BÀI GIẢNG: THU THẬP VÀ PHÂN LOẠI DỮ LIỆU
CHUYÊN ĐỀ: DỮ LIỆU VÀ BIỂU ĐỒ
MÔN: TOÁN 8 (KẾT NỐI TRI THỨC)
GIÁO VIÊN: ĐỖ VĂN BẢO

MỤC TIÊU

- ✓ *Phân biệt được các loại dữ liệu.*
- ✓ *Nắm được cách phân loại dữ liệu.*
- ✓ *Thực hiện và lý giải việc thu thập dữ liệu.*

Vai trò của số liệu trong thực tế:

- Cung cấp thông tin: Số liệu là nguồn thông tin cơ bản cho các quyết định trong kinh tế. Nhờ có số liệu, các quyết định về chính sách kinh tế có thể được đưa ra dựa trên các thông tin đầy đủ và chính xác hơn.

- Phân tích và đánh giá: Số liệu cung cấp dữ liệu để phân tích và đánh giá hiện trạng kinh tế. Những số liệu này được xử lý và phân tích để tìm ra những xu hướng và mối quan hệ giữa các biến số kinh tế.

- Dự báo và kế hoạch hóa: Só liệu cũng được sử dụng để dự báo và lập kế hoạch cho tương lai. Những dự báo này có thể được sử dụng để đưa ra các quyết định chiến lược trong kinh doanh hoặc để định hướng cho chính sách kinh tế của một quốc gia.

- Theo dõi hiệu quả: Số liệu cũng được sử dụng để đo lường hiệu quả của các chương trình và chính sách kinh tế. Việc theo dõi các chỉ số kinh tế có thể giúp cho các quyết định kinh tế được điều chỉnh để đạt được mục tiêu đề ra.

- Đào tạo và nghiên cứu: Số liệu cũng được sử dụng để đào tạo và nghiên cứu trong lĩnh vực kinh tế. Việc sử dụng số liệu cho phân tích và nghiên cứu giúp tăng hiểu biết và kiến thức về kinh tế.

Tóm lại, số liệu đóng vai trò vô cùng quan trọng trong việc quản lý và phát triển kinh tế, giúp cung cấp thông tin, phân tích và đánh giá, dự báo và kế hoạch hóa, theo dõi hiệu quả, đào tạo và nghiên cứu. Ngoài ra, số liệu cũng đóng vai trò quan trọng trong việc đánh giá và dự báo tình hình kinh tế.

I. THU THẬP DỮ LIỆU

HĐ1. Nêu các phương pháp thu thập dữ liệu đã được học. Mỗi phương pháp cho một ví dụ.

Cách giải:

Thu thập dữ liệu có thể là trực tiếp hoặc gián tiếp.

- Thu thập dữ liệu trực tiếp là việc thu thập dữ liệu thông qua quan sát, làm thí nghiệm, lập bảng hỏi, phỏng vấn,...

- Thu thập dữ liệu gián tiếp là việc thu thập dữ liệu từ những nguồn có sẵn như sách, báo, mạng Internet,...

Để có thể đưa ra các kết luận hợp lí, dữ liệu thu thập được phải đảm bảo tính đại diện cho toàn bộ đối tượng đang được quan tâm.

Ví dụ 1. Để thu thập dữ liệu sau, ta nên làm thế nào? Đó là thu thập dữ liệu trực tiếp hay gián tiếp?

a) Dữ liệu về xếp hạng FIFA của bóng đá nam Việt Nam trong thời gian gần đây.

b) Số liệu về sự phát triển chiều cao của một giống cây mới theo thời gian.

Cách giải:

a) Cách tốt nhất ta vào website của Liên đoàn Bóng đá Thế giới (FIFA) tại địa chỉ fifa.com/fifa-world-ranking/vie để thu thập. Đây là phương pháp thu thập dữ liệu gián tiếp.

b) Để có số liệu về sự phát triển chiều cao của một giống cây mới theo thời gian, ta trồng cây và định kì đo chiều cao, ghi lại kết quả. Đây là phương pháp thu thập dữ liệu trực tiếp.

Luyện tập 1. Hãy cho biết phương pháp thu thập dữ liệu trong mỗi trường hợp sau là trực tiếp hay gián tiếp.

a) Nam vào website của Tổng cục Thống kê và ghi lại số quận/huyện của các tỉnh/thành phố thuộc đồng bằng Bắc Bộ.

b) Thầy giáo dạy Giáo dục thể chất đã đo và ghi lại thời gian chạy cự li 1000 m của các bạn học sinh khối 8

Cách giải:

a) Đây là phương pháp thu thập dữ liệu gián tiếp

b) Đây là phương pháp thu thập dữ liệu trực tiếp

Chú ý:

- Một trong những phương pháp thu thập dữ liệu trực tiếp hiệu quả hiện nay là khảo sát online.
- Người ta sử dụng các ứng dụng phần mềm để tạo bộ câu hỏi và thu thập dữ liệu trên môi trường trực tuyến. Người được khảo sát cũng dễ dàng cung cấp thông tin bằng cách truy cập một đường link hay quét một mã QR code được gửi từ người khảo sát.
- Các ứng dụng phần mềm này có thể tổng hợp kết quả ngay lập tức sau khi người được khảo sát cung cấp câu trả lời, rất dễ dàng và hiệu quả. Một số ứng dụng thường được sử dụng như Google Forms (Biểu mẫu), Microsoft Office Form, Survey Monkey, Slido, Mentimeter,...

II. PHÂN LOẠI DỮ LIỆU

HĐ2. Cho hai dãy dữ liệu sau về 5 học sinh

(A) Chiều cao (đơn vị: cm): 128,1; 132,9; 125,7; 131,3; 133,6

(B) Số môn thể thao học sinh biết chơi: 2; 1; 5; 2; 3

a) Hai dãy dữ liệu này có phải là số liệu không?

b) Đo chiều cao (kí hiệu là h) một học sinh khác và hỏi về số môn thể thao (kí hiệu là n) mà em đó biết chơi.

+ h có thể nhận gía trị bất kì lớn hơn 120 cm và nhỏ hơn 150 cm được không?

+ n có thể nhận được giá trị lớn hơn 3 và nhỏ hơn 4 được không?

Cách giải:

a) Hai dãy dự liệu dạng số được gọi là số liệu,

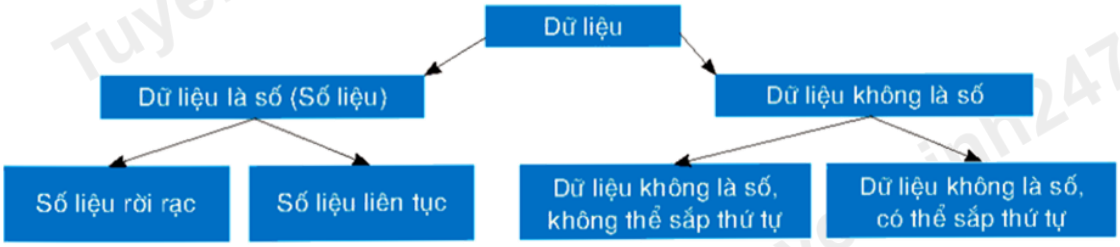
b) h có thể nhận gía trị bất kì lớn hơn 120 cm và nhỏ hơn 150 cm

n không thể nhận được giá trị lớn hơn 3 và nhỏ hơn 4.

- Số liệu có thể nhận giá trị tùy ý trong một khoảng nào đó được gọi là số liệu liên tục.
- Số liệu không phải là số liệu liên tục được gọi là số liệu rời rạc.

Chú ý:

- Dạng hay gặp của số liệu liên tục là số liệu thu được từ các phép đo như chiều cao, cân nặng, nhiệt độ,…
- Dạng hay gặp của số liệu rời rạc là số liệu đếm số phần tử của một tập hợp nào đó, chẳng hạn số học sinh trong lớp học, số sản phẩm một công nhân làm được trong ngày…



Ví dụ 2. Với mỗi câu hỏi sau, Quỳnh đã hỏi ba bạn và ghi lại câu trả lời.

a) Nhà bạn có bao nhiêu chiếc ti vi? Kết quả: 5; 52; 2

b) Bạn mất bao nhiêu thời gian (đơn vị: giờ) để hoàn thành bài tập về nhà? Kết quả: 1,5; 2,3; 1,9

Mỗi dãy dữ liệu trên thuộc loại nào? Chỉ ra giá trị không hợp kí nếu có

Cách giải:

a) Dữ liệu thu được là số liệu rời rạc. Giá trị 52 không hợp lí

b) Dữ liệu thu được là số liệu liên tục.

Chú ý:

- Các dữ liệu nào thể hiện số lượng, thời gian, khoảng cách, kích thước, cân nặng, thể tích,... hay kết quả của một phép tính thì thuộc loại dữ liệu là số (số liệu).
- Số liệu có thể nhận giá trị tuy ý trong một khoảng nào đó được gọi là số liệu liên tục. Số liệu không phải là số liệu liên tục được gọi là só liệu rời rạc.
- Các dứ liệu thể hiện ý kiến, quan điểm bằng từ ngữ, văn bản hay danh từ chung, danh từ riêng, âm thanh, hình ảnh... thuộc loại dữ liệu không là số. Ví dụ, dữ liệu tên các tuyến xe buýt (Xe buýt 01, 15, ...) là các danh từ nên là dữ liệu không là số.
- Dữ liệu không là số có thể phân thành hai loại, loại không thể sắp thứ tự (chẳng hạn dữ liệu về sở thích môn học: Toán, Văn, Lịch sử,...) và loại có thể sắp thứ tự (chẳng hạn dữ liệu đánh giá chất lượng một dịch vụ nào đó có các mức Kém, Trung bình, Tốt, Rất tốt).

Luyện tập 2. Với mỗi câu hỏi sau, An đã hỏi 5 bạn và ghi lại câu trả lời.

a) Bạn nặng bao nhiêu kilogam? Kết quả: 48; 51; 46; 145; 48

b) Tên bạn có bao nhiêu chữ cái? Kết quả: 4; 5; 6; 3; 5.

Mỗi dãy dữ liệu trên thuộc loại nào? Chỉ ra giá trị không hợp lí nếu có.

Cách giải:

a) Dữ liệu thu được là số liệu liên tục. Giá trị 145 không hợp lí.

b) Dữ liệu thu được là số liệu rời rạc.

Vận dụng. Em muốn ước lượng thời gian tự học ở nhà (đơn vị: giờ) của các bạn trong lớp. Hãy đưa ra cách thu thập dữ liệu và xác định xem dữ liệu thu được thuộc loại nào.

Cách giải:

Cách thu thập trực tiếp: hỏi các bạn trong lớp thời gian tự học ở nhà.

Thời gian tự học: 1 giờ; 1,5 giờ; 3,2 giờ.... là dữ liệu liên tục

Bài 5.1. Dữ liệu thu được trong mỗi câu hỏi sau thuộc loại nào?

a) Bạn cao bao nhiêu?

b) Mạng điện thoại bạn đang dung là gì?

c) Gia đình bạn có bao nhiêu người dưới 18 tuổi?

Cách giải:

a) Dữ liệu thu được là số liệu liên tục.

b) Dữ liệu thu được không phải là số liệu

c) Dữ liệu thu được là số liệu rời rạc.

Bài 5.2. Ghép cặp cho phù hợp:

a. Số liệu rời rạc.	A. Kết quả đánh giá của 5 bạn về để kiểm tra học kì I môn Toán: Khó, Rất khó, Trung bình, Dễ, Khó.
b. Số liệu liên tục.	B. Nhiệt độ (°C) tại Nha Trang trong 5 ngày đầu tháng 6 là: 23,2; 25,7; 31,4; 27,3; 28,6.
c. Dữ liệu không là số, không thể sắp thứ tự.	C. Số hoạt động hè mà các bạn trong tổ tham gia: 2; 1; 3; 0; 4.
d. Dữ liệu không là số, có thể sắp thứ tự.	D. Tên các môn thể thao mà các bạn yêu thích nhất: Bóng đá, Cầu lông, Cờ vua, Võ thuật, Bóng bàn.

Cách giải:

a – C; b – B; c – D; d – A

Bài 5.3. Nên sử dụng phương pháp thu thập dữ liệu nào để thu được mỗi dữ liệu sau:

a) Tên của 10 quốc gia có diện tích lớn nhất.

b) Ý kiến của các bạn về địa điểm đi tham quan trong tuần tới.

c) Chiều cao của các cây giống sau 6 tháng trồng.

Cách giải:

a) Phương pháp thu thập dữ liệu gián tiếp

b) Phương pháp thu thập dữ liệu trực tiếp

c) Phương pháp thu thập dữ liệu trực tiếp

BT1. Xác định phương pháp thu thập dữ liệu là trực tiếp hay gián tiếp trong các trường hợp sau và giải thích tại sao.

- Chỉ số khối cơ thể BMI (Body Mass Index) có thể cho thấy bạn đang có mức cân nặng bình thường so với chiều cao hay béo phì, thừa cân, thiếu cân hay suy dinh dưỡng. Chỉ số BMI được tính theo công thức: $BMI = \frac{m}{h^2}$, trong đó m là khối lượng cơ thể (theo ki-lô-gam), h là chiều cao cơ thể (theo mét).

- Đối với người lớn, chỉ số BMI từ 18,5 - 24,9 nằm trong mức cân nặng bình thường hoặc khỏe mạnh. Chỉ số BMI từ 25,0 trở lên là thừa cân, trong khi chỉ số BMI dưới 18,5 là thiếu cân.

a) Hải Anh thực hiện cân và đo 5 người lớn trong đại gia đình để tính chỉ số BMI của từng người.

b) An Huy dựa vào chỉ số chiều cao, cân nặng trong hồ sơ khám sức khoẻ của mẹ để tính chỉ số BMI của mẹ.

c) Diệu Hân tìm kiếm thông tin các nước có chỉ số BMI trung bình của người trưởng thành cao nhất thế giới trên Internet.

d) Hà My làm một poster và chia sẻ trên trang Facebook cá nhân, sau đó thống kê số lượt tương tác cảm xúc (thích, thả tim,...), số lượt chia sẻ, số lượt bình chọn để báo cáo với giáo viên.

Cách giải:

a) Phương pháp thu thập dữ liệu trực tiếp.

b) Phương thu thập gián tiếp.

c) Phương thu thập gián tiếp.

d) Phương pháp thu thập dữ liệu trực tiếp.

BT2. Để thu thập dữ liệu sau ta nên sử dụng phương pháp thu thập dữ liệu nào?

a) Dữ liệu về 10 nước đông dân nhất thế giới hiện nay.

b) Loại phương tiện đến trường của các bạn trong lớp.

c) Số lần giơ tay phát biểu của các bạn học sinh trong các giờ học.

d) Số lần vô địch SEA Games của đội tuển bóng đá Việt Nam tính đến năm 2023.

Cách giải:

a) Phương pháp thu thập dữ liệu gián tiếp.

b) Phương thu thập trực tiếp.

c) Phương thu thập trực tiếp.

d) Phương pháp thu thập dữ liệu gián tiếp.

BT3. Dữ liệu thu được trong mỗi câu hỏi sau thuộc loại nào?

a) Số lượng bài tập về nhà của mỗi môn học lớp 8 là bao nhiêu?

b) Kích thước chiều dài, chiều rộng của bàn học là bao nhiêu?

c) Ba môn học yêu thích nhất của học sinh lớp 8A là những môn nào?

d) Từ trường Đại học Sư Phạm Hà Nội đến trường Đại học Bách Khoa có những tuyến xe buýt số bao nhiêu?

Cách giải:

a) Số liệu rời rạc.

b) Số liệu liên tục.

c) Không phải dữ liệu số.

d) Không phải dữ liệu số.

BT4. Hãy phân loại phương pháp thu thập dữ liệu và loại dữ liệu thu được trong các trường hợp sau và xác định dữ liệu thuộc loại nào? (Trực tiếp, gián tiếp, số liệu rời rạc, số liệu liên tục, dữ liệu không phải là số).

a) Số lượng giải thưởng Nobel của 10 quốc gia có nhiều nhà khoa học đạt giải nhất.

b) Gia Hân thu thập tên các thần tượng âm nhạc của các bạn trong lớp mình.

c) Phương Anh thu thập thông tin điểm chuẩn vào lớp 10 của một số trường trên địa bàn tỉnh để chuẩn bị kế hoạch ôn tập thi vào 10 trong năm học tới.

Cách giải:

a) Thu thập gián tiếp và số liệu rời rạc.

b) Thu thập trực tiếp và không phải số liệu số.

c) Thu thập gián tiếp và số liệu liên tục.