

Ôn tập chương VI

A. Lý thuyết

1. Số gần đúng

Trong thực tế cuộc sống cũng như trong khoa học kỹ thuật, có nhiều đại lượng mà ta không thể xác định được giá trị chính xác. Mỗi dụng cụ hay phương pháp đo khác nhau có thể sẽ cho ra các kết quả khác nhau. Vì vậy kết quả thu được thường chỉ là những **số gần đúng**.

2. Sai số tuyệt đối và sai số tương đối

2.1. Sai số tuyệt đối

Nếu a là số gần đúng của số đúng \bar{a} thì $\Delta_a = |\bar{a} - a|$ được gọi là **sai số tuyệt đối** của số gần đúng a .

* Độ chính xác:

Trên thực tế ta thường không biết số đúng \bar{a} nên không thể tính được chính xác Δ_a . Khi đó, ta thường tìm cách khống chế sai số tuyệt đối Δ_a không vượt quá mức $d > 0$ cho trước:

$$\Delta_a = |\bar{a} - a| \leq d \text{ hay } a - d \leq \bar{a} \leq a + d.$$

Khi đó, ta nói a là số gần đúng của số đúng \bar{a} với **độ chính xác** d .

Quy ước viết gọn: $\bar{a} = a \pm d$.

2.2. 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 Δ_a và $|a|$, tức

$$\text{là } \delta_a = \frac{\Delta_a}{|a|}.$$

Nếu $\bar{a} = a \pm d$ thì $\Delta_a \leq d$. Do đó $\delta_a \leq \frac{d}{|a|}$. Nếu δ_a hay $\frac{d}{|a|}$ càng nhỏ thì chất lượng của phép đo đạc hay tính toán càng cao.

Chú ý: Người ta thường viết sai số tương đối dưới dạng phần trăm.

3. Số quy tròn

3.1. Quy tắc làm tròn số

Quy tắc làm tròn số đến một hàng nào đó (gọi là hàng quy tròn):

- + Nếu chữ số sau hàng quy tròn nhỏ hơn 5 thì ta thay nó và các chữ số bên phải nó bởi chữ số 0.
- + Nếu chữ số sau hàng quy tròn lớn hơn hoặc bằng 5 thì ta cũng làm như trên nhưng cộng thêm 1 đơn vị vào chữ số hàng quy tròn.

Chú ý:

- + 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 quy tròn. Ta có thể nói độ chính xác của số quy tròn bằng nửa đơn vị của hàng quy tròn.
- + Khi quy tròn số đúng \bar{a} đến một hàng nào đó thì ta nói số gần đúng a nhận được là chính xác đến hàng đó. Ví dụ số gần đúng của π chính xác đến hàng phần trăm là 3,14.

3.2. Xác định số quy tròn của số gần đúng với độ chính xác cho trước

Các bước xác định **số quy tròn** của số gần đúng a với độ chính xác d cho trước:

Bước 1: Tìm hàng của chữ số khác 0 đầu tiên bên trái của d .

Bước 2: Quy tròn số a ở hàng gấp 10 lần hàng tìm được ở Bước 1.

3.3. Xác định số gần đúng của một số với độ chính xác cho trước

Để tìm số gần đúng a của số đúng \bar{a} với độ chính xác d , ta thực hiện các bước sau:

Bước 1: Tìm hàng của chữ số khác 0 đầu tiên bên trái của d .

Bước 2: Quy tròn \bar{a} đến hàng tìm được ở trên.

4. Bảng số liệu

Dựa vào các thông tin đã biết và sử dụng mối liên hệ toán học giữa các số liệu, ta có thể phát hiện ra được số liệu *không chính xác* trong một số trường hợp.

5. Biểu đồ

Ta có thể biểu diễn số liệu thống kê dưới dạng biểu đồ.

Một số dạng biểu đồ thường gặp: biểu đồ cột, biểu đồ cột kép, biểu đồ quạt, biểu đồ tranh,...

Quan sát các biểu đồ ta có thể đưa ra các nhận xét về số liệu thống kê.

6. Số trung bình

6.1. Công thức tính số trung bình

- Giả sử ta có một mẫu số liệu là x_1, x_2, \dots, x_n .

Số trung bình (hay **số trung bình cộng**) của mẫu số liệu này, kí hiệu là \bar{x} , được tính bởi công thức

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

- Giả sử mẫu số liệu được cho dưới dạng bảng tần số

Giá trị	x_1	x_2	...	x_k
---------	-------	-------	-----	-------

Tần số	n_1	n_2	\dots	n_k
---------------	-------	-------	---------	-------

Khi đó, công thức tính số trung bình trở thành

$$\bar{x} = \frac{n_1 x_1 + n_2 x_2 + \dots + n_k x_k}{n}.$$

Trong đó $n = n_1 + n_2 + \dots + n_k$. Ta gọi n là **cỡ mẫu**.

Chú ý: Nếu kí hiệu $f_k = \frac{n_k}{n}$ là tần số tương đối (hay còn gọi là tần suất) của x_k trong mẫu

số liệu thì số trung bình còn có thể biểu diễn là: $\bar{x} = f_1 x_1 + f_2 x_2 + \dots + f_k x_k$.

6.2. Ý nghĩa của số trung bình

Số trung bình của mẫu số liệu được dùng làm đại diện cho các số liệu của mẫu. Nó là một số đo xu thế trung tâm của mẫu đó.

7. Trung vị và tứ phân vị

7.1. Trung vị

7.1.1 Định nghĩa và cách tính số trung vị

Khi các số liệu trong mẫu số liệu chênh lệch nhau quá lớn, ta dùng một đặc trưng khác của mẫu số liệu, gọi là **trung vị** để so sánh các mẫu số liệu với nhau.

Trung vị được định nghĩa như sau:

Sắp xếp mẫu số liệu theo thứ tự không giảm, ta được:

$$x_1 \leq x_2 \leq \dots \leq x_n.$$

Trung vị của mẫu, kí hiệu là M_e , là giá trị ở chính giữa dãy x_1, x_2, \dots, x_n . Cụ thể:

- Nếu $n = 2k + 1$, $k \in \mathbb{N}$ (tức n là số tự nhiên lẻ), thì trung vị của mẫu $M_e = x_{k+1}$.

- Nếu $n = 2k$, $k \in \mathbb{N}$ (tức n là số tự nhiên chẵn), thì trung vị của mẫu $M_e = \frac{1}{2}(x_k + x_{k+1})$.

7.1.2 Ý nghĩa của số trung vị

Trung vị được dùng để đo xu thế trung tâm của mẫu số liệu. Trung vị là giá trị nằm ở chính giữa của mẫu số liệu theo nghĩa: luôn có ít nhất 50% số liệu trong mẫu lớn hơn hoặc bằng trung vị và ít nhất 50% số liệu trong mẫu nhỏ hơn hoặc bằng trung vị. Khi trong mẫu xuất hiện thêm một giá trị rất lớn hoặc rất nhỏ thì số trung bình sẽ bị thay đổi đáng kể nhưng trung vị thì ít thay đổi.

7.2. Tứ phân vị

- Trung vị chia mẫu thành hai phần. Trong thực tế người ta cũng quan tâm đến trung vị của mỗi phần đó. Ba trung vị này được gọi là **tứ phân vị** của mẫu.

Sắp xếp mẫu số liệu theo thứ tự không giảm, ta được:

$$x_1 \leq x_2 \leq \dots \leq x_n.$$

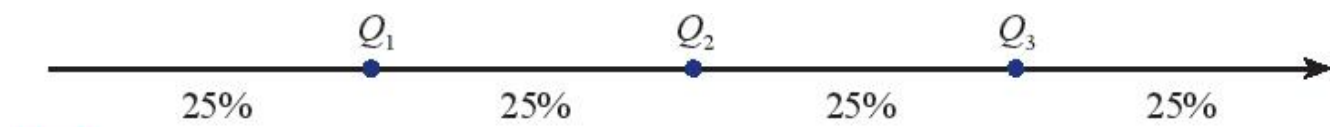
Tứ phân vị của một mẫu số liệu gồm ba giá trị, gọi là **tứ phân vị thứ nhất, thứ hai và thứ ba** (lần lượt kí hiệu là Q_1 , Q_2 , Q_3). Ba giá trị này chia tập hợp dữ liệu đã sắp xếp thành bốn phần đều nhau. Cụ thể:

- Giá trị tứ phân vị thứ hai, Q_2 , chính là số trung vị của mẫu.
- Giá trị tứ phân vị thứ nhất, Q_1 , là trung vị của nửa số liệu đã sắp xếp bên trái Q_2 (không bao gồm Q_2 nếu n lẻ).
- Giá trị tứ phân vị thứ ba, Q_3 , là trung vị của nửa số liệu đã sắp xếp bên phải Q_2 (không bao gồm Q_2 nếu n lẻ).

- **Ý nghĩa của tứ phân vị**

Các điểm tứ phân vị 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 chia khoảng 25% tổng số liệu đã thu thập được.

Tứ phân vị thứ nhất Q_1 còn được gọi là tứ phân vị dưới và đại diện cho nửa mẫu số liệu phía dưới. Tứ phân vị thứ ba Q_3 , còn được gọi là tứ phân vị trên và đại diện cho nửa mẫu số liệu ở phía trên.



8. Một

Cho mẫu số liệu dưới dạng bảng tần số. Giá trị có tần số lớn nhất được gọi là **mốt** của mẫu số liệu và kí hiệu là M_o .

Ý nghĩa của mốt: Mốt đặc trưng cho giá trị xuất hiện nhiều nhất trong mẫu.

Chú ý: Một mẫu số liệu có thể có rất nhiều mốt. Khi tất cả các giá trị trong mẫu số liệu có tần số xuất hiện bằng nhau thì mẫu số liệu đó không có mốt.

9. Khoảng biến thiên và khoảng tứ phân vị

9.1. Khoảng biến thiên và khoảng tứ phân vị

Sắp xếp mẫu số liệu theo thứ tự không giảm, ta được:

$$x_1 \leq x_2 \leq \dots \leq x_n.$$

• **Khoảng biến thiên** của một mẫu số liệu, kí hiệu là R , là hiệu giữa giá trị lớn nhất và giá trị nhỏ nhất của mẫu số liệu đó, tức là:

$$R = x_n - x_1.$$

• **Khoảng tứ phân vị**, kí hiệu là Δ_Q , là hiệu giữa Q_3 và Q_1 , tức là:

$$\Delta_Q = Q_3 - Q_1.$$

9.2. Ý nghĩa của khoảng biến thiên và khoảng tứ phân vị:

Khoảng biến thiên đặc trưng cho độ phân tán của toàn bộ mẫu số liệu.

Khoảng tứ phân vị đặc trưng cho độ phân tán của một nửa các số liệu, có giá trị thuộc đoạn từ Q_1 đến Q_3 trong mẫu.

Khoảng tứ phân vị không bị ảnh hưởng bởi các giá trị rất lớn hoặc rất bé trong mẫu.

9.3. Giá trị ngoại lệ

Khoảng tứ phân vị được dùng để xác định các *giá trị ngoại lệ* trong mẫu, đó là các giá trị quá nhỏ hay quá lớn so với đa số các giá trị của mẫu. Cụ thể, phần tử x trong mẫu là giá trị ngoại lệ nếu $x > Q_3 + 1,5\Delta_Q$ hoặc $x < Q_1 - 1,5\Delta_Q$.

Sự xuất hiện của các giá trị ngoại lệ làm cho số trung bình và phạm vi của mẫu thay đổi lớn. Do đó, khi mẫu có giá trị ngoại lệ, người ta thường sử dụng trung vị và khoảng tứ phân vị để đo mức độ tập trung và mức độ phân tán của đa số các phần tử trong mẫu số liệu.

10. Phương sai và độ lệch chuẩn

10.1. Công thức tính phương sai và độ lệch chuẩn

* Giả sử ta có một mẫu số liệu là x_1, x_2, \dots, x_n .

• **Phương sai** của mẫu số liệu này, kí hiệu là S^2 , được tính bởi công thức:

$$S^2 = \frac{1}{n} \left[(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2 \right],$$

trong đó \bar{x} là số trung bình của mẫu số liệu.

• Căn bậc hai của phương sai được gọi là **độ lệch chuẩn**, kí hiệu là S .

Chú ý: Có thể biến đổi công thức tính phương sai ở trên thành:

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

Trong thống kê, người ta cũng quan tâm đến phương sai hiệu chỉnh, kí hiệu là \hat{s}^2 , được tính bởi công thức:

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

* Giả sử mẫu số liệu được cho dưới dạng bảng tần số:

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

Khi đó, công thức tính phương sai trở thành:

$$S^2 = \frac{1}{n} \left[n_1 (x_1 - \bar{x})^2 + n_2 (x_2 - \bar{x})^2 + \dots + n_k (x_k - \bar{x})^2 \right],$$

trong đó $n = n_1 + n_2 + \dots + n_k$.

Có thể biến đổi công thức tính phương sai trên thành

$$S^2 = \frac{1}{n} (n_1 x_1^2 + n_2 x_2^2 + \dots + n_k x_k^2) - \bar{x}^2.$$

10.2. Ý nghĩa của phương sai và độ lệch chuẩn

Phương sai là trung bình cộng của các bình phương độ lệch từ mỗi giá trị của mẫu số liệu đến số trung bình.

Phương sai và độ lệch chuẩn được dùng để đo mức độ phân tán của các số liệu trong mẫu quanh số trung bình. Phương sai và độ lệch chuẩn càng lớn thì các giá trị của mẫu càng cách xa nhau (có độ phân tán lớn).

B. Bài tập tự luyện

Bài 1. Cho số $\sqrt{5} = 2,236067977\dots$

a) Hãy quy tròn $\sqrt{5}$ đến hàng phần trăm.

b) Hãy tìm số gần đúng của $\sqrt{5}$ với độ chính xác 0,005.

Hướng dẫn giải

a) Quy tròn số $\sqrt{5}$ đến hàng phần trăm ta được số gần đúng là 2,24.

Vậy $\sqrt{5} \approx 2,24$ (quy tròn đến hàng phần trăm).

b) Hàng của chữ số khác 0 đầu tiên bên trái của độ chính xác 0,005 là hàng phần nghìn.

Quy tròn $\sqrt{5}$ đến hàng phần nghìn ta được số gần đúng là 2,236.

Vậy $\sqrt{5} \approx 2,236$ với độ chính xác 0,005.

Bài 2. 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

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

Bài 3. Hãy viết số quy tròn của số gần đúng trong những trường hợp sau:

a) 3678008 ± 1000 ;

b) $21,02345 \pm 0,001$.

Hướng dẫn giải

a) 3678008 ± 1000

Hàng lớn nhất của độ chính xác $d = 1000$ là hàng nghìn, nên ta quy tròn đến hàng phần chục nghìn.

Vậy số quy tròn trong trường hợp này là 3680000.

b) $21,02345 \pm 0,001$

Hàng lớn nhất của độ chính xác $d = 0,001$ là hàng phần nghìn, nên ta quy tròn đến hàng phần trăm.

Vậy số quy tròn cần tìm là 21,02.

Bài 4. Một tam giác có ba cạnh đo được như sau: $a = 5,4 \text{ cm} \pm 0,2 \text{ cm}$; $b = 7,2 \text{ cm} \pm 0,2 \text{ cm}$ và $c = 9,7 \text{ cm} \pm 0,1 \text{ cm}$. Tính chu vi của tam giác đó.

Hướng dẫn giải

Chu vi của tam giác là:

$$\begin{aligned} P &= a + b + c = (5,4 \pm 0,2) + (7,2 \pm 0,2) + (9,7 \pm 0,1) \\ &= (5,4 + 7,2 + 9,7) \pm (0,2 + 0,2 + 0,1) \\ &= 22,3 \pm 0,5 \text{ (cm)}. \end{aligned}$$

Vậy chu vi của tam giác đã cho là $P = 22,3 \text{ cm} \pm 0,5 \text{ cm}$.

Bài 5. Một đội gồm 15 thợ điêu khắc được chia đều vào 3 tổ. Trong một ngày, mỗi người thợ làm được 2 hoặc 3 sản phẩm. Cuối ngày, đội trưởng thống kê lại số sản phẩm mà mỗi tổ làm được ở bảng sau:

Tổ	1	2	3
Số sản phẩm	10	15	17

Đội trưởng đã thống kê đúng chưa? Tại sao?

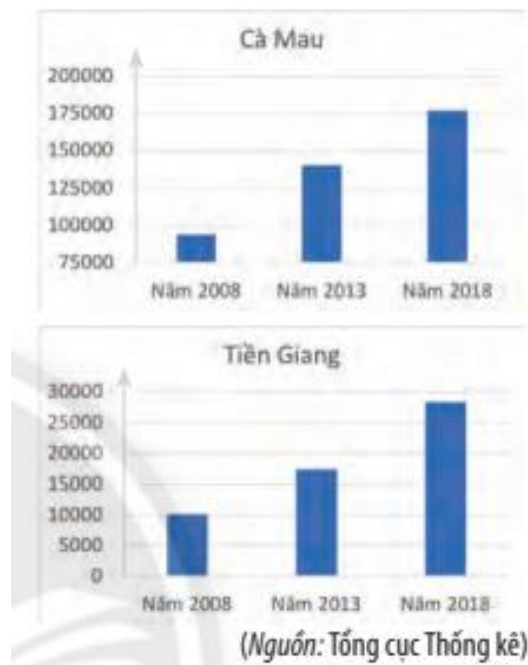
Hướng dẫn giải

Mỗi tổ có $15 : 3 = 5$ người thợ.

Trong một ngày, mỗi người thợ làm được 2 hoặc 3 sản phẩm nên mỗi tổ làm được từ 10 đến 15 sản phẩm. Do đó, bảng trên ghi Tổ 4 làm được 17 sản phẩm là không chính xác.

Vậy đội trưởng thống kê chưa đúng.

Bài 6. Sản lượng nuôi tôm phân theo địa phương của các tỉnh Cà Mau và Tiền Giang được thể hiện ở hai biểu đồ sau (đơn vị: tấn):



a) Hãy cho biết các phát biểu sau là đúng hay sai:

- Sản lượng nuôi tôm mỗi năm của tỉnh Tiền Giang đều cao hơn tỉnh Cà Mau.
- Ở tỉnh Cà Mau, sản lượng nuôi tôm năm 2018 tăng gấp hơn 4 lần so với năm 2008.
- Ở tỉnh Tiền Giang, sản lượng nuôi tôm năm 2018 tăng gấp hơn 2,5 lần so với năm 2008.

b) Để so sánh sản lượng nuôi tôm của hai tỉnh Cà Mau và Tiền Giang, ta nên sử dụng loại biểu đồ nào?

Hướng dẫn giải

a)

i. Quan sát biểu đồ ta thấy:

Sản lượng nuôi tôm mỗi năm ở Tiền Giang đều thấp hơn 30 000 tấn, sản lượng nuôi tôm mỗi năm ở Cà Mau đều cao hơn 75 000 tấn.

Do đó sản lượng nuôi tôm mỗi năm của tỉnh Cà Mau đều cao hơn rất nhiều so với tỉnh Tiền Giang.

Vậy phát biểu i. là sai.

ii. Ở tỉnh Cà Mau:

- Sản lượng nuôi tôm năm 2018 là 175 000 tấn.

- Sản lượng nuôi tôm năm 2008 khoảng hơn 90 000 tấn.

$$\text{Vì } \frac{175000}{90000} \approx 2.$$

Do đó sản lượng nuôi tôm năm 2018 ở tỉnh Cà Mau tăng khoảng gần 2 lần so với năm 2008.

Vậy phát biểu ii. là sai.

iii. Ở tỉnh Tiền Giang:

- Sản lượng nuôi tôm năm 2018 khoảng 29 000 tấn.

- Sản lượng nuôi tôm năm 2008 là 10 000 tấn.

$$\text{Vì } \frac{29000}{10000} = 2,9.$$

Do đó sản lượng nuôi tôm năm 2018 ở tỉnh Tiền Giang tăng gấp khoảng 2,9 ($> 2,5$) lần so với năm 2008.

Vậy phát biểu iii. là đúng.

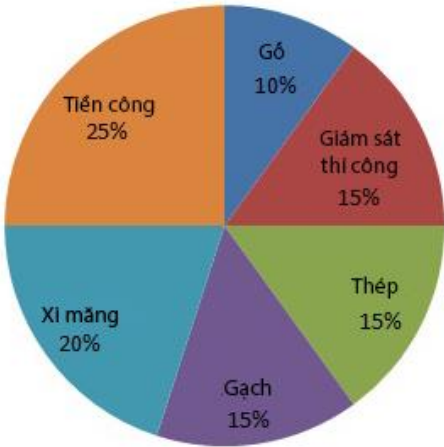
b) Để so sánh sản lượng nuôi tôm của hai tỉnh Cà Mau và Tiền Giang, ta nên sử dụng loại biểu đồ cột ghép.

Bài 7. Hoa vẽ biểu đồ biểu thị tỉ lệ chi phí xây dựng nhà ở của gia đình theo bảng thống kê dưới đây:

Loại chi phí	Số tiền (triệu đồng)
Tiền công	250
Gỗ	100
Giám sát thi công	150
Thép	150
Gạch	200
Xi măng	150



Chi Phí Xây Dựng Nhà



Bạn hãy cho biết biểu đồ Hoa vẽ đã chính xác chưa. Nếu chưa thì cần điều chỉnh lại như thế nào cho đúng?

Hướng dẫn giải

Theo bảng thống kê thì các loại chi phí giám sát thi công, thép và xi măng bằng nhau nên trên biểu đồ quạt, hình quạt biểu diễn tỉ lệ giám sát thi công, thép và xi măng phải bằng nhau. Do đó biểu đồ bạn Hoa vẽ chưa chính xác.

Ta cần đổi chỗ phần chữ chú thích trên biểu đồ của xi măng cho gạch thì biểu đồ chính xác.

Bài 8. Hãy tìm số trung bình, tứ phân vị và một của mẫu số liệu sau:

56; 45; 65; 45; 56; 78; 100; 78; 78.

Hướng dẫn giải

Cỡ mẫu: $n = 9$.

Số trung bình: $\bar{x} = \frac{1}{9}(56 + 45 + 65 + 45 + 56 + 78 + 100 + 78 + 78) \approx 66,78$.

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

45; 45; 56; 56; 65; 78; 78; 78; 100.

Vì cỡ mẫu là 9, là số lẻ nên tứ phân vị thứ hai là $Q_2 = 65$.

Tứ phân vị thứ nhất là trung vị của mẫu: 45; 45; 56; 56. Do đó $Q_1 = \frac{1}{2}(45 + 56) = 50,5$.

Tứ phân vị thứ ba là trung vị của mẫu: 78; 78; 78; 100. Do đó $Q_3 = \frac{1}{2}(78 + 78) = 78$.

Giá trị 78 có tần số lớn nhất nên một của mẫu số liệu là $M_o = 78$.

Bài 9. Hãy tìm số trung bình, trung vị và một của mẫu số liệu sau:

Giá trị	20	25	30	35
Tần số	2	3	5	7

Hướng dẫn giải

Cỡ mẫu $n = 2 + 3 + 5 + 7 = 17$.

Số trung bình: $\bar{x} = \frac{1}{17}(2.20 + 3.25 + 5.30 + 7.35) = 30.$

Sắp xếp các số liệu đã cho theo thứ tự không giảm, ta được:

20; 20; 25; 25; 25; 30; 30; 30; 30; 30; 35; 35; 35; 35; 35; 35; 35.

Vì cỡ mẫu là 17 là số lẻ nên trung vị là $M_e = 30$.

Giá trị 35 có tần số lớn nhất nên mốt của mẫu số liệu là $M_o = 35$.

Bài 10. Trong một cuộc thi nghề, người ta ghi lại thời gian hoàn thành một sản phẩm của một số thí sinh ở bảng sau:

Thời gian (đơn vị: phút)	5	6	7	8	35
Số thí sinh	1	3	5	2	1

a) Hãy tìm số trung bình, tứ phân vị và mốt của thời gian thi nghề của các thí sinh trên.

b) Năm ngoái, thời gian thi của các thí sinh có số trung bình và trung vị đều bằng 7. Bạn hãy so sánh thời gian thi nói chung của các thí sinh trong hai năm.

Hướng dẫn giải

a) Cỡ mẫu là $n = 1 + 3 + 5 + 2 + 1 = 12$.

Số trung bình là: $\bar{x} = \frac{1.5 + 3.6 + 5.7 + 2.8 + 1.35}{12} \approx 9,08.$

Số thí sinh là trong thời gian 7 phút là nhiều nhất nên mốt của mẫu là $M_o = 7$.

Sắp xếp các giá trị của mẫu theo thứ tự không giảm, ta được:

5; 6; 6; 6; 7; 7; 7; 7; 7; 8; 8; 35.

Vì cỡ mẫu là số chẵn nên tứ phân vị thứ hai là $Q_2 = \frac{1}{2}(7 + 7) = 7.$

Tứ phân vị thứ nhất là trung vị của mẫu: 5; 6; 6; 6; 7; 7. Do đó $Q_1 = 6$.

Tứ phân vị thứ ba là trung vị của mẫu: 7; 7; 7; 8; 8; 35. Do đó $Q_3 = 7,5$.

b) Dựa theo số trung bình, vì $9,08 > 7$ nên thời gian thi của các thí sinh năm nay nhiều hơn năm ngoái.

Dựa theo trung vị, thì cả hai năm trung vị đều bằng nhau và bằng 7 nên thời gian của các thí sinh trong hai năm là ngang nhau.

Vì trong mẫu số liệu của năm nay có số liệu 35 lớn hơn so với các số liệu còn lại rất nhiều, do đó ta dùng trung vị để so sánh sẽ phù hợp hơn.

Vậy thời gian thi nói chung của các thí sinh trong hai năm là ngang nhau.

Bài 11. Người ta đã tiến hành thăm dò ý kiến của khách hàng về các mẫu 1, 2, 3, 4 của một loại sản phẩm mới được sản xuất ở nhà máy. Dưới đây là bảng phân bố tần số theo số phiếu bình chọn tín nhiệm cho các mẫu kể trên.

Mẫu	1	2	3	4	Cộng
Tần số	195	300	356	149	1000

a) Tìm mốt của mẫu số liệu.

b) Trong sản xuất, nhà máy nên ưu tiên cho mẫu nào?

Hướng dẫn giải

a) Quan sát bảng tần số ta thấy mẫu 3 có tần số lớn nhất nên mốt của mẫu số liệu đã cho là $M_0 = 3$.

b) Vì $M_0 = 3$ nên trong sản xuất, nhà máy nên ưu tiên cho mẫu 3.

Bài 12. Hãy tìm độ lệch chuẩn, khoảng biến thiên, khoảng tứ phân vị và các giá trị ngoại lệ (nếu có) của mẫu số liệu sau: 6; 8; 3; 4; 5; 6; 7; 2; 4.

Hướng dẫn giải

Số trung bình: $\bar{x} = \frac{6+8+3+4+5+6+7+2+4}{9} = 5.$

Phương sai mẫu số liệu là:

$$S^2 = \frac{1}{9} (6^2 + 8^2 + 3^2 + 4^2 + 5^2 + 6^2 + 7^2 + 2^2 + 4^2) - 5^2 = \frac{10}{3}.$$

Độ lệch chuẩn mẫu số liệu là: $S = \sqrt{S^2} = \sqrt{\frac{10}{3}} = \frac{\sqrt{30}}{3}.$

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

2; 3; 4; 4; 5; 6; 6; 7; 8.

Khoảng biến thiên của mẫu là: $R = 8 - 2 = 6.$

Vì cỡ mẫu là 9 là số lẻ nên tứ phân vị thứ hai là $Q_2 = 5.$

Tứ phân vị thứ nhất là trung vị của mẫu: 2; 3; 4; 4. Do đó $Q_1 = 3,5.$

Tứ phân vị thứ ba là trung vị của mẫu: 6; 6; 7; 8. Do đó $Q_3 = 6,5.$

Khoảng tứ phân vị của mẫu là: $\Delta_Q = 6,5 - 3,5 = 3.$

Ta có: $Q_3 + 1,5\Delta_Q = 6,5 + 1,5 \cdot 3 = 11$ và $Q_1 - 1,5\Delta_Q = 3,5 - 1,5 \cdot 3 = -1.$

Do đó mẫu số liệu không có giá trị ngoại lệ.

Bài 13. Hai lớp 10A, 10B của một trường Trung học phổ thông đồng thời làm bài thi môn Toán theo cùng một đề thi. Kết quả thi được trình bày ở hai bảng phân bố tần số sau đây:

Điểm thi Toán của lớp 10A

Điểm thi	5	6	7	8	9	10	Cộng
Tần số	3	7	12	14	3	1	40

Điểm thi Toán của lớp 10B

Điểm thi	6	7	8	9	Cộng
Tần số	8	18	10	4	40

a) Tính các số trung bình cộng, phương sai, độ lệch chuẩn của các mẫu số liệu đã cho.

b) Xét xem kết quả làm bài thi môn Toán ở lớp nào đồng đều hơn?

Hướng dẫn giải

a)

* Lớp 10A:

$$\text{Số trung bình: } \overline{x}_A = \frac{1}{40} (3 \cdot 5 + 7 \cdot 6 + 12 \cdot 7 + 14 \cdot 8 + 3 \cdot 9 + 1 \cdot 10) = 7,25.$$

Phương sai mẫu số liệu:

$$S_A^2 = \frac{1}{40} (3 \cdot 5^2 + 7 \cdot 6^2 + 12 \cdot 7^2 + 14 \cdot 8^2 + 3 \cdot 9^2 + 1 \cdot 10^2) - 7,25^2 = 1,2875.$$

$$\text{Độ lệch chuẩn: } S_A = \sqrt{S_A^2} = \sqrt{1,2875} \approx 1,135.$$

* Lớp 10B:

$$\text{Số trung bình: } \overline{x}_B = \frac{1}{40} (8 \cdot 6 + 18 \cdot 7 + 10 \cdot 8 + 4 \cdot 9) = 7,25.$$

Phương sai mẫu số liệu:

$$S_B^2 = \frac{1}{40} (8 \cdot 6^2 + 18 \cdot 7^2 + 10 \cdot 8^2 + 4 \cdot 9^2) - 7,25^2 = 0,7875.$$

$$\text{Độ lệch chuẩn: } S_B = \sqrt{S_B^2} = \sqrt{0,7875} \approx 0,887.$$

b) Vì $0,887 < 1,135$ nên $S_B < S_A$ hay độ lệch chuẩn của mẫu số liệu lớp 10B nhỏ hơn lớp 10A.

Vậy kết quả làm bài thi của học sinh lớp 10B đồng đều hơn.

Bài 14. Kết quả điều tra mức lương hàng tháng của một số công nhân của hai nhà máy A và B được cho ở bảng sau (đơn vị: triệu đồng):

Công nhân nhà máy A	4	5	5	47	5	6	4	4	
Công nhân nhà máy B	2	9	9	8	10	9	9	11	9

a) Hãy tìm số trung bình, mốt, tứ phân vị và độ lệch chuẩn của hai mẫu số liệu lấy từ nhà máy A và nhà máy B.

b) Hãy tìm các giá trị ngoại lệ trong mỗi mẫu số liệu trên. Công nhân nhà máy nào có mức lương cao hơn? Tại sao?

Hướng dẫn giải

a)

* Nhà máy A:

+ Số trung bình mức lương hàng tháng: $\bar{x}_A = \frac{4 + 5 + 5 + 47 + 5 + 6 + 4 + 4}{8} = 10.$

+ Giá trị 4 và 5 có tần số lớn nhất nên mốt của mẫu số liệu ở nhà máy A là 4 và 5.

+ Sắp xếp các số liệu theo thứ tự không giảm, ta được:

4; 4; 4; 5; 5; 5; 6; 47.

Vì cỡ mẫu là 8 là số chẵn nên tứ phân vị thứ hai là $Q_{2A} = 5.$

Tứ phân vị thứ nhất là trung vị của mẫu: 4; 4; 4; 5. Do đó $Q_{1A} = 4.$

Tứ phân vị thứ ba là trung vị của mẫu: 5; 5; 6; 47. Do đó $Q_{3A} = 5,5.$

+ Phương sai mẫu:

$$S_A^2 = \frac{1}{8} (4^2 + 5^2 + 5^2 + 47^2 + 5^2 + 6^2 + 4^2 + 4^2) - 10^2 = 196.$$

+ Độ lệch chuẩn: $S_A = \sqrt{S_A^2} = \sqrt{196} = 14$.

* Nhà máy B:

+ Số trung bình mức lương hàng tháng: $\bar{x}_B = \frac{2+9+9+8+10+9+9+11+9}{9} \approx 8,4$.

+ Giá trị 9 có tần số lớn nhất nên một của mẫu số liệu ở nhà máy B là 9.

+ Sắp xếp các số liệu theo thứ tự không giảm, ta được:

2; 8; 9; 9; 9; 9; 9; 10; 11.

Vì cỡ mẫu là 9 là số lẻ nên tứ phân vị thứ hai là $Q_{2B} = 9$.

Tứ phân vị thứ nhất là trung vị của mẫu: 2; 8; 9; 9. Do đó $Q_{1B} = 8,5$.

Tứ phân vị thứ ba là trung vị của mẫu: 9; 9; 10; 11. Do đó $Q_{3B} = 9,5$.

+ Phương sai mẫu:

$$S_B^2 = \frac{1}{9} (2^2 + 8^2 + 9^2 + 9^2 + 9^2 + 9^2 + 9^2 + 10^2 + 11^2) - 8,4^2 = 6,55.$$

+ Độ lệch chuẩn: $S_B = \sqrt{S_B^2} = \sqrt{6,55} \approx 2,6$.

b)

+ Khoảng tứ phân vị của mẫu số liệu ở nhà máy A là: $\Delta_{QA} = 5,5 - 4 = 1,5$.

Ta có: $Q_{3A} + 1,5\Delta_{QA} = 5,5 + 1,5 \cdot 1,5 = 7,75$ và $Q_{1A} - 1,5\Delta_{QA} = 4 - 1,5 \cdot 1,5 = 1,75$.

Do đó giá trị ngoại lệ trong mẫu số liệu ở nhà máy A là 47.

+ Khoảng tứ phân vị của mẫu số liệu ở nhà máy B là: $\Delta_{QB} = 9,5 - 8,5 = 1$.

Ta có: $Q_{3B} + 1,5\Delta_{QB} = 9,5 + 1,5 \cdot 1 = 11$ và $Q_{1B} - 1,5\Delta_{QB} = 8,5 - 1,5 \cdot 1 = 7$.

Do đó giá trị ngoại lệ trong mẫu số liệu ở nhà máy B là 2.

+ Quan sát các số liệu tính được ở câu a), ta thấy

- Số trung bình mức lương hàng tháng của công nhân ở nhà máy A cao hơn nhà máy B.
- Phương sai mẫu và độ lệch chuẩn mẫu số liệu ở nhà máy A cao hơn nhà máy B nên mức lương hàng tháng của công nhân nhà máy A có độ phân tán cao hơn nhà máy B, do đó mức lương của công nhân nhà máy B ổn định hơn nhà máy A.
- Mức lương xuất hiện nhiều nhất trong mẫu A là 4 và 5 triệu đồng, nhà máy B là 9 triệu đồng.

Do đó, ta có thể khẳng định công nhân nhà máy A có mức lương cao hơn (đều và ổn định hơn).