

## Chương III - THỐNG KÊ

Thống kê là một khoa học được ứng dụng rộng rãi trong các hoạt động kinh tế, xã hội. Ta vẫn thường nghe nói đến thống kê dân số, thống kê sản lượng đạt được hàng năm của một ngành sản xuất, của một xí nghiệp,... Ta cũng thường thấy các biểu đồ trên báo chí, trong các cuộc triển lãm, trên vô tuyến truyền hình,... Qua nghiên cứu, phân tích các thông tin thu thập được, khoa học thống kê cùng với các khoa học kỹ thuật khác giúp cho ta biết được tình hình các hoạt động, diễn biến của các hiện tượng, từ đó dự đoán các khả năng có thể xảy ra, góp phần phục vụ lợi ích con người ngày càng tốt hơn. Trong chương này, ta sẽ bước đầu làm quen với Thống kê mô tả, một bộ phận của khoa học thống kê.

### §1. Thu thập số liệu thống kê, tần số

Các số liệu thu thập được khi điều tra sẽ được ghi lại ra sao ?

#### 1. Thu thập số liệu, bảng số liệu thống kê ban đầu

*Ví dụ* : Khi điều tra về số cây trồng được của mỗi lớp trong dịp phát động phong trào Tết trồng cây, người điều tra lập bảng dưới đây (bảng 1) :

STT	Lớp	Số cây trồng được	STT	Lớp	Số cây trồng được
1	6A	35	11	8A	35
2	6B	30	12	8B	50
3	6C	28	13	8C	35
4	6D	30	14	8D	50
5	6E	30	15	8E	30
6	7A	35	16	9A	35
7	7B	28	17	9B	35
8	7C	30	18	9C	30
9	7D	30	19	9D	30
10	7E	35	20	9E	50

*Bảng 1*

Việc làm trên của người điều tra là *thu thập số liệu* về vấn đề được quan tâm. Các số liệu trên được ghi lại trong một bảng, gọi là *bảng số liệu thống kê ban đầu* (bảng 1).

**?1** *Hãy quan sát bảng 1 để biết cách lập một bảng số liệu thống kê ban đầu trong các trường hợp tương tự. Chẳng hạn như điều tra số con trong từng gia đình (ghi theo tên các chủ hộ) trong một xóm, một phường,...*

Tuỳ theo yêu cầu của mỗi cuộc điều tra mà các bảng số liệu thống kê ban đầu có thể khác nhau. Ví dụ : Bảng điều tra dân số nước ta tại thời điểm 1/4/1999 phân theo giới tính, phân theo thành thị, nông thôn trong từng địa phương (đơn vị là nghìn người) (trích theo tài liệu của Tổng cục Thống kê (bảng 2)) :

Địa phương \ Số dân	Tổng số	Phân theo giới tính		Phân theo thành thị, nông thôn	
		Nam	Nữ	Thành thị	Nông thôn
Hà Nội	2672,1	1336,7	1335,4	1538,9	1133,2
Hải Phòng	1673,0	825,1	847,9	568,2	1104,8
Hưng Yên	1068,7	516,0	552,7	92,6	976,1
Hà Giang	602,7	298,3	304,4	50,9	551,8
Bắc Kạn	275,3	137,6	137,7	39,8	235,5
...	...	...	...	...	...

**Bảng 2**

## 2. Dấu hiệu

### a) Dấu hiệu, đơn vị điều tra

**?2** *Nội dung điều tra trong bảng 1 là gì ?*

Vấn đề hay hiện tượng mà người điều tra quan tâm tìm hiểu gọi là *dấu hiệu* (thường được kí hiệu bằng các chữ in hoa X, Y,...).

Dấu hiệu X ở bảng 1 là *số cây trồng được của mỗi lớp*, còn mỗi lớp là một *đơn vị điều tra*.

**?3** *Trong bảng 1 có bao nhiêu đơn vị điều tra ?*

### b) Giá trị của dấu hiệu, dãy giá trị của dấu hiệu

Mỗi lớp (đơn vị) trồng được một số cây ; chẳng hạn lớp 7A trồng 35 cây, lớp 8D trồng 50 cây (bảng 1).

Như vậy ứng với mỗi đơn vị điều tra có một số liệu, số liệu đó gọi là một giá trị của dấu hiệu. Số các giá trị (không nhất thiết khác nhau) của dấu hiệu đúng bằng số các đơn vị điều tra (thường được kí hiệu là  $N$ ).

Trong ví dụ trên thì các giá trị ở cột thứ ba của bảng 1 (kể từ bên trái sang) gọi là dãy giá trị của dấu hiệu  $X$  (số cây trồng được của mỗi lớp).

**?4** Dấu hiệu  $X$  ở bảng 1 có tất cả bao nhiêu giá trị? Hãy đọc dãy giá trị của  $X$ .

### 3. Tần số của mỗi giá trị

Tiếp tục quan sát bảng 1.

**?5** Có bao nhiêu số khác nhau trong cột số cây trồng được? Nêu cụ thể các số khác nhau đó.

**?6** Có bao nhiêu lớp (đơn vị) trồng được 30 cây (hay giá trị 30 xuất hiện bao nhiêu lần trong dãy giá trị của dấu hiệu  $X$ )? Hãy trả lời câu hỏi tương tự như vậy với các giá trị 28, 50.

Mỗi giá trị có thể xuất hiện một hoặc nhiều lần trong dãy giá trị của dấu hiệu. Số lần xuất hiện của một giá trị trong dãy giá trị của dấu hiệu được gọi là tần số của giá trị đó.

Giá trị của dấu hiệu thường được kí hiệu là  $x$  và tần số của giá trị thường được kí hiệu là  $n$ . Cần phân biệt  $n$  (tần số của một giá trị) với  $N$  (số các giá trị). Cũng như vậy, cần phân biệt  $X$  (kí hiệu đối với dấu hiệu) và  $x$  (kí hiệu đối với giá trị của dấu hiệu).

**?7** Trong dãy giá trị của dấu hiệu ở bảng 1 có bao nhiêu giá trị khác nhau? Hãy viết các giá trị đó cùng tần số của chúng.

- Các số liệu thu thập được khi điều tra về một dấu hiệu gọi là số liệu thống kê. Mỗi số liệu là một giá trị của dấu hiệu.
- Số tất cả các giá trị (không nhất thiết khác nhau) của dấu hiệu bằng số các đơn vị điều tra.
- Số lần xuất hiện của một giá trị trong dãy giá trị của dấu hiệu là tần số của giá trị đó.

► **Chú ý :**

– Ta chỉ xem xét, nghiên cứu các dấu hiệu mà giá trị của nó là các số ; tuy nhiên cần lưu ý rằng : không phải mọi dấu hiệu đều có giá trị là số.

*Ví dụ :* Khi điều tra về sự ham thích đối với bóng đá của một nhóm học sinh thì ứng với một bạn nào đó trong nhóm, người điều tra phải ghi lại mức độ ham thích của bạn ấy theo một trong các mức đã quy định, chẳng hạn : rất thích, thích, không thích.

– Trong trường hợp chỉ chú ý tới các giá trị của dấu hiệu thì bảng số liệu thống kê ban đầu có thể chỉ gồm các cột số. Chẳng hạn, từ bảng 1 ta có bảng 3 dưới đây :

35	30	28	30	30	35	28	30	30	35
35	50	35	50	30	35	35	30	30	50

**Bảng 3**

### Bài tập

1. Lập bảng số liệu thống kê ban đầu cho một cuộc điều tra nhỏ về một dấu hiệu mà em quan tâm (điểm một bài kiểm tra của mỗi em trong lớp, số bạn nghỉ học trong một ngày của mỗi lớp trong trường, số con trong từng gia đình sống gần nhà em,...).
2. Hàng ngày, bạn An thử ghi lại thời gian cần thiết để đi từ nhà đến trường và thực hiện điều đó trong 10 ngày. Kết quả thu được ở bảng 4 :

Số thứ tự của ngày	1	2	3	4	5	6	7	8	9	10
Thời gian (phút)	21	18	17	20	19	18	19	20	18	19

**Bảng 4**

- a) Dấu hiệu mà bạn An quan tâm là gì và dấu hiệu đó có tất cả bao nhiêu giá trị ?
- b) Có bao nhiêu giá trị khác nhau trong dãy giá trị của dấu hiệu đó ?
- c) Viết các giá trị khác nhau của dấu hiệu và tìm tần số của chúng.

## Luyện tập

3. Thời gian chạy 50 mét của các học sinh trong một lớp 7 được thầy giáo dạy Thể dục ghi lại trong hai bảng 5 và 6 :

Số thứ tự của học sinh nam	Thời gian (giây)
1	8,3
2	8,5
3	8,5
4	8,7
5	8,5
6	8,7
7	8,3
8	8,7
9	8,5
10	8,4
11	8,5
12	8,4
13	8,5
14	8,8
15	8,8
16	8,5
17	8,7
18	8,7
19	8,5
20	8,4

**Bảng 5**

Số thứ tự của học sinh nữ	Thời gian (giây)
1	9,2
2	8,7
3	9,2
4	8,7
5	9,0
6	9,0
7	9,0
8	8,7
9	9,2
10	9,2
11	9,2
12	9,0
13	9,3
14	9,2
15	9,3
16	9,3
17	9,3
18	9,0
19	9,2
20	9,3

**Bảng 6**

Hãy cho biết :

- a) Dấu hiệu chung cần tìm hiểu (ở cả hai bảng)
- b) Số các giá trị của dấu hiệu và số các giá trị khác nhau của dấu hiệu (đối với từng bảng)
- c) Các giá trị khác nhau của dấu hiệu và tần số của chúng (đối với từng bảng).

4. Chọn 30 hộp chè một cách tùy ý trong kho của một cửa hàng và đem cân, kết quả được ghi lại trong bảng 7 (sau khi đã trừ khối lượng của vỏ) :

Khối lượng chè trong từng hộp (tính bằng gam)		
100	100	101
100	101	100
98	100	100
98	102	98
99	99	102
100	101	101
100	100	100
102	100	100
100	100	99
100	99	100

**Bảng 7**

Hãy cho biết :

- Dấu hiệu cân tìm hiểu và số các giá trị của dấu hiệu đó
- Số các giá trị khác nhau của dấu hiệu
- Các giá trị khác nhau của dấu hiệu và tần số của chúng.