



CÁC SỐ ĐẶC TRƯNG CỦA MẪU SỐ LIỆU KHÔNG GHÉP NHÓM

BÀI 13. CÁC SỐ ĐẶC TRƯNG ĐO XU THẾ TRUNG TÂM



LÝ THUYẾT.

1. 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.

▶ **Luyện tập 1.** Bảng sau cho biết thời gian chạy cự li 100m của các bạn trong lớp (đơn vị giây):

Thời gian	12	13	14	15	16
Số bạn	5	7	10	8	6

Hãy tính thời gian chạy trung bình cự li 100m của các bạn trong lớp.

Giải: Thời gian chạy trung bình cự li 100m của các bạn trong lớp là

$$\bar{x} = \frac{12 \cdot 5 + 13 \cdot 7 + 14 \cdot 10 + 15 \cdot 8 + 16 \cdot 6}{36} \approx 14,08.$$

b. Trung vị

Để tìm trung vị 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 là giá trị chia đôi mẫu số liệu, nghĩa là trong mẫu số liệu được sắp xếp theo thứ tự không giảm thì giá trị trung vị ở vị trí chính giữa. Trung vị không bị ảnh hưởng bởi giá trị bất thường trong khi số trung bình bị ảnh hưởng bởi giá trị bất thường.

Luyện tập 2. Chiều dài (đơn vị feet) của 7 con cá voi trưởng thành được cho như sau:

48 53 51 31 53 112 52.

Tìm số trung bình và trung vị của mẫu số liệu trên. Trong hai số đó, số nào phù hợp hơn để đại diện cho chiều dài của 7 con cá voi trưởng thành này?

Giải:

+ Chiều dài trung bình của 7 con cá voi trưởng thành là

$$\frac{48+53+51+31+53+112+52}{7} \approx 57,14 \text{ (feet)}.$$

+ Sắp thứ tự dãy số liệu thành dãy không giảm: 31 48 51 52 53 53 112

Trung vị của dãy số là số 52.

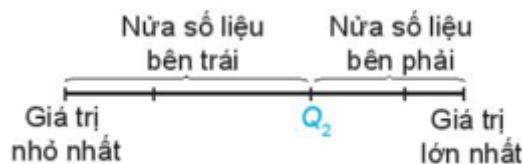
Trong hai số trên, số trung vị phù hợp hơn để đại diện cho chiều dài của 7 con cá voi trưởng thành này.

2. TỨ PHÂN VỊ

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.



Hình 5.3b

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 (hình 5.3a).



Hình 5.3a. Các tứ phân vị



VÍ DỤ: Hàm lượng Natri (đơn vị miligam, $1mg = 0,001g$) trong 100 g một số loại ngũ cốc được cho như sau:

0	340	70	140	200	180	210	150	100	130
140	180	190	160	290	50	220	180	200	210.

Hãy tìm các tứ phân vị. Các phân vị này cho ta thông tin gì?

Giải

- Sắp xếp các giá trị này theo thứ tự không giảm:

0 50 70 100 130 140 140 150 160 180 180 180 190 200 200 210 210 220 290 340.

Hai giá trị chính giữa

- Vì $n = 20$ là số chẵn nên Q_2 là trung bình cộng của hai giá trị chính giữa:

$$Q_2 = (180 + 180) : 2 = 180.$$

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

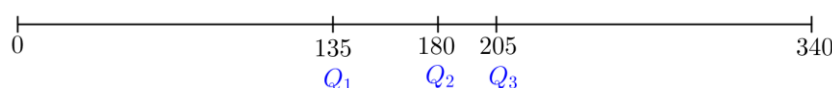
0 50 70 100 130 140 140 150 160 180.

và ta tìm được $Q_1 = (130 + 140) : 2 = 135$.

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

180 180 190 200 200 210 210 220 290 340.

và tìm được $Q_3 = (200 + 210) : 2 = 205$.



Hình 5.4. Hình ảnh về sự phân bố của mẫu số liệu

Các tứ phân vị cho ta hình ảnh phân bố của mẫu số liệu. Khoảng cách từ Q_1 đến Q_2 là 45 trong khi khoảng cách từ Q_2 đến Q_3 là 25. Điều này cho thấy mẫu số liệu tập trung mật độ cao ở bên phải Q_2 và mật độ thấp ở bên trái Q_2 (H.5.4).

Luyện tập 3. Bảng sau đây cho biết số lần học tiếng Anh trên Internet trong một tuần của một số học sinh lớp 10:

Số lần	0	1	2	3	4	5
Số học sinh	2	4	6	12	8	3

Hãy tìm các tứ phân vị cho mẫu số liệu này.

Giải:

- Vì $n = 35$ là số lẻ nên trung vị là số thứ 18: $Q_2 = 3$.



- Bên trái Q_2 có 17 số liệu nên trung vị của nửa này là số thứ 9: $Q_1 = 2$.
- Bên phải Q_2 có 17 số liệu nên trung vị của nửa này là số thứ 27: $Q_3 = 4$.

3. MÔT

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

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



BÀI TẬP SÁCH GIÁO KHOA.

5.7. Tìm số trung bình, trung vị, môđ và tứ phân vị của mỗi mẫu số liệu sau đây:

a) Số điểm mà năm vận động viên bóng rổ ghi được trong một trận đấu:

9 8 15 8 20

b) Giá của một số loại giày (đơn vị nghìn đồng):

350 300 650 300 450 500 300 250.

c) Số kênh được chiếu của một số hãng truyền hình cáp:

36 38 33 34 32 30 34 35.

Giải

a) Số trung bình là $\frac{8.2 + 9 + 15 + 20}{5} = 12$.

Sắp xếp số liệu theo thứ tự không giảm: 8 8 9 15 20.

Trung vị là 9.

Số 8 xuất hiện nhiều nhất nên môđ là 8.

Tứ phân vị $Q_1 = 8$; $Q_2 = 9$; $Q_3 = 17.5$.

b) Số trung bình là $\frac{250 + 300.3 + 350 + 450 + 500 + 650}{8} = 387.5$.

Sắp xếp số liệu theo thứ tự không giảm: 250 300 300 300 350 450 500 650.

Trung vị là 325.

Môđ là 300.

Tứ phân vị $Q_1 = 300$; $Q_2 = 325$; $Q_3 = 475$.

c) Số trung bình là $\frac{30 + 32 + 33 + 34.2 + 35 + 36 + 38}{8} = 34$.

Sắp xếp số liệu theo thứ tự không giảm: 30 32 33 34 34 35 36 38.

Trung vị là 34.

Môđ là 34.

Tứ phân vị $Q_1 = 32.5$; $Q_2 = 34$; $Q_3 = 35.5$.



5.8. Hãy chọn số đặc trưng đo xu thế trung tâm của mỗi mẫu số liệu sau. Giải thích và tính giá trị của số đặc trưng đó.

a) Số mặt trăng đã biết của các hành tinh:

Hành tinh	Thu ý tinh	Kim tinh	Trái Đ ất	Hoả tinh	Mộc tinh	Thổ t i n h	Thiên Vương tinh	Hải Vươn g tinh
Số mặt trăng	0	0	1	2	63	34	27	13

(Theo NASA)

b) Số đường chuyền thành công trong một trận đấu của một số cầu thủ bóng đá:

32 24 20 14 23.

c) Chỉ số IQ của một nhóm học sinh: 60 72 63 83 68 74 90 86 74 80.

d) Các sai số trong phép đo: 10 15 18 15 14 13 42 15 12 14 42.

Giải

a) Chọn số đặc trưng là tứ phân vị, vì các số liệu không đồng đều nhau, nhiều số liệu trong mẫu chênh lệch lớn so với trung vị.

Sắp xếp số liệu theo thứ tự không giảm 0 0 1 2 13 27 34 63 .

Tứ phân vị $Q_1 = 0.5$; $Q_2 = 7.5$; $Q_3 = 30.5$.

b) Chọn số đặc trưng là số trung bình, các giá trị không lặp lại.

Số trung bình là $\frac{32 + 24 + 20 + 14 + 23}{5} = 22.6$.

c) Chọn số đặc trưng là trung bình, vì các số liệu gần nhau. Số trung bình là:

$\frac{60 + 63 + 68 + 72 + 74 + 80 + 83 + 86 + 90}{10} = 75$.

d) Chọn số đặc trưng là trung vị, vì có số 42 lớn bất thường. Trung vị là 15.

5.9. Số lượng học sinh giỏi Quốc gia năm học 2018 - 2019 của 10 trường Trung học phổ thông được cho như sau: 0 0 4 0 0 0 10 0 6 0.

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

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

Giải

a) Số trung bình là $\frac{0.7 + 4 + 6 + 10}{10} = 2$.

Sắp xếp số liệu theo thứ tự không giảm 0 0 0 0 0 0 0 4 6 10.

Số 0 xuất hiện nhiều nhất nên mốt là 0 .



Tứ phân vị $Q_1 = 0$; $Q_2 = 0$; $Q_3 = 4$.

- b) Tứ phân vị thứ nhất và trung vị trùng nhau do mẫu có 10 số liệu mà số 0 đã xuất hiện 7 lần.

5.10. Bảng sau đây cho biết số chỗ ngồi của một số sân vận động được sử dụng trong Giải Bóng đá Vô địch Quốc gia Việt Nam năm 2018 (số liệu gần đúng).

Sân vận động	Cẩm phả	Thiên Trường	Hàng Đẫy	Thanh Hoá	Mỹ Đình
Chỗ ngồi	20 120	21 315	23 405	20 120	37 546

(Theo vov.vn)

Các giá trị số trung bình, trung vị, một bị ảnh hưởng như thế nào nếu bỏ đi số liệu chỗ ngồi của Sân vận động Quốc gia Mỹ Đình?

Giải:

Số trung bình là $\frac{20120 + 21315 + 23405 + 20120 + 37546}{5} = 24501.2$.

Sắp xếp số liệu theo thứ tự không giảm 20120 20120 21315 23405 37546.

một là 20120.

Trung vị 21315.

Nếu bỏ số liệu chỗ ngồi của Sân vận động Quốc gia Mỹ Đình

Số trung bình là $\frac{20120 + 21315 + 23405 + 20120}{4} = 21240$.

Sắp xếp số liệu theo thứ tự không giảm 20120 20120 21315 23405.

một là 20120.

Trung vị 20717.5.

Vậy nếu bỏ số liệu chỗ ngồi của Sân vận động Quốc gia Mỹ Đình thì một giữ nguyên, số trung bình và trung vị sẽ thay đổi.



HỆ THỐNG BÀI TẬP TỰ LUẬN

PHƯƠNG PHÁP

Để xác định các số đặc trưng đo xu thế trung tâm ta làm như sau:

- ♦ Liệt kê số liệu theo thứ tự không giảm.
- ♦ Tính số trung bình, trung vị, một, tứ phân vị theo yêu cầu của bài toán.
- ♦ Kết luận về mẫu số liệu

Câu 1: Bảng số liệu sau cho biết số lượng học sinh đạt học lực giỏi năm học 2022 – 2023 của một số tỉnh

Tỉnh	Thanh Hóa	Hà Nội	Hải Phòng	Bạc Liêu	Nam Định	Thái Bình	Quảng Ninh	Bình Phước	Bình Dương
------	-----------	--------	-----------	----------	----------	-----------	------------	------------	------------



Số lượng HS	2324	4305	3524	2324	4322	3846	4305	2542	4305
-------------	------	------	------	------	------	------	------	------	------

Giá trị về số trung bình, trung vị thay đổi thế nào nếu bỏ đi số lượng học sinh giỏi của tỉnh Bạc Liêu và Thái Bình

Lời giải

$$\text{Số trung bình} = \frac{2324 + 4305 + 3524 + 2324 + 4322 + 3846 + 4305 + 2542 + 4305}{9} = 3533$$

Sắp xếp theo thứ tự không giảm:

2324 2324 2542 3524 3846 4305 4305 4305 4322

Trung vị là 3846

Khi bỏ đi số lượng học sinh giỏi của hai tỉnh Bạc Liêu và Thái Bình thì:

$$\text{Số trung bình} = \frac{2324 + 4305 + 3524 + 4322 + 4305 + 2542 + 4305}{7} = \frac{21305}{7}$$

Sắp xếp theo thứ tự không giảm:

2324 2542 3524 4305 4305 4305 4322

Trung vị là 4305

Vậy khi bỏ đi số lượng học sinh giỏi của hai tỉnh Bạc Liêu và Thái Bình thì cả số trung bình và trung vị đều thay đổi.

Câu 2: Hãy chọn số đặc trưng đo xu thế trung tâm của mỗi mẫu số liệu sau giải thích và tính giá trị của số đặc trưng đó với bảng số liệu sau cho biết số lượng học sinh đạt điểm 10 ở các môn học của một trường THPT.

Môn	Toán	Lý	Hóa	Sinh	Anh	Địa	Sử
Số lượng HS	80	70	30	25	5	7	2

Lời giải

Chọn số đặc trưng là tứ phân, vì các số liệu không đều nhau, nhiều số liệu trong mẫu chênh lệch lớn so với trung vị

Sắp xếp theo thứ tự không giảm

2 5 7 25 30 70 80

Tứ phân vị $Q_1 = 5, Q_2 = 25, Q_3 = 70$.

Câu 3: Số lượng học sinh đăng kí xét tuyển đại học năm học 2022 – 2023 của một số tỉnh như sau: 9000 12000 7000 13200 14120 23003 33422 11234 12037 7564
Hãy tìm các tứ phân vị. Các tứ phân vị cho ta thông tin gì?

Lời giải

Sắp xếp theo thứ tự không giảm:

7000 7564 9000 11234 12000 12037 13200 14120 23003 33422



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

$$Q_2 = \frac{(1200+12037)}{2} = 12018,5$$

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

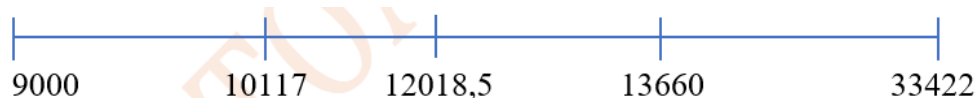
7000 7564 9000 11234 12000 12018,5

Và tìm được $Q_1 = \frac{(9000+11234)}{2} = 10117$

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

12018,5 12037 13200 14120 23003 33422

và tìm được $Q_3 = \frac{(13200+14120)}{2} = 13660$



Các tứ phân vị cho ta hình ảnh phân bố của mẫu số liệu. Khoảng cách từ Q_1 đến Q_2 là 1901,5 trong khi khoảng cách từ Q_2 đến Q_3 là 1641,5 điều này cho thấy mẫu số liệu tập chung cao hơn ở bên trái Q_2 và thấp hơn ở bên phải Q_2 .

