

ÔN TẬP CHƯƠNG VI. THỐNG KÊ

TOÁN 10 – HỌC KỲ 1

NĂM HỌC 2024 – 2025

TỔ TOÁN – TRƯỜNG THPT HỒ THỊ BÌ



I. Các số đặc trưng đo xu thế trung tâm của mẫu số liệu

1. Số trung bình

- Giả sử ta có một mẫu số liệu là $x_1, x_2, x_3, \dots, x_n$.

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

Khi đó, công thức tính số trung bình trở thành:

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

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

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

I. Các số đặc trưng đo xu thế trung tâm của mẫu số liệu

1. Số trung bình

- 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.

2. Số trung vị

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

I. Các số đặc trưng đo xu thế trung tâm của mẫu số liệu

2. Số trung vị

- **Ý nghĩa của 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. 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.

3. Tứ phân vị

Sắp xếp mẫu số liệu theo thứ tự không giảm, ta được: $x_1 \leq x_2 \leq \dots \leq x_n$.

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:

I. Các số đặc trưng đo xu thế trung tâm của mẫu số liệu

3. Tứ phân vị

- 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 chứa khoảng 25% tổng số số liệu đã thu thập được. Tứ phân vị thứ nhất còn gọi là tứ phân vị dưới, đại diện cho nửa mẫu số liệu phía dưới; tứ phân vị thứ ba còn gọi là tứ phân vị trên, đại diện cho nửa mẫu số liệu phía trên.

I. Các số đặc trưng đo xu thế trung tâm của mẫu số liệu

4. Mốt

- Cho một 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_0 .
- **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 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.

II. Các số đặc trưng đo mức độ phân tán của mẫu số liệu

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: $\Delta_Q = Q_3 - Q_1$.

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.

II. Các số đặc trưng đo mức độ phân tán của mẫu số liệu

1. Khoảng biến thiên và khoảng tứ phân vị

- **Giá trị ngoại lệ:** khoảng tứ phân vị dùng để xác định các giá trị ngoại lệ trong mẫu, đó là các giá trị quá nhỏ hoặc 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 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.

II. Các số đặc trưng đo mức độ phân tán của mẫu số liệu

2. 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:

$$S^2 = \frac{1}{n} \left[(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2 \right], \text{ trong đó } \bar{x} \text{ là số trung bình của mẫu số liệu.}$$

Công thức trên cũng có thể viết như sau: $S^2 = \frac{1}{n} (x_1^2 + x_2^2 + \dots + x_n^2) - \bar{x}^2$.

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:

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

II. Các số đặc trưng đo mức độ phân tán của mẫu số liệu

2. Phương sai và độ lệch chuẩn

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

III. Bài tập củng cố

Bài 1. Độ tuổi của 22 cầu thủ ở đội hình xuất phát của hai đội bóng đá được ghi lại ở bảng sau:

a) Hãy tìm số trung bình, mốt, độ lệch chuẩn và tứ phân vị của tuổi mỗi cầu thủ của từng đội bóng.

b) Tuổi của các cầu thủ ở đội bóng nào đồng đều hơn? Tại sao?

Hướng dẫn giải.

a) Bảng tần số của tuổi cầu thủ của đội bóng A:

Tuổi	20	21	23	24	25	26	28	29
Tần số	1	1	1	3	2	1	1	1

Đội A	Đội B
28	32
24	20
26	19
25	21
25	28
23	29
20	21
29	22
21	29
24	19
24	29

III. Bài tập củng cố

Bài 1. Hướng dẫn giải.

a) Bảng tần số của tuổi cầu thủ của đội bóng B:

Tuổi	19	20	21	22	28	29	32
Tần số	2	1	2	1	1	3	1

Tuổi trung bình của các cầu thủ đội bóng A là

$$\bar{x}_A = \frac{20.1 + 21.1 + 23.1 + 24.3 + 25.2 + 26.1 + 28.1 + 29.1}{11} \approx 24,45$$

Tuổi trung bình của các cầu thủ đội bóng B là

$$\bar{x}_B = \frac{19.2 + 20.1 + 21.2 + 22.1 + 28.1 + 29.3 + 32.1}{11} \approx 22,54$$

Dựa vào bảng tần số, ta có một của đội A là 24 và một của đội B là 29.

Đội A	Đội B
28	32
24	20
26	19
25	21
25	28
23	29
20	21
29	22
21	29
24	19
24	29

III. Bài tập củng cố

Bài 1. Hướng dẫn giải.

Bảng tần số của tuổi cầu thủ đội bóng A

Tuổi	20	21	23	24	25	26	28	29
Tần số	1	1	1	3	2	1	1	1

Bảng tần số của tuổi cầu thủ đội bóng B

Tuổi	19	20	21	22	28	29	32
Tần số	2	1	2	1	1	3	1

Độ lệch chuẩn của tuổi mỗi cầu thủ ở đội bóng A là

$$S_A = \sqrt{\frac{1}{11} \cdot (1 \cdot 20^2 + 1 \cdot 21^2 + 1 \cdot 23^2 + 3 \cdot 24^2 + 2 \cdot 25^2 + 1 \cdot 26^2 + 1 \cdot 28^2 + 1 \cdot 29^2) - (24,45)^2} \approx 2,58$$

Độ lệch chuẩn của tuổi mỗi cầu thủ ở đội bóng B là

$$S_B = \sqrt{\frac{1}{11} \cdot (2 \cdot 19^2 + 1 \cdot 20^2 + 2 \cdot 21^2 + 1 \cdot 22^2 + 1 \cdot 28^2 + 3 \cdot 29^2 + 1 \cdot 32^2) - (22,54)^2} \approx 10,58$$

III. Bài tập củng cố

Bài 1.

Hướng dẫn giải.

Bảng tần số của tuổi cầu thủ đội bóng A

Tuổi	20	21	23	24	25	26	28	29
Tần số	1	1	1	3	2	1	1	1

Trung vị: $M_e = x_6 = 25$; Tứ phân vị thứ nhất: $Q_1 = x_3 = 23$; Tứ phân vị thứ ba: $Q_3 = x_9 = 26$

Bảng tần số của tuổi cầu thủ đội bóng B

Tuổi	19	20	21	22	28	29	32
Tần số	2	1	2	1	1	3	1

Trung vị: $M_e = x_6 = 22$; Tứ phân vị thứ nhất: $Q_1 = x_3 = 20$; Tứ phân vị thứ ba: $Q_3 = x_9 = 29$

III. Bài tập củng cố

Bài 1.

Hướng dẫn giải.

b) Ta có:
$$S_A = \sqrt{\frac{1}{11} \cdot (1.20^2 + 1.21^2 + 1.23^2 + 3.24^2 + 2.25^2 + 1.26^2 + 1.28^2 + 1.29^2) - (24,45)^2} \approx 2,58$$

$$S_B = \sqrt{\frac{1}{11} \cdot (2.19^2 + 1.20^2 + 2.21^2 + 1.22^2 + 1.28^2 + 3.29^2 + 1.32^2) - (22,54)^2} \approx 10,58$$

Vì độ lệch chuẩn của tuổi của mỗi cầu thủ ở đội bóng A nhỏ hơn độ lệch chuẩn của tuổi của mỗi cầu thủ ở đội bóng B nên tuổi các cầu thủ ở đội bóng A đều hơn.

III. Bài tập củng cố

Bài 2. Bạn Châu cân lần lượt 50 quả vải thiều Thanh Hà được lựa chọn ngẫu nhiên từ vườn nhà mình và được kết quả như sau:

Cân nặng (đơn vị: gam)	Số quả
8	1
19	10
20	19
21	17
22	3

- Hãy tìm số trung bình, trung vị, mốt của mẫu số liệu trên.
- Hãy tìm độ lệch chuẩn, khoảng biến thiên, khoảng tứ phân vị và giá trị ngoại lệ của mẫu số liệu trên.

III. Bài tập củng cố

Hướng dẫn giải.

$$\text{a) } \bar{x} = \frac{8 \cdot 1 + 19 \cdot 10 + 20 \cdot 19 + 21 \cdot 17 + 22 \cdot 3}{1 + 10 + 19 + 17 + 3} = 20,02$$

Cỡ mẫu là 50. Khi sắp xếp các số liệu theo thứ tự không giảm thì số liệu thứ 25, 26 lần lượt là 20 và 20 nên trung vị của mẫu số liệu là $M_e = \frac{20 + 20}{2} = 20$

Giá trị 20 có tần số lớn nhất nên $M_0 = 20$.

b) Khoảng biến thiên $R = 22 - 8 = 14$.

$$\text{Độ lệch chuẩn } S = \sqrt{\frac{1}{50}(1 \cdot 8^2 + 10 \cdot 19^2 + 19 \cdot 20^2 + 17 \cdot 21^2 + 3 \cdot 22^2) - (20,02)^2} \approx 1,9$$

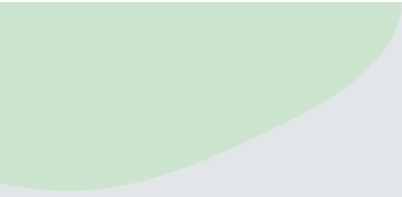
Tứ phân vị thứ nhất $Q_1 = x_{13} = 20$; tứ phân vị thứ ba $Q_3 = x_{38} = 21$

Khoảng tứ phân vị là $\Delta_Q = Q_3 - Q_1 = 21 - 20 = 1$.

Ta có $Q_3 + 1,5 \cdot \Delta_Q = 21 + 1,5 = 22,5$ và $Q_1 - 1,5 \Delta_Q = 20 - 1,5 = 18,5$

nên giá trị ngoại lệ của mẫu số liệu trên là 8 (vì $8 < Q_1 - 1,5 \Delta_Q$).

Cân nặng (đơn vị: gam)	Số quả
8	1
19	10
20	19
21	17
22	3



— ***BÀI GIẢNG ĐẾN ĐÂY LÀ KẾT THÚC!***

Chúc các em học thật tốt!

