निश्चयता गुणांक (Coefficient of Determination)

भौगोलिक अध्ययनों में तत्वों की पारस्परिक सम्बद्धता के अध्ययन का विशेष महत्व है। इसमें चरों के मध्य परस्पर निर्भरता का अध्ययन किया जाता है। ऐसे अध्ययनों में सामान्यत: स्वतन्त्र चर (X) का आश्रित चर (Y) पर प्रभाव का आगणन किया जाता है। सहसम्बन्ध गुणक की गणना से ही तत्वों की परस्पर निर्भरता पूर्णत: स्पष्ट नहीं हो पाती है, इसलिये Coefficient of Determination की गणना अधिक उपादेय होती है, जो X चर पर Y चर की आश्रितता को स्पष्ट रूप से व्यक्त करता है। सहसम्बन्ध सम्बन्धी सभी अध्ययनों में शोधकर्ता को निश्चयता गुणक की गणना अवश्य करनी चाहिये। इससे X चर पर Y चर का क्या प्रभाव है, यह स्पष्ट रूप से जाना जा सकता है, क्योंकि यथार्थ धरातल पर Y चर पर केवल X चर का एकल प्रभाव नहीं पड़ता है। उस घटना के लिये कई तत्व उत्तरदायी होते हैं। इसलिये यह जानना आवश्यक है कि अध्ययन हेतु चयनित X चर का, Y चर पर क्या प्रभाव है और निश्चयता गुणक से बहुत आसानी से यह तथ्य विश्लेषित किया जा सकता है। इसकी गणना बहुत आसान है। कार्ल पियर्सन द्वारा प्रस्तुत सहसम्बन्ध गुणक विधि से आगणित (r) के आधार पर इसकी गणना की जाती है।

यथा

$$C.D. = \gamma^2$$

उदाहरणार्थ-यदि किसी सहसम्बन्ध अध्ययन में गुणक 0.7 आता है, जो उच्च सहसम्बन्ध को प्रदर्शित करता है लेकिन स्वतंन्त्र चर के रूप X Variable का आश्रित चर Y Variable पर पड़ने वाले प्रभाव को वह उतना सही रूप में अभिव्यक्त नहीं करता है जितना की निश्चयता गुणांक। इस 0.72 सहसम्बन्ध गुणक के आधार पर परिगणित निश्चयता गुणक 0.51 होगा जो यह व्यक्त करता है कि आश्रित चर (Y) से स्वतन्त्र चर (X) , 51 प्रतिशत अंक तक प्रभावित अथवा व्याख्यायित है। शेष 49% प्रभाव हेतु अन्य चरों का निरीक्षण व उनसे सहसम्बन्ध का मापन आवश्यक हो जाता है। Y चर की व्याख्या में X चर का योगदान 51 प्रतिशत है, अर्थात रबी फसलों के उत्पादन (Y) की व्याख्या में सिंचाई X का योगदान 51 प्रतिशत है, शेष 49 प्रतिशत व्याख्या उत्पादन को प्रभावित करने वाले अन्य तत्वों के अनुशीलन पर आधारित होगी। इसका तात्पर्य यह है कि उच्च सहसम्बन्ध पाये जाने के बावजूद Y चर पर X चर का प्रभाव मात्र 51% है।

सहसम्बन्ध गुणांक की सार्थकता परीक्षण-

भौगोलिक अध्ययनों एवं शोध में विभिन्न तत्वों की परस्पर सम्बद्धता एवं निर्भरता का अध्ययन प्रमुखता के साथ किया जाता हैं। कार्य-कारण सम्बन्ध विश्लेषण में एक तत्व का दूसरे तत्व पर प्रभाव का अध्ययन अपेक्षित होता है, लेकिन सहसम्बन्ध गुणक की गणना के पश्चात यह जानना आवश्यक है कि दोनो चरों में पाये जाने वाला सहसम्बन्ध सार्थक है अथवा नहीं। यह आवश्यक नहीं है कि सहसम्बन्ध गुणक की मात्रा अधिक होने पर सहसम्बन्ध सार्थक ही

हो। यह प्रतिदर्शी की संख्या और उसके आकार पर भी आश्रित होता है। फिशर (Fisher) महोदय ने दो प्रतिदर्शी चरो के बीच सहसम्बन्ध गुणांक की सार्थकता का परीक्षण करने हेतु टी (t) अनुपात का निम्नवत प्रयोग किया है-

$$t = r\sqrt{\frac{n-2}{1-r^2}}$$

यहाँ n= प्रतिदर्शों की संख्या एवं r सहसम्बन्ध गुणांक को व्यक्त करते है।

इस सन्दर्भ में d.f. की गणना निम्न सूत्र से की जाती है- d.f. = n− 2 दिये गये प्रश्न में प्रतिदर्शों की संख्या n=8 है। अत: d.f.= 8 होगा और 8 d.f. पर .05 एवं .01 विश्वास स्तर पर सारणीमान क्रमश: 0.632 एवं 0.765 है। चूँकि परिगणित r का मान 0.765 से भी अधिक है अत: दोनों प्रतिदर्श (variables) चरों X और Y में सहसम्बन्ध 0.01 विश्वास स्तर पर सार्थक है। दूसरे शब्दों में कहा जा सकता है कि इन दोनो चरों के सहसम्बन्ध में 99 प्रतिशत सार्थकता विद्यमान है।

समाश्रयण विश्लेषण (Regression Analysis)

समाश्रयण विश्लेषण, सांख्यिकी विज्ञान की एक उपयोगी विधि है जिसमें द्विचरीय विश्लेषण (Bi-Variant Analysis) में प्रयुक्त तत्वों की परस्पर निर्भरता की प्रकृति एवं मात्रा, क्षेत्रीय इकाईयों के आधार पर अध्ययन एवं विश्लेषण किया जा सकता है। जैसा कि नाम से स्पष्ट है, इसमें सहसम्बन्ध अध्ययन हेतु प्रयुक्त दोनों तत्वों के परस्पर आश्रयण (निर्भरता) का ज्ञान प्राप्त किया जाता है। सहसम्बन्ध अध्ययन में स्वतन्त्र चर (X) एवं आश्रित चर (Y) को क्रमश: प्रभावित करने वाले एवं प्रभाव ग्रहण करने वाले तत्व के रूप में जाना जाता है, जबकि समाश्रयण में X एवं Y तत्वों को समान प्रभाव वाला व्यवहृत करते हुये दोनों की परस्पर निर्भरता का अध्ययन किया जाता है। समाश्रयण एवं सहसम्बन्ध विश्लेषण में कतिपय विभिन्नताओं के होते हुये भी इनमें परस्पर गहन सम्बन्ध विद्यमान है। समाश्रयण विश्लेषण में अध्ययनित क्षेत्र के प्रत्येक इकाई में दोनो तत्वों की परस्पर निर्भरता का ज्ञान प्राप्त करने के साथ इसका मानचित्रण भी किया जा सकता है, जबकि सहसम्बन्ध अध्ययन में प्रयुक्त चरों के बीच सहसम्बन्ध की दिशा एवं उसकी एकल मात्रा का ज्ञान प्राप्त होता है। इसलिये भौगोलिक अध्ययनों में, जहाँ स्थानिक प्रतिरूपों का विशेष महत्व है, समाश्रयण विश्लेषण अधिक उपयोगी एवं महत्वपूर्ण है। इसमें अध्ययनित क्षेत्र के अन्तर्गत प्रत्येक इकाई में दोनों चरों की परस्पर निर्भरता की गणना एवं उसके आधार पर इसका मानचित्रण दिया जाता है। इस प्रकार समाश्रयण विश्लेषण के द्वारा एक ज्ञात चर (X या Y) में परिवर्तन के आधार पर दूसरे चर (Y या X) में परिवर्तन की मात्रा को ज्ञात किया जाता है, जिसके द्वारा दूसरे चर के विषय में भविष्येक्षण भी किया जा सकता है। दो

चरों के समाश्रयण की प्रवृत्ति को ग्राफ पर प्रकीर्णन आरेख के माध्यम से रेखीय रूप में प्रदर्शित किया जाता है, जिसे समाश्रयण रेखा (Regression Line) अथवा Line of the Best Fit भी कहा जाता है।

उदाहरण

तालिका : 6.9

क्षेत्र	सिंचित क्षेत्र हजार हे० में(X)	द्विफसली क्षेत्र हजार हे० में(Y)	XY	X^2	Y^2	Y'	X'	Y–Y'
A	3	5	15	9	25	13.52	15.0	–8.52
B	5	17	85	25	289	15.0	23.88	2
C	8	15	120	64	225	17.22	22.4	–2.22
D	12	24	228	144	576	20.18	29.06	3.82
E	6	20	120	36	400	15.74	26.1	4.26
F	5	22	110	25	484	15.0	27.58	7.0
G	3	12	36	9	144	13.52	20.18	–1.52
H	2	10	20	4	100	12.78	18.7	–2.78
I	4	15	60	16	225	14.26	22.4	0.74
J	2	10	20	4	100	12.78	18.7	–2.78
N=10	$\bar{x} = 50$	$y = 150$	$\Sigma xy =$	$\Sigma x^2 =$	$\Sigma y^2 =$			
	$\bar{x} = 5$	$\bar{y} = 15$	814	336	2568			

समाश्रयण समीकरण

$$y' = a+bx$$

$$b = \frac{\Sigma xy - N\overline{X}\ \overline{Y}}{X^2 - N(\overline{X})^2}$$

$$= \frac{814 - 10\times5\times15}{336 - 10\times5\times5}$$

$$= \frac{814 - 750}{336 - 250}$$

$$= \frac{64}{86} = 0.74$$

$$a = \bar{y} - b\bar{x}$$
$$= 15 - 0.74 \times 5$$
$$= 15 - 3.7$$
$$= 11.3$$

इस प्रकार उक्त उदाहरण हेतु समाश्रयण समीकरण होगा-

$$y' = a + bx$$
$$= 11.3 + 0.74x$$

विभिन्न क्षेत्रों के लिये y का ऑकलित मान

$$y' = a + bx$$

A= 11.3 + .74 × 3 = 13.52

B= 11.3 + .74 × 5 = 15.0

C= 11.3 + .74 × 8= 17.22

D= 11.3 + .74 ×12 = 20.18

E= 11.3 + .74 × 6 = 15.74

F = 11.3 + .74 × 5 = 15.0

G = 11.3 + .74 × 3 = 13.52

H = 11.3 + .74 × 2= 12.78

I = 11.3 + .74 × 4 = 14.26

J = 11.3 + .74 × 2 = 12.78

इसी प्रकार विभिन्न क्षेत्रों हेतु X का ऑकलित मान (X') भी ज्ञात किया जा सकता है जो निम्नवत है।

$$X' = a + by$$

A= 11.3 + .74 × 5 = 15.0.

B= 11.3 + .74 × 17= 23.88

C= 11.3 + .74 × 15= 22.4

D= 11.3 + .74 ×24 = 29.06

E= 11.3 + .74 × 20 = 26.1

F = 11.3 + .74 × 22 = 27.58

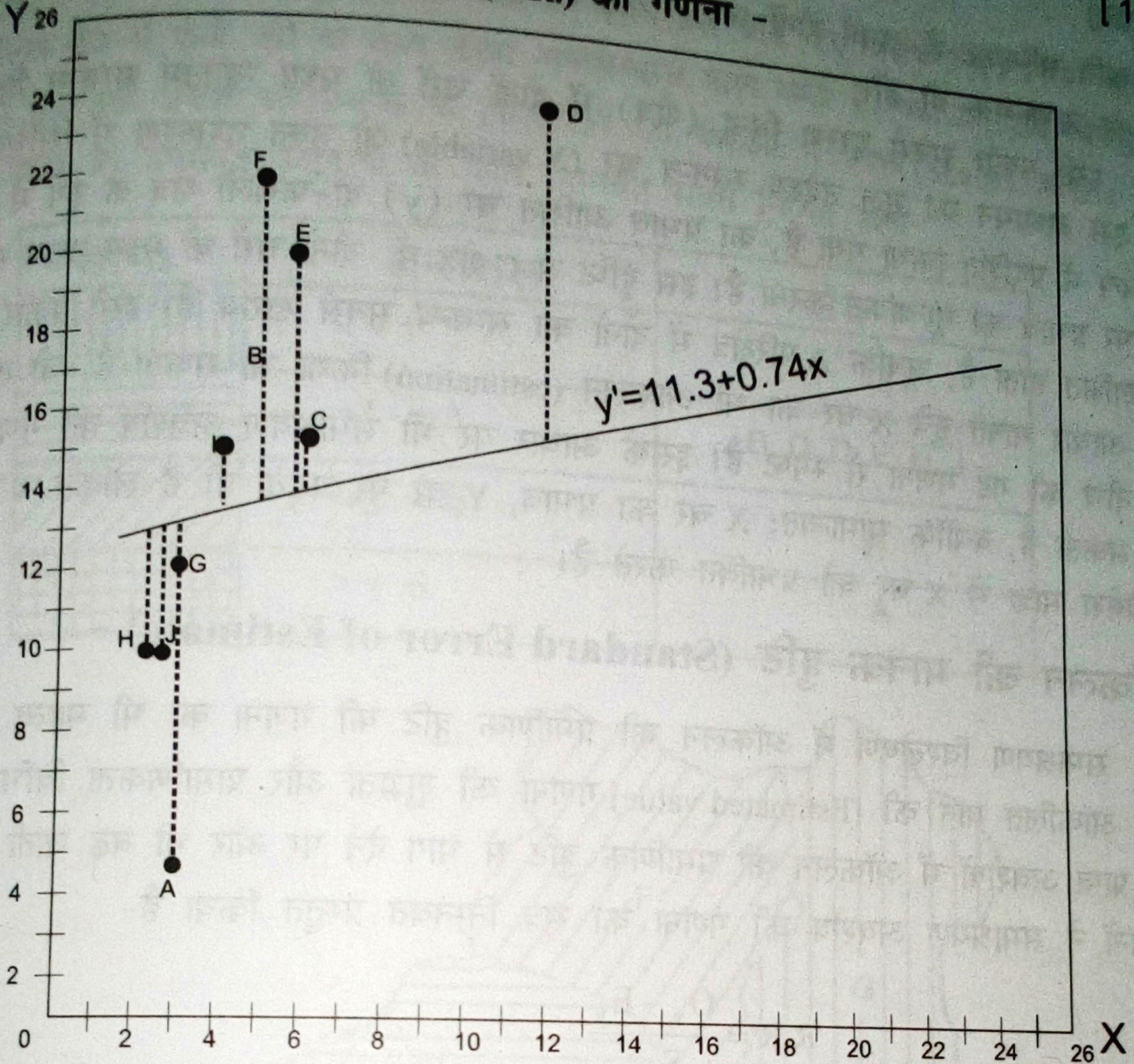
G = 11.3 + .74 × 12 = 20.18

H = 11.3 + .74 × 10= 18.7

I = 11.3 + .74 × 15 = 22.4

J = 11.3 + .74 × 10 = 18.7

Residuals from Regression (R.R.) की गणना -



$$R.R = \frac{Y_{cn} - Y_n}{Sy_e} = \frac{O_y - E_y}{Sy_e}$$

y' = a+bx	y' = a+ by
= 11.3 +0 .74×2	= 11.3×0.74×10
= 12.78	= 18.70

उक्त तालिका में समाश्रयण विश्लेषण का अध्ययन प्रस्तुत है जिसमें दिये गये क्षेत्र (X) में सिंचित क्षेत्र एवं द्वि–फसली क्षेत्र (Y) के मध्य परस्पर निर्भरता का अध्ययन किया गया है। समाश्रयण विश्लेषण में उक्त दोनों तत्वों की परस्पर निर्भरता/ अन्तर्सम्बन्ध को अभिव्यक्त करने के लिये समाश्रयण अवशेष की गणना की गई है, जिसमें A,C, G, H और J क्षेत्रों में ऋणात्मक अवशेष प्राप्त हुये है, जबकि B, D, E, F एवं J क्षेत्रों में धनात्मक अवशेष प्राप्त है। धनात्मक अवशेषों को over estimated और ऋणात्मक अवशेषों को under estimated के रूप में जाना जाता है। यह तथ्य इसके आरेखीय प्रदर्शन में देखा जा सकता है। over estimated बिन्दु आरेख में Line of the Best Fit (y = a+bx) के ऊपर अवस्थित है जबकि under estimated बिन्दु इस रेखा के नीचे अवस्थित है। क्षेत्र का प्रतिनिधित्व करते हुये जो बिन्दु Line of Best Fit या समाश्रयण रेखा

के अति सन्निकट हैं, उनमें दोनों चरों में सर्वोत्तम अन्तर्सम्बन्ध विद्यमान है, चाहे वह धनात्मक हो या ऋणात्मक ही हों।

इसी प्रकार सबसे दूरस्थ बिन्दु (क्षेत्र) में दोनों चरों के मध्य न्यूनतम सम्बन्ध विद्यमान है। इस अध्ययन का मूल उद्देश्य स्वतन्त्र चर (X variable) जो उक्त उदाहरण में सिंचित क्षेत्र के रुप में प्रदर्शित किया गया है, का प्रभाव आश्रित चर (y) दो-फसली क्षेत्र के रुप में प्रस्तुत है, पर प्रभाव का मूल्यांकन करना है। इस दृष्टि से C क्षेत्र में दोनों चरों के मध्य उत्तम सम्बन्ध परिलक्षित होता है, जबकि A परिक्षेत्र में दोनो का सम्बन्ध सबसे खराब है। इसी प्रकार y चर को आधार मानते हुये X चर का भी आकलन (estimation) किया जा सकता है, जो तालिका के नीचे की गई गणना से स्पष्ट है। इसके आधार पर भी समाश्रयण अवशेष की गणना की जा सकती है, क्योंकि सामान्यतः X चर का प्रभाव, Y चर पर पड़ता ही है लेकिन X चर भी कमोवेश मात्रा में X चर को प्रभावित करते है।

ऑकलन की मानक त्रुटि (Standard Error of Estimate) -

समाश्रयण विश्लेषण में ऑकलन की प्रमाणिक त्रुटि की गणना का भी महत्व है। इस ज्ञात आंकलित मान की (Estimated value) गणना की शुद्धता और प्रामाणिकता विभिन्न क्षेत्रों हेतु प्राप्त अवशेषों में ऑकलन की प्रमाणिक त्रुटि से भाग देने पर और भी बढ़ जाती है, कई विद्वानों ने समाश्रयण अवशेष की गणना का सूत्र निम्नवत प्रस्तुत किया है-

$$R.R = \frac{O_y - E_y}{S_{EE}}$$

ऑकलन की मानक त्रुटि की गणना निम्नवत सूत्र से की जाती है-

$$SEE = \sqrt{\frac{\Sigma(y-\hat{y})^2}{N-1}}$$

$$= \sqrt{\frac{318}{9}}$$

$$= \sqrt{3.53}$$

$$= 1.83$$

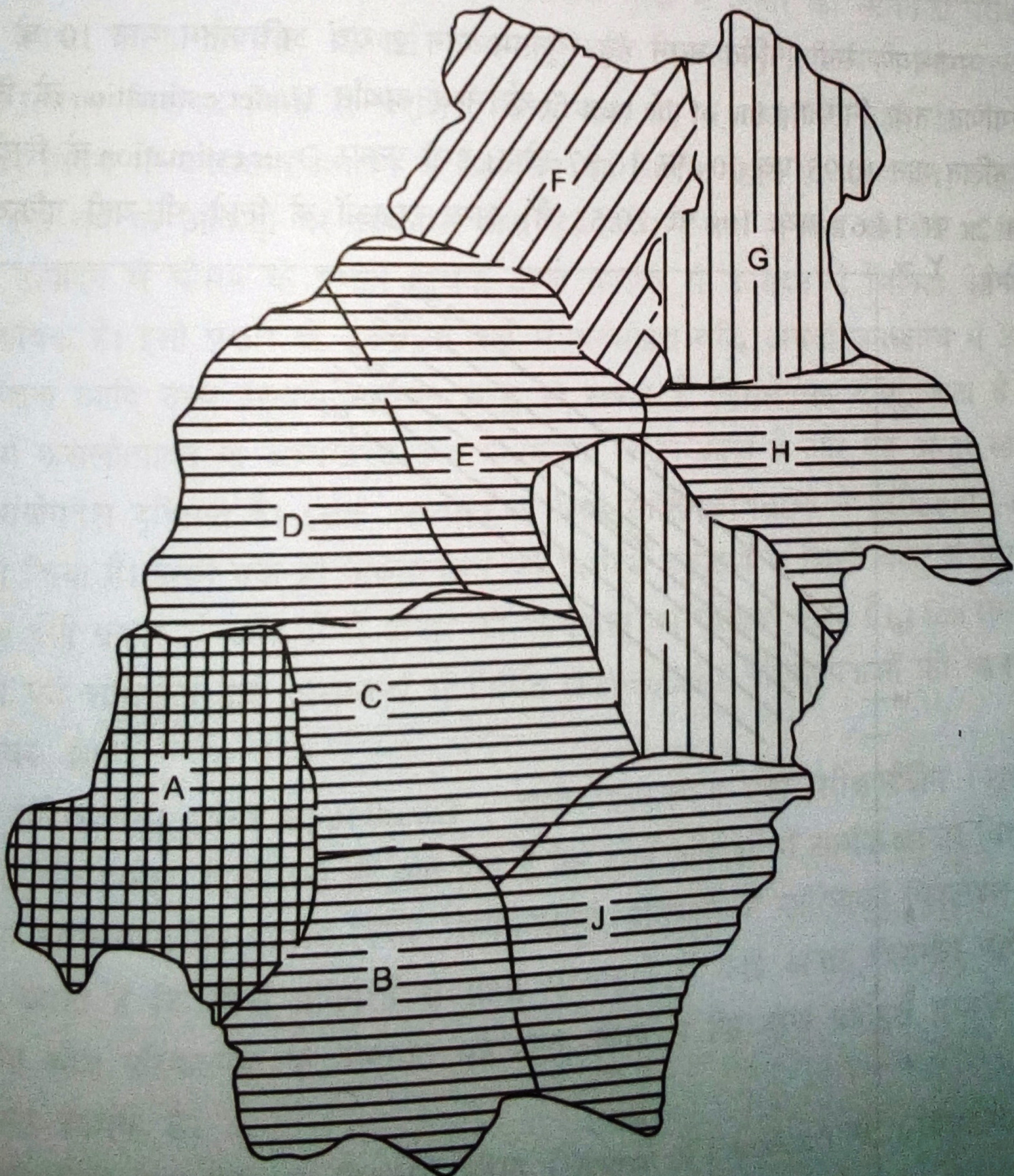
इस प्रकार विभिन्न क्षेत्रों में समाश्रयण अवशेष प्रभावी रूप से निम्नवत प्राप्त होगा।

क्षेत्र	A	B	C	D	E	F	G	H	I	J
समाश्रण अवशेष	−4.69	1.09	-1.21	2.8	2.25	3.70	-0.83	-1.51	0.40	-1.51

उक्त गणना के आधार पर इसका मानचित्रीय प्रदर्शन भी किया जा सकता है। मानचित्रीय प्रदर्शन हेतु उपरोक्त गणना से प्राप्त समाश्रयण अवशेष का संवर्ग बनाकर उसे आसानी से

तत्सम्बन्धी मानचित्र पर प्रस्तुत किया जाता है। इसमें इस बात का विशेष ध्यान दिया जाता है कि जिस क्षेत्र में दोनो चरों के मध्य उत्तम अन्तसम्बन्ध पाया जाता है, वहाँ का शेड अपेक्षाकृत हल्का एवं इसी प्रकार सहसम्बन्ध के निम्न स्तर की दशाओं में आभा (Shade) क्रमशः गाढ़ा होता जाता है। उपरोक्त उदाहरण के लिये निम्नवत संवर्ग निर्मित होगा-

आभा	अन्तराल	क्षेत्र
(ऊर्ध्वाधर रेखाएँ)	< + 1	G I
(क्षैतिज रेखाएँ)	+1 - + 3	B, D, D, E, H, J
(तिरछी रेखाएँ)	+3 - + 6	F
(जालीदार)	> + 6	A



समाश्रयण विश्लेषण का आरेखीय प्रदर्शन-

चूँकि समाश्रयण रेखा का न्यूनतम एवं अधिकतम मान 2x एवं 10x पर क्रमशः 12.78 एवं 18.70 परिगणित है, अतः Under Estimation तथा Estimation की गणना निम्न रूप में की जा सकती है-

	Under Estimation		Over Estimation	
1 SEE =	12.78 – 1.83 18.70 – 1.83	10.95 16.87	12.78 + 1.83 18.70 + 1.83	14.44 20.53
2 SEE =	12.78 – 3.66 18.70 – 3.66	9.12 15.04	12.78 + 3.66 18.70 + 3.66	16.44 22.36
3 SEE =	12.78 – 5.49 18.70 – 5.49	7.29 13.21	12.78 + 5.49 18.70 + 5.49	18.27 24.19

समाश्रयण रेखा निर्मित करने हेतु न्यूनतम मान 2 एवं अधिकतम मान 10 के आधार पर परिगणित तथा निर्मित Line of the best fit के नीचे अर्थात Under estimation के लिये 2x पर आकलित मान 10.95 एवं 10x पर 16.87 होगा। इसी प्रकार Over estimation के लिये आकलित मान 2x पर 14.61 तथा 10x पर 20.53 है। अन्य रेखाओं के लिये भी यही प्रक्रिया अपनायी जायेगी।