

CHƯƠNG V. THỐNG KÊ

BÀI 1. BẢNG PHÂN BỐ TẦN SỐ, TẦN SUẤT

A. KIẾN THỨC CƠ BẢN CẦN NẮM

I. ÔN TẬP

1. Số liệu thống kê

- **Thống kê** là khoa học về các phương pháp thu thập, tổ chức, trình bày, phân tích và xử lý số liệu.
- **Dấu hiệu (điều tra)** là một vấn đề hay hiện tượng nào đó mà người điều tra quan tâm tìm hiểu. Mỗi đối tượng điều tra gọi là một **đơn vị điều tra**. Mỗi đơn vị điều tra có một số liệu, số liệu đó gọi là giá trị của dấu hiệu trên đơn vị điều tra đó.
- Một tập con hữu hạn các đơn vị điều tra được gọi là một **mẫu**. Số phần tử của một mẫu được gọi là **kích thước mẫu**. Các giá trị của dấu hiệu thu được trên mẫu được gọi là một **mẫu số liệu** (mỗi giá trị như thế còn gọi là một số liệu của mẫu).
- Nếu thực hiện điều tra trên mọi đơn vị điều tra thì đó là điều tra toàn bộ. Nếu chỉ điều tra trên một mẫu thì đó là điều tra mẫu.

2. Tần số

Tần số của giá trị x_i là số lần lặp lại của giá trị x_i trong mẫu số liệu.

II. Tần suất

Tần suất f_i của giá trị x_i là tỷ số giữa tần số n_i và kích thước mẫu N hay $f_i = \frac{n_i}{N}$.

Người ta thường viết tần suất dưới dạng phần trăm.

Bảng phân bố tần số - tần suất.

- **Bảng phân bố tần số** (gọi tắt là **bảng tần số**) được trình bày như sau:

Bảng ngang

Giá trị (x)	x_1	x_2	x_3	..	x_m	
Tần số (n)	n_1	n_2	n_3	..	n_m	$N = \sum_{i=1}^m n_i$

Bảng dọc

Giá trị (x)	Tần số (n)
x_1	n_1
x_2	n_2
x_3	n_3
..	
x_m	n_m

Cộng	$N = \sum_{i=1}^m n_i$
------	------------------------

Trên hàng tần số, người ta dành một ô để ghi kích thước mẫu N hàng tổng các tần số (tức N

$$= \sum_{i=1}^m n_i).$$

- **Bảng phân bố tần suất** (gọi tắt là *bảng tần suất*) được trình bày như sau:

Bảng ngang

Giá trị (x)	x ₁	x ₂	x ₃	..	x _m	
Tần suất (%)	f ₁	f ₂	f ₃	..	f _m	100%

Bảng dọc

Giá trị (x)	Tần suất (n)
x ₁	f ₁
x ₂	f ₂
x ₃	f ₃
..	
x _m	f _m
Cộng	100%

- **Bảng phân bố tần số - tần suất** (gọi tắt là *bảng tần số - tần suất*).

Bảng dọc

Giá trị (x)	x ₁	x ₂	x ₃	..	x _m	
Tần số (n)	n ₁	n ₂	n ₃	..	x _m	$N = \sum_{i=1}^m n_i$
Tần suất %	f ₁	f ₂	f ₃	..	f _m	

Bảng dọc

Giá trị (x)	Tần số (n)	Tần suất (n)
x ₁	n ₁	f ₁
x ₂	n ₂	f ₂
x ₃	n ₃	f ₃
..	..	..
x _m	n _m	f _m
Cộng	$N = \sum_{i=1}^m n_i$	100%

III. Bảng phân bố tần số - tần suất ghép lớp.

Nếu kích thước mẫu số liệu khá lớn, thì người ta thường chia số liệu thành nhiều lớp dưới dạng $[a; b]$ hay $[a; b)$ (thường có độ dài các lớp bằng nhau). Khi đó tần số của lớp $[a; b]$ là số giá trị

$x_i \in [a; b]$ (hay $x_i \in [a; b)$) xuất hiện trong lớp đó. Tần suất của lớp $[a; b]$ là $f = \frac{n}{N}$ trong đó n là tần số của lớp $[a; b]$ và N là kích thước mẫu.

- Bảng phân bố tần suất ghép lớp được xác định tương tự như trên.

- Giá trị đại diện của lớp $[a; b]$ là $c = \frac{a + b}{2}$

Bảng phân bố tần số, tần suất ghép lớp

Lớp giá trị (x)	Tần số (n)	Tần suất (n)
$[x_1, x_2)$	n_1	f_1
$[x_2, x_3)$	n_2	f_2
$[x_3, x_4)$	n_3	f_3
..	..	..
$[x_k, x_m]$	n_m	f_m
Cộng	$N = \sum_{i=1}^m n_i$	100%

B. CÂU HỎI TRẮC NGHIỆM

Câu 1: Để điều tra số con trong mỗi gia đình của một chung cư gồm 100 gia đình, người ta chọn ra 20 gia đình ở tầng 1 và thu được mẫu số liệu sau đây.

3 2 4 1 3 5 1 1 2 3 1 2 1 3 4 1 1 2 3 5

Dấu hiệu ở đây là gì?

A. Số gia đình ở tầng 1.

B. Số người trong mỗi gia đình.

C. Số tầng ở trong khu chung cư.

D. Số con trong mỗi gia đình.

Lời giải

Chọn D.

Dấu hiệu là: Số con trong mỗi gia đình.

Câu 2: Điểm kiểm tra học kì môn Toán của các học sinh lớp 10A cho ở bảng dưới đây.

Điểm	3	4	5	6	7	8	9	10
Tần số	1	2	5	8	6	10	7	2

Hỏi lớp 10A có bao nhiêu học sinh?

A. 40.

B. 39.

C. 41.

D. Không tính được.

Lời giải

Chọn C.

Số học sinh lớp 10A là: $1 + 2 + 5 + 8 + 6 + 10 + 7 + 2 = 41$ (học sinh)

Câu 3: Điều tra thời gian hoàn thành một sản phẩm của 20 công nhân, người ta thu được mẫu số liệu sau (thời gian tính bằng phút).

10	12	13	15	11	13	16	18	19	21
23	21	15	17	16	15	20	13	16	11

Kích thước mẫu là bao nhiêu?

A. 10.

B. 20.

C. 200.

D. 300.

Lời giải

Chọn B

Ta có 20 số liệu thống kê nên ta có kích thước mẫu là 20.

Câu 4: Điểm thi học kì I của lớp 10A được ghi lại trong bảng sau.

8	6,5	7	5	5,5	8	4	5	7
8	4,5	10	7	8	6	9	6	8
6	6	2,5	8	8	7	4	10	6
9	6,5	9	7,5	7	6	6	3	6
6	9	5,5	7	8	6	5	6	4

Số các giá trị khác nhau của dấu hiệu trong bảng trên là.

A. 14.

B. 13.

C. 12.

D. 11.

Lời giải

Chọn B

Nhìn vào bảng ta thấy có 13 giá trị khác nhau.

Câu 5: 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 6: Thống kê điểm môn Toán trong một kì thi của 500 em học sinh ở một trường phổ thông thấy số bài được điểm 9 chiếm tỉ lệ 4,0%. Hỏi tần số của giá trị $x_i = 9$ là bao nhiêu?

A. 10.

B. 20.

C. 30.

D. 40.

Lời giải

Chọn B

Ta có tần số của giá trị $x_i = 9$ là $n_i = \frac{4 \times 500}{100} = 20$.

Câu 7: Thống kê về điểm thi môn Toán trong một kì thi của 450 em học sinh trong một kì thi ở một trường phổ thông. Người ta thấy có 99 bài được điểm 7. Hỏi tần suất của giá trị $x_i = 7$ là bao nhiêu?

A. 7%.

B. 22%.

C. 45%.

D. 50%.

Lời giải

Chọn B

Ta có tần suất của giá trị $x_i = 7$ là $f_i = \frac{99 \times 100\%}{450} = 22\%$

Câu 8: 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 lớp 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 9: Tuổi thọ của 30 bóng đèn được thử được thống kê theo bảng sau. Hãy điền số thích hợp vào dấu ô chứa dấu (*).

Tuổi thọ (giờ)	Tần số	Tần suất (%)
1150	13	10
1160	6	20
1170	*	40
1180	6	20
1190	3	10
Cộng	30	100%

- A. 3. B. 6. C. 9. D. 12.

Lời giải

Chọn D

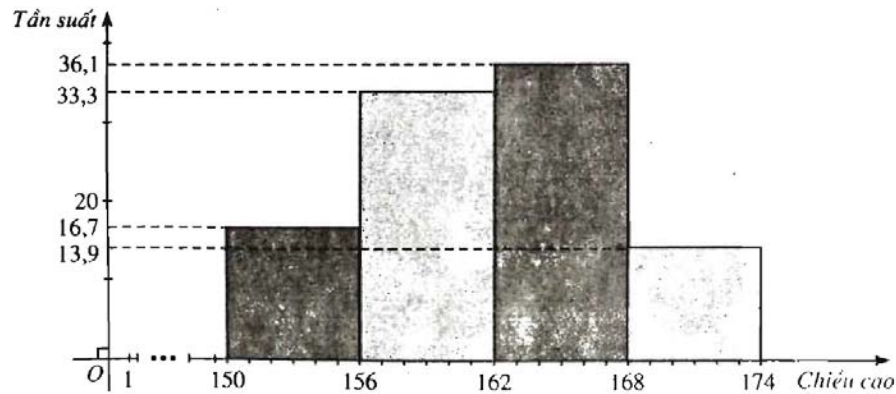
Ta có tần số của giá trị 1170 là $\frac{30 \times 40\%}{100\%} = 12$.

BÀI 2. BIỂU ĐỒ

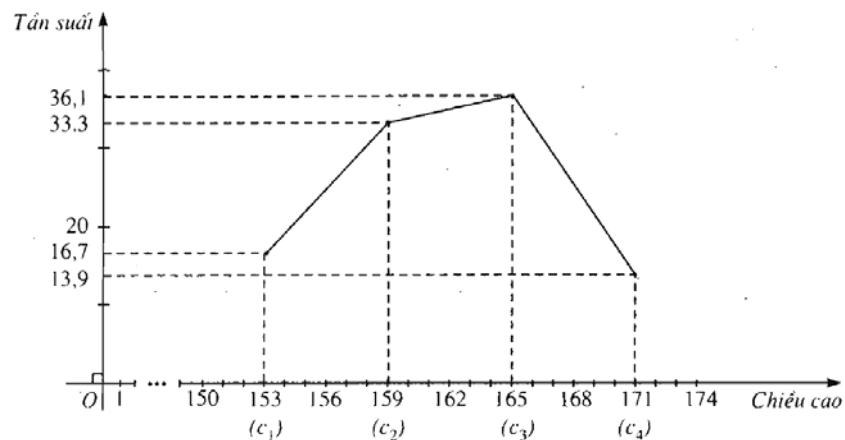
A. KIẾN THỨC CƠ BẢN CẦN NẮM

I. BIỂU ĐỒ TẦN SUẤT HÌNH CỘT VÀ ĐƯỜNG GẤP KHÚC TẦN SUẤT

1. Biểu đồ tần suất hình cột



2. Đường gấp khúc tần suất



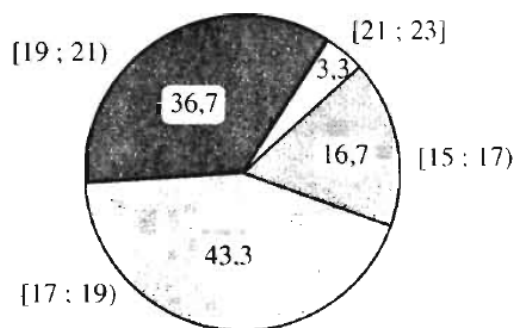
II. BIỂU ĐỒ HÌNH QUẠT

Nhiệt độ trung bình của tháng 12 tại thành phố Vinh từ 1961 đến 1990 (30 năm).

Lớp nhiệt độ ($^{\circ}\text{C}$)	Tần suất (%)
[15 ; 17)	16,7
[17 ; 19)	43,3
[19 ; 21)	36,7
[21 ; 23]	3,3
Cộng	100 (%)

Bảng 6

Bảng phân bố tần số tần suất ghép lớp có thể miêu tả bằng biểu đồ hình quạt



Hình 36 b)

Hình 36b miêu tả bảng 6.

B. CÂU HỎI TRẮC NGHIỆM

Câu 1: Bảng phân bố tần số sau đây ghi lại số vé không bán được trong 40 buổi chiếu phim:

Lớp	[0;5);	[5;10)	[10;15)	[15;20)	[20;25)	[25;30)	Cộng
Tần số	3	4	2	20	5	6	40

Ta vẽ biểu đồ tần suất hình cột với 6 cột hình chữ nhật, các đáy tương ứng là [0;5);[5;10);[10;15);[15;20);[20;25);[25;30). Mỗi lớp trên trục Ox dài 5cm, 1% trên trục Oy dài 1cm. Hỏi cột cao nhất có diện tích là:

A. 25

B. 50

C. 62,5

D. 250

Lời giải

Chọn D

Cột cao nhất là cột có đáy [15;20), có tần suất là 50% $\Rightarrow$ diện tích cột có đáy [15;20) là $50.5=250$

Câu 2: Với mỗi tỉnh người ta ghi lại số phần trăm những trẻ mới sinh có khối lượng dưới 2500g. Sau đây là kết quả khảo sát ở 43 tỉnh trong một năm (đơn vị %)

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

7,6	7,0	7,1	7,4	8,6	6,7	7,9		
-----	-----	-----	-----	-----	-----	-----	--	--

Ta vẽ biểu đồ tần số hình cột với 5 cột hình chữ nhật, các đáy tương ứng là [4,5;5,5);[5,5;6,5);[6,5;7,5);[7,5;8,5);[8,5;9,5). Hỏi cột nào có chiều cao lớn nhất

- A.** [4,5;5,5) **B.** [5,5;6,5) **C.** [6,5;7,5) **D.** [7,5;8,5)

Lời giải

Chọn C

Từ dãy số liệu ta có bảng phân bố tần số -tần suất ghép lớp như sau:

Lớp	[4,5;5,5)	[5,5;6,5)	[6,5;7,5)	[7,5;8,5)	[8,5;9,5)	Cộng
Tần số	9	6	17	8	3	43
Tần suất (%)	20,93	13,95	39,53	18,60	6,98	100 (%)

Nhìn vào bảng ta thấy hình chữ nhật đáy [6,5;7,5) có chiều cao 17 là lớn nhất

Câu 3: Chọn 36 học sinh nam của một trường THPT và đo chiều cao của họ ta thu được mẫu số liệu sau (đơn vị centimet):

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				

Ta vẽ biểu đồ hình quạt với 5 lớp: [159,5;162,5); [162,5;165,5); [165,5;168,5); [168,5;171,5); [171,5;174,5)

Hình quạt nào có diện tích lớn nhất?

- A.** [159,5;162,5) **B.** [162,5;165,5) **C.** [165,5;168,5) **D.** [168,5;171,5)

Lời giải

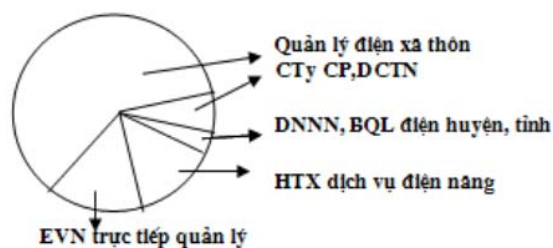
Chọn B

Ta có bảng phân bố tần số ghép lớp như sau

Lớp	Tần số
[159,5;162,5)	6
[162,5;165,5)	12
[165,5;168,5)	10
[168,5;171,5)	5
[171,5;174,5)	3
Cộng	36

Từ đó ta thấy lớp [162,5;165,5) có tần số 12 là cao nhất, nên có diện tích lớn nhất

Câu 4: Cơ cấu quản lý kinh doanh điện nông thôn thể hiện qua biểu đồ hình quạt như hình vẽ.



Cơ cấu quản lý điện nào lớn nhất?

A. Quản lý điện xã thôn.

B. EVN Trực tiếp quản lý

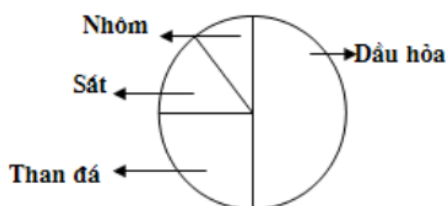
B. HTX dịch vụ điện năng

D. DNNN, BQL điện huyện, tỉnh

Lời giải

Chọn A

Câu 5: Biểu đồ hình quạt của thống kê giá trị xuất khẩu của nước ta về dầu hỏa là 800 triệu USD. Hỏi giá trị xuất khẩu của than đá là bao nhiêu triệu USD



A. 100

B. 200

C. 250

D. 400

Lời giải

Chọn D

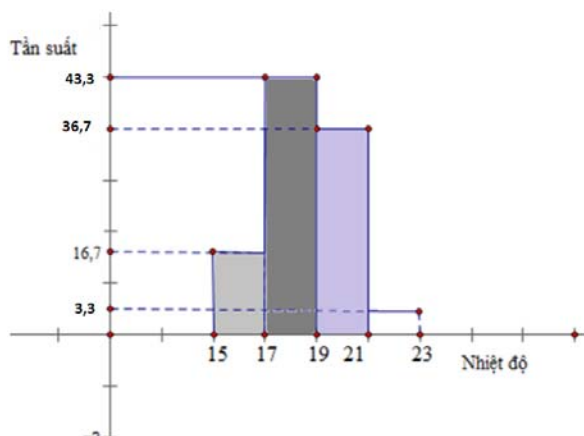
Dựa vào biểu đồ ta thấy dầu hỏa chiếm 50%. Vậy tổng giá trị xuất khẩu của tất cả 4 mặt hàng là:

$$800:50\%=1600 \text{ (triệu USD)}$$

Dựa vào biểu đồ, ta thấy than đá chiếm 25%. Vậy giá trị xuất khẩu về than đá của nước ta là:

$$1600.25\%=400 \text{ (triệu USD)}$$

Câu 6: Cho biểu tần suất hình cột của thống kê nhiệt độ thành phố Vinh từ năm 1961 đến 1990 (30 năm) như hình vẽ. Hỏi Lớp nhiệt độ $[17;19)$ chiếm bao nhiêu %?



A. 16,7 %

B. 43,3%

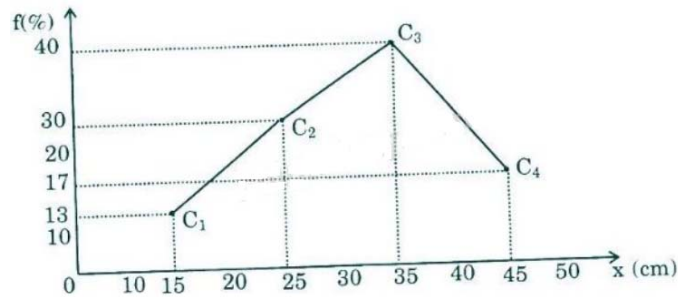
C. 36,7%

D. 3,3%

Lời giải

Chọn B

Câu 7: Cho biểu đồ đường gấp khúc tần suất của thống kê chiều dài của 60 lá dương xỉ trưởng thành như hình vẽ. Hỏi tần suất phần trăm của lớp $[20;30)$ là?



A. 13,3%

B. 30%

C. 40%

D. 16,7%

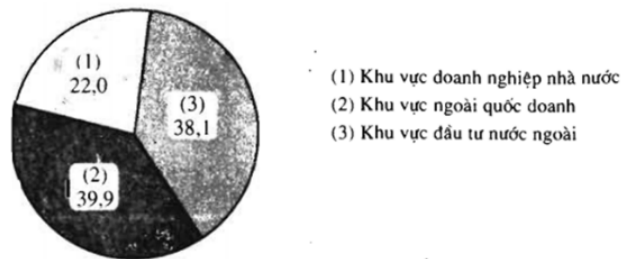
Lời giải

Chọn B

Giá trị đại diện của lớp [20;30) là 25, nhìn vào biểu đồ ta thấy lớp [20;30) chiếm 30%

Câu 8: Cho biểu đồ hình quạt về cơ cấu giá trị sản xuất công nghiệp trong nước năm 1999, phân theo thành phần

Kinh tế. Hỏi khu vực doanh nghiệp nhà nước chiếm bao nhiêu %?



A. 22%

B. 38,1

C. 39,9%

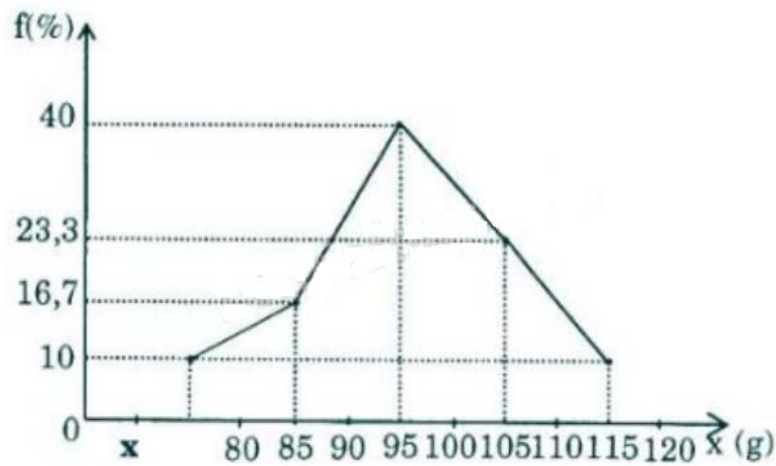
D. Đáp án khác

Lời giải

Chọn A

Câu 9: Cho biểu đồ đường gấp khúc tần suất của thống kê khối lượng 30 củ khoai tây được thu hoạch ở nông

trường T (đơn vị gam). Hỏi lớp [90;100) có tần suất bao nhiêu %?



A. 10%

B. 20%

C. 40%

D. 30%

Lời giải

Chọn C

Giá trị đại diện của lớp [90;100) là 95, dựa vào biểu đồ ta thấy lớp [90;100) có tần suất 40%

Câu 10: Bảng cơ cấu giá trị sản xuất công nghiệp trong nước năm 2000 phân theo ngành kinh tế (%) như sau

Các thành phần kinh tế	Tỉ trọng (%)
Khu vực doanh nghiệp nhà nước	23,5%
Khu vực ngoài quốc doanh	32,2 %
Khu vực đầu tư nước ngoài	44,3%
Cộng	100%

Nếu vẽ biểu đồ hình quạt thì khu vực có diện tích lớn nhất là:

A. Khu vực doanh nghiệp nhà nước

B. Khu vực ngoài quốc doanh

C. Khu vực đầu tư nước ngoài

D. Không có khu vực nào cả

Lời giải

Chọn C

BÀI 3. SỐ TRUNG BÌNH. SỐ TRUNG VỊ. MỐT

A. KIẾN THỨC CƠ BẢN CẦN NẮM

I. SỐ TRUNG BÌNH CỘNG (HAY SỐ TRUNG BÌNH)

• Với mẫu số liệu kích thước N là $\{x_1, x_2, \dots, x_N\}$: $\bar{x} = \frac{\sum_{i=1}^k x_i}{N} = \frac{x_1 + x_2 + \dots + x_k}{N}$

• Với mẫu số liệu được cho bởi bảng phân bố tần số: $\bar{x} = \frac{\sum_{i=1}^k n_i x_i}{N} = \frac{n_1 x_1 + n_2 x_2 + \dots + n_k x_k}{N}$

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

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

II. 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 $\frac{N+1}{2}$ nếu N lẻ;

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

III. MỐT

Mốt của một bảng phân bố tần số là giá trị có tần số lớn nhất và được kí hiệu là M_0 .

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 một làm đại diện. Một mẫu số liệu có thể có nhiều một.

B. CÂU HỎI TRẮC NGHIỆM

Câu 1: Cho bảng thống số liệu thông kê điểm kiểm tra 1 tiết môn Toán của 40 học sinh như sau:

Điểm	3	4	5	6	7	8	9	10	Cộng
Số học sinh	2	3	7	18	3	2	4	1	40

Số trung vị (M_e) và một (M_0) của bảng số liệu thống kê trên là

A. $M_e = 8; M_0 = 40$.

B. $M_e = 6; M_0 = 18$.

C. $M_e = 6; M_0 = 6$.

D. $M_e = 7; M_0 = 6$

Lời giải

Chọn C

Theo công thức trung vị đối với N chẵn thì ta có:

Số đứng vị trí $\frac{N}{2}$ là 6 và số đứng vị trí $\frac{N}{2} + 1$ là 6.

Vậy số trung vị $M_e = 6$.

$M_0 = 6$ do số điểm 6 có tần suất xuất hiện nhiều nhất là 18 lần.

Câu 2: Bạn An đạt được điểm môn Toán như sau: điểm hệ số 1: 7; 9; 8; 8; 8, điểm hệ số 2: 7; 8; 8, điểm thi học kỳ (hệ số 3): 8. Điểm trung bình môn Toán của An là

A. $\approx 8,1$.

B. $\approx 7,6$.

C. $\approx 7,9$.

D. $\approx 7,7$.

Lời giải

Chọn C

Điểm trung bình môn toán của An là: $\frac{7+9+8+8+8+7.2+8.2+8.2+8.3}{1+1+1+1+1+2+2+2+3} = \frac{55}{7} \approx 7,9$.

Câu 3: Số trung bình của dãy số liệu 1; 1; 2; 3; 3; 4; 5; 5; 6; 7; 8; 9; 9; 9 gần đúng với giá trị nào nhất trong các giá trị sau?

A. 5,14.

B. 5,15.

C. 5.

D. 6

Lời giải

Chọn A

Số trung bình của dãy số liệu 1; 1; 2; 3; 3; 4; 5; 5; 6; 7; 8; 9; 9; 9 là

$$x_{TB} = \frac{1+1+2+3+3+4+5+5+6+7+8+9+9+9}{14} = \frac{36}{7} \approx 5,142857.$$

Câu 4: Thời gian chạy 50m của 20 học sinh được ghi lại trong bảng dưới đây:

Thời gian (giây)	8,3	8,4	8,5	8,7	8,8
Tần số	2	3	9	5	1

Số trung bình cộng thời gian chạy của học sinh là:

A. 8,54.

B. 4.

C. 8,50.

D. 8,53

Lời giải

Chọn D

$$\bar{x} = \frac{(8,3.2+8,4.3+8,5.9+8,7.5+8,8.1)}{20} = 8,53.$$

Câu 5: Điểm kiểm tra của 24 học sinh được ghi lại trong bảng sau:

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

Tìm một của điểm điều tra

A. 2.

B. 7.

C. 6.

D. 9

Lời giải

Chọn C

Điểm	1	2	3	4	5	6	7	8	9	
Tần số	1	3	3	1	2	5	3	3	3	N=24

Ta thấy điểm 6 có tần số lớn nhất nên $M_0 = 6$.

Câu 6: Cho bảng phân bố tần số khối lượng 30 quả trứng gà của một rổ trứng gà:

Khối lượng (g)	Tần số
25	3
30	5
35	10
40	6
45	4
50	2
Cộng	30

Số trung vị là

A. 37,5.

B. 40.

C. 35.

D. 75

Lời giải

Chọn C

Ta thấy N chẵn nên số trung vị là: $M_e = \frac{35+35}{2} = 35$

Câu 7: Có 100 học sinh tham dự kì thi học sinh giỏi Hóa (thang điểm 20). Kết quả như sau:

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

Số trung bình cộng là

A. $\bar{x} = 15,20$.

B. $\bar{x} = 15,21$.

C. $\bar{x} = 15,23$.

D. $\bar{x} = 15,25$

Lời giải

Chọn C

$$\bar{x} = \frac{(9.1 + 10.1 + 11.3 + 12.5 + 13.8 + 14.13 + 15.19 + 16.24 + 17.14 + 18.10 + 19.2)}{100} = 15,23.$$

Câu 8: Có 100 học sinh tham dự kì thi học sinh giỏi Hóa (thang điểm 20). Kết quả như sau:

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

Số trung vị là:

A. $M_e = 15$.

B. $M_e = 15,50$.

C. $M_e = 16$.

D. $M_e = 16,5$

Lời giải

Chọn B

Ta thấy N=100 chẵn nên số trung vị là: $M_e = \frac{15+16}{2} = 15,5$.

Câu 9: Bảng phân bố tần số- tần suất ghép lớp điểm thi của 32 học sinh trong kì thi Tiếng Anh (thang điểm 100) như sau:

Lớp điểm	Tần số n_i	Tần suất f_i %	Đại diện c_i	$n_i c_i$	$n_i c_i^2$
[40;50)	4	13	45	180	8100
[50;60)	6	19	55	330	18150
[60;70)	10	31	65	650	42250
[70;80)	6	19	75	450	33750
[80;90)	4	13	85	340	28900
[90;100]	2	6	95	190	18050
N	32	100%		2140	149200

Số điểm trung bình là:

A. $\bar{x} = 66,88$.

B. $\bar{x} = 68,68$.

C. $\bar{x} = 88,66$.

D. $\bar{x} = 68,88$

Lời giải

Chọn A

$$\bar{x} = \frac{4.45 + 6.55 + 10.65 + 6.75 + 4.85 + 2.95}{32} = 66,88 \text{ hoặc tính } \bar{x} = \frac{\sum n_i c_i}{N} = \frac{2140}{32} = 66,88.$$

Câu 10: 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?

A. 97,5.

B. 92,5.

C. 95,5.

D. 97,8

Lời giải

Chọn A

Gọi x là số điểm trong lần kiểm tra cuối mà Minh cần đạt được để được cấp chứng chỉ. Ta có số điểm qua 5 lần thi của Minh là $64,5 \cdot 5 = 322,5$ suy ra:

$$\frac{x + 322,5}{6} = 70 \Rightarrow x = 70 \cdot 6 - 322,5 = 97,5$$

BÀI 4. PHƯƠNG SAI VÀ ĐỘ LỆCH CHUẨN

A. KIẾN THỨC CƠ BẢN CẦN NẮM

I. PHƯƠNG SAI

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 $\{x_1, x_2, \dots, x_N\}$:

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

- 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 $N = n_1 + n_2 + \dots + n_k$)

II. ĐỘ LỆCH CHUẨN

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. CÂU HỎI TRẮC NGHIỆM

Câu 1: Nếu đơn vị của số liệu là kg thì đơn vị của phương sai là ?

- A. kg B. kg^2 C. Không có đơn vị D. kg^3

Lời giải

Chọn B

Theo định nghĩa phương sai: Giả sử ta có một mẫu số liệu kích thước N là $\{x_1, \dots, x_N\}$,

Phương sai của mẫu số liệu này, kí hiệu là s^2 , được tính bởi công thức sau:

$$s^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2. \text{ Do đó nếu đơn vị của mẫu là } kg \text{ thì đơn vị của phương sai là } kg^2$$

Câu 2: Cho bảng phân bố tần số ghép lớp:

Các lớp giá trị của X	[50; 52)	[52; 54)	[54; 56)	[56; 58)	[58; 60)	Cộng
Tần số n_i	15	20	45	15	5	100

Mệnh đề đúng là:

- A. Giá trị trung tâm của lớp [50;53) là 52
- B. Tần số của lớp [58;60) là 95
- C. Tần số của lớp [52;54) là 35
- D. Số 50 không phụ thuộc vào lớp [54;56)

Lời giải

Chọn D

Dựa vào định nghĩa tần số và giá trị trung tâm ta loại A,B,C. Do đó ta chọn D

Câu 3: Chọn đáp án đúng trong những đáp án sau: Độ lệch chuẩn là gì ?

- A. Bình phương của phương sai
- B. Một nửa của phương sai
- C. Căn bậc hai của phương sai
- D. Căn bậc ba của phương sai

Lời giải

Chọn C

Căn bậc hai của phương sai được gọi là độ lệch chuẩn

Câu 4: Điều tra về một khối 10 ta có kết quả như sau: ?

Nhóm	Chiều cao (cm)	Số học sinh
1	[150;152)	5
2	[152;154)	18
3	[154;156)	40
4	[156;158)	26
5	[158;160)	8
6	[160;162)	3
		N=100

Giá trị đại diện của nhóm thứ tư là ?

- A. 156,5
- B. 157,5
- C. 157
- D. 158

Lời giải

Chọn C

Ta gọi trung điểm của $x_4 = \frac{156+158}{2} = 157$ giá trị đại diện của nhóm thứ 4

Câu 5: Cho dãy số liệu thống kê 1,2,3,4,5,6,7,8 Độ lệch chuẩn của dãy số liệu thống kê gần bằng ?

- A. 2.30
- B. 3.30
- C. 4.30
- D. 5.30

Lời giải

Chọn A

$$\bar{x} = \frac{1}{8}(1.1 + 2.1 + 3.1 + 4.1 + 5.1 + 6.1 + 7.1 + 8.1) = 4,5$$

$$s^2 = \frac{1}{8}[1.(1-4,5)^2 + 1.(2-4,5)^2 + \dots + 1.(8-4,5)^2] = 5,25$$

$$\Rightarrow s = \sqrt{s^2} \approx 2,30$$

Câu 6: Tỉ số giữa tần số và kích thước mẫu người ta gọi là ?

- A.** Một **B.** Phương sai **C.** Tần suất **D.** Trung vị

Lời giải

Chọn C

Tần số là tỉ số giữa tần số và kích thước mẫu

Câu 7: Cho dãy số liệu thống kê 10,8,6,2,4 Độ lệch chuẩn của mẫu là?

- A.** 2.8 **B.** 8 **C.** 6 **D.** 2.4

Lời giải

Chọn A

$$\bar{x} = \frac{1}{5}(10.1 + 8.1 + 6.1 + 2.1 + 4.1) = 6$$

$$s^2 = \frac{1}{5}[1.(10-6)^2 + 1.(8-6)^2 + 1.(6-6)^2 + 1.(2-6)^2 + 1.(4-6)^2] = 8$$

$$\Rightarrow s = \sqrt{s^2} \approx 2,8$$

Câu 8: Cho bảng số liệu ghi lại điểm của 40 học sinh bài kiểm tra một tiết môn Toán

Điểm	3	4	5	6	7	8	9	10	Cộng
Số học sinh	2	3	7	18	3	2	4	1	40

Mốt của bảng số liệu trên là ?

- A.** $M_0 = 40$ **B.** $M_0 = 18$ **C.** $M_0 = 6$ **D.** $M_0 = 7$

Lời giải

Chọn C

Mốt của dấu hiệu là giá trị có tần số lớn nhất

Câu 9: 100 học sinh tham dự giải toán (thang điểm là). Kết quả được cho trong bảng sau:

Điểm (x)	9	10	11	12	13	14	15	16	17	18	19
Tần số (n)	1	1	3	5	8	13	19	24	14	10	2

Trung bình cộng của bảng số liệu trên là ?

- A.** 15 **B.** 15,23 **C.** 15.5 **D.** 16

Lời giải

Chọn C

$$\bar{x} = \frac{1}{100}(9.1 + 10.1 + 11.3 + 12.5 + \dots + 19.2) = 15,23$$

Câu 10: Điều tra về học sinh khối 10 ta có kết quả sau:

Nhóm	Chiều cao (cm)	Số học sinh
1	[150;152)	5
2	[152;154)	18
3	[154;156)	40
4	[156;158)	26
5	[158;160)	8
6	[160;162)	3
		N=100

Độ lệch chuẩn là ?

A. 0.78

B. 1.28

C. 2.17

D. 1.73

Lời giải

Chọn C

$$\bar{x} = \frac{1}{100}(5.151 + 18.153 + 40.155 + 26.157 + 8.159 + 3.161) = 155,46$$

$$s^2 = \frac{1}{100} \sum_{i=1}^6 n_i (c_i - \bar{x})^2 \approx 4,71$$

$$\Rightarrow s = \sqrt{s^2} \approx 2,17$$