इसलिए (9.3a) से, हम पाते हैं

$$\frac{\partial E}{\partial a} = 0 \quad \text{और} \quad \frac{\partial E}{\partial b} = 0 \quad \text{---} (9.4)$$

$$\Rightarrow \quad \Sigma y = na + b\Sigma x \quad \text{------} (9.5)$$

$$\text{और} \quad \Sigma xy = a\Sigma x + b\Sigma x^2 \quad \text{------} (9.6)$$

ये समीकरण a और b को आकलित करने के लिए प्रसामान्य समीकरण कहे जाते हैं। परिमाण Σx, Σx^2, Σy, Σxy प्राप्त किए जा सकते हैं n बिन्दुओं $(x_1, y_1), (x_2, y_2) \ldots (x_n, y_n)$ से और इसलिए समीकरणों (9.5) और (9.6) को एक साथ हल कर सकते हैं a और b के लिए पाने के लिए :

$$a = \frac{(\Sigma x^2)(\Sigma y) - (\Sigma x)(\Sigma xy)}{n\Sigma x^2 - (\Sigma x)^2} \quad \text{----} (9.7)$$

$$\text{और} \quad b = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{n\Sigma x^2 - (\Sigma x)^2} \quad \text{----} (9.8)$$

a और b के इन मूल्यों को प्रतिस्थापित करने पर (9.7) और (9.8) से (9.1) में, ~~y का x पर वांछित समीकरण~~, प्रतीपगमन की रेखा का वांछित समीकरण y का x पर पाते हैं।

प्रतीपगमन की रेखा का ~~की~~ समीकरण y का x पर प्राप्त किया जा सकता है बहुत व्यवस्थित और आकलन रूप में x̄, ȳ, σx, σy के क्रम में और $r = r_{xy}$ जैसा नीचे वर्णित है

(9.5) के दोनों तरफ n से भाग देने पर, कुल जोड़ों की संख्या, हम पाते हैं

$$\frac{1}{n}\Sigma y = a + b \cdot \frac{1}{n}\Sigma x \Rightarrow \bar{y} = a + b\bar{x} \quad \text{------} (9.9)$$

इसका अर्थ है कि ~~a~~ उपयुक्तता की रेखा i.e. y का x पर प्रतीपगमन बिन्दु $(\bar{x}, \bar{y})$ से गुजरती है। या दूसरे शब्दों में, बिन्दु $(\bar{x}, \bar{y})$ y का x पर प्रतीपगमन की रेखा है।

$$(9.8) \text{ से, हम पाते हैं :} \quad b = \frac{cov.(x, y)}{\sigma_x^2} \quad \text{------} (9.10)$$

हम पाते हैं कि समीकरण ~~सं~~ (9.1) ढलान - अंतः खंड रूप (slope-intercept) में है i.e., $y = mx + c$। इसलिए b प्रतीपगमन की रेखा का ढलान y का x पर प्रदर्शित करता है। आगे, ऊपर (9.9) में सिद्ध किया गया है कि यह रेखा (i.e. y का x पर प्रतीपगमन की रेखा) बिन्दु $(\bar{x}, \bar{y})$ से गुजरती है। इसलिए, ढलान - बिन्दु का उपयोग कर एक बिन्दु का समीकरण का रूप, y का x पर प्रतीपगमन की रेखा का वांछित समीकरण हो जाता है :

$$y - \bar{y} = b(x - \bar{x}) \quad \text{------} (9.11)$$

$$\text{या,} \quad y - \bar{y} = \frac{cov.(x, y)}{\sigma_x^2} \cdot (x - \bar{x}) \quad \text{------} (9.12)$$

नियम 7.1

परन्तु $r = r_{xy} = \frac{Cov.(x,y)}{\sigma_x \cdot \sigma_y} \Rightarrow Cov.(x,y) = r\,\sigma_x \cdot \sigma_y$

$\therefore \quad y - \bar{y} = r \cdot \frac{\sigma_y}{\sigma_x}(x - \bar{x})$ ---

टिप्पणियाँ. 1. (9.4) से हमें :

$\Sigma(y - a - bx) = 0 \quad \Rightarrow \quad \Sigma(y - y_e) = 0$

जहाँ y_e. y का प्राक्कलित मूल्य है x के दिये मूल्य के लिए, x पर प्रतीपगमन रेखा (9.1) से ली गई है।

2. y का x पर प्रतीपगमन की रेखा बिन्दु $(\bar{x}, \bar{y})$ से गुजरती है।

3. रैखिक और गैर-रैखिक प्रतीपगमन [उपनति मान (trend)] को विस्तार से अध्याय - 11 में वर्णित की गई है — काल श्रेणी उपनति मानों के निर्धारण के लिए।

9.3.2 x का y पर प्रतीपगमन की रेखा :-

x का y पर प्रतीपगमन की रेखा ऐसी रेखा है जो x का प्राक्कलन करती है y के दिये गये मूल्य के लिए। यह न्यूनतम वर्गों से प्राप्त की जाती है x-अक्ष के समानान्तर त्रुटियों के वर्गों के योग न्यूनतम करके (देखें चित्र 9.2)। शुरू करके समीकरण के साथ को

$x = A + By$ ---------- (9.15)

और x के प्राक्कलनों की त्रुटियों के वर्गों के योगफल को न्यूनतम करते x के दिये गये मूल्यों के बीच के विचलनों और उनके प्राक्कलनों को जो प्रतीप की रेखा द्वारा x का y पर दिया जाता है विद. (9.15), i.e. न्यूनतम करने पर

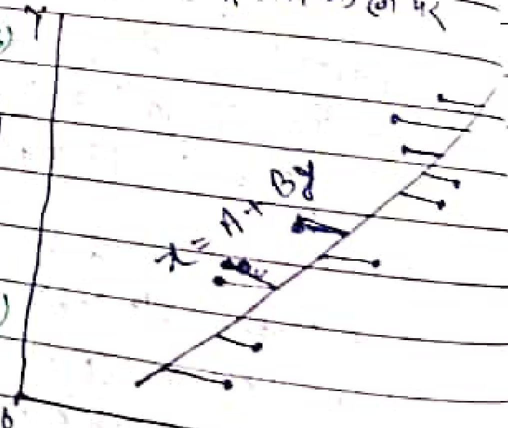
$E = \Sigma(x - A - By)^2$ ---- (9.16)

हमलोग A और B को प्राक्कलित करने के लिए प्रसामान्य समीकरण पायेंगे जैसे :-

$\Sigma x = nA + B\Sigma y$ और

$\Sigma xy = A\Sigma y + B\Sigma y^2$ ---- (9.17)

(9.17) को A और B के लिए एक साथ हल करने पर, हम पायेंगे



$A = \frac{(\Sigma y^2)(\Sigma x) - (\Sigma y)(\Sigma xy)}{n\Sigma y^2 - (\Sigma y)^2}$ ---- (9.18)

चित्र - 9.2.

और $B = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{n\Sigma y^2 - (\Sigma y)^2}$ ---- (9.19)

A और B के इन मूल्यों को (9.15) में प्रतिस्थापित करने पर, हम वांछित समीकरण पायेंगे प्रतीपगमन की रेखा का x का y पर।

इकाई: प्रतीपगमन, विश्लेषण (Linear Regression Analysis)

1.1 प्रस्तावना :-

'प्रतीपगमन' शब्द का शाब्दिक या शब्दकोषीय अर्थ है 'पीछे लौटना या औसत मान की ओर लौटना'। इस शब्द का प्रथम बार प्रयोग एक ब्रिटिश जीवविज्ञानिक सर फ्रांसिस गाल्टन ने 19वीं शताब्दी के उत्तरार्द्ध में कुछ मापनाओं के सम्बन्ध में किया जिसमें उन्होंने आकलन की रेखा पर किया [illegible] उनके अनुसार लम्बे माता-पिताओं की संतान प्रत्यावर्तित या पीछे प्रतीपगमन कर जनसंख्या की औसत ऊँचाई की ओर लौट जाती है। उन्होंने लगभग [illegible] जोड़ा पिताओं और पुत्रों के ऊँचाईयों के बीच सम्बन्धों का अध्ययन किया और इसके प्राप्त निष्कर्षों को अपने शोध लेख में प्रकाशित किया "पैतृक ऊँचाई में मध्यमता की ओर प्रतीपगमन (Regression towards Mediocrity in Hereditary Stature)"। उनके अध्ययन की रोचक विशेषताएँ थीं:

(i) लम्बे पिताओं के लम्बे पुत्र होते हैं और छोटे पिताओं के छोटे पुत्र

(ii) पुत्रों की औसत ऊँचाई लम्बे पिताओं के समूह के पिताओं से कम होती है और पुत्रों की औसत ऊँचाई छोटे कद के पिताओं के समूह के पिताओं से अधिक होती है।

दूसरे शब्दों में, गाल्टन के अध्ययन ने प्रकट किया कि असामान्य रूप से लम्बे या छोटे कद के माताओं पिताओं की संतान प्रवृत्त होती हैं प्रत्यावर्तित या पीछे लौटकर जनसंख्या के औसत ऊँचाई की ओर, एक घटना जिसे गाल्टन ने वर्णित किया मध्यमता की ओर प्रतीपगमन (Regression to Mediocrity)।

उन्होंने निष्कर्ष निकाला कि अगर किसी समूह के पिताओं की औसत ऊँचाई 'a' सेंटीमीटर ऊपर (नीचे) है सामान्य औसत ऊँचाई से तब उनके पुत्रों की औसत ऊँचाई $(a \times r)$ सेंटीमीटर ऊपर (नीचे) होगी सामान्य औसत ऊँचाई से जहाँ r सह-सम्बंध गुणांक है दिये पिताओं के समूह की ऊँचाईयों और उनके पुत्रों की ऊँचाईयों के बीच। इस स्थिति में सह-सम्बंध धनात्मक है और चूँकि $|r| \leq 1$ हमें $a \times r \leq a$। यह उपर्युक्त (ii) के निष्कर्ष का समर्थन करता है।

परन्तु आजकल प्रतीपगमन शब्द जैसा सांख्यिकी में प्रयुक्त होती है का बहुत विस्तृत दृष्टिकोण है बगैर जैविकी के संदर्भ को प्रतीपगमन विश्लेषण का, सामान्य अर्थ में, अर्थ है आकलन या पूर्वानुमान एक चर के अज्ञात मूल्य का दूसरे चर के ज्ञात मूल्य से। यह एक बहुत महत्वपूर्ण सांख्यिकीय साधन है जो लगभग सभी विज्ञानों में विस्तृत रूप से प्रयुक्त होती है - प्राकृतिक, सामाजिक और भौतिकी। यह विशेष रूप से व्यावसायिक और अर्थशास्त्र के अध्ययनों में

दो या अधिक चरों के संबंधों के अध्ययन में प्रयुक्त होती है जो जिस से सम्बन्धित हैं: माँग और पूर्ति वक्रों के आकलन में, लागत, उत्पादन और खपत फलनों इत्यादि।

पूर्वानुमान या आकलन एक बहुत बड़ी समस्या है मानवीय क्रियाकलाप के लगभग सभी क्षेत्रों में। आकलन या पूर्वानुमान के लिए अधिकतम उत्पादन का, खपत, कीमतों, लागतों, बिक्री, लाभ, आय इत्यादि का अत्यधिक महत्त्व की हैं एक व्यापारी या अर्थशास्त्री के लिए। जनसंख्या आकलन और जनसंख्या प्रक्षेपण अपरिहार्य हैं एक अर्थव्यवस्था की प्रभावी योजना के लिए। फार्मास्युटिकल प्रतिष्ठानों को रुचि नई दवाओं के प्रभाव का रोग पर प्रभाव का अध्ययन या आकलन में होती है। प्रतीपगमन विश्लेषण सांख्यिकीय तकनीकों में एक बहुत महत्त्व की है इस तरह के पूर्वानुमानों करने के लिए। एम एम ब्लेयर के शब्दों में "प्रतीपगमन विश्लेषण गणितीय माप है दो या अधिक चरों के बीच औसत संबंधों की इसकी वास्तविक इकाईयों के क्रम में।"

हमलोग दैनानुदिनी जिंदगी में कई अंतर-संबंधित घटनाएं के अवलोकन करते हैं। उदाहरण के लिए, एक फसल की उपज वर्षा पर निर्भर करती है, एक उत्पाद की कीमत या मूल्य निर्भर है उत्पादन और विज्ञापन पर, एक विशेष उत्पाद की माँग इसके कीमत पर निर्भर करती है, एक परिवार का खर्च इसकी आय पर निर्भर करती है, और आगे। प्रतीपगमन जो दो चरों तक सीमित ~~साधारण~~ सरल प्रतीपगमन कहलाती है। बहुत बार एक विशेष संवृत्ति बहुविध कारकों से प्रभावित हो सकती है। एक समय में दो से अधिक चरों के अध्ययन के लिए प्रतीपगमन विश्लेषण बहुविध प्रतीपगमन कहलाती है। परन्तु इस अध्याय में हमलोग अपने को सिर्फ सरल प्रतीपगमन तक परिसीमित रखेंगे।

प्रतीपगमन विश्लेषण में दो प्रकार के चर होते हैं। वह चर जिसका मूल्य प्रभावित होता है या अनुमान लगाया जाता है निर्भर चर कहलाता है और वह चर जो मूल्यों को प्रभावित करता है और अनुमान-पूर्वानुमान के लिए प्रयुक्त होती है, स्वतंत्र चर कहलाती है। प्रतीपगमन विश्लेषण में स्वतंत्र चर समाश्रय (regressor) या भविष्यवक्ता (predictor) या व्याख्याकार (explanator) भी कहलाती है जबकि निर्भर चर समाश्रयी (regressand) या वर्णित चर (explained variable) भी कहलाती है।

9.2 रैखिक और गैर-रैखिक प्रतीपगमन

अगर दिया गया द्विचर समंक एक आलेख पत्र पर खींचा जाए, विक्षेप चित्र पर प्राप्त बिन्दुएं कमोबेश एक वक्र के चारों ओर संकेन्द्रित होंगी, जो 'प्रतीपगमन का वक्र' कहलाती है। अधिकतर ऐसी वक्र सरल नहीं होता है और बहुत भ्रामक है और कभी-कभी जटिल भी। प्रतीपगमन वक्र का गणितीय समीकरण, सामान्यतः प्रतीपगमन

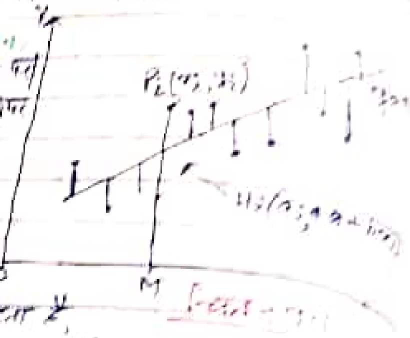
(14)

7.1 [illegible] प्रतीपगमन का आकलन [illegible]

माना कि $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, [illegible] के [illegible] को x और y पर। माना कि

$y = a + bx$ (7.1)

प्रतीपगमन की रेखा (अनुमानित) [illegible] किसी [illegible] बिन्दु $P_i(x_i, y_i)$ [illegible] [illegible] आकलन का विचलन [illegible] रेखा (7.1) [illegible] $P_i H_i$ है। अब, H_i का x-निर्देशांक, [illegible] P_i का है, x_i और [illegible] H_i [illegible] रेखा (7.1) पर पड़ता है, H_i का y-निर्देशांक (Co-ordinate) [illegible] लिया जाता है [illegible] इसलिए, P_i के लिए आकलन का विचलन (त्रुटि) [illegible]



$$P_i H_i = P_i M - H_i M$$

$$= y_i - (a + bx_i) \quad \ldots (7.2)$$

यह त्रुटि है (y-अक्ष के समानान्तर) [illegible] बिन्दु के लिए। त्रुटियाँ सभी बिन्दुओं के लिए [illegible] [illegible] के ऊपर पड़ती है, त्रुटि धनात्मक होगी और बिन्दु [illegible] के नीचे [illegible] हैं, त्रुटि ऋणात्मक होगी।

न्यूनतम वर्गों के सिद्धान्त के अनुसार, [illegible] b का निर्धारण करना है (7.1) में ऐसा कि आकलनों की [illegible] वर्गों का योग न्यूनतम हो दूसरे शब्दों में, [illegible] न्यूनतम करना है

$$E = \sum_{i=1}^{n} P_i H_i^2 = \sum_{i=1}^{n} (y_i - a - bx_i)^2 \quad (7.3)$$

a और b में विचरणता के संदर्भ में।

हम E को निम्न प्रकार से भी लिख सकते हैं

$$E = \sum (y - y_e)^2 = \sum (y - a - bx)^2$$

जहाँ y_e, y का प्राक्कलित मूल्य है जैसा (7.1) से लिया जाए x के दिये गये मान के लिए और योग चिन्ह (Σ) श्रेणीयों के n जोड़े के लिए लिया गया है।

अवकलन गणित (Differential Calculus) के उच्चिष्ठ निम्निष्ठ (Maxima and minima) सिद्धान्त का प्रयोग कर, E को आत्यंतिक (उच्चतम या निम्नतम) बनाने के लिए अगर इसका आंशिक अवकलज (Partial derivatives) a और b के सापेक्ष अलग-अलग गायब हो जाते हैं, a और b के [illegible]