

ইউনিট ৬ সহসম্পর্ক ও নির্ভরণ

ইউনিট ৬ সহসম্পর্ক ও নির্ভরণ (Correlation and Regression)

পূর্বের অধ্যয়নগুলোতে আমরা একটি চলকের বিভিন্ন মান সম্বন্ধে আলোচনা করেছি এবং কেন্দ্রীয় প্রবণতা, বিস্তার ইত্যাদির সাহায্যে কীভাবে উক্ত চলকের মানগুলোর বিশ্লেষণ করা যায় বিস্তারিত বর্ণনা করেছি। বাস্তবক্ষেত্রে আমাদেরকে দুই বা ততোধিক পরস্পর সম্পর্কযুক্ত চলক নিয়ে কাজ করতে হয়। যেমন কোন শ্রেণির ছাত্রদের পরিসংখ্যান ও গণিত বিষয়ে প্রাপ্ত নম্বরের মধ্যে সম্পর্ক নির্ণয় দুই বা ততোধিক চলকের মধ্যে পরস্পর সম্পর্কে পরিসংখ্যানের ভাষার সহসম্পর্ক বলা হয়। আবার অনেক ক্ষেত্রে দুটি চলকের মধ্যে একটি অপরটির ওপর নির্ভরশীল হতে পারে। যেমন একটি জমির শস্য উৎপাদন ঐ জমিতে সার প্রয়োগের পরিমাণের ওপর নির্ভর করতে পারে। সুতরাং কোন একটি চলকের নির্দিষ্ট মানের জন্য অপর চলকের গড় মানের যে পরিবর্তন হয় সেটাই হচ্ছে নির্ভরণ। এখানে একটি স্বাধীন চলক অন্যটি নির্ভরশীল চলক। সহসম্পর্ক এবং নির্ভরণ সহজ এবং বহুবিধ হতে পারে। শুধুমাত্র দুটি চলক নিয়ে কাজ করলে বলা হয় সহজ এবং দুটির অধিক চলক নিয়ে কাজ করলে বলা হয় বহুদা। এ ইউনিটে বিভিন্ন পাঠে সহসম্পর্ক ও নির্ভরণ নিয়ে বিস্তারিতভাবে আলোচনা করা হয়েছে।

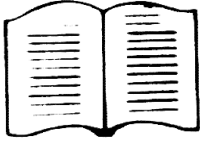
পাঠ ৬.১ সহসম্পর্ক (Correlation)



এ পাঠ শেষে আপনি-

- সহসম্পর্কের সংজ্ঞা বলতে পারবেন।
- সহসম্পর্কের সহগ কাকে বলে তা বলতে ও লিখতে পারবেন।
- সহসম্পর্কের প্রকারভেদ সম্পর্কে বলতে পারবেন।
- সহসম্পর্ক নির্ণয় করতে পারবেন।
- সহসম্পর্কের ব্যবহার সম্পর্কে বলতে পারবেন।
- অনুক্রমিক সহসম্পর্কের সহগ সম্পর্কে বলতে ও ব্যাখ্যা করতে পারবেন।
- সহসম্পর্ক সম্পর্কিত বিভিন্ন সমস্যার সমাধান করতে পারবেন।

সহসম্পর্ক ও সহসম্পর্কের সহগ (Correlation and correlation coefficient)



দুই বা ততোধিক চলকের মধ্যে রৈখিক সম্পর্কেই সহসম্পর্ক বলে এবং পরিসংখ্যানিক পদ্ধতিতে সহসম্পর্কের মাত্রার পরিমাপকেই সহসম্পর্কের সহগ বলে এবং একে r দ্বারা চিহ্নিত করা হয়।

সহসম্পর্কের সহগ নির্ণয়ের সূত্র: কার্ল পিয়ারসন (K. Pearson) সহসম্পর্কের সহগ নির্ণয়ের সূত্র প্রদান করেন।

ধরা যাক $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, x এবং চলকের n জোড়া মান। এক্ষেত্রে সহসম্পর্কের সহগ নির্ণয়ের সূত্র হচ্ছে -

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}}$$

$$\text{or, } r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

$$\text{or, } r = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sqrt{\left[\sum x_i^2 - \frac{(\sum x_i)^2}{n} \right] \left[\sum y_i^2 - \frac{(\sum y_i)^2}{n} \right]}}$$

$$\text{or, } r = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{\{(\sum x_i^2 - n \bar{x}^2)(\sum y_i^2 - n \bar{y}^2)\}^{1/2}}$$

$$\text{এখানে } \bar{x} = \text{চলকের গাণিতিক গড়} = \frac{\sum x_i}{n}$$

$$\bar{y} = \text{y চলকের গাণিতিক গড়} = \frac{\sum y_i}{n}$$

যে পরিসংখ্যান পদ্ধতির সাহায্যে দুই বা ততোধিক রৈখিক স্বাধীন চলকের সম্পর্কের মাত্রা নির্ণয় ও বিশ্লেষণ করতে পারি তাকে সহসম্পর্ক বিশ্লেষণ বলা হয়।

গহসম্পর্ক একটি এককবিহীন সংখ্যা এবং এর মান +১ থেকে -১ পর্যন্ত হতে পারে, অর্থাৎ $-1 \leq r \leq 1$

যে পরিসংখ্যান পদ্ধতির সাহায্যে দুই বা ততোধিক রৈখিক স্বাধীন চলকের সম্পর্কের মাত্রা নির্ণয় ও বিশ্লেষণ করতে পারি তাকে সহসম্পর্ক বিশ্লেষণ (Correlation Analysis) বলা হয়।

সহসম্পর্কের প্রকারভেদ

সহসম্পর্ক বিভিন্ন প্রকারের হতে পারে। সাধারণত ২(দুই) প্রকারের সহসম্পর্ক হতে পারে, যথা-

- ধনাত্মক সহসম্পর্ক এবং
- ঋণাত্মক সহসম্পর্ক

ধনাত্মক সম্পর্ক

যখন দুটি চলকের মানই পরস্পর একই দিকে পরিবর্তন হয় সেক্ষেত্রে এদের সহসম্পর্ক হবে ধনাত্মক। এখানে সহ সম্পর্ক সহগের মান ০ চেয়ে বেশি হবে এবং ১ এর সমান অথবা কম হবে। অর্থাৎ $0 \leq r \leq 1$

উদাহরণ $x : 10 \ 13 \ 15 \ 18 \ 20$
 $y : 12 \ 15 \ 18 \ 21 \ 23$

এখানে, $r = 0.8885$ যাহা $0 \leq r \leq 1$ এর মধ্যে অবস্থিত। অতএব X ও Y এর সম্পর্ক ধনাত্মক।

আবার যখন দুটি চলকের মান একটি অপরটির বিপরীত দিকে পরিবর্তন হয় তখন তাদের সম্পর্ক হবে ঋনাত্মক সহসম্পর্ক।

ঋনাত্মক সহসম্পর্ক

আবার যখন দুটি চলকের মান একটি অপরটির বিপরীত দিকে পরিবর্তন হয় তখন তাদের সম্পর্ক হবে ঋনাত্মক সহসম্পর্ক। এক্ষেত্রে সহসম্পর্ক সহগের মান -১ থেকে ০ পর্যন্ত হতে পারে অর্থাৎ $-1 < r < 0$

উদাহরণ $x : 20 \ 30 \ 40 \ 50 \ 60$
 $y : 50 \ 45 \ 30 \ 25 \ 20$

এখানে, $r = -0.98$ যাহা $-1 \leq r \leq 0$ এর মধ্যে অবস্থিত। অতএব X ও Y এর সম্পর্ক ঋণাত্মক।

উদাহরণ

নিম্নে কোন শ্রেণির ৫ জন ছাত্রের পরিসংখ্যান ও গণিতের প্রাপ্ত নম্বর দেয়া হলো। সহসম্পর্কের সহগ নির্ণয় করুন:

ক্রমিক নং	:	১	২	৩	৪	৫
পরিসংখ্যান	:	৪৭	৩৫	১৭	২৪	৪৭

গণিত : ৪০ ২৫ ৩৫ ৩৮ ৪২

সমাধান

সূত্র ব্যবহার করার জন্য নিম্নলিখিত সারণির প্রয়োজন। মনেকরি x পরিসংখ্যানের নম্বর, y গণিতের নম্বর

x	y	x ²	y ²	xy
47	40	2209	1600	1880
35	25	1225	625	875
17	35	289	1225	595
24	38	576	1444	912
47	48	2209	2304	2256
$\Sigma x = 170$	$\Sigma y = 186$	$\Sigma x^2 = 6508$	$\Sigma y^2 = 7198$	$\Sigma xy = 6518$

এখন,

$$\begin{aligned}
 r &= \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}} \\
 &= \frac{5618}{\sqrt{6508 \times 7198}} \\
 &= \frac{5618}{\sqrt{46844584}} \\
 &= \frac{6518}{6844.310338} \\
 &= 0.9523
 \end{aligned}$$

∴ পরিসংখ্যান ও গণিতের সম্পর্কের সহগ $r = 0.9523$

এখানে, $\Sigma xy = 6518$; $\Sigma x^2 = 6508$; $\Sigma y^2 = 7198$

গহসম্পর্ক সহগের বৈশিষ্ট্য (Characteristic of correlaton coefficient)

- দুটি চলকের সহসম্পর্ক থাকলে r -এর মান -১ থেকে +১ পর্যন্ত হবে এবং সহসম্পর্ক না থাকলে r -এর মান শূন্য হবে।
- এটি একটি একক বিহীন সংখ্যা।
- এটি মূলবিন্দুও মাপনী হতে স্বাধীন অর্থাৎ নির্ভরশীল নয়।
- এটি দুটি চলকের সম্পর্কের মাত্রা নির্ণয় করে কিন্তু কী কারণে এরা সম্পর্কিত বলতে পারে না।
- একই শক্তি চলক দুটিকে প্রভাবিত না করলে এদের মধ্যে সহসম্পর্ক হয় না। যেমন মানুষের ওজন এবং গরুর ওজনের মধ্যে কোন সম্পর্ক নেই।
- $r_{xy} = r_{yx}$.
- চলক দুটি স্বাধীন হলে সহসম্পর্ক শূন্য (০) হবে।

সহসম্পর্কের ব্যবহার (Uses of correlation)

- রাষ্ট্রীয়, সামাজিক, অর্থনৈতিক ক্ষেত্রে গবেষণা করতে গেলে সহসম্পর্কের ব্যাপক ব্যবহার করতে হয়।
- সমাজ বিজ্ঞান ও মনস্তত্ত্বের ক্ষেত্রে অনুক্রমিক সহসম্পর্কের সহগ প্রচুর ব্যবহৃত হয়।
- দুটি চলকের মধ্যে সম্পর্ক আছে জানা থাকলে ঐ চলকগুলোর ভবিষ্যৎ সম্পর্ক সম্বন্ধেও পূর্বাভাস দেয়া যায়। যেমন, চাউলের উৎপাদন এবং মূল্য এ দুটি সম্পর্কযুক্ত। এ ক্ষেত্রে অন্য সব কিছু স্থির থাকলে চাউলের বেশি উৎপাদন হলে দাম কমবে সে সম্পর্কে পূর্বাভাস পাওয়া যায়।

অনুক্রমিক সহসম্পর্কের সহগ (Rank correlation coefficient)

চলকদ্বয়ের বিভিন্ন মানের ক্রমসংখ্যা পাওয়া গেলে এদের মধ্যে সহসম্পর্কের মাত্রাকেই অনুক্রমিক সহসম্পর্কের সহগ বলে।

চার্লস এডওয়ার্ড স্পেয়ারম্যান ১৯০৪ সালে সহসম্পর্কের সহগকে এরূপ নামকরণ করেন। যখন চলকদ্বয়ের পরিমাণ গুণগত হয় সেক্ষেত্রে এদের মান বের করা কঠিন তবে গুণানুসারে এদেরকে ক্রমানুযায়ী সাজান যায়। অর্থাৎ চলকদ্বয়ের বিভিন্ন মানের ক্রমসংখ্যা পাওয়া গেলে এদের মধ্যে সহসম্পর্কের মাত্রাকেই অনুক্রমিক সহসম্পর্কের সহগ বলে। উদাহরণস্বরূপ দুজন পরীক্ষক কয়েকজন ছাত্রের একটি বিষয়ের পরীক্ষার ওপর নম্বর দিলে একই ছাত্র ভিন্ন নম্বর পেতে পারে কেননা পরীক্ষক দুজনের ভিন্ন মত ও মেধা যাচাইয়ের ভিন্ন ভিন্ন পদ্ধতি থাকতে পারে। এক্ষেত্রে দুজন পরীক্ষক প্রদত্ত নম্বর না নিয়ে কে প্রথম কে দ্বিতীয় ইত্যাদি ক্রমে সাজিয়ে দুজন পরীক্ষকের মতের ও মিলের সম্পর্ক বের করা হয়। ধরা যাক n সংখ্যক ছাত্রের মান ২টি বৈশিষ্ট্য x এবং y ক্রমের মান অনুযায়ী প্রদত্ত হলো-

x এর ক্রম: x_1, x_2, \dots, x_n

y এর ক্রম: y_1, y_2, \dots, y_n

$$\therefore \bar{x} = \bar{y} = \frac{1}{n} (1 + 2 + \dots + n) = \frac{n(n+1)}{2n} = \frac{n+1}{2}$$

$$\therefore \sum (x_i - \bar{x})^2 = \sum (y_i - \bar{y})^2 = \sum X_i^2 - n\bar{x}^2 = \sum y_i^2 - n\bar{y}^2$$

$$= 1^2 + 2^2 + \dots + n^2 - n \left(\frac{n+1}{2} \right)^2$$

$$= \frac{n(n+1)(2n+1)}{6} - \frac{n(n+1)^2}{4}$$

$$= \frac{n(n-1)}{3}$$

মনেকরি x এবং y এর ক্রমের পার্থক্য d_i

অর্থাৎ $d_i = x_i - y_i$

$$\therefore d_i = (x_i - \bar{x}) - (y_i - \bar{y})$$

$$\therefore \sum d_i^2 = \sum (x_i - \bar{x})^2 + \sum (y_i - \bar{y})^2 - 2\sum (x_i - \bar{x})(y_i - \bar{y})$$

$$\text{আমরা জানি, } r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

$$\text{বা, } r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} = \frac{\sum d_i^2}{n(n-1)}$$

$$\therefore \frac{\sum d_i^2}{n} = \frac{\sum(n - 1)}{n} - r \frac{n - 1}{n}$$

$$\text{or, } \frac{\sum d_i^2}{n} = \frac{\sum(n - 1)(1 - r)}{n}$$

$$\text{or, } r = 1 - \frac{\sum d_i^2}{n(n-1)}$$

কার্ল পিয়ার্সনের সহসম্পর্কের মতই অনুক্রমিক সহসম্পর্কের ব্যাখ্যা করা হয়।

উপরের সূত্রের মাধ্যমে সহসম্পর্কের সহগ নির্ণয় করতে হলে নিম্নলিখিত পদক্ষেপ নিতে হবে।

- x এবং y চলকদ্বয়ের তথ্যমানকে ক্রমের মান অনুযায়ী সাজাতে হবে। যদি কোন ক্ষেত্রে কোন চলকের দুটি মান একই ক্রমের মান হয় যেমন একটি চলকের দুটি মানের ক্রমের মান ৫, ৬ এক্ষেত্রে মান দুটির ক্রমের মান হবে $\frac{৫ + ৬}{২} = ৫.৫$ এবং পরবর্তী ক্রমের মান হবে ৭।
- চলকদ্বয়ের ক্রমের মান বের করার পর একটি মান থেকে অন্যটির মান বিয়োগ করতে হবে। অর্থাৎ $d_i = x_i - y_i$ এখানে x_i এবং y_i প্রথম ও দ্বিতীয় চলকের ক্রমের মান।
- বিয়োগফল d_i কে বর্গ করে যোগ করতে হবে।
- নিম্নের সূত্রের মাধ্যমে ক্রমের মানের সহসম্পর্কের সহগ নির্ণয় করতে হবে।

$$R = \frac{\sum d_i^2}{n(n-1)}$$

উদাহরণ

ধরায়াক, পাঁচজন ছাত্রের পরিসংখ্যান বিষয়ের ওপর পরীক্ষায় অভ্যন্তরীণ ও বহিরাগত পরীক্ষক নিম্নলিখিত নম্বর দিলেন। এদের ক্রমের মানের সহসম্পর্কের সহগ বের করুন।

অভ্যন্তরীণ পরীক্ষকী : ১৮ ৩৭ ২৩ ৩৩ ২৭ ৩১

বহিরাগত পরীক্ষকু : ৩৩ ৩৫ ৩১ ৩৩ ৩০ ২৮

সমাধান

ক্রমের মানের সহসম্পর্কের সহগ নির্ণয় করার জন্য নিম্নের সারণি তৈরি করা হলো।

x	ক্রমের মান x_i	y	ক্রমের মান y_i	$d_i = x_i - y_i$	d_i^2
১৮	৬	৩৩	২.৫	৩.৫	১২.২৫
৩৭	১	৩৫	১	০	০
২৩	৫	৩১	৪	১	১
৩৩	২	৩৩	২.৫	-০.৫	.২৫
২৭	৪	৩০	৫	-১	১
৩১	৩	২৮	৬	-৩	৯
					$\sum d_i^2 = ২৩.৫০$

$$r = 1 - \frac{6 \times 23.50}{6(6^2 - 1)} = 0.33$$

∴ অনুক্রমিক সংশ্লেষাঙ্ক $r = 0.33$

অনুক্রমিক সংশ্লেষাঙ্কের সুবিধা

- এটি বুঝতে ও নির্ণয় করতে সহজ।
- উপাত্ত গুণবাচক হলে এ পদ্ধতি ব্যবহার করা হয়।
- চলক দুটি মান ক্রম সংখ্যানুযায়ী থাকলে এ পদ্ধতিতেই কেবলমাত্র সংশ্লেষাঙ্ক নির্ণয় করা যায়।

উদাহরণ ১

নিচের উপাত্তটির ক্ষেত্রে সংশ্লেষাঙ্ক নির্ণয় করুন।

x :	২	২	৪	৫	৬
y :	৫	৩	২	৬	৭

সমাধান

x	y	x^2	y^2	xy
2	5	4	25	10
2	3	4	9	6
4	2	16	4	8
5	6	25	36	30
6	7	36	49	42
$\sum x = 19$	$\sum y = 23$	$\sum x^2 = 85$	$\sum y^2 = 123$	$\sum xy = 96$

আমরা জানি,

$$r = \frac{\sum xy - \frac{\sum x \sum y}{N}}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{N} \right] \left[\sum y^2 - \frac{(\sum y)^2}{N} \right]}}$$

$$\text{or, } r = \frac{96 - \frac{19 \times 23}{5}}{\sqrt{\left[85 - \frac{(19)^2}{5} \right] \left[123 - \frac{(23)^2}{5} \right]}}$$

$$\text{or, } r = \frac{8.6}{\sqrt{12.8 \times 17.22}} = \frac{8.6}{14.8} = 0.58$$

∴ নির্ণেয় সংশ্লেষাঙ্ক $r = 0.58$

উদাহরণ ২

নিচের উপাত্তটির অনুক্রমিক সংশ্লেষাঙ্ক নির্ণয় করুন।

x :	১৫	৩৪	২০	৩০	২৮
y :	৩০	৩২	২৮	৩০	২৫

সমাধান

x	ক্রম সংখ্যা X_i	Y	ক্রম সংখ্যা Y_i	$d_i = X_i - Y_i$	d_i
১৫	৬	৩০	২.৫	৩.৫	১২.২৫
৩৪	১	৩২	১	০	০
২০	৫	২৮	৪	১	১
৩০	২	৩০	২.৫	-০.৫০	০.২৫
২৪	৪	২৭	৫	-১	১
২৮	৩	২৫	৬	-৩	৯
					$\sum d^2 = ২৩.৫০$

আমরা জানি,

অনুক্রমিক সংশ্লেষাঙ্ক

$$r = 1 - \frac{6 \sum d^2}{n(n^2 - 1)} \quad ; n = 6$$

$$\text{OR, } r = 1 - \frac{6 \times 23.50}{6(36 - 1)}$$

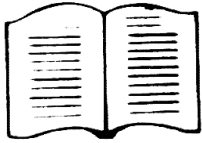
$$\text{OR, } r = 1 - \frac{141}{210} = \frac{69}{210} = 0.33$$

∴ অনুক্রমিক সংশ্লেষাঙ্ক, $r = 0.33$ ।



অনুশীলন (Activity) : বাংলাদেশ কৃষি বিশ্ববিদ্যালয়ে অনুষ্ঠিত একটি পরীক্ষায় কৃষি পরিসংখ্যান ও কৃষি অর্থনীতি বিষয়ে ৯ জন ছাত্রের শতকরা প্রাপ্ত নম্বর নিম্নে দেয়া হলো। অনুক্রমিক সহসম্পর্ক সহগ নির্ণয় করুন।

ছাত্র	১	২	৩	৪	৫	৬	৭	৮	৯
পরীক্ষা									
কৃষি অর্থনীতি	৪১	৪৫	৫০	৬৮	৪৭	৭৭	৯০	৯৮	৮০
কৃষি পরিসংখ্যান	৬০	৬৩	৬১	৪৮	৮৫	৫৬	৫৩	৯১	৯৪



সারমর্ম : দুই বা ততোধিক চলকের মধ্যে রৈখিক সম্পর্কেই সহসম্পর্ক বলে এবং পরিসংখ্যানিক পদ্ধতিতে সহসম্পর্কের মাত্রার পরিমাপকেই সহসম্পর্কের সহগ বলে এবং একে ৭ দ্বারা চিহ্নিত করা হয়। কার্ল পিয়ারসন সহসম্পর্কের সহগ নির্ণয়ের সূত্র প্রদান করেন। সহসম্পর্ক একটি এককবিহীন সংখ্যা এবং এর মান +১ থেকে -১ পর্যন্ত হতে পারে, অর্থাৎ $-1 \leq r \leq 1$ । যে পরিসংখ্যান পদ্ধতির সাহায্যে দুই বা ততোধিক রৈখিক স্বাধীন চলকের সম্পর্কের মাত্রা নির্ণয় ও বিশ্লেষণ করতে পারি তাকে সহসম্পর্ক বিশ্লেষণ বলা হয়। সাধারণত ২(দুই) প্রকারের সহসম্পর্ক হতে পারে, যথা- ধনাত্মক সহসম্পর্ক এবং ঋণাত্মক সহসম্পর্ক। দুটি চলকের সহসম্পর্ক থাকলে এর সহগের মান -১ থেকে +১ পর্যন্ত হবে এবং সহসম্পর্ক না থাকলে এর মান শূন্য হবে।



পাঠ্যের মূল্যায়ন ৬.১

সঠিক উত্তরের পাশে টিক চিহ্ন (✓) দিন।

- ১। দুই বা ততোধিক চলকের মধ্যে রৈখিক সম্পর্ককে কী বলা হয়?
 - ক) সম্ভাবনা
 - খ) ভেদাঙ্ক
 - গ) সহসম্পর্ক
 - ঘ) পরিমিত ব্যবধান

- ২। সহসম্পর্কের মাত্রা পরিমাপ পদ্ধতিকে কী বলা হয়?
 - ক) বিস্তার পরিমাপ
 - খ) সহসম্পর্ক সহগ
 - গ) বিভেদাঙ্ক
 - ঘ) মধ্যক

- ৩। নিচের কোনটিতে সহসম্পর্ক সহগের মান সর্বদা অবস্থিত?
 - ক) $-1 \leq r \leq +1$
 - খ) $-1 \geq r \leq +1$
 - গ) $-1 \geq r \geq +1$
 - ঘ) $-1 \leq r \geq +1$

- ৪। চলক দুটি স্বাধীন হলে সহসম্পর্ক সহগের মান কোনটি হবে?
 - ক) $r \geq +1$
 - খ) $r = 0$
 - গ) $r \leq -1$
 - ঘ) $r \geq -1$

- ৫। কোনটির ক্ষেত্রে দু'ধরনের অনুক্রমকে সম্পর্ক সামঞ্জস্য অনুক্রমিক সহসম্পর্ক সহগ বলা হয়?
 - ক) $r = 0$
 - খ) $r = 1$
 - গ) $r = -1$
 - ঘ) $r = 1.2$

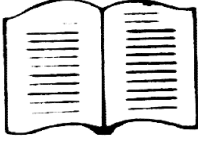
পাঠ ৬.২ নির্ভরণ (Regression)



এ পাঠ শেষে আপনি -

- নির্ভরণ বলতে কী বোঝায় তা বলতে ও লিখতে পারবেন।
- নির্ভরণের ব্যবহার সম্পর্কে বলতে পারবেন।
- সহসম্পর্ক ও নির্ভরণ বিশ্লেষণের পার্থক্য নির্ণয় করতে পারবেন।
- নির্ভরণ রেখা সম্বন্ধে বলতে পারবেন।
- নির্ভরণ সম্পর্কিত বিভিন্ন সমস্যার সমাধান করতে পারবেন।

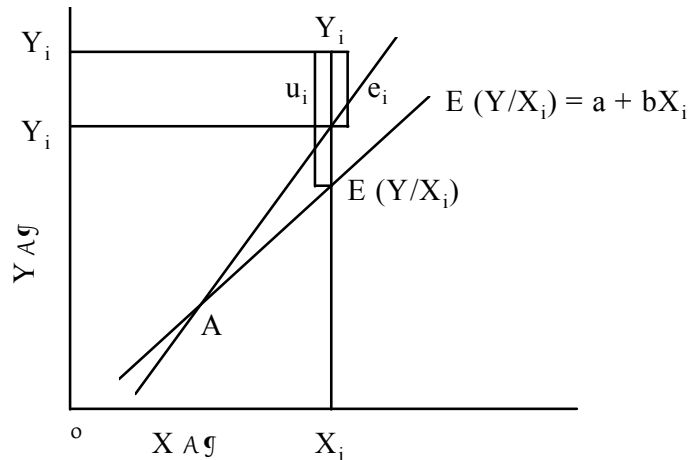
নির্ভরণ (Regression)



সহসম্পর্কের মাধ্যমে দুটি চলকের মধ্যে রৈখিক সম্পর্ক নির্ণয় করা যায় এবং সম্পর্কের মাত্রা বের করা যায়। কিন্তু একটির পরিবর্তন অন্যটির ওপর কতটুকু নির্ভরশীল তা সহসম্পর্কের মাধ্যমে জানা যায় না। একটি চলকের একক পরিবর্তনের ফলে অন্যটির পরিবর্তনের হার কতটুকু তা জানা প্রয়োজন। নির্ভরণের মাধ্যমেই এই পরিবর্তনের হার সম্পর্কে জানা যায়। এক্ষেত্রে দুটি চলকের মধ্যে একটি নির্ভরশীল বা অধীন চলক এবং অন্যটি স্বাধীন চলক। স্বাধীন চলকের মান পরিবর্তন করলে সাধারণত অন্যটির পরিবর্তন হয় এবং গড়পড়তা কী পরিবর্তন হবে তা নির্ভরণ দ্বারা নির্দেশিত। যেমন কোন জমিতে সারের পরিমাণের পরিবর্তন করলে ঐ জমির ফসলের উৎপাদনের পরিবর্তন বাড়ে। বাবা-মার উচ্চতা বেশি হলে সন্তানের উচ্চতা বেশি হবে অর্থাৎ সন্তানের উচ্চতা বাবা-মার উচ্চতার সাথে সম্পর্কিত। এভাবে আরও অনেক উদাহরণ দেয়া যেতে পারে। এসমস্ত উদাহরণে একটি চলকের পরিবর্তনের সাথে অন্যটির গড় পরিবর্তন কতটুকু তা জানার জন্যই নির্ভরণের প্রয়োগ করতে হয়।

নির্ভরণ রেখা (Regression curve)

আমরা আগেই বলেছি নির্ভরণ এ একটি অধীন চলক এবং অন্যগুলো স্বাধীন চলক। যে চলক সম্বন্ধে ব্যাখ্যা করতে হয় সেটাই নির্ভরশীল চলক এবং একে সাধারণত Y দ্বারা সূচিত করা হয় এবং বাকী চলকগুলোকে স্বাধীন চলক বলে এবং এদেরকে সাধারণত X দ্বারা সূচিত করা হয়। উদাহরণস্বরূপ ফসলের উৎপাদনকে যদি Y ধরা হয় তবে এর সাথে সম্পৃক্ত সারের পরিমাণকে X বলা যেতে পারে। X এর পরিবর্তনের সাথে Y এর কী পরিবর্তন হয় এটা ব্যাখ্যা করা প্রয়োজন। X কে X অক্ষের দিকে এবং Y কে Y অক্ষের দিকে নিলে একটি বিক্ষিপ্ত চিত্র পাওয়া যাবে যা নিচে দেখানো হলো।



চিত্র- নির্ভরণ রেখা

এখানে Y এর মানগুলোর গড় মানসমূহের ভিত্তিতে একটি সরল রেখা আঁকা সম্ভব এবং যে রেখাটি পাওয়া যাবে তাকেই রৈখিক নির্ভরণ রেখা বলে।

গড় মান থেকে অন্যান্য মানের পার্থক্যকে বিচ্যুতি বলে। নির্ভরণ রেখাকে নিম্নোক্ত সমীকরণের মাধ্যমে লেখা যায়, $Y = A + BX$

এবং নির্ভরণ মডেল কে লেখা যায়, $Y = A + BX + e$

এখানে B হলো নির্ভরণ সহগ A হলো স্থিরক এবং e হলো বিচ্যুতি। আমাদেরকে A এবং B এর মান নির্ধারণ করতে হবে। যেহেতু তথ্যবিশ্বের সকল মান নিয়ে বিশ্লেষণ করা পরিসংখ্যানিকভাবে সম্ভব নয় এবং নমুনায়নের মাধ্যমে উপাত্ত সংগ্রহ করে সেটা দিয়ে বিশ্লেষণ করা হয়।

ধরা যাক, $(X_i, Y_i) i = 1, 2, \dots, n$ X এবং Y দুটি চলকের নমুনা তথ্যমান। এক্ষেত্রে

$Y_i = a + bX_i$ হচ্ছে নির্ভরণ রেখা এবং

$Y_i = a + bX_i + e_i$ হচ্ছে নির্ভরণ মডেল।

এখানে b হচ্ছে Y এর ওপর X এর প্রভাব কতটুকু তার পরিমাণ। অর্থাৎ b হচ্ছে X এর ওপর Y এর নির্ভরণ সহগ (regression co-efficient) X এবং Y এর রৈখিক সম্পর্কের প্রভাব অনুযায়ী b এর মান শূন্য, ধনাত্মক অথবা ঋণাত্মক হতে পারে।

a হচ্ছে স্থিরক, X এর প্রভাব বাদ দিয়ে Y এর যে মান সেটাই a , এটা একটা উদাহরণ দিয়ে বুঝান যেতে পারে। ধরা যাক, কোন জমিতে সার দিলে ফসলের উৎপাদন বাড়বে কিন্তু সার প্রয়োগ না করলেও ফসলের উৎপাদন হবে। এখানে সারের প্রভাব ছাড়া ফসলের যে উৎপাদন সেটাই হচ্ছে a ।

নির্ভরণ মডেলের ক্ষেত্রে নিম্নলিখিত অনুমানগুলো ধরে নিতে হবে।

- X এর মানসম n নির্দিষ্ট হতে হবে।
- X এর ওপর Y এর নির্ভরণ রৈখিক হবে।
- X এর প্রত্যেকটি মানের জন্য Y এর শর্তাধীন বিন্যাস পরিমিত হবে।
- বিচ্যুতিসমূহ স্বাধীনভাবে পরিমিত বিন্যাসে বিন্যস্ত থাকবে এবং এদের গড় মান শূন্য এবং ভেদাংক সমান হবে।

বিচ্যুতিসমূহের বর্গের সমষ্টি ন্যূনতম করে যে পদ্ধতির মাধ্যমে a এবং b এর মান বের করা যায় তাকেই ন্যূনতম বর্গ পদ্ধতি বলে।

নির্ভরণ বিশ্লেষণের আসল উদ্দেশ্য হচ্ছে a এবং b এর মান বের করা। ন্যূনতম বর্গ পদ্ধতির মাধ্যমে a এবং b এর মান বের করা যায়। বিচ্যুতিসমূহের বর্গের সমষ্টি b ন্যূনতম করে যে পদ্ধতির মাধ্যমে a এবং b এর মান বের করা যায় তাকেই ন্যূনতম বর্গ পদ্ধতি বলে। আমরা জানি

$$Y_i = a + bX_i + e_i$$

$$\text{বিচ্যুতি } e_i = (Y_i - a - bX_i)$$

$$\therefore \sum e_i^2 = \epsilon = \sum (Y_i - a - bX_i)^2 = \text{ন্যূনতম}$$

এখানে a ও b এর প্রসঙ্গে ϵ কে আংশিক অন্তরকলন করলে ফল শূন্য হবে অর্থাৎ

$$\frac{\partial \epsilon}{\partial a} = 0 \text{ এবং } \frac{\partial \epsilon}{\partial b} = 0$$

$$\therefore \frac{\partial \epsilon}{\partial a} = 0 \Rightarrow \sum Y_i - na - b \sum X_i = 0$$

$$\text{বা, } \sum Y_i = na + b \sum X_i = 0 \dots\dots\dots (১)$$

$$\therefore \frac{\partial \epsilon}{\partial b} = 0 \Rightarrow \sum X_i Y_i - a \sum X_i - b \sum X_i^2 = 0$$

$$\text{or, } \sum X_i Y_i = a \sum X_i + b \sum X_i^2 \dots\dots\dots (২)$$

সমীকরণ (১) এবং (২) কে পরিমিত সমীকরণ বলা হয় এবং এদেরকে সমাধান করেই a এবং b এর মান বের করা যায়।

সমীকরণ দুটিকে আমরা n দ্বারা ভাগ করে পাই

$$\bar{Y} = a + b\bar{X} \dots\dots\dots (৩)$$

$$\text{এবং } \frac{\sum X_i Y_i}{n} = a\bar{X} + b \frac{\sum X_i^2}{n} \dots\dots\dots (৪)$$

(৩) নং সমীকরণে \bar{X} দ্বারা গুণ করার পর (৪) নং সমীকরণ থেকে বিয়োগ করলে আমরা পাই

$$b \left(\frac{\sum X_i^2}{n} - \bar{X}^2 \right) = \frac{\sum X_i Y_i}{n} - \bar{X}\bar{Y} \dots\dots\dots (৫)$$

$$\therefore b = \frac{\frac{\sum X_i Y_i}{n} - \bar{X}\bar{Y}}{\frac{\sum X_i^2}{n} - \bar{X}^2} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

$$\therefore b = \frac{Sp(XY)}{Sq(x)} \dots\dots\dots (৬)$$

নির্ভরণ সহগকে সাধারণত b_{yx} দ্বারা প্রকাশ করা হয়।

এখানে b কে X এর ওপর Y এর নির্ভরণ সহগ বলা হয় এবং সাধারণত b_{yx} দ্বারা নির্দেশিত হয়।

সমীকরণ (৩) এ b এর মান বসিয়ে a এর মান পাওয়া যাবে,

$$a = \bar{Y} - b\bar{X}$$

$$\therefore a = \bar{Y} - \frac{Sp(XY)}{Sq(X)} \times \bar{X} \dots\dots\dots (৭)$$

সুতরাং নির্ভরণ রেখাটি হবে নিম্নরূপ -

$$Y_c = \left\{ \bar{Y} - \frac{sp(XY)}{sq(X)} \times \bar{X} \right\} + \frac{sp(XY)}{sq(X)} \times X$$

$$\therefore Y_c = \bar{Y} - \frac{sp(XY)}{sq(X)} (X - \bar{X}) \dots \dots \dots (৩)$$

উদাহরণ

নিম্নে ৭ জন শ্রমিকের আয় এবং ব্যয়ের উপাত্ত দেয়া হলো। উপাত্তটির নির্ভরণ সমীকরণ তৈরি করণ এবং নির্ভরণ রেখা নির্ণয় করণ।

x : আয় (টাকায়)	১০	১২	১১	১৫	২০	১৩	১৪
y : ব্যয় (টাকায়)	১৫	১৬	১৬	১৮	১৮	১৫	১৩

সমাধান

x	y	x ²	xy
10	15	100	150
12	16	144	192
11	16	121	176
15	18	225	270
20	18	400	360
13	15	169	195
14	13	196	182
$\Sigma x = 95$	$\Sigma y = 111$	$\Sigma x^2 = 1355$	$\Sigma xy = 1525$

$$b_{yx} = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{N}}{\Sigma x^2 - \frac{(\Sigma x)^2}{N}} \quad \text{GL: } b, N = 7$$

$$= \frac{1525 - \frac{95 \times 111}{7}}{1355 - \frac{(95)^2}{7}}$$

$$= \frac{1525 - 1506.43}{1355 - 1289.28}$$

$$= \frac{18.57}{65.72} = 0.28$$

$$\therefore b_{yx} = 0.28$$

$$a = \bar{y} - b\bar{x}$$

$$GL\neq b,$$

$$\bar{y} = \frac{\sum y}{n}$$

$$\bar{x} = \frac{\sum x}{n}$$

$$\therefore a = 15.86 - 0.28 \times 13.95$$

$$= 15.86 - 3.80$$

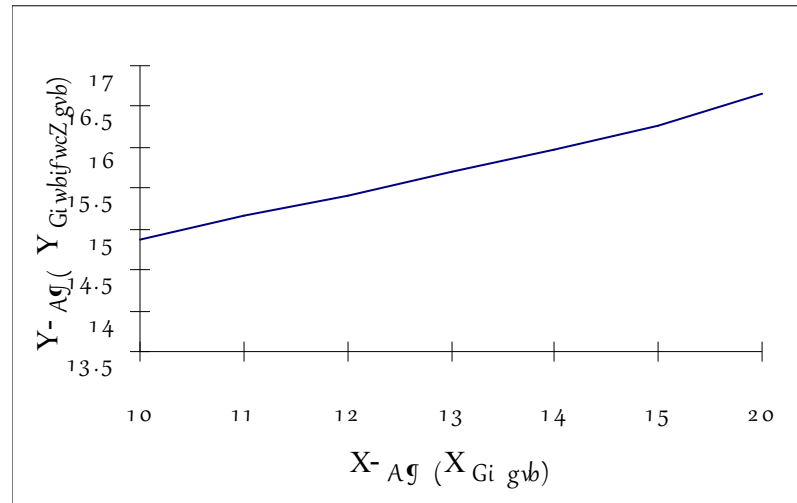
$$= 12.06$$

$$Y_c = 12.06 + 0.28x$$

উক্ত নির্ভরণ সমীকরণ অনুসারে y এর নিরূপিত মান নিরূপ

x	১০	১১	১২	১৩	১৪	১৫	২০
y	১৪.৮৬	১৫.১৬	১৫.৪২	১৫.৭০	১৫.৯৮	১৬.২৬	১৬.৬৬

অতএব নির্ভরণ রেখা-



চিত্র- নির্ভরণ রেখা

সহসম্পর্ক এবং নির্ভরণের মধ্যে পার্থক্য

নিম্নে সহসম্পর্ক এবং নির্ভরণ সম্পর্কে তুলনামূলকভাবে আলোচনা করা হলো-

সহসম্পর্ক	নির্ভরণ
১) সহসম্পর্ক বলতে দুই বা ততোধিক চলকের মধ্যে রৈখিক সম্পর্ককে বুঝায়।	১) নির্ভরণ হচ্ছে একটি স্বাধীন চলকের ওপর আর একটি নির্ভরশীল চলকের নির্ভরণ।
২) সহসম্পর্কে চলকসমূহের সম্পর্কের কারণ ও প্রভাব বিশ্লেষণ করা হয় না শুধুমাত্র এদের মাত্রা নির্ণয় করা হয়।	নির্ভরণের কাজই হলো চলকের সম্পর্কের কারণ ও প্রভাব বিশ্লেষণ করা অর্থাৎ স্বাধীন চলকের বিভিন্ন মানের জন্য নির্ভরশীল চলকের গড় মান কত বাহির করা।
৩) X এবং Y দুটি চলক হলে X এর সাথে Y এর সম্পর্কের সহগ এবং Y এর সাথে X এর সম্পর্কের সহগ সমান অর্থাৎ $r_{xy} = r_{yx}$	৩) X এবং Y দুটি চলক হলে X এর ওপর Y এর নির্ভরণের সহগ Y এর ওপর X এর নির্ভরণে সহগ সমান নয় অর্থাৎ $b_{xy} \neq b_{yx}$
৪) সহসম্পর্কের সহগের মান - ১ থেকে +১ পর্যন্ত হতে পারে।	নির্ভরণ সহগের মান - ∞ থেকে + ∞ পর্যন্ত হতে পারে।
৫) সহসম্পর্কের সহগ মূল বিন্দু এবং মাপনীর ওপর স্বাধীন।	নির্ভরণের সহগ মূলবিন্দু পরিবর্তনের ওপর স্বাধীন কিন্তু মাপনীর ওপর নির্ভরশীল।
৬) সহসম্পর্কে ক্ষেত্রে X এবং Y দুটি দৈব চলক।	৬) নির্ভরণের ক্ষেত্রে শুধুমাত্র নির্ভরশীল চলক Y দৈব চলক। স্বাধীন চলক X দৈব চলক নাও হতে পারে।
৭) $r_{xy} = \frac{S_{xy}}{\sqrt{S_x \cdot S_y}}$	৭) $b_{yx} = \frac{S_{xy}}{S_x} = r \cdot \frac{S_y}{S_x}$



অনুশীলন (Activity) : ১. নিম্নের উপাত্তে বৃষ্টিপাতের পরিমাণ ও রবিশস্যের উৎপাদনের পরিমাণ দেয়া হলো। উৎপাদনে ও বৃষ্টিপাতের উপাত্ত থেকে নির্ভরণ সহগ ও রেখা নির্ণয় করুন।

বৃষ্টিপাত	২০	২২	২৪	২৬	২৮	৩০	৩২
উৎপাদন	১৫	১৮	২০	৩২	৪০	৩৯	৪০

২. দুটি চলক x ও y এর ১০০ জোড় মানের ক্ষেত্রে নিচের তথ্য পাওয়া গেল

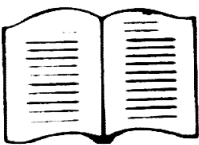
$$\sum x = 125000, \sum x^2 = 1585000$$

$$\sum y = 8000, \sum y^2 = 685100$$

$$\sum xy = 1009825$$

i) x এর ওপর y এর নির্ভরণ সহগ নির্ণয় করুন।

ii) y এর ওপর x এর নির্ভরণ সমীকরণ নির্ণয় করুন।



সারমর্ম : সহসম্পর্ক বলতে দুই বা ততোধিক চলকের মধ্যে রৈখিক সম্পর্ককে বুঝায়। নির্ভরণ হচ্ছে একটি স্বাধীন চলকের ওপর আর একটি নির্ভরশীল চলকের নির্ভরণ। বিচ্যুতিসমূহের বর্গের সমষ্টি ন্যূনতম করে যে পদ্ধতির মাধ্যমে a এবং b এর মান বের করা যায় তাকেই ন্যূনতম বর্গ পদ্ধতি বলে। নির্ভরণ সহগকে সাধারণত b_{yx} দ্বারা প্রকাশ করা হয়।



পাঠ্যোত্তর মূল্যায়ন ৬.২

সঠিক উত্তরের পাশে টিক (✓) দিন

- ১। দুটি নির্ভরশীল চলকের সম্পর্ক নির্ণয়কে কী বলা হয়?
 - ক) সহসম্পর্ক
 - খ) নির্ভরণ
 - গ) সহসম্পর্ক সহগ
 - ঘ) অনুক্রমিক সহসম্পর্ক সহগ

- ২। নির্ভরণ সমীকরণের ধ্রুবক নির্ণয় করতে কোন পদ্ধতির সাহায্য নেয়া হয়?
 - ক) যোজন বিধি
 - খ) নূন্যতম বর্গ পদ্ধতি
 - গ) বিয়োজন পদ্ধতি
 - ঘ) পুণরায়ন পদ্ধতি

- ৩। কোনটির ওপর নির্ভরশীল চলকের নির্ভরণ সহগ নির্ভর করে?
 - ক) মূল বিন্দুর পরিবর্তনের ওপর
 - খ) মাপনীর ওপর
 - গ) মূল বিন্দু ও মাপনী উভয়ের ওপর
 - ঘ) উপরের কোনটিই নয়

- ৪। নির্ভরণ সহগের মান কোনটি?
 - ক) $b_{xy} \geq -1$
 - খ) $-\alpha \leq b_{xy} \leq +\alpha$
 - গ) $b_{xy} \geq +1$
 - ঘ) $b_{xy} = i$ (অলীক সংখ্যা)



চূড়ান্ত মূল্যায়ন - ইউনিট ৬

সংক্ষিপ্ত ও রচনামূলক প্রশ্নাবলী

- ১। সহসম্পর্কের সংজ্ঞা লিখুন। সহসম্পর্ক কত প্রকার তাদের সম্পর্কে লিখুন।
- ২। সহসম্পর্কের সহগ ব্যাখ্যা করুন। সহসম্পর্কের বৈশিষ্ট্যগুলো লিখুন।
- ৩। অনুক্রমিক সহসম্পর্কের সংজ্ঞা লিখুন। কোন কোন ক্ষেত্রে অনুক্রমিক সহসম্পর্ক ব্যবহার করা হয় লিখুন।
- ৪। নির্ভরণের সংজ্ঞা লিখুন। নির্ভরণ রেখা কিভাবে অঙ্কন করা যায় লিখুন।
- ৫। নির্ভরণ রেখার সংজ্ঞা লিখুন। সহসম্পর্ক ও নির্ভরণের তুলনামূলক আলোচনা করুন।
- ৬। ঋণাত্মক ও ধনাত্মক সহসম্পর্কের সহগ এবং সংজ্ঞা লিখুন।
- ৭। দুজন বিচারক দশজন শ্রমিককে দক্ষতা অনুসারে নিম্নোক্ত অনুক্রমিক মান দিলে অনুক্রমিক সংশ্লেষাঙ্ক নির্ণয় করুন।

শ্রমিক	:	A	B	C	D	E	F	G	H	I	J
১ম বিচারক	:	৪	৮	৬	৭	১	৩	২	৫	১০	৯
২য় বিচারক	:	৩	৯	৬	৫	১	২	৪	৭	৮	১০

- ৮। ১০ জন দম্পতির বয়সের উপাত্ত নিম্নে দেয়া হলো। প্রদত্ত উপাত্ত থেকে সহসম্পর্কের সহগ নির্ণয় করুন ও মন্তব্য করুন।

স্ত্রীর বয়স (x)	১৮	২০	২০	২৪	২২	২৪	২৭	২৪	২১	২৫
স্বামীর বয়স (y)	২২	২৪	২৬	২৬	২৭	২৭	২৮	২৮	২৯	৩০

- ৯। দেয়া আছে,

$$\sum x = 150000, \sum x^2 = 2272500$$

$$\sum y = 90000, \sum y^2 = 4936000$$

$$\sum xy = 10522500 \text{ এবং } n = 1000$$

$$\text{নির্ণয় করুন i) } \bar{x} = ?, \text{ ii) } \bar{y} = ?, \text{ iii) } r = ?$$

- ১০। প্রদত্ত দুটি চলকের উপাত্ত থেকে x এর ওপর y এর নির্ভরণ সমীকরণ নির্ণয় করে x = ১০ হলে y এর অনুমিত মান নির্ণয় করুন।

x:	১	২	৩	৪	৫	৬	৭	৮
y:	৬	১	০	০	১	২	১	৫



উত্তরমালা - ইউনিট ৬

পাঠ ৬.১

১। গ ২। খ ৩। ক ৪। খ ৫। ক

পাঠ ৬.২

১। খ ২। খ ৩। খ ৪। খ

