

§3. CÁC SỐ ĐẶC TRUNG CỦA MẪU SỐ LIỆU (3 tiết)

I. MỤC TIÊU

Giúp học sinh :

Về kiến thức. Nhớ được công thức tính các số đặc trưng của mẫu số liệu như trung bình, số trung vị, mốt, phương sai và độ lệch chuẩn và hiểu được ý nghĩa của các số đặc trưng này.

Về kĩ năng. Biết cách tính các số trung bình, số trung vị, mốt, phương sai, độ lệch chuẩn.

II. NHỮNG ĐIỀU CẦN LƯU Ý

1) Học sinh nhiều khi thuộc công thức, biết cách tính các đặc trưng mẫu nói trên nhưng không hiểu ý nghĩa và không biết dùng các số đó vào việc gì. Việc hiểu ý nghĩa của các số đặc trưng là rất quan trọng. Giáo viên khi dạy cần chú ý để cho học sinh hiểu rõ ý nghĩa của các số đặc trưng trong những tình huống thực tế.

2) Số trung bình, số trung vị, mốt cho ta những thông tin cô đọng về mẫu số liệu và chúng thường được dùng làm đại diện cho mẫu số liệu.

Khi trong mẫu số liệu có những giá trị dị thường (chênh lệch quá lớn hoặc quá bé so với các giá trị khác) thì số trung vị làm đại diện tốt hơn. Thông thường, khi trong mẫu không có những giá trị dị thường như vậy, thì số trung vị và số trung bình khá gần nhau (Xem thêm tiểu mục 1 trong mục *Bổ sung kiến thức* của bài này).

3) Lưu ý rằng để tính số trung vị, trước hết cần sắp xếp các số liệu trong mẫu theo thứ tự tăng dần.

4) Phương sai là số đo mức độ phân tán của các giá trị trong mẫu số liệu quanh số trung bình. Nhìn vào công thức tính phương sai ta thấy nó là *trung bình cộng của tổng bình phương các khoảng cách từ mỗi số liệu tới số trung bình*. Vì thế, người ta còn gọi phương sai là độ lệch bình phương trung bình. Khi số liệu có đơn vị thì đơn vị của phương sai là bình phương đơn vị của số liệu.

5) Độ lệch chuẩn là căn bậc hai của phương sai. Do đó, nó cũng là một số đo mức độ phân tán các giá trị trong mẫu số liệu quanh số trung bình. Khi số liệu có đơn vị thì độ lệch chuẩn có cùng đơn vị đo với đơn vị của số liệu.

6) Trong trường hợp mẫu số liệu cho dưới dạng bảng phân bố tần số ghép lớp, ta có thể tính (xấp xỉ) số trung bình cũng như phương sai bằng cách lấy trung điểm mỗi lớp làm đại diện. Khi đó, ta coi giá trị đại diện này là có tần số bằng tần số của lớp đó. Cần lưu ý rằng số trung bình, phương sai tính được theo cách đó chỉ là *số trung bình, phương sai gần đúng so với cách tính đúng theo định nghĩa từ các số liệu ban đầu* (vì ta đã thay các số liệu trong khoảng bởi giá trị đại diện). Cách tính này có sai số khá nhỏ, chấp nhận được. Không nên cho rằng đây là một công thức mới để tính số trung bình và phương sai trong trường hợp mẫu số liệu được cho dưới dạng bảng phân bố tần số - tần suất ghép lớp.

III. GỢI Ý VỀ DẠY HỌC

1) *Gợi ý về phân phối thời gian*

Bài này có thời lượng 3 tiết. Một tiết dành cho khái niệm số trung bình, số trung vị và mốt. Hai tiết dành cho khái niệm phương sai, độ lệch chuẩn. Việc tính số trung bình học sinh đã được làm quen từ lớp 7. Điểm mới ở đây là tính số trung bình cho trường hợp mẫu số liệu được cho dưới dạng bảng phân bố tần số ghép lớp và tính phương sai, độ lệch chuẩn.

2) *Gợi ý về đồ dùng dạy học*. Sử dụng máy tính bỏ túi phổ thông (làm các phép tính cộng, trừ, nhân, chia, khai căn) hoặc máy tính CASIO *fx-500MS* (máy này có lập trình sẵn để tính số trung bình, phương sai, độ lệch chuẩn sau khi ta nhập số liệu vào). Giáo viên cần có một thời gian thích hợp trong tiết luyện tập để hướng dẫn học sinh sử dụng máy tính bỏ túi.

3) *Gợi ý các hoạt động trên lớp và trả lời câu hỏi*

H1 a) $M_e = 70$.

b) $\bar{x} \approx 42,32$. Số trung bình xấp xỉ số trung vị.

H2 Số liệu đứng thứ 18 là 165, đứng thứ 19 là 166. Do vậy, số trung vị là

$$M_e = \frac{165 + 166}{2} = 165,5.$$

H3 Điểm trung bình các môn học của An là 8,1. Điểm trung bình các môn học của Bình xấp xỉ 8,1 (chính xác đến hàng phần chục). An và Bình có điểm trung bình xấp xỉ nhau. Tuy nhiên, An học đều các môn còn Bình thì học giỏi các môn tự nhiên và học trung bình ở các môn xã hội. Câu hỏi "Theo em, bạn nào học khá hơn?" là câu hỏi mở. Giáo viên để học sinh tự do trình bày ý kiến, lí lẽ của mình.

Giáo viên hướng dẫn học sinh tính số trung bình, phương sai và độ lệch chuẩn trên máy tính bỏ túi. Khi giải một số bài toán tính phương sai, độ lệch chuẩn, giáo viên cần yêu cầu học sinh phải trình bày đầy đủ các bước tính toán theo như mẫu trong SGK, không được chỉ ghi đáp số.

IV. GỢI Ý TRẢ LỜI CÂU HỎI VÀ BÀI TẬP

9. a) $\bar{x} = 15,23$.

b) $M_e = 15,5$ (vì số liệu đứng thứ năm mươi là 15, số liệu đứng thứ năm mươi mốt là 16). Mốt là 16. Như vậy, có khoảng một nửa số học sinh có điểm dưới 15,5 và số học sinh đạt 16 điểm là nhiều nhất.

c) $s^2 \approx 3,96$. $s \approx 1,99$.

10. Ta có bảng sau :

Lớp	Giá trị đại diện	Tần số
[10 ; 19]	14,5	1
[20 ; 29]	24,5	14
[30 ; 39]	34,5	21
[40 ; 49]	44,5	73
[50 ; 59]	54,5	42
[60 ; 69]	64,5	13
[70 ; 79]	74,5	9
[80 ; 89]	84,5	4
[90 ; 99]	94,5	2
		$N = 179$

Từ đó, khối lượng trung bình của một củ khoai tây là $\bar{x} \approx 48,35$ g.

$s^2 \approx 194,64$. $s \approx 13,95$ g.

11. $\bar{x} \approx 2,35$. $s^2 \approx 1,57$. $s \approx 1,25$.

V. BỔ SUNG KIẾN THỨC

1. Chọn số nào làm đại diện : số trung bình, số trung vị hay mốt ?

Nói chung, số trung bình thường hay được sử dụng để làm đại diện cho mẫu số liệu. Tuy nhiên, tùy từng yêu cầu cụ thể mà người ta quan tâm tới việc dùng đại diện nào. Trong một số tình huống, dùng mốt hay số trung vị làm đại diện thì hợp lí hơn.

Ví dụ 1

Một cửa hàng đồ điện tử gia dụng bán năm loại tivi với giá tiền mỗi chiếc tương ứng là 1 ; 2 ; 3 ; 4 ; 5 (triệu đồng). Trong năm vừa qua có 1285 lượt khách mua các mặt hàng trên với bảng số liệu sau :

Giá tiền	1	2	3	4	5
Số chiếc bán được	256	350	500	104	75

Số trung bình xấp xỉ là 2,527 triệu đồng, mốt là 3 triệu đồng.

Ý nghĩa. Một chiếc tivi ở cửa hàng được bán với giá trung bình 2,527 triệu đồng. Cục thuế thì quan tâm nhất tới giá trị này để xác định doanh thu của cửa hàng. Song điều mà người chủ cửa hàng quan tâm lại là : Loại tivi nào nhiều người mua nhất ? Đó là loại tivi giá 3 triệu đồng. Như vậy, điều mà người chủ cửa hàng quan tâm nhất là mốt của mẫu số liệu trên.

Ví dụ 2

Một công ti tư nhân thuê sáu người A, B, C, D, E, F với mức lương hàng tháng như sau (đơn vị : USD ; F là giám đốc điều hành) :

Nhân viên	A	B	C	D	E	F
Tiền lương	56	60	70	120	120	450

Lương trung bình của mỗi nhân viên là 146 USD. Tuy nhiên, con số này không thể lấy làm đại diện cho mức lương trung bình của công ti vì nó cao hơn lương của tất cả năm người chỉ trừ giám đốc. Số trung vị

$M_e = \frac{70 + 120}{2} = 95$ (USD). Số trung vị phản ánh tốt mức lương trung bình của công ti.

2. Tính số trung vị cho bảng phân bố tần số ghép lớp

Trong trường hợp mẫu số liệu cho dưới dạng bảng phân bố tần số ghép lớp, ta có thể tính (xấp xỉ) số trung bình cũng như phương sai bằng cách lấy trung điểm mỗi đoạn (hay nửa khoảng) xác định lớp làm giá trị đại diện. Giá trị đại diện này được xem là có tần số bằng tần số của lớp đó.

Tuy nhiên, để tính số trung vị cho bảng phân bố tần số ghép lớp, trong đó mỗi lớp là một nửa khoảng thì phức tạp hơn. Giả sử mẫu số liệu với kích thước N được chia làm m lớp theo các nửa khoảng C_1, C_2, \dots, C_m có độ dài bằng nhau (kí hiệu độ dài này là h). Tần số của lớp C_i là n_i ($i = 1, 2, \dots, m$). Số trung vị được tính xấp xỉ theo cách sau :

i) Đầu tiên, ta xác định số nguyên dương k thoả mãn điều kiện

$$n_1 + n_2 + \dots + n_{k-1} < \frac{N}{2} \leq n_1 + n_2 + \dots + n_k.$$

Nghĩa là, k là số nguyên dương nhỏ nhất thoả mãn

$$\frac{N}{2} \leq n_1 + n_2 + \dots + n_k.$$

ii) Giả sử $C_k = [a_k ; a_{k+1})$. Khi đó, số trung vị được tính xấp xỉ theo công thức

$$M_e \approx a_k + \frac{h \left(\frac{N}{2} - n_1 - n_2 - \dots - n_{k-1} \right)}{n_k}.$$

Ví dụ 3

Cân 96 quả táo (đơn vị : gam), ta có được mẫu số liệu được trình bày trong bảng phân bố tần số ghép lớp sau đây.

Lớp	Tần số
[85,5 ; 87,5)	13
[87,5 ; 89,5)	16
[89,5 ; 91,5)	14
[91,5 ; 93,5)	27
[93,5 ; 95,5)	14
[95,5 ; 97,5)	12
	$N = 96$

Hãy tính số trung vị.

Giải. Ta có $\frac{N}{2} = \frac{96}{2} = 48$.

Vì $13 + 16 + 14 = 43 < 48 < 70 = 13 + 16 + 14 + 27$ nên $k = 4$.

Ta có

$$C_4 = [91,5 ; 93,5) \text{ và độ dài } h = 2.$$

Do đó, số trung vị là $M_e \approx 91,5 + \frac{2 \cdot (48 - 43)}{27} \approx 91,87$ (gam).

3. Biên độ và độ lệch trung bình

Ngoài phương sai ra, người ta cũng còn dùng hai chỉ tiêu khác để đo mức độ phân tán của các số liệu trong mẫu. Đó là biên độ và độ lệch trung bình.

a) Biên độ. *Biên độ* là hiệu số giữa giá trị lớn nhất và giá trị bé nhất của số liệu trong mẫu.

Ví dụ 4

Trong ví dụ 6 SGK, điểm trung bình các môn học của An có biên độ là $9 - 7 = 2$ còn điểm trung bình các môn học của Bình có biên độ là $10 - 5 = 5$.

Các giá trị lớn nhất và bé nhất của mẫu trong nhiều trường hợp cho ta những thông tin quan trọng nhưng cũng có nhiều trường hợp các giá trị này chỉ là ngoại lệ, chúng cho ta rất ít thông tin.

b) Độ lệch trung bình. *Độ lệch trung bình* (ĐLTB) là trung bình cộng khoảng cách từ các số liệu tới số trung bình

$$\text{ĐLTB} = \frac{1}{N} \sum_{i=1}^N |x_i - \bar{x}|.$$

Ví dụ 5

Trong ví dụ 6 SGK, ĐLTB điểm trung bình các môn học của An là

$$\frac{|8 - 8,1| + |7,5 - 8,1| + \dots + |9 - 8,1|}{11} = \frac{4,6}{11} \approx 0,42.$$

ĐLTB điểm trung bình các môn học của Bình là

$$\frac{|8,5 - 8,1| + |9,5 - 8,1| + \dots + |10 - 8,1|}{11} = \frac{15,5}{11} \approx 1,41.$$

Người ta ít dùng độ lệch trung bình vì việc tính toán nó không thuận tiện bằng phương sai.