**Chủ đề 7. MỘT SỐ YẾU TỐ THỐNG KÊ**

**A. KIẾN THỨC CẦN NHỚ**

**I. CÁC SỐ ĐẶC TRƯNG ĐO XU THẾ TRUNG TÂM CHO MẪU SÓ LIỆU GHÉP NHÓM**

**1. Số trung bình cộng (số trung bình)**

Cho mẫu số liệu ghép nhóm như ở *Bảng 1.*

 Trung điểm  của nửa khoảng (tính bằng trung bình cộng của hai đầu mút) ứng với nhóm  là giá trị đại diện của nhóm đó.

 Số trung bình cộng của mẫu số liệu ghép nhóm, kí hiệu  được tính theo công thức:



***Ý nghĩa:*** Số trung bình cộng của mẫu số liệu ghép nhóm có thể làm đại diện cho vị trí trung tâm của mẫu số liệu đó khi các số liệu trong mẫu ít sai lệch với số trung bình cộng.

**2. Trung vị**

Cho mẫu số liệu ghép nhóm như ở *Bảng 2.* Giả sử nhóm  là nhóm đầu tiên có tần số tích luỹ lớn hơn hoặc bằng  tức là  nhưng  Ta gọi  lần lượt là đầu mút trái, độ dài, tần số của nhóm  là tần số tích luỹ của nhóm 

*Trung vị* của mẫu số liệu ghép nhóm, kí hiệu  được tính theo công thức sau:



*Quy ước:* 

***Ý nghĩa:*** Trung vị của mẫu số liệu có thể dùng để đại diện cho mẫu số liệu đó.

**3. Tứ phân vị**

Cho mẫu số liệu ghép nhóm như ở *Bảng 2.*

 Giả sử nhóm  là nhóm đầu tiên có tần số tích luỹ lớn hơn hoặc bằng  tức là  nhưng  Ta gọi  lần lượt là đầu mút trái, độ dài, tần số của nhóm  là tần số tích luỹ của nhóm 

Tứ phân vị thứ nhất  được tính theo công thức sau:



 Tứ phân vị thứ hai  bằng trung vị 

 Giả sử nhóm  là nhóm đầu tiên có tần số tích luỹ lớn hơn hoặc bằng  tức là  nhưng  Ta gọi  lần lượt là đầu mút trái, độ dài, tần số của nhóm  là tần số tích luỹ của nhóm 

Tứ phân vị thứ ba  được tính theo công thức sau:



***Ý nghĩa:*** Tứ phân vị  của mẫu số liệu chia mẫu số liệu đó thành bốn phần, mỗi phần chứa  giá trị.

**4. Mốt**

Cho mẫu số liệu ghép nhóm như ở *Bảng 2.*

Giả sử nhóm  là nhóm có tần số lớn nhất. Ta gọi  lần lượt là đầu mút trái, độ dài, tần số của nhóm  lần lượt là tần số của nhóm  nhóm  Mốt của mẫu số liệu ghép nhóm, kí hiệu  được tính theo công thức sau:



*Quy ước:* 

***Ý nghĩa:*** Mốt của mẫu số liệu ghép nhóm có thể dùng để đo xu thế trung tâm của mẫu số liệu đó.

**II. CÁC SỐ ĐẶC TRƯNG ĐO MỨC ĐỘ PHÂN TÁN CHO MẪU SỐ LIỆU GHÉP NHÓM**

**1. Khoảng biến thiên**

Cho mẫu số liệu ghép nhóm như ở *Bảng 3,* trong đó  và  là các số nguyên dương

Gọi  lần lượt là đầu mút trái của nhóm 1, đầu mút phải của nhóm 

Hiệu  được gọi là khoảng biến thiên của mẫu số liệu ghép nhóm đó.

***Ý nghĩa***

 Khoảng biến thiên của mẫu số liệu ghép nhóm đo mức độ phân tán cho mẫu số liệu đó. Khoảng biến thiên càng lớn thì mẫu số liệu càng phân tán.

 Trong các đại lượng đo mức độ phân tán của mẫu số liệu ghép nhóm, khoảng biến thiên là đại lượng dễ hiểu, dễ tính toán. Tuy nhiên, do khoảng biến thiên chỉ sử dụng hai giá trị  và  của mẫu số liệu nên đại lượng đó dễ bị ảnh hưởng bởi các giá trị bất thường.

**2. Khoảng tứ phân vị**

Cho mẫu số liệu ghép nhóm như ở *Bảng 2.*

Gọi  là tứ phân vị của mẫu số liệu đó. Ta gọi hiệu  là khoảng tứ phân vị của mẫu số liệu đó.

***Ý nghĩa:*** Khoảng tứ phân vị của mẫu số liệu ghép nhóm giúp xác định các giá trị bất thường của mẫu đó. Khoảng tứ phân vị thường được sử dụng thay cho khoảng biến thiên vì nó loại trừ hầu hết giá trị bất thường của mẫu số liệu và nó không bị ảnh hưởng bởi các giá trị bất thường đó.

**3. Phương sai và độ lệch chuẩn của mẫu số liệu ghép nhóm**

Cho mẫu số liệu ghép nhóm như ở Bảng 1.

 Gọi  là số trung bình cộng của mẫu số liệu đó.

Số  được gọi là phương sai của mẫu số liệu đó.

 Căn bậc hai (số học) của phương sai được gọi là độ lệch chuẩn của mẫu số liệu ghép nhóm, kí hiệu là  nghĩa là 

***Ý nghĩa***

 Phương sai (độ lệch chuẩn) của mẫu số liệu ghép nhóm được dùng để đo mức độ phân tán của mẫu số liệu ghép nhóm đó:

 Độ lệch chuẩn có cùng đơn vị vơi 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 số trung bình cộng) của các số liệu trong mẫu đó sẽ thấp hơn.

**B. MỘT SỐ VÍ DỤ**

**Dạng 1. Câu trắc nghiệm nhiều phương án lựa chọn**

*Mỗi câu hỏi thí sinh chỉ chọn một phương án.*

***Ví dụ 1.*** Người ta tiến hành phỏng vấn 40 người về một mẫu quần mới. Người phóng vấn yêu cầu cho điểm mẫu quần đó theo thang điểm là 100. Kết quả được trình bày theo mẫu số liệu ghép nhóm được cho ở *Bảng 4.* Trung vị của mẫu số liệu ghép nhóm đó là:

 **A.** 75. **B.** 70,8.

 **C.** 78,8. **D.** 74,8.

*Giải*

Số phần tử của mẫu là  Ta có:  mà  Suy ra nhóm 3 là nhóm đầu tiên có tần số tích luỹ lớn hơn hoặc bằng 20. Xét nhóm 3 có   và nhóm 2 có 

Trung vị của mẫu số liệu đó là:  Chọn 

***Ví dụ 2.*** Xét mẫu số liệu ghép nhóm được cho ở *Bảng 4.* Phương sai của mẫu số liệu ghép nhóm đó là:

 **A.** 9,08. **B.** 82,4375. **C.** 74,75. **D.** 50.

*Giải*

Số trung bình cộng cua mẫu số liệu đó là:



Phương sai của mẫu số liệu đó là:



 Chọn 

**Dạng 2. Câu trắc nghiệm đúng sai**

*Trong mỗi ý a), b), c), d) ở mỗi câu, thí sinh chọn đúng hoặc sai.*

***Ví dụ 3.*** *Bảng 5* biểu diễn mẫu số liệu ghép nhóm về số tiền (đơn vị: nghìn đồng) mà 60 khách hàng mua sách ở một cửa hàng trong một ngày.

a) Số trung bình cộng của mẫu số liệu trên là 65 (nghìn đồng).

b) Trung vị của mẫu số liệu trên là 66,8 (nghìn đồng).

c) Tứ phân vị nhất  của mẫu số liệu trên là 60,8 (nghìn đồng).

d) Mốt của mẫu số liệu trên là 65 (nghìn đồng).

*Giải*

Số trung bình cộng của mẫu số liệu ghép nhóm trên là:

 (nghìn đồng).

Số phần tử của mẫu là  Ta có:  mà  Suy ra nhóm 3 là nhóm đầu tiên có tần số tích luỹ lớn hơn hoặc bằng 30. Xét nhóm 3 có   và nhóm 2 có 

Trung vị của mẫu số liệu đó là:  (nghìn đồng).

Ta có:  mà  Suy ra nhóm 3 là nhóm đầu tiên có tần số tích luỹ lớn hơn hoặc bằng 15. Xét nhóm 3 có  và nhóm 2 có  Tứ phân vị thứ nhất  của mẫu số liệu đó là:  (nghìn đồng).

Ta thấy nhóm 3 là nhóm có tần số lớn nhất với  Nhóm 2 có tần số  nhóm 4 có tần số 

Mốt của mẫu số liệu đó là:  (nghìn đồng).

 Đáp án: a) **S,** b) **Đ,** c) **Đ,** d) **S.**

***Ví dụ 4.*** Kết quả kiểm tra môn Tiếng Anh (cùng đề) của học sinh hai lớp  và  được cho lần lượt bởi mẫu số liệu ghép nhóm ở *Bảng 6, Bảng 7.*



a) Số trung bình cộng của hai mẫu số liệu trên bằng nhau.

b) Độ lệch chuẩn của mẫu số liệu lớp  nhỏ hơn 2.

c) Phương sai của mẫu số liệu lớp  lớn hơn 3.

d) Điểm thi của học sinh lớp  đồng đều hơn lớp 

*Giải*

Số trung bình cộng của mẫu số liệu lớp  là:



Số trung bình cộng của mẫu số liệu lớp  là:



Suy ra số trung bình cộng của hai mẫu số liệu trên bằng nhau.

Phương sai của mẫu số liệu lớp  là:



Độ lệch chuẩn của mẫu số liệu lớp  là:  và 

Phương sai của mẫu số liệu lớp  là:



và 

Vì  nên điểm thi của học sinh lớp  đồng đều hơn lớp 

 Đáp án: a) **Đ,** b) **S,** c) **Đ,** d) **Đ.**

**Dạng 3. Câu trắc nghiệm trả lời ngắn**

***Ví dụ 5.*** Mẫu số liệu dưới đây ghi lại tốc độ của 40 ô tô khi đi qua một trạm đo tốc độ (đơn vị: 

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 49 | 42 | 51 | 55 | 45 | 60 | 53 | 55 | 44 | 65 |
| 52 | 62 | 41 | 44 | 57 | 56 | 68 | 48 | 46 | 53 |
| 63 | 49 | 54 | 61 | 59 | 57 | 47 | 50 | 60 | 62 |
| 48 | 52 | 58 | 47 | 60 | 55 | 45 | 47 | 48 | 61 |

Sau khi ghép nhóm mẫu số liệu trên thành sáu nhóm ứng với sáu nửa khoảng:

 

thì trung vị của mẫu số liệu ghép nhóm nhận được bằng ( là phân số tối giản). Khi đó giá trị của  bằng bao nhiêu?

*Giải*

Lập mẫu số liệu ghép nhóm bao gồm cả tần số tích luỹ như ở *Bảng 8.*

Số phần tử của mẫu là  Ta có:

 mà  Suy ra

nhóm 3 là nhóm đầu tiên có tần số tích luỹ 1ớn hơn hoặc bằng 20. Xét nhóm 3 có  và nhóm 2 có 

Trung vị của mẫu số liệu ghép nhóm đó là:  Suy ra 

***Ví dụ 6.*** *Bảng 9* biểu diễn mẫu số liệu ghép nhóm về nhiệt độ không khí trung bình các tháng trong năm 2021 tại Hà Nội (đơn vị: độ C) (*Nguồn: Niên giám Thống kê 2021, NXB Thống kê, 2022*). Phương sai của mẫu số liệu đó bằng bao nhiêu (làm tròn kết quả đến hàng phần mười)?

*Giải*

Số trung bình cộng của mẫu số liệu đó là:



Phương sai của mẫu số liệu đó là:



**C. BÀI TẬP TỰ LUYỆN**

**Dạng 1. Câu trắc nghiệm nhiều phương án lựa chọn**

*Mỗi câu hỏi thí sinh chỉ chọn một phương án.*

**Câu** **1.** *Bảng 10* biểu diễn mẫu số liệu ghép nhóm về doanh thu (tỉ USD) của 20 hãng xe ô tô có doanh thu cao nhất thế giới năm 2023.

(*Nguồn: Business Research Insights, wiki*)

Tứ phân vị thứ ba  của mẫu số liệu đó bằng

 **A.** 300. **B.** 100.

 **C.** 275. **D.** 175.

**Câu 2.** *Bảng 11* biểu diễn mẫu số liệu ghép nhóm về chi tiêu bình quân (đơn vị: USD) của một lượt khách quốc tế đến Việt Nam phân theo 27 quốc tịch năm 2019. (*Nguồn: https://www.gso.gov.vn*) Độ lệch chuẩn của mẫu số liệu đó nằm trong khoảng nào dưới đây?

 **A.**  **B.** 

 **C.**  **D.** 

**Dạng 2. Câu trắc nghiệm đúng sai**

*Trong mỗi ý a), b), c), d) ở mỗi câu, thí sinh chọn đúng hoặc sai.*

**Câu 3.** Bảng 12 cho ta bảng tần số ghép nhóm về số liệu thống kê tì lệ che phủ rừng (đơn vị: %) của 60 tỉnh, thành phố ở Việt Nam (không bao gồm Hưng Yên, Vĩnh Long, Cần Thơ) tính đến ngày 

(*Nguồn: https://bandolamnghiep.com*)

a) Tỉ lệ che phủ rừng trung bình trên một tỉnh, thành phố được thống kê ở trên là lớn hơn 

b) Trung vị của mẫu số liệu trên là 

c) Có 20 tỉnh, thành phố có tỉ lệ che phủ rừng nhỏ hơn 

d) Mốt của mẫu số liệu trên là 

**Câu 4.** Bạn An và bạn Bình làm thí nghiệm trồng cây. Mỗi bạn trồng 40 cây cần tấy trong cốc, phần gốc của các cây khi bắt đầu trồng đều dài  *Bảng 13* và *Bảng 14* lần lượt biểu diễn mẫu số liệu ghép nhóm về sổ liệu thống kê chiều cao của các cây (đơn vị: centimét) mà bạn An và bạn Bình trồng sau 5 tuần.



a) Chiều cao trung bình của mỗi cây do hai bạn An và Bình trồng không bằng nhau.

b) Khoảng biến thiên của cả hai mẫu số liệu trên là 20.

c) Khoảng tứ phân vị của mẫu số liệu ở *Bảng 13* là 5,5.

d) Chiều cao của các cây mà bạn Bình trồng đồng đều hơn các cây mà bạn An trồng.

**Dạng 3. Câu trắc nghiệm trả lời ngắn**

**Câu 5.** *Bảng 15* cho ta bảng tần số ghép nhóm về số liệu thống kê chiều dài đường bờ biển (đơn vị: kilômét) của 28 tỉnh, thành phố có giáp biển ở Việt Nam. (*Nguồn: https:///vi.wikipedia.org*)

Trung vị của mẫu số liệu đó bằng bao nhiêu (làm tròn kết quả đến hàng đơn vị)?

**Câu 6.** *Bảng 16* cho ta bảng tần số ghép nhóm về số liệu thống kê chiều cao (đơn vị: mét) của 40 núi cao nhất Đông Nam Á. (*Nguồn: https://vi.wikipedia.ong*)

Độ lệch chuẩn của mẫu số liệu đó bằng bao nhiêu (làm tròn kết quả đến hàng đơn vị)?

**D. LỜI GIẢI - HƯỚNG DẪN - ĐÁP SỐ**

**Dạng 1. Câu trắc nghiệm nhiều phương án lựa chọn**

**Câu 1.** D.

**Câu 2.** B.

**Dạng 2. Câu trắc nghiệm đúng sai**

Câu 3. Tỉ lệ che phủ rừng trung bình trên một tỉnh, thành phố là:



Trung vị của mẫu số liệu đó là:



Theo bảng thống kê thì có 17 tỉnh, thành phố có tỉ lệ che phủ rừng nhỏ hơn  Mốt của mẫu số liệu đó là:



 Đáp án: a) **Đ,** b) **Đ,** c) **S,** d) **S.**

**Câu 4.** Chiều cao trung bình của cây do bạn An trồng là: 

Chiều cao trung bình của cây do bạn Bình trồng là: 

Suy ra chiều cao trung bình của mỗi cây do hai bạn An và Bình trồng là bằng nhau.

Khoảng biến thiên của cả hai mẫu số liệu là 

Xét mẫu số liệu ở *Bảng 13.*

 Tứ phân vị thứ nhất  của mẫu số liệu đó là:



 Tứ phân vị thứ ba  của mẫu số liệu đó là:



Suy ra khoảng tứ phân vị của mẫu số liệu ở Bảng 13 là 

Phương sai của mẫu số liệu ở Bảng 13 là: 

Phương sai của mẫu số liệu ở Bảng 14 là: 

Suy ra  Vậy chiều cao của các cây mà bạn An trồng đồng đều hơn các cây mà bạn Bình trồng.

 Đáp án: a) **S,** b) **Đ,** c) **Đ,** d) **S.**

**Dạng 3. Câu trắc nghiệm trả lời ngắn**

**Câu 5.** Trung vị của mẫu số liệu đó bằng khoảng 

**Câu 6.** Độ lệch chuẩn của mẫu số liệu đó bằng khoảng 