



আমার বিজ্ঞকথা



টপিকঃ

ডেটা সায়েন্স

০০১-০১৫ Numerical
Data (পর্ব-৭)





০০১-০১৫ Numerical Data (পর্ব-৭)

আমরা Continuous Numerical Data এর ক্ষেত্রে Dispersion এর range, variance, standard deviation শিখেছি, এবার Continuous Numerical Data এর ক্ষেত্রে Dispersion এর IQR শিখবো।

Continuous Numerical Data এর ক্ষেত্রে IQR:

Percentile:

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, 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]

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

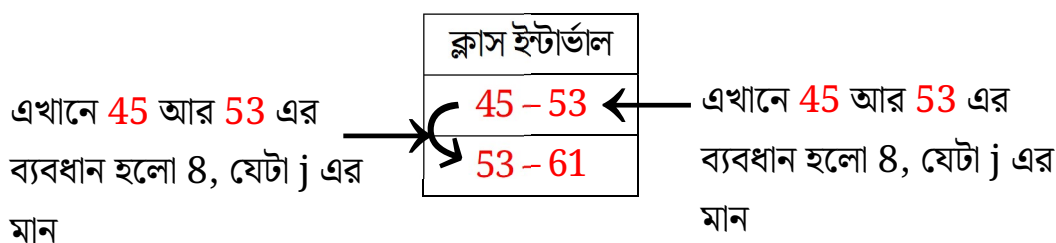
অর্থাৎ ৭টি ক্লাস ইন্টারভাল তৈরি হবে।

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

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	

$k = 7$
এখানে ৭ টি ক্লাস ইন্টারভাল রয়েছে

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

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

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

Continuous Data এর ক্ষেত্রে g-তম Percentile এর লোকেশন এর সূত্রঃ-

$$L_g = \frac{g \sum_{i=1}^k f_i}{100}$$

এবং g-তম Percentile এর সূত্রঃ-

$$P_g = L + \left(\frac{L_g - m}{f} \times j \right)$$

এখানে,

L = g-তম পারসেন্টাইলের Lower value

L_g = g-তম Percentile এর লোকেশন

m = g-তম পারসেন্টাইলের পূর্ববর্তী ক্লাসের
কিউমুলেটিভ ফ্রিকুয়েন্সি

f = g-তম পারসেন্টাইলের ফ্রিকুয়েন্সি

j = class width

এখন,

ক্লাস ইন্টারভাল	মধ্যমান (x_i)	ফ্রিকুয়েন্সি (f_i)	কিউমুলেটিভ ফ্রিকুয়েন্সি	$f_i x_i$
45 – 53	49	14	14	686
53 – 61	57	10	14 + 10 = 24	570
61 – 69	65	9	24 + 9 = 33	585
69 – 77	73	1	33 + 1 = 34	73
77 – 85	81	8	34 + 8 = 42	648
85 – 93	89	9	42 + 9 = 51	801
93 – 101	97	9	51 + 9 = 60	873
Total	$\sum_{i=1}^k x_i = 511$	$\sum_{i=1}^k f_i = 60$		$\sum_{i=1}^k f_i x_i = 4236$

এখন, মনে করুন এই দশম শ্রেণীর “ঘ” শাখার ৬০ জন শিক্ষার্থীর পদার্থবিজ্ঞান পরীক্ষার নম্বরের মধ্যে আমরা P_{53} (53 percentile) বের করতে চাই। তাহলে,

$$L_{53} = \frac{53 \times \sum_{i=1}^k f_i}{100} = \frac{53 \times 60}{100} = 31.8$$

এখন 31.8 হচ্ছে 53 – 61 এর কিউমুলেটিভ ফ্রিকুয়েন্সির 24 এর চেয়ে বড় আর 61 – 69 এর কিউমুলেটিভ ফ্রিকুয়েন্সির 33 এর চেয়ে ছোট।

সুতরাং এই $L_{53} = 31.8$ ভ্যালুর ক্ষেত্রে বলা যায় যে 61 – 69 হচ্ছে P_{53} (53 percentile) এর ক্লাস। আর 53 – 61 হচ্ছে P_{53} (53 percentile) এর পূর্ববর্তী ক্লাস।

তাহলে, $L = 61$ [যেহেতু 61 – 69 হচ্ছে P_{53} (53 percentile) এর ক্লাস আর এই ক্লাসের Lower value হলো 61]

$m = 24$ [যেহেতু 53 – 61 হচ্ছে P_{53} (53 percentile) এর পূর্ববর্তী ক্লাস আর এই ক্লাসের কিউমুলেটিভ ফ্রিকুয়েন্সি হলো 24]

$f = P_{53}$ (53 percentile) এর ক্লাস **61 – 69** এর ফ্রিকুয়েন্সি = 9

ক্লাস ইন্টারভাল	মধ্যমান (x_i)	ফ্রিকুয়েন্সি (f_i)	কিউমুলেটিভ ফ্রিকুয়েন্সি	$f_i x_i$	স্ট্যাটাস
45 – 53	49	14	14	686	
53 – 61	57	10	14 + 10 = 24 [m] [$L_{53} > 24$]	570	P_{53} (53 percentile) এর পূর্ববর্তী ক্লাস
61 – 69 [L = 61]	65	9 [f]	24 + 9 = 33 [$L_{53} < 33$]	585	P_{53} (53 percentile) এর ক্লাস
69 – 77	73	1	33 + 1 = 34	73	
77 – 85	81	8	34 + 8 = 42	648	
85 – 93	89	9	42 + 9 = 51	801	
93 – 101	97	9	51 + 9 = 60	873	
Total	$\sum_{i=1}^k x_i = 511$	$\sum_{i=1}^k f_i = 60$		$\sum_{i=1}^k f_i x_i = 4236$	

অতএব, P_{53} (53 percentile) হলো:-

$$P_{53} = L + \left(\frac{L_{53} - m}{f} \times j \right) = \mathbf{61} + \left(\frac{31.8 - \mathbf{24}}{9} \times 8 \right) = 67.93$$

এভাবেই আমরা percentile বের করা শিখে গেছি।

Quartiles:

25 Percentile (P_{25}) কে first quartile (Q_1) বলা হয়। 50 Percentile (P_{50}) কে second quartile (Q_2) বলা হয়। 75 Percentile (P_{75}) কে third quartile (Q_3) বলা হয়।

IQR এর ক্ষেত্রে আমাদেরকে first quartile (Q_1) আর third quartile (Q_3) বের করতে হবে।

অতএব, first quartile (Q_1) ও third quartile (Q_3) এর ক্ষেত্রে,

$$L_{25} = \frac{25 \times \sum_{i=1}^k f_i}{100} = \frac{25 \times 60}{100} = 15$$

$$L_{75} = \frac{75 \times \sum_{i=1}^k f_i}{100} = \frac{75 \times 60}{100} = 45$$

তাহলে এক্ষেত্রে -

ক্লাস ইন্টারভাল	মধ্যমান (x_i)	ফ্রিকুয়েন্সি (f_i)	কিউমুলেটিভ ফ্রিকুয়েন্সি	$f_i x_i$	স্ট্যাটাস
45 – 53	49	14	14 [m_1] [$L_{25} > 14$]	686	25 Percentile (P_{25}) বা first quartile (Q_1) এর পূর্ববর্তী ক্লাস
53 – 61 [$L_{q1} = 53$]	57	10 [f_{q1}]	14 + 10 = 24 [$L_{25} < 24$]	570	25 Percentile (P_{25}) বা first quartile (Q_1) এর ক্লাস
61 – 69	65	9	24 + 9 = 33	585	
69 – 77	73	1	33 + 1 = 34	73	
77 – 85	81	8	34 + 8 = 42 [m_3] [$L_{75} > 42$]	648	75 Percentile (P_{75}) বা third quartile (Q_3) এর পূর্ববর্তী ক্লাস
85 – 93 [$L_{q3} = 85$]	89	9 [f_{q3}]	42 + 9 = 51 [$L_{75} < 51$]	801	75 Percentile (P_{75}) বা third quartile (Q_3) এর ক্লাস
93 – 101	97	9	51 + 9 = 60	873	
Total	$\sum_{i=1}^k x_i = 511$	$\sum_{i=1}^k f_i = 60$		$\sum_{i=1}^k f_i x_i = 4236$	

অতএব, first quartile (Q_1) হলোঃ-

$$Q_1 = P_{25} = L_{q1} + \left(\frac{L_{25} - m_1}{f_{q1}} \times j \right) = 53 + \left(\frac{15 - 14}{10} \times 8 \right) = 53.80$$

অতএব, third quartile (Q_3) হলোঃ-

$$Q_3 = P_{75} = L_{q3} + \left(\frac{L_{75} - m_3}{f_{q3}} \times j \right) = 85 + \left(\frac{45 - 42}{9} \times 8 \right) = 87.67$$

IQR:

Interquartile Range কে সংক্ষেপে IQR বলা হয়। Interquartile Range হলো First Quartile (Q_1) আর Third Quartile (Q_3) এর মধ্যে পার্থক্য। Interquartile Range এর সূত্রঃ

$$IQR = Q_3 - Q_1$$

তাহলে দশম শ্রেণীর “ঘ” শাখার ৬০ জন শিক্ষার্থীর পদার্থবিজ্ঞান পরীক্ষার নম্বর ডেটাসেটের IQR হলোঃ-

$$IQR = Q_3 - Q_1 = 87.67 - 53.80 = 33.87$$

তো আমরা Continuous Numerical Data এর Interquartile Range বের করা শেখার মাধ্যমে আমরা পুরোপুরিভাবে Continuous Numerical Data এর Dispersion শেখা শেষ করেছি। ধন্যবাদ।