



KẾT NỐI TRI THỨC
VỚI CUỘC SỐNG

LỚP TOÁN THẦY CƯ- TP HUẾ

CS 1: Trung tâm MASTER EDUCATION- 25 THẠCH HẪN

CS 2: Trung Tâm 133 Xuân 68

CS 3: Trung tâm 168 Mai Thúc Loan

CS4: Trung Tâm THPT Nguyễn Trường Tộ

BÀI GIẢNG CHƯƠNG 3 TOÁN 11-KNTT

CÁC ĐẶC TRƯNG ĐO XU THẾ TRUNG TÂM



TÀI LIỆU DÀNH CHO HỌC SINH LỚP TOÁN THẦY CƯ-TP HUẾ

(Chiêu sinh thường xuyên, bổ trợ kiến thức kịp thời)

CHƯƠNG III: CÁC SỐ ĐẶC TRƯNG ĐO XU THẾ TRUNG TÂM CỦA MẪU SỐ LIỆU GHÉP NHÓM

BÀI 8: MẪU SỐ LIỆU GHÉP NHÓM

A. TÓM TẮT KIẾN THỨC CƠ BẢN CẦN NẮM

1. GIỚI THIỆU VỀ MẪU SỐ LIỆU GHÉP NHÓM

Mẫu số liệu ghép nhóm là mẫu số liệu cho dưới dạng bảng tần số của các nhóm số liệu. Mỗi nhóm số liệu là tập hợp gồm các giá trị của số liệu được ghép nhóm theo một tiêu chí xác định. Nhóm số liệu thường được cho dưới dạng [a; b), trong đó a là đầu mút trái, b là đầu mút phải.

Nhận xét.

- Mẫu số liệu ghép nhóm được dùng khi ta không thể thu thập được số liệu chính xác hoặc do yêu cầu của bài toán mà ta phải biểu diễn mẫu số liệu dưới dạng ghép nhóm để thuận lợi cho việc tổ chức, đọc và phân tích số liệu.

- Trong một số trường hợp, nhóm số liệu cuối cùng có thể lấy đầu mút bên phải.

Ví dụ 1. Mẫu số liệu sau cho biết phân bố theo độ tuổi của dân số Việt Nam năm 2019.

Độ tuổi	Dưới 15	Từ 15 đến dưới 65 tuổi	Từ 65 tuổi trở lên
Số người	23 371 882	65 420 451	7 416 651

(Theo: Báo cáo số liệu tổng điều tra dân số năm 2019)

a) Mẫu số liệu đã cho có là mẫu số liệu ghép nhóm hay không?

b) Nêu các nhóm và tần số tương ứng. Dân số Việt Nam năm 2019 là bao nhiêu?

Lời giải

a) Mẫu số liệu đã cho là mẫu số liệu ghép nhóm.

b) Có ba nhóm là: Dưới 15 tuổi, Từ 15 đến dưới 65 tuổi, Từ 65 tuổi trở lên.

Có 23371882 người dưới 15 tuổi; 65420451 người từ 15 đến dưới 65 tuổi và 7416651 người từ 65 tuổi trở lên.

Dân số Việt Nam năm 2019 là $23371882 + 65420451 + 7416651 = 96208984$ người.

Luyện tập 1. Cho mẫu số liệu ghép nhóm về thời gian (phút) đi từ nhà đến nơi làm việc của các nhân viên một công ty như sau:

Thời gian	[15; 20)	[20; 25)	[25; 30)	[30; 35)	[35; 40)	[40; 45)	[45; 50)
Số nhân viên	6	14	25	37	21	13	9

Đọc và giải thích mẫu số liệu này.

Lời giải

Mẫu số liệu được chia thành 7 nhóm

- Có 6 nhân viên có thời gian đi từ nhà đến nơi làm việc là từ 15 đến dưới 20 phút

- Có 14 nhân viên có thời gian đi từ nhà đến nơi làm việc là từ 20 đến dưới 25 phút

- Có 25 nhân viên có thời gian đi từ nhà đến nơi làm việc là từ 25 đến dưới 30 phút

- Có 37 nhân viên có thời gian đi từ nhà đến nơi làm việc là từ 30 đến dưới 35 phút
- Có 21 nhân viên có thời gian đi từ nhà đến nơi làm việc là từ 35 đến dưới 40 phút
- Có 13 nhân viên có thời gian đi từ nhà đến nơi làm việc là từ 40 đến dưới 45 phút
- Có 9 nhân viên có thời gian đi từ nhà đến nơi làm việc là từ 45 đến dưới 50 phút

2. GHÉP NHÓM MẪU SỐ LIỆU

Để chuyển mẫu số liệu không ghép nhóm sang mẫu số liệu ghép nhóm, ta làm như sau:

Bước 1. Chia miền giá trị của mẫu số liệu thành một số nhóm theo tiêu chí cho trước.

Bước 2. Đếm số giá trị của mẫu số liệu thuộc mỗi nhóm (tần số) và lập bảng thống kê cho mẫu số liệu ghép nhóm.

Chú ý:

- Độ dài của nhóm $[a ; b)$ là $b - a$
- Không nên chia thành quá nhiều nhóm hoặc quá ít nhóm. Các nhóm không giao nhau, các nhóm nên có độ dài như nhau và tổng độ dài các nhóm lớn hơn khoảng biến thiên.

Ví dụ 2. Bảng thống kê sau cho biết thời gian chạy (phút) của 30 vận động viên (VĐV) trong một giải chạy Marathon.

Thời gian	129	130	133	134	135	136	138	141	142	143	144	145
Số VĐV	1	2	1	1	1	2	3	3	4	5	2	5

Hãy chuyển mẫu số liệu trên sang mẫu số liệu ghép nhóm gồm sáu nhóm có độ dài bằng nhau và bằng 3.

Lời giải

Giá trị nhỏ nhất là 129 , giá trị lớn nhất là 145 nên khoảng biến thiên là $145 - 129 = 16$.

Tổng độ dài của sáu nhóm là 18. Để cho đối xứng, ta chọn đầu mút trái của nhóm đầu tiên là 127,5 và đầu mút phải của nhóm cuối cùng là 145,5 ta được các nhóm là $[127,5;130,5)$, $[130,5;133,5)$,..., $[142,5;145,5)$. Đếm số giá trị thuộc mỗi nhóm, ta có mẫu số liệu ghép nhóm như sau:

Thời gian	$[127,5; 130,5)$	$[130,5; 133,5)$	$[133,5; 136,5)$	$[136,5; 139,5)$	$[139,5; 142,5)$	$[142,5; 145,5)$
Số VĐV	3	1	4	3	7	12

Luyện tập 2. Cân nặng (kg) của 35 người trưởng thành tại một khu dân cư được cho như sau:

43 51 47 62 48 40 50 62 53 56 40 48 56 53 50 42 55
 52 48 46 45 54 52 50 47 44 54 55 60 63 58 55 60 58 53.

Chuyển mẫu số liệu trên thành dạng ghép nhóm, các nhóm có độ dài bằng nhau, trong đó có nhóm $[40; 45)$.

Lời giải

Cân nặng	$[40;45)$	$[45;50)$	$[50;55)$	$[55;60)$	$[60;65)$
Số người	5	7	11	7	5

GV: TRẦN ĐÌNH CỬ - 0834332133

Vận dụng. Một công ty may quần áo đồng phục học sinh cho biết cỡ áo theo chiều cao của học sinh được tính như sau:

Chiều cao (cm)	[150;160)	[160; 167)	[167; 170)	[170; 175)	[175; 180)
Cỡ áo	S	M	L	XL	XXL

Công ty muốn ước lượng tỉ lệ các cỡ áo khi may cho học sinh lớp 11 đã đo chiều cao của 36 học sinh nam khối 11 của một trường và thu được mẫu số liệu sau (đơn vị là centimét):

160 161 161 162 162 162 163 163 163 164 164 164 164
 165 165 165 165 165 166 166 166 166 167 167 168 168
 168 168 169 169 170 171 171 172 172 174.

- a) Lập bảng tần số ghép nhóm của mẫu số liệu với các nhóm đã cho ở bảng trên.
- b) Công ty may 500 áo đồng phục cho học sinh lớp 11 thì nên may số lượng áo theo mỗi cỡ là bao nhiêu chiếc?

Lời giải

a)

Chiều cao (cm)	[160;167)	[167;170)	[170;175)
Số học sinh	22	8	6

b) Tỉ lệ học sinh mặc vừa cỡ M là: $(22 : 36) \times 100 = 61.11\%$

Tỉ lệ học sinh mặc vừa cỡ L là: $(8 : 36) \times 100 = 22.22\%$

Tỉ lệ học sinh mặc vừa cỡ XL là: $(6 : 36) \times 100 = 16.67\%$

Số lượng áo cỡ M nên may là: $61.11\% \times 500 = 306$

Số lượng áo cỡ L nên may là: $22.22\% \times 500 = 111$

Số lượng áo cỡ XL nên may là: $16.67\% \times 500 = 83$

B. GIẢI BÀI TẬP SÁCH GIÁO KHOA

Bài 3.1. Trong các mẫu số liệu sau, mẫu nào là mẫu số liệu ghép nhóm? Đọc và giải thích mẫu số liệu ghép nhóm đó.

- a) Số tiền mà sinh viên chi cho thanh toán cước điện thoại trong tháng.

Số tiền (nghìn đồng)	[0; 50)	[50; 100)	[100; 150)	[150; 200)	[200; 250)
Số sinh viên	5	12	23	17	3

- b) Thống kê nhiệt độ tại một địa điểm trong 40 ngày, ta có bảng số liệu sau:

Nhiệt độ (°C)	[19; 22)	[22; 25)	[25; 28)	[28; 31)
Số ngày	7	15	12	6

Lời giải

Cả hai mẫu số liệu đều là mẫu số liệu ghép nhóm

a) Có 5 sinh viên chi dưới 50 nghìn đồng cho việc thanh toán cước điện thoại trong tháng
 Có 12 sinh viên chi từ 50 đến dưới 100 nghìn đồng cho việc thanh toán cước điện thoại trong tháng
 Có 23 sinh viên chi từ 100 đến dưới 150 nghìn đồng cho việc thanh toán cước điện thoại trong tháng
 Có 17 sinh viên chi từ 150 đến dưới 200 nghìn đồng cho việc thanh toán cước điện thoại trong tháng
 Có 3 sinh viên chi từ 200 đến dưới 250 nghìn đồng cho việc thanh toán cước điện thoại trong tháng
 Như vậy, đa số sinh viên chi từ 100 đến dưới 150 nghìn đồng mỗi tháng cho cước điện thoại và có ít sinh viên chi trên 200 nghìn đồng cho cước điện thoại mỗi tháng

b) Có 7 ngày có nhiệt độ từ 19°C đến dưới 22°C

Có 15 ngày có nhiệt độ từ 22°C đến dưới 25°C

Có 12 ngày có nhiệt độ từ 25°C đến dưới 28°C

Có 6 ngày có nhiệt độ từ 28°C đến dưới 31°C

Bài 3.2. Số sản phẩm một công nhân làm được trong một ngày được cho như sau:

18 25 39 12 54 27 46 25 19 8 36 22
 20 19 17 44 5 18 23 28 25 34 46 27 16.

Hãy chuyển mẫu số liệu sang dạng ghép nhóm với sáu nhóm có độ dài bằng nhau.

Lời giải

Giá trị nhỏ nhất là: 5

Giá trị lớn nhất là 54

Do đó khoảng biến thiên là $54 - 5 = 49$

Để chia thành 6 nhóm với độ dài bằng nhau ta lấy điểm đầu mút phải trái của nhóm đầu tiên là 3 và đầu mút phải của nhóm cuối cùng là 57 với độ dài mỗi nhóm là 9 .

Ta được mẫu số liệu ghép nhóm như sau:

Số sản phẩm	[3;12)	[12;21)	[21;30)	[30;39)	[39;48)	[48;57)
Số công nhân	2	8	8	2	4	1

Bài 3.3. Thời gian ra sân (giờ) của một số cựu cầu thủ ở giải ngoại hạng Anh qua các thời kì được cho như sau:

653 632 609 572 565 535 516 514 508 505 504 504 503 499 496 492.

(Theo: <https://www.premierleague.com/>)

Hãy chuyển mẫu số liệu trên sang dạng ghép nhóm với bảy nhóm có độ dài bằng nhau.

Lời giải

Giá trị lớn nhất là: 653

Giá trị bé nhất là: 492

Khoảng biến thiên là: $653 - 492 = 161$

Để chia thành 7 nhóm có độ dài bằng nhau, ta lấy điểm đầu mút trái của nhóm đầu tiên là 492, điểm đầu mút phải của nhóm cuối là 653 với độ dài mỗi nhóm là 23

Ta có mẫu số liệu ghép nhóm sau:

Thời gian ra sân (giờ)	[492;515)	[515;538)	[538;561)	[561;584)	[584;607)	[607;630)	[630;653]
Số cầu thủ	9	2	0	2	0	1	2

C. LUYỆN TẬP

Câu 1: Độ dài của 60 lá dương xỉ trưởng thành được cho bằng bảng phân bố tần số ghép lớp như sau.

Số TT	Lớp của độ dài (cm)	Tần số
1	[10;20)	8
2	[20;30)	18
3	[30;40)	24
4	[40;50)	10
	Cộng	60

Hỏi số lá có chiều dài từ 30cm đến 50cm chiếm bao nhiêu phần trăm?

- A.** 50%. **B.** 56%. **C.** 56,7%. **D.** 57%.

Lời giải

Chọn C

Ta có tần suất của lớp [30;40) là $\frac{24 \times 100}{60} = 40$.

Tần suất của lớp [40;50) là $\frac{10 \times 100}{60} \approx 16,7$.

Vậy số lá có chiều dài từ 30cm đến 50cm chiếm $40\% + 16,7\% = 56,7\%$.

Câu 2: Khối lượng của 30 củ khoai tây thu hoạch ở một nông trường được thống kê như bảng sau.

Lớp khối lượng (gam)	Tần số
[70;80)	3
[80;90)	6
[90;100)	12
[100;110)	6
[110;120)	3
Cộng	30

Tần suất ghép nhóm của lớp [100;110) là.

- A.** 20%. **B.** 40%. **C.** 60%. **D.** 80%.

Lời giải

Chọn A

Ta có tần suất ghép lớp của lớp [100;110) là $\frac{6 \times 100\%}{30} = 20\%$.

Câu 3: Cho các số liệu thống kê ghi trong bảng sau

Thành tích chạy 50m của học sinh lớp 10A ở trườngng Trung học phổ thông C (đơn vị : giây)

6,3	6,2	6,5	6,8	6,9	8,2	8,6
6,6	6,7	7,0	7,1	7,2	8,3	8,5
7,4	7,3	7,2	7,1	7,0	8,4	8,1
7,1	7,3	7,5	7,5	7,6	8,7	
7,6	7,7	7,8	7,5	7,7	7,8	

Bảng 1

a) Lập bảng phân bố tần số ghép nhóm và bảng phân bố tần suất ghép nhóm, với các nhóm $[6,0;6,5); [6,5;7,0); [7,0;7,5); [7,5;8,0); [8,0;8,5); [8,5;9,0]$.

b) Trong lớp 10A, số học sinh chạy 50m hết từ 7 giây đến dưới 8,5 giây chiếm bao nhiêu phần trăm ?

Giải

a) Từ các số liệu thống kê đã cho, ta xác định được :

Tần số của các lớp

$$n_1 = 2; \quad n_2 = 5; \quad n_3 = 10;$$

$$n_4 = 9; \quad n_5 = 4; \quad n_6 = 3.$$

Tần suất của các lớp

$$f_1 \approx 6,06\%; \quad f_2 \approx 15,15\%; \quad f_3 \approx 30,30\%;$$

$$f_4 \approx 27,27\%; \quad f_5 \approx 12,12\%; \quad f_6 = 9,10\%.$$

Từ đó ta có bảng phân bố tần số ghép nhóm

Thành tích chạy 50 m của học sinh lớp 10 A ở trường Trung học phổ thông C

Lớp thời gian chạy (giây)	Tần suất (%)
$[6,0 ; 6,5)$	6,06
$[6,5 ; 7,0)$	15,15
$[7,0 ; 7,5)$	30,30
$[7,5 ; 8,0)$	27,27
$[8,0 ; 8,5)$	12,12
$[8,5 ; 9,0]$	9,10
Cộng	100 (%)

b) $30,30\% + 27,27\% + 12,12\% = 69,69\%$.

Trả lời : 69,69%.

BÀI 9: CÁC SỐ ĐẶC TRƯNG ĐO XU THỂ TRUNG TÂM

A. TÓM TẮT KIẾN THỨC CƠ BẢN CẦN NẮM

1. SỐ TRUNG BÌNH CỦA MẪU SỐ LIỆU GHÉP NHÓM

Số trung bình của mẫu số liệu ghép nhóm kí hiệu là \bar{x} .

$$\bar{x} = \frac{m_1x_1 + \dots + m_kx_k}{n}$$

trong đó, $n = m_1 + \dots + m_k$ là cỡ mẫu và $x_i = \frac{a_i + a_{i+1}}{2}$ (với $i = 1, \dots, k$) là giá trị đại diện của nhóm $[a_i; a_{i+1})$

Ví dụ 1. Tìm cân nặng trung bình của học sinh lớp 11D cho trong Bảng 3.5.

Cân nặng	[40,5; 45,5)	[45,5; 50,5)	[50,5; 55,5)	[55,5; 60,5)	[60,5; 65,5)	[65,5; 70,5)
Số học sinh	10	7	16	4	2	3

Bảng 3.5. Cân nặng của học sinh lớp 11D

Lời giải

Trong mỗi khoảng cân nặng, giá trị đại diện là trung bình cộng của giá trị hai đầu mút nên ta có bảng sau:

Cân nặng (kg)	43	48	53	58	63	68
Số học sinh	10	7	16	4	2	3

Tổng số học sinh là $n = 42$. Cân nặng trung bình của học sinh lớp 11D là

$$\bar{x} = \frac{10 \cdot 43 + 7 \cdot 48 + 16 \cdot 53 + 4 \cdot 58 + 2 \cdot 63 + 3 \cdot 68}{42} \approx 51,81(\text{kg}).$$

Luyện tập 1. Tìm hiểu thời gian xem ti vi trong tuần trước (đơn vị: giờ) của một số học sinh thu được kết quả sau:

Thời gian (giờ)	[0; 5)	[5; 10)	[10; 15)	[15; 20)	[20; 25)
Số học sinh	8	16	4	2	2

Tính thời gian xem ti vi trung bình trong tuần trước của các bạn học sinh này.

Lời giải

Thời gian (giờ)	[0;5)	[5;10)	[10;15)	[15;20)	[20;25)
Số học sinh	8	16	4	2	2
Giá trị đại diện	2.5	7.5	12.5	17.5	22.5

Thời gian xem ti vi trung bình trong tuần trước của các bạn học sinh này là:

$$\bar{x} = \frac{8 \times 2.5 + 16 \times 7.5 + 4 \times 12.5 + 2 \times 17.5 + 2 \times 22.5}{8 + 16 + 4 + 2 + 2} = 8.4375 \text{ (giờ)}$$

Ý nghĩa. Số trung bình của mẫu số liệu ghép nhóm xấp xỉ cho số trung bình của mẫu số liệu gốc, nó cho biết vị trí trung tâm của mẫu số liệu và có thể dùng để đại diện cho mẫu số liệu.

2. TRUNG VỊ CỦA MẪU SỐ LIỆU GHÉP LỚP

Để tính trung vị của mẫu số liệu ghép nhóm, ta làm như sau:

Bước 1. Xác định nhóm chứa trung vị. Giả sử đó là nhóm thứ $p : [a_p; a_{p+1})$.

Bước 2. Trung vị là $M_e = a_p + \frac{\frac{n}{2} - (m_1 + \dots + m_{p-1})}{m_p} \cdot (a_{p+1} - a_p)$,

trong đó n là cỡ mẫu, m_p là tần số nhóm p . Với $p = 1$, ta quy ước $m_1 + \dots + m_{p-1} = 0$.

Ví dụ 2. Thời gian (phút) truy cập Internet mỗi buổi tối của một số học sinh được cho trong bảng sau:

Thời gian (phút)	[9,5; 12,5)	[12,5; 15,5)	[15,5; 18,5)	[18,5; 21,5)	[21,5; 24,5)
Số học sinh	3	12	15	24	2

Tính trung vị của mẫu số liệu ghép nhóm này.

Lời giải

Cỡ mẫu là $n = 3 + 12 + 15 + 24 + 2 = 56$.

Gọi x_1, \dots, x_{56} là thời gian vào Internet của 56 học sinh và giả sử dãy này đã được sắp xếp theo thứ tự tăng dần.

Khi đó, trung vị là $\frac{x_{28} + x_{29}}{2}$. Do 2 giá trị x_{28}, x_{29} thuộc nhóm $[15,5; 18,5)$ nên nhóm này chứa trung vị.

Do đó, $p = 3; a_3 = 15,5; m_3 = 15; m_1 + m_2 = 3 + 12 = 15; a_4 - a_3 = 3$ và ta có

$$M_e = 15,5 + \frac{\frac{56}{2} - 15}{15} \cdot 3 = 18,1$$

Luyện tập 2. Ghi lại tốc độ bóng trong 200 lần giao bóng của một vận động viên môn quần vợt cho kết quả như bảng bên. Tính trung vị của mẫu số liệu ghép nhóm này.

Tốc độ v (km/h)	Số lần
$150 \leq v < 155$	18
$155 \leq v < 160$	28
$160 \leq v < 165$	35
$165 \leq v < 170$	43
$170 \leq v < 175$	41
$175 \leq v < 180$	35

Lời giải

Cỡ mẫu là: $n = 18 + 28 + 35 + 43 + 41 + 35 = 200$

Gọi x_1, x_2, \dots, x_{200} là tốc độ giao bóng của 200 lần và giả sử dãy này được sắp xếp theo thứ tự tăng dần. Khi đó trung vị là $\frac{x_{100} + x_{101}}{2}$.

Do hai giá trị x_{100}, x_{101} thuộc nhóm $[165; 170)$ nên nhóm này chứa trung vị.

Suy ra, $p = 4; a_4 = 165; m_4 = 43; m_1 + m_2 + m_3 = 18 + 28 + 35 = 81; a_5 - a_4 = 5$ và ta có

$$M_e = 165 + \frac{\frac{200}{2} - 81}{43} \times 5 = 167.21$$

Ý nghĩa. Trung vị của mẫu số liệu ghép nhóm xấp xỉ cho trung vị của mẫu số liệu gốc, nó chia mẫu số liệu thành hai phần, mỗi phần chứa 50 % giá trị.

Để tính **tứ phân vị thứ nhất** Q_1 của mẫu số liệu ghép nhóm, trước hết ta xác định nhóm chứa Q_1 , giả sử đó là nhóm thứ $p : [a_p; a_{p+1})$. Khi đó,

$$Q_1 = a_p + \frac{\frac{n}{4} - (m_1 + \dots + m_{p-1})}{m_p} \cdot (a_{p+1} - a_p),$$

trong đó, n là cỡ mẫu, m_p là tần số nhóm p , với $p = 1$ ta quy ước $m_1 + \dots + m_{p-1} = 0$.

Để tính **tứ phân vị thứ ba** Q_3 của mẫu số liệu ghép nhóm, trước hết ta xác định nhóm chứa Q_3 . Giả sử đó là nhóm thứ $p : [a_p; a_{p+1})$. Khi đó,

$$Q_3 = a_p + \frac{\frac{3n}{4} - (m_1 + \dots + m_{p-1})}{m_p} \cdot (a_{p+1} - a_p),$$

trong đó, n là cỡ mẫu, m_p là tần số nhóm p , với $p = 1$ ta quy ước $m_1 + \dots + m_{p-1} = 0$.

Tứ phân vị thứ hai Q_2 chính là trung vị M_e .

Vi dụ 3. Tìm tứ phân vị thứ nhất Q_1 và tứ phân vị thứ ba Q_3 của mẫu số liệu ghép nhóm cho trong Ví dụ 2.

Lời giải

Cỡ mẫu là $n = 56$.

Tứ phân vị thứ nhất Q_1 là $\frac{x_{14} + x_{15}}{2}$. Do x_{14}, x_{15} đều thuộc nhóm $[12,5; 15,5)$ nên nhóm này chứa Q_1 .

Do đó, $p = 2; a_2 = 12,5; m_2 = 12; m_1 = 3, a_3 - a_2 = 3$ và ta có

$$Q_1 = 12,5 + \frac{\frac{56}{4} - 3}{12} \cdot 3 = 15,25$$

Với tứ phân vị thứ ba Q_3 là $\frac{x_{42} + x_{43}}{2}$. Do x_{42}, x_{43} đều thuộc nhóm $[18,5;21,5)$ nên nhóm này chứa Q_3 .

Do đó, $p = 4; a_4 = 18,5; m_4 = 24; m_1 + m_2 + m_3 = 3 + 12 + 15 = 30; a_5 - a_4 = 3$ và ta có

$$Q_3 = 18,5 + \frac{\frac{3.56}{4} - 30}{24} \cdot 3 = 20$$

Nhận xét. Ta cũng có thể xác định nhóm chứa tứ phân vị thứ r nhờ tính chất: có khoảng $\left(\frac{r \cdot n}{4}\right)$ giá trị nhỏ hơn tứ phân vị này.

Luyện tập 3. Tìm tứ phân vị thứ nhất và tứ phân vị thứ ba cho mẫu số liệu ghép nhóm ở Luyện tập 2.

Tốc độ v (km/h)	Số lần
$150 \leq v < 155$	18
$155 \leq v < 160$	28
$160 \leq v < 165$	35
$165 \leq v < 170$	43
$170 \leq v < 175$	41
$175 \leq v < 180$	35

Lời giải

Cỡ mẫu: $n = 200$

Tứ phân vị thứ nhất Q_1 là $\frac{x_{50} + x_{51}}{2}$. Do x_{50}, x_{51} đều thuộc nhóm $[160;165)$ nên tứ phân vị thứ nhất thuộc nhóm $[160;165)$. Do đó, $p = 3; a_3 = 160; m_3 = 35; m_1 + m_2 = 18 + 28 = 46; a_4 - a_3 = 5$ và ta có:

$$Q_1 = 160 + \frac{\frac{200}{4} - 46}{35} \times 5 = 160.57$$

Tứ phân vị thứ ba Q_3 là $\frac{x_{150} + x_{151}}{2}$. Do x_{150}, x_{151} đều thuộc nhóm $[170;175)$ nên tứ phân vị thứ ba thuộc nhóm $[170;175)$. Do đó,

$$p = 5; a_5 = 170; m_5 = 41; m_1 + m_2 + m_3 + m_4 = 18 + 28 + 35 + 43 = 124; a_6 - a_5 = 5$$

và ta có:

$$Q_3 = 170 + \frac{\frac{600}{4} - 124}{41} \times 5 = 173.17$$

Ý nghĩa. Các tứ phân vị của mẫu số liệu ghép nhóm xấp xỉ cho các tứ phân vị của mẫu số liệu gốc, chúng chia mẫu số liệu thành 4 phần, mỗi phần chứa 25% giá trị.

4. MỘT CỬA MẪU SỐ LIỆU GHÉP NHÓM

GV: TRẦN ĐÌNH CỬ - 0834332133

Để tìm một của mẫu số liệu ghép nhóm, ta thực hiện theo các bước sau:

Bước 1. Xác định nhóm có tần số lớn nhất (gọi là nhóm chứa một), giả sử là nhóm $j: [a_j; a_{j+1})$

Bước 2. Một được xác định là: $M_o = a_j + \frac{m_j - m_{j-1}}{(m_j - m_{j-1}) + (m_j - m_{j+1})} \cdot h$

trong đó m_j là tần số của nhóm j (quy ước $m_0 = m_{k+1} = 0$) và h là độ dài của nhóm.

Lưu ý.

- Người ta chỉ định nghĩa một cho mẫu ghép nhóm có độ dài các nhóm bằng nhau. Một mẫu có thể không có một hoặc có nhiều hơn một một.

- Khi tần số của các nhóm số liệu bằng nhau thì mẫu số liệu ghép nhóm không có một.

Ví dụ 4. Bảng số liệu ghép nhóm sau cho biết chiều cao (cm) của 50 học sinh lớp 11A.

Khoảng chiều cao (cm)	[145; 150)	[150; 155)	[155; 160)	[160; 165)	[165; 170)
Số học sinh	7	14	10	10	9

Tính một của mẫu số liệu ghép nhóm này. Có thể kết luận gì từ giá trị tính được?

Lời giải

Tần số lớn nhất là 14 nên nhóm chứa một là nhóm [150; 155). Ta có, $j = 2, a_2 = 150, m_2 = 14, m_1 = 7, m_3 = 10, h = 5$. Do đó

$$M_o = 150 + \frac{14 - 7}{(14 - 7) + (14 - 10)} \cdot 5 \approx 153,18.$$

Số học sinh có chiều cao khoảng 153,18cm là nhiều nhất.

Luyện tập 3. Thời gian (phút) để học sinh hoàn thành một câu hỏi thi được cho như sau:

Thời gian (phút)	[0,5; 10,5)	[10,5; 20,5)	[20,5; 30,5)	[30,5; 40,5)	[40,5; 50,5)
Số học sinh	2	10	6	4	3

Tìm một của mẫu số liệu ghép nhóm này.

Lời giải

Tần số lớn nhất là 10 nên nhóm chứa một là nhóm [10,5; 20,5). Ta có,

$$j = 2, a_2 = 10,5, m_2 = 10, m_1 = 2, m_3 = 6, h = 10.$$

$$\text{Do đó: } M_o = 10,5 + \frac{10 - 2}{(10 - 2) + (10 - 6)} \times 10 = 17,16$$

Ý nghĩa. Một của mẫu số liệu ghép nhóm xấp xỉ cho một của mẫu số liệu gốc, nó được dùng để đo xu thế trung tâm của mẫu số liệu.

Vận dụng: Hãy tính các số đặc trưng cho mẫu số liệu trong Bảng 3.1 và giải thích ý nghĩa của các giá trị thu được.

GV: TRẦN ĐÌNH CỬ - 0834332133

Một cửa hàng đã ghi lại số tiền bán xăng cho 35 khách hàng đi xe máy. Mẫu số liệu gốc có dạng: x_1, x_2, \dots, x_{35} trong đó x_i là số tiền bán xăng cho khách hàng thứ i . Vì một lí do nào đó, cửa hàng chỉ có mẫu số liệu ghép nhóm dạng sau:

Số tiền (nghìn đồng)	[0; 30)	[30; 60)	[60; 90)	[90; 120)
Số khách hàng	3	15	10	7

Bảng 3.1. Số tiền khách hàng mua xăng

Lời giải

Số tiền (nghìn đồng)	[0;30)	[30;60)	[60;90)	[90;120)
Số khách hàng	3	15	10	7
Giá trị đại diện	15	45	75	105

Số trung bình là: $\bar{x} = \frac{3 \times 15 + 15 \times 45 + 10 \times 75 + 7 \times 105}{3 + 15 + 10 + 7} = 63$

Cỡ mẫu là: $n = 3 + 15 + 10 + 7 = 35$

Ý nghĩa: Xấp xỉ bằng số trung bình của mẫu số liệu gốc, cho biết vị trí trung tâm của mẫu số liệu và đại diện cho mẫu số liệu

Trung vị là x_{18} thuộc nhóm [30;60), do đó $p = 2; a_2 = 30; m_2 = 15; m_1 = 3, a_3 - a_2 = 30$

và ta có: $M_e = 30 + \frac{\frac{35}{2} - 3}{15} \times 30 = 59$

Ý nghĩa: Xấp xỉ bằng trung vị của mẫu số liệu gốc, chia mẫu số liệu thành 2 phần, mỗi phần chứa 50% giá trị

Tứ phân vị thứ nhất Q_1 là x_9 thuộc nhóm [30;60), do đó $p = 2; a_2 = 30; m_2 = 15; m_1 = 3; a_3 - a_2 = 30$ và ta

có: $Q_1 = 30 + \frac{\frac{35}{4} - 3}{15} \times 30 = 41.5$

Ý nghĩa: Có 25% số giá trị nhỏ hơn 41.5

Tứ phân vị thứ ba Q_3 là x_{27} thuộc nhóm [60;90),

do đó $p = 3; a_3 = 60; m_3 = 10; m_1 + m_2 = 3 + 15 = 18; a_4 - a_3 = 30$ và ta có:

$Q_3 = 30 + \frac{\frac{3 \times 35}{4} - 18}{10} \times 30 = 54.75$

Ý nghĩa: Có 75% số giá trị nhỏ hơn 54.75%

Tần số lớn nhất là 15 nên một thuộc nhóm [30;60). Ta có,

$$j = 2; a_2 = 30; m_2 = 15; m_1 = 3; m_3 = 10; h = 30,$$

$$\text{do đó: } M_o = 30 + \frac{15-3}{(15-3)+(15-10)} \times 30 = 51.18$$

Ý nghĩa: Xấp xỉ bằng một của mẫu số liệu gốc, được dùng để đo xu thế trung tâm của mẫu số liệu

B. GIẢI BÀI TẬP SÁCH GIÁO KHOA

Bài 3.4. Quãng đường (km) từ nhà đến nơi làm việc của 40 công nhân một nhà máy được ghi lại như sau:

5 3 10 20 25 11 13 7 12 31 19 10 12 17 18 11 32 17 16 2
7 9 7 8 3 5 12 15 18 3 12 14 2 9 6 15 15 7 6 12.

- a) Ghép nhóm dãy số liệu trên thành các khoảng có độ rộng bằng nhau, khoảng đầu tiên là [0;5) . Tìm giá trị đại diện cho mỗi nhóm.
- b) Tính số trung bình của mẫu số liệu không ghép nhóm và mẫu số liệu ghép nhóm. Giá trị nào chính xác hơn?
- c) Xác định nhóm chứa một của mẫu số liệu ghép nhóm thu được.

Lời giải

a)

Quãng đường (km)	[0;5)	[5;10)	[10;15)	[15;20)	[20;25)	[25;30)	[30;35)
Số công nhân	6	10	11	9	1	1	2
Giá trị đại diện	2.5	7.5	12.5	17.5	22.5	27.5	32.5

b) Với mẫu số liệu không ghép nhóm:

$$\bar{x} = (5 + 3 + 10 + 20 + 25 + 11 + 13 + 7 + 12 + 31 + 19 + 10 + 12 + 17 + 18 + 11 + 32 + 17 + 16 + 2 + 7 + 9 + 7 + 8 + 3 + 5 + 12 + 15 + 18 + 3 + 12 + 14 + 2 + 9 + 6 + 15 + 15 + 7 + 6 + 12) : 40 = 11.9$$

Với mẫu số liệu ghép nhóm: $\bar{x} = \frac{2.5 \times 6 + 7.5 \times 10 + 12.5 \times 11 + 17.5 \times 9 + 22.5 \times 1 + 27.5 \times 1 + 32.5 \times 2}{40} = 12.5$

Số trung bình của mẫu số liệu không ghép nhóm chính xác hơn

c) 11 là tần số lớn nhất nên nhóm chứa một là [10;15)

Bài 3.5. Tuổi thọ (năm) của 50 bình ắc quy ô tô được cho như sau:

Tuổi thọ (năm)	[2; 2,5)	[2,5; 3)	[3; 3,5)	[3,5; 4)	[4; 4,5)	[4,5; 5)
Tần số	4	9	14	11	7	5

- a) Xác định một và giải thích ý nghĩa.
- b) Tính tuổi thọ trung bình của 50 bình ắc quy ô tô này.

GV: TRẦN ĐÌNH CỬ - 0834332133

Lời giải

a) 14 là tần số lớn nhất nên một thuộc nhóm [3;3.5), ta có $j = 3, a_3 = 3, m_3 = 14, m_2 = 9, m_4 = 11, h = 0.5$

Do đó: $M_o = 3 + \frac{14 - 9}{(14 - 9) + (14 - 11)} \times 0.5 = 3.31$

b)

Tuổi thọ (năm)	[2;2.5)	[2.5;3)	[3;3.5)	[3.5;4)	[4;4.5)	[4.5;5)
Tần số	4	9	14	11	7	5
Giá trị đại diện	2.25	2.75	3.25	3.75	4.25	4.75

Tuổi thọ trung bình: $\bar{x} = \frac{4 \times 2.25 + 9 \times 2.75 + 14 \times 3.25 + 11 \times 3.75 + 7 \times 4.25 + 5 \times 4.75}{50} = 3.48$

Bài 3.6. Điểm thi môn Toán (thang điểm 100 , điểm được làm tròn đến 1) của 60 thí sinh được cho trong bảng sau:

Điểm	0 - 9	10 - 19	20 - 29	30 - 39	40 - 49
Số thí sinh	1	2	4	6	15
Điểm	50 - 59	60 - 69	70 - 79	80 - 89	90 - 99
Số thí sinh	12	10	6	3	1

a) Hiệu chỉnh để thu được mẫu số liệu ghép nhóm dạng Bảng 3.2.

b) Tìm các tứ phân vị và giải thích ý nghĩa của chúng.

Lời giải

a)

Điểm	[0;10)	[10;20)	[20;30)	[30;40)	[40;50)	[50;60)	[60;70)	[70;80)	[80;90)	[90;100)
Số thí sinh	1	2	4	6	15	12	10	6	3	1

b) Cỡ mẫu $n = 60$

Tứ phân vị thứ nhất Q_1 là $\frac{x_{15} + x_{16}}{2}$. Do x_{15}, x_{16} đều thuộc nhóm [40;50) nên nhóm này chứa Q_1 .

Do đó, $p = 5; a_5 = 40, m_5 = 15, m_1 + m_2 + m_3 + m_4 = 1 + 2 + 4 + 6 = 13, 4$

$4a_6 - a_5 = 10$ và ta có:

$$Q_1 = 40 + \frac{\frac{60}{4} - 13}{15} \times 10 = 41.33$$

GV: TRẦN ĐÌNH CỬ - 0834332133

Ý nghĩa: Có 25% số giá trị nhỏ hơn 41.3

Tứ phân vị thứ hai tức M_e là $\frac{x_{30} + x_{31}}{2}$. Do x_{30}, x_{31} đều thuộc nhóm $[50;60)$ nên nhóm này chứa M_e .

Do đó, $p = 6; a_6 = 50, m_6 = 12$ $m_1 + m_2 + m_3 + m_4 + m_5 = 1 + 2 + 4 + 6 + 15 = 28$ $a_7 - a_6 = 10$ và ta có:

$$M_e = 50 + \frac{\frac{60}{2} - 28}{12} \times 10 = 51.67$$

Ý nghĩa: Có 50% số giá trị nhỏ hơn 51.67

Tứ phân vị thứ ba Q_3 là $\frac{x_{45} + x_{46}}{2}$. Do x_{45}, x_{46} đều thuộc nhóm $[60;70)$ nên nhóm này chứa Q_3 .

Do đó, $p = 7; a_7 = 60, m_7 = 10$ $m_1 + m_2 + m_3 + m_4 + m_5 + m_6 = 1 + 2 + 4 + 6 + 15 + 12 = 40$ $a_8 - a_7 = 10$ và ta có:

$$Q_3 = 60 + \frac{\frac{60 \times 3}{2} - 40}{10} \times 10 = 65$$

Ý nghĩa: Có 75% số giá trị nhỏ hơn 65

Tứ phân vị thứ hai tức M_e là $\frac{x_{30} + x_{31}}{2}$. Do x_{30}, x_{31} đều thuộc nhóm $[50;60)$ nên nhóm này chứa M_e .

Do đó, $p = 6; a_6 = 50, m_6 = 12$

$$m_1 + m_2 + m_3 + m_4 + m_5 = 1 + 2 + 4 + 6 + 15 = 28$$

$a_7 - a_6 = 10$ và ta có:

$$Q_3 = 60 + \frac{\frac{60 \times 3}{2} - 40}{10} \times 10 = 65$$

Ý nghĩa: Có 75% số giá trị nhỏ hơn 65

Bài 3.7. Phòng vấn một số học sinh khối 11 về thời gian (giờ) ngủ của một buổi tối, thu được bảng số liệu ở bên.

a) So sánh thời gian ngủ trung bình của các bạn học sinh nam và nữ.

b) Hãy cho biết 75% học sinh khối 11 ngủ ít nhất bao nhiêu giờ?

Thời gian	Số học sinh nam	Số học sinh nữ
[4; 5)	6	4
[5; 6)	10	8
[6; 7)	13	10
[7; 8)	9	11
[8; 9)	7	8

Lời giải

a)

Thời gian	Giá trị đại diện	Số học sinh nam	Số học sinh nữ
[4;5)	4.5	6	4
[5;6)	5.5	10	8
[6;7)	6.5	13	10
[7;8)	7.5	9	11
[8;9)	8.5	7	8

Thời gian ngủ trung bình của các bạn nam $\bar{x}_{nam} = \frac{4.5 \times 6 + 5.5 \times 10 + 6.5 \times 13 + 7.5 \times 9 + 8.5 \times 7}{6 + 10 + 13 + 9 + 7} = 6.52$

Thời gian ngủ trung bình của các bạn nữ $\bar{x}_{nữ} = \frac{4.5 \times 4 + 5.5 \times 8 + 6.5 \times 10 + 7.5 \times 11 + 8.5 \times 8}{4 + 8 + 10 + 11 + 8} = 6.77$

$6.77 > 6.52$. Như vậy thời gian ngủ trung bình của các bạn nữ nhiều hơn các bạn nam

b) Cỡ mẫu $n = 86$

Tứ phân vị thứ nhất Q_1 là $\frac{x_{21} + x_{22}}{2}$. Do x_{21}, x_{22} đều thuộc nhóm [5;6) nên nhóm này chứa Q_1 . Do đó,

$p = 2, a_2 = 5, m_2 = 18; m_1 = 10, a_3 - a_2 = 1$ và ta có:

$$Q_1 = 5 + \frac{\frac{86}{4} - 10}{18} \times 1 = 5.64$$

Vậy 75% học sinh khối 11 ngủ ít nhất 5.64 giờ

BÀI TẬP CUỐI CHƯƠNG 3

A. TRẮC NGHIỆM

Khảo sát thời gian tập thể dục trong ngày của một số học sinh khối 11 thu được mẫu số liệu ghép nhóm sau:

Thời gian (phút)	[0; 20)	[20; 40)	[40; 60)	[60; 80)	[80; 100)
Số học sinh	5	9	12	10	6

Bảng 3.6

Câu 3.8: Giá trị đại diện của nhóm [20;40) là

- A. 10. B. 20. C. 30. D. 40.

Lời giải

Chọn C

Câu 3.9: Mẫu số liệu ghép nhóm này có số một là

- A. 0. B. 1. C. 2. D. 3.

Lời giải

Chọn B

Vì đây là mẫu số liệu ghép nhóm và tần số các nhóm khác nhau nên có 1 một

Câu 3.10: Nhóm chứa một của mẫu số liệu này là

- A. [20;40) . B. [40;60) . C. [60;80) . D. [80;100)

Lời giải

Chọn B

Tần số của nhóm [40;60) lớn nhất (=12) nên một thuộc nhóm [40;60)

Câu 3.11: Nhóm chứa tứ phân vị thứ nhất là

- A. [0;20) . B. [20;40) . C. [40;60) . D. [60;80) .

Lời giải

Chọn B

Cỡ mẫu $n = 42$.

Tứ phân vị thứ nhất Q_1 là $\frac{x_{10} + x_{11}}{2}$.

Do x_{11}, x_{12} đều thuộc nhóm [20;40) nên nhóm này chứa Q_1

Câu 3.12: Nhóm chứa trung vị là

- A. [0;200) . B. [20;40) . C. [40;60) . D. [60;80) .

Lời giải

Chọn C

Cỡ mẫu $n = 42$

Trung vị M_e là $\frac{x_{21} + x_{22}}{2}$. Do x_{21}, x_{22} đều thuộc nhóm [40;60) nên nhóm này chứa trung vị

B. TỰ LUẬN

Bài 3.13. Cơ cấu dân số Việt Nam năm 2020 theo độ tuổi được cho trong bảng sau:

GV: TRẦN ĐÌNH CỬ - 0834332133

Độ tuổi	Dưới 5 tuổi	5 - 14	15 - 24	25 - 64	Trên 65
Số người (triệu)	7,89	14,68	13,32	53,78	7,66

(Theo: <http://ourworldindata.org>)

Chọn 80 là giá trị đại diện cho nhóm trên 65 tuổi. Tính tuổi trung bình của người Việt Nam năm 2020.

Lời giải

Độ tuổi	Dưới 5 tuổi	5-14	15 - 24	25 - 64	Trên 65
Số người (triệu)	7.89	14.68	13.32	53.78	7.66
Giá trị đại diện	2.5	9.5	19.5	44.5	80

Tuổi trung bình của người Việt Nam năm 2020:

$$\bar{x} = \frac{7.89 \times 2.5 + 14.68 \times 9.5 + 13.32 \times 19.5 + 53.78 \times 44.5 + 7.66 \times 80}{7.89 + 14.68 + 13.32 + 53.78 + 7.66} = 35$$

Bài 3.14. Người ta ghi lại tuổi thọ của một số con ong cho kết quả như sau:

Tuổi thọ (ngày)	[0; 20)	[20; 40)	[40; 60)	[60; 80)	[80; 100)
Số lượng	5	12	23	31	29

Tìm mốt của mẫu số liệu. Giải thích ý nghĩa của giá trị nhận được.

Lời giải

Tần số lớn nhất là 31 nên nhóm chứa mốt là [60;80). Ta có:

$$j = 4, a_4 = 60, m_4 = 31, m_3 = 23, m_5 = 29, h = 20.$$

$$\text{Do đó } M_o = 60 + \frac{31 - 23}{(31 - 23) + (31 - 29)} \times 20 = 76$$

Ý nghĩa: Đa số các con ong có tuổi thọ là 76 ngày

Bài 3.15. Một bảng xếp hạng đã tính điểm chuẩn hoá cho chỉ số nghiên cứu của một số trường đại học ở Việt Nam và thu được kết quả sau:

Điểm	Dưới 20	[20; 30)	[30; 40)	[40; 60)	[60; 80)	[80; 100)
Số trường	4	19	6	2	3	1

Xác định điểm ngưỡng đề đưa ra danh sách 25% trường đại học có chỉ số nghiên cứu tốt nhất Việt Nam.

Lời giải

GV: TRẦN ĐÌNH CỬ - 0834332133

Điểm ngưỡng để đưa ra danh sách 25% trường đại học có chỉ số nghiên cứu tốt nhất Việt Nam là tứ phân vị thứ ba

Ta có: cỡ mẫu $n = 35$

Tứ phân vị thứ ba Q_3 là x_{27} mà x_{27} thuộc nhóm $[30;40)$ nên nhóm này chứa Q_3 . Do đó, $p = 3, a_3 = 30, m_3 = 6, m_1 + m_2 = 4 + 19 = 23, a_4 - a_3 = 10$ và ta có:

$$Q_3 = 30 + \frac{\frac{3 \times 35}{4} - 23}{6} \times 10 = 35.42$$

Vậy để đưa ra danh sách 25% trường đại học có chỉ số nghiên cứu tốt nhất Việt Nam ta lấy các trường có điểm chuẩn hóa trên 35.42