

### §3. CÁC SỐ ĐẶC TRƯNG CỦA MẪU SỐ LIỆU

#### A. TÓM TẮT LÝ THUYẾT.

##### 1. Số trung bình

- Với mẫu số liệu kích thước N là  $x_1, x_2, \dots, x_N$  :

$$\bar{x} = \frac{\sum_{i=1}^N x_i}{N} = \frac{x_1 + x_2 + \dots + x_N}{N}$$

- Với mẫu số liệu được cho bởi bảng phân bố tần số:

$$\bar{x} = \frac{\sum_{i=1}^N n_i x_i}{N} = \frac{n_1 x_1 + n_2 x_2 + \dots + n_k x_k}{N}$$

- Với mẫu số liệu được cho bởi bảng phân bố tần số ghép lớp:

$$\bar{x} = \frac{\sum_{i=1}^N n_i c_i}{N} = \frac{n_1 c_1 + n_2 c_2 + \dots + n_k c_k}{N} \quad (c_i \text{ là giá trị đại diện của lớp thứ } i)$$

##### 2. Số trung vị

Giả sử ta có một mẫu gồm N số liệu được sắp xếp theo thứ tự không giảm (hoặc không tăng). Khi đó **số trung vị**  $M_e$  là:

- Số đứng giữa nếu N lẻ;
- Trung bình cộng của hai số đứng giữa (số thứ  $\frac{N}{2}$  và  $\frac{N}{2} + 1$ ) nếu N chẵn.

##### 3. Mốt

**Mốt** của một bảng phân bố tần số là giá trị có tần số lớn nhất và được kí hiệu là  $M_o$ .

**Chú ý:** – Số trung bình của mẫu số liệu được dùng làm đại diện cho các số liệu của mẫu.

– Nếu các số liệu trong mẫu có sự chênh lệch quá lớn thì dùng số trung vị làm đại diện cho các số liệu của mẫu.

– Nếu quan tâm đến giá trị có tần số lớn nhất thì dùng mốt làm đại diện. Một mẫu số liệu có thể có nhiều mốt.

##### 4. Phương sai và độ lệch chuẩn

Để đo mức độ chênh lệch (**độ phân tán**) giữa các giá trị của mẫu số liệu so với số trung bình ta dùng **phương sai**  $s^2$  và **độ lệch chuẩn**  $s = \sqrt{s^2}$ .

- Với mẫu số liệu kích thước N là  $x_1, x_2, \dots, x_N$  :

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

- Với mẫu số liệu được cho bởi bảng phân bố tần số, tần suất:

$$\begin{aligned} s^2 &= \frac{1}{N} \sum_{i=1}^k n_i (x_i - \bar{x})^2 = \frac{1}{N} \sum_{i=1}^k n_i x_i^2 - \frac{1}{N^2} \left( \sum_{i=1}^k n_i x_i \right)^2 \\ &= \sum_{i=1}^k f_i (x_i - \bar{x})^2 = \sum_{i=1}^k f_i x_i^2 - \left( \sum_{i=1}^k f_i x_i \right)^2 \end{aligned}$$

- Với mẫu số liệu được cho bởi bảng phân bố tần số, tần suất ghép lớp:

$$\begin{aligned} s^2 &= \frac{1}{N} \sum_{i=1}^k n_i (c_i - \bar{x})^2 = \frac{1}{N} \sum_{i=1}^k n_i c_i^2 - \frac{1}{N^2} \left( \sum_{i=1}^k n_i c_i \right)^2 \\ &= \sum_{i=1}^k f_i (c_i - \bar{x})^2 = \sum_{i=1}^k f_i c_i^2 - \left( \sum_{i=1}^k f_i c_i \right)^2 \end{aligned}$$

( $c_i, n_i, f_i$  là giá trị đại diện, tần số, tần suất của lớp thứ  $i$ ;

$N$  là số các số liệu thống kê  $N = n_1 + n_2 + \dots + n_k$ )

**Chú ý:** Phương sai và độ lệch chuẩn càng lớn thì độ phân tán (so với số trung bình) của các số liệu thống kê càng lớn.