सह-सम्बन्ध
(Correlation)

सांख्यिकी में सह-सम्बन्ध के अन्तर्गत यह ज्ञात किया जाता है कि दो या दो से अधिक समंक-श्रेणियों के चर-मूल्यों में कोई पारस्परिक सम्बन्ध है अथवा नहीं है और यदि कोई पारस्परिक सम्बंध है तो उसकी दिशा व परिमाण क्या है। यदि दो समंक-श्रेणियों के चर-मूल्य स्वतन्त्र रूप से घटते बढ़ते हैं अर्थात् एक श्रेणी के चर-मूल्यों में वृद्धि या ह्रास का दूसरी श्रेणी के चर-मूल्यों की वृद्धि या ह्रास पर कोई प्रभाव नहीं होता है तो उन समंक-श्रेणियों में सह-सम्बन्ध का अभाव माना जायेगा। इसके विपरीत यदि एक श्रेणी के चर-मूल्य का परिवर्तन दूसरी श्रेणी के चर-मूल्य को प्रभावित करता है तो वे दोनों समंक-श्रेणियाँ सह-सम्बन्धित कही जायेंगी।

सह-सम्बन्ध दो प्रकार का होता है—(i) धनात्मक (positive) या प्रत्यक्ष (direct) सह-सम्बन्ध तथा (ii) ऋणात्मक (negative), विलोम (inverse) या अप्रत्यक्ष सह-सम्बन्ध। जब एक श्रेणी के चर-मूल्य में वृद्धि होने पर दूसरी श्रेणी के चर-मूल्य में भी वृद्धि होती है अथवा एक में कमी आने पर दूसरे में भी कमी आती है तो चर-मूल्यों के इस सह-सम्बन्ध को धनात्मक कहा जायेगा। इसके विपरीत यदि वे चर-मूल्य इस प्रकार सम्बन्धित हैं कि एक चर-मूल्य में वृद्धि होने पर दूसरे चर-मूल्य में कमी होती है या एक चर-मूल्य में कमी होने पर दूसरे में वृद्धि होती है तो वे चर-मूल्य ऋणात्मक सह-सम्बंध वाले माने जायेंगे। धनात्मक व ऋणात्मक सह-सम्बंधों का अन्तर स्पष्ट करने के उद्देश्य से नीचे दो सारणियाँ दी गयी हैं। सारणी 24.62 में प्रदर्शित दोनों दशाएँ धनात्मक सह-सम्बन्ध को तथा सारणी 24.63 की दोनों दशाएँ ऋणात्मक सह-सम्बन्ध को प्रकट करती हैं।

सारणी 24.62 **धनात्मक सह-सम्बन्ध**

प्रथम दशा		द्वितीय दशा	
X श्रेणी	Y श्रेणी	X श्रेणी	Y श्रेणी
10	40	30	55
15	43	25	50
20	48	20	48
25	50	15	43
30	55	10	40

सारणी 24.63 **ऋणात्मक सह-सम्बन्ध**

प्रथम दशा		द्वितीय दशा	
X श्रेणी	Y श्रेणी	X श्रेणी	Y श्रेणी
10	55	30	40
15	50	25	43
20	48	20	48
25	43	15	50
30	40	10	55

सह-सम्बन्ध के परिमाण को सह-सम्बन्ध गुणांक (coefficient of correlation) के द्वारा व्यक्त करते हैं। सह-सम्बन्ध के अभाव में सह-सम्बन्ध गुणांक का मान शून्य होता है तथा सह-सम्बन्ध होने पर इसका मान $+1$ तथा -1 के मध्य कोई भी मूल्य हो सकता है। सह-सम्बन्ध गुणांक का मान $+1$ होने की स्थिति में पूर्ण धनात्मक सह-सम्बन्ध (perfect positive correlation) तथा -1 होने की दशा में पूर्ण ऋणात्मक सह-सम्बन्ध (perfect negative correlation) माना जाता है।

कभी-कभी सह-सम्बन्ध गुणांक के शाब्दिक विवेचन में 'उच्च', 'मध्यम' व 'निम्न' शब्दों का प्रयोग किया जाता है। उच्च स्तरीय धनात्मक सह-सम्बन्ध, मध्यम स्तरीय धनात्मक सह-सम्बन्ध तथा निम्न स्तरीय धनात्मक सह-सम्बन्ध उस दशा में कहे जाते हैं जब सह-सम्बन्ध गुणांक का मान क्रमशः $+0.75$ से 1, $+0.25$ से $+0.75$ तथा 0 से अधिक व 0.25 से कम होता है। इसी प्रकार सह-सम्बन्ध गुणांक का मान -0.75 से -1, -0.25 से -0.75 तथा 0 से -0.25 के मध्य होने की दशा में क्रमशः उच्च स्तरीय ऋणात्मक सह-सम्बन्ध, मध्यम स्तरीय ऋणात्मक सह-सम्बन्ध तथा निम्न स्तरीय ऋणात्मक सह-सम्बन्ध शब्दावली का प्रयोग करते हैं।

सह-सम्बन्ध ज्ञात करने की कुछ प्रमुख विधियों को नीचे समझाया गया है।

[I] कार्ल पियर्सन की सह-सम्बन्ध गुणांक विधि (Karl Pearson's coefficient of correlation method)

दो समंक-श्रेणियों में सह-सम्बन्ध ज्ञात करने की सांख्यिकीय विधियों में कार्ल पियर्सन की सह-सम्बन्ध गुणांक विधि को

सर्वोत्तम माना जाता है। यह गुणांक समान्तर माध्य एवं मानक विचलन पर आधारित है तथा इससे सह-सम्बन्ध की मात्रा एवं दिशा दोनों का बोध हो जाता है। कार्ल पियर्सन के सह-सम्बन्ध गुणांक को ज्ञात करने के लिये परिस्थितिनुसार प्रत्यक्ष अथवा लघु विधि का प्रयोग करते हैं।

1. प्रत्यक्ष विधि (Direct method)—कार्ल पियर्सन के सह-सम्बन्ध गुणांक का मूल सूत्र नीचे लिखा गया है।

$$r = \frac{\Sigma d_x d_y}{N \sigma_x \sigma_y}$$

इस सूत्र में r सह-सम्बन्ध गुणांक है; $\Sigma d_x d_y$ दो समंक-श्रेणियों (X तथा Y) के समान्तर माध्यों से निकाले गये विचलनों (अर्थात् $X-\overline{X}=d_x$ तथा $Y-\overline{Y}=d_y$) के गुणनफलों का योग है, N युग्म-पदों की संख्या है तथा $\sigma_x\sigma_y$ दोनों श्रेणियों के मानक विचलनों का गुणनफल है। उपर्युक्त सूत्रों को दो अन्य रूपों में निम्न प्रकार प्राप्त किया जा सकता है :

$$\text{(i)} \quad r = \frac{\Sigma d_x d_y}{N\sqrt{\frac{\Sigma d_x^2}{N}} \times \sqrt{\frac{\Sigma d_y^2}{N}}}$$

(मूल सूत्र का द्वितीय रूप)

यहाँ X तथा Y श्रेणियों के मानक विचलन के सूत्रों को मूल सूत्र में $\sigma_x \sigma_y$ के स्थान पर लिख दिया गया है।

(ii) मूल सूत्र के तृतीय रूप को निम्न प्रकार प्राप्त किया जा सकता है :

$$r = \frac{\Sigma d_x d_y}{N\sigma_x\sigma_y}$$

$$= \frac{\Sigma d_x d_y}{N\sqrt{\frac{\Sigma d_x^2}{N}} \times \sqrt{\frac{\Sigma d_y^2}{N}}}$$

$$= \frac{\Sigma d_x d_y}{N\sqrt{\frac{\Sigma d_x^2 \times \Sigma d_y^2}{N \times N}}}$$

$$= \frac{\Sigma d_x d_y}{\frac{N}{N}\sqrt{\Sigma d_x^2 \times \Sigma d_y^2}}$$

$$= \frac{\Sigma d_x d_y}{\sqrt{\Sigma d_x^2 \times \Sigma d_y^2}}$$

(मूल सूत्र का तृतीय रूप)

यहाँ यह उल्लेखनीय है कि सह-सम्बन्ध गुणांक ज्ञात करने की प्रत्यक्ष विधि में उपर्युक्त तीनों सूत्रों में किसी का भी प्रयोग किया जा सकता है परन्तु गणना-कार्य की सरलता के विचार से अन्तिम या तृतीय सूत्र को प्रयोग में लाना हितकर रहता है। इन सूत्रों में प्रयुक्त संकेताक्षरों के मान निम्न प्रकार परिकलित करते हैं :

(1) सर्वप्रथम दोनों समंक-श्रेणियों के अलग-अलग समान्तर माध्य अर्थात् $\overline{X}$ तथा $\overline{Y}$ ज्ञात करते हैं।

(2) $(X-\overline{X})$ तथा $(Y-\overline{Y})$ सूत्रों की सहायता से दोनों श्रेणियों के विभिन्न पद-मूल्यों का सम्बन्धित समान्तर माध्य से विचलन अर्थात् d_x व d_y ज्ञात करते हैं।

(3) प्रत्येक श्रेणी के विचलन मूल्यों का वर्ग करते हैं तथा इन विचलन वर्गों को अलग-अलग जोड़कर $\Sigma d_x{}^2$ व $\Sigma d_y{}^2$ के मान निकालते हैं।

(4) दोनों श्रेणियों के सम्बन्धित विचलन मूल्यों (अर्थात् d_x व d_y) की गुणा करते हैं तथा इन गुणनफलों को जोड़कर $\Sigma d_x d_y$ का मान प्राप्त कर लेते हैं।

(5) यदि सह-सम्बन्ध गुणांक के मूल सूत्र को प्रयोग में लाना हो

तो $\sigma_x = \sqrt{\frac{\Sigma d_x^2}{N}}$

तथा $\sigma_y = \sqrt{\frac{\Sigma d_y^2}{N}}$ सूत्रों की

सहायता से दोनों श्रेणियों के अलग-अलग मानक विचलन (अर्थात् σ_x व σ_y के मान) ज्ञात कर लेने चाहिएँ।

(6) इस प्रकार ज्ञात किये गये मानों को ऊपर बतलाये गये किसी भी सूत्र में रखकर गणना करने पर कार्ल पियर्सन के सह-सम्बन्ध गुणांक का मान प्राप्त हो जाता है।

उदाहरण (62) दस कृषि-फार्मों में के पूँजी निवेश तथा लाभ के निम्नांकित समंकों में कार्ल पियर्सन का सह-सम्बन्ध गुणांक ज्ञात कीजिये :

पूँजी-निवेश (000 रु०)	लाभ (000 रु०)
100	30
90	22
80	20
70	14
60	15
50	10
40	5
30	8
20	4
10	2

हल—

सारणी 24.64

कार्ल पियर्सन के सह-सम्बंध गुणांक की गणना (प्रत्यक्ष विधि)

पूँजी-निवेश (X)			लाभ (Y)			d_x व d_y का गुणनफल
पूँजी (000 रु०)	$\overline{X}$= 55 से विचलन	विचलन का वर्ग	लाभ (000 रु०)	$\overline{Y}$= 13 से विचलन	विचलन का वर्ग	
X	$X-\overline{X} = d_x$	d_x^2	Y	$Y-\overline{Y} = d_y$	d^2_y	d_xd_y
100	+45	2025	30	+17	289	765
90	+35	1225	22	+ 9	81	315
80	+25	625	20	+ 7	49	175
70	+15	225	14	+ 1	1	15
60	+ 5	25	15	+ 2	4	10
50	− 5	25	10	− 3	9	15
40	−15	225	5	− 8	64	120
30	−25	625	8	− 5	25	125
20	−35	1225	4	− 9	81	315
10	−45	2025	2	−11	121	495
$\Sigma X=550$ N=10		$\Sigma d_x^2=8250$	$\Sigma Y=130$ N=10		$\Sigma d_y^2=724$	$\Sigma d_xd_y=2350$

$\overline{X}=\frac{\Sigma X}{N} = 550/10 = 55$ रुपये

$\overline{Y}=\frac{\Sigma Y}{N} = 130/10 = 13$ रुपये

∴ कार्ल पियर्सन का सह-सम्बन्ध गुणांक अथवा

$$r = \frac{\Sigma d_x d_y}{\sqrt{\Sigma d_x^2 \times \Sigma d_y^2}} = \frac{2350}{\sqrt{8250 \times 724}}$$

$$= \frac{2350}{\sqrt{5973000}} = \frac{2350}{2443.97}$$

$$= +0.96$$

अत: पूँजी निवेश तथा लाभ में उच्च स्तरीय धनात्मक सह-सम्बन्ध है।

2. लघु विधि (Short-cut method)—यदि किसी समंक-श्रेणी में समान्तर माध्य का मान पूर्णांकों में प्राप्त नहीं होता तो ऊपर समझायी गयी प्रत्यक्ष विधि के द्वारा सह-सम्बन्ध गुणांक का परिकलन कुछ कठिन हो जाता है। अत: ऐसी दशा में लघु-विधि के द्वारा गणना-कार्य को अपेक्षाकृत सरल बनाया जा सकता है। इस विधि में निम्न प्रकार गणनाएँ की जाती हैं :

(1) सर्वप्रथम दोनों श्रेणियों में अलग-अलग सुविधाजनक मूल्यों को कल्पित माध्य मान लेते हैं। X श्रेणी के कल्पित माध्य को A_x तथा Y श्रेणी के कल्पित माध्य को A_y संकेताक्षर से प्रकट करते हैं।

(2) इन कल्पित माध्यों से सम्बन्धित समंक-श्रेणी के विभिन्न पद-मूल्यों का विचलन (अर्थात् d_x व d_y) ज्ञात करते हैं। इस कार्य के लिये पद-मूल्य में कल्पित माध्य को घटाया जाता है अर्थात् $X - A_x = d_x$ तथा $Y - A_y = d_y$ ।

(3) विचलनों के अलग-अलग योग, अर्थात् Σd_x व Σd_y के मान निकालते हैं।

(4) विचलनों का वर्ग करते हैं तथा इन वर्गों को जोड़कर Σd_x^2 व Σd_y^2 के मान निकालते हैं।

(5) दोनों समंक-श्रेणियों के सम्बन्धित विचलन-मूल्यों की आपस में गुणा करते है तथा इन गुणनफलों को जोड़कर $\Sigma d_x d_y$ का मान ज्ञात करते हैं।

(6) अन्त में नीचे लिखे गये दो सूत्रों में किसी के भी द्वारा कार्ल पियर्सन के सह-सम्बन्ध गुणांक की गणना की जा सकती है परन्तु प्रथम सूत्र के प्रयोग में अधिक सरलता रहती है :

प्रथम सूत्र,

$$r = \frac{\Sigma d_x d_y - \frac{\Sigma d_x \times \Sigma d_y}{N}}{\sqrt{\left[\Sigma d_x^2 - \frac{(\Sigma d_x)^2}{N}\right]\left[\Sigma d_y^2 - \frac{(\Sigma d_y)^2}{N}\right]}}$$

द्वितीय सूत्र,

$$r = \frac{N \times \Sigma d_x d_y - (\Sigma d_x \times \Sigma d_y)}{\sqrt{[N \times \Sigma d_x^2 - (\Sigma d_x)^2][N \times \Sigma d_y^2 - (\Sigma d_y)^2]}}$$

उदाहरण (63) निम्नांकित समंकों में कार्ल पियर्सन की लघु विधि के द्वारा सह-सम्बन्ध गुणांक ज्ञात कीजिये :

श्रेणी A	श्रेणी B
50	22
54	25
56	34
59	28
60	26
62	30
61	32
65	30
67	28
71	34
71	36
74	40

हल—

सारणी 24.65 कार्ल पियर्सन के सह-सम्बंध गुणांक की गणना (लघु विधि)

श्रेणी A			श्रेणी B			विचलन-युग्मों का गुणनफल
पद-मूल्य	कल्पित माध्य (A_x = 60) से विचलन	विचलन का वर्ग	पद-मूल्य	कल्पित माध्य (A_y = 30) से विचलन	विचलन का वर्ग	
X	d_x	d_x^2	Y	d_y	d_y^2	dxdy
50	−10	100	22	− 8	64	+ 80
54	− 6	36	25	− 5	25	+ 30
56	− 4	16	34	+ 4	16	− 16
59	− 1	1	28	− 2	4	+ 2
60	0	0	26	− 4	16	0
62	+ 2	4	30	0	0	0
61	+ 1	1	32	+ 2	4	+ 2
65	+ 5	25	30	0	0	0
67	+ 7	49	28	− 2	4	− 14
71	+11	121	34	+ 4	16	+ 44
71	+11	121	36	+ 6	36	+ 66
74	+14	196	40	+10	100	+140
N = 12	Σd_x = +30	Σd_x^2 = 670	N = 12	Σd_y = +5	Σd_y^2 = 285	$\Sigma d_x d_y$ = +334

प्रथम सूत्र के अनुसार सह-सम्बन्ध गुणांक अथवा

$$r = \frac{\Sigma d_x d_y - \frac{\Sigma d_x \times \Sigma d_y}{N}}{\sqrt{\left[\Sigma d_x^2 - \frac{(\Sigma d_x)^2}{N}\right]\left[\Sigma d_y^2 - \frac{(\Sigma d_y)^2}{N}\right]}}$$

$$= \frac{334 - \frac{30 \times 5}{12}}{\sqrt{\left[670 - \frac{(30)^2}{12}\right]\left[285 - \frac{(5)^2}{12}\right]}}$$

$$= \frac{334 - 12.5}{\sqrt{(670 - 75)(285 - 2.083)}}$$

$$= \frac{321.5}{\sqrt{595 \times 282.917}} = \frac{321.5}{\sqrt{168335.6}} = \frac{321.5}{410.287}$$

$$= +0.783$$

अतः दोनों श्रेणियों में उच्च धनात्मक सह-सम्बन्ध है।

द्वितीय सूत्र के अनुसार सह-सम्बन्ध गुणांक का मान निम्न प्रकार परिकलित किया जा सकता है :

$$r = \frac{N \times \Sigma d_x d_y - (\Sigma d_x \times \Sigma d_y)}{\sqrt{[N \times \Sigma d_x^2 - (\Sigma d_x)^2]\,[N \times \Sigma d_y^2 - (\Sigma d_y)^2]}}$$

$$= \frac{12 \times 334 - (30 \times 5)}{\sqrt{[12 \times 670 - (30)^2]\,[12 \times 285 - (5)^2]}}$$

$$= \frac{4008 - 150}{\sqrt{(8040 - 900)\,(3420 - 25)}}$$

$$= \frac{3858}{\sqrt{7140 \times 3395}} = \frac{3858}{\sqrt{24240300}} = \frac{3858}{4923.44}$$

$$= +0.783$$

उदाहरण (64) आयु तथा खेलने की आदत सम्बन्धी निम्नांकित समंकों में कार्ल पियर्सन का सह-सम्बन्ध गुणांक ज्ञात कीजिये :

आयु	छात्रों की संख्या	नियमित खिलाड़ी
15	250	200
16	200	150
17	150	90
18	120	48
19	100	30
20	80	12

हल—इस प्रश्न को हल करने के लिये पहले निम्नांकित ढंग से अलग-अलग आयु के छात्रों में नियमित खिलाड़ियों का प्रतिशत ज्ञात किया जायेगा :

$\because$ 250 छात्रों में नियमित खिलाड़ी $= 200$

$\therefore$ 100 छात्रों में नियमित खिलाड़ी $= \dfrac{200 \times 100}{250}$

$= 80$

उपर्युक्त ढंग से गणना करने पर ज्ञात होता है कि 15, 16, 17, 18, 19 व 20 वर्ष की आयु वाले छात्रों में नियमित खिलाड़ियों का प्रतिशत क्रमशः $(200 \times 100)/250 = 80$, $(150 \times 100)/200 = 75$, $(90 \times 100)/150 = 60$, $(48 \times 100)/120 = 40$, $(30 \times 100)/100 = 30$ तथा $(12 \times 100)/80 = 15$ आता है। अब आयु के समंकों को X श्रेणी के पद-मूल्य तथा नियमित खिलाड़ियों के प्रतिशतों को Y श्रेणी के पद-मूल्य मानते हुए निम्न प्रकार गणना कीजिये (सारणी 24.66) :

सारणी 24.66 **कार्ल पियर्सन के सह-सम्बंध गुणांक की गणना (लघु विधि)**

X श्रेणी (आयु)			Y श्रेणी (खेलने की आदत)			d_x व d_y का गुणनफल
आयु	कल्पित माध्य ($A_x = 17$) से विचलन	विचलन का वर्ग	खिलाड़ी %	कल्पित माध्य ($A_y = 60$) से विचलन	विचलन का वर्ग	
X	d_x	d_x^2	Y	d_y	d_y^2	d_xd_y
15	−2	4	80	+20	400	−40
16	−1	1	75	+15	225	−15
17	0	0	60	0	0	0
18	+1	1	40	−20	400	−20
19	+2	4	30	−30	900	−60
20	+3	9	15	−45	2025	−135
N = 6	$\Sigma d_x = +3$	$\Sigma d_x^2 = 19$	N = 6	$\Sigma d_y = -60$	$\Sigma d_y^2 = 3950$	$\Sigma d_xd_y = -270$

कार्ल पियर्सन का सह-सम्बन्ध गुणांक अथवा

$$r = \frac{\Sigma d_x d_y - \dfrac{\Sigma d_x \times \Sigma d_y}{N}}{\sqrt{\left[\Sigma d_x^2 - \dfrac{(\Sigma d_x)^2}{N}\right]\left[\Sigma d_y^2 - \dfrac{(\Sigma d_y)^2}{N}\right]}} = \frac{-270 - \dfrac{3\,(-60)}{6}}{\sqrt{\left[19 - \dfrac{(3)^2}{6}\right]\left[3950 - \dfrac{(-60)^2}{6}\right]}}$$

$$= \frac{-270-(-30)}{\sqrt{\left[19-\frac{9}{6}\right]\left[3950-\frac{3600}{6}\right]}}$$

$$= \frac{-270+30}{\sqrt{(19-1.5)(3950-600)}}$$

$$= \frac{-240}{\sqrt{17.5\times 3350}} = \frac{-240}{\sqrt{58625}} = \frac{-240}{242.13}$$

$$= -0.99$$

अतः उच्च कोटि के इस ऋणात्मक सह-सम्बन्ध गुणांक से स्पष्ट होता है कि आयु बढ़ने के साथ-साथ खेलने की आदत में कमी हुई है।

[II] स्पियरमैन की कोटि-अन्तर विधि

(Spearman's rank difference method)

चार्ल्स स्पियरमैन (Charles Spearman) नामक सांख्यिकी-विद् ने व्यक्तिगत समंक-श्रेणियों के विभिन्न पद-मूल्यों की कोटियों (rank) के आधार पर सह-सम्बन्ध गुणांक ज्ञात करने की एक सरल विधि प्रतिपादित की थी, जिसे उनके नाम पर स्पियरमैन की कोटि-अन्तर विधि कहते हैं। जिस प्रकार प्राप्तांकों की संख्या के आधार पर परीक्षार्थियों को प्रथम, द्वितीय, तृतीय आदि क्रम या स्थान (position) देते हैं ठीक उसी तरह किसी समंक-श्रेणी के विभिन्न पद-मूल्यों को, उनके आकार (size) या मान के अनुसार 1, 2, 3, 4, 5 आदि कोटियाँ प्रदान की जा सकती हैं। उदाहरणार्थ, यदि किसी समंक-श्रेणी में 10, 8, 3, 7 व 15 कोई पाँच पद-मूल्य हैं तो स्पष्ट है कि इन पद-मूल्यों की कोटियाँ क्रमशः 2, 3, 5, 4 व 1 होंगी।

यहाँ यह संकेत करना आवश्यक है कि कभी-कभी किसी श्रेणी में एक ही मूल्य की दो या दो से अधिक बार पुनरावृत्ति हो जाती है। ऐसे मूल्य या मूल्यों को औसत कोटि (average rank) प्रदान करते हैं। उदाहरणार्थ, मान लीजिये किसी श्रेणी में 75, 65, 65, 55, 45, 45, 45 व 35 कोई आठ पद-मूल्य हैं तो स्पष्ट है कि सबसे बड़े मूल्य अर्थात् 75 की कोटि 1 होगी। 75 से अगला छोटा मूल्य 65 है जो श्रेणी में दो बार आया है अतः इस मूल्य की कोटि निर्धारित करने के लिये अगली दो कोटियों का औसत निकाला जायेगा जो $(2+3)/2=2.5$ होगा। इस तरह दोनों स्थानों पर 65 की कोटि 2.5 मानी जायेगी। 55 की कोटि 4 होगी क्योंकि 3 तक की कोटियों को पहले गिना जा चुका है। 55 के बाद छोटा मूल्य 45 है जो श्रेणी में तीन स्थानों पर आया है अतः इस मूल्य की कोटि, चौथी कोटि से अगली तीन कोटियों का औसत अर्थात् $(5+6+7)/3=6$ होगी। अन्तिम मूल्य 35 को 8वीं कोटि दी जायेगी क्योंकि 7 तक की कोटियों को पहले गिना जा चुका है।

स्पियरमैन की विधि के अनुसार निम्न प्रकार कोटि सह-सम्बन्ध गुणांक ज्ञात करते हैं :

(1) सर्वप्रथम प्रत्येक समंक-श्रेणी के विभिन्न पद-मूल्यों को उनके आकार के अनुसार कोटियाँ (ranks) देते हैं।

(2) X तथा Y श्रेणियों के तत्सम्बन्धी पद-मूल्यों की कोटियों का अन्तर (D) ज्ञात करते हैं। इसके लिये X श्रेणी की कोटि में तत्सम्बन्धी मूल्य की Y श्रेणी में लिखी कोटि को घटाया जाता है, अर्थात्

$$D = (X \text{ श्रेणी में कोटि} - Y \text{ श्रेणी में कोटि})$$

(3) इन प्रकार प्राप्त कोटि-अन्तर के मानों का वर्ग (D^2) करते हैं तथा इन वर्गों को जोड़कर ΣD^2 का मान निकाल लेते हैं।

(4) अब यदि किसी भी श्रेणी में दो या दो से अधिक पद-मूल्यों की कोटियाँ समान नहीं हैं अर्थात् X या Y श्रेणी में किसी भी मूल्य की पुनरावृत्ति नहीं हुई है तो निम्नांकित सूत्र की सहायता से स्पियरमैन के कोटि सह-सम्बन्ध गुणांक की गणना की जा सकती है :

$$\rho = 1 - \frac{6\,[\Sigma D^2]}{N^3 - N}$$

इस सूत्र में ग्रीक वर्णमाला का ρ (rho) अक्षर कोटि सह-सम्बन्ध गुणांक को, ΣD^2 कोटि अन्तर के वर्गों के योग को तथा N पद-युग्मों की संख्या को प्रकट करता है।

जैसा कि हम ऊपर पढ़ चुके हैं, किसी समंक-श्रेणी में 1 या 1 से अधिक पद-मूल्यों की 2 या 2 से अधिक बार पुनरावृत्ति हो सकती है तथा इस प्रकार के समान आकार वाले मूल्यों को औसत के आधार पर समान कोटि प्रदान करते हैं। इस प्रकार की समंक-श्रेणियों में कोटि सह-सम्बन्ध ज्ञात करने के लिये उपर्युक्त सूत्र को निम्न प्रकार संशोधित करना अनिवार्य होता है :

$$\rho = 1 - \frac{6\,[\Sigma D^2 + \frac{1}{12}(m^3 - m) + \cdots]}{N^3 - N}$$

इस सूत्र में m उन पद-मूल्यों की संख्या को प्रकट करता है जिनकी कोटि समान है। वस्तुतः उपर्युक्त संशोधित सूत्र को केवल

उस दशा में प्रयोग किया जा सकता है जब दोनों श्रेणियों में केवल किसी एक पद-मूल्य की पुनरावृत्ति हुई हो। यदि इन श्रेणियों में दो पद-मूल्यों की पुनरावृत्ति होती है तो संशोधित सूत्र में अंकित बिन्दुओं के स्थान पर पुनः $1/12(m^3-m)$ लिखना आवश्यक हो जायेगा। संक्षेप में, सूत्र को संशोधित करने से पूर्व यह देख लेते हैं कि दोनों श्रेणियों में कुल कितने पद-मूल्यों की पुनरावृत्ति हुई है तथा जितने पद-मूल्यों की पुनरावृत्ति होती है उतनी ही बार सूत्र में $1/12(m^3-m)$ जोड़ देते हैं।

उदाहरण (65) निम्नांकित आँकड़ों से कोटि-अन्तर विधि के द्वारा सह-सम्बन्ध गुणांक ज्ञात कीजिये :

X श्रेणी	Y श्रेणी
8	84
36	51
98	91
25	60
75	68
82	62
92	86
62	58
65	35
39	49

हल—

सारणी 24.67 **कोटि सह-सम्बन्ध की गणना**

X श्रेणी		Y श्रेणी		कोटि अन्तर	कोटि-अन्तरों के वर्ग
मूल्य	कोटि	मूल्य	कोटि		
	X		Y	X–Y = D	D^2
8	10	84	3	7	49
36	8	51	8	0	0
98	1	91	1	0	0
25	9	60	6	3	9
75	4	68	4	0	0
82	3	62	5	−2	4
92	2	86	2	0	0
62	6	58	7	−1	1
65	5	35	10	−5	25
39	7	49	9	−2	4
N=10					$\Sigma D^2 = 92$

चूँकि प्रत्येक श्रेणी में दिये गये मूल्यों का आकार एक दूसरे से भिन्न है अर्थात् कोई भी मूल्य एक से अधिक स्थानों पर नहीं आया है अतः कोटि सह-सम्बन्ध के सूत्र को संशोधित करने की कोई आवश्यकता नहीं है।

∴ कोटि सह-सम्बन्ध गुणांक अथवा

$$\rho = 1 - \frac{6(\Sigma D^2)}{N^3 - N}$$

$$= 1 - \frac{6 \times 92}{(10)^3 - 10}$$

$$= 1 - \frac{552}{1000 - 10}$$

$$= 1 - \frac{552}{990} = 1 - 0.5576$$

$$= +\ 0.442$$

उदाहरण (66) निम्नांकित सारणी में 10 ग्रामों की सिंचित भूमि तथा चावल के अर्न्तगत बोये गये क्षेत्र के समंक दिये गये हैं। कोटि-अन्तर विधि के द्वारा सिंचित क्षेत्र व चावल क्षेत्र के मध्य सह-सम्बन्ध गुणांक ज्ञात कीजिये।

ग्राम	सिंचित क्षेत्र (हेक्टेअर)	चावल क्षेत्र (हेक्टेअर)
A	48	13
B	33	13
C	40	24
D	9	6
E	16	15
F	16	4
G	65	20
H	24	9
I	16	6
J	57	19

हल—

सारणी 24.68 **कोटि सह-सम्बन्ध की गणना**

ग्राम	X श्रेणी		Y श्रेणी		कोटि अन्तर (X–Y)	कोटि अन्तरों का वर्ग
	सिंचित क्षेत्र (हेक्टेअर)	कोटि	चावल क्षेत्र (हेक्टेअर)	कोटि		
		X		Y	D	D^2
A	48	3	13	5.5	−2.5	6.25
B	33	5	13	5.5	−0.5	0.25
C	40	4	24	1	3.0	9.00
D	9	10	6	8.5	1.5	2.25
E	16	8	15	4	4.0	16.00
F	16	8	4	1.0	−2.0	4.00
G	65	1	20	2	−1.0	1.00
H	24	6	9	7	−1.0	1.00
I	16	8	6	8.5	−0.5	0.25
J	57	2	19	3	−1.0	1.00
N = 10						$\Sigma D^2 = 41$

उपर्युक्त गणना कार्य से दो तथ्यों पर प्रकाश पड़ता है — **प्रथम**, दोनों श्रेणियों में कुल तीन मूल्यों (अर्थात् 16, 13 व 6) की पुनरावृत्ति हुई है अतः कोटि सह-सम्बन्ध गुणांक के संशोधित सूत्र के अंश में, ΣD^2 में तीन बार $1/12\ (m^3 - m)$ जोड़ने के पश्चात् 6 की गुणा की जायेगी। **द्वितीय**, चूँकि 16, 13 व 6 मूल्यों की पुनरावृत्ति क्रमशः 3, 2 व 2 बार हुई है इसलिये तीन बार लिखे गये $1/12(m^3 - m)$ में m का मान पहले कोष्ठक में 3, दूसरे कोष्ठक में 2 व तीसरे कोष्ठक में 2 रखा जायेगा।

∴ कोटि सह-सम्बन्ध गुणांक का संशोधित सूत्र,

$$\rho = 1 - \frac{6[\Sigma D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m)]}{N^3 - N}$$

$$= 1 - \frac{6[41 + \frac{1}{12}(3^3 - 3) + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(2^3 - 2)]}{10^3 - 10}$$

$$= 1 - \frac{6[41 + \frac{1}{12}(24) + \frac{1}{12}(6) + \frac{1}{12}(6)]}{1000 - 10}$$

$$= 1 - \frac{6(41 + 2 + 0.5 + 0.5)}{990}$$

$$= 1 - \frac{264}{990}$$

$$= 1 - 0.267$$

$$= +0.733$$

उदाहरण (67) निम्नांकित सूचना से जनसंख्या के घनत्व एवं मृत्यु-दर में स्पियरमैन की कोटि-अन्तर विधि के द्वारा सह-सम्बन्ध गुणांक की गणना कीजिये :

प्रदेश	क्षेत्रफल (वर्ग किमी)	जनसंख्या	मृत्यु संख्या
A	200	40,000	480
B	150	75,000	1200
C	120	72,000	1080
D	80	20,000	280

हल—इस प्रश्न को हल करने से पूर्व दो बातों को समझ लेना चाहिए—**प्रथम**, जनसंख्या का घनत्व ज्ञात करने के लिये कुल जनसंख्या में क्षेत्रफल का भाग दिया जाता है। उदाहरणार्थ, इस प्रश्न के A प्रदेश में जनसंख्या का प्रति वर्ग किमी घनत्व 40,000/200 = 200 मनुष्य है। **द्वितीय**, मृत्यु-दर (death rate) निकालने के लिये मृत्यु-संख्या (number of deaths) में 1000 की गुणा करके, गुणनफल को कुल जनसंख्या से विभाजित कर देते हैं। अतः A प्रदेश में मृत्यु-दर (480 × 1000)/40,000 = 12 मनुष्य प्रति हजार है।

दूसरे शब्दों में,

$$\text{प्रति वर्ग किमी घनत्व} = \frac{\text{कुल जनसंख्या}}{\text{कुल क्षेत्रफल (वर्ग किमी)}}$$

$$\text{मृत्यु दर} = \frac{\text{मृत्यु संख्या}}{\text{कुल जनसंख्या}} \times 1000$$

सारणी 24.69 कोटि सह-सम्बन्ध गुणांक की गणना

प्रदेश	क्षेत्रफल (वर्ग किमी)	जनसंख्या	X श्रेणी		Y श्रेणी		कोटि-अन्तर	कोटि-अन्तर का वर्ग
			प्रति वर्ग किमी घनत्व	कोटि	प्रति हजार मृत्यु-दर	कोटि		
				X		Y	D	D^2
A	200	40,000	200	4	12	4	0	0
B	150	75,000	500	2	16	1	1	1
C	120	72,000	600	1	15	2	−1	1
D	80	20,000	250	3	14	3	0	0
N=4								$\Sigma D^2=2$

∴ कोटि सह-सम्बन्ध गुणांक अथवा

$$\rho = 1 - \frac{6[\Sigma D^2]}{N^3 - N} = 1 - \frac{6 \times 2}{(4)^3 - 4}$$

$$= 1 - \frac{12}{64-4} = 1 - \frac{1}{5} = \frac{4}{5}$$

$$= +\ 0.80$$

[III] संगामी विचलन विधि

(Concurrent deviation method)

गणना-कार्य के विचार से सह-सम्बन्ध ज्ञात करने की सांख्यिकीय विधियों में संगामी विचलन विधि सबसे सरल है। जैसा कि हम आगे पढ़ेंगे, इस विधि में विचलन-मूल्यों के आकार (size) के बजाय केवल उनके धनात्मक या ऋणात्मक चिह्नों के अनुसार गणना करते हैं। इसी कारणवश संगामी विचलन गुणांक (coefficient of concurrent deviation) से सह-सम्बन्ध की मात्रा का सही-सही ज्ञान नहीं होता। इस दोष ने संगामी विचलन विधि के प्रयोग को बहुत सीमित कर दिया है। दूसरे शब्दों में, दो समंक-श्रेणियों में केवल सह-सम्बन्ध की दिशा ज्ञात करने के लिये इस विधि का प्रयोग हो सकता है। इस विधि में निम्न प्रकार गणनाएँ की जाती हैं :

(1) सर्वप्रथम दोनों श्रेणियों के प्रत्येक पद-मूल्य का, उसके पूर्ववर्ती मूल्य से विचलन निकालते हैं तथा इस विचलन को परिस्थितिनुसार केवल (+), (–) या (=) के बीजगणितीय चिह्न के द्वारा प्रकट करते हैं अर्थात् विचलन का अंकों में मान नहीं लिखा जाता। इस प्रकार यदि कोई पद-मूल्य अपने पूर्ववर्ती पद-मूल्य की तुलना में बड़ा, छोटा अथवा बराबर है तो उसके विचलन को क्रमशः +, – अथवा = के चिह्न से प्रकट किया जायेगा। चूँकि श्रेणी के प्रथम मूल्य से पहले और कोई मूल्य नहीं होता है अतः इस मूल्य का कोई विचलन नहीं बतलाया जा सकता। इस प्रकार दोनों श्रेणियों के प्रथम पद-मूल्य को छोड़ देने के कारण विचलन-युग्मों की संख्या (n),. पद-युग्मों की संख्या (N) से 1 कम हो जाती है अर्थात् $n = N - 1$ होता है।

(2) दोनों श्रेणियों के तत्सम्बन्धी विचलन-चिह्नों की गुणा करके धनात्मक चिह्नों की संख्या गिन लेते हैं। यह संख्या **संगामी विचलनों की संख्या** (number of concurrent deviation) कहलाती है तथा इसे संगामी-विचलन-गुणांक ज्ञात करने के सूत्र में अंग्रेज़ी वर्णमाला के छोटे c अक्षर से प्रकट करते हैं।

(3) अन्त में निम्नांकित सूत्र का प्रयोग करते हैं :

$$r_c = \pm\sqrt{\pm\frac{(2c - n)}{n}}$$

इस सूत्र में, r_c का अर्थ संगामी विचलन गुणांक, c का अर्थ संगामी विचलनों की संख्या तथा n का अर्थ विचलन-युग्मों की संख्या समझना चाहिए।

यहाँ यह बात ध्यान देने योग्य है कि ऋणात्मक मूल्यों का वर्गमूल निकालना असम्भव होता है। अतः यदि $(2c-n)$ का मान ऋणात्मक है तो उपर्युक्त सूत्र में वर्गमूल के चिह्न से पहले और बाद में, दोनों स्थानों पर (–) का चिह्न अंकित कर देते हैं। ऐसा करने से $2c-n$ का मान धनात्मक हो जाता है तथा उत्तर पर भी कोई विपरीत प्रभाव नहीं पड़ता। यदि $2c-n$ का मान धनात्मक है तो वर्गमूल के चिह्न से पहले या बाद में कोई अतिरिक्त + या – का चिह्न लगाने की आवश्यकता नहीं होती।

उदाहरण (68) 11 श्रमिकों की दैनिक मजदूरी व दैनिक व्यय के निम्नांकित समंकों से संगामी विचलन विधि के द्वारा आय व व्यय के मध्य सह-सम्बन्ध ज्ञात कीजिये :

श्रमिक	दैनिक मजदूरी (रु०)	दैनिक व्यय (रु०)
A	65	60
B	40	55
C	35	50
D	75	56
E	63	30
F	80	70
G	35	40
H	20	35
I	80	80
J	60	75
K	50	80

सारणी 24.70 **सह-सम्बन्ध की गणना (संगामी विचलन विधि)**

श्रमिक	X श्रेणी		Y श्रेणी		विचलन चिह्नों का गुणनफल
	दैनिक आय (रु०) X	विचलन-चिह्न	दैनिक व्यय (रु०) Y	विचलन-चिह्न	
A	65		60		
B	40	–	55	–	+
C	35	–	50	–	+
D	75	+	56	+	+
E	63	–	30	–	+
F	80	+	70	+	+
G	35	–	40	–	+
H	20	–	35	–	+
I	80	+	80	+	+
J	60	–	75	–	+
K	50	–	80	+	–
N = 11		n = 10			c = 9

∴ संगामी विचलन गुणांक या,

$$r_c = \sqrt{\frac{(2c-n)}{n}} = \sqrt{\frac{(2\times 9)-10}{10}}$$

$$= \sqrt{\frac{18-10}{10}} = \sqrt{\frac{8}{10}} = \sqrt{0.8}$$

$$= +\ 0.894$$

अतः दैनिक आय तथा व्यय में उच्च स्तर का धनात्मक सह-सम्बन्ध है।

उदाहरण (69) किसी वस्तु की विभिन्न वर्षों में माँग (demand) तथा आपूर्ति (supply) के निम्नांकित समंकों से संगामी विचलन विधि के द्वारा सह-सम्बन्ध गुणांक ज्ञात कीजिये:

वर्ष	माँग	आपूर्ति
1984	25	27
1985	23	36
1986	22	39
1987	34	39
1988	33	32
1989	32	32
1990	33	40
1991	32	33
1992	33	59
1993	34	36
1994	36	44
1995	38	36

हल–

सारणी 24.71 **सह-सम्बन्ध की गणना (संगामी विचलन विधि)**

वर्ष	X श्रेणी		Y श्रेणी		विचलन चिह्नों का गुणनफल
	माँग X	विचलन-चिह्न	आपूर्ति Y	विचलन-चिह्न	
1984	25		27		
1985	23	−	36	+	−
1986	22	−	39	+	−
1987	34	+	39	=	
1988	33	−	32	−	+
1989	32	−	32	=	
1990	33	+	40	+	+
1991	32	−	33	−	+
1992	33	+	59	+	+
1993	34	+	36	−	−
1994	36	+	44	+	+
1995	38	+	36	−	−
N = 12		n = 11			c = 5

∴ संगामी विचलन-गुणांक या,

$$r_c = \sqrt{\frac{(2c - n)}{n}} = \sqrt{\frac{(2 \times 5) - 11}{11}}$$

$$= \sqrt{\frac{10 - 11}{11}} = \sqrt{-\frac{1}{11}}$$

$$= -\sqrt{-\frac{(-1)}{11}}$$

(∵ $2c - n$ का मान ऋणात्मक है।)

$$= -\sqrt{\frac{1}{11}} = -\sqrt{0.09}$$

$$= -0.3$$

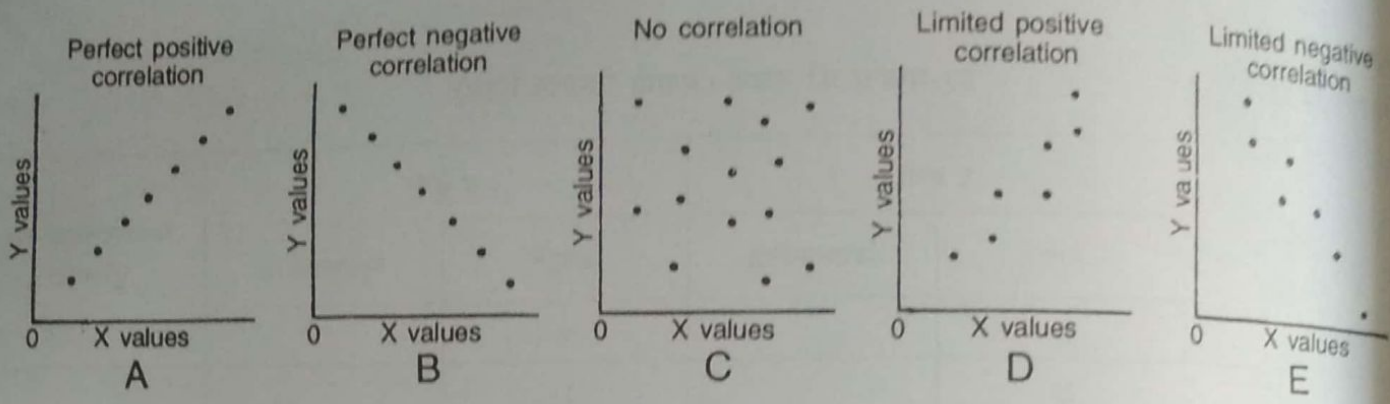
अतः माँग व आपूर्ति में मध्यम स्तर का ऋणात्मक सह-सम्बन्ध है।

[IV] प्रकीर्ण आरेख के द्वारा सह-सम्बन्ध का प्रदर्शन (Representation of correlation by scatter diagram)

प्रकीर्ण आरेख में विभिन्न इकाइयों (units) की उनके X तथा Y मूल्यों के अनुसार ग्राफ पेपर पर बिन्दुओं द्वारा स्थिति प्रदर्शित की जाती है। प्रकीर्ण आरेख की रचना करने के लिये ग्राफ पेपर अथवा सादे कागज़ पर खींचे गये भुजाक्ष (X-axis) पर X मूल्यों के मान तथा कोटि अक्ष (Y-axis) पर Y मूल्यों के मान समान अन्तर पर लिख देते हैं जैसा कि साधारण आलेखों की रचना में किया जाता है। इसके पश्चात् प्रत्येक इकाई की स्थिति उसके X तथा Y मूल्यों के अनुसार एक बिन्दु द्वारा अंकित करते हैं। इस प्रकार समग्र में जितनी इकाइयाँ दी होती हैं प्रकीर्ण आरेख में उतने ही बिन्दु अंकित हो जाते हैं।

प्रकीर्ण आरेख में बिन्दुओं की स्थिति देखकर X तथा Y मूल्यों के सह-सम्बन्ध की मात्रा व प्रकृति के बारे में अनुमान लगाया जा सकता है। यदि समस्त बिन्दु आरेख के निचले बायें कोने से ऊपरी दायें कोने की ओर एक सीधी रेखा में फैले हैं तो X तथा Y मूल्यों में पूर्ण धनात्मक सह-सम्बन्ध होगा (चित्र 24.8A)। यदि ये बिन्दु निचले दायें कोने से ऊपरी बायें कोने की ओर एक सीधी रेखा में पाये जायें तो X तथा Y मूल्यों में पूर्ण ऋणात्मक सह-सम्बन्ध होगा (चित्र 24.8B)। जब प्रकीर्ण आरेख में बिन्दु चारों ओर फैले होते हैं तथा उनके वितरण से उपर्युक्त किसी भी दिशा का बोध नहीं होता तो उन मूल्यों में सह-सम्बन्ध का अभाव माना जायेगा (चित्र 24.8C)।

यदि समस्त बिन्दु एक सीधी रेखा में न हों परन्तु उनकी प्रवृत्ति उपर्युक्त दोनों दिशाओं में से किसी भी एक दिशा में फैलने की पायी जाये तो मूल्यों में सीमित सह-सम्बन्ध होगा। आरेख के निचले बायें कोने से ऊपरी दायें कोने की ओर को बिन्दुओं के फैलने की प्रवृत्ति सीमित धनात्मक सह-सम्बन्ध को तथा निचले दायें कोने से ऊपरी बायें कोने की ओर को बिन्दुओं के फैलने



चित्र 24.8—सह-सम्बंध के प्रतिरूप।

की प्रवृत्ति सीमित ऋणात्मक सह-सम्बन्ध को प्रदर्शित करती है (चित्र 24.8 D व E)। आरेख में अंकित बिन्दु आपस में जितने अधिक निकट होंगे उतनी ही मूल्यों में सह-सम्बन्ध की मात्रा अधिक होगी।

प्रतीपगमन विश्लेषण
(Regression Analysis)

हम प्रायः यह अनुभव करते हैं कि जीवन के प्रत्येक क्षेत्र में कुछ बातों का आपस में गहरा सम्बन्ध होता है। वर्षा एवं मृदा-अपरदन (soil erosion), अनावृष्टि (drought) तथा अकाल, सिंचाई की सुविधा व फसलों का चयन, उर्वरकों का प्रयोग एवं खाद्य-उत्पादन, किसी वस्तु की माँग व आपूर्ति, आय तथा व्यय, स्वास्थ्य सेवाएँ एवं मृत्यु-दर तथा पूंजी व लाभ इसके कतिपय उदाहरण हैं। इस प्रकार के तथ्यों से सम्बन्धित चर-मूल्यों के पारस्परिक सम्बन्धों का विश्लेषण व विवेचन करने के लिये सांख्यिकी में दो प्रकार की विधियाँ—(i) सह-सम्बन्ध विश्लेषण (correlation analysis) तथा (ii) प्रतीपगमन विश्लेषण, अपनायी जाती हैं। इन दोनों विधियों से अलग-अलग प्रकार की जानकारी प्राप्त होती है। सह-सम्बन्ध से दो चरों के पारस्परिक सह-सम्बन्ध की मात्रा (degree) का पता चलता है तथा प्रतीपगमन इस सम्बन्ध की प्रकृति (nature) अर्थात् कारण-प्रभाव सम्बन्ध (cause-effect relationship) की व्याख्या करता है। यही इन विधियों का मुख्य अन्तर है। जैसा कि हम पहले पढ़ चुके हैं, सह-सम्बन्ध गुणांक, जिसका मान +1 से −1 तक हो सकता है, दो चर-मूल्यों में विद्यमान धनात्मक अथवा ऋणात्मक सह-सम्बन्ध की मात्रा को प्रकट करता है। परन्तु इस गुणांक से यह पता नहीं चलता कि एक चर के मूल्य में परिवर्तन होने पर दूसरे चर के मूल्य पर क्या प्रभाव पड़ेगा। इसके विपरीत प्रतीपगमन में एक चर को स्वतन्त्र (independent) तथा दूसरे को आश्रित (dependent) मानकर, स्वतन्त्र चर-मूल्य से आश्रित चर-मूल्य का प्रकार्यात्मक सम्बन्ध (functional relationship) ज्ञात करते हैं। यह प्रकार्यात्मक सम्बन्ध विभिन्न प्रकार का हो सकता है परन्तु इनमें रेखीय प्रकार का प्रकार्यात्मक सम्बन्ध सबसे सरल है। साधारण प्रतीपगमन (simple regression) में इसी रेखीय प्रकार पर विचार करते हैं। उपर्युक्त विवरण से स्पष्ट है कि दो चर-मूल्यों के पारस्परिक सम्बन्ध की मात्रा ज्ञात करने के लिये सह-सम्बन्ध विधि को प्रयोग में लाते हैं तथा एक चर-मूल्य के आधार पर दूसरे चर-मूल्य का पूर्वानुमान (forecast) लगाने में प्रतीपगमन तकनीक का प्रयोग किया जाता है।

प्रतीपगमन या समाश्रयण रेखाएँ
(Regression Lines)

प्रतीपगमन रेखाएँ वे सर्वोपयुक्त रेखाएँ (lines of the best fit) होती हैं जिनसे दो सम्बन्धित श्रेणियों के पारस्परिक माध्य सम्बन्ध का बोध होता है। दूसरे शब्दों में, ये रेखाएँ एक श्रेणी के माध्य मूल्यों के समकक्ष दूसरी श्रेणी के माध्य मूल्यों को स्पष्ट करती हैं। इस प्रकार प्रतीपगमन रेखाओं के द्वारा एक चर के माध्य मूल्य के आधार पर दूसरे चर का माध्य मूल्य ज्ञात किया जा सकता है। दो सम्बन्धित श्रेणियों के लिये दो प्रतीपगमन रेखाएँ खींची जाती हैं। एक रेखा के द्वारा Y का X पर प्रतीपगमन (regression of Y on X) तथा दूसरी के द्वारा X का Y पर प्रतीपगमन (regression of X on Y) प्रकट होता है। इन प्रतीगमन रेखाओं के समीकरण निम्नांकित हैं :

Y के X पर प्रतीपगमन का समीकरण :

$$Y-\overline{Y} = r\frac{\sigma_y}{\sigma_x}(X-\overline{X})$$

X के Y पर प्रतीपगमन का समीकरण :

$$X-\overline{X} = r\frac{\sigma_x}{\sigma_y}(Y-\overline{Y})$$

उपरोक्त समीकरणों में $\overline{X}$ तथा $\overline{Y}$ क्रमशः X तथा Y मूल्यों के समान्तर माध्यों को प्रकट करते हैं, r सह-सम्बन्ध गुणांक को,

σ_x X श्रेणी के मानक विचलन को σ_y Y श्रेणी के मानक विचलन को प्रकट करता है। पहले समीकरण से X के दिये गये मूल्यों के आधार पर Y के मूल्य अनुमानित किये जाते हैं तथा दूसरे समीकरण से Y के दिये गये मूल्यों के आधार पर X के मूल्य अनुमानित किये जाते हैं।

उदाहरण (70) निम्नांकित आँकड़ों से प्रतीपगमन रेखाएँ ज्ञात कीजिये तथा उन्हें ग्राफ पेपर पर प्रदर्शित कीजिये :

X मूल्य	4	5	6	7	8
Y मूल्य	8	10	12	12	13

हल—उपरोक्त समीकरणों में लिखे संकेताक्षरों के मान ज्ञात करने के लिये निम्नलिखित क्रिया की जायेगी :

सारणी 24.72

X	$\overline{X}$ = 6 से विचलन (X–$\overline{X}$) या dx	विचलन मूल्य का वर्ग (d_x^2)	Y	$\overline{Y}$ = 11 से विचलन (Y–$\overline{Y}$) या dy	विचलन मूल्य का वर्ग (d_y^2)	विचलन मूल्यों का गुणनफल (d_xd_y)
4	−2	4	8	−3	9	+6
5	−1	1	10	−1	1	+1
6	0	0	12	+1	1	0
7	+1	1	12	+1	1	+1
8	+2	4	13	+2	4	+4
N = 5		10	N = 5		16	12
ΣX = 30		Σd_x^2	ΣY = 55		Σd_y^2	Σd_xd_y

X मूल्यों का समान्तर माध्य या $\overline{X}$

$$= \frac{\Sigma X}{N} = \frac{30}{5} = 6$$

Y मूल्यों का समान्तर माध्य या $\overline{Y}$

$$= \frac{\Sigma Y}{N} = \frac{55}{5} = 11$$

X मूल्यों का मानक विचलन

$$\sigma_x = \sqrt{\frac{\Sigma d_x^2}{N}} = \sqrt{\frac{10}{5}}$$

$$= \sqrt{2} = 1.41$$

Y मूल्यों का मानक विचलन

$$\sigma_y = \sqrt{\frac{\Sigma d_y^2}{N}} = \sqrt{\frac{16}{5}}$$

$$= 1.78$$

X तथा Y का सह-सम्बन्ध गुणांक

$$r = \frac{\Sigma d_xd_y}{N\sigma_x\sigma_y} = \frac{\Sigma d_xd_y}{\sqrt{\Sigma d_x^2 \times \Sigma d_y^2}}$$

$$= \frac{12}{\sqrt{10 \times 16}} = \frac{12}{\sqrt{160}}$$

$$= +0.948$$

X के Y पर प्रतीपगमन का समीकरण—

$$X - \overline{X} = r\frac{\sigma_x}{\sigma_y}(Y - \overline{Y})$$

मान रखने पर—

$$X - 6 = .948 \times \frac{1.41}{1.78}(Y - 11)$$

$$= .75\ (Y - 11) = .75Y - 8.25$$

$$X = .75Y - 8.25 + 6$$

$$X = .75Y - 2.25$$

अब इस समीकरण में दिये हुए Y मूल्यों को रखकर उनके सम्बन्धित X मूल्य निम्न प्रकार ज्ञात किये जायेंगे :

$$X = .75\ Y - 2.25$$

X के अनुमानित मूल्य, यदि

Y = 8, तो $X = .75 \times 8 - 2.25 = 3.75$

Y = 10, तो $X = .75 \times 10 - 2.25 = 5.25$

Y = 12, तो $X = .75 \times 12 - 2.25 = 6.75$

Y = 13, तो $X = .75 \times 13 - 2.25 = 7.5$

Y के X पर प्रतीपगमन का समीकरण—

$$Y - \overline{Y} = r\frac{\sigma_y}{\sigma_x}(X - \overline{X})$$

$$Y - 11 = \frac{.948 \times 1.78}{1.41}(X - 6)$$

$$=1.19\ (X-6)\ =1.19\ X-7.14$$

$$Y=1.19\ X-7.14+11$$

$$Y=1.19\ X+3.86$$

उपरोक्त समीकरण में दिये हुए X मूल्यों को रखकर उनके सम्बन्धित Y मूल्य निम्न प्रकार ज्ञात किये जायेंगे :

$$Y=1.19X+3.86$$

Y के अनुमानित मूल्य, यदि

X=4, तो $Y=1.19\times4+3.86=\ 8.62$

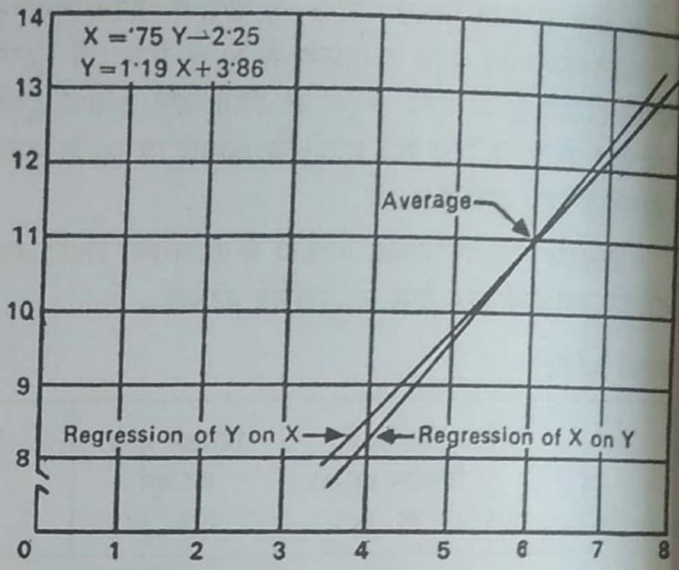
X=5, तो $Y=1.19\times5+3.86=\ 9.81$

X=6, तो $Y=1.19\times6+3.86=11.00$

X=7, तो $Y=1.19\times7+3.86=12.19$

X=8, तो $Y=1.19\times8+3.86=13.38$

ऊपर दिये हुए Y के मूल्यों (8, 10, 12, 13) के अनुसार X के अनुमानित मूल्यों (3.75, 5.25. 6.75 तथा 7.5) से ग्राफ पेपर पर X की Y पर प्रतीपगमन रेखा तथा X के दिये हुए मूल्यों (4, 5, 6, 7, 8) के अनुसार Y के अनुमानित मूल्यों (8.62, 9.81, 11.00, 12.19 तथा 13.38) से ग्राफ पेपर पर Y की X पर दूसरी प्रतीपगमन रेखा खींची जा सकती है (चित्र 24.9)।



चित्र 24.9—प्रतीपगमन रेखाएँ।