

আমার বিজ্ঞকথা



টপিকঃ

ডেটা সায়েন্স

০০১-০১৬ Numerical
Data (পর্ব-৮)





০০১-০১৬ Numerical Data (পর্ব-৮)

আমরা Continuous Numerical Data এর Central Tendency, Dispersion এগুলো শেষ করেছি। এরপর আমরা Continuous Numerical Data এর দুইটি জিনিস শিখবো। Mean Absolute Deviation (\bar{x}_{ad} বা μ_{ad}) এবং Median Absolute Deviation (\tilde{x}_{ad})। Sample এর ক্ষেত্রে Mean Absolute Deviation কে \bar{x}_{ad} দ্বারা প্রকাশ করানো হয়েছে, আর Population এর ক্ষেত্রে Mean Absolute Deviation কে μ_{ad} দ্বারা প্রকাশ করানো হয়েছে। আমরা আগেও বলেছি যে Mean Absolute Deviation আর Median Absolute Deviation এরাও Dispersion এর অংশ।

Continuous Numerical Data এর ক্ষেত্রে Mean Absolute Deviation এবং Median Absolute Deviation:

Mean Absolute Deviation:

দশম শ্রেণীর “ঘ” শাখার ৬০ জন শিক্ষার্থীর পদার্থবিজ্ঞান পরীক্ষার নম্বর নিচে দেওয়া হলো —

53, 97, 87, 46, 46, 53, 91, 66, 49, 49, 91, 51, 49, 81, 67, 56, 54, 98, 95, 48, 65,
56, 48, 80, 88, 84, 50, 63, 60, 67, 66, 78, 98, 52, 56, 79, 90, 67, 61, 95, 86, 45,
48, 99, 68, 71, 85, 87, 58, 59, 93, 57, 93, 97, 82, 48, 50, 80, 90, 78

[আমরা আগেই “০০১-০১০ Numerical Data (পর্ব-২)” এ বলেছিলাম যে - দশম শ্রেণীর “ঘ” শাখার ৬০ জন শিক্ষার্থীর পদার্থবিজ্ঞান পরীক্ষার নম্বরের ডেটাসেট হলো Population]

এক্ষেত্রে আমাদের ফ্রিকুয়েন্সি ডিস্ট্রিবিউশন টেবিল বানানোর নিয়ম হলো, আগে sorting করা।

45, 46, 46, 48, 48, 48, 48, 49, 49, 49, 50, 50, 51, 52, 53, 53, 54, 56, 56, 56, 57,
58, 59, 60, 61, 63, 65, 66, 66, 67, 67, 67, 68, 71, 78, 78, 79, 80, 80, 81, 82, 84,
85, 86, 87, 87, 88, 90, 90, 91, 91, 93, 93, 95, 95, 97, 97, 98, 98, 99

আমরা এক্ষেত্রে লক্ষ্য করতে পাচ্ছি যে maximum value = 99 এবং minimum value = 45।

অতএব, পরিসর (Range) = maximum value – minimum value = 99 – 45 = 54

এরপর আমরা স্ট্রাজেসের ফর্মুলা (Sturges' Formula) দিয়ে class number বের করবো। এখানে “k” হলো class number এবং N হলো উপাত্ত সংখ্যা। $N = 60$ [৬০ জন শিক্ষার্থী]

$$k = 1 + 3.322 \log N = 1 + 3.322 \log(60) = 6.91 \approx 7$$

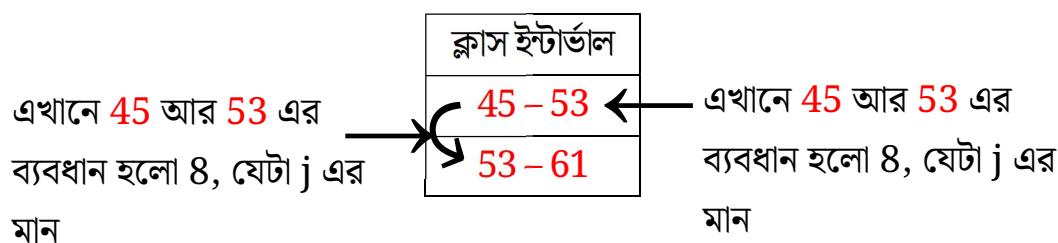
অর্থাৎ ৭টি ক্লাস ইন্টারভাল তৈরি হবে।

এরপর আমরা class width ‘j’ বের করে পাই।

$$j = \frac{\text{Range}}{k} = \frac{54}{7} = 7.7 \approx 8$$

অর্থাৎ প্রত্যেক ক্লাস ইন্টারভালের ব্যবধান ৮ হবে। এবং এক ক্লাস ইন্টারভালের lower limit থেকে তার পরবর্তী ক্লাস ইন্টারভালের lower limit এর ব্যবধান ৮ হবে।

ঠিক এরকম,



এখন আমরা $k = 7$, $j = 8$ অনুযায়ী ফ্রিকুয়েন্সি ডিস্ট্রিবিউশন টেবিল বানানো হলোঃ

$k = 7$
এখানে 7 টি
ক্লাস ইন্টারভাল
রয়েছে

ক্লাস ইন্টারভাল	ট্যালি মার্ক	ফ্রিকুয়েন্সি	কিউমুলেটিভ ফ্রিকুয়েন্সি
45 – 53		14	14
53 – 61		10	14 + 10 = 24
61 – 69		9	24 + 9 = 33
69 – 77		1	33 + 1 = 34
77 – 85		8	34 + 8 = 42
85 – 93		9	42 + 9 = 51
93 – 101		9	51 + 9 = 60
Total		60	

বিঃ দ্রঃ পরীক্ষার নাম্বার 101 হয় না, কিন্তু এখানে ক্লাস ইন্টারভাল width এর কারণে 101 হয়ে গেছে।

এখন, এই ডেটা টেবিলের মধ্যমান (x_i) আর ফ্রিকুয়েন্সি (f_i) বের করে পাইঃ

ক্লাস ইন্টারভাল	মধ্যমান (x_i)	ফ্রিকুয়েন্সি (f_i)	$f_i x_i$
45 – 53	49	14	686
53 – 61	57	10	570
61 – 69	65	9	585
69 – 77	73	1	73
77 – 85	81	8	648
85 – 93	89	9	801
93 – 101	97	9	873
Total	$\sum_{i=1}^k x_i = 511$	$\sum_{i=1}^k f_i = 60$	$\sum_{i=1}^k f_i x_i = 4236$

$$\text{মধ্যমান} = \frac{\text{Lower Limit} + \text{Upper Limit}}{2}$$

উদাহরণঃ

$$45 - 53 \text{ এর মধ্যমান} = \frac{45+53}{2} = 49$$

$$53 - 61 \text{ এর মধ্যমান} = \frac{53+61}{2} = 57$$

- Mean Absolute Deviation এর ফর্মুলাঃ

Sample এর ক্ষেত্রে Mean Absolute Deviation এর সূত্র –

$$\bar{x}_{ad} = \frac{1}{n} \sum_{i=1}^k (f_i |x_i - \bar{x}|)$$

Population এর ক্ষেত্রে Mean Absolute Deviation এর সূত্র –

$$\mu_{ad} = \frac{1}{N} \sum_{i=1}^k (f_i |x_i - \mu|)$$

এখানে x_i হলো মধ্যমান।

কোনো ডেটাসেটের ক্ষেত্রে Population বা Sample উল্লেখ না থাকলে আমরা সাধারণভাবে Sample mean বের করি, এবং Sample Mean Absolute Deviation বের করি। কিন্তু আমরা আগেই বলেছি যে দশম শ্রেণীর “ঘ” শাখার ৬০ জন শিক্ষার্থীর পদার্থবিজ্ঞান পরীক্ষার নম্বরের ডেটাসেট হলো Population। তাই এক্ষেত্রে আমরা Population mean (μ) এবং Population Mean Absolute Deviation (μ_{ad}) বের করবো।

এই দশম শ্রেণীর “ঘ” শাখার ৬০ জন শিক্ষার্থীর পদার্থবিজ্ঞান পরীক্ষার নম্বরের Population mean আমরা আগেই বের করেছি।

$$\text{Population mean, } \mu = 70.6$$

সুতরাং,

ক্লাস ইন্টারভাল	মধ্যমান (x_i)	$ x_i - \mu $	ফ্রিকুয়েন্সি (f_i)	$f_i x_i - \mu $
45 – 53	49	-21.6	14	-302.4
53 – 61	57	-13.6	10	-136
61 – 69	65	-5.6	9	-50.4
69 – 77	73	2.4	1	2.4
77 – 85	81	10.4	8	83.2
85 – 93	89	18.4	9	165.6
93 – 101	97	26.4	9	237.6
Total	$\sum_{i=1}^k x_i = 511$		$\sum_{i=1}^k f_i = 60 = N$	$\sum_{i=1}^k (f_i x_i - \mu) = 0$

Sample এর ক্ষেত্রে $\sum_{i=1}^k f_i$ হলো n । Population এর ক্ষেত্রে $\sum_{i=1}^k f_i$ হলো N ।

কারণ ফ্রিকুয়েন্সির সমষ্টি হলো উপাত্তের সংখ্যা।

এখন, দশম শ্রেণীর “ঘ” শাখার ৬০ জন শিক্ষার্থীর পদার্থবিজ্ঞান পরীক্ষার নম্বরের Mean Absolute Deviation

$$\mu_{ad} = \frac{1}{N} \sum_{i=1}^k (f_i |x_i - \mu|) = \frac{1}{60} \times 0 = 0$$

অর্থাৎ দশম শ্রেণীর “ঘ” শাখার ৬০ জন শিক্ষার্থীর পদার্থবিজ্ঞান পরীক্ষার নম্বরের Mean Absolute Deviation হলো ০.

Median Absolute Deviation:

আমরা “০০১-০১৩ Numerical Data (পর্ব-৫)” এ জেনেছি যে দশম শ্রেণীর “ঘ” শাখার ৬০ জন শিক্ষার্থীর পদার্থবিজ্ঞান পরীক্ষার নম্বরের Median হলো ৬১ – ৬৯ ক্লাস।

এখন,

ক্লাস ইন্টারভাল	মধ্যমান (x_i)	ফ্রিকুয়েন্সি (f_i)	কিউমুলেটিভ ফ্রিকুয়েন্সি (cf_i)
৪৫ – ৫৩	৪৯	১৪	১৪
৫৩ – ৬১	৫৭	১০	১৪ + ১০ = ২৪
৬১ – ৬৯	৬৫	৯	২৪ + ৯ = ৩৩
৬৯ – ৭৭	৭৩	১	৩৩ + ১ = ৩৪
৭৭ – ৮৫	৮১	৮	৩৪ + ৮ = ৪২
৮৫ – ৯৩	৮৯	৯	৪২ + ৯ = ৫১
৯৩ – ১০১	৯৭	৯	৫১ + ৯ = ৬০
Total	$\sum_{i=1}^k x_i = 511$	$\sum_{i=1}^k f_i = 60$	

আমরা “০০১-০১৩ Numerical Data (পর্ব-৫)” এ বের করেছি যে median, $\tilde{x} = 67$ কিন্তু Median Absolute Deviation বের করার ক্ষেত্রে median হবে $\tilde{x} = 65$

Median Absolute Deviation এর সূত্রঃ

$$\tilde{x}_{ad} = \text{median}(|x_i - \tilde{x}|)$$

$$\tilde{x}_{ad} = \text{median}(|x_1 - \tilde{x}|, |x_2 - \tilde{x}|, \dots, |x_k - \tilde{x}|)$$

তাহলে দশম শ্রেণীর “ঘ” শাখার ৬০ জন শিক্ষার্থীর পদার্থবিজ্ঞান পরীক্ষার নম্বরের Median Absolute Deviation হবেঃ

$$\tilde{x}_{ad} = \text{median}(|49 - \tilde{x}|, |57 - \tilde{x}|, |65 - \tilde{x}|, |73 - \tilde{x}|, |81 - \tilde{x}|, |89 - \tilde{x}|, |97 - \tilde{x}|)$$

$$\Rightarrow \tilde{x}_{ad} = \text{median}(|49 - 65|, |57 - 65|, |65 - 65|, |73 - 65|, |81 - 65|, |89 - 65|, |97 - 65|)$$

$$\Rightarrow \tilde{x}_{ad} = \text{median}(16, 8, 0, 8, 16, 24, 32)$$

$$\Rightarrow \tilde{x}_{ad} = \text{median}(0, 8, 8, 16, 16, 24, 32) \text{ [Sorting the data]}$$

$$\Rightarrow \tilde{x}_{ad} = 16$$

সুতরাং দশম শ্রেণীর “ঘ” শাখার ৬০ জন শিক্ষার্থীর পদার্থবিজ্ঞান পরীক্ষার নম্বরের Median Absolute Deviation হবে, $\tilde{x}_{ad} = 16$