BÀI GIẢNG: CÁC SỐ ĐẶC TRƯNG ĐO XU THẾ TRUNG TÂM CỦA MẪU SỐ LIỆU GHÉP NHÓM

CHUYÊN ĐỀ: NGUYÊN HÀM. TÍCH PHÂN. ỨNG DỤNG

MÔN: TOÁN 11

GIÁO VIÊN: NGÔ VĂN TOẢN

MỤC TIÊU

- ✓ *Ôn tập thống kê lớp 11*
- ✓ *Khoảng biến thiêm khoảng tứ phân vị, phương sai, độ lệch chuẩn của mẫu số liệu ghép nhóm*

ÔN TẬP CHUYÊN ĐỀ THỐNG KÊ LỚP 11

I. MẪU SỐ LIỆU GHÉP NHÓM

Bảng tần số ghép nhóm bao gồm cả giá trị đại diện, tần số tích lũy được lập như ở bảng:

Nhóm	Tần số	Giá trị đại diện	Tần số tích lũy
$[a_1;a_2)$	n_1	$x_1=(a_1+a_2):2$	$cf_1=n_1$
$[a_2;a_3)$	n_2	$x_2=(a_2+a_3):2$	$cf_2=n_1+n_2$
...	...	...	...
$[a_m;a_{m+1})$	n_m	$x_m=(a_m+a_{m+1}):2$	$cf_m=n_1+n_2+\ldots+n_m$
	n		

II. SỐ TRUNG BÌNH CỘNG (SỐ TRUNG BÌNH)

Công thức:

Số trung bình cộng của mẫu số liệu ghép nhóm, kí hiệu $\overline{x}$ được tính theo công thức:

$$\overline{x}=\frac{n_1x_1+n_2x_2+\ldots+n_mx_m}{n}$$

III. TRUNG VỊ

Cách tìm trung vị của mẫu số liệu ghép nhóm có n giá trị

Bước 1: Tính $\frac{n}{2}$, tính tần số tích lũy của các nhóm

Bước 2: Tìm nhóm đầu tiên theo thứ tự có tần số tích lũy lớn hơn hoặc bằng $\frac{n}{2}$

(giả sử là nhóm thứ $k:[a_k;a_{k+1})$)

Bước 3: Ta tìm các thông số:

$r = a_k$ là giá trị đầu mút trái của nhóm k

$d = a_{k+1} - a_k$ là độ dài của nhóm k

n_k là tần số của nhóm k

cf_{k-1} là tần số tích lũy của nhóm $k-1$ (nhóm ngay trước nhóm k)

Bước 4: Thay vào công thức: $M_e = r + \left(\dfrac{\dfrac{n}{2} - cf_{k-1}}{n_k} \right).d$

IV. TỨ PHÂN VỊ

Cách tìm tứ phân vị Q_1, Q_2, Q_3

- Tứ phân vị thứ nhất Q_1

Giả sử nhóm k là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng $\dfrac{n}{4}$, tức là $cf_{k-1} < \dfrac{n}{4}$ nhưng $c_k \geq \dfrac{n}{4}$

Ta gọi r, d, n_k lần lượt là đầu mút trái, độ dài, tần số của nhóm k; cf_{k-1} là tần số tích lũy của nhóm $k-1$

$$Q_1 = r + \left(\frac{\dfrac{n}{4} - cf_{k-1}}{n_k} \right).d$$

- Tứ phân vị thứ ba Q_3

Giả sử nhóm k là nhóm đầu tiên có tần số tích lũy lớn hơn hoặc bằng $\dfrac{3n}{4}$, tức là $cf_{k-1} < \dfrac{3n}{4}$ nhưng $c_k \geq \dfrac{3n}{4}$

Ta gọi r, d, n_k lần lượt là đầu mút trái, độ dài, tần số của nhóm k; cf_{k-1} là tần số tích lũy của nhóm $k-1$

$$Q_3 = r + \left(\frac{\dfrac{3n}{4} - cf_{k-1}}{n_k} \right).d$$

V. MỐT

Cách tìm mốt của mẫu số liệu ghép nhóm

Bước 1: Xác định nhóm có tần số lớn nhất (giả sử là nhóm thứ k)

Bước 2: Ta xác định các thông số sau:

$r = a_k$ là giá trị đầu mút trái của nhóm k

$d = a_{k+1} - a_k$ là độ dài của nhóm k

n_k; n_{k-1}; n_{k+1} lần lượt là tần số của nhóm thứ k; thứ $k-1$ (liền trước nhóm k); thứ $k+1$ (liền sau nhóm k

Bước 3: Thay vào công thức

$$M_o = r + \left(\frac{n_k - n_{k-1}}{2n_k - n_{k-1} - n_{k+1}} \right).g$$

Chú ý: Khi $k = 0$ thì $n_0 = 0$

Khi $k = m$ thì $n_{m+1} = 0$

Ví dụ: Khi thống kê chỉ số đường huyết (đơn vị: mmol/L) của 28 người cao tuổi trong một lần đo, ta được kết quả sau:

7,5	7,2	7,5	7,1	7,9	7,4	7,0
7,1	7,5	7,2	7,1	8,0	7,9	7,7
7,5	7,6	7,7	7,2	7,6	7,6	7,5
7,3	7.4	7,2	7,1	7,2	7,1	7,0

a) Lập bảng tần số ghép nhóm bao gồm cả tần số tích lũy có năm nhóm ứng với năm nửa khoảng:

$[7,0;\ 7,2), [7,2;\ 7,4), [7,4;\ 7,6), [7,6;\ 7,8), [7,8;\ 8,0)$

b) Xác định các số đặc trưng đo xu thế trung tâm cho mẫu số liệu ghép nhóm đó (làm tròn các kết quả đến hàng phần mười).

Cách giải:

a)

Nhóm	Tần số
$[7,0;7,2)$	**7**
$[7,2;7,4)$	6
$[7,4;7,6)$	7
$[7,6;7,8)$	5
$[7,8;8,0)$	3
	$n = 28$

b)

Nhóm	Tần số	Tần số tích lũy	
$[7,0;7,2)$	7	7	7,1
$[7,2;7,4)$	6	13	7,3
$[7,4;7,6)$	7	20	7,5
$[7,6;7,8)$	5	25	7,7
$[7,8;8,0)$	3	28	7,9
	$n = 28$		

Trung bình cộng: $\overline{x} = \frac{7.7,1 + 6.7,3 + 7.7,5 + 5.7,7 + 3.7,9}{28} \approx 7,44$

Tứ phân vị thứ nhất:

Ta có: $\frac{n}{4}=7$ nên nhóm 1 là nhóm đầu tiên

Khi đó $Q_1=7+\frac{7-0}{7}.0,2=7,2$

Tứ phân vị thứ hai:

Ta có: $\frac{n}{2}=14$

Khi đó $Q_2=7,4+\frac{14-13}{7}.0,2\approx 7,43$

Tứ phân vị thứ ba:

Ta có: $\frac{3n}{4}=21$

Khi đó $Q_3=7,6+\frac{21-20}{5}.0,2=7,64$

Mốt:

$$M_{01}=7,0+\frac{7-0}{2.7-0-6}.0,2=7,175$$

$$M_{02}=7,4+\frac{7-6}{2.7-6-5}.0,2\approx 7,46$$

CHUYÊN ĐỀ THỐNG KÊ LỚP 12

I. KHOẢNG BIẾN THIÊN

Xét mẫu số liệu ghép nhóm như sau:

Nhóm	$[a_1;a_2)$	$[a_2;a_3)$	$\cdots$	$[a_m;a_{m+1})$	
Tần số	n_1	n_2	$\cdots$	n_m	n

Khi đó $R=a_{m+1}-a_1$ được gọi là khoảng biến thiên của mẫu số liệu ghép nhóm trên.

Chú ý:

- Khoảng biến thiên của mẫu số liệu ghép nhóm luôn lớn hơn hoặc bằng khoảng biến thiên của mẫu số liệu gốc.

- Ta có thể chọn khoảng biến thiên của mẫu số liệu gốc là khoảng biến thiên của mẫu số liệu ghép nhóm.

Ý nghĩa:

- Khoảng biến thiên của mẫu số liệu ghép nhóm đo mức độ phân tán của mẫu số liệu đó.

- Khoảng biến thiên dễ bị ảnh hưởng bởi các giá trị bất thường nên chưa phản ánh đúng mức độ phân tán của phần lớn các số liệu trong mẫu.

II. KHOẢNG TỨ PHÂN VỊ

Xét mẫu số liệu ghép nhóm như sau:

Nhóm	$[a_1;a_2)$	$[a_2;a_3)$	$\cdots$	$[a_m;a_{m+1})$	
Tần số	n_1	n_2	$\cdots$	n_m	n
Tần số tích lũy	$cf_1=n_1$	$cf_2=n_1+n_2$	$\cdots$	$cf_m=n_1+n_2+\ldots+n_m$	

Gọi Q_1, Q_2, Q_3 là tứ phân vị của mẫu số liệu đó.

Khi đó, ta gọi $\Delta_Q=Q_3-Q_1$ là khoảng tứ phân vị của mẫu số liệu đó.

Ý nghĩa:

- Khoảng tứ phân vị của mẫu số liệu ghép nhóm là giá trị xấp xỉ cho khoảng tứ phân vị của mẫu số liệu gốc và có thể dùng để đo mức độ phân tán của nữa giữa của mẫu số liệu.
- Khoảng tứ phân vị của mẫu số liệu ghép nhóm càng nhỏ thì dữ liệu càng tập trung xung quanh trung vị
- Khoảng tứ phân vị giúp xác định các giá trị bất thường của mẫu số liệu và thường dùng để thay cho khoảng biến thiên của mẫu để đo mức độ phân tán vì không bị ảnh hưởng bởi các giá trị bất thường đó.

III. PHƯƠNG SAI, ĐỘ LỆCH CHUẨN

Xét mẫu số liệu ghép nhóm như sau:

Nhóm	$[a_1;a_2)$	$[a_2;a_3)$	$\cdots$	$[a_m;a_{m+1})$	
Tần số	n_1	n_2	$\cdots$	n_m	n

Gọi $\overline{x}$ là số trung bình cộng của mẫu số liệu trên. Khi đó:

+) Phương sai:

$$s^2=\frac{n_1\left(x_1-\overline{x}\right)^2+n_2\left(x_2-\overline{x}\right)^2+\ldots+n_m\left(x_m-\overline{x}\right)^2}{n}$$

+) Độ lệch chuẩn: $s=\sqrt{s^2}$

Ý nghĩa:

- Phương sai (độ lệch chuẩn) của mẫu số liệu ghép nhóm xấp xỉ cho phương sai (độ lệch chuẩn) của mẫu số liệu gốc và được dùng để đo mức độ phân tán của mẫu số liệu ghép nhóm.
- Độ lệch chuẩn có cùng đơn vị với đơn vị của mẫu số liệu.
- Khi hai mẫu số liệu ghép nhóm có cùng đơn vị đo và có số trung bình cộng bằng nhau (hoặc xấp xỉ nhau), mẫu số liệu nào có độ lệch chuẩn nhỏ hơn thì mức độ phân tán (so với trung bình cộng) của các số liệu trong mẫu đó sẽ thấp hơn.

ĐỀ BÀI

Câu 1: Xét mẫu số liệu ghép nhóm cho bởi bảng sau. Khoảng biến thiên của mẫu số liệu ghép nhóm bằng

Nhóm	Tần số
$[8;11)$	5
$[11;14)$	6
$[14;17)$	8
$[17;20)$	7
$[20;23)$	4
	$n=30$

A. 3. **B.** 8. **C.** 31. **D.** 15.

Câu 2: Biểu diễn mẫu số liệu ghép nhóm về cân nặng (đơn vị: kilogam) của 40 học sinh một lớp 12 ở một trường trung học phổ thông được cho dưới bảng sau. Khoảng biến thiên của mẫu số liệu ghép nhóm đó là

Nhóm	Tần số
$[36;40)$	4
$[40;44)$	7
$[44;48)$	9
$[48;52)$	12
$[52;56)$	6
$[56;60)$	2
	$n=40$

A. 24. **B.** 4. **C.** 12. **D.** 6.

Câu 3: Kết quả đo chiều cao của 100 cây keo ba năm tuổi tại một công trường được cho ở bảng sau. Khoảng tứ phân vị của mẫu số liệu ghép nhóm bằng

Chiều cao (m)	$[8,4;8,6)$	$[8,6;8,8)$	$[8,8;9,0)$	$[9,0;9,2)$	$[9,2;9,4)$
Số cây	5	12	25	44	14

A. $0,886$. **B.** $0,115$. **C.** $0,826$. **D.** $0,286$.

Câu 4: Bảng số liệu ghép nhóm tổng lượng mưa (đơn vị: mm) đo được vào tháng 7 từ năm 2005 đến 2024 tại một trạm quan trắc đặt ở Cà Mau. Khoảng tứ phân vị của mẫu số liệu ghép nhóm bằng

Lượng mưa	$[150;225)$	$[225;300)$	$[300;375)$	$[375;450)$	$[450;525)$
Số năm	3	5	3	6	3

A. 375. **B.** 170. **C.** 225. **D.** 200.

Câu 5: Bảng dưới đây thống kê cân nặng của 45 học sinh lớp 10 tại một trường THPT của Thành phố Hà Nội. Tìm phương sai, độ lệch chuẩn của mẫu số liệu.

Nhóm (Số cân nặng)	Giá trị đại diện	Tần số
$[40;44)$	42	8
$[44;48)$	46	12
$[48;52)$	50	8
$[52;56)$	54	10
$[56;62)$	58	7
		45

A. 375. **B.** 170. **C.** 225. **D.** 200.

Câu 6: Bảng dưới đây thống kê cân nặng của 40 học sinh lớp 11A trong một trường trung học phổ thông (đơn vị: kilogam)

Xét tính đúng – sai của các mệnh đề sau?

a) Khoảng biến thiên của mẫu số liệu ghép nhóm là $R=60$

b) Khoảng tứ phân vị của mẫu số liệu ghép nhóm là $\Delta Q=14,5$

c) Số trung bình cộng của mẫu số liệu ghép nhóm là $\overline{x}=56$

d) Phương sai của mẫu số liệu ghép nhóm là $s^2=128$

Nhóm	Tần số	Tần số tích lũy
$[30;40)$	2	2
$[40;50)$	10	12
$[50;60)$	16	28
$[60;70)$	8	36
$[70;80)$	2	38
$[80;90)$	2	40
	$n=40$	

HƯỚNG DẪN GIẢI

Câu 1: Xét mẫu số liệu ghép nhóm cho bởi bảng sau. Khoảng biến thiên của mẫu số liệu ghép nhóm bằng

Nhóm	Tần số
$[8;11)$	5
$[11;14)$	6
$[14;17)$	8
$[17;20)$	7
$[20;23)$	4
	$n=30$

A. 3. **B.** 8. **C.** 31. **D.** 15.

Cách giải:

Trong mẫu số liệu ghép nhóm đó, ta có: đầu mút trái của nhóm 1 là $a_1=8$, đầu mút phải của nhóm 5 là $a_6=23$

Vậy khoảng biến thiên của mẫu số liệu ghép nhóm đó là $R=a_6-a_1=23-8=15$

Chọn D.

Câu 2: Biểu diễn mẫu số liệu ghép nhóm về cân nặng (đơn vị: kilogam) của 40 học sinh một lớp 12 ở một trường trung học phổ thông được cho dưới bảng sau. Khoảng biến thiên của mẫu số liệu ghép nhóm đó là

Nhóm	Tần số
$[36;40)$	4
$[40;44)$	7
$[44;48)$	9
$[48;52)$	12
$[52;56)$	6
$[56;60)$	2
	$n=40$

A. 24. **B.** 4. **C.** 12. **D.** 6.

Cách giải:

Trong mẫu số liệu ghép nhóm đó, ta có: đầu mút trái của nhóm 1 là $a_1=36$, đầu mút phải của nhóm 6 là $a_7=60$

Vậy khoảng biến thiên của mẫu số liệu ghép nhóm đó là $R=a_7-a_1=60-36=24\,(kg)$

Chọn D.

Câu 3: Kết quả đo chiều cao của 100 cây keo ba năm tuổi tại một công trường được cho ở bảng sau. Khoảng tứ phân vị của mẫu số liệu ghép nhóm bằng

Chiều cao (m)	$[8,4;8,6)$	$[8,6;8,8)$	$[8,8;9,0)$	$[9,0;9,2)$	$[9,2;9,4)$
Số cây	5	12	25	44	14

A. $0,886$. **B.** $0,115$. **C.** $0,826$. **D.** $0,286$.

Cách giải:

Ta có: $\frac{n}{4}=25$

Do đó nhóm 3 là nhóm chứa tứ phân vị thứ nhất $Q_1=8,8+\frac{25-17}{25}.0,2=8,864$

Lại có: $\frac{3n}{4}=75$

Do đó nhóm 4 là nhóm chứa tứ phân vị thứ ba $Q_3=9,0+\frac{75-42}{44}.0,2=9,15$

Khi đó $\Delta_Q=Q_3-Q_1=0,286$

Chọn D.

Câu 4: Bảng số liệu ghép nhóm tổng lượng mưa (đơn vị: mm) đo được vào tháng 7 từ năm 2005 đến 2024 tại một trạm quan trắc đặt ở Cà Mau. Khoảng tứ phân vị của mẫu số liệu ghép nhóm bằng

Lượng mưa	$[150;225)$	$[225;300)$	$[300;375)$	$[375;450)$	$[450;525)$
Số năm	3	5	3	6	3

A. 375. **B.** 170. **C.** 225. **D.** 200.

Cách giải:

Ta có: $\frac{n}{4}=5$. Do đó nhóm 2 là nhóm chứa tứ phân vị thứ nhất $Q_1=225+\frac{5-3}{5}.75=255$

Lại có: $\frac{3n}{4}=15$. Do đó nhóm 4 là nhóm chứa tứ phân vị thứ ba $Q_3=375+\frac{15-11}{6}.75=425$

Khi đó $\Delta_Q=Q_3-Q_1=425-255=170$

Chọn D.

Câu 5: Bảng dưới đây thống kê cân nặng của 45 học sinh lớp 10 tại một trường THPT của Thành phố Hà Nội. Tìm phương sai, độ lệch chuẩn của mẫu số liệu.

Nhóm (Số cân nặng)	Giá trị đại diện	Tần số
$[40;44)$	42	8
$[44;48)$	46	12
$[48;52)$	50	8
$[52;56)$	54	10
$[56;62)$	58	7
		45

A. 375. **B.** 170. **C.** 225. **D.** 200.

Cách giải:

Ta có: Số trung bình của mẫu số liệu ghép nhóm là: $\overline{x}=\frac{8.42+12.46+8.50+10.54+7.58}{45}=49,64$

$s^2=\frac{8.(42-49,64)^2+12.(46-49,64)^2+8.(50-49,64)^2+10.(54-49,64)^2+7.(58-49,64)^2}{45}\approx 29,03$

$s=\sqrt{s^2}\approx 5,39$

Chọn D.

Câu 6: Bảng dưới đây thống kê cân nặng của 40 học sinh lớp 11A trong một trường trung học phổ thông (đơn vị: kilogam)

Nhóm	Tần số	Tần số tích lũy
$[30;40)$	2	2
$[40;50)$	10	12
$[50;60)$	16	28
$[60;70)$	8	36
$[70;80)$	2	38
$[80;90)$	2	40
	$n=40$	

Xét tính đúng – sai của các mệnh đề sau?

a) Khoảng biến thiên của mẫu số liệu ghép nhóm là $R=60$

b) Khoảng tứ phân vị của mẫu số liệu ghép nhóm là $\Delta Q=14,5$

c) Số trung bình cộng của mẫu số liệu ghép nhóm là $\overline{x}=56$

d) Phương sai của mẫu số liệu ghép nhóm là $s^2=128$

Cách giải:

a) $R=90-30=60$

b) Ta có: $\frac{n}{4}=10$

Do đó nhóm 2 là nhóm chứa tứ phân vị thứ nhất

$Q_1=40+\frac{10-2}{10}.10=48$

Lại có: $\frac{3n}{4}=30$

Do đó nhóm 3 là nhóm chứa tứ phân vị thứ ba

$Q_3=60+\frac{30-28}{8}.10=62,5$

Khi đó $\Delta Q=Q_3-Q_1=14,5$

c) Đúng

d) Sai