

# CÁC SỐ ĐẶC TRƯNG ĐO MỨC ĐỘ PHÂN TÁN CỦA MẪU SỐ LIỆU GHÉP NHÓM

## §1. KHOẢNG BIẾN THIÊN VÀ KHOẢNG TƯ PHÂN VỊ

### A TÓM TẮT LÝ THUYẾT

#### 1. Khoảng biến thiên

**Định nghĩa 1.1.** Cho mẫu số liệu ghép nhóm:

|        |              |     |                  |     |                  |
|--------|--------------|-----|------------------|-----|------------------|
| Nhóm   | $[a_1; a_2)$ | ... | $[a_i; a_{i+1})$ | ... | $[a_k; a_{k+1})$ |
| Tần số | $m_1$        | ... | $m_i$            | ... | $m_k$            |

Bảng 3.1. Mẫu số liệu ghép nhóm

Trong đó các tần số  $m_1 > 0, \dots, m_k > 0$  và  $n = m_1 + \dots + m_k$  là cỡ mẫu.

Khoảng biến thiên của mẫu số liệu ghép nhóm trên là  $R = a_{k+1} - a_1$ . Chỉ ra rằng khoảng biến thiên của mẫu số liệu ghép nhóm trong Bảng 3.1 lớn hơn khoảng biến thiên của mẫu số liệu gốc.

**Ý nghĩa.**

- ✔ Khoảng biến thiên của mẫu số liệu ghép nhóm xấp xỉ cho khoảng biến thiên của mẫu số liệu gốc.
- ✔ Khoảng biến thiên được dùng để đo mức độ phân tán của mẫu số liệu ghép nhóm.
- ✔ Khoảng biến thiên càng lớn thì mẫu số liệu càng phân tán.

#### Ví dụ 1

Thống kê thời gian sử dụng mạng xã hội trong ngày của các bạn Tổ 1, Tổ 2 lớp 12A, được kết quả như sau:

|                          |           |            |            |            |
|--------------------------|-----------|------------|------------|------------|
| Thời gian sử dụng (phút) | $[0; 10)$ | $[10; 30)$ | $[30; 60)$ | $[60; 90)$ |
| Số học sinh Tổ 1         | 2         | 4          | 3          | 1          |
| Số học sinh Tổ 2         | 5         | 1          | 3          | 0          |

Tìm khoảng biến thiên cho thời gian sử dụng mạng xã hội của học sinh mỗi tổ và giải thích ý nghĩa.

## Ví dụ 2

Thời gian hoàn thành bài kiểm tra môn Toán của các bạn trong lớp 12C được cho trong bảng sau:

|                  |          |          |          |          |
|------------------|----------|----------|----------|----------|
| Thời gian (phút) | [25; 30) | [30; 35) | [35; 40) | [40; 45) |
| Số học sinh Tổ 1 | 8        | 16       | 4        | 2        |

- Tính khoảng biến thiên  $R$  cho mẫu số liệu ghép nhóm trên.
- Nếu biết học sinh hoàn thành bài kiểm tra sớm nhất mất 27 phút và muộn nhất mất 43 phút thì khoảng biến thiên của mẫu số liệu gốc là bao nhiêu?

## 2. Khoảng tứ phân vị

**Định nghĩa 1.2.** Tứ phân vị thứ  $r$  là  $Q_r = a_p + \frac{r \cdot n}{4} - (m_1 + \dots + m_{p-1})}{m_p} (a_{p+1} - a_p)$ , trong đó  $[a_p; a_{p+1})$  là nhóm chứa tứ phân vị thứ  $r$  với  $r = 1, 2, 3$ . Cho mẫu số liệu ghép nhóm:

|        |              |     |                  |     |                  |
|--------|--------------|-----|------------------|-----|------------------|
| Nhóm   | $[a_1; a_2)$ | ... | $[a_i; a_{i+1})$ | ... | $[a_k; a_{k+1})$ |
| Tần số | $m_1$        | ... | $m_i$            | ... | $m_k$            |

Bảng 3.1. Mẫu số liệu ghép nhóm

Khoảng tứ phân vị của mẫu số liệu ghép nhóm, kí hiệu  $\Delta_Q$ , là hiệu số giữa tứ phân vị thứ ba  $Q_3$  và tứ phân vị thứ nhất  $Q_1$  của mẫu số liệu đó, tức là  $\Delta_Q = Q_3 - Q_1$ .

**Ý nghĩa.**

- ✔ Khoảng tứ phân vị của mẫu số liệu ghép nhóm xấp xỉ cho khoảng tứ phân vị của mẫu số liệu gốc.
  - ✔ Khoảng tứ phân vị cũng được dùng để đo mức độ phân tán của mẫu số liệu ghép nhóm.
  - ✔ Khoảng tứ phân vị càng lớn thì mẫu số liệu càng phân tán.
- ⚠ Do khoảng tứ phân vị của mẫu số liệu ghép nhóm chỉ phụ thuộc vào nửa giữa của mẫu số liệu, nên không bị ảnh hưởng bởi các giá trị bất thường và có thể dùng đại lượng này để loại giá trị bất thường.

## B VÍ DỤ MINH HỌA

### Ví dụ 1

Thời gian chờ khám bệnh của các bệnh nhân tại phòng khám X được cho trong bảng sau:

| Thời gian (phút) | [0; 5) | [5; 10) | [10; 15) | [15; 20) |
|------------------|--------|---------|----------|----------|
| Số bệnh nhân     | 3      | 12      | 15       | 8        |

- Tìm khoảng tứ phân vị của mẫu số liệu ghép nhóm này.
- Từ một số liệu về thời gian chờ khám bệnh của các bệnh nhân tại phòng khám Y người ta tính được khoảng tứ phân vị bằng 9,23. Hỏi thời gian chờ của bệnh nhân tại phòng khám nào phân tán hơn?

### Ví dụ 2

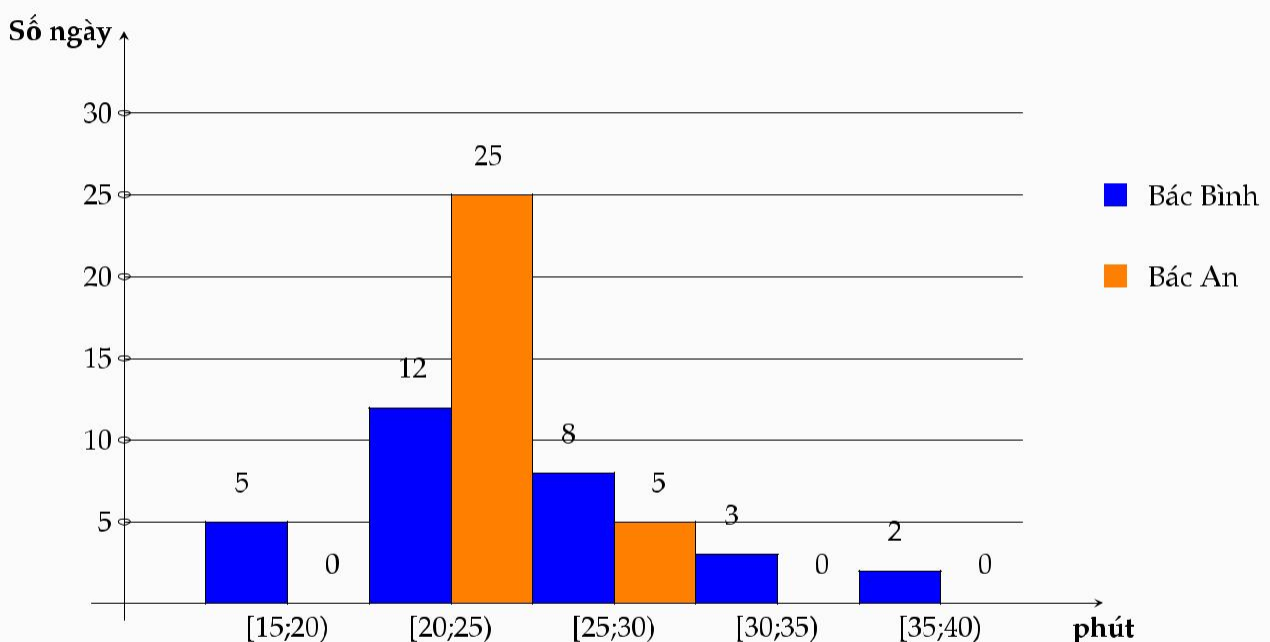
Một người ghi lại thời gian đàm thoại của một số cuộc gọi cho kết quả như bảng sau:

| Thời gian $t$ (phút) | Số cuộc gọi |
|----------------------|-------------|
| $0 \leq t < 1$       | 8           |
| $1 \leq t < 2$       | 17          |
| $2 \leq t < 3$       | 25          |
| $3 \leq t < 4$       | 20          |
| $4 \leq t < 5$       | 10          |

Tính khoảng tứ phân vị của mẫu số liệu ghép nhóm trên.

### Ví dụ 3

Biểu đồ dưới đây thống kê thời gian tập thể dục buổi sáng mỗi ngày trong tháng 9/2022 của bác Bình và bác An.



Hãy so sánh khoảng tứ phân vị của mẫu số liệu ghép nhóm về thời gian tập thể dục buổi sáng mỗi ngày của bác Bình và bác An.