

BÀI GIẢNG GIÁO KHOA

Thầy giáo : Nguyễn Quốc Tùng

Toán 11

Bài 8

MẪU SỐ LIỆU GHÉP NHÓM

1. Giới thiệu về mẫu số liệu ghép nhóm

Trong thực tế, khi đối mặt với một tập dữ liệu quá lớn hoặc dữ liệu liên tục (như chiều cao, cân nặng, thời gian), việc liệt kê từng giá trị đơn lẻ là không khả thi. Chúng ta cần **ghép nhóm** chúng.

Khái niệm: Mẫu số liệu ghép nhóm là mẫu số liệu được cho dưới dạng bảng tần số của các nhóm số liệu.

Cấu trúc một nhóm: Thường có dạng $[a; b)$, trong đó:

a: Đầu mút trái (giá trị nhỏ nhất của nhóm).

b: Đầu mút phải (giá trị lớn nhất không thuộc nhóm đó mà thuộc nhóm kế tiếp).

2. Các bước lập bảng tần số ghép nhóm

Để chuyển một mẫu số liệu thô sang dạng ghép nhóm, chúng ta thực hiện các bước sau:

Tìm khoảng biến thiên: $R = x_{\max} - x_{\min}$.

Chọn số nhóm (k): Thường từ 5 đến 10 nhóm.

Xác định độ dài mỗi nhóm (L): $L \approx \frac{R}{k}$.

Xác định các đầu mút: Đảm bảo giá trị nhỏ nhất nằm trong nhóm đầu tiên và giá trị lớn nhất nằm trong nhóm cuối cùng.

Đếm tần số: Xác định số lượng giá trị thuộc mỗi nhóm.

3. Các đại lượng đặc trưng

Khi làm việc với mẫu số liệu ghép nhóm, chúng ta cần lưu ý các khái niệm sau:

Giá trị đại diện của nhóm $[a_i; a_{i+1})$:

$$x_i = \frac{a_i + a_{i+1}}{2}$$

(Đây là trung điểm của nhóm, dùng để tính toán số trung bình cộng sau này).

Tần số của nhóm (n_i): Số giá trị rơi vào nhóm đó.

Tần số tích lũy (cf_i): Tổng tần số của nhóm đó và các nhóm đứng trước nó.

4. Ví dụ minh họa

Đề bài: Khảo sát thời gian tự học trong ngày (giờ) của 20 học sinh:

1, 2, 3, 3, 4, 4, 4, 5, 5, 5, 6, 6, 7, 7, 8, 9, 9, 10, 11, 12

Bảng tần số ghép nhóm:

Nhóm (Thời gian - giờ)	Tần số (Số học sinh)	Giá trị đại diện (xi)
[0; 3)	2	1.5
[3; 6)	9	4.5
[6; 9)	4	7.5
[9; 12)	4	10.5
[12; 15)	1	13.5
Tổng	20	

Ví dụ : Khảo sát chiều cao của học sinh (Dữ liệu có khoảng biến thiên nhỏ)

Mẫu số liệu thô: Đo chiều cao (đơn vị: cm) của 15 học sinh lớp 11A:

152, 155, 158, 160, 162, 163, 165, 168, 170, 172, 174, 175, 178, 180, 182

Cách tạo mẫu số liệu ghép nhóm:

Tìm giá trị nhỏ nhất và lớn nhất: $x_{min} = 152$, $x_{max} = 182$.

Chia nhóm: Ta có thể chia thành 6 nhóm, mỗi nhóm có độ dài là 5 cm, bắt đầu từ 150.

Bảng ghép nhóm:

Nhóm chiều cao (cm)	Tần số (Số học sinh)
[150; 155)	1
[155; 160)	2
[160; 165)	3
[165; 170)	2
[170; 175)	3
[175; 180)	1
[180; 185)	3
Tổng	15

5. Lưu ý

Lựa chọn nhóm: Các nhóm phải liên tiếp và không chồng lấn lên nhau.

Ý nghĩa: Ghép nhóm giúp ta thấy được **phân bố** của dữ liệu (ví dụ: học sinh tập trung học nhiều nhất ở khoảng 3-6 giờ) nhưng sẽ làm **mất đi độ chính xác** tuyệt đối của từng số liệu riêng lẻ.

