2.5 সহ-সম্বন্ধ এবং সহগাঙ্ক [Correlation and Co-efficient of Correlation]

সহ-সম্বন্ধ হল কোনো দুটি চলের (Variable) মধ্যে এমন একটি সম্বন্ধ যার মাধ্যমে চল দুটির পারস্পরিক নির্ভরতার মাত্রা নির্ণয় করা যেতে পারে। *Guilford*-এর মতে 'যখন দুটি শ্রেণি বা ঘটনার মধ্যে একটি পরিমাণগত বা গুণগত সম্বন্ধ থাকে, তখন সেই সম্বন্ধ নির্ধারণ, পরিমাপ ও সঠিকভাবে প্রকাশ করার উপযুক্ত পরিসংখ্যানমূলক সূত্র হল সহ-সম্বন্ধ।

(When there is a relationship of a quantitative or qualitative nature between two sets of phenomena, the appropriate statistical test for discovering and measuring the relationship and expresing it in a precise way is known as correlation).

যদি দুটি চলরাশি (Variable) পরস্পরের সাথে এমন ভাবে সম্পর্কিত হয় যে একটির পরিবর্তন ঘটলে অপরটিও পরিবর্তিত হয় তখন বলা হয় দুটি রাশির মধ্যে সহগতি রয়েছে। যেমন ঃ একটি বৃত্তের ব্যাসের কমা বাড়ার সঙ্গে সঙ্গে তার পরিধি কমে বাড়ে অর্থাৎ ব্যাস ও পরিধি পরস্পরের সঙ্গে সম্পর্কিত, শিক্ষাক্ষেত্রে বুদ্ধি ও গণিতে পারদর্শিতা পরস্পরের মধ্যে সম্পর্কিত ইত্যাদি।

যে গাণিতিক সূচক দ্বারা দুটি চলের মধ্যকার সহগতির পরিমাপ ও গুণগত বৈশিষ্ট্য (অর্থাৎ ধনাত্মক, ঋণাত্মক বা শূন্য) প্রকাশ করা হয় তাকেই বলে সহগতির সহগাঙ্ক (Co-efficient of Correlation)। সহগাঙ্ককে সাধারণত r দ্বারা প্রকাশ করা হয়। r এর মান –1 থেকে +1 পর্যন্ত হতে পারে।

Karl Pearson দুটি চলের মধ্যে সহ-সম্বন্ধ গুণাঙ্ক নির্ণয়ের পদ্ধতি প্রবর্তন করেন। পিয়ারসন পদ্ধতিতে যখন সহ-সম্বন্ধ নির্ণয় করা হয় তখন সর্বদা ধরে নেওয়া হয় ওই দুটি চলের মধ্যে সরলরৈখিক সম্পর্ক আছে। এই সম্বন্ধকে Product moment coefficient বলে। এর মান '–'1 থেকে '+'1 পর্যন্ত বিস্তৃত হতে পারে। যখন দুটি চলের মধ্যে একটির মানের বৃদ্ধি বা হ্রাসের ফলে অপরটির মান যথাক্রমে সমপরিমাণ বৃদ্ধি বা হ্রাস পায় অর্থাৎ চলরাশিদ্বয়ের মানের পরিবর্তন সমমুখী হয় তখন সহগতি গুণাঙ্ক +1 হয়। যেমন একটি শ্রেণির ছাত্রদের I.Q ও পরীক্ষার ফলাফল একই হারে একই দিকে পরিবর্তিত হলে তখন তাদের মধ্যে সহ-সম্বন্ধ গুণাঙ্ক হবে +1। আবার যদি একটি চলের মানের হ্রাস বা বৃদ্ধিতে অপরটির মান যথাক্রমে সম পরিমাণ বৃদ্ধি বা হ্রাস পায় অর্থাৎ চলরাশিদ্বয়ের মানের পরিবর্তন বিপরীতমুখী হয় তখন সহসম্বন্ধ গুণাঙ্ক –1 ধরা হয়। যেমন কোনো একটি শ্রেণিতে ছাত্ররা বাংলাতে ভালো নম্বর পেয়েছে এবং আনুপাতিক হারে অঙ্কে খারাপ নম্বর পেয়েছে। এক্ষেত্রে উভয় বিষয়ের মধ্যে সহ-সম্বন্ধ গুণাঙ্ক হল –1.

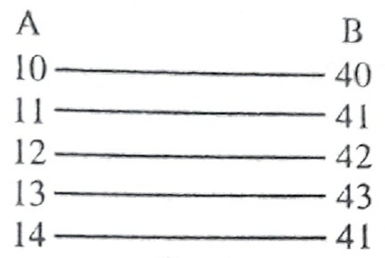
আবার দুটি চলরাশির মধ্যে এমন সম্পর্ক থাকে যে একটির মানের পরিবর্তনের উপর অপরটির মান নির্ভরশীল নয়। তাহলে সেরূপ ক্ষেত্রে রাশি দুটি পরস্পর সম্বন্ধহীন (uncorrelated) হবে। এক্ষেত্রে সহ-সম্বন্ধ গুণাঙ্ক হল 0. যেমন ছাত্রদের উচ্চতার সঙ্গে তাদের অঙ্কের ফলাফলের কোনো সম্পর্ক নেই।

তাই সহ-সম্বন্ধের গুণাঙ্কের মান –1 থেকে +1 পর্যন্ত যে-কোনো মান হতে পারে।

সহগতি তিন ধরনের হতে পারে—(1) ধনাত্মক সহগতি (Positive Correlation), (2) ঋণাত্মক সহগতি (Negative Correlation) (3) শূন্য সহগতি (Zero Correlation)।

ধনাত্মক সহগতি ঃ

যদি দুটি চলরাশির একটি বৃদ্ধি পেলে অপরটি বৃদ্ধি পায় আবার একটি হ্রাস পেলে অপরটিও হ্রাস পায় তখন চলরাশি দুটির মধ্যে ধনাত্মক সহগতি আছে বলা হয়। যখন দুটি চলের মধ্যে পরিপূর্ণ ধনাত্মক সহগতি থাকে অর্থাৎ একটি চল রাশির যে কোনো পরিমাণ পরিবর্তনে অপরটির সমপ্রকৃতির সমপরিমাণ পরিবর্তন হয় তখন চলরাশি দুটির মধ্যে সহগতির সহগাঙ্ক (r) হয় +1। ধরা যাক কোনো বিশেষ পরিস্থিতিতে কোনো একটি শ্রেণির ছাত্রদের বুদ্ধ্যঙ্ক যে হারে বাড়ছে তাদের প্রাপ্ত পরীক্ষার নম্বরও সেই হারে বাড়ছে। এক্ষেত্রে বুদ্ধ্যঙ্ক ও পরীক্ষার ফলাফলের মধ্যে সহগতির সহগাঙ্ক +1। তাপমাত্রা ও থার্মোমিটারে পারদ স্তম্ভের উচ্চতার মধ্যে সহগতির সহগাঙ্ক +1।



(চিত্র -১)

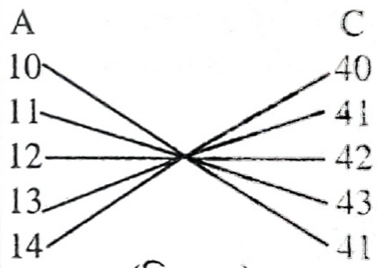
চিত্র -১-এ A স্তম্ভে ও B স্তম্ভে একদল ছাত্রের দুটি বিষয়ে নম্বর আছে। চিত্রে দেখা যাচ্ছে যারা A বেশি পেয়েছে তারা B তেও বেশি পেয়েছে। এক্ষেত্রে A-র পরিবর্তনের সঙ্গে B-এর সমপরিমাণ পরিবর্তন হয়েছে। অর্থাৎ এক্ষেত্রে r = +1।

ঋণাত্মক সহগতি ঃ

যখন দুটি চলরাশির মধ্যে বিপরীত সম্পর্ক থাকে অর্থাৎ একটি চলরাশির বৃদ্ধিতে অপরটি হ্রাস পায় তখন রাশি দুটির মধ্যে ঋণাত্মক সহগতি আছে বলা হয়।

যখন দুটি চলের মধ্যে সর্বক্ষেত্রে সম্পূর্ণ বিপরীত সম্পর্ক থাকে তখন চলরাশির দুটির মধ্যে সহগতির সহগাঙ্ক r-এর মান -1 হয়।

যেমন ঃ কোনো বিশেষ পরিস্থিতিতে দেখা গেল, যে সব ছাত্ররা ইংরাজিতে ভালো নম্বর পেয়েছে তারা আনুপাতিক হারে অঙ্কে খারাপ নম্বর পেয়েছে বা যারা অঙ্কে ভালো নম্বর পেয়েছে, আনুপাতিক হারে তারা ইংরাজিতে খারাপ নম্বর পেয়েছে। এখানে ইংরাজি ও অঙ্কের মধ্যে সহগতির সহগাঙ্ক -1। চিত্র - ২-এ ঋণাত্মক সহগতি দেখানো হল।

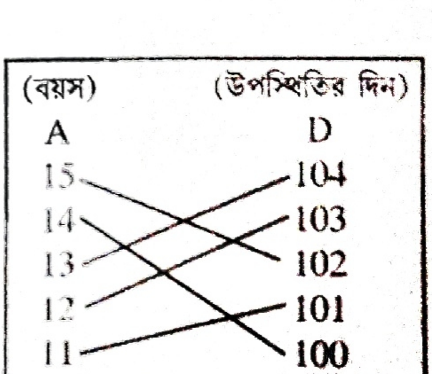


(চিত্রে -২)

শূন্য সহগতি ঃ

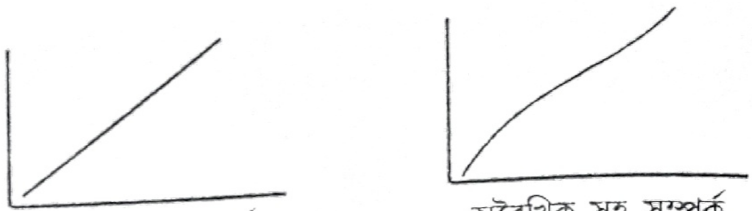
যখন দুটি চলের মধ্যে একটির কোনো পরিবর্তন অপরকে প্রভাবিত করতে পারে না তখন রাশি দুটির মধ্যে শূন্য সহগতি আছে বলা হয়। এক্ষেত্রে সহগতির সহগাঙ্ক r-এর মান শূন্য।

চিত্র - ৩-এ দেখা যাচ্ছে ছাত্রদের বয়সের সঙ্গে উপস্থিত দিনের কোনো সম্পর্ক নেই। এক্ষেত্রে সহগতির সহগাঙ্ক প্রায় শূন্যের কাছাকাছি।



(চিত্রে -৩)

সহ-সম্বন্ধ রৈখিক (linear) বা অরৈখিক (non linear) হতে পারে। দুটি চলের মানের পরিবর্তনের অনুপাত ধ্রুবক হলে তাদের সহ-সম্বন্ধ রৈখিক। অপর দিকে চল দুটির মানের অনুপাত ধ্রুবক না হলে তাদের সহ-সম্বন্ধের মানকে অরৈখিক বলে। যেমন মানুষের বয়স ও শারীরিক শক্তির মধ্যে সহ-সম্বন্ধ অরৈখিক।



রৈখিক সহ সম্পর্ক   অরৈখিক সহ সম্পর্ক

সহগতির সহগাঙ্ক নির্ণয়ের বিভিন্ন পদ্ধতি আছে। এর মধ্যে রৈখিক সম্পর্কের ক্ষেত্রে দুটি পদ্ধতি রয়েছে। যেমন—(1) প্রোডাক্ট মোমেন্ট পদ্ধতি (Product Moment Method) এবং (2) র‍্যাঙ্ক পার্থক্য পদ্ধতি (Rank Difference Method)।

(1) প্রোডাক্ট মোমেন্ট পদ্ধতি (Product Moment Method) ঃ

রৈখিক সম্পর্ক যুক্ত দুটি চলের মধ্যে সহগতির সহগাঙ্ক নির্ণয়ের জন্য যে পদ্ধতি রয়েছে তাকে প্রোডাক্ট মোমেন্ট পদ্ধতি বলে। এই পদ্ধতিতে সহগতির সহগাঙ্ককে 'r' দিয়ে চিহ্নিত করা হয়।

পিয়ারসনের প্রোডাক্ট মোমেন্ট গুণাঙ্ক (Co-efficient correlation) হল এমন একটি অনুপাত যা একটি চলের পরিবর্তনের সাথে অপর চলের যে পরিবর্তন হচ্ছে তার সীমা প্রকাশ করে (Product moment coefficient of correlation is a kind of ratio that express the extent to which change in the one variable is accompanied by change in another variable)। এই পদ্ধতিতে মান নির্ণয়ের বিভিন্ন পদ্ধতি আছে তার মধ্যে কয়েকটি হল ঃ

অবিন্যস্ত স্কোরের সহগাঙ্ক ঃ X ও Y যদি দুটি চলক হয় এবং M_1 ও M_2 যদি ওই দুই চল রাশির গড় হয় এবং শ্রেণি দুটিতে স্কোর সংখ্যা যদি N হয় তবে x হল X শ্রেণির প্রতিটি রাশিকে M_1 দ্বারা বিয়োগ করে যে চ্যুতি পাওয়া যায় এবং অনুরূপে y হল Y শ্রেণির প্রতি রাশিকে M_2 দ্বারা বিয়োগ করে যে চ্যুতি পাওয়া যায়। চ্যুতি দ্বয়ের গুণফল xy। এইভাবে সব xy যোগ করে পাওয়া যায় Σxy এরপর X শ্রেণির চল রাশির S.D হয় σ_x ও Y শ্রেণির চলরাশির S.D হয় σ_y তখন প্রোডাক্ট মোমেন্ট পদ্ধতি অনুযায়ী সহগতির সহগাঙ্ক নির্ণয়ের সূত্র হল—

$$(r_{xy}) = \frac{\sum xy}{N\sigma_x\sigma_y} \quad [x = X - M_1;\ y = Y - M_2]$$

x = X চল শ্রেণির যে কোনো স্কোর

y = Y চল শ্রেণির যে কোনো স্কোর

σ_x, σ_y = যথাক্রমে X ও Y চলক শ্রেণির S.D.

প্রত্যক্ষভাবে স্কোর থেকে ঃ (From original Score)

$$r_{xy} = \frac{N\sum XY - \sum X \sum Y}{\sqrt{[N\sum X^2 - (\sum X)^2] \times [N\sum Y^2 - (\sum Y)^2]}}$$

যেখানে N = স্কোর সংখ্যা

ΣXY = X শ্রেণির স্কোর ও Y শ্রেণির স্কোরগুলির গুণফল-এর যোগফল

ΣX = X শ্রেণির স্কোরসমূহের যোগফল ও ΣY = Y শ্রেণির স্কোর সমূহের যোগফল

ΣX^2 ও ΣY^2 হল যথাক্রমে X ও Y শ্রেণির স্কোরগুলির বর্গের সমষ্টি

উদাহরণ ঃ নিম্নলিখিত চলক দুটির মধ্যে সহগতির সম্পর্ক মূল স্কোর থেকে নির্ণয় করো। (Calculate the 'r' belower the following two variables from original scale)

Subject	Test-I X	Test-II Y	X^2	Y^2	XY
A	50	22	2500	484	1100
B	54	25	2916	625	1350
C	56	34	3136	1156	1904
D	59	28	3481	784	1652
E	60	26	3600	676	1560
F	62	30	3844	900	1860
G	61	32	3721	1024	1952
H	65	30	4225	900	1950
I	67	28	4449	784	1876
J	71	34	5041	1156	2414
K	71	36	5041	1296	2556
L	74	40	5478	1600	2960
	ΣX=750	ΣY^2=365	ΣX^2=47470	ΣY=11385	ΣXY=23134

আমরা জানি মূল স্কোর থেকে 'r' নির্ণয়ের সূত্র—

$$r = \frac{N\sum XY - \sum X \times \sum Y}{\sqrt{\left[N\left(\sum X^2\right)-\left(\sum X^2\right)\right]\left[N\left(\sum Y^2\right)-\left(\sum Y^2\right)\right]}}$$

উপরোক্ত মান বসিয়ে আমরা পাই—

$$r = \frac{23134 \times 12 - 750 \times 365}{\sqrt{[12 \times 47470 - 562500][12 \times 11385 - 133225]}}$$

$= \cdot 68$

$\therefore\ r = \cdot 68$

যখন 'বিচ্যুতি' শ্রেণির গড় থেকে (When deviations are from Means of variables)

Calculate r from ungrouped score when deviations are ..... from Mean of the score. (অবিচ্ছিন্ন স্কোরের গড় থেকে যখন বিচ্যুতি নেওয়া হয় তখন 'r' নির্ণয় করার পদ্ধতি)

Subject	Test-I X	Test-II Y	x	y	x^2	y^2	xy
A	50	22	–12·5	–8·4	156·25	70·56	105·00
B	54	25	–8·5	–5·4	72·25	29·16	45·90
C	56	34	–6·5	3·6	42·25	12·96	–23·40
D	59	28	–3·5	–2·4	12·25	5·76	8·40
E	60	26	–2·5	–4·4	6·25	19·36	11·00
F	62	30	– ·5	– ·4	·25	·16	–2·40
G	61	32	–1·5	1·6	2·25	2·56	–1·00
H	65	30	2·5	– ·4	6·25	·10	–10·80
I	67	28	4·5	–2·4	20·25	5·76	30·60
J	71	34	8·5	3·6	72·25	12·96	47·60
K	71	36	8·5	5·6	72·25	31·36	110·40
L	74	40	11·5	9·6	132·25	92·16	0
	750	**365**			**569·00** (Σx^2)	**282·92** (Σy^2)	**321·50** (Σxy)

আমরা জানি,

$$r = \frac{\Sigma xy}{N.\sigma x.\sigma y} = \frac{321\cdot 50}{12} = \cdot 79 \text{ অর্থাৎ, } r = \cdot 79$$

অবিচ্ছিন্ন স্কোরের কল্পিত গড় পদ্ধতি বিবেচনা করে 'r' নির্ণয় (Calculation of r from deviation from assumed Mean) ঃ

Calculate 'r' from the following two distribution considering deviation from assumed Mean.

Subject	Test-I X	Test-II Y	x′	y′	x'^2	y'^2	x′y′
A	50	22	–10	–8	100	64	80
B	54	25	–6	–5	36	25	30
C	56	34	–4	4	16	16	–16
D	59	28	–1	–2	1	4	2
E	60	26	0	–4	0	16	0
F	62	30	2	0	4	0	0
G	61	32	1	2	1	4	2
H	65	30	5	0	25	0	0
I	67	28	7	–2	49	4	–14
J	71	34	11	4	121	16	44
K	71	36	11	6	121	36	66
L	74	40	14	10	196	100	140
					670 ($\Sigma x'^2$)	**285** ($\Sigma y'^2$)	**334** ($\Sigma x'y'$)

AMX = 60·0
MX = 62·5

AMY = 30·0
MY = 30·4

Cx = 2·5
Cx^2 = 6·25

Cy = 4
Cy^2 = ·16

আমরা জানি এক্ষেত্রে,

$$r = \frac{\frac{\sum x'y'}{N} - CxCy}{\sigma y' \sigma' x} \qquad \sigma' x = \sqrt{\frac{670}{12} - 6 \cdot 25} = 7 \cdot 04$$

$$r = \frac{\frac{334}{12} - 1 \cdot 00}{7 \cdot 04 \times 4 \cdot 86} \qquad \sigma' y = \sqrt{\frac{285}{12} - \cdot 16} = 4 \cdot 86$$

$= \cdot 78$ অর্থাৎ $rxy = \cdot 78$

বিন্যস্ত স্কোরের পিয়ারসনের 'r' নির্ণয় করার উপায় (Calculation of Pearson's 'r' from Group Data) ঃ

N যখন বড়ো বা মাঝারি সাইজের তখন Scattergram-এর সাহায্যে 'r' নির্ণয় করা উচিত। এক্ষেত্রে 'r' নির্ণয়ের সূত্র হল ঃ

$$r = \frac{N\sum fxy - (\sum fx)(\sum fy)}{\sqrt{[N\sum fx^2 - (\sum fx)^2][N(\sum fy^2 - (\sum fy)^2]}}$$

নিম্নে এর একটি উদাহরণ দেওয়া হল ঃ

Mathematics Test Score (X)

Language Test Score (Y)	12-13	14-15	16-17	18-19	20-21	22-23	24-25	f	y	fy	fy²	Σfxy
35-37					0 1 0		6 1 6	2	+3	6	18	6
32-34					0 6 0	2 3 6		9	+2	18	36	6
29-31		-3 1 -3	-2 2 -4	-1 6 -6	0 8 0	1 1 1		18	+1	18	18	−12
26-28		0 4 0	0 4 0	0 6 0	0 11 0	0 4 0	0 1 0	30	0	0	0	0
23-25	4 2 8	3 1 3	2 6 12	1 5 5	0 4 0	-1 1 -1		19	−1	−19	10	27
20-22	6 3 18	4 1 4	2 1 2					5	−2	−10	20	42
f	5	7	13	17	30	9	2	83		13	111	69
x	−4	−3	−2	−1	0	+1	+2					
fx	−20	−21	−26	−17	0	9	4	−71				
fx²	80	63	52	+7	0	9	8	229				
fxy	32	12	12	1	0	6	6	69				

$$r = \frac{(83)(69) - (-71)(13)}{\sqrt{[83 \times 229 - (-71)^2][83 \times 111 - (13)^2]}}$$

$$= \frac{5727 - 923}{\sqrt{[19007 - 5041][9213 - 169]}}$$

$$= \frac{4804}{(13966)(90 \cdot 44)}$$

$$= 0 \cdot 43$$

Σfx, Σfx^2, Σfy, Σfy^2 এবং Σfxy নিম্নোক্তভাবে নির্ণয় করা হয়—

Step I :

শিক্ষার্থীর Language Test স্কেলকে Scattergram-এর ডান দিকে 'f' column-এ বিন্যস্ত করা হল। যে interval-এ সবচেয়ে অধিক frequency আছে সেখানে Language Test স্কেলের কল্পিত গড় ধরা হয় এবং row বরাবর একটা মোট লাইন টানা হয়। এখানে দেখা যাচ্ছে '26-28' এই Interval-এ পরিসংখ্যান সবচেয়ে বেশি অর্থাৎ কল্পিত গড় হল '27' এবং 'y'গুলি (কল্পিত গড় থেকে বিচ্যুতি) এর প্রেক্ষিতেই নির্ণয় করা হয়। এবার fy এবং fy^2 নির্ণয় করা হয়।

Step II :

85 জন শিক্ষার্থীর গণিতের স্কোর Scattergram-এর নাচে 'f' row-এ বিন্যস্ত করা হয়। এখানে পূর্বের মতো যেখানে সর্বোচ্চ পরিমাণে 'f' আছে সেখানে কল্পিত গড় ধরা হয় এবং মোটা লাইন অঙ্কন হয়। এখানে সেই intervalটা হল '20-21' অর্থাৎ গড় হল 20·5। এর থেকেই x বিচ্যুতিগুলি নেওয়া হয় এবং fx, fx^2 row-তে বিন্যস্ত করা হয়।

Step III :

নির্দিষ্ট cell বা কক্ষে যে 'f' আছে তার সঙ্গে x ও y গুণ করে fxy নির্ণয় করা হয়।

Step IV :

এর পরে সূত্র প্রয়োগ করে Correlation নির্ণয় করা হয়।

পিয়ারসনের প্রোডাক্ট মোমেন্ট পদ্ধতির অসুবিধা :

(1) সহগাঙ্ক নির্ণয়ের এই পদ্ধতি দীর্ঘ ও শ্রমসাধ্য।

(2) শিক্ষামূলক পরিমাপের ক্ষেত্রে আমরা যে তথ্য পাই তা সব সময় সাংখ্যমানের নাও হতে পারে সেক্ষেত্রে এই পদ্ধতি অসুবিধাজনক।

(3) শিক্ষার্থীদের যোগ্যতাকে গুণগত দিকে প্রকাশ করলে (প্রথম, দ্বিতীয়, তৃতীয় ইত্যাদি) এই পদ্ধতিতে সহগাঙ্ক নির্ণয় করা যায় না।

(4) 'r'-এর মান রৈখিক সম্পর্কের ক্ষেত্রে প্রযোজ্য। রৈখিক সম্পর্ক না থাকলে এই মান সহগতির প্রকৃত মান ব্যক্ত করে না।

(2) র‍্যাঙ্ক পার্থক্য পদ্ধতি (Rank Difference Method) :

যে সব চল রাশির মানসমূহকে সংখ্যা দ্বারা প্রকাশ করা যায় না যেমন বুদ্ধি, সৌন্দর্য প্রভৃতি গুণগত বৈশিষ্ট্যসমূহের ক্ষেত্রে দুটি চলের পারস্পরিক সম্পর্কের প্রকৃতিকে 1, 2, 3 ... ইত্যাদি র‍্যাঙ্ক

বা অবস্থানের প্রেক্ষিতে বিভিন্ন ব্যক্তিসমূহের নির্দিষ্ট বৈশিষ্ট্যের সাপেক্ষে ক্রমপ্রকাশ করা হয় তখন এই পদ্ধতি প্রয়োগ করা হয়। এইভাবে বৈশিষ্ট্যকে ক্রম অনুযায়ী বিন্যাসকরণকে Ranking বলে এবং যে সংখ্যা দিয়ে কোনো পদের বা ব্যক্তির ক্রম স্থির করা হয় তা ওই পদের সারিতে অবস্থানকে চিহ্নিত করে। এই ভাবে র‍্যাঙ্ক দুটি চলের দুটি শ্রেণি তৈরি করা হয়। এর পর চল দুটির র‍্যাঙ্কের পার্থক্য নির্ণয় করা হয়। এই পার্থক্য নির্ণয়কে কাজে লাগিয়ে র‍্যাঙ্ক পার্থক্য পদ্ধতিতে দুটি চলের মধ্যে সহগতি সহগাঙ্ক (Rank co-efficient of cerrelation by differences network) নির্ণয় করা যায়। একে Rank difference method বলে। ***Spearman*** এক্ষেত্রে সহগতিকে ρ দ্বারা চিহ্নিত করেন। এক্ষেত্রে যে সূত্রটি প্রযোজ্য সেটি হল ঃ

$$\rho = 1 - \frac{6\sum D^2}{N(N^2-1)}$$

$D = R_1 - R_2$ (দুটি র‍্যাঙ্কের পার্থক্য গাণিতিক প্রক্রিয়া বিবেচনা করা হয় না।)

N = মোট স্কোর সংখ্যা।

ρ এর মান −1 থেকে +1 এর মধ্যে থাকে। যখন এর মান হয় +1 তখন দুটি সারিতে ক্রম অনুযায়ী প্রত্যেক পদের Rank একই থাকে এবং −1 হলে ক্রম অনুযায়ী প্রত্যেক পদের অবস্থান একই ক্রমে বিপরীতমুখী হবে।

র‍্যাঙ্ক পার্থক্য পদ্ধতির সুবিধা ঃ

(1) গুণগত ও পরিমাণগত উভয় স্কোর এর ক্ষেত্রে সহগতির সহগাঙ্ক নির্ণয়ে ব্যবহার করা যায়।

(2) এই পদ্ধতিতে ρ (সহগতির সহগাঙ্ক) অপেক্ষাকৃত সহজে নির্ণয় করা যায়। এই কারণে মনোবিজ্ঞান, শিক্ষাতত্ত্ব ও সমাজশিক্ষায় এর অধিক ব্যবহার হয়।

Example 1 :

Compute the correlation between following data using Rank difference method.

x	15	18	22	17	19	20	16	14	17	22
y	41	40	42	50	45	38	46	41	40	39

Ans :

প্রথমে আমরা Data-গুলিকে Rank-এ বিন্যস্ত করব। সবচেয়ে বেশি স্কোরকে 1 Rank দেওয়া হয়। এইভাবে পরপর '2', '3' ইত্যাদি Rank দেওয়া হয়। একই স্কোর একাধিকবার থাকলে গড় র‍্যাঙ্ক হয়। উদাহরণস্বরূপ 'X' চলকে 22 স্কোর দুবার আছে। একই স্কোর হওয়ায় '2' Rank করা হয় না। গড় Rank করা হয়। এখানে গড় হল $\frac{1+2}{2} = 1\cdot5$ । উভয়কেই 1·5 Rank দেওয়া হয়। পরবর্তী স্কোরের অর্থাৎ '20' Rank হবে '3'।

X	Y	Rx	Ry	D	D^2
15	41	9	6·5	2·5	6·25
18	40	5	8	3	9·00
22	42	1·5	5	3·5	12·25
17	50	6·5	1	5·5	30·25
19	45	4	4	0	0
20	38	3	10	7	49·00
16	46	8	3	5	25·00
14	41	10	6·5	3·5	12·25
17	49	6·5	2	4·5	20·25
22	39	1·5	9	7·5	56·25
					$\Sigma D^2 = 220{\cdot}50$

সূত্র ব্যবহার করে পাই ঃ

$$\rho = 1 - \frac{6\sum D2}{N(N^2-1)} = 1 - \frac{6 \times 220\cdot50}{10(10^2-1)}$$

$$= 1 - \frac{1323}{10 \times 99} = 1 - \frac{1323}{990} = 1 - 1\cdot336$$

$$= -0\cdot336$$

$$= -\cdot336$$

Example 2 :

Use the rank difference method to compute the correlation co-efficient of the following distribution and comment on it.

x	10	12	16	18	22	17	19	20	19	18	15	21
y	12	10	19	15	19	15	17	21	22	13	11	18

Ans :

X	Y	Rx	Ry	D=Rx–Ry	D^2
10	12	12	10	2.0	4.00
12	10	11	12	1.0	1.00
16	19	9	3.5	5.5	30.25
18	15	6.5	7.5	1.0	1.00
22	19	1	3.5	2.5	6.25
17	15	8	7.5	0.5	0.25
19	17	4.5	6	1.5	2.25
20	21	3	2	1.0	1.00
19	22	4.5	1	3.5	12.25
18	13	6.5	9	2.5	6.25
15	11	10	11	1.0	1.00
21	18	2	5	3.0	9.00
					$\Sigma D^2 = 74.50$

$$r = 1 - \frac{6\sum D^2}{N(N^2-1)} = 1 - \frac{6 \times 74 \cdot 5}{12(12^2-1)} \quad [\text{এক্ষেত্রে } N=12]$$

$$= 1 - \frac{470}{12 \times 143} = 1 - 0.26 = 0.74$$

এক্ষেত্রে সহগতির সহগাঙ্কের মান + 0.74 এবং এই মান ধনাত্মক এবং উচ্চমান সম্পন্ন। তাই X এর মান বৃদ্ধিতে Y এর মানও বৃদ্ধি পাবে। আবার X এর মান হ্রাসে Y এর মানও হ্রাস পাবে।

সহগতি সহগাঙ্ক 'r' এর তাৎপর্য ঃ

গিলফোর্ড সহগতি সহগাঙ্ক গুণাঙ্কের মাত্রার উপর ভিত্তি করে সহ-সম্বন্ধকে নিম্নলিখিত শ্রেণিগুলিতে ভাগ করেন ঃ

সহগাঙ্ক (r এর মান)	সম্বন্ধ
(i) .00 থেকে ±0.20	খুবই কম
(ii) ±0.21 থেকে ±0.40	কম
(iii) ±0.41 থেকে ±0.60	সাধারণ
(iv) ±0.61 থেকে ±0.80	অধিক
(v) ±0.81 থেকে ±0.99	খুবই বেশি
(vi) ±1.00	পরিপূর্ণ সহসম্বন্ধ

সহসম্পর্কের মান নির্ধারণে কোন্ কৌশলটি গ্রহণ করা হবে তা বিচার করা হয় (a) চলগুলির প্রকৃতি অনুযায়ী, (b) যে উদ্দেশ্যে সহগতি নির্ণয় করা হয় তার পরিপ্রেক্ষিতে।

শিক্ষাতত্ত্ব, মনোবিজ্ঞান প্রভৃতি বিষয়গুলির তাৎপর্য নির্ণয়ের ক্ষেত্রে সহসম্পর্কের মানগুলি অত্যন্ত গুরুত্বপূর্ণ।

অনুশীলনী

1. What is coefficient of correlation? How coefficient of correlation is calculated? What is product moment method of finding out correlation?
2. What is rank difference method of finding out co-efficient of correlation? What is its advantages and delimitation?
3. Define the term correlation. Define co-efficient of correlation and state its uses in the field of Education.
4. What do you understand by (i) positive (ii) negative (iii) zero correlation illustrate with example in the field of Education.
5. Find out the co-efficient of correlation of the following two set of scores of 10 student by (i) Rank difference Method (ii) Product moment Method & comment on it.

Students :	A	B	C	D	E	F	G	H	I	J
Score I	32	38	48	43	40	22	41	69	35	64
Score II	30	31	38	43	33	11	27	76	40	59