खंड 3
परिमाणात्मक आँकड़ों का विश्लेषण

ignou
THE PEOPLE'S
UNIVERSITY

इकाई 7 एक विचर डेटा विश्लेषण (UNIVARIATE DATA ANALYSIS)

संरचना

7.0 उद्देश्य

7.1 विषय प्रवेश

7.2 अन्वेषी डेटा विश्लेषण (Exploratory Data Analysis)

- 7.2.1 पाँच-संख्या सारांश तथा छोटे भू-खंड (The 5-No.Summary and Boxplots)
- 7.2.2 बॉक्स प्लॉट बनाने की प्रक्रिया तथा इससे प्राप्त निष्कर्ष
- 7.2.3 उदाहरण

7.3 आनुमानिक सांख्यिकी : आधारभूत संकल्पनाएँ तथा निर्णयन में केंद्रीय प्रवृत्ति तथा परिक्षेपण/प्रकीर्णन (Dispersions) के मापों का महत्त्व

- 7.3.1 प्राचल तथा सांख्यिकी
- 7.3.2 यादृच्छिक चर तथा प्रायिकता बंटन
- 7.3.3 प्रसामान्य बंटन
- 7.3.4 मानक प्रसामान्य बंटन
- 7.3.5 व्यावहारिक समस्याओं के हल में मानक प्रसामान्य बंटन के अनुप्रयोग
- 7.3.6 परिकल्पना परीक्षण (Hypothesis Testing)

7.4 आनुमानिक आंकड़े/सांख्यिकी : बिंदु आकलन तथा समष्टि प्राचलों के लिए विश्वास्यता अंतराल तय करना

- 7.4.1 समष्टि माध्य के लिए बिंदु आकलन
- 7.4.2 माध्य के लिए विश्वास्यता अंतराल (Confidence Intervals) जब समष्टि मानक विचलन (σ) ज्ञात हो
- 7.4.3 माध्य के लिए विश्वास्यता अंतराल जब समष्टि मानक विचलन (σ) ज्ञात न हो
- 7.4.4 अनुपातों के लिए विश्वास्यता अंतराल
- 7.4.5 प्रसरणों तथा मानक विचलनों के लिए विश्वास्यता अंतराल

7.5 सार-संक्षेप

7.6 संदर्भ ग्रंथादि

7.7 बोध प्रश्नों के उत्तर अथवा संकेत

7.0 उद्देश्य

इस इकाई के अध्ययन के पश्चात्, आप निम्नलिखित से भली-भाँति अवगत हो जाएंगेः

- आनुमानिक सांख्यिकी के उद्देश्यों से;
- अन्वेषी डेटा विश्लेषण की संकल्पना से;
- प्रसामान्य बंटन की विशेषताओं तथा व्यावहारिक जीवन में सामने आने वाली सांख्यिकी समस्याओं के हल में इनके अनुप्रयोगों से;

- परिकल्पना परीक्षण की संकल्पना से; तथा इस प्रक्रिया में तथा उनके विनिमय में सामने आने वाली त्रुटियों के प्रकारों से; तथा
- समष्टि प्राचलों के लिए विश्वास्यता अंतराल तय करने की प्रक्रिया से।

7.1 विषय प्रवेश

इस पाठ्यक्रम की इकाई 2 में, आपका परिचय सांख्यिकी की उस शाखा से करवाया गया जो कि डेटा के संक्षिप्तीकरण तथा वर्णन से संबंधित है तथा जिसे वर्णनात्मक सांख्यिकी कहते हैं। आपने अनेक वर्णनात्मक मापों का परिकलन करना सीखा जैसे कि औसत, प्रसरण, शतमक, इत्यादि। इससे यह स्पष्ट होता है कि वर्णनात्मक सांख्यिकी का क्षेत्र केवल संख्याओं/प्रेक्षणों के सारणीकरण तथा उनके आलेखीय प्रस्तुतीकरण तक ही सीमित नहीं है। यह इकाई मूलतः डेटा के विश्लेषण पर केंद्रित है। हम अन्वेषी डेटा विश्लेषण से प्रारंभ करेंगे, जिसमें हम डेटा/आंकड़ों की प्रमुख विशेषताओं को जानने का प्रयास करेंगे। इसके पश्चात् हम अपना रुख और कठिन आनुमानिक सांख्यिकी की ओर करेंगे जिसमें समष्टि के बारे में निष्कर्ष ज्ञात करने की तथा प्राप्त निष्कर्षों की विश्वसनीयता को मापने की विधियाँ सम्मिलित होती हैं।

7.2 अन्वेषी डेटा विश्लेषण (EXPLORATORY DATA ANALYSIS)

पारंपरिक सांख्यिकी में, आंकड़ों को एक बारंबारता बंटन के रूप में व्यवस्थित किया जाता है। इस बंटन की सहायता से, बंटन का आकार या प्रकृति जानने के लिए, आयतचित्र, बारंबारता बहुभुज, ओजाइव/तोरण जैसे विभिन्न ग्राफ/आलेख बनाए जा सकते हैं। इसके साथ ही, डेटा का सारांश ज्ञात करने के लिए माध्य तथा मानक विचलन जैसे विभिन्न आंकड़े प्राप्त किए जा सकते हैं। पारंपरिक विश्लेषण का उद्देश्य डेटा की प्रकृति के बारे में विभिन्न अनुमानों की पुष्टि करना है।

अन्वेषी डेटा विश्लेषण (EDA) में हमारा उद्देश्य यह जानने के लिए डेटा की जाँच/परीक्षण करना होता है कि डेटा के बारे में क्या जानकारी प्राप्त की जा सकती है जैसे कि डेटा का मध्य बिंदु/केंद्र या विस्तार इत्यादि। EDA में प्रयोग किया जाने वाला केंद्रीय प्रवृत्ति का माप माध्यिका तथा विचरण का माप, अंतश्चतुर्थक परिसर $Q_3 - Q_1$ होता है। EDA में डेटा को बॉक्सप्लाट (जिसे बॉक्स-एंड-व्हिस्कर प्लॉट भी कहा जाता है)। अन्वेषी डेटा विश्लेषण का विकास जॉन तुकी द्वारा किया गया तथा यह उनकी पुस्तकें 'Exploratory data analysis' में प्रस्तुत किया गया।

7.2.1 पाँच-संख्या सारांश तथा छोटे भू-खंड

एक बॉक्स प्लॉट किसी डेटा समुच्चय को आलेखीय रूप में निरूपित करने के लिए किया जाता है। इस प्लॉट में पाँच विशिष्ट मान शामिल होते हैं :

1) डेटा समुच्चय का सबसे छोटा मान (अर्थात् न्यूनतम)
2) Q_1

3) माध्यिका (अर्थात् Q_2)
4) Q_3
5) डेटा समुच्चय का सबसे बड़ा मान (अर्थात्, अधिकतम)

ये मान डेटा समुच्चय का फाइब-नंबर समरी (पाँच-संख्या सारांश) कहलाते हैं।

7.2.2 बॉक्स प्लॉट बनाने की प्रक्रिया तथा इससे प्राप्त निष्कर्ष

एक डेटा समुच्चय का बॉक्स प्लॉट, एक ऐसा आलेख है जिसमें एक क्षैतिज रेखा डेटा की न्यूनतम मान वाली प्रविष्टि से Q_1 तक खींची जाती है, एक क्षैतिज रेखा Q_3 से डेटा की अधिकतम मान वाली प्रविष्टि तक खींची जाती है तथा एक बॉक्स बनाया जाता है जिसकी ऊर्ध्वाधर भुजाएं Q_1 तथा Q_3 से होकर जाती हैं तथा बॉक्स के अंदर एक ऊर्ध्वाधर रेखा माध्यिका Q_2 से होकर जाती है। नीचे इस विधि को चरणबद्ध रूप में दिया गया है :

1) डेटा मानों के लिए फाइब-नंबर समरी, अर्थात् डेटा मानों को न्यूनतम तथा अधिकतम मान, Q_1 और Q_3 तथा माध्यिका, ज्ञात करें।

2) एक क्षैतिज अक्ष ऐसे पैमाने के साथ बनाएं जिसमें डेटा के न्यूनतम तथा अधिकतम मान सम्मिलित हों।

3) एक बॉक्स बनाएं जिसकी ऊर्ध्वाधर भुजाएं Q_1 तथा Q_3 से होकर जाती हों। माध्यिका से होकर जाती हुई ऊर्ध्वाधर रेखा खींचें।

4) डेटा के न्यूनतम मान से बॉक्स की बायीं भुजा तक एक रेखा खीचें। इसी प्रकार डेटा के अधिकतम मान से बॉक्स की दायीं भुजा तक एक रेखा खीचें।

बॉक्स प्लॉट से प्राप्त होने वाली जानकारी को नीचे सारबद्ध किया गया है :

- यदि माध्यिका बॉक्स के केंद्र के समीप हो, तो बंटन लगभग सममित होता है।
- यदि माध्यिका बॉक्स के केंद्र के बायीं ओर स्थित हो, तो बंटन धनात्मक रूप से विषम होता है।
- यदि माध्यिका बॉक्स के केंद्र के दायीं ओर स्थित हो, तो बंटन ऋणात्मक रूप से विषम होता है।
- यदि रेखाएं लगभग समान लंबाई की हों तो बंटन लगभग सममित होता है।
- यदि दायीं रेखा, बायीं रेखा से बड़ी हो तो बंटन धनात्मक रूप से विषम होता है।
- यदि बायीं रेखा, दायीं रेखा से बड़ी हो तो बंटन ऋणात्मक रूप से विषम होता है।

यदि दो या दो से अधिक डेटा समुच्चयों के बॉक्स प्लॉट, एक ही अक्ष पर बनाए जाएं, तो बंटनों की तुलना की जा सकती है। औसतों की तुलना के लिए, माध्यिका की स्थिति का प्रयोग करें। परिवर्तनशीलता की तुलना के लिए अंतश्चतुर्थक परिसर (Q_3–Q_1) का, अर्थात् बक्सों की लंबाई का प्रयोग करें।

7.2.3 उदाहरण

एक आहार विशेषज्ञ की रुचि, असली पनीर में सोडियम की मात्रा की तुलना पनीर के एक विकल्प में सोडियम की मात्रा से करने में है। दो यादृच्छिक प्रतिदर्शों के आंकड़े नीचे दर्शाए गए हैं।

असली पनीर				**पनीर का विकल्प**			
310	420	45	40	270	180	250	290
220	240	180	90	130	260	340	310

हल :

चरण 1: Q_1, MD तथा Q_3 ज्ञात करें।

40 45 90 180 220 240 310 420

Q_1 MD Q_3

$$Q_1 = \frac{45+90}{2} = 67.5 \qquad MD = \frac{180+220}{2} = 200$$

$$Q_3 = \frac{240+310}{2} = 275$$

चरण 2: पनीर के विकल्प के लिए Q_1, MD तथा Q_3 ज्ञात करें।

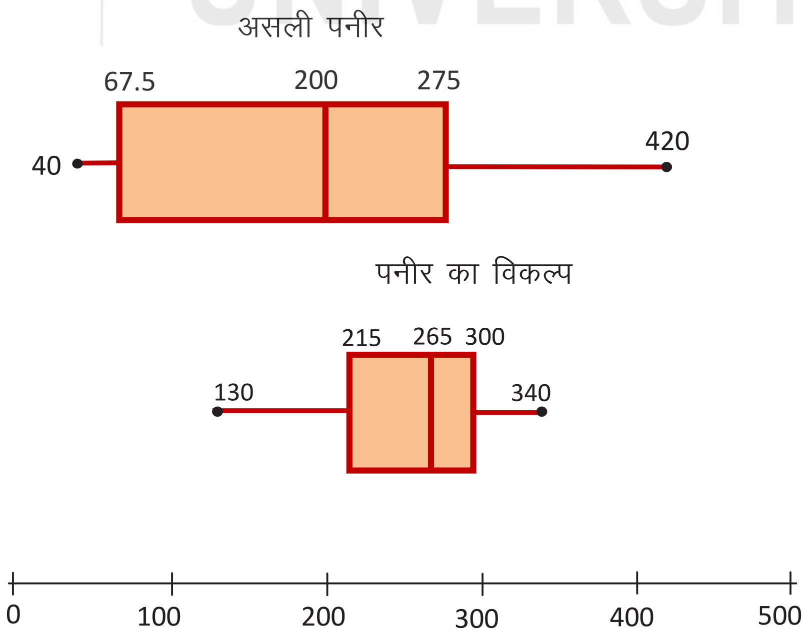
130 180 250 260 270 290 310 340

Q_1 MD Q_3

$$Q_1 = \frac{180+250}{2} = 215 \qquad MD = \frac{260+270}{2} = 265$$

$$Q_3 = \frac{290+310}{2} = 300$$

चरण 3 : एक ही आलेख में दोनों बंटनों को बॉक्स प्लॉट बनाएं।



चरण 4: दोनों प्लॉटस् की तुलना करें। यह पूर्णतया स्पष्ट है कि पनीर के विकल्प के बंटन की माध्यिका असली पनीर के बंटन की माध्यिका से अधिक है। दूसरी ओर, असली पनीर के बंटन की परिवर्तनशीलता या उसका विस्तार पनीर के विकल्प के बंटन की परिवर्तनशीलता से अधिक है।

बोध प्रश्न 1

1) पदों 'फाइव-नंबर समरी' तथा 'बॉक्सप्लॉट' को परिभाषित कीजिए।

....

....

....

....

....

2) एक बॉक्सप्लॉट की सहायता से संबंधित बंटन की विशेषताओं की व्याख्या कीजिए।

....

....

....

....

7.3 आनुमानिक सांख्यिकी : आधारभूत संकल्पनाएँ तथा निर्णयन में केंद्रीय प्रवृत्ति तथा परिक्षेपण/प्रकीर्णन के मापों का महत्त्व

वर्णनात्मक तथा अनुमानिक सांख्यिकी एक-दूसरे से संबंधित हैं। ***वर्णनात्मक सांख्यिकी*** में डेटा का संग्रहण, संगठन, संक्षेपण तथा प्रस्तुतीकरण सम्मिलित होते हैं; दूसरी ओर ***आनुमानिक सांख्यिकी*** प्रतिदर्श डेटा के प्रयोग से डेटा समष्टि के बारे में अनुमान लगाने से संबंधित है। ***समष्टि*** से हमारा तात्पर्य है उन सभी व्यक्तियों या वस्तुओं का संग्रह जिनमें अन्वेषक प्राथमिक रूप से, अपने सांख्यिकीय अध्ययन के लिए, रुचि रखता है। वर्णनात्मक सांख्यिकी – जब समष्टियों पर लागू की जाए – तो उसमें आलेखों, चार्टों तथा तालिकाओं का निर्माण/की रचना तथा विभिन्न वर्णनात्मक मापों जैसे कि औसत, परिवर्तनशीलता के मापों तथा शतमक इत्यादि का परिकलन, जिसके बारे में आपने पिछली इकाई में पढ़ा, सम्मिलित होते हैं। हालाँकि सामान्यतः पूरी समष्टि के डेटा तक हमारी पहुँच नहीं होती तथा हमारी समष्टि के बारे में जानकारी उसके एक प्रतिदर्श पर आधारित होती है। उदाहरण के लिए, शोधकर्ता की रुचि एक राज्य के सभी विद्यार्थियों द्वारा सभी परीक्षाओं में प्राप्त अंकों में हो सकती है। परंतु ये संभव नहीं होता कि पूरे राज्य के सभी विद्यार्थियों द्वारा सभी परीक्षाओं में प्राप्त अंकों का अध्ययन किया जा सके, अतः शोधकर्ता को विद्यार्थियों को एक छोटे

प्रतिदर्श (जैसे कि 100 विद्यार्थियों का एक प्रतिदर्श) तक सीमित रखना पड़ता है, जो कि राज्य के सभी विद्यार्थियों की समष्टि को निरूपित करने के लिए प्रयोग किया जाता है। अतः, सांख्यिकीय समष्टि का कोई प्रतिदर्श, समष्टि की इकाइयों या उसके अवयवों का उपसमुच्चय होता है जो अन्वेषण के दौरान एकत्रित/संकलित किया जाता है। आनुमानिक सांख्यिकी इन प्रतिदर्शों के आधार पर उन समष्टियों के बारे में व्यापक निष्कर्ष जैसे कि बिंदु आकलन, अंतराल आकलन तथा परिकल्पना परीक्षण, इत्यादि ज्ञात करती हैं जिनसे वे प्रतिदर्श लिये गये हैं।

इससे पहले कि आनुमानिक सांख्यिकी विधियों द्वारा किसी प्रतिदर्श से प्राप्त जानकारी का प्रयोग अध्ययन किए जा रहे विषय की जांच के संपूर्ण विश्लेषण के लिए किया जा सके, प्रतिदर्श द्वारा प्राप्त जानकारी का वर्णनात्मक सांख्यिकी की विधियों द्वारा प्रबंधन तथा संक्षेपण किया जाना आवश्यक है। आनुमानिक सांख्यिकी की विधियों को ठीक प्रकार से समझने के लिए कुछ आवश्यक संकल्पनाओं की चर्चा हम अगले भाग में कर रहे हैं।

7.3.1 प्राचल तथा सांख्यिकी

जाँचाधीन समष्टि की विशेषताओं को संख्यात्मक ***प्राचलों*** जैसे कि माध्य, मानक विचलन इत्यादि जिनका अध्ययने आपने इकाई 2 में किया, से व्यक्त किया जा सकता है। एक शोध अध्ययन का उद्देश्य सामान्यतः इन प्राचलों के मानों की जाँच करना होता है जो कि अज्ञात होते हैं तथा इसलिए इनका अनुमान एक प्रतिदर्श के आधार पर लगाया जाता है। किसी समष्टि प्राचल का प्रतिदर्श के आधार पर अनुमान एक ***प्रतिदर्शज*** (statistic) कहलाता है जिसका प्रयोग अज्ञात प्राचल का अनुमान लगाने के लिए किया जाता है।

उदाहरण के लिए, प्रतिदर्श माध्य ($\bar{X}$) तथा प्रतिदर्श मानक विचलन S, प्रतिदर्श डेटा से प्राप्त प्रतिदर्शज हैं जिनका प्रयोग संबंधित समष्टि प्राचलों, समष्टि माध्य (μ) तथा समष्टि मानक विचलन (σ) का अनुमान लगाने के लिए किया जा सकता है।

परिभाषा : एक प्राचल, समष्टि का संक्षिप्त विवरण होता है। एक प्रतिदर्शज, एक प्रतिदर्श का ज्ञात संक्षिप्त वर्णन है जो प्राचलों के बारे में अनुमान लगाने के लिए प्रयोग किया जा सकता है।

कृपया ध्यान दें : प्रतिदर्शजों ($\bar{X}$ तथा S) के मान चुने गए प्रतिदर्श पर निर्भर करते हैं जबकि समष्टि के प्राचल (μ तथा σ) स्थिर रहते हैं। इस अर्थ में प्रतिदर्शज ***यादृच्छिक चर*** कहलाते हैं। अगले भाग में हम यादृच्छिक चर के अर्थ, उसके प्रकार तथा उससे संबद्ध प्रायिकता बंटनों पर विचार करेंगे।

7.3.2 यादृच्छिक चर तथा प्रायिकता बंटन

यादृच्छिक का अर्थ होता है जिसका पूर्वानुमान न लगाया जा सके; एक ***यादृच्छिक चर*** जिसे सामान्यतः बड़े अक्षरों X, Y, इत्यादि से व्यक्त किया जाता है, एक यादृच्छिक घटना का परिणाम होता है (एक ऐसी प्रक्रिया जिसके परिणाम का अनुमान निश्चितता के साथ नहीं लगाया जा सकता जैसे कि एक सिक्के को उछालना या

पासे को लुढ़काना) यादृच्छिक चर संतत भी हो सकते हैं और असंतत भी। एक असंतत यादृच्छिक चर द्वारा लिए जाने वाले मानों की संख्या केवल गणनीय हो सकती है अर्थात्, ये चर अपने दो मानों के बीच आने वाले सभी मान नहीं लेते। उदाहरण के लिए, यदि एक पासे को पाँच बार फेंकने पर, पासे की ऊपरी फलक पर किसी सम अंक (2, 3 या 6) के आने की संख्या, एक यादृच्छिक अचर है जो कि मानो 0, 1, 2, 3, 4, 5 में से कुछ भी हो सकता है। यह चर 1.2, 2.5, 4,6 इत्यादि मान नहीं लेता। दूसरी ओर, एक संतत यादृच्छिक चर अपने दो मानों के बीच के सभी मान लेता है। संतत चरों के उदाहरण है : वयस्क पुरुषों की ऊँचाई, चूहों के शरीर का तापमान, वयस्कों के कोलेस्ट्रॉल का स्तर।

प्रायिकता आनुमानिक सांख्यिकी में एक महत्त्वपूर्ण भूमिका निभाती है। एक यादृच्छिक घटना में दोहराए गए पर्यवेक्षणों का एक लंबा अनुक्रम होता है, जहाँ यह आवश्यक नहीं होता कि प्रत्येक पर्यवेक्षण कोई विशिष्ट परिणाम ही दे। किसी विशिष्ट परिणाम के आने की प्रायिकता, पुनरावृत्त पर्यवेक्षणों की एक बड़ी संख्या में उस परिणाम वाले पर्यवेक्षणों का अनुपात होती है। उदाहरण के लिए, एक पासे को फेंकने पर सम अंक प्राप्त की प्रायिकता $\frac{3}{6}$ या $\frac{1}{2}$ होगी।

एक यादृच्छिक चर X का प्रायिकता बंटन, X के संभव मानों को उनसे संबद्ध प्रायिकताओं को सूचीबद्ध करता है। उदाहरण के लिए, एक असंतत यादृच्छिक चर X का प्रायिकता बंटन, जहाँ X एक 6 फलक वाले सममित पासे के फेंकने पर प्राप्त होने वाले अंक को निरूपित करता है, इस प्रकार है –

X:	1	2	3	4	5	6
P(X):	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

यदि हम एक यादृच्छिक चर के संभव परिणामों को x_i से व्यक्त करें जहाँ $i = 1, 2, 3 \ldots k$ है तथा इन मानों के संबद्ध प्रायिकता को $P(X = x_i) = P(x_i)$ से व्यक्त करें, तो निम्नलिखित कथन सत्य होंगे :

i) $0 \le P(x_i) \le 1$, अर्थात् प्रत्येक प्रायिकता 0 से लेकर 1 तक हो सकती है;

ii) $\sum_{i=1}^{k} P(x_i) = 1$ अर्थात्, कुल प्रायिकता 1 के बराबर होती है।

दूसरी ओर, एक संतत यादृच्छिक अचर X, किन्हीं दो संख्याओं a और b के बीच के अंतराल में प्रत्येक मान लेता है। अर्थात् a और b के मध्य के अंतराल में स्थित किन्हीं भी दो बिंदुओं x_1 तथा x_2 के लिए (जहाँ $x_1 \le x_2$ हो, जिससे $[x_1, x_2]$ भी एक अंतराल बन जाए) यादृच्छिक चर X का मान अंतराल $[x_1, x_2]$ में स्थित होने की प्रायिकता 0 और 1 के बीच की कोई संख्या होगी तथा X का मान [a, b] में होने की प्रायिकता 1 होगी :

$$0 \le P(x_1 \le X \le x_2) \le 1 \text{ तथा } P(a \le X \le b) = 1$$

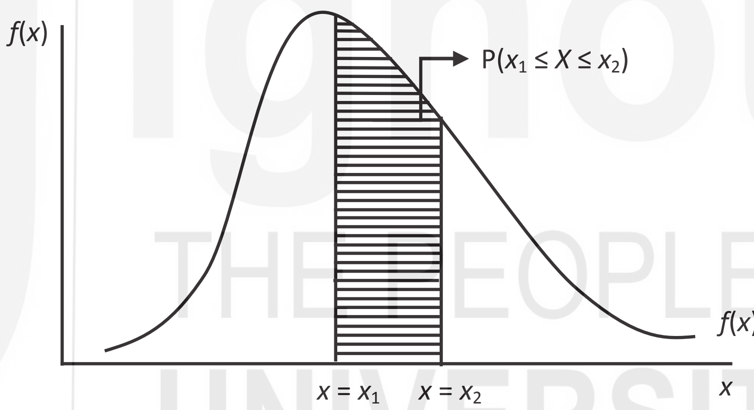
एक असंतत् यादृच्छिक चर के प्रायिकता बंटन की तरह से एक संतत यादृच्छिक चर के बंटन को एक तालिका के रूप में प्रस्तुत नहीं किया जा सकता। इसे एक फलन के रूप में या एक आलेख के रूप में प्रस्तुत किया जाता है। आलेखीय रूप में, एक

संतत यादृच्छिक चर का प्रायिकता बंटन एक प्रायिकता घनत्व वक्र के रूप में दर्शाया जाता है (चित्र 7.1 देखें)। एक ***प्रायिकता घनत्व वक्र*** एक निष्कोण संतत वक्र होता है जिसके नीचे का क्षेत्रफल ठीक 1 के बराबर होता है। वह फलन जो इस वक्र को परिभाषित करता है ***प्रायिकता घनत्व फलन $f(x)$*** कहलाता है जो निम्नलिखित गुणधर्मों को परिभाषित करता है :

i) अंतराल $[a, b]$ में स्थित प्रत्येक बिंदु x के लिए $f(x) \geq 0$ होता है

ii) सभी $x_1, x_2 \in [a, b]$ के लिए जहाँ $x_1 \leq x_2$ है $P(x_1 \leq X \leq x_2)$ = प्रायिकता घनत्व वक्र के नीचे, कोटि $X = x_1$ तथा $X = x_2$ के मध्य स्थित क्षेत्रफल के बराबर होती है। इसे अवकल गणित के पदों में $\int_{x_1}^{x_2} f(x)dx$ से प्राप्त किया जाता है। ध्यान रहे कि $P(X = x_1) = 0$ होता है अर्थात् प्रायिकता घनत्व वक्र और $X = x_1$ पर स्थित ऊर्ध्वाधर रेखा के मध्य का क्षेत्रफल 0 होता है। अतः, एक संतत यादृच्छिक चर के लिए, किसी एक बिंदु पर प्रायिकता सदैव 0 होती है।*

iii) प्रायिकता घनत्व वक्र के नीचे का कुल क्षेत्रफल 1 के बराबर होता है, अर्थात् $\int_a^b f(x)dx = 1$ होता है।



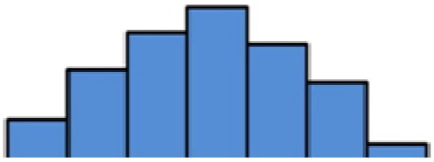
चित्र 7.1 : प्रायिकता घनत्व वक्र

7.3.3 प्रसामान्य बंटन

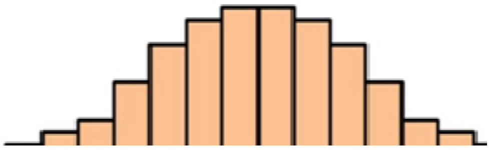
एक संतत यादृच्छिक चर के लिए सबसे महत्त्वपूर्ण प्रायिकता बंटन प्रसामान्यतः प्रायिकता बंटन होता है। सांख्यिकीय विश्लेषण में इस बंटन का अत्याधिक प्रयोग होता है क्योंकि यह अनेक संतत चरों के बंटनों को निरूपित करता है। उदाहरण के लिए, यदि एक शोधकर्ता 100 वयस्क स्त्रियों का एक यादृच्छिक प्रतिदर्श चुनता है तथा उनकी ऊँचाई के आंकड़े एकत्रित करके एक हिस्टोग्राम आयतचित्र बनाता है, तो उसे चित्र 7.2(क) जैसा आलेख प्राप्त होता है। अब, यदि शोधकर्ता प्रतिदर्श को बड़ा कर दे तथा वर्गों की चौड़ाई घटा दें, तो प्राप्त आयतचित्र ऐसे दिखेंगे जैसे कि चित्रों 7.2(ख) तथा (ग) में दर्शाए गए हैं। इस प्रकार, यदि भारत की सभी वयस्क स्त्रियों की ऊँचाई नापना संभव हो तथा उनका आलेख बनाया जाए तो आयतचित्र एक प्रसामान्य बंटन को रूप ले लेगा जैसा चित्र 7.2(घ) में दर्शाया गया है।

---

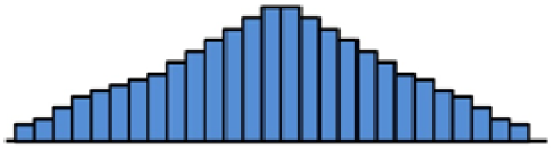
* आपने निश्चित समाकल के बारे में सेमेस्टर 1 के पाठ्यक्रम 'अर्थशास्त्र में गणितीय विधियाँ –I (BECC-103) में पढ़ा।



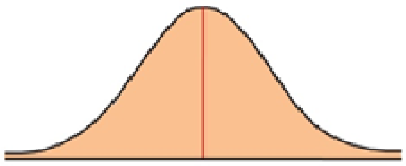
(क) 100 स्त्रियों का यादृच्छिक प्रतिदर्श



(ख) प्रतिदर्श को बड़ा किया गया तथा
वर्ग की चौड़ाई को घटाया गया



(क) प्रतिदर्श को और अधिक बढ़ाया गया तथा
वर्ग की चौड़ाई को और अधिक घटाया गया

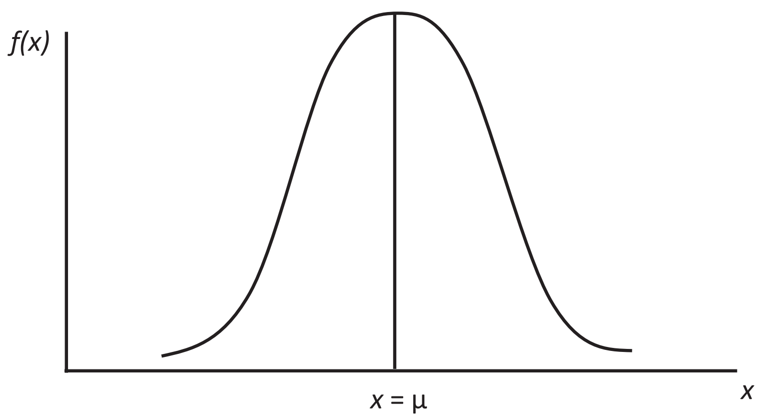


(घ) समष्टि के लिए प्रसामान्य बंटन

चित्र 7.2

आलेखीय रूप में एक प्रसामान्य बंटन एक घंटी के आकार (घंटाकार) से निरूपित किया जाता है (चित्र 7.3 देखें)। एक अंतत चर के प्रायिकता बंटन की तरह ही, एक संतत यादृच्छिक चर के मानों को भी अंक या प्रेक्षण कहा जाता है तथा इन्हें x-अक्ष पर निरूपित किया जाता है। किसी भी अंतराल में, वक्र के नीचे का क्षेत्रफल, उस अंतराल में स्थित प्रेक्षणों का अनुपात होता है। माध्य (μ) प्रेक्षणों के वास्तविक या सामान्य मान को निरूपित करता है और इससे लिए गए विचलन त्रुटियाँ कहलाती हैं। माध्य के आस पास के अंतराल में, जहाँ वक्र सबसे अधिक ऊँचा होता है, प्रेक्षण अधिक सघनता से उपस्थित होते हैं। वक्र के छोरों की ओर, ऊँचाई कम होती है, जैसे-जैसे हम माध्य से दूर जाते हैं प्रेक्षण कम (सघन) होते जाते हैं अर्थात्, प्रेक्षणों की माध्य के समीप स्थित होने की संभावना अधिक होती है तथा माध्य से दूर होने की कम। प्रसामान्य बंटन वाले एक संतत यादृच्छिक चर X को परिभाषित करने वाले दो प्राचल होते हैं – माध्य μ तथा मानक विचलन σ। माध्य μ तथा मानक विचलन σ वाले एक संतत यादृच्छिक चर X ~ के प्रसामान्य बंटन को $X \sim N(\mu, \sigma)$ के रूप में व्यक्त किया जाता है तथा इसका प्रायिकता घनत्व फलन $f(x)$

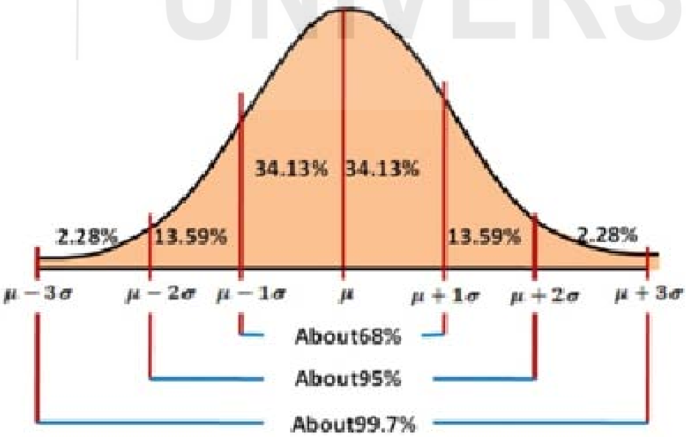
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad , -\infty < x < \infty \text{ होता है।}$$



चित्र 7.3 : प्रसामान्य बंटन वक्र

एक प्रसामान्य बंटन वक्र का आकार और इसकी स्थिति इसके दोनों प्राचलों (μ तथा σ) पर निर्भर करती है। माध्य का मान प्रसामान्य की स्थिति को नियत करता है कि इसका केंद्र कहाँ होगा जबकि मानक विचलन का मान इसके फैलाव को निर्धारित करता है; σ जितना अधिक होगा, वक्र उतना ही अधिक फैला हुआ या चपटा होगा। किसी प्रसामान्य बंटन के गुणधर्म, जो इसे विशेष बनाते हैं, इस प्रकार हैं :

1) एक प्रसामान्य वक्र घंटाकार होता है।
2) इसके माध्य, माध्यिका तथा बहुलक समान होते हैं तथा बंटन के केंद्र पर स्थित होते हैं।
3) एक प्रासामान्य वक्र एकबहुलकी होता है अर्थात् इसका केवल एक ही बहुलक होता है।
4) यह वक्र माध्य के सापेक्ष सममित होता है अर्थात् इसका आकार माध्य (केंद्र) से होकर जाती हुई एक ऊर्ध्वाधर के दोनों ओर एक समान होता है।
5) यह वक्र संतत होता है अर्थात् इसमें कोई रिक्तियाँ अथवा छिद्र नहीं होते।
6) यह वक्र x-अक्ष को कभी भी स्पर्श नहीं करता। सैद्धांतिक रूप से, इस वक्र को दोनों ओर जितना चाहे बढ़ा लिया जाए, यह कभी भी x-अक्ष को स्पर्श नहीं करेगा – पर यह निरंतर x-अक्ष के समीप और समीप आता चला जाएगा।
7) एक प्रसामान्य बंटन वक्र के नीचे का कुल क्षेत्रफल 1 के बराबर या 100% होता है। यह तथ्य असामान्य प्रतीत हो सकता है परंतु इसे अवकल गणित के प्रयोग से सिद्ध किया जा सकता है।
8) प्रसामान्य वक्र के नीचे माध्य के दोनों ओर 1 मानक विचलन की दूरी तक स्थित भाग का क्षेत्रफल लगभग 0.68 या 68% होता है; 2 मानक विचलन की दूरी तक स्थित भाग का क्षेत्रफल लगभग 0.95 या 95%; तथा 3 मानक विचलन की दूरी तक स्थित भाग का क्षेत्रफल लगभग 0.997 या 99.7% होता है। नीचे दिए चित्र में प्रत्येक क्षेत्र का क्षेत्रफल दर्शाया गया है।



चित्र 7.4

इसका सरल भाषा में अर्थ यह है कि लगभग 68.26% प्रेक्षण प्रसामान्य वक्र के नीचे $\mu \pm \sigma$ परिसर में स्थित होते हैं या दूसरे शब्दों में $P(\mu - \sigma \leq X \leq \mu + \sigma) = 0.68$। इसी प्रकार $P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = 0.95$ तथा $P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 0.997$.

प्रसामान्य बंटन का प्रयोग करने का औचित्य

जैसा कि हमने प्रारंभ में ही बताया कि अन्वेषी विश्लेषण में हमारा प्रमुख उद्देश्य होता है एक प्रतिदर्श के आधार पर मूल समष्टि के बारे में अनुमान लगाना। कोई भी समष्टि माध्य तथा प्रसरण जैसे प्राचलों से परिलक्षित की जाती है तथा मूल समष्टि से चुना गया एक प्रतिदर्श हमें समष्टि के प्राचलों के प्रतिदर्शज उपलब्ध करवाता है। हम यह भी जानते हैं कि एक प्रतिदर्श के माध्य के उसी समष्टि से लिए गए एक-दूसरे प्रतिदर्श के माध्य से अलग होने की संभावना है। अतः, प्रतिदर्श माध्य एक यादृच्छिक चर है, जो कि अनेक अलग-अलग मान ले सकता है। इन मानों के साथ संबद्ध प्रायिकताओं को भी ले लिया जाएं तो हमें एक बंटन प्राप्त होता है। इस बंटन को प्रतिदर्श माध्यों का बंटन कहते हैं। जब प्रतिदर्श पर्याप्त रूप से बड़े हों तो इस बंटन का आकार लगभग एक प्रसामान्य बंटन वक्र की तरह हो जाता है (चाहे मूल समष्टि कैसी भी हो)। इससे हम यह पाते हैं कि यदि मूल समष्टि का बंटन प्रसामान्य हो तो प्रतिदर्श माध्यों का बंटन ठीक प्रसामान्य होगा। परंतु, यदि मूल समष्टि का बंटन प्रसामान्य न हो तो प्रतिदर्श माध्यों का बंटन लगभग प्रसामान्य केवल तभी होगा जब प्रतिदर्श माप पर्याप्त रूप से बड़ा हो। यह परिणाम हमें एक अत्यंत महत्त्वपूर्ण प्रमेय से प्राप्त होता है जिसे सैंट्रल लिमिट थ्योरम/केंद्रीय सीमा प्रमेय कहते हैं।

केंद्रीय सीमा प्रमेय (Central Limit Theorem)

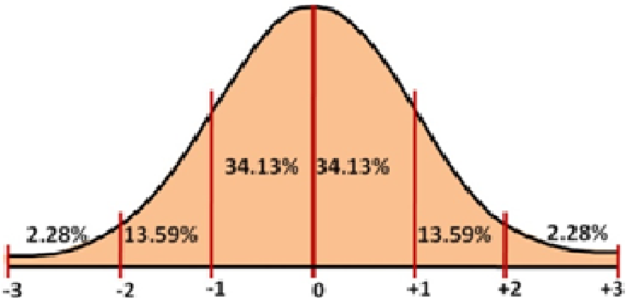
संक्षेप में, केंद्रीय सीमा प्रमेय के अनुसार, यदि एक यादृच्छिक चर n स्वतंत्र, समान रूप से वितरित, अप्रसामान्य यादृच्छिक चरों का योग है, तो जैसे-जैसे n की ओर जाता है, इस यादृच्छिक चर का बंटन एक प्रसामान्य बंटन की ओर जाता है। नीचे दिये उपप्रमेय केंद्रीय सीमा प्रमेय से प्राप्त किया जा सकता है (परंतु इसको सिद्ध करना इस पाठ्यक्रम के विषय क्षेत्र से बाहर है) :

जैसे-जैसे प्रतिदर्श मान n बढ़ता है प्रतिदर्श माध्यों का बंटन एक प्रसामान्य बंटन की ओर बढ़ता है, तब भी जब मूल समष्टि जिसमें से ये प्रतिदर्श चुने गए हों, प्रसामान्य न हो।

केंद्रीय सीमा प्रमेय हमें प्रसामान्य बंटन के अनेक यादृच्छिक घटनाओं से जुड़े होने का कारण बतलाती है।

7.3.4 मानक प्रसामान्य बंटन

क्योंकि प्रत्येक प्रसामान्य बंटन वाले चर का अपना माध्य तथा मानक विचलन होता है, इन वक्रों का आकार तथा स्थिति इन प्राचलों के मानों के अनुसार भिन्न-भिन्न होगी। अतः, व्यावहारिक अनुप्रयोगों में हमें प्रत्येक चर के लिए, वक्र के नीचे के क्षेत्रफल की एक तालिका बनानी पड़ेगी। इस स्थिति को सरल बनाने के लिए, सांख्यिकीविद् मानक प्रसामान्य बंटन का प्रयोग करते हैं। मानक प्रसामान्य बंटन एक 0 माध्य तथा 1 मानक विचलन वाला प्रसामान्य बंटन होता है।



चित्र 7.5

7.3.5 व्यावहारिक समस्याओं के हल में मानक प्रसामान्य बंटन के अनुप्रयोग

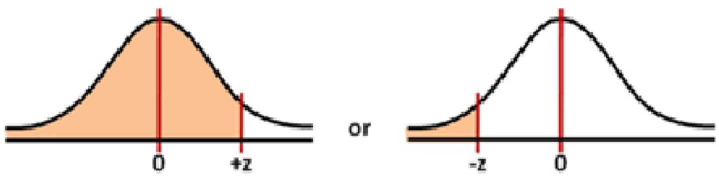
ऊपर दिए चित्र में, वक्र के नीचे प्रत्येक खंड में दिए हुए मान, उस खंड के क्षेत्रफल का अनुपात दर्शाते हैं। सभी प्रसामान्य बंटन वाले चर निम्नलिखित सूत्र के प्रयोग द्वारा एक मानक प्रसामान्य चर में परिवर्तित किए जा सकते हैं :

$Z = X$ का मान – माध्य / मानक विचलन अथवा $z = \frac{X-\mu}{\sigma}$

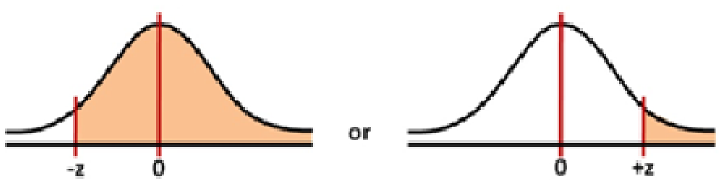
जैसे कि पहले कहा गया है, एक मानक प्रसामान्य बंटन वक्र के नीचे के क्षेत्रफल का प्रयोग व्यावहारिक समस्याओं को हल करने के लिए किया जाता है, जैसे कि ऐसी वयस्क स्त्रियों का प्रतिशत जिनका मान 5 फुट 4 इंच तथा 5 फुट 7 इंच के बीच में हो अथवा एक नई बैटरी के 4 साल से अधिक चलने की प्रायिकता। अतः, z के किसी भी मान के लिए, मानक प्रसामान्य बंटन वक्र के नीचे का क्षेत्रफल निकालने की विधि को समझना आवश्यक है। एक बार X के मान, ऊपर दिए सूत्र के अनुसार परिवर्तित कर लिए जाएं तो वे z मान कहलाते हैं। z मान वास्तव में यह दर्शाता है कि उसके संगत X मान, माध्य से कितने मानक विचलन दूर है। मानक प्रसामान्य विचर से संबंधित z-तालिका , – 3.49 से लेकर 3.49 तक z के किसी भी मान के लिए, मानक प्रसामान्य वक्र के नीचे का (4 दशमलव स्थानों तक) क्षेत्रफल, सार्थकता के विभिन्न स्तरों पर दर्शाती है।

मानक प्रसामान्य बंटन के प्रयोग से समस्याओं/प्रश्नों को हल करने की एक विधि नीचे दी तालिका में दर्शायी गई है :

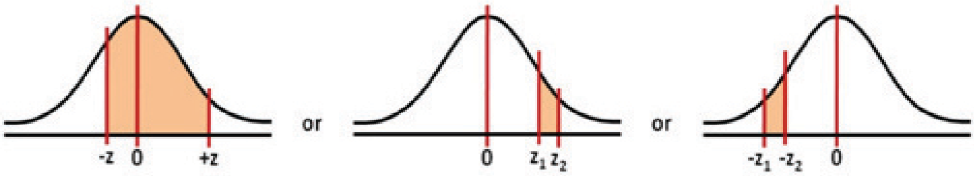
1. किसी z मान के बाईं ओर का क्षेत्रफल जानने के लिए : तालिका में z का मान ढूँढें तथा उसके संगत क्षत्रफल का प्रयोग करें।



2. किसी z मान के दाईं ओर का क्षेत्रफल जानने के लिए : तालिका में z का मान ढूँढें तथा उसके संगत क्षेत्रफल को 1 में से घटाएँ।



3. दो मानों के बीच का क्षेत्रफल जानने के लिए : दोनों z मानों के संगत क्षेत्रफल देखें तथा उन्हें घटाएँ



चित्र 7.6

तालिका के सामान्यतः 0 से लेकर किसी +z तक का क्षेत्रफल दर्शाती है। अतः, किसी +z से दाईं ओर का क्षेत्रफल 0.5 –, 0 से +z के क्षेत्रफल के बराबर होगा। यही क्षेत्रफल –z के बाईं ओर का भी होगा। किसी दिए हुए +z के बाईं ओर का क्षेत्रफल 0.5 + 0 से +z तक के क्षेत्रफल के बराबर होगा।

7.3.6 परिकल्पना परीक्षण

शोधकर्ता अनेक प्रकार के प्रश्नों के उत्तर खोजना चाहते हैं, जैसे कि शिक्षाविद् यह जानना चाह सकता है कि क्या पढ़ाने की एक नई तकनीक पारंपरिक तकनीक से बेहतर है। इस प्रकार के प्रश्न का उत्तर सांख्यिकीय परिकल्पना परीक्षण के माध्यम से पाया जा सकता है, जो कि समष्टि के बारे में किए गए दावे का मूल्यांकन करने के लिए निर्णय लेने की प्रक्रिया है। किसी परिकल्पना परीक्षण के लिए, शोधकर्ता को अध्ययनगत समष्टि को परिभाषित करने की, एक विशिष्ट परिकल्पना को वर्णित/कथन निश्चित करने की, एक सार्थकता स्तर तय करने की, समष्टि में से एक प्रतिदर्श का चयन करने की, डेटा संग्रहित करने की, सांख्यिकीय परीक्षण के लिए आवश्यक परिकलन करने की तथा एक निष्कर्ष तक पहुंचने की आवश्यकता होती है। अन्वेषी सांख्यिकी में, परिकल्पना परीक्षण में प्रयोग की जाने वाली विधियाँ हैं: (i) पारंपरिक विधि, (ii) p-मान विधि तथा (iii) विश्वास्यता अंतराल विधि।

परिकल्पना परीक्षण के विभिन्न चरण

पारंपरिक विधि

प्रत्येक परिकल्पना परीक्षण स्थिति एक परिकल्पना के कथन से प्रारंभ होती है। एक सांख्यिकी परिकल्पना किसी समष्टि प्राचल के बारे में एक अनुमान होता है। यह अनुमान सही भी हो सकता है और गलत भी। प्रत्येक स्थिति में दो प्रकार की सांख्यिकीय परिकल्पनाएँ हो सकती हैं :

i) निराकरणीय परिकल्पना

ii) वैकल्पिक परिकल्पना

i) **निराकरणीय परिकल्पना,** जिसे प्रतीकात्मक रूप में H_0 से व्यक्त किया जाता है, एक सांख्यिकीय परिकल्पना है जिसमें यह माना जाता है कि एक प्राचल तथा एक विशिष्ट मान में कोई अंतर नहीं है या यह कि दो प्राचलों में कोई अंतर नहीं है।

ii) वैकल्पिक परिकल्पना, जिसे प्रतीकात्मक रूप में H_1 से व्यक्त किया जाता है, एक सांख्यिकीय परिकल्पना है जिसमें यह माना जाता है कि एक प्राचल तथा एक विशिष्ट मान में अंतर है या यह कि दो प्राचलों के मानों में अंतर है।

यह स्पष्ट करने के लिए कि परिकल्पनाओं को किस प्रकार व्यक्त करना चाहिए हम तीन अलग-अलग सांख्यिकीय अध्ययनों को उदाहरणों के तौर पर प्रस्तुत करेंगे।

i) स्थिति A : एक आयुर्विज्ञान शोधकर्ता यह जानना चाहता है कि क्या एक नई चिकित्सा/औषधि के कोई अवांछनीय दुष्प्रभाव होंगे। शोधकर्ता की चिंता विशेष रूप से औषधि लेने वाले मरीजों की नाड़ी दर/स्पंद दर के प्रति है। औषधि लेने के पश्चात् एक मरीज़ की नाड़ी दर बढ़ेगी या कम होगी या उसमें कोई परिवर्तन नहीं होगा? क्योंकि शोधकर्ता को यह ज्ञात है कि जिस समष्टि का अध्ययन किया जा रहा है उसमें औसत नाड़ी दर 82 स्पंद/बीट्स प्रति मिनट है, तो इस स्थिति के लिए परिकल्पनाएँ होंगी : H_0: $\mu = 82$ तथा H_1:$\mu \neq 82$। निराकरणीय परिकल्पना यह निर्दिष्ट करती है कि माध्य में कोई परिवर्तन नहीं होगा तथा वैकल्पिक परिकल्पना का कथन है कि माध्य अलग होगा। इस परीक्षण को द्विपुच्छ परीक्षण (two-tailed test) कहते हैं, क्योंकि औषधि के संभव दुष्प्रभाव स्पंद दर का बढ़ना या घटना दोनों हो सकते हैं।

ii) स्थिति B : एक रसायनशास्त्री एक वाहन बैटरी का जीवनकाल बढ़ाने के लिए एक योगज तैयार करता है। यदि बिना इस योगज के वाहन बैटरी की जीवनकाल 36 महीने है, तो परिकल्पनाएँ होंगी : H_0: $\mu = 36$ तथा H_1: $\mu > 36$ । इस स्थिति में रसायनशास्त्री की रुचि केवल बैटरियों का जीवनकाल बढ़ाने में है, अतः वैकल्पिक परिकल्पना को, औसत 36 महीने से अधिक है, लिया गया है। यह दायाँ-पुच्छ परीक्षण (right-tailed test) कहलाता है, क्योंकि हमारी रुचि केवल जीवनकाल बढ़ाने में है।

iii) स्थिति C : एक ठेकेदार घरों में एक विशेष प्रकार का तापावरोधन का प्रयोग करके तापन (Heating) के बिल कम करना चाहता है। यदि मासिक तापन बिलों का औसत रु. 780 है, तो तापावरोधक के प्रयोग के द्वारा तापन के खर्च से संबंधित परिकल्पनाएँ होंगी : H_0: $\mu = 780$ तथा H_1: $\mu < 780$। यह एक बायाँ-पुच्छ परीक्षण (left-tailed test) है, क्योंकि ठेकेदार की रुचि केवल तापन के खर्च को कम करना है। परिकल्पनाओं को ठीक से व्यक्त करने के लिए, शोधकर्ताओं को अनुमान या दावे को शब्दों में गणितीय प्रतीकों में परिवर्तित करने की आवश्यकता होती है इसके लिए मूलभूत संकेत जिनका प्रयोग किया जाता है, इस प्रकार है :

समान है	$=$	अधिक है	$>$
समान नहीं है	$\neq$	कम है	$<$

निराकरणीय (Null) तथा वैकल्पिक (Alternative) परिकल्पनाओं को एक साथ लिखा जाता है तथा निराकरणीय परिकल्पना में 'समान है' के चिन्ह का प्रयोग किया जाता है जैसा कि नीचे दी तालिका में दर्शाया गया है (यहाँ k किसी विशिष्ट संख्या को निरूपित करता है)

द्वि-पुच्छ परीक्षण	**दायाँ-पुच्छ परीक्षण**	**बायाँ-पुच्छ परीक्षण**
$H_0 = k$	$H_0 = k$	$H_0 = k$
$H \neq k$	$H_0 > k$	$H_0 < k$

परिकल्पनाओं को तय करने के पश्चात् शोधकर्ता अध्ययन की रूपरेखा तैयार करता है। शोधकर्ता उपयुक्त सांख्यिकीय परीक्षण चुनता है, सार्थकता का एक उपयुक्त स्तर चुनता है तथा अध्ययन के संचालन की एक योजना तैयार करता है। उदाहरण के

लिए, स्थिति A में, शोधकर्ता मरीज़ों का एक प्रतिदर्श लेगा जिन्हें वह औषधि दी जाएगी जिसका परीक्षण करना है। औषधि का असर होने तक का उचित समय देकर, शोधकर्ता प्रत्येक रोगी की स्पंद दर को नापेगा।

यह ध्यान रहे कि जब किसी समष्टि में से एक विशिष्ट आकार के प्रतिदर्श चुने जाते हैं तो इन प्रतिदर्शों के माध्य, समष्टि माध्य के अनुसार बदलेंगे तथा जब प्रतिदर्श आकार 30 या उससे अधिक होगा तो प्रतिदर्श माध्यों का बंटन लगभग प्रसामान्य होगा। अतः, यदि निराकरणीय परिकल्पना सत्य हो, तो भी अधिकतर स्थितियों में, रोगियों के प्रतिदर्श की स्पंद दर का माध्य, समष्टि माध्य 82 स्पंद/बीट्स प्रति मिनट के ठीक बराबर नहीं होगा।

अब दो संभावनाएँ हैं या तो निराकरणीय परिकल्पना सत्य है, और प्रतिदर्श माध्य तथा समष्टि माध्य के बीच का अंतर संयोगवश है; या फिर निराकरणीय परिकल्पना असत्य है, अर्थात् प्रतिदर्श एक ऐसी समष्टि से लिया गया है जिसका माध्य 82 स्पंद/बीट्स प्रति मिनट न होकर कोई अन्य मान है जो हमें ज्ञात नहीं है। प्रतिदर्श माध्य, समष्टि माध्य से जितना दूर होगा, हमारे पास निराकरणीय परिकल्पना को अस्वीकार करने का उतना ही ठोस साक्ष्य होगा। प्रतिदर्श के एक ऐसी समष्टि से होने की जिसका माध्य 82 है, प्रायिकता कम होती चली जाती है यदि दोनों माध्यों के बीच दूरी या उनके बीच के अंतर का निरपेक्ष मान बढ़ता जाए। यदि प्रतिदर्श की औसत पल्स दर, मान लीजिए 83 है तो शोधकर्ता संभवतः इस निष्कर्ष पर पहुँचेगा कि यह अंतर संयोगवश है तथा वह निराकरणीय परिकल्पना को अस्वीकार नहीं करेगा। परंतु, यदि प्रतिदर्श माध्य, मान लीजिए 90 आए तो इसकी संभावना बहुत अधिक है कि शोधकर्ता इस निष्कर्ष पर पहुंचेगा कि औषधि का प्रयोग करने वालों की पल्स दर में वृद्धि हुई है और निराकरणीय परिकल्पना को अस्वीकार कर देगा। अब प्रश्न यह है कि परिकल्पना के स्वीकार या अस्वीकार होने की बीच की रेखा कहाँ पर है? यह निर्णय शोधकर्ता की भावना या सहजबोध पर निर्भर नहीं होता, यह निर्णय एक सांख्यिकी निर्णय है। अर्थात् यह निर्णय महत्त्वपूर्ण/सार्थक होना चाहिए और केवल संयोग पर आधारित नहीं होना चाहिए। यहीं पर सांख्यिकीय परीक्षणों तथा सार्थकता स्तर की संकल्पनाओं के प्रयोग की आवश्यकता होती है।

इस प्रकार के सांख्यिकीय परीक्षण में, माध्य का परिकलन प्रतिदर्श से प्राप्त आंकड़ों के आधार पर किया जाता है तथा इसकी तुलना समष्टि माध्य से की जाती है। तत्पश्चात् निराकरणीय परिकल्पना को अस्वीकार करने या न करने इस सांख्यिकी परीक्षण से प्राप्त मान के आधार पर किया जाता है। यदि अंतर महत्त्वपूर्ण/सार्थक होता है, तो निराकरणीय परिकल्पना को अस्वीकार कर दिया जाता है। यदि यह अंतर महत्त्वपूर्ण नहीं होता तो निराकरणीय परिकल्पना को अस्वीकार नहीं किया जाता। परिकल्पना-परीक्षण में चार संभव परिणाम हो सकते हैं। वास्तविकता में, निराकरणीय परिकल्पना सत्य भी हो सकती है और असत्य भी तथा एक प्रतिदर्श से प्राप्त आंकड़ों के आधार पर लिया गया निर्णय इसे अस्वीकार करना भी हो सकता है तथा अस्वीकार न करना भी। नीचे दिए हुए चित्र में चार संभव परिणाम दिए गए हैं। ध्यान दे कि सही उत्तर प्राप्त करने की दो संभावनाएँ हैं तथा गलत उत्तर प्राप्त करने की भी दो संभावनाएँ हैं।

	H_0 सत्य है	H_0 असत्य है
H_0 अस्वीकार किया गया	प्रकार I की त्रुटि α	सही निर्णय $1-\beta$
H_0 अस्वीकार नहीं किया गया	सही निर्णय $1-\alpha$	**प्रकार II** की त्रुटि β

यदि निराकरणीय परिकल्पना सत्य है तथा वह अस्वीकार कर दिया जाए तो इस त्रुटि को प्रकार 1 की त्रुटि कहते हैं। उदाहरण के लिए, स्थिति A में यह संभव है कि औषधि सभी समष्टि के सभी उपयोगकर्ताओं की पल्स दर/स्पंद दर में कोई महत्त्वपूर्ण परिवर्तन न करती हो, परंतु संयोग से, प्रतिदर्श में लिए गए व्यक्तियों की स्पंद दर में परिवर्तन कर दे। इस स्थिति में शोधकर्ता निराकरणीय परिकल्पना को अस्वीकार कर देगा जबकि वास्तव में वह सही है, अतः α प्रकार 1 की गलती/त्रुटि हो जाएगी। दूसरी ओर, यह भी हो सकता है कि औषधि प्रतिदर्श में लिए गए व्यक्तियों की स्पंद दर में परिवर्तन न करे, परंतु जब समष्टि के अन्य व्यक्तियों को यह औषधि दी जाए तो उनकी स्पंद दर को काफी बढ़ा दे या कम कर दे। इस स्थिति में शोधकर्ता, प्रतिदर्श में प्राप्त आंकड़ों के आधार पर, निराकरणीय परिकल्पना को अस्वीकार नहीं करेगा। अतः यह एक प्रकार II की त्रुटि हो जाएगी।

स्थिति B में, यह संभव है कि योजक, समष्टि की वाहन बैटरियों का जीवनकाल में कोई ध्यान देने योग्य वृद्धि न करें, परंतु चुने गए प्रतिदर्श की बैटरियों के जीवनकाल में महत्त्वपूर्ण वृद्धि दर्शाए। ऐसे में निराकरणीय परिकल्पना अस्वीकृत हो जाएगी। स्थिति B में यह संभव है कि योजक समष्टि में वाहन बैटरियों का जीवनकाल अधिक न बढ़ाए परंतु, प्रतिदर्श में बैटरियों का जीवनकाल बढ़ा दे। ऐसे में, निराकरणीय परिकल्पना अस्वीकार हो जाएगी जबकि यह सही होगी। यह एक *प्रकार-I की त्रुटि* होगी। दूसरी ओर, यह भी संभव है कि योजक प्रतिदर्श में चुनी गई बैटरियों पर कोई प्रभाव न दर्शाए, परंतु जब इसे समष्टि की बैटरियों में व्यापक रूप से प्रयोग किया जाए तो वह उनका जीवनकाल पर्याप्त रूप से बढ़ा दें। ऐसे में शोधकर्ता, प्रतिदर्श से प्राप्त जानकारी के आधार पर निराकरणीय परिकल्पना को अस्वीकार कर देगा। अतः, यह *प्रकार-II की त्रुटि* हो जाएगी।

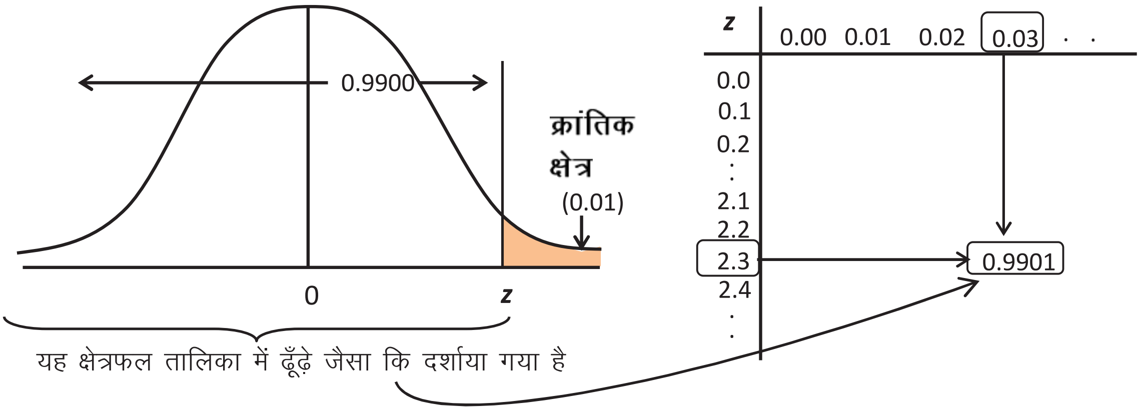
परंतु किसी निराकरणीय परिकल्पना को अस्वीकार करने या अस्वीकार न करने से कुछ सिद्ध नहीं होता। यह निर्णय प्रायिकताओं के आधार पर लिया जाता है। अर्थात्, जब प्रतिदर्श से प्राप्त माध्य तथा परिकल्पिक माध्य में अंतर बढ़ा हो, तो निराकरणीय परिकल्पना संभवतया सत्य नहीं होगा। प्रश्न यह है कि कितना बड़ा अंतर निराकरणीय परिकल्पना को अस्वीकार करने के लिए आवश्यक है? इसके लिए हमें सार्थकता के स्तर का उपयोग करने की आवश्यकता पड़ती है। सार्थकता का स्तर प्रकार-I की त्रुटि होने की अधिकतम प्रायिकता है। इस प्रायिकता को α (ग्रीक अक्षर अल्फा) से व्यक्त यिका जाता है। अर्थात् P (प्रकार I की त्रुटि)= α. प्रकार II की त्रुटि को ग्रीक अक्षर β से व्यक्त किया जाता है अर्थात्, P (प्रकार II की त्रुटि)= β अधिकतर परिकल्पना-

परीक्षण स्थितियों में β का परिकलन सरल नहीं होता; तथापि α तथा β इस प्रकार संबंधित होते हैं कि एक के कम होने से दूसरा बढ़ता है।

साख्यिकीविद्, सामान्यतः, तीन स्वेच्छ (arbitrary) सार्थकता स्तरों के प्रयोग पर सहमत होते हैं : 0.10 स्तर, 0.05 स्तर तथा 0.01 स्तर। अर्थात् यदि निराकरणीय परिकल्पना अस्वीकार की जाती है तो प्रकार I की त्रुटि होने की प्रायिकता 10%, 5% तथा 1% होगी और यह इस पर आधारित है कि सार्थकता के कौन-से स्तर का प्रयोग किया गया है। दूसरे शब्दों में, इस तथ्य को हम इस प्रकार रख सकते हैं : जब $\alpha = 0.10$ है, तो एक सही निराकरणीय परिकल्पना के अस्वीकृत किए जाने की संभावना 10% है; जब $\alpha = 0.05$ है, तो एक सही निराकरणीय परिकल्पना के अस्वीकृत किए जाने की संभावना 5% है तथा जब $\alpha = 0.01$ है, तो एक सही निराकरणीय परिकल्पना के अस्वीकृत किए जाने की संभावना 1% है; एक परिकल्पना-परीक्षण स्थिति में, शोधकर्ता यह तय करता है कि सार्थकता के कौन-से स्तर का प्रयोग करना है। यह कोई भी स्तर हो सकता है और इसका निर्धारण प्रकार I की त्रुटि की गंभीरता पर निर्भर करता है। एक सार्थकता स्तर का चयन करने के पश्चात् उपयुक्त परीक्षण की तालिका से एक क्रांतिक मान (critical value) का चयन किया जाता है। उदाहरण के लिए, यदि Z-परीक्षण का प्रयोग किया जा रहा है तो क्रांतिक मान ज्ञात करने के लिए Z-तालिका का प्रयोग किया जाता है। क्रांतिक मान से क्रांतिक/निराकरण क्षेत्र तथा Non-critical क्षेत्र का निर्धारण होता है।

एक एक-पुच्छ परीक्षण (one tailed test) के लिए क्रांतिक मान माध्य के दाईं ओर या बाईं ओर हो सकता है। इसकी स्थिति वैकल्पिक परिकल्पना में प्रयुक्त असमता चिन्ह पर निर्भर करती है। उदाहरण के लिए, स्थिति B में, जहाँ रसायनशास्त्री की रुचि वाहन बैटरियों के जीवनकाल को बढ़ाने में है, वैकल्पिक परिकल्पना H_1: $\mu>36$ है। क्योंकि यहाँ असमता चिन्ह > है, निराकरणीय परिकल्पना केवल तभी अस्वीकृत होगी जब प्रतिदर्श माध्य 36 से बड़ा हो। अतः क्रांतिक मान माध्य के दाईं ओर होना चाहिए। अतः इस परीक्षण को दायाँ-पुच्छ परीक्षण कहा जाता है।

क्रांतिक मान ज्ञात करने के लिए, शोधकर्ता को एक अल्फा स्तर चुनना पड़ता है। स्थिति B में, मान लीजिए कि शोधकर्ता $\alpha = 0.01$ का चयन करता है। अब शोधकर्ता को Z-तालिका से Z का एक ऐसा मान ज्ञात करना होगा जिससे Z मान के दाईं ओर का क्षेत्रफल 1% हो तथा बाईं ओर का क्षेत्रफल 99% हो, जैसा नीचे चित्र 7.7(क) में दर्शाया गया है। अब, शोधकर्ता को Z-तालिका E में 0.9900 के सबसे समीप का क्षेत्रफल मान ढूँढना पड़ेगा जैसा कि चित्र 7.7(ख) में दर्शाया गया है।

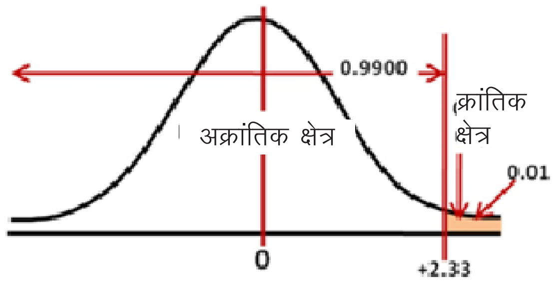


(क) क्रांतिक क्षेत्र

(ख) Z-तालिका E से क्रांतिक मान

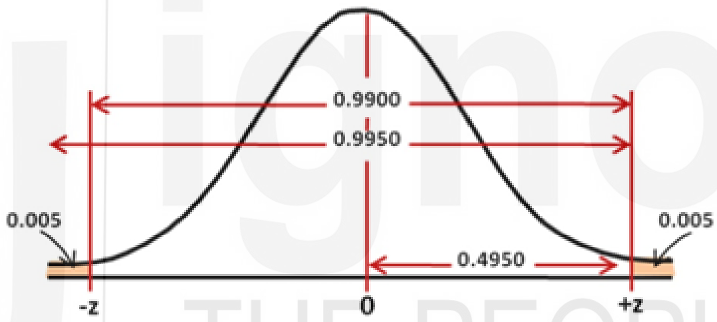
चित्र 7.7

नीचे दिए चित्र 7.8 में क्रांतिक तथा अक्रांतिक क्षेत्र तथा क्रांतिक मान दर्शाए गए हैं।



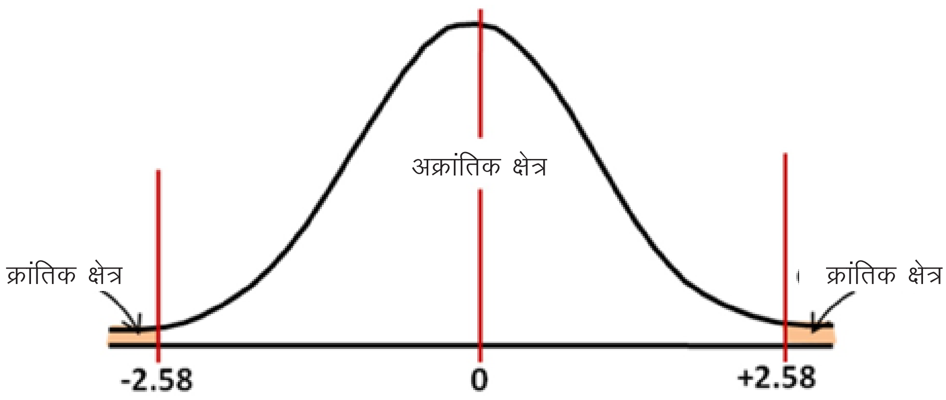
चित्र 7.8

जब एक शोधकर्ता एक द्वि-पुच्छ परीक्षण का प्रयोग करता है, जैसे कि स्थित A में है, तो निराकरणीय परिकल्पना को किसी भी दिशा में, माध्य से अधिक या माध्य से कम, अंतर अधिक होने पर अस्वीकृत किया जा सकता है। परंतु द्वि-पुच्छ परीक्षण में क्रांतिक क्षेत्र को दो बराबर भागों में बांटना पड़ता है। यदि $\alpha = 0.01$ है तो आधा क्षेत्रफल या 0.005, माध्य के दाईं ओर तथा आधा क्षेत्र माध्य के बाईं ओर होना चाहिए, जैसा कि चित्र 7.9 में दर्शाया गया है।



चित्र 7.9

इस स्थिति में, दाईं ओर का Z मान, क्षेत्रफल 0.0050 के संगत Z मान ढूँढकर ज्ञात किया जाता है। यह Z मान −2.57 तथा −2.58 के मध्य में क्षेत्रफलों 0.0049 तथा 0.0051 के संगत स्थित है। −2.57 तथा −2.58 का औसत $[(-2.57) + (-2.58)] / 2 = -2.575$ है। अतः, यदि Z मान 3 दशमलव स्थानों तक चाहिए तो −2.575 का प्रयोग किया जाता है और यदि Z के मान को 2 दशमलव स्थानों तक ही रखा है तो −2.58 का प्रयोग किया जा सकता है। दाईं ओर, 0.99 + 0.005 अर्थात् 0.9950 के संगत Z मान ज्ञात करने की आवश्यकता है। पुनः यह मान 2.57 तथा 2.58 के मध्य में आता है, अतः −2.575 या 2.58 का प्रयोग किया जा सकता है। क्रांतिक मान −2.58 तथा +2.58 है जैसा कि चित्र 7.10 में दर्शाया गया है।

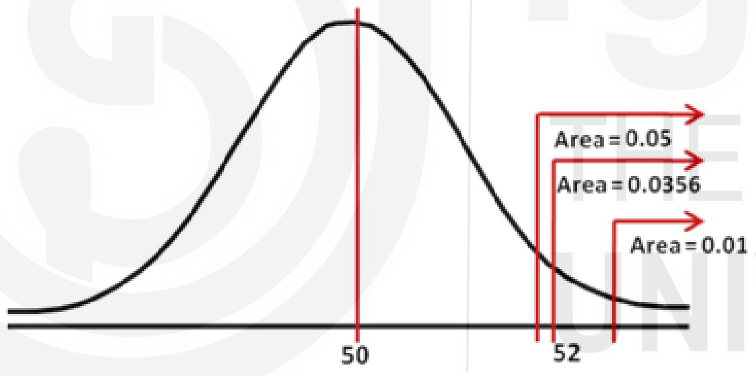


चित्र 7.10

परिकल्पना परीक्षण की p-मान विधि

सामान्यतः, हम परिकल्पना का परीक्षण α स्तरों 0.05 या 0.01 या फिर 0.10 पर करते हैं। लेकिन कई कंप्यूटर स्टैटिस्टिकल पैकेजिस् में परिकल्पना परीक्षण के लिए एक p-मान दिया होता है। यह p-मान (या प्रायिकता मान) एक प्रतिदर्श के लिए प्रतिदर्शज (जैसे कि माध्य) या वैकल्पिक परिकल्पना की दिशा में एक अधिक चरम प्रतिदर्शज प्राप्त करने की प्रायिकता है, जबकि निराकरणीय परिकल्पना सही/सत्य हो। दूसरे शब्दों में, p-मान, मानक प्रसामान्य बंटन वक्र के नीचे का या किसी अन्य वक्र के नीचे का यदि कोई और सांख्यिकीय परीक्षण का प्रयोग किया जा रहा हो वास्तविक क्षेत्रफल है, जो कि विशिष्ट प्रतिदर्शज या एक अधिक चरम प्रतिदर्शज के प्राप्त होने की प्रायिकता को निरूपित करता है जबकि निराकरणीय परिकल्पना सही हो।

उदाहरण के लिए, मान लीजिए एक वैकल्पिक परिकल्पना H_1: $\mu > 50$ है तथा एक प्रतिदर्श का माध्य $\bar{X} = 52$ है। यदि कंप्यूटर एक सांख्यिकी परीक्षण का p-मान 0.0356 प्रिंट करता है और यदि एक दिए हुए प्रतिदर्श माप तथा मानक विचलन के लिए वास्तविक समष्टि माध्य 50 है, तो प्रतिदर्श माध्य 52 या अधिक प्रापत करने की प्रायिकता 0.0356 है। P-मान और α मान के बीच संबंध की व्याख्या इस प्रकार की जा सकती है। P =0.0356 के लिए निराकरणीय परिकल्पना $\alpha = 0.05$ में अस्वीकृत हो जाएगी परंतु $\alpha = 0.01$ पर नहीं। यह नीचे चित्र 7.11 में स्पष्ट किया गया है।



चित्र 7.11

जब परिकल्पना परीक्षण द्वि-पुच्छ हो तो, एक पुच्छ के क्षेत्रफल को दोगुना कर देना चाहिए। एक द्वि-पुच्छ के लिए यदि $\alpha = 0.05$ है तथा एक पुच्छ में क्षेत्रफल 0.0356 है, तो p-मान 2(0.0356) = 0.0712 होगा। संक्षेप में, यदि p-मान α से कम हो, तो निराकरणीय परिकल्पना को अस्वीकार करें। यदि p-मान α से बड़ा हो तो निराकरणीय समीकरण को अस्वीकार न करें। Z-परीक्षण के लिए p-मान, मानक प्रसामान्य बंटन तालिका (Z-तालिका) के प्रयोग से ज्ञात किया जा सकता है। पहले Z-परीक्षण मान के संगत, मानक प्रसामान्य बंटन वक्र के नीचे का क्षेत्रफल ज्ञात कीजिए। एक बाएं-पुच्छ परीक्षण के लिए, तालिका में दिए क्षेत्रफल का प्रयोग करें; एक दाएं-पुच्छ परीक्षण के लिए 1.0000 – (घटा) तालिका में दिए क्षेत्रफल का प्रयोग करें। एक द्वि-पुच्छ परीक्षण के लिए p-मान ज्ञात करने के लिए पुच्छ में प्राप्त क्षेत्रफल को दोगुना करें।

विश्वास्यता अंतराल तथा परिकल्पना परीक्षण

विश्वास्यता अंतराल तथा परिकल्पना परीक्षण परस्पर संबंधित हैं। जब किसी परिकल्पना परीक्षण स्थिति में निराकरणीय परिकल्पना अस्वीकृत होती है तो उसी सार्थकता स्तर का प्रयोग करते हुए प्राप्त होने वाले माध्य के विश्वास्यता अंतराल में परिकल्पित माध्य सम्मिलित नहीं होगा। इसी प्रकार जब निराकरणीय परिकल्पना अस्वीकृत नहीं होती, माध्य के इसी विश्वास्यता अंतराल में परिकल्पित माध्य सम्मिलित होगा। विश्वास्यता अंतराल की संकल्पना तथा इसकी रचना की चर्चा इसी इकाई के भाग 7.4 में की गई।

प्रकार II की त्रुटि तथा एक परीक्षण की घात (Type II Error and the Power of a Test)

जैसा कि ऊपर कहा गया है, परिकल्पना परीक्षण में दो प्रकार की त्रुटियों की संभावना होती है : प्रकार I की त्रुटि तथा प्रकार II की त्रुटि। प्रकार I की त्रुटि तब होती है जब निराकरणीय परिकल्पना सही हो परंतु उसे अस्वीकृत कर दिया जाए। एक सार्थकता स्तर के चयन से जैसे कि 0.05 या 0.01, शोधकर्ता प्रकार I की त्रुटि होने की प्रायिकता का निर्धारण कर सकता है। दूसरी ओर यदि निराकरणीय परिकल्पना को अस्वीकृत न किया जाए तो या तो यह परिणाम ठीक होगा या प्रकार II की त्रुटि हो सकती है। प्रकार II की त्रुटि तब होती है जब निराकरणीय परिकल्पना सत्य न हो परंतु फिर भी उसे अस्वीकार न किया जाए। प्रकार II की त्रुटि होने की प्रायिकता को β से व्यक्त किया जाता है।

β का मान ज्ञात करना सरल नहीं होता। यह अनेक कारकों पर निर्भर करता है जिसमें प्रतिदर्श के आकार, समष्टि के मानक विचलन इत्यादि के मान तथा जाँचाधीन प्राचल के परिकल्पिक मान तथा प्राचल के सही मान के बीच का वास्तविक अंतर। शोधकर्ता का नियंत्रण इनमें से केवल दो कारकों पर होता है, α तथा प्रतिदर्श के माप का चयन। समष्टि का मानक चलन भी कभी-कभी ज्ञात होता है या उसका अनुमान लगाया जा सकता है। मुख्य समस्या जो सामने आती है वह है परिकल्पिक प्राचल तथा प्राचल के सही मान के मध्य वास्तविक अंतर ज्ञात करना। यदि यह अंतर ज्ञात हो जाए तो प्राचल का मान भी ज्ञात हो जाएगा और यदि प्राचल ज्ञात हो जाएं तो किसी प्रकार के परिकल्पना परीक्षण की आवश्यकता नहीं होगी। यद्यपि β का मान परिकलित करना कठिन है परंतु इसका अर्थ यह नहीं है कि इसे नज़रअंदाज कर दिया जाए। सामान्यतः शोधकर्ता β का माप न्यूनतम करने की या ऐसा कह सकते हैं कि $1 - \beta$ का मान अधिकतम करने का प्रयास करते हैं जो कि एक परीक्षण की घात/पावर कहलाता है। किसी परीक्षण की घात वास्तविक प्राचल के विभिन्न मानों के लिए परिकलित की जा सकती है तथा उसे आलेखित करके परीक्षण की घात की जाँच की जा सकती है। किसी परीक्षण की घात, प्राचलों में वास्तविक अंतर, यदि कोई हो तो का पता लगाने में परीक्षण की संवेदनशीलता का माप होती है।

किसी परीक्षण की घात एक प्रायिकता है, अतः इसका मान 0 से लेकर 1 तक हो सकता है। घात जितनी अधिक होगी, परीक्षण प्राचलों के बीच वास्तविक अंतर यदि कोई हो तो, का पता लगाने में उतना ही अधिक संवेदनशील होगा। जैसा कि कहा गया है, एक परीक्षण की घात, प्रकार II की त्रुटि होने की प्रायिकता पर निर्भर करती है और क्योंकि β का परिकलन सरल नहीं होता, किसी परीक्षण की घात का

परिकलन सरल नहीं होता। तथापि कुछ दिशा-निर्देश हैं जिनका प्रयोग किसी परीक्षण की घात से संबंधित सांख्यिकीय अध्ययन के संचालन के लिए किया जा सकता है।

- जब आप किसी सांख्यिकीय अध्ययन का संचालन कर रहे हों तो उस परीक्षण का प्रयोग करें जिसकी घात, प्रदत्त/अध्ययनाधीन आंकड़ों के लिए सबसे अधिक हो।
- किसी परीक्षण की घात बढ़ाने का एक तरीका यह भी है कि प्रतिदर्श का आकार बड़ा चुना जाए। एक बड़े आकार का प्रतिदर्श, माध्य का मानक विचलन को कम कर देगा जिसके फलस्वरूप β का मान कम हो जाएगा।

इन दोनों विधियों का प्रयोग पर्याप्त सावधानी के साथ करना चाहिए : α बढ़ाने से पहले शोधकर्ता को प्रकार I की त्रुटि हो जाने के परिणामों पर विचार कर लेना चाहिए। यदि ये परिणाम, प्रकार II की त्रुटि होने से अधिक गंभीर हों, तो α को नहीं बढ़ाना चाहिए। इसी प्रकार, प्रतिदर्श माप बढ़ाने के भी अपने परिणाम हो सकते हैं। इनमें अध्ययन के लिए आवश्यक खर्च में बढ़ोतरी तथा आंकड़ों को सारणीबद्ध करने में लगने वाले समय में वृद्धि सम्मिलित हैं। जब इन परिणामों की संभावना हो तो प्रतिदर्श माप को बढ़ाना व्यावहारिक नहीं होगा। α और β के बीच संबंध तथा यह रहस्योद्घाटन कि ऐसा कोई जादुई तरीका या सांख्यिकीय परीक्षण नहीं है जो यह सुनिश्चित कर सके कि जब H_0 की वैधता के बारे में निर्णय लिया जाए तो वह शत-प्रतिशत सही होगा। चाहे निर्णय H_0 को अस्वीकार करने का हो या अस्वीकार न करने का, दोनों स्थितियों में परिणाम के गलत होने की संभावना है। अतः, हमारा लक्ष्य यही होता है कि प्रकार I तथा प्रकार II, दोनों प्रकार की त्रुटियों की प्रायिकताओं को जितना संभव हो सके कम से कम रखने का प्रयास करें।

बोध प्रश्न 2

1) 'प्राचल' तथा 'प्रतिदर्शज' की परिभाषा दीजिये।

....

....

....

2) एक प्रसामान्य बंटन की विशेषताओं तथा व्यावहारिक समस्याओं में इसके प्रयोगों की व्याख्या कीजिए।

....

....

....

3) z-विचर क्या है तथा सांख्यिकी समस्याओं में इसके अनुप्रयोग क्या हैं?

....

....

....

4) परिकल्पना परीक्षण की विधियाँ कौन-कौन सी हैं?

....

....

....

....

5) परिकल्पना परीक्षण में प्रकार I तथा प्रकार II की त्रुटियों की तथा उनके बीच संबंध तथा समंजन की व्याख्या कीजिए।

....

....

....

....

6) किसी परीक्षण की घात क्या होती है तथा इसका प्रकार II की त्रुटि से क्या संबंध है?

....

....

....

....

....

7.4 आनुमानिक आंकड़े/सांख्यिकी : बिंदु आकलन तथा समष्टि प्राचलों के लिए विश्वास्यता अंतराल तय करना

आनुमानिक सांख्यिकी का एक पहलू आकलन है, जो कि एक प्रतिदर्श से प्राप्त जानकारी से एक प्राचल के आकलन की प्रक्रिया है। *इस भाग में, समष्टि माध्य, अनुपात, प्रसरण तथा मानक विचलन के आकलन की सांख्यिकीय विधियों की चर्चा की गई है।*

7.4.1 समष्टि माध्य के लिए बिंदु आकलन

मान लीजिए एक कॉलेज शिक्षक, इस सेमेस्टर में उसकी कक्षाओं में पढ़ने वाले विद्यार्थियों की सांख्यिकी में औसत अंकों का आकलन करना चाहता है। शिक्षक 100 विद्यार्थियों का एक यादृच्छिक प्रतिदर्श का चयन कर सकता है तथा इन विद्यार्थियों के अंकों का औसत ज्ञात कर सकता है। मान लीजिए यह औसत 65 है जबकि अधिकतम अंक 100 है। इस प्रतिदर्श माध्य से, शिक्षक यह निष्कर्ष निकाल सकता है कि सभी विद्यार्थियों के अंकों का औसत 65 है। इस प्रकार के आकलन को बिंदु आकलन कहते हैं। प्रतिदर्श माप (अर्थात् प्रतिदर्शज) समष्टि मापों (अर्थात् प्राचलों) के

आकलन के लिए प्रयोग किए जाते हैं। ये प्रतिदर्शज आकलक कहलाते हैं। एक आकलक के लिए अच्छा तथा विश्वसनीय आकलक होने के लिए, उसमें निम्नलिखित गुणधर्म होने आवश्यक हैं :

- आकलक एक निष्पक्ष आकलक होना चाहिए अर्थात् दिए हुए माप के प्रतिदर्शों से प्राप्त आकलनों को प्रत्याशित मान या माध्य, आकलित किए जा रहे प्राचलों के बराबर होना चाहिए।
- आकलक संगत/अविरोध होना चाहिए। एक अविरोधी आकलक के लिए जैसे-जैसे प्रतिदर्श का आकार बढ़ता है, आकलक का मान, प्राचल के आकलित मान की ओर जाता है।
- आकलक तुलनात्मक रूप से एक प्रभावशाली/कुशल आकलक होना चाहिए। अर्थात् उन सभी प्रतिदर्शजों में से जो कि एक प्राचल का आकलन करने के लिए प्रयोग किए जा सकते हैं, तुलनात्मक रूप से कुशल आकलक का प्रसरण सबसे कम होता है।

आप पूछ सकते हैं कि केंद्रीय प्रवृत्ति के अन्य माप जैसे कि माध्यक तथा बहुलक, समष्टि माध्य के आकलन के लिए क्यों नहीं प्रयोग किए जाते। इसका कारण यह है कि जब एक ही समष्टि से अनेक प्रतिदर्श चुने जाते हैं प्रतिदर्शों के माध्यों में परिवर्तन अन्य प्रतिदर्शजों (जैसे कि माध्यकों तथा बहुलकों) की अपेक्षा कम होता है। इसलिए, प्रतिदर्श माध्य को समष्टि माध्य का सबसे अच्छा आकलन माना जाता है।

7.4.2 माध्य के लिए विश्वास्यता अंतराल जब समष्टि मानक विचलन (σ) ज्ञात हो

आप यह प्रश्न पूछ सकते हैं कि एक बिंदु आकलन कितना अच्छा होता है? इस प्रश्न का उत्तर यह है कि हमारे पास यह जानने का कोई तरीका नहीं है कि एक विशिष्ट बिंदु आकलन, समष्टि माध्य के कितना समीप है। इस कारण से, सांख्यिकीविद् एक और प्रकार के आकलन की अपेक्षाकृत अधिक पसंद करते हैं, जिसे अंतराल आकलन कहते हैं।

एक प्राचल का अंतराल आकलन, एक अंतराल या मानों का एक परिसर होता है जिसका प्रयोग प्राचल का आकलन करने के लिए किया जाता है। इस आकलन अंतराल में, वह प्राचल जिसका आकलन किया जाना है सम्मिलित हो भी सकता है और नहीं भी।

एक अंतराल आकलन में, प्राचल के लिए निर्दिष्ट किया जाता है कि वह दो मानों के बीच में स्थित है। उदाहरण के लिए, सभी विद्यार्थियों के औसत अंकों का अंतराल आकलन $60 < \mu < 65$ हो सकता है अर्थात् $\pm$ 5 अंक। या तो अंतराल में प्राचल होगा या नहीं। विश्वास्यता की कोटि (जो सामान्यतः प्रतिशत में व्यक्त की जाती है), एक अंतराल आकलन बनाने से पूर्व भी निर्धारित की जा सकती है। उदाहरण के लिए, आप 95% विश्वस्त होना चाह सकते हैं कि अंतराल में वास्तविक समष्टि माध्य सम्मिलित हो। अब एक अन्य प्रश्न उठता है। 95% ही क्यों? 99% या 99.5% क्यों नहीं? यदि आप और अधिक आश्वस्त होना चाहते हैं जैसे कि 99% या 99.5%

विश्वस्त, तो आपको एक बड़ा अंतराल लेना पड़ेगा। इसके लिए अधिक विश्वस्त होने के लिए अंतराल में वास्तविक समष्टि माध्य सम्मिलित हो, हमें अंतराल को बड़ा करना पड़ता है।

एक प्राचल के अंतराल आकलन का विश्वास्यता स्तर प्राचल के अंतराल आकलन में स्थित/सम्मिलित होने की प्रायिकता है, यह मानते हुए कि प्रतिदर्शों की एक बड़ी संख्या चुनी गयी है तथा उसी/एक ही प्राचल पर आकलन प्रक्रिया की पुनरावृत्ति की गई है। विश्वास्यता अंतराल प्राचल का एक विशिष्ट अंतराल आकलन है जिसका निर्धारण एक प्रतिदर्श से प्राप्त आंकड़ों के प्रयोग से तथा आकलन के एक विशिष्ट विश्वास्यता स्तर के प्रयोग से किया जाता है। इस प्रकार प्राप्त अंतराल, विश्वास्यता अंतराल कहलाते हैं। सामान्यतः तीन विश्वास्यता अंतरालों का प्रयोग किया जाता है : 90%, 95% तथा 99% विश्वास्यता अंतराल।

केंद्रीय सीमा प्रमेय पर आधारित संक्षिप्त सह-ज्ञान से उत्पन्न व्याख्या के अनुसार जब प्रतिदर्श का आकार बड़ा होता है, तो समष्टि से लिए गए एक ही आकार के प्रतिदर्शों में से लगभग 95% प्रतिदर्श माध्य, समष्टि माध्य के ±1.96 मानक त्रुटि के अंतराल अर्थात्

$$\mu \pm 1.96 \left(\frac{\sigma}{\sqrt{n}}\right)$$

में स्थित होंगे।

अब, यदि एक विशिष्ट प्रतिदर्श माध्य चुना जाए, मान लीजिए $\bar{X}$ तो अंतराल $\mu \pm 1.96 \left(\frac{\sigma}{\sqrt{n}}\right)$ में $\bar{X}$ स्थित होगा, इसकी प्रायिकता 95% है। इसी प्रकार नीचे दिए अंतराल

$$\bar{X} \pm 1.96 \left(\frac{\sigma}{\sqrt{n}}\right)$$

में μ स्थित होगा, इसकी प्रायिकता 95% है, जैसा कि आगे हम सिद्ध करेंगे। हम इस कथन को इस प्रकार भी व्यक्त कर सकते हैं–

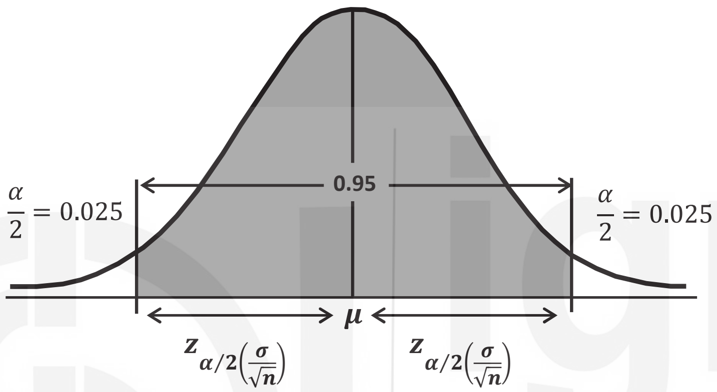
$$P\left\{\bar{X} - 1.96 \left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \bar{X} + 1.96 \left(\frac{\sigma}{\sqrt{n}}\right)\right\} = 0.95$$

अतः, आप 95% विश्वस्त हो सकते हैं कि समष्टि माध्य उसी अंतराल में स्थित होगा, जब चरों के मान समष्टि में प्रसामान्य रूप से बंटित थे। 95% विश्वास्यता अंतराल के लिए प्रयुक्त मान 1.96 है। 99% विश्वास्यता अंतराल के लिए सूत्र में 1.96 के स्थान पर मान 2.58 का प्रयोग किया जाता है। ये मान एक तालिका से प्राप्त होते हैं जो कि मानक प्रसामान्य बंटन पर आधारित है। सांख्यिकी में विश्वास्यता अंतरालों के व्यापक सूत्र में, संकेत $Z_{\alpha/2}$ का प्रयोग होता है। ग्रीक अक्षर α (अल्फा) मानक बंटन वक्र के दोनों पुच्छों के कुल क्षेत्रफल को निरूपित करता है तथा $\alpha/2$ इनमें से प्रत्येक पुच्छ के क्षेत्रफल को निरूपित करता है। α तथा विश्वास्यता स्तर के बीच संबंध यह है कि कथित विश्वास्यता स्तर, $1 - \alpha$ के दशमलव मान का प्रतिशत समतुल्य है तथा विलोमतः। यदि 95% विश्वास्यता अंतराल ज्ञात करना हो, तो $\alpha = 0.05$ होगा क्योंकि 1

$- 0.05 = 0.95$ है या 95%। जब $\alpha = 0.01$ तो $1 - \alpha = 1 - 0.01 = 0.99$, तथा हमें 99% विश्वास्यता अंतराल प्राप्त करते हैं।

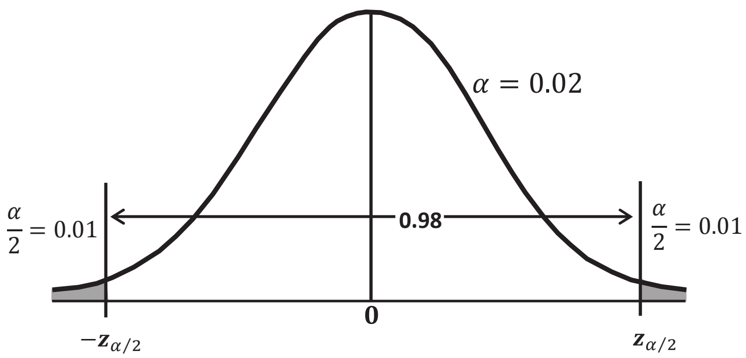
$$\bar{X} - z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \bar{X} + z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

एक 90% विश्वास्यता अंतराल के लिए, $z_{\alpha/2} = 1.65$ होता है; एक 95% विश्वास्यता अंतराल के लिए $z_{\alpha/2}= 1.96$ होता है तथा 99% विश्वास्यता अंतराल के लिए, $z_{\alpha/2}=$ 2.58 होता है। व्यंजक $z_{\alpha/2}(\sigma/\sqrt{n})$ आकलन की अधिकतम त्रुटि कहलाता है (इसे त्रुटि सीमांत भी कहते हैं)। आकलन की अधिकतम त्रुटि, एक प्राचल के बिंदु आकलन तथा प्राचल के वास्तविक मान के बीच का अधिकतम संभव अंतर है। एक विशिष्ट मान, मान लीजिए $\alpha = 0.05$ के लिए, 95% प्रतिदर्श माध्य, समष्टि माध्य के दोनों ओर इसी त्रुटि मान के अंदर स्थित होंगे, जैसा कि नीचे चित्र 7.12 में दर्शाया गया।



चित्र 7.12 : $\overline{X}'$ के मानों का बंटन

क्योंकि कभी-कभी सांख्यिकी में (90, 95 तथा 99 प्रतिशत के अतिरिक्त भी) अन्य विश्वास्यता अंतराल भी प्रयोग किए जाते हैं, अतः $Z_{\alpha/2}$ के मान कैसे ज्ञात किए जाते हैं, इसकी चर्चा करना आवश्यक है। जैसा कि पहले भी कहा गया है, ग्रीक अक्षर α, प्रसामान्य बंटन के दोनों पुच्छों के कुल क्षेत्रफल को निरूपित करता है। α का मान, 1 में से वांछित विश्वास्यता स्तर के दशमलव समतुल्य को घटाकर प्राप्त किया जाता है। उदाहरण के लिए, यदि आप 98% विश्वास्यता अंतराल ज्ञात करना चाहते हैं, जो आप 98% को 0.98 के रूप में लिखेंगे तथा $\alpha = 1 - 0.98$ या 0.02 प्राप्त करेंगे। अब $\alpha/2$, α को z से विभाजित करके प्राप्त किया जाता है। अतः, $\alpha/2$, 0.02/2 अर्थात् 0.01 है। अंततः, $Z_{0.01}$ वह z मान है जो हमें मानक प्रसामान्य बंटन वक्र की दाईं पुच्छ के नीचे का क्षेत्रफल 0.01 के बराबर होगा। यह नीचे दिए चित्र 7.13 में यह दर्शाया गया है।



चित्र 7.13 : मानक प्रसामान्य बंटन

जब मूल चर प्रसामान्य रूप से बंटित हो तथा σ ज्ञात है, मानक प्रसामान्य बंटन का प्रयोग विश्वास्यता अंतराल ज्ञात करने के लिए किया जा सकता है, चाहे प्रतिदर्श का आकार कुछ भी हो।

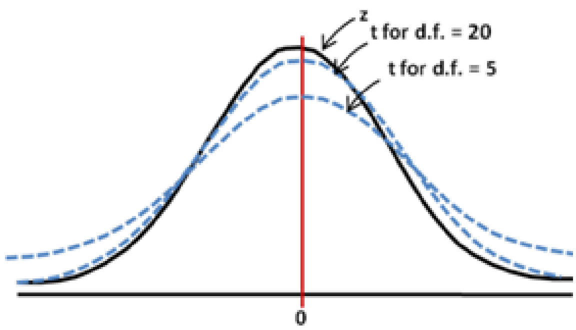
7.4.3 माध्य के लिए विश्वास्यता अंतराल जब समष्टि मानक विचलन (σ) ज्ञात न हो

अधिकतर स्थितियों में σ का मान ज्ञात नहीं होता। अतः, माध्य का आकलन S, अर्थात् प्रतिदर्श के मानक विचलन के प्रयोग द्वारा किया जाता है। S का प्रयोग, विशेषकर तब किया जाता है जब प्रतिदर्श का आकार छोटा हो तथा विश्वास्यता अंतराल को एक दिए हुए स्तर पर रखने के लिए, इस अंतराल में $Z_{\alpha/2}$ के मान से बड़े क्रांतिक मान किए जाते हैं, जैसे कि 95% । ये मान अक्सर t बंटन से लिए जाते हैं। t बंटन मानक प्रसामान्य बंटन से निम्नलिखित प्रकार से अलग है :

- प्रसरण 1 से अधिक है।
- t बंटन वास्तव में स्वतंत्रता की कोटि पर आधारित वक्र-कुल है (नीचे दिए बॉक्स में देखें), जो कि प्रतिदर्श के आकार से संबंधित है।
- जैसे-जैसे प्रतिदर्श का आकार बढ़ता है, t बंटन मानक प्रसामान्य बंटन की ओर जाता है।

स्वतंत्रता की कोटि की संकल्पना

स्वतंत्रता की कोटि उन मानों की संख्या होती है जो एक प्रतिदर्श माप (प्रतिदर्शज) के परिकलन के पश्चात् परिवर्तित होने के लिए स्वतंत्र होते हैं तथा यह खोजकर्ता को यह बताती है कि जब एक बंटन स्वतंत्रता की अलग-अलग कोटियों वाले वक्रों का एक कुल हो, तो किस विशिष्ट वक्र का प्रयोग किया जाए। उदाहरण के लिए, यदि 5 मानों का माध्य 10 है, तो 5 मानों में मान परिवर्तित होने के लिए स्वतंत्र होते हैं। परंतु जब 4 मान चुन लिए गए हों, पाँचवाँ मान एक विशिष्ट संख्या होगी जिससे कुल योग 50 हो जाए क्योंकि $10 \times 5 = 50$ होता है। अतः, स्वतंत्रता की कोटि 5 – 1 = 4 है और यह मान हमें बताता है कि कौन-से t वक्र का प्रयोग किया जाए।



चित्र 7.14

प्रतीक d.f. का प्रयोग स्वतंत्रता की कोटि के लिए किया जाता है। माध्य के किसी विश्वास्यता अंतराल के लिए स्वतंत्रता की कोटि, प्रतिदर्श का आकार में से 1 घटाकर प्राप्त की जाती है। अर्थात् d.f.= $n - 1$। जैसे-जैसे प्रतिदर्श का आकार बढ़ता है, t बंटन मानक प्रसामान्य बंटन की ओर जाता है।

t-बंटन का प्रयोग करने के लिए, प्रतिदर्श, साधारण यादृच्छिक प्रतिदर्श होने चाहिए तथा वह समष्टि जिसमें से प्रतिदर्श लिए गए हैं, वह प्रसामान्य या लगभग प्रसामान्य रूप से बंटित हो अथवा प्रतिदर्श का आकार 30 या उससे अधिक हो।

t-बंटन के प्रयोग से माध्य के आस-पास एक विश्वास्यता अंतराल ज्ञात करने का सूत्र नीचे दिया गया है :

$$\bar{X} - t_{\alpha/2}\left(\frac{S}{\sqrt{n}}\right) < \mu < \bar{X} + t_{\alpha/2}\left(\frac{S}{\sqrt{n}}\right)$$

स्वतंत्रता की कोटि है $n - 1$.

$t_{\alpha/2}$ के मान t-बंटन की तालिका से ज्ञात किए जाते हैं, सबसे ऊपर की पंक्ति जिसको विश्वास्यता अंतराल शीर्षक दिया गया है, इन मानों को ज्ञात करने में प्रयोग की जाती है।

उदाहरण : 95% विश्वास्यता अंतराल के लिए $t_{\alpha/2}$ मान ज्ञात कीजिए जबकि प्रतिदर्श का आकार 22 हो। यहाँ d.f. = 22 – 1 या 21 है। इसलिए बाएं स्तंभ में 21 को खोजें तथा विश्वास्यता अंतराल वाली पंक्ति में 95% खोजें। जहाँ यह पंक्ति तथा स्तंभ मिलते हैं, वहाँ पर दिया हुआ मान 2.080 है जैसा कि नीचे दर्शाया गया है। यह $t_{\alpha/2}$ का अपेक्षित मान है।

t-बंटन तालिका

	विश्वास्यता अंतराल	50%	80%	90%	95%	98%	99%
d.f.	एक पुच्छ α	0.25	0.10	0.05	0.025	0.01	0.005
	द्वि-पुच्छ α	0.50	0.20	0.10	0.05	0.02	0.01
1 2 3 ⋮							
21					2.080	2.518	2.831
$(z)^{\infty}$		0.674	1.282[a]	1.645[b]	1.960	2.326[c]	2.576[d]

7.4.4 अनुपातों के लिए विश्वास्यता अंतराल

अनुपात प्रतिदर्शों या समष्टियों से प्राप्त किए जा सकते हैं। अनुपातिक संकेतन में प्रयोग किए जाने वाले चिन्ह हैं :

P = समष्टि अनुपात

$\hat{p}$ (इसे "p hat" पढ़ा जाता है) प्रतिदर्श अनुपात

एक प्रतिदर्श अनुपात के लिए,

$$\hat{p} = \frac{X}{n} \quad \text{तथा} \quad \hat{q} = \frac{n - X}{n} \quad \text{या} \quad \hat{q} = 1 - \hat{p}$$

जहाँ X = उन प्रतिदर्श इकाइयों की संख्या है जिनमें अपेक्षित विशेषताएँ हैं तथा n = प्रतिदर्श का आकार है।

उदाहरण के लिए, एक अध्ययन में 200 व्यक्तियों से यह पूछा गया कि क्या वे अपनी नौकरी या व्यवसाय से संतुष्ट हैं; 162 ने कहा कि 'हाँ' वे संतुष्ट हैं। इस स्थिति में $n = 200$, $X = 162$ तथा $\hat{p} = X/n = 162/200 = 0.81$।

यह कहा जा सकता है कि इस प्रतिदर्श के लिए 0.81 या जिनका सर्वेक्षण किया गया, उनमें से 81% अपनी नौकरी अथवा व्यवसाय से संतुष्ट थे। प्रतिदर्श अनुपात $\hat{p} = 0.81$ है।

जब $\hat{p}$ तथा $\hat{q}$ दशमलवों या भिन्नों मे दिए हों, तो $\hat{p} + \hat{q} = 1$ होगा। जब $\hat{p}$ तथा $\hat{q}$ प्रतिशतों में दिए हों तो $\hat{p} + \hat{q} = 100\%$ होगा। अतः, हम पाते हैं कि $\hat{q} = 1 - \hat{p}$ या $\hat{p} = 1 - \hat{q}$ है जब $\hat{p}$ तथा $\hat{q}$ दशमलव या भिन्न रूप में है। व्यवसाय संतुष्टि से संबंधित प्रतिदर्श सर्वेक्षण के लिए भी यह देखा जा सकता है कि $\hat{q} = 1 - \hat{p}$, या $1 - 0.81 = 0.19$. है। इसी प्रकार का तर्क समष्टि अनुपातों में भी लागू होता है, अर्थात् $p = 1 - q$, $q = 1 - p$ तथा $p + q = 1$ होता है, जहाँ p तथा q दशमलव या भिन्न रूप में व्यक्त है। जब p तथा q प्रतिशतों के रूप में व्यक्त हों तो $p + q = 100\%$, $p - 100\% = q$ तथा $q = 100\% - p$ होता है।

जैसा कि माध्यों के लिए किया गया, एक समष्टि अनुपात के लिए भी बिंदु तथा अंतराल आकलन भी प्रतिदर्श अनुपात के प्रयोग से प्राप्त किए जा सकते हैं। p (समष्टि अनुपात) एक बिंदु आकलन के लिए, $\hat{p}$ (प्रतिदर्श अनुपात) का प्रयोग किया जाता है। तीन गुणधर्मों के आधार पर, एक अच्छा आकलक निष्पक्ष, अविरोधी तथा तुलनात्मक रूप से कुशल/प्रभावशाली होता है। परंतु जैसा कि माध्यों के लिए भी सत्य है, यह तय करना संभव नहीं है कि p का बिंदु आकलन कितना अच्छा है। अतः, सांख्यिकीविद् एक अनुपात के लिए भी अंतराल आकलन करते हैं तथा इसके लिए एक प्रायिकता निर्दिष्ट करते हैं कि समष्टि अनुपात अंतराल में सम्मिलित होगा। एक विशिष्ट p के लिए विश्वास्यता अंतराल $\hat{p}$ के प्रतिदर्श बंटन पर आधारित होता है। जब प्रतिदर्श का आकार n बड़ा होता है, $\hat{p}$ का प्रतिदर्श बंटन लगभग प्रसामान्य होता है तथा इसका माध्य p तथा मानक विचलन $\sqrt{\frac{pq}{n}}$ होता है। जहाँ $q = 1 - p$ है। अनुपात के आस-पास विश्वास्य अंतरालों को मापदंडों $n\hat{p} \geq 5$ तथा $n\hat{q} \geq 5$ पर खरा उतरना चाहिए। अनुपात के एक विशिष्ट विश्वास्यता अंतराल के लिए सूत्र नीचे दिया गया है :

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

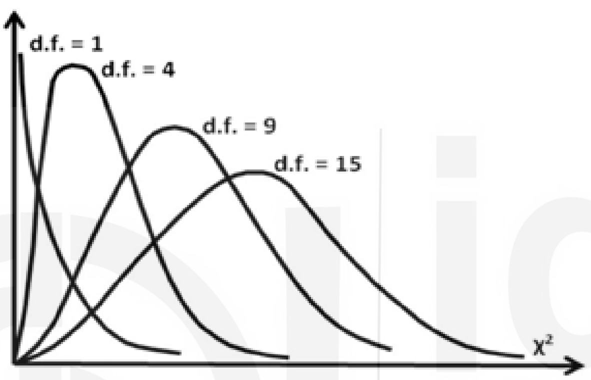
जब $n\hat{p}$ और $n\hat{q}$ 5 के बराबर या अधिक हैं।

7.4.5 प्रसरणों तथा मानक विचलनों के लिए विश्वास्यता अंतराल

सांख्यिकी में, एक चर के प्रसरण तथा मानक विचलन भी उतने ही महत्त्वपूर्ण हैं जितना कि माध्य। उदाहरण के लिए, जब ऐसे उत्पादों का निर्माण किया जा रहा है जो एक साथ फिट होते हैं (जैसे कि पाइप), यह महत्त्वपूर्ण हो जाता है कि उत्पादों

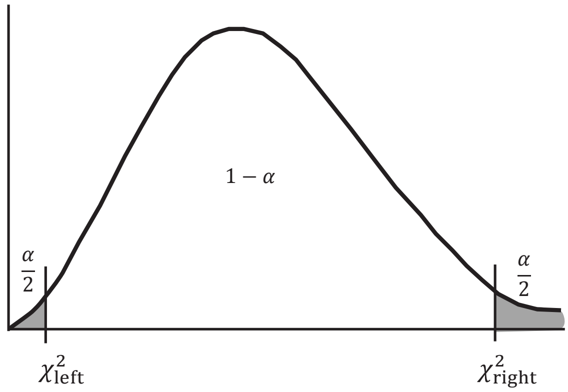
(जैसे कि पाईपों) के व्यासों में परिवर्तन/विविधता जितना संभव हो सके कम-से-कम हो; अन्यथा वे एक-दूसरे में फिट नहीं होंगी और उन्हें फेंकना पड़ेगा। दवाइयों के निर्माण में, गोलियों में दवा के प्रसरण तथा मानक विचलन की यह सुनिश्चित करने में महत्त्वपूर्ण भूमिका होती है कि रोगियों को दवा की उचित मात्रा मिलें। इन कारणों से, प्रसरणों तथा मानक विचलनों के विश्वास्यता अंतराल आवश्यक है।

इन विश्वास्यता अंतरालों को ज्ञात करने के लिए, एक नए सांख्यिकीय बंटन की आवश्यकता है। इसे काई-स्क्वेयर बंटन कहते हैं। काई-स्क्वेयर चर, t चर के समान ही होता है। यह बंटन भी स्वतंत्रता की कोटि की संख्या पर आधारित एक वक्र-कुल है। काई-स्क्वेयर को चिन्ह χ^2 से व्यक्त किया जाता है (एक ग्रीक अक्षर है, जिसे काई ("ki") कहा जाता है)। इस बंटन के कई वक्र नीचे, उनके संगत स्वतंत्रता की कोटि के साथ दर्शाए गए हैं।



चित्र 7.15

काई-स्क्वेयर बंटन $(n-1)\ S^2/\sigma^2$ के मानों से प्राप्त किया जाता है जब यादृच्छिक प्रतिदर्श एक ऐसी प्रसामान्य बंटन वाली समष्टि से चुने गए हों जिसका प्रसरण σ^2 हो। एक काई-स्क्वेयर चर ऋणात्मक नहीं हो सकता तथा ये बंटन दाईं ओर विषम होते हैं। लगभग स्वतंत्रता की 100 कोटि के लिए, काई-स्क्वेयर बंटन कुछ हद तक सममित हो जाते हैं। प्रत्येक काई-स्क्वेयर बंटन के नीचे का क्षेत्रफल 1.00 या 100% के बराबर होता है। एक मानक तालिका स्वतंत्रता की विभिन्न कोटियों के संगत काई-स्क्वेयर बंटन के मान बताती है। ये मान विश्वास्यता अंतरालों के सूत्रों के हर में प्रयोग किए जाते हैं क्योंकि यह बंटन सममित नहीं है, सूत्र में दो भिन्न मानों का प्रयोग किया जाता है। एक मान तालिका के बाईं ओर से तथा दूसरा दाईं ओर से लिया जाता है। उदाहरण के लिए, 95% विश्वास्यता अंतराल के संगत तालिका मान ज्ञात करने के लिए, हमें पहले 95% को दशमलव में परिवर्तित करके उसे 1 में से घटाना पड़ता है $(1 - 0.95 = 0.05)$ । इसके बाद हमें इस प्राप्त संख्या को 2 से विभाजित करना होगा $(\alpha/2 = 0.05/2 = 0.025)$। यह तालिका के दाईं ओर का स्तंभ है जिसमें से χ^2 दाएँ के मान प्राप्त होते हैं। χ^2 बाएँ के मान प्राप्त करने के लिए $\alpha/2$ का मान 1 में से घटाएं $(1 - 0.05/2 = 0.975)$ । नीचे दिए चित्र 7.16 में इस बिंदु को स्पष्ट किया गया है :



अंततः, स्वतंत्रता की कोटि के संगत उपयुक्त पंक्ति खोजें। इसी प्रकार की प्रक्रिया 90% या 99% विश्वास्यता अंतराल के लिए मान ज्ञात करने के लिए अपनाई जाती है।

उदाहरणः यदि n = 25 है तो एक 90% विश्वास्यता अंतराल के लिए χ^2दायाँ तथा χ^2 बायाँ के मान ज्ञात कीजिए χ^2दायाँ ज्ञात करने के लिए 1 में से 0.90 घटाएं तथा प्राप्त मान 1 – 0.90 =0.10 को 2 से भाग करें। ऐसा करने पर हमें 0.05 प्राप्त होता है। χ^2बायाँ ज्ञात करने के लिए 1 में से 0.05 घटाएं : 1 – 0.05 = 0.95 है। अतः, 0.95 तथा 0.05 स्तंभों तो 24 d.f. के संगत पंक्ति का प्रयोग करें जैसा कि नीचे दर्शाया गया है :

काई-स्क्वेयर बंटन

स्वतंत्रता की कोटि	α									
	0.995	**0.99**	**0.975**	**0.95**	**0.90**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**
1										
2										
⋮										
24				13.848			36.415			
				χ^2_{left}			χ^2_{right}			

उत्तर है

$$\chi^2_{right} = 36.415$$

$$\chi^2_{left} = 13.848$$

σ^2 तथा σ के आकलन, क्रमशः S^2 तथा S हैं। प्रसरणों तथा मानक विचलनों के विश्वास्यता अंतराल ज्ञात करने के लिए, यह माना जाता है कि चर प्रसामान्य रूप से बंटित है। विश्वास्यता अंतराल के सूत्र नीचे दिए गए हैं।

एक मानक विचलन के लिए विश्वास्यता अंतराल ज्ञात करने का सूत्र

$$\sqrt{\frac{(n-1)s^2}{\chi^2_{right}}} < \sigma < \sqrt{\frac{(n-1)s^2}{\chi^2_{left}}}$$

एक प्रसरण के लिए विश्वास्यता अंतराल ज्ञात करने का सूत्र

$$\frac{(n-1)s^2}{\chi^2_{right}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{left}}$$

$$d.f. = n - 1$$

अब तक यह स्पष्ट हो गया होगा कि आनुमानिक सांख्यिकी का एक महत्त्वपूर्ण पक्ष आकलन (estimation) है। एक प्राचल के दो प्रकार के आकलन होते हैं : बिंदु आकलन तथा अंतराल आकलन, जिनकी चर्चा हम इस भाग में पर्याप्त रूप से कर चुके हैं।

बोध प्रश्न 3

1) बिंदु आकलन तथा अंतराल आकलन में क्या अंतर है? एक अच्छे आकलन के गुणधर्म क्या हैं?

....

....

....

....

....

2) Z बंटन की व्याख्या कीजिए इस बंटन की मान्यताओं के अंतर्गत तथा माध्य और अनुपात के लिए विश्वास्यता अंतरालों की परिभाषा दीजिए।

....

....

....

....

....

3) t बंटन का प्रयोग कब किया जाता है? t बंटन का प्रयोग करते हुए माध्य तथा अनुपातों के लिए विश्वास्यता अंतराल ज्ञात कीजिए।

....

....

....

....

....

....

4) काई-स्क्वेयर (chi-square) बंटन क्या होता है? मानक विचलन के विश्वास्यता अंतराल ज्ञात करने के लिए इसका प्रयोग कैसे किया जाता है?

....

....

....

....

....

7.5 सार-संक्षेप

इस इकाई में हमने देखा कि किस प्रकार बॉक्स प्लॉट तथा फाइव-नंबर समरी (जो कि अन्वेषी डेटा विश्लेषण के भाग हैं) का प्रयोग डेटा की जाँच यह देखने के लिए किया जाता है कि उससे क्या निष्कर्ष निकलता है।

इसमें प्रसामान्य बारंबरता बंटन के लिए, जिसका प्रयोग वास्तविक व्यावहारिक समस्याओं को हल करने में व्यापक रूप में किया जाता है, केंद्रीय प्रवृत्ति के मापों तथा प्रकीर्णन के महत्त्व की व्याख्या की गई।

डेटा विश्लेषण का एक महत्त्वपूर्ण भाग/पक्ष आकलन है। एक प्राचल के दो प्रकार के आकलन हैं : बिंदु आकलन तथा अंतराल आकलन। एक बिंदु आकलन एक विशिष्ट मान होता है। समष्टियों के प्राचलों का आकलन करने में; समष्टि से एक यादृच्छिक प्रतिदर्श चुनना तथा एक ऐसा प्रतिदर्शज चुनना तथा उसका परिकलना करना, जो प्राचल का सबसे अच्छा आकलन हो, सम्मिलित होता है। एक अच्छा आकलन निष्पक्ष, अविरोधी तथा तुलनात्मक रूप से कुशल होना चाहिए। बिंदु आकलनों के साथ समस्या यह है कि इसकी शुद्धता निर्धारित नहीं की जा सकती। इस कारण से, सांख्यिकीविद् अंतराल आकलन का प्रयोग करना अधिक पसंद करते हैं। प्रतिदर्श मान के आस-पास एक अंतराल का परिकलन करके, सांख्यिकीविद् 95% या 99% (या किसी अन्य प्रतिशत के बराबर) विश्वस्त हो सकते हैं कि उनके आकलन अंतराल में वास्तविक प्राचल सम्मिलित है। विश्वास्यता स्तर शोधकर्ता द्वारा निर्धारित किया जाता है। विश्वास्यता स्तर जितना अधिक होगा, आकलन का अंतराल भी उतना ही अधिक बड़ा (चौड़ा) होगा। माध्य, अनुपात तथा मानक विचलन और प्रसरण के विश्वास्यता अंतराल तय करने के लिए अनेक प्रायिकता बंटनों का प्रयोग किया जाता है। कौन-सा प्रायिकता बंटन अधिक उपयुक्त रहेगा, यह समष्टि मानक विचलन, प्रतिदर्श के आकार, स्वतंत्रता की कोटि इत्यादि की उपलब्धता पर निर्भर करता है।

7.6 संदर्भ ग्रंथादि

1) Blaikie, N. (2003). *Analyzing quantitative data: From description to explanation.* Sage Publication.

2) Bryman, A., & Cramer, D. (1994). *Quantitative data analysis for social scientists* (rev. Taylor & Frances/Routledge.

3) Cramer, D. (2003). *Advanced quantitative data analysis.* McGraw-Hill International.

4) Creswell, J. W. (2002). *Educational research: Planning, conducting, and evaluating quantitative data.* Prentice Hall.

5) Suen, H. K., &Ary, D. (2014). *Analyzing quantitative behavioral observation data.* Psychology Press.

7.7 बोध प्रश्नों के उत्तर अथवा संकेत

बोध प्रश्न 1

1) उपभाग 7.2.1 देखें।

2) उपभाग 7.2.2 देखें।

बोध प्रश्न 2

1) उपभाग 7.3.1 देखें।

2) उपभाग 7.3.3 तथा 7.3.4 देखें।

3) उपभाग 7.3.5 देखें।

4) उपभाग 7.3.6 देखें।

5) उपभाग 7.3.6 देखें।

6) उपभाग 7.3.6 में शीर्षक ''प्रकार-II की त्रुटि तथा एक परीक्षण की घात'' का अध्ययन करें।

बोध प्रश्न 3

1) उपभाग 7.4.1 तथा 7.4.2 देखें।

2) उपभाग 7.4.2 देखें।

3) उपभाग 7.4.3 तथा 7.4.4 देखें।

4) उपभाग 7.4.5 देखें।

इकाई 8 द्वि-विचर आँकड़ों का विश्लेषण (BIVARIATE DATA ANALYSIS)

संरचना

8.0 उद्देश्य

8.1 विषय प्रवेश

8.2 प्रकीर्ण चित्रांकन और सहसंबंध (Scatter Plots and Correlation)

8.3 द्वि-विचर आंकड़ों में संबंध के 'स्तर' का विश्लेषण – 'सहसंबंध' (Analysing Extent of Relationship in Bivariate Data Analysis – Correlation)

8.3.1 सहसंबंध की संकल्पना

8.3.2 सहसंबंध गुणांक

8.3.3 सहसंबंध गुणांक के महत्त्व की कसौटी

8.3.4 सहसंबंध तथा कारण-प्रभाव (उद्भावन)

8.4 द्वि-विचर आंकड़ों में कारण-प्रभाव संबंध का विश्लेषण – प्रतीपगमन (Regression)

8.4.1 सर्वोत्तम संगति रेखा (Line of Best Fit)

8.4.2 प्रतीपगमन रेखा का समीकरण

8.4.3 प्रतीपगमन गुणांक

8.4.4 प्रतीपगमन समीकरणों की पूर्वाकलनीयता

8.4.5 निर्धारण का गुणांक (Coefficient of Determination)

8.5 प्रतीपगमन प्रतिमान की विश्वस्तता– अनुमान की मानक त्रुटि (Standard Error of Estimate)

8.5.1 अनुमान की मानक त्रुटि : संकल्पना और इसका प्राक्कलन

8.5.2 पूर्वानुमेयता अंतराल (Prediction Interval)

8.6 दो औसतों, दो अनुपातों तथा दो विचरणों के बीच अंतरों की जाँच

8.6.1 दो औसतों के बीच अंतरों की समीक्षा : z- तथा t-कसौटी

8.6.2 दो अनुपातों में अंतरों की समीक्षा : z-कसौटी

8.6.3 दो विचरणों में अंतरों की समीक्षा : F-कसौटी

8.7 विचरणों का विश्लेषण

8.8 सार-संक्षेप

8.9 संदर्भ ग्रंथादि

8.10 बोध प्रश्नों के उत्तर अथवा संकेत

8.0 उद्देश्य

द्वि-विचर विश्लेषण पर इस इकाई का अध्ययन करने के बाद आप इन विषयों को समझ कर इनकी व्याख्या करने और

क) इनसे जुड़े प्रश्नों के उत्तर देने में समर्थ हो जाएंगे,

- क्या दो चरों के बीच कोई संबंध है?

- यदि हाँ, तो वह संबंध कितना सशक्त है?
- उनके बीच किस प्रकार का संबंध है?
- इन चरों के संबंध के विषय में किस प्रकार की भविष्यवाणी की जा सकती है?

ख) समझाना

- सहसंबंध तथा कारण-प्रभाव संबंध में भेद क्या है?
- प्रतीपगमन विश्लेषण का प्रयोग कर द्वि-विचर आंकड़ों में कारण-प्रभाव संबंध पर चर्चा, और
- कतिपय मान्यताओं के आधार पर किसी समष्टि से प्राप्त प्रतिदर्श के आधार पर उक्त समष्टि में कारण-प्रभाव संबंध की पूर्वाकलनीयता।

8.1 विषय प्रवेश

इकाई 2 तथा 7 में हमने एकल चर (ऊँचाई, भार, आयु, अंक, मजदूरी आदि) विषयक जानकारी को सारबद्ध करने के लिए वर्णन एवं अन्वेषणात्मक सांख्यिकीय विधियों पर चर्चा की थी, साथ ही किसी समष्टि से निष्कर्षित प्रतिदर्श की विशेषताओं (सांख्यिकों) का प्रयोग कर उस समष्टि के प्राचलों का अनुमान लगाना भी सीखा था। वर्तमान इकाई में, हम द्वि-विचर पर चर्चा करेंगे, अर्थात् दो चरों के युग्मों के संयुक्त आवंटन को समझाने वाली वर्णनात्मक विधियों को जानेंगे। दो चरों के बीच संबंध का अध्ययन करने वाली विधियों को ही हम द्वि-विचर विश्लेषण कहते हैं। किसी संदर्भ विषय की इकाइयों के संभव युग्मों का आवंटन किसी स्वरूप विशेष का हो सकता है। जब हम दो चरों के मानों के युग्मों का अवलोकन या मापन करते हैं तो उन चरों के संयुक्त व्यवहार को समझने की उत्कंठा स्वाभाविक है। निष्कर्षणात्मक सांख्यिकी के एक अन्य आयाम का नाता यह जानने से होता है कि क्या दो या अधिक परिमाणात्मक चरों के मानों के बीच कोई अंतर्संबंध भी विद्यमान है।

उदाहरण: एक व्यापारी यह जानने का उत्सुक होगा कि क्या किसी मास में उसकी बिक्री पर विज्ञापन व्यय का कुछ प्रभाव पड़ा है। शिक्षकों की रुचि यह जानने की होगी कि पढ़ाई के घंटों का शिक्षार्थी के प्राप्तांकों पर प्रभाव क्या है? चिकित्सा शोधकर्ता कैफ़ीन के हृदय पर दुष्प्रभाव या फिर व्यक्ति की आयु और रक्तचाप के बीच संबंध जानने के इच्छुक हो सकते हैं। ये तो उन प्रश्नों के कुछ उदाहरण मात्र हैं जिनके उत्तर हमें द्वि-विचर आंकड़ा विश्लेषण से प्राप्त हो सकते हैं।

भाग 8.0 की सूची के पहले प्रश्न के उत्तर के लिए सांख्यिकीविद् उन चरों के बीच संबंध होने के लिए एक सांख्यिकीय मापक का प्रयोग करते हैं और उसके मान के आधार पर उस संबंध की 'सशक्ता' का आकलन करते हैं। इस मापक को सहसंबंध गुणांक कहते हैं। उदाहरण के लिए, हृदय रोग में योगदान करने वाले अनेक चर हैं, जैसे कि, व्यायाम का अभाव, धूम्रपान, आनुवांशिक कारण, आयु, तनाव एवं खान-पान। इनमें से कुछ अपेक्षाकृत अधिक महत्त्वपूर्ण होते हैं, अतः रोगी की सहायता करने के

इच्छुक चिकित्सक को इस बात का ज्ञान होना चाहिए कि कौन-से कारक अधिक महत्त्वपूर्ण हैं।

इसी प्रकार भाग 8.0 के दूसरे प्रश्न को लेते हैं : एक सामान्य संबंध में तो एक चर को स्वतंत्र या पूर्वाकलनकारी तथा दूसरे को निर्भर या प्रतिक्रियात्मक नाम दिया जाता है। साधारण प्रतीपगमन सरल संबंध निरूपण करता है – अर्थात् यह स्वतंत्र चर का प्रयोग कर निर्भर चर का पूर्वाकलन करता है। एक प्रबंधक यह जानने का उत्सुक हो सकता है कि कंपनी के विक्रय विभाग में बिताए गए वर्षों की संख्या का किसी कर्मी द्वारा की गई बिक्री के परिमाण पर कोई प्रभाव है या नहीं। ऐसे अध्ययन में भी एक सरल संबंध का निरूपण होना है– अनुभव के वर्ष और बिक्री का परिमाण। सरल संबंध धनात्मक या फिर ऋणात्मक भी हो सकता है। एक धनात्मक संबंध की दशा में दोनों चरों में एक साथ वृद्धि या कमी होती है। किंतु ऋणात्मक संबंध की अवस्था में एक चर में वृद्धि के साथ दूसरे के मान में कमी (या फिर इसके विपरीत) ही दिखाई देती है।

अंत में, भाग 8.0 का चौथा प्रश्न पूर्वाकलन (predictions) या भविष्यवाणी के प्रकार से जुड़ा है। प्रायः जीवन के सभी क्षेत्रों में नित्यप्रति कुछ न कुछ पूर्वाकलन होते ही रहते हैं। उदाहरण के रूप में, मौसम, शेयर कीमतों, बिक्री का परिमाण, पैट्रोल की कीमतों या फिर खेलकूद के परिणामों विषयक पूर्व अनुमान की बात की जा सकती है। कुछ पूर्वानुमान अधिक सटीक रहते हैं, कुछ कम – यह संबद्ध चरों के बीच संबंध की शक्ति (strength) का परिणाम रहता है। दो चरों के बीच संबंध जितना सशक्त होगा, पूर्वाकलन भी उतना ही अधिक 'सटीक' होगा।

8.2 प्रकीर्ण चित्रांकन और सहसंबंध (SCATTER PLOTS AND CORRELATION)

सरल द्वि-विचर अध्ययनों के शोधकर्ता किसी अध्ययन वस्तु (या व्यक्ति) के विषय में दो चरों के परिमाणों का मापन करता है तथा यह जानने का प्रयास करता है कि क्या उनके चरों के बीच कोई सहसंबंध है। उदाहरण के रूप में, शोधकर्ता निम्न चरों के उपयुक्त आकार के प्रतिदर्शों के आंकड़े लेकर उनके बीच संबंध का आकलन कर सकता है :

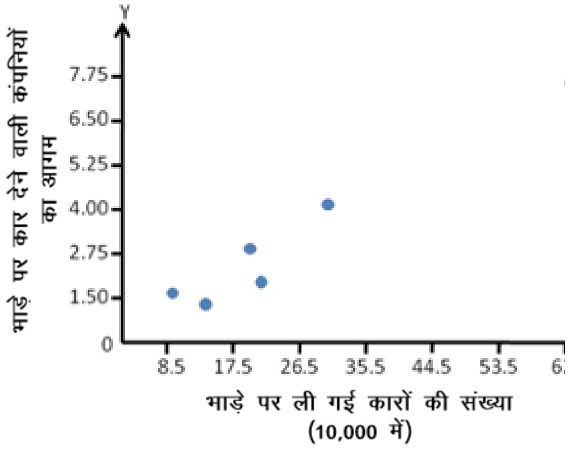
i) भाड़े पर ली गई कारों की संख्या तथा भाड़े पर कार देने वाली कंपनीज़ की की आगम (revenue)

ii) किसी परीक्षा में अनुपस्थिति तथा सकल प्राप्तांक

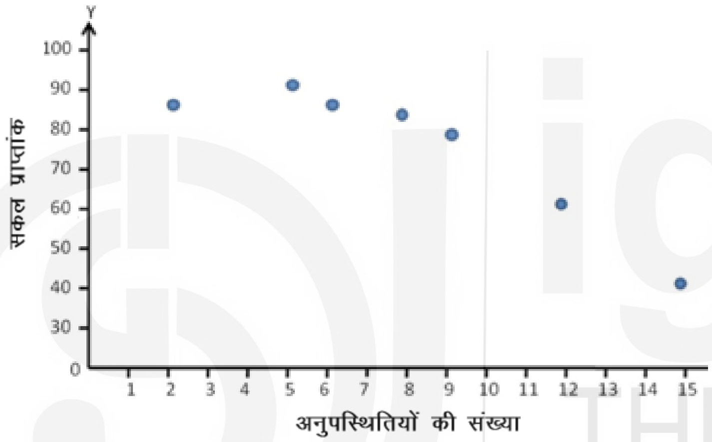
iii) प्रतिसप्ताह व्यायाम में लगाए गए घंटे तथा दूध उपभोग आदि।

ऐसे अध्ययनों में स्वतंत्र एवं निर्भर चरों का निर्धारण कर उनके युग्मों को एक प्रकीर्ण चित्र पर बिंदुओं द्वारा दर्शाया जाता है। यह चित्र x तथा y चरों के युग्मों को बिंदुओं द्वारा दर्शाता है। स्वतंत्र चर के मान क्षैतिज तथा निर्भर चर के मान हम ऊर्द्धव अक्ष

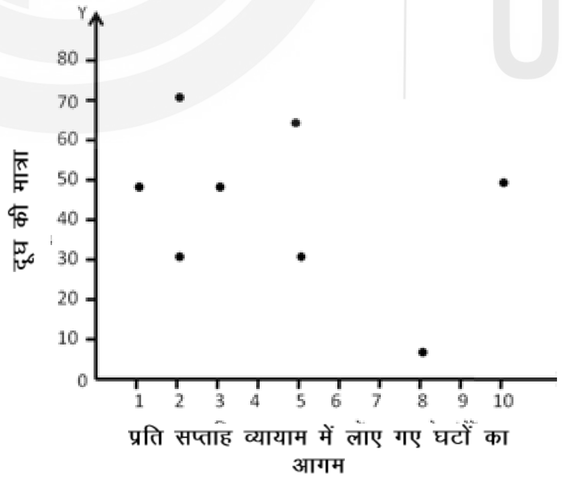
पर दिखा रहे हैं। हम नीचे तीन अध्ययनों की जानकारियाँ अक्षों को अंकित करते हुए दिखा रहे हैं :



चित्र 8.1 : प्रकीर्ण चित्र-1



चित्र 8.2 : प्रकीर्ण चित्र-2



चित्र 8.3 : प्रकीर्ण चित्र-3

इस चित्रांकन के बाद संबंध के प्रकार का निर्धारण होता है। उदाहरण के लिए, प्रकीर्ण चित्र-1 से हमें धनात्मक संबंध का संकेत मिलता है क्योंकि भाड़े पर ली गई कारों की संख्या के साथ-साथ आगम में वृद्धि हो रही है। प्रकीर्ण चित्र-2 एक ऋणात्मक संबंध दर्शाता है क्योंकि जैसे अनुपस्थिति संख्या में वृद्धि होती है प्राप्तांक भी कम हो जाते हैं और अंत में प्रकीर्ण चित्र-3 के बिंदुओं में किसी भी प्रवृत्ति के न होने का संकेत मिलता

है कि उन चरों के बीच किसी प्रकार विशेष का संबंध नहीं है। पहले दो प्रकीर्ण चित्रों में रैखिक (संपूर्ण तो नहीं) संबंध होने के संकेत मिलते हैं।

बोध प्रश्न 1

1) एक प्रकीर्ण चित्र क्या होता है, यह क्या दर्शाता है?

....

....

....

....

....

2) ऊपर उद्धृत तीन प्रकीर्ण चित्र क्या प्रदर्शित कर रहे हैं?

....

....

....

....

....

8.3 द्वि-विचर आंकड़ों में संबंध के 'स्तर' का विश्लेषण – 'सहसंबंध'

8.3.1 सहसंबंध की संकल्पना

द्वि-विचर विश्लेषण में प्रायः आंकड़े अंतर्संबंधित पाए जाते हैं। उदाहरण के लिए, स्वास्थ्य विज्ञान में हम रक्तचाप तथा आयु, किन्हीं पोषकों के उपभोग और वज़न में वृद्धि तथा कुल आय और चिकित्सा व्यय के बीच संबंधों का अध्ययन करते हैं। इस प्रकार के संबंधों का स्तर या उनकी 'सशक्तता' की समीक्षा सहसंबंध विश्लेषण से हो सकती है।

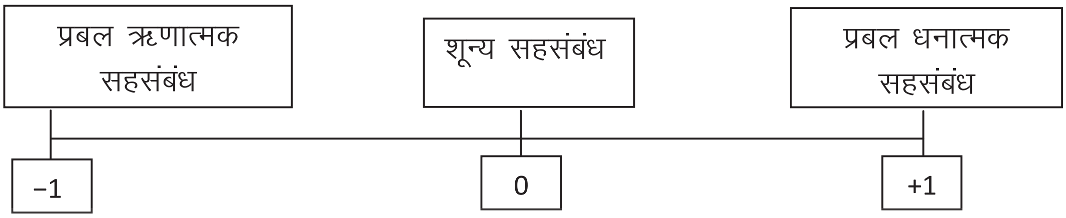
सहसंबंध वह सांख्यिकीय विश्लेषण है जो दो चरों के उतार-चढ़ावों को परस्पर संबंधित संदर्भों में मापता है। यह 'संबंध' शब्द महत्त्वपूर्ण है। यह चरों के बीच किसी संपर्क सूत्र की उपस्थिति का आभास कराता है। यह संबंध की निकटता का द्योतक है। सहसंबंध कारण और प्रभाव रूपी संबंधी नहीं है। कीमत और आपूर्ति, आय और व्यय सहसंबंधित हैं। भौतिक एवं सामाजिक विज्ञानों में सहसंबंध विश्लेषण का प्रायः प्रयोग होता है। अर्थशास्त्री कीमत, परिमाण, लागतों-बिक्री-कीमत, आय और व्यय के बीच संबंधों का अध्ययन सहसंबंध के माध्यम से करते हैं। कहीं अधिक महत्त्व इस बात का है कि यह प्रतीपगमन की संकल्पना के आधार की भी रचना करता है।

8.3.2 सहसंबंध गुणांक

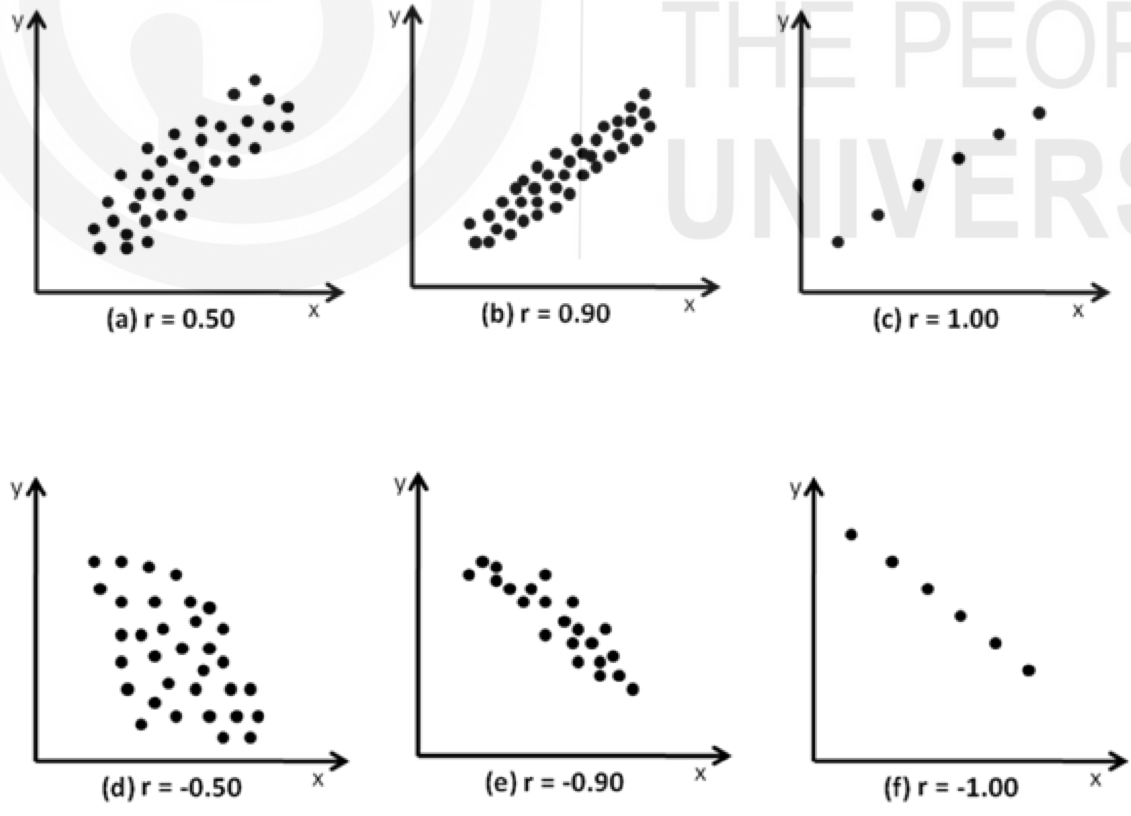
सहसंबंध गुणांक कई प्रकार के होते हैं। इस उपभाग में हम पीयरसन के **गुणित आघूर्ण सहसंबंध गुणांक (PPMC– Pearson Product Moment Correlation**

Coefficient) का वर्णन कर रहे हैं। यह सांख्यिकीविद् कार्ल पीयरसन की देन है जो इस क्षेत्र में शोध के अग्रणी रहे हैं।

प्रतिदर्श आंकड़ों का प्रयोग कर आकलित ***सहसंबंध गुणांक*** *दो चरों के बीच रैखिक संबंध के बल या प्राबल्य* (strength) *तथा दिशा का मापन करता है। प्रतिदर्श सहसंबंध गुणांक को 'r' द्वारा दर्शाया जाता है जबकि समष्टि के इसी माप को ग्रीक भाषा के अक्षर ρ (रो) द्वारा दिखाते हैं।* इस गुणांक का मान –1 से +1 के बीच हो सकता है। यदि दो चरों के बीच प्रबल धनात्मक सहसंबंध हो तो इसका मान +1 के निकट होगा जबकि प्रबल ऋणात्मक सहसंबंध का मान –1 के निकट होता है। यदि उक्त चरों के बीच कोई सहसंबंध नहीं हो तो गुणांक का आकलित मान शून्य के निकट होगा।



हम अगले चित्रों में सहसंबंध गुणांकों के मानों तथा उन चरों के प्रकीर्ण चित्रों के बीच का संबंध दिखा रहे हैं। ध्यान दें कि जैसे-जैसे चित्र a, b तथा c में गुणांक के मान में वृद्धि होती है आंकड़ा युग्म भी संबंध की सबलता को दर्शाने लगते हैं – वे एक सरल रेखा के निकट होते चले जाते हैं। यही बात ऋणात्मक गुणांक मान के लिए चित्र d, e और f में दिखाई गई है। यह भी सहसंबंध के प्राबल्य में क्रमिक वृद्धि दर्शा रहे हैं।



चित्र 8.4: सहसंबंध गुणांक का आकलन

पीयरसन के इस गुणांक के आकलन के लिए कई परस्पर भिन्न सूत्रों का प्रयोग किया जाता है। ये इस प्रकार हैं :

i) $r = \frac{Cov(X,Y)}{\sigma_X.\sigma_Y}$

ii) $r = \frac{\sum xy}{n\,\sigma_X.\sigma_Y}$

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

जहाँ n कुल प्रेक्षण (observation), $x = X - \bar{X}$, $y = Y - \bar{Y}$ हैं।

σ_X और σ_Y क्रमशः X तथा Y के मानक विचलन हैं जिनके सूत्र इस प्रकार हैं :

$\sigma_X = \sqrt{\frac{\sum(X-\bar{X})^2}{n}}$ or $\sqrt{\sum X^2 - \frac{(\sum X)^2}{n}}$ तथा $\sigma_y = \sqrt{\frac{\sum(Y-\bar{Y})^2}{n}}$ अथवा $\sqrt{\sum Y^2 - \frac{(\sum Y)^2}{n}}$

$Cov\ (X,Y)$ X तथा Y के बीच सहविचरण है और इसका सूत्र है : $Cov(X,Y) = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{n}$ अथवा $\sum XY - \frac{\sum X \sum Y}{n}$.

तीसरा सूत्र गणना की दृष्टि से सरल है, क्योंकि यहां मानक विचलनों के अलग से आकलन की ज़रूरत ही नहीं रहती। इस विधि के सोपान इस प्रकार हैं :

1) दोनों शृंखलाओं, X तथा Y के औसत आकलित करें
2) शृंखलाओं की प्रविष्टियों के अपने-अपने औसत से विचरण आकलित करें : $x = X - \bar{X}$, $y = Y - \bar{Y}$
3) इन विचलनों के वर्गित मानों के योग करें जिन्हें क्रमशः $\sum x^2$, $\sum y^2$ द्वारा दर्शाया जाता है।
4) X के विचलन को Y के समकक्ष विचलन से गुणा कर लें। इन गुणनफलों का योग कर उसे 'n' से भाग कर दें।
5) अब निम्न सूत्र में विभिन्न आकलित मान रख दें :

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

इसी सूत्र का सरलीकृत रूप है :

$$r = \frac{n\,(\sum XY) - (\sum X)(\sum Y)}{\sqrt{[n(\sum X^2) - (\sum X)^2][n(\sum Y^2) - (\sum Y)^2]}}$$

उपर्युक्त सहसंबंध गुणांक की ये विशेषताएँ हैं :

- 1 : सहसंबंध गुणक –1 से +1 के बीच रहता है अर्थात् $-1 \le r \le +1$
- 2 : 'r' अक्षमूल (origin) तथा पैमाने से स्वतंत्र रहता है।
- 3 : यह एक निरपेक्ष अंक है, मापन की इकाई से स्वतंत्र है।

- 4 : स्वतंत्र चर अ-सहसंबंधित होते हैं किंतु विलोम अनिवार्य नहीं होता।
- 5 : सहसंबंध गुणंक दो प्रतीपगमन गुणांकों का ज्यॉमितीय औसत है।
- 6 : X तथा Y का सहसंबंध गुणांक सममित (symmetric) होता है, अर्थात् $r_{XY} = r_{YX,}$

किंतु इस गुणांक की कुछ **त्रुटियां या सीमाएं** भी हैं।

- यह सहसंबंध गुणांक तो चरों के बीच रैखिक संबंध ही मानता है, चाहे ऐसा हो, या नहीं।
- चरों के चरम मानों का गुणांक पर अनावश्यक रूप से अधिक प्रभाव रहता है।
- सहसंबंध का होना कारण-प्रभाव संबंध का सत्यापन नहीं करता।

8.3.3 सहसंबंध गुणांक के महत्त्व की कसौटी

जैसा कि हम पहले ही बता चुके हैं, गुणांक का 'परिसर' –1 से +1 के बीच रहता है। यदि मान +1 या –1 के निकट हो तो प्रबल रैखिक संबंध होता है। यदि r का मान शून्य के निकट हो तो यह रैखिक संबंध की दुर्बलता या अनुपस्थिति का संकेत होता है। हम r का अनुमान प्रतिदर्श के आंकड़ों के आधार पर करते हैं। अतः r के शून्य नहीं होने पर दो संभावनाएं होंगी : या तो यह मान इतना उच्च है कि चरों के बीच महत्त्वपूर्ण रैखिक संबंध की बात स्वीकार कर ली जाए, या फिर 'r' का ये मान किसी संयोग का परिणाम मात्र हो सकता है।

इस विषय में किसी निर्णय तक पहुंचने के लिए हम संकल्पना सत्यापन पद्धति का प्रयोग करते हैं जिसमें ये निम्न सोपान होते हैं :

- प्रथम सोपान : संकल्पना निरूपण
- द्वितीय सोपान : निर्णायक मानों का आकलन
- तृतीय सोपान : कसौटी मान का आकलन
- चतुर्थ सोपान : निष्कर्ष निकालना
- पंचम सोपान : परिणाम को सारबद्ध करना

समष्टि सहसंबंध गुणांक का आकलन X तथा Y के सभी संभव युग्मों का प्रयोग कर किया जाता है; इसे ग्रीक अक्षर ρ (रो) द्वारा दिखाते हैं। औपचारिक रूप से ρ किसी समष्टि के सभी संभव (X,Y) युग्मों के आधार पर आकलित किया जाता है। ऐसी अवस्था में उस समष्टि से निष्कर्षित प्रतिदर्श के सहसंबंध गुणांक (r) को ρ का 'अनुमानक' माना जा सकता है, यदि निम्नलिखित शर्तें सत्य हों :

- चर X तथा Y के बीच रैखिक संबंध है
- चर यादृच्छिक हैं
- दोनों चरों का प्रसामान्य यादृच्छिक द्वि-विचर आवंटन है।

द्वि-विचर प्रसामान्य आवंटन का अर्थ है कि (X, Y) के किसी भी युग्म के लिए प्रत्येक X हेतु Y के मानों का एक घटाकर आवंटन होगा। इसी प्रकार प्रत्येक Y के लिए X के मानों का आवंटन भी घटाकर होगा।

संकल्पना सत्यापन में इनमें से एक बात सत्य होती है :

H_0: $\rho = 0$ इसे शून्यक संकल्पना कहते है– इसका अभिप्राय है कि X तथा Y चरों के बीच समष्टि में कोई सहसंबंध नहीं है।

H_1: $\rho \neq 0$ इस वैकल्पिक संकल्पना का अर्थ है कि इन चरों के बीच समष्टि में महत्त्वपूर्ण सहसंबंध है।

किसी स्तर विशेष α पर शून्यक संकल्पना को अस्वीकार किये जाने का अर्थ है कि ρ का मान शून्य से महत्त्वपूर्ण रूप में भिन्न है। यदि संकल्पना अस्वीकार नहीं की जाती तो अर्थ होगा कि ρ का मान शून्य से महत्त्वपूर्ण रूप में भिन्न नहीं है– जो भी अंतर दिखाई दे रहा है वह संभवतः संयोग मात्र है।

सहसंबंध गुणांक की महत्त्वपूर्णता की जाँच की कई विधियाँ विद्यमान हैं किंतु हम यहाँ केवल t-कसौटी पर चर्चा करेंगे। इस t-कसौटी के लिए सूत्र प्रकार है :

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

जहाँ स्वतंत्रता कोटि अंक $n-2$ है।

यहाँ द्विपुच्छ (two tailed) निर्णायक मानों का प्रयोग होता है जो t-तालिका से पाए जाते हैं। यह भी आवश्यक है कि X तथा Y दोनों चर प्रसामान्य रूप से आवंटित समष्टियों से उद्भूत (come) हुए हों।

उदाहरण : निम्न प्रतिदर्श आँकड़ों से सहसंबंध गुणांक का आकलन कर उसकी महत्त्वशीलता की जाँच करें।

कंपनी	कारें (दस हजार में)	आगम (बिलियनों में)
A	63.0	7.0
B	29.0	3.9
C	20.8	2.1
D	19.1	2.8
E	13.4	1.4
F	8.5	1.5

उत्तर : r का आकलन

कंपनी	कार X (दस हजार में)	आगम Y (बिलियन में)	XY	X^2	Y^2
A	63.0	7.0	441.00	3969.00	49.00
B	29.0	3.9	113.10	841.00	15.21
C	20.8	2.1	43.68	432.64	4.41
D	19.1	2.8	53.48	364.81	7.84
E	13.4	1.4	18.76	179.56	1.96
F	8.5	1.5	2.75	72.25	2.25
	$\sum X = 153.8$	$\sum Y = 18.7$	$\sum XY = 682.77$	$\sum X^2 = 5859.26$	$\sum Y^2 = 80.67$

सूत्र में आकलित मान रखकर 'r' का मान ज्ञात करें।

$$r = \frac{n(\sum XY) - (\sum X)(\sum Y)}{\sqrt{[n(\sum X^2) - (\sum X)^2][n(\sum Y^2) - (\sum Y)^2]}}$$

$$= \frac{(6)(682.77) - (153.8)(18.7)}{\sqrt{[(6)(5859.26) - (153.8)^2][(6)(80.67) - (18.7)^2]}} = 0.982$$

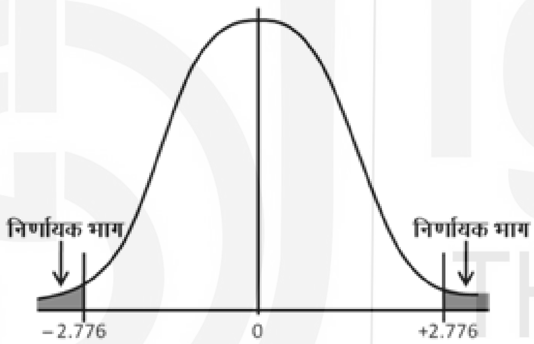
सहसंबंध गुणांक का मान कार भाड़े पर देने वाली एजेंसी के पास कारों की संख्या और उसकी वार्षिक आगम के बीच एक 'सशक्त सहसंबंध' का संकेत दे रहा है।

सहसंबंध गुणांक की महत्त्वशीलता

सोपान 1 : संकल्पना स्पष्टीकरण

$H_0: \rho = 0$ और $H_1: \rho \neq 0$

सोपान 2 : निर्णायक मानों का आकलन करें। क्योंकि $\alpha = 0.05$ और यहाँ $6 - 2 = 4$ स्वतंत्रता कोटि अंक है, तालिका से प्राप्त निर्णायक मान हैं ± 2.776

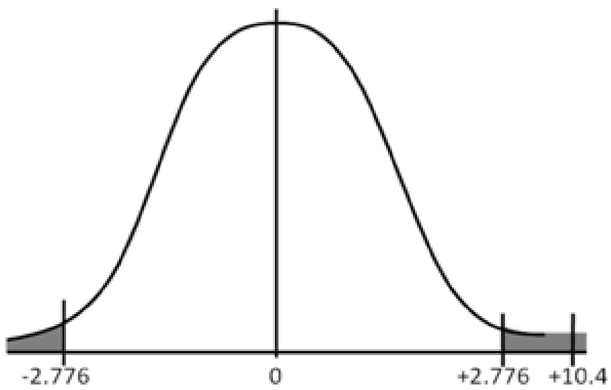


चित्र 8.5

सोपान 3 : कसौटी मान का आकलन

$$t = r\sqrt{\frac{n-2}{1-r^2}} = 0.982\sqrt{\frac{6-2}{1-(0.982)^2}} = 10.4$$

सोपान 4 : यहाँ निर्णय करना है। शून्यक संकल्पना को अस्वीकार कर दें (वह है, $\rho = 0$), क्योंकि कसौटी मान नीचे दिखाए गए निर्णायक क्षेत्र में आता है।



चित्र 8.6

सोपान 5 : परिणाम को सारबद्ध करें। कार भाड़े पर देने वाली एजेंसी के पास कारों की संख्या और उसकी वार्षिक आगम के बीच महत्त्वपूर्ण संबंध है।

8.3.4 सहसंबंध तथा कारण-प्रभाव (उद्भावन)

स्वतंत्र चर X तथा निर्भर चर Y के बीच रैखिक संबंध के स्वरूप को बहुत स्पष्ट रूप से समझ लेना अनिवार्य होगा। जब एक संकल्पना परीक्षण चरों के बीच महत्त्वपूर्ण रैखिक सहसंबंध का संकेत दे रहा हो तो शोधकर्ता को रैखिक संबंध की सभी संभावनाओं पर विचार कर लेना चाहिए। एक शून्यक संकल्पना की किसी मान विशेष हेतु निरस्ति का अर्थ है निम्न पाँच में से एक संभावना का होना :

1) चरों के बीच एक प्रत्यक्ष कारण-प्रभाव संबंध होना। दूसरे शब्दों में, X ही Y का कारक है। उदाहरणः जल किसी पादप की वृद्धि का, विष मृत्यु का और ताप हिम के पिघलने का कारण है।

2) चरों के बीच वस्तुतः कोई विलोम कारण-प्रभाव संबंध है, अर्थात् Y ही चर X का कारक है। उदाहरण : एक शोधकर्ता को प्रतीत होता है कि कॉफी का अत्यधिक उपभोग व्यक्ति को धैर्यहीन बना देता है किंतु यहां शोधकर्ता विलोम की संभाव्यता पर विचार नहीं कर रहा। संभव है कि अधिक अधीर व्यक्ति अपने चित्त को शांत करने के लिए अधिक कॉफी पी रहा हो।

3) उन चरों के बीच प्रकट हो रहा संबंध वस्तुतः किसी तीसरे चर के उन दोनों पर एक साथ प्रभाव का परिणाम भी हो सकता है। यदि ग्रीष्म ऋतु में डूबकर मरने की घटनाओं और शीतल पेय के डब्बों के उपभोग के बीच सहसंबंध सांख्यिकीय दृष्टि से महत्त्वपूर्ण हो सकता है। किंतु इन डूबकर मर जाने की घटनाओं के लिए शीतल पेय के अधिक उपभोग को 'कारण' तो नहीं माना जा सकता। ये दोनों ही चर मौसम के उच्च ताप और आर्द्रता के साथ सहसंबंधित हो सकते हैं।

4) अनेक चरों के बीच अंतर्संबंधों में बहुत जटिलता भी हो सकती है। उदाहरण : एक शोधकर्ता को छात्रों के स्कूल और कॉलेज शिक्षा में प्राप्त अंकों के बीच महत्त्वपूर्ण संबंध दिखाई दे सकता है किंतु बौद्धिक ज्ञान, अध्ययन के घंटे, अभिभावकों के प्रभाव, संप्रेरणा, आयु और शिक्षक वर्ग जैसे अनेक कारक इन पर अपने प्रभाव छोड़ सकते हैं।

5) संबंध संयोग मात्र भी हो सकता है। शोधकर्ता को व्यायामकर्ताओं की संख्या में वृद्धि तथा अपराधियों की संख्या में वृद्धि के बीच महत्त्वपूर्ण सहसंबंध दिखाई दे सकता है। किंतु सहज विचारशीलता ही यह स्पष्ट कर देती है इन दो चरों के मानों के बीच सहसंबंध संयोग मात्र ही होगा।

जब दो चरों के बीच प्रबल सहसंबंध हो तो हमारा उपर्युक्त (3) किसी तीसरे चर के प्रभाव की संभावना की ओर भी संकेत करता है। यदि यह तीसरा चर शोधकर्ता को अज्ञात है— अर्थात् इस पर शोध में ध्यान नहीं दिया गया हो तो इसे एक मंडराते हुए चर का नाम दिया जाता है। शोधकर्ता को ऐसे चरों की पहचान कर उनके प्रभावों को नियंत्रित करने का प्रयास करना चाहिए। एक बार पुनः इस बात को दोहरा लेना आवश्यक है कि उच्च सहसंबंध गुणांक का अर्थ निश्चित रूप से कारण-प्रभाव संबंध नहीं होता। किसी मंडरा रहे चर के प्रभाव या पूर्णतः संयोग मात्र की संभावना को

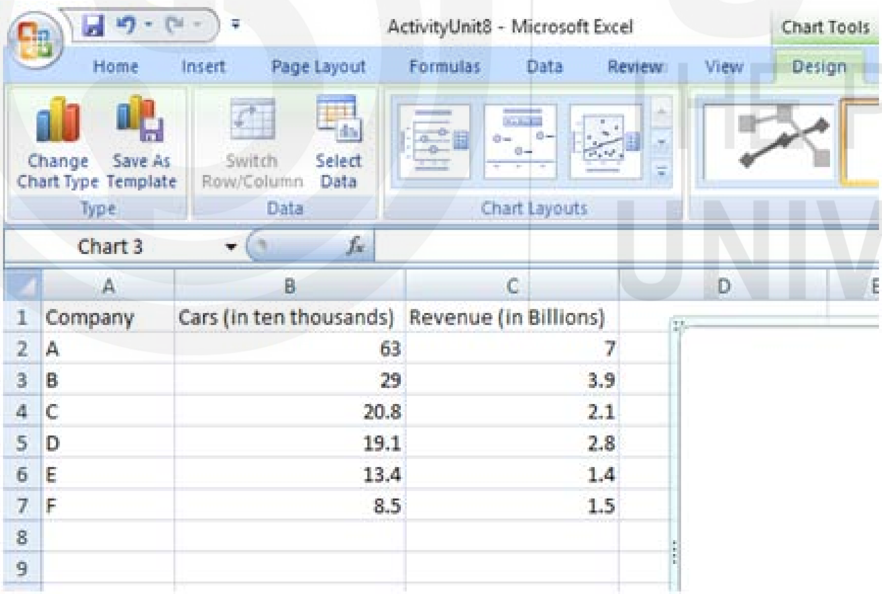
सीधे से नकारना उचित नहीं होगा। साथ ही, यदि एक या दोनों चरों के मान वास्तविक आंकड़े नहीं बल्कि औसत मान हों तो विशेष सावधानी की आवश्यकता रहेगी। औसतों का प्रयोग करना गलत नहीं है – किंतु उनके निष्कर्षों को वैयक्तिक स्तर पर मान्य समझना उचित नहीं होगा – क्योंकि औसत आकलन विधि पहले ही वैयक्तिक स्तर के आंकड़ों के उच्चावचनों का शमन कर चुकी होगी। ऐसी दशा में सहसंबंध गुणांक का मान वास्तविकता से कहीं उच्च हो जाएगा। अतः जब एक शून्यक संकल्पना को नकारा जाता है तो शोधकर्ता को सभी संभावनाओं पर विचार कर 'उपयुक्त' निष्कर्ष का चयन करना चाहिए। ध्यान रहे सहसंबंध का अर्थ निश्चित उद्भावन (causation) नहीं है।

अभ्यास कार्य-1 : प्रकीर्ण चित्रण

उपभाग 8.3.3 के उदाहरण के आंकड़ों का प्रकीर्ण (scatter) चित्र बनाएं।

यह कार्य कैसे करेंगे

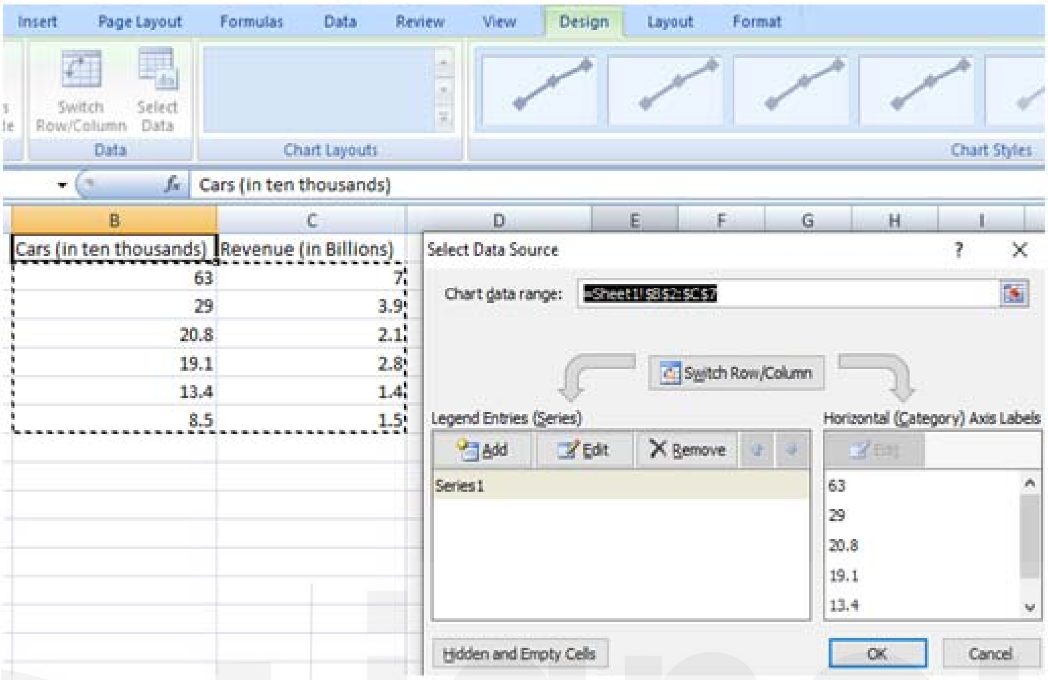
- आंकड़ों को एक स्प्रैडशीट में अंकित करें (देखें चित्र 8.6)
- प्रकीर्ण चित्र बनाने के लिए ***Insert*** दबाकर रिबन के चित्र समूह में से ***scatter*** का चयन करें– उसकी सूची में से ***Scatter with only Markers*** चुनें।
- एक 'खाली' खाका दिखाई देगा, साथ ही Tabs में आपको Chart Tools दिखाई देंगे : Design, Layout और Format (चित्र 8.7)।



चित्र 8.7 : खाली प्रकीर्ण चित्र का अंकन

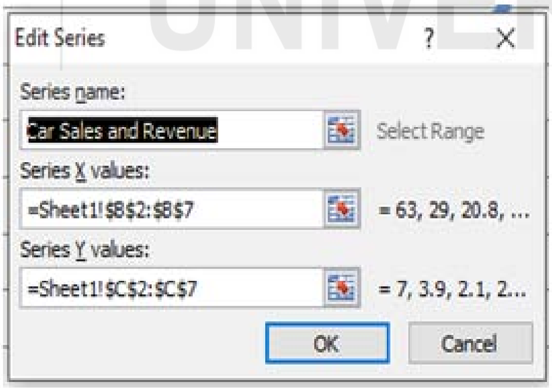
- अब *Design* tab में से *Select Data* विकल्प को *Data Group* में से चुन लें। आपको *Select Data Source* का वार्ता बॉक्स दिखाई देगा जिसमें Chart Data Range का text बॉक्स अधिक प्रकाशित होगा। अपने माउस के संकेतक को तालिका की प्रथम प्रविष्टि दर्शा रहे कोष्ठक (हमारे उदाहरण में B2 जहाँ 63 मान अंकित है) पर ले जाएं। वहां माउस का बायां बटन दबाएं और उसे दबाए हुए ही संकेतक को तालिका की अंतिम प्रविष्टि का मान दिखाने वाले कोष्ठक C-7 तक ले जाएं (इस उदाहरण में यह मान 1.5 है)। अब Chart Data Range में दिखाई

देगा : =*Sheet1!$B$2:$C$7* (देखें चित्र 8.8)। ध्यान दें कि कोष्ठक B2 से C7 तक, जिन्हें चुना गया है, अब बिंदु रेखाओं से घिरे दिखाई देते हैं।

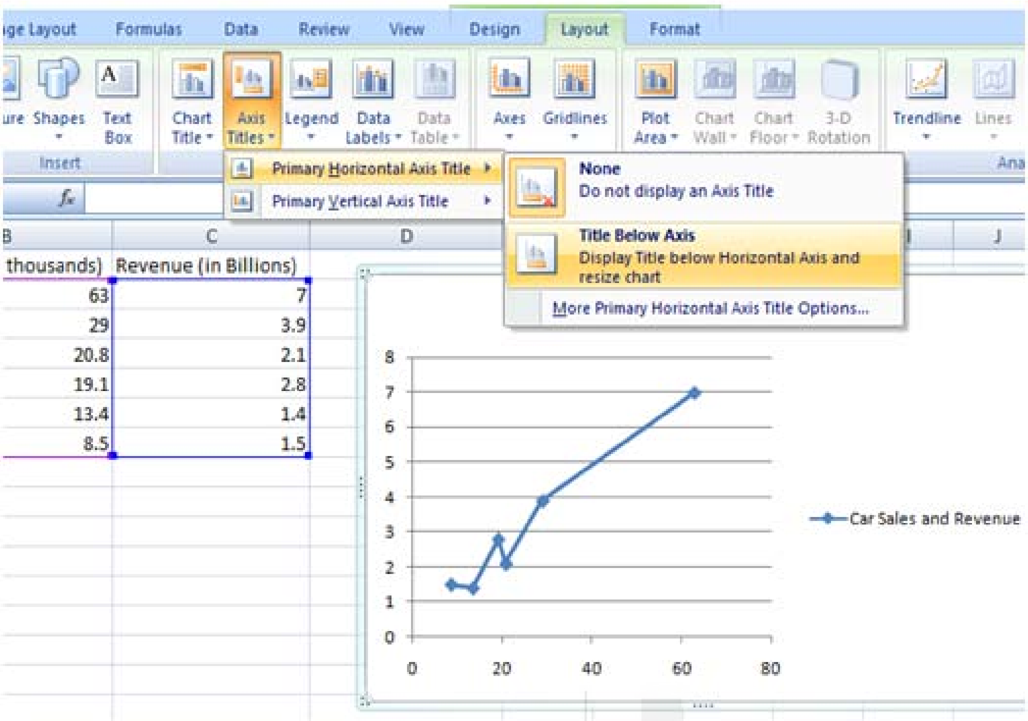


चित्र 8.8 : Data Range B2 से C7 का चयन

- कृपया ध्यान दें यहां *Legend Entries (Series)* Area में *Series1* दिखाई गई है। इसे बदलने के लिए *Edit* बटन दबाएं। अब आपको *Edit Series* वार्ता बॉक्स दिखाई देगा। इस बॉक्स में आप *Series name* text box में "*Car Sales and Revenue*" टंकित करें। एक बार X तथा Y मानों को देख लें और फिर OK बटन दबाएं। आप फिर से चित्र 8.8 पर ही पहुँच जाएंगे– बस अंतर यही होगा कि *Series1* के स्थान पर *Car Sales and Revenue* लिखा दिखाई देगा। अब *Select Data Source* dialog box वार्ता बॉक्स में OK बटन दबाएं।

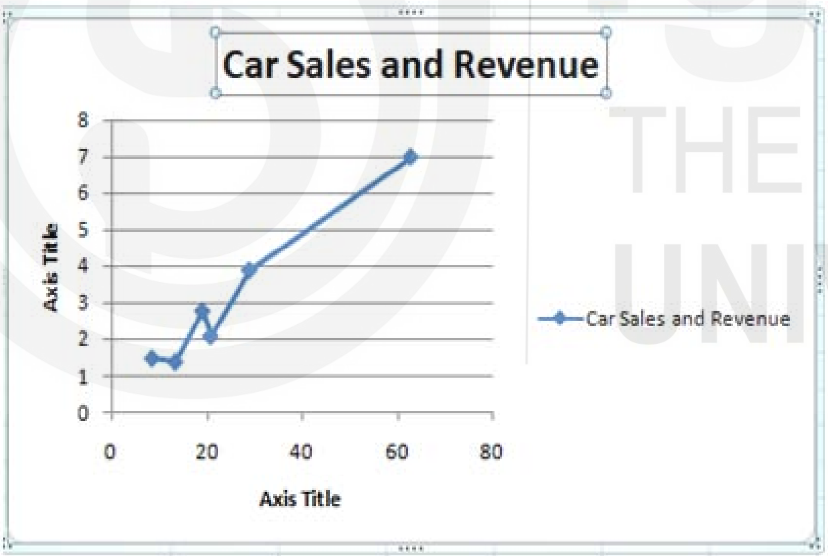


- अब आपको एक प्रकीर्ण चित्र दिखाई देगा जिसमें Title and Legend Entry तथा अक्ष मान भी हैं (देखें चित्र 8.9)। अक्ष शीर्षक अंकित करने के लिए *Layout* tab पर जाकर रिबन में Select *Axis Title* बटन दबाएं। फिर *Primary Horizontal Axis* चुनें और *Title Below Axis* का चयन करें, जैसा कि चित्र 8.9 में दिखाया गया है। इसी प्रकार ऊर्ध्व अक्ष शीर्षक के लिए *Primary Vertical Axis Title* और *Rotated Titles* चुनें। इससे शीर्षक लिखने के लिए स्थान का सृजन हो जाएगा। वास्तविक शीर्षक तो आपको वहां टाइप करने होंगे।



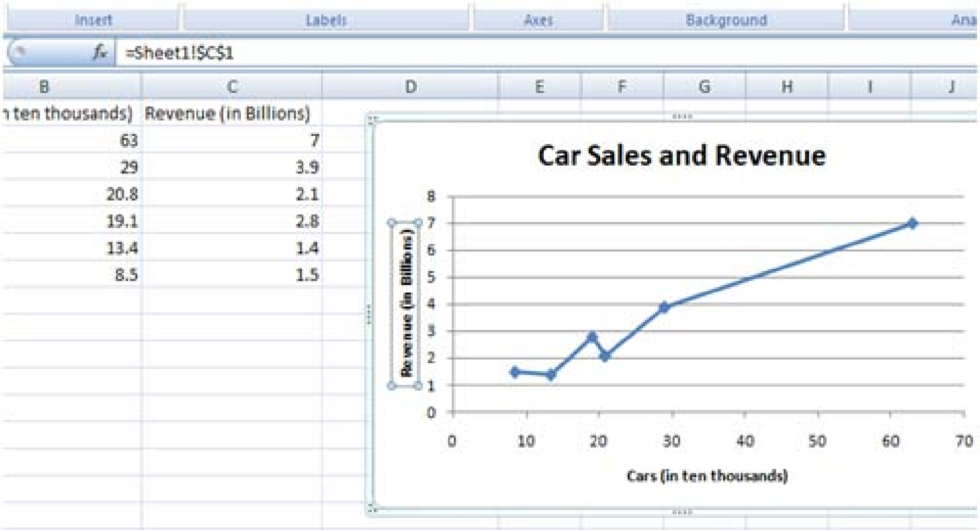
चित्र 8.9 : अक्ष शीर्षक भरने की प्रक्रिया

- दोनों शीर्षक को चुनने के बाद आपका scatter plot इस प्रकार परिवर्तित हो जाएगा।



- आपको केवल वास्तविक शीर्षक ही भरने हैं, Legend नहीं, क्योंकि यहां केवल एक ही श्रृंखला है। अतः Legend को चुनकर उसे delete कर दें। Axis Title चुनें और Formula कोष्ठक में = B1 टाइप करें, यदि कोष्ठक B1 में अक्ष शीर्षक हो (देखें चित्र 8.10)। इसी में हमारे चित्र का अंतिम स्वरूप दिखाई दे रहा है।

- एक बात पर और ध्यान दें। चित्र 8.10 में चित्र में Vertical Axis चुना गया है और संबद्ध मान को Formula bar में *=Sheet1!$C$1* दिखाया गया है। यह कोष्ठक C1 को ही इंगित करता है। इसी में ऊर्ध्व अक्ष का शीर्षक है। आप C1 में यदि कोई परिवर्तन करते हैं तो वह चित्र में शीर्षक के परिवर्तन में दिखाई देने लगेगा।



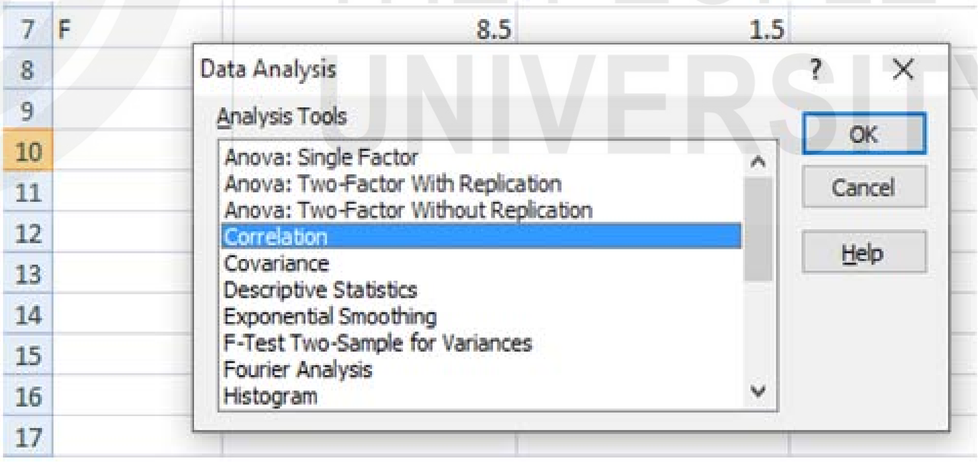
चित्र 8.10

अभ्यास कार्य 2 : अभ्यास कार्य-1 के आंकड़ों में सहसंबंध गुणांक का आकलन करें।

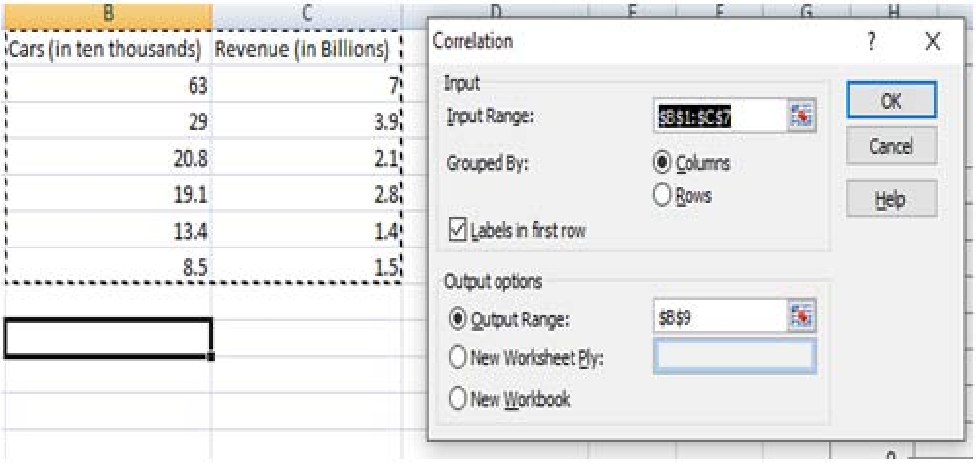
यहाँ चित्र 8.10 में दिखाए गए आंकड़ों के बीच सहसंबंध गुणांक का आकलन कीजिए।

यह कार्य कैसे करेंगे?

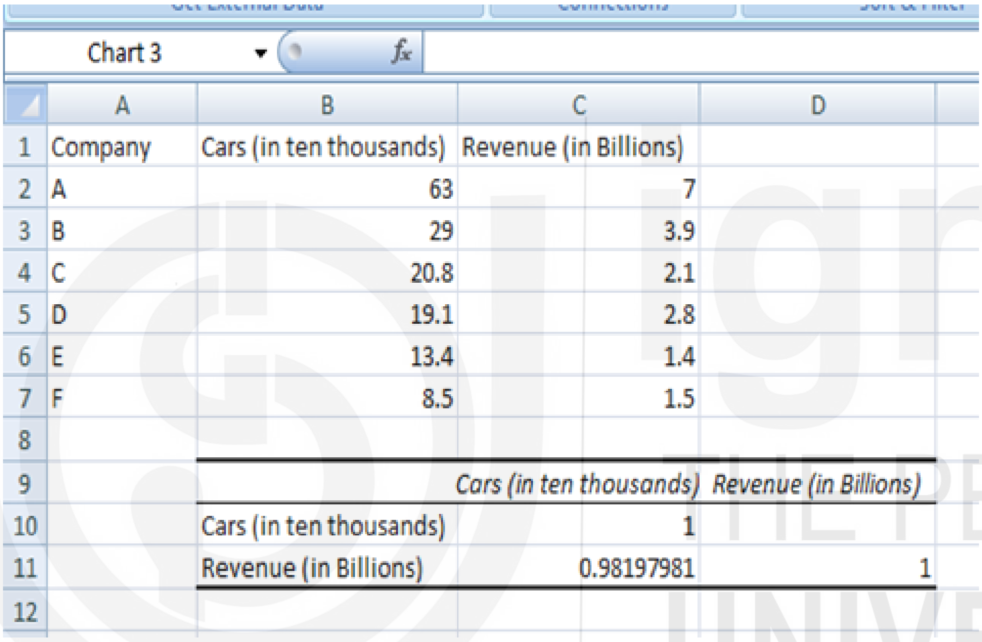
- *Data* Tab से *Analysis* (सबसे दाहिना विकल्प) चुनें। इससे *Data Analysis* का वार्ता बॉक्स खुल जाएगा। यहां आप *Correlation* चुनकर *OK* का बटन दबाएं। आपको अपने स्क्रीन पर कुछ ऐसा दिखाई देगा :



- *Correlation* वार्ता बॉक्स में Input Range में *$B$1:$C$7* टाइप करें तथा पहली पंक्ति में check box Labels की जाँच करें। हमारी आंकड़ा तालिका में पहली पंक्ति में शीर्षक दिए गए हैं। Grouped By विकल्प को Columns पर ही रखें क्योंकि प्रत्येक स्तंभ में एक ही चर के आंकड़े हैं। परिणाम या output range को *$B$9* अंकित करें (अगले चित्र में यह वार्ता बॉक्स दर्शाया गया है) और OK बटन दबाएं।



- अब चित्र 8.11 जैसा सहसंबंध गुणांक आव्यूह आपके स्क्रीन पर आ जाएगा।



	A	B	C	D
1	Company	Cars (in ten thousands)	Revenue (in Billions)	
2	A	63	7	
3	B	29	3.9	
4	C	20.8	2.1	
5	D	19.1	2.8	
6	E	13.4	1.4	
7	F	8.5	1.5	
8				
9			*Cars (in ten thousands)*	*Revenue (in Billions)*
10		Cars (in ten thousands)	1	
11		Revenue (in Billions)	0.98197981	1
12				

चित्र 8.11 : सहसंबंध गुणांक

ध्यान दें कि आगम और कारों के बीच सहसंबंध का मान 0.98197981 है। इसी भाग में इसके आकलित मान से तुलना करके देखें।

बोध प्रश्न 2

1) द्वि-विचर आंकड़ा विश्लेषण में सहसंबंध की संकल्पना समझाएं।

....

....

....

....

....

....

2) सहसंबंध गुणांक की परिभाषा करें तथा दो चरों के सहविचरण के साथ इसका संबंध समझाएं।

....

....

....

....

3) सहसंबंध गुणांक की आकलन विधि समझाएं और इस गुणांक की विशेषताएं और त्रुटियां/सीमाएं भी बताएं।

....

....

....

....

....

4) t-कसौटी का प्रयोग करते हुए सहसंबंध गुणांक के महत्त्व की जाँच की प्रक्रिया समझाएं।

....

....

....

....

....

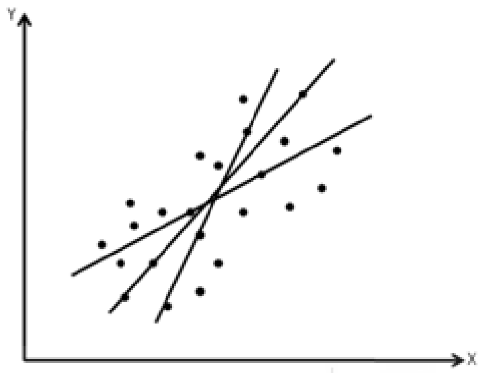
8.4 द्वि-विचर आंकड़ों में कारण-प्रभाव संबंध का विश्लेषण– प्रतीपगमन

दो चरों के संबंध का अध्ययन करते समय हम सबसे पहले उन आंकड़ों का प्रकीर्ण (Scatter plot) चित्र अंकित करते हैं। यह चित्र संबंध की प्रकृति बताता है। यहां तीन संभावनाएं होती है : एक 'धनात्मक रैखिक संबंध', 'ऋणात्मक रैखिक संबंध' या फिर 'कोई प्रेक्षणीय संबंध नहीं''। प्रकीर्ण चित्रांकन के बाद अगला कदम सहसंबंध गुणांक का आकलन करना और गुणांक की महत्त्वशीलता का परीक्षण करना है। यदि सहसंबंध गुणांक का मान महत्त्वपूर्ण है तो अगला कदम एक चर के मान के अनुसार दूसरे के मानों का अनुमान लगाना होगा। इस प्रकार अनुमानित मान को निर्भर या अनुमानित चर तथा दूसरा स्वतंत्र या 'अनुमानक' चर कहा जाता है। उदाहरण के लिए, यदि हमें यह ज्ञात हो कि विज्ञापन और बिक्री में सहसंबंध है तो हम विज्ञापन व्यय के प्रत्येक स्तर पर बिक्री के अपेक्षित आंकड़े या फिर निश्चित बिक्री के आंकड़ों के लिए आवश्यक विज्ञापन व्यय का आकलन करेंगे। प्रतीपगमन दो या अधिक चरों के बीच मूल आकड़ों की अपनी इकाइयों में औसत संबंध का मापन है। यह ध्यान रखना आवश्यक है कि यदि सहसंबंध महत्त्वहीन है तो फिर प्रतीपगमन रेखा का आकलन और उसके आधार पर कोई भी पूर्वाकलन करना अर्थहीन हो जाता है। प्रतीपगमन

रेखा का उद्देश्य शोधकर्ता को प्रवृत्तियों को देखने में समर्थ बनाकर आंकड़ों के आधार पर पूर्वाकलन करने योग्य बनाना है।

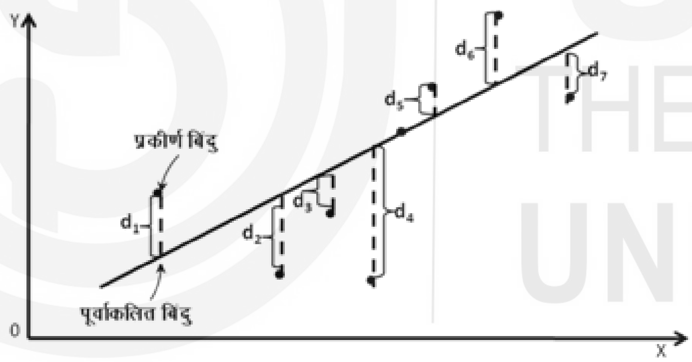
8.4.1 सर्वोत्तम संगति (fit) रेखा

निम्न चित्र में दो चरों का प्रकीर्ण चित्र दिखाया गया है। यहां यह भी दिखाया गया है कि इन बिंदुओं के आस-पास से अनेक सरल रेखाएं खींची जा सकती हैं।



चित्र 8.12

एक प्रकीर्ण चित्र प्राप्त होने पर आपको सबसे अधिक निकटस्थ या संगतिपूर्ण रेखा खींचने में समर्थ होना चाहिए। सर्वोत्तम संगति का अर्थ है कि प्रत्येक बिंदु से रेखा पर ऊर्ध्व लम्बों के वर्गों का योगफल न्यूनतम होना चाहिए।



चित्र 8.13

आपको सर्वोत्तम संगति की रेखा खींचने का कारण भी जान लेना चाहिए। हमें X के प्रत्येक मान के संगत Y का मान पूर्वाकलित करना है। अतः यह पूर्वाकलित बिंदु पथ हमारे प्रकीर्ण बिंदुओं के जितना अधिक निकट होगा, हमारा पूर्वाकलन उतना ही सटीक होगा। (जब r धनात्मक हो तो यह रेखा दाहिनी ओर उठती हुई होगी और जब r ऋणात्मक हो तो यह दाहिनी ओर ढलवां होगी)। न्यूनतम वर्ग विधि से प्राप्त रेखा सर्वोत्तम संगत रेखा जाना जाती है क्योंकि यह निम्नलिखित नियमों का पालन करती है :

क) विभिन्न अवलोकित (प्रकीर्ण) बिंदुओं के प्रतीपगमन रेखा से अंतरों का बीजगणितीय योगफल शून्य के समान होगा। अर्थात् :

$$\sum(Y - Y_e) = 0$$

जहाँ Y प्रेक्षित मान है तथा Y_e प्रतीपगमन रेखा से प्राप्त मान है।

ख) इन अंतरों के वर्गों का योगफल किसी भी अन्य रेखा से अंतरों के योगफल से कम है। अर्थात् :

$$\sum(Y - Ye)^2 < \sum (Y - Ai)^2$$

जहाँ Ai किसी अन्य रेखा से संगत मान हैं।

ग) प्रतीपगमन रेखा (सर्वोत्तम संगति रेखा) X तथा Y चरों के औसत मानों के बिंदु $(\bar{X}, \bar{Y})$ से गुजरती है।

8.4.2 प्रतीपगमन रेखा का समीकरण

दो चरों के संदर्भ में दो प्रतीपगमन समीकरण इस प्रकार हैं : Y के X पर प्रतीपगमन के लिए :

$Y = a + bX$ और

X के Y पर प्रतीपगमन के लिए

$$X = a' + b'Y$$

यहां X तथा Y तो हमारे चर हैं तथा a, b और a′, b′ वे स्थिरांक या गुणांक हैं जिनका निर्धारण किया जाना है। इन गुणांकों के निर्धारण के लिए हमें प्रत्येक प्रतीपगमन समीकरण से दो-दो समीकरण व्युत्पन्न करने होंगे। उन्हें हम प्रामाणिक समीकरण कहते हैं।

पहले प्रतीपगमन समीकरण $Y= a + bX$ के लिए प्रामाणिक समीकरण हैं:

$$\sum Y = na + b\sum X \quad \text{तथा}$$

$$\sum XY = a\sum X + b\sum X^2$$

इन प्रामाणिक समीकरणों से a, b तथा a′b′ के मान ज्ञात किये जा सकते हैं।

$$\sum X = na' + b'\sum Y \quad \text{तथा}$$

$$\sum XY = a'\sum Y + b'\sum Y^2$$

इसी प्रकार दूसरे प्रतीपगमन समीकरण $X = a' + b'Y$ के प्रामाणिक समीकरण होंगे :

8.4.3 प्रतीपगमन गुणांक

उपर्युक्त प्रामाणिक समीकरण युग्मों को युग्पद रूप में हल कर हमें प्रतीपगमन गुणांक प्राप्त होते हैं। इस विधि से हल करते हुए हमें प्रतीपगमन रेखाओं के ढाल के अनुमान क्रमशः इन सूत्रों के रूप में मिल जाते हैं:

$$b = b_{YX} = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{\sum(X-\bar{X})^2} = \frac{\sum xy}{\sum x^2}$$

$$b' = b'_{XY} = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{\sum(Y-\bar{Y})^2} = \frac{\sum xy}{\sum y^2}$$

जहाँ, $x = X - \bar{X}$ और $y = Y - \bar{Y}$

कुछ बीजगणितीय प्रयोगों तथा b_{yx} को σ_y और b'_{xy} को σ_x से गुणा कर हमें वैकल्पिक सूत्र भी मिल सकते हैं। यथा :

$$b = b_{yx} = r\frac{\sigma_y}{\sigma_x}$$

$$b' = b_{xy} = r\frac{\sigma_x}{\sigma_y}$$

ढाल गुणांक आकलित करने के बाद हम प्रथम प्रामाणिक समीकरणों में उनका प्रयोग कर अंतःखंड गुणांकों का आकलन कर सकते हैं। अतः $a = \bar{Y} - b\bar{X}$ और $a' = \bar{X} - b'\bar{X}$।

प्रतीपगमन गुणांकों की विशेषताएँ :

1) प्रतीपगमन रेखा सदैव उस बिंदु से गुजरती है जिसका X अक्षांश X मानों का औसत है और Y अक्षांश Y मानों का औसत, अर्थात् $(\bar{X}, \bar{Y})$.

2) दोनों प्रतीपगमन गुणांकों का चिन्ह समान होगा– वे या तो धनात्मक होंगे या ऋणात्मक।

3) सहसंबंध गुणांक दोनों प्रतीपगमन गुणांकों का ज्यामितिक औसत मान होता है, अर्थात्

$$r = \pm\sqrt{b_{XY} b_{YX}}$$

4) सहसंबंध गुणांक और प्रतीपगमन गुणांकों के चिन्ह एक समान होंगे। यदि r धनात्मक है तो b_{XY} और b_{YX} भी धनात्मक होंगे। इसी प्रकार r ऋणात्मक होने पर दोनों ढाल गुणांक भी ऋणात्मक ही होंगे। इसका कारण यही है कि दोनों सूत्रों में अंश के पद समान होते हैं और वही r तथा प्रतीपगमन गुणांकों के चिन्ह का निर्धारण करते हैं। 'हर' के पद तो वर्गरूपी होने के कारण सदैव धनात्मक ही रहते हैं।

5) यदि एक प्रतीपगमन गुणांक इकाई से अधिक हो तो दूसरा अनिवार्यतः इकाई से कम होगा।

6) प्रतीपगमन गुणांक अक्ष केंद्र से निरपेक्ष होते हैं किंतु मापन के पैमाने से नहीं।

7) b_{XY} तथा b_{YX} का गणितीय औसत सहसंबंध गुणांक से अधिक या उसके समान होता है। अर्थात् :

$$\frac{b_{XY} + b_{YX}}{2} \geq r$$

8) यदि $r = 0$, तो चर असंबद्ध होते हैं। इस दशा में ये दोनों प्रतीपगमन रेखाएँ लंबकोणिक हो जाएंगी।

9) यदि $r = \pm 1$, तो दोनों रेखाएं परस्पर आच्छादित होती हैं।

10) प्रतीपगमन रेखाओं के बीच का कोण चरों के बीच निर्भरता की कोटि दर्शाता है।

11) दोनों प्रतीपगमन रेखाएँ $(\bar{X}, \bar{Y})$, अर्थात् X तथा Y के औसत बिंदु पर प्रतिच्छेदन करती हैं।

8.4.4 प्रतीपगमन समीकरणों की पूर्वाकलनीयता

प्रतीपगमन गुणांकों के उपर्युक्त निष्कर्षित मानों का प्रयोग कर हम प्रतीपगमन समीकरणों को निम्न प्रकार व्यावहारिक स्वरूप में भी लिख सकते हैं :

Y के X पर प्रतीपगमन के लिए :

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

X के Y पर प्रतीपगमन के लिए :

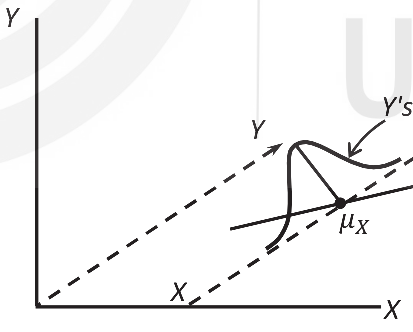
$$X - \bar{X} = r\frac{\sigma_x}{\sigma_y}(Y - \bar{Y})$$

ये दो सूत्र पूर्णतः पृथक्-पृथक् रेखाएं दर्शाते हैं। दूसरे शब्दों में Y का X पर प्रतीपगमन X का फलन है, जिसे Y = F(X) लिखा जा सकता है। इसी प्रकार X का Y पर प्रतीपगमन X = F(Y) होगा। ये चर X तथा Y पर परस्पर प्रतिस्थापनीय नहीं हैं। इसका कारण यही है कि Y के X पर प्रतीपगमन में Y निर्भर चर होता है तथा X स्वतंत्र चर। अर्थात् हम X के विभिन्न दिए गए मानों के लिए Y के अनुमान आकलित कर सकते हैं। इसी प्रकार X के Y पर प्रतीपगमन समीकरण से हम Y के प्रदत्त मानों के आधार पर X के मानों के अनुमान लगा सकते हैं।

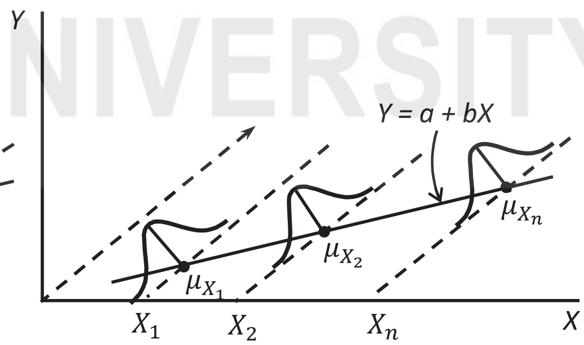
स्वतंत्र चर में एक इकाई परिवर्तन पर निर्भर चर में हुए परिवर्तन को सीमांत परिवर्तन कहा जाता है। प्रतीपगमन रेखा में ढाल गुणांक इसी सीमांत परिवर्तन को दर्शाता है। यदि r का मान 0 से अधिक दूर नहीं हो तो Y का श्रेष्ठतम पूर्वानुमान अवलोकित Y आंकड़ों का औसत मान ही होता है। पूर्वाकलन की वैधता के लिए सहसंबंध गुणांक का मान महत्त्वपूर्ण होना चाहिए। साथ ही, ये मान्यताएं भी पूरी होनी चाहिएं :

1. स्वतंत्र चर X के किसी भी मान विशेष के लिए निर्भर चर Y के मानों का आवंटन प्रतीपगमन रेखा के गिर्द प्रसामान्य स्वरूप का होता है। यही बात चित्र 8.14(क) में दिखाई गई है।

2. स्वतंत्र चर के प्रत्येक मान के लिए निर्भर चर के आवंटन का मानक विचलन एक समान रहना चाहिए, देखें 8.14(ख)।



चित्र 8.14(क)



चित्र 8.14(ख)

(क) निर्भर चर का आवंटन प्रसामान्य है। (ख) $\sigma_1 = \sigma_2 = \cdots = \sigma_n$

प्रभावकारी या बहिष्शायी (Outlier) अवलोकन की पहचान

बहिष्शायी की पहचान के लिए प्रकीर्ण चित्र की जाँच करें। यह ऐसे बिंदु हैं जो अधिकांश अन्य बिंदुओं से बहुत दूर दिखाई देते हैं। ऐसे कुछ बिंदु प्रतीपगमन रेखा के समीकरण को भी प्रभावित कर सकते हैं। ऐसी दशा में इन्हें प्रभावकारी अवलोकनों की संज्ञा दी जाती है। प्रकीर्ण चित्र में बहिष्शायी लगने वाले बिंदुओं की प्रतीपगमन रेखा पर प्रभाव के लिए जांच करनी चाहिए। एक प्रभावकारी बिंदु पूरी रेखा को अपनी ओर खीचंता हुआ दिखाई देता है। इस प्रवृत्ति की जाँच के लिए प्रतीपगमन रेखा का पहले

ऐसे बिंदु को शामिल करते हुए रेखांकन किया जाता है और फिर उसके बिना भी रेखाकंन करते हैं। यदि दूसरा चित्र पहले की अपेक्षा बहुत भिन्न हो तो उस बिंदु को प्रभावकारी माना जाता है। प्रायः क्षैतिज दिशा के बहिष्शायिओं के प्रभावकारी सिद्ध होने की संभावना अधिक होती है।

आंकड़ों के विश्लेषण में प्रभावकारी अवलोकनों को समाहित करने के विषय में तो अध्ययनकर्ता को अपने विवेक का प्रयोग करना पड़ता है। यदि उसे प्रतीत हो कि कोई अवलोकन विशेष महत्त्वपूर्ण नहीं है तो उसे छोड़ देना अच्छा रहता है ताकि वह अध्ययन के निष्कर्षों को अनावश्यक रूप से प्रभावित नहीं कर पाए। यदि उस अवलोकन को आवश्यक माना जाए तो उसके आसपास के कुछ अन्य मानों के अवलोकन भी एकत्र कर अध्ययन में समाहित करने चाहिए ताकि उसका प्रभाव बहुत स्पष्ट रूप से हमारे समीकरणों में प्रतिबिंबित हो सके।

8.4.5 निर्धारण का गुणांक

हमने पिछले उपभागों में जाना कि यदि सहसंबंध गुणांक महत्त्वपूर्ण है तो प्रतीपगमन समीकरणों का आकलन हो सकता है। साथ ही स्वतंत्र चर के विभिन्न मानों के अनुरूप निर्भर चर के मानों का प्राक्कलन भी हो सकता है। प्रतीपगमन समीकरण की पूर्वानुमेयता का निर्णय उसके निर्धारण के गुणांक द्वारा हो जाता है।

प्रतीपगमन प्रतिमान से संबंधित अनेक प्रकार के विचरणों की परिभाषा कर लेना उपयोगी रहेगा। इस काल्पनिक प्रतीपगमन प्रतिमान पर विचार करें :

X	1	2	3	4	5
Y	10	8	12	16	20

प्रतीपगमन समीकरण है : $Y_e = 4.8 + 2.8X$ और $r = 0.919$। प्रतिदर्श में Y के मान क्रमशः 10, 8, 12, 16 और 20 हैं। Y_e द्वारा निर्दिष्ट आकलित मान हम उक्त समीकरण में X के विभिन्न मान रखकर ज्ञात कर सकते हैं। उदाहरण के लिए, जहाँ $X = 1$ वहां $Y_e = 4.8 + 2.8X = 4.8 + (2.8)(1) = 7.6$

अब हमारे पास प्रत्येक X के लिए एक अवलोकित Y तथा एक आकलित Y_e मान इकट्ठे हो गए हैं। पिछले भाग से याद करें कि यदि आकलित मान अवलोकित मानों के निकटवर्ती हों तो संगति या 'fit' बेहतर होता है– अर्थात् r का मान 1 या –1 के निकट होता है।

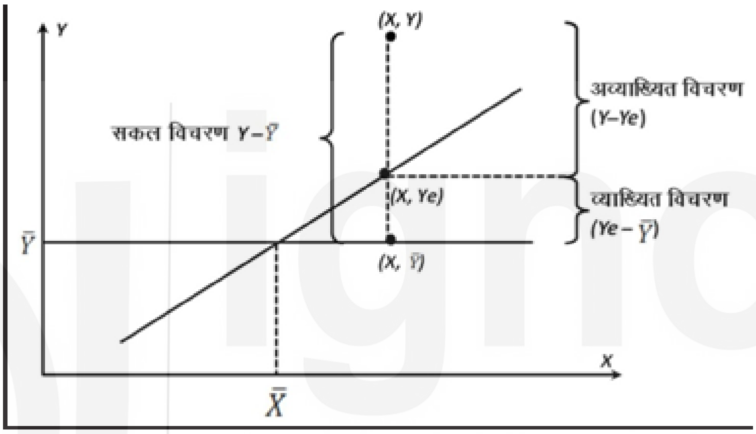
सकल विचरण $\sum(Y - \bar{Y})^2$ प्रत्येक बिंदु के औसत से ऊर्ध्व अंतर के वर्गों का योग है। इस विचरण को दो भागों में बांटा जा सकता है : जो X तथा Y के बीच संबंध का परिणाम है और मात्र संयोग का परिणाम है। संबंध के कारण विचरण को पूर्वाकलित Y_e से अंतर माना जाता है– यह होगा $\sum(Y_e - \bar{Y})^2$ । इसे 'व्याख्यित (explained) विचरण' का नाम दिया जाता है। संबंध द्वारा विचरण के अधिकांश भाग की व्याख्या हो जानी चाहिए। r का मान –1 या +1 के जितना निकट हो, उतना अच्छा रहता है। हमारी आकलित रेखा उतनी ही बिंदुओं के निकटस्थ होगी – अर्थात् $\sum(Y_e - \bar{Y})^2$ उतना ही $\sum(Y - \bar{Y})^2$ के निकट होगा। यदि सभी बिंदु प्रतीपगमन रेखा

पर ही व्यवस्थित हों तो ये दोनों पद समान ही होंगे (क्योंकि प्रत्येक मामले में Y_e = Y)।

दूसरी ओर केवल संयोग के कारण उत्पन्न अंतरों के वर्ग का योगफल, $\sum(Y - Y_e)^2$ अव्याख्यित रह गया विचरण है। यह हमारे आकलित संबंध का परिणाम नहीं है। जब यह विचरण-अंश न्यून होता है तो r का मान –1 या फिर +1 के निकट पहुँच जाता है। यदि सभी बिंदु प्रतीपगमन रेखा पर हों तो यह विचरण अंश शून्य हो जाएगा। अतः सकल विचरण इन दोनों विचरण अंशों– व्याख्यित एवं अव्याख्यित का योग होता है। अर्थात्

$$\sum(Y - \bar{Y})^2 = \sum(Y_e - \bar{Y})^2 + \sum(Y - Y_e)^2$$

चित्र 8.15 इन्हीं मानों को दर्शा रहा है :



चित्र 8.15

अतः निर्धारण गुणांक व्याख्यित विचरण का सकल विचरण से अनुपात होगा। अर्थात्

r^2 =व्याख्यित विचरण / सकल

विचरण $= \frac{\sum(Y_e - \bar{Y})^2}{\sum(Y - \bar{Y})^2}$

यह निर्धारण गुणांक निर्भर चर में हुए उस परिवर्तन को मापता है जिसकी स्वतंत्र चर के माध्यम से आकलित प्रतीपगमन रेखा व्याख्या कर देती है। इसका आकलन करना r का वर्ग करने की विधि से अधिक सरल रहता है। यदि r = 0.90 तो r^2 = 0.81 – यह 81% के समान है। इसका अर्थ है कि स्वतंत्र चर के विचरण द्वारा निर्भर चर के 81 प्रतिशत विचरण की व्याख्या हो जाती है। शेष 0.19 अर्थात् 19% विचरण की व्याख्या नहीं हो पाई है। इसे निर्धारण-गुणांक को घटाकर प्राप्त किया जाता है। जैसे-जैसे r का मान शून्यगामी होता है, r^2 उससे कहीं अधिक तेज़ी से घटने लगता है। यदि r = 0.6 हो तो निर्धारण गुणांक r^2 = 0.36 ही रह जाएगा – अर्थात् निर्भर चर के मात्र 36% विचरण की व्याख्या हमारा स्वतंत्र चर कर पाएगा।

बोध प्रश्न 3

1) प्रतीपगमन विश्लेषण में सर्वश्रेष्ठ संगति (best fit) रेखा से आपका क्या अभिप्राय है?

....

....

....

....

....

2) दो चरों के संदर्भ में प्रतीपगमन समीकरणों, उनके प्रतीपगमन गुणांकों तथा उनकी विशेषताओं की परिभाषा करें।

....

....

....

....

....

3) आप प्रतीपगमन समीकरणों की पूर्वानुमेयता से क्या समझते हैं? वैध पूर्वाकलनों के लिए आवश्यक मान्यताओं की संक्षिप्त व्याख्या करें।

....

....

....

....

....

4) निर्धारण गुणंक की परिभाषा करें और प्रतीपगमन विश्लेषण में इसका महत्त्व समझाएं।

....

....

....

....

8.5 प्रतीपगमन प्रतिमान की विश्वस्तता – अनुमान की मानक त्रुटि (RELIABILITY OF REGRESSION MODEL – STANDARD ERROR OF ESTIMATE)

जब किसी विशेष X द्वारा Y_e का पूर्वानुमान किया जाता है तो उसे पूर्वानुमान बिंदु कहा जाता है। किंतु किसी भी पूर्वानुमान के गिर्द एक पूर्वाकलन अंतराल की रचना वैसे ही हो सकती है जैसे समष्टि औसत के अनुमान के लिए इकाई 7 में की गई थी। यह पूर्वाकलन अंतराल ''अनुमान की मानक त्रुटि'' नामक सांख्यिक का प्रयोग करता है।

8.5.1 अनुमान की मानक त्रुटि : संकल्पना और इसका प्राक्कलन

अनुमान की मानक त्रुटि जिसे s_{est} द्वारा दर्शाया जाता है अवलोकित Y मानों के पूर्वाकलित मानों के गिर्द अंतरों का मानक विचलन ही है। इस मानक त्रुटि का सूत्र इस प्रकार है :

$$s_{est} = \sqrt{\frac{\sum(Y - Y_e)^2}{n-2}}$$

यह मानक त्रुटि मानक विचलन जैसी ही होती है। किंतु इसके आकलन में औसत का प्रयोग नहीं होता। जैसा कि सूत्र से ही स्पष्ट है कि इसे अव्याख्यित विचरण को n–2 से भाग देकर ज्ञात किया जाता है। अतः पूर्वाकलित मान वास्तविक अवलोकनों के जितना अधिक निकट होंगे यह मानक त्रुटि उतनी ही कम होगी।

उदाहरण :

एक शोधकर्ता ने निम्न आंकड़ें एकत्र कर यह निर्धारित किया है कि प्रतिलिपि मशीन की आयु तथा उसकी मासिक रखरखाव की लागत में महत्त्वपूर्ण संबंध है। उस द्वारा आकलित प्रतीपगमन रेखा है : $Y_e = 55.57 + 8.13X$। अनुमान की मानक त्रुटि ज्ञात करें।

मशीन	आयु (X) (वर्षों में)	महीने की लागत (Y)$
A	1	62
B	2	78
C	3	70
D	4	90
E	5	93
F	6	103

हल :

चरण 1 : निम्न प्रकार की तालिका बनाएं :

X	Y	Y_e	$(Y - Y_e)$	$(Y - Y_e)^2$
1	62			
2	78			
3	70			
4	90			
5	93			
6	100			

चरण 2 : प्रतीपगमन समीकरण $Y_e = 55.57 + 8.13X$ का प्रयोग कर प्रत्येक X के लिए Y_e का आकलन कर उसे उपर्युक्त तालिका में यथास्थान लिखें (स्तंभ Y_e)।

चरण 3 : प्रत्येक Y में से संबद्ध Y_e घटाकर स्तंभ $(Y - Y_e)$ में लिखें।

चरण 4 : पिछले चरण 3 की संख्याओं का वर्ग कर $(Y - Y_e)^2$ द्वारा इंगित स्तंभ में लिखें।

चरण 5 : अंतिम स्तंभ के मानों का योग करें। हम पूर्ण रूप से भरी तालिका नीचे दर्शा रहे हैं :

X	Y	Y_e	$(Y - Y_e)$	$(Y - Y_e)^2$
1	62	63.70	−1.70	2.89
2	78	71.83	6.17	38.0689
3	70	79.96	−9.96	99.2016
4	90	88.09	1.91	3.6481
5	93	88.09	4.91	24.1081
6	103	104.35	−1.35	1.8225
				169.7392

चरण 6 : सूत्र में यथास्थान उपर्युक्त तालिका के मान रखकर गणना करें :

$$s_{est} = \sqrt{\frac{\sum(Y - Y_e)^2}{n-2}}$$

$$= \sqrt{\frac{169.7392}{6-2}} = 6.51$$

इस मामले में अवलोकित मानों का पूर्वाकलित मानों से मानक विचलन 6.51 है।

8.5.2 पूर्वानुमेयता अंतराल (Prediction Interval)

प्रतीपगमन विश्लेषण का मूल उद्देश्य प्रतिदर्श के आकलन के आधार पर समष्टि प्रतीपगमन फलन के विषय में पूर्वानुमान लगाना है। अतः शोधकर्ता समष्टि प्रतीपगमन गुणांकों के बिंदु अनुमान आकलित करने का प्रयास तो करता ही है किंतु इन बिंदु अनुमानकों की विश्वस्तता की स्थापना कर पाना सदैव संभव नहीं होता। अतः किसी विश्वस्ति स्तर के साथ ऐसे अंतरालों की परिभाषा करना उपयोगी रहता है, जिनमें अनुमानित मान भी समाहित होने का कुछ भरोसा किया जा सके– अर्थात् जिन्हें निर्णयकर्ता स्वीकार कर पाए।

अनुमान की मानक त्रुटि का प्रयोग कर हम Y_e के गिर्द पूर्वानुमेयता अंतराल (विश्वस्ति अंतराल की ही भांति) ज्ञात कर सकते हैं। जब हम प्रतीपगमन समीकरण में कोई X मान रखते हैं तो हमें Y का एक बिंदु अनुमान (Y_e के रूप में) मिलता है। उदाहरण के लिए, प्रतिलिपि मशीन की आयु और मासिक रखरखाव की लागत की प्रतीपगमन रेखा $Y_e = 55.57 + 8.13X$ के अनुसार एक 3 वर्ष पुरानी मशीन की रखरखाव लागत का अनुमान $Y_e = 55.57 + 8.13(3) =$ \$79.96 होगा। यह एक बिंदु अनुमान है और हम नहीं जानते कि यह कितना भरोसेमंद है। किंतु इस अनुमान के गिर्द हम एक विश्वस्ति अंतराल अवश्य बना सकते हैं। एक महत्त्वपूर्णता स्तर (α) चुनकर हम $(1-\alpha) \times 100\%$ विश्वास से कह सकते हैं कि उक्त अंतराल में दिए हुए X के मान से संबद्ध Y मानों का वास्तविक औसत भी समाहित होगा।

कारण यही है कि प्रतीपगमन रेखा के निर्धारण में पूर्वाकलन त्रुटियों के कुछ कारण अवश्य होते हैं। एक तो अनुमान की मानक त्रुटि s_{est} ही है। दो अन्य त्रुटियां ढाल और Y_e के अंतःखंड (Intercept) के आकलन में रहती हैं क्योंकि ज़रा-सा अलग यादृच्छिक प्रतिचयन लेने पर ही आकलित प्रतीपगमन समीकरण भी बदल जाता है।

इकाई 7 में चर्चित आवंटन का प्रयोग करते हुए किसी Y_e मान के गिर्द पूर्वानुमेयता अंतराल इस प्रकार होगा :

$$Y_e - t\alpha_{/2}\, s_{est}\sqrt{1+\frac{1}{n}+\frac{n(X-\bar{X})^2}{n\sum X^2-(\sum X)^2}} < Y$$
$$< Y_e + t\alpha_{/2}\, s_{est}\sqrt{1+\frac{1}{n}+\frac{n(X-\bar{X})^2}{n\sum X^2-(\sum X)^2}}$$

जहाँ $n-2$ स्वतंत्रता की कोटि (degree of freedom) को इंगित करता है।

उदाहरण :

पिछले अनुभाग की अनुमान की मानक त्रुटि के आकलन की समस्या में 3 वर्ष पुरानी मशीन की मासिक रखरखाव लागत के लिए 95% पूर्वानुमान अंतराल आकलित करें ($t\alpha_{/2} = 2.776$)।

हल :

चरण 1 : ज्ञात करें $\sum X, \sum X^2$ और $\bar{X}$

$\sum X = 21$, $\sum X^2 = 91$ और $\bar{X} = \frac{21}{6} = 3.5$

चरण 2 : $X = 3$ के लिए Y_e ज्ञात करें

$Y_e = 55.57 + 8.13(3) = 79.96$

चरण 3 : पिछले अनुच्छेद की भांति s_{est} ज्ञात करें।

चरण 4 : सूत्र में विभिन्न मान रखें और हल करें, 95% के लिए :

$t_{\alpha/2} = 2.776$, स्वतंत्रता कोटि $= 6 - 2 = 4$

$$Y_e - t\alpha_{/2}\, s_{est}\sqrt{1+\frac{1}{n}+\frac{n(X-\bar{X})^2}{n\sum X^2-(\sum X)^2}} < Y < Y_e +$$
$$t\alpha_{/2}\, s_{est}\sqrt{1+\frac{1}{n}+\frac{n(X-\bar{X})^2}{n\sum X^2-(\sum X)^2}}$$

$$79.96 - (2.776)(6.51)\sqrt{1+\frac{1}{6}+\frac{6(3-3.5)^2}{6(91)-(21)^2}} < Y < 79.96 +$$
$$(2.776)(6.51)\sqrt{1+\frac{1}{6}+\frac{6(3-3.5)^2}{6(91)-(21)^2}}$$

$$79.96 - (2.776)(6.51)(1.087) < Y < 79.96 + (2.776)(6.51)(1.087)$$

$$79.96 - 19.64 < Y < 79.96 + 19.64$$

$$60.32 < Y < 99.6$$

अतः आप 95% भरोसे के साथ कह सकते हैं कि अंतराल $60.53<Y<99.39$ में ही Y का वास्तविक मान भी होगा।

अभ्यास कार्य 3 : व्यावहारिक कार्य-2 के आंकड़ों के लिए प्रतीपगमन रेखा का आकलन करें

इसी अनुभाग के वर्षों में (आयु) X तथा मासिक रखरखाव लागत Y के बीच प्रतीपगमन रेखा का आकलन करें।

यह कार्य कैसे करेंगे?

- *Data* टैब से *Analysis* का विकल्प रिबन से चुन लें। इससे *Data Analysis* संवाद बॉक्स खुल जाएगा। यहां *Regression* चुनकर *OK* बटन दबाएं। Regression परिणाम इस SUMMARY OUTPUT में दिखाई देगा :

Summary Output

Regression Statistics	
Multiple R	0.9255044
R Square	0.8565585
Adjusted R Square	0.8206981
Standard Error	6.5141985
Observations	6

- ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	1013.594	1013.594	23.885929	0.0081
Residual	4	169.7391	42.43478		
Total	5	1183.333			

	Coefficients	*Standard Error*	*t Stat*	*p-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	55.565217	6.149984	9.035018	0.0008313	38.49	72.64	38.4901237	72.64031
Age x (Years)	8.1304348	1.663576	4.887323	0.0081177	3.5116	12.749	3.51160674	12.74926

आप इसी अनुभाग में चर्चित गुणांकों से तुलना करके विश्लेषण कर सकते हैं। अभी आप अन्य सांख्यिकों पर ध्यान न दें।

बोध प्रश्न 4

1) प्रतीपगमन विश्लेषण में 'अनुमान की मानक त्रुटि' से आपका क्या तात्पर्य है? इसका अनुमान किस प्रकार लगाया जाता है?

....

....

....

....

2) t-आवंटन प्रयोग करते हुए अनुमान की मानक त्रुटि का पूर्वानुमेयता अंतराल आकलित करें।

....

....

....

....

8.6 दो औसतों, दो अनुपातों तथा दो विचरणों के बीच अंतरों की जाँच

ऐसी अनेक स्थितियाँ होती हैं जहाँ शोधकर्ता दो समष्टि औसतों की जांच करना चाहते हैं – यहां एक प्रायोगिक तथा दूसरी नियंत्रण समूह की औसत होती है। उदाहरण के लिए, दो अलग-अलग बस टायरों में यह भी तुलना की जा सकती है कि किसका जीवन कितना है, अर्थात् कब घिस जाता है। दो अलग-अलग ब्रांड के उर्वरकों में से कौन-सा पौधों की वृद्धि में अधिक सहायक है या फिर यह तुलना करने का प्रयास हो सकता है कि खांसी के दो औषधि युक्त शर्बतों में से कौन-सा अधिक प्रभावोत्पादक है? यहां भी इकाई 7 में चर्चित z-कसौटी आधारित संकल्पनाओं का ही दो औसतों और अनुपातों तथा F-कसौटी का प्रयोग दो विचरणों की तुलना के लिए किया जाता है।

8.6.1 दो औसतों के बीच अंतरों की समीक्षा : z- तथा t-कसौटी

मान लें कि एक शोधकर्ता किसी सामुदायिक कॉलेज में नर्स का प्रशिक्षण प्रारंभ करने वाले व्यक्तियों की औसत आयु की तुलना किसी विश्वविद्यालय में नर्स का प्रशिक्षण प्रारंभ करने वालों की औसत आयु से कर रहा है। यहां नर्सिंग प्रशिक्षण के लिए जाने वाले सभी व्यक्तियों की औसत आयु में शोधकर्ता को कोई रुचि नहीं है, वह तो बस उक्त दो समूहों के बीच तुलना करना चाहता है। उसका शोध प्रश्न है : क्या सामुदायिक कॉलेज में नर्सिंग में प्रवेश लेने वालों की औसत आयु किसी विश्वविद्यालय में प्रवेश लेने वाले नर्सिंग प्रशिक्षु से भिन्न है? यहां संकल्पनाएं होंगी :

$$H_0: u_1 = u_2$$

$$H_1: u_1 \neq u_2$$

यहाँ

u_1 = सामुदायिक कॉलेज में नर्सिंग प्रशिक्षुओं की प्रवेश के समय औसत आयु; और

u_2 = किसी विश्वविद्यालय में नर्सिंग प्रशिक्षुओं की प्रवेश के समय औसत आयु है।

इन संकल्पनाओं को लिखने की दूसरी विधि है :

$H_0: u_1 - u_2 = 0$

$H_1: u_1 - u_2 \neq 0$

यदि समष्टि औसतों में अंतर नहीं हों तो उन्हें घटाने पर हमें शून्य ही प्राप्त होगा। यदि वे समान नहीं हों तो अंतर शून्य से इतर होगा। संकल्पना निरूपण की दोनों विधियाँ ठीक होती हैं। दो औसतों में अंतर की समीक्षा का सैद्धांतिक आधार दो प्रतिदर्शों का चयन कर उनकी औसतों के बीच तुलना करने वाला ही है। समष्टि की औसत पूर्वविदित होना आवश्यक नहीं है। समष्टियों से सभी संभव प्रतिदर्श युग्मों पर विचार किया जाता है।

$\bar{X}_1 - \bar{X}_2$ के विचरण का मान $\bar{X}_1$ तथा $\bar{X}_2$ के निजी विचरणों का योगफल होता है। अर्थात्

$$\sigma^2_{\bar{X}_1 - \bar{X}_2} = \sigma^2_{\bar{X}_1} + \sigma^2_{\bar{X}_2}$$

जहाँ $\sigma^2_{\bar{X}_1} = \frac{\sigma_1^2}{n_1}$ और $\sigma^2_{\bar{X}_2} = \frac{\sigma_2^2}{n_2}$

अतः $\bar{X}_1 - \bar{X}_2$ का मानक विचलन होगा :

$$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

दो स्वतंत्र समष्टियों से औसतों की तुलना के लिए z-कसौटी सूत्र इस व्यापक स्वरूप पर आधारित है :

$$z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

जहाँ $\bar{X}_1 - \bar{X}_2$ अवलोकित अंतर है और अपेक्षित अंतर $\mu_1 - \mu_2$। उस दशा में शून्य होना चाहिए जहाँ $\mu_1 = \mu_2$ और अंत में अंतर का मानक विचलन होगा :

$$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

दो प्रतिदर्श औसतों के बीच अंतर संयोग मात्र हो सकता है, उस दशा में शून्यक संकल्पना निरस्त नहीं की जाएगी और शोधकर्ता यह स्वीकार कर लेगा कि दोनों समष्टियाँ मूलतः समान ही हैं। यहां अंतर महत्त्वपूर्ण नहीं है। दूसरी ओर यदि अंतर महत्त्वपूर्ण हो तो शून्यक संकल्पना अमान्य कर दी जाती है तथा शोधकर्ता यह निष्कर्ष बता सकता है कि दोनों समष्टियां परस्पर भिन्न हैं।

दो औसतों के बीच अंतर के लिए विश्वस्ति अंतराल भी निर्धारित किया जा सकता है। यदि आप शून्य अंतर की संकल्पना लेकर चल रहे हैं और विश्वस्ति अंतराल में शून्य भी हो तो शून्यक संकल्पना को अमान्य नहीं किया जाएगा। यदि अंतराल में शून्य नहीं

हो तो उक्त संकल्पना अमान्य हो जाएगी। हम निम्न सूत्र का प्रयोग कर दो औसतों के बीच अंतर के लिए विश्वस्ति अंतराल ज्ञात कर सकते हैं :

$$(\bar{X}_1 - \bar{X}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

दो औसतों के बीच अंतर की जांच के लिए z-कसौटी का प्रयोग उस समय किया गया था जब समष्टि मानक विचलन ज्ञात थे और चरों के आवंटन प्रसामान्य या प्रसामान्य–प्रायः थे अथवा दोनों प्रतिदर्शों के आकार 30 या अधिक ही थे। यदि σ_1 और σ_2 हों, किंतु प्रतिदर्श आकार n_1 तथा n_2 बड़े हो (30 से) तो हम उनके अनुमानों s_1^2 तथा s_2^2 का प्रयोग कर सकते हैं और z-सांख्यिक इस सूत्र जैसा हो जाएगा :

$$z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

8.6.2 दो अनुपातों में अंतरों की समीक्षा : z-कसौटी

कुछ परिवर्तनों के साथ दो अनुपातों की समानता की जाँच के लिए z-कसौटी का प्रयोग हो सकता है। उदाहरण : शोधकर्ता यह जानने का इच्छुक हो सकता है कि क्या ऐसे पुरुषों का अनुपात जो नियमित व्यायाम करते हैं महिलाओं की अपेक्षा कम है? या क्या निजी कंप्यूटर स्वामी छात्रों का अनुपात उनसे भिन्न है जिनके पास निजी कंप्यूटर नहीं है?

दो अनुपातों में $\hat{p}_1 = \frac{X_1}{n_1}$ p_1 तथा $\hat{p}_2 = \frac{X_2}{n_2}$ p_2 का अनुपात लगाने के लिए प्रयोग करते हैं। उनके अंतर की मानक त्रुटि होगी :

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\sigma_{p_1}^2 + \sigma_{p_2}^2} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

जहाँ $\sigma_{p_1}^2$ तथा $\sigma_{p_2}^2$ उन अनुपातों के विचरण हैं तथा $q_1 = 1 - p_1$, $q_2 = 1 - p_2$ तथा n_1 और n_2 क्रमशः प्रतिदर्श आकार हैं।

z-कसौटी प्रयोग करने के लिए दो शर्तें पूरी होना आवश्यक है : (1) प्रतिदर्श परस्पर स्वतंत्र हों; तथा (2) n_1p_1 और n_1q_1 के मान 5 या उससे अधिक हों, यही बात n_2p_2 और n_2q_2 पर भी लागू हो।

दो अनुपातों में अंतर के लिए विश्वस्ति अंतराल सूत्र होगा :

$$(\hat{p}_1 - \hat{p}_2) - z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} < p_1 - p_2 < (\hat{p}_1 - \hat{p}_2) + z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

8.6.3 दो विचरणों में अंतरों की समीक्षा : F-कसौटी

सांख्यिकीविद् दो औसतों के साथ-साथ दो विचरणों या मानक विचलनों की तुलना भी करना चाहते हैं। उदाहरण : क्या दो नगरों में किसी महीने में तापमानों के विचरण भिन्न हैं? या फिर क्या पुरुषों एवं महिलाओं में कोलेस्ट्रोल के स्तर में विचरण भिन्न हैं? दो विचरणों का मानक विचलनों की तुलना के लिए प्रायः F-कसौटी का प्रयोग किया जाता है।

यदि समान विचरणों वाली दो समष्टियों $(\sigma_1^2 = \sigma_2^2)$ से दो प्रतिदर्श स्वतंत्र रूप से निकाले जाते हैं और उनके s_1^2 तथा s_2^2 की तुलना के लिए $\frac{s_1^2}{s_2^2}$ अनुपात पर विचार किया जाता है तो विचरणों का प्रतिदर्श आवंटन F-आवंटन कहलाता है।

F-आवंटन की विशेषताएँ ये होती हैं :

- F का मान कभी ऋणात्मक नहीं होगा, क्योंकि विचरण सदैव धनात्मक या फिर शून्य होता है।
- आवंटन में धनात्मक वैषम्य होता है।
- F का औसत मान लगभग एक इकाई के समान होता है।
- F-आवंटन उन वक्रों का समूह है जो अंश तथा हर के विचरणों की स्वतंत्रता कोटियों पर निर्भर होता है।

F-सांख्यिक का सूत्र यह है : $F = \frac{s_1^2}{s_2^2}$

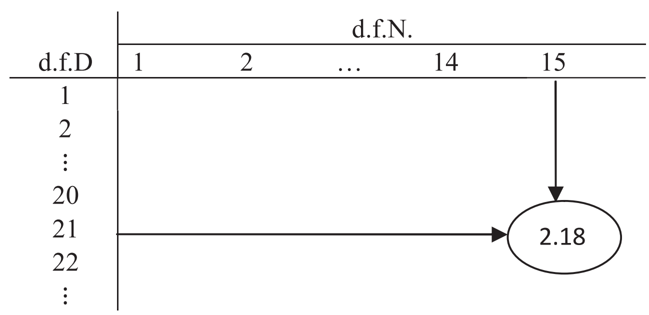
किंतु ध्यान रहे कि क्रम कोई भी हो, उच्चतर विचरण को ही अंश माना जाता है। F-कसौटी में स्वतंत्रता कोटि के दो पद होते हैं $n_1 - 1$, 'अंश' के लिए तथा $n_2 - 1$, 'हर' के लिए। यहां n_1 उस प्रतिदर्श का आकार है जिसका आकलित विचरण अधिक है।

F-आवंटन की तालिका α = 0.005, 0.01, 0.025, 0.05 तथा 0.10 के लिए निर्णायक मान (critical values) दर्शाती है। इस प्रकार से प्रत्येक α के लिए एक अलग F-तालिका होती है। ये एक-पुच्छ मान हैं। यदि आप द्वि-पुच्छ समीक्षा कर रहे हैं तो आपको α/2 मान देखने होंगे। उदाहरण के लिए, α = 0.05 की द्वि-पुच्छ समीक्षा के लिए 0.05/2 = 0.025 वाली F-तालिका का प्रयोग करना होगा।

उदाहरण : दक्षिण पुच्छ (right tailed) F-कसौटी का निर्णायक मान ज्ञात करें जबकि α = 0.05 अंश की स्वतंत्रता कोटि (d.f.N.) =15 तथा हर की स्वतंत्रता कोटि (d.f.D.) = 21।

हल : यह एक दक्षिण पुच्छ की जांच का मामला है जहां α = 0.05। अतः 0.05 तालिका का प्रयोग करें। d.f.N. को शीर्षक में तथा d.f.D. को स्तंभ में दर्शाया जाता है। यह चित्र 2.16 में इंगित 2.18 है।

$$\boldsymbol{\alpha = 0.05}$$



आपको पहले भी बताया गया है F-कसौटी के सूत्र में उच्चतर विचरण को ही अंश माना जाता है। यदि द्वि-पुच्छ जांच कर रहे हों तो α को विभाजित करते हैं तथा दो मान होने के बावजूद दक्षिण पुच्छ का ही प्रयोग किया जाता है। कारण यही है कि F का मान सदैव एक इकाई या उससे अधिक होना चाहिए।

बोध प्रश्न 5

1) उदाहरण सहित बताएं कि एक ही समष्टि से या भिन्न समष्टियों से निष्कर्षित दो प्रतिदर्श सांख्यिकों में अंतर की जांच क्यों आवश्यक होती है?

....

....

....

....

2) दो औसतों के बीच अंतर के लिए z तथा t दोनों कसौटियों का प्रयोग क्यों किया जाता है?

....

....

....

....

3) F-आवंटन की विशेषताओं का वर्णन करें और इसे प्रयोग करते हुए भिन्न समष्टियों से प्राप्त दो विचरण मानों के बीच अंतर की समीक्षा के लिए सूत्र का निरूपण करें।

....

....

....

....

8.7 विचरणों का विश्लेषण

पिछले भाग में दो विचरणों की तुलना में प्रयुक्त F-कसौटी को तीन या अधिक औसतों की तुलना के लिए भी प्रयोग किया जा सकता है। इस तकनीक को विचरण विश्लेषण ANOVA कहते हैं।

यह उन दावों की जांच के लिए प्रयोग होता है जहां तीन या अधिक औसत विद्यमान हों। उदाहरण : एक अध्येता यह जानना चाहता है कि Fortran, Basic तथा Pascal का प्रयोग कर किसी समस्या को हल करने वाले तीन समूहों के छात्रों द्वारा लिए गए औसत समय अलग-अलग हैं। यहां शोधकर्ता ANOVA विधि का प्रयोग करेगा। जब

तीन या अधिक औसतों की तुलना हों वहां निम्न कारणों से z और t कसौटियों का प्रयोग नहीं करना चाहिए :

- जब आप एक बार में केवल दो औसतों पर विचार करते हैं तो शेष औसतों की अनदेखी भी हो रही होती है। किंतु F-कसौटी में सभी एक साथ विचार होता है।
- दूसरे, युग्मानुसार तुलनाएं करने पर सत्य होने पर भी किसी संकल्पना (Null hypothesis) को अमान्य करने की संभावना बढ़ जाती है। कारण इतना ही है कि जितनी अधिक t-जांच होगी, संयोग मात्र से ही महत्त्वपूर्ण अंतर पाने की संभावना भी उतनी ही अधिक हो जाएगी।
- तीसरे, जितनी अधिक औसतें होंगी, उतनी ही अधिक t-जाँचें करनी होंगी। उदाहरण के लिए, 3 औसतों में 3 बार युग्मानुसार जांच होगी, किंतु 5 में तो 10 बारे युग्मानुसार जांच आवश्यक होगी तो 10 औसतों में युग्मानुसार जांचों की संख्या 45 हो जाएगी।

तीन या अधिक औसतों के लिए F-कसौटी के लिए मान्यताएँ इस प्रकार हैं :

- जिन समष्टियों से प्रतिदर्श लिए गए हैं उनके आवंटन प्रसामान्य या प्रसामान्य-प्रायः हों।
- प्रतिदर्श परस्पर स्वतंत्र हों
- समष्टियों के विचरण समान हों।

एक बात पर गौर करें – हम भले ही तीन या अधिक औसतों की तुलना करना चाहते हैं किंतु इस F-कसौटी के अनुप्रयोग में औसतों के स्थान पर विचरणों का ही इस्तेमाल किया जाता है। F-कसौटी के साथ समष्टि विचरण के दो अलग-अलग अनुमान तैयार किए जाते हैं। पहले अनुमान को समूहों के बीच विचरण करते हैं और यह औसतों के विचरण का प्रयोग करता है। दूसरा अनुमान समूह के भीतर विचरण कहा जाता है। यहां सारे आंकड़ों का प्रयोग कर विचरण का आकलन किया जाता है। यह समूहों की औसतों के अंतर से प्रभावित नहीं होता। यदि औसतों में कोई अंतर नहीं है तो समूहों के बीच तथा समूह का विचरण भी समान प्रायः होगा। अतः F का आकलित मान इकाई के निकट ही होगा। शून्यक संकल्पना अमान्य नहीं होगी। किंतु यदि औसतों के बीच महत्त्वपूर्ण अंतर हों तो समूहों के बीच का विचरण समूह के विचरण से अधिक हो जाएगा, F का मान इकाई से अधिक होगा और शून्यक संकल्पना अमान्य हो जाएगी। यहां हम विचरणों की तुलना कर रहे हैं। इसीलिए इस विधि को विचरण विश्लेषण ANOVA कहते हैं। तीन या अधिक औसतों में अंतरों की समीक्षा के लिए हमें इन संकल्पनाओं का प्रयोग करना चाहिए।

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_k$$

H_1: = कम से कम एक औसत अन्यों से भिन्न है। इस F-सांख्यिक में d.f.N. = k – 1 जहाँ k समूहों की संख्या है तथा d.f.D. = N – k जहां N सभी समूहों के प्रतिदर्शों के आकार का योग है : $N = n_1 + n_2 + \dots\dots\dots + n_k$ । प्रतिदर्शो का आकार समान होना आवश्यक नहीं है। औसतों की तुलना की F-कसौटी सदैव दक्षिण-पुच्छीय ही होगी। हमारा निम्नलिखित उदाहरण ANOVA की आकलन विधि समझ रहा है।

उदाहरण : एक शोधकर्ता उच्च रक्तचाप पीड़ितों के रक्तचाप को कम करने की तीन पृथक्-पृथक् विधियों का प्रयोग कर देखना चाहता है। उन सभी रोगियों को यादृच्छिक

रूप से तीन समूहों में बांट दिया जाता है (पहला समूह दवा लेता है, दूसरा व्यायाम करता है तथा तीसरा एक विशेष आहार पद्धति अपनाता है। चार सप्ताह बाद सभी के रक्तचाप में हुई कमी का मापन किया जाता है। $\alpha = 0.05$ पर यह जांच की जाती है है कि तीनों समूहों की औसतों में कोई अंतर नहीं है। आंकड़े निम्न तालिका में दिखाए गए हैं :

औषधि-प्रयोग / चिकित्सा	व्यायाम	भोजन
10	6	5
12	8	9
9	3	12
15	0	8
13	2	4
$\bar{X}_1 = 11.8$	$\bar{X}_2 = 3.8$	$\bar{X}_3 = 7.6$
$s_1^2 = 5.7$	$s_2^2 = 10.2$	$s_3^2 = 10.3$

हल :

सोपान 1 : संकल्पनाएं बताएं और किए जा रहे दावे की पहचान करें

$H_0: \mu_1 = \mu_2 = \mu_3$ (दावा)

H_1: कम से कम एक औसत अन्यों से भिन्न है।

सोपान 2 : निर्णायक मान ज्ञात करें k = 3 और N = 15,

d.f.N = k – 1 = 3 – 1 = 2

d.f.D = N – k = 15 – 3 = 12

निर्णायक मान C, तालिका H के अनुसार, 3.89 है। यहां $\alpha = 0.05$

सोपान 3 : यहां वर्णित विधि का प्रयोग कर कसौटी मान का आकलन करें :

क) प्रत्येक प्रतिदर्श के औसत और विचरण ज्ञात करें (ये मान उपर्युक्त तालिका में आंकड़ों के अंतर में दिखाए गए हैं)

ख) सामूहिक (बृहत्र) औसत आकलित करें, यह $\bar{X}_{GM}$ सभी प्रतिदर्शों में सम्मिलित इकाइयों का औसत है

$$\bar{X}_{GM} = \frac{\sum X}{N} = \frac{10+12+9+\cdots+4}{15} = \frac{116}{15} = 7.73$$

यदि प्रतिदर्शों के आकार समान हों तो $\bar{X}_{GM}$ को हम समूहवार $\bar{X}$ का योगकर उसे k से विभाजित कर जान जाते हैं।

ग) समूहों के बीच विचरण s_B^2 का आकलन इस सूत्र द्वारा करें :

$$s_B^2 = \frac{\sum n_i(\bar{X}_i - \bar{X}_{GM})^2}{k-1}$$

$$= \frac{5(11.8-7.73)^2 + 5(3.8-7.73)^2 + 5(7.6-7.73)^2}{3-1}$$

$$= \frac{160.13}{2} = 80.07$$

ध्यान दें : यह सूत्र औसतों के बीच विचरण का आकलन करने के लिए प्रतिदर्शों के आकार को भार स्वरूप प्रयोग करता है और औसतों में अंतर पर विचार करता है?

घ) समूचे समूह का विचरण s_w^2 इस प्रकार आकलित करेंगे :

$$s_w^2 = \frac{\sum(n_i - 1)s_i^2}{\sum(n_i - 1)}$$

$$= \frac{(5-1)(5.7) + (5-1)(10.2) + (5-1)(10.3)}{(5-1) + (5-1) + (5-1)}$$

$$= \frac{104.80}{12} = 8.73$$

नोट : यह सूत्र विभिन्न विचरणों की भारित औसत विधि से सकल स्तर पर विचरण का आकलन करता है। यह औसतों के बीच अंतरों का उपयोग नहीं करता।

च) F-सांख्यिक का आकलन करें

$$F = \frac{s_B^2}{s_w^2} = \frac{80.07}{8.73} = 9.17$$

सोपान 4 : अब निर्णय करें। यहां 9.17 > 3.89, अतः शून्यक संकल्पना अमान्य हो जाती है।

सोपान 5 : सभी परिणामों को संक्षिप्त रूप में तालिकाबद्ध करें। दावे को अस्वीकार करने हेतु अनेक साक्ष्य हैं और हमारा निष्कर्ष है कि कम से कम एक औसत अवश्य अन्यों से भिन्न है।

व्यावहारिक कार्य 4 : ANOVA

इसी अनुच्छेद के आंकड़ों पर ANOVA जाँच करें :

यह कैसे करेंगे?

- उदाहरण में दिए गए आंकड़ों (रक्तचाप में कमी) को स्प्रैडशीट पर अंकित करें।
- *Data* Tab से चिन्ह *Analysis* चुनें (यह रिबन पर सबसे अंत में है)। इससे *Data Analysis* संवाद बॉक्स खुल जाएगा। यहां *Anova: Single Factor* चुनें और *OK* बटन दबाएं। अगले संवाद बॉक्स में range *$A$2:$C$7* चुनें (हमारे आंकड़े इसी परिसर में है – देखें चित्र 8.16)। प्रथम पंक्ति बॉक्स में Label चुनें। परिणाम परिसर *$E$2* चुनें और *OK* बटन दबा दें। अब चित्र 8.16 में ही आंकड़े और परिणाम दिखाई दे पाएंगे।

	A	B	C	D	E	F	G	H	I	J	K
1	**Decrease in Blood Pressure**										
2	Medication	Exercise	Diet		Anova: Single Factor						
3	10	6	5								
4	12	8	9		SUMMARY						
5	9	3	12		*Groups*	*Count*	*Sum*	*Average*	*Variance*		
6	15	0	8		Medication	5	59	11.8	5.7		
7	13	2	4		Exercise	5	19	3.8	10.2		
8					Diet	5	38	7.6	10.3		
9											
10											
11					ANOVA						
12					*Source of Variation*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
13					Between Groups	160.1	2	80.067	9.167939	0.003831	3.885294
14					Within Groups	104.8	12	8.7333			
15											
16					Total	264.9	14				
17											

चित्र 8.16 : एनोवा के प्रयोग से प्राप्त परिणाम

आप इन परिणामों की इसी अनुच्छेद में पहले आकलित अनुमानों से तुलना कर सकते हैं।

बोध प्रश्न 6

1) ANOVA की परिभाषा करें। यह तकनीक कब प्रयोग होती है?

....

....

....

....

2) ANOVA में *z*- और t-कसौटियों को नहीं बल्कि F-कसौटी का प्रयोग क्यों करते हैं?

....

....

....

....

3) ANOVA के सूत्र का निरूपण करें और इसकी स्वतंत्रता कोटि की ध्यान से व्याख्या करें।

....

....

....

....

8.8 सार-संक्षेप

द्वि-विचर आंकड़ा विश्लेषण वास्तविक विश्व के परिवेश में दो चरों के बीच संबंधों की व्याख्या करता है। किसी संबंध की उपस्थिति के निर्धारण की विधि सहसंबंध और प्रतीपगमन जैसे सांख्यिकीय उपस्कर हैं। यहां संबंध की प्रबलता और दिशा का ज्ञान सहसंबंध गुणांक का मान प्रदान करता है। इसके मान –1 से +1 तक (दोनों चरम मानों सहित) हो सकते हैं। यह गुणंक +1 या –1 के जितना निकट होता है, उतना ही प्रबल रैखिक संबंध होता है। यदि मान +1 या –1 हो तो संपूर्ण रूप से रैखिक संबंध होता है।

दो चरों में धनात्मक संबंध का अर्थ है कि स्वतंत्र चर के निम्न मानों के साथ निर्भर चर के मान भी निम्न होंगे और बड़े स्वतंत्र चर मानों के साथ निर्भर चर के मान भी बड़े होंगे। दूसरी ओर, ऋणात्मक सहसंबंध का अर्थ है स्वतंत्र चर के मानों में वृद्धि पर निर्भर चर के मान कम हो जाते हैं तो स्वतंत्र चर के निम्न मानों के साथ संबंधित निर्भर चर मान बड़े पाए जाते हैं। दो चरों के बीच संबंध रैखिक या वक्रीय हो सकता है। उसके स्वरूप के रेखांकन के लिए हम पहले चर युग्मों को प्रकीण चित्र के रूप में अंकित करते हैं। यदि संबंध रैखिक है तो आंकड़ों का अनुमान एक सरल रेखा से हो सकता है – जिसे प्रतीपगमन रेखा कहा जाता है। इसी को सर्वश्रेष्ठ संगति रेखा भी कहतेहैं। r का मान +1 या –1 के जितना निकट होगा, प्रकीर्ण बिंदु इस रेखा के उतने ही निकटस्थ होंगे।

8.9 संदर्भ ग्रंथादि

1) Sharma, A.K. (2005). *Text Book of Correlations and Regression*, Discovery Publishing House, New Delhi.

2) O'Brien, D., P. Sharkey Scott, (2012). "*Correlation and Regression*", in Approaches to Quantitative Research – A Guide for Dissertation.

3) Jeremy Miles, Mark Shevlin (2000). *Applying Regression and Correlation: A Guide for Students and Researchers*, Published in UK.

4) Thomas Archdeacon (1994). *Correlation and Regression Analysis: A Historian's Guide*, Paperback, University of Wisconsin Press; USA.

5) R S N Pillai and Bagavathi (2016). *Statistics: Theory and Practice*, S Chand Publishing; Eight edition.

8.10 बोध प्रश्नों के उत्तर अथवा संकेत

बोध प्रश्न 1

1) भाग 8.2 देखें।

2) भाग 8.2 देखें।

बोध प्रश्न 2

1) उपभाग 8.3.1 देखें।

2) उपभाग 8.3.2 देखें।

3) उपभाग 8.3.2 देखें।

4) उपभाग 8.3.3 देखें।

बोध प्रश्न 3

1) उपभाग 8.4.1 देखें।

2) उपभाग 8.4.2 और 8.4.3 देखें।

3) उपभाग 8.4.4 देखें।

4) उपभाग 8.4.5 देखें।

बोध प्रश्न 4

1) उपभाग 8.5.1 देखें।

2) उपभाग 8.5.2 देखें।

बोध प्रश्न 5

1) उपभाग 8.6.1 देखें।

2) उपभाग 8.6.1 देखें।

3) उपभाग 8.6.3 देखें।

बोध प्रश्न 6

1) भाग 8.7 देखें।

2) भाग 8.7 देखें।

3) भाग 8.7।

इकाई 9 बहु-विचर आंकड़ों का विश्लेषण (MULTIVARIATE DATA ANALYSIS)

संरचना

9.0 उद्देश्य

9.1 विषय प्रवेश : विशेषताएँ और अनुप्रयोग

9.2 बहु-विचर विश्लेषण क्या है?

9.2.1 बहु-विचर विश्लेषण : सांख्यिकीय स्वरूप

9.2.2 बहु-विचर विश्लेषण : कुछ मूल संकल्पनाएँ

9.3 बहु-विचर तकनीकों का वर्गीकरण

9.4 बहु-विचर तकनीकों के प्रकार भेद

9.4.1 प्रमुख घटक एवं सांझा कारक विश्लेषण (Principal Components and Common Factor Analysis)

9.4.2 बहुपद प्रतीपगमन एवं बहुपद सह-संबंध (Multiple Regression)

9.4.3 बहुपद विवेचक विश्लेषण और लॉजिस्टिक प्रतीपगमन (Multiple Discriminant Analysis (MDA) and Logistic Regression)

9.4.4 शास्त्रिक सहसंबंध विश्लेषण (Canonical Correlation Analysis)

9.4.5 विचरण का बहुपद विश्लेषण (Multivariate Analysis of Variance)

9.4.6 सहबद्धता विश्लेषण (Conjoint Analysis)

9.4.7 संगुच्छ विश्लेषण (Cluster Analysis)

9.4.8 प्रत्यक्षज्ञानात्मक निरूपण (Perceptual Mapping)

9.4.9 संगतता विश्लेषण (Correspondence Analysis)

9.4.10 संरचनात्मक समीकरण प्रतिमानन (Structural Equation Modelling)

9.5 बहु-विचर तकनीकों एवं उनकी व्याख्या हेतु मार्ग दर्शिकाएँ

9.5.1 व्यावहारिक एवं सांख्यिकीय महत्त्व, दोनों का निरूपण करें

9.5.2 ध्यान रखें कि प्रतिदर्श का आकार परिणामों को प्रभावित करता है

9.5.3 अपने आँकड़ों को समझकर कार्य करें

9.5.4 प्रतिमान में कल्पना-लाघव (Parsimony) हेतु प्रयासरत् रहें

9.5.5 अपनी गलतियों को जान कर सुधारें

9.5.6 अपने परिणामों का सत्यापन करें

9.5.7 बहु-विचर विश्लेषण में सॉफ्टवेयर अनुप्रयोग

9.6 बहु-विचर प्रतिमानन के प्रति एक संरचित दृष्टिकोण

9.7 सार-संक्षेप

9.8 बोध प्रश्नों के उत्तर अथवा संकेत

9.9 संदर्भ ग्रंथादि

9.0 उद्देश्य

बहु-विचर विश्लेषण पर इस इकाई का अध्ययन करने के उपरांत, आप सांख्यिकीय दृष्टि से इन प्रश्नों को समझकर इनका उत्तर दे पाएंगे :

- यह व्याख्या कर पाएंगे कि बहु-विचर विश्लेषण क्या है और कहां इसका अनुप्रयोग उपयुक्त रहता है;
- मापन पैमानों, मापन त्रुटियों और बहु-विचार विश्लेषण पर उनके प्रभावों को समझ पाएंगे;
- बहु-विचर विश्लेषण में सम्मिलित तकनीकों की परिभाषाएँ कर पाएंगे;
- यह तय कर पाएंगे कि किस शोध समस्या के लिए कौन-सी बहु-विचर तकनीक उपयुक्त होगी;
- बहु-विचर विश्लेषण के अनुप्रयोग और व्याख्या के दिशा-निर्देशों पर चर्चा कर पाएंगे; तथा
- बहु-विचर प्रतिमान सृजन की विधि समझा पाएंगे।

9.1 विषय-प्रवेश : विशेषताएँ और अनुप्रयोग

इकाई 7 और 8 में आपका परिचय क्रमशः एक-विचर एवं द्वि-विचर आंकड़ा विश्लेषण की सांख्यिकीय विधियों से हुआ है। किंतु आपको यह भी अनुभूति होगी कि सामाजिक विज्ञान के विविधि विषयों, विशेषकर अर्थशास्त्र में तो प्रायः एक या दो विचर आधारित सांख्यिकी तकनीकें पर्याप्त नहीं होतीं। यहाँ तो दो से अधिक विचरों में परस्पर अनुक्रियाएँ चलती हैं, वे परस्पर पोषण करती हैं और इसी कारण से उन अवस्थाओं की व्याख्या कर पाना बहुत जटिल हो जाता है।

परिणामस्वरूप, पिछले पचास वर्षों में, अनेक सांख्यिकीविद्ों ने बहु-विचर विश्लेषण विधियों के विकास में योगदान दिए हैं। आज उन तकनीकों का अर्थशास्त्र, समाजशास्त्र, मनोवैज्ञानिक, कृषि, नृ-शास्त्र, जीव विज्ञान और चिकित्साशास्त्र आदि में व्यापक अनुप्रयोग हो रहा है। ये तकनीकें विशेषकर उन संदर्भों में अधिक उपयोगी पाई गई हैं जहाँ सामाजिक, मनोवैज्ञानिक, चिकित्सीय एवं आर्थिक आँकड़ों के पारस्परिक प्रभाव दिखाई देते हैं तथा सुदृढ़ प्रायिकता आधारित प्रतिमानों का प्रयोग करना उपयुक्त नहीं लगता। उच्च गति से कार्य करने वाले इलैक्ट्रॉनिक कंप्यूटरों के विकास ने तो इन बहु-विचर तकनीकों के प्रयोग को बहुत बढ़ावा दिया है।

बहु-विचर तकनीकें मुख्यतः आंकड़ों पर आधारित एवं वास्तविकता से संबंधित होती हैं। इनमें जटिल आंकड़ों के विश्लेषण की क्षमता होती है। अतः अधिकांश प्रायोगिक और व्यावहारिक शोध कार्यों में अधिक वास्तविकतापूर्ण परिणाम पाने के लिए हम बहु-विचर विश्लेषण तकनीकों का प्रयोग करते हैं। ये तकनीकें आँकड़ों के विश्लेषण के साथ-साथ विभिन्न प्रकार से निर्णय प्रक्रिया में भी सहायक होती हैं। उदाहरणतः एक कॉलेज प्रवेश प्रतियोगिता पर विचार करें जिसमें अभ्यर्थियों को कई परीक्षाएँ देनी होती हैं और अनेक विषयों में उच्चतम अंक योग पाने वालों को ही प्रवेश मिल पाता है। यह विधि उचित तो प्रतीत होती है किंतु प्रायः इसमें ऐसे कुछ विषयों के प्रति अभिनति भी

हो सकती है जिनमें मानक विचलन अधिक हो। यहाँ बहु-विचर तकनीकों का उपयोग कर प्रवेश की उचित कसौटी विकसित हो सकती है। हम चिकित्सा क्षेत्र से भी उदाहरण ले सकते हैं। रोगियों की अनेक रक्तचाप और कोलेस्ट्रॉल आदि की जाँच की जाती है। ऐसी प्रत्येक जाँच के परिणामों का अपना महत्त्व होता है किंतु विभिन्न जाँच परिणामों में परस्पर और विभिन्न समय पर उनके अपने-अपने परिणामों में संभावित सह-संबंधों को समझ लेना उपयुक्त निदान निष्कर्ष पाने तथा सही चिकित्सा निर्धारित करने में बहुत उपयोगी होगा। इन परिस्थितियों में बहु-विचर तकनीकें बहुत सहायक हो सकती हैं। कहा जा सकता है कि, ''यदि शोधकर्ता निष्कर्ष पाना चाहता है तो कोई बहु-विचर सांख्यिकीय तकनीक ही उपयुक्त होगी।'' यह इकाई बहु-विचर विश्लेषण का एक सरल सिंहावलोकन प्रस्तुत कर रही है। यह भी आग्रह किया जा रहा है कि बहु-विचर विश्लेषण न केवल शोध विश्लेषण को प्रभावित करेगा बल्कि निर्णय करने एवं समस्या का निदान निवारण करने के लिए शोध प्रकल्पना और आंकड़ों के संकलन पर भी इसकी छाप रहती है। इस विधि में अनेक बातें एकल या द्वि-विचर विश्लेषण से मिलती-जुलती हो सकती हैं किंतु बहु-विचर की ओर संक्रमण कुछ निर्णायक कारकों पर निर्भर होता है। इस संक्रमण का उद्धरण देने के लिए इस इकाई में विभिन्न बहु-विचर तकनीकों का वर्गीकरण किया गया है। फिर भी विस्तृत विवरण को बहु-विचर प्रतीपगमन एवं सह-संबंध तक सीमित रखना इकाई के कलेवर की विवशता है। एक प्रारंभिक पाठ्यक्रम होने के नाते यहां बहु-विचर विश्लेषण के सभी आयामों की व्याख्या कर पाना संभव नहीं होगा। हाँ, विभिन्न उपलब्ध तकनीकों के प्रयोग, प्रतिमान संरचना, आकलन और परिणामों की व्याख्या के लिए संरचनात्मक दिशा-निर्देश हम अवश्य दे रहे हैं।

9.2 बहु-विचर विश्लेषण क्या है?

9.2.1 बहु-विचर विश्लेषण : सांख्यिकीय स्वरूप

बहु-विचर विश्लेषण में अन्वेषणा या शोधगत व्यक्तियों/पदार्थों विषयक अनेक मापनों का एक साथ विश्लेषण किया जाता है। कुछ शोधकर्ता दो से अधिक विचरों के बीच संबंधों के विश्लेषण में ही इन तकनीकों का प्रयोग कर लेते हैं। अन्य वहीं इनका प्रयोग करते हैं जहाँ अनेक विचर हों और उन सबके प्रसामान्य बहु-विचर आवंटन के अनुसरण की मान्यता संभव हो। सटीक रूप से बहु-विचरता वहीं होती है जहाँ सभी विचर यादृच्छिक हों तथा इस प्रकार परस्पर संबंधित हो कि उनके विभिन्न प्रभावों का अलग-अलग अनुमान कर व्याख्या कर पाना संभव नहीं हो।

कुछ विद्वानों का मत है कि बहु-विचर विश्लेषण का ध्येय विभिन्न विचरों में संबंध की कोटि (उनके भारित संयोग) का मापन, व्याख्या और पूर्वाकलन करना है। अतः इसका बहु-विचरीय स्वरूप केवल विचर संख्या या अवलोकनों की बहुलता में नहीं होता — यह तो उनके बहुलतापूर्ण संयोजनों में निहित होता है। इस इकाई में हम दोनों प्रकार की तकनीकों से आपका परिचय करा रहे हैं क्योंकि हमारा विचार है कि बहु-विचर तकनीकों का ज्ञान इस विश्लेषण विधि को समझने की ओर पहला कदम होगा।

9.2.2 बहु-विचर विश्लेषण : कुछ मूल संकल्पनाएँ

यद्यपि बहु-विचर विश्लेषण भी एकल एवं द्वि-विचर विधियों पर ही आधारित है, फिर भी विश्लेषण को दो से अधिक चरों पर विस्तारित करते समय कुछ विशेष संकल्पनाओं और मुद्दों पर ध्यान देना आवश्यक हो जाता है। ये विचर की आधारिक संकल्पना से लेकर प्रयुक्त मापन पैमानों तथा जाँच/परीक्षण में महत्त्व के स्तर और विश्वस्तता के स्तर के मुद्दे हैं।

विचर

बहु-विचर विश्लेषण की आधारभूत इकाई 'विचर' है– यह चरों के आंकड़ा आधारित भार मानों सहित उनके रैखिक संयोजन का ही नाम है। चरों का निरूपण शोधकर्ता द्वारा किया जाता है, किंतु उनके भार मान बहु-विचर तकनीकों द्वारा इस प्रकार निर्धारित होते हैं कि किसी लक्ष्य विशेष की प्राप्ति हो सके। हम 'n' भारित चरों (X_1 to X_n) के संयोजन से निर्मित विचर को गणितीय रूप में इस प्रकार दर्शा सकते हैं :

विचर मान= $w_1X_1+w_2X_2+w_3X_3+ \ldots\ldots+W_nX_n$

यह विभिन्न चरों के संपूर्ण समुच्चय का एक ही संख्या में समाया हुआ मान है जो बहु-विचर विश्लेषण के किसी लक्ष्य विशेष को सर्वोपयुक्त रूप में प्राप्त कर सकता है। यहाँ यह विचर विभिन्न स्वतंत्र चरों तथा निर्भर चर के बीच सह-संबंध को अधिकतम कर रहा है। विवेचक विश्लेषण में विचर का निरूपण इस प्रकार किया जाता है कि प्रत्येक अवलोकन के सूचक मान ऐसे रहें कि विभिन्न अवलोकन वर्गों में अधिकतम भेद स्पष्ट हो सकें। प्रत्येक दशा में विचर विश्लेषण की बहुविचरता के स्वरूप को ही प्रगृहित कर रहा है। अतः हमारी तकनीक विषयक चर्चा में अनेक प्रकार से यह विचर का स्वरूप ही विश्लेषण का केंद्र बिंदु रहेगा।

मापन के पैमाने

हम इकाई 2 में पहले ही मापीय, गैर-मापीय, संख्यावाची तथा अनुक्रमवाची पैमानों पर चर्चा कर चुके हैं। ये मापन के पैमाने हमारे उपलब्ध आंकड़ों के लिए उपयुक्त बहु-विचर तकनीक का निर्धारण करने में बहुत उपयोगी होंगे– हम निर्भर और स्वतंत्र, दोनों प्रकार के चरों का ध्यान रखते हुए इन पैमानों का चयन करते हैं।

मापन की त्रुटि तथा बहु-विचर मापन

यह माना जाता है कि प्रत्येक अवलोकन वास्तविक मान तथा एक यादृच्छिक त्रुटि और किसी व्यवस्थित त्रुटि का योगफल होता है। यादृच्छिक त्रुटि पूरे प्रतिदर्श में चर के मान को यादृच्छिक रूप से प्रभावित करने वाले कारकों का परिणाम होती है। उदाहरण के लिए, किसी व्यक्ति की मनःस्थिति (या मूड) किसी भी समय पर उसके निष्पादन को बढ़ाने या घटाने का काम कर सकती है। व्यवस्थित त्रुटि प्रतिदर्श के सभी सदस्यों को एक ही प्रकार से प्रभावित करती है। ऐसी त्रुटियों को न्यूनतम रखने के सभी प्रयासों के साथ-साथ कई बार शोधकर्ता समायोजित पैमानों जैसे बहु-विचर मापक भी विकसित कर लेता है – यहां उक्त पैमाने को ध्यान में रखते हुए अनेक

बार अवलोकन किया जाता है तथा समान प्रायः परिणामों की औसत का प्रयोग कर संयुक्त मापक जिन्हें 'संयुक्त सूचक' भी कहते हैं विकसित किए जाते हैं।

मापन त्रुटि और न्यून विश्वस्तता सीधे से दिखाई नहीं देतीं क्योंकि ये अवलोकित चरों के मानों में ही निहित होती है। मापन त्रुटियों के कारण परिणामों की गुणवत्ता न्यून होना आवश्यक नहीं होता किंतु यह अवलोकित अंतर्संबंधों को दूषित कर बहु-विचर तकनीकों को कुछ क्षीण अवश्य कर सकती हैं।

सांख्यिकीय त्रुटियों के प्रकार और सांख्यिकीय सामर्थ्य

इन संकल्पनाओं पर भी इकाई 4 में चर्चा हो चुकी है। ये यहाँ भी बहुत सार्थक होती हैं। यहाँ विश्लेषणकर्ता को प्रतिदर्श त्रुटि का निरूपण करना होता है यह बहु-विचर विश्लेषण के परिणामों की व्याख्या करने या उनसे सांख्यिकीय निष्कर्ष प्राप्त करने में बहुत आवश्यक होता है। इस प्रकार, एकल एवं द्वि-विचर विश्लेषण की संकल्पनाओं का विस्तार कर हम अगले भाग में एक वर्गीकरण विधि प्रस्तुत कर रहे हैं जो शोध उद्देश्यों एवं आँकड़ों के प्रकार भेद के अनुसार उपयुक्त तकनीकों के चयन में सहायक रहेगी।

बोध प्रश्न 1

1) 'बहु-विचर विश्लेषण' की परिभाषा करें। इसकी विशेषताएँ और अनुप्रयोगों पर भी चर्चा करें।

....

....

....

....

2) मापन का पैमाना, मापन त्रुटि, सांख्यिकीय त्रुटि तथा सांख्यिकीय सामर्थ्य से बहु-विचर विश्लेषण किस प्रकार एवं क्यों प्रभावित होता है?

....

....

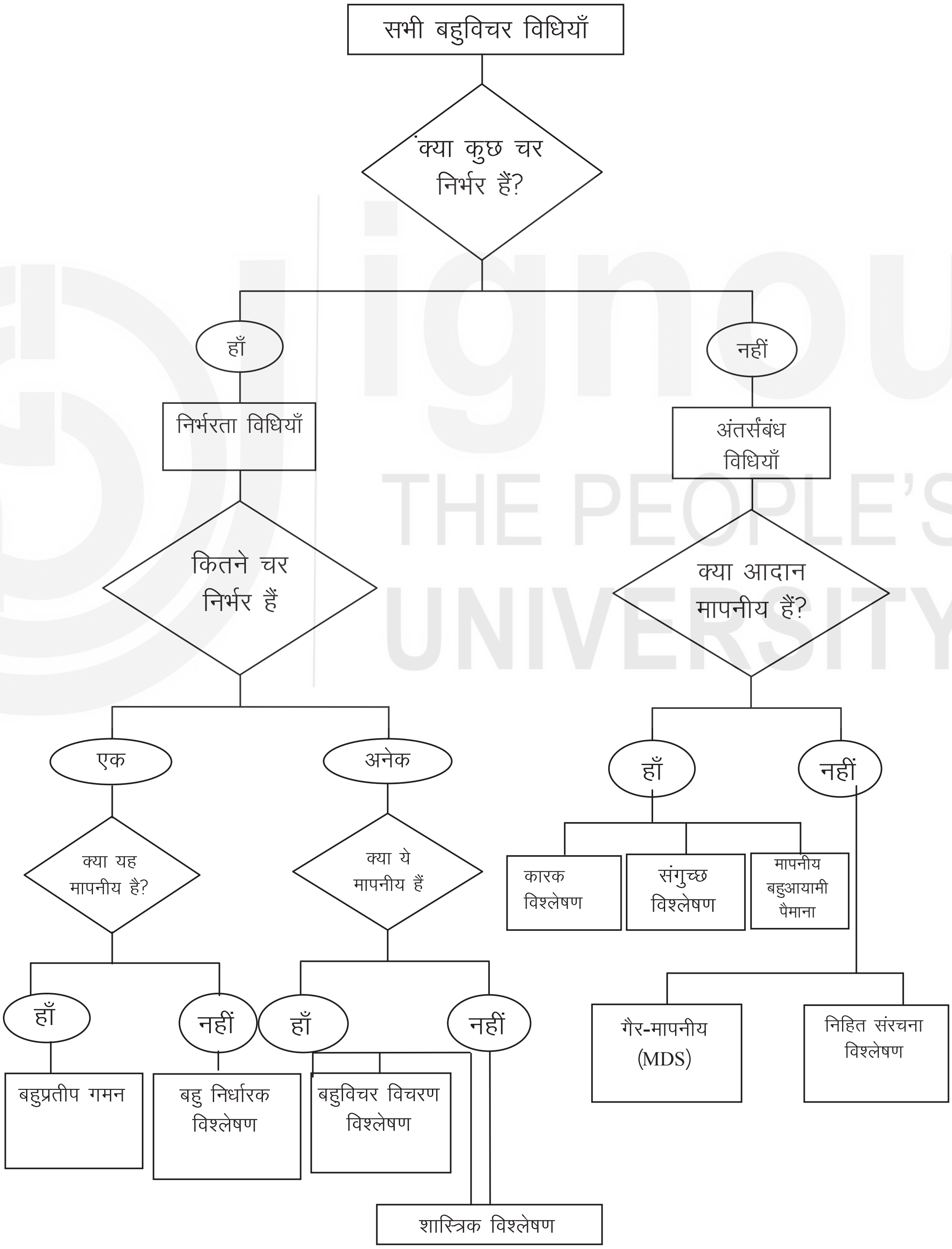
....

9.3 बहु-विचर तकनीकों का वर्गीकरण

आज अनेक प्रकार की बहु-विचर तकनीकें उपलब्ध हैं जिन्हें दो मुख्य वर्गों में बाँटा जा सकता है : निर्भरता विधियाँ और अंतर्निर्भरता विधियाँ। इस वर्गीकरण का आधार यही है कि क्या कुछ संबद्ध चर अन्यों पर निर्भर हैं? यदि हाँ तो निर्भरता विधि होगी। यदि नहीं तो हमें अंतर्निर्भरता विधि का प्रयोग करना होगा। बहु-विचर तकनीकों के स्वरूप को समझने के लिए दो और प्रश्न भी उपयोगी रहते हैं। प्रथम, यदि कुछ चर निर्भर हैं तो उनकी संख्या कितनी है। दूसरा प्रश्न है कि क्या आँकड़े मापनीय हैं? दूसरे शब्दों में, क्या आँकड़े परिमाणात्मक है और उन्हें अंतराल या अनुपातिक पैमाने पर मापा गया है या वे गुणवाची हैं जिन्हें अभिहित (nominal) या क्रमवाची पैमाने के अनुसार एकत्र किया गया है। इन सभी प्रश्नों के उत्तर ही संदर्भ विशेष में प्रयोग होने वाली तकनीकों

का निर्धारण करेंगे। कुछ महत्त्वपूर्ण बहु-विचर तकनीकों का चित्र 9.1 में प्रवाह चित्रांकन के रूप में दिया गया है।

हमारे पास दो प्रकार की बहु-विचर तकनीके हैं : एक जिनमें निर्भर एवं स्वतंत्र, दोनों प्रकार के चरों पर आंकड़े होते हैं तथा दूसरी वे जिनमें अनेक चर हैं किंतु किसी प्रकार की (एक दिश) निर्भरता नहीं होती। पहले वर्ग में बहुचर प्रतीपगमन बहु निर्धारक विश्लेषण, बहु-विचर विचरण विश्लेषण तथा शास्त्रिक विश्लेषण शामिल हैं। दूसरे वर्ग में, हम कारक विश्लेषण, संगुच्छ विश्लेषण बहुआयामी पैमाना (MDS), जिनमें मापनीय एवं गैर-मापनीय दोनों सम्मिलित हैं और निहित संरचना विश्लेषण आदि को समाहित करते हैं।



बोध प्रश्न 2

1) बहु-विचर तकनीकों के वर्गीकरण का आधार समझाइए।

....

....

....

....

....

9.4 बहु-विचर तकनीकों के प्रकार भेद

आंकड़ा विश्लेषण की बहु-विचर तकनीकों का समूह निरंतर प्रसार कर रहा है और पिछले भाग में बताए गए संभावित शोध वर्गीकरण से ही यह बात स्पष्ट हो जाती है। पूर्व स्थापित एवं उदीप्यमान तकनीकों में निम्नलिखित शामिल हैं :

1) प्रमुख घटक और सांझा कारक विश्लेषण (Principal Components and Common Factor Analysis)
2) बहुपद प्रतीपगमन और बहुपद सहसंबंध (Multiple Regression and Multiple Correlation)
3) बहुपद विवेचक विश्लेषण और लॉजिस्टिक प्रतीपगमन (Multiple Discriminant Analysis and Logistic Regression)
4) शास्त्रिक सहसंबंध विश्लेषण (Canonical Correlation Analysis)
5) बहु-विचर विचरण विश्लेषण (Multivariate Analysis of Variance)
6) संयुक्त विश्लेषण (Conjoint Analysis)
7) संगुच्छ विश्लेषण (Cluster Analysis)
8) प्रत्यक्ष ज्ञानात्मक निरूपण (Perceptual Mapping)
9) संगतता विश्लेषण (Correspondence Analysis)
10) संरचनात्मक समीकरण प्रतिमान (Structural Equation Modelling)

आइये, अब संक्षेप में इन तकनीकों की परिभाषाएँ करते हुए इनके उपयोग के ध्येय जानने का प्रयास करें।

9.4.1 प्रमुख घटक एवं सांझा कारक विश्लेषण

ये दोनों सांख्यिकीय विधियाँ बड़ी संख्या में चरों के बीच अंतर्संबंधों के विश्लेषण और किन्हीं सांझे कारक आयामों के आधार पर उन चरों की व्याख्या में उपयोगी होती हैं। उद्देश्य बहुत से चरों में निहित जानकारी को बहुत कम चरों या कारकों में इस प्रकार समाहित करना है कि सूचनाओं की हानि न्यूनतम रहे। संदर्भित चरों की संरचना का संख्यात्मक अनुमान प्रदान कर कारक विश्लेषण योगित पैमाने निर्मित करने का आधार बन जाता है।

9.4.2 बहुपद प्रतीपगमन एवं बहुपद सह-संबंध

हमने आठवीं इकाई में द्विपद विश्लेषण पर चर्चा में साधारण रैखिक प्रतीपगमन तथा सहसंबंध समझाए थे। साधारण रैखिक प्रतीपगमन के समीकरण में एक स्वतंत्र चर x तथा एक ही निर्भर चर y_e होता है :

$$y_e = a + bx$$

बहुपद प्रतीपगमन में निर्भर चर तो एक ही रहता है किंतु स्वतंत्र चर अनेक होते हैं। समीकरण होता है :

$$y_e = a + b_1x_1 + b_2x_2 + b_3x_3 + \ldots\ldots\ldots + b_kx_k$$

जहाँ $x_1, x_2, \ldots, x_k$ स्वतंत्र चर हैं। बहुपद प्रतीपगमन सहसंबंध R का आकलन यह जानने के लिए किया जाता है कि क्या उन सभी स्वतंत्र चरों के समूह तथा निर्भर चर के बीच महत्त्वपूर्ण संबंध विद्यमान है। बहुपद प्रतीपगमन का प्रयोग वहां होता है जहाँ सांख्यिकीविद् यह मान रहा हो कि अनेक स्वतंत्र चर निर्भर चर में परिवर्तनों में योगदान कर रहे हैं। यह विश्लेषण किसी एक स्वतंत्र चर पर निर्भरता की अवस्था की अपेक्षा कहीं बेहतर पूर्वाकलन करने में उपयोगी हो सकता है। बहुपद प्रतीपगमन के दो अच्छे उदाहरण हो सकते हैं : एक, स्टोर का प्रबंधक यह जानने का प्रयास कर सकता है कि विज्ञापन पर व्यय तथा स्टोर में वस्तुओं की प्रदर्शनी के लिए प्रयोग किए गए तल क्षेत्र का किसी वस्तु की बिक्री पर क्या प्रभाव हो रहा है। दूसरे एक समाजविज्ञानी यह जानने का इच्छुक हो सकता है कि किसी बच्चों के भार पर उनके टेलीविज़न देखने और वीडियो गेम खेलने में व्यतीत समय का क्या प्रभाव हो रहा है। बहुपद प्रतीपगमन की मान्यताएँ साधारण प्रतीपगमन जैसी ही होती हैं :

1) स्वतंत्र चर के किसी भी मान विशेष के लिए y चर का आवंटन प्रसामान्य होता है (प्रसामान्यता की मान्यता)
2) y-चर का विचरण या मानक विचलन प्रत्येक स्वतंत्र चर के लिए समान रहता है (सम-विचरण मान्यता)
3) निर्भर चर का स्वतंत्र चरों के साथ रैखिक संबंध है (रैखिकता मान्यता)
4) स्वतंत्र चर सहसंबंधित नहीं हैं (बहु रैखिकता हीनता की मान्यता)
5) y-चर के मान स्वतंत्र होते हैं (स्वतंत्रता की मान्यता)

साधारण प्रतीपगमन की भांति बहुपद प्रतीपगमन में भी स्वतंत्र एवं निर्भर चरों के बीच संबंध की प्रबलता का मापन एक सहसंबंध गुणांक से किया जाता है। यह इसे बहुपद सहसंबंध कहा जाता है और R^2 द्वारा दर्शाया जाता है। इस R^2 का मान 0 से 1 तक हो सकता है, यह कभी ऋणात्मक नहीं होता। यदि R^2 का मान इकाई के निकट हो तो यह संबंध की प्रबलता का परिचायक है जबकि शून्य के निकट होने का अर्थ संबंध की दुर्बलता होता है। R^2 का मान सभी स्वतंत्र चरों पर निर्भर होता है – इसे सभी 'वैयक्तिक' सहसंबंध गुणांकों के मान का प्रयोग कर आकलित किया जा सकता है। द्वि-चर संदर्भ में बहुपद सहसंबंध सूत्र इस प्रकार होगा :

$$R = \sqrt{\frac{r_{yx_1}^2 + r_{yx_2}^2 - 2r_{yx_1}.r_{yx_2}.r_{x_1x_2}}{1 - r_{x_1x_2}^2}}$$

$$R^2 = \frac{\text{व्याख्यित विचरण}}{\text{सकल विचरण}} = \frac{\sum(y_e - \bar{y})^2}{\sum(y - \bar{y})^2}$$

साधारण प्रतीपगमन की ही भाँति R^2 को बहुपद निर्धारक गुणांक कहते हैं। यह प्रतीपगमन प्रतिमान द्वारा व्याख्यित विचरण का अनुपात है। पद $1 - R^2$ अव्याख्यित विचरण है – इसे ही त्रुटि या शेष विचरण कहा जाता है।

R के महत्त्व की जाँच के लिए एक F कसौटी का प्रयोग होता है। यहां पूर्व कल्पनाएँ हैं :

$$H_0 : \rho = 0 \quad \text{तथा} \quad H_1 : \rho \neq 0$$

जहाँ ρ समष्टि का बहुपद सहसंबंध है।

F कसौटी का सूत्र है :

$$F = \frac{R^2/(k-1)}{(1 - R^2)/(n-k)}$$

जहाँ n आंकड़ा समूहों ($x_1, x_2, \ldots\ldots\ldots\ldots y$) तथा k स्वतंत्र चरों की संख्या है।

अंश की स्वतंत्रता कोटि है $= n - k$ तथा 'हर' की स्वतंत्रता कोटि है $= n - k - 1$.

समंजित R^2

R^2 का मान n पर तथा चरों की संख्या k पर निर्भर है। इसी कारण सांख्यिकीविद् एक समंजित R^2 भी आकलित करते हैं, जिसका सूत्र है :

$$R_{adj}^2 = 1 - \frac{(1 - R^2)(n-1)}{n - k - 1}$$

यह समंजित R^2 असमंजित से कम होता है। इसमें इस बात का हिसाब हो जाता है कि n और k की निकटता के कारण R का मान कृत्रिम रूप से अधिक हो सकता है। इसका कारण चरों के बीच सही संबंध नहीं बल्कि प्रतिचयन की त्रुटियां हो सकती है। यह इसलिए होता है कि प्रतीपगमन समीकरण की व्युत्पत्ति में सभी चरों में यादृच्छिक परिवर्तनों को परस्पर एकयुत रूप में प्रयोग किया जाता है। यदि निर्भर चर का प्रत्येक स्वतंत्र चर से सहसंबंध शून्य हो तो भी प्रतिचयन की त्रुटियों के कारण बहुपद सहसंबंध का मान शून्य से अधिक हो सकता है। अतः बहुपद प्रतीपगमन विश्लेषण में R^2 तथा R^2_{adj} दोनों ही प्रदर्शित किए जाते हैं।

9.4.3 बहुपद विवेचक विश्लेषण (MDA) और लॉजिस्टिक प्रतीपगमन

लॉजिस्टिक प्रतीपगमन तथा बहुपद विवेचक विश्लेषण वे सांख्यिकीय विधियाँ हैं जिनके माध्यम से विभिन्न सह-विचरकों तथा एक निश्चयात्मक परिणाम के बीच साहचर्य का मूल्यांकन किया जा सकता है। चिकित्साशास्त्र एवं सामाजिक विज्ञान के शोध में इन दोनों विधियों का व्यापक रूप से प्रयोग होता रहा है।

यदि एकल निर्भर चर किन्हीं सुनिश्चित मानों में से किसी एक का प्रग्रहण करता हो (जैसे कि पुरुष और स्त्री या फिर बहुत सारों में से एक (जैसे, उच्च, मध्यम, न्यून) धारण करता हो और इसी कारण इनके मान सांख्यिक हों तो MDA को उपयुक्त बहु-विचर तकनीक माना जाता है।

आँकड़ों के वर्णन तथा एक द्विपदीय चर तथा एक या अधिक सांख्यिक, क्रमवाची, अंतराल या अनुपात मापित स्वतंत्र चरों के संबंध की व्याख्या के लिए लॉजिस्टिक प्रतीपगमन विधि का प्रयोग किया जाता है। यह प्रतीपगमन इस प्रकार के प्रश्नों की समीक्षा कर सकता है : व्यक्ति के अधिक भार में एक पाउंड की और वृद्धि या प्रतिदिन उसके धूम्रपान में एक पैकेट की वृद्धि से उसके फेफड़ों में कैंसर होने (हाँ या नहीं) की प्रायिकता किस प्रकार परिवर्तित होगी?

या फिर शरीर के भार, कैलोरी उपभोग, वसा उपभोग तथा अध्ययन में भागीदार की आयु का 'हृदय घात' पर प्रभाव होता है (हाँ या नहीं)।

9.4.4 शास्त्रिक सहसंबंध विश्लेषण (CCA)

यह शास्त्रिक सहसंबंध विश्लेषण चरों के दो समुच्चयों के बीच रैखिक निर्भरता का परिमाण आकलित करने की विधि है। अधिक सटीक रूप से CCA विधि दोनों समुच्चयों से एक आंकड़ा समुच्चय के चरों के रैखिक संयोजन इस प्रकार ज्ञात करता है कि उन संयोजनों के बीच अधिकतम सहसंबंध हो। इस प्रकार प्राप्त सहसंबंध को ही शास्त्रिक सहसंबंध कहते हैं। यह सहसंबंध के विचार को दो से अधिक चरों तक विस्तृत कर देता है।

9.4.5 विचरण का बहुपद विश्लेषण (MANOVA)

t कसौटी का ध्येय इस संभाविता का आकलन होता है कि दो समूहों के औसत मान औसतों के एक ही प्रतिदर्श आवंटन से प्राप्त किए गए हैं। इस t कसौटी का बहु-विचर समतुल्य होटेलिंग की T^2 कसौटी है। यह कसौटी इस बात की जांच करती है कि क्या दो समूहों के औसत मानों के सदिश एक ही प्रतिदर्श आवंटन से व्युत्पन्न हुए हैं। MANOVA का ध्येय भी यही जानना होता है। जैसे होटेलिंग T^2 एक ही बड़े समूह से दो यादृच्छिक औसत सदिश पाने की संभाव्यता का मापन करता है, वैसे ही MANOVA उसी समूह में से दो या अधिक यादृच्छिक औसत पाने की कुल संभाव्यता का मापन प्रदान करता है।

मुख्य रूप से दो स्थितियों में MANOVA विधि का प्रयोग होता है। एक तो वहां जब अनेक सहसंबंधित निर्भर चर विद्यमान हों तथा शोधकर्ता इन चरों पर अनेक युग्मस्तरीय परीक्षण आकलन करने के स्थान पर उन सभी चरों पर एक सांझा सांख्यिकीय परीक्षण करना चाहता हो। दूसरे (अनेक दशाओं में) यह जानने का ध्येय अधिक महत्त्वपूर्ण हो सकता है कि स्वतंत्र चर किस प्रकार से निर्भर चरों पर प्रभाव-चित्रण कर रहे हैं।

उदाहरण के लिए, शोधकर्ता का ध्येय विभिन्न चिकित्सा विधियों के रोगी की व्यग्रता के विभिन्न प्रकारों पर प्रभावों की समीक्षा करना हो सकता है – जैसे कि परीक्षण व्यग्रता, जीवन में छोटी-छोटी तनावपूर्ण बातों को लेकर व्यग्रता और तथाकथित सामान्य (रूप से निरंतर प्रवाहमान) व्यग्रता। यहां स्वतंत्र चर (IV) विभिन्न तीन स्तरों

वाली चिकित्साएँ हैं : (वि-संवेदन, शिथिलन और प्रतीक्षा-सूची नियंत्रण) विभिन्न व्यक्तियों पर इनमें से सभी विधियाँ यादृच्छिक रूप से अपनाई जाती हैं और फिर उनके तीनों प्रकारों के व्यग्रता स्तरों पर मापन किया जाता है। इनके मापांक ही निर्भर चर (DVs) का कार्य करते हैं। यहाँ MANOVA में पूछा जाता है कि क्या इन व्यग्रता मापकों के किसी संयोजन में चिकित्सा विधि के फलन रूप में कोई परिवर्तन आते हैं।

MANOVA सांख्यिकी दृष्टि से विवेचक विश्लेषण जैसा ही होता है। दोनों में अंतर आग्रह का ही होता है। MANOVA का आग्रह औसत अंतरों तथा विभिन्न समूहों में इन अंतरों के सांख्यिकीय महत्त्व पर होता है। विवेचक विश्लेषण का आग्रह पूर्वाकलन समूह की सदस्यता पर तथा उन आयामों पर होता है जिनमें विभिन्न समूह पृथकतापूर्ण होते हैं।

9.4.6 सहबद्धता विश्लेषण

यह विश्लेषण विधि 1970 के दशक काल से विकसित हुई है। यह व्यवसायों को यह बूझने में सहायक होती है कि यदि ग्राहकों को कोई सांझी पेशकश की जाती है तो वे उसे पेशकश के विभिन्न घटकों (उत्पादों एवं सेवाओं) के बीच समप्रत्ययन करने के किन परोक्ष नियमों का पालन करते हैं तथा उन नियमों को कितना-कितना मान प्रदान करते हैं। लोग किस प्रकार अपना निर्णय करते हैं इस बात को सटीक रूप से जानकर आप अपने उत्पाद एवं सेवाओं का एक ऐसे संयोजन की पेशकश कर सकते हैं जो ग्राहकों को बहुत लुभावना लगे और कंपनी को कम लागत पर स्पर्धी बाज़ार में अपना हिस्सा बढ़ाने में भी सहायक रहे।

9.4.7 संगुच्छ विश्लेषण

संगुच्छ विश्लेषण ऐसा संगवेषणात्मक विश्लेषण है जो आंकड़ों में अंतर्निहित संरचनाओं की पहचान करने का प्रयास करता है। इसे खंड विभाजन या वर्ग विभाजन विधि भी कहा जाता है। यहाँ विभिन्न अवलोकनों, भागीदारों, उत्तर देने वालों के सजातीय उपसमूहों की पहचान करते हैं यदि कोई उपसमूह पूर्व परिभाषित नहीं है तो संगुच्छ विश्लेषण उसे स्पष्ट निरूपण प्रदान करता है। अपने संगवेषणात्मक स्वरूप के कारण यह निर्भर और स्वतंत्र चरों के बीच कोई भेद नहीं करता।

9.4.8 प्रत्यक्षज्ञानात्मक निरूपण

यह निरूपण एक ऐसी दृश्यांकन तकनीक है जो यह दर्शाती है कि औसत लक्षित उपभोक्ता बाज़ार में विभिन्न स्पर्धी उत्पादों के 'स्थान' को क्या समझाता है। दूसरे शब्दों में यह उपस्कर उपभोक्ता की अनुभूतियों और सहजबोध को एक चित्र में अंकित करता है। यहां शब्द 'प्रत्यक्षज्ञानात्मक' को अनुभूतियों एवं उपभोक्ता के सहज ज्ञान से प्रग्रहित किया गया है जो स्पर्धी वस्तुओं और उनके अभिलक्षणों से संबंधित हैं।

इस विश्लेषण में प्रायः प्रयुक्त परिभाषाएं ये हैं : *''प्रत्यक्षज्ञानात्मक चित्र यह मापते हैं कि उपभोक्ता के मन में विभिन्न वस्तुओं की छवियाँ किस प्रकार बसती हैं और इन*

छवियों को वस्तुओं के अभिलक्षणों के अनुसार अंकित रेखाचित्र के रूप में अंकित किया जाता है।''

''एक प्रत्यक्षज्ञान मानचित्र एक दृश्यात्मक स्वरूप में अंतरों के माध्यम से उपभोक्ता की अनुभूतियों और वरीयताओं को दर्शाता है।''

ध्यान दें कि दोनों ही परिभाषाओं में इस बात पर बल दिया गया है कि जानकारी को दृश्य रूप में दिखाया गया है।

9.4.9 संगतता विश्लेषण

संगतता विश्लेषण की सांख्यिकीय तकनीकें तिर्यक तालिकाओं को रेखाचित्र के रूप में दिखाती हैं (इन तालिकाओं को क्रॉस टैब्स या 'आक्समिकता' तालिकाएं भी कहते हैं)। यह उस समय होता है जब किन्हीं घटनाओं को एक से अधिक वर्ग समूहों में रखना संभव हो। उदाहरणः बाज़ार अनुसंधान में वस्तुओं और स्थानों के वर्ग, चिकित्सीय अन्वेषण में रोग और चिकित्सा के वर्ग आदि।

9.4.10 संरचनात्मक समीकरण प्रतिमानन (SEM)

इस तकनीक में प्रत्येक निर्भर चर के लिए एक अलग संबंध सूत्र अनुमित होता है। सरलतम रूप के SEM एक साथ अनुमान की जा रही पृथक-पृथक बहुपद प्रतीपगमन समीकरण शृंखला के लिए सर्वोपयुक्त और दक्षतम आकलन तकनीक होती है।

बहु-विचर तकनीकों की यह संक्षिप्त झलक उनके विविधतापूर्ण स्वरूप एवं तथा सभी प्रकार के वास्तविक जीवन के संदर्भों में उनकी उपादेयता को दिखा रही है। किंतु इनके प्रयोग के लिए न केवल तकनीकों को बल्कि उन परिस्थितियों को भी सटीक रूप से जान लेना ज़रूरी होगा जहां इनका प्रयोग किया जाना है। अतः अवस्था और प्रयोग किए जाने वाली तकनीक के बीच संगति की स्पष्ट रूप से स्थापना करना अत्यावश्यक है। इस कार्य में अगले भाग में बहु-विचर तकनीक तथा उनकी व्याख्या के लिए बताए गए दिशा-निर्देश बहुत उपयोगी सिद्ध होंगे।

बोध प्रश्न 3

1) बहुत ही उदीप्यमान (energing) और परीक्षित बहु-विचर तकनीकों की सूची बनाएं।

....

....

....

2) बहुपद प्रतीपगमन विश्लेषण की आधारिक मान्यताओं का संक्षेप में वर्णन करें।

....

....

....

3) ANOVA तथा MANOVA में क्या अंतर है?

....

....

....

....

....

9.5 बहु-विचर तकनीकों एवं उनकी व्याख्या हेतु मार्ग दर्शिकाएं

9.5.1 व्यावहारिक और सांख्यिकीय महत्त्व, दोनों का निरूपण करें

कई बार शोधकर्ता व्यावहारिक उपादेयता को भुलाकर केवल सांख्यिकीय महत्त्व पर ही निर्भर हो जाते हैं, जबकि परिणाम भ्रामक होते हैं। उदाहरण के लिए, मान लें कि इस बात की 0 से 100 की प्रायिकता आकलित करने के लिए प्रतीपगमन आकलन किया जाता है कि कोई व्यक्ति पुनः किसी फर्म से खरीदारी करने आएगा। अध्ययन के बाद परिणामों को 0.05 पर महत्त्वपूर्ण पाया जाता है। फर्म के अधिकारी बहुत खुश हो अपनी व्यापार युक्ति को इन परिणामों के अनुरूप ढालने में व्यस्त हो जाते हैं। किंतु वे भूल गए हैं कि परिणाम बहुत ही विचित्र थे– यहां तक कि 0.05 महत्त्व पर भी उस व्यक्ति के दुबारा आने की संभाव्यता ±20 प्रतिशत ही थी। अतः महत्त्वपूर्ण सांख्यिकीय संबंध में 40 प्रतिशत का अंतराल था। यदि किसी ग्राहक के दुबारा आने की संभावना 50 प्रतिशत आंकी गई हो तो वास्तव में उसके आने की संभाव्यता 30% से 70% हो सकती थी। ये तो ऐसे स्तर हैं जिनको किसी कार्य नीति का आधार बनाना सुरक्षित नहीं होगा। यदि प्रबंधकों/अधिकारियों ने परिणामों के व्यावहारिक महत्त्व को समझने का प्रयास किया होता तो उन्हें समझ आ जाता कि उस प्रतीपगमन संबंध के और परिष्कार की आवश्यकता थी (तभी उसे व्यापार नीति/युक्ति का आधार बनाना उपयुक्त होता)।

9.5.2 ध्यान रखें कि प्रतिदर्श का आकार परिणामों को प्रभावित करता है

हमारी इकाई 7 की चर्चा यह अच्छी प्रकार से समझा चुकी है कि प्रतिदर्श बड़े हों या छोटे, दोनों दशाओं में प्रतिदर्श के आकार का सांख्यिकीय महत्त्व के आकार पर बहुत प्रभाव रहता है।

9.5.3 अपने आंकड़ों को समझकर कार्य करें

बहु-विचर तकनीकों का प्रयोग उन जटिल संबंधों के अन्वेषण के लिए किया जाता है जिनका निरूपण सहज स्पष्ट नहीं हो। अतः एकल या द्वि-विचर विश्लेषण की भाँति बिना किसी समीक्षा के परिणामों को स्वीकार कर लेना घातक सिद्ध हो सकता है। बहिष्शायियों (outliers), मान्यता उल्लंघन तथा विलुप्त आंकड़ों के प्रभाव अनेक चरों

तक फैलकर परिणामों पर गंभीर प्रभाव डाल सकते हैं। अतः यहाँ आंकड़ों की बहुत ही गंभीरतापूर्वक पूर्व समीक्षा करने की आवश्यकता होती है।

9.5.4 प्रतिमान में कल्पना-लाघव (Parsimony) हेतु प्रयासरत् रहें

बहु-विचर तकनीकें अनेकों चरों को विश्लेषण में समाहित करने का प्रयास करती हैं। किंतु इसका अर्थ यह नहीं होना चाहिए कि इनके प्रयोग से पूर्व सैद्धांतिक दृष्टि से प्रतिमान के निरूपण पर ध्यान ही न दिया जाए। यह तो सदैव महत्त्वपूर्ण होता है कि किसी निर्णायक पूर्वाकलक चर की उपेक्षा नहीं की जाए (जिसे विनिर्देशन-त्रुटि भी कहा जाता है), किंतु दो और कारण भी है जिनके आधार पर शोधकर्ता को अनावश्यक रूप से कारकों का अंबार लगाकर उनमें से चयन करने का काम तकनीकों के भरोसे नहीं छोड़ देना चाहिए। ये हैं : (i) प्रायः अनर्थक चरों की उपस्थिति से तकनीकों के प्रतिदर्श आंकड़ों के साथ संगति कर जाने की क्षमता में तो वृद्धि होती है, किंतु यह परिणामों को उक्त प्रतिदर्श पर अतिनिर्भर बना देने की 'लागत' पर होता है। ऐसे निष्कर्ष समष्टि के लिए मान्य समझना उचित नहीं रहता; (ii) यद्यपि अनर्थक चर सार्थक चरों के गुणांकों के अनुमानों को अभिनतिपूर्ण तो नहीं बना पाते, फिर भी बहु रैखिकता में वृद्धि कर वे सार्थक चरों के सटीक प्रभाव को भी कुछ धुंधला अवश्य कर सकते हैं। हम जानते हैं कि बहु रैखिकता किसी चर के प्रभाव का अन्य चरों द्वारा पूर्वाकलन हो सकने की कोटि को दर्शाती है। अतः यदि बहुरैखिकता में वृद्धि होती है तो किसी भी चर के प्रभाव को अनन्य रूप से परिभाषित कर पाना कठिन हो जाता है। अनर्थक चरों को समाहित करने से बहुरैखिकता में वृद्धि अवश्य होती है और इसके कारण व्याख्या कर पाना अधिक कठिन हो जाता है।

9.5.5 अपनी गलतियों को जान कर सुधारें

शोधकर्ता को पूर्वाकलन की त्रुटियों को बहु-विचर विश्लेषण की विफलता मानकर उन्हें कम करने में नहीं जुट जाना चाहिए। इसे तो प्राप्त परिणामों की प्रतिपुष्टि की समीक्षा का प्रारंभ बिंदु तथा अव्याख्यित रह गए शेष संबंधों का प्रतीक मानना चाहिए।

9.5.6 अपने परिणामों का सत्यापन करें

बहु-विचर विश्लेषण की जटिल संबंधों की पहचान करने की क्षमता का अर्थ यह भी हो सकता है कि हमारे परिणाम किसी प्रतिदर्श विशेष तक सीमित रह जाएं, वह समष्टि तक विस्तार करने योग्य नहीं रहे। अतः शोधकर्ता को इनमें से किसी न किसी विधि का प्रयोग कर अपने परिणामों का सत्यापन अवश्य करना चाहिए :

- प्रतिदर्श को दो उपप्रतिदर्शों में विभाजित कर एक को प्रतिमान आकलन तथा दूसरे को सटीकता के पूर्वाकलन हेतु प्रयोग करके देखें।
- कोई अन्य प्रतिदर्श एकत्र कर यह देखें कि पहले प्रतिदर्श के निष्कर्ष अन्य प्रतिदर्शों के लिए मान्य हैं या नहीं।
- कम से काम चलाने की बूट स्ट्रैप्पिंग विधि का प्रयोग कर दिए हुए प्रतिदर्श में से ही अनेक उपप्रतिदर्श निकाल कर प्रत्येक के लिए अलग से आकलन करके

देखें। फिर उन सबसे प्राप्त अनुमानों के औसत का प्रयोग कर संपूर्ण प्रतिमान के गुणांकों का आकलन करके देखें।

और अंत में, एक बार फिर केवल 'महत्त्वपूर्ण' प्रतिमान के आकलन पर आग्रह नहीं रहना चाहिए – यह भी ध्यान रखना ही चाहिए कि आकलित प्रतिमान समष्टि के लिए भी उपयोगी हो। अतः केवल प्रतिदर्श के लिए सर्वोपयुक्त होने पर नहीं बल्कि ऐसे प्रतिमान की रचना पर ध्यान देना चाहिए जो समष्टि का सर्वश्रेष्ठ निरूपण कर सकता हो।

9.5.7 बहु-विचर विश्लेषण में सॉफ्टवेयर अनुप्रयोग

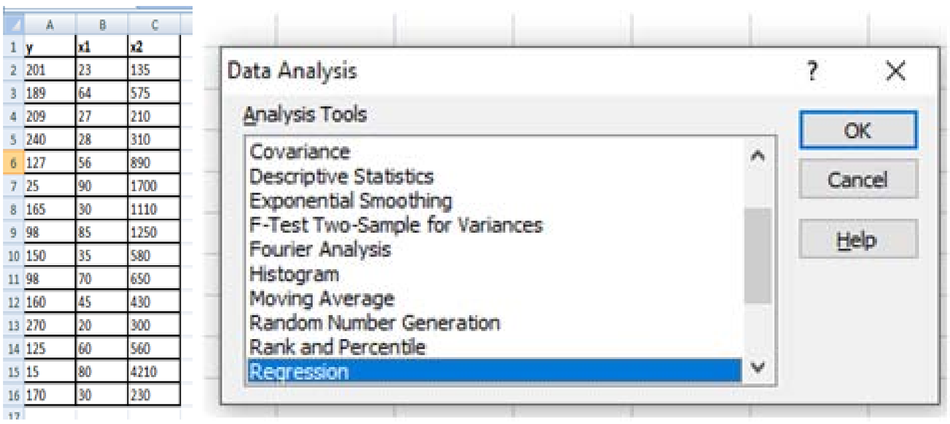
कंप्यूटर सॉफ्टवेयर द्वारा बहु-विचर विश्लेषण करना बहुत सरल होता है। यह आपको कुछ परिणाम अवश्य देगा। किंतु मुख्य बात यही है कि क्या वह परिणाम सार्थक भी है? इसकी जांच के लिए आपके पास बहु-विचर विश्लेषण के प्रति जो ज्ञानपूर्ण अंतर्दृष्टि होनी चाहिए, उसका विकास कर पाना वर्तमान पाठ्यक्रम की सीमा से परे ही है। किंतु यहां दिया जा रहा उदाहरण आपको कंप्यूटर सॉफ्टवेयर द्वारा बहु-विचर विश्लेषण करना अवश्य सिखा सकता है।

उदाहरणः इन आंकड़ों का प्रयोग कर y को निर्भर तथा x1, x2 को स्वतंत्र चर मानते हुए प्रतीपगमन परिणाम आकलित करें :

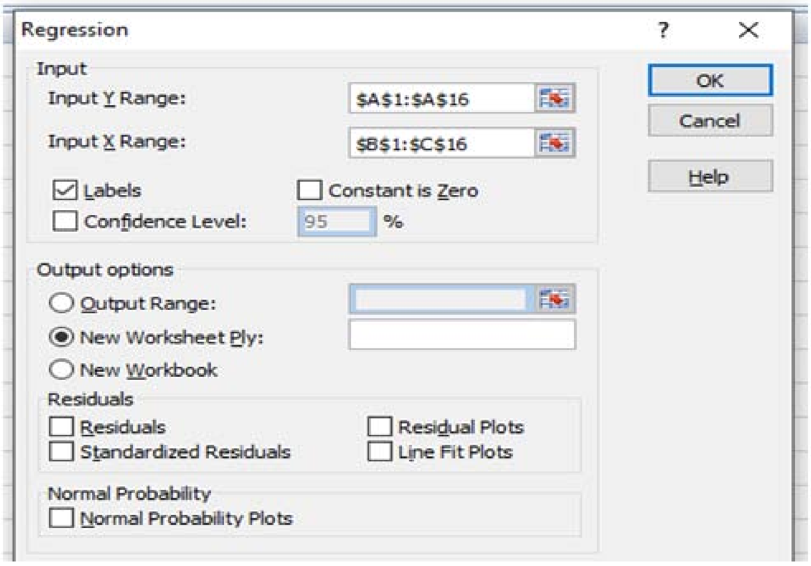
y	201	189	209	240	127	25	165	98	150	98	160	270	125	15	170
x_1	23	64	27	28	56	90	30	85	35	70	45	20	60	80	30
x_2	135	575	210	310	890	1700	1110	1250	580	650	430	300	560	4210	230

समाधान : ध्यान दें कि x_1 तथा x_2 के परिसर (range) एक जैसे नहीं हैं। यदि ऐसे आंकड़ों से वास्ता पड़े तो पहले उन्हें सम-प्रायः मान प्रदान करना उचित रहता है। इस उदाहरण में आप सीधे MS-Excel में बहुपद प्रतीपगमन का प्रयोग कर सकते हैं।

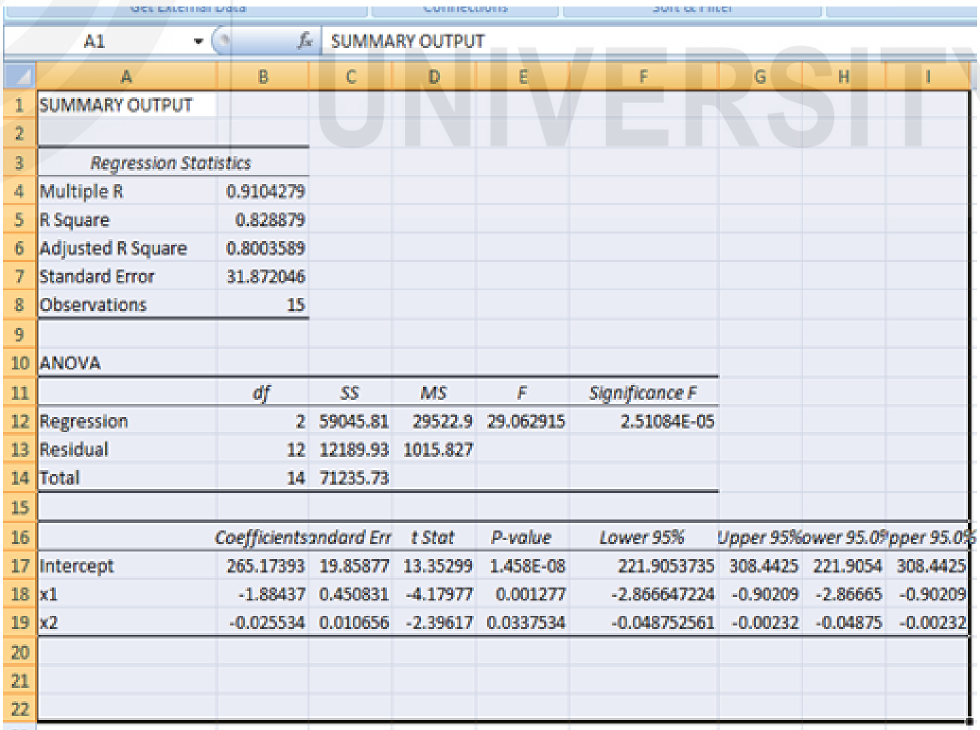
आप पहले आंकड़ो को MS-Excel में प्रविष्ट करें। निम्न चित्र यही दिखा रहा है :



प्रतीपगमन के लिए *Data* tab से *Data Analysis* का विकल्प चुनें। आपको ऊपर दिखाया गया Analysis Tools पृष्ठ दिखाई देगा। उसमें *Regression* का चयन करने पर यह पृष्ठ दिखाई देगा :



यहाँ input Y range के चयन पर गौर करें। यह आपके आंकड़ों में y चर का परिसर है। ध्यान रहे कि Input X-range में स्तंभ B और C, दोनों के परिसर को चुनना है। इससे excel को ज्ञान हो जाता है कि एक से अधिक (यहां दो) स्वतंत्र चर हैं। कृपया यह भी ध्यान दें कि Check box में Label के आगे (√) चिन्ह आ गया है। यह इसलिए किया गया है क्योंकि Excel में आंकड़ा तालिका की प्रथम पंक्ति में चरों के label (y, x_1, x_2) ही अंकित हैं। साथ ही New Worksheet Ply विकल्प का चयन भी किया जा चुका है। इससे प्रतीपगमन के आंकड़ों के लिए एक अलग स्प्रेडशीट खुल जाएगी। अब आप OK बटन दबा सकते हैं। एक नया स्प्रेडशीट पृष्ठ यह दिखाते हुए खुल जाएगा :



	A	B	C	D	E	F	G	H	I
1	SUMMARY OUTPUT								
2									
3	*Regression Statistics*								
4	Multiple R	0.9104279							
5	R Square	0.828879							
6	Adjusted R Square	0.8003589							
7	Standard Error	31.872046							
8	Observations	15							
9									
10	ANOVA								
11		*df*	*SS*	*MS*	*F*	*Significance F*			
12	Regression	2	59045.81	29522.9	29.062915	2.51084E-05			
13	Residual	12	12189.93	1015.827					
14	Total	14	71235.73						
15									
16		*Coefficients*	*andard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
17	Intercept	265.17393	19.85877	13.35299	1.458E-08	221.9053735	308.4425	221.9054	308.4425
18	x1	-1.88437	0.450831	-4.17977	0.001277	-2.866647224	-0.90209	-2.86665	-0.90209
19	x2	-0.025534	0.010656	-2.39617	0.0337534	-0.048752561	-0.00232	-0.04875	-0.00232
20									
21									
22									

प्रतीपगमन परिणाम आंकड़े काफी विस्तृत हैं – किंतु आप अभी केवल F तथा p-मान ही देखें। यहां ये दोनों ही 0.05 से कम हैं। साथ ही, सहसंबंध गुणंक R तथा R-square भी इकाई के निकट हैं। यह भी ध्यान दें कि आकलित अंतःखंड पद 265.17393 बहुत उच्च है तथा x_1 का गुणांक −1.88437 है किंतु x_2 का गुणांक −0.025534 है। क्या यह चर संख्याओं के मानों में अंतर का परिणाम है?

आप अभी तो इस प्रकार के सभी प्रश्नों के उत्तर नहीं दे पाएंगे, किंतु यह उदाहरण सॉफ्टवेयर के अनुप्रयोग में एक प्रारंभिक बिंदु के रूप में अवश्य सहायक हो सकता है।

9.6 बहु-विचर प्रतिमानन के प्रति एक संरचित दृष्टिकोण

एक षट्पदीय प्रतिमान निरूपण विधि बहुविचर विश्लेषण को विकसित करने उसकी व्याख्या तथ सत्यापन के लिए एक रूपरेखा प्रदान करती है। ये 6 पद इस प्रकार हैं:

पद 1 : शोध समस्या, उसके उद्देश्य और प्रयोग की जाने वाली बहु-विचर तकनीक की परिभाषा करें।

पद 2 : एक विश्लेषण योजना विकसित करें।

पद 3 : बहु-विचर विश्लेषण की आधारिक मान्यताओं का मूल्यांकन करें।

पद 4 : बहु-विचर प्रतिमान का आकलन करें और प्रतिमान की सकल स्तरीय संगति की समीक्षा करें।

पद 5 : चरों की व्याख्या करें।

पद 6 : बहु-विचर प्रतिमान का सत्यापन करें।

यहां इन 6 पदों को प्रस्तुत करने का ध्येय आपको किसी नम्यताहीन (rigid) प्रक्रिया से बांधना नहीं है। ये श्रृंखलाबद्ध पद प्रतिमान रचनाक्रम को ही स्पष्ट कर रहे हैं। यह दृष्टिकोण एक सुपरिभाषित शोध प्रकल्पना पर केंद्रित है। यह एक सैद्धांतिक प्रतिमान से प्रारंभ होती है जिसमें समीक्षा किए जाने वाले संबंध सूत्रों का विवरण दिया जाता है। एक बार संकल्पना निरूपण के बाद व्यावहारिक मुद्दों पर ध्यान दिया जा सकता है, जिनमें किसी बहु-विचर तकनीक विशेष का चयन और शोध कार्य को क्रियान्वित करना मुख्य हैं। महत्त्वपूर्ण परिणाम पाने के बाद अगला कदम विचर पर ध्यान देते हुए व्याख्या पर केंद्रित होता है। अंत में, निदान उपाय यह निश्चित करते हैं कि प्रतिमान केवल प्रतिदर्श के आंकड़ों के लिए नहीं बल्कि उसे यथासंभव अधिकाधिक विस्तृत पटल पर भी लागू किया जा सकता है।

बोध प्रश्न 4

1) बहु-विचर विश्लेषण परिणामों की व्याख्या करते समय जिन बिंदुओं का ध्यान रखना होता है उनकी सूची बनाइए।

....

....

....

....

....

2) परिणामों के व्यावहारिक और सांख्यिकीय महत्त्व के बीच क्या अंतर होता है?

....

....

....

....

....

3) बहु-विचर विश्लेषण करते समय शोधकर्ता को भ्रामक परिणामों से बचने के लिए क्या सतर्कता बरतनी चाहिए?

....

....

....

....

....

9.7 सार-संक्षेप

अनेक चरों वाले आंकड़ों के विश्लेषण के लिए कुछ समय से बहु-विचर तकनीकें बहुत सशक्त उपस्कर के रूप में उभर रही हैं। इसका मुख्य कारण यही है कि एकल विचर विश्लेषणों की श्रृंखला से सकल परिणामों की भ्रामक व्याख्या हो जाने का जोखिम रहता है। एकल विचर विश्लेषण में विभिन्न चरों के सहसंबंधों और अन्य अंतर्निर्भरताओं पर ध्यान नहीं दिया जाना ही इसका प्रमुख कारण है। बहु-विचर विश्लेषण की तकनीकें तो आंकड़ों के भारी परिमाण में निहित जानकारी को सरल रूप में प्रस्तुत करती हैं। दूसरे शब्दों में, ये तकनीकें अवलोकनों की एक विशाल और विविधतापूर्ण राशि को थोड़े से संयुक्त सूचकों में समाहित कर देती हैं जो शोध विषयक आंकड़ों में निहित जानकारी का यथासंभव संपूर्ण प्रतिनिधित्व कर सकें। अतः इन तकनीकों का सही योगदान तो विशाल राशि में जटिल जानकारी को सरलीकृत प्रस्तुति प्रदान करना ही है। इकाई का समापन सॉफ्टवेयर आधारित बहुपद प्रतीपगमन और सहसंबंध आकलन से किया गया है।

9.8 संदर्भ ग्रंथादि

1) B.F.J. (1994). *Multivariate Statistical Methods–A Primer*, Chapman and Hall, London.

2) Everitt B.S. and Dunn G. (2001). *Applied Multivariate Data Analysis*, Arnold, London.

3) Härdle, W., Simar, L. (2007). *Applied Multivariate Statistical Analysis*, Springer.

4) Hair, J. F. Jr. (1995). *Multivariate Data Analysis*, 4th ed. Prentice-Hall.

5) Harris R.J. (1985). *A Primer in Multivariate Statistics*, Academic Press, New York. Manly.

6) Johnson, Richard A.; Wichern, Dean W. (2007). *Applied Multivariate Statistical Analysis*, 6th edition,. Prentice Hall.

7) Laurence G. Grimm, PhD, and Paul R. Yarnold (Edited), *Reading and Understanding Multivariate Statistics*, American Psychological Association, Washington, USA.

8) Sharma, S. (1996). *Applied Multivariate Techniques*, University of South California, John Wiley & Sons, Inc.

9) Tabachnick B., Fidell, L. (2007). *Using Multivariate Statistics*, 5th edition Pearson Education. Inc.

9.9 बोध प्रश्नों के उत्तर अथवा संकेत

बोध प्रश्न 1

1) भाग 9.1 तथा उपभाग 9.2.1 देखें।

2) भाग 9.1 तथा उपभाग 9.2.1 देखें।

बोध प्रश्न 2

1) उपभाग 9.3 देखें।

बोध प्रश्न 3

1) भाग 9.4 देखें।

2) उपभाग 9.4.3 देखें।

3) उपभाग 9.4.5 देखें।

बोध प्रश्न 4

1) उपभाग 9.5.1 से 9.5.6 देखें।

2) उपभाग 9.5.1 देखें।

3) भाग 9.6 देखें।

ignou
THE PEOPLE'S
UNIVERSITY