

*Thống Kê & Xác suất*

THU THẬP VÀ PHÂN LOẠI DỮ LIỆU

**A. KIẾN THỨC TRỌNG TÂM.**

**1. Thu thập dữ liệu.**

* Có nhiều cách để thu thập dữ liệu, chẳng hạn : quan sát, lập phiếu điều tra (phiếu hỏi), tiến hành phỏng vấn,… hoặc thu thập từ những nguồn có sẵn như sách, báo, trang web, các phương tiện thông tin đại chúng,…

**2. Phân loại và tổ chức dữ liệu**

* *Có những dữ liệu thống kê là số (số liệu), những dữ liệu này còn gọi là dữ liệu định lượng.*
* *Có những dữ liệu thống kê không phải là số, những dữ liệu này còn gọi là dữ liệu định tính.*
* *Dữ liệu định lượng được biểu diễn bằng số thực.*
* *Dữ liệu định tính được biểu diễn bằng từ, chữ cái, kí hiệu,….*
* *Việc phân loại dữ liệu thống kê phụ thuộc vào những tiêu chí đưa ra, hay nói cách khác, phụ thuộc vào mục đích phân loại.*

**3. Tính hợp lí của dữ liệu.**

* Để đánh giá tính hợp lí của dữ liệu, ta cần đưa ra các tiêu chí đánh giá, chẳng hạn như dữ liệu phải:
* Đúng định dạng.
* Nằm trong phạm vi dự kiến.
* Phải có tính đại diện đối với vấn đề cần thống kê.
* Để đánh giá tính hợp lí của dữ liệu, ta có thể dựa vào mối liên hệ toán học đơn giản giữa các số liệu.
* Để đảm bảo tính hợp lí, dữ liệu cần phải đáp ứng đúng các tiêu chí toán học đơn giản như:
* Tổng tất cả các số liệu thành phần phải bằng số liệu của toàn thể.
* Số lượng của bộ phận phải nhỏ hơn số lượng của toàn thể.

**B. CÁC DẠNG BÀI TẬP VÀ PHƯƠNG PHÁP GIẢI**

|  |
| --- |
| Dạng 1: Thu thập dữ liệu. |

**Ví dụ 1.** Một cửa hàng bán điện thoại muốn tìm hiểu về loại điện thoại những khách hàng yêu thích khi vào cửa hàng. Theo em, cửa hàng có thể thu thập những thông tin đó bằng cách nào ?

*Lời giải:* Người bán hàng quan sát, hỏi xem vị khách mua yêu thích loại điện thoại và lập phiếu hỏi để điều tra.

**Ví dụ** :

|  |  |
| --- | --- |
| Tên vị khách | Loại điện thoại |
| Nguyễn Văn A | Iphone |
| ………….. | ………… |
| ………….. | …………. |

|  |
| --- |
| Dạng 2: Phân loại và tổ chức dữ liệu. |

**Ví dụ 2.**

Thu thập thông tin về sự phân bố dân cư Việt Nam năm 2020 được dữ liệu thống kê sau :

* Các vùng : Đồng bằng sông Hồng, Trung du và miền núi phía Bắc, Bắc Trung Bộ và Duyên hải miền Trung, Tây Nguyên, Đông Nam Bộ, Đồng bằng sông Cửu Long.
* Dân số của các vùng đó lần lượt là *1078; 134; 212; 109; 779; 424*.

(*đơn vị : người / km*).

Hãy phân loại các dữ liệu đó dựa trên tiêu chí định tính và định lượng.

*Lời giải:*

*Tiêu chí định lượng :* 1078; 134; 212; 109; 779; 424. (*đơn vị : người / km*).

*Tiêu chí định tính :* Đồng bằng sông Hồng, Trung du và miền núi phía Bắc, Bắc Trung Bộ và Duyên hải miền Trung, Tây Nguyên, Đông Nam Bộ, Đồng bằng sông Cửu Long.

**Ví dụ 3.**

Trong môn Sinh học 7 của *NXB giáo dục Việt Nam* học sự đa dạng thế giới động vật như: *Trùng roi, Trùng biến hình, Trùng giày, Thủy tức, Sán lá gan, Giun đũa, Giun đất, Trai sông, Tôm sông, Nhện,Châu chấu, Cá chép, Ếch đồng, Thằn lằn bóng đuôi dài, Chim bồ câu, Thỏ, Chuột túi.*

Hãy phân nhóm những động vật bên trên theo tiêu chí sau để biết sự đa dạng của thế giới động vật :

|  |  |
| --- | --- |
| Ngành động vật nguyên sinh | Tên các động vật (liệt kê cụ thể). |
| Ngành ruột khoang | Tên các động vật (liệt kê cụ thể). |
| Ngành giun | Tên các động vật (liệt kê cụ thể). |
| Ngành thân mềm | Tên các động vật (liệt kê cụ thể). |
| Ngành chân khớp | Tên các động vật (liệt kê cụ thể). |
| Ngành động vật có xương sống | Tên các động vật (liệt kê cụ thể). |

*Lời giải:*

|  |  |
| --- | --- |
| Ngành động vật nguyên sinh | Trùng roi, Trùng biến hình, Trùng giày. |
| Ngành ruột khoang | Thủy tức. |
| Ngành giun | Sán lá gan, Giun đũa, Giun đất |
| Ngành thân mềm | Trai sông |
| Ngành chân khớp | Tôm sông, Châu chấu, Nhện |
| Ngành động vật có xương sống | Cá chép, Ếch đồng, Thằn lằn bóng đuôi dài, Chim bồ câu, Thỏ, Chuột túi. |

|  |
| --- |
| Dạng 3: Tính hợp lí của dữ liệu. |

**Ví dụ 4.**

Bạn Hà đưa ra kết quả điều tra về sự ưa thích các loại màu sắc của 120 học sinh được cho trên biểu đồ hình quạt tròn bên. Hỏi những số liệu mà bạn Hà nêu ra trong biểu đồ hình quạt tròn đã chính xác chưa? Vì sao?

*Lời giải:*

Những số liệu mà bạn Hà nêu ra trong biểu đồ hình quạt tròn chưa chính xác. Vì tổng phần trăm 1 hình tròn là 100% mà tổng tỉ số phần trăm của hình bên là 110% nên số liệu không hợp lí.

**Ví dụ 5.**

Một trường THCS phát động phong trào biểu diễn văn nghệ nhân ngày nhà giáo Việt Nam 20 -11 cho học sinh tất cả các lớp . Biết rằng mỗi lớp chỉ biểu diễn tối đa 3 tiết mục . Bảng thống kê số tiết mục đăng kí tham gia diễn văn nghệ của từng lớp như sau :

|  |  |
| --- | --- |
| Lớp | Số tiết mục |
| 6A | 2 |
| 6B | 1 |
| 6C | 3 |
| 7A | 5 |
| 7B | 3 |
| 7C | 2 |
| 8A | 2 |
| 8B | 2 |
| 9A | 3 |
| 9B | 2 |

Theo em số liệu nào trong bảng trên là không hợp lí? Vì sao?

*Lời giải:*

Ta thấy mỗi lớp chỉ được tham gia tối đa 3 tiết mục mà số liệu bảng trên đưa ra lớp 7A 5 tiết mục vì thế bảng số liệu trên đưa ra là không hợp lí .

**C. BÀI TẬP VẬN DỤNG**

**Bài 1.**

Lớp trưởng lớp 8A thống kê số bạn có sở thích môn thể thao trong một lớp như sau :

* Các môn thể thao : *Bóng đá, cầu lông, cờ vua, đá cầu.*
* Số học sinh yêu thích lần lượt là : *40 , 19 , 22 , 10 .*

Hãy phân loại các dữ liệu đó dựa trên tiêu chí : dữ liệu định tính, dữ liệu định lượng.

*Lời giải:*

Dữ liệu định tính : *Bóng đá, cầu lông, cờ vua, đá cầu.*

Dữ liệu định lượng : *40 , 19 , 22 , 10 .*

**Bài 2.**

Thu thập thông tin tỉ lệ phần trăm phân bố đàn lợn của một số vùng như sau:

* Các vùng : *Đồng bằng Cửu Long; Đồng bằng sông Hồng, Trung du và MN phía Bắc , Bắc Trung Bộ và DHMT, Tây Nguyên; Đông Nam Bộ.*
* Tỉ lệ phần trăm của các vùng lần lượt là : *9%; 21%; 26%; 20%; 8%; 16%.*

Hãy phân loại các dữ liệu đó dựa trên tiêu chí : dữ liệu định tính, dữ liệu định lượng.

*Lời giải:*

Dữ liệu định tính : *Đồng bằng Cửu Long; Đồng bằng sông Hồng, Trung du và MN phía Bắc , Bắc Trung Bộ và DHMT, Tây Nguyên; Đông Nam Bộ.*

Dữ liệu định lượng : *9%; 21%; 26%; 20%; 8%; 16%.*

**Bài 3.**

Bạn Nga đã sưu tầm tên những loại cây. Kết quả sưu tầm được như sau : *Bồ công anh; Cây sầu riêng; Cây cam; Cỏ mần trầu; Cây nhọ nồi; Cây mận; Cây ngải cứu.*



a/ Bạn Nga sưu tầm được bao nhiêu cây ?

b/ Hãy sắp xếp các loại cây mà bạn Nga đã sưu tầm theo những nhóm sau :

Nhóm 1: Các loại cây ăn quả.

Nhóm 2. Các loại cây thuốc dùng để trị bệnh.

*Lời giải:*

a/ Bạn Nga sưu tầm được bao nhiêu 7 loại cây.

b/ Sắp xếp các loại cây mà bạn Nga đã sưu tầm theo những nhóm :

Nhóm 1: Các loại cây ăn quả: Cây cam; cây sầu riêng; Cây mận.

Nhóm 2. Các loại cây thuốc dùng để trị bệnh : Cây ngải cứu; cây mần trầu; cây bò công anh; cây nhọ nồi.

**Bài 4.** Thông tin về sự yêu thích các môn học của 120 em học sinh khối 8 được cho bởi biểu đồ hình quạt tròn sau: :

Theo em thông tin đó đã hợp lí chưa ? Vì sao ?

*Lời giải:*

Theo em thông tin về sự yêu thích các môn học của 120 em học sinh khối 8 được cho bởi biểu đồ hình quạt là chưa hợp lí. Vì tổng phần trăm 1 hình tròn là 100% mà tổng tỉ số phần trăm của hình bên là 90% nên số liệu không hợp lí.

**Bài 5.** Thông tin về kết quả xếp loại học lực của lớp 8A học kì I được cho biểu đồ hình quạt tròn bên dưới. Biết lớp 8A có 40 học sinh .Trong đó có 2 học sinh xếp loại giỏi , 9 học sinh xếp loại tiên tiến.

Theo em với thông tin đó có chính xác không ? Giải thích ?

*Lời giải:*

Thông tin về kết quả xếp loại học lực của lớp 8A học kì I được cho biểu đồ hình quạt tròn bên dưới là chưa chính xác. Vì : Số học sinh xếp loại tiên tiến là : 40 . 20 % = 8 (học sinh) mà thông tin đưa ra là 9 học sinh nên thông tin đó chưa chính xác.

**Bài 6.** Lượng mưatrung bình của 4 thángở hai tỉnh A và B *(đơn vị : mm)* được người lập thể hiện qua biểu đồ cột kép ở hình bên dưới . Biết tổng lượng mưa trung bình của hai tỉnh với mỗi tháng không quá 100mm.



Theo em với số liệu được cho trong biểu đồ cột kép số liệu không hợp lí ở tháng mấy ? Giải thích ?

*Lời giải:*

Theo em với số liệu được cho trong biểu đồ cột kép số liệu không hợp lí ở tháng 8 .

Vì : tổng lượng mưa trung bình của hai tỉnh với mỗi tháng không quá 100mm. Mà tháng 8 tổng lượng mưa trung bình của hai tỉnh là :150 mm. Nên số liệu cho không hợp lí.

**Bài 7.** Để đảm bảo kì thi cuối HKI lớp 8 một cách khách quan nhà trường chia phòng . Lớp 8 có 105 học sinh được chia thành 5 phòng. Bảng thống kê số học sinh ở mỗi phòng như sau :

|  |  |
| --- | --- |
| Phòng | Số học sinh |
| 1 | 21 |
| 2 | 21 |
| 3 | 20 |
| 4 | 22 |
| 5 | 22 |

Theo em số liệu đã cho trong bảng thống kê trên có hợp lí không ? Giải thích ?

*Lời giải:*

Theo em bảng số liệu đã cho trong bảng thống kê trên là chưa hợp lí . Vì : Lớp 8 có 105 học sinh mà lúc chia phòng tổng số học sinh các phòng là 106 nên số liệu thống kê đó không hợp lí.