

# 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

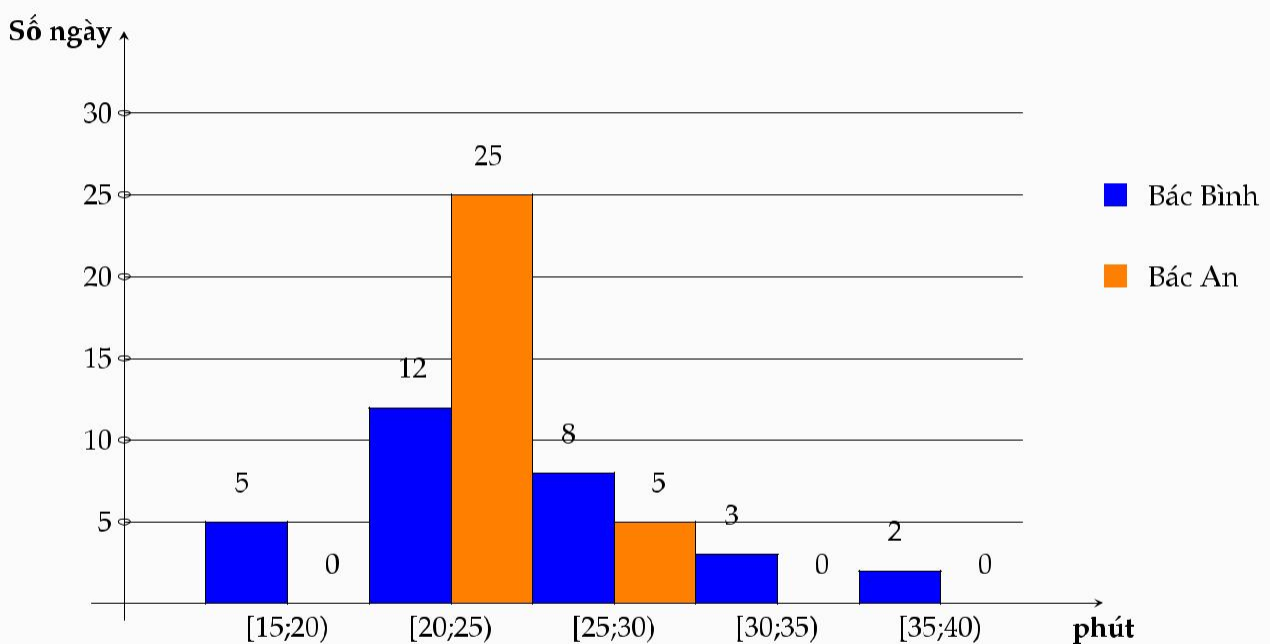
Một người ghi lại thời gian đàm thoại của một số cuộc gọi cho kết quả như bảng sau:

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.

#### Ví dụ 4

Hàng ngày ông Thắng đều đi xe buýt từ nhà đến cơ quan. Dưới đây là bảng thống kê thời gian của 100 lần ông Thắng đi xe buýt từ nhà đến cơ quan.

Thời gian (phút)	[15; 18)	[18; 21)	[21; 24)	[24; 27)	[27; 30)	[30; 33)
Số lượt	22	38	27	8	4	1

- Hãy tìm khoảng tứ phân vị của mẫu số liệu ghép nhóm trên.
- Biết rằng trong 100 lần đi trên, chỉ có đúng một lần ông Thắng đi hết hơn 29 phút. Thời gian của lần đi đó có phải là giá trị ngoại lệ không?

#### Ví dụ 5

Hãy tìm khoảng biến thiên và khoảng tứ phân vị của mẫu số liệu ghép nhóm ở Ví dụ 4 sau khi đã loại bỏ các giá trị ngoại lệ. Em có nhận xét gì về khoảng biến thiên, khoảng tứ phân vị vừa tìm được và khoảng biến thiên, khoảng tứ phân vị ban đầu?

#### Ví dụ 6

Bạn Trang thống kê lại chiều cao (đơn vị: cm) của các bạn học sinh nữ lớp 12C và lớp 12D ở bảng sau.

Chiều cao \ Lớp	[155; 160)	[160; 165)	[165; 170)	[170; 175)	[175; 180)	[180; 185)
12C	2	7	12	3	0	1
12D	5	9	8	2	1	0

Hãy so sánh độ phân tán của nửa giữa hai mẫu số liệu chiều cao của các học sinh nữ lớp 12C và 12D.

#### Ví dụ 7

Giả sử kết quả khảo sát hai khu vực A và B về độ tuổi kết hôn của một số phụ nữ vừa lập gia đình được cho ở bảng sau

Tuổi kết hôn	[19; 22)	[22; 25)	[25; 28)	[28; 31)	[31; 34)
Số phụ nữ khu vực A	10	27	31	25	7
Số phụ nữ khu vực B	47	40	11	2	0

- Hãy tìm khoảng biến thiên và khoảng tứ phân vị của từng mẫu số liệu ghép nhóm ứng với mỗi khu vực A và B.
- Nếu so sánh theo khoảng tứ phân vị thì phụ nữ ở khu vực nào có độ tuổi kết hôn đồng đều hơn?



# C PHÂN LOẠI VÀ PHƯƠNG PHÁP GIẢI TOÁN

## Dạng 1 Tìm khoảng biến thiên của mẫu số liệu ghép nhóm

- Xác định  $a_1$  là giá trị đầu mút trái của nhóm đầu tiên và  $a_{k+1}$  là giá trị đầu mút phải của nhóm cuối cùng có chứa dữ liệu (tần số khác 0).
- Khoảng biến thiên của mẫu số liệu ghép nhóm là  $R = a_{k+1} - a_1$ .

## 1. HỆ THỐNG BÀI TẬP

**PHẦN I. Câu trắc nghiệm nhiều phương án lựa chọn. Mỗi câu hỏi học sinh chỉ chọn một phương án.**

### Câu 1

Bảng 3.2 biểu thị kết quả điều tra thời gian sử dụng Internet hằng ngày của một số người.

Bảng 3.2. Thời gian sử dụng Internet hằng ngày

Thời gian (phút)	[30;60)	[60;90)	[90;120)	[120;150)	[150;180)
Số người	2	4	10	5	3

Khoảng biến thiên của mẫu số liệu đã cho bằng

- (A) 150.                      (B) 160.                      (C) 140.                      (D) 170.

### Câu 2

Bảng 3.3 thống kê thành tích nhảy xa của một số học sinh lớp 12. Khoảng biến thiên thành tích nhảy xa của số học sinh này.

Bảng 3.3. Thành tích nhảy xa của một số học sinh lớp 12

Thành tích (cm)	[150;180)	[180;210)	[210;240)	[240;270)	[270;300)
Số học sinh	3	5	28	14	8

- (A) 150.                      (B) 160.                      (C) 140.                      (D) 170.

### Câu 3

Để chuẩn bị mở một trung tâm thể dục thể thao, anh Dũng đã tiến hành điều tra tuổi thọ của máy chạy bộ do hai hãng X, Y sản xuất. Bảng 3.4 biểu thị hai mẫu số liệu mà anh thu thập được qua Internet. Gọi khoảng biến thiên của tuổi thọ máy chạy bộ do hãng X và hãng Y sản xuất tương ứng là  $R_X$  và  $R_Y$ . Chọn phương án đúng?

Bảng 3.4. Tuổi thọ của máy chạy bộ (đơn vị: năm)

Tuổi thọ	[2;4)	[4;6)	[6;8)	[8;10)	[10;12)
Số máy của hãng X	7	20	36	20	17
Số máy của hãng Y	0	20	35	35	10

- (A)  $R_X > R_Y$ .                      (B)  $R_X = R_Y$ .                      (C)  $R_X < R_Y$ .                      (D)  $R_X = 12$ .