

# Bài 14. CÁC SỐ ĐẶC TRƯNG ĐO MỨC ĐỘ PHÂN TÁN CHO MẪU SỐ LIỆU KHÔNG GHÉP NHÓM

| Fanpage: Nguyễn Bảo Vương

## PHẦN A. LÝ THUYẾT

### I. Khoảng biến thiên. Khoảng tứ phân vị

#### 1. Định nghĩa

- Trong một mẫu số liệu, khoảng biến thiên là hiệu số giữa giá trị lớn nhất và giá trị nhỏ nhất của mẫu số liệu đó.

Ta có thể tính khoảng biến thiên  $R$  của mẫu số liệu theo công thức sau:  $R = x_{\max} - x_{\min}$ , trong đó  $x_{\max}$  là giá trị lớn nhất,  $x_{\min}$  là giá trị nhỏ nhất của mẫu số liệu đó.

- Giả sử  $Q_1, Q_2, Q_3$  là tứ phân vị của mẫu số liệu. Ta gọi hiệu  $\Delta_Q = Q_3 - Q_1$  là khoảng tứ phân vị, của mẫu số liệu đó.

*Chú ý:* Khoảng tứ phân vị của mẫu số liệu còn gọi là khoảng trải giữa (tiếng Anh là InterQuartile Range -  $IQR$ ) của mẫu số liệu đó.

*Ví dụ 1.* Mẫu số liệu thống kê chiều cao (đơn vị: mét) của 15 cây bạch đàn là:

6,3 6,6 7,5 8,2 8,3 7,8 7,9 9,0 8,9 7,2 7,5 8,7 7,7 8,8 7,6 (2)

a) Tìm khoảng biến thiên của mẫu số liệu (2).

b) Tìm khoảng tứ phân vị của mẫu số liệu (2).

**Giải**

a) Trong mẫu số liệu (2), số lớn nhất là 9,0 và số bé nhất là 6,3. Vậy khoảng biến thiên của mẫu số liệu (2) là:

$$R = x_{\max} - x_{\min} = 9,0 - 6,3 = 2,7(m).$$

b) Sắp xếp các số liệu của mẫu (2) theo thứ tự tăng dần, ta được:

6,3 6,6 7,2 7,5 7,5 7,6 7,7 7,8 7,9 8,2 8,3 8,7 8,8 8,9 9,0

Do đó  $Q_1 = 7,5(m); Q_2 = 7,8(m); Q_3 = 8,7(m)$ .

Vậy khoảng tứ phân vị của mẫu số liệu (2) là:  $\Delta_Q = Q_3 - Q_1 = 8,7 - 7,5 = 1,2(m)$ .

#### 2. Ý nghĩa

a) **Ý nghĩa của khoảng biến thiên:** Khoảng biến thiên của mẫu số liệu phản ánh sự "dao động", "sự dàn trải" của các số liệu trong mẫu đó. Khoảng biến thiên được sử dụng trong nhiều tình huống thực tiễn, chẳng hạn: tìm ra sự phân tán điểm kiểm tra của một lớp học hay xác định phạm vi giá cả của một dịch vụ ...

Theo cách nhìn như ở trong vật lí, ở đó biên độ dao động phản ánh khoảng cách từ điểm cân bằng đến điểm xa nhất của dao động, nếu coi số trung bình cộng là "điểm cân bằng" của mẫu số liệu thì khoảng biến thiên của mẫu số liệu có thể xem như hai lần biên độ dao động của các số trong mẫu đó quanh điểm cân bằng.

Trong các đại lượng đo mức độ phân tán của mẫu số liệu, khoảng biến thiên là đại lượng dễ hiểu, dễ tính toán và tương đối tốt đối với các mẫu số liệu nhỏ. Tuy nhiên, do khoảng biến thiên chỉ sử dụng hai giá trị

$x_{\max}$  và  $x_{\min}$  của mẫu số liệu nên đại lượng đó chưa diễn giải đầy đủ sự phân tán của các số liệu trong mẫu.

Ngoài ra, giá trị của khoảng biến thiên sẽ bị ảnh hưởng bởi các giá trị bất thường của mẫu số liệu đó. Trong những trường hợp như vậy, khoảng biến thiên của mẫu số liệu không phản ánh chính xác độ dàn trải của mẫu số liệu.

b) **Ý nghĩa của khoảng tứ phân vị:** Khoảng tứ phân vị là một đại lượng cho biết mức độ phân tán của nửa giữa mẫu số liệu và có thể giúp xác định các giá trị bất thường của mẫu số liệu đó. Khoảng tứ phân vị thường được sử dụng thay cho khoảng biến thiên vì nó loại trừ hầu hết giá trị bất thường của mẫu số liệu.

## II. Phương sai

### 1. Định nghĩa

Cho mẫu số liệu thống kê có  $n$  giá trị  $x_1, x_2, \dots, x_n$  và số trung bình cộng là  $\bar{x}$ .

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}$$

Ta gọi số  $s^2$  là phương sai của mẫu số liệu trên.

*Nhận xét*

- Khi có các số liệu bằng nhau, ta có thể tính phương sai theo công thức sau:

+ Phương sai của mẫu số liệu thống kê trong bảng phân bố tần số là:

$$s^2 = \frac{n_1(x_1 - \bar{x})^2 + n_2(x_2 - \bar{x})^2 + \dots + n_k(x_k - \bar{x})^2}{n}$$

Giá trị	$x_1$	$x_2$	...	$x_k$
Tần số	$n_1$	$n_2$	...	$n_k$

trong đó  $n = n_1 + n_2 + \dots + n_k$ ;  $\bar{x}$  là số trung bình cộng của các số liệu đã cho.

+ Phương sai của mẫu số liệu thống kê trong bảng phân bố tần số tương đối là:

$$s^2 = f_1(x_1 - \bar{x})^2 + f_2(x_2 - \bar{x})^2 + \dots + f_k(x_k - \bar{x})^2,$$

Giá trị	$x_1$	$x_2$	...	$x_k$
Tần số tương đối	$f_1$	$f_2$	...	$f_k$

trong đó  $\bar{x}$  là số trung bình cộng của các số liệu đã cho.

- Trong thực tế, người ta còn dùng công thức sau để tính phương sai của một mẫu số liệu:

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n - 1}$$

, trong đó:  $x_i$  là giá trị của quan sát thứ  $i$ ;  $\bar{x}$  là giá trị trung bình và  $n$  là số quan sát trong mẫu số liệu đó.

### 2. ý nghĩa

*Nhận xét:* Phương sai  $s^2$  đánh giá mức độ phân tán của các số liệu thống kê (so với số trung bình cộng).

Phương sai là số đặc trưng đo mức độ phân tán của mẫu số liệu.

**Ví dụ 2.** Xét mẫu số liệu thống kê kết quả 5 bài kiểm tra môn Toán của bạn Huy là:

6 7 7 8 7 (4). Còn của bạn Dũng là 8 6 7 5 9 (3)

Số trung bình cộng của mẫu số liệu (4) là:  $\bar{x} = 7$ .

a) Tính phương sai của mẫu số liệu (4).

b) So sánh phương sai của mẫu số liệu (4) với phương sai của mẫu số liệu (3) Từ đó cho biết bạn nào có kết quả kiểm tra môn Toán đồng đều hơn.

**Giải**

a) Gọi phương sai của hai mẫu số liệu (3) và (4) lần lượt là  $s_D^2, s_H^2$ . Ta có:  $s_D^2 = 2$ ;

$$s_H^2 = \frac{(6-7)^2 + (7-7)^2 + (7-7)^2 + (8-7)^2 + (7-7)^2}{5} = \frac{2}{5} = 0,4.$$

b) Do  $s_H^2 = 0,4 < s_D^2 = 2$  nên bạn Huy có kết quả kiểm tra môn Toán đồng đều hơn bạn Dũng.

## III. Độ lệch chuẩn

### 1. Định nghĩa

Căn bậc hai của phương sai gọi là *độ lệch chuẩn* của mẫu số liệu thống kê.

**Nhận xét:** Vì đơn vị đo của phương sai là bình phương đơn vị đo của số liệu thống kê, trong khi độ lệch chuẩn lại có cùng đơn vị đo với số liệu thống kê, nên khi cần chú ý đến đơn vị đo thì ta sử dụng độ lệch chuẩn.

**Ví dụ 3.** Bảng sau thống kê nhiệt độ (đơn vị:  $^{\circ}C$ ) ở Thành phố Hồ Chí Minh ngày 03/6/2021 sau một số lần đo.

Giờ đo	1h	4h	7h	10h	13h	16h	19h	22h
Nhiệt độ ( $^{\circ}C$ )	27	26	28	32	34	35	30	28

a) Viết mẫu số liệu thống kê nhiệt độ nhận được từ bảng.

b) Tính số trung bình cộng, phương sai và độ lệch chuẩn của mẫu số liệu đó (làm tròn kết quả đến hàng phần trăm).

**Giải**

a) Mẫu số liệu thống kê nhiệt độ nhận được từ bảng là: 27 26 28 32 34 35 30 28

b) Nhiệt độ trung bình là:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8}{8} = \frac{27 + 26 + 28 + 32 + 34 + 35 + 30 + 28}{8} = 30 (^{\circ}C).$$

Phương sai của mẫu số liệu đó là:

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + (x_4 - \bar{x})^2 + (x_5 - \bar{x})^2 + (x_6 - \bar{x})^2 + (x_7 - \bar{x})^2 + (x_8 - \bar{x})^2}{8}$$

$$= \frac{(-3)^2 + (-4)^2 + (-2)^2 + 2^2 + 4^2 + 5^2 + 0^2 + (-2)^2}{8} = \frac{78}{8} = 9,75.$$

Độ lệch chuẩn của mẫu số liệu đó là:  $s = \sqrt{9,75} \approx 3,12 (^{\circ}C)$

## 2. Ý nghĩa

Cũng như phương sai, khi hai mẫu số liệu thống kê có cùng đơn vị đo và có số trung bình cộng bằng nhau (hoặc xấp xỉ nhau), mẫu số liệu nào có độ lệch chuẩn nhỏ hơn thì mức độ phân tán (so với số trung bình cộng) của các số liệu trong mẫu đó sẽ thấp hơn. Độ lệch chuẩn là số đặc trưng đo mức độ phân tán của mẫu số liệu thống kê có cùng đơn vị đo.

## IV. Tính hợp lý của số liệu thống kê

Ta có thể sử dụng các số đặc trưng đo mức độ phân tán cho mẫu số liệu không ghép nhóm để chỉ ra được những số liệu bất thường của mẫu số liệu đó. Ta thường sử dụng khoảng tứ phân vị để xác định số liệu bất thường của mẫu số liệu. Cụ thể như sau:

Giả sử  $Q_1, Q_2, Q_3$  là tứ phân vị của mẫu số liệu và hiệu  $\Delta_Q = Q_3 - Q_1$  là khoảng tứ phân vị của mẫu số liệu

đó. Một giá trị trong mẫu số liệu được coi là một giá trị bất thường nếu nó nhỏ hơn  $Q_1 - \frac{3}{2}\Delta_Q$  hoặc lớn hơn

$Q_3 + \frac{3}{2}\Delta_Q$ . Như vậy, khoảng tứ phân vị cho ta cách nhận ra giá trị bất thường của mẫu số liệu.

**Ví dụ 4.** Nêu các giá trị bất thường của mẫu số liệu (7) thống kê sau:

5      6      19      21      22      23      24      25      26      27      28      29      30      31      32  
       33      34      48      49

**Giải**

Mẫu số liệu (7) có tứ phân vị là  $Q_1 = 22; Q_2 = 27; Q_3 = 32$ .

Suy ra  $\Delta_Q = Q_3 - Q_1 = 32 - 22 = 10$ .

Các giá trị 5,6 (nhỏ hơn  $Q_1 - \frac{3}{2}\Delta_Q = 22 - \frac{3}{2} \cdot 10 = 7$ ) và các giá trị 48,49 (lớn hơn

$Q_3 + \frac{3}{2}\Delta_Q = 32 + \frac{3}{2} \cdot 10 = 47$ ) là các giá trị bất thường của mẫu số liệu (7).

**Chú ý:** Ta cũng có thể xác định số liệu bất thường của mẫu số liệu bằng số trung bình cộng và độ lệch chuẩn. Cụ thể như sau:



- A. kg.                                      B.  $\text{kg}^2$ .                                      C. Không có đơn vị.                                      D.  $\frac{\text{kg}}{2}$ .

**Câu 8.** Tìm phát biểu đúng về phương sai của một mẫu số liệu.

- A. Phương sai được sử dụng làm đại diện cho các số liệu của mẫu.  
 B. Phương sai được sử dụng để đánh giá mức độ phân tán của các số liệu thống kê (so với số trung bình).  
 C. Phương sai được tính bằng tổng số phần tử của một mẫu số liệu.  
 D. Phương sai là số liệu xuất hiện nhiều nhất (số liệu có tần số lớn nhất) trong bảng các số liệu thống kê.

**Câu 9.** Theo kết quả thống kê điểm thi giữa kỳ 2 môn toán khối 11 của một trường THPT, người ta tính được phương sai của bảng thống kê đó là  $s_x^2 = 0,573$ . Độ lệch chuẩn của bảng thống kê đó bằng

- A. 0,812.                                      B. 0,757.                                      C. 0,936.                                      D. 0,657.

**Câu 10.** Cho mẫu số liệu  $x_1, x_2, \dots, x_N$  có số trung bình là  $\bar{x}$ . Phương sai được tính theo công thức nào trong các công thức sau

- A.  $\frac{1}{N} \sum_{i=1}^N x_i$ .                                      B.  $\sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})}$ .                                      C.  $\sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2}$ .                                      D.  $\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$ .

**Câu 11.** Phương sai của dãy số 2;3;4;5;6 là

- A.  $S_x^2 = 4$ .                                      B.  $S_x^2 = \sqrt{2}$ .                                      C.  $S_x^2 = 2$ .                                      D.  $S_x^2 = -2$ .

**Câu 12.** Khoảng tứ phân vị của dãy số 2;3;4;5;6 là

- A.  $\Delta_Q = 3$ .                                      B.  $\Delta_Q = \sqrt{2}$ .                                      C.  $\Delta_Q = 2$ .                                      D.  $\Delta_Q = -2$ .

**Câu 13.** Thống kê điểm kiểm tra môn toán (thang điểm 10) của một nhóm gồm 6 học sinh ta có bảng số liệu sau:

Tên học sinh	Kim	Son	Ninh	Binh	Việt	Nam
Điểm	9	8	7	10	8	9

Tìm độ lệch chuẩn  $\delta$  của bảng số liệu trên (làm tròn đến hàng phần trăm).

- A.  $\delta \approx 0,92$ .                                      B.  $\delta \approx 0,95$ .                                      C.  $\delta \approx 0,96$ .                                      D.  $\delta \approx 0,91$ .

**Câu 14.** Có 100 học sinh tham dự kì thi học sinh giỏi Toán (thang điểm 20). Kết quả cho trong bảng sau:

Điểm (x)	9	10	11	12	13	14	15	16	17	18	19
Tần số	1	1	3	5	8	13	19	24	14	10	2

Khi đó **độ lệch chuẩn** là

- A. 1,98.                                      B. 3,96.                                      C. 15,23.                                      D. 1,99.

**Câu 15.** Điểm thi của lớp 10C của một trường Trung học Phổ Thông được trình bày ở bảng phân bố tần số sau:

Điểm thi	5	6	7	8	9	10	
Tần số	7	5	10	12	4	2	$n = 40$

Phương sai của bảng phân bố tần số đã cho là:

- A. 0,94.                                      B. 3,94.                                      C. 2,94.                                      D. 1,94.

**Câu 16.** Theo dõi thời gian làm một bài toán (tính bằng phút) của 40 học sinh, giáo viên lập được bảng sau:

Thời gian ( $x$ )	4	5	6	7	8	9	10	11	12	
Tần số ( $n$ )	6	3	4	2	7	5	5	7	1	$N = 40$

Phương sai của mẫu số liệu trên gần với số nào nhất?

- A. 6.                                      B. 12.                                      C. 40.                                      D. 9.

**Câu 17.** Cho dãy số liệu thống kê: 1; 2; 3; 4; 5; 6; 7. Phương sai của các số liệu thống kê là

- A. 1.                                      B. 2.                                      C. 3.                                      D. 4.

**Câu 18.** Cho dãy số liệu thống kê: 1; 2; 3; 4; 5; 6; 7. Khoảng biến thiên là

- A. 1.                                      B. 2.                                      C. 3.                                      D. 6.

**Câu 19.** Số liệu thống kê  $^{100}$  học sinh tham gia kì thi học sinh giỏi toán (thang điểm 20). Kết quả được thống kê trong bảng sau:

Điểm	9	10	11	12	13	14	15	16	17	18	19	
Tần số	1	1	3	5	8	13	19	24	14	10	2	$N = 100$

Tính độ lệch chuẩn của bảng số liệu thống kê.

- A. 2,01.                                      B. 1,89.                                      C. 1,98.                                      D. 1,99.

**Câu 20.** Cho mẫu số liệu thống kê {1; 2; 3; 4; 5; 6; 7; 8; 9}. Tính (gần đúng) độ lệch chuẩn của mẫu số liệu trên?

- A. 2,45.                                      B. 2,58.                                      C. 6,67.                                      D. 6,0.

**Câu 21.** Cho mẫu số liệu thống kê {1; 2; 3; 4; 5; 6; 7; 8; 9}. Tìm khoảng tứ phân vị của mẫu số liệu trên?

- A. 2.                                      B. 5.                                      C. 3.                                      D. 4.

**Câu 22.** Một cửa hàng bán gạo, thống kê số  $^{kg}$  gạo mà cửa hàng bán mỗi ngày trong 30 ngày, được bảng tần số:

Bảng tần số	
Số kg gạo	Tần số
100	7
120	4
130	2
160	8
180	3
200	2
250	4
<b>Tổng</b>	<b>30</b>

Phương sai của bảng số liệu gần đúng với giá trị nào dưới đây nhất?

- A. 155.                                      B. 2318.                                      C. 3325.                                      D. 1234.

**Câu 23.** Sản lượng lúa (tạ) của 40 thửa ruộng thí nghiệm có cùng diện tích được trình bày trong bảng phân bố tần số sau đây:

Sản lượng	20	21	22	23	24
Tần số	5	8	11	10	6

Phương sai của mẫu số liệu là:

- A.  $s_x^2 = 1,5$                       B.  $s_x^2 = 1,24$                       C. 1,54                      D. 22,1

**Câu 24.** Điểm kiểm tra giữa kỳ 2 của một học sinh lớp 10 như sau: 2, 4, 6, 8, 10. Phương sai của mẫu số liệu trên là bao nhiêu?

- A. 6                      B. 8                      C. 10                      D. 40

**Câu 25.** Điểm kiểm tra giữa kỳ 2 của một học sinh lớp 10 như sau: 2, 4, 6, 8, 10. Khoảng biến thiên của mẫu số liệu trên là bao nhiêu?

- A. 6                      B. 8                      C. 10                      D. 40

**Câu 26.** Cho thống kê điểm thi môn toán trong một kì thi của 400 em học sinh. Người ta thấy có 72 bài được điểm 5. Hỏi tần suất của giá trị  $x_i = 5$  là bao nhiêu

- A. 72%                      B. 36%                      C. 10%                      D. 18%

**Câu 27.** Cho bảng số liệu điểm thi học kì 2 của 40 học sinh lớp 10A (thang điểm là 10):

Điểm	5	6	7	8	9	10	
Tần số	5	12	8	9	4	2	N=40

Tính phương sai  $S_x^2$

- A.  $S_x^2 = 1,784$                       B.  $S_x^2 = 1,874$                       C.  $S_x^2 = 1,847$                       D.  $S_x^2 = 1,748$

**Câu 28.** Điểm thi môn Toán lớp 10A<sub>2</sub> của một Trường trung học phổ thông được trình bày ở bảng phân bố tần số sau

Điểm thi	5	6	7	8	9	10	
Tần số	7	5	10	12	4	2	$n = 40$

Trong các giá trị dưới đây, giá trị nào gần nhất với phương sai của bảng phân bố tần số trên?

- A. 0,94                      B. 3,94                      C. 2,94                      D. 1,94