

§3. CÁC SỐ ĐẶC TRƯNG CỦA MẪU SỐ LIỆU

A. TÓM TẮT LÝ THUYẾT.

1. Số trung bình

- Với mẫu số liệu kích thước N là x_1, x_2, \dots, x_N :

$$\bar{x} = \frac{\sum_{i=1}^N x_i}{N} = \frac{x_1 + x_2 + \dots + x_N}{N}$$

- Với mẫu số liệu được cho bởi bảng phân bố tần số:

$$\bar{x} = \frac{\sum_{i=1}^N n_i x_i}{N} = \frac{n_1 x_1 + n_2 x_2 + \dots + n_k x_k}{N}$$

- Với mẫu số liệu được cho bởi bảng phân bố tần số ghép lớp:

$$\bar{x} = \frac{\sum_{i=1}^N n_i c_i}{N} = \frac{n_1 c_1 + n_2 c_2 + \dots + n_k c_k}{N} \quad (c_i \text{ là giá trị đại diện của lớp thứ } i)$$

2. Số trung vị

Giả sử ta có một mẫu gồm N số liệu được sắp xếp theo thứ tự không giảm (hoặc không tăng). Khi đó số trung vị M_e là:

– Số đứng giữa nếu N lẻ;

– Trung bình cộng của hai số đứng giữa (số thứ $\frac{N}{2}$ và $\frac{N}{2} + 1$) nếu N chẵn.

3. Mốt

Mốt của một bảng phân bố tần số là giá trị có tần số lớn nhất và được kí hiệu là M_o .

Chú ý: – Số trung bình của mẫu số liệu được dùng làm đại diện cho các số liệu của mẫu.

– Nếu các số liệu trong mẫu có sự chênh lệch quá lớn thì dùng số trung vị làm đại diện cho các số liệu của mẫu.

– Nếu quan tâm đến giá trị có tần số lớn nhất thì dùng mốt làm đại diện. Một mẫu số liệu có thể có nhiều mốt.

4. Phương sai và độ lệch chuẩn

Để đo mức độ chênh lệch (**độ phân tán**) giữa các giá trị của mẫu số liệu so với số trung bình ta dùng **phương sai** s^2 và **độ lệch chuẩn** $s = \sqrt{s^2}$.

- Với mẫu số liệu kích thước N là x_1, x_2, \dots, x_N :

$$\begin{aligned} s^2 &= \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 = \frac{1}{N} \sum_{i=1}^N x_i^2 - \frac{1}{N^2} \left(\sum_{i=1}^N x_i \right)^2 \\ &= \overline{x^2} - (\bar{x})^2 \end{aligned}$$

- Với mẫu số liệu được cho bởi bảng phân bố tần số, tần suất:

$$s^2 = \frac{1}{N} \sum_{i=1}^k n_i (x_i - \bar{x})^2 = \frac{1}{N} \sum_{i=1}^k n_i x_i^2 - \frac{1}{N^2} \left(\sum_{i=1}^k n_i x_i \right)^2$$
$$= \sum_{i=1}^k f_i (x_i - \bar{x})^2 = \sum_{i=1}^k f_i x_i^2 - \left(\sum_{i=1}^k f_i x_i \right)^2$$

- Với mẫu số liệu được cho bởi bảng phân bố tần số, tần suất ghép lớp:

$$s^2 = \frac{1}{N} \sum_{i=1}^k n_i (c_i - \bar{x})^2 = \frac{1}{N} \sum_{i=1}^k n_i c_i^2 - \frac{1}{N^2} \left(\sum_{i=1}^k n_i c_i \right)^2$$
$$= \sum_{i=1}^k f_i (c_i - \bar{x})^2 = \sum_{i=1}^k f_i c_i^2 - \left(\sum_{i=1}^k f_i c_i \right)^2$$

(c_i, n_i, f_i là giá trị đại diện, tần số, tần suất của lớp thứ i ;

N là số các số liệu thống kê $N = n_1 + n_2 + \dots + n_k$)

Chú ý: Phương sai và độ lệch chuẩn càng lớn thì độ phân tán (so với số trung bình) của các số liệu thống kê càng lớn.

B. CÁC DẠNG TOÁN VÀ PHƯƠNG PHÁP GIẢI.

➤ DẠNG TOÁN : XÁC ĐỊNH CÁC SỐ ĐẶC TRƯNG CỦA MẪU SỐ LIỆU .

1. Các ví dụ minh họa.

Ví dụ 1: Cho các số liệu thống kê về sản lượng chè thu được trong 1 năm (kg/sào) của 20 hộ gia đình

111	112	112	113	114	114	115	114	115	116
112	113	113	114	115	114	116	117	113	115

a) Lập bảng phân bố tần số - tần suất

b) Tìm số trung bình, trung vị, mốt

Lời giải

Bảng phân bố tần số - tần suất:

Giá trị x	Tần số	Tần suất (%)
111	1	5
112	3	15
113	4	20
114	5	25
115	4	20
116	2	10
117	1	5
	N=20	100

b) * Số trung bình:

$$\bar{x} = \frac{1}{20} (1.111 + 3.112 + 4.113 + 5.114 + 4.115 + 2.116 + 1.117) = 113,9$$

* Số trung vị: Do kích thước mẫu $N = 20$ là một số chẵn nên số trung vị là trung bình cộng của hai giá trị đứng thứ $\frac{N}{2} = 10$ và $\frac{N}{2} + 1 = 11$ đó là 114 và 114.

Vậy $M_e = 114$

*Mốt: Do giá trị 114 có tần số lớn nhất là 5 nên ta có: $M_o = 114$.

Ví dụ 2: Để khảo sát kết quả thi tuyển sinh môn Toán trong kì thi tuyển sinh đại học năm vừa qua của trường A, người điều tra chọn một mẫu gồm 100 học sinh tham gia kì thi tuyển sinh đó. Điểm môn Toán (thang điểm 10) của các học sinh này được cho ở bảng phân bố tần số sau đây.

Điểm	0	1	2	3	4	5	6	7	8	9	10	
Tần số	1	1	3	5	8	13	19	24	14	10	2	N=100

a) Tìm mốt, số trung vị.

b) Tìm số trung bình, phương sai và độ lệch chuẩn (chính xác đến hàng phần trăm).

Lời giải

a) Ta có giá trị có tần số lớn nhất $M_o = 7$

Kích thước mẫu là số chẵn nên số trung vị là trung bình cộng của hai số đứng giữa

$$\text{Vậy } M_e = \frac{6 + 7}{2} = 6,5$$

b) Ta có số trung bình cộng là

$$\bar{x} = \frac{n_1x_1 + n_2x_2 + \dots + n_kx_k}{N} = \frac{0.1 + 1.1 + 2.3 + \dots + 10.2}{100} = 6,23$$

$$\text{Ta có } \sum_{i=1}^k n_i x_i^2 = 4277, \sum_{i=1}^k n_i x_i = 623$$

$$\text{Suy ra phương sai là } s^2 = \frac{1}{N} \sum_{i=1}^k n_i x_i^2 - \frac{1}{N^2} \left(\sum_{i=1}^k n_i x_i \right)^2 = \frac{4277}{100} - \left(\frac{623}{100} \right)^2 = 3,96$$

Do đó độ lệch chuẩn là $S \approx 1,99$.

Ví dụ 3: Tiền lãi (nghìn đồng) trong 30 ngày được khảo sát ở một quầy bán báo.

81	37	74	65	31	63	58	82	67	77	63	46	30	53	73
51	44	52	92	93	53	85	77	47	42	57	57	85	55	64

a) Hãy lập bảng phân bố tần số và tần suất theo các lớp như sau:

[29,5; 40,5), [40,5; 51,5), [51,5; 62,5), [62,5; 73,5), [73,5; 84,5), [84,5; 95,5]

b) Tính số trung bình cộng, phương sai, độ lệch chuẩn ?

Lời giải

a) Bảng phân bố tần số và tần suất là

Lớp tiền lãi	Tần số	Tần suất
[29,5;40,5)	3	10%
[40,5;51,5)	5	17%
[51,5;62,5)	7	23%
[62,5;73,5)	6	20%
[73,5;84,5)	5	17%
[84,5;95,5]	4	13%
N	30	100%

b) Ta có $\bar{x} = \frac{\sum_{i=1}^N n_i c_i}{N}$ nên

$$\bar{x} = \frac{3.35 + 5.46_2 + 7.57 + 6.68 + 5.79 + 4.90}{30} = 63,23$$

Ta có $\sum_{i=1}^k n_i c_i^2 = 128347$, $\sum_{i=1}^k n_i c_i = 1897$

Suy ra phương sai là

$$s^2 = \frac{1}{N} \sum_{i=1}^k n_i c_i^2 - \frac{1}{N^2} \left(\sum_{i=1}^k c_i x_i \right)^2 = \frac{128347}{100} - \left(\frac{1897}{100} \right)^2 \approx 279,78$$

Do đó độ lệch chuẩn là $S \approx 16,73$.

Ví dụ 4: Cho mẫu số liệu gồm bốn số tự nhiên khác nhau và khác 0, biết số trung bình là 6 và số trung vị là 5. Tìm các giá trị của mẫu số liệu đó sao cho hiệu của giá trị lớn nhất và giá trị nhỏ nhất của mẫu số liệu đạt giá trị nhỏ nhất.

Lời giải

Giả sử các giá trị của mẫu số liệu là a, b, c, d với $0 < a < b < c < d$, $a, b, c, d \in N$

$$\text{Ta có } M_e = \frac{b+c}{2} = 5 \Rightarrow b+c = 10$$

Lớp tiền lãi	Tần số	Giá trị đại diện c_i
[29,5;40,5)	3	35
[40,5;51,5)	5	46
[51,5;62,5)	7	57
[62,5;73,5)	6	68
[73,5;84,5)	5	79
[84,5;95,5]	4	90

Mà $\bar{x} = 6 \Rightarrow a + b + c + d = 24 \Rightarrow a + d = 14$

Ta có $\begin{cases} a < b < c \\ b + c = 10 \end{cases} \Rightarrow \begin{cases} b > 1 \\ 10 > 2b \end{cases}$ hay $1 < b < 5$ mà $b \in N \Rightarrow b \in \{2; 3; 4\}$

• Nếu $b = 2$ thì $c = 8$, mà $0 < a < b, a \in N \Rightarrow a = 1, d = 13$

Khi đó các giá trị của mẫu số liệu là 1;2;8;13

• Nếu $b = 3$ thì $c = 7$, mà $0 < a < b, a \in N \Rightarrow \begin{cases} a = 1 \Rightarrow d = 13 \\ a = 2 \Rightarrow d = 12 \end{cases}$

Khi đó có hai mẫu số liệu thỏa đề bài có giá trị là 1;3;7;13 và 2;3;7;12

• Nếu $b = 4$ thì $c = 6$, mà $0 < a < b, a \in N \Rightarrow \begin{cases} a = 1 \Rightarrow d = 13 \\ a = 2 \Rightarrow d = 12 \\ a = 3 \Rightarrow d = 11 \end{cases}$

Khi đó có ba mẫu số liệu thỏa đề bài có giá trị là 1;4;6;13, 2;4;6;12 và 3;4;6;11

Suy ra với mẫu số liệu có các giá trị là 3;4;6;11 thì hiệu của giá trị lớn nhất và giá trị nhỏ nhất của mẫu số liệu đạt giá trị nhỏ nhất.

3. Bài tập luyện tập.

Bài 5.8: Đo chiều cao (cm) của 40 học sinh nam ở một trường THPT, người ta thu được mẫu số liệu sau:

176	167	165	164	144	176	162	175	149	144
176	166	166	163	156	170	161	176	148	143
175	174	175	146	157	170	165	176	152	142
163	173	175	147	160	170	169	176	168	141

a) Lập bảng phân bố tần số - tần suất ghép lớp theo chiều cao của học sinh với các lớp: [141;146], [147;152], ..., [171;176].

b) Dựa vào bảng phân bố tần số ghép lớp trên, tính chiều cao trung bình, phương sai và độ lệch chuẩn của mẫu số liệu đã cho.

Bài 5.9: Có 100 học sinh tham dự kỳ thi học sinh giỏi môn toán, kết quả được cho trong bảng sau: (thang điểm là 20)

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

a) Tính số trung bình và số trung vị.

b) Tính phương sai và độ lệch chuẩn.

Bài 5.10: Có tài liệu về tuổi nghề của công nhân hai tổ trong một xí nghiệp cơ khí như sau:

Tổ I 2 2 5 7 9 9 9 10 10 11 12

Tổ II 2 3 4 4 4 5 5 7 7 8

Trong mỗi tổ, tính tuổi nghề bình quân, số một và số trung vị?

Bài 5.11: Thống kê điểm kiểm tra toán của lớp 10C, giáo viên bộ môn thu được số liệu:

Điểm	0	1	2	3	4	5	6	7	8	9	10	
Tần số	1	1	1	5	6	7	11	5	4	2	2	N = 45

Tính: Số trung bình, số trung vị, phương sai và độ lệch chuẩn (chính xác đến hàng phần chục)

Bài 5.12: Để được cấp chứng chỉ A- Anh văn của một trung tâm ngoại ngữ, học viên phải trải qua 6 lần kiểm tra trắc nghiệm, thang điểm mỗi lần kiểm tra là 100, và phải đạt điểm trung bình từ 70 điểm trở lên. Qua 5 lần thi Minh đạt điểm trung bình là 64,5 điểm. Hỏi trong lần kiểm tra cuối cùng Minh phải đạt ít nhất là bao nhiêu điểm để được cấp chứng chỉ?

Bài 5.13: Cho hai bảng phân bố tần số mô tả kết quả điểm thi môn Toán của hai lớp 10A và 10B của một trường (Hai lớp làm cùng một đề) như sau:

Bảng 1: Điểm thi của lớp 10A

Điểm	1	3	4	5	6	7	8	
Tần số	1	3	4	8	10	3	1	N=30

Bảng 2: Điểm thi của lớp 10B

Điểm	1	2	3	4	5	6	7	8	9	
Tần số	1	2	3	4	6	7	3	3	1	N=30

a) Tính phương sai của từng bảng.

b) Nhận xét lớp nào có điểm thi môn Toán đồng đều hơn, vì sao?

Bài 5.14: Người ta đã thống kê số gia cầm bị tiêu hủy trong vùng dịch của 6 xã A, B, ..., F như sau (đơn vị: nghìn con):

Xã	A	B	C	D	E	F
Số lượng gia cầm bị tiêu hủy	12	27	22	15	45	5

Tính số trung vị, số trung bình, phương sai và độ lệch chuẩn (chính xác đến hàng trăm) của bảng số liệu thống kê trên

Bài 5.15: Tiến hành một cuộc thăm dò về số cân nặng của mỗi học sinh nữ lớp 10 trường THPT A, người điều tra chọn ngẫu nhiên 30 học sinh nữ lớp 10 và đề nghị các em cho biết số cân nặng của mình. Kết quả thu được ghi lại trong bảng sau (đơn vị là kg):

43	50	43	48	45	40	38	48	45	50	43	45	48	43	38
40	43	48	40	43	45	43	50	40	50	43	45	50	43	45

a) Lập bảng phân bố tần số và tần suất (chính xác đến hàng phần trăm).

b) Tính số trung bình; số trung vị và một.

Bài 5.16: Điểm kiểm tra môn toán của hai học sinh An và Bình được ghi lại như sau :

An	9	8	4	10	3	10	9	7
Bình	6	7	9	5	7	8	9	9

- Tính điểm trung bình của mỗi học sinh .
- Tính phương sai và độ lệch chuẩn về điểm của mỗi học sinh (chính xác đến hàng phần trăm).
- Học sinh nào có kết quả ổn định hơn? Vì sao ?