quan sát, thu thập dữ liệu một cách tương đối chính xác, phân tích một cách khoa học, để đưa đến những kết luận đúng đắn.

Ví dụ: Trong lĩnh vực kế toán sử dụng các phương pháp thống kê để lấy ra các mẫu cho mục đích kiểm toán. Tài chính sử dụng các phương pháp thống kê để lựa chọn giữa các khoản đầu tư khác nhau và để theo dõi các xu hướng của các thước đo tài chính qua thời gian. Các nhà quản lý sử dụng các phương pháp thống kê để cải tiến chất lượng của các sản phẩm được sản xuất hoặc các dịch vụ kinh doanh. Marketing sử dụng các phương pháp thống kê để ước lượng tỷ lệ khách hàng ưa thích một sản phẩm nào đó hơn một sản phẩm khác và tại sao họ lại thích hơn và để rút ra các kết luận về một chiến lược marketing để làm tăng doanh số bán hàng.

2. Một số khái niệm trong thống kê

2.1. Tổng thể thống kê, đơn vị tổng thể và mẫu

Tổng thể thống kê

Tổng thể thống kê là tập hợp các đơn vị (hay phân tử) thuộc hiện tượng nghiên cứu, cần được quan sát, thu thập và phân tích theo một hoặc một số đặc trưng nào đó.

Ví dụ: Muốn tính thu nhập trung bình của một hộ gia đình ở Bình Thuận thì tổng thể nghiên cứu sẽ là toàn bộ các hộ gia đình ở Bình Thuận;

Đơn vị tổng thể

Đơn vị tổng thể là các đơn vị (hay phần tử) tạo thành tổng thể thống kê.

Ví dụ: Muốn tính chiều cao trung bình của sinh viên nam Trường Đại học X thì đơn vị tổng thể là các sinh viên nam của Trường Đại học X.

Mẫu

Mẫu là một số đơn vị tổng thể được chọn ra từ tổng thể thống kê theo một phương pháp lấy mẫu nào đó. Các đặc trưng mẫu được sử dụng để suy rộng ra các đặc trưng của tổng thể thống kê.

Ví dụ: 100 sinh viên nam được lựa chọn để tính chiều cao; 50 khách hàng đi mua sắm ở siêu thị đã được hỏi để hoàn thành một cuộc điều tra về sự thỏa mãn của khách hàng.

2.2. Tiêu thức

Tiêu thức (biến) là đặc điểm của đơn vị tổng thể được chọn ra để nghiên cứu tùy theo mục đích nghiên cứu.

Ví dụ: Khi nghiên cứu nhân khẩu, mỗi nhân khẩu có những tiêu thức như: giới tính, độ tuổi, trình độ học vấn, nghề nghiệp, dân tộc, tôn giáo,...; Khi nghiên cứu các doanh nghiệp, mỗi doanh nghiệp có các tiêu thức như: loại hình doanh nghiệp, số lượng nhân viên, vốn cố định, vốn lưu động, có xuất khẩu hay không,...

Tiêu thức được chia thành hai loại:

- Tiêu thức định tính: là tiêu thức phản ánh tính chất hay loại hình của đơn vị tổng thể, không có biểu hiện trực tiếp bằng con số.

Ví dụ: giới tính, nghề nghiệp, tình trạng hôn nhân, dân tộc, tôn giáo, loại hình doanh nghiệp, ...

- Tiêu thức định lượng: là tiêu thức có biểu hiện trực tiếp bằng con số.

Ví dụ: tuổi, chiều cao, cân nặng, năng suất làm việc của công nhân, quy mô vốn của doanh nghiệp, ...

2.3. Lượng biến

Các trị số cụ thể khác nhau của tiêu thức định lượng gọi là lượng biến.

Ví dụ: tuổi là tiêu thức định lượng; lượng biến là: 18 tuổi, 20 tuổi, 30 tuổi, ...

Lượng biến được chia thành hai loại:

+ Lượng biến rời rạc: có thể nhận các giá trị cụ thể và luôn luôn có khoảng cách giữa các giá trị.

Ví dụ: Số nhân viên trong một doanh nghiệp, số môn thi lại của một sinh viên, số nhân khẩu trong một gia đình, điểm môn Thống kê ứng dụng (làm tròn đến 0,5), ...

+ Lượng biến liên tục: có thể nhận bất kỳ một giá trị nào trong một khoảng nhất định trên trục số.

Ví dụ: trọng lượng, chiều cao của sinh viên; năng suất của một loại cây trồng,...

2.4. Tham số tổng thể

Tham số tổng thể: là giá trị quan sát được của tổng thể và dùng để mô tả đặc trưng của hiện tượng nghiên cứu.

Ví dụ: trung bình tổng thể (μ), tỷ lệ tổng thể (p), phương sai tổng thể (σ)

2.5. Chỉ tiêu thống kê

Chỉ tiêu thống kê là con số có ý nghĩa và nội dung trong điều kiện thời gian và không gian xác định.

Chỉ tiêu thống kê có thể phân biệt thành hai loại:

- Chỉ tiêu khối lượng: là các chỉ tiêu biểu hiện qui mô của tổng thể.

Ví dụ: dân số của một thành phố, số doanh nghiệp trong một ngành, vốn định, vốn lưu động của một doanh nghiệp, số sinh viên của một trường đại học, ...

- Chỉ tiêu chất lượng: là các chỉ tiêu biểu hiện tính chất, trình độ phổ biến, quan hệ so sánh trong tổng thể.

Ví dụ: giá thành đơn vị sản phẩm là một chỉ tiêu chất lượng, nó biểu hiện quan hệ so sánh giữa tổng giá thành và số lượng sản phẩm sản xuất ra, đồng thời nó phản ánh tính chất phổ biến về mức chi phí cho một đơn vị sản phẩm đã được sản xuất ra. Tương tự, các chỉ tiêu năng suất lao động, tiền lương trung bình của một nhân viên, tỉ lệ nhân viên có bằng đại học trong doanh nghiệp, ... là các chỉ tiêu chất lượng.

Các chỉ tiêu chất lượng mang ý nghĩa phân tích, trị số của nó được xác định chủ yếu từ việc so sánh giữa các chỉ tiêu khối lượng.

2.6. Thang đo

Thang đo (scales of measurement): là công cụ để đo lường các đối tượng nghiên cứu.

Trong thống kê sử dụng 4 cấp bậc đo lường theo mức độ thông tin tăng dần đó là thang đo: định danh, thứ bậc, khoảng và tỉ lệ.

2.6.1. Thang đo định danh (Nominal scale)

Thang đo định danh sử dụng cho các biến định tính, mỗi số (giá trị) tượng trưng cho một tên gọi, loại, nhãn hiệu, địa danh; sự hơn kém của các con số không thể hiện sự hơn kém của các tính chất;

Thang đo định danh không có ý nghĩa về mặt lượng mặc dù nó có thể được ký hiệu bằng các con số.

Ví dụ: Giới tính, nam ký hiệu là số 1, nữ ký hiệu là số 2. Giữa các con số ở đây không có quan hệ hơn kém, chỉ dùng để đếm tần số xuất hiện của các biểu hiện.

Chúng ta hay gặp thang đo định danh trong các câu hỏi về thông tin cá nhân của từng người hay của thông tin về doanh nghiệp.

Ví dụ: Tình trạng hôn nhân của anh (chị):

1. Độc thân
2. Có gia đình
3. Ly dị
4. Trường hợp khác

2.6.2. Thang đo thứ bậc (Ordinal scale)

Thang đo thứ bậc sử dụng cho biến định tính; mỗi con số đại diện cho một tính chất, sự hơn kém của các con số thể hiện sự hơn kém của các tính chất; thể hiện sự xếp hạng, có mối quan hệ so sánh giữa các đối tượng.

Chúng ta hay gặp thang đo thứ bậc trong các câu hỏi dạng so sánh.

Ví dụ: Thu nhập hàng tháng của anh (chị):

1. Dưới 5 triệu đồng
2. Từ 5 đến 10 triệu đồng
3. Trên 10 triệu đồng

2.6.3. Thang đo khoảng (Interval scale)

Thang đo khoảng sử dụng cho biến định lượng; thang đo này xếp hạng các đối tượng đo nhưng khoảng cách bằng nhau trên đặc điểm của đối tượng; bao gồm cả thông tin từ thang đo thứ bậc cũng như cho phép so sánh sự khác biệt giữa các đối tượng; số 0 không mang giá trị thật.

Ví dụ: Nhiệt độ, $32^{\circ}\text{C} > 30^{\circ}\text{C}$ và $80^{\circ}\text{C} > 78^{\circ}\text{C}$. Sự chênh lệch giữa 32°C và 30°C cũng giống như sự chênh lệch giữa 80°C và 78°C .

Như vậy, thang đo khoảng cho phép đo lường chính xác sự khác nhau giữa hai giá trị bất kỳ.

Chúng ta hay gặp thang đo khoảng trong các câu hỏi phỏng vấn dạng đánh giá.

Ví dụ: Hãy đánh giá mức độ thân thiện của nhân viên bán hàng tại siêu thị:

Rất không thân thiện

Rất thân thiện

1

2

3

4

5

2.6.4. Thang đo tỷ lệ (Ratio scale)

Thang đo tỷ lệ sử dụng cho biến định lượng; là loại thang đo cao nhất chứa đựng tất cả nội dung của thang đo định danh, thứ tự và khoảng; giá trị 0 mang giá trị thật. Trong thang đo tỷ lệ, có thể nhận dạng hoặc phân loại đối tượng, xếp hạng đối tượng và so sánh sự khác biệt.

Ví dụ: Bạn A có 5 triệu đồng và bạn B có 10 triệu đồng. Như vậy số tiền của bạn B gấp đôi số tiền của bạn A. Nếu đổi sang USD thì số tiền của bạn B vẫn gấp đôi số tiền của bạn A. Nếu số tiền của bạn A bị mất thì bạn A còn 0 đồng. Số 0 ở đây là một trị số thật. Như vậy, tiền tệ là loại thang đo tỷ lệ.

3. Tổng hợp và trình bày dữ liệu thống kê

3.1. Phương pháp chọn mẫu:

3.1.1. Chọn mẫu ngẫu nhiên đơn giản

Mỗi đơn vị trong mẫu được lựa chọn với sự ngẫu nhiên như nhau. Phương pháp này cho phép các đơn vị tổng thể được chọn vào mẫu với cơ hội bằng nhau. Để thực hiện chọn mẫu ngẫu nhiên đơn giản, phải chuẩn bị danh sách các đơn vị của tổng thể cần thu thập dữ liệu. Với phương pháp chọn mẫu này, có thể thực hiện việc lấy đơn vị mẫu ra bằng nhiều cách như bốc thăm, quay số hay dùng số ngẫu nhiên.

Ví dụ: Bảng số liệu chiều cao (cm) được quan sát từ 21 sinh viên:

STT	Chiều cao	STT	Chiều cao	STT	Chiều cao
1	155	8	159	15	156
2	165	9	164	16	159
3	157	10	162	17	161
4	158	11	155	18	163
5	162	12	154	19	168
6	155	13	163	20	167
7	157	14	157	21	154

Chọn mẫu ngẫu nhiên gồm 4 phần tử: ở 4 vị trí 5, 8, 11, 20.

Vậy mẫu gồm các phần tử: { 162, 159, 167 }

3.1.2. Chọn mẫu ngẫu nhiên theo hệ thống

Quy trình thực hiện chọn mẫu ngẫu nhiên theo hệ thống:

- Chuẩn bị danh sách chọn mẫu, xếp thứ tự theo một quy ước nào đó, đánh số thứ tự cho các đơn vị trong danh sách. Tổng số đơn vị trong danh sách là N .

- Xác định cỡ mẫu n muốn chọn

- Xác định khoảng cách chọn mẫu $k = N/n$

- Trong k đơn vị đầu tiên ta chọn ngẫu nhiên ra 1 đơn vị. Đây là đơn vị mẫu đầu tiên, các đơn vị mẫu tiếp theo được lấy cách đơn vị này 1 khoảng là $k, 2k, 3k, \dots$

Như vậy, chọn mẫu ngẫu nhiên theo hệ thống là phương pháp chọn mẫu trong đó các đơn vị mẫu chọn ra cách nhau 1 khoảng là k đơn vị.

Ví dụ: Cho $N = 60$; $n = 10$; $k = N/n = 6$;

Chọn 1 số ngẫu nhiên trong khoảng từ vị trí 1 đến 6; ví dụ là vị trí 4; ta có các đơn vị mẫu ở các vị trí: 4; 10; 16; 22; 28; 34; 40;

3.1.3. Chọn mẫu ngẫu nhiên phân cụm

Tổng thể được chia thành nhiều cụm, mỗi cụm xem như một tổng thể con, đại diện cho toàn bộ tổng thể. Sau đó lấy ngẫu nhiên một mẫu các cụm và tiến hành nghiên cứu tất cả các đơn vị trong các cụm đã được lấy ra.

Ví dụ: Các cụm có thể được phân chia theo các đường ranh giới địa lý như: các tỉnh, thành phố, các quận, huyện, các phường, xã, các cụm dân cư, các tổ dân phố hay các khu vực bán hàng.

3.1.4. Chọn mẫu ngẫu nhiên phân tầng

Tổng thể nghiên cứu được chia thành các tầng lớp, dựa trên kích thước và độ phân tán của mỗi tầng đó, mục đích là để các giá trị của các đối tượng tổng thể thuộc cùng một tầng càng ít khác nhau càng tốt. Sau đó các đơn vị mẫu được chọn từ các tầng này theo các phương pháp chọn mẫu ngẫu nhiên đơn giản hoặc chọn mẫu ngẫu nhiên hệ thống.

Ví dụ: Một doanh nghiệp muốn lựa chọn một mẫu 32 nhân viên từ một tổng thể 800 nhân viên để ước lượng chi phí tài trợ khám răng của công ty. Trong số các nhân viên của doanh nghiệp bao gồm: 25% là cán bộ quản lý và 75% không phải là cán bộ quản lý. Chúng ta sẽ lựa chọn phương pháp chọn mẫu ngẫu nhiên phân tầng để cho mẫu có sự đại diện hợp lý của các nhân viên.

3.2. Bảng phân phối tần số

Bảng phân phối tần số là một bảng trong đó các dữ liệu được sắp xếp thành các tổ theo thứ tự tăng dần.

Phân tổ thống kê là phân chia tổng thể nghiên cứu thành các tổ có tính chất khác nhau dựa trên một hay một số tiêu thức nào đó.

Ví dụ: Khi thống kê về dân số, căn cứ vào tiêu thức “giới tính” để chia dân số thành hai tổ: nam và nữ; căn cứ vào tiêu thức “tuổi” để chia dân số thành nhiều tổ có độ tuổi khác nhau.

* Các bước khi tiến hành phân tổ:

• Xác định số tổ cần chia: $k = (2.n)^{1/3}$

• Xác định khoảng cách tổ: $h = \frac{x_{\max} - x_{\min}}{k}$

x_{\max} là giá trị lớn nhất của tập dữ liệu; x_{\min} là giá trị nhỏ nhất của tập dữ liệu

• Xác định giới hạn dưới và giới hạn trên của các tổ:

+ Khi xác định giới hạn dưới của tổ đầu tiên: phải đảm bảo giá trị của nó phải nhỏ hơn hoặc bằng giá trị x_{\min}

+ Giới hạn trên của tổ cuối cùng: phải đảm bảo giá trị của nó phải lớn hơn hoặc bằng giá trị x_{\max}

+ Với các tổ liên tiếp nhau, giá trị cận trên của tổ trước chính là giá trị cận dưới của tổ liền kề sau.

Ví dụ 1: (Trường hợp giới hạn dưới của tổ đầu tiên bằng với x_{\min})

Dữ liệu của một mẫu điều tra về tuổi của 30 sinh viên đang học năm 1 lớp vừa làm vừa học chuyên ngành Quản trị kinh doanh như sau:

28 23 30 24 19 21 39 22 22 31
 37 33 20 30 35 21 26 27 25 29
 27 21 25 28 26 29 29 22 32 37

Hãy lập bảng phân phối tần số cho mẫu dữ liệu trên?

Giải:

* Kích thước mẫu: $n = 30$

* Số tổ: $k = (2.n)^{1/3} = (2.30)^{1/3} = 3,9 \approx 4$

* Khoảng cách tổ: $h = \frac{x_{\max} - x_{\min}}{k} = \frac{39 - 19}{4} = 5$

Bảng phân phối tần số:

Tổ	Độ tuổi	Tần số (Số SV)
1	19 - 24	9
2	24 - 29	10
3	29 - 34	8
4	34 - 39	3
Tổng		30

Ví dụ 2: (Trường hợp giới hạn dưới của tổ đầu tiên nhỏ hơn x_{\min})

Một doanh nghiệp kinh doanh thương mại có 28 cửa hàng bán lẻ trong thành phố. Số liệu về mức tiêu thụ (triệu đồng) của 28 cửa hàng được ghi chép và báo cáo hàng tuần, trong tuần thứ nhất của tháng 5 năm 2021 người ta thấy tình hình tiêu thụ của 28 cửa hàng này như sau:

60 57,5 52,4 50,9 50,2 53,3 50,1 43,3 42,5 41,7
 41,1 45,8 47,2 46,9 56,9 47,5 38,8 50,3 37,6 38,9
 52,3 49,2 47,5 47 49,6 46,3 49,8 36,8

Hãy lập bảng phân phối tần số cho mẫu dữ liệu trên?

Giải:

* Kích thước mẫu: $n = 28$

* Số tổ: $k = (2.n)^{1/3} = (2.28)^{1/3} = 3,8 \approx 4$

* Khoảng cách tổ: $h = \frac{x_{\max} - x_{\min}}{k} = \frac{60 - 36,8}{4} = 5,8 \approx 6$

Ở bài này, $x_{\min} = 36,8$, nếu chọn cận dưới là 36,8 thì tổ đầu tiên là: 36,8 – 41,8 thì số lẻ sẽ khiến người đọc rối mắt và khó cảm nhận nên chúng ta có thể lùi cận dưới của tổ đầu tiên xuống giá trị 36.

Bảng phân phối tần số:

Tổ	Mức tiêu thụ (triệu đồng/tuần)	Tần số (Số cửa hàng)
1	36 - 42	6
2	42 - 48	9
3	48 - 54	10
4	54 - 60	3
Tổng		28

Khi thiết lập được bảng phân phối tần số, chúng ta có thể xây dựng thêm cột tần số tích lũy, tần suất, tần suất tích lũy:

- Tần số tích lũy: thể hiện số quan sát có giá trị bé hơn (hoặc bằng) giới hạn trên của tổ mà nó nằm cùng hàng. Tần số tích lũy được tính bằng cách cộng dồn các tần số từ trên xuống cho đến vị trí tương ứng với biểu hiện mà ta đang muốn tính tần số tích lũy.

- Tần suất: là tỷ lệ phần trăm giữa tần số của mỗi tổ với tổng số trường hợp quan sát của mẫu.

- Tần suất tích lũy: thể hiện tỷ lệ phần trăm số quan sát có giá trị bé hơn (hoặc bằng) giới hạn trên của tổ mà nó nằm cùng hàng. Tần suất tích lũy được tính bằng cách cộng dồn các tần suất từ trên xuống cho đến vị trí tương ứng với biểu hiện mà ta đang muốn tính tần suất tích lũy.

Ví dụ 3: Từ bảng phân phối tần số ở ví dụ 1, hãy lập bảng phân phối tần số tích lũy, tần suất, tần suất tích lũy:

Tổ	Độ tuổi	Tần số (Số SV)	Tần số tích lũy	Tần suất (%)	Tần suất tích lũy (%)
1	19 - 24	9	9	30	30
2	24 - 29	10	19	33,33	63,33
3	29 - 34	8	27	26,67	90
4	34 - 39	3	30	10	100
Tổng		30		100	

3.3. Biểu đồ

3.3.1. Biểu đồ cột (Bar chart)

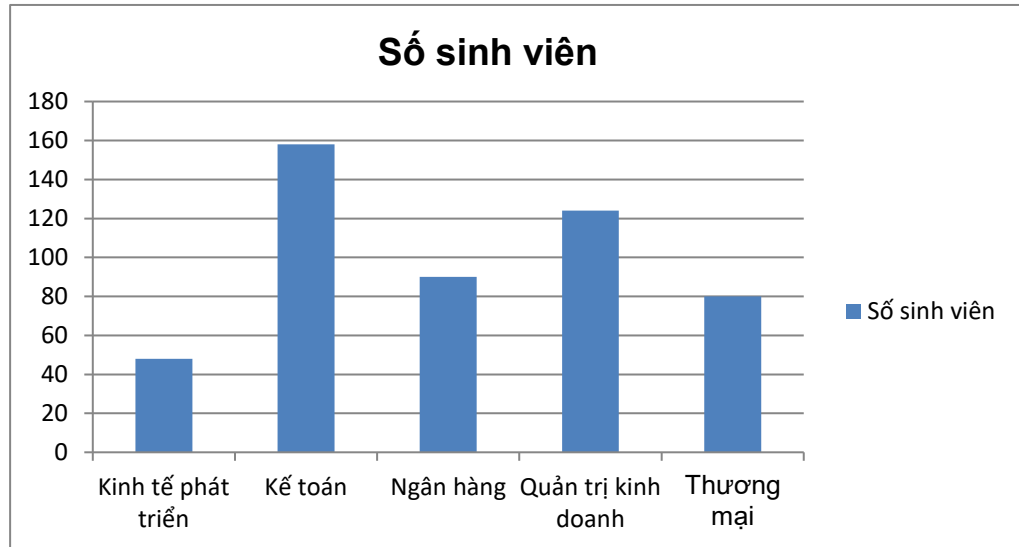
Trong biểu đồ cột, mỗi cột chỉ ra một nhóm, chiều cao của cột biểu diễn số lượng, tần số hay tỷ lệ phần trăm các giá trị rơi vào trong nhóm đó.

Biểu đồ cột chủ yếu dùng cho các dữ liệu định tính dạng phân loại, giữa các cột của biểu đồ có khoảng cách vì mỗi cột là một biểu hiện của tiêu thức định tính.

Ví dụ: Khảo sát 500 sinh viên tại khoa Kinh tế của một trường Đại học với 5 chuyên ngành được tổng hợp số liệu trong bảng phân phối tần số sau:

Ngành học	Số sinh viên (người)
Kinh tế phát triển	48
Kế toán	158
Ngân hàng	90
Quản trị kinh doanh	124
Thương mại	80
Tổng	500

Biểu đồ cột thể hiện số sinh viên theo các ngành học như sau:

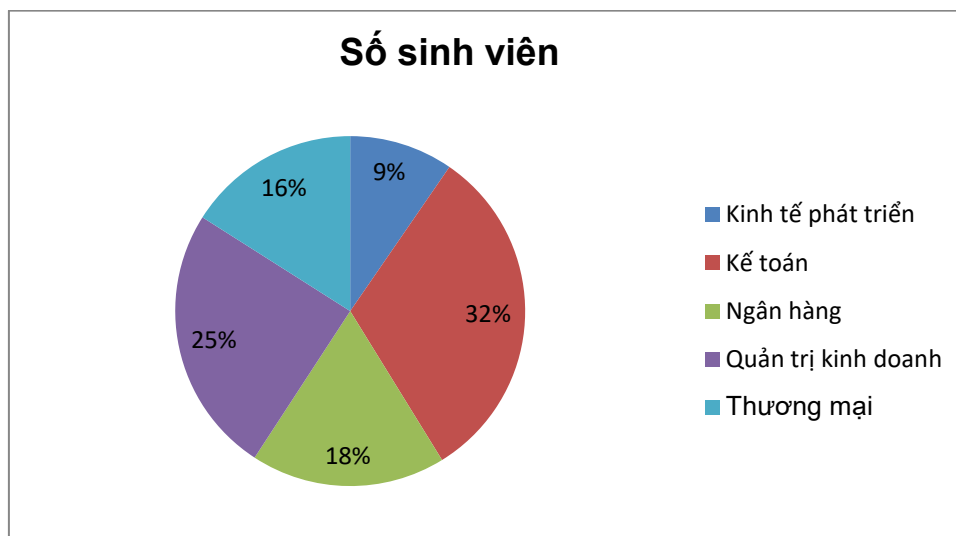


3.3.2. Biểu đồ hình tròn (Pie chart)

Biểu đồ hình tròn được chia thành các lát cắt hình rẻ quạt biểu diễn các nhóm phân loại, diện tích mỗi lát cắt tương đương với tỷ lệ phần trăm trong mỗi nhóm và mang một màu khác nhau.

• Tóm lại, chúng ta nên sử dụng dạng biểu đồ cột hay biểu đồ tròn là phụ thuộc vào ý định thể hiện thông tin của người vẽ biểu đồ. Nếu mục đích là để so sánh các phân loại thì nên sử dụng biểu đồ cột. Nếu muốn quan sát tỷ lệ của mỗi nhóm phân loại so với toàn bộ đối tượng thì nên sử dụng biểu đồ hình tròn.

Ví dụ : Biểu đồ hình tròn thể hiện số sinh viên theo các ngành học như sau:



Trong biểu đồ trên, 9% sinh viên chọn học ngành Kinh tế phát triển. Do đó, trong việc xây dựng biểu đồ hình tròn, góc 360° tạo nên hình tròn nhân với 0,09 được kết quả là một lát cắt của hình tròn chiếm $32,4^{\circ}$ của hình tròn.

3.3.3. Biểu đồ thân và lá (Stem and leaf display)

Biểu đồ thân và lá (hay còn gọi là biểu đồ nhánh – lá) là sự tóm tắt và trình bày dữ liệu cho thấy hình dáng phân phối dữ liệu như thế nào và xu hướng tập trung của dữ liệu ở đâu. Quy tắc lập biểu đồ là mỗi giá trị của dữ liệu định lượng sẽ được tách thành 2 phần: phần thân và phần lá. Các chữ số bên trái (ở hàng trăm hoặc hàng chục) đóng vai trò là phần thân, chữ số bên phải (ở hàng chục hoặc hàng đơn vị) đóng vai trò là phần lá; các dữ liệu vào trong nhóm phần thân sao cho các giá trị trong mỗi nhóm phần lá mở rộng sang bên phải trên mỗi dòng.

Ví dụ: Giả sử 15 sinh viên trong lớp ăn trưa tại một cửa hàng thức ăn nhanh. Dữ liệu về số tiền chi tiêu cho bữa trưa như sau (đơn vị tính nghìn đồng):

54 48 43 55 49 66 35 49 63 55
73 85 61 48 59

Hãy lập biểu đồ thân và lá cho dữ liệu trên.

Giải:

Sắp xếp các giá trị theo thứ tự tăng dần:

35 43 48 48 49 49 54 55 55 59
61 63 66 73 85

Sử dụng cột hàng chục làm thân và cột hàng đơn vị làm lá:

35 được biểu diễn thành →

43, 48, 48, 49, 49 được biểu diễn thành →

Thân	Lá
3	5
4	3 8 8 9 9
5	4 5 5 9
6	1 3 6
7	3
8	5

Từ biểu đồ thân và lá trên có thể kết luận:

- Số tiền chi tiêu nhỏ nhất là 35 nghìn đồng
- Số tiền chi tiêu lớn nhất là 85 nghìn đồng
- Phần lớn dữ liệu tập trung trong khoảng từ 43 đến 59 nghìn đồng

- Chỉ có một sinh viên chi tiêu dưới 40 nghìn đồng và chỉ có hai sinh viên chi tiêu trên 70 nghìn đồng.

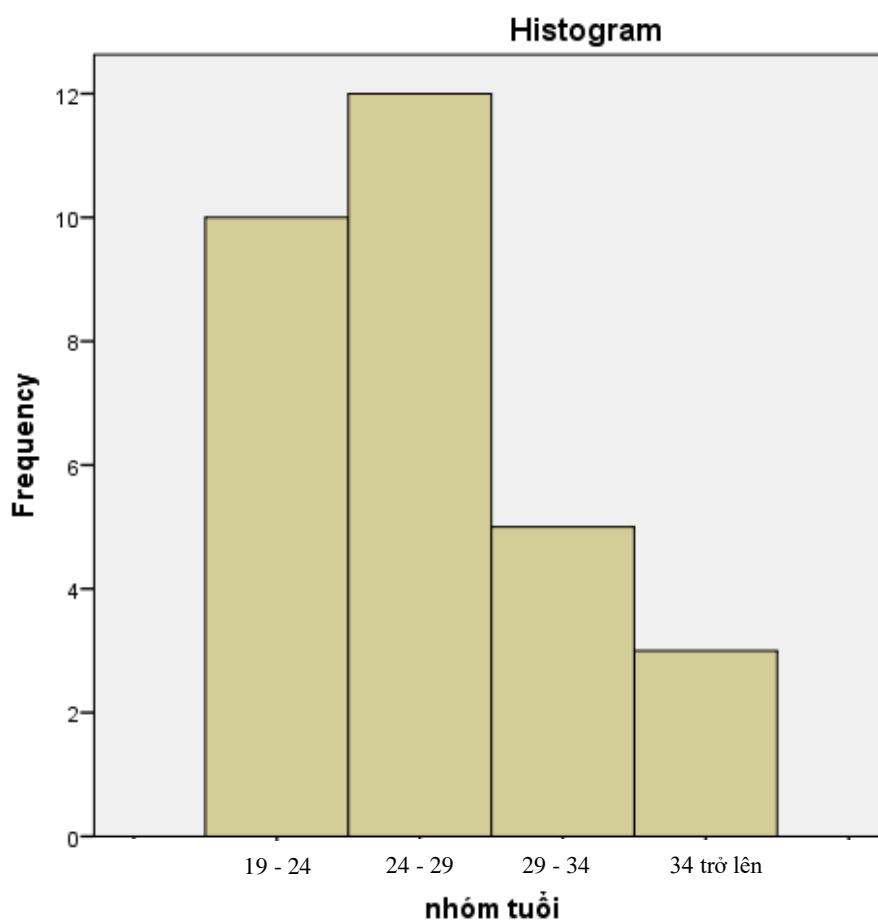
3.3.4. Biểu đồ phân phối tần số(Histogram)

Biểu đồ phân phối tần số là một biểu đồ cột cho các dữ liệu định lượng đã được phân tổ, trong đó các tần số được trình bày trên các cột riêng biệt. Trong biểu đồ phân phối tần số không có khoảng cách giữa các cột liên tiếp nhau như trong biểu đồ cột của các dữ liệu định tính. Chúng ta có thể mô tả các tiêu thức nghiên cứu trên trục nằm ngang; còn trục đứng thể hiện tần số của các giá trị trong mỗi khoảng cách tổ.

Ví dụ: Bảng phân phối tần số của 30 sinh viên đang học năm 1 lớp vừa làm vừa học chuyên ngành Quản trị kinh doanh, như sau:

Tổ	Nhóm tuổi	Tần số (Số SV)
1	19 - 24	9
2	24 - 29	10
3	29 - 34	8
4	34 - 39	3
Tổng		30

Vẽ biểu đồ phân phối tần số như sau:



• **Chú ý:** Nếu chúng ta vẽ biểu đồ phân phối với thông tin của trực đứng là tần suất (đơn vị %) thì biểu đồ phân phối tần số lúc này sẽ trở thành *biểu đồ phân phối tần suất* (có hình dáng hoàn toàn giống với biểu đồ phân phối tần số).

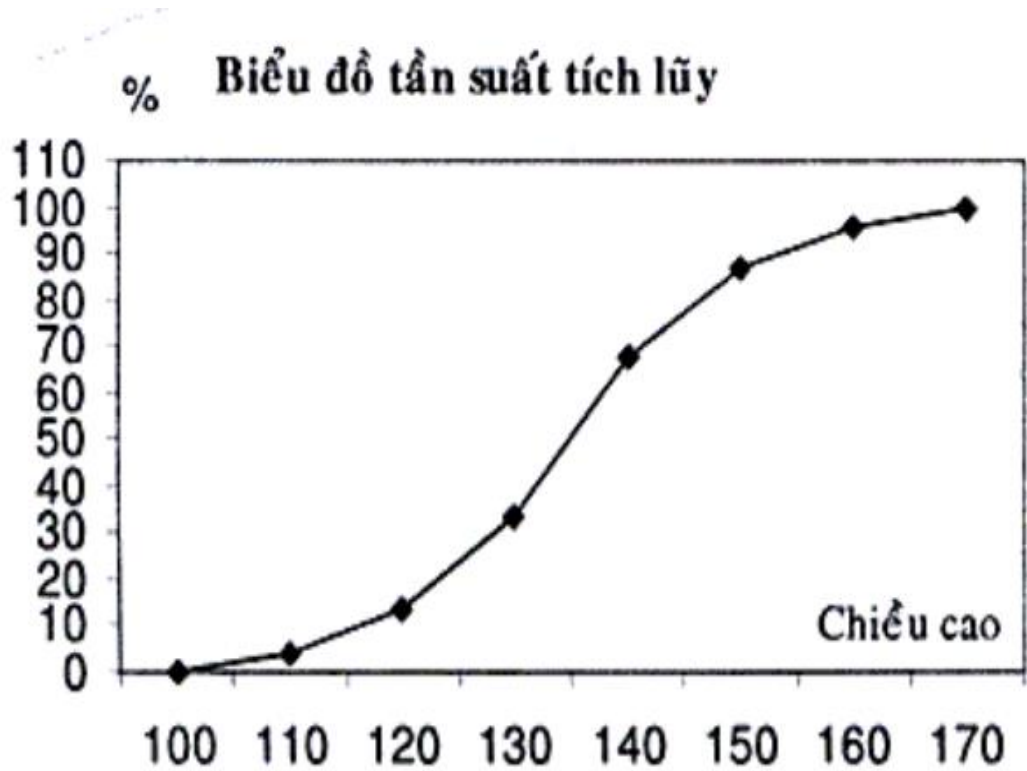
3.3.5. Biểu đồ tần suất tích lũy

Biểu đồ tần suất tích lũy mô tả tiêu thức nghiên cứu dọc theo trục ngang và tần suất tích lũy dọc theo trục đứng.

Ví dụ: Chiều cao của 500 trẻ em ở một trường học được phân tổ và lập thành bảng như sau:

Chiều cao (cm)	Số trẻ	Tần suất (%)	Tần suất tích lũy (%)
100 - 110	20	4	4
110 - 120	48	9,6	13,6
120 - 130	100	20	33,6
130 - 140	170	34	67,6
140 - 150	98	19,6	87,2
150 - 160	44	8,8	96
160 - 170	20	4	100
Tổng	500	100	

Vẽ biểu đồ tần số tích lũy:



Căn cứ vào biểu đồ tần suất tích lũy trên, ta xác định khoảng 50% trẻ em có chiều cao dưới 135 cm.

BÀI TẬP CHƯƠNG 1

Bài 1: Công ty điện lực X xây dựng bảng câu hỏi nhằm nghiên cứu khách hàng, trong đó có các câu hỏi về các yếu tố như sau:

1. Giới tính của người chủ gia đình.
2. Tuổi của người chủ gia đình.
3. Nghề nghiệp của người chủ gia đình.
4. Số người có trong gia đình.
5. Có sử dụng máy nước nóng năng lượng mặt trời (có/không).
6. Số lượng máy điều hòa nhiệt độ sử dụng trung bình mỗi ngày (qui đổi thành máy lạnh 1 CV).
7. Thời gian trung bình sử dụng máy điều hòa nhiệt độ mỗi ngày.
8. Số lượng ti vi.
9. Thời gian sử dụng tính trung bình cho 1 ti vi mỗi ngày.
10. Thu nhập/tháng (triệu đồng) của người chủ gia đình
 < 5 5 – 7 7 – 10 ≥ 10
11. Tổng thu nhập mỗi tháng của tất cả những người trong gia đình.
12. Số tiền điện trung bình trong hóa đơn mỗi tháng.

Hãy xác định loại thang đo và loại biến - định tính/định lượng – rời rạc hay liên tục - thể hiện qua các yếu tố trên.

***Bài 2:** Để kiểm tra chất lượng sản phẩm, người ta tiến hành đo chiều dày (cm) của 40 sản phẩm, kết quả thu được như sau:

20,0	20,3	20,2	20,1	20,0	21,0	20,9	20,5
21,0	21,0	20,5	22,0	21,2	20,4	20,8	20,9
21,2	21,2	22,0	20,8	20,4	20,3	20,6	20,7
21,0	22,0	20,1	20,4	20,5	21,2	21,8	20,9
21,2	22,0	21,3	21,0	21,4	20,3	20,9	21,0

Yêu cầu:

1. Trình bày số liệu theo phương pháp Thân và Lá (chọn nhánh bằng hai chữ số đầu).
2. Bằng cách phân nhóm có khoảng cách đều, hãy trình bày lại số liệu trên: lập bảng phân phối tần số, tần suất.
3. Tính tần số tích lũy, tần suất tích lũy.
4. Vẽ biểu đồ phân phối tần số, tần số tích lũy, tần suất tích lũy.

Bài 3: Trong chương trình quốc gia nghiên cứu thể chất của người Việt Nam, số liệu về chiều cao (cm) của một mẫu gồm 50 thanh niên tại một địa phương ghi nhận được như sau:

164	155	160	162	172	171	162	160	162	159
160	158	166	172	158	163	165	164	161	158
160	170	168	157	168	166	160	162	163	167
171	164	167	158	159	160	163	167	168	159
160	162	170	168	164	160	168	165	165	160

Yêu cầu:

1. Phân lớp dữ liệu trên thành 5 nhóm với khoảng cách đều nhau: lập bảng phân phối tần số, tần suất.
2. Tính tần số tích lũy, tần suất tích lũy.

3. Vẽ biểu đồ phân phối tần số, tần số tích lũy, tần suất tích lũy.

Bài 4: Điều tra doanh số hàng tháng (triệu đồng) của 25 doanh nghiệp kinh doanh một ngành hàng, ta thu được số liệu sau đây:

97	93	94	108	102
102	103	100	115	116
111	117	117	116	117
113	115	123	129	124
122	128	122	132	130

Yêu cầu:

1. Trình bày số liệu theo phương pháp Thân và Lá.
2. Bằng cách phân nhóm có khoảng cách đều, hãy trình bày lại số liệu trên: lập bảng phân phối tần số, tần suất.
3. Tính tần số tích lũy, tần suất tích lũy.
4. Vẽ biểu đồ phân phối tần số, tần số tích lũy, tần suất tích lũy.

Bài 5: Có kết quả điểm thi của một nhóm sinh viên được vẽ thành biểu đồ thân và lá như sau (kết quả từ phần mềm SPSS chạy ra):

Frequency	Stem & Leaf
1	1 . 5
2	2 . 33
1	3 . 2
3	4 . 555
6	5 . 446689
3	6 . 067
2	7 . 58
1	8 . 0
1	9 . 0

Stem width: 1,00 Each leaf: 1 cases

Yêu cầu:

1. Cho biết nhóm này có bao nhiêu sinh viên?
2. Người có kết quả kém nhất trong nhóm có điểm là bao nhiêu? Người có kết quả tốt nhất đạt bao nhiêu điểm?
3. Từ biểu đồ này hãy lập bảng tần số trên cơ sở phân tổ đều (chọn số tổ là 4)
4. Vẽ biểu đồ phân phối tần số mô tả lại điểm của nhóm sinh viên này.