

## THỐNG KÊ

### A. NHỮNG KIẾN THỨC CẦN NHỚ

- Một *dấu hiệu* là một vấn đề nào đó mà người điều tra quan tâm. Mỗi đối tượng điều tra gọi là một *đơn vị điều tra*. Mỗi đơn vị điều tra tương ứng với một số liệu gọi là *giá trị của dấu hiệu* trên đơn vị điều tra đó.
- Một tập con hữu hạn các đơn vị điều tra gọi là một *mẫu*. Tập hợp các số liệu thu được sau khi điều tra trên mẫu gọi là một *mẫu số liệu*.
- Bảng phân bố tần số gồm hai dòng (hoặc hai cột). Dòng (cột) đầu ghi các giá trị khác nhau của mẫu số liệu. Dòng (cột) thứ hai ghi tần số (số lần xuất hiện của mỗi giá trị trong mẫu số liệu) tương ứng. Nếu bổ sung một dòng (cột) thứ ba ghi tần suất (tỉ số % giữa tần số và kích thước mẫu) thì ta có bảng phân bố tần số - tần suất.
- Khi số liệu được ghép thành lớp, mỗi lớp gồm các số liệu nằm trong một đoạn (hay nửa khoảng) nào đó, ta có bảng phân bố tần số (tần số - tần suất) ghép lớp.
- Số trung bình được tính bởi công thức

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i.$$

- Phương sai được tính bởi công thức

$$s^2 = \frac{1}{N} \sum_{i=1}^N x_i^2 - \frac{1}{N^2} \left( \sum_{i=1}^N x_i \right)^2.$$

$$s^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2.$$

Độ lệch chuẩn  $s$  là căn bậc hai của phương sai.

- Nếu mẫu số liệu được cho dưới dạng một bảng phân bố tần số thì

$$\bar{x} = \frac{1}{N} \sum_{i=1}^m n_i x_i ;$$

$$s^2 = \frac{1}{N} \sum_{i=1}^m n_i x_i^2 - \frac{1}{N^2} \left( \sum_{i=1}^m n_i x_i \right)^2 ;$$

trong đó  $n_i$  là tần số của số liệu  $x_i$  ( $i = 1, 2, \dots, m$ ),  $\sum_{i=1}^m n_i = N$ .

- Nếu mẫu số liệu được cho dưới dạng bảng tần số ghép lớp (Bảng 7a, 7b SGK) thì

$$\bar{x} \approx \frac{1}{N} \sum_{i=1}^m n_i x_i ;$$

$$s^2 \approx \frac{1}{N} \sum_{i=1}^m n_i x_i^2 - \frac{1}{N^2} \left( \sum_{i=1}^m n_i x_i \right)^2 ;$$

trong đó  $n_i$  là tần số của lớp thứ  $i$ ,  $x_i$  là giá trị đại diện của lớp thứ  $i$ , tức là trung điểm của đoạn (hay nửa khoảng) ứng với lớp thứ  $i$  ( $i = 1, 2, \dots, m$ ).

- Số trung vị là giá trị thứ  $\frac{N+1}{2}$  của mẫu số liệu nếu  $N$  lẻ và là trung bình cộng của giá trị thứ  $\frac{N}{2}$  và  $\frac{N}{2} + 1$  nếu  $N$  chẵn (khi xếp các giá trị của mẫu số liệu theo thứ tự tăng dần). Số trung vị được kí hiệu là  $M_e$ .
- Mốt là giá trị có tần số lớn nhất. Mốt được kí hiệu là  $M_o$ .