

CHUYÊN ĐỀ 9. KHOẢNG BIẾN THIÊN VÀ KHOẢNG TỨ PHÂN VỊ

PHẦN A. LÝ THUYẾT VÀ VÍ DỤ

NHẮC LẠI LÝ THUYẾT TOÁN 11

1. Ở lớp 10, chúng ta đã biết cách tính các số đặc trưng của mẫu số liệu không ghép nhóm (còn gọi là mẫu số liệu gốc) $x_1, x_2, \dots, x_n$. Tuy nhiên, trong nhiều trường hợp ta không có mẫu số liệu gốc mà chỉ có mẫu số liệu ghép nhóm dạng

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

Khi đó, các số đặc trưng của mẫu số liệu ghép nhóm xấp xỉ cho các số đặc trưng của mẫu số liệu gốc. Các số đặc trưng đo xu thế trung tâm của mẫu số liệu ghép nhóm gồm số trung bình, trung vị, tứ phân vị và mótt.

$$\bar{x} = \frac{m_1 x_1 + \dots + m_k x_k}{n}$$

2. Số trung bình của mẫu số liệu ghép nhóm là

trong đó, $n = m_1 + \dots + m_k$ là tổng số quan sát (còn gọi là cỡ mẫu) và $x_i = \frac{a_i + a_{i+1}}{2}$ gọi là giá trị đại diện của nhóm $[a_i; a_{i+1})$. Số trung bình của mẫu số liệu ghép nhóm xấp xỉ cho số trung bình của mẫu số liệu gốc, nó cho biết vị trí trung tâm của mẫu số liệu và có thể dùng để đại diện cho mẫu số liệu.

3. Để tính trung vị M_e của mẫu số liệu ghép nhóm ta làm như sau:

Bước 1. Xác định nhóm chứa trung vị. Giả sử đó là nhóm thứ $j: [a_j; a_{j+1})$.

$$M_e = a_j + \frac{\frac{n}{2} - (m_1 + \dots + m_{j-1})}{m_j} \cdot (a_{j+1} - a_j)$$

Bước 2. Trung vị là

trong đó n là cỡ mẫu. Với $j=1$ ta quy ước $m_1 + \dots + m_{j-1} = 0$. Trung vị chính là tứ phân vị thứ hai Q_2 . Trung vị của mẫu số liệu ghép nhóm xấp xỉ cho trung vị của mẫu số liệu gốc, nó chia mẫu số liệu thành hai phần, mỗi phần chứa 50% giá trị.

4. Để tính tứ phân vị thứ nhất Q_1 của mẫu số liệu ghép nhóm trước hết ta xác định nhóm chứa

Q_1 , giả sử đó là nhóm thứ $p: [a_p; a_{p+1})$. Khi đó,
$$Q_1 = a_p + \frac{\frac{n}{4} - (m_1 + \dots + m_{p-1})}{m_p} \cdot (a_{p+1} - a_p)$$

trong đó n là cỡ mẫu, với $p=1$ ta quy ước $m_1 + \dots + m_{p-1} = 0$.

Để tính tứ phân vị thứ ba Q_3 của mẫu số liệu ghép nhóm trước hết ta xác định nhóm chứa Q_3 . Giả

sử đó là nhóm thứ $p: [a_p; a_{p+1})$. Khi đó,
$$Q_3 = a_p + \frac{\frac{3n}{4} - (m_1 + \dots + m_{p-1})}{m_p} \cdot (a_{p+1} - a_p)$$

trong đó n là cỡ mẫu, với $p=1$ ta quy ước $m_1 + \dots + m_{p-1} = 0$.

Để xác định nhóm chứa tứ phân vị thứ $r (r=1,2,3)$ ta có thể dựa vào tính chất có khoảng $\frac{nr}{4}$ số giá trị nhỏ hơn tứ phân vị đó.

Các tứ phân vị Q_1, Q_2, Q_3 của mẫu số liệu ghép nhóm xấp xỉ cho các tứ phân vị của mẫu số liệu gốc, chúng chia mẫu số liệu thành 4 phần, mỗi phần chứa 25% giá trị.

5. Để tìm một của mẫu số liệu ghép nhóm, ta thực hiện theo các bước sau:

Bước 1. Xác định nhóm có tần số lớn nhất (gọi là nhóm một), giả sử là nhóm $j: [a_j; a_{j+1})$.

$$M_0 = a_j + \frac{(m_j - m_{j-1})}{(m_j - m_{j-1}) + (m_j - m_{j+1})} \cdot h,$$

Bước 2. Một được xác định là

trong đó h là độ rộng của nhóm và ta quy ước $m_0 = m_{k+1} = 0$.

Một của mẫu số liệu ghép nhóm xấp xỉ cho một của mẫu số liệu gốc, nó được dùng để đo xu thế trung tâm của mẫu số liệu.

Lưu ý. Người ta chỉ định nghĩa một cho mẫu ghép nhóm có độ dài các nhóm bằng nhau. Một mẫu có thể không có một hoặc có nhiều hơn 1 một. Khi tần số của các nhóm bằng nhau thì mẫu số liệu ghép nhóm không có một.

6. Đối với dữ liệu rời rạc, người ta thường cho các nhóm dưới dạng $k_1 - k_2$ trong đó $k_1, k_2 \in \mathbb{N}$.

Nhóm $k_1 - k_2$ được hiểu là nhóm gồm các giá trị $k_1, k_1 + 1, \dots, k_2$. Khi đó, ta cần hiệu chỉnh mẫu dữ liệu ghép nhóm trước khi thực hiện tính toán các số đặc trưng bằng cách hiệu chỉnh nhóm $k_1 - k_2$ thành nhóm $[k_1 - 0,5; k_2 + 0,5)$.

1. KHOẢNG BIẾN THIÊN

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

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

Khoảng biến thiên của mẫu số liệu ghép nhóm trên là $R = a_{k+1} - a_1$.

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 bảng 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.

Giải

Gọi R_1, R_2 tương ứng là khoảng biến thiên của mẫu số liệu ghép nhóm về thời gian sử dụng mạng xã hội trong ngày của các bạn Tổ 1 và Tổ 2.

Ta có: $R_1 = 90 - 0 = 90$ và $R_2 = 60 - 0 = 60$.

Do $R_1 > R_2$ nên ta có thể kết luận rằng thời gian sử dụng mạng xã hội trong ngày của các bạn Tổ 1 phân tán hơn thời gian sử dụng mạng xã hội của các bạn Tổ 2.

2. KHOẢNG TỨ PHÂN VỊ

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

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

Khoảng tứ phân vị của mẫu số liệu ghép nhóm, kí hiệu là Δ_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.

Nhận xét. 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.

Ví dụ 2: 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

a) Tìm khoảng tứ phân vị của mẫu số liệu ghép nhóm này.

b) Từ một mẫu 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?

Giải

a) Cỡ mẫu là $n = 3 + 12 + 15 + 8 = 38$. Gọi $x_1, \dots, x_{38}$ là thời gian chờ khám bệnh của 38 bệnh nhân này và giả sử rằng dãy số liệu gốc này đã được sắp xếp theo thứ tự tăng dần. Tứ phân vị thứ nhất của mẫu số liệu gốc là x_{10} nên nhóm chứa tứ phân vị thứ nhất là nhóm $[5; 10)$ và ta có:

$$Q_1 = 5 + \left[\frac{\frac{38}{4} - 3}{12} \right] \cdot 5 \approx 7,71.$$

Tứ phân vị thứ ba của mẫu số liệu gốc là x_{29} nên nhóm chứa tứ phân vị thứ ba là nhóm $[10;15)$ và ta có:

$$Q_3 = 10 + \left[\frac{\frac{3 \cdot 38}{4} - 15}{15} \right] \cdot 5 = 14,5.$$

Vậy khoảng tứ phân vị của mẫu số liệu ghép nhóm là: $\Delta_Q = Q_3 - Q_1 \approx 14,5 - 7,71 = 6,79$.

b) Do $\Delta_Q = 6,79 < 9,23$ nên thời gian chờ của bệnh nhân tại phòng khám Y phân tán hơn thời gian chờ của bệnh nhân tại phòng khám X .