

A. Lý thuyết

1. Số gần đúng

Trong nhiều trường hợp, ta không biết hoặc khó biết số đúng (kí hiệu là \bar{a}) mà chỉ tìm được giá trị xấp xỉ nó. Giá trị này được gọi là **số gần đúng**, kí hiệu là a .

Chú ý:

Ta có thể sử dụng máy tính cầm tay để tìm giá trị gần đúng của các biểu thức chứa các số vô tỉ như π , \sqrt{a} , $\sqrt[3]{a}$, ...

2. Sai số tuyệt đối và sai số tương đối

a) Sai số tuyệt đối

Giá trị $\Delta_a = |a - \bar{a}|$ phản ánh mức độ sai lệch giữa số đúng \bar{a} và số gần đúng a , được gọi là **sai số tuyệt đối** của số gần đúng a .

Chú ý:

+ Trên thực tế, nhiều khi ta không biết \bar{a} nên cũng không biết Δ_a . Tuy nhiên, ta có thể đánh giá được Δ_a không vượt quá một số dương d nào đó.

+ Nếu $\Delta_a \leq d$ thì $a - d \leq \bar{a} \leq a + d$, khi đó ta viết $\bar{a} = a \pm d$ và hiểu là số đúng \bar{a} nằm trong đoạn $[a - d; a + d]$. Do đó d càng nhỏ thì a càng gần \bar{a} nên d được gọi là **độ chính xác của số gần đúng**.

+ Trong các phép đo, độ chính xác d của số gần đúng bằng một nửa đơn vị của thước đo. Chẳng hạn, một thước đo có chia vạch đến xentimét thì mọi giá trị đo nằm giữa 6,5 cm và 7,5 cm đều được coi là 7 cm. Vì vậy, thước đo có thang đo càng nhỏ thì cho giá trị đo càng chính xác.

b) Sai số tương đối

Sai số tương đối của số gần đúng a , kí hiệu là δ_a , là tỉ số giữa sai số tuyệt đối và $|a|$, tức là $\delta_a = \frac{\Delta_a}{|a|}$.

Nhận xét:

Nếu $\bar{a} = a \pm d$ thì $\Delta_a \leq d$, do đó $\delta_a \leq \frac{d}{|a|}$. Nếu $\frac{d}{|a|}$ càng nhỏ thì chất lượng của phép đo hay tính toán càng cao. Người ta thường viết sai số tương đối dưới dạng phần trăm.

3. Quy tròn số gần đúng

Số thu được sau khi thực hiện làm tròn số được gọi là **số quy tròn**. Số quy tròn là một số gần đúng của số ban đầu.

Chú ý:

- Đối với chữ số hàng làm tròn:
 - + Giữ nguyên nếu chữ số ngay bên phải nó nhỏ hơn 5;
 - + Tăng 1 đơn vị nếu chữ số ngay bên phải nó lớn hơn hoặc bằng 5.
- Đối với chữ số sau hàng làm tròn:
 - + Bỏ đi nếu ở phần thập phân;
 - + Thay bởi các chữ số 0 nếu ở phần số nguyên.

Nhận xét:

- + Khi thay số đúng bởi số quy tròn đến một hàng nào đó thì sai số tuyệt đối của số quy tròn không vượt quá nửa đơn vị của hàng làm tròn.

+ Cho số gần đúng a với độ chính xác d . Khi được yêu cầu làm tròn số a mà không nói rõ làm tròn đến hàng nào thì ta làm tròn số a đến hàng thấp nhất mà d nhỏ hơn 1 đơn vị của hàng đó.

4. Số trung bình và trung vị

a) Số trung bình

Số trung bình (số trung bình cộng) của mẫu số liệu x_1, x_2, \dots, x_n , kí hiệu là \bar{x} được tính bằng công thức:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

Chú ý:

Trong trường hợp mẫu số liệu cho dưới dạng bảng tần số thì số trung bình được tính theo công thức:

$$\bar{x} = \frac{m_1 x_1 + m_2 x_2 + \dots + m_k x_k}{n}$$

trong đó m_k là tần số của giá trị x_k và $n = m_1 + m_2 + \dots + m_k$.

Ý nghĩa: Số trung bình là giá trị trung bình cộng của các số trong mẫu số liệu, nó cho biết vị trí trung tâm của mẫu số liệu và có thể dùng để đại diện cho mẫu số liệu.

b) Trung vị

Trong trường hợp mẫu số liệu có giá trị bất thường (rất lớn hoặc rất bé so với đa số các giá trị khác), người ta không dùng số trung bình để đo xu thế trung tâm mà dùng trung vị.

Để tìm trung vị (kí hiệu là Me) của một mẫu số liệu, ta thực hiện như sau:

- + Sắp xếp các giá trị trong mẫu số liệu theo thứ tự không giảm.
- + Nếu số giá trị của mẫu số liệu là số lẻ thì giá trị chính giữa của mẫu là trung vị. Nếu là số chẵn thì trung vị là trung bình cộng của hai giá trị chính giữa của mẫu.

Ý nghĩa: Trung vị không bị ảnh hưởng bởi giá trị bất thường trong khi số trung bình bị ảnh hưởng bởi giá trị bất thường. Vì vậy, khi mẫu số liệu có giá trị bất thường, người ta thường dùng trung vị đại diện cho các số liệu thống kê.

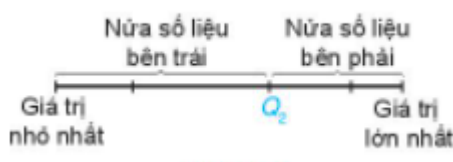
5. Tứ phân vị

Tứ phân vị dùng để xác định ngưỡng để phân loại các số liệu có trong mẫu số liệu.

Để tìm các tứ phân vị của mẫu số liệu có n giá trị, ta làm như sau:

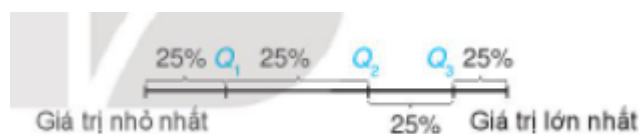
- + Sắp xếp mẫu số liệu theo thứ tự không giảm.
- + Tìm trung vị. Giá trị này là Q_2 .
- + Tìm trung vị của nửa số liệu bên trái Q_2 (không bao gồm Q_2 nếu n lẻ). Giá trị này là Q_1 .
- + Tìm trung vị của nửa số liệu bên phải Q_2 (không bao gồm Q_2 nếu n lẻ). Giá trị này là Q_3 .

Q_1 , Q_2 , Q_3 được gọi là các **tứ phân vị** của mẫu số liệu.



Chú ý: Q_1 được gọi là tứ phân vị thứ nhất hay tứ phân vị dưới, Q_3 được gọi là tứ phân vị thứ ba hay tứ phân vị trên.

Ý nghĩa: Các điểm Q_1 , Q_2 , Q_3 chia mẫu số liệu đã sắp xếp theo thứ tự từ nhỏ đến lớn thành bốn phần, mỗi phần đều chứa 25% giá trị.



6. Một

Mốt của mẫu số liệu là giá trị xuất hiện với tần số lớn nhất.

Ý nghĩa: Có thể dùng một để đo xu thế trung tâm của mẫu số liệu khi mẫu số liệu có nhiều giá trị trùng nhau.

Nhận xét:

+ Mốt có thể không là duy nhất. Chẳng hạn, với mẫu số liệu sau:

6 9 5 9 9 7 5 5 9 5

Ta thấy các số 5; 9 đều xuất hiện với số lần lớn nhất (4 lần) nên mẫu số liệu này có hai mốt là 5 và 9.

+ Khi các giá trị trong mẫu số liệu xuất hiện với tần số như nhau thì mẫu số liệu không có mốt. Chẳng hạn, với mẫu số liệu sau:

6 8 6 7 8 6 7 7 8

Ta thấy các giá trị 6; 7; 8 trong mẫu số liệu đều xuất hiện với tần số như nhau (3 lần) nên mẫu số liệu này không có mốt.

+ Mốt còn được định nghĩa cho mẫu dữ liệu định tính (dữ liệu không phải là số). Ví dụ trong buổi biểu quyết chọn một trong ba bạn Hoa, Bình, Tú làm bí thư của lớp 10C, bạn thư ký của lớp đã tổng kết được kết quả biểu quyết như sau:

Tên bạn được chọn	Hoa	Bình	Tú
Số phiếu	18	12	10

Trong mẫu dữ liệu này, số phiếu chọn “bạn Hoa” nhiều nhất, được gọi là mốt.

7. Khoảng biến thiên và khoảng tứ phân vị

a) Khoảng biến thiên

Khoảng biến thiên, kí hiệu là R , là hiệu số giữa giá trị lớn nhất và giá trị nhỏ nhất trong mẫu số liệu.

Ý nghĩa: Khoảng biến thiên dùng để đo độ phân tán của mẫu số liệu. Khoảng biến thiên càng lớn thì mẫu số liệu càng phân tán.

Nhận xét: Sử dụng khoảng biến thiên có ưu điểm là đơn giản, dễ tính toán song khoảng biến thiên chỉ sử dụng thông tin của giá trị lớn nhất và giá trị nhỏ nhất mà bỏ qua thông tin từ tất cả các giá trị khác. Do đó, khoảng biến thiên rất dễ bị ảnh hưởng bởi các giá trị bất thường.

b) Khoảng tứ phân vị

Khoảng tứ phân vị, kí hiệu là Δ_Q , là hiệu số giữa tứ phân vị thứ ba và tứ phân vị thứ nhất, tức là:

$$\Delta_Q = Q_3 - Q_1.$$

Về bản chất, khoảng tứ phân vị là khoảng biến thiên của 50% số liệu chính giữa của mẫu số liệu đã sắp xếp.

Ý nghĩa: Khoảng tứ phân vị cũng là một số đo độ phân tán của mẫu số liệu. Khoảng tứ phân vị càng lớn thì mẫu số liệu càng phân tán.

Chú ý: Một số tài liệu gọi khoảng biến thiên là *biên độ* và khoảng tứ phân vị là *độ trải giữa*.

8. Phương sai và độ lệch chuẩn

Khoảng biến thiên chỉ sử dụng thông tin của giá trị lớn nhất và giá trị nhỏ nhất của mẫu số liệu (bỏ qua thông tin của tất cả các giá trị khác). Khoảng tứ phân vị chỉ sử dụng thông tin của 50% số liệu chính giữa. Có một vài số đặc trưng khác đo độ phân tán sử dụng thông tin của tất cả các giá trị trong mẫu số liệu. Hai trong số đó là *phương sai* và *độ lệch chuẩn*.

Cụ thể với mẫu số liệu x_1, x_2, \dots, x_n , nếu gọi số trung bình là \bar{x} thì với mỗi giá trị x_i , độ lệch của nó so với giá trị trung bình là $x_i - \bar{x}$.

- **Phương sai** là giá trị $s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}$.

- Căn bậc hai của phương sai, $s = \sqrt{s^2}$, được gọi là **độ lệch chuẩn**.

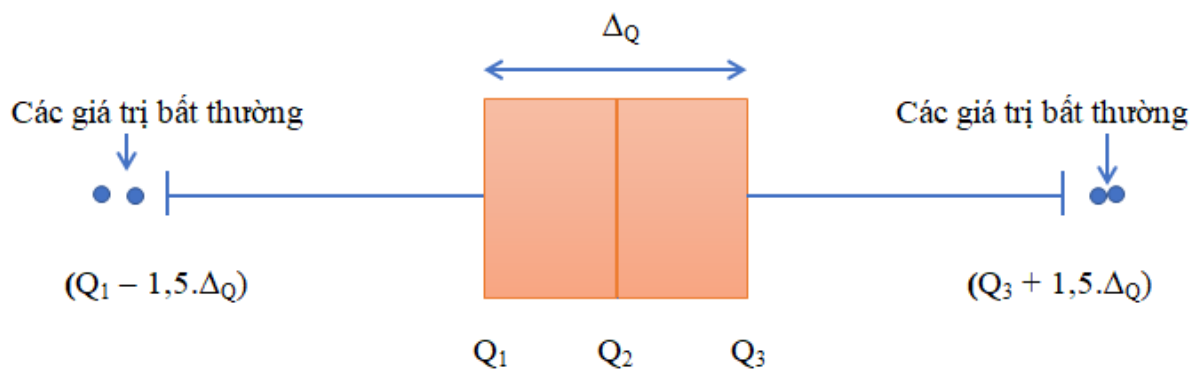
Chú ý: Người ta còn sử dụng đại lượng để đo độ phân tán của mẫu số liệu:

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

Ý nghĩa: Nếu số liệu càng phân tán thì phương sai và độ lệch chuẩn càng lớn.

9. Phát hiện số liệu bất thường hoặc không chính xác bằng biểu đồ hộp

Trong mẫu số liệu thống kê, có khi ta sẽ gặp phải những giá trị quá lớn hoặc quá nhỏ so với đa số các giá trị khác. Những giá trị này được gọi là **giá trị bất thường**. Chúng xuất hiện trong mẫu số liệu có thể do nhầm lẫn hay sai sót nào đó. Ta có thể dùng biểu đồ hộp để phát hiện những giá trị bất thường này.



Các giá trị lớn hơn $Q_3 + 1,5 \cdot \Delta_Q$ hoặc bé hơn $Q_1 - 1,5 \cdot \Delta_Q$ được xem là **giá trị bất thường**.

B. Bài tập tự luyện

B1. Bài tập trắc nghiệm

Câu 1. Doanh thu của một cửa hàng tạp hoá trong 5 ngày được cho bởi số liệu: 2,3; 2,5; 3,1; 2,0; 2,3 (đơn vị: triệu đồng). Khoảng biến thiên của mẫu số liệu là

A. 2,3;

B. 1,1;

C. 2,0;

D. 3,1.

Hướng dẫn giải

Đáp án đúng là: B

Ta có khoảng biến thiên là hiệu số giữa giá trị lớn nhất bằng 3,1 và giá trị nhỏ nhất bằng 2,0 của dãy số liệu.

Vậy khoảng biến thiên bằng $R = 3,1 - 2,0 = 1,1$.

Câu 2. Sản phẩm bình quân trong một giờ của công nhân trong 10 ngày liên tiếp của công ty A được thống kê bởi dãy số liệu: 30; 40; 32; 40; 50; 45; 42; 42; 45; 50. Tìm tứ phân vị của mẫu số liệu

A. $Q_1 = 30$; $Q_2 = 42$; $Q_3 = 45$;

B. $Q_1 = 40$; $Q_2 = 45$; $Q_3 = 50$;

C. $Q_1 = 30$; $Q_2 = 42$; $Q_3 = 50$;

D. $Q_1 = 40$; $Q_2 = 42$; $Q_3 = 45$.

Hướng dẫn giải

Đáp án đúng là: D

Ta sắp xếp mẫu số liệu theo thứ tự không giảm: 30; 32; 40; 40; 42; 42; 45; 45; 50; 50.

Vì $n = 10$ là số chẵn nên Q_2 là trung bình cộng của hai số chính giữa:

$$Q_2 = (42 + 42) : 2 = 42$$

Ta tìm Q_1 là trung vị nửa số liệu bên trái Q_2 : 30; 32; 40; 40; 42 gồm 5 giá trị, và ta tìm được $Q_1 = 40$.

Ta tìm Q_3 là trung vị nửa số liệu bên phải Q_2 : 42; 45; 45; 50; 50 gồm 5 giá trị và ta tìm được $Q_3 = 45$.

Vậy tứ phân vị $Q_1 = 40$; $Q_2 = 42$; $Q_3 = 45$.

Câu 3. Số đo áo của 20 học sinh lớp 10A được thống kê như sau: 8; 9; 10; 8; 7; 9; 8; 10; 9; 9; 8; 10; 7; 9; 8; 10; 9; 8; 9; 7. Tìm một của mẫu số liệu này

A. 10;

B. 8;

C. 9;

D. 7.

Hướng dẫn giải

Đáp án đúng là: C

Ta có bảng sau:

Số đo áo	7	8	9	10
Số học sinh	3	6	7	4

Dựa vào bảng trên ta thấy số áo học sinh mặc nhiều nhất là áo số 9 (7 học sinh) nên một bằng 9.

Vậy $M_0 = 9$.

Câu 4. Một cửa hàng dép da đã thống kê cỡ dép của một số khách hàng nam cho kết quả như sau: 39; 38; 39; 40; 41; 41; 43; 37; 38; 40; 43; 41; 42; 41; 42. Tìm trung vị của mẫu số liệu trên

A. 37;

B. 39;

C. 41;

D. 43.

Hướng dẫn giải

Đáp án đúng là: C

Ta sắp xếp số liệu theo thứ tự không giảm: 37; 38; 38; 39; 39; 40; 40; 41; 41; 41; 41; 42; 42; 43; 43.

Vì $n = 15$ là số lẻ nên số trung vị là số chính giữa của dãy số liệu. Vậy trung vị $Q_2 = 41$.

Câu 5. Năng xuất lúa của 4 xã được thống kê bởi mẫu số liệu: 36; 38; 34; 40 (đơn vị: tạ/ha). Độ lệch chuẩn của mẫu số liệu

A. 1,23;

B. 2,03;

C. 2,21;

D. 2,24.

Hướng dẫn giải

Đáp án đúng là: D

Giá trị trung bình của mẫu số liệu là $\bar{x} = \frac{36 + 38 + 40 + 34}{4} = 37$

Ta có bảng sau

Giá trị	Độ lệch	Bình phương độ lệch
36	$36 - 37 = -1$	1
38	$38 - 37 = 1$	1
34	$34 - 37 = -3$	9

40	$40 - 37 = 3$	9
Tổng		20

Vì có 4 giá trị nên $n = 4$. Do đó $s^2 = \frac{20}{4} = 5$

Do đó $s = \sqrt{5} = 2,24$.

Câu 6. Số đo chiều cao (đơn vị cm) của học sinh trong tổ 1 lớp 10A cho kết quả như sau: 156; 159; 162; 165; 163; 159; 155; 160. Chiều cao trung bình của học sinh tổ 1 là:

- A. 159;
- B. 162,425;
- C. 159, 875;
- D. 160.

Hướng dẫn giải

Đáp án đúng là: C

Ta có $\bar{x} = \frac{156 + 159 + 162 + 165 + 163 + 159 + 155 + 160}{8} = 159,875$.

Vậy chiều cao trung bình của học sinh tổ 1 là 159,875 cm.

Câu 7. Cho mẫu số liệu 5; 6; 7; 8; 9. Phương sai của mẫu số liệu trên là

- A. 2;
- B. 5;
- C. 7;
- D. 6.

Hướng dẫn giải

Đáp án đúng là:

Số trung bình của mẫu số liệu là: $\bar{x} = \frac{5 + 6 + 7 + 8 + 9}{5} = 7$

Ta có bảng sau:

Giá trị	Độ lệch	Bình phương độ lệch
5	$5 - 7 = -2$	4

6	$6 - 7 = -1$	1
7	$7 - 7 = 0$	0
8	$8 - 7 = 1$	1
9	$9 - 7 = 2$	4
Tổng		10

Mẫu số liệu có 5 giá trị nên $n = 5$. Do đó phương sai là: $s^2 = \frac{10}{5} = 2$.

Câu 8. Số học sinh trong 4 tổ của lớp 10A là 9; 10; 8; 9. Độ lệch chuẩn của mẫu số liệu là

- A. 0,75;
- B. 0,1825;
- C. 0,1475;
- D. 0,433.

Hướng dẫn giải

Đáp án đúng là: D

Số trung bình của mẫu số liệu là: $\bar{x} = \frac{9+10+8+9}{4} = 9$

Ta có bảng sau

Giá trị	Độ lệch	Bình phương độ lệch
9	$9 - 9 = 0$	0
10	$10 - 9 = 1$	1
8	$8 - 9 = -1$	1
9	$9 - 9 = 0$	0
Tổng		2

Mẫu số liệu có 4 giá trị nên $n = 4$. Do đó phương sai là: $s^2 = \frac{2}{4} = 0,5$.

Độ lệch chuẩn là: $s = \sqrt{0,5} \approx 0,71$.

Câu 9. Tiền thưởng cho 5 công nhân trong một công ty được thống kê bởi mẫu số liệu: 3; 3; 5; 5; 4 (đơn vị: triệu đồng). Phương sai của mẫu số liệu.

A. 0,6;

B. 0,7;

C. 0,8;

D. 0,8.

Hướng dẫn giải

Đáp án đúng là: C

Giá trị trung bình của mẫu số liệu là $\bar{x} = \frac{3+3+5+5+4}{5} = 4$

Ta có bảng sau

Giá trị	Độ lệch	Bình phương độ lệch
3	$3 - 4 = -1$	1
3	$3 - 4 = -1$	1
5	$5 - 4 = 1$	1
5	$5 - 4 = 1$	1
4	$4 - 4 = 0$	0
Tổng		4

Vì có 5 giá trị nên $n = 5$. Do đó $s^2 = \frac{4}{5} = 0,8$.

Câu 10. Mẫu số liệu sau đây cho biết sĩ số của 12 lớp ở một trường trung học như sau: 45; 43; 46; 41; 40; 40; 42; 41; 45; 45; 43; 42. Khoảng tứ phân vị của mẫu số liệu là

A. 2,5;

B. 4;

C. 1,5;

D. 5.

Hướng dẫn giải

Đáp án đúng là: B

Ta sắp xếp mẫu số liệu theo thứ tự không giảm: 40; 40; 41; 41; 42; 42; 43; 43; 45; 45; 45; 46.

Vì $n = 12$ là số chẵn nên Q_2 là trung bình cộng của hai số chính giữa:

$$Q_2 = (42 + 43) : 2 = 42,5$$

Ta tìm Q_1 là trung vị nửa số liệu bên trái Q_2 : 40; 40; 41; 41; 42; 42 gồm 6 giá trị, hai phần tử chính giữa là 41, 41 do đó $Q_1 = (41 + 41) : 2 = 41$.

Ta tìm Q_3 là trung vị nửa số liệu bên phải Q_2 : 43; 43; 45; 45; 45; 46 gồm 6 giá trị, hai phần tử chính giữa là 45, 45 do đó $Q_3 = (45 + 45) : 2 = 45$.

Vậy khoảng tứ phân vị $\Delta_Q = 45 - 41 = 4$.

B2. Bài tập tự luận

Câu 1. Giải thích kết quả “Đo độ sâu của mực nước biển ở một vùng cho kết quả là 2136 ± 5 m” và thực hiện làm tròn số gần đúng.

Hướng dẫn giải

Độ sâu thực tế của mực nước biển \bar{a} là số đúng. Tuy không biết \bar{a} nhưng ta biết kết quả đo đạc là 2136 m nên 2136 là số gần đúng cho \bar{a} . Độ chính xác $d = 5$ (m).

Vì độ chính xác đến hàng đơn vị ($d = 5$) nên ta làm tròn đến hàng chục theo quy tắc làm tròn. Do đó số quy tròn của a là 2140.

Câu 2. Các nhà vật lý sử dụng ba phương pháp đo hằng số Hubble lần lượt cho kết quả như sau:

$$67,31 \pm 0,96;$$

$$67,90 \pm 0,55;$$

$$67,74 \pm 0,46.$$

Phương pháp nào chính xác nhất tính theo sai số tương đối?

Hướng dẫn giải

Với phương pháp đo thứ nhất: $a_1 = 67,31$ và $d_1 = 0,96$, do đó sai số tương đối là:

$$\delta_1 \leq \frac{d_1}{|a_1|} = \frac{0,96}{67,31} \approx 1,4\%$$

Với phương pháp đo thứ hai: $a_2 = 67,90$ và $d_2 = 0,55$, do đó sai số tương đối là:

$$\delta_2 \leq \frac{d_2}{|a_2|} = \frac{0,55}{67,90} \approx 0,81\%$$

Với phương pháp đo thứ ba: $a_3 = 67,74$ và $d_3 = 0,46$, do đó sai số tương đối là:

$$\delta_3 \leq \frac{d_3}{|a_3|} = \frac{0,46}{67,74} \approx 0,68\%$$

Vì $0,68\% < 0,81\% < 1,4\%$ nên $\delta_3 < \delta_2 < \delta_1$.

Do đó phương pháp đo thứ ba là chính xác nhất tính theo sai số tương đối.

Câu 3. Làm tròn số 4372,8 đến hàng chục và 8,125 đến hàng phần trăm rồi tính sai số tuyệt đối của số quy tròn.

Hướng dẫn giải

+ Số quy tròn của số 4372,8 đến hàng chục là 4370. Sai số tuyệt đối là

$$\Delta = |4370 - 4372,8| = 2,8.$$

+ Số quy tròn của số 8,125 đến hàng phần trăm là 8,13. Sai số tuyệt đối là

$$\Delta' = |8,13 - 8,125| = 0,005.$$

Câu 4. Tìm số trung bình, trung vị, một và tứ phân vị của mỗi mẫu số liệu sau đây:

a) Tuổi thọ (tính theo năm) của 6 loại bóng đèn được ghi lại như sau:

$$19 \quad 20 \quad 31 \quad 18 \quad 26 \quad 24$$

b) Số cân nặng (tính theo kg) của 9 học sinh lớp 10D được ghi lại như sau:

$$37 \quad 37 \quad 40 \quad 42 \quad 45 \quad 40 \quad 42 \quad 42 \quad 55$$

Hướng dẫn giải

a) Tuổi thọ trung bình của 6 loại bóng đèn là: $\frac{19+20+31+18+26+24}{6} = 23$

(năm)

+ Ta tìm trung vị của mẫu số liệu:

Sắp xếp các giá trị đã cho theo thứ tự không giảm:

18 19 20 24 26 31

Hai giá trị chính giữa

Dãy trên có hai giá trị chính giữa lần lượt là 20 và 24.

Vậy trung vị của mẫu số liệu là: $\frac{20+24}{2} = 22$.

+ Vì các giá trị 18; 19; 20; 24; 26; 31 trong mẫu số liệu đều xuất hiện với tần số như nhau (1 lần) nên mẫu số liệu này không có mốt.

+ Ta tìm tứ phân vị của mẫu số liệu:

Giá trị Q_2 chính là trung vị của mẫu số liệu. Do đó ta có $Q_2 = 22$.

Ta tìm Q_1 là trung vị của nửa số liệu bên trái Q_2 :

18 19 20
 ↓
Giá trị chính giữa

Giá trị chính giữa lúc này là 19. Ta suy ra $Q_1 = 19$.

Ta tìm Q_3 là trung vị của nửa số liệu bên phải Q_2 :

24 26 31
 ↓
Giá trị chính giữa

Giá trị chính giữa lúc này là 26. Ta suy ra $Q_3 = 26$.

Vậy các tứ phân vị của mẫu số liệu là $Q_1 = 19$; $Q_2 = 22$; $Q_3 = 26$.

b) Số cân nặng trung bình của 9 học sinh lớp 10D là:

$$\frac{37.2 + 40.2 + 42.3 + 45 + 55}{6} \approx 63,33 \text{ (kg)}$$

+ Ta tìm trung vị của mẫu số liệu:

Sắp xếp các giá trị đã cho theo thứ tự không giảm:

37 37 40 40 42 42 42 45 55
↓
Giá trị chính giữa

Trung vị của mẫu số liệu là: 42.

+ Ta thấy giá trị 42 xuất hiện với tần số cao nhất (3 lần) nên một của mẫu số liệu này là 42.

+ Ta tìm tứ phân vị của mẫu số liệu:

Giá trị Q_2 chính là trung vị của mẫu số liệu. Do đó ta có $Q_2 = 42$.

Ta tìm Q_1 là trung vị của nửa số liệu bên trái Q_2 :

37 37 40 40

Hai giá trị chính giữa

Dãy trên có hai giá trị chính giữa là 37 và 40. Do đó $Q_1 = (37 + 40) : 2 = 38,5$.

Ta tìm Q_3 là trung vị của nửa số liệu bên phải Q_2 :

42 42 45 55

Hai giá trị chính giữa

Dãy trên có hai giá trị chính giữa là 42 và 45. Do đó $Q_3 = (42 + 45) : 2 = 43,5$.

Vậy các tứ phân vị của mẫu số liệu là $Q_1 = 38,5$; $Q_2 = 42$; $Q_3 = 43,5$.

Câu 5. Số lượng học sinh giỏi Quốc gia năm học 2019 – 2020 của 10 trường Trung học phổ thông được cho như sau:

0 0 9 0 0 4 0 0 0 6

a) Tìm số trung bình, một và các tứ phân vị của mẫu số liệu trên.

b) Giải thích tại sao tứ phân vị thứ nhất và trung vị trùng nhau.

Hướng dẫn giải

a) Số lượng học sinh trung bình là: $\frac{0.7 + 9 + 4 + 6}{10} = 1,9$ (học sinh)

+ Ta tìm trung vị của mẫu số liệu:

Sắp xếp các giá trị đã cho theo thứ tự không giảm:

0 0 0 0 0 0 0 4 6 9

Hai giá trị chính giữa

Dãy trên có hai giá trị chính giữa cùng bằng 0.

Vậy trung vị của mẫu số liệu cũng bằng 0.

+ Ta thấy giá trị 0 xuất hiện với tần số cao nhất (7 lần) nên một của mẫu số liệu này là 0.

+ Ta tìm tứ phân vị của mẫu số liệu:

Giá trị Q_2 chính là trung vị của mẫu số liệu. Do đó ta có $Q_2 = 0$.

Ta tìm Q_1 là trung vị của nửa số liệu bên trái Q_2 :

0 0 0 0 0
 ↓

Giá trị chính giữa

Giá trị chính giữa lúc này là 0. Ta suy ra $Q_1 = 0$.

Ta tìm Q_3 là trung vị của nửa số liệu bên phải Q_2 :

0 0 4 6 9
 ↓
Giá trị chính giữa

Giá trị chính giữa lúc này là 4. Ta suy ra $Q_3 = 4$.

Vậy các tứ phân vị của mẫu số liệu là $Q_1 = 0$; $Q_2 = 0$; $Q_3 = 4$.

b) Vì mẫu số liệu tập trung với mật độ cao ở bên trái Q_2 , tức là trung vị, nên tứ phân vị thứ nhất Q_1 và trung vị Q_2 trùng nhau.

Câu 6. Sức chứa của một số sân vận động tại Việt Nam đã được thống kê và ghi lại như trong bảng sau (số liệu gần đúng):

Sân vận động	Mỹ Đình	Cần Thơ	Thiên Trường	Việt Trì
Sức chỗ ngồi	40 000	50 000	30 000	20 000

Các giá trị số trung bình, trung vị và một bị ảnh hưởng như thế nào nếu bỏ đi số liệu sức chứa của sân vận động Cần Thơ?

Hướng dẫn giải

+ Trước khi bỏ đi số liệu sức chứa của sân vận động Cần Thơ, ta có:

Sức chỗ ngồi trung bình là:

$$\frac{40\,000 + 50\,000 + 30\,000 + 20\,000}{4} = 35\,000.$$

Sắp xếp các giá trị đã cho theo thứ tự không giảm:

$$20\ 000 \quad \underbrace{30\ 000 \quad 40\ 000} \quad 50\ 000$$

Hai giá trị chính giữa

Dãy trên có hai giá trị chính giữa lần lượt là 30 000 và 40 000.

Do đó trung vị của mẫu số liệu là: $\frac{30\ 000 + 40\ 000}{2} = 35\ 000$.

Trong mẫu dữ liệu trên, sức chứa ở sân vận động Cần Thơ là lớn nhất.

Do đó mốt của mẫu dữ liệu trên là “Cần Thơ”.

+ Sau khi bỏ đi số liệu sức chứa của sân vận động Cần Thơ, ta có:

Sức chỗ ngồi trung bình là:

$$\frac{40\ 000 + 30\ 000 + 20\ 000}{3} = 30\ 000.$$

Sắp xếp các giá trị đã cho theo thứ tự không giảm:

$$20\ 000 \quad 30\ 000 \quad 40\ 000$$

↓

Giá trị chính giữa

Dãy trên có giá trị chính giữa lần lượt là 30 000.

Do đó trung vị của mẫu số liệu là: 30 000.

Trong mẫu dữ liệu trên, sức chứa ở sân vận động Mỹ Đình là lớn nhất.

Do đó mốt của mẫu dữ liệu trên là “Mỹ Đình”.

+ Nếu bỏ đi số liệu sức chứa của sân vận động Cần Thơ, ta thấy:

Giá trị số trung bình và trung vị đều giảm 5000 (giảm từ 35 000 xuống còn 30 000).

Mốt của dữ liệu cũng bị thay đổi từ “Cần Thơ” sang “Mỹ Đình”.

Câu 7. Mỗi khẳng định sau đúng hay sai?

- a) Nếu các giá trị của mẫu số liệu càng tập trung quanh giá trị trung bình thì độ lệch chuẩn càng lớn;
- b) Khoảng biến thiên chỉ sử dụng thông tin của giá trị lớn nhất và bé nhất, bỏ qua thông tin của các giá trị còn lại;
- c) Khoảng tứ phân vị có sử dụng thông tin của các giá trị lớn nhất và giá trị bé nhất;
- d) Các số đo độ phân tán đều không âm.

Hướng dẫn giải

a) Sai.

Nếu các giá trị của mẫu số liệu đều tập trung quanh giá trị trung bình thì độ lệch của mỗi giá trị $x_i - \bar{x}$ càng nhỏ.

Ta suy ra $(x_i - \bar{x})^2$ cũng càng nhỏ.

Do đó tổng các bình phương độ lệch cũng càng nhỏ.

Mà độ lệch chuẩn tỉ lệ thuận với tổng các bình phương độ lệch.

Vậy trong trường hợp này, độ lệch chuẩn càng nhỏ.

b) Đúng.

Khoảng biến thiên chỉ sử dụng thông tin của giá trị lớn nhất và giá trị nhỏ nhất mà bỏ qua thông tin từ tất cả các giá trị khác. Do đó, khoảng biến thiên rất dễ bị ảnh hưởng bởi các giá trị bất thường.

c) Sai.

Khoảng tứ phân vị chỉ sử dụng 50% số liệu chính giữa của mẫu số liệu đã sắp xếp. Vậy khoảng biến thiên không sử dụng giá trị lớn nhất và giá trị bé nhất.

d) Đúng.

Khoảng biến thiên là hiệu của giá trị lớn nhất và giá trị bé nhất. Do đó khoảng biến thiên luôn không âm.

Khoảng tứ phân vị là hiệu của tứ phân vị thứ ba và tứ phân vị thứ nhất. Do đó khoảng tứ phân vị luôn không âm.

Phương sai là tổng các bình phương các độ lệch nên luôn không âm.

Độ lệch chuẩn là căn bậc hai của phương sai nên độ lệch chuẩn luôn không âm.

Câu 8. Mẫu số liệu sau đây cho biết cân nặng của 10 trẻ sơ sinh (đơn vị kg):

2,8 2,5 3,1 3,3 2,9 3,5 4,1 2,8 3,0 3,3

Hãy tính khoảng biến thiên, khoảng tứ phân vị và độ lệch chuẩn cho mẫu số liệu trên.

Hướng dẫn giải

+ Khoảng biến thiên:

Số cân nặng cao nhất, thấp nhất lần lượt là 4,1 (kg) và 2,5 (kg). Do đó khoảng biến thiên là $R = 4,1 - 2,5 = 1,6$ (kg).

+ Khoảng tứ phân vị:

Trước hết, ta sắp xếp mẫu số liệu theo thứ tự không giảm:

2,5 2,8 2,8 2,9 3,0 3,1 3,3 3,3 3,5 4,1

Mẫu số liệu trên gồm 10 giá trị, hai phần tử chính giữa là 3,0 và 3,1.

Do đó trung vị là $Q_2 = (3,0 + 3,1) : 2 = 3,05$.

Nửa số liệu bên trái là 2,5; 2,8; 2,8; 2,9; 3,0 gồm 5 giá trị. Do đó trung vị là số ở vị trí chính giữa $Q_1 = 2,8$.

Nửa số liệu bên phải là 3,1; 3,3; 3,3; 3,5; 4,1 gồm 5 giá trị, Do đó trung vị là số ở vị trí chính giữa $Q_3 = 3,3$.

Vậy khoảng tứ phân vị cho mẫu số liệu là $\Delta_Q = Q_3 - Q_1 = 3,3 - 2,8 = 0,5$.

+ Độ lệch chuẩn:

Số trung bình của mẫu số liệu là:

$$\bar{x} = \frac{2,5 + 2,8.2 + 2,9 + 3,0 + 3,1 + 3,3.2 + 3,5 + 4,1}{10} = 3,13$$

Ta có bảng sau:

Giá trị	Độ lệch	Bình phương độ lệch
2,8	$2,8 - 3,13 = -0,33$	0,1089
2,5	$2,5 - 3,13 = -0,63$	0,3969
3,1	$3,1 - 3,13 = -0,03$	0,0009
3,3	$3,3 - 3,13 = 0,17$	0,0289
2,9	$2,9 - 3,13 = -0,23$	0,0529
3,5	$3,5 - 3,13 = 0,37$	0,1369
4,1	$4,1 - 3,13 = 0,97$	0,9409
2,8	$2,8 - 3,13 = -0,33$	0,1089
3,0	$3,0 - 3,13 = -0,13$	0,0169
3,3	$3,3 - 3,13 = 0,17$	0,0289
Tổng		1,821

Mẫu số liệu gồm 10 giá trị nên $n = 10$. Do đó phương sai là: $s^2 = \frac{1,821}{10} = 0,1821$.

Độ lệch chuẩn là: $s = \sqrt{0,1821} \approx 0,4267$.

Câu 9. Tỷ lệ thất nghiệp từ năm 2006 – 2020 tại Việt Nam (đơn vị %) được cho như sau:

0,00	2,03	0,00	1,74	1,11	1,00	1,03
1,32	1,26	1,85	1,85	1,87	1,16	2,04
2,39						

Hãy tìm các giá trị bất thường (nếu có) của mẫu số liệu trên.

Hướng dẫn giải

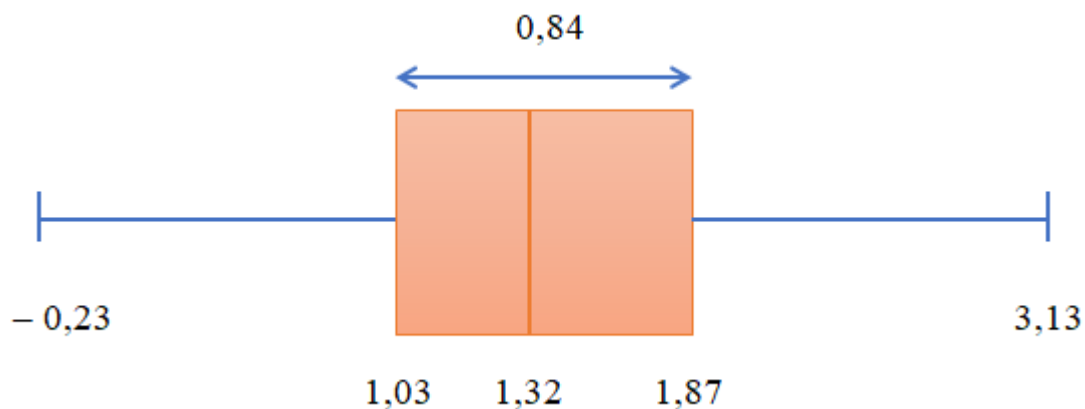
Sắp xếp các số liệu theo thứ tự không giảm ta có:

0,00; 0,00; 1,00; 1,03; 1,11; 1,16; 1,26; 1,32; 1,74; 1,85; 1,85; 1,87; 2,03; 2,04; 2,39.

Từ mẫu số liệu trên, ta tính được $Q_2 = 1,32$; $Q_1 = 1,03$ và $Q_3 = 1,87$. Do đó khoảng tứ phân vị là:

$$\Delta_Q = Q_3 - Q_1 = 1,87 - 1,03 = 0,84.$$

Biểu đồ hộp cho mẫu số liệu này là:



Ta có $Q_1 - 1,5.\Delta_Q = -0,23$ và $Q_3 + 1,5.\Delta_Q = 3,13$ nên trong mẫu số liệu trên không có giá trị nào được xem là bất thường.