इकाई 1 मूल अवधारणाएँ

इकाई की रूपरेखा

1.0 उद्देश्य

इस इकाई के अध्ययन के बाद आप :

- हमारे जीवन में सांख्यिकी के महत्त्व को समझ सकेंगे;
- सांख्यिकी के अध्ययन में प्रयोग होने वाली मूल अवधारणाओं की जानकारी प्राप्त कर सकेंगे;
- सांख्यिकी को एकवचन तथा बहुवचन के रूपों में परिभाषित कर सकेंगे; और
- सांख्यिकी के उपयोगों तथा दुरुपयोगों को समझ सकेंगे।

1.1 प्रस्तावना

आधुनिक युग में सांख्यिकी एक सुप्रसिद्ध शब्द बन गया है चाहे विभिन्न व्यक्ति इसको विभिन्न रूपों में समझते है। आज का शिक्षित व्यक्ति अनिवार्य रूप से सांख्यिकी का व्यक्ति है जोकि इसके व्यापक अर्थ को समझता है तथा इसे अपने जीवन में विभिन्न प्रकार से प्रयोग करता है। उदाहरणार्थ, समाचार तथा टी.वी. के माध्यम द्वारा प्रतिदिन हमें जनसंख्या, विनिमय दर उच्चावचन, कीमत

वृद्धिदर, दिन तथा रात का तापमान (सामान्य से अधिक या कम; शताब्दी या गत तीस वर्षों, आदि, में न्यूनतम तथा अधिकतम), इत्यादि विषयों पर परिमाणात्मक जानकारी प्राप्त होती रहती है। इसलिए समाज को समझने के लिए आवश्यक है कि

क) जो कहा जा रहा है उसको मापा जाए;

ख) इसको संख्याओं के रूप में लिखा जाए अर्थात् मात्राएँ किलोग्राम में अण्डों की संख्या दर्जनों में आदि; तथा

ग) परिमाणात्मक सूचना से निष्कर्ष निकालने के लिए इनका प्रयोग तथा नीति मापदंडों के बारे में सुझाव देने की योग्यता रखना।

यह कहना अनावश्यक है कि यदि जो कुछ कहा जा रहा है वह माप कर अंकों के रूप में न लिखा जा सके तो इसको समझ पाना कठिन है तथा इस प्रकार हमारा ज्ञान अपर्याप्त तथा असन्तुष्ट रहेगा। अत: सांख्यिकी में एक प्रकार की अंकात्मक सूचना होती है जिसे समंक या आँकड़े कहा जाता है। उदाहरणार्थ, एक व्यक्ति यह कह सकता है कि उसने एक भारतीय उद्योग में शिक्षित तथा अशिक्षित श्रमिकों की अनुपस्थिति सांख्यिकी (अर्थात् परिमाणात्मक सूचना) का अध्ययन किया है तथा पाया है कि अशिक्षित श्रमिकों अनुपस्थिति आपतन (incidence) अधिक है। यहाँ पर वह अंकात्मक सूचना की बात कर रहा है जिसे तकनीकी रूप में समंक कहते है।

समंकों के अन्य उदाहरण :

क) भारत में जनसंख्या विस्फोट की परिस्थिति है, जनसंख्या वृद्धि की वार्षिक दर लगभग 2% है।

ख) कक्षा XIIA के विद्यार्थियो का परीक्षा परिणाम XIIB के विद्यार्थियों की तुलना में अच्छा है क्योंकि XIIA के विद्यार्थियों के प्राप्त अंक XIIB के विद्यार्थियों के औसत अंकों की तुलना में 25% अधिक है।

ग) भारत का विदेशी मुद्रा भण्डार स्वतंत्रता के बाद अब तक सबसे ज़्यादा है। वर्ष 2004 में यह 118 अरब डॉलर है।

विद्यार्थी इस प्रकार अनेक उदाहरण स्वयं पा सकते हैं।

सांख्यिकी का इतिहास

सांख्यिकी (Statistics) शब्द "स्टैटिस्टिक" (statistik) शब्द का आधुनिक रूप है जोकि इटैलियन शब्द "स्टैटिस्टा" (Statista) से प्राप्त किया गया है तथा जिसका अर्थ "राजनेता" (statesman) होता है। प्रोफेसर गॉटफ्रॉयड आचेनवाल (Gott Fried Achenwall) ने इस शब्द का प्रयोग 18वीं शताब्दी में किया था। डॉ० ई० ए० डब्ल्यू जीम्मरमैन ने इस शब्द को इंगलेंड में परिचित कराया।

प्राचीन सरकारों के अभिलेखों में सांख्यिकीय सूचना जनसंख्या के कुछ पहलुओं, भूमि अभिलेख, विभिन्न पक्षों की फौजी ताकत, किसी महामारी में मृत्यकों की संख्या आदि को दर्शाती है। शायद इसीलिए सांख्यिकी को राजाओं का विज्ञान कहा जाता था। लेकिन मानवता के विकास के साथ सांख्यिकी के प्रयोग तथा समझ में वृद्धि हुई तथा अब ज्ञान के ऐसे क्षेत्र के बारे में सोचना कठिन है जिसमें सांख्यिकी की आवश्यकता न हो। वास्तव में यह विश्लेषण का एक महत्त्वपूर्ण हथियार बन चुका है।

1.2 सांख्यिकी का अर्थ

अब हम सांख्यिकी शब्द के अर्थ की जानकारी प्राप्त करने का प्रयास करते हैं। विभिन्न व्यक्तियों के लिए इसका अर्थ भिन्न होता है। एक सामान्य व्यक्ति के लिए सांख्यिकी का अर्थ एक आर्थिक

व्यवसायिक या अन्य वैज्ञानिक गतिविधि से सम्बन्धित आँकड़ों का ढेर आलेख या चित्र होता है। लेकिन एक विशेषज्ञ के लिए यह आँकड़ों के ढेर के अतिरिक्त, एक अन्वेषण की विधि भी हो सकता है। इस प्रकार सांख्यिकी शब्द के प्रयोग दो रूपों, अर्थात्, एकवचन तथा बहुवचन, में किया जाता है। इन दोनों रूपों की व्याख्या निम्नलिखित है।

1.2.1 बहुवचन के रूप में सांख्यिकी

बहुवचन के रूप में सांख्यिकी का अर्थ एक परिमाणात्मक सूचना का ढेर होता है जिसे समंक कहते हैं। उदाहरणार्थ, भारतीय सरकार द्वारा प्रत्येक दस वर्ष बाद किए गए जनसंख्या संगणना से प्राप्त सूचना के आधार पर हम देश की जनसंख्या या जनसांख्यिकीय विशेषताओं के बारे में बात करते हैं। इसी प्रकार हम एक विश्वविद्यालय में पिछले दस वर्षों में नामांकित विद्यार्थियों की सांख्यिकी (समंक) एकत्र कर सकते हैं। इसके अतिरिक्त, भारतीय सरकार के लगभग सभी मंत्रालयों द्वारा, उनकी गतिविधियों से सम्बन्धित, समंक एकत्र किए जाते है।

समंकों को सांख्यिकीय आँकड़े भी कहते हैं। होरेस सेक्राइस्ट ने बहुवचन के रूप में सांख्यिकी की निम्नलिखित व्याख्या की है।

"समंक से हमारा अभिप्राय तथ्यों के उन समूहों से है अनगिनत कारणों से अत्यधिक प्रभावित होते हैं, संख्याओं में व्यक्त किए जाते है, एक उचित शुद्धता स्तर के अनुसार गिने या अनुमानित किए जाते है, किसी पूर्व-निश्चित उद्देश्य के लिए एक व्यवस्थित ढंग से एकत्र किए जाते है और जिन्हे एक दूसरे से सम्बन्धित रूप में प्रस्तुत किया जाता है।"

बहुवचन के रूप में सांख्यिकी की उपरोक्त परिभाषा इसकी निम्नलिखित विशेषताओं का उल्लेख करती है :

क) ***समंक संख्यात्मक तथ्य होते हैं :*** किसी सर्वेक्षण से प्राप्त सूचना को समंक तभी कहा जा सकता है जब वह अंकों द्वारा प्रस्तुत की गई हो। यह समंक या तो अभिलक्षण (जैसे व्यक्ति का कद, भार आदि) के माप द्वारा या गिनती द्वारा, जब अभिलक्षण (जैसे ईमानदारी, धुम्रपान की आदत, सुन्दरता आदि) को मापना संभव न हो, प्राप्त किए जा सकते हैं।

ख) ***समंक तथ्यों के समूह होते हैं :*** एक या असंबद्ध अंकों को समंक नही कहा जा सकता। उदाहरणार्थ, सुरेन्द्र द्वारा विश्वविद्यालय परीक्षा में 65% अंक प्राप्त किए गए है, समंक नही कहलाता। लेकिन, विश्वविद्यालय के 3 लाख विद्यार्थियों जिनके औसत अंक 55% थे, में से सुरेन्द्र ने 65% अंक प्राप्त किए, समंक कहलाते हैं। इस प्रकार किसी सांख्यिकीय सर्वेक्षण की परिधि, जैसे उत्पादन, रोज़गारी, मज़दूरी तथा आय में एक अंक को समंक नहीं कहा जा सकता।

ग) ***समंक विविध कारणों से पर्याप्त सीमा तक प्रभावित होते हैं :*** भौतिक विज्ञान में एक विशेष घटना पर विभिन्न प्रकार की शक्तियों के प्रभावों को अलग-अलग करना सम्भव होता है। लेकिन सांख्यिकी में एकत्रित अंक अनेक कारणों तथा शक्तियों से प्रभावित होते हैं। किसी वर्ष में गेहूँ का उत्पादन अनेक कारणों से प्रभावित होता है जैसे सिंचाई सुविधाओं की उपलब्धता, भूमि की किस्म, जोत की विधि, बीज की किस्म, उर्वरक की मात्रा आदि। इसके अतिरिक्त कुछ ऐसे कारण भी हो सकते हैं जिनकी पहचान करना भी कठिन कार्य हो सकता है।

घ) ***समंक संख्याओं में व्यक्त किए जाते हैं :*** किसी तथ्य के अंकों द्वारा कथन को समंक कहते है। गुणात्मक तथ्य जैसे, एक विद्यालय के विद्यार्थी किसी दूसरे विद्यालय के विद्यार्थियों की तुलना में अधिक मेधावी है, को समंक नहीं माना जा सकता। इसके विपरीत 'पहले विद्यालय के विद्यार्थियों के प्राप्तांक का औसत 90% है तथा दूसरे विद्यालय के विद्यार्थियों के प्राप्तांक

का औसत 60% है तथा दोनों विद्यालयों में क्रमश: 80% तथा 50% विद्यार्थी प्रथम श्रेणी में उत्तीर्ण हुए है, एक सांख्यिकीय कथन है।

ङ) ***समंक उचित कोटी की शुद्धता सहित संकलित या अनुमानित किए जाते हैं :*** समंकों के संकलन या अनुमान में एक उचित कोटि की शुद्धता बनाए रखना अनिवार्य होता है। किसी सर्वेक्षण में आवश्यक शुद्धता दो बातों पर निर्भर होती है, सर्वेक्षण की प्रकृति तथा उद्देश्य, समय तथा साधनों की उपलब्धता। अत: समंको की उचित कोटि की शुद्धता का चुनाव उपरोक्त विचारों के आधार पर किया जाना आवश्यक होता है। एक बार चयन किए गए शुद्धता स्तर को समस्त सर्वेक्षण में समान रूप से बनाए रखना चाहिए।

च) ***समंक एक व्यवस्थित रूप में संकलित किए जाते है :*** समंकों के संकलन से पूर्व सर्वेक्षण का उद्देश्य जानना अति आवश्यक होता है। किसी सर्वेक्षण का उद्देश्य सुनिश्चित तथा सुपरिभाषित होना चाहिए। तत्पश्चात् समंकों का संकलन सुनियोजित तथा व्यवस्थित विधि द्वारा किया जाना चाहिए। सुनियोजित विधि तैयार करते समय इन प्रश्नों, जैसे संगणना तथा निदर्शन विधियों में कौन-सी प्रयोग की जाएगी, समंकों के संकलन, वर्गीकरण प्रस्तुतिकरण एवं विश्लेषण की कौन-सी विधियाँ प्रयोग होंगी इत्यादि, के उत्तर जानना अति आवश्यक हैं। इस विषय का विस्तार से विवेचन एकक 2 में उल्लेखित है।

छ) ***समंक एक-दूसरे से सम्बन्धित रूप में प्रस्तुत किए जाने चाहिए :*** केवल तुलनात्मक समंको का कुछ अर्थ होता है। असम्बन्धित एवं अतुलनात्मक अंकों को समंक नहीं कहते। वह मात्र अंक होते हैं। उदाहरणार्थ, एक कक्षा के विद्यार्थियों की ऊँचाइयाँ तथा वजन के अंकों का उनके माता-पिता की आय तथा योग्यता से कोई सम्बन्ध नहीं होता। तुलना की दृष्टि से अंकों का सजातीय होना आवश्यक है, अर्थात् इनका सम्बन्ध समान विषय या वर्ग या घटना से होना चाहिए, जैसे, एक कक्षा के विद्यार्थियों का जेब खर्च आवश्यक रूप से उनके माता-पिता की आय से सम्बन्धित होता है। देहली में प्याज तथा आलू की कीमतों का भारत के अन्य शहरों में इनकी कीमतों से सम्बन्ध होता है।

अत: यह कहना गलत नहीं होगा कि 'सभी समंक तथ्यों के संख्यात्मक कथन होते हैं किन्तु सभी संख्यात्मक कथन समंक नही होते'। केवल वही संख्यात्मक कथन समंक कहलाते है जिनमें उपर्युक्त सभी लक्षण या अधिकांश विद्यमान हों।

1.2.2 एकवचन के रूप में सांख्यिकी

एकवचन के रूप में, सांख्यिकी का तात्पर्य *सांख्यिकीय विधियों* से होता है जिसका अर्थ समंको के संकलन, संक्षेपण, प्रस्तुतिकरण विश्लेषण तथा निर्वचन की विधियों के निरन्तर वृद्धमान ढाँचे से है। सरल भाषा में सांख्यिकी, अन्य विषयों जैसे गणित, अर्थशास्त्र आदि की भाँति एक वैज्ञानिक विषय है।

कुछ प्रसिद्ध विद्वानों द्वारा सांख्यिकी की परिभाषाएँ निम्नलिखित है :

ए० एल० बाउले ने सांख्यिकी की कुछ परिभाषाएँ दी हैं लेकिन उनमें से कोई भी पूर्ण तथा संतोषजनक नही है। फिर भी उनकी दो परिभाषाएँ विचारणीय हैं। एक परिभाषा के अनुसार 'सांख्यिकी गणना का विज्ञान कहा जा सकता है'। यहाँ पर वह सांख्यिकी के केवल समंक संकलन दृष्टिकोण पर प्रकाश डालता है जोकि निस्सन्देह महत्त्वपूर्ण है। अन्य परिभाषा के अनुसार 'सांख्यिकी सामाजिक व्यवस्था को मापने का विज्ञान है...' उसके विचार में 'सांख्यिकी को सही अर्थ में औसतों का विज्ञान कहा जा सकता है'।

यद्यपि मापन संकलन तथा औसत (समान्तर, गुणोत्तर तथा हरात्मक माध्य; बहुलक तथा माध्यिका) आदि का सांख्यिकी में बहुत महत्त्व है, फिर भी, जैसा कि हम बाद में अध्ययन करेंगे, ये सांख्यिकी के मात्र विषय नहीं है।

क्रॉक्सटन तथा काउडेन द्वारा सांख्यिकी की सरल तथा सुस्पष्ट परिभाषा इस प्रकार की है, 'सांख्यिकी को संख्यात्मक समंकों के संकलन, प्रस्तुतिकरण, विश्लेषण तथा निर्वचन के रूप में परिभाषित किया जा सकता है।'

ये दोनों परिभाषाएँ, निम्नलिखित पाँच महत्त्वपूर्ण पक्षों पर बल देती हैं जोकि वास्तव में, *सांख्यिकीय विधियों* का कार्य क्षेत्र है।

क) **समंकों का संकलन** : समंकों का संकलन, किसी सांख्यिकी अन्वेषण का प्रथम चरण होता है। क्योंकि समंक सांख्यिकीय विश्लेषण का आधार होते है, इस लिए इनका संकलन बहुत सावधानी से किया जाना चाहिए। दोषपूर्ण समंको के परिणाम गलत हो सकते हैं तथा लाभदायक होने की बजाय अधिक हानिकारक हो सकते हैं। समंकों के संकलन के दो स्रोत हो सकते हैं :

i) *प्राथमिक स्रोत* : जहाँ पर विभिन्न विधियों (जिनका विस्तृत विवेचन एकक 2 भाग 4 में किया गया है।) द्वारा अन्वेषण समंको का संकलन करता है।

ii) *द्वितीयक स्रोत* : जहाँ पर समंक किसी प्रकाशित या अप्रकाशित स्रोत से प्राप्त किए जाते हैं। अर्थात् किसी अन्य व्यक्ति अथवा संस्था द्वारा संकलित समंक होते हैं। इस प्रकार संकलित समंको द्वारा अन्वेषक के समय, मुद्रा तथा प्रयास की बचत होती है, लेकिन इनका प्रयोग बड़े ध्यानपूर्वक तरीके से किया जाना आवश्यक होता है। इसका विस्तृत विवेचन इकाई 2, भाग 2.5 में किया जाएगा।

ख) **समंकों का विन्यास** : द्वितीयक स्रोत से प्राप्त समंक पहले से ही व्यवस्थित होते हैं, जैसे भारतीय संगणना से प्राप्त जनसंख्या समंक। इन संमको में अपनी आवश्यकतानुसार थोड़े बहुत परिवर्तन किए जा सकते हैं। इसके विपरीत प्राथमिक स्रोत से प्राप्त समंक अव्यवस्थित होते है तथा इनसे कुछ तथ्य प्राप्त करने हेतु इनको व्यवस्थित करना आवश्यक होता है। यह प्रक्रिया निम्नलिखित चरणों द्वारा पूर्ण की जाती है।

1) *सम्पादन* : इस प्रक्रिया द्वारा संकलित समंको में विद्यमान त्रुटियों तथा असंगतियों को दूर किया जाता है।

2) *समंकों का वर्गीकरण* : यह चरण समंको के सम्पादन के बाद आता है। इस चरण में किसी समान अभिलक्षण के अनुसार समंकों को व्यवस्थित किया जाता है। प्राय प्रत्यार्थियों (respondent) द्वारा प्रदान सूचना को मास्टर पत्रों (sheets) में लिखित किया है। उदाहरणार्थ उडीसा राज्य के धातु सम्बन्धित अभियान्त्रिकी उद्योगों का सर्वेक्षण करके पूँजी संरचना, विभिन्न प्रकार की वस्तुओं का उत्पादन, अकुशल, अर्ध-कुशल तथा कुशल श्रमिकों, लागत तथा कीमत संरचना, तकनीकी पहलु आदि के बारे में समंक संकलित कर सकते है। यह सूचना मास्टर पत्रों में लिखी जा सकती है। विस्तृत विवेचन इकाई 3 में दिया गया है।

3) *सारणीकरण* : यह विन्यास प्रक्रिया का अन्तिम चरण है। मास्टर या सांकेतिक (coded) पत्रों में दी गई सूचना को बारम्बारता बंटन या सारणियों अर्थात् स्तम्भ तथा पंक्तियों के रूप में लिखा जाता है। इसका विस्तृत विवेचन इकाई 3 में दिया गया है।

4) *समंकों का प्रस्तुतिकरण* : समंकों के विन्यास तथा सारणीकरण के उपरांत उनको चित्रों तथा आलेखों में प्रस्तुत किया जाता है। इस प्रकार के प्रस्तुतिकरण समंकों की विभिन्न परिस्थितियों की तुलना में सहायक होते है। समंकों का प्रस्तुतिकरण प्रायः दो प्रकार से किया जाता है :

(i) सांख्यिकीय सारणी, तथा (ii) आलेख; जिसका विस्तृत विवेचन इकाई 3 में दिया गया है।

5) *समंकों का विश्लेषण* : किसी सांख्यिकीय अन्वेषण का यह महत्त्वपूर्ण चरण है। सांख्यिकी के इस पाठ्यक्रम के बृहत अंश का उपयोग समंकों के विश्लेषण तथा उनके निष्कर्ष प्राप्त करने की विधियों से सम्बन्धित होता है। विश्लेषण के उपकरणों का विस्तृत विवेचन बाद की इकाइयों में किया गया है। संक्षिप्त रूप में ये उपकरण निम्नलिखित है।

सांख्यिकीय विश्लेषण के उपकरण

I) **सैद्धान्तिक सांख्यिकी**

अ) ***एक चर विश्लेषण***

i) केन्द्रीय प्रवृति के माप। इसमें समान्तर माध्य$(\bar{X})$ गुणोत्तर माध्य (GM) तथा हरात्मक माध्य (HM) जैसे गणितीय माध्य तथा बहुलक (M_o), मध्यका (M_d) तथा अन्य विभाजन मूल्य जैसे चतुर्थक (Q), अष्टमक (O), दशमक (D) तथा शतमक (P) सम्मिलित होते हैं।

ii) प्रकीर्णन के माप। इनमें विस्तार (R), चतुर्थक विचलन (QD), माध्य विचलन (MD), मानक विचलन आदि सम्मिलित होते हैं।

iii) विषमता के माप (S_K), जिनमें कार्ल पीयरसन का माप, राबर्ट बावली का माप तथा परिघात पर आधारित (β_1 गुणांक) माप सम्मिलित हैं।

iv) परिघात पर आधारित प्रथुशीर्षत्व (β_2 गुणांक) माप।

v) प्रायिकता तथा प्रायिकता बंटन जैसे द्विपद बंटन प्वॉयसन बंटन तथा प्रसामान्य बंटन।

ब) ***द्विचर विश्लेषण***

इस विश्लेषण में दो प्रकार के चर होते हैं जैसे उर्वरक की मात्रा (x) तथा उत्पादन की मात्रा (y), जहाँ पर यह ज्ञात है कि उत्पादन की मात्रा उर्वरक की मात्रा से प्रभावित होती है। इस सन्दर्भ में हम रेखीय सहसम्बन्ध (r_{xy}) तथा प्रतीपगमन विश्लेषण का विवेचन इकाई 8 तथा 9 में करेंगे।

II) **व्यावहारिक सांख्यिकी**

यहाँ हम सांख्यिकी के उपकरणों के उपयोग द्वारा अपने दैनिक जीवन के कुछ महत्त्वपूर्ण पहलुओं का विश्लेषण करते हैं। इनमें मुख्यत:

क) काल श्रेणियाँ

ख) सूचकांक

ग) जीवन-संबंधी सांख्यिकी

घ) आनुमानिक सांख्यिकी जैसे परिकल्पना जाँच आदि सम्मिलित होते हैं।

6) *समंकों का निर्वचन* : यह सांख्यिकीय अन्वेषण का अन्तिम लेकिन महत्त्वपूर्ण चरण होता है। इस चरण के समापन के लिए उच्च स्तर की योग्यता, कार्यकुशलता तथा अनुभव आवश्यक होता है। त्रुटिपूर्ण निर्वचन होने पर अन्वेषण का उद्देश्य ही समाप्त हो जाता है। हमारी नीतियाँ तथा कार्य इस बात पर निर्भर होते है कि समंकों का निर्वचन कितना सही है। इसी के आधार पर वालिस तथा रॉबर्ट के अनुसार सांख्यिकी को 'अनिश्चितता के समक्ष विवेकपूर्ण निर्णय लेने की विधियों का समूह' कहा गया है।

1.2.3 प्रतिदर्शज शब्द का अर्थ

आप प्रति माह अपने परिवार के उपभोग हेतु कुछ मात्रा में गेहूँ खरीदते होंगे। 100 किलोग्राम की एक बोरी में गेहूँ की किस्म जाँच आप कैसे करते है? सैद्धान्तिक रूप में इसकी दो विधियाँ हैं :

अ) संगणना विधि, जिसमें गेहूँ के एक-एक दाने का परीक्षण किया जाता है। एकक 2 में आप यह अध्ययन करेंगे कि यह विधि किस प्रकार महँगी, अधिक समय लेने वाली, उबाऊ तथा अनावश्यक है, क्योंकि निदर्शन विधि द्वारा लगभग वैसे ही परिणाम प्राप्त किए जा सकते हैं।

ब) निदर्शन विधि, जिसमें गेहूँ के दानों का प्रतिदर्श लेकर जाँच की जाती है। यदि आप जाँच से संतुष्ट है तो इस मान्यता पर कि बोरी में सभी दाने एक ही किस्म के हैं, आप गेहूँ खरीद लेते है।

एक समष्टि के अभिलक्षण के माप को *प्राचल* कहते हैं। इसके विपरीत समष्टि से प्राप्त प्रतिदर्श के माध्य, मानकविचलन आदि को *प्रतिदर्शज* कहते हैं। ये प्रतिदर्शज *प्राचलों* के *आकलक* कहलाते हैं।

बोध प्रश्न 1

1) क्या निम्नलिखित कथन सत्य हैं? दो या तीन पंक्तियों में कारण लिखिए।
   - i) आधुनिक व्यक्ति के लिए सांख्यिकी का कोई उपयोग नहीं है।
   - ii) सांख्यिकी तथा प्रतिदर्शज का एक ही अर्थ होता है।
   - iii) एक वचन के रूप में सांख्यिकी का अर्थ सांख्यिकीय विधियों से होता है।
   - iv) सांख्यिकी को सही अर्थ में औसतों का विज्ञान कहा जा सकता है।
   - v) समंकों को संख्याओ में लिखा जाना अनिवार्य नहीं होता।

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

2) दैनिक जीवन में सांख्यिकी के उपयोग के पाँच उदाहरण दीजिए।

....

....

3) सांख्यिकी, समंक तथा प्रतिदर्शज शब्दों के प्रयोग द्वारा ऐसे तीन अलग-अलग वाक्य बनाइए जिससे इनमें अन्तर स्पष्ट हो जाय।

4) एक आलू भरी बोरी में से आप आलू की किस्म की जाँच कैसे करेंगे? अपने उत्तर में आप समष्टि, प्रतिदर्श, प्रतिदर्शज तथा प्राचल शब्दों का प्रयोग करें।

....

....

....

....

....

....

....

....

....

....

1.3 सांख्यिकी का महत्त्व

हमें अनुभाग 1.1 से पता चलता है कि सांख्यिकी हमारे दैनिक जीवन में कितनी उपयोगी है। वास्तव में इसकी आवश्यकता अन्वेषण के प्रत्येक क्षेत्र में होती है तथा प्रत्येक समस्या की समझ तथा समाधान के लिए इसका ज्ञान आवश्यक है। सांख्यिक एच.जी. वैल्स के अनुसार, 'किसी कार्यकुशल नागरिक के लिए एक दिन सांख्यिकीय सोच उतनी ही आवश्यक हो जाएगी जितनी उसके पढने तथा लिखने के लिए होती है।' इसके अतिरिक्त सांख्यिकी का उपयोग व्यापक है। ए.एल. बाउले के अनुसार सांख्यिकी को किसी एक विज्ञान तक सीमित नहीं किया जा सकता।' अत: विभिन्न क्षेत्रों में सांख्यिकी का विभिन्न प्रकार से उपयोग तथा निर्वचन किया जाता है।

1.3.1 सांख्यिकी तथा अर्थशास्त्र

सांख्यिकी तथा अर्थशास्त्र का सम्बन्ध बहुत पुराना है। सर विलियम पैटी ने 1690 में 'पोलिटिकल अरिथमेटिक' नामक एक पुस्तक लिखी जिसमें सांख्यिकी तथा अर्थशास्त्र का प्रयोग किया गया। 19वीं शताब्दी के अन्त में आलफ्रेड मार्शल ने यह कहा कि, 'समंक वे तिनके है जिनसे प्रत्येक अन्य अर्थशास्त्री की भाँति मुझे ईंटें बनानी पड़ती हैं'।

20वीं शताब्दी के अर्थशास्त्रियों के अधिकतर सिद्धान्त सांख्यिकीय अन्वेषण - प्रयोगाश्रित (empirical) मानव आचरण (human behaviour) पर न कि विश्लेषण की निगमनात्मक (deductive) पद्धति - पर आधारित है। जे. एम. केन्स, वी. पेरेटो तथा अन्य अर्थशास्त्रियों द्वारा सांख्यिकी का अत्याधिक प्रयोग किया गया। हाल में सांख्यिकी तथा अर्थशास्त्र का ऐसा अन्तर्मिश्रण हुआ है कि एक नया विषय 'इकनोमेट्रिक्स' विकसित हो गया। आर्थिक विश्लेषण में माध्य, मानक विचलन, प्रतीपगमन विश्लेषण, प्रसामान्य बंटन प्रतिचयन सिद्धान्त इत्यादि का अत्यधिक उपयोग होता है। इसके अतिरिक्त सांख्यिकी के अन्य महत्त्वपूर्ण उपयोग निम्नलिखित है। यह सूची केवल निदर्शी है तथा सम्पूर्ण नहीं है।

1) राष्ट्रीय आय का आकलन तथा विश्लेषण।

2) आगत-निर्गत विश्लेषण।

3) उत्पादन फलन का प्रयोगाश्रित विश्लेषण।

4) भारतीय रिज़र्व बैंक बुलेटिनों में दिए गए वित्तीय समंक।

5) जनसंख्या तथा जनांकिकी विशेषताएँ जैसे मृत्युदर, जन्मदर, जीवन प्रत्याशा आदि के सांख्यिकीय अध्ययन।

6) बाज़ार संरचनाओं जैसे एकाधिकार अल्पाधिकार आदि के सांख्यिकीय अध्ययन।

7) समष्टि-आर्थिक चरों जैसे कीमत स्तर, रोज़गार स्तर मुद्रा पूर्ति आदि के सांख्यिकीय अध्ययन

8) पर्याप्त तथा विश्वसनीय सूचना के अभाव में विकसित अर्थव्यवस्थाओं की आर्थिक विकास प्रक्रियाओं को समझना तथा प्रचालित करना असंभव होगा। पर्याप्त तथा विश्वसनीय समंको के बिना आर्थिक नियोजन संभव नहीं है।

1.3.2 सांख्यिकी तथा व्यवसाय

सांख्यिकी व्यवसाय में भी सहायता करती है। एक प्रगतिशील व्यवसायिक संस्था के लिए लागत, आगम, लाभ श्रम तथा पूँजी, विपणन आदि का विश्लेषण अनिवार्य होता है। व्यावसायिक नियोजन में माँग सम्बन्धी बाज़ार सर्वेक्षण पर आधारित व्यवसायिक पूर्वानुमान, अनुकल्प किस्मों (substitute brands) की उपलब्धता, विभिन्न किस्मों के बारे में उपभोगताओं का मत, उपभोगता अधिमान आदि सम्मिलित होते हैं। काल शृंखलाओं के प्रयोग द्वारा व्यवसायिक गतिविधि पर उपनति, मौसमी, चक्रीय तथा अनियमित विचरणों के प्रभावों को अलग-अलग किया जा सकता है। (इकाई 11 देखिए)

सांख्यिकीय विधियाँ व्यवसाय में व्यवसायिक नीतियों तथा उत्पादन क्षेत्र, वित्त, कार्मिक, लेखा विधि तथा कोटि नियंत्रण सम्बन्धी नीतियों के निर्धारण में उपयोगी होती है। आधुनिक व्यवसायिक फर्मे अपने विपणन प्रोत्साहन तथा उत्पादन निष्पादन प्रदर्शन हेतु आलेख चार्ट तथा चित्रों का भरपूर उपयोग करती है।

1.3.3 सांख्यिकी तथा भौतिक विज्ञान

सांख्यिकी का प्रयोग भौतिक विज्ञानों जैसे भौतिकी, भू-विज्ञान, खगोल-विज्ञान, जीव-विज्ञान, चिकित्सा-विज्ञान आदि में बहुत ही उपयोगी सिद्ध हुआ है। आधुनिक चिकित्सक द्वारा इलाज रोगी के रोग के उपचार हेतू उसके बारे में विभिन्न प्राचलों की सूचना पर आधारित होता है। इनमें शरीर ताप का आचरण रक्त दबाव रक्त-चीनी स्तर, ई.सी.जी. आदि सम्मिलित होते हैं। जब शल्य चिकित्सा करनी हो तो चिकित्सक के लिए ऐसी सूचना अति आवश्यक होती है।

इसके अतिरिक्त, एक नई औषधि प्रस्तुत करने से पूर्व, इसके चूहों, बन्दरों, तथा खरगोशों पर प्रभाव के संमक एकत्रित किए जाते हैं। यदि इनके परिणाम सांख्यिकीय दृष्टिकोण से सन्तोषजनक पाएं जाएं तो मनुष्यों पर भी प्रयोग किए जाते हैं। औषधि की प्रभावोत्पादकता का अध्ययन सांख्यिकीय विधि के अनुसार किया जाता है। उदाहरणार्थ, अन्वेषकों की रुचि यह जानने में हो सकती है कि क्या एक नए मच्छर के कारण मलेरिया के लिए क्यूनीन अब भी प्रभावशाली है। इसके लिए वह 1000 रोगियों का यादृच्छिक प्रतिचयन करके प्रयोग कर सकते हैं। यदि सफलता का प्रतिशत बहुत अधिक है तो यह कहा जा सकता है कि मलेरिया के उपचार हेतु क्यूनीन अभी भी प्रभावशाली है।

इसी प्रकार के सांख्यिकीय अध्ययन अन्य भौतिक विज्ञानों में किए जाते है। शायद यह कहना कि कोई भी वैज्ञानिक अध्ययन क्षेत्र ऐसा नहीं है जहाँ पर सांख्यिकी का प्रयोग नहीं होता, अतिशयोक्ति नही होगा। गासियन के 'त्रुटियों का सामान्य नियम' का उपयोग सितारों तथा ग्रहों के चलन के अध्ययन के लिए किया गया। अत: बाउले का यह कथन कि 'सांख्यिकी किसी भी समय या किसी भी परिस्थिति में उपयोगी सिद्ध हो सकती है,' बिल्कुल उपयुक्त है।

1.3.4 सांख्यिकी तथा गणित

सांख्यिकी तथा गणित के सम्बन्ध की जानकारी 17वीं शताब्दी से है। विभिन्न सांख्यिकीय विधियाँ प्रायिकता सिद्धान्त पर आधारित है। पिछले लगभग 100 वर्षों में सांख्यिकी तथा गणित इतने निकट आ गए कि एक नया विषय गणितीय सांख्यिकी विकसित हो गया।

1.3.5 सांख्यिकी तथा सामाजिक विज्ञान

इसी प्रकार विद्वानों द्वारा सांख्यिकी का उपयोग शिक्षा, राजनीति शास्त्र, भूगोल, मनोविज्ञान, मानव विज्ञान आदि में बढता जा रहा है। सभी सार्वजनिक मत सर्वेक्षण सांख्यिकी पर आधारित होते हैं। अन्य क्षेत्र जहाँ पर सांख्यिकी उपयोगी है उनमें बीमा युद्ध / प्रतिरक्षा की तैयारियाँ, सूचकांक, महँगाई भत्ता सम्बन्धी सूत्र आदि सम्मिलित है।

1.4 सांख्यिकी के दुरुपयोग

जैसा ऊपर बताया जा चुका है, कि सांख्यिकी ज्ञान के लगभग सभी क्षेत्रों के लिए अनिवार्य है, लेकिन इसके निहित स्वार्थियों द्वारा दुरुपयोग किए जाने की सम्भावना होती है। यह स्वार्थी, जैसे सत्ताधारी दल, पूर्व निर्धारित अनुकूल परिणाम प्राप्त करने हेतु समंकों को प्रभावित कर सकते हैं। इसके विभिन्न दुरुपयोगों के कारण सांख्यिकी को अनैतिक विज्ञान भी कहा जाता है। विभिन्न तथ्यों को तोड़-मरोड़ कर बुरे उद्देश्य सहित प्रस्तुत किया जा सकता है। जब राज्य या निहित स्वार्थियों का समंको के संकलन तथा प्रस्तुतिकरण पर एकाधिकार हो तो ऐसा करना सरल हो जाता है।

निस्संदेह इन सभी के कारण सांख्यिकी के बारे में आशंकाएँ उत्पन्न हुई है जैसे :

i) 'सांख्यिकी द्वारा कुछ भी सिद्ध किया जा सकता है'।

ii) 'समंक पहले दर्जे की झूठ होते है'।

iii) 'झूठ तीन प्रकार के होते हैं - झूठ, सफेद झूठ तथा समंक'।

iv) 'समंक झूठों का इन्द्रधनुष होता है'।

सांख्यिकीय निष्कर्षों का भ्रांत निर्वचन अनर्थकारी हो सकता है। इसके बारे में, एक गणितज्ञ ने पाया कि क्योंकि उसके परिवार के सदस्यों की औसत ऊँचाई नदी की औसत गहराई से अधिक है, नदी को पार करना सुरक्षित है। लेकिन नदी के पार करने पर केवल वह ही बचा तथा बाकी अन्य सदस्य डूब गए, क्योंकि उसकी असामान्य ऊँचाई के कारण परिवार की औसत ऊँचाई अधिक थी।

1.5 सांख्यिकी की सीमाएँ

जैसा कि हम इकाई के भाग 1.3 में पहले बता चुके हैं कि एच० जी० वैल्स के अनुसार, ' किसी कार्यकुशल नागरिक के लिए एक दिन सांख्यिकीय सोच उतनी ही आवश्यक हो जाएगी जितनी उसके पढ़ने तथा लिखने के लिए होती है'। लेकिन समंक कोई अल्लादीन का चिराग नहीं होते जिनसे सभी प्रकार के कार्य किए जा सकें। इसकी सीमाओं की निम्नलिखित सूची बताने योग्य है :

प्रथम, सांख्यिकीय विश्लेषण विचाराधीन चरों के प्रकार पर निर्भर करता है। गुणात्मक समंक जैसे सुन्दरता, स्वास्थ्य, सद्भाव, ईमानदारी आदि का विश्लेषण नहीं किया जा सकता। लेकिन इसके लिए गुण साहचर्य के विश्लेषण के रूप में अप्रत्यक्ष प्रयास किए गए हैं। यहाँ पर हम ईमानदारी जैसे अभिलक्षणों को मापने की बजाय केवल उनकी संख्या की गणना करते हैं।

द्वितीय, सांख्यिकी केवल समूह से सम्बन्धित होती है। इस समूह को बनाने वाले व्यक्तिगत इकाई का कोई महत्त्व नहीं होता। उदाहरणार्थ भारत का एक राज्य अन्य राज्यों की तुलना में धनी हो सकता है लेकिन इसके कुछ व्यक्ति गरीब राज्यों के कुछ व्यक्तियों की तुलना में गरीब हो सकते हैं। इस प्रकार औसतें कभी-कभी भ्रामक हो सकती हैं।

तृतीय, सांख्यिकीय निष्कर्ष गणितीय दृष्टि से शत-प्रतिशत सटीक नहीं होते। जाने-अनजाने में लिए गए गलत निदर्शनों के परिणाम संयोगवश अनुकूल हो सकते हैं।

चतुर्थ, जैसा कि भाग 1.4 में जिक्र किया गया है, औसत जैसे सांख्यिकीय मापों का त्रुटिपूर्ण निर्वचन हो सकता है तथा अनर्थ सिद्ध हो सकता है।

बोध प्रश्न 2

1) निम्नलिखित तथ्यों पर 3-4 वाक्यों में टिप्पणी कीजिए।

   i) समंक केवल अर्थशास्त्र तथा व्यवसाय तक ही सीमित होते हैं।

   ii) झूठ, सफेद झूठ के बाद सांख्यिकी की बारी आती है।

   iii) एक देश में प्रतिव्यक्ति आय 4050 रुपये है, जिसका अर्थ यह होता है कि प्रत्येक व्यक्ति इतनी आय प्राप्त कर रहा है।

   iv) अर्थशास्त्र तथा गणित के मिश्रण को इकनोमैट्रिक्स कहते हैं।

   ....

   ....

   ....

   ....

   ....

   ....

   ....

   ....

   ....

   ....

   ....

   ....

   ....

2) सांख्यिकी के अर्थशास्त्र तथा व्यवसाय में उपयोगों के पाँच उदाहरण दीजिए।

   ....

   ....

   ....

   ....

   ....

   ....

   ....

   ....

   ....

   ....

   ....

   ऐसे चार क्षेत्रों का जिक्र कीजिए जिनमें सांख्यिकी मुख्य रूप में प्रयोग की जाती है।

   ....

   ....

....

....

....

....

....

....

....

....

4) निम्नलिखित पदों की व्याख्या कीजिए :

i) गणितीय सांख्यिकी

ii) अनैतिक विज्ञान के रूप में सांख्यिकी

iii) अल्लादीन के चिराग के रूप में सांख्यिकी

....

....

....

....

....

....

....

....

....

....

....

....

1.6 सारांश

पढ़ने तथा लिखने की भाँति एक आधुनिक व्यक्ति को सांख्यिकी का ज्ञान अवश्य होना चाहिए। एक वचन के रूप में सांख्यिकी का अर्थ सांख्यिकीय विधियों से होता है जिनका उद्देश्य समंको का संकलन व्यवस्थितकरण, प्रस्तुतिकरण, विश्लेषण तथा निर्वचन करना होता है। बहुवचन के रूप में सांख्यिकी का अर्थ मात्र सूचना के ढेर से होता है, जैसे जनसंख्या समंक।

प्रतिदर्शज का अर्थ प्रतिदर्श से प्राप्त किए गए एक आकलक से होता है जिसका उद्देश्य समष्टि प्राचल का अनुमान करना होता है।

ज्ञान के लगभग सभी क्षेत्रों में सांख्यिकी की उपयोगिता है। इसकी अर्थशास्त्र तथा व्यवसाय में विशेष उपयोगिता है। अर्थशास्त्र, सांख्यिकी तथा गणित को मिलाकर एक विषय इकनोमैट्रिक्स का जन्म हुआ है।

संख्यिकी की अत्यधिक उपयोगिता के बावजूद कुछ अनैतिक व्यक्तियों (मुख्यत: राजनीतिज्ञों) द्वारा इसका दुरुपयोग किए जाने के कारण इसको सफेद झूठ से भी नीचे स्तर तक पहुँचा दिया है। परिणामत:, कभी कभी, इसको अनैतिक विज्ञान भी कहा गया है।

1.7 शब्दावली

सांख्यिकी	:	बहुवचन के रूप में इसका अर्थ संख्याओं के समुच्चय से होता है जिनको प्राय: सांख्यिकीय समंक कहते हैं।
सांख्यिकी	:	एकवचन के रूप में इसका अर्थ समंकों का संकलन प्रस्तुतिकरण, विश्लेषण तथा निर्वचन होता है।
प्रतिदर्शज	:	प्रतिदर्श से परिकलित माप जैसे समान्तर माध्य, माध्यिका, गुणोत्तर माध्य, मानक विचलन आदि को प्रतिदर्शज कहते हैं। आकलन सिद्धान्त में इसे आकलक भी कहा जाता है।
प्राचल	:	यह समष्टि के सभी परिमाणों द्वारा परिकलित माप जैसे समान्तर माध्य, माध्यिका, गुणोत्तर माध्य, मानक विचलन आदि होते हैं।
परिमाणात्मक समंक	:	यह परिमेय अभिलक्षण पर प्राप्त सूचना होती है। इस प्रकार के समंक संख्याओं में उपलब्ध होते हैं।
गुणात्मक समंक	:	यह अपरिमेय अभिलक्षण, जैसे ईमानदारी, सुन्दरता, रंग जाति आदि, पर सूचना होती है।
समष्टि	:	यह किसी अन्वेषण के अन्तर्गत आने वाली सभी इकाइयों का समूह होता है।
प्रतिदर्श	:	यह समष्टि का एक भाग होता है जो इसके एक या अधिक अभिलक्षणों के अध्ययन के लिए प्राप्त किया जाता है।
संगणना	:	अन्वेषण की वह विधि जिसमें समष्टि की सभी इकाइयों से सूचना एकत्रित की जाय।
निदर्शन	:	अन्वेषण की वह विधि जिसमें सूचना केवल प्रतिदर्शित इकाइयों से एकत्रित की जाय।

1.8 कुछ उपयोगी पुस्तकें

Elhance, D. N. and V. Elhance, 1988, *Fundamentals of Statistics,* Kitab Mahal, Allahabad.

Nagar, A. L. and R. K. Dass, 1983, *Basic Statistics,* Oxford University Press, Delhi.

Mansfield, E., 1991, *Statistics for Business and Economics : Methods and Applications,* W. W. Norton and Co.

Yule, G. U. and M. G. Kendall, 1991, *An Introduction to the Theory of Statistics,* Universal Books, Delhi.

1.9 बोध प्रश्नों के उत्तर या संकेत

बोध प्रश्न 1

1) i) असत्य, ii) असत्य, iii) सत्य, iv) सत्य, v) असत्य।
2) भाग 1.1. देखिए।
3) अनुभाग 1.2.1 से 1.2.3 देखिए।
4) अनुभाग 1.2.3 देखिए।

1) i) भाग 1.3 देखिए।
   ii) तथा iii) के लिए भाग 1.4 देखिए।
   iv) अनुभाग 1.3.1 देखिए।
2) अनुभाग 1.3.1 और 1.3.2 देखिए।
3) भाग 1.3 देखिए।
4) i) अनुभाग 1.3.4 देखिए।
   ii) भाग 1.4 देखिए।
   iii) भाग 1.5 देखिए।

1.10 पारिभाषिक शब्दावली

समंक	:	data or statistics
समष्टि	:	population or universe
प्रतिदर्श	:	sample
प्रतिदर्शज	:	statistic or estimator
समान्तर माध्य	:	arithmetic mean
माध्यिका	:	median
बहुलक	:	mode
गुणोत्तर माध्य	:	geometric mean
हरात्मक माध्य	:	harmonic mean
काल श्रेणी	:	time series
सूचकांक	:	index numbers

इकाई 2 समंक संकलन विधियाँ

इकाई की रूपरेखा

2.0 उद्देश्य

इस इकाई के अध्ययन के बाद आप :

- समंकों की अवधारणा तथा प्रकृति की व्याख्या कर सकेंगे;
- सांख्यिकीय अन्वेषण में समंकों के महत्त्व की पहचान कर सकेंगे;
- सर्वेक्षण तकनीकों को समझ सकेंगे; और
- द्वितीयक समंकों के उपयोगों तथा सीमाओं को समझ सकेंगे।

2.1 प्रस्तावना

अपने जीवन के विभिन्न क्षेत्रों में हमें समस्याओं का सामना करना पड़ता है, जिसके कारण हम उनके समाधान के बारे में सोचने के लिए बाध्य होते हैं। जब भी हम एक सम्मुख समस्या के समाधान के लिए वास्तव में गम्भीर होते हैं तो सोचने की प्रक्रिया आरम्भ हो जाती है। इस सोचने की प्रक्रिया को सांख्यिकीय सोच या सांख्यिकीय अन्वेषण कहते हैं। इस प्रक्रिया को कुछ सूचना, अधिमानत: परिमाणात्मक सूचना, की आवश्यकता होती है जिनको सांख्यिकीय सूचना या समंक कहते है।

एक सांख्यिकीय अन्वेषण में प्रथम कदम समंकों का संकलन होता है। हर बार अन्वेषक को शुरू से चलना आवश्यक नहीं होता। उसे अन्य व्यक्तियों द्वारा किए गए आविष्कारों का उपयोग करने का प्रयास करना चाहिए। इसके द्वारा लागत, प्रयास तथा समय की बचत होती है।

जैसा इकाई 1 (अनुभाग 1.2.1) में विवेचन किया गया, कि समंकों का अर्थ सम्बन्धित परिमाणात्मक सूचना से होता है। यह एक पूर्वनिर्धारित उद्देश्य के लिए कितनी ही संख्या में सम्बन्धित प्रेक्षणों का समूह होता है। हम बिक्री प्रतिरूप, क्षीण बिक्री के दिनों, प्रतियोगी उत्पादों का प्रभाव, आय आचरण तथा अन्य सम्बन्धित बातों के अध्ययन के लिए विक्रेता तथा विक्रेता समूहों द्वारा, दिल्ली के विभिन्न क्षेत्रों में, सप्ताह के विभिन्न दिनों में, बेचे गए टेलीविज़नों की संख्या के बारे में सूचना एकत्रित कर सकते हैं। इस प्रकार एकत्रित सूचना को समंक समुच्चय तथा एक प्रेक्षण को समंक बिन्दु कहते हैं।

बिना किसी उद्देश्य के प्राप्त की गई किसी भी प्रकार की सूचना का कोई उपयोग नहीं होता। उदाहरणार्थ राम का कद 5 फुट 6 इंच है या श्याम की मासिक आय पहली जनवरी 2004 को 15000 रुपये है, समंक नहीं है। सभी परिमाणात्मक सूचनाएँ सांख्यिकीय नहीं होती। विलग माप सांख्यिकीय समंक नहीं होते। एकवचन के रूप में सांख्यिकी का सम्बन्ध एक विशेष समस्या के समाधान हेतु प्रासंगिक समंकों के संकलन से होता है। सिम्पसन तथा काफ्का के अनुसार *'समंकों का स्वत: कोई अर्थ नहीं होता; उनके अस्तित्व का आधार तब होता है जब कोई समस्या सामने हो'*।

2.2 समंक संकलन का उद्देश्य

संकलित समंकों को निम्नलिखित तीन प्रकार से वर्गीकृत किया जा सकता है:

अ) परिमाणात्मक तथा गुणात्मक समंक

ब) निदर्शन तथा संगणना समंक

स) प्राथमिक तथा द्वितीयक समंक

अ) ***परिमाणात्मक तथा गुणात्मक समंक*** : परिमाणात्मक समंक उस सूचना समुच्चय को कहते है जिसको मापा जा सकता है तथा किसी मानक इकाई, जैसे रुपये, किलोग्राम, लीटर आदि, द्वारा अभिव्यक्त किया जा सकता है। उदाहरणार्थ, एक कक्षा के विद्यार्थियों का जेबखर्च तथा उनके माता-पिता की आय को रुपयों में अभिव्यक्त किया जा सकता है; गेहूँ के उत्पादन या आयात को किलोग्राम या क्विंटलों में अभिव्यक्त किया जा सकता है; एक वर्ष में भारत में पेट्रोल तथा डीजल की खपत को लीटर में आदि।

इसके विपरीत गुणात्मक समंकों का मापना संभव नहीं होता अर्थात् इनको माप की मानक इकाइयों जैसे रुपये, किलोग्राम, लिटर आदि में अभिव्यक्त नहीं किया जा सकता। इसका कारण यह है कि ये आंख का रंग, चमड़ी का रंग, ईमानदारी, अच्छा या बुरा आदि विशेषताएँ या अभिलक्षण होते हैं, जिनका माप करना संभव नहीं है। इनको गुण (attributes) भी कहा जाता है। इस परिस्थिति में एक विशेष गुण वाली इकाइयों की केवल गणना करना संभव होता है।

ब) ***निदर्शन तथा संगणना समंक*** : जैसा इकाई 1 के अनुभाग 1.2.3 में विवेचन किया था कि, समंकों का संकलन संगणना या निदर्शन विधि से किया जा सकता है। निदर्शन विधि द्वारा संकलित सूचना को निदर्शन संमक तथा संगणना विधि द्वारा संकलित सूचना को संगणना संमक कहते है। भारत में प्रत्येक 10 वर्ष बाद जनसंख्या संगणना संमक संकलित किए जाते हैं।

स) ***प्राथमिक तथा द्वितीयक समंक*** : जैसा अनुभाग 1.2.2 में संक्षिप्त विवेचन किया गया था कि, प्राथमिक समंक अन्वेषक द्वारा, विभिन्न विधियों के प्रयोग से, स्वयं संकलित किए जाते हैं। इन विधियों का विवेचन इस इकाई के भाग 2.3.3 तथा 2.4 में किया जाएगा। इस प्रकार के समंक अपरिष्कृत रूप में होते हैं जिनका उपयोग करने से पहले परिष्करण आवश्यक है। इसके विपरीत द्वितीयक समंक किसी विद्यमान प्रकाशित या अप्रकाशित स्रोत, अर्थात् किसी अन्य व्यक्ति द्वारा संकलित समंकों, से प्राप्त किए जाते हैं।

किसी समस्या के सांख्यिकीय समाधान का पहला चरण समंकों का संकलन होता है। संकलित समंकों के उपयुक्त रूपान्तर तथा विश्लेषण के पश्चात् समष्टि के बारे में निष्कर्ष निकाले जाते हैं। ये निष्कर्ष निम्नलिखित में से एक या दोनों हो सकते हैं :

अ) समष्टि के एक या अधिक प्राचलों या समष्टि की प्रकृति का आकलन करना। यह आकलन सिद्धान्त का विषय है, जिसका विवेचन खंड 7 में किया जाएगा।

ब) परिकल्पना की जाँच करना। समष्टि की प्रकृति या प्राचलों के बारे में कथन को परिकल्पना कहते हैं।

2.3 समंकों का संकलन

किसी सांख्यिकीय अन्वेषण में विश्वसनीय तथा पर्याप्त समंक सूचना आवश्यक होती है। इस इकाई के वर्तमान यह तथा उत्तरवर्ती अनुभागों में समंकों के संकलन की विधियाँ दी गई हैं।

2.3.1 सांख्यिकीय अन्वेषण – आयोजन तथा संचालन

विश्वसनीय तथा पर्याप्त समंकों के संकलन के लिए एक सांख्यिकीय सर्वेक्षण का सुविचारित आयोजन तथा संचालन अनिवार्य होता है। ऐसा न होने की परिस्थिति में इसके परिणाम दोषपूर्ण तथा बेकार हो सकते हैं। इन परिणामों से लाभ की तुलना में अधिक हानि हो सकती है। निम्नलिखित अनुभाग में आयोजन के पहलु का विवेचन करने का प्रयास किया गया है।

सांख्यिकीय समंकों का संकलन, सर्वेक्षण या एक प्रयोग करके किया जा सकता है। सामाजिक विज्ञानों जैसे अर्थशास्त्र तथा व्यवसाय में सर्वेक्षणों का अधिक प्रचलन होता है। जबकि प्राकृतिक / भौतिक विज्ञानों में सूचना प्रायः प्रयोगों द्वारा प्राप्त की जाती है। सर्वेक्षण में सम्मिलित विभिन्न व्यक्तियों या इकाइयों के प्रेक्षण से प्राप्त समंक, अत्याधिक संख्या में अनियंत्रित कारकों से प्रभावित होते हैं। उदाहरणार्थ, एक देश में मजदूरी अनेक कारकों, जैसे मज़दूर की कार्य कुशलता, शिक्षा स्तर, तथा लिंग; प्रशिक्षण तथा अनुभव; तथा कुछ देशों में मज़दूर की जाति, से प्रभावित होती है। भारत में नीची जाति तथा ऐतिहासिक दृष्टि से अप्राधिकृत (underprivileged) लोगों, जैसे भंगी, को सामाजिक कारणों से भी कम मज़दूरी दी जाती है।

यह बताना रुचिकर है कि भौतिक विज्ञानों में प्रयोगों द्वारा प्राप्त समंक भी अनन्य अनियंत्रित कारकों से प्रभावित होते है, चाहे ये उपयोग नियंत्रित परिस्थितियों में भी किए गए हों। यहाँ पर अनियंत्रित कारकों के मुख्य कारण – प्रयोग करने वाले व्यक्ति का पक्षपात, मापन यंत्र की प्रकृति तथा शुद्धता आदि हो सकते हैं। किसी सांख्यिकीय सर्वेक्षण को दो चरणों में विभाजित किया जा सकता है।

i) आयोजन चरण

ii) संचालन चरण

2.3.2 आयोजन चरण – सांख्यिकीय अन्वेषण के अपेक्षित गुण

प्राथमिक या द्वितीयक स्रोत से समंकों के संकलन से पूर्व एक अन्वेषक को निम्नलिखित बातों की जानकारी आवश्यक होती है।

i) अन्वेषण का उद्देश्य तथा कार्यक्षेत्र क्या है?

इस प्रश्न के संतोषजनक उत्तर के अभाव में अन्वेषक को सही दिशा प्राप्त नहीं हो सकती। यदि अन्वेषण से सम्बन्धित समंकों का संकलन नहीं होता तो मुद्रा तथा प्रयास दोनों की हानि होती है। केवल इतना ही नहीं, अन्वेषक को यह भी जानकारी होनी चाहिए कि कितने समंकों की

आवश्यकता है जिससे केवल आवश्यक समंक ही संकलित हों। उदाहरणार्थ, यदि हम एक राज्य में गेहूँ उत्पादन के प्रतिरूप से सम्बन्धित समंक संकलित करना चाहते है तो हमें, भूमि की किस्म, कृषि आगतें, सम्बन्धित किसानों का शिक्षा स्तर, भूमि सुधार की उपस्थिति या अनुपस्थिति के दोष, कृषि वित्त की उपलब्धता तथा लागतें, विपणन व्यवस्था आदि, से सम्बन्धित समंकों की आवश्यकता होती है।

ii) सूचना का स्रोत क्या होगा?

अन्वेषक को प्राथमिक स्रोत, जहाँ पर उसे समंक स्वयं संकलित करने होते हैं, तथा द्वितीयक स्रोत, जहाँ पर वह पहले से संकलित समंकों का उपयोग करता है, में से चयन करना होता है।

iii) अन्वेषण की प्रकृति क्या होगी?

अर्थात्, अन्वेषक को निम्नलिखित चयन करने होते हैं :

1) *संगणना या निदर्शन अन्वेषण :* संगणना विधि में उसे समष्टि की प्रत्येक इकाई की जाँच करनी होती है जबकि निदर्शन विधि में केवल प्रतिदर्श में सम्मिलित इकाइयों की जाँच की जाती है। उदाहरणार्थ, संगणना विधि में वह एक गाँव के व्यक्तियों की जाँच करता है जबकि निदर्शन विधि में वह केवल कुछ व्यक्तियों की जाँच करता है।

2) *प्रत्यक्ष या अप्रत्यक्ष अन्वेषण :* प्रत्यक्ष अन्वेषण में मात्रात्मक प्रेक्षण प्रत्यक्ष रूप में प्राप्त किए जाते हैं जैसे टेलिविज़नों की बिक्री तथा रुपयों में विज्ञापन व्यय। इसके विपरीत एक अप्रत्यक्ष अन्वेषण में, विद्यार्थियों की योग्यता की जाँच करने के लिए उनके द्वारा प्राप्त अंको का उपयोग किया जाता है।

3) *मूल या आवृत्तीय अन्वेषण : पहली* बार किए गए अन्वेषण को मूल अन्वेषण कहते हैं जबकि बार-बार किया गया अन्वेषण आवृत्तीय अन्वेषण कहलाता है। उदाहरणार्थ, भारत में जनसंख्या संगणना प्रत्येक 10 वर्ष बाद की जाती है। ये सभी अन्वेषण आपस में सम्बन्धित होने अनिवार्य हैं।

4) *प्रकट या गोपनीय अन्वेषण :* प्रकट अन्वेषण के परिणाम जैसे जनसंख्या तथा राष्ट्रीय आय संमक, जनता को बता दिए जाते हैं। इसके विपरीत बहुत से सरकारी अन्वेषणों के परिणाम, राष्ट्रीय सुरक्षा कारणों से, गोपनीय होते हैं जैसे प्रतिरक्षा परिमाणविक शक्ति, अन्तरिक्ष अनुसंधान तथा विकास सम्बन्धी आदि समंक।

iv) अन्वेषण या गणना की सांख्यिकी इकाई क्या होगी?

सांख्यिकीय इकाई एक गुण या गुणों का समूह होता है जिसे परम्परागत रूप से चुना गया है ताकि उनको रखने वाले व्यक्ति या वस्तुएँ, किसी सांख्यिकीय अन्वेषण के लिए, गिनी या मापी जा सकें। इस प्रकार सांख्यिकीय इकाई, व्यक्ति या वस्तु का एक अभिलक्षण या अभिलक्षणों का समुच्चय होता है जिनका सूचना प्राप्त करने के लिए प्रेक्षण किया जाता है। उदाहरणार्थ, एक व्यक्ति के विभिन्न लक्षण उसकी आय, कद, वज़न आदि, हो सकते हैं। सांख्यिकीय इकाई की परिभाषा का अर्थ एक व्यक्ति या वस्तु के उन अभिलक्षणों का विशेष विवरण होता है जिससे समंकों का संकलन किया जाना है।

यह बताना अनिवार्य है कि एक सांख्यिकीय इकाई के प्रेक्षण का परिणाम एक संख्या हो सकता है जोकि गणना या मापन से प्राप्त होती है। यदि संख्या मापन द्वारा प्राप्त हुई है तो इसकी मापन इकाई बताना भी अनिवार्य है। संकलित समंकों में एकरूपता रखने के लिए सांख्यिकीय इकाई तथा मापन इकाई का निर्धारण आवश्यक होता है।

v) **शुद्धता की कोटि क्या होगी?**

विभिन्न आर्थिक तथा व्यवसायिक अध्ययनों में पूर्ण शुद्धता न तो संभव होती है तथा न आवश्यक होती है। जनसंख्या समंकों में अन्तिम व्यक्ति तक शुद्धता की आवश्यकता नहीं है। उदाहरणार्थ, भारत की जनसंख्या 98, 89, 70, 510 या 98,89,00,000 लिखी जाए तो इससे कोई खास अंतर नहीं होता। लेकिन, समंकों के संकलन की विभिन्न विधियों में से चयन शुद्धता की कोटि पर निर्भर होता है। इसके अतिरिक्त, एक बार निर्धारित शुद्धता की कोटि को सारे सर्वेक्षण में बनाए रखना चाहिए।

2.3.3 कार्यान्वयन चरण

यह चरण आयोजन चरण के बाद होता है। इस चरण में योजना को लागू किया जाता है जिसमें निम्नलिखित कार्य सम्मिलित होते हैं :

1) केन्द्रीय प्रशासनिक कार्य प्रणाली का स्थापन करना जोकि अन्वेषण से सम्बन्धित प्रश्नों का ग्रन्थाकार तैयार करती है, जिसे प्रश्नसूची या प्रश्नावली कहते हैं। अन्वेषण की प्रकृति तथा आकार के अनुसार बड़े भोगोलिक क्षेत्रों को सम्मिलित करने के लिए यह शाखा कार्यालयों की स्थापना का निर्णय भी लेती है।

2) कार्यक्षेत्र के कर्मचारियों, जिनको प्रश्नकर्त्ता या अन्वेषक या परिगणक कहते हैं, का चयन व प्रशिक्षण करना। जैसा भाग 2.4 में बताया गया है, ये अन्वेषक प्रत्यार्थी के पास विभिन्न तरीकों से पहुँचते हैं। अन्वेषक का प्रशिक्षण भली-भाँति होना चाहिए तथा इनका ईमानदार तथा मेहनती होना आवश्यक है। इस चरण में कोई भी त्रुटि अन्वेषण की पूरी प्रक्रिया को संकट में डालकर भ्रामक परिणाम दे सकती है। सर्वेक्षण से श्रेष्ठ परिणाम प्राप्त करने के लिए कार्यक्षेत्र के कर्मचारियों को प्रत्यार्थियों की भाषा की जानकारी, धैर्य तथा उनसे सूचना प्राप्त करने का कौशल होना अनिवार्य है।

3) कार्यक्षेत्र कर्मचारियों का पर्यविक्षण आवश्यक है जिससे यह सुनिश्चित हो सके कि सूचना वास्तव में प्रत्यार्थियों से प्राप्त की गई है न कि अन्वेषकों द्वारा अपने होटल के कमरे में बैठकर, प्रश्नावली मनगढ़न्त तरीके से भरी गई है। इसके अतिरिक्त कार्यक्षेत्र में कुछ विशेषज्ञ भी होने **आवश्यक है जिससे अन्वेषकों के सम्मुख आने वाली समस्याओं का समाधान हो सके। सर्वेक्षण करते समय अप्रतिसंवेदी (non-response) की समस्या बड़ी सामान्य होती है। जो निम्नलिखित प्रकार से हो सकती है :**

   अ) सूची में लिखे गए प्रत्यार्थी की अनुपलब्धता। ऐसी परिस्थिति में इस प्रत्यार्थी को किसी अन्य प्रत्यार्थी से प्रतिस्थापित नहीं किया जाना चाहिए क्योंकि इससे प्रतिदर्श का यादृच्छिक लक्षण प्रभावित हो सकता है तथा अन्वेषण के परिणाम पक्षपाती हो सकते हैं।

   ब) अप्रतिसंवेदी के कारण प्रश्नावली के कुछ प्रश्न या उनके अंशों के उत्तर न प्राप्त होने की संभावना होती है। इनको अन्वेषक द्वारा पूरा नहीं किया जाना चाहिए।

4) समंकों को व्यवस्थित करने के बाद उनका विश्लेषण किया जाता है। विश्लेषण विधियों की व्याख्या बाद के खंडों में की गई है। आजकल इस कार्य के लिए कम्प्यूटर उपलब्ध हैं।

5) समंकों के विश्लेषण के बाद सांख्यिकीय अन्वेषण की विस्तृत रिपोर्ट तैयार की जाती है जिसमें इसके मुख्य: निष्कर्षों का जिक्र होता है। मुख्य निष्कर्ष तथा नीति परामर्श, रिपोर्ट के अन्त में भी लिखे जाते हैं।

2.3.4 प्राथमिक तथा द्वितीयक समंक

इस सन्दर्भ में मुख्य प्रश्न यह है कि समंक कैसे तथा कहाँ से प्राप्त किए जाएँ? समंकों को निम्नलिखित दो प्रकार के अन्वेषणों द्वारा प्राप्त किया जा सकता है।

1) प्रत्यक्ष अन्वेषण, जिसका अर्थ यह है कि अन्वेषण के अन्तर्गत आने वाली इकाइयों के प्रेक्षण से प्राप्त सूचना प्राप्त की जाती है। जैसा पहले बताया जा चुका है, यह समंक प्राप्त करने का मूल स्रोत या प्राथमिक समंकों का स्रोत होता है जोकि प्रेक्षण या प्रश्न पूछकर किया जाता है। प्रेक्षण विधि में हम एक घटना को घटित होते हुए देखते हैं जैसे एक दिन व रात में नई दिल्ली के विजय चौंक से गुजरने वाले वाहनों की संख्या। दूसरी विधि में हम प्रश्नावली द्वारा प्रत्यार्थियों से प्रश्न पूछते हैं जोकि डाक या व्यक्तिगत रूप से भेजी जाती है। यह विधि मुद्रा, समय तथा प्रयास की दृष्टि से खर्चीली होती है।

2) **द्वितीयक स्रोत के द्वारा अन्वेषण, जिसमें पहले से संकलित समंकों से सूचना प्राप्त की जाती है। द्वितीयक समंक अन्य लोगों या संस्थाओं द्वारा संकलित होते हैं जैसे सरकारी संस्थाएँ, आई० एम० एफ०, आई० बी० आर० डी० जैसी अन्तरराष्ट्रीय संस्थाएँ, अन्य देशों, निजी तथा सरकारी अन्वेषण संस्थाएँ, भारतीय रिज़र्व बैंक तथा अन्य बैंक आदि।** द्वितीयक समंकों के स्रोतों को मोटे तौर पर दो भागों मे बाँटा जा सकता है : प्रकाशित स्रोत तथा अप्रकाशित स्रोत।

अ) **प्रकाशित स्रोत**

i) सरकार के विभिन्न स्तरों – केन्द्रीय, राज्य, केन्द्रीय शासित प्रदेश तथा संघ – पर सरकारी प्रकाशन।

ii) विदेशों में सरकारी प्रकाशन।

iii) अन्तरराष्ट्रीय संस्थाओं जैसे आई० एम० एफ०, यूनेस्को, डब्ल्यू० एच० ओ० आदि के अधिकारिक प्रकाशन।

iv) विख्यात समाचारपत्र तथा पत्रिकाएँ (स्वदेशी तथा विदेशी) जैसे कामर्स, कैपिटल, इकनोमिक टाइम्स, इकनोमिका आदि।

v) भारतीय रिज़र्व बैंक तथा अन्य बैंकों, भारतीय जीवन बीमा निगम, व्यवसाय संघ, स्टाक एक्सचेंज, चैम्बर ऑफ कॉमर्स आदि के आधिकारिक प्रकाशन।

vi) विख्यात अर्थशास्त्रियों, शोधकर्त्ता, विश्वविद्यालय, पूछताछ आयोगों आदि की रिपोर्ट।

भारत में प्रकाशित समंकों के अन्य स्रोत निम्नलिखित हैं :

i) *केन्द्रीय सांख्यिकीय संगठन (CSO) :* यह राष्ट्रीय आय, बचत, पूँजी निर्माण आदि पर समंक प्रकाशित करता है। प्रकाशन का नाम राष्ट्रीय लेखा सांख्यिकी (National Accounts Statistics) है।

ii) *राष्ट्रीय प्रतिदर्श सर्वेक्षण संगठन (NSSO) :* यह संगठन सांख्यिकी एवं कार्यक्रम कार्यान्वयन मंत्रालय के अन्तर्गत है तथा राष्ट्रीय अर्थव्यवस्था के विभिन्न पहलुओं जैसे कृषि, उद्योग, श्रम, उपभोग व्यय आदि पर समंक प्रकाशित करता है।

iii) *भारतीय रिज़र्व बैंक के प्रकाशन (RBI) :* यह वित्तीय समंक प्रकाशित करता है। मुद्रा तथा वित्त पर रिपोर्ट, भारतीय रिज़र्व बैंक बुलेटिन, भारतीय बैंकों से सम्बन्धित सांख्यिकीय सारणी आदि इसके मुख्य प्रकाशन हैं।

iv) *श्रम ब्यूरो* : इसके मुख्य प्रकाशन भारतीय श्रम सांख्यिकी, भारतीय श्रम वर्ष पुस्तिका, भारतीय श्रम पत्रिका आदि हैं।

ब) **अप्रकाशित स्रोत**

i) पूछताछ कमेटियों की अप्रकाशित रिपोर्ट।

ii) अन्वेषकों की रिपोर्ट।

iii) व्यवसायिक संस्थाओं, श्रम संगठन तथा चैम्बर ऑफ कॉमर्स आदि के पास उपलब्ध अप्रकाशित सामग्री।

बोध प्रश्न 1

1) निम्नलिखित पदों की व्याख्या कीजिए :
(प्रत्येक उत्तर तीन व्याक्यों से अधिक नहीं होना चाहिए)

i) प्रश्नावली ii) अप्रतिसंवेदी
iii) परिकल्पना iv) सांख्यिकीय इकाई
v) सांख्यिकीय अन्वेषण vi) प्रश्न सूची

....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....

2) निम्नलिखित पदों में भेद स्पष्ट कीजिए
(प्रत्येक उत्तर चार व्याक्यों से अधिक नहीं होना चाहिए)

i) समंक, सांख्यिकीय समंक तथा सांख्यिकी
ii) समंक समुच्चय तथा समंक बिन्दु
iii) प्राथमिक तथा द्वितीयक समंक
iv) परिमाणात्मक तथा गुणात्मक समंक

v) प्रतिदर्श तथा संगणना

vi) सांख्यिकीय अन्वेषण का आयोजन तथा संचालन

vii) सर्वेक्षण तथा प्रयोग

viii) प्रत्यक्ष अन्वेषण तथा द्वितीयक स्रोत अन्वेषण

3) सूचना के विभिन्न स्रोत क्या होते हैं?

....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....

2.4 प्राथमिक समंकों का संकलन – सर्वेक्षण विधि

जब अन्वेषक को यह विश्वास हो जाए कि प्राथमिक समंकों के लाभ मुद्रा, प्रयास तथा समय लागतों से अधिक हैं, तो उसे इनका संकलन करना चाहिए। प्राथमिक समंकों के संकलन के लिए वह निम्नलिखित विधियों में से किसी एक का उपयोग कर सकता है।

1) प्रत्यक्ष व्यक्तिगत अन्वेषण
2) अप्रत्यक्ष मौखिक अन्वेषण
3) स्थानीय सूचनाओं का उपयोग
4) प्रश्नावली विधि

1) प्रत्यक्ष व्यक्तिगत अन्वेषण

इस विधि में अन्वेषक व्यक्तिगत रूप से प्रत्यार्थियों से सूचना प्राप्त करता है। वह सूचना प्राप्ति के लिए व्यक्तिगत सम्पर्क स्थापित करता है। इस विधि में अन्वेषक के बारे निम्नलिखित आशाएँ की जाती हैं।

i) उसे विनम्र, निष्पक्ष तथा चतुर होना चाहिए।

ii) उसे स्थानीय परिस्थितियों तथा रीति-रिवाजों का ज्ञान होना चाहिए ताकि वह अपनी पहचान प्रत्यार्थियों में से एक के रूप में दे सके।

iii) वह अच्छी प्रेक्षण शक्ति सहित बुद्धिमान होना चाहिए।

iv) सूचना प्राप्त करने के लिए उसे सरल तथा अर्थपूर्ण प्रश्न पूछने चाहिए।

यह विधि केवल गहन अन्वेषण के लिए उपयुक्त होती है। यह मुद्रा, समय तथा प्रयास की दृष्टि से खर्चीली है। इसके अतिरिक्त अन्वेषक के पक्षपात को निकाला जाना कठिन है, जिससे अन्वेषण को अत्यधिक हानि हो सकती है। यदि अन्वेषक में उपरोक्त गुण नहीं हैं तो यह विधि बिल्कुल बेकार सिद्ध हो सकती है।

2) अप्रत्यक्ष मौखिक अन्वेषण

इस विधि का प्राय: तब उपयोग किया जाता है जब विभिन्न कारणों से प्रत्यार्थी उत्तर देने में अनिच्छुक होते हैं। इस विधि में सूचना किसी गवाह या अन्य ऐसे व्यक्ति से प्राप्त की जाती है जो उस घटना से प्रत्यक्ष एवं अप्रत्यक्ष रूप में सम्बन्धित हो तथा इसके बारे में उसे पर्याप्त ज्ञान हो। इन सूचना देने वाले व्यक्तियों की निम्नलिखित विशेषताएँ होनी चाहिएँ :

i) उन्हें घटना की पूरी जानकारी होनी चाहिए।

ii) वे इसके बारे में निष्ठापूर्वक तथा ईमानदारी से बताने के लिए तैयार हों।

iii) वे पक्षपात एवं द्वेष न रखते हों।

iv) वे अन्वेषण के वास्तविक अभिप्राय के अनुकूल उत्तर देने में समर्थ हों।

3) स्थानीय सूचनाओं का उपयोग

इस विधि में अन्वेषक स्थानीय समाचारों तथा पत्रिकाओं का उपयोग करते हैं। यह सूचना स्थानीय पत्रकार न कि अन्वेषक द्वारा एकत्रित की हुई होती है। अत: इस विधि द्वारा प्रर्याप्त तथा विश्वसनीय परिणाम प्राप्त नहीं होते। यह विधि कम खर्चीली है लेकिन इसका उपयोग ऐसे अन्वेषण में नहीं किया जाना चाहिए जहाँ शुद्धता की ऊँची कोटि की आवश्यकता हो।

4) प्रश्नावली विधि

प्राथमिक समंकों के संकलन की यह सबसे महत्त्वपूर्ण तथा व्यवस्थित विधि है। इस विधि में अन्वेषण से सम्बन्धित प्रश्नों की सूची तैयार की जाती है, जिसे प्रश्नावली कहते हैं। प्रश्नावली के दो भाग होते हैं:

i) सामान्य आरंभिक भाग, जिसमें प्रत्यार्थी की पहचान सम्बन्धित प्रश्न जैसे नाम, पता, टेलिफोन नम्बर, शैक्षिक योग्यता, व्यवसाय आदि पूछे जाते हैं।

ii) मुख्य प्रश्न भाग, जिसमें अन्वेषण से सम्बन्धित प्रश्न होते हैं। ये प्रश्न विभिन्न अन्वेषणों के लिए भिन्न हो सकते हैं।

प्रश्नावली तैयार करने का कार्य बड़ा विशिष्ट होता है जिसको अनुभव द्वारा प्राप्त किया सकता है। अत: प्रश्नावली तैयार करते समय कुछ अनुभवी व्यक्तियों को साथ रखना आवश्यक है। प्रश्नावली तैयार करते समय निम्नलिखित महत्त्वपूर्ण बातों का ध्यान रखना आवश्यक होता है :

1) लोगों से वांछित रूप में तथा पर्याप्त शुद्धता सहित सूचना प्राप्त करना एक कठिन कार्य है। कुछ शंकाओं के कारण व्यक्ति सूचना देने के लिए स्वत: तैयार नहीं होते। बहुधा वे अपूर्ण तथा त्रुटिपूर्ण सूचना देते हैं। अत: उनको विश्वास में लेना आवश्यक होता है। उनको यह विश्वास दिलाया जाना चाहिए कि उनके द्वारा दी गई सूचना गोपनीय रखी जाएगी तथा इसका कोई भी अंश कर अधिकारियों या अन्य सरकारी संस्थाओं को नहीं दिया जाएगा।

2) जब सूचना देना कानूनी बंधन न हो तो प्रत्यार्थी को प्रार्थना या चतुर तर्कों द्वारा उत्तर देने के लिए प्रेरित किया जाना चाहिए। उनको यह बताया जाना चाहिए कि अन्वेषण के परिणामों

का उपयोग ऐसी नीतियों के बनाने में होगा जिनसे उनको लाभ होगा। अतः यह स्पष्ट है कि अन्वेषण के लिए अच्छी बिक्री कला आवश्यक है।

3) व्यक्तिगत प्रश्न, जिनसे प्रत्यार्थी को परेशानी हो, नहीं पूछे जाने चाहिएँ। उदाहरणार्थ, क्या आप आयकर की चोरी करते है या कालाबाज़ारी करते है या स्मगलिंग करते हैं? आदि प्रश्न नहीं पूछे जाने चाहिए।

4) भावनाओं को चोट पहुँचाने वाले प्रश्न नहीं पूछे जाने चाहिएँ। इनमें जुए की आदतें, सेक्स सम्बन्धी आदतें, ऋणग्रस्तता आदि सम्मिलित होती हैं।

5) ऐसे प्रश्न जिनमें जटिल परिकलन निहित हो, नहीं पूछे जाने चाहिएँ क्योंकि प्रत्यार्थी की इनमें रुचि तथा योग्यता होना आवश्यक नहीं है। इन परिस्थितियों में अन्वेषक को

   i) प्रत्यार्थी से तुलनपत्र, लाभ तथा हानि लेखा तथा सामान सूची (inventory) प्राप्त करके अपेक्षित सूचना का परिकलन स्वयं करना चाहिए।

   ii) सरल तथा अप्रत्यक्ष प्रश्न पूछकर अपेक्षित सूचना प्राप्त करनी चाहिए।

6) ऐसे प्रश्न पूछे जाने चाहिए जिससे प्रत्यार्थी द्वारा दी गई सूचना की सच्चाई की जाँच हो सके। उदाहरणार्थ, प्रशासनिक, उत्पादन, भण्डार, विपणन विभागों में विभिन्न प्रकार की श्रमिकों की जानकारी पर प्रश्न पूछकर फैक्ट्री की कुल मज़दूरी व्यय सूचना की जाँच की जा सकती है।

7) जहाँ तक संभव हो प्रश्न हाँ/नहीं प्रकार के होने चाहिए। ये सरल तथा सुनिश्चित होते हैं तथा इनके उत्तर में कम समय लगता है। बाद में इनका सारणीयन भी सरल होता है। उदाहरणार्थ:

   क्या आप विवाहित हैं? हाँ / नहीं

   सही उत्तर पर निशान (3) लगाएँ।

8) प्रश्न संक्षिप्त तथा सुस्पष्ट होने चाहिए अर्थात् ये अस्पष्ट तथा भ्रामक नहीं होने चाहिए। जहाँ तक संभव हो, प्रश्नों के वैकल्पिक उत्तर दिये जाएँ तथा प्रत्यार्थी को उनमें से एक पर, जो वह ठीक समझता है, निशान लगाने के लिए कहें। जब उत्तरों की सूची सम्पूर्ण न हो तो एक रिक्त लाइन छोड़ देनी चाहिए जिसके पहले "अन्य, यदि कोई है तो", लिखा हुआ होना चाहिए। इसको समझने के लिए हम एक उदाहरण लेते हैं। एक प्रश्न का उदाहरण निम्नलिखित है :

   लोग अपने मताधिकार का प्रयोग क्यों नहीं करते?

   सही उत्तर पर निशान (3) लगाइए।

   क) वे अनपढ़ हैं तथा मत के महत्त्व को नहीं समझते।

   ख) उनके ख्याल में लाखों मतों में से एक मत न देने से कोई अन्तर नहीं होता।

   ग) मतदान शिविर उनके घरों से बहुत दूर हैं।

   घ) उन्हें स्थानीय बदमाशों तथा हिंसा से डर लगता है।

   ङ) वे सरकार से खुश नहीं हैं तथा विरोध प्रकट करने के लिए मतदान नहीं करते।

   च) जब तक कोई मुद्रा प्रलोभन न हो, वे वोट नहीं डालते।

छ) अन्य कोई कारण, कृपया बताएँ : ....

प्रश्नों तथा उत्तरों का यह रूप समंकों के व्यवस्थितिकरण तथा सारणीकरण में सहायक होता है।

9) प्रश्नों की संख्या अधिक नहीं होनी चाहिए क्योंकि इससे प्रत्यार्थी में नीरसता की भावना उत्पन्न होती है। समय तथा रुचि के अभाव में प्रत्यार्थी अधिक प्रश्नों के उत्तर देने के इच्छुक नहीं होते।

परिवार नियोजन पर प्रश्नावली का एक प्रतिदर्श निम्नलिखित है।

परिवार नियोजन पर सर्वेक्षण

1. नाम :

2. पिता / पति का नाम :

3. घर का पता :

4. कार्यस्थान :

5. आयु : 6. पुरुष / स्त्री :

7. धर्म : 8. दूरभाष संख्या :

9. व्यवसाय : i) स्वयं : ii) पत्नी / पति :

10. सभी स्रोतों से पारिवारिक वार्षिक आय :

11. शैक्षिणिक योग्यता : (सही उत्तर पर निशान (3) लगाएँ।)

i) अनपढ़ ii) प्राथमिक स्तर

iii) मध्य स्तर iv) माध्यमिक

v) उच्चतर माध्यमिक vi) स्नातक

vii) स्नातकोत्तर

12. पत्नी / पति की शैक्षिणिक योग्यता : (सही उत्तर पर निशान (3) लगाएँ।)

i) अनपढ़ ii) प्राथमिक स्तर

iii) मध्य स्तर iv) माध्यमिक

v) उच्चतर माध्यमिक vi) स्नातक

vii) स्नातकोत्तर

13. वैवाहिक जीवन, वर्षों में : ____________________

14. उत्पन्न बच्चों की संख्या : लड़कियाँ ________ लड़के ________

15. जीवित बच्चों की संख्या : लड़कियाँ ________ लड़के ________

16. बच्चों में अन्तर, वर्षों में

i) शादी तथा प्रथम बच्चा : ______________

ii) प्रथम तथा द्वितीय बच्चा : ______________

iii) द्वितीय तथा तृतीय बच्चा : ______________

iv) तृतीय तथा चतुर्थ बच्चा : ______________

v) चतुर्थ तथा पाँचवाँ बच्चा : ______________

vi) पाँचवाँ तथा छठा बच्चा : ______________

17. क्या आप परिवार नियोजन के पक्ष में हैं? (हाँ / नहीं)

18. यदि नहीं, तो कारण बताएँ :

i) बच्चे प्राकृतिक उपहार हैं : (हाँ / नहीं)

ii) परिवार नियोजन मेरे धर्म के विरुद्ध है : (हाँ / नहीं)

iii) परिवार नियोजन का अर्थ है अनजन्मे बच्चे की हत्या : (हाँ / नहीं)

iv) बच्चों की संख्या मेरे भाग्य का हिस्सा है : (हाँ / नहीं)

v) अन्य कोई कारण, कृपया बताइए : ______________

19. यदि आप परिवार नियोजन के पक्ष में हैं, कारण बताएँ :

i) छोटा परिवार सुखी परिवार : (हाँ / नहीं)

ii) दो बच्चों को सरलता से नियंत्रित किया जा सकता है : (हाँ / नहीं)

iii) दो बच्चों की उपयुक्त शिक्षा तथा पोषण हो सकता है : (हाँ / नहीं)

iv) जीवन में कष्ट कम होते हैं : (हाँ / नहीं)

v) माता के स्वास्थ्य पर बुरा प्रभाव नहीं पड़ता : (हाँ / नहीं)

vi) अन्य कोई कारण, कृपया बताइए : ______________

20. अपने बच्चों की आयु, शिक्षण स्तर तथा स्वास्थ्य स्थिति की जानकारी दीजिए।

क्र.सं.	नाम	आयु	शिक्षण स्तर	स्वास्थ्य स्थिति (*)
1)	______	______	______	______
2)	______	______	______	______
3)	______	______	______	______
4)	______	______	______	______

(* कमजोर, सामान्य से निम्न, या उत्तम बताइए।)

प्रत्यार्थी के पास प्रश्नावली सहित किस प्रकार पहुँचना है?

इसके लिए हमारे पास तीन विधियाँ उपलब्ध हैं :

1) प्रश्नावली को डाक द्वारा भेजा जाय। इसके साथ एक अग्रसारण पत्र भी भेजा जाना चाहिए जिसमें प्रत्यार्थी, समाज या राष्ट्र के लिए सर्वेक्षण की व्याख्या की गई हों तथा इसको भरके

भेजने के सहयोग की प्रार्थना की गई हो। फिर आप उत्तरों की प्रतीक्षा कीजिए। प्राय: यह पाया गया है कि उत्तर बहुत कम प्राप्त होते हैं।

2) प्रश्नावली को अन्वेषकों के माध्यम द्वारा भेजिए जोकि प्रत्यार्थियों से प्रश्न पूछ कर स्वयं सूचना लिखेंगे। यह विधि चाहे खर्चीली है, लेकिन अच्छी है। यह प्रत्यार्थी को प्रश्न समझने में सहायक है। सुस्ती तथा गैरजिम्मेवारी का क्षेत्र कम होने के कारण इस विधि के अच्छे परिणाम मिलते हैं। एक चतुर तथा बुद्धिमान अन्वेषक अच्छे परिणाम पाने में सक्षम होता है।

3) प्रश्नावली को डाक द्वारा भेजने के पश्चात् अन्वेषक को क्षेत्र में भेज दीजीए। यह विधि सबसे उत्तम होती है क्योंकि इसमें दोनों विधियों के गुण सम्मिलित होते हैं। निस्संदेह, यह खर्चीली है। यह विस्तृत अध्ययनों के लिए बहुत उपयोगी है। खर्चीली होने के कारण प्राय: इस विधि का उपयोग सरकार द्वारा किया जाता है, जिसके पास साधनों की कोई कमी नहीं होती।

2.5 द्वितीयक समंकों का संकलन

जैसा अनुभाग 2.3.4 में बताया गया था कि प्रत्यक्ष अन्वेषण चाहे वांछनीय होता लेकिन मुद्रा, प्रयास तथा समय की दृष्टि से खर्चीला होता है। विकल्पत:, सूचना द्वितीयक स्रोत से भी प्राप्त की जा सकती है। इसका अर्थ किसी अन्य संस्था द्वारा संकलित समंकों से सूचना प्राप्त करना होता है। तकनीकी तौर पर इन समंकों को द्वितीयक समंक कहते हैं।

2.5.1 द्वितीयक समंकों की सीमाएँ

हालाँकि द्वितीयक स्रोत मुद्रा, समय तथा प्रयास की दृष्टि से सस्ता होता है, इन समंकों का उपयोग बड़े ध्यानपूर्वक किया जाना चाहिए। समंक विशाल तथा विश्वसनीय होने वांछनीय है तथा प्रयोग किए गए पारिभाषिक शब्द तथा परिभाषाएँ वर्तमान अन्वेषण की परिभाषाओं के सदृश होने चाहिएँ। समंकों की उपयुक्तता की जाँच वर्तमान अन्वेषण तथा मूल अन्वेषण की प्रकृति तथा कार्यक्षेत्र की तुलना द्वारा की जा सकती है। यदि द्वितीयक समंक निष्पक्ष, बुद्धिमान एवं प्रशिक्षित अन्वेषक द्वारा संकलित किए गए हों तो विश्वसनीय होते हैं। इन समंकों की समयावधि की भी उपयुक्त जाँच की जानी चाहिए। कॉनोर ने ठीक ही कहा है कि, ''समंक विशेषत: अन्य व्यक्ति के (द्वारा एकत्रित) समंक प्रयोगकर्त्ता के लिए कमियों से पूर्ण होते हैं''। अर्थात् यह आवश्यक नहीं है कि, द्वितीयक समंकों के उपयोग से पहले, अन्वेषक को इनसे मुद्रा, समय तथा प्रयास की बचत के लाभों की तुलना त्रुटिपूर्ण निष्कर्षों की हानियों से कर लेनी चाहिए। समंकों का उपयोग सुरक्षित है या नहीं की जाँच इनकी पर्याप्तता, उपयुक्तता तथा विश्वसनीयता की जाँच द्वारा की जानी चाहिए।

अत: द्वितीयक समंकों के उपयोग से पहले हमें यह जाँचना आवश्यक है कि क्या यह :

i) विश्वसनीय हैं,

ii) उपयुक्त हैं, तथा

iii) पर्याप्त हैं?

स्पष्टत:, समंकों की विश्वसनीयता किन्हीं भी समंकों के लिए आवश्यक होती हैं। तथा ऐसा द्वितीयक समंकों के लिए और आवश्यक है। इस दृष्टि से प्रयोगकर्ता को सन्तुष्ट होना अनिवार्य है। उसे इस बात की जँच कर लेनी चाहिए कि समंक एक विश्वसनीय स्रोत से तथा विश्वसनीय, निष्पक्ष तथा प्रशिक्षित अन्वेषक द्वारा संकलित किए गए हैं। दूसरे, यह भी जानना चाहिएं कि क्या समंक उसी प्रकार के वर्ग से प्राप्त किए गए है या नहीं। तीसरे, क्या समय बीतने के कारण वर्ग समूह की आदत, रीति-रिवाज तथा फैशन में कोई अन्तर आया है नहीं। निस्संदेह बिल्कुल वैसी ही परिस्थितियों की आशा नहीं की जानी चाहिए।

समंकों की उपयुक्तता एक अन्य अपेक्षा है। अन्वेषक को यह विश्वास हो जाना चाहिए कि उसके द्वारा उपयोग किए जाने वाले समंक अन्वेषण के लिए उपयुक्त हैं। उसे स्रोत के प्राचलों, जैसे व्यक्तियों का वर्ग, भौगोलिक क्षेत्र, अवधारणाओं की परिभाषा, मापन इकाई, समय तथा अन्य, की तुलना अन्वेषण के प्राचलों से कर लेना चाहिए। इसके अतिरिक्त, उपयुक्तता के लिए, लक्ष्यों तथा उद्देश्यों की तुलना भी कर लेनी चाहिए।

द्वितीयक समंकों का विश्वसनीय तथा उपयुक्त होने के साथ-साथ पर्याप्त होना भी आवश्यक होता है। अतः अन्वेषण की आवश्यकता की तुलना में उपलब्ध समंक पर्याप्त होने वांछनीय हैं। उदाहरणार्थ, एक राज्य के उपभोग स्वरूप के समंक बड़े तथा कस्बों के समंकों से प्राप्त नहीं किए जा सकते।

बोध प्रश्न 2

1) कारण सहित बताइए कि क्या निम्नलिखित वाक्य ठीक हैं या नहीं?
   i) द्वितीयक समंक प्राथमिक समंकों से श्रेष्ठ होते हैं।
   ii) 1991 की भारतीय जनसंख्या संगणना से प्राप्त समंक प्राथमिक स्रोत है।
   iii) द्वितीयक समंकों को बिना जाँच के स्वीकृत नहीं किया जाना चाहिए।
   iv) एक लम्बी सूची वाले प्रश्नों की प्रश्नावली उचित होती है।
   v) सभी सर्वेक्षण तकनीकों में से प्रश्नावली विधि श्रेष्ठ होती है।

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

2) जब प्राथमिक समंकों के संकलन की प्रत्यक्ष व्यक्तिगत अन्वेषण विधि प्रयोग की जानी हो तो अन्वेषक की दो महत्त्वपूर्ण विशेषताओं के बारे में जानकारी दीजिए।

....

....

....

....

....

....

....

....

....

3) विभिन्न सर्वेक्षण तकनीकों की व्याख्या कीजिए ।

....

....

....

....

....

....

....

....

....

....

....

....

4) इस कथन पर टिप्पणी कीजिए : "हमें सदैव द्वितीयक समंक उपयोग करने चाहिए"।

....

....

....

....

....

....

....

....

....

....

....

2.6 सारांश

समंक / सांख्यिकी परिमाणात्मक सूचना होती है तथा इसका प्रतिदर्श या संगणना; प्राथमिक या द्वितीयक समंक के रूप में भेद किया जा सकता है। एक अन्वेषण के संचालन के लिए समंकों की आवश्यकता होती है जोकि नए सिरे से या द्वितीयक स्रोत से प्राप्त किए जा सकते हैं। इन दोनों के लिए सांख्यिकीय सर्वेक्षण की आवश्यकता होती है जिसके आयोजन चरण तथा संचालन चरण होते हैं। आयोजन चरण में अन्वेषक को प्राथमिक स्रोत या द्वितीयक स्रोत, संगणना या प्रतिदर्श अन्वेषण, सांख्यिकीय इकाई की प्रकृति, तथा मापन इकाई, शुद्धता की कोटि आदि के बारे में निर्णय करने होते हैं।

संचालन चरण में अन्वेषक का मुख्य कार्य प्रशासनिक व्यवस्था स्थापित करना, कार्यक्षेत्र कर्मचारियों की नियुक्ति तथा प्रशिक्षण तथा समंक संकलन की सारी प्रक्रिया का निरीक्षण करना होता हैं।

प्रकाशित तथा अप्रकाशित स्रोतों से प्राप्त द्वितीयक समंकों का उपयोग सावधानी से किया जाना चाहिए क्योंकि इनमें विभिन्न कमियाँ होती हैं। सभी सर्वेक्षण तकनीकों में से प्रश्नावली विधि अति महत्त्वपूर्ण है। प्रश्नावली में प्रासंगिक प्रश्न होते हैं जोकि सरल, सुस्पष्ट तथा हाँ / नहीं प्रकार के होने चाहिए। व्यक्तिगत तथा भावनाओं को चोट पहुँचाने वाले प्रश्न नहीं पूछे जाने चाहिए।

2.7 शब्दावली

समंक बिन्दु : यह एक व्यक्ति या वस्तु से प्राप्त प्रेक्षण होता है।

समंक समुच्चय : यह सभी समंक बिन्दुओं का समुच्चय होता है।

संगणना समुच्चय : समष्टि की सभी इकाइयों के प्रेक्षण से प्राप्त समंक होते हैं।

प्रतिदर्श समंक : प्रतिदर्श में सम्मिलित इकाइयों के प्रेक्षण से प्राप्त समंक होते हैं।

प्राथमिक समंक : एक विचाराधीन अन्वेषण के अन्तर्गत इकाइयों के प्रेक्षण से प्राप्त समंक होते हैं।

द्वितीयक समंक : किसी अन्य संस्था द्वारा संकलित समंकों से प्राप्त समंक होते हैं।

प्रश्नावली या प्रश्न सूची : वर्तमान अन्वेषण से सम्बन्धित प्रश्नों की सूची होती हैं।

सांख्यिकीय अन्वेषण : यह एक अन्वेषण होता है जिसकी जाँच के लिए संख्याओं में सूचना की आवश्यकता होती है।

सांख्यिकीय सर्वेक्षण : यह एक दी हुई समस्या के अन्तर्गत आने वाली सभी या प्रतिदर्श में सम्मिलित इकाइयों के प्रेक्षण से समंक प्राप्त करनी की विधि होती है।

सांख्यिकीय इकाई : एक व्यक्ति या वस्तु का एक या अधिक अभिलक्षण होता है जिनका समंक संकलन के लिए प्रेक्षण किया जाता है।

प्रत्यार्थी : ये सूचना देने वाला व्यक्ति होता है।

अन्वेषक : ये प्रत्यार्थी से सूचना प्राप्त करने वाला व्यक्ति होता है।

परिकल्पना : यह समष्टि के बारे में एक कथन होता हैं।

परिकल्पना जाँच : संकलित समंको के आधार पर किसी परिकल्पना की वैधता की जाँच करना।

2.8 कुछ उपयोगी पुस्तकें

Elhance, D. N. and V. Elhance, 1988, *Fundamentals of Statistics*, Kitab Mahal, Allahabad.

Nagar, A. L. and R. K. Dass, 1983, *Basic Statistics*, Oxford University Press, Delhi.

Mansfield, E., 1991, *Statistics for Business and Economics : Methods and Applications*, W. W. Norton and Co.

Yule, G. U. and M. G. Kendall, 1991, *An Introduction to the Theory of Statistics*, Universal Books, Delhi.

2.9 बोध प्रश्नों के उत्तर या संकेत

बोध प्रश्न 1

1) i), ii) तथा vi) के लिए अनुभाग 2.3.3 देखिए।
   iii) के लिए भाग 2.2 देखिए।
   iv) के लिए अनुभाग 2.3.2 देखिए।
   v) के लिए अनुभाग 2.3.1 देखिए।

2) i) भाग 2.1 देखिए।
   ii) भाग 2.2 देखिए।
   iii), iv), v), viii) भाग 2.2 देखिए।
   vi) अनुभाग 2.3.2 देखिए।
   vii), ix) अनुभाग 2.3.1 देखिए।
   x) अनुभाग 2.3.4 देखिए।

3) अनुभाग 2.3.4 देखिए।

बोध प्रश्न 2

1) i), ii), iv) गलत।
   iii), v) सही।

2) भाग 2.4 देखिए।

3) भाग 2.4 देखिए।

4) अनुभाग 2.3.4 देखिए।

2.10 पारिभाषिक शब्दावली

चर	:	variable
गुण	:	attribute
संगणना	:	census
प्राथमिक समंक	:	primary data
द्वितीयक समंक	:	secondary data
प्रश्नावली	:	questionnaire
सांख्यिकीय अन्वेषण	:	statistical inqury
सांख्यिकीय सर्वेक्षण	:	statistical survey
सांख्यिकीय इकाई	:	statistical unit
प्रत्यार्थी	:	respondent
अन्वेषक	:	investigator
परिकल्पना	:	hypothesis

इकाई 3 समंकों का सारणीयन तथा आलेखी प्रस्तुतिकरण

इकाई की रूपरेखा

3.0 उद्देश्य

इस इकाई के अध्ययन के बाद आप को निम्नलिखित के बारे में जानकारी मिलेगी :

- समंकों के संकलित हो जाने के बाद सांख्यिकीय अन्वेषण के चरण;

- समंकों के संगठन (वर्गीकरण तथा विन्यास) तथा संक्षेपण की विधियाँ;
- बारंबारता बंटन तथा इनके विभिन्न रूप; और
- सांख्यिकीय समंकों के प्रस्तुतिकरण की विभिन्न विधियाँ जैसे सारणी आलेख, आरेख, चित्रलेख आदि।

3.1 प्रस्तावना

इससे पहली इकाई में हमने समंकों के संकलन, सांख्यिकीय सर्वेक्षण या द्वितीयक स्रोत, की विधियों का विवेचन किया था। प्राथमिक स्रोत में संकलित समंक प्राय: अव्यवस्थित होते हैं। शुरू में वे हजारों प्रश्नावलियों में होते है। इनको समझने के लिए इनका संगठन (अर्थात् वर्गीकरण तथा विन्यास) तथा संक्षेपण आवश्यक है। इसके लिए हम विभिन्न विधियों का उपयोग कर सकते हैं, जैसे प्रश्नावलियों से सूचना को मास्टर पत्रों में लिखना आदि। इन पत्रों से हम संक्षिप्त सारणी तैयार कर सकते हैं। आधुनिक युग में समंकों के शीघ्र संगठन तथा संक्षेपण के लिए कम्प्यूटर का उपयोग किया जा सकता है। ऐसे कम्प्यूटर सोफटवेयर उपलब्ध हैं जो कि विभिन्न प्रकार के आलेख तथा आरेख बनाने में सहायक होते हैं। समंकों का संख्यात्मक संक्षेपण भी किया जा सकता है। इसके लिए हम संक्षेपण माप जैसे प्रथम कोटि के केन्द्रीय प्रवृत्ति के माप (समान्तर, गुणोत्तर तथा हरात्मक माध्य, बहुलक तथा माध्यिका); द्वितीय कोटि के केन्द्रीय प्रवृत्ति के माप या प्रकीर्णन के माप (परिसर, चतुर्थक विचलन, माध्य विचलन तथा मानक विचलन); द्विचर विश्लेषण में साहचर्य के माप (सहसंबंध तथा समाश्रयण); सूचकांक इत्यादि उपयोग करते हैं। इस इकाई में हम सारणी तथा आलेखों द्वारा समंकों के संक्षेपण का विवचेन करेंगे। संख्यात्मक संक्षेपण का विवेचन बाद के खंडों (2, 3 तथा 4) में किया जाएगा। यहाँ यह बताना आवश्यक है कि अच्छा संक्षेपण तथा प्रस्तुतिकरण स्वयं में लक्ष्य नहीं है। वास्तव में यह समंकों के उपयोगी विश्लेषण तथा निर्वचन का पहला चरण है। फिर भी अच्छा प्रस्तुतिकरण महत्त्वपूर्ण तथ्यों को सामने लाने तथा उनकी तुलना करने में सहायक होता है। समंकों को बोलने योग्य बना कर उनका बुद्धिमता से उपयोग किया जा सकता है।

3.2 सांख्यिकीय अन्वेषण के चरण

जैसा कि इकाई 2 के भाग 2.3 में पढ़ा था, सांख्यिकीय सर्वेक्षण या अन्वेषण दो चरणों, आयोजन तथा संचालन, में किया जाता है। इस इकाई में हम संचालन पहलू पर ध्यान केन्द्रित करेंगे। इसमें समंकों का सरल क्रम-विन्यास (आरोही तथा अवरोही क्रम) बारंबारता क्रम-विन्यास तथा संतत बारंबारता बंटन इत्यादि; तथा सारणी तथा आलेखों के रूप में संगठन तथा संक्षेपण सम्मिलित होता है।

3.3 समंकों का विन्यास

संकलित समंकों का ढेर प्राय: बड़ा अबोधगम्य तथा उबाने वाला होता है। यह बिल्कुल नीरस तथा सरलता से समझ न आने वाला होता है। उदाहरणार्थ, 1000 परिवारों की मासिक आय के समंक दिए होने पर उनको समझना कठिन है। लेकिन यदि उनकी औसत मासिक आय 2540 रुपये दी हो तो इसकी अन्य समंकों से तुलना की जा सकती है।

समंकों के विश्लेषण तथा निर्वचन का पहला कदम उनका वर्गीकरण तथा सारणीयन होता है। समंकों का एक समान अभिलक्षण के अनुसार विन्यास, वर्गीकरण कहलाता है। इसके विपरीत समंकों का प्रमुख अभिलक्षणों के अनुसार पंक्तियों तथा स्तम्भों में सुव्यवस्थित प्रस्तुतिकरण सारणीयन कहलाता है। इकाई 2 में हमने परिवार नियोजन पर एक प्रश्नावली तैयार की थी। मान लिया यह प्रश्नावली नई दिल्ली की XYZ कालोनी के C-III ब्लाक के 50 परिवारों से सूचना प्राप्त करने के लिए उपयोग की गई। यह सूचना सारणी 3.1 तथा 3.2 में दी हुई है। क्या हम इसे समझ सकते हैं?

सारणी 3.1

प्रति परिवार बच्चों की संख्या, C-III ब्लाक, XYZ कालोनी, नई दिल्ली

2	0	1	5	3	1	2	1	0	2
4	3	2	2	3	4	1	0	2	3
1	4	2	3	1	2	5	4	1	3
2	1	3	2	3	4	1	2	3	1
4	5	2	1	1	0	3	2	0	2

सारणी 3.2

50 परिवारों की मासिक आय, C-III ब्लाक, XYZ कालोनी, नई दिल्ली

547	622	691	684	567	586	680	578	583	578
708	544	528	540	730	541	720	698	763	633
640	637	598	631	618	692	600	650	604	640
646	654	689	736	731	844	798	712	772	820
678	663	800	692	700	781	658	798	709	720

जैसा कि पहले बताया जा चुका है, ऊपर दिए हुए अपरिष्कृत समंकों के ढेर को समझने के लिए इनका वर्गीकरण तथा विन्यास आवश्यक है। यह एक सरल क्रम विन्यास या बारंबारता क्रम-विन्यास (असंतत बारंबारता बंटन) या संतत बारंबारता बंटन तैयार करके किया जा सकता है। इस पहलु की व्याख्या अनुभाग 3.3.1, 3.3.2 तथा 3.3.3 में की गई है।

3.3.1 सरल क्रम-विन्यास

यह दिए हुए अपरिष्कृत (एक विचर) संमकों का आरोही तथा अवरोही क्रम में विन्यास होता है। आरोही क्रम में प्रेक्षणों का विन्यास, आकार के बढ़ते क्रम किया जाता है। उदाहरणार्थ, संख्याएँ 3, 5, 7, 8, 9, 10 आरोही क्रम में हैं। अवरोही क्रम इसका विपरीत होता है। उदाहरणार्थ, संख्याएँ 10, 9, 8, 7, 5, 3 अवरोही क्रम में हैं।

हम सारणी 3.1 से दोनों प्रकार के सरल क्रम-विन्यास तैयार कर सकते हैं। निम्नलिखित सारणी में समंकों का आरोही विन्यास किया गया है। इस विन्यास से यह स्पष्ट है कि न्यूनतम मान 0 तथा उच्चतम मान 5 है।

सारणी 3.3

प्रति परिवार बच्चों की संख्या, C-III ब्लाक, XYZ कालोनी, नई दिल्ली

सरल क्रम विन्यास-आरोही क्रम

0	0	0	0	0	1	1	1	1	1
1	1	1	1	1	1	1	2	2	2
2	2	2	2	2	2	2	2	2	2
2	3	3	3	3	3	3	3	3	3
3	4	4	4	4	4	4	5	5	5

इस विन्यास के बाद सारणी 3.1 में दिए हुए समंकों का कुछ अर्थ समझ में आने लगता है। इस विन्यास से संभवत: यह निष्कर्ष निकाले जा सकते हैं कि 5 परिवारों में बच्चे नहीं है, प्रत्येक 12 परिवारों में 1 बच्चा है, प्रत्येक 14 परिवारों में 2 बच्चे हैं, प्रत्येक 10 परिवारों में 3 बच्चे हैं, प्रत्येक 6 परिवारों में 4 बच्चे हैं तथा प्रत्येक 3 परिवारों में 5 बच्चे हैं।

3.3.2 बारंबारता क्रम-विन्यास या असंतत बारंबारता बंटन

यहाँ सरल क्रम विन्यास, जैसे 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1,1 आदि, की भाँति विभिन्न प्रेक्षणों को बार-बार नहीं लिखा जाता। हम एक प्रेक्षण कितनी बार घटित होता है (अर्थात् बारंबारता) की गणना करते हैं। उदाहरणार्थ, सारणी 3.3 में प्रेक्षण 4, छ: बार घटित हुआ। अत: 4 की बारंबारता 6 है। सारणी 3.3 में दिए हुए सरल क्रम विन्यास का बारंबारता क्रम विन्यास, सारणी 3.4 में दिया गया है।

बारंबारता क्रम-विन्यास एक सांख्यिकीय सारणी होती है। जिसमें विभिन्न प्रेक्षणों को उनके आकार के अनुसार क्रमबद्ध करके उनकी क्रमश: बारंबारताओं सहित लिखा जाता है।

सारणी 3.4

बच्चों की संख्या	:	0	1	2	3	4	5	योग
परिवारों की संख्या	:	5	12	14	10	6	3	50

जब प्रेक्षणों की संख्या अधिक हो तो गणना का कार्य मिलान दंड़ (Tally bar) के उपयोग द्वारा किया जाता है। इस विधि में चर के हर संभव मान एक स्तम्भ में लिख दिए जाते हैं। प्रत्येक प्रेक्षण के लिए एक मिलान दण्ड (।) उसके मान के सम्मुख लगाया जाता है। प्रत्येक पाँचवें प्रेक्षण का दण्ड पहले चार दण्डों को काटता हुआ दिखाया जाता है जैसे (卌)। इस प्रकार हम पाँच-पाँच प्रेक्षणों के समूह तैयार कर लेते हैं जोकि अन्त में गणना को सरल बना देते हैं। इस प्रकार, 14 बार घटित हुआ प्रेक्षण (卌 卌 ||||) प्रकार से लिखा जाता है। यह ध्यान रखना अति आवश्यक है कि मिलान पत्र प्रेक्षण का मिलान दण्ड लगाने के तुरन्त बाद उस प्रेक्षण पर (×) या (√) चिह्न लगा देना चाहिए जिससे वह दोबारा गिना न जा सके। सारणी 3.1 के समंक, बारंबारता बंटन के रूप में पुन: सारणी 3.5 में लिखे गए हैं।

सारणी 3.5

प्रति परिवार बच्चों की संख्या का बारंबारता बंटन

बच्चों की संख्या	मिलान पत्र	बारंबारता
0	卌	5
1	卌 卌 ||	12
2	卌 卌 ||||	14
3	卌 卌	10
4	卌 |	6
5	|||	3
योग		50

3.3.3 संतत या वर्गीकृत बारंबारता बंटन

1, 2, 3, 4, 5, 20, 40 आदि संख्याएँ असंतत होती है। तथा इनका उपयोग तब किया जाता है जहाँ पर दो लगातार संख्याओं के मध्य मान संभव नहीं होता। जैसा की बच्चों की संख्या के उदाहरण में, यह संभव नहीं है तथा हास्यकर है कि एक विशेष परिवार में बच्चों की संख्या 2.083 या 2.1 या 2.75 है। एक परिवार में 2 या 3 बच्चे हो सकते हैं। भाग 3.3 में दिए गए अपरिष्कृत समंकों के दो उदाहरणों में से बच्चों की संख्या (सारणी 3.1) एक असंतत समंकों का उदाहरण है जबकि मासिक आय (सारणी 3.2) एक संतत चर का उदाहरण है जिससे संतत समंक प्राप्त होते हैं।

इस भाग में हम 50 परिवारों की मासिक आय के अपरिष्कृत समंकों से संतत या वर्गीकृत बारंबारता बंटन तैयार करने की विधि का अध्ययन करेंगे।

एक वर्गीकृत बारंबारता बंटन तैयार करने के लिए हम समंकों के परिसर, अर्थात् सबसे बड़े तथा सबसे छोटे प्रेक्षणों का अन्तर, को विभिन्न परस्पर अपवर्जी तथा नि:शेष अन्तरालों, जिनको वर्ग-अन्तराल कहते हैं, में बाँटा जाता है। इसके बाद प्रत्येक वर्ग अन्तराल की बारंबारता की गणना करके उसके सम्मुख लिख दिया जाता है।

सारणी 3.6

परिवारों की मासिक आय का बारंबारता बंटन

मासिक आय (रुपये में)	मिलान पत्र	परिवारों की संख्या (बारंबारता)
500 - 550	IIII	5
550 - 600	IIII I	6
600 - 650	IIII IIII	10
650 - 700	IIII IIII II	12
700 - 750	IIII IIII	9
750 - 800	IIII	5
800 - 850	III	3
योग		50

सारणी 3.6 में हमने एक संतत या वर्गीकृत बारंबारता बंटन तैयार करने का अभ्यास किया है जिसमें समंकों को नियंत्रित रूप में लाने के लिए परिवार की आय चर को विभिन्न वर्गों में बाँटा गया। लेकिन किसी वर्गीकृत बारंबारता बंटन के तैयार करने से पहले हमें निम्नलिखित प्रश्नों के उत्तर जान लेने आवश्यक होते हैं।

1) वर्ग अंतरालों की संख्या कितनी होनी चाहिए?
2) प्रत्येक वर्ग अन्तराल की चौड़ाई कितनी होनी चाहिये?
3) वर्ग सीमाएँ कैसे निर्दिष्ट की जाएँगी?

1) वर्ग अंतरालों की संख्या कितनी होनी चाहिए?

यद्यपि बनाए जाने वाले वर्गों की संख्या के बारे में कोई ठोस नियम नहीं है। फिर भी इनकी संख्या न तो बहुत ही कम तथा न बहुत अधिक होनी चाहिए। यदि वर्गों की संख्या बहुत कम है, अर्थात् प्रत्येक वर्ग की चौड़ाई अधिक है, तो वर्गीकरण के कारण सूचना की क्षति होने की संभावना होती है। इसके विपरीत, वर्गों की अधिक संख्या होने पर बंटन खण्डित प्रतीत होता है जिसके कारण आचरण अस्पष्ट रहने की संभावना होती है। अनुभव के आधार पर यह पाया गया है कि किसी भी परिस्थिति में वर्गों की न्यूनतम संख्या 5 या 6 से कम तथा अधिकतम संख्या 20 से अधिक नहीं होनी चाहिए।

वर्गों की संख्या ज्ञात करने के लिए प्राय: निम्नलिखित सूत्र दिया जाता है :

वर्गों की संख्या $= 1 + 3.222 \times \log_{10} N$, जहाँ पर N, कुल प्रेक्षणों की संख्या है। 50 परिवारों के आय के अपरिष्कृत समंकों के उदाहरण में वर्गों की संख्या इस प्रकार से परिकलित की जा सकती है।

$$\text{वर्गों की संख्या} = 1 + 3.222 \times \log_{10} 50 = 1 + 3.322 \times 1.6990$$
$$= 1 + 5.644 = 6.644 \approx 7.$$

2) प्रत्येक वर्ग अंतराल की चौड़ाई कितनी होनी चाहिए?

जहाँ तक संभव हो सभी वर्ग अंतराल समान चौड़ाई के होने चाहिए। लेकिन जब एक समान वर्ग अंतरालों पर आधारित बांरबारता बंटन द्वारा प्रेक्षणों का नियमित प्रतिरूप प्रदर्शित नहीं होता तो प्रेक्षणों को असमान चौड़ाई के वर्गों में बाँटना आवश्यक हो सकता है। नियमित आचरण प्रतिरूप से अर्थ यह है कि सिरे वाले वर्गों को छोड़कर ऐसे कोई वर्ग न हों जिनमें शून्य या कुछ ही प्रेक्षण हो तथा उनके संलग्न वर्ग में प्रेक्षणों का केन्द्रीयकरण हो।

एक वर्ग की सन्निकट चौड़ाई, निम्नलिखित सूत्र द्वारा ज्ञात की जा सकती है :

$$\text{एक वर्ग की सन्निकट चौड़ाई} = \frac{\text{सबसे बड़ा प्रेक्षण} - \text{सबसे छोटा प्रेक्षण}}{\text{वर्ग अंतरालों की संख्या}}$$

फिर भी, वर्ग अंतरालों की चौड़ाई के विषय में अन्तिम निर्णय निम्नलिखित बातों को ध्यान में रखते हुए करना चाहिए :

i) जहाँ तक संभव हो चौड़ाई 5 का गुणज होनी चाहिए क्योंकि 5, 10, 15... आदि संख्याओं को समझना सरल होता है।

ii) वर्ग का मध्य बिन्दु ज्ञात करना सुविधाजनक होना चाहिए।

iii) एक वर्ग में प्रेक्षण एक समान बंटित होने चाहिए।

3) वर्ग सीमाएँ कैसे निर्दिष्ट की जाएँगी?

एक वर्ग अंतराल के सबसे छोटे तथा सबसे बड़े प्रेक्षणों को वर्ग सीमाएँ कहते है। इनको क्रमश: वर्ग की निम्न तथा उच्च सीमाएँ कहते है। क्योंकि एक वर्ग का मध्य मान जो कि माध्य मानक विचलन आदि के परिकलन में प्रयोग होता है, इन वर्ग सीमाओं से प्राप्त किया जाता है। अत: इनको सुस्पष्ट रूप में परिभाषित करना अति आवश्यक है। वर्ग सीमाओं को परिभाषित करते समय निम्नलिखित बातों को ध्यान में रखना चाहिए :

i) पहले वर्ग की निम्न सीमा का समंकों के न्यूनतम प्रेक्षण के बराबर होना आवश्यक नहीं है। वास्तव में यह न्यूनतम प्रेक्षण से कम या बराबर हो सकती है। इसी प्रकार अंतिम वर्ग की उच्च सीमा समंकों के अधिकतम प्रेक्षण से अधिक या बराबर हो सकती है।

ii) वर्ग की न्यूनतम सीमा 0 या 5 का गुणज रखना सुविधाजनक होता है।

iii) वर्ग सीमाएँ ऐसी होनी चाहिएँ कि वर्ग में प्रेक्षण समान रूप से बंटित हों।

वर्ग सीमाएँ निम्नलिखित किसी भी विधियों द्वारा परिभाषित की जा सकती हैं।

1) अपवर्जी विधि, तथा
2) समावेशी विधि।

1) अपवर्जी विधि

इस विधि में एक वर्ग की उच्च सीमा अनुसरण वर्ग की निम्न सीमा के बराबर होती है। विभिन्न वर्ग अंतरालों को परस्पर अपवर्जी रखने के लिए यह निर्णय लिया जाता है कि वे प्रेक्षण जिनका मान निम्न सीमा से अधिक या बराबर तथा उच्च सीमा से कम हो, को इस वर्ग में सम्मिलित किया जाता है। उदाहरणार्थ 500 - 550 के वर्ग में वह सभी प्रेक्षण सम्मिलत होंगे जिनके मान 500 या अधिक है लेकिन 550 से कम है। एक प्रेक्षण जिसका मान 550 है 550 - 600 के वर्ग में सम्मिलित किया जाएगा।

अपवर्जी वर्ग अंतरालों का मुख्य लाभ यह है कि इससे समंकों की संतत बनाए रखना संभव होता है क्योंकि एक वर्ग की उच्च सीमा अनुसरण वर्ग की निम्न सीमा के बराबर होती है। हमारे मासिक आय के उदाहरण (सारणी 3.6) में 5 परिवारों की आय 550 - 550 रुपये अर्थात् 500 - 549 रुपये है तथा 6 परिवारों की आय 550 - 600 रुपये अर्थात् 550 से 599 है आदि। इस परिकल्पना के आधार पर हम इस बारंबारता बंटन को निम्नलिखित रूप में भी लिख सकते हैं।

सारणी 3.7

अपवर्जी वर्ग अंतराल

मासिक आय (रुपये)	परिवारों की संख्या (बारंबारता)
500 लेकिन 550 से कम	5
550 लेकिन 600 से कम	6
600 लेकिन 650 से कम	10
650 लेकिन 700 से कम	12
700 लेकिन 750 से कम	9
750 लेकिन 800 से कम	5
800 लेकिन 850 से कम	3
योग	50

2) **समावेशी विधि**

इस विधि में वे सभी प्रेक्षण उस वर्ग में सम्मिलित किए जाते हैं जिनका मान निम्न सीमा के बराबर या अधिक लेकिन उच्च सीमा से कम या बराबर होता है। निम्नलिखित सारणी 3.8 में 549 रुपये आय को वर्ग 500 से 549 में सम्मिलित किया गया है। इस प्रकार 550 रुपये आय स्वत: वर्ग 550 से 599 में सम्मिलित हो जाएगी। क्योंकि एक वर्ग की उच्च सीमा अनुसरण वर्ग की निम्न सीमा के बराबर नहीं है। अत: प्रेक्षणों के वर्गीकरण में भ्रांति की कोई संभावना नहीं होती।

सारणी 3.8

समावेशी वर्ग अंतराल

मासिक आय (रुपये)	परिवारों की संख्या (बारंबारता)
500 - 549	5
550 - 599	6
600 - 649	10
650 - 699	12
700 - 749	9
750 - 799	5
800 - 849	3
योग	50

अपवर्जी तथा समावेशी विधियों में चयन इस बात पर निर्भर होता है कि क्या चर संतत है जैसे आय, ऊँचाई, वज़न आदि या असंतत है, जैसे परिवार में बच्चों की संख्या आदि। संतत चर के लिए

बारंबारता बंटन अपवर्जी विधि से तैयार करना वांछनीय होता है। इसके विपरीत असंतत चर जैसे परिवार में बच्चों की संख्या या प्रथम श्रेणी में उत्तीर्ण होने वाले विद्यार्थियों की संख्या के लिए बारंबारता बंटन समावेशी विधि द्वारा तैयार किया जाता है।

एक वर्ग का मध्य मान

जब वर्ग अंतराल अपवर्जी हों तो वर्ग का मध्य बिन्दु या वर्ग संकेत उसकी निम्न तथा उच्च सीमाओं का समान्तर माध्य होता है। लेकिन समावेशी वर्ग अन्तराल होने पर एक वर्ग की उच्च सीमा तथा अनुसरण वर्ग की निम्न सीमा में अन्तर होता है। इस अन्तर को समाप्त करने के लिए इसका आधा उच्च सीमा में जोड़ दिया जाता है तथा इतना ही निम्न सीमा में से घटा दिया जाता है। इस प्रकार प्राप्त नई सीमाओं को वर्ग परिसीमाएँ कहते हैं।

सारणी 3.8 में दिए हुए समावेशी अंतरालों की वर्ग परिसीमाएँ सारणी 3.9 में दी गई है।

सारणी 3.9

मासिक आय (रुपये)	परिवारों की संख्या (बारंबारता)
499.5 - 549.5	5
549.5 - 599.5	6
599.5 - 649.5	10
649.5 - 699.5	12
699.5 - 749.5	9
749.5 - 799.5	5
799.5 - 849.5	3
योग	50

3.3.4 बारंबारता बंटन के विभिन्न रूप

इस अनुभाग में हम निम्नलिखित बारंबारता बंटनों के अर्थ का परिचय देगें।

1) विवृत-छोर बारंबारता बंटन

2) असमान वर्ग अंतराल सहित बारंबारता बंटन

3) संचयी बारंबारता बंटन

4) सापेक्षिक बारंबारता बंटन

1) विवृत-छोर बारंबारता बंटन

वह बारंबारता बंटन जिसका कम से कम एक छोर विवृत हो, विवृत-छोर बारंबारता बंटन कहलाता है। ऐसे बारंबारता बंटन की पहले वर्ग की निम्न सीमा या अंतिम वर्ग की उच्चसीमा या दोनों निश्चित नहीं होती। इनके लिए "कम" या "से कम" तथा "अधिक" या "**से अधिक**" शब्दों का प्रयोग किया जाता है। निम्न सीमा, "से कम" लिखे जाने का अर्थ चर का मान $-\infty$ तक हो सकता है। इसी प्रकार उच्च सीमा, "से अधिक" लिखे जाने का अर्थ चार का मान $+\infty$ तक हो सकता है। इस प्रकार के बारंबारता बंटन का उदाहरण सारणी 3.10 में दिया गया है।

सारणी 3.10

विवृत-छोर वर्ग बारंबारता

वर्ग	बारंबारता
25 से कम	1
25 - 30	3
30 - 40	5
40 - 50	2
50 तथा अधिक	1
योग	12

सारणी 3.11

असमान वर्ग बारंबारता

वर्ग	बारंबारता
20 - 25	1
25 - 30	3
30 - 40	5
40 - 55	2
55 - 60	1
योग	12

2) असमान वर्ग अंतराल सहित बारंबारता बंटन

एक बारंबारता बंटन के सभी अंतराल समान होने अनिवार्य नहीं होते। असमान वर्ग अंतराल सहित बंटन सारणी 3.11 में दिया हुआ है। इसमें पहले दूसरे तथा पाँचवें वर्गों का अंतराल 5 है। जबकि तीसरे का 10 है तथा चौथे का 15 है। जैसा कि हम इकाई 4 में पढ़ेंगे इस प्रकार के बंटन में बहुलक प्रातिनिधिक मान नहीं होता, अत: यह परिभाषित नहीं होता।

3) संचयी बारंबारता बंटन

सारणी 3.6 में दिए हुए समंकों के संदर्भ में हम, मान लिया, निम्नलिखित प्रश्न पूछते हैं :

1) कितने परिवारों की मासिक आय रुपये 700 या इससे कम है?

2) कितने परिवारों की मासिक आय रुपये 600 या इससे अधिक है?

एक उपयुक्त संचयी बारंबारता बंटन तैयार करके उपरोक्त प्रश्नों के उत्तर सरलता से दिए जा सकते है। पहले प्रश्न के उत्तर के लिए हमें "से कम" संचयी बारंबारता बंटन तैयार करना होगा। तथा दूसरे प्रश्न के उत्तर के लिए हमें "से अधिक" बारंबारता बंटन की आवश्यकता होगी। ये बंटन क्रमश: सारणी 3.12 तथा 3.13 में दिए गए हैं।

सारणी 3.12

'से कम' संचयी बारंबारता बंटन

मासिक आय (रुपये)	बारंबारता		
	सरल		संचयी
550 से कम	5		5
600 से कम	6	5+6	11
650 से कम	10	5+6+10	21
700 से कम	12	5+6+10+12	33
750 से कम	9	5+6+10+12+9	42
800 से कम	5	5+6+10+12+9+5	47
850 से कम	3	5+6+10+12+9+5+3	50

सारणी 3.13

'से अधिक' संचयी बारंबारता बंटन

मासिक आय (रुपये)	बारंबारता		
	सरल		संचयी
500 से अधिक	5	3+5+9+12+10+6+5	50
550 से अधिक	6	3+5+9+12+10+6	45
600 से अधिक	10	3+5+9+12+10	39
650 से अधिक	12	3+5+9+12	29
700 से अधिक	9	3+5+9	17
750 से अधिक	5	3+5	8
800 से अधिक	3		3

4) सापेक्षिक बारंबारता बंटन

अभी तक हमने एक मान या वर्ग के घटित होने की संख्या को बारंबारता के रूप में अभिव्यक्त किया था। ये बारंबारताएँ कुल प्रेक्षणों की संख्या के भिन्न प्रतिशत के रूप में भी लिखी जा सकती हैं। जिनको सापेक्षिक बारंबारताएँ कहते है। सापेक्षिक बारंबारता बंटन की रचना सारणी 3.14 में प्रदर्शित की गई है।

सारणी 3.14

50 परिवारों की मासिक आय का सापेक्षिक बारंबारता बंटन

वर्ग	बारंबारता	सापेक्षिक बारंबारता	
		भिन्न के रूप में	प्रतिशत के रूप में
500 - 549	5	$5 \div 50 = 0.10$	$0.10 \times 100 = 10$
550 - 599	6	$6 \div 50 = 0.12$	$0.12 \times 100 = 12$
600 - 649	10	$10 \div 50 = 0.20$	$0.20 \times 100 = 20$
650 - 699	12	$12 \div 50 = 0.24$	$0.24 \times 100 = 24$
700 - 749	9	$9 \div 50 = 0.18$	$0.18 \times 100 = 18$
750 - 799	5	$5 \div 50 = 0.10$	$0.10 \times 100 = 10$
800 - 849	3	$3 \div 50 = 0.06$	$0.06 \times 100 = 06$
योग	50	1	100

उपरोक्त सारणी से यह स्पष्ट है कि सापेक्षिक बारंबारताओं का योग 1 या 100 (प्रतिशत के रूप में) होता है।

बोध प्रश्न 1

1) निम्नलिखित में, कम से कम दो विभेदीकरण के कारण देकर, भेद कीजिए।

   i) असंतत तथा संतत बारंबारता बंटन

   ii) सरल तथा संचयी बारंबारता बंटन

   iii) अपवर्जी तथा समावेशी वर्ग अंतराल

   iv) सरल तथा बारंबारता क्रम विन्यास

   ....

....

....

....

....

....

....

....

....

....

2) निम्नलिखित शब्दों की उदाहरण देकर व्याख्या कीजिए।

i) अवर्गीकृत संमक

ii) वर्ग संकेत

iii) विवृत छोर वर्ग

iv) वर्ग सीमाएँ

v) वर्ग परिसीमाएँ

vi) वर्ग बारंबारता

vii) मिलान दण्ड

viii) सापेक्षिक बारंबारता

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

3) एक महाविद्यालय के निम्न मध्य वर्ग के विद्यार्थियों के मासिक जेब खर्च का काल्पनिक बारंबारता बंटन बनाइए। इससे सापेक्षिक बारंबारता बंटन तैयार कीजिए।

....

....

....
....
....
....
....
....
....
....

4) एक बारंबारता बंटन तैयार करने के लिए निम्नलिखित के बारे में निर्णय करते समय किन बातों को ध्यान में रखना चाहिए?

i) वर्गों की संख्या तथा

ii) वर्ग अंतराल का आकार

....
....
....
....
....
....
....
....
....

5) निम्नलिखित बारंबारता बंटन द्वारा "से कम" तथा "से अधिक" संचयी बारंबारता तैयार कीजिए :

वर्ग	:	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70
बारंबारता	:	5	8	10	12	8	7

....
....
....
....
....
....
....
....
....
....
....
....

6) प्रश्न 5 में दिए गए समंकों के लिए एक सापेक्षिक बारंबारता बंटन तैयार कीजिए।

....

....

....

....

....

....

....

....

....

....

....

....

3.4 समंकों का सारणीयन

समंकों के संतोषजनक संकलन तथा विन्यास की भाँति इनका अच्छा प्रस्तुतिकरण भी महत्त्वपूर्ण होता है। वास्तव में, समंकों के संतोषजनक संकलन तथा विन्यास के बाद इनका अच्छा प्रस्तुतिकरण अनिवार्य होता है। लेकिन अच्छा प्रस्तुतिकरण अपने आप में कोई लक्ष्य नहीं होता। यह संतोषजनक विश्लेषण एवं निर्वचन के लिए अनिवार्य हो सकता है। समंकों का एक संतोषजनक प्रस्तुतिकरण कई प्रकार से सहायक होता है। पहला, यह समंकों में विद्यमान महत्त्वपूर्ण तथ्यों को केन्द्रीभूत करने में सहायक होता है। दूसरे, यह समंकों की तुलना में सहायक होता है। अन्त में, यह सांख्यिकीय सूचना की सहज समझ तथा बुद्धिमता से उपयोग करने में सहायक होता है।

सांख्यिकीय समंकों के प्रस्तुतिकरण को हम तीन शीर्षकों के अंतर्गत अध्ययन करेंगे।

i) सारणी द्वारा

ii) आलेख विधियाँ जिनमें रेखा आलेख, आयत चित्र, बारंबारता बहुभुज तथा वक्र, तथा संचयी बारंबारता वक्र सम्मिलित हैं।

iii) ज्यामितिक रूप, चित्र तथा सांख्यिकीय मानचित्र जिसमें वृत्तचित्र, दण्ड चित्र, क्षेत्रफल तथा परिमा चित्र आदि सम्मिलित होते हैं।

इस अनुभाग में हम सांख्यिकीय प्रस्तुतिकरण का अध्ययन करेंगे।

3.4.1 सारणियों का अर्थ तथा प्रकार

एक सांख्यिकीय सारणी, एक पूर्व निश्चित तथा सुस्पष्ट उद्देश्य सहित, सांख्यिकीय समंकों का स्तम्भों तथा पंक्तियों में सुव्यवस्थित विन्यास होती है। किसी सारणी में समंकों के क्षैतिज विन्यास को पंक्ति तथा उर्ध्वाधर विन्यास को स्तंभ कहा जाता है। सारणी में दी हुई सूचना की व्यवस्था के लिए इसकी पंक्तियों तथा स्तंभों को उपयुक्त स्थूणों तथा शीर्षकों या (उपशीर्षकों) द्वारा निर्दिष्ट किया जाता है। समंकों का सारणीयन सरल, नियोजित, सुस्पष्ट तथा तर्कसंगत होना चाहिए।

सारणी 3.15 में देश X का देश B को तीन वर्षों 1995, 1996 तथा 1997 में निर्यात तथा आयात के कल्पित समंक दिए हुए हैं।

सारणी 3.15

देश X का देश B को निर्यात तथा आयात (1995-1997) (करोड़ रुपये)

देश	1995		1996		1997*	
	आयात	निर्यात	आयात	निर्यात	आयात	निर्यात
A	60	70	65	75	70	65
B	50	60	60	65	65	60
C	40	30	40	40	42	50
D	45	42	60	55	63	55
योग	195	202	225	235	240	230

नोट : * शीघ्र अनुमानित संख्याएँ।

स्रोत : व्यापार पत्रिका, 1998, X देश का विदेश व्यापार मंत्रालय।

स्पष्टत: ऊपर दी गई सारणी का उद्देश्य देश X का शेष विश्व से आयात तथा निर्यात को दर्शाना है। सारणी की प्रत्येक प्रविष्टि एक पंक्ति तथा स्तम्भ से संबधित होती है। उदाहरणार्थ, दूसरी पंक्ति तथा चौथे स्तम्भ के प्रतिच्छेद पर दी हुई प्रविष्टि से ज्ञात होता है कि 1996 में देश X ने देश B से 60 करोड़ रुपये के मूल्य की वस्तुएँ तथा सेवाएँ आयातित की हैं। इस संख्या की तुलना आयात-निर्यात की अन्य संख्याओं से करके महत्त्वपूर्ण निर्वचन किए जा सकते हैं।

सारणियों के प्रकार

मूलत: सारणियाँ दो प्रकार की होती हैं :

1) सन्दर्भ सारणी या सामान्य उद्देश्य वाली सारणी
2) विषय सारणी या विशेष उद्देश्य वाली सारणी

1) *सन्दर्भ सारणियाँ* एक सामान्य उद्देश्य वाली सारणियाँ होती हैं। जिनका उद्देश्य विस्तृत सांख्यिकीय सूचना प्रदान करना होता है। इन सारणियों से हम अपनी आवश्यकतानुसार सूचना प्राप्त कर सकते हैं (अर्थात् द्वितीयक स्रोत)। विभिन्न सरकारी विभागों, मंत्रालयों, भारतीय रिजर्व बैंक, आर्थिक सर्वेक्षणों आदि द्वारा तैयार सारणियाँ सन्दर्भ सारणियाँ होती है। इसका अन्य महत्त्वपूर्ण उदाहरण भारतीय प्रधान पंजीयक द्वारा तैयार जनसंख्या संगणना की सारणियाँ हैं जिनमें भारत की जनसांख्यिकीय विशेषताओं से संबन्धित विस्तृत सूचना होती है। विद्यार्थियों को 'आर्थिक सर्वेक्षण', जो कि प्रतिवर्ष भारतीय संघ के बजट के साथ प्रकाशित होता है, का नवीनतम अंक पढ़ने की सलाह दी जाती है। इससे आप भारत के यू.एस.ए., यू.के. रूस, कनाड़ा तथा जर्मनी से, पिछले तीन या चार वर्षों, में आयात-निर्यात की सारणी तैयार कीजिए।

2) *विषय सारणियाँ* एक विशेष प्रकार की सारणियाँ होती है। इनका आकार छोटा होता है तथा यह सन्दर्भ सारणियों से बनाई जाती है। इनका उद्देश्य एक विशेष पहलू का विश्लेषण या एक विशेष प्रश्न का उत्तर प्रदान करना होता है। उदाहरण के लिए, हम जनसंख्या संगणना सारणियों में से मुम्बई तथा दिल्ली में रहने वाले लोगों की वह संख्याएँ प्राप्त कर सकते हैं जो विभिन्न मातृभाषाएँ बोलते हैं, जिनके विभिन्न धर्म हैं तथा जो भारत के विभिन्न राज्यों से आते है। इसी प्रकार भारतीय रिज़र्व बैंक के विभिन्न प्रकाशनों से हम सारणी के रूप में, पिछले दस वर्षों, में मुद्रा पूर्ति, ब्याज की दर तथा बैंक दर आदि की सूचना प्राप्त कर सकते है।

भाग 3.3 में दी गई सारणियों की भाँति, सारणियाँ सरल तथा एकधा हो सकती हैं जिनमें केवल एक चर, जैसे आय, होता है। विकल्पत: इसे एक विचर बारंबारता बंटन कहते हैं। इसके अतिरिक्त सारणियाँ जटिल एवं द्विधा, बहुधा आदि हो सकती हैं जिनमें दो या अधिक संबधित अभिलक्षणों को सम्मिलित किया जाता है।

3.4.2 सारणी के अंग

समंकों की प्रकृति तथा सारणीयन के उद्देश्यानुसार एक सारणी के अंग दूसरी सारणी के अंगों से भिन्न हो सकते हैं। फिर भी कुछ अंग समान होते है जैसे :

1) **सारणी संख्या** : सारणी की पहचान के लिए, विशेष रूप से जब किसी विश्लेषण में एक से अधिक सारणियाँ हों, इसकी संख्या की आवश्यकता होती है। यह संख्या सारणी के उपरी भाग पर लिखी जाती है।

2) **सारणी का शीर्षक** : सारणी के शीर्षक द्वारा सारणी में दी गई सूचना के बारे जानकारी प्राप्त होती है। यह सारणी संख्या के बाद लिखा जाता है। इसका उद्देश्य निम्नलिखित प्रकार के प्रश्नों का उत्तर देना होता है :

   i) सारणी में क्या है?
   ii) सारणी में कहाँ है?
   iii) एक विशेष सूचना कब घटित हुई?
   iv) एक विशेष सूचना का कैसे विन्यास किया गया है?

सारणी 3.15 में दी गई आयात-निर्यात सूचना के संदर्भ में उपरोक्त प्रश्नों के उत्तर निम्नलिखित है :

i) सारणी में देश X के आयात-निर्यात के मान दिये गए है।
ii) सारणी में दी गई सूचना चार देशों A, B, C तथा D से निर्यात तथा आयात को दर्शाती हैं।
iii) ये आयात तथा निर्यात 1995, 1996 तथा 1997 वर्षों में घटित हुए।
iv) आयात-निर्यात सूचना का वर्ष तथा देशों के अनुसार विन्यास किया गया हैं।

शीर्षक के बारे में क्या करें तथा क्या न करें

इसके लिए बड़े वाक्य का प्रयोग न करें। शीर्षक संक्षिप्त तथा उपयुक्त होना चाहिए। शीर्षक को मोटे शब्दों में लिखा जाना चाहिए। शीर्षक का एक से अधिक अर्थ व्यक्त नहीं होना चाहिए। 'सारणी में प्रस्तुत है ...' या 'समंको की विस्तृत तुलना ...' व्यजनों को लिखने से बचना चाहिए। शीर्षक एक टेलीग्राम की सूचना की भाँति होना चाहिए।

3) **शीर्ष टिप्पणी** : इसको प्रारंभिक टिप्पणी भी कहते हैं। यह शीर्षक के नीचे लिखी जाती है। यह विषय-वस्तु तथा मापन इकाई, जैसे रुपये या लाख टन या हजार गाँठ आदि को दर्शाती है। इसे कोष्ठकों में रखना चाहिए तथा सारणी के उपर दाईं ओर शीर्षक के नीचे लिखा जाना चाहिए। यहाँ यह बताना आवश्यक है कि सभी सारणियों में शीर्ष टिप्पणी की आवश्यकता नही होती।

4) **अनुशीर्षक** : इसका उपयोग पंक्तियों को निर्दिष्ट करने के लिए किया जाता है। यह सारणी के बाईं ओर के स्तम्भ में लिखे जाते है। अनुशीर्षक दो प्रकार के होते है।

   i) शीर्ष-अनुशीर्षक विभिन्न प्रकार की अनुशीर्षक प्रविष्टियों की प्रकृति के बारे में जानकारी देता है।
   ii) एक अनुशीर्षक पंक्ति में दी हुई प्रविष्टियों के बारे में जानकारी देता है।

5) **उपशीर्षक** : एक उपशीर्षक सारणी के स्तम्भ में दी हुई प्रविष्टियों के बारे में जानकारी देता है। एक उपशीर्षक के कई शीर्ष स्तम्भ हो सकते है। तथा प्रत्येक शीर्ष स्तम्भ भी कुछ उपशीर्षक स्तम्भों में विभाजित हो सकता है। उदाहरणार्थ एक महाविद्यालय के विद्यार्थियों को छात्रावासी तथा ग़ैर छात्रावासियों में वर्गीकृत किया जा सकता है तथा प्रत्येक वर्ग को स्त्री तथा पुरुषों में वर्गीकृत किया जा सकता है। इस प्रकार हमें प्रथम, द्वितीय तथा तृतीय वर्षों (मान लिया) में पुरुष छात्रावासियों आदि, प्रकार की जानकारी पाने में सहायता मिलती है।

6) **सारणी का मुख्य अंग** : इसे सारणी का क्षेत्र भी कहते हैं तथा यह इसका सबसे महत्त्वपूर्ण तथा बृहत अंग होता है। इसमें प्रांसगिक सूचना दी हुई होती है जिसके बारे में संकेत शीर्षक में अन्तर्विष्ट होता है। सारणी 3.15 के रूप में दिए हुए उदाहरण में शीर्षक द्वारा यह संकेत मिलता है कि सारणी में देश X की तीन वर्षों की आयात-निर्यात सम्बन्धी संख्यात्मक सूचना दी हुई है।

7) **पाद टिप्पणी** : यह एक प्रकार का स्पष्टीकरण होता है जो सारणी के नीचे लिखा जाता है। इसका उद्देश्य समंकों की सीमाओं या किसी विशेष चूक या त्रुटि के बारे में सावधान करना होता है। उदाहरणार्थ, सारणी 3.15 में पाद टिप्पणी यह बताती है कि दी हुई संख्याएँ अन्तिम नहीं है इसी प्रकार नवीनतम जनसंख्या संगणना में "जम्मू तथा कश्मीर रहित" जैसी की पाद टिप्पणी हो सकती है।

8) **समंकों का स्रोत** : यह सारणी का अंतिम लेकिन महत्त्वपूर्ण अंग होता है। इससे सारणी में दिए हुए समंकों की प्रामाणिकता की जानकारी मिलती है। इसके अतिरिक्त यदि पाठक चाहे तो उसे समंकों की जाँच तथा और अधिक समंक प्राप्त करने का अवसर मिलता है।

उपरोक्त बातों को ध्यान में रखते हुए एक काल्पनिक सारणी का आकार नीचे दिया गया हैं।

सारणी 3.16

(...................... **शीर्षक** ....)

शीर्ष टिप्पणी :
(करोड़ रुपयों में)

शीर्ष-अनुशीर्षक	उप-शीर्षक			
अनुशीर्षक	स्तम्भ शीर्षक I		स्तम्भ शीर्षक II	
	उपशीर्ष	उपशीर्ष	उपशीर्ष	उपशीर्ष
	सारणी	**का मुख्य**	**अंग**	
योग				

पाद टिप्पणी :
स्रोत :

3.4.3 सारणियों का महत्त्व

संख्यात्मक सूचना प्रस्तुति के अन्य रूपों की तुलना में सारणी के रूप में प्रस्तुति श्रेष्ठ होती है। प्रथम, सारणी में दिए हुए समंकों की समझ तथा निर्वचन सरल होती है। दूसरे, विभिन्न अभिलक्षणों की तुलना शीघ्र हो सकती है, उदाहरणार्थ, क्या तीनों वर्षों में आयात निर्यात से अधिक है? या क्या निर्यात में वृद्धि हो रही है? तीसरे इसके द्वारा और आगे अन्वेषण का अवसर प्राप्त होता है। चौथे, शाब्दिक कथन की तुलना में इसका मानव चित पर स्थायी प्रभाव होता है। यह कहने की

आवश्यकता नहीं कि मानवीय अन्वेषण के लगभग सभी क्षेत्रों में सांख्यिकीय सारणियों का बहुत उपयोग किया जाता है।

बोध प्रश्न 2

1) निम्नलिखित में भेद स्पष्ट कीजिए।

   i) उप-शीर्षक, शीर्ष-अनुशीर्षक तथा अनुशीर्षक
   ii) एकधा तथा द्विधा सारणियाँ
   iii) संदर्भ सारणी तथा विषय सारणी
   iv) स्तंभ प्रविष्टि तथा पंक्ति प्रविष्टि
   v) शीर्ष टिप्पणी तथा पाद टिप्पणी

....
....
....
....
....
....
....
....
....
....
....
....
....
....
....

2) इस कथन पर टिप्पणी कीजिए : "जो संबंध एक प्रस्ताव का उसके शीर्षक से होता है वही संबंध सारणी का उसके शीर्षक से होता है"।

....
....
....
....
....
....
....
....
....
....
....

3) सांख्यिकीय सारणी के विभिन्न अंगों के बारे में जानकारी दीजिए।

....

....

....

....

....

....

....

....

....

....

....

4) एक महाविद्यालय के विद्यार्थियों में निम्नलिखित सूचना को दर्शाने के लिए एक द्विधा सारणी तैयार कीजिए :

i) प्रथम, द्वितीय तथा तृतीय वर्ष के अनुसार वर्गीकरण
ii) छात्रावासी तथा ग़ैर-छात्रावासी वर्गीकरण
iii) स्त्री तथा पुरुष विद्यार्थी वर्गीकरण

इसके लिए कुछ काल्पनिक संख्याएँ स्वयं ले लीजिए।

....

....

....

....

....

....

....

....

....

....

....

....

....

....

3.5 समंकों का आलेखी प्रस्तुतिकरण

सारणियों के अतिरिक्त, सांख्यिकीय समंकों को विभिन्न प्रकार के आलेखों द्वारा भी प्रस्तुत किया जा सकता है। शीघ्र तथा संक्षिप्त सूचना को संप्रेषित करने के लिए आलेख बहुत उपयोगी होते हैं। उतनी ही आसानी तथा कार्य कुशलता सहित ये समय तथा स्थान के अनुसार समंकों की तुलना में सहायक होते हैं। यह चाक्षुष उपकरण (visual aids) होते है तथा व्यक्ति के चित्त पर शक्तिशाली प्रभाव

डालते हैं। प्राय: यह कहा जाता है कि "एक चित्र 1000 शब्दों के बराबर होता है"। यह समंकों द्वारा संप्रेषित तथ्यों की ओर पाठक का ध्यान आकर्षित करते हैं। इसके अतिरिक्त, इनके द्वारा किसी माप का अनुमान करने में सहायता मिल सकती है तथा हमारे उत्तरों की चित्रीय जाँच भी की जा सकती है।

समंकों का आलेखीय प्रस्तुतिकरण, विभिन्न प्रकार से चाहे जितना भी उपयोगी हो, समंकों की व्याख्या का केवल वैकल्पिक तरीका है। यह किसी प्रकार से अन्य प्रस्तुतिकरण के रूपों तथा सांख्यिकीय विश्लेषण की अतिरिक्त विधियों का प्रतिस्थापक नहीं है। आलेखी प्रस्तुतिकरण की कुछ विधियाँ निम्नलिखित हैं।

3.5.1 रेखा आलेख

एक समतल पर चार चतुर्थांश होते हैं लेकिन अर्थशास्त्र में हम प्राय: चित्र केवल प्रथम चतुर्थांश में ही बनाते हैं। जिसमें X-अक्ष तथा Y-अक्ष पर मापी गई दोनों मात्राएँ धनात्मक होती हैं। आर्थिक मात्राएँ जैसे कीमत, माँग तथा पूर्ति, राष्ट्रीय आय, उपभोग, उत्पादन, तथा इस प्रकार के अन्य चर अऋणात्मक (≥ 0) होते हैं।

हम एक माँग तालिका को आलेख पर अंकित करें। विभिन्न बिन्दुओं को मिलानेवाली परिणामी वक्र, संतत मानते हुए, एक रेखा आलेख कहलाता है। जो कीमत तथा माँगी गई मात्रा के बीच संबंध को व्यक्त करता है। अर्थशास्त्र में इस रेखा आलेख को माँग वक्र कहते हैं। यह ध्यान रहे कि कीमत को Y-अक्ष पर तथा मात्रा को X-अक्ष पर लिया जाता है। सारणी 3.17 में दिए हुए समंकों का माँग वक्र चित्र 3.1 में दिया गया है।

सारणी 3.17

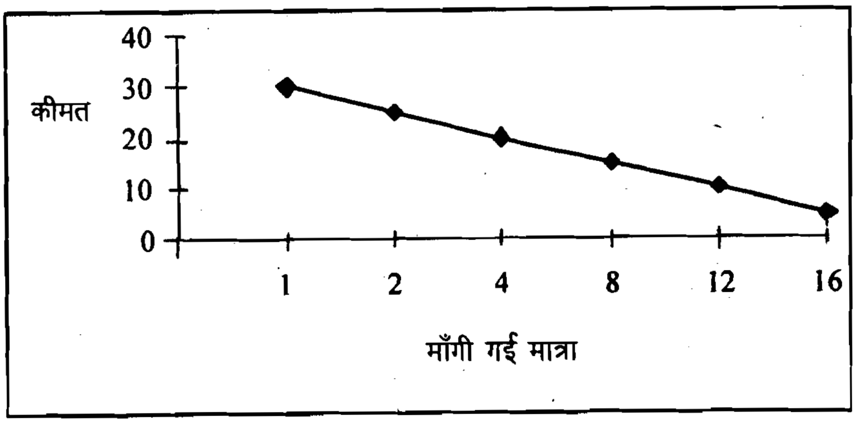
माँग सारणी

X की कीमत	X की माँग
5	16
10	12
15	8
20	4
25	2
30	1

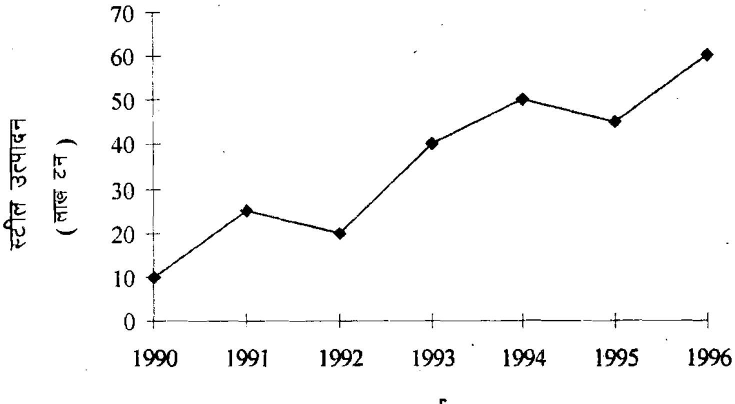
सारणी 3.18

काल श्रेणी संभक

वर्ष	स्टील का उत्पादन (लाख टन)
1990	10
1991	25
1992	20
1993	40
1994	50
1995	45
1996	60



चित्र 3.1



चित्र 3.2 : स्टील के उत्पादन का कालिक चित्र

रेखा आलेख का उपयोग किसी आर्थिक चर, जैसे समय के साथ स्टील के उत्पादन, में परिवर्तन को दर्शाने के लिए किया जा सकता है। अन्य शब्दों में, यदि दो में से एक चर समय (मास, वर्ष आदि) है तो आलेख को काल श्रेणी आलेख या कालिक चित्र कहते हैं। एक काल श्रेणी में एक आर्थिक चर का समय के साथ संबंध व्यक्त होता है। सारणी 3.18 में काल श्रेणी समंकों का उदाहरण दिया गया है। वर्षों को X-अक्ष तथा स्टील उत्पादन को Y-अक्ष पर लेते हुए काल श्रेणी आलेख चित्र 3.2 में दिया गया है।

3.5.2 आयत चित्र, बारंबारता बहुभुज तथा बारंबारता वक्र

आयत चित्र वर्गीकृत समंकों को दर्शाने वाला अति सामान्य आलेख होता है। यह साथ-साथ खड़ी हुई आयतों का समूह होता है जिसकी चौड़ाई वर्ग अंतराल के बराबर तथा ऊँचाई वर्ग की बारंबारता की आनुपाती होती है। इसकी विशेषताएँ निम्नलिखित हैं :

i) यह एक आयताकार चित्र होता है।

ii) क्योंकि आयतों की निश्चित चौड़ाई तथा ऊँचाई होती है, आयत चित्र एक द्वि-विमीय चित्र होता है। एक आयत की चौड़ाई वर्ग अंतराल के बराबर तथा

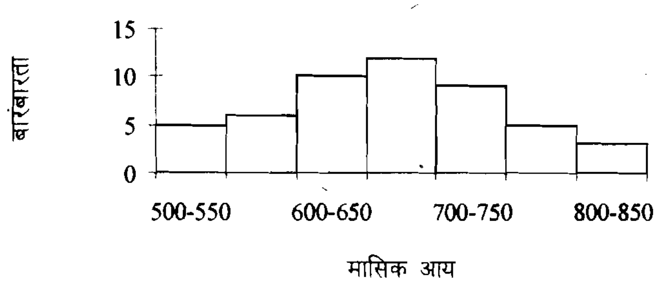
$$\text{ऊँचाई} = \frac{\text{वर्ग बारंबारता} \times \text{समंकों में न्यूनतम वर्ग अंतराल}}{\text{वर्ग अंतराल}}$$

iii) **प्रत्येक आयत का क्षेत्रफल उससे संबंधित बारंबारता का अनुपाति** होता है।

आयत चित्र की रचना

सारणी 3.6 में दिए गए बारंबारता बंटन को आरेख पत्र पर अंकित करने के लिए, हम X-अक्ष पर वर्ग अंतराल जैसे 500 – 550, 550 – 600 आदि अंकित कर देते हैं। इसी प्रकार हम Y-अक्ष पर बारंबारताओं को अंकित करते हैं। क्योंकि सभी वर्गों का अंतराल बराबर है। प्रत्येक आयत की ऊँचाई उससे संबंधित वर्ग की बारंबारता के बराबर ली जाएगी। यह आयत चित्र, 3.3 में दर्शाया गया है।

आयत चित्र



चित्र 3.3

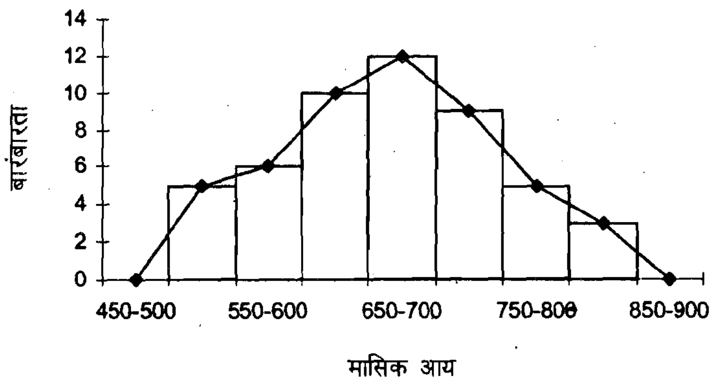
आयत चित्र के लाभ

1) विभिन्न आयतों की चौड़ाई बंटन में वर्गों की प्रकृति को दर्शाती है अर्थात् वर्ग समान अंतराल के है या नहीं।

2) एक आयत का क्षेत्रफल वर्ग की अनुपातिक बारंबारता को दर्शाता है।

बारंबारता बहुभुज

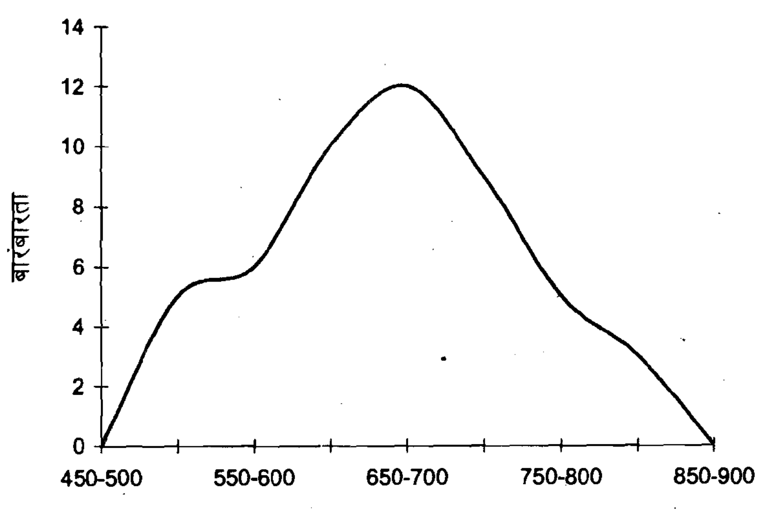
बारंबारता बहुभुज शब्द बहुभुज से लिया गया है। जिसका अर्थ बहुत भुजाओं वाला चित्र होता है। सांख्यिकी में इसका अर्थ बारंबारता बंटन का आलेख है। आयात चित्र की विभिन्न आयतों के शिखर के मध्य बिन्दुओं को सरल रेखाओं से जोड़ने पर हमें बारंबारता बहुभुज प्राप्त होता है, जैसा चित्र 3.4 में दिर्शाया गया है। इस चित्र में बहुभुज तथा आयत चित्र के अंतर्गत क्षेत्रफल बराबर रखने के लिए इच्छाधीन दो वर्ग, शुन्य बारंबारता सहित, बंटन के दोनों सिरों पर सम्मिलत किए जाते हैं। यह बारंबारता बहुभुज चित्र 3.4 में दर्शाया गया है।



चित्र 3.4 : बारंबारता बहुभुज

बारंबारता वक्र

यदि बहुभुज के लिए प्राप्त बिन्दुओं को एक निष्कोण वक्र द्वारा मिला दिया जाए तो एक बारंबारता वक्र प्राप्त होता है, जैसा चित्र 3.5 में दर्शाया गया है।



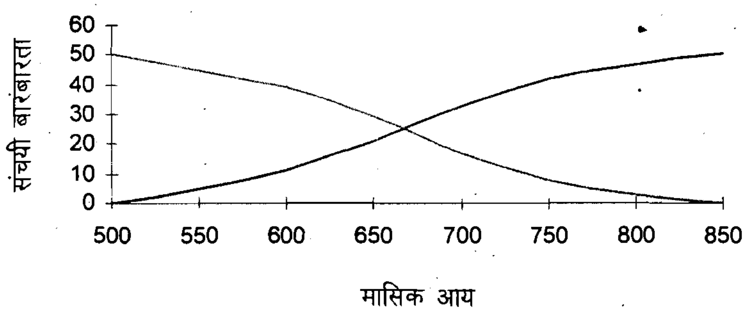
चित्र 3.5 : बारंबारता वक्र

3.5.3 संचयी बारंबारता वक्र-तोरण

एक संचयी बारंबारता वक्र के आलेख को संचयी बारंबारता वक्र या तोरण कहते हैं। जिस प्रकार एक संचयी बारंबारता वक्र 'से कम' या 'सें अधिक' हो सकते हैं, उसी प्रकार 'से कम' या 'सें अधिक' तोरण हो सकते हैं।

तोरणों का उपयोग विभिन्न विभाजन मूल्यों के आलेखी निर्धारण के लिए किया जाता है। हम दी हुई सीमाओं के मध्य, प्रेक्षणों का प्रतिशत भी ज्ञात कर सकते हैं। सारणी 3.12 तथा 3.13 में दिए हुए संचयी बारंबारता बंटन, चित्र 3.6 में दर्शाए गए हैं।

यह बात ध्यान रखने योग्य है कि 'से कम' तोरण के लिए हम एक वर्ग अंतराल '500 से कम', शून्य बारंबारता सहित और सम्मिलित कर लेते हैं। इसी प्रकार 'से अधिक' तोरण के लिए हम एक वर्ग अंतराल '900 से अधिक', शून्य बारंबारता सहित सम्मिलत कर लेते हैं।



चित्र 3.6 : 'से कम' तथा 'से अधिक' तोरण

3.6 समंकों का आरेखी प्रस्तुतिकरण

एक आलेख सांख्यिकीय समंकों का चाक्षुण प्रस्तुतिकरण होता है। आरेख से अर्थ दण्ड, वर्ग, वृत्त, मानचित्र, चित्र, मानारेख (cartogram) आदि से होता है। आरेख तथा आलेख में अंतर होता है। जबकि आरेख का उपयोग केवल प्रस्तुतिकरण होता है, आलेख का उपयोग प्रस्तुतिकरण के अतिरिक्त विश्लेषण के लिए भी हो सकता है।

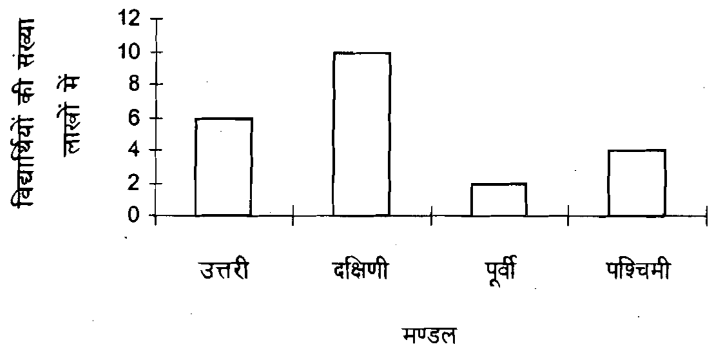
3.6.1 एक विमीय आरेख

इनको दण्ड आरेख भी कहते हैं। दण्ड एक मोटी रेखा को कहते हैं जो कि पाठक का ध्यान आकर्षित करने के लिए मोटी की जाती है। दण्ड की ऊँचाई चर के मान का प्रतीक होती है, जबकि चौड़ाई, अतः क्षेत्रफल, का कोई महत्त्व नहीं होता। यह आयत चित्र से भिन्न होता है, जिसमें आयत की चौड़ाई तथा ऊँचाई का महत्त्व होता है। इसके अतिरिक्त दण्ड आरेख के दण्डों के बीच में बराबर अंतर रखा जाता है। जबकि आयत चित्र की आयतें साथ-साथ, बिना अंतर, आरेखित की जाती हैं। अन्ततः, आयत चित्र में आयत सदैव खड़ी हुई होती हैं जबकि दण्ड आरेख में ये खड़ी या पड़ी हुई हो सकती हैं। एक दण्ड आरेख की रचना हम निम्नलिखित सरल उदाहरण द्वारा करेंगे।

सारणी 3.19 : एक देश के चार मण्डलों में विद्यार्थियों की संख्या

मण्डल	विद्यार्थियों की संख्या (लाखों में)
उत्तरी	6
दक्षिणी	10
पूर्वी	2
पश्चिमी	4

ऊपर दिए हुए समंकों का दण्ड आरेख चित्र 3.7 में दर्शाया गया है। दण्डों में विभिन्न प्रकार के रंग भर कर या छायाकरणों द्वारा सुन्दर बनाया जा सकता है। यह अन्वेषक की रुचि पर निर्भर करता है।



चित्र 3.7 : दण्ड आरेख

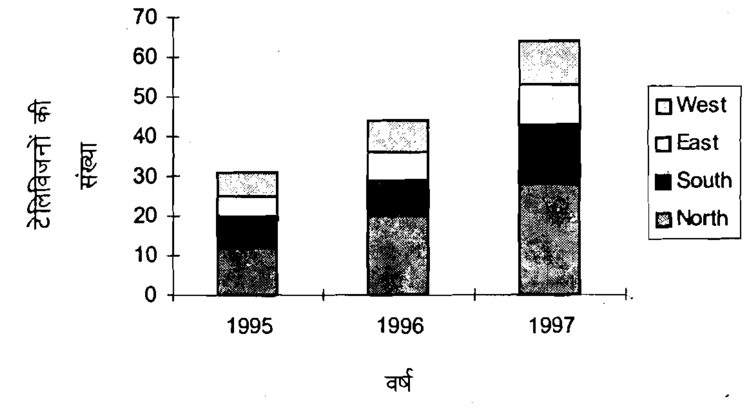
अन्तर्विभक्त या घटक दण्ड आरेख

जब एक घटना के विभिन्न घटकों के तुलनात्मक मानों की तुलना को प्रदर्शित करना हो तो अन्तर्विभक्त दण्ड आरेख का उपयोग किया जाता है। इस आरेख में, प्रत्येक घटना के दण्ड को विभिन्न अंगों में विभक्त किया जाता है। प्रत्येक अंग द्वारा अधिकृत दण्ड का हिस्सा, कुल योग में, इसके अंश को निर्दिष्ट करता है। विभिन्न दण्डों के प्रविभाजन एक ही क्रम में किए जाने आवश्यक हैं तथा इनको एक दूसरे से अलग दिखाने के लिए विभिन्न रंग या छाया का उपयोग किया जाना चाहिए। सारणी 3.20 में दिए हुए, टेलिविज़न के विक्रय संबंधि, कल्पित समंकों का एक अन्तर्विभक्त दण्ड आरेख चित्र 3.8 में दर्शाया गया है।

सारणी 3.20

टेलिविज़न का मण्डलानुसार विक्रय (1995-97)

मण्डल	बिक्री (लाखों में)		
	1995	1996	1997
उत्तरी	12	20	28
दक्षिणी	8	9	15
पूर्वी	5	7	10
पश्चिमी	6	8	11
योग	31	44	64



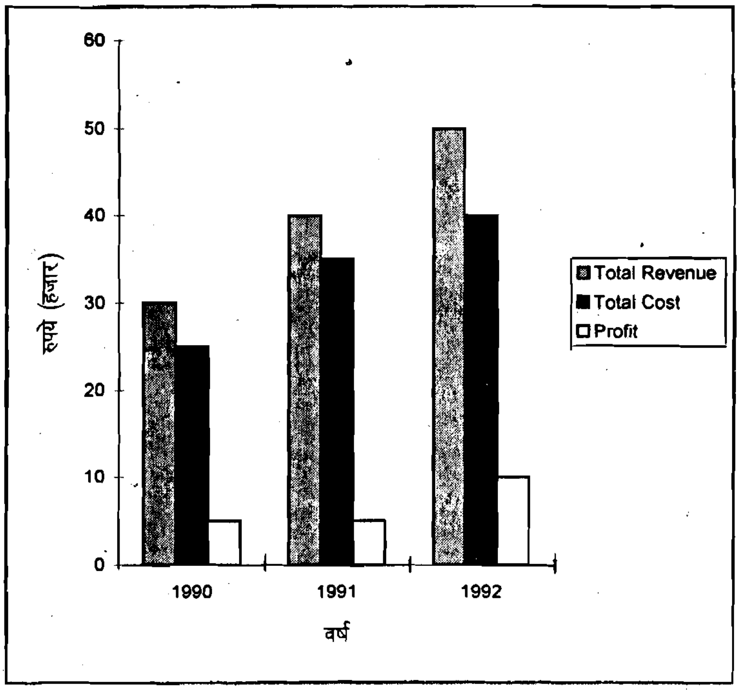
चित्र 3.8 : अन्तर्विभक्त या घटक दण्ड आरेख

जब दो या दो से अधिक समंक समुच्चयों की तुलना दर्शानी हो तो इस आरेख का उपयोग किया जाता है। किसी समय, स्थान या संबंधित घटनाओं के दण्डों के समुच्चय साथ-साथ (बिना किसी अन्तर के) आरेखित किए जाते हैं। विभिन्न दण्डों के अलग-अलग रंग या छाया द्वारा पहचाना जाता है। सारणी 3.21 में दिए गए कल्पित समंकों का एक बहुदण्ड आरेख चित्र 3.9 में दर्शाया गया है।

सारणी 3.21

मै० XYZ का कुल आगम, कुल लागत तथा लाभ (1990-92) (हजार रुपये)

वर्ष	कुल आगम	कुल लागत	लाभ
1990	30	25	5
1991	40	35	5
1992	50	40	10



चित्र 3.9 : बहुदण्ड आरेख

3.6.2 द्विविमीय आरेख या क्षेत्रफल आरेख

एक विमिय आरेख में दण्ड की केवल ऊँचाई का महत्त्व होता है, तथा इसकी चौड़ाई अन्वेषक की सुविधा या रुचि के अनुसार चयन की जाती सकती है। लेकिन द्विविमीय आरेखों में क्षेत्रफल का अधिक महत्त्व होता है। इसीलिए इनको क्षेत्रफल आरेख भी कहा जाता है। क्षेत्रफल आरेख तीन प्रकार के होते हैं :

i) आयत, जिसका क्षेत्रफल आयत की चौड़ाई गुणा लंबाई (या ऊँचाई) के बराबर होता है।

ii) वर्ग, जिसका क्षेत्रफल भुजा (आधार) के वर्ग के बराबर होता है।

iii) वृत्त, जिसका क्षेत्रफल πr^2 (जहाँ $\pi = 22 / 7$ तथा r = वर्ग की त्रिज्या) के बराबर होता है।

हम विश्वविद्यालय में शिक्षकों के तीन वर्गों के काल्पनिक समंकों से ये तीनों प्रकार के आलेख तैयार करेंगे।

सारणी 3.22

विश्वविद्यालय शिक्षकों का औसत वेतन (1.1.1998)

वर्ग	औसत वेतन (रुपयों में)
प्रोफेसर	25000
रीडर	16000
लेक्चरर	9000

i) आयत बनाने के लिए हम एक समान आधार, मान लिया 100 ले लेते हैं। तदनुसार, ऊँचाई ज्ञात करने के लिए :

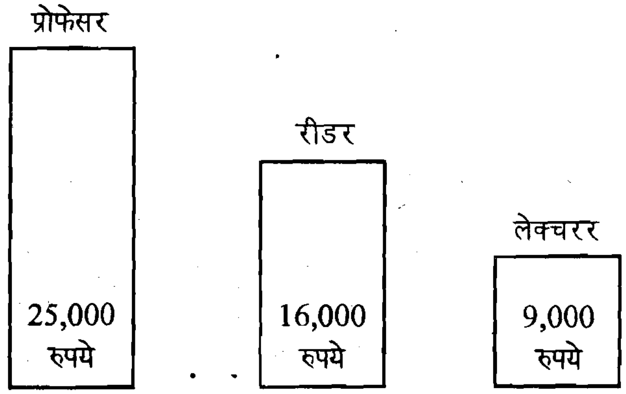
1) 25000 रुपये वेतन = 100 (आधार) × 250 (ऊँचाई)

2) 16000 रुपये वेतन = 100 (आधार) × 160 (ऊँचाई)

3) 9000 रुपये वेतन = 100 (आधार) × 90 (ऊँचाई)

हम अनुमाप 2 से.मी. = 100 ले लेते हैं। इस प्रकार पहली आयत का आयाम (dimension) 2 से.मी. × 5 से.मी., दूसरी आयत का आयाम 2 से.मी. × 3.2 से.मी. तथा तीसरी आयत का आयाम 2 से.मी. × 1.8 से.मी. होगा। इन आयतों को बनाने पर क्षेत्रफल आरेख प्राप्त होगा। (चित्र 3.10)

विश्वविद्यालय शिक्षकों का औसत वेतन (रुपयों में)



अनुमाप : 2 से.मी. = 100 से.मी.

चित्र 3.10 : क्षेत्रफल आरेख (आयत)

ii) वर्ग बनाने के लिए हम विभिन्न वेतनों का वर्गमूल ले लेते हैं।

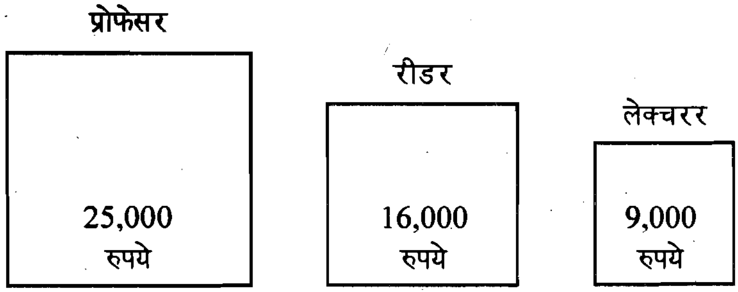
1) $\sqrt{25000} = 158.114$

2) $\sqrt{16000} = 126.491$

3) $\sqrt{9000} = 94.868$

हम अनुमाप 1 से.मी.= 50 ले लेते हैं। इस प्रकार पहले वर्ग की भुजा लगभग 3.2 (158.114/50) सेमी, दूसरे वर्ग की भुजा 2.53 से.मी. तथा तीसरे वर्ग की भुजा 1.9 से.मी. होगी। ये वर्ग चित्र 3.11 में दर्शाए गए हैं।

विश्वविद्यालय शिक्षकों का औसत वेतन (रुपयों में)



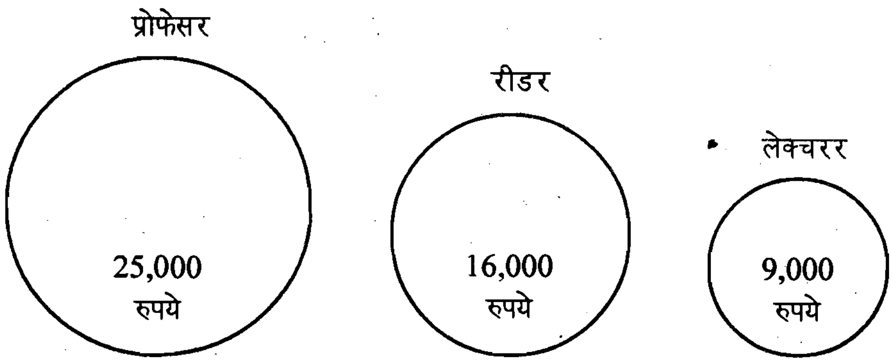
अनुमाप : 2 से.मी. = 50 से.मी.

चित्र 3.11 : क्षेत्रफल आरेख (वर्ग)

iii) वृत्त तैयार करने के लिए हम उनकी त्रिज्याओं के वर्ग उनके क्षेत्रफलों के आनुपातिक ले लेते हैं। अर्थात् 25000 : 16000 : 9000 या 25 : 16 : 9 । ऐसा करना इस बात पर आधारित है कि वृत्त का क्षेत्रफल उसकी त्रिज्या के वर्ग का आनुपातिक होता है। यदि तीन वृत्तों की त्रिज्या क्रमशः r_1, r_2 तथा r_3 है तो हम $r_1^2 : r_2^2 : r_3^2 = 25 : 16 : 9$ या $r_1 : r_2 : r_3 = 5 : 4 : 3$ लिख सकते हैं।

अनुमाप : 1 से.मी. = 2.5 इकाई लेने पर तीन वृत्तों की त्रिज्याएँ क्रमशः 2.0, 1.6 तथा 1.2 से. मी. होंगी। ये वृत्त, चित्र 3.12 में दर्शाए गए हैं।

विश्वविद्यालय शिक्षकों का औसत वेतन (रुपये में)



अनुमाप : 1 से.मी. = 2.5 इकाई

चित्र 3.12 : क्षेत्रफल आरेख (वृत्त)

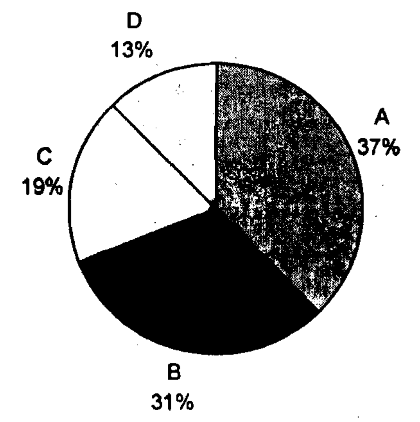
3.6.3 वृत्तारेख या वृत्त चित्र

इसको कोणीय आरेख भी कहते है। इसका उपयोग दिए हुए समंकों के प्रतिशत विघटनों को दर्शाने के लिए किया जाता है। उदाहरणार्थ, एक देश का विश्व के विभिन्न देशों को निर्यात आनुपातिक या प्रतिशत के रूप में व्यक्त किया जा सकता है। इन अनुपातों या प्रतिशतों को निम्नलिखित सूत्र की सहायता द्वारा कोणों में परिवर्तित किया जा सकता है।

$$\frac{\text{विघटन का अंश}}{\text{कुल}} \times 360^\circ$$

सारणी 3.23 : देश X का A, B, C तथा D देशों को 1990 में निर्यात

देश	निर्यात	प्रतिशत अंश	कोण की डिग्री
A	300	$(300 \times 100) \div 800 = 37.5$	135°
B	250	$(250 \times 100) \div 800 = 31.25$	112.5°
C	150	$(150 \times 100) \div 800 = 18.75$	67.5°
D	100	$(100 \times 100) \div 800 = 12.5$	45°
योग	800	100	360°



चित्र 3.13 : X के निर्यात का वृत्तारेख

वृत्तारेख की रचना के चरण

1) सभी विघटनों का योग कीजिए।
2) उपविघटन का कुल योग में अनुपात या प्रतिशत ज्ञात कीजिए। इसको 360° से गुणा करके उपविघटन का कोण (डिग्री में) ज्ञात कीजिए।
3) एक उपयुक्त आकार का वृत्त बनाइए।
4) वृत्त के केन्द्र पर विभिन्न कोण बनाइए। पहले सबसे बड़ा कोण बनाना सुविधाजनक होता है।
5) विभिन्न खण्डों को विभिन्न प्रकार के रंग या छाया से दर्शाइए।
6) विभिन्न खण्डों के प्रतिशत मानो को आरेख में लिखिए।

3.6.4 त्रिविमीय आरेख

ये आरेख अधिक लोकप्रिय नहीं है अत: इनका उपयोग बहुत कम किया जाता है। क्योंकि ये आरेख त्रिविमिय होते हैं (जिसमें लंबाई, चौड़ाई तथा ऊँचाई), इनसे आयतन प्राप्त होता है। इनका रूप बॉक्स, घन, खंड, गोला तथा बेलन हो सकता है। जब प्रेक्षणों में अंतर बहुत अधिक हो तो त्रिविमीय आरेख बहुत उपयोगी होते हैं। हम केवल घन के द्वारा समंकों की प्रस्तुति की व्याख्या करेंगे जिसके लिए हमें निम्नलिखित कार्य करने होते हैं।

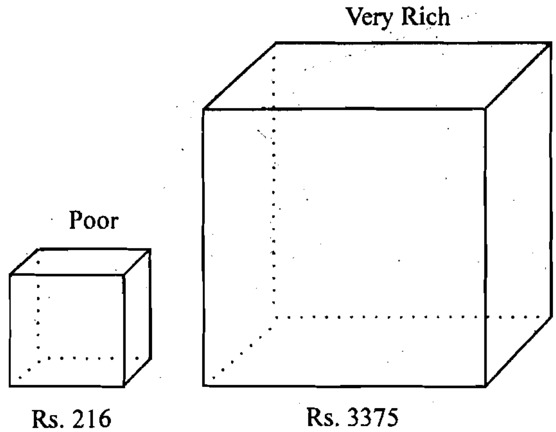
1) प्रत्येक संख्या का घनमूल ज्ञात कीजिए।
2) एक सुविधाजनक अनुमाप, अधिमानत: सेण्टीमीटरों में लीजिए।
3) घन बनाइए जिसकी विमा, निम्नलिखित में दो परिवारों, गरीब तथा बहुत धनी का उदाहरण लेकर, परिकलित की गई है।

सारणी 3.24

आय वर्ग	आय (रुपये)	घनमूल	घन की भुजा
1. गरीब	216	$\sqrt[3]{216}$	1.5 से.मी.
2. बहुत धनी	3375	$\sqrt[3]{3375}$	3.75 से.मी.

अनुमाप : 1 से.मी. = 4 इकाई

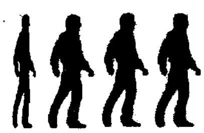
4) अब दो घन जिनकी भुजाएँ क्रमशः 1.5 तथा 3.75 से.मी. हों, बनाइए।



चित्र 3.14 : गरीब तथा बहुत धनी के आय स्तर (रुपये में)

3.6.5 चित्रलेख तथा सांख्यिकीय मानचित्र

इनको मानारेखा भी कहते हैं। अन्य आलेखी प्रस्तुति की तुलना में एक सामान्य जन के लिए चित्र अधिक आकर्षक होते हैं। लेकिन ये हर परिस्थिति में उपयुक्त नहीं होते। यह किसी राज्य की जनसंख्या संबंधी या किसी बड़े शहर, जैसे दिल्ली या मुम्बई, में वाहनों की संख्या संबंधी तथ्यों के लिए उपयोग हो सकता है। व्यक्ति का चित्र बनाकर जनसंख्या को दर्शाया जा सकता है। यहाँ पर भी अनुमाप का उपयोग किया जाता है। हम एक लाख व्यक्तियों को एक व्यक्ति के चित्र द्वारा निरूपित कर सकते हैं। इस प्रकार 3.5 लाख व्यक्तियों को 3.5 व्यक्ति चित्रों द्वारा दर्शाया जा सकता है। जैसा चित्र 3.15 में दिया गया है।



चित्र 3.15 : चित्रलेख तथा सांख्यिकीय मानचित्र

चित्र लेखों का मुख्य दोष यह है कि यह केवल समीपवर्ती मानों को ही निरूपित करते हैं। अधिक परिशुद्ध प्रस्तुति के लिए दण्ड आरेख अधिमान्य होते हैं।

बोध प्रश्न 3

1) निम्नलिखित में, कम से कम दो विभेदीकरण के कारण देकर, भेद कीजिए।

   i) आयत चित्र तथा कालिक चित्र
   ii) आयत चित्र तथा दण्ड आरेख
   iii) आयत चित्र तथा बारंबारता बहुभुज
   iv) 'से कम' तथा 'से अधिक' तोरण
   v) वृत्तारेख तथा वृत्त

....
....
....
....
....
....
....
....
....
....
....
....
....

2) निम्नलिखित समंकों से अन्तर्विभक्त दण्ड आरेख तथा वृत्तारेख तैयार कीजिए।

पुस्तकों पर व्यय

शैक्षिक वर्ष	अर्थशास्त्र	वाणिज्य शास्त्र	गणित	भाषाएँ	योग
1996-97	5200	10000	5000	4800	25000
1997-98	8000	14000	7000	6000	35000

....
....
....
....
....
....
....
....
....
....
....
....
....

3) निम्नलिखित पदों की व्याख्या कीजिए :

i) रेखा आरेख
ii) दण्ड आरेख
iii) अन्तर्विभक्त या घटक दण्ड आरेख
iv) बहुदण्ड आरेख
v) क्षेत्रफल आरेख
vi) आयतन आरेख

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

4) कोष्ठकों में दिए हुए शब्दों में से उपयुक्त शब्द द्वारा रिक्त स्थान भरिए।

i) वृत्तारेख को .... आरेख भी कहते हैं। (दण्ड, कोणिक, बहुदण्ड)

ii) जब दण्ड खड़े हुए हो तो चर को .... पर मापा जाता है। (X-अक्ष, समतल, Y-अक्ष)

iii) दण्ड आरेख, आयत, वर्ग, वृत्त तथा वृत्तारेख समंकों की प्रस्तुति का .... रूप है। (रेखागणित, अंकगणित, क्षैतिज)

iv) एक आयत चित्र की सभी आयतों के शिखर के मध्यबिन्दुओं को मिलाने पर हमें .... प्राप्त होता है।(एक तोरण, एक बारंबारता वक्र, एक बारंबारता बहुभुज)

v) 'से अधिक' संचयी बारंबारता बंटन के आलेख को 'से अधिक' .... भी कहतें है। (तोरण, बारंबारता बहुभुज, बारंबारता वक्र)

vi) एक सारणी का उपशीर्षक सारणी के .... में दिए हुए समंकों का नाम होता है। (पंक्ति, स्तम्भ, पाद टिप्पणी)

5) क्या निम्नलिखित कथन सत्य या असत्य हैं? यदि असत्य हैं तो सत्य कथन क्या होना चाहिए।

i) एक चित्र 1000 शब्दों के बराबर होता है।

ii) वर्ग तथा वृत्त का क्षेत्रफल आरेखों के उदाहरण है।

iii) एक चर वाले समंकों की प्रस्तुति के लिए केवल एक खड़ा दण्ड हो सकता है।

iv) साधारण बारंबारता बंटन का आलेख तोरण कहलाता है।

v) काल श्रेणी आलेख को कालिक चित्र कहलाता है।

vi) आयात चित्र तथा दण्ड चित्र समान होते है।

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

3.7 सारांश

संकलित समंक, संख्याओं का असंगठित तथा जटिल ढेर होता है। इनसे कुछ अर्थपूर्ण निष्कर्ष निकालने के लिए इनका सुव्यवस्थित विन्यास आवश्यक है। यह कई प्रकार से किया जा सकता है जैसे सरल तथा बारंबारता क्रम विन्यास द्वारा, असंतत तथा संतत बारंबारता बंटन द्वारा आदि।

कभी-कभी 'से कम' या 'से अधिक' संचयी बारंबारता बंटन को तैयार करना उपयोगी होता है। पहले को बनाने के लिए हम उपर से बारंबारताओं का आनुक्रमिक योग करते हैं तथा दूसरे के लिए हम नीचे से बारंबारताओं का आनुक्रमिक योग करते हैं।

समंकों के संकलन तथा संक्षेपण के पश्चात् उनका अच्छा प्रस्तुतिकरण महत्त्वपूर्ण होता है। एक अच्छा प्रस्तुतिकरण समंकों के तथ्यों को सामने लाता है परिणामस्वरूप उनका बुद्धिमता से उपयोग तथा तुलना करना संभव होता है। यह सारणी, रेखा आलेख, आयत चित्र, बारंबारता बहुभुज तथा बारंबारता वक्र; 'से कम' तथा 'से अधिक' तोरण; रेखागणित रूप – एक, दो तथा त्रिविमीय आरेख जैसे दण्ड आरेख, आयत, वर्ग, वृत्त, घन तथा वृत्तारेख; सांख्यिकीय मानचित्र आदि बनाकर किया जा सकता है। आरेखों का उपयोग करते समय उनकी सीमाओं का ध्यान रखना आवश्यक होता है। आरेख किसी समस्या के बारे में केवल अस्पष्ट बोध प्रदान करते हैं तथा केवल सीमित अभिलक्षणों को ही दर्शा सकते हैं। आलेखी प्रस्तुतिकरण के विपरीत आरेखी प्रस्तुतिकरण की मुख्य सीमा यह है कि इसको विश्लेषण के उपकरण के रूप में उपयोग नहीं किया जा सकता। आलेखी विधि का शुद्धता स्तर प्रायः गणितीय विधि के शुद्धता स्तर से निम्न होता है।

3.8 शब्दावली

समंकों का संक्षेपण : यह समंकों के असंगठित तथा जटिल ढेर का वर्गीकरण तथा विन्यास करने की प्रक्रिया है जिससे उनको तुलना तथा विश्लेषण के उपयुक्त बनाया जा सके।

क्रम विन्यास : यह समंकों का आरोही या अवरोही विन्यास होता है। इसको सरल क्रम विन्यास भी कहते हैं।

बारंबारता क्रम विन्यास : इस क्रम विन्यास को चर के विभिन्न संभावित मानों के समक्ष उनकी क्रमशः बारंबारताओं को लिखकर तैयार किया जाता है।

असंतत बारंबारता बंटन	:	जब चर केवल असंतत मान जैसे 1, 2, 3 ... लेता हो तो बंटन को असंतत बंटन कहते हैं। असंतत चर के उदाहरण, परिवार में बच्चों की संख्या, एक विश्वविद्यालय में विद्यार्थियों की संख्या आदि होते हैं।
सतत बारंबारता बंटन	:	एक संतत बारंबारता बंटन में चर का दो संख्याओं के बीच में कोई भी मान हो सकता है। जैसे ऊँचाई, वज़न आय, तापमान आदि।
समावेशी वर्ग अंतराल	:	ऐसा वर्ग अंतराल जिसमें वे सभी प्रेक्षण सम्मिलत होते हैं जो वर्ग सीमाओं के बराबर या बीच में होते हैं।
अपवर्जी वर्ग अंतराल	:	ऐसा वर्ग अंतराल जिसमें वे सभी प्रेक्षण सम्मिलत होते हैं जिनका मान निम्न सीमा के बराबर या अधिक तथा उच्च सीमा से कम होता है।
विवृत-छोर वर्ग	:	ऐसा वर्ग जिसकी एक सीमा निश्चित न हो।
बारंबारता बहुभुज	:	यह बारंबारता बंटन को चित्रित करने के लिए एक बहुभुज चित्र होती है जो आयत चित्र या बंटन से प्रत्यक्ष बनाई जा सकती है।
बारंबारता वक्र	:	यह एक बारंबारता बंटन का निष्कोणीय आलेख होता है जो कि बारंबारता बहुभुज से, स्वतंत्र हाथ अनुरेखन द्वारा, इस प्रकार प्राप्त किया जाता है जिससे दोनों के अन्तर्गत क्षेत्रफल लगभग बराबर रहे।
वर्ग तथा वर्ग सीमाएँ	:	यह एक निश्चित आकारों का समूह होता है जिसके दो सिरे, जो वर्ग सीमाएँ या वर्ग परिसीमाएँ कहलाती हैं, होते हैं।
वर्ग परिसर	:	इसको वर्ग अंतराल भी कहते हैं, तथा यह वर्ग की दो सीमाओं, उच्च तथा निम्न सीमा, का अंतर होता है। इसे वर्ग की चौड़ाई भी कहते हैं।
मध्य बिन्दु	:	इसको मध्यमान भी कहते हैं, तथा यह वर्ग की दो सीमाओं का माध्य होता है। इसका स्थान वर्ग के मध्य में होता है।
सापेक्षिक बारंबारता बंटन	:	एक बारंबारता बंटन जिसमें प्रत्येक मान की बारंबारता कुल प्रेक्षणों की संख्या का प्रतिशत होती है।
संचयी बारंबारता बंटन	:	एक असंतत या संतत बंटन की सरल बारंबारताओं के आनुक्रमिक योग द्वारा प्राप्त बंटन संचयी बारंबारता बंटन होता है। यह योग उपर से ('से कम' बारंबारता बंटन के लिए) या नीचे से ('से अधिक' बारंबारता बंटन के लिए) किया जा सकता है।
तोरण	:	यह संचयी बारंबारता बंटन का आलेख होता है। 'से कम' संचयी बारंबारताओं से 'से कम' तोरण तथा 'से अधिक' बारंबारताओं से 'से अधिक' तोरण प्राप्त होता है।
सारणीयन	:	यह समंकों का पंक्तियों तथा स्तंभों में सुव्यवस्थित विन्यास होता है।

उपशीर्षक	:	यह सारणी के स्तंभ में प्रस्तुत समंकों का लेबल होता है। इसको कक्ष शीर्ष भी कहते हैं। इसमें एक से अधिक स्तंभ शीर्ष हो सकते हैं।
अनुशीर्षक	:	यह भी सारणी का एक अंग होता है। इसमें शीर्ष अनुशीर्षक तथा अनुशीर्षक होते हैं। यह दोनों सारणी के बाईं ओर लिखे जाते हैं। इनके द्वारा पंक्ति शीर्षों की व्याख्या होती है।
सारणी का मुख्य अंग	:	निःसंदेह यह सारणी का महत्त्वपूर्ण अंग होता है तथा इसमें संख्यात्मक सूचना दी होती है जिसका संकेत सारणी के शीर्षक से प्राप्त होता है। यह सारणी का क्षेत्रफल भी कहलाता है।
रेखा आलेख	:	यह आलेख X तथा Y के निर्देशांकों से प्राप्त बिन्दुओं को सरल रेखा से मिलाने पर प्राप्त होता है।
कालिक चित्र	:	काल श्रेणी के रेखा आलेख को कालिक चित्र कहते हैं जैसे 1950 से स्टील का उत्पादन।
आयत चित्र	:	यह साथ-साथ खड़ी हुई आयतों का समूह होता है जिसमें आयतों का क्षेत्रफल बारंबारताओं का आनुपातिक होता है।
दण्ड आरेख	:	चर के विभिन्न मानों के अनुरूप खींची गई मोटी रेखाओं के समूह को दण्ड आरेख कहते हैं। यह आयत चित्र से भिन्न होता है जिसमें आयत की चौड़ाई का महत्त्व होता है।
सरल तथा अंतर्विभक्त दण्ड आरेख	:	सरल दण्ड आरेख द्वारा केवल एक चर को प्रस्तुत किया जा सकता है जबकि अंतर्विभक्त दण्ड आरेख का उपयोग एक घटना के विभिन्न घटकों को प्रस्तुत करने के लिए किया जाता है।
वृत्तारेख	:	यह एक वृत्त होता है जिसको उपभागों में बाँटकर किसी संख्या के अनुपातों को दर्शाया जाता है। इसे वृत्त चित्र भी कहते हैं।
क्षेत्रफल आरेख	:	इनको द्विविमीय आरेख भी कहते हैं। इनमें आरेख की लंबाई तथा चौड़ाई का महत्त्व होता है। इसीलिए इनको क्षेत्रफल आरेख भी कहते हैं। यह आयत या वर्ग या वृत्त हो सकता है।
आयतन आरेख	:	इनको त्रिविमीय आरेख भी कहते हैं। इनको तैयार करने में लंबाई, चौड़ाई तथा ऊँचाई का उपयोग होता है। यह बक्स, घन, खण्ड, गोला, तथा बेलन के आकार का होता है।
चित्रलेख	:	इसमें समंकों को चित्रों के उपयोग द्वारा प्रस्तुत किया जाता है।

3.9 कुछ उपयोगी पुस्तकें

Elhance, D. N. and V. Elhance, 1988, *Fundamentals of Statistics,* Kitab Mahal, Allahabad.

Mansfield, E., 1991, *Statistics for Business and Economics : Methods and Applications,* W. W. Norton and Co.

Nagar, A. L. and R. K. Dass, 1983, *Basic Statistics,* Oxford University Press, Delhi.

Yule, G. U. and M. G. Kendall, 1991, *An Introduction to the Theory of Statistics,* Universal Books, Delhi.

3.10 बोध प्रश्नों के उत्तर या संकेत

बोध प्रश्न 1

1) i) अनुभाग 3.2.2 तथा 3.2.3 देखिए।

   ii) अनुभाग 3.3.4 देखिए।

   iii) अनुभाग 3.3.3 देखिए।

   iv) अनुभाग 3.3.1 तथा 3.3.2 देखिए।

2) अपने आस-पास की घटनाओं पर आधारित उदाहरण दीजिए। शब्दों के उपयुक्त अर्थ के लिए भाग 3.3 देखिए।

4) अनुभाग 3.3.3 देखिए।

5) अनुभाग 3.3.4 (iii) देखिए।

6) अनुभाग 3.3.4 (iv) देखिए।

बोध प्रश्न 2

1) सारणी 3.16 तथा अनुभाग 3.4.2 देखिए।

2) अनुभाग 3.4.2 (2) देखिए।

3) सारणी 3.16 देखिए।

4) यह एक से अधिक प्रकार से किया जा सकता है। यहाँ एक दिया हुआ है। आप और बनाएँ।

X Y महाविद्यालय के विद्यार्थियों का वर्गीकरण

वर्ष	छात्रावासी		गैर छात्रावासी	
	पुरुष	स्त्री	पुरुष	स्त्री
प्रथम वर्ष				
द्वितीय वर्ष				
तृतीय वर्ष				

बोध प्रश्न 3

4) i) कोणिक

   ii) Y-अक्ष

इकाई 4 केन्द्रीय प्रवृत्ति के माप

इकाई की रूपरेखा

4.0 उद्देश्य

इस इकाई के अध्ययन के पश्चात् आपको निम्नलिखित जानकारी प्राप्त हो सकेगी :

- दिये हुए समंक समुच्चय द्वारा संख्यात्मक मात्राओं, जैसे समान्तर माध्य, माध्यिका, बहुलक, गुणोत्तर माध्य, हरात्मक माध्य, का परिकलन; और
- किस माप का उपयोग किस परिस्थिति में किया जाता है।

4.1 प्रस्तावना

इससे पहली इकाई में हमने अपरिष्कृत समंकों का कुछ वर्ग अन्तरालों में वर्गीकरण तथा सारणी या आलेख के रूप में प्रस्तुतिकरण द्वारा संक्षेपण का विवेचन किया था। सारणी या आरेख, प्रेक्षणों के बंटन का मोटे रूप में बोध कराते है। प्राय: हमें बंटनों की तुलना करनी पड़ती है। अत: सारणियों तथा आरेखों द्वारा इनकी तुलना कठिन होती है। यदि हम समंकों की व्याख्या के लिए एक संख्या ज्ञात कर सके तो विभिन्न बंटनों की तुलना करना बड़ा सुविधाजनक कार्य हो जाता है।

इस कार्य के लिए केन्द्रीय प्रवृत्ति का माप एक महत्त्वपूर्ण प्रतिदर्शज होता है। केन्द्रीय प्रवृत्ति के पाँच मुख्य माप हैं। यह सामान्तर माध्य, गुणोत्तर माध्य, हरात्मक माध्य, माध्यिका तथा बहुलक हैं। निम्नलिखित में आप इन सभी के बारे में जानकारी प्राप्त करेंगे।

4.2 केन्द्रीय प्रवृत्ति के माप

इकाई 3 में विवेचन किये गये बारंबारता बंटन में हमने यह पाया कि एक अन्वेषण के प्रेक्षणों की एक केन्द्रीय मान के आसपास समूह बनाने की प्रवृत्ति होती है। इस प्रक्रिया को केन्द्रीय प्रवृत्ति कहते है। हमारी रुचि इस प्रकार के केन्द्रीय मान को ज्ञात करने में है। एक बारंबारता बंटन की केन्द्रीय प्रवृत्ति के कई माप हो सकते हैं तथा मापों द्वारा हमें ऐसी संख्याएँ प्राप्त होती हैं जो कि बारंबारता बंटन का संक्षेपण करती हैं।

4.2.1 समांतर माध्य

सबसे प्रचलित केन्द्रीय प्रवृत्ति के माप को औसत या समांतर माध्य या केवल माध्य (जब अस्पष्टता की संभावना न हो) कहते हैं। इसको ज्ञात करने के लिए प्रतिदर्श के सभी मानों को जोड़कर इसे प्रेक्षणों की संख्या से विभाजित किया जाता है। समांतर माध्य का संकेतन चर के संकेत के उपर रेखिका (Bar) लगाकर किया जाता है। इस प्रकार $\overline{X}$ का उपयोग प्रतिदर्श में X के मानों के माध्य के लिए किया जाता है। यदि प्रतिदर्श में X के एक विशेष मान X_i की बारंबारता f_i है तो इसका X के मानों के कुल योग में योगदान f_iX_i के बराबर होता है। इस प्रकार हम X के मानों का समांतर माध्य निम्नलिखित सूत्र द्वारा ज्ञात कर सकते हैं।

$$\overline{X} = \frac{1}{N}(f_1X_1 + f_2X_2 + \cdots + f_nX_n) = \frac{\sum_{i=1}^{n} f_iX_i}{N}, \quad \text{जहाँ पर } N = \sum_{i=1}^{n} f_i$$

जब प्रेक्षण वर्ग अंतरालों में वर्गीकृत हों, जैसा कि संतत चर के लिए होता है, तो एक वर्ग में आने वाले व्यष्टि प्रेक्षणों की पृथक् रूप में पहचान नहीं की जा सकती, अत: इस वर्ग के व्यष्टि प्रेक्षणों का कुल योग में योगदान परिकलित नहीं किया जा सकता है। इस कठिनाई के समाधान के लिए यह मानलिया जाता है कि एक वर्ग में प्रत्येक प्रेक्षण का मान इस वर्ग के मध्यबिन्दु के बराबर है। इस प्रक्रिया द्वारा परिकलित माध्य, वास्तविक माध्य जो कि अपरिष्कृत समंकों से परिकलित किया गया हो, से भिन्न होता है। इसके कारण हमें वर्गीकरण संशुद्धि की आवश्यकता हो सकती है।

उदाहरण 1

सारणी 4.1 में दिये हुए असंतत बारंबारता बंटन का माध्य परिकलित कीजिए।

सारणी 4.1 : 100 गृहों का, आकार के अनुसार, बारंबारता बंटन

गृह का आकार (X_i)	बारंबारता (f_i)
1	3
2	16
3	25
4	33
5	12
6	7
7	2
8	2
योग	100

$$\overline{X} = \frac{\sum_{i=1}^{n} f_i X_i}{N} = \frac{1\times3+2\times16+3\times25+4\times33+5\times12+6\times7+7\times2+8\times2}{100} = \frac{374}{100} = 3.74$$

अत: 100 गृहों पर आधारित औसत गृह आकार = 3.74 है।

उदाहरण 2

सारणी 4.2 में दिये हुए वर्ग बारंबारता बंटन का माध्य परिकलित कीजिये।

सारणी 4.2 : 100 गृहों का, खाद्य सामग्री पर औसत मासिक व्यय के अनुसार बारंबारता बंटन

व्यय वर्ग (रुपये)	बारंबारता
262.5 - 286.5	1
286.5 - 310.5	14
310.5 - 334.5	16
334.5 - 358.5	28
358.5 - 382.5	26
382.5 - 406.5	15
योग	100

माध्य के परिकलन के लिए हम निम्नलिखित सारणी तैयार करते हैं।

वर्ग अंतराल (रुपये) (0)	मध्य बिन्दु (X_i) (1)	बारंबारता (f_i) (2)	$f_i X_i$ (3)
262.5 - 286.5	274.5	1	274.5
286.5 - 310.5	298.5	14	4179.0
310.5 - 334.5	322.5	16	5160.0
334.5 - 358.5	346.5	28	9702.0
358.5 - 382.5	370.5	26	9633.0
382.5 - 406.5	394.5	15	5917.5
योग		100	34866.0

अत: गृहों का खाद्य सामग्री पर औसत मासिक व्यय $\overline{X} = \frac{34866}{100} = 348.66$ रुपये है।

इस उदाहरण द्वारा यह पता चलता है कि स्तंभ (3) को प्राप्त करने के लिए हमें स्तंभ (1) तथा (2) के संगत मानों को गुणा करना होता है तथा बहुधा, प्रत्येक गुणन के लिए, ये परिकलन कठिन होते हैं। इन परिकलनों को निम्नलिखित रूपांतरण द्वारा सरल बनाया जा सकता है।

$i = 1,2, \text{-----} n$ के लिए हम

$u_i = \frac{X_i - A}{h}$ या $X_i = A + hu_i$ लिखते हैं। अत: $\overline{X} = A + h\overline{u}$.

यहाँ पर A को कल्पित माध्य $h\bar{u}$ तथा $\bar{X}$ को प्राप्त करने के लिए इसका संशोधन पद कहते हैं। A तथा h का ऐसा चयन किया जाता है कि जिससे $\bar{u}$ का परिकलन सरल हो जाय। प्राय: A को X के उस मान के बराबर लिया जाता है जिसकी बारंबारता अधिकतम हो। जब स्तंभ (1) में X के आनुक्रमिक मान समान दूरी पर हों, तो h का मान X के दो आनुक्रमिक मानों के अंतर के बराबर लिया जाता है। यदि वर्गों के अन्तराल समान हैं तो दो आनुक्रमिक मध्य बिन्दुओं का अंतर प्रत्येक वर्ग अंतराल के बराबर होता है।

इस विधि की व्याख्या के लिए हम सारणी 4.2 में दिये हुए गृहों के खाद्य सामग्री पर औसत मासिक व्यय के समंकों का माध्य पुन: परिकलित करते हैं। A तथा h के उपयोग द्वारा हम सारणी 4.3 तैयार करते हैं, जहाँ पर

A = अधिकतम बारंबारता वाले वर्ग का मध्यबिन्दु = 346.5 तथा

h = सभी वर्गों में विद्यमान समान अंतराल = 24

अत: $u_i = \dfrac{X_i - 346.5}{24}$

सारणी 4.3 : सारणी 4.2 के बारंबारता बंटन के माध्य का परिकलन

वर्ग अंतराल (रु०)	मध्य बिन्दु (X_i)	$u = \dfrac{X - 346.5}{24}$	बारंबारता (f)	fu
262.5 - 286.5	274.5	– 3	1	– 3
286.5 - 310.5	298.5	– 2	14	– 28
310.5 - 334.5	322.5	– 1	16	– 16
334.5 - 358.5	346.5	0	28	0
358.5 - 382.5	370.5	1	26	26
382.5 - 406.5	394.5	2	15	30
योग			100	9

सारणी द्वारा हम $\bar{X} = A + h \times \dfrac{\sum_{i=1}^{n} fu}{N} = 346.5 + 24 \times \dfrac{9}{100} = Rs.\ 348.66$ ज्ञात कर सकते हैं।

समांतर माध्य की विशेषताएँ

1) *एक दिये हुए प्रेक्षण समुच्चय में प्रेक्षणों के समांतर माध्य से लिए गए विचलनों का बीजगणितीय योग शून्य होता है।*

मान लिया $X_1, X_2, \ldots\ldots X_n$ प्रेक्षणों की संख्या n है जिनकी बारंबारताएँ क्रमश: $f_1, f_2, \ldots\ldots f_n$ हैं। गणितीय दृष्टिकोण से इस विशेषता का अर्थ है कि $\sum_{i=1}^{n} f_i(X_i - \bar{X}) = 0$ जहाँ पर $X_i - \bar{X}$, i वें प्रेक्षण का माध्य से विचलन है।

इस विशेषता को हम निम्नलिखित विधि से सिद्ध कर सकते हैं।

$$\sum_{i=1}^{n} f_i(X_i - \overline{X}) = \sum_{i=1}^{n} f_i X_i - \overline{X}\sum_{i=1}^{n} f_i = \sum_{i=1}^{n} f_i X_i - N.\overline{X} = 0$$, अत: परिणाम।

2) *यदि एक प्रेक्षण समुच्चय में प्रेक्षणों के विचलन समांतर माध्य से लिए गए हों तो विचलन वर्गों का योग न्यूनतम होता है।*

गणितीय दृष्टिकोण से इस विशेषता का अर्थ यह है कि $S = \sum_{i=1}^{n} f_i(X_i - A)^2$ तब न्यूनतम होगा जब $A = \overline{X}$ हो, यहाँ पर A एक मनमाना माध्य है।

इस विशेषता का प्रमाण इस संकेत पर आधारित है कि S का आकार A के मान पर निर्भर करता है। अर्थात् चर S को चर A का फलन कहा जा सकता है। हमारी रुचि A के ऐसे मान को ज्ञात करना है जिसके लिए S का मान न्यूनतम हो। कलन के उपयोग द्वारा यह मान समीकरण $\frac{dS}{dA} = 0$ द्वारा दिया गया ऐसा मान है कि $\frac{d^2S}{dA^2} > 0$ रहे।

(याद रहे कि फलन के न्यूनतम मान के लिए प्रथम अवकलन शून्य तथा द्वितीय अवकलन धनात्मक होता है।)

A के सापेक्ष S का अवकलन लेकर उसको शून्य के बराबर लिखने पर, हमें निम्नलिखित प्राप्त होता है।

$$\frac{dS}{dA} = -2\sum_{i=1}^{n} f_i(X_i - A) = 0$$

$$\Rightarrow \sum_{i=1}^{n} f_i X_i - A\sum_{i=1}^{n} f_i = 0$$

या $$A = \frac{\sum_{i=1}^{n} f_i X_i}{\sum_{i=1}^{n} f_i} = \overline{X}$$

हम यह भी दर्शा सकते है कि जब $A = \overline{X}$ हो तो $\frac{d^2S}{dA^2} > 0$ होगा।

4.2.2 माध्यिका

माध्यिका से अर्थ उस केन्द्र बिन्दु से होता है जो एक बंटन को दो बराबर भागों में बाँटता है अर्थात् यह प्रेक्षणों के समुच्चय में मध्यवर्ती मान होता है। पहले हम असंतत चर के उदाहरण द्वारा इसकी व्याख्या करेंगे। मान लीजिए हमारे पास 5 पृथक् प्रेक्षण 2, 4, 9, 12, 19 हैं जो वर्धमान क्रम में व्यवस्थित हैं। यहाँ पर मध्यवर्ती मान 9 है क्योंकि बराबर संख्या में प्रेक्षण इससे कम तथा इससे अधिक हैं। अत: 2, 4, 9, 12, 19 की माध्यिका 9 है। हम एक और 6 पृथक् समंकों के समुच्चय पर विचार करते हैं: 3, 8, 15, 25, 35, 43; यहाँ पर किसी मान जो कि 15 तथा 25 के बीच हो, के लिए समान संख्या में प्रेक्षण इससे कम तथा अधिक हैं। इसलिए 15 तथा 25 के बीच का कोई मान माध्यिका के रूप में उपयोग किया जा सकता है। अद्वितीय (unique) माध्यिका को परिभाषित करने के लिए 15 तथा 25 का मध्यमान लेने की परंपरा है। अत: 3, 8, 15, 25 ,35, 43 की माध्यिका 20 होगी।

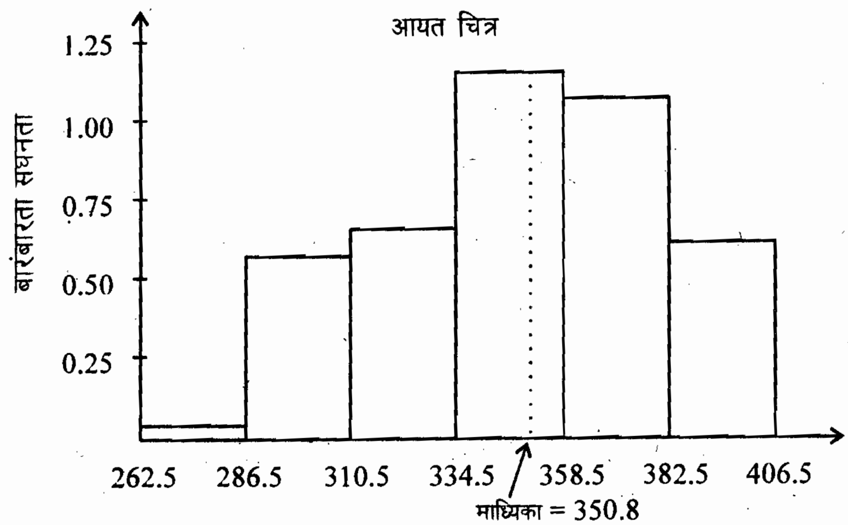
प्राय: व्यावहारिक परिस्थितियों में समंकों के समुच्चय में अपृथक् (non-distinct) प्रेक्षण होते हैं, जिनके कारण माध्यिका ज्ञात करने में कठिनाई हो सकती है। जैसा कि 5 प्रेक्षणों के समुच्चय : 2, 9, 9, 12, 19 ; से स्पष्ट है, इन परिस्थितियों में ऐसा मध्यवर्ती मान या केन्द्र बिन्दु ज्ञात करना, जो बंटन को दो बराबर भागों में बाँट दें, हमेशा संभव नहीं होता, अत: माध्यिका कि विधिवत परिभाषा में इन कठिनाइयों का ध्यान रखा जाना चाहिए।

बंटन की माध्यिका वह बिन्दु या केन्द्रीय मान होता है जिस तक (यह मान या इससे न्यून मान) प्रेक्षणों की संख्या कम से कम 50 प्रतिशत हो जिस तक जिससे अधिक (यह मान तथा इससे अधिक मान) प्रेक्षणों की संख्या कम से कम 50 प्रतिशत हो।

इस परिभाषा तथा अंतराल, जिसमें प्रत्येक मान माध्यिका होता है, के मध्य बिन्दु की परंपरा के आधार पर बंटन की माध्यिका को हमेशा अद्वितीय रूप में परिभाषित किया जा सकता है। अत: 2, 9, 9, 12, 19 प्रेक्षणों की माध्यिका 9 है, क्योंकि 5 में से 3 प्रेक्षणों (60%) में अधिकतम मान 9 है तथा 5 में से 4 प्रेक्षणों (80%) में न्यूनतम मान 9 है।

सारणी 4.1 में दिए गए बारंबारता बंटन की माध्यिका ज्ञात करने के लिए हमें यह पता चलता है कि 77% गृहों का आकार 4 या इसे कम है तथा 56% गृहों का आकार 4 या इससे अधिक है। अत: बंटन की माध्यिका 4 है।

संतत चर के वर्ग-बारंबारता बंटन की माध्यिका को सहचारी आयात चित्र, जिसमें आयत की ऊँचाई वर्ग की बारंबारता सघनता के बराबर होती है, द्वारा सरलता से समझा जा सकता है। ऐसे आयात चित्र में एक आयात का क्षेत्रफल उसके संगत वर्ग बारंबारता के बराबर होता है। इस परिस्थिति में माध्यिका, किसी एक वर्ग में वह बिन्दु होती है जिसके बाईं तथा दाईं ओर के क्षेत्रफल प्रत्येक 50% होते हैं। सर्वप्रथम हम उस वर्ग (जिसको माध्यिका वर्ग कहते हैं) का पता करते हैं जिसकी दाहिनी परिसीमा तक कुल क्षेत्रफल कम से कम 50% है। इसके बाद, माध्यिका के परिकलन के लिए, हम इस वर्ग की न्यून परिसीमा में वर्ग की वह लंबाई जोड़ देते हैं, जो कि 50% क्षेत्रफल के लिए तुलनात्मक बारंबारता के अनुपात में होती है। माध्यिका वर्ग को संचयी बारंबारता के परिकलन द्वारा, यह पहचान करके की N/2 वाँ प्रेक्षण किस वर्ग में है, सरलता से ज्ञात किया जा सकता है। इस विधि की व्याख्या सारणी 4.2 में दिये गये गृहों के औसत मासिक व्यय के समंकों की माध्यिका के परिकलन द्वारा की गई है।



चित्र 4.1 : मासिक गृह व्यय (रुपये)

वर्ग परिसीमा 334.5 तक क्षेत्रफल 31 है तथा 358.5 तक 59 है। अतः माध्यिका वर्ग 334.5 – 358.5 है अर्थात् माध्यिका इस वर्ग में स्थित है। अब हम इस वर्ग में ऐसा बिन्दु ज्ञात करना चाहते हैं कि 334.5 से उस बिन्दु तक क्षेत्रफल (50 – 31) = 19 हो। यहाँ पर यह ध्यान रहे कि 334.5 तक क्षेत्रफल 31 है। चूँकि वर्ग 334.5 – 358.5 पर आयत का क्षेत्रफल 28 तथा इस वर्ग का अंतराल 24 है, हमें 19 इकाई क्षेत्रफल के लिए 24 का 19/28वाँ हिस्सा चाहिए, जो कि 19/28 × 24 = 16.3 होगा। अतः माध्यिका 334.5 + 16.3 = 350.8 होगी। यहाँ पर यह ध्यान दें कि वर्ग 350.8 – 358.5 का क्षेत्रफल (28 – 19) = 9 इकाई है तथा 350.8 के दाईं ओर कुल क्षेत्रफल 9 + 26 + 15 = 50 इकाई है, जैसा कि होना चाहिए।

उपरोक्त विधि के आधार पर, हम माध्यिका के परिकलन के लिए निम्नलिखित सूत्र लिख सकते हैं।

$$M_d = l_m + \frac{\frac{N}{2} - C}{f_m} \times h,$$ जहाँ पर

l_m माध्यिका वर्ग अर्थात् वह वर्ग जिसमें माध्यिका है, की न्यून सीमा,

N कुल बारंबारता,

C माध्यिका वर्ग से पहले आनेवाले वर्गों की संचयी बारंबारता (उपरोक्त उदाहरण में $C = 31$),

f_m माध्यिका वर्ग की बारंबारता, तथा

h माध्यिका वर्ग का अंतराल है।

4.2.3 बहुलक

जैसा कि पहले बताया जा चुका है, प्रायः प्रेक्षणों में एक केन्द्रीय मान के आस-पास समूह बनाने की प्रवृत्ति होती है। इस प्रवृत्ति के एक सरल माप को बहुलक कहते हैं।

एक असंतत चर के लिए बहुलक या बहुलकता मान से अर्थ चर के उस मान से होता है, जिसकी बारंबारता अधिकतम होती है। बहुलक से अर्थ बहुमत नहीं होता अर्थात् इसका अर्थ यह नहीं होता कि अधिकतर (50% से अधिक) प्रेक्षणों के मान बहुलकता मान के बराबर हैं।

सारणी 4.1 द्वारा हमें ज्ञात होता है कि गृह के आकार का बहुलक या बहुलकता मान 4 है, क्योंकि इसकी बारंबारता अधिकतम है।

हमारे पास समंकों के ऐसे समुच्चय हो सकते हैं, जिनके लिए अद्वितीय बहुलक की परिभाषा नहीं की जा सकती, अर्थात् बंटन के कई बहुलक हैं। अपरिष्कृत समंक जिनमें 7 काल्पनिक प्रेक्षणों के मान 4, 3, 4, 1, 2, 5, 3 हैं, के दो बहुलक 3 तथा 4 हैं। दो बहुलक वाले बंटन को द्विबहुलक (bimodal) बंटन कहा जाता है। प्रायः बंटनों का एक ही बहुलक होता है या ये एक बहुलकी होते हैं।

जब सतत चर के प्रेक्षण, जैसे खाद्य सामग्री पर गृहों का व्यय, अपरिष्कृत रूप में हों तो किन्हीं दो प्रेक्षणों के मान बराबर होने की कोई संभावना नहीं होती तथा इस परिस्थिती में बहुलक माप का कोई अर्थ नहीं होता। लेकिन जब इन अपरिष्कृत समंकों को वर्गों में परिवर्तित किया जाता है तो समंकों की केन्द्रीय प्रवृत्ति सामने आ जाती है। वर्गीकृत समंकों के लिए अधिकतम बारंबारता वाले वर्ग को बहुलक वर्ग कहा जाता है। चूँकि बड़े अंतराल वाले वर्ग में प्रेक्षणों की संख्या, छोटे अंतराल

वाले वर्ग की प्रेक्षणों की संख्या से अधिक होने की संभावना होती है, इसलिए बहुलक वर्ग की अर्थपूर्ण परिभाषा के लिए विभिन्न वर्गों के अंतराल समान होने आवश्यक होते हैं।

असंतत समंकों का बहुलक ज्ञात करना सरल होता है। लेकिन संतत समंकों के बहुलक का परिकलन निम्नलिखित सूत्र के उपयोग द्वारा किया जाता है।

$$M_o = l_m + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times h,$$ जहाँ पर

l_m बहुलक वर्ग, वह वर्ग जिसमें बहुलक है, की निम्न परिसीमा,

$\Delta_1 (= f_m - f_{m-1})$ बहुलक वर्ग तथा इसके पहले वर्ग की बारंबारताओं का अंतर,

$\Delta_2 (= f_m - f_{m+1})$ बहुलक वर्ग तथा इससे अगामी वर्ग की बारंबारताओं का अंतर, तथा

h बहुलक वर्ग का अंतराल है।

सारणी 4.2 में दिए गए बंटन में अधिकतम बारंबारता 28, वर्ग 334.5 – 358.5 की है। अत: बहुलकता वर्ग 334.5 – 358.5 होगा।

इस प्रकार $l_m = 334.5,\ \Delta_1 = 28 - 16 = 12, \Delta_2 = 28 - 26 = 2$ तथा $h = 24$.

अत: $$M_o = 334.5 + \frac{12}{12+2} \times 24 = 355.07$$

जब बारंबारता बंटन शिखर प्रबल हो तो बहुलक इसकी केन्द्रीय प्रवृत्ति का उपयोगी माप होता है और यदि बारंबारता बंटन सपाट हो तो बहुलक के माप की कोई उपयोगिता नहीं होती।

बोध प्रश्न 1

1) एक औद्योगिक नगर के एक क्षेत्र में 250 परिवारों के आकार का बारंबारता बंटन निम्नलिखित है :

परिवार आकार	बारंबारता
1	4
2	22
3	25
4	45
5	52
6	41
7	36
8	15
9	7
10	3
योग	250

माध्य, माध्यिका तथा बहुलक ज्ञात कीजिए।

....

....

....

....

....

....

....

....

....

....

2) निम्नलिखित बारंबारता बंटन का माध्य, माध्यिका तथा बहुलक परिकलित कीजिए :

6 वर्ष के 309 बच्चों के आई० क्यू० (I.Q.) का बारंबारता बंटन

आई० क्यू०	बारंबारता
160 - 169	2
150 - 159	3
140 - 149	7
130 - 139	19
120 - 129	37
110 - 119	79
योग	309

....

....

....

....

....

....

....

....

....

....

4.3 केन्द्रीय प्रवृत्ति के अन्य माप

समांतर माध्य, माध्यिका तथा बहुलक के अतिरिक्त केन्द्रीय प्रवृत्ति के अन्य माप भी होते हैं, जो अपेक्षाकृत इतने महत्त्वपूर्ण नहीं होते लेकिन कुछ विशेष परिस्थितियों में बड़े ही उपयुक्त होते हैं। इनको *गुणोत्तर माध्य* तथा *हरात्मक माध्य* कहते हैं। इनका विवेचन अनुभाग 4.3.1 में किया गया है।

इसके अतिरिक्त प्राय: हम यह पाते हैं कि सभी प्रेक्षणों का महत्त्व बराबर नहीं होता। ऐसी परिस्थिति में हम साधारण माध्य के स्थान पर भारित माध्य – समांतर, गुणोत्तर या हरात्मक – का उपयोग करते हैं। इसका विवेचन अनुभाग 4.3.2 में किया जायेगा।

4.3.1 गुणोत्तर तथा हरात्मक माध्य

प्राय: हमारा संपर्क समय से संबंधित समंकों से होता है अर्थात् कालश्रेणी समंकों से, जो कि सारणी 4.1 तथा 4.2 में दिए गए एक समय बिन्दु पर समंकों से भिन्न होते हैं। इन कालाश्रित समंकों में हमारी रुचि प्राय: समय के साथ इनके परिवर्तन के रूप को जानने में होती है। निम्नलिखित समंकों के दो समुच्चयों पर ध्यान दीजिए।

समुच्चय I :	1000	1100	1200	1300	1400	1500	1600
समुच्चय II :	1100	1210	1331	1464	1611	1772	1949

पहला समुच्चय एक कर्मचारी के मूल वेतन (रुपयों में) की तरह लगता है जो कि, 100 रुपये की वार्षिक वृद्धि सहित, 7 वर्षों के लिए दिया हुआ है।

दूसरा समुच्चय कर्मचारी के सकल (Gross) वेतन की तरह लगता है। दोनों समुच्चयों में वार्षिक वृद्धि इस प्रकार है :

समुच्च्य I :	100	100	100	100	100	100
समुचच्य II :	110	121	133	147	161	177

समुच्चय I का समांतर माध्य 100 है तथा समुच्चय II का समांतर माध्य 141.5 है। इस औसत वृद्धि के आधार पर यदि हम दोनों समुच्चयों की संख्याएँ, उनके शुरू के मान लेकर, परिकलित करें तो हमें निम्नलिखित प्राप्त होगा :

समुच्चय I :	1000	1100	1200	1300	1400	1500	1600
समुच्च्य II :	1100	1241.5	1383	1524.5	1666	1807.5	1949

यहाँ पर समांतर माध्य का प्रयोग समुच्चय I के लिए उपयुक्त है, लेकिन समुच्चय II के लिए नहीं, क्योंकि दोनों समुच्चयों (दिया हुआ तथा परिकलित) में संख्याओं की श्रेणियाँ भिन्न हैं। समुच्चय I की संख्याओं में वृद्धि एक स्थिर मात्रा में हुई है जबकि समुच्चय II की संख्याओं में वृद्धि एक निश्चित दर पर हुई है। समुच्चय I की संख्याएँ समांतर श्रेणी में हैं इसलिए औसत वृद्धि की व्याख्या के लिए समांतर माध्य उपयुक्त है। इसी प्रकार, समुच्चय II की संख्याएँ गुणोत्तर श्रेणी में है तथा औसत वृद्धि दर की व्याख्या के लिए गुणोत्तर माध्य उपयुक्त होगा।

n संख्याओं, $X_1, X_2, \ldots X_n$ के लिए गुणोत्तर माध्य (GM) इन संख्याओं के गुणनफल का nवाँ मूल होता है।

$$\text{गुणोत्तर माध्य } (G.M.) = (X_1 X_2 \cdots X_n)^{\frac{1}{n}} = \left[\prod_{i=1}^{n} X_i\right]^{\frac{1}{n}}$$

स्पष्टत:, यदि सभी संख्याएँ धनात्मक न हों तो गुणोत्तर माध्य परिभाषित नहीं होता। $G.M.$ का लघु (log) लेने पर

$$\text{लघु } (GM) = \left(\frac{1}{n}\right)(\text{लघु } X_1 + \text{लघु } X_2 + \cdots + \text{लघु } X_n) = \frac{1}{n}\sum_{i=1}^{n} \text{लघु } X_i$$

इससे यह पता चलता है कि लघु सारणी के उपयोग द्वारा गुणोत्तर माध्य का परिकलन किया जा सकता है। यहाँ पर ध्यान दें की लघु X के मानों के समांतर माध्य का प्रतिलघु (antilog), गुणोत्तर माध्य होता है। समंकों के समुच्चय II में सकल वेतन में वृद्धि 11% प्रतिवर्ष है। लेकिन, व्यवहार में, वृद्धि या कमी किसी निश्चित दर पर नहीं होती तथा औसत वृद्धि दर ज्ञात करना अर्थपूर्ण हो

सकता है। सामान्यत: औसत वृद्धि दर ज्ञात करने के लिए गुणोत्तर माध्य का उपयोग समांतर माध्य के उपयोग से अधिक उपयुक्त होता है। इसलिए, विभिन्न प्रकार के कीमत सूचकांक, उपभोगता कीमत सूचकांक आदि में गुणोत्तर माध्य का उपयोग किया जाता है।

अंत में, हम केन्द्रीय प्रवृत्ति के एक और माप की व्याख्या करेंगे। इसको हरात्मक माध्य (Harmonic Mean या *H.M.*) कहते हैं। जैसा कि दिए हुए उदाहरण से स्पष्ट होगा, कई परिस्थितियों में यह माध्य स्वत: ही प्राप्त हो जाता है। एक व्यापारी, प्रत्येक माह के शुरू में, 5000 रुपये मूल्य के बराबर सामान का संग्रहण करता है। 5 उत्तरोत्तर महीनों में वस्तु की प्रति इकाई दर (रुपयों में) इस प्रकार है: 10.75, 11.80, 14.00, 11.45 तथा 12.00। व्यापारी पिछले 5 महीनों में संचित सामान की औसत प्रति इकाई कीमत जानना चाहता है। यह परिकलन निम्नलिखित में प्रस्तुत है :

मास	व्यय की गई राशि (रुपयों में)	प्रति इकाई दर (रुपयों में)
1	5000	10.75
2	5000	11.80
3	5000	14.00
4	5000	11.45
5	5000	12.00
योग	25000	

कुल संग्रहण की प्रति इकाई कीमत = (कुल व्यय की गई राशि) ÷ (कुल क्रय की गई मात्रा)

$$= \frac{5 \times 5000}{\frac{5000}{10.75} + \frac{5000}{11.80} + \frac{5000}{14.00} + \frac{5000}{11.45} + \frac{5000}{12.00}}$$

$$= \frac{5}{\frac{1}{10.75} + \frac{1}{11.80} + \frac{1}{14.00} + \frac{1}{11.45} + \frac{1}{12.00}}$$

$$= \frac{1}{\frac{1}{5}\left(\frac{1}{10.75} + \frac{1}{11.80} + \frac{1}{14.00} + \frac{1}{11.45} + \frac{1}{12.00}\right)} = 11.91$$

अंतिम व्यंजक व्युत्क्रमों के समांतर माध्य का व्युत्क्रम है तथा इसको हरात्मक माध्य (*H.M.*) कहते हैं। X के मानों के एक समुच्चय, $X_1, X_2, \ldots X_n$, के लिए हरात्मक माध्य की परिभाषा निम्नलिखित है :

$$H.M. = \frac{n}{\frac{1}{X_1} + \frac{1}{X_2} + \ldots + \frac{1}{X_n}} = \frac{n}{\sum_{i=1}^{n} \frac{1}{X_i}}$$

यदि एक भी प्रेक्षण शून्य हो तो हरात्मक माध्य परिभाषित नहीं होता।

यदि व्यापारी, प्रतिमास के शुरू में, दी हुई कीमतों पर 5000 रुपये मूल्य के बराबर सामान के स्थान पर, 3000 इकाइयों का संग्रहण करता तो उपयुक्त औसत, समांतर माध्य होगा।

इसकी जाँच के लिए हम लिखते हैं :

औसत कीमत = (कुल व्यय की गई राशि) ÷ (कुल क्रय की गई मात्रा)

$$= \frac{3000\times10.75+3000\times11.80+3000\times14.00+3000\times11.45+3000\times12.00}{3000\times5}$$

$$= \frac{10.75+11.80+14.00+11.45+12.00}{5} = \text{दी हुई कीमतों का सामांतर माध्य}$$

4.3.2 भारित माध्य

बहुत से व्यवहारिक अनुप्रयोगों में भारित माध्य (समांतर, गुणोत्तर तथा हरात्मक), अभारित या साधारण माध्य की तुलना में एक घटना को भली भाँति दर्शाते हैं।

उदाहरण के लिए उपभोगता कीमत सूचकांक के परिकलन में सभी वस्तुओं का महत्त्व समान नहीं होता। ईंधन की कीमत में वृद्धि, कृषि वस्तुओं की कीमतों में वृद्धि की तुलना में उपभोगता कीमत सूचकांक को अधिक प्रभावित कर सकती है। शेयर बाज़ार में कुछ कम्पनियों के शेयर ही बाज़ार के प्रवृत्ति के निर्धारक हो सकते हैं। इस प्रकार की परिस्थितियों में भारित माध्य अधिक उपयुक्त होते हैं। भारित माध्य प्राप्त करने के लिए प्रत्येक X_i के साथ एक भार w_i को संलग्न किया जाता है तथा इसका माध्य ठीक उसी प्रकार परिकल्नित किया जाता है जैसे कि w_i, संकेत रूप में, X_i की बारंबारता हों। भारित माध्य के विभिन्न सूत्र निम्नलिखित हैं :

$$\text{भारित समांतर माध्य} = \frac{\sum_{i=1}^{n} w_i X_i}{\sum_{i=1}^{n} w_i}$$

$$\text{भारित गुणोत्तर माध्य} = \left(\prod_{i=1}^{n} X_i^{w_i}\right)^{\frac{1}{\sum w_i}}, \text{ तथा}$$

$$\text{भारित हरात्मक माध्य} = \frac{\sum_{i=1}^{n} w_i}{\sum_{i=1}^{n} \frac{w_i}{X_i}}$$

यदि प्रत्येक $w_i = 1$ हो तो भारित माध्य साधारण माध्य बन जाता है।

4.3.3 संयुक्त माध्य

यदि हमारे पास विभिन्न समूहों या प्रतिदर्शों के परिकलित माध्य हों तो कई बार हमारी रुचि समग्र माध्य ज्ञात करने में हो सकती है। इस प्रकार के समग्र माध्य को संयुक्त माध्य (pooled mean) कहते हैं।

मान लिया कि $m_1, m_2 \ldots\ldots m_r$, r समांतर(या गुणोत्तर या हरात्मक) माध्य हैं, जो कि क्रमशः $n_1, n_2 \ldots\ldots n_r$, प्रेक्षणों से परिकलित किए गए हैं। तब

$$\text{संयुक्त समांतर माध्य} = \frac{1}{n}\sum_{i=1}^{r} m_i n_i$$

$$\text{संयुक्त गुणोत्तर माध्य} = \left(\prod_{i=1}^{r} m_i^{n_i}\right)^{\frac{1}{n}} \text{ तथा}$$

$$\text{संयुक्त हरात्मक माध्य} = \frac{n}{\sum_{i=1}^{r} \frac{n_i}{m_i}}$$

यहाँ पर $n = n_1 + n_2 + \ldots + n_r$ हैं।

संयुक्त तथा भारित माध्यों के व्यंजकों की समरूपता पर ध्यान दीजिए।

4.3.4 केन्द्रीय प्रवृत्ति के माप का चयन

हम पहले ही बता चुके हैं कि किन परिस्थितियों में एक माध्य (समांतर, गुणोत्तर, हरात्मक) बाकी दो माध्यों की तुलन्ा में अधिक उपयुक्त होता है। लेकिन, यदि हमारे समंकों के विभिन्न वर्गों के बाएँ या दाएँ छोर खुले है अर्थात् 'c_i तक' तथा / या 'c_{k-1} या अधिक' प्रकार के हैं, तो इन वर्गों के मध्य बिन्दु ज्ञात करना संभव नहीं होता। अतः ऐसी परिस्थिति में कोई माध्य परिकलित नहीं किया जा सकता। लेकिन, इन परिस्थितियों में, माध्यिका या बहुलक के परिकलन में कोई कठिनाई नहीं होती। इसके विपरीत यहाँ पर, माध्य की तरह, संयुक्त माध्यिका या संयुक्त बहुलक का परिकलन नहीं किया जा सकता। इनके परिकलन के लिए हमारे पास समग्र समंकों का समुच्चय होना आवश्यक है। इन कठिनाइयों का संबंध माप की उपयुक्तता से न होकर केवल परिकलन की कठिनाइयों से है।

समंकों का आलेखी निरूपण अधिक आकर्षक होता है। इसलिए इस परिस्थिति में माध्यिका या बहुलक अधिक उपयोगी रहते हैं क्योंकि आलेखों द्वारा, बिना परिकलन किए, इनके अशोधित (crude) मान ज्ञात किए जा सकते हैं। इसके अतिरिक्त, आलेखों में तुलना तथा संचारण (communication) के लिए भी माध्यिका तथा बहुलक सरल अवधारणाएँ हैं। लेकिन, इस प्रकार के आलेखों की तुलना बड़ी सावधानी से की जानी चाहिए, क्योंकि पुनरावृत प्रतिचयन (repeated sampling) में यह पाया गया है कि समांतर माध्य की तुलना में माध्यिका कम स्थायी होती है।

उन समंकों के लिए, जिनका बंटन प्रसामान्य (normal) बंटन जैसा होता है, जिसमें एक शिखर होता है तथा इस शिखर से यह दोनों ओर सममितत: (symmetrically) कम होता जाता है, हम माध्य, माध्यिका या बहुलक का उपयोग कर सकतें है। इस प्रकार के बंटन में ये तीनों माप बराबर होते है।

यहाँ यह समझना आवश्यक है कि केन्द्रीय प्रवृत्ति के उपयुक्त माप का चयन ही समंकों के विश्लेषण का उद्देश्य नहीं है तथा इस दिशा में अभी बहुत कुछ करना बाकी है। उदाहरण के लिए, यह कहना की खाद्य सामग्री पर गृहों का औसत मासिक व्यय 348.66 रुपये है, पर्याप्त नहीं है क्योंकि इससे यह पता नहीं चलता कि क्या बहुत बड़ी संख्या में गृहों का औसत मासिक व्यय बहुत कम है या कुछ गृहों का मासिक व्यय बहुत अधिक है। इस प्रकार के प्रश्नों के उत्तर बाद में किये जाने वाले विश्लेषणों से प्राप्त होंगे।

4.4 शतमक

शतमक की धारणा का विवेचन सारणी 4.2 में दिए हुए, गृहों के खाद्य सामग्री पर औसत मासिक व्यय, समंकों के उपयोग द्वारा किया जाएगा। शतमक के उपयोग द्वारा दो प्रकार के प्रश्नों के उत्तर दिए जा सकते हैं। उदाहरण के लिए, कितने प्रतिशत गृहों का खाद्य सामग्री पर औसत मासिक व्यय 350.80 रुपये तक है? या निम्न 50% गृहों का खाद्य सामग्री पर अधिकतम औसत मासिक व्यय क्या है? सारणी 4.2 में माध्यिका के परिकलन पर ध्यान देने से यह ज्ञात होता है कि पहले प्रश्न का उत्तर दूसरे प्रश्न में दी हुई संख्या है, अर्थात् निम्न 50% गृहों का अधिकतम औसत मासिक व्यय 350.80 रुपये है। रुचि के अनुकुल हमारी एक अंतकीय (cut-off) बिन्दु के नीचे प्रतिशत ज्ञात करने की इच्छा हो सकती है; गरीबी रेखा के निर्धारण में हमारी रुचि इस रेखा के नीचे वाले प्रतिशत में हो सकती है। दूसरे प्रकार के प्रश्नों में हमारी रुचि जनसंख्या के निम्न 10% या उच्च 5% की स्थिति ज्ञात करना हो सकती है। इस प्रकार के प्रश्नों के उत्तर शतमक के उपयोग के द्वारा दिए जा सकते हैं।

4.1.1 शतमक : परिभाषा तथा परिकलन

किसी दिए हुए प्रतिशत v के लिए v वाँ शतमक P_v, अध्ययन के अंतर्गत चर का ऐसा मान होता है कि कम से कम v प्रतिशत प्रेक्षण P_v से कम या इसके बराबर रहें तथा कम से कम $(100 - v)$ प्रतिशत प्रेक्षण P_v से अधिक या बराबर रहें। उदाहरण के लिए सारणी 4.1, गृहों के आकार का बंटन, में v के 78 से 89 तक किसी मान के लिए $P_v = 5$ है।

वर्गीकृत समंकों के लिए, शतमकों को, संचयी बारंबारता बंटन द्वारा अच्छी तरह से समझा जा सकता है। मान लिया, X से कम या बराबर प्रेक्षणों का अनुपात $F(X)$ है। इसी प्रकार कोई दिया हुआ मान X_0, बंटन का $100F(X_0)$वाँ शतमक होगा। सारणी 4.2 में वर्ग परिसीमाओं के लिए $F(286.5) = 0.1$, $F(310.5) = 0.15$, $F(334.5) = 0.31$, $F(358.5) = 0.59$ तथा $F(382.5) = 0.85$ है। अत: $286.5 = P_{10}$, $310.5 = P_{15}$, $334.5 = P_{31}$, $358.5 = P_{59}$, तथा $382.5 = P_{85}$ होगा। यह ध्यान दीजिए कि 262.5 (पहले वर्ग की निम्न परिसीमा) से कम कोई संख्या इस शून्य-वाँ शतमक होगी तथा 406.5 (अंतिम वर्ग की उच्च परिसीमा) से अधिक 100वाँ शतमक होगी।

4.1.2 चतुर्थक तथा दशमक

उपयोग के अनुसार, कुछ विशेष शतमकों के विभिन्न नाम हो सकते हैं। प्रत्येक 25वाँ शतमक एक चतुर्थक कहलाता है तथा दसवाँ शततमक एक दशमक कहलाता है।

उदाहरण के लिए :

25 वाँ शतमक $= P_{25} = Q_1$ प्रथम चतुर्थक

50 वाँ शततमक $= P_{50} = Q_2$ द्वितीय चतुर्थक

75 वाँ शततमक $= P_{75} = Q_3$ तृतीय चतुर्थक

10 वाँ शततमक $= P_{10} = d_1$ प्रथम दशमक

20 वाँ शततमक $= P_{20} = d_2$ द्वितीय दशमक इत्यादि, तथा

$P_{50} = Q_2 = d_5 =$ माध्यिका होती है।

Q_1 तथा Q_3 के सूत्र, माध्यिका सूत्र के सदृश होते हैं। इनको प्रत्यक्ष रूप में इस प्रकार लिखा जा सकता है :

$$Q_1 = l_{Q_1} + \frac{\frac{N}{4} - C}{f_{Q_1}} \times h, \text{ तथा}$$

$$Q_3 = l_{Q_3} + \frac{\frac{3N}{4} - C}{f_{Q_3}} \times h,$$

जहाँ पर C, प्रथम (या तृतीय) चतुर्थक वर्ग से पहले वर्गों की संचयी बारंबारता को सूचित करता है तथा h इसका अंतराल है।

इसी प्रकार हम किसी भी विभाजन मान का सूत्र लिख सकते हैं। उदाहरण के लिए, 40वें शतमक का सूत्र निम्नलिखित होगा।

$$P_{40} = l_{P_{40}} + \frac{\frac{40N}{100} - C}{f_{P_{40}}} \times h$$

जब प्रतिशत के स्थान पर अनुपात प्रयोग किए जाएँ तो शतमक को भिन्नक (fractile) कहते हैं। उदाहरण के लिए P_{30} को 0.3 भिन्नक कहते हैं।

जिस प्रकार केन्द्रीय प्रवृत्ति के माप द्वारा बंटन के बारे में पूर्ण जानकारी प्राप्त नहीं हो सकती, उसी प्रकार बंटन के प्रकीर्णन की व्याख्या के लिए बहुत से शतमकों की आवश्यकता हो सकती है। इसीलिए प्रकीर्णन के सरल माप की आवश्यकता महसूस की गई। यही अगली इकाई में विवेचन का विषय भी है।

बोध प्रश्न 2

1) निम्नलिखित में पाँच वस्तुओं की कीमत (अनुपातों में) तथा संगत भार दिए हुए हैं। भारित समांतर तथा गुणोत्तर माध्य परिकलित कीजिए।

वस्तु	कीमत अनुपात	भार
1	2.20	30
2	1.85	25
3	1.80	22
4	2.05	13
5	1.75	10

....

....

....

....

....

....

....

....

2) पाँच राष्ट्रीयकृत बैंकों का उपार्जन (करोड़ रुपयों में) निम्नलिखित है :
217.40, 330.5, 682.55, 1263.59, 2249.63
इन उपार्जनों का गुणोत्तर माध्य ज्ञात कीजिए।

3) पुरुषों की शादी के समय आयु का बंटन निम्नलिखित था।

आयु (वर्षों में)	पुरुषों की संख्या
18 - 20	5
20 - 22	18
22 - 24	28
24 - 26	37
26 - 28	24
28 - 30	22

शादी के समय पर (i) औसत आयु, (ii) बहुलकता आयु, (iii) माध्यिका आयु, (iv) तृतीय चतुर्थक, (v) छठा दशमक, तथा (vi) 19वाँ शतमक ज्ञात कीजिए।

4) एक कारखाने में एक मिस्त्री 15 दिनों में एक मशीन का निर्माण करता है, दूसरा मिस्त्री इसको 18 दिनों में, तीसरा मिस्त्री 30 दिनों में तथा चौथा मिस्त्री इसको 90 दिनों में निर्मित करता है। एक मशीन के निर्माण में औसत दिनों की संख्या ज्ञात कीजिए। आप कौन-सा माध्य प्रयोग करेंगे तथा क्यों?

....

....

....

....

....

....

....

....

....

....

....

5) तीन विभिन्न मुद्रा राशियों पर 10%, 12% तथा 15% प्रतिवर्ष की दर से, वार्षिक सरल ब्याज समान हैं। कुल निवेशित राशि पर औसत उपार्जन प्रतिशत दर क्या होगी?

....

....

....

....

....

....

....

....

....

....

4.5 सारांश

इस इकाई में आपने केन्द्रीय प्रवृत्ति के विभिन्न मापों के परिकलन के बारे में अध्ययन किया। इन मापों को, मोटे तौर पर, दो भागों – गणितीय माध्य तथा स्थितीय माध्य – में विभाजित किया जा सकता है। बहुलक, माध्यिका, चतुर्थक, शतमक आदि को, स्थितीय माध्य कहते हैं जबकि सामांतर माध्य, गुणोत्तर माध्य तथा हरात्मक माध्य गणितीय माध्य कहलाते हैं। अनुपात या आनुपातिक वृद्धि दरों का औसत ज्ञात करने के लिए गुणोत्तर माध्य सबसे उपयुक्त होता है जबकि समांतर या हरात्मक माध्य का उपयोग, दी हुई शर्त के अनुसार, कीमत, गति आदि का औसत ज्ञात करने के लिए किया जाता है।

4.6 शब्दावली

समांतर माध्य	:	एक समुच्चय के प्रेक्षित मानों के योग को प्रेक्षणों की संख्या से विभाजित करने पर समांतर माध्य या औसत प्राप्त होता है।
बारंबारता बंटन	:	समंकों का बारंबारता बंटन के रूप में विन्यास, समंकों के ढेर में निहित प्रतिरूप की व्याख्या करता है।
गुणोत्तर माध्य	:	यह चर के n मानों के गुणनफल का nवाँ मूल होता है।
हरात्मक माध्य	:	यह समुच्चय के प्रेक्षणों के व्युत्क्रमों के समांतर माध्य का व्युत्क्रम होता है।
माध्यिका	:	एक समुच्चय के प्रेक्षणों को परिमाण कोटि (order of magnitude) के अनुसार व्यवस्थित करने पर इनके मध्यवर्ती मान को माध्यिका कहते हैं।
बहुलक	:	यह प्रेक्षणों के समुच्चय में वह मान होता है, जिसकी बारंबारता अधिकतम होती है।

4.7 कुछ उपयोगी पुस्तकें

Elhance, D. N. and V. Elhance, 1988, *Fundamentals of Statistics,* Kitab Mahal, Allahabad.

Nagar, A. L. and R. K. Dass, 1983, *Basic Statistics,* Oxford University Press, Delhi.

Mansfield, E., 1991, *Statistics for Business and Economics : Methods and Applications,* W. W. Norton and Co.

Yule, G. U. and M. G. Kendall, 1991, *An Introduction to the Theory of Statistics,* Universal Books, Delhi.

4.8 बोध प्रश्नों के उत्तर या संकेत

बोध प्रश्न 1

1) 5.1, 5, 5

2) 108.48, 108.41, 111.42

बोध प्रश्न 2

1) 1.96 रुपये; 1.95 रुपये

2) 6.74.31 करोड़ रुपये

3) (i) 25.83 वर्ष (ii) 24.82 वर्ष (iii) 24.86 वर्ष
(iv) 27.30 वर्ष (v) 25.59 वर्ष (vi) 28.79 वर्ष

4) समांतर माध्य, 38.25 दिन

5) हरात्मक माध्य 12%

4.9 पारिभाषिक शब्दावली

केन्द्रीय प्रवृत्ति के माप	:	measures of central tendency
चतुर्थक	:	quartile
दशमक	:	decile
प्रकीर्णन के माप		measures of disperson
बहुलक		mode
भारित माध्य		weighted mean
शतमक		percentile
संयुक्त माध्य		pooled mean
प्रसामान्य बंटन		normal distribution
भिन्नक		fractile

इकाई 5 प्रकीर्णन के माप

इकाई की रूपरेखा

5.0 उद्देश्य

इस इकाई का अध्ययन करने के बाद आपको निम्नलिखित के संबंध में जानकारी मिल सकेगी:

- प्रकीर्णन की अवधारणा;
- समंक समुच्च्य के प्रकीर्णन का संख्यात्मक परिकलन करना;
- चेबाईचेव्य (Chebychev) की असमता;
- विचरण गुणांक का परिकलन करना; तथा
- समंकों के कुछ बंटनों के सांद्रता माप को ज्ञात करना।

5.1 प्रस्तावना

इकाई 4 में हमने विभिन्न केन्द्रीय प्रवृत्ति के मापों जैसे सामान्तर माध्य, माध्यिका, बहुलक, गुणोत्तर माध्य तथा हरात्मक माध्य आदि, का विवेचन किया था। लेकिन कई परिस्थितियों में यह माप समंकों के बंटन को प्रर्याप्त रूप में निरूपित करने में असमर्थ होते हैं। उदाहरण के लिए समंकों के निम्नलिखित समुच्च्यों पर ध्यान दीजिए :

समुच्चय क : 2, 5, 17, 17, 44
समुच्चय ख : 17, 17, 17, 17, 17
समुच्चय ग : 13, 14, 17, 17, 24

इन सभी समुच्चयों के माध्य, माध्यिका तथा बहुलक के संख्यात्मक मान समान अर्थात् 17 हैं, लेकिन ये भिन्न हैं। जबकि समुच्चय ख के सभी प्रेक्षण बराबर हैं, समुच्चय क के प्रेक्षण बहुत भिन्न हैं। अतः, समंकों की इस भिन्नता को मापने के लिए, हमें एक और माप की आवश्यकता है।

इस इकाई में आप चर/चरों के समंकों के, विचरण परिसर में, बंटन के बारे में निष्कर्ष प्राप्त करने के लिए अवधारणाओं तथा प्रविधियों के उपयोग के बारे में सीखेंगे।

5.2 प्रकीर्णन की अवधारणा

प्रकीर्णन शब्द का उपयोग समंकों के विषमांगता (Heterogeneity) की मात्रा को सूचित करने के लिए किया जाता है। विभिन्न प्रेक्षण आपस में किस सीमा तक भिन्न है, इसको बताने वाला यह मुख्य अभिलक्षण होता है। यदि प्रेक्षणों के समुच्चय में सभी प्रेक्षण बराबर हैं (जैसा समुच्चय ख), तो इनका प्रकीर्णन शून्य होगा। प्रेक्षणों में भिन्नता जितनी अधिक होगी, प्रकीर्णन उतना ही अधिक होगा (इस प्रकार समुच्चय क में प्रकीर्णन समुच्चय ग से अधिक होना चाहिए)। प्रकीर्णन के माप का उद्देश्य व्यष्टि प्रेक्षणों में औसत भिन्नता की सीमा को संख्यात्मक रूप में अभिव्यक्त करना होता है।

प्रकीर्णन के कई माप होते हैं। इनका विवेचन निम्नलिखित है :

5.2.1 परिसर

प्रकीर्णन के सभी मापों में परिसर माप सरलतम होता है। प्रेक्षणों के अधिकतम तथा न्यूनतम मानों के अंतर को परिसर कहते हैं। इस प्रकार समुच्चय क में दिये गए समंकों का परिसर $44 - 2 = 42$ है। इसी प्रकार समुच्चय ख में परिसर $17 - 17 = 0$ तथा समुच्चय ग में परिसर 11 है। अब हम वर्गीकृत समंकों के परिसर की जानकारी प्राप्त करते हैं। सारणी 4.2 (पहली इकाई में देखिए) में, परिसर $406.5 - 262.5 = 144$ रुपये है। यहाँ यह ध्यान दें कि वर्गीकृत समंकों के लिए प्रेक्षणों के अधिकतम या न्यूनतम मानों की पहचान करना संभव नहीं होता। इसलिए यहाँ बंटन की दो चरम परिसीमाओं के अंतर को परिसर कहते हैं।

यह बात अंतःदर्शी है कि यदि हम छोटे आकार का प्रतिदर्श लें तो केन्द्रीय प्रवृत्ति के कारण, प्रेक्षणों के अपने बहुलक के आसपास रहने की बड़ी संभावना होती है। यदि प्रतिदर्श के आकार में वृद्धि होती है तो इसमें कम संभावनीय या चरममान भी सम्मितिल हो जाते हैं जिसका अर्थ यह होगा कि प्रतिदर्श के आकार में वृद्धि के साथ परिसर आकार में भी वृद्धि हो जाती है। इसके साथ हम यह भी जानते है कि पुनरावृत प्रतिचयनों में प्रतिदर्शों का आकार सामान रहने पर भी परिसर में अत्यधिक परिर्वतन होता है। इन सभी बातों के बावजूद परिसर एक ऐसा माप है, जिसको सरलता से समझा तथा परिकलित किया जा सकता है।

5.2.2 अंतःचतुर्थक परिसर

क्योंकि परिसर केवल दो चरम मानों पर ही निर्भर होता है, इसलिए प्रकीर्णन के माप के रूप में यह बंटन की ठीक प्रकार से व्याख्या नहीं करता। समंकों के समुच्चय में एक मान, जो बहुत बड़ा या छोटा है तथा प्रेक्षणों के सामान्य प्रतिरूप से भिन्न है, परिसर को बड़ा कर देता है। उदाहरण के लिए समुच्चय क में, एक बहुत बड़े प्रेक्षण 44 के कारण, परिसर $(44 - 2 = 42)$ बहुत बड़ा हो गया है। इस प्रकार के प्रेक्षणों से बचने के लिए, विशेषकर जब समंकों में केंन्द्रीय प्रवृत्ति प्रबल हो, अंतःचतुर्थक परिसर, प्रकीर्णन का उपयोगी माप होता है। इसकी परिभाषा इस प्रकार की जाती है :

अंत:चतुर्थक परिसर $= Q_3 - Q_1 = P_{75} - P_{25}$

अंत:चतुर्थक परिसर मध्यवर्ती 50% प्रेक्षणों का परिसर होता है। यदि प्रेक्षण माध्यिका के आसपास सघन हैं, अर्थात् माध्यिका के निकट प्रबल बहुलक है तो परिसर के आधे की तुलना में अंत:चतुर्थक परिसर छोटा होगा। यदि समंक सपाट हैं, जिनमें कोई केन्द्रीय प्रवृत्ति विद्यमान नहीं है, तो यह माप बड़ा होगा तथा परिसर के आधे के निकट होगा।

सारणी 4.1 के समंकों के लिए $P_{75} = 4$ तथा $P_{25} = 3$ है। अत: अंत:चतुर्थक परिसर = 4 – 3 = 1 होगा। क्योंकि यहाँ परिसर 7 है, इसलिए गृहों के आकार में प्रबल केन्द्रीय प्रवृत्ति विद्यमान है।

सारणी 4.2, खाद्य सामग्री पर औसत मासिक व्यय के बंटन के लिए $P_{25} = 325.50$ रुपये तथा $P_{75} = 377.88$ रुपये है। अत: अंत:चतुर्थक परिसर = 377.88 – 325.50 = 52.38 रुपये होगा। इसकी तुलना में परिसर में 146.00 रुपये है, जो कि अंत: परिसर का 2.79 गुणा है। यह इस बात का सूचक है कि खाद्य सामग्री पर औसत मासिक व्यय के समंकों में केन्द्रीय प्रवृत्ति इतनी प्रबल नहीं है।

5.2.3 माध्य विचलन

जबकि परिसर दो चरम प्रेक्षणों पर निर्भर होता है, अंत:चतुर्थक परिसर मध्यवर्ती 50% के दो चरम प्रेक्षणों पर आधारित होता है। अत: हम प्रत्येक माप में प्रेक्षणों के बंटन के बारे में कुछ न कह कर केवल प्रेक्षणों के न्युनतम (या P_{25}) तथा अधिकतम (या P_{75}) के बीच प्रतिशत की बात करते हैं। समंकों के प्रकीर्णन का निर्धारण, अनेक संभावनाओं में से एक, प्रेक्षणों के किसी केन्द्रीय मान से विचलन के उपयोग द्वारा किया जा सकता है। केन्द्रीय प्रवृत्ति के माप के रूप में अधिकतर माध्य का उपयोग किया जाता है, अत: हम प्रेक्षणों के विचलन प्राय: इसी माप से परिकलित करते हैं। प्रकीर्णन के माप को ज्ञात करने के लिए इन विचलनों को उपयुक्त प्रकार से संयुक्त किया जाता है।

माध्य विचलन, प्रत्येक प्रेक्षण पर आधारित विचलनों के समांतर माध्य के रूप में, प्रत्येक प्रेक्षण मान को समान भार (महत्त्व) देता है।

$X_1, X_2, \ldots X_n$ प्रेक्षणों के लिए, यदि हम $\bar{X}$ से सामान्य अंतर को विचलन लें तो i वें प्रेक्षण के लिए विचलन $X_i - \bar{X}$ होगा, जहाँ पर $\bar{X}$ प्रेक्षणों का माध्य है। इन विचलनों का माध्य

$$\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right) = \frac{1}{n}\sum_{i=1}^{n}X_i - \frac{\bar{X}}{n}\sum_{i=1}^{n}1 = \bar{X} - \bar{X} = 0 \text{ है।}$$

क्योंकि विचलनों को सामान्य अंतर के रूप में लेने से कोई माप प्राप्त नहीं होता, इसलिए माध्य विचलन के लिए निरपेक्ष (absolute) अंतरों का उपयोग किया जाता है।

माध्य विचलन $= \frac{1}{n}\sum_{i=1}^{n}\left|X_i - \bar{X}\right|$, जहाँ पर दो रेखिकाएँ यह व्यक्त करती हैं कि इनके बीच दो संख्याओं के अंतर का चिह्न धनात्मक लिया जाना है। उदाहरण के लिए 2 – 4 = 2 लिया जायेगा, आदि।

असंतत तथा संतत बारंबारता समंकों के लिए, सूत्र को इस प्रकार लिखा जाता है :

माध्य विचलन = $\frac{1}{N}\sum_{i=1}^{n} f_i|x_1 - \bar{x}|$ जहाँ पर $N = \sum_{i=1}^{n} f_i$, तथा X_i पृथक् (distinct) प्रेक्षण हैं तथा असंतत प्रकार के बंटन के लिए X_i की बारंबारता f_i है। यदि बंटन संतत है तो X_i का मान i वें वर्ग का मध्यबिन्दु होता है तथा इस वर्ग की बारंबारता f_i होती है। इस प्रकार के माप की आवश्यकता निम्नलिखित उदाहरण द्वारा स्पष्ट हो जाएगी।

दो समंक समुच्चयों के लिए परिकलित संक्षेपण मान निम्नलिखित है :

	समंक समुच्चय I	समंक समुच्चय II
प्रेक्षणों की संख्या	7	7
P_{25}	7	7
माध्यिका $= P_{50}$	12	12
P_{75}	17	17
परिसर	20	20
अंत:चतुर्थक परिसर	10	10
माध्य	12	12

यहाँ पर केवल इन मापों के आधार पर, समंक समुच्च्यों पर ध्यान दिए बिना, ऐसा प्रतीत होता है कि दो व्यक्तियों ने समंकों के एक ही समुच्चय से ये परिकलन किए हैं। लेकिन वास्तव में दोनों समंक समुच्चय निम्नलिखित है :

समंक समुच्चय I	:	3	7	8	12	14	17	23
समंक समुच्चय II	:	2	7	11	12	13	17	22

इसी प्रकार, हम समंकों के कई और समुच्चय बना सकते हैं जो आपस में बहुत भिन्न हों तथा उनके माप उपरोक्त मापों के मानों के बराबर हों। यह तुलना इस बात की सूचक है कि हमें और अतिरिक्त मापों की आवश्यकता है तथा माध्य विचलन उनमें से एक है। इसका यह अर्थ कदापि नहीं लेना चाहिए कि उपरोक्त मापों के साथ माध्य विचलन को सम्मिलित करने से समंक समुच्चय की पूर्ण व्याख्या हो जाती है।

समंक समुच्चय I के लिए

माध्य विचलन = $\frac{1}{7}$ (|3-12| + |7-12| + |8-12| + |12-12| + |14-12| + |17-12| + |23-12|)

$$= \frac{9+5+4+0+2+5+11}{7} = \frac{36}{7} = 5.14$$

समंक समुच्चय II के लिए

माध्य विचलन = $\frac{1}{7}$ (|2 -12| + |7-12| + |11-12| + |12-12| + |13-12| + |17 -12| + |22 -12|)

$$= \frac{10+5+1+0+1+5+10}{7} = \frac{32}{7} = 4.57$$

अत: समंक समुच्चय I में प्रेक्षणों का प्रकीर्णन, समंक समुच्चय II से अधिक है।

अब हम गृह आकार तथा गृहों के खाद्य सामग्री पर औसत मासिक व्यय के समंकों के माध्य विचलन का परिकलन करेंगे।

सारणी 4.1 में गृह आकार के लिए समंकों के लिए माध्य = $\bar{x}$ = 3.74 है।

$$\text{माध्य विचलन} = \frac{1}{N}\sum_{i=1}^{n} f_i |X_i - \overline{X}|$$

$$= \frac{1}{100}(3|1-3.74| + 16|2-3.74| + \cdots + 2|8-3.74|) = \frac{109.12}{100} = 1.0912$$

सारणी 4.2 में खाद्य सामग्री पर औसत मासिक व्यय के बंटन के लिए माध्य = $\overline{X}$ = 348.66 रुपये है।

$$\text{माध्य विचलन (रुपयों में)} = \frac{1}{100}(2|274.5-348.66| + \cdots + 15|394.5-348.66|) = \frac{2510.88}{100} = 25.11$$

अभी तक हमने माध्य से माध्य विचलन पर विचार किया है। इसी प्रकार से हम माध्यिका या बहुलक से माध्य विचलन को परिभाषित कर सकते हैं।

5.2.4 प्रसरण तथा मानक विचलन

प्रसरण तथा मानक विचलन प्राय: उपयोग किये जाने वाले प्रकीर्णन के माप हैं। प्रसरण का तो इतना अधिक उपयोग होता है कि बहुधा इसको ही प्रकीर्णन कहा जाता है। प्रसरण व्यष्टि विचलनों को उपयुक्त रूप में संयुक्त करने वाला ऐसा माप है जो माध्य विचलन की भाँति प्रत्येक प्रेक्षण को समान भार (महत्त्व) देता है। प्रसरण में प्रेक्षण तथा माध्य के अंतर के वर्ग को व्यष्टि विचलन कहते हैं। क्योंकि निरपेक्ष अंतर की अपेक्षा अंतर के वर्ग का उपयोग (विशेषत: विधिवत् गणित में) अधिक सरल होता है, इसलिए इनका उपयोग बड़ा ही लोकप्रिय है। यह प्रेक्षणों के माध्य से अंतरों के वर्ग का माध्य होता है। अपरिष्कृत समंकों से प्रसरण का परिकलन निम्नलिखित सूत्र द्वारा किया जाता है :

$$\text{प्रसरण} = \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

असंतत तथा संतत बारंबारता समंकों के लिए सूत्र इस प्रकार है :

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{n} f_i (X_i - \overline{X})^2, \text{ जहाँ पर } N = \sum_{i=1}^{n} f_i \text{ है।}$$

यदि माप का पैमाना समान हो तो वे प्रेक्षण जिनका प्रसरण 2 (उदाहरणार्थ) है, अन्य प्रेक्षणों की तुलना में जिनका प्रसरण 2 से अधिक है, कम प्रकीर्ण (dispersed) कहलाते हैं। किसी बंटन की उसके केन्द्रीय प्रवृत्ति के माप तथा प्रकीर्णन माप द्वारा व्याख्या करते हुए यह आवश्यक है कि दोनों माप एक ही इकाई में व्यक्त हों। माध्य तथा माध्य विचलन की एक ही इकाई होती है। लेकिन, क्योंकि प्रसरण के परिकलन में प्रत्येक विचलन का वर्ग किया जाता है, इसलिए प्रसरण की इकाई प्रेक्षण की इकाई का वर्ग होती है।

प्रसरण पर आधारित तथा इतना या इससे अधिक लोकप्रिय, प्रकीर्णन का एक अन्य माप होता है, जिसकी माप की इकाई प्रेक्षण की इकाई के समान होती है। इस माप को मानक विचलन कहते हैं। मानक विचलन, प्रसरण का धनात्मक वर्गमूल होता है। इसको σ से सूचित किया जाता है।

सारणी 4.1 में गृहों के आकार के लिए मानक विचलन का परिकलन निम्नलिखित है :

$$\sigma^2 = \frac{1}{100}\left[3(1-3.74)^2 + 16\ (2-3.74)^2 + \ldots + 2(8-3.74)^2\right] = \frac{199.24}{100} = 1 \text{ तथा}$$

$$\sigma = 1.4115$$

इसी प्रकार सारणी 4.2 में दिए हुए गृहों के खाद्य सामग्री पर औसत मासिक व्यय के समंकों के लिए प्रसरण (s^2), वर्ग रुपयों में, इस प्रकार होता है।

$$\sigma^2 = \frac{1}{100}\left[2(274.50-348.66)^2 + \ldots + 15(394.5-348.66)^2\right]$$

$$= \frac{95725.437}{100} = 957.25 \text{ तथा } \sigma = 30.94 \text{ रुपये}$$

परिकलन की सुविधा के लिए, प्रसरण सूत्र को वैकल्पिक रूप में भी लिखा जा सकता है।

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2 \text{ या}$$

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{n} f_i\left(X_i - \overline{X}\right)^2 = \frac{1}{N}\sum_{i=1}^{n} f_i X_i^2 - \overline{X}^2$$

इन सूत्रों द्वारा हम पाते है कि

प्रसरण = मानों के वर्ग का माध्य – माध्य का वर्ग

इन सूत्रों के उपयोग द्वारा आप सारणी 4.1 तथा 4.2 में दिए हुए समंकों के प्रसरण परिकलन कर सकते हैं तथा पूर्व प्राप्त परिणामों की जाँच कर सकते हैं।

जैसा इकाई 4 में माध्य के परिकलन के लिए किया गया था, X_i को $u_i = \frac{X_i - A}{h}$ में रूपांतरित करके प्रसरण के परिकलन को अत्यधिक सरल बनाया जा सकता है।

यह ध्यान दीजिए कि, क्योंकि

$$u_i - \overline{u} = \frac{X_i - A}{h} - \frac{\overline{X} - A}{h} = \frac{X_i - \overline{X}}{h} \text{ है, इसलिए हम}$$

$X_i - \overline{X} = h(u_i - \overline{u})$ लिख सकते हैं। अतः

$$\sigma_X^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\left\{h(u_i - \overline{u})\right\}^2 = h^2\sigma_u^2$$

जहाँ पर X_i मानों का प्रसरण σ_X^2 है तथा u_i मानों का प्रसरण σ_u^2 है। क्योंकि u के मानों का आकार छोटा होता है, इसलिए इनका प्रसरण परिकलित सरल होता है। X के मानों का प्रसरण उपरोक्त सूत्र के उपयोग द्वारा प्राप्त किया जा सकता है।

इस विधि द्वारा हम सारणी 4.2 में दिए हुए समंकों के प्रसरण परिकलन करते हैं।

यदि हम $u_i = \frac{X_i - 346.5}{24}$ लिखें तो u के मान $-3, -2, -1, 0, 1, 2,$ होंगे जिनकी क्रमशः बारंबारताएँ 1, 14, 16, 28, 26, 15 हैं।

$$u \text{ के मानों का माध्य } \bar{u} = \frac{-3\times1 - 2\times14 - 1\times16 + 0\times28 + 1\times26 + 2\times15}{100} = 0.09$$

$$u \text{ के मानों के वर्गों का माध्य } \frac{\sum_{i=1}^{n} f_i u_i^2}{N} = \frac{9\times1 + 4\times14 + 1\times16 + 0\times28 + 1\times26 + 4\times15}{100} = 1.67$$

इस प्रकार $\sigma^2 = 1.67 - (0.09)^2 = 1.6619$ तथा

$$\sigma_X^2 = (24)^2 . (1.6619) = 957.25 \text{ है।}$$

हालांकि X से u में रूपांतर, परिकलन की सुविधा के लिए किया गया है, लेकिन एक महत्त्वपूर्ण बात सामने आयी है। यह ध्यान दें कि $\sigma_u^2 = 1.6619$ है तथा $\sigma_X^2 = 957.25$ है, जहाँ एक साधारण रैखिक रूपांतरण द्वारा X से u को प्राप्त किया गया है, अर्थात् X के मूल एवं पैमाने में परिवर्तन द्वारा प्राप्त किया गया है। इस प्रकार के स्वभाविक उदाहरण, वज़न के लिए पौण्ड तथा किलोग्राम, द्रव्यों के आयतन के लिए गैलन तथा लीटर आदि हैं। क्योंकि एक किलोग्राम =2.2046 पौण्ड, इसलिए 5 किलोग्राम मानक विचलन, जब किलोग्राम में मापा जाय =11.023 पौण्ड मानक विचलन, जब पौण्ड में मापा जाय। इसी प्रकार क्योंकि एक लीटर = 0.22 गैलन, इसलिए 5 लीटर मानक विचलन, जब लीटर में मापा जाय = 1.1 गैलन मानक विचलन, जब गैलन में मापा जाय। अतः, जबकि प्रसरण तथा मानक विचलन द्वारा प्रेक्षणों के प्रसार को मापा जाना चाहिए, मापों की इकाई पर निर्भरता के कारण, इनसे कुछ अधिक प्राप्त नहीं किया जा सकता।

प्रेक्षणों के प्रसार के संदर्भ में केवल एक बहुत ही उपयोगी परिणाम, जो कि माध्य तथा मानक विचलन पर आधारित है तथा माप की इकाई पर निर्भर नहीं है, चेबाइचेव्य (Chebychev) द्वारा किया गया है (अनुभाग 5.3.1 देखें)।

बोध प्रश्न 1

1) प्रकीर्णन क्या होता है? इसको मापने की प्रचलित विधियाँ क्या हैं?

....

....

....

....

....

....

....

....

....

....

....

2) दस विद्यार्थियों की कक्षा में एक कमज़ोर विद्यार्थी के अंक अन्य विद्यार्थियों के औसत अंको से 25 कम हैं। यह दर्शाइए कि इन सभी विद्यार्थियों द्वारा प्राप्त अंकों का मानक विचलन कम से कम 7.5 है। यदि यह मानक विचलन वास्तव में 12.0 हो तो, कमज़ोर विद्यार्थी को छोड़कर बाकी विद्यार्थियों द्वारा प्राप्त अंकों का मानक विचलन ज्ञात कीजिए।

....

....

....

....

....

....

....

....

....

....

....

3) निम्नलिखित समंक, एक दुकानदार द्वारा उत्तरोत्तर 15 दिनों में अर्जित लाभ दर्शाते हैं:

116, 87, 91, 81, 98, 102, 97, 100, 105, 101, 115, 98, 102, 98, 93

समंकों का परिसर, माध्य से माध्य विचलन तथा मानक विचलन ज्ञात कीजिए।

....

....

....

....

....

....

....

....

....

....

....

4) निम्नलिखित समंकों से समान्तर माध्य, मानक विचलन तथा माध्य विचलन का परिकलन कीजिए :

अंक	4 - 5	6 - 7	8 - 9	10 - 11	12 - 13	14 - 15	योग
बारंबारता	4	10	20	15	8	3	60

....

....

5) एक विद्यार्थी द्वारा 100 प्रेक्षणों का माध्य तथा मानक विचलन क्रमशः 40 तथा 5.1 परिकलित किया गया। बाद में यह पता चला कि उसने गलती से एक प्रेक्षण को 40 के स्थान पर 50 लिया था। सही मानक विचलन ज्ञात कीजिए।

5.3 प्रकीर्णन तथा मानक विचलन में संबंध

आप यह अध्ययन कर चुके हैं कि यदि समंकों के समुच्चय में सभी मान आपने माध्य के आस-पास हैं तो इनमें प्रकीर्णन या प्रसरण की मात्रा कम होती है। इसके विपरीत उन समंक समुच्चयों में, जिनमें कुछ मान अपने माध्य से अधिक दूरी पर स्थित हैं, प्रकीर्णन की मात्रा अधिक होती है। चेबाइचेव्य (Chebychev) के प्रमेय द्वारा एक उपयोगी नियम, जो प्रकीर्णन तथा मानक विचलन में संबंध की व्याख्या करता है, निम्नलिखित है।

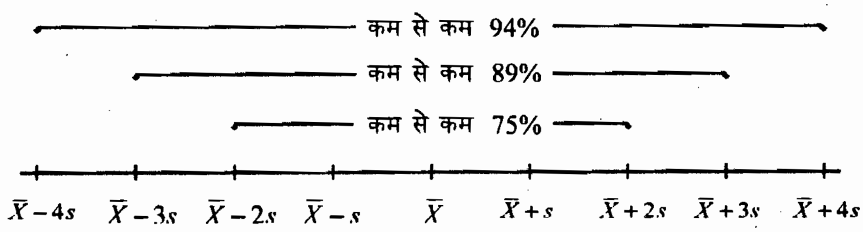
5.3.1 चेबाइचेव्य प्रमेय

किसी समंक समुच्चय तथा एक धनात्मक k (>1) के लिए, माध्य सें दोनों ओर k मानक

तक आनेवाले प्रेक्षणों का अनुपात अवश्य ही कम से कम $1-\frac{1}{k^2}$ होता है।

यह प्रमेय k के उन धनात्मक मानों जो एक से कम हैं के लिए उपयोगी नहीं है क्योंकि $1-\frac{1}{k^2}$ का अधिकतम मान शून्य हो सकता है। k के अन्य मानों के लिए न्यूनतम अनुपात सरलता से परिकलित किया जा सकता है। उदाहरण के लिए, माध्य से 1.5 मानक विचलन तक प्रेक्षणों का न्युनतम अनुपात अवश्यंभावी $1-\frac{1}{1.5^2}=0.556$ या 55.6% होगा।

चेबाइचेव्य प्रमेय पर आधारित समंकों का प्रकीर्णन निम्नलिखित चित्र में दर्शाया गया है। सारणी 4.1 में दिए हुए गृह आकार के समंकों के लिए $\bar{X}=3.74$ तथा $\sigma=1.4115$ है। यदि हम $k=2$ लेते हैं तो हम यह कह सकते हैं कि कम से कम $\left[\left(1-\frac{1}{2^2}\right)\times 100\right]=75\%$ गृहों का आकार $3.74\ \pm 2\times 1.4115$, अर्थात् 0.917 तथा 6.563 के बीच अवश्यम्भावी है।



चित्र 5.1

सारणी 4.2 में दिए हुए गृहों के खाद्य सामग्री पर औसत मासिक व्यय के बंटन के लिए, $\bar{Y}$ = 348.66 रुपये, $s=30.94$ रुपये है। अत: कम से कम 56% गृहों ($k = 1.5$) का औसत मासिक व्यय 302.25 तथा 395.07 रुपये के बीच अवश्य होगा।

5.3.2 बंटन का आकार

बहुत सी परिस्थितियों में, प्रणालीमूलक (methodological) अध्ययनों के लिए माध्य तथा प्रकीर्णन के मापों द्वारा बंटन की पर्याप्त व्याख्या हो जाती है। फिर भी, व्यावहारिक परिस्थितियों में, विशेषकर आय, व्यय, आर्थिक परिसंपत्तियों जैसे आर्थिक चरों के लिए, जो धनात्मक होते हैं, बंटन की व्याख्या के लिए अन्य माप भी उपयोग किये जाते है। इस प्रकार के दो माप विचरण गुणांक तथा सान्द्रता अनुपात हैं। आर्थिक चरों के बंटन की असमानताओं के आवश्यक माप के रूप में इन मापों का अध्ययन हम अब करेंगे।

5.3.3 विचरण गुणांक

आइए, हम दो गाँवों में गृहों की आर्थिक स्थिति की तुलना करने का प्रयास करें। इन दोनों गाँवों में गृहों द्वारा मासिक कैलोरी (calorie) अंतर्ग्रहण की संक्षेपण संख्याएँ निम्नलिखित हैं :

	गाँव	
	क	ख
गृहों की संख्या (n)	817	561
माध्य कैलोरी अंतर्ग्रहण ($\bar{x}$)	2417	2235
कैलोरी अंतर्ग्रहण का मानक विचलन (σ)	418	232

हमारा प्रश्न यह ज्ञात करना है कि कौन से गाँव में अंतर्ग्रहण की दृष्टि से असमानता अधिक है? गाँव 'ख' की तुलना में माध्य कैलोरी अंतर्ग्रहण गाँव 'क' में अधिक है, तथा इसका मानक विचलन तथा गृहों की संख्या भी अधिक है। वास्तव में गाँव 'क' में, गाँव 'ख' की तुलना में अधिक संख्या

में गरीब गृह हो सकते हैं और इस प्रकार गाँव 'क' के गृहों में अधिक असमानताएँ हो सकती हैं। इन असमानताओं की मात्रा के माप के एक सूचकांक को विचरण गुणांक कहते हैं। इसकी परिभाषा इस प्रकार है :

$$\text{विचरण गुणांक } (C.V.) = \frac{\sigma}{\overline{X}} \times 100$$

चूँकि σ तथा $\overline{X}$ की इकाई समान होती है, इसलिए विचरण गुणांक माप की इकाई से मुक्त होता है तथा मापन इकाई के चयन से प्रभावित नहीं होता।

$$\text{गाँव 'क' के लिए विचरण गुणांक} = \frac{418}{2417} \times 100 = 17.29 \text{ तथा}$$

$$\text{गाँव 'ख' के लिए विचरण गुणांक} = \frac{232}{2235} \times 100 = 10.38 \text{ है।}$$

क्योंकि गाँव 'क' का गुणांक, गाँव 'ख' के विचरण गुणांक की तुलना में अधिक है। गाँव 'क' में असमानता अधिक हैं।

असमानताओं के परिमाण की तुलना के लिए हम $\frac{17.29 - 10.38}{10.38} \times 100 = 66.57$ परिकलित करते हैं। जिसका अर्थ यह है कि गाँव 'ख' की तुलना में, गाँव 'क' में असमानताएँ 65.57% अधिक हैं।

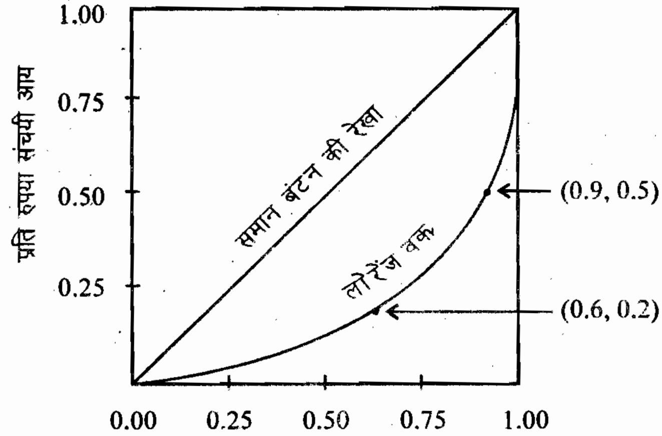
5.3.4 सान्द्रता अनुपात

उपरोक्त अनुभाग में हमने प्रत्येक गाँव में असमानता की मात्रा का अध्ययन किये बिना, दो गाँवों में विद्यमान असमानताओं की तुलना की है। यदि किसी बंटन का दायाँ छोर लंबा हो तो यह इस बात को व्यक्त करता है कि बंटन में कुछ व्यक्तियों के पास अधिक हिस्सा है अर्थात् अधिक जनसंख्या का हिस्सा बहुत कम है। इसको समझने के लिए हम एक काल्पनिक अर्थव्यवस्था में आय के वितरण का अध्ययन करते हैं। मान लीजिए कि अर्थव्यवस्था में व्यक्तियों के तीन वर्ग – उच्च वर्ग, मध्य वर्ग तथा निम्न वर्ग हैं तथा इन वर्गों में जनसंख्या क्रमशः 10%, 30% तथा 60% अंश है। मान लीजिए कि निम्न वर्ग को राष्ट्रीय आय का 20%, मध्य वर्ग को 30% तथा उच्च वर्ग को बाकी 50% प्राप्त होता है। इन समंकों को प्रतिशत संचयी बारंबारता बंटन के रूप में प्रस्तुत किया जा सकता है। अतः 60% जनसंख्या के पास कुल आय का 20% है, निम्न 90% के पास कुल आय का (20 + 30) = 50% है तथा स्पष्टतः 100% जनसंख्या के आय का 100% है। यदि हम एक आलेख पत्र के क्षैतिज अक्ष पर प्रतिशत संचयी बारंबारता तथा उर्ध्वाधर अक्ष पर प्रतिशत संचयी आय को लेकर (0, 0), (60, 20), (90, 50) तथा (100, 100) को अंकित करें तो इन बिन्दुओं को मिलाने वाली वक्र को सान्द्रता वक्र या लारेंज वक्र (Lorenz Curve) कहते हैं। बिन्दु (0, 0) तथा (100, 100) को मिलाने वाली सरल रेखा को समान बंटन की रेखा कहते हैं। यह रेखा इस बात को व्यक्त करती है कि आय में अंश का अनुपात तथा इसको प्राप्त करने वाले व्यक्तियों का अनुपात बराबर है। समान बंटन की रेखा तथा सान्द्रता वक्र के बीच क्षेत्रफल को सान्द्रता का क्षेत्रफल (area of concentration) कहते हैं। यह सान्द्रता की कोटि का सूचक होता है। जितना क्षेत्रफल अधिक होगा उतनी ही सान्द्रता अधिक होगी।

असमानता का गुणांक

हम, ऊपर दिये गये बिन्दुओं के निर्देशांक, प्रतिशत के रूप में न लेकर प्रति इकाई के रूप में ले लेते हैं। अतः ये निर्देशांक (0, 0), (0.60, 0.20), (0.90, 0.50) तथा (1.00, 1.00) के रूप में लिखे जाते हैं। तत्पश्चात् आय बंटन असमानता गुणांक को सान्द्रता क्षेत्रफल ÷ त्रिभुज का क्षेत्रफल, के रूप में परिभाषित किया जाता है। क्योंकि त्रिभुज का क्षेत्रफल 0.5 (क्योंकि

½ ×1 ×.1 = 0.5) है, तो विभिन्न बिन्दुओं के निर्देशांक प्रति इकाई लेने पर, असमानता गुणांक, सान्द्रता क्षेत्रफल का दुगुना होता है।



चित्र : 5.1 लॉरेंज वक्र

बोध प्रश्न 2

1) स्विटज़रलैंड में 1968 से 1980 तक अशोधित जन्म दर प्रति 1000 व्यक्तियों की संख्याएँ निम्नलिखित हैं :

अशोधित जन्म दर (*X*) : 17.1, 16.5, 15.8, 15.2, 14.3, 13.6, 12.9, 12.3, 11.7, 11.5, 11.3, 11.3, 11.6

प्रसरण, मानक विचलन तथा विचरण गुणांक परिकलित कीजिए।

....

....

....

....

....

....

....

....

....

....

2) निम्नलिखित सारणी में महिला शिक्षकों की आयु (जैसा कि अभिलेखों में प्रकाशित है), का बंटन दिया हुआ है :

आयु वर्ग (वर्षों में)	महिला शिक्षकों की संख्या
15 - 19	3
20 - 24	13
25 - 29	21
30 - 34	15
35 - 39	5
40 - 44	4
45 - 49	2

i) विचरण गुणांक, तथा

ii) 26 से 33 वर्ष की आयु के बीच शिक्षकों की संख्या का परिकलन कीजिए।

....

....

....

....

....

....

....

....

....

....

....

....

....

....

5.4 सारांश

इस इकाई में आपने प्रकीर्णन के मापों के बारे में अध्ययन किया है। प्रकीर्णन के अति महत्त्वपूर्ण माप प्रसरण, मानक विचलन तथा सान्द्रता अनुपात हैं। आपने दोनों प्रकार के समंकों (वर्गीकृत एवं अवर्गीकृत) के लिए इन मापों का परिकलन करना भी सीखा है। जब दो बंटनों के माध्य भिन्न हों या इनकी मापन इकाइयाँ भिन्न हों तो, इनके प्रकीर्णन की तुलना के लिए विचरण गुणांक का उपयोग किया जाता है।

5.5 शब्दावली

विचरण गुणांक : यह प्रकीर्णन का सापेक्षिक माप है जोकि मापन इकाई से स्वतंत्र होता है। इसके विपरीत मानक विचलन, प्रकीर्णन का निरपेक्ष माप है।

माध्य विचलन : यह प्रेक्षणों के माध्य या अन्य निर्धारित माप (जैसे माध्यिका, बहुलक) से निरपेक्ष विचलनों का सामांतर माध्य होता है।

परिसर : यह प्रेक्षणों के अधिकतम तथा न्यूनतम प्रेक्षणों का अंतर होता है।

प्रसरण : यह प्रेक्षणों के माध्य से विचलनों के वर्गों का समान्तर माध्य होता है।

मानक विचलन : यह प्रसरण का धनात्मक वर्गमूल होता है।

5.6 कुछ उपयोगी पुस्तकें

Elhance, D. N. and V. Elhance, 1988, *Fundamentals of Statistics,* Kitab Mahal, Allahabad.

Nagar, A. L. and R. K. Dass, 1983, *Basic Statistics*, Oxford University Press, Delhi.

Mansfield, E., 1991, *Statistics for Business and Economics : Methods and Applications*, W. W. Norton and Co.

Yule, G. U. and M. G. Kendall, 1991, *An Introduction to the Theory of Statistics*, Universal Books, Delhi.

5.7 बोध प्रश्नों के उत्तर या संकेत

बोध प्रश्न 1

1) आप स्वयं कीजिए
2) 9.9
3) 35, 6.46, 8.85
4) 9.23, 2.49, 2.03
5) 5.0

बोध प्रश्न 2

1) 4.085, 2.021, 15.004 %
2) 23.47%, 25 (पूर्णांकित संख्या)

5.8 पारिभाषिक शब्दावली

अंत:चतुर्थक परिसर	:	interquartile range
प्रसरण	:	variance
माध्य विचलन	:	mean deviation
मानक विचलन	:	standard deviation
विचरण गुणांक	:	coefficient of variation
सान्द्रता अनुपात	:	concentration ratio
परिसर	:	range

इकाई 6 विषमता तथा प्रथुशीर्षत्व के माप

इकाई की रूपरेखा

6.0 उद्देश्य

इस इकाई के अध्ययन के पश्चात् आप :

- सम्मित तथा विषम बंटनों में भेद कर सकेंगे;
- एक बंटन में विषमता के माप के लिए विभिन्न गुणांकों का परिकलन कर सकेंगे;
- चपटे, सामान्य तथा नुकीले शीर्ष वाले बंटनों में भेद कर सकेंगे; और
- प्रथुशीर्षत्व के गुणांक का परिकलन कर सकेंगे।

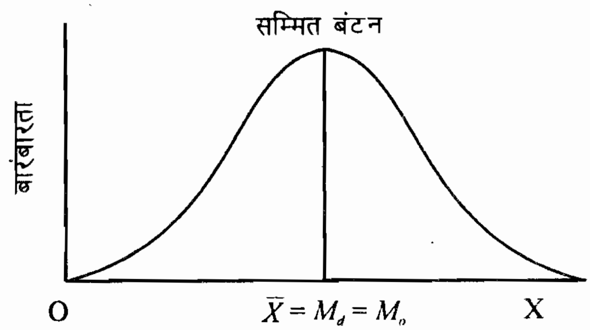
6.1 प्रस्तावना

इस इकाई में आप विभिन्न प्रकार के आकार वाले बंटनों में भेद करने की विधियों का अध्ययन करेंगे। यह एक विचर समंकों के संक्षेपण संबंधी अंतिम इकाई है। इस इकाई में आपको विषमता तथा प्रथुशीर्षत्व की अवधारणाओं से परिचित कराया जाएगा। क्योंकि केन्द्रीय प्रवृत्ति तथा प्रकीर्णन के मापों द्वारा एक बंटन की पूर्ण व्याख्या नहीं हो पाती, इसलिए, विषमता तथा प्रथुशीर्षत्व मापों के अध्ययन की आवश्यकता होती है। ऐसे बंटनों, जिनकी प्रकृति तथा बनावट बिलकुल भिन्न हों तथा केन्द्रीय प्रवृत्ति तथा प्रकीर्णन के माप समान हों, पाना संभव है। अत: केन्द्रीय प्रवृत्ति तथा प्रकीर्णन के मापों के अतिरिक्त कुछ और मापों की आवश्यकता महसूस होती है। परिणामत:, इस इकाई में हम दो मापों, अर्थात् विषमता तथा प्रथुशीर्षत्व के मापों का विवेचन करेंगे।

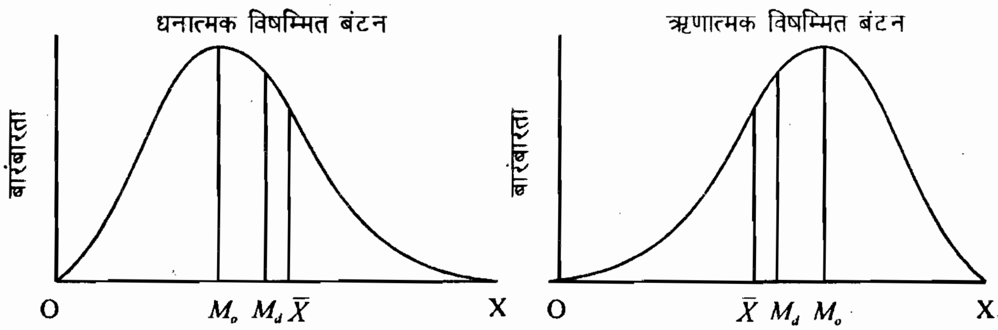
6.2 विषमता की अवधारणा

एक बंटन में सम्मितता की कमी को विषमता कहते हैं। एक सम्मित बंटन में माध्य, माध्यिका तथा बहुलक आपस में बराबर होते हैं तथा माध्य पर भुजमान (ordinate) बंटन को ऐसे दो भागों में

विभाजित करता है कि एक भाग दूसरे का दर्पण प्रतिरूप होता है चित्र (6.1)। यदि इस प्रकार के बंटन में कुछ बड़े (छोटे) आकार के प्रेक्षण सम्मिलत कर दिये जाएँ तो इसका दायाँ (बायाँ) सिरा लंबा हो जाता है। इस प्रकार के प्रेक्षणों को चरम प्रेक्षण कहते हैं। बंटन के दाईं ओर चरम प्रेक्षणों की विद्यमानता इसे धनात्मक विषम्मित बना देती है तथा केन्द्रीय प्रवृत्ति के तीन मापों में भिन्नता उत्पन्न हो जाती है। धनात्मक विषमता बंटन में माध्य > माध्यिका > बहुलक होता है। इसके विपरीत बंटन के बाईं ओर चरम प्रेक्षणों की विद्यमानता इसे ऋणात्मक विषम्मित बंटन बना देती है तथा माध्य < माध्यिका < बहुलक होता है। धनात्मक विषम्मित तथा ऋणात्मक विषम्मित बंटनों को चित्र 6.2 में दर्शाया गया है।



चित्र 6.1



चित्र 6.2

विषमता की दिशा तथा परिमाण का माप विभिन्न प्रकार से किया जा सकता है। इस इकाई में हम विषमता के चार मापों का विवेचन करेंगे।

6.2.1 कार्ल पीयरसन का विषमता गुणांक

चित्र 6.2 में आपने देखा कि एक विषम्मित बंटन में माध्य, माध्यिका तथा बहुलक बराबर नहीं होते। कार्ल पीयरसन का विषमता गुणांक, एक विषम्मित बंटन के माध्य के बहुलक से विचलन पर आधारित है।

क्योंकि एक सम्मित बंटन में माध्य = बहुलक होता है। इसलिए (माध्य – बहुलक) को विषमता का निरपेक्ष माप माना जा सकता है। एक बंटन की विषमता का निरपेक्ष माप, मापन इकाई पर निर्भर होता है। उदाहरण के लिए यदि माध्य = 2.45 मीटर तथा बहुलक = 2.14 मीटर है, तो विषमता

का निरपेक्ष माप 2.45 – 2.14 = 0.31 मीटर होगा। इसी बंटन के लिए यदि हम मापन इकाई सेंटीमीटर ले लें, तो विषमता का निरपेक्ष माप 245 – 214 = 31 सेंटीमीटर होगा। इस समस्या के समाधान के लिए पीयरसन द्वारा विषमता का तुलनात्मक माप परिभाषित किया गया।

एक तुलनात्मक माप मापन इकाई से स्वतंत्र होता है। तुलनात्मक माप को गुणांक भी कहते हैं। कार्ल पीयरसन का विषमता गुणांक S_k निम्नलिखित है :

$$S_k = \frac{\text{माध्य} - \text{बहुलक}}{\text{मानक विचलन}}$$

S_k का चिह्न तथा आकार, क्रमशः विषमता की दिशा तथा परिमाण की जानकारी प्रदान करते हैं। यदि $S_k > 0$ है तो बंटन धनात्मक विषम्मित तथा यदि $S_k < 0$ है तो बंटन ऋणात्मक विषम्मित होता है।

उपरोक्त विवेचन में हमने देखा कि S_k का मान बहुलक पर आधारित है। यदि किसी बंटन में बहुलक परिभाषित नहीं है तो S_k का मान ज्ञात करना संभव नहीं है। इस परिस्थिति में हम माध्य, माध्यिका तथा बहुलक के बीच प्रयोगाश्रित संबंध के उपयोग द्वारा S_k का मान ज्ञात कर सकते हैं। प्रयोगाश्रित संबंध के अनुसार, एक मामूली से विषम्मित बंटन में

माध्य – बहुलक ≈ 3 (माध्य – माध्यिका) होता है।

अतः, माध्यिका के उपयोग होने पर कार्ल पीयरसन का विषमता गुणांक निम्नलिखित है :

$$S_k = \frac{3\left(\text{माध्य} - \text{माध्यिका}\right)}{\text{मानक विचलन}}$$

उदाहरण 6.1 :

निम्नलिखित समंकों से कार्ल पीयरसन का विषमता गुणांक परिकलित कीजिए :

सारणी 6.1

ऊँचाई (इंचों में)	व्यक्तियों की संख्या
58	10
59	18
60	30
61	42
62	35
63	28
64	16
65	8

माध्य तथा मानक विचलन के परिकलन की सारणी

ऊँचाई (X)	$u = X - 61$	व्यक्तियों की संख्या (f)	fu	fu^2
58	−3	10	−30	90
59	−2	18	−36	72
60	−1	30	−30	30
61	0	42	0	0
62	1	35	35	35
63	2	28	56	112
64	3	16	48	144
65	4	8	32	128
योग		187	75	611

माध्य $= 61 + \frac{75}{187} = 61.4$

मानक विचलन $= \sqrt{\frac{611}{187} - \left(\frac{75}{187}\right)^2} = 1.76$

बहुलक ज्ञात करने के लिए हम ध्यान देते हैं कि ऊँचाई एक संतत चर है। ऊँचाई के माप यह मान कर (उदाहरण के लिए) किए गए हैं कि वह माप जो 58 से अधिक लेकिन 58.5 से कम हैं, को 58 इंच तथा वह माप जो 58.5 से अधिक लेकिन 59 से कम है, को 59 इंच लिया जाना है। अत: दिए हुए समंकों को इस प्रकार लिखते हैं।

ऊँचाई (इंचों में)	व्यक्तियों की संख्या
57.5 - 58.5	10
58.5 - 59.5	18
59.5 - 60.5	30
60.5 - 61.5	42
61.5 - 62.5	35
62.5 - 63.5	28
63.5 - 64.5	16
64.5 - 65.5	8

निरीक्षण द्वारा, बहुलक वर्ग 60.5 – 61.5 है।

अत: $l_m = 60.5$, $\Delta_1 = 42 - 30 = 12, \Delta_2 = 42 - 35 = 7$ तथा $h = 1$

$\therefore$ बहुलक $= 60.5 + \frac{12}{12+7} \times 1 = 61.13$

अत: कार्ल पीयरसन का विषमता गणांक $S_k = \frac{61.4 - 61.13}{1.76} = 0.153$

इस प्रकार, बंटन जरा सा धनात्मक विषम्मित है।

6.2.2 बाउले का विषमता माप

यह माप चतुर्थकों पर आधारित है। एक सम्मित बंटन में Q_1 तथा Q_3, माध्य से समान दूरी पर होते हैं। इस प्रकार $(Q_3 - M_d) - (M_d - Q_1)$ को विषमता का निरपेक्ष माप माना जा सकता है।

विषमता का तुलनात्मक माप, जिसे बाउले का विषमता गुणांक (S_Q) भी कहते हैं, निम्नलिखित है :

$$S_Q = \frac{(Q_3 - M_d) - (M_d - Q_1)}{(Q_3 - M_d) + (M_d - Q_1)}$$

$$= \frac{Q - 2M_d + Q_1}{Q_3 - Q_1}$$

सारणी 6.1 में दिए हुए ऊँचाई समंकों के बाउले गुणांक का परिकलन निम्नलिखित है।

ऊँचाई (इंचों में)	व्यक्तियों की संख्या (f)	संचयी बारंबारता
57.5 - 58.5	10	10
58.5 - 59.5	18	28
59.5 - 60.5	30	58
60.5 - 61.5	42	100
61.5 - 62.5	35	135
62.5 - 63.5	28	163
63.5 - 64.5	16	179
64.5 - 65.5	8	187

Q_1 का परिकलन :

क्योंकि $\frac{N}{4} = 46.75$, प्रथम चतुर्थक वर्ग $59.5 - 60.5$ होगा। इसलिए

$l_{Q_1} = 59.5,\ C = 28,\ f_{Q_1} = 30$ तथा $h = 1$

$$\therefore \quad Q_1 = 59.5 + \frac{46.75 - 28}{30} \times 1 = 60.125$$

माध्यिका $M_d\ (Q_2)$ का परिकलन :

क्योंकि $\frac{N}{2} = 935$, माध्यिका वर्ग $60.5 - 61.5$ होगा। इसलिए

$l_m = 60.5,\ C = 58, f_m = 42$ तथा $h = 1$

$$\therefore \quad M_d = 60.5 + \frac{93.5 - 58}{42} \times 1 = 61.345$$

Q_3 का परिकलन :

क्योंकि $\frac{3N}{4} = 140.25$, तृतीय चतुर्थक वर्ग 62.5 – 63.5 होगा। इसलिए

$$l_{Q_3} = 62.5, C = 135, = 28 \text{ तथा } h = 1$$

$$\therefore \quad Q_3 = 62.5 + \frac{140.25 - 135}{28} \times 1 = 62.688$$

अतः बाउले गुणांक $S_Q = \frac{62.688 - 2 \times 61.345 + 60.125}{62.688 - 60.125} = 0.048$

6.2.3 कैली का विषमता माप

क्योंकि बाउले के विषमता गुणांक में बंटन के दोनों सिरों पर 25% प्रेक्षण छूट जाते हैं, यह माप मध्य 50% प्रेक्षणों पर आधारित होता है। बाउले गुणांक के सुधार के रूप में, कैली द्वारा P_{10} तथा P_{90} पर आधारित एक माप का सुझाव दिया गया, जिससे बंटन के प्रत्येक सिरे पर केवल 10% प्रेक्षण ही छूट सकें।

कैली का विषमता गुणांक S_P, निम्नलिखित है:

$$S_P = \frac{(P_{90} - P_{50}) - (P_{50} - P_{10})}{(P_{90} - P_{50}) + (P_{50} - P_{10})}$$

$$= \frac{P_{90} - 2.P_{50} + P_{10}}{P_{90} - P_{10}}$$

ध्यान दीजिए कि $P_{50} = M_d$ (माध्यिका) होता है।

सारणी 6.1 में दिए गए समंकों के लिए S_P का परिकलन निम्नलिखित है:

P_{10} का परिकलन :

क्योंकि $\frac{10N}{100} = \frac{10 \times 187}{100} = 18.7$, 10वाँ शतमक वर्ग 58.5 – 59.5 में होगा। इसलिए

$$l_{P_{10}} = 58.5, C = 10, f_{P_{10}} = 18 \text{ तथा } h = 1$$

$$\therefore \quad P_{10} = 58.5 + \frac{18.7 - 10}{18} \times 1 = 58.983$$

P_{90} का परिकलन :

क्योंकि $\frac{90N}{100} = \frac{90 \times 187}{100} = 168.3$, 90वाँ शतमक वर्ग 63.5 – 64.5 में होगा। इसलिए

$$l_{P_{90}} = 63.5, C = 163, f_{P_{90}} = 16 \text{ तथा } h = 1$$

$$\therefore \quad P_{90} = 63.5 + \frac{168.3 - 163}{16} \times 1 = 63.831$$

$$\text{अत: कैली का गुणांक } S_P = \frac{63.831-2\times 61.345+58.983}{63.831-58.983}=0.026$$

यहाँ पर यह बताना आवश्यक है कि चाहे गुणांक S_k, S_Q तथा S_P तुलनीय नहीं हैं, लेकिन विषमता की अनुपस्थिति में प्रत्येक का मान शून्य होता है।

बोध प्रश्न 1

1) निम्नलिखित समंकों द्वारा कार्ल पीयरसन का विषमता गुणांक परिकलित कीजिए :

प्रतिदिन व्यय (रुपयों में)	0-20	20-40	40-60	60-80	80-100
परिवारों की संख्या	13	25	27	19	16

....

....

....

....

....

....

....

....

....

....

2) निम्नलिखित संख्याएँ, 285 कम्पनियों के पूँजी आकार से संबंधित हैं :

पूँजी (लाख रुपये में)	1-5	6-10	11-15	16-20	21-25	26-30	31-35	योग
कम्पनियों की संख्या	20	27	29	38	48	53	70	285

बाउले तथा कैली के विषमता गुणांकों का परिकलन तथा परिणामों का विवेचन कीजिए।

....

....

....

....

....

....

....

....

....

....

....

3) एक बारंबारता बंटन के लिए निम्नलिखित माप परिकलित किए गए :

माध्य = 50, विचरण गुणांक = 35% तथा
कार्ल पियरसन का विषमता गुणांक = – 0.25

बंटन का मानक विचलन, बहुलक तथा माध्यिका परिकलित कीजिए।

....

....

....

....

....

....

....

....

....

....

....

....

6.3 परिघात

एक बंटन का μ_r द्वारा सूचित r वाँ परिघात निम्नलिखित होता है।

$$\mu_r = \frac{1}{N}\sum_{i=1}^{n} f_i\left(X_i - \overline{X}\right)^r \quad \text{जहाँ पर } r = 0, 1, 2, 3, 4, \ldots\ldots\ldots \text{ है।}$$

इस प्रकार माध्य से r वाँ परिघात, प्रेक्षणों के उनके माध्य से विचलनों कि r वीं घात का माध्य होता है। विस्तृत रूप में

$$\text{यदि } r = 0, \text{ तो } \mu_0 = \frac{1}{N}\sum_{i=1}^{n} f_i\left(X_i - \overline{X}\right)^0 = 1,$$

$$\text{यदि } r = 1, \text{ तो } \mu_1 = \frac{1}{N}\sum_{i=1}^{n} f_i\left(X_i - \overline{X}\right) = 0,$$

$$\text{यदि } r = 2, \text{ तो } \mu_2 = \frac{1}{N}\sum_{i=1}^{n} f_i\left(X_i - \overline{X}\right)^2 = \sigma^2,$$

$$\text{यदि } r = 3, \text{ तो } \mu_3 = \frac{1}{N}\sum_{i=1}^{n} f_i\left(X_i - \overline{X}\right)^3 \text{ इत्यादि।}$$

इन परिघातों को *केन्द्रीय परिघात* भी कहते हैं।

इसके अतिरिक्त हम किसी मनमाने माध्य से अपरिष्कृत परिघातों को भी परिभाषित कर सकते हैं।

मान लिया कि A एक मनमाना माध्य है, तब A से r वाँ परिघात इस प्रकार परिभाषित किया जाता है।

$$\mu'_r = \frac{1}{N}\sum_{i=1}^{n} f_i(X_i - A)^r \text{ , } r = 0, 1, 2, 3, \ldots\ldots$$

जब $A = 0$ हो तो विभिन्न परिघात मूल बिन्दु से होते हैं।

परिघात पर आधारित विषमता माप

यह माप इस विशेषता पर आधारित है कि एक सम्मित बंटन में, सभी विषम क्रमित केन्द्रीय परिघात शून्य होते हैं।

क्योंकि प्रत्येक बंटन के लिए $\mu_1 = 0$ होता है, इसलिए निम्नतम क्रमिक केन्द्रीय परिघात जोकि विषमता का निरपेक्ष माप हो सकता है, μ_3 लिया जाता है।

इसके अतिरिक्त विषमता गुणांक, जो कि मापन इकाई से स्वतंत्र होता है, α_3 द्वारा सूचित किया जाता है।

$$\alpha_3 = \frac{\mu_3}{\sigma^3} = \pm\sqrt{\beta_1} = \gamma_1$$

जहाँ पर β_1 तथा γ_1 को क्रमशः प्रथम बीटा तथा प्रथम गामा गुणांक कहते हैं। जैसा कि आप अगले अनुभाग में पढ़ेंगे, β_2 प्रथुशीर्षत्व का माप होता है।

प्रायः विषमता को $\beta_1 = \frac{\mu_3^2}{\mu_2^3}$, के रूप में मापा जाता है, जहाँ पर विषमता का चिह्न μ_3 के चिह्न द्वारा निर्धारित होता है।

उदाहरण 6.2 :

निम्नलिखित समंकों द्वारा परिघात विषमता गुणांक (β_1) का परिकलन कीजिए।

प्राप्त अंक	0-10	10-20	20-30	30-40	40-50	50-60	60-70
बरंबारता	6	12	22	24	16	12	8

माध्य, मानक विचलन तथा μ_3 के परिकलन की सारणी

वर्ग अंतराल	बारंबारता (f)	मध्य मान (X)	$u = \frac{X-35}{10}$	fu	fu^2	fu^3
0 - 10	6	5	–3	–18	54	–162
10 - 20	12	15	–2	–24	48	–96
20 - 30	22	25	–1	–22	22	–22
30 - 40	24	35	0	0	0	0
40 - 50	16	45	1	16	16	16
50 - 60	12	55	2	24	48	96
60 - 70	8	65	3	24	72	216
योग	100			0	260	48

क्योंकि $\sum_{\cdot} fu = 0$ है, इसलिए बंटन का माध्य 35 होगा।

इसके अतिरिक्त द्वितीय परिघात, μ_2, प्रसरण (σ^2) के बराबर, तथा इसका घनात्मक वर्गमूल मानक विचलन (σ) के बराबर होगा।

$$\mu_2 = \frac{260}{100} \times 100 = 260\text{, तथा}$$

$$\text{मानक विचलन } \sigma = \sqrt{260} = 16.12$$

$$\mu_3 = \frac{48}{100} \times 1000 = 480$$

अत: $\beta_1 = \dfrac{(480)^2}{(260)^3} = 0.01$ है।

क्योंकि μ_3 का चिह्न धनात्मक तथा β_1 का मान छोटा है, दिया हुआ बंटन ज़रा सा धनात्मक विषम्मित है।

ऊपर दिए गये उदाहरण में यदि बंटन का माध्य 35 की तरह सुविधाजनक संख्या नहीं है तो विभिन्न परिघातों का परिकलन कार्य कठिन हो सकता है। विकल्पत:, पहले हम अपरिष्कृत परिघातों का परिकलन करके उनको केन्द्रीय परिघातों में परिवर्तित कर सकते हैं।

अपरिष्कृत परिघातों का केन्द्रीय परिघातों में परिर्वतन

केन्द्रीय परिघात μ_r को इस प्रकार लिखा जा सकता है :

$$\mu_r = \frac{1}{N}\sum_{i=1}^{n} f_i\left(X_i - \bar{X}\right)^r = \frac{1}{N}\sum_{i=1}^{n} f_i\left[(X_i - A) - \left(\bar{X} - A\right)\right]^r$$

$$= \frac{1}{N}\sum_{i=1}^{n} f_i\left[(X_i - A) - \mu_1'\right]^r \quad \left(\text{क्योंकि } \mu_1' = \frac{1}{N}\sum f_i(X_i - A) = \bar{X} - A\right)$$

बाइनोमियल प्रमेय द्वारा विस्तार करने पर

$$= \frac{1}{N}\sum_{i=1}^{n} f_i\left[{}^rC_0(X_i - A)^r \mu_1'^0 - {}^rC_1(X_i - A)^{r-1}\mu_1' + {}^rC_2(X_i - A)^{r-2}\mu_1'^2 - \cdots\cdots\right]$$

$$= \frac{1}{N}\sum_{i=1}^{n} f_i(X_i - A)^r - {}^rC_1\frac{1}{N}\sum_{i=1}^{n} f_i(X_i - A)^{r-1}\mu_1' + {}^rC_2\frac{1}{N}\sum_{i=1}^{n} f_i(X_i - A)^r \mu_1'^2 - \cdots\cdots$$

उपरोक्त को इस प्रकार भी लिखा जा सकता है।

$$\mu_r = \mu_r' - {}^rC_1\mu_{r-1}'\mu_1' + {}^rC_2\mu_{r-2}'\mu_1'^2 - {}^rC_3\mu_{r-3}'\mu_1'^3 + \cdots\cdots$$

r का मान 2, 3, 4 इत्यादि लेने पर

$$\mu_2 = \mu_2' - {}^2C_1\mu_1'^2 + {}^2C_2\mu_0'\mu_1'^2 = \mu_2' - \mu_1'^2 \quad (\text{क्योंकि } \mu'_0 = 1)$$

$$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 3\mu_1'^3 - \mu_1'^3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3$$

$$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 4\mu_1'^4 + \mu_1'^4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4$$

उदाहरण 6.3 :

निम्नलिखित समंकों के लिए माध्य के अंतर्गत पहले चार परिघातों का परिकलन कीजिए :

वर्ग अंतराल	0-10	10-20	20-30	30-40
बारंबारता (f)	1	3	4	2

अपरिष्कृत परिघातों के परिकलन की सारणी ($A = 25$ लिया गया है)

वर्ग अंतराल	f	मध्य मान (X)	$u = \frac{X-25}{10}$	fu	fu^2	fu^3	fu^4
0 - 10	1	5	–2	–2	4	–8	16
10 - 20	3	15	–1	–3	3	– 3	3
20 - 30	4	25	0	0	0	0	0
30 - 40	2	35	1	2	2	2	2
योग	10			–3	9	–9	21

उपरोक्त सारणी द्वारा

$$\mu_1' = \frac{-3\times10}{10} = -3,$$

$$\mu_2' = \frac{9\times10^2}{10} = 90,$$

$$\mu_3' = \frac{-9\times10^3}{10} = -900 \text{ तथा}$$

$$\mu_4' = \frac{21\times10^4}{10} = 21000$$

माध्य से परिघात

$\mu_1 = 0$, (परिभाषा द्वारा)

$$\mu_2 = 90 - 9 = 81,$$

$$\mu_3 = -900 - 3\times90\times(-3) + 2\times(-3)^3 = -900 + 810 - 54 = -144 \text{ तथा}$$

$$\mu_4 = 21000 - 4\times(-900)\times(-3) + 6\times90\times(-3)^2 - 3\times(-3)^4$$

$$= 21000 - 10800 + 4860 - 243 = 14817$$

बोध प्रश्न 2

1) निम्नलिखित बंटन के लिए माध्य के अंतर्गत पहली चार परिघातों का परिकलन कीजिए। β_1 के परिकलन द्वारा विषमता की प्रकृति पर टिप्पणी कीजिए।

अंक	0 - 20	20 - 40	40 - 60	60 - 80	80 - 100
बरंबारता	8	28	35	17	12

....

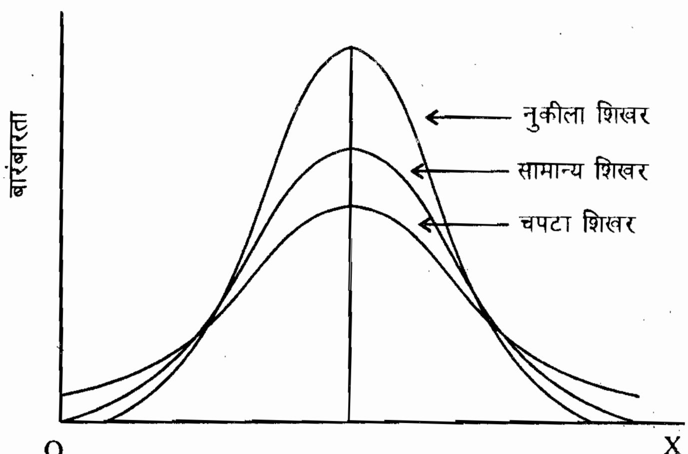
....

....

2) एक बंटन के लिए, मान 3 के अंतर्गत, पहले तीन परिघात क्रमशः 2, 10 तथा 30 हैं। $\bar{X}$, μ_2, μ_3 तथा β_1 का परिकलन कीजिए। विषमता की प्रकृति पर टिप्पणी कीजिए।

6.4 प्रथुशीर्षत्व की अवधारणा

बंटन के आकार के एक और माप को प्रथुशीर्षत्व कहते हैं। जबकि विषमता द्वारा बंटन की बारंबारता वक्र में सम्मितता की कमी को मापा जाता है, प्रथुशीर्षत्व द्वारा बारंबारता वक्र की शिखर के नुकीलेपन को मापा जाता है। विभिन्न बारंबारता वक्रों को, उनके शिखर के आकार के आधार पर, तीन वर्गों में विभाजित किया जा सकता है। इन आकारों को नुकीला शिखर (Leptokurtic), सामान्य शिखर (Mesokurtic) तथा चपटा शिखर (Platykurtic) कहते हैं, जैसा चित्र 6.3 में दर्शाया गया है।



चित्र 6.3

प्रथुशीर्षत्व का माप

प्रथुशीर्षत्व को कार्ल पीयरसन के गुणांक $\beta_2 = \frac{\mu_4}{\mu_2^2}$, द्वारा मापा जाता है। सामान्य शिखर वक्र के लिए β_2 का मान 3 होता है।

यदि $\beta_2 > 3$ हो तो वक्र का शिखर सामान्य से अधिक नुकीला होता है। इसी प्रकार $\beta_2 < 3$ होने पर वक्र को चपटा शिखर वक्र कहते हैं।

उदाहरण 6.4 :

एक बंटन के पहले चार केन्द्रीय परिघात क्रमश: 0, 2.5, 0.7 तथा 18.75 हैं। बंटन की विषमता तथा प्रथुशीर्षत्व की जाँच कीजिए।

विषमता की जाँच के लिए हम β_1 का परिकलन करते हैं।

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(0.7)^2}{(2.5)^3} = 0.031$$

क्योंकि $\mu_3 > 0$ तथा β_1 का मान छोटा है, अत: बंटन जरा सा धनात्मक विषम्मित है।

प्रथुशीर्षत्व को β_2 गुणांक द्वारा मापा जाता है।

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{18.75}{(2.5)^2} = 3.0$$

अत: बारंबारता वक्र एक सामान्य शिखर वक्र है।

बोध प्रश्न 3

1) निम्नलिखित समंकों द्वारा प्रथम चार केन्द्रीय परिघातों का परिकलन कीजिए। दोनों बीटा (β) गुणांक भी ज्ञात कीजिए।

मान	5	10	15	20	25	30	35
बारंबारता	8	15	20	32	23	17	5

....

....

....

....

....

....

....

....

....

....

2) एक बंटन के पहले चार परिघात क्रमशः 1, 4, 10 तथा 46 हैं। विषमता तथा प्रथुशीर्षत्व के परिघात गुणांकों का परिकलन करके बंटन की प्रकृति पर टिप्पणी कीजिए।

....

....

....

....

....

....

....

....

....

....

....

....

....

6.5 सारांश

इस इकाई में विषमता तथा प्रथुशीर्षत्व के बारे में अध्ययन किया गया है। इन दोनों अवधारणाओं का उपयोग एक बंटन के आकार के बारे में जानकारी प्राप्त करने के लिए किया गया है। बंटन में सम्मितता की कमी को विषमता कहते हैं। जबकि प्रथुशीर्षत्व किसी बंटन की बारंबारता वक्र के शिखर के नुकीलेपन का माप है।

6.6 शब्दावली

विषमता : सम्मितता से विचलन को विषमता कहते हैं।

***r* वीं कोटि की परिघात** : यह प्रेक्षणों के विचलनों की *r* वीं घात का सामान्तर माध्य होता है।

प्रथुशीर्षत्व गुणांक : यह एक बारंबारता वक्र के शिखर के नुकीलेपन का माप होता है।

6.7 कुछ उपयोगी पुस्तकें

Elhance, D. N. and V. Elhance, 1988, *Fundamentals of Statistics*, Kitab Mahal, Allahabad.

Nagar, A. L. and R. K. Dass, 1983, *Basic Statistics*, Oxford University Press, Delhi.

Mansfield, E., 1991, *Statistics for Business and Economics : Methods and Applications*, W. W. Norton and Co.

Yule, G. U. and M. G. Kendall, 1991, *An Introduction to the Theory of Statistics*, Universal Books, Delhi.

6.8 बोध प्रश्नों के उत्तर या संकेत

बोध प्रश्न 1

1) 0.237

2) – 0.12, – 0.243

3) 17.5, 54.38, 51.46

बोध प्रश्न 2

1) 0, 499.64, 2579.57, 589111.61, 0.053, विषमता धनात्मक है।

2) 5, 6, –14, 0.907; क्योंकि μ_3 ऋणात्मक है, बंटन ऋणात्मक विषम्मित है।

बोध प्रश्न 3

1) 0, 59.99, – 50.18, 8356.64, 0.012 (ऋणात्मक विषम्मित), 2.32 (चपटा शिखर)।

2) 0, 3; अत: बंटन सम्मित तथा सामान्य शिखर वाला है। ऐसे बंटन को प्रसामान्य बंटन भी कहते हैं।

6.9 पारिभाषिक शब्दावली

सम्मित बंटन	:	symmetrical distribution
परिघात	:	moments
प्रथुशीर्षत्व	:	kurtosis
विषमता	:	skewness
चरम प्रेक्षण	:	extreme observations

इकाई 7 द्विचर आंकड़ों की प्रस्तुति

इकाई की रूपरेखा

7.0 उद्देश्य

इस इकाई को पढ़ने के बाद, आप :

- विभिन्न प्रकार के चरों के बीच के अंतर को स्पष्ट कर सकेंगे;
- बारंबारता बंटनों के रूप में द्विचर आंकड़ों को प्रस्तुत कर सकेंगे; और
- सीमांत (marginal) और सप्रतिबंध (conditional) बंटनों की संकल्पनाओं को व्यक्त कर सकेंगे।

7.1 प्रस्तावना

द्विचर शब्द का प्रयोग उन परिस्थितियों की व्याख्या के लिए किया जाता है, जिसमें प्रत्येक व्यष्टि के दो अभिलक्षणों का माप किया जाता है। जैसे स्कूल में विद्यार्थियों की ऊँचाई (X_i) और भार (Y_i) संबंधी माप। इस मामले में पादांक (subscript) i संबद्ध विद्यार्थी को दर्शाता है। अतः उदाहरण के तौर पर; X_5, Y_5, पाँचवें विद्यार्थी की ऊँचाई और भार को दर्शाएगा।

दो अभिलक्षण, दो चरों द्वारा निरूपित किए जाते हैं। इन दो चरों के एक साथ मापे गए आंकड़ों को द्विचर आंकड़े (bivariate data) कहा जाता है। प्रत्येक व्यष्टि के प्रेक्षण, युग्म रूप में होते हैं, जिनमें प्रत्येक मान एक चर को निरूपित करता है। जैसे (X_1, Y_1), (X_2, Y_2),..., (X_n, Y_n)। इन द्विचर आंकड़ों जब अधिक संख्या में उपलब्ध हों तो इनका द्विधा सारणी के रूप में संक्षेपण आवश्यक हो जाता है। इस सारणी को *द्विचर बारंबारता बंटन* (bivariate frequency distribution) कहते हैं।

आइए अब पिछले अनुभाग में चर्चित विभिन्न प्रकार के चरों के लिए द्विचर बारंबारता बंटन बनाने की क्रियाविधि पर चर्चा करें।

7.2 विभिन्न प्रकार के चर

हम चर का प्रयोग कैसे करेंगे (जैसे आंकड़ों को प्रस्तुत करना या संक्षेपण आदि करना) यह मूल रूप से चर की प्रकृति पर निर्भर करता है। अब हम सांख्यिकीय उद्देश्यों के लिए तीन प्रकार के चर अर्थात नामिक (nominal), क्रमसूचक (ordinal) और संख्यात्मक (numerical) चरों के बीच के अंतर को स्पष्ट करेंगे।

नामिक चर वह है जिसके गुणात्मक मान होते हैं और जिनमें किसी भी प्रकार का क्रम आधारित संबंध नहीं होता। जैसे लिंग नामिक चर है तो केवल गुणात्मक मानों अर्थात स्त्री और पुरुषों से संबंधित है और 'स्त्री' और 'पुरुष' की स्थिति में कोई क्रमबद्धता नहीं है। गुणात्मक चर को गुण (attribute) भी कहते हैं। यहाँ, हम प्रेक्षणों को श्रेणियों में विभाजित कर सकते हैं। लेकिन, हम यह नहीं कह सकते कि एक श्रेणी दूसरी से उच्च है।

क्रमसूचक चर वह है जिसके गुणात्मक मान हैं और जिनमें क्रमबद्धता होती है। जैसे, उदाहरण के रूप में, शिक्षा क्रमसूचक चर है और जिसके साक्षर, निरक्षर जैसे गुणात्मक मान हैं और जिनमें साक्षर पर माध्यमिक से नीचे, उच्च माध्यमिक, स्नातक तथा स्नातकोत्तर श्रेणियां बन सकती हैं। यदि इसी संदर्भ में हम उच्च शैक्षिक स्तर की बात करें तो हमारा आशय उल्लिखित पहली श्रेणी से अंतिम श्रेणी की ओर बढ़ने से है।

संख्यात्मक चर वह है जिसके मात्रात्मक/परिमाणात्मक मान होते हैं। संख्यात्मक चर दो तरह के हो सकते हैं : असतत् (discrete) और सतत् (continuous)। असतत् चर वह है जो निश्चित वियुक्त बिंदुओं पर ही मान धारण करता है। जैसे किसी परिवार में बच्चों की संख्या असतत् चर है, जहां मान हैं 0, 1, 2,...। वियुक्त मानों की संख्या जिन्हें असतत् चर धारण कर सकता है, आवश्यक नहीं है कि वह *परिमित* (finite) हो। दूसरी ओर, सतत् चर, किसी अंतराल में कोई भी मान ले सकता है। जैसे, ऊँचाई सतत् चर है जो अंतराल में संकल्पनात्मक रूप से कोई भी मान ले सकता है, जैसे 0 से 200 सेंटीमीटर के बीच का कोई भी मान ।

7.3 नामिक और क्रमसूचक चरों की प्रस्तुति

एक चर वाले आंकड़ों के निरूपण की विधि का अध्ययन आप पहले कर चुके हैं। आइए अब दो चरों पर विचार करें। आइए किसी कॉलेज के विद्यार्थियों पर विचार करें और दो चरों का अर्थात प्रत्येक विद्यार्थी के लिंग और मातृभाषा का पता लगाएं। ध्यान दें कि विचाराधीन दोनों चर नामिक हैं : लिंग चर के दो मान, पुरुष (M) और महिला (F) और मातृभाषा के मान, जैसे हिन्दी (H), बंगाली (B) तमिल (T) और अन्य (O) है। इस प्रकार के आंकड़ों का समुच्चय इस प्रकार होगा;

(H,M), (B,M) (T,M), (O,M), (H,F), (B,F), (T,F), (O,F)

जहाँ (H,M) हिन्दी भाषी पुरुष को इंगित करता है। सूचना की गरिमा को बिना खोए, हम इन आंकड़ों का संक्षेपण कर सकते हैं। इसे हमने सारणी 7.1 में दर्शाया है।

सारणी 7.1 में दो चरों, मातृभाषा एवं लिंग का *संयुक्त बारंबारता बंटन* प्रस्तुत है। इस प्रकार की सारणी में (जहाँ पर चर नामिक, क्रमसूचक या संख्यात्मक हो सकते हैं), दो चरों

के स्तर या मानों के संचय की इस सारणी को "कोष्ठ" (cell) कहते हैं तथा इसकी बारंबारता को *कोष्ठ बारंबारता* कहा जाता है। उदाहरण के लिए आप सारणी 7.1 देख सकते हैं, जिसमें (पुरुष, तमिल) सारणी का एक कोष्ठ है और जिसकी कोष्ठ बारंबारता 367 है। इसका अर्थ है कि कॉलेज के 367 विद्यार्थी तमिल पुरुष हैं। प्रत्येक चर के अलग बंटन को दर्शाने की बजाए, संयुक्त बंटन, यहां और अधिक जानकारी प्रस्तुत करता है।

सारणी 7.1 : मातृभाषा और लिंग का द्विचर बारंबारता बंटन

मातृ भाषा	लिंग		कुल
	पुरुष	महिला	
हिन्दी	456	523	979
बंगाली	234	221	455
तमिल	367	387	754
अन्य	350	401	751
कुल	1407	1532	2939

प्रत्येक चर के पृथक बंटन को **सीमांत बंटन** कहते हैं। आप सारणी के अंतिम स्तम्भ में देखें जिसमें पुरुष और महिला बारंबारता का योग है : यह मातृभाषा का सीमांत बंटन है। इसी प्रकार सारणी की अंतिम पंक्ति में लिंग का सीमांत बंटन दिया हुआ है। संयुक्त बंटन द्वारा हमें चरों के बीच के संबंध के अध्ययन में सहायता मिलती है।

इस उदाहरण में हमारी रुचि यह हो सकती है कि क्या विभिन्न मातृभाषा समूहों में लिंग अनुपात भिन्न है। इसके लिए प्रत्येक भाषा के लिए पुरुष तथा महिला का अनुपात परिकलित किया जाता है। इस प्रकार के बंटनों को सप्रतिबंध बंटन (conditional distribution) कहते हैं। ये इस प्रकार प्राप्त किए जाते हैं; मातृभाषा हिन्दी के लिए : पुरुष $0.4658\left(=\frac{456}{979}\right)$; महिला $0.5342\left(=\frac{523}{979}\right)$। इसी तरह अन्य सप्रतिबंध बंटन हैं; बंगाली के लिए : 0.5143; 0.4857, तमिल के लिए 0.4867; 0.5133; अन्य के लिए: 0.4660; 0.5340।

आइए, अब 'लिंग के सप्रतिबंध बंटनों' की तुलना, 'लिंग के सीमांत बंटन' से करें। सारणी 7.1 की अंतिम पंक्ति से हम देखते हैं कि लिंग का सीमांत बंटन $0.4787\left(=\frac{1407}{2939}\right)$; $0.5213\left(=\frac{1532}{2939}\right)$ है। इसकी पहले से प्राप्त सप्रतिबंध बंटन से तुलना करें। यह विविध मातृभाषाओं वाले समूहों में लिंग बंटन में पाये जाने वाले अंतर को समझने में हमारी सहायता करेगा।

इसी प्रकार के परिकलन इन चरों की भूमिका को उल्टा करके किए जा सकते हैं। पुरुषों और महिलाओं के लिए मातृभाषा का सप्रतिबंध बंटन परिकलित किया जा सकता है, जोकि इस प्रकार है, पुरुषों के लिए होगा; हिन्दी $0.3241\left(=\frac{456}{1407}\right)$; बंगाली 0.1663; तमिल

0.2608; अन्य 0.2488। इसी तरह से महिला; $0.3414\left(=\frac{523}{1532}\right)$; 0.1443; 0.2526; 0.2617। इनकी तुलना मातृभाषा के सीमांत बंटन से की जा सकती है, जोकि इस प्रकार है; $0.3331\left(=\frac{979}{2939}\right)$; 0.1548; 0.2565; 0.2555। अगर इनमें से एक चर क्रम सूचक हो तो भी ये संकल्पनाएँ तथा धारणाएं ऐसी ही रहती हैं।

अगर हम कारण–प्रभाव संबंध का अध्ययन करना चाहते हैं तो हमारी रुचि स्वतंत्र चर (कारण) के प्रत्येक मान के लिए आश्रित चर (प्रभाव) के सप्रतिबंध बंटनों में होगी। उदाहरण के लिए, अगर शिक्षा स्तर और व्यवसाय में संबंध–जानने के लिए अध्ययन किया जाना है तो हम शिक्षा–स्तर को कारण तथा व्यवसाय को प्रभाव के रूप में ले सकते हैं और इस प्रकार प्रत्येक शिक्षा स्तरों के लिए सप्रतिबंध बंटनों का अध्ययन कर सकते हैं और इन सप्रतिबंध बंटनों की तुलना भी कर सकते हैं।

अतः जब चर केवल नामिक तथा क्रमसूचक हो तो संयुक्त बारंबारता बंटन का सारणी के रूप में निर्माण तथा प्रस्तुतीकरण बहुत ही सरल होता है। इन सारणियों से संबद्धित सीमांत और सप्रतिबंध–बंटनों से बहुत उपयोगी सूचना प्राप्त की जाती है। लेकिन अगर एक चर के बहुत से स्तर हैं तो सारणी के रूप में इनका सुव्यवस्थित प्रस्तुतीकरण कठिन होता है। इसके अतिरिक्त, अगर चर के एक स्तर के अनुरूप बारंबारता कम है तो इनसे संबद्धित सीमांत और सप्रतिबंध तुलनात्मक बारंबारताएं विश्वसनीय नहीं होंगी। अतः इनका परिकलन, प्रस्तुतीकरण तथा विश्लेषण लाभकारी नहीं होगा। इन परिस्थितियों में चर के कुछ स्तरों का संयोजन (pool) करके, इनका एक साथ प्रस्तुतीकरण, परिकलन और विश्लेषण किया जाता है। उपर्युक्त उदाहरण में मातृभाषा का स्तर "अन्य" इस प्रकार का उदाहरण है, जहां पर वे सभी मातृभाषाएं जिनकी बारंबारता कम है, एक समूह "अन्य" में संयोजित की गई है।

बोध प्रश्न 1

1) निम्नलिखित में से प्रत्येक में यह निर्णय कीजिए कि क्या चर नामिक, क्रमसूचक या संख्यात्मक प्रकृति का है?

   क) 'भारत में आने वाले' पर्यटक के देश की नागरिकता

   ख) एक विद्यार्थी द्वारा परीक्षा में प्राप्त कोटि (grade) जोकि इस प्रकार वर्गीकृत है : A^+, A, B, C तथा D

   ग) एक व्यक्ति की आयु

   घ) शेयर बाज़ार में किसी सार्वजनिक कंपनी के शेयर की कीमत

   च) एक विदेशी पर्यटक का भारत के बारे में विचार, जिसको इस प्रकार वर्गीकृत किया गया है : अद्भुत, अच्छा, औसत, बुरा, भयंकर।

   ....

   ....

   ....

   ....

2) उचित उदाहरण की सहायता से, सीमांत और सप्रतिबंध बंटनों की संकल्पनाओं का वर्णन कीजिए।

....

....

....

....

....

7.4 संख्यात्मक चरों की प्रस्तुति

आइए अब ऐसी स्थिति का विचार करें जहां संख्यात्मक चरों से द्विचर बारंबारता बनाई जाती है। शुरू मे, हम दो चर–एक नामिक और एक संख्यात्मक चर–पर विचार करेंगे। बाद में हम ऐसी स्थिति पर विचार करेंगे जहां दोनों चर संख्यात्मक हैं।

सारणी 7.2 : व्यवसाय और बच्चों की संख्या का द्विचर बारंबारता बंटन

बच्चों की संख्या	व्यवसाय (t)					योग
	बेरोज़गार	अकुशल श्रमिक	कुशल श्रमिक	स्वयं–रोज़गारी	पेशेवर	
x	1	2	3	4	5	
0	10	15	10	12	5	52
1	35	25	17	18	25	120
2	22	33	45	40	43	183
3	11	40	48	58	30	187
4	8	22	12	11	8	61
5	3	11	18	8	1	41
कुल	89	146	150	147	112	644

7.4.1 एक चर संख्यात्मक असंतत्

सारणी 7.2 में प्रस्तुत आँकड़ों में जहां पिता का व्यवसाय (t) एक नामिक चर है और बच्चों की संख्या (x) एक संख्यात्मक चर है। व्यवसाय की पाँच श्रेणियां हैं। बच्चों की संख्या वाला चर 0 से 5 तक के मान लेता है। यहां द्विचर बारंबारता सारणी का निर्माण ठीक उसी तरह का है, जिसकी हमने पिछले अनुभाग में चर्चा की थी।

आइए अब सीमांत बंटन और सप्रतिबंध बंटन के अभिकलन पर ध्यान दें। इस उदाहरण में, आपकी रुचि 'स्वतंत्र चर' (व्यवसाय) का 'आश्रित चर' (बच्चों की संख्या) पर प्रभाव का अध्ययन करना हो सकता है। पहले की तरह यहां पर भी हम विभिन्न व्यवसाय स्तरों के लिए बच्चों की संख्या के सप्रतिबंध तथा सीमांत बंटन परिकलित कर सकते हैं। लेकिन यहां, उपर्युक्त उदाहरणों के विपरीत, बच्चों की संख्या एक संख्यात्मक चर है, अतः इसके परिमाणों का प्रयोग सप्रतिबंध तथा सीमांत बंटनों के अन्य संक्षेपण मापों के परिकलन के लिए भी किया जा सकता है। इस तरह, सप्रतिबंध बंटनों की तुलना के अतिरिक्त हम इनके समांतर माध्य या बहुलक की तुलना भी कर सकते हैं। ध्यान रखिए कि हमने

वर्गीकृत बारंबारता बंटन (इकाई 4 देखें) के समांतर माध्य के परिकलन का अध्ययन किया हुआ है। मान लीजिए, X_t वें कोष्ठ की बारंबारता f_{xt} है जहां पर $X = 0, 1, 2, 3, 4, 5$ एवं $t = 1, 2, 3, 4, 5$ अर्थात X बच्चों की संख्या को निरुपित करता है तथा t व्यवसाय वर्ग को निरुपित करता है। मान लीजिए, t वें व्यवसाय वर्ग की उपांत बारंबारता f_t है।

$$\text{अर्थात}\quad f_t = \sum_{x=0}^{5} f_{xt}$$

ये सीमांत बारंबारता सारणी की अंतिम पंक्ति में दी हुई संख्याएं हैं। इस प्रकार t वें व्यवसाय वर्ग के सप्रतिबंध बंटन का ***सप्रतिबंध समांतर माध्य*** का सूत्र होगा

$$\overline{X}_t = \frac{1}{f}\sum_{x=0}^{5} f_{xt} X \qquad t = 1, 2, 3, 4, 5$$

उदाहरण के लिए, बेरोज़गारों के लिए सप्रतिबंध बंटन का समांतर माध्य है

$$\frac{10\times 0 + 35\times 1 + 22\times 2 + 11\times 3 + 8\times 4 + 3\times 5}{89} = 1.79$$

और अकुशल श्रमिकों के लिए है

$$\frac{15\times 0 + 25\times 1 + 33\times 2 + 40\times 3 + 22\times 4 + 11\times 5}{146} = 2.42$$

इसी तरह से आप अन्य व्यावसायिक समूहों के लिए सप्रतिबंध बंटनों के समांतर माध्य की जांच कर सकते हैं। ये हैं; कुशल श्रमिक : 2.59; स्वयं रोज़गारः 2.42; व्यावसायिकः 2.12

सीमांत बंटन का समांतर माध्य अर्थात सभी व्यवसायों का कुल समांतर माध्य होगा

$$\frac{52\times 0 + 120\times 1 + 183\times 2 + 187\times 3 + 61\times 4 + 41\times 5}{644} = 2.32$$

आइए अब इस बारंबारता बंटन से बहुलक ज्ञात करें। जैसा कि आप जानते हैं, बहुलक वह मान है जो सर्वाधिक उभर कर सामने आता है (देखें इकाई 4)। इस परिभाषा के प्रयोग से, सप्रतिबंध बंटन के बहुलक हैं :

बेरोज़गार : 1; अकुशल श्रमिक : 3; कुशल श्रमिक : 3; स्वयं रोज़गार : 3; व्यावसायिकः 2

सीमांत बंटन से प्राप्त बहुलक– अर्थात सभी व्यवसायों का कुल बहुलक 3 है। यहां पर यह ध्यान दीजिए कि बच्चों की संख्या एक चर है जिसके मान केवल पूर्ण संख्याएं होती हैं और इसके माध्यों को दशमलव संख्याओं में प्रस्तुत करने का कोई अर्थ नहीं होता। यहां पर इन माध्यों का पूर्ण संख्याओं में निकटन (round off) करने पर सभी 2 के बराबर हो जाती है, जिससे विभिन्न व्यवसाय वर्गों में अंतर का पता करना संभव नहीं होता।

7.4.2 एक चर संख्यात्मक एवं सतत्

जब चरों में से एक सतत् हो और प्रेक्षणों की संख्या बहुत अधिक न हो तो प्रायः आंकड़ों को उसी रूप में, जिस रूप में संकलित किए गए हैं, प्रस्तुत किया जाता है। इसके बाद, सांख्यिकीय विश्लेषण के लिए आवश्यक परिकलन किए जाते हैं। उदाहरण के लिए, अगर

हम एक बैंक के पुरुष तथा महिला अर्थशास्त्रियों के वेतन की जाँच करना चाहते हैं तथा इस संबंध में अगर सिर्फ 15 प्रेक्षण है तो इसको (सारणी 7.3 की भांति) सरल रूप में प्रस्तुत किया जा सकता है।

तालिका 7.3

एक बैंक के 15 अर्थशास्त्रियों की लिंग के अनुसार वार्षिक आय (रुपयों में) के आँकड़े

पुरुष	महिला
45120	80505
72580	75012
80912	60045
120100	40010
30042	35010
80045	---
81250	---
105505	---
111005	---
60123	---

लेकिन अगर प्रेक्षणों की संख्या अधिक है तो आंकड़ों का इस प्रकार प्रस्तुतीकरण असुविधाजनक होता है। मान लीजिए, हमारे पास सार्वजनिक और निजी क्षेत्रों में कार्यरत 709 अधिकारियों के वार्षिक वेतन के आंकड़े है और हम इनका अध्ययन करना चाहते हैं। यहां हमारे पास नामिक चर दो श्रेणियों वाले कर्मचारियों के रूप में हैं और जो इस बात पर निर्भर करता है कि ये निजी क्षेत्र या सार्वजनिक क्षेत्र में से किससे संबंधित हैं। हमारे पास 'आय' के रूप में संख्यात्मक सतत् चर है जिसे 14 वर्ग अंतरालों में बांटा जा सकता है। इसके लिए द्विचर बारंबारता सारणी निर्मित की जा सकती है (देखिए सारणी 7.4)

सारणी 7.4

सार्वजनिक और निजी क्षेत्र के अधिकारियों के वार्षिक वेतन (रुपयों में) का बारंबारता बंटन

		सेक्टर में संख्या		कुल का प्रतिशत	
वार्षिक आय	मध्यमान	निजी	सार्वजनिक	निजी	सार्वजनिक
45-50	47.5	84	0	13.6	0.0
50-55	52.5	31	11	5.0	12.2
55-60	57.5	135	12	21.8	13.3
60-65	62.5	115	12	18.6	13.3
65-70	67.5	73	15	11.8	16.7
70-75	72.5	77	8	12.4	8.9
75-80	77.5	31	5	5.0	5.6
80-85	82.5	13	5	2.1	5.6
85-90	87.5	18	7	2.9	7.8
90-95	92.5	32	3	5.2	3.3
95-100	97.5	4	8	0.6	8.9
100-105	102.5	2	3	0.3	3.3
105-110	107.5	1	1	0.2	1.1
110-115	112.5	3	0	0.5	0.0
	$f_i \rightarrow$	619	90	100.0	100.0

सारणी 7.4 के अंतिम दो स्तम्भों में बारंबारता बंटन (प्रतिशत रूप में) दिया हुआ है जो दो क्षेत्रों की तुलना करने में सहायक है। जैसा कि सारणी 7.2 में हमने अध्ययन किया था, यहां भी दो क्षेत्रों के दो वेतनों के समांतर माध्य की तुलना करना भी उपयोगी हो सकता है। मान लीजिए jt वें कोष्ठ की बारंबारता f_{jt} है जहां $j = 1, 2, ...k;\ t = 1, 2, ... l;$ तथा x चर के j वें वर्ग का मध्यबिंदु X_j है। सारणी 7.4 में $k = 14; l = 2, X$ (आय) के मध्यमान सारणी के दूसरे स्तम्भ में दिए हुए हैं। मान लीजिए $\bar{X}_t = \frac{1}{f_t}\sum_{j=1}^{k} f_{jt} X_j$ सारणी की अंतिम पंक्ति में दी गई सीमांत बारंबारता हैं। t वें क्षेत्र में x के सप्रतिबंध समांतर माध्य का सूत्र इस प्रकार लिखा जा सकता है :

$$\bar{X}_t = \frac{1}{f_1}\sum_{j=1}^{k} f_{ft}\, X_j$$

इस सूत्र द्वारा परिकलित समांतर माध्य :

निजी क्षेत्र : 64.83 ('000 रुपए) तथा सार्वजनिक क्षेत्र : 72.17 ('000 रुपए) हैं।

सांख्यिकीय विश्लेषण के कुछ निश्चित उद्देश्यों के लिए प्रत्येक समूह के लिए परिकलित प्रसरणों को प्राप्त करना भी उपयोगी होगा।

ध्यान में लाएं कि इकाई 5 में हमने देखा था कि वर्गीकृत बारंबारता द्वारा प्रसरण का सूत्र इस प्रकार होता है;

$$\sigma^2_{jt} = \frac{\sum_{f=1}^{k} f_{jt} X_j^2}{f_t} - \bar{X}_t^2$$

इस सूत्र द्वारा परिकलित प्रसरण के मान इस प्रकार है : निजी क्षेत्र : 163.37;

सार्वजनिक क्षेत्र : 232.74।

7.4.3 दोनों चर संख्यात्मक सतत्

आइए अब ऐसी स्थिति पर विचार करें जहां दोनों चर सतत् हैं। यदि प्रेक्षणों की संख्या काफी अधिक है तो हम दोनों चरों के लिए, वर्ग अंतरालों के प्रयोग से द्विचर बारंबारता बंटन दर्शा सकते हैं। सारणी 7.5 में हमने इस प्रकार के उदाहरण को दर्शाया है जहां 99 परिवारों के प्रतिदर्श से आंकड़ों की प्राप्ति की गई है और जहां Y वर्ष में मनोरंजन पर किए जाने वाले पारिवारिक खर्च (रुपयों में) को दर्शाता है और X परिवार की कुल वार्षिक आय (रुपयों में) को दर्शाता है।

इस प्रकार की परिस्थिति में हम स्वतंत्र चर X (आय) के प्रत्येक वर्ग के लिए आश्रित चर Y (मनोरंजन पर व्यय) के सप्रतिबंध बंटनों की जांच कर सकते हैं। लेकिन इस सारणी में चूंकि कुल प्रतिदर्श का आकार छोटा है, इसलिए बहुत से खानों की बारंबारताएं शून्य हैं। इस प्रकार के बंटन अधिक उपयोगी नहीं होते। फिर भी यहां पर प्रत्येक आय वर्ग के लिए मनोरंजन व्यय के सप्रतिबंध बंटनों का समांतर माध्य तथा अन्य संक्षेपण परिमाप परिकलित किए जाते हैं। इसके साथ ही, मनोरंजन पर व्यय के सीमांत के परिमाप भी परिकलित किए जा सकते हैं। सप्रतिबंध बंटनों के परिकलित समांतर माध्य सारणी 7.6 में दिए गए हैं।

सारणी 7.5

वार्षिक पारिवारिक आय तथा मनोरंजन पर वार्षिक पारिवारिक व्यय का द्विचर बारंबारता बंटन

मनोरंजन पर आय			वार्षिक आय ('00 रुपयों में) (yt)									
			25-80	80-135	135-190	190-245	245-300	300-355	355-410	410-465	465-520	520-575
			मध्य–मान									
(00Rs.)	मध्य–मान	x_j ↓	52.5	107.5	162.5	217.5	272.5	327.5	382.5	437.5	492.5	547.5
	y_t →		–6	–5	–4	–3	–2	–1	0	1	2	3
45-50	47.5	5	-	-	-	-	-	-	-	-	-	1
40-45	42.5	4	-	-	-	-	-	-	1	-	1	-
35-40	37.5	3	-	-	-	-	1	-	-	1	2	1
30-35	32.5	2	-	-	-	-	-	-	-	4	3	2
25.30	27.5	1	-	-	-	-	3	4	4	5	6	1
20-25	22.5	0	-	-	-	-	-	5	7	12	1	1
15-20	17.5	-1	-	-	-	1	4	8	1	1	-	-
10-15	12.5	-2	-	1	3	1	1	-	-	-	-	-
5-10	7.5	-3	-	4	4	-	-	-	-	-	-	-
0-5	2.5	-4	4	-	-	-	-	-	-	-	-	-

सारणी 7.6

वार्षिक पारिवारिक आय के प्रत्येक वर्ग के लिए मनोरंजन पर वार्षिक पारिवारिक व्यय के माध्य ('00 रुपए में)

वार्षिक आय	मनोरंजन पर औसत वार्षिक व्यय
25-80	2.50
80-135	8.50
135-190	9.65
190-245	15.00
245-300	22.50
300-355	21.30
355-410	25.19
410-465	25.76
465-520	30.96
520-575	33.33

बोध प्रश्न 2

1) मान लीजिए आपके पास भारत में विदेशी पर्यटकों के देश की नागरिकता और उनके द्वारा व्यय की गई मुद्रा की मात्रा के आंकड़े हैं, जोकि 2000 पर्यटकों के सर्वेक्षण के आधार पर उनके वापिस जाते समय प्राप्त किए गए हैं। यह बताइए कि इन आंकड़ों को आप सारणीबद्ध रूप में किस प्रकार प्रस्तुत करेंगे?

....

....

....

....

2) निम्नलिखित सारणी में एक शहर के तीन इलाकों से लिए गए प्रतिदर्शों द्वारा खाद्य पदार्थों पर मासिक व्यय का द्विचर बारंबारता बंटन दिया हुआ है। प्रत्येक इलाके के लिए व्यय का सप्रतिबंध बंटन ज्ञात कीजिए तथा इनके सप्रतिबंध माध्य एवं प्रसरण परिकलित कीजिए।

तीन इलाकों में खाद्य पदार्थों पर मासिक पारिवारिक व्यय का बारंबारता बंटन

खाद्य पर व्यय (रुपयों में)	इलाका		
	A	B	C
<250	122	2	0
251-500	100	5	1
501-750	75	11	3
751-1000	59	25	18
1001-1250	34	39	27
1251-1500	23	56	56
1501-2000	12	67	89
>2000	0	45	114

....

....

....

....

7.5 सारांश

चर तीन तरह के होते हैं : नामिक, क्रमसूचक और संख्यात्मक। नामिक चरों को श्रेणीबद्ध किया जा सकता है। दूसरी तरफ, क्रमसूचक चरों को हम क्रमबद्ध कर सकते हैं। संख्यात्मक चर, असतत् और सतत् हो सकते हैं और इनके परिमाणात्मक मान होते हैं।

इस इकाई में हमने उपर्युक्त चरों को द्विचर बारंबारता बंटन के रूप में प्रस्तुत किया। हमने इन चरों के सीमांत और सप्रतिबंध बंटनों को भी परिकलित किया।

7.6 शब्दावली

द्विचर आंकड़े (Bivariate Data) : ऐसे आंकड़े जिनमें प्रत्येक व्यष्टि के दो अभिलक्षणों का माप किया जाता है। जैसे, प्रत्येक शिक्षित व्यक्ति की आय तथा शिक्षा प्राप्त करने के वर्षों की संख्या।

सीमांत बंटन (Marginal Distribution) : इसका अर्थ द्विधा या बहुधा सारणी के पंक्ति योग या स्तंभ योग के बंटन से होता है।

7.7 कुछ उपयोगी पुस्तकें

Nagar, A.L. and R.K. Das, 1989 : *Basic Statistics,* Oxford University Press, Delhi.

Goon, A.M.: M.K. Gupta and B. Dasgupta, 1987 : *Basic Statistics,* The World Press Pvt. Ltd., Kolkata.

7.8 बोध प्रश्नों के उत्तर या संकेत

बोध प्रश्न 1

1) (i) नामिक (ii) क्रमसूचक (iii) संख्यात्मक (vi) संख्यात्मक (v) नामिक

2) अनुभाग 7.3 पढ़िए और उत्तर दीजिए।

बोध प्रश्न 2

1) खर्च की गई राशि संख्यात्मक सतत चर है जबकि देश की नागरिकता नामिक चर है। अतः खर्च की जाने वाली राशि के लिए, आपको वर्ग अंतराल बनाने होंगे और देश की नागरिकता को श्रेणियों में दर्शाना होगा।

   इसी आधार पर आप द्विचर बारंबारता बंटन निर्मित कर सकते हैं जहां प्रत्येक कोश के अंतर्गत आने वाले पर्यटकों की संख्या का पता चलता है।

2) देखें अनुभाग 7.3

इकाई 8 सहसंबंध विश्लेषण

इकाई की रूपरेखा

8.0 उद्देश्य

इस इकाई को पढ़ने के बाद, आपः

- प्रकीर्ण आरेख को बना सकेंगे;
- दो चरों के बीच के सहसंबंध को माप सकेंगे;
- सहसंबंध गुणांक को परिकलित कर सकेंगे;
- कोटि सहसंबंध गुणांक को परिकलित कर सकेंगे; और
- निर्धारित कर सकेंगे कि क्या दो चर सहसंबंधित हैं।

8.1 प्रस्तावना

पिछली इकाई में हमने बारंबारता बंटनों के रूप में द्विचर आंकड़ों की प्रस्तुति की प्रविधियों की चर्चा की। इस इकाई में हम सहसंबंध की संकल्पना को समझेंगे जो दो चरों के साहचर्य की शक्ति के परिमाप से संबंधित है। जब हम द्विचर आंकड़ों के समुच्चय से सहसंबंध के परिमाप का परिकलन करते हैं तो हमारी रुचि, चरों के बीच सहसंबंध की **गहनता** और **दिशा** पर केन्द्रित होती है।

बहुत से चरों के सांख्यिकीय अध्ययनों में प्रायः दो प्रकार की समस्याएं होती हैं। कुछ समस्याओं के अध्ययन में हमारी रुचि यह जानने में होती है कि चरों में किस प्रकार का परस्पर संबंध है। इस प्रकार की समस्याओं का समाधान सहसंबंध प्रविधियों के प्रयोग द्वारा किया जाता है। उदाहरण के लिए, एक अर्थशास्त्री की रुचि विभिन्न कंपनियों के शेयरों की कीमतों में संबंध का अध्ययन करना हो सकती है, इसके लिए वह सहसंबंध प्रविधियों का प्रयोग कर सकता है।

दूसरे प्रकार की समस्याओं में मूल रुचि Y में होती है तथा हमें यह जानना होता है कि अन्य चर, Y के बारे में क्या सूचना प्रदान करते हैं। इस प्रकार की समस्याओं का समाधान समाश्रयण (regression) प्रविधियों के प्रयोग द्वारा किया जाता है। उदाहरण के लिए, एक अर्थशास्त्री की रुचि इस बात में हो सकती है कि किसी कार्यरत व्यक्ति की आय किन कारकों से निर्धारित होती है, विशेष रूप से उसकी रुचि यह जानना हो सकती है कि शिक्षा, अनुभव, बाज़ार मांग आदि की व्यक्ति के वेतन के निर्धारण में क्या भूमिका है। इसके लिए वह समाश्रयण प्रविधियों के प्रयोग द्वारा शिक्षा, अनुभव आदि पर आधारित वेतन का प्रागुक्ति (prediction) सूत्र ज्ञात कर सकता है।

इस इकाई में हम सहसंबंध पर चर्चा कर रहे हैं। समाश्रयण विश्लेषण पर हम अगली इकाई में विचार करेंगे।

8.2 प्रकीर्ण आरेख

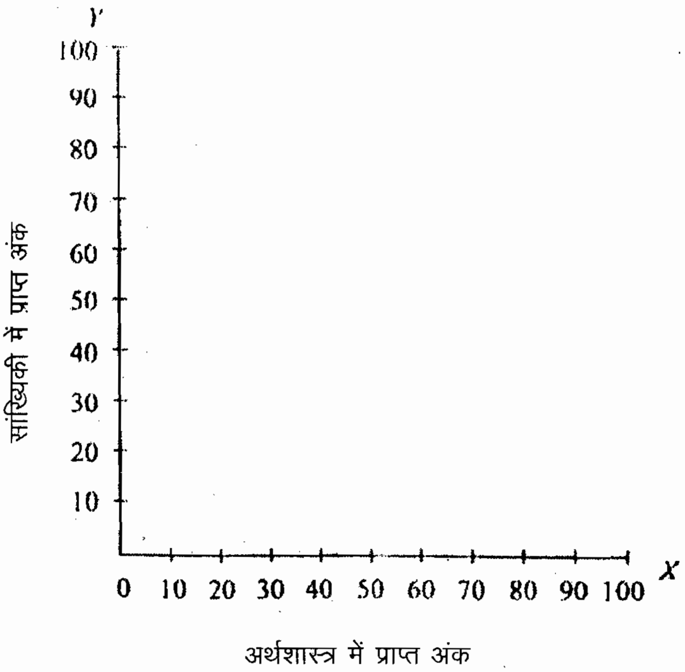
पहले हम बताएंगे कि दो चरों में संबंध का अध्ययन किस प्रकार किया जाता है। एक शिक्षक की रुचि कक्षा के 20 विद्यार्थियों की सांख्यिकी तथा अर्थशास्त्र में योग्यता के संबंध का अध्ययन करना हो सकती है। इसके लिए वह पिछली अर्ध–सत्रीय परीक्षा में इन विद्यार्थियों द्वारा इन विषयों में प्राप्त अंकों के आंकड़े संकलित करता है। इस प्रकार के कुछ आंकड़े सारणी 8.1 में प्रस्तुत किए गए हैं।

सारणी 8.1 : विद्यार्थियों द्वारा सांख्यिकी तथा अर्थशास्त्र में प्राप्त अंक

क्रम	प्राप्त अंक		क्रम	प्राप्त अंक	
संख्या	सांख्यिकी	अर्थशास्त्र	संख्या	सांख्यिकी	अर्थशास्त्र
1	82	64	11	76	58
2	70	40	12	76	66
3	34	35	13	92	72
4	80	48	14	72	46
5	66	54	15	64	44
6	84	56	16	86	76
7	74	62	17	84	52
8	84	66	18	60	40
9	60	52	19	82	60
10	86	82	20	90	60

इस प्रकार के आंकड़ों का आलेखी निरूपण एक उपयोगी विधि है, जोकि दो चरों के बीच संबंध की प्रकृति तथा रूप के अध्ययन में सहायक होती है। आलेखी निरूपण द्वारा यह पता किया जा सकता है कि क्या चरों में अध्ययन करने लायक कोई संबंध है या नहीं, अगर है तो क्या वह रैखिक है या अरैखिक। इसके लिए मान लीजिए हम सांख्यिकी में प्राप्त अंकों को X से सूचित करते हैं तथा अर्थशास्त्र में प्राप्त अंकों को Y से सूचित करते हैं तथा सारणी 8.1 के आंकड़ों को X, Y समतल पर अंकित करते हैं। इस कार्य के लिए हम किसको X तथा किसको Y लें, कोई अर्थ नहीं रखता। इस प्रकार के अंकन को **प्रकीर्ण आरेख** (scatter diagram) कहते हैं। चित्र 8.1 में सारणी 8.1 के आंकड़ों का प्रकीर्ण आरेख दिया गया है।

सारणी 8.1 और चित्र 8.1 की जांच द्वारा यह पता चलता है कि X तथा Y में धनात्मक संबंध है अर्थात X के बड़े मान Y के बड़े मानों के साथ तथा X के छोटे मान, Y के छोटे मानों के साथ सहचारी हैं। इसके अतिरिक्त, बिंदुओं एक सरल रेखा के दोनों ओर प्रकीर्ण दिखाई देते हैं। अतः X तथा Y के बीच रैखिक संबंध प्रतीत होता है, लेकिन यह संबंध पूर्ण (perfect) नहीं है, क्योंकि इस प्रकार के संबंध में विचलन मौजूद है। वास्तव में, इस रैखिक संबंध की शक्ति का परिमाप प्राप्त करना बड़ा ही उपयोगी होगा।



चित्र 8.1 : सांख्यिकी और अर्थशास्त्र में प्राप्त अंकों का प्रकीर्ण आरेख

8.3 सहप्रसरण

एक चर वाली स्थिति में हमने प्रसरण की संकल्पना का अध्ययन किया है, जो इस तरह परिभाषित है

$$\sigma_x^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 \qquad ...(8.1)$$

ऊपर पादांक X का प्रयोग यह दर्शाने के लिया किया कि σ_x^2 हमें X के प्रसरण को दर्शाता है। इसी ढंग से हम σ_y^2 को Y में प्रसरण के रूप में और σ_x और σ_y को क्रमशः X और Y में मानक विचलन के रूप में दर्शा सकते हैं।

जैसा कि आप जानते हैं प्रसरण, माध्य से प्रकीर्णन का परिमाप होता है। द्विचर आंकड़ों की स्थिति में हमें ऐसे एकल अंक तक पहुंचना है जो दोनों चरों में अपने संबद्ध माध्यों से विचलन को प्रस्तुत करेगा। इस उद्देश्य के लिए हमने संकल्पना सहप्रसरण (covariance) का प्रयोग किया और जिसे इस प्रकार परिभाषित करते है;

$$\sigma_{xy} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y}) \qquad ...(8.2)$$

आपको याद होगा कि मानक विचलन सदैव धनात्मक होता है, क्योंकि यह प्रसरण के धनात्मक वर्गमूल के रूप में परिभाषित किया जाता है। सहप्रसरण की स्थिति में, दो पद $(X_i - \overline{X})$ और $(Y_i - \overline{Y})$ हैं जो X से $\overline{X}$ और Y से $\overline{Y}$ में विचलन को निरूपित करते हैं। इसके अलावा $(X_i - \overline{X})$ धनात्मक या ऋणात्मक हो सकता है जो इस बात पर निर्भर करता है कि X_i का मूल्य $\overline{X}$ से कम या अधिक है। इसी तरह $(Y_i - \overline{Y})$ भी धनात्मक या ऋणात्मक हो सकता है। यह आवश्यक नहीं है कि जब भी $(X_i - \overline{X})$ धनात्मक हो, $(Y_i - \overline{Y})$ भी धनात्मक होगा। इसलिए, गुणनफल $(X_i - \overline{X})(Y_i - \overline{Y})$ धनात्मक या ऋणात्मक में से कोई भी हो सकता है। $(X_i - \overline{X})(Y_i - \overline{Y})$ के धनात्मक मान से आशय् है कि जब भी $X_i > \overline{X}$, तब $Y_i > \overline{Y}$ होगा। अतः X_i का उच्च मान, Y_i में उच्च मान से सापेक्षिक रूप से संबद्ध है। दूसरी तरफ $(X_i - \overline{X})(Y_i - \overline{Y}) < 0$ से आशय् है कि X_i में निम्न मान Y_i में सापेक्षिक रूप से उच्च मान से संबद्ध है। जब हम सभी प्रेक्षणों से इन्हें जोड़ते हैं और प्रेक्षणों की संख्या से विभाजित करते हैं तो हमें ऋणात्मक या धनात्मक मान की प्राप्ति हो सकती है। इसलिए सहप्रसरण के ऋणात्मक और धनात्मक अर्थात दोनों तरह के मान हो सकते हैं।

जब X और Y के बीच सहप्रसरण ऋणात्मक $(\sigma_{xy} < 0)$ है तो हम कह सकते हैं कि दोनों चरों के बीच संबंध विपरित है। इसी तरह $(\sigma_{xy} > 0)$, X और Y के बीच धनात्मक संबंध को दर्शाता है। सहप्रसरण की मुख्य सीमा है कि यह माप की इकाई से अलग नहीं है। इसका अर्थ है कि जब हम चरों के मात्रक बदलते हैं तो हमें σ_{xy} के लिए अलग मान प्राप्त होंगे। जैसा कि (8.2) में दिया है σ_{xy} के परिकलन में प्रायः काफी संख्याएं शामिल होती हैं। इसलिए आगे, इसे इस प्रकार व्युत्पन्न किया गया है

$$\sigma_{xy} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y}) = \frac{1}{n}\sum_{i=1}^{n}(X_iY_i - \overline{X}Y_i - X_i\overline{Y} + \overline{X}\,\overline{Y})$$

इसे और अधिक सरल बनाने पर हम पाते हैं

$$\sigma_{xy} = \frac{1}{n}\sum_{i=1}^{n}X_i Y_i - \frac{1}{n}\sum_{i=1}^{n}\overline{X}Y_i - \frac{1}{n}\sum_{i=1}^{n}X_i\overline{Y} + \frac{1}{n}\sum_{i=1}^{n}\overline{X}\,\overline{Y}$$

चूंकि $\frac{1}{n}\sum_{i=1}^{n}\overline{X}Y_i = \frac{1}{n}\sum_{i=1}^{n}X_i\overline{Y} = \overline{X}\,\overline{Y}$

अतः $$\sigma_{xy} = \frac{1}{n}\sum_{i=1}^{n}X_i Y_i - \overline{X}\,\overline{Y} \qquad \dots (8.3)$$

8.4 सहसंबंध गुणांक

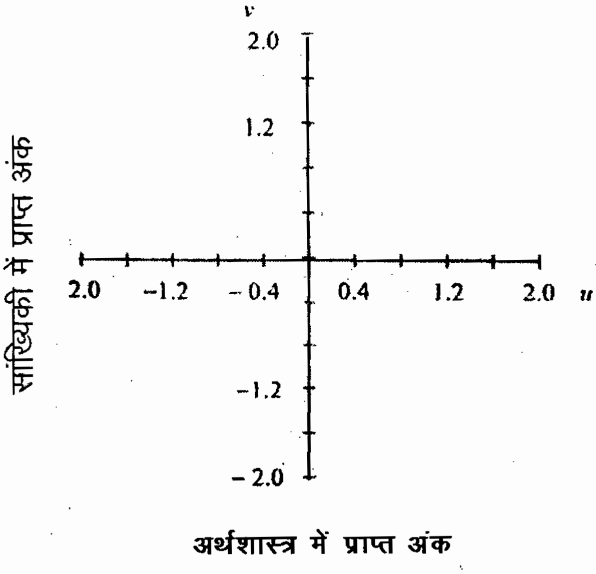
अब हमें X और Y को बीच के रैखिक संबंध का परिमाप प्राप्त करना है। रैखिक संबंध की शक्ति का ऐसा परिमाप प्राप्त करना है जोकि चर के माप के लिए प्रयोग किए गए पैमाने से स्वतंत्र हो, वांछनीय होता है। उदाहरण के लिए, अगर हम ऊंचाई और वजन में संबंध को मापना चाहते हैं, तो चाहे हम ऊँचाई को इंचों में मापे या सेंटीमीटरों में तथा वज़न को

पाउंड में मापें या किलोग्राम में, हमें वही परिमाप प्राप्त होना चाहिए। इसी प्रकार, अगर तापमान एक चर है तो, चाहे वह सेल्सियस में है या फारेनहाइट में है, इससे विश्लेषण में कोई अंतर नहीं आना चाहिए। यह स्थिति प्रत्येक चर के मानकीकरण द्वारा प्राप्त की जा सकती अर्थात $\frac{X_i - \overline{X}}{\sigma_x}$ और $\frac{Y_i - \overline{Y}}{\sigma_y}$ पर विचार करके, जहां $\overline{X}$ और $\overline{Y}$ क्रमशः X और Y के माध्य हैं और σ_x और σ_y प्रतिदर्श मानक विचलन।

मान लीजिए, हम इन मानकीकृत चरों को क्रमशः : u तथा v से सूचित करते हैं। हम यह भी जानते हैं कि (X_i, Y_i) पवें विद्यार्थी द्वारा क्रमशः सांख्यिकी तथा अर्थशास्त्र में प्राप्त अंकों को सूचित करता है। i का मान 1 से n तक होता है। हमारे उदाहरण में n का मान 20 है। इसी प्रकार, मान लीजिए (u_i, v_i) iवें विद्यार्थी द्वारा प्राप्त मानकीकृत अंको को सूचित करता है। यहां माध्य तथा मानक विचलन के सूत्रों को पुनः स्मरण करने पर :

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i;\ \sigma_x^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2;$$

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i;\ \sigma_y^2 = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2;$$



चित्र 8.2 : सांख्यिकी और अर्थशास्त्र में प्राप्त मानकीकृत अंकों का प्रकीर्ण आरेख

चित्र 8.2 मानकीकृत चरों u तथा v में प्रकीर्ण आरेख है। मान लीजिए इस उदाहरण में हम दो प्रकार के अंकों में धनात्मक साहचर्य का प्रेक्षण करते हैं, मोटे तौर पर अगर एक विषय में प्राप्त अंक बड़ा है तो दूसरे विषय में भी प्राप्त अंक बड़ा होगा और अगर एक विषय में प्राप्त अंक कम है तो दूसरे विषय में प्राप्त अंक भी कम होगा। इस दृष्टि से अधिकतर बिंदु या तो पहले चतुर्थांश में हैं या फिर तीसरे चतुर्थांश में है। पहला चतुर्थांश उन परिस्थितियों को व्यक्त करता है, जहां दोनों विषयों में प्राप्त अंक अपने माध्यों से अधिक हैं और तीसरा चतुर्थांश उन परिस्थितियों को व्यक्त करता है, जहां दोनों विषयों में प्राप्त अंक अपने माध्यों से कम हैं। दूसरे तथा चौथे चतुर्थांश में केवल कुछ बिंदु है जोकि उन परिस्थितियों को व्यक्त करते हैं, जहां एक विषय में माध्य से अधिक तथा दूसरे विषय में माध्य से कम अंक

प्राप्त है। अतः u और v का गुणनफल, संबंध की शक्ति का उपयुक्त सूचक है। यह गुणनफल पहले तथा तीसरे चुतर्थांश में धनात्मक तथा दूसरे एवं चौथे चतुर्थांश में ऋणात्मक है। अतः u और v के सभी बिंदुओं के लिए औसत गुणनफल को X तथा Y के बीच रैखिक संबंध की शक्ति का उपयुक्त माप लिया जा सकता है। इस परिमाप को X और Y में सहसंबंध गुणांक कहते हैं, जिसको प्रायः r_{xy} या केवल r से सूचित किया जाता है। अन्य प्रकार के सहसंबंध गुणांक से भेद करने के लिए इसको पियर्सन का गुणन–आघूर्ण सहसंबंध गुणांक (Pearson's Product-Moment Correlation Coefficient) भी कहते हैं।

अतः r के लिए सूत्र है

$$r = \frac{1}{n}\sum_{i=1}^{n} u_i v_i \qquad \text{... (8.4)}$$

यदि हम उपर्युक्त (8.4) में X और Y चरों को प्रतिस्थापित करते हैं

$$r = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{\sigma_x}\right)\left(\frac{Y_i-\bar{Y}}{\sigma_y}\right) = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)\left(Y_i-\bar{Y}\right)}{\sigma_x\sigma_y}$$

उपर्युक्त व्यंजन में, पद

$$\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)\left(Y_i-\bar{Y}\right)$$

X और Y के बीच सहप्रसरण कहलाता है और इसको (σ_{xy}) से सूचित किया जाता है।

अतः सहसंबंध गुणांक का सूत्र है;

$$r = \frac{\sigma_{xy}}{\sigma_x \times \sigma_x} \qquad \text{... (8.5)}$$

इनमें $\bar{X}, \bar{Y}, \sigma_x, \sigma_y$ के सूत्रों को जोड़ने पर यह बनता है

$$r = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)\left(Y_i-\bar{Y}\right)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i-\bar{Y}\right)^2}} = \frac{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)\left(Y_i-\bar{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}\sqrt{\sum_{i=1}^{n}\left(Y_i-\bar{Y}\right)^2}} \qquad \text{...(8.6)}$$

या वैकल्पिक रूप से होगा

$$r = \frac{n\sum_{i=1}^{n}\left(X_i - Y_i\right) - \sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{\sqrt{\left[n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2\right]\left[n\sum_{i=1}^{n} Y_i^2 - \left(\sum_{i=1}^{n} Y_i\right)^2\right]}} \qquad \text{...(8.7)}$$

आइए अब दुबारा सारणी 8.1 में दिए गए आंकड़ों पर ध्यान दें और r के मान ज्ञात करें। आप r के मान के लिए (8.4), (8.5), (8.6) या (8.7) में से कोई सूत्र का प्रयोग कर सकते हैं। चूंकि उपर्युक्त सभी सूत्र एक ही संकल्पना से व्युत्पन्न है, इसलिए सूत्र चाहे कोई भी हो r के लिए हमें एक जैसे मान की ही प्राप्त होगी। सारणी 8.1 में दर्शित आंकड़ों के लिए हमने (8.4) और (8.7) के प्रयोग से इसे परिकलित किया है। इस उद्देश्य के लिए हमनें सारणी 8.2 निर्मित की है।

सारणी 8.2 : सहसंबंध गुणांक का परिकलन

प्रेषण सं.	X	Y	X^2	Y^2	XY
1	82	64	6724	4096	5248
2	70	40	4900	1600	2800
3	34	35	1156	1225	1190
4	80	48	6400	2304	3840
5	66	54	4356	2916	3564
6	84	56	7056	3136	4704
7	74	62	5476	3844	4588
8	84	66	7056	4356	5544
9	60	52	3600	2704	3120
10	86	82	7396	6724	7052
11	76	58	5776	3364	4408
12	76	66	5776	4356	5016
13	92	72	8464	5184	6624
14	72	46	5184	2116	3312
15	64	44	4096	1936	2816
16	86	76	7396	5776	6536
17	84	52	7056	2704	4368
18	60	40	3600	1600	2400
19	82	60	6724	3600	4920
20	90	60	8100	3600	5400
Total	**1502**	**1133**	**116292**	**67141**	**87450**

सारणी 8.2 से, हम देखते हैं कि

$$\sum_{i=1}^{20} X_i = 1502; \overline{X} = 75.1;$$

$$\sum_{i=1}^{20} Y_i = 1133; \overline{Y} = 56.65;$$

$$\sum_{i=1}^{20} X_i^2 = 116292; \sigma_x^2 = \frac{1}{20}\left[116292 - \frac{1502^2}{20}\right] = 174.59; \sigma_x = 13.21;$$

$$\sum_{i=1}^{20} Y_i^2 = 67141; \sigma_y^2 = \frac{1}{20}\left[67141 - \frac{1133^2}{20}\right] = 147.83; \sigma_y = 12.16;$$

$$\sum X_i Y_i = 87450; \sigma_{xy} = \frac{1}{20}\left[87450 - \frac{1502 \times 1133}{20}\right] = 118.09$$

$$r = \frac{118.09}{13.21 \times 12.16} = 0.735$$

आइए अब सूत्र (8.7) का प्रयोग करें। अब हमें मिलता है

$$r = \frac{20 \times 87450 - 1502 \times 1133}{\sqrt{(20 \times 116292 - 1502^2)(20 \times 67141 - 1133^2)}} = 0.735$$

अतः हम देखते हैं कि दोनों सूत्र, सहसंबंध गुणांक r के एक जैसे मान प्रदान करते हैं। आप स्वयं भी जांच कर सकते हैं कि सूत्र (8.5) के प्रयोग से r का समान मान प्राप्त किया जाता है। आपको निम्नलिखित मानों की आवश्यकता होगी :

$$\Sigma(X_i - \overline{X})^2, \Sigma(Y_i - \overline{Y})^2 \text{ और } \Sigma(X_i - \overline{X})(Y_i - \overline{Y})$$

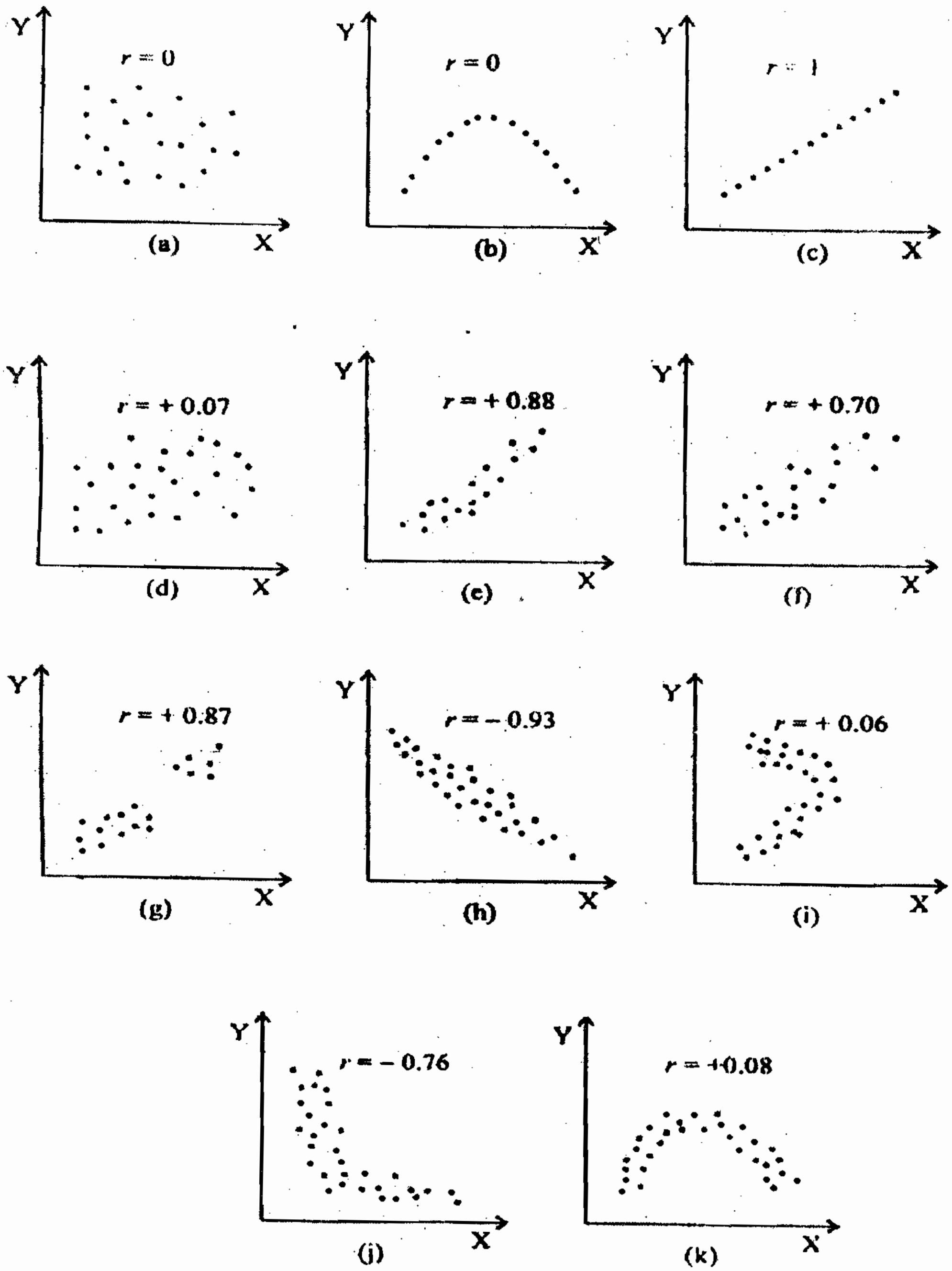
एक सारणी में $(X_i - \overline{X}), (Y_i - \overline{Y}), (X_i - \overline{X})^2, (Y_i - \overline{Y})^2$ और $(X_i - \overline{X})(Y_i - \overline{Y})$ पर आप पाँच स्तम्भ प्राप्त कर सकते हैं और उनका योगफल ज्ञात कर सकते हैं।

8.5 सहसंबंध गुणांक की व्याख्या

यह एक गणितीय तथ्य है कि उपर्युक्त परिभाषित r का मान हमेशा -1 तथा $+1$ के बीच रहता है जब X और Y में पूर्ण रैखिक संबंध हो तो r का चरममान -1 या $+1$ प्राप्त होता है। जब X तथा Y में विपरीत संबंध हो तो मान -1 होता है एवं जब संबंध प्रत्यक्ष हो तो मान $+1$ होता है। जब X और Y में कोई संबंध न हो तो मान 0 होता है।

चित्र 8.3, r के विविध मानों के लिए प्रकीर्ण आरेख के उदाहरणों को उजागर करता है। चित्र 8.3 (क) स्थिति $r = 0$ का प्रकीर्ण आरेख है जहां X और Y के बीच कोई संबंध नहीं है। चित्र 8.3 (ख) भी $r = 0$ की परिस्थिति में प्रकीर्ण आरेख का उदाहरण है यहां पर X और Y में संबंध तो दिखाई देता है, पर वह रैखिक नहीं है। यहां पहले X में वृद्धि के साथ Y में भी वृद्धि होती है। लेकिन बाद में, X में वृद्धि के साथ Y में कमी होती है जोकि एक द्विघात संबंध है। परिणामस्वरूप इस परिस्थिति में सहसंबंध गुणांक शून्य है। अतः सहसंबंध गुणांक केवल रैखिक संबंध का परिमाप होता है। अगर हम व्यक्तियों की आयु तथा वज़न को अंकित करें तो इस प्रकार का प्रकीर्ण आरेख प्राप्त होगा। चित्र 8.3 (ग) ऐसे प्रकीर्ण आरेख का उदाहरण है जहां X और Y के बीच पूर्ण धनात्मक रैखिक संबंध है। अगर हम व्यक्तियों की इंचों में लंबाई को उनकी सेंटीमीटरों में ऊंचाई के सम्मुख अंकित करें तो इस प्रकार का प्रकीर्ण आरेख प्राप्त होगा; क्योंकि इस परिस्थिति में $Y = 2.54X$ जहां पर Y इंचों में ऊंचाई को तथा X सेंटीमीटरों में ऊँचाई को सूचित करता है, एक पूर्ण रैखिक संबंध है। चित्र 8.3 (घ) से चित्र 8.3 (ड़) तक प्रकीर्ण आरेख r के अन्य मानों के लिए है। इन प्रकीर्ण आरेखों द्वारा हमें संबंध की प्रकृति तथा r के मान के बारे में जानकारी प्राप्त होती है।

उपर्युक्त विवरण द्वारा यह प्रतीत होता है कि विद्यार्थियों द्वारा सांख्यिकी तथा अर्थशास्त्र में प्राप्त किए अंकों के बीच $r = 0.74$, एक संतोषजनक किस्म के रैखिक संबंध का सूचक है। इस प्रकार चरों के बीच संबंध या साहचर्य का परिमापन प्राकृतिक तथा समाजशास्त्रियों को उन तथ्यों को समझने में सहायक होता है जिनकी यह जांच कर रहे हैं। इस प्रकार के उदाहरण में एक शिक्षा मनोविज्ञानी विभिन्न विषयों में प्राप्त अंकों के बीच सहसंबंध गुणांक को परिकलित कर सकता है तथा इन गुणांको का और सांख्यिकीय विश्लेषण करके तथा मनोवैज्ञानिक प्रविधियों के प्रयोग द्वारा एक सिद्धांत बना सकता है जोकि विद्यार्थियों को विभिन्न विषयों में अच्छा बनाने के लिए मानसिक एवं अन्य योग्यताओं की जानकारी दे सकता है।



चित्र 8.3 : सहसंबंध–गुणांक के विभिन्न मानों के लिए प्रकीर्ण आरेख

याद रखें

- सहसंबंध गुणांक, X और Y के बीच रैखिक संबंध को दर्शाता है। इसलिए यदि X और Y के बीच गूढ़ गैर–रैखिक संबंध होगा तो सहसंबंध–गुणांक निम्न हो सकता है।
- सहसंबंध गुणांक स्केल और मूल बिंदु (origin) से स्वतंत्र होता है। यदि हम एक (या दोनों) चरों में से किसी स्थिरांक को घटाते हैं तो सहसंबंध गुणांक अपरिवर्तित रहेगा। इसी तरह, यदि हम किसी स्थिरांक से एक (या दोनों) चरों को विभाजित करते हैं तो सहसंबंध गुणांक बदलेगा नहीं।
- सहसंबंध गुणांक का मान –1 और +1 के बीच रहता है।

दो चरों में रैखिक संबंध की मौजूदगी का अर्थ यह नहीं लेना चाहिए कि उन दोनों में कारण–प्रभाव का संबंध है। जैसे, अगर आप पैट्रोल और चॉकलेट पर पारिवारिक खर्च के बीच सहसंबंध परिकलित करें तो आपको इनमें सहसंबंध का मान अधिक बड़ा प्राप्त हो सकता है, जोकि इन चरों में बड़ी ऊँची कोटि के रैखिक संबंध का सूचक है। लेकिन यह आवश्यक नहीं है कि अधिक कार प्रयोग के कारण लोग अधिक चॉकलेट खरीदते हैं। दोनों वस्तुएँ विलासिता की वस्तुएँ है तथा धनी परिवार इनको खरीद सकते हैं, जबकि निर्धन परिवार नहीं खरीद सकते। इस प्रकार, यहाँ सहसंबंध के ऊँचे होने के कारण प्रत्येक चर तथा आय के बीच ऊँचा सहसंबंध होता है। एक और उदाहरण पर विचार कीजिए। मान लीजिए पिछले बीस वर्षों से आप एक भारतीय की औसत ऊँचाई तथा उसके द्वारा टेलीविज़न देखने का प्रति सप्ताह औसत समय के आंकड़ें प्राप्त कर रहे हैं। यह संभव है कि आप इनमें गहरा धनात्मक सहसंबंध पाएँ। लेकिन इसका अर्थ यह नहीं है कि अधिक समय तक टेलीविज़न देखने से व्यक्ति की ऊँचाई में वृद्धि होती है या अधिक ऊँचाई वाले व्यक्ति अधिक समय तक टेलीविज़न देखते हैं। वास्तव में, इन दोनों चरों में समय के साथ वृद्धि होने की प्रवृत्ति होती है, जोकि उनमें ऊँचे सहसंबंध द्वारा प्रतिबिंबित होती है। इस प्रकार के दो चरों के बीच सहसंबंध, जोकि उनके ऊपर किसी तीसरे चर के प्रभाव के कारण प्राप्त होता है (न कि उनमें प्रत्यक्ष रैखिक कारण–प्रभाव संबंध के कारण), को मिथ्या सहसंबंध (spurious correlation) कहते हैं।

सहसंबंध परिकलन के बारे में एक और बात का ज्ञान होना चाहिए। प्रतिदर्श द्वारा परिकलित अन्य मात्राओं की तरह, सहसंबंध भी प्रत्येक प्रतिदर्श के लिए भिन्न होता है तथा परिकलित सहसंबंध गुणांक के प्रयोग के लिए इन उच्चावचनों का हिसाब रखना भी आवश्यक होता है। यहाँ हम प्रविधियों की व्याख्या नहीं करेंगे। दो चरों के बीच रैखिक संबंध की मौजूदगी, यानि उनमें ऊँचा सहसंबंध वास्तविक है या मिथ्या, इस प्रकार की जानकारी एक चर द्वारा दूसरे चर की *प्रागुक्ति* में सहायक होती है। इन प्रागुक्ति प्रविधियों की जाँच हम इकाई 9 में करेंगे।

बोध प्रश्न 1

1) निम्नलिखित परिणामों द्वारा r का मान परिकलित कीजिए:

$n = 10, \sum X = 125; \sum X^2 = 1585; \sum Y = 80; \sum Y^2 = 650; \sum XY = 1007$

....

....

....

....

....

....

2) पति और पत्नी की आयु के लिए सहसंबंध गुणांक का परिकलन कीजिए।

पति की आयु:	23	27	28	29	30	31	33	35	36	39
पत्नी की आयु:	18	22	23	24	25	26	28	29	30	32

....

....

3) समान प्रकार से संसाधित मिश्रित इस्पात के नमूने, जिनमें निकल के प्रतिशत की जाँच उनकी मज़बूती के साथ की गई है, के परिणाम निम्नलिखित हैं :

मज़बूती	47	50	52	54	56	58	59	60	60	62	64	65	66
निकल का प्रतिशत	2.7	2.7	2.8	2.8	2.9	3.2	3.2	3.3	3.5	3.6	3.7	3.7	3.8

मज़बूती तथा निकल की मात्रा में सहसंबंध परिकलित कीजिए तथा परिणाम पर टिप्पणी कीजिए।

4) X और Y में सहसंबंध गुणांक का निर्धारण कीजिए–

X:	5	7	9	11	13	15
Y:	1.7	2.4	2.8	3.4	3.7	4.4

5) निम्नलिखित सारणी में बहुत से वर्षों के लिए बचत बैंक जमा (बिलियन डालरों में) और हड़ताल एवं तालाबंदी (हज़ारों में) के आंकड़े दिए हुए हैं। संबंध गुणांक परिकलित करके परिणाम पर टिप्पणी कीजिए।

बचत जमा :	5.1	5.4	5.5	5.9	6.4	6.0	7.2
हड़ताल एवं तालाबंदी :	3.8	4.4	3.3	3.6	3.3	2.3	1.0

8.6 कोटि सहसंबंध गुणांक

अगर विचाराधीन दोनों चर संख्यात्मक हैं तथा इनमें संबंध रैखिक हैं तो उपर्युक्त सहसंबंध गुणांक या पियर्सन का गुणन आघूर्ण सहसंबंध गुणांक उपयुक्त होता है। लेकिन ऐसी स्थितियाँ भी हो सकती हैं जहाँ चर संख्यात्मक नहीं होते लेकिन लक्षणों के आधार पर विभिन्न मदों को श्रेणीबद्ध (अर्थात् क्रमसूचक) किया जा सकता है। कभी–कभी मूल चरों के मापनीय होने पर भी उनको कोटि में परिवर्तित किया जाता है और साहचर्य माप परिकलित किया जाता है। उदाहरण के लिए, उस परिस्थिति पर विचार कीजिए, जिसमें दो परीक्षकों द्वारा 10 परीक्षार्थियों की, मौखिक परीक्षा के आधार पर जाँच करनी है। उस परिस्थिति में परीक्षार्थियों के अंक निर्धारित करना कठिन हो सकता है, लेकिन परीक्षकों के लिए उनको उनकी योग्यता के क्रम को कोटि करना सरल हो सकता है। इन परिणामों के प्रयोग से पहले यह ज्ञात करना उपयुक्त होगा कि क्या परीक्षकों द्वारा की गई कोटि (ranking) में उचित सामंजस्य है। इसके लिए दो परीक्षकों के बीच साहचर्य का माप परिकलित किया जा सकता है। इस परिस्थिति में सहसंबंध गुणांक का प्रयोग उपयुक्त नहीं है। यहाँ पर **स्पीयरमैन का कोटि सहसंबंध गुणांक** प्रयोग किया जा सकता है।

सारणी 8.3: दो परीक्षकों द्वारा 10 परीक्षार्थियों की कोटि का निर्धारण

क्र.सं.	परीक्षक I	परीक्षक II	D_i	अंतर D_i^2
1	6.0	6.5	–0.5	0.25
2	2.0	3.0	–1.0	1.00
3	8.5	6.5	2.0	4.0
4	1.0	1.0	0.0	0.00
5	10.0	2.0	8.0	64.00
6	3.0	4.0	–1.0	1.00
7	8.5	9.5	–1.0	1.00
8	4.0	5.0	–1.0	1.00
9	5.0	8.0	–30	9.00
10	7.0	9.5	–2.5	6.25
			$\sum D_i = 0$	$\sum D_i^2 = 87.50$

आइए अब सारणी 8.3 में दिए गए आँकड़ों पर विचार करें। यहाँ पर कुछ कोटियों में साम्य हे। इन साम्यावस्थाओं को एक ही कोटि इस प्रकार दी जाती है कि इनका योग उतना ही रहे, जितना साम्य न होने पर होता। उदाहरण के लिए, अगर दो अवस्थाओं की समान कोटि 6 है तो प्रत्येक को 6.5 कोटि दी जाती है तथा किसी अवस्था को 6 या 7 कोटि नहीं दी जाती। इसी प्रकार अगर 5 कोटि की तीन अवस्थाएँ हैं तो प्रत्येक को 6 की कोटि दी जाती है तथा किसी अवस्था को 5 या 7 की कोटि नहीं दी जाती। *स्पीयरमैन का कोटि–सहसंबंध गुणांक,* जिसको स्पीयरमैन का रो (Rho) कहा जाता है, ρ से सूचित किया जाता है। यह दोनों प्रकार की कोटि के अंतर D पर आधारित होता है। यदि दोनों कोटियाँ पूर्णतया संपाती हैं तो हर स्थिति में D_i शून्य होगा। D_i का मान जितना अधिक होगा, दो कोटियों के बीच का अंतर भी उतना ही अधिक होगा और साहचर्य कम होगा।

अतः D_i के आकार द्वारा साहचर्य का माप किया जा सकता है। चूंकि विभिन्न व्यष्टियों के लिए D का योग हमेशा शून्य होता है इसलिए D के आधार पर एक अकेला सूचकांक ज्ञात करने के लिए D की धनात्मकता या ऋणात्मकता को दूर करना होगा तथा सिर्फ D के आकार को ही लेना होगा। स्पीयरमैन के ρ में यह कार्य D का वर्ग लेकर किया जाता है। फिर भी यहाँ पर $\sum_{i=1}^{n} D_i^2$ का मान अधिक होगा या कम, यह n पर अर्थात व्यष्टियों की संख्या पर निर्भर करता है। व्याख्या के लिए हम इसको अधिकतम संभव मान, जोकि सिर्फ n पर निर्भर है, द्वारा भाग करके एक अनुपात बना सकते हैं। यह अधिकतम मान $\frac{n(n^2-1)}{6}$ है। इस प्रकार, $\frac{6 \times \sum_{i=1}^{n} D_i^2}{n(n^2-1)}$ का मान पूर्ण साहचर्य के लिए शून्य है और साहचर्य न होने पर 2 के बराबर है। लेकिन हम इसको अन्य प्रकार से रखना पसंद करेंगे। इसके लिए हम इसको एक में से घटा देते हैं। अतः

$$\rho = 1 - \frac{6 \times \sum_{i=1}^{n} D_i^2}{n(n^2-1)} \qquad \text{... (8.8)}$$

आइए सारणी 8.3 में दिए गए आंकड़ों से ρ का मान परिकलित करें;

$$\rho = 1 - \frac{6 \times 87.5}{10\,(10^2-1)} = 1 - \frac{525}{990} = 1 - 0.53 = 0.47$$

कार्ल पियर्सन के सहसंबंध गुणांक की भांति, स्पीयरमैन कोटि सह संबंध में कोटियों का पूर्ण मेल है तो मान +1 और पूर्णतया मेल न होने की स्थिति में मान –1 और कोटियों के बीच संबंध न होने की स्थिति में शून्य मान को व्यक्त करेगा।

जब चर नामिक, क्रमसूचक तथा दूसरे प्रकार के हों तो साहचर्य के अन्य उपयुक्त माप प्रयोग किए जा सकते हैं। लेकिन यहाँ पर हम उनका विवेचन नहीं करेंगे।

बोध प्रश्न 2

1) एक प्रतियोगिता में दो निर्णायकों ने 8 प्रतियोगियों A, B, C, D, E, F, G और H को अपने वरीयता क्रम के अनुसार निम्नलिखित तालिका में दी गई कोटियाँ दी हैं। कोटि सहसंबंध गुणांक ज्ञात कीजिए।

	A	B	C	D	E	F	G	H
पहला निर्णायक	5	2	8	1	4	6	3	7
दूसरा निर्णायक	4	5	7	3	2	8	1	6

....

....

....

....

2) दो परीक्षाओं में विद्यार्थियों के एक समूह को निम्नलिखित कोटियों के लिए सहसंबंध गुणांक परिकलित कीजिए। इस परिणाम से आप क्या निष्कर्ष निकालते हैं?

रोल नं.	1	2	3	4	5	6	7	8	9	10
बी.कॉम परीक्षा में कोटि	1	5	8	6	7	4	2	3	9	10
एम.कॉम परीक्षा में कोटि	2	1	5	7	6	3	4	8	10	9

....

....

....

....

....

3) तीन निर्णायकों A, B, C ने एक संगीत प्रतियोगिता में दस प्रतियोगियों को निम्नलिखित क्रम में कोटिबद्ध किया:

A द्वारा कोटि	1	6	5	10	3	2	4	9	7	8
B द्वारा कोटि	3	5	8	4	7	10	2	1	6	9
C द्वारा कोटि	6	4	9	8	1	2	3	10	5	7

निर्णायकों का कौन सा युग्म संगीत की सामान्य रुचि के निकटतम सादृश्य है? कोटि सहसंबंध विधि के प्रयोग द्वारा विवेचन कीजिए।

....

....

....

....

....

4) दस विद्यार्थियों द्वारा गणित और सांख्यिकी में निम्नलिखित अंक प्राप्त किए गए। कोटि सहसंबंध गुणांक परिकलित कीजिए।

विद्यार्थी (रोल नं.)	1	2	3	4	5	6	7	8	9	10
गणित में अंक	78	36	98	25	75	82	90	62	65	29
संख्यिकी में अंक	84	51	91	60	68	62	86	58	53	47

....

....

....

....

....

....

8.7 सारांश

इस इकाई में आपने प्रकीर्ण आरेख और सहप्रसरण के बारे में जानकारी प्राप्त की। इसके अलावा आपने सहसंबंध गुणांक और कोटि सहसंबंध गुणांक का भी अध्ययन किया जो दो चरों के बीच सहसंबंध या रैखिक साहचर्य की समीपता को, बिना कारण–प्रभाव संबंध पर ध्यान दिए, दर्शाता है।

8.8 शब्दावली

सहसंबंध विश्लेषण (Correlation Analysis) : इससे अर्थ दो यादृच्छिक चरों के बीच साहचर्य का परिमाप है। जो दो यादृच्छिक चर इस प्रकार के हैं कि एक में परिवर्तन से दूसरे से संबंधित तरीके से परिवर्तन होता है तो इनको सहसंबंधित कहते हैं। जो चर स्वतंत्र होते हैं, वे सहसंबंधित नहीं होते। सहसंबंध गुणांक –1 तथा +1 के बीच एक संख्या होती है। यह प्रेक्षणों के बहुत से युग्मों, जिनको बिंदु (X,Y) से सूचित किया जाता है, से परिकलित किया जाता है। जब गुणांक का मान +1 है तो इसका अर्थ पूर्ण धनात्मक सहसंबंध, गुणांक का मान –1 है तो इसका अर्थ पूर्ण ऋणात्मक सहसंबंध तथा गुणांक का मान 0 है तो इसका अर्थ कोई सहसंबंध नहीं होता है।

सहप्रसरण (Covariance): यह दो चरों का उनके माध्य से प्रथम गुणन आघूर्ण (First product Moment) होता है। इसे परिकलित करने का सूत्र है :

$$\frac{1}{n}\sum\left(X_i-\overline{X}\right)\left(Y_i-\overline{Y}\right. \quad \text{या} \quad \frac{1}{n}\left(\sum X_iY_i-\frac{\sum X_iY_i}{n}\right)$$

जहाँ X, Y प्रत्येक चर के मान हैं तथा 'n' प्रेक्षणों की संख्या है।

कोटि सहसंबंध गुणांक (Rank Correlation Coefficient)

बहुत सी परिस्थितियों में चरों का माप प्राप्त करना, सुविधाजनक अथवा कम खर्चीला नहीं होता। कई बार तो यह संभव नहीं होता। ऐसी स्थिति में उनको क्रम के अनुसार कोटिबद्ध किया जाता है। इन परिस्थितियों में कोटि सहसंबंध गुणांक का प्रयोग किया जा सकता है। जब चरों में अरैखिक संबंध हो तो भी कोटि सहसंबंध गुणांक उपयुक्त होता है।

प्रकीर्ण आरेख (Scatter Diagrams): ऐसा आरेख है जो दो चरों X और Y के बीच संयुक्त परिवर्तन को दर्शाता है। प्रत्येक व्यष्टि को एक बिंदु द्वारा निरूपित किया जाता है जिसके साधारण आयताकार अक्षों पर निर्देशांक, चरों के मान होते हैं। इस प्रकार n प्रेक्षणों को समुच्चय, आरेख पर n बिंदु प्रदान करता है। इन बिंदुओं का प्रकीर्ण X तथा Y के बीच संबंध को दर्शाता है।

8.9 कुछ उपयोगी पुस्तकें

Nagar, A.L. and R.K. Das, 1989 : *Basic Statistics,* Oxford University Press, Delhi.

Goon, A.M., M.K. Gupta and B. Dasguprta, 1987 : *Basic Statistics,* The World Press Pvt. Ltd., Calcutta.

8.10 बोध प्रश्नों के उत्तर या संकेत

बोध प्रश्न 1

1) +0.47
2) +0.996
3) +0.98
4) –0.84

बोध प्रश्न 2

1) 2/3
2) +0.64
3) –0.21, +0.64, –0.30
4) +0.82

इकाई 9 समाश्रयण विश्लेषण

इकाई रूपरेखा

9.0 उद्देश्य

इस इकाई को पढ़ने के बाद, आपः

- समाश्रयण की संकल्पना को समझा सकेंगे;
- न्यूनतम वर्ग विधि-को व्यक्त कर सकेंगे;
- रैखिक समाश्रयण की सीमाओं की पहचान कर सकेंगे;
- दिए गए आंकड़ों पर रैखिक समाश्रयण निदर्शों को लागू कर सकेंगे; और
- प्रागुक्ति के लिए समाश्रयण समीकरण का प्रयोग कर सकेंगे।

9.1 प्रस्तावना

पिछली इकाई में आपने देखा कि सहसंबंध गुणांक दो चरों के बीच कारण और प्रभाव संबंध को प्रतिबिंबित नहीं करता। अतः हम एक चर के लिए दिए हुए मान के अनुरूप अन्य चर के मान की प्रागुक्ति नहीं कर सकते। लेकिन समाश्रयण विश्लेषण (regression aralysis) के माध्यम से हम इस दोष को दूर करते हैं। इस इकाई में हम समाश्रयण विश्लेषण की चर्चा करेंगे जिससे चरों के बीच के संबंध को गणितीय समीकरण के रूप में अभिव्यक्त किया जाता है। इसमें हम मान लेते हैं कि एक चर कारण है और दूसरा प्रभाव। आपको याद होना चाहिए कि समाश्रयण एक सांख्यिकीय उपकरण है जो चरों के बीच के संबंध

को समझने में सहायक होता है और जो स्वतंत्र चर के ज्ञात मानों से आश्रित चर के अज्ञात मानों की प्रागुक्ति करता है।

9.2 समाश्रयण की संकल्पना

समाश्रयण विश्लेषण में दो प्रकार के चर होते हैं : i) आश्रित (या वर्णित) चर, और ii) स्वतंत्र (या व्याख्यात्मक) चर। जैसा कि इनके नाम से इंगित है, स्वतंत्र चर से आश्रित चर का विवरण दिया जाता है।

समाश्रयण विश्लेषण के सरलतम मामले में, एक आश्रित चर और एक स्वतंत्र चर होता है। आइए मान लेते हैं कि परिवार का उपभोग व्यय, परिवार की आय से संबंधित है। जैसे, मान लेते हैं कि पारिवारिक आय बढ़ने के साथ–साथ खर्च में भी बढ़ोतरी होती है। इस संदर्भ में उपभोग व्यय आश्रित चर है और पारिवारिक आय स्वतंत्र चर है।

आमतौर पर हम आश्रित चर को Y और स्वतंत्र चर को X से दर्शाते हैं। मान लीजिए हमने पारिवारिक सर्वेक्षण किया और X और Y में n प्रेक्षण युग्मों को इकठ्ठा किया। अब हमारा अगला चरण, X और Y के बीच के संबंध की प्रकृति का पता लगाना है। X और Y के बीच का संबंध अलग–अलग रूपों का हो सकता है। आम व्यवहार में इस संबंध को किसी गणितीय समीकरण से अभिव्यक्त किया जाता है। इन समीकरणों में से सरलतम, रैखिक समीकरण है। इसका अर्थ है कि X और Y के बीच का संबंध सरल रेखा में है और इसे रैखिक समाश्रयण कहते हैं। जब समीकरण (सरल रेखा न होकर) वक्रों को दर्शाता है तो इसे अरैखिक या वक्ररेखी समाश्रयण (non-linear regression) कहते हैं।

अब प्रश्न उठता है कि, 'समीकरण के रूप की पहचान हम कैसे करते हैं ?' इसके लिए कोई विशेष नियम नहीं है। समीकरण का स्वरूप हमारी तार्किक सोच और कल्पनाशक्ति पर आधारित है। लेकिन प्रकीर्ण आरेख बनाने के लिए, हम X और Y चरों को ग्राफ पर खींच सकते हैं। प्रकीर्ण आरेख से हमें ग्राफ कागज़ पर बिंदुओं की स्थिति का पता चल जाता है जिससे समीकरण के रूप को पहचाना जा सकता है। यदि बिंदु लगभग सीधी रेखा में हैं तो रैखिक समीकरण बनेगा। दूसरी तरफ यदि बिंदु सीधी रेखा में नहीं हैं, बल्कि वक्र के रूप में है तो इसे उपयुक्त अरैखिक समीकरण बनेगा।

अब हमें एक बात और तय करनी है और वह है आश्रित और स्वतंत्र चरों की पहचान करना। यह बात भी दुबारा तर्क और विश्लेषण के उद्देश्य पर आधारित है कि क्या Y, X पर निर्भर है या X, Y पर निर्भर है। अतः आंकड़ों के एक ही समुच्चय से दो समाश्रयण समीकरणों की प्राप्ति की जा सकती है। ये हैं : i) Y को X पर आश्रित मान लिया गया है (इसे X रेखा पर Y के रूप में माना जाता है), और ii) X को Y पर आश्रित मान लिया गया है (इसे Y रेखा पर X के रूप में माना जाता है)।

समाश्रयण विश्लेषण को उन अवस्थाओं में भी प्रयोग किया जा सकता है जहां एक आश्रित चर को बहुत से स्वतंत्र चरों की संख्या से समझाया जाता है। ऐसे मामले को बहु–समाश्रयण (multiple regression) कहते हैं। उच्च समाश्रयण निदर्शों में बहुत से आश्रित और स्वतंत्र चर हो सकते हैं।

अब तक आप सोच रहे होंगे कि 'समाश्रयण' शब्द का प्रयोग क्यों किया गया है, क्योंकि इसका अर्थ तो घटाना या कम करना होता है। यह नाम एक घटना के साथ जुड़ा हुआ है, जोकि उस समय प्रेक्षित की गई जब इन धारणाओं को विकसित किया जा रहा था। पिता की ऊँचाई (X) तथा बेटे की ऊँचाई (Y) के संबंध में एक अध्ययन में यह प्रेक्षित किया

गया कि सबसे ऊँचे पिताओं के बेटों की औसत ऊँचाई इन पिताओं की औसत ऊँचाई से कम होने के प्रवृति है। इस तरह सबसे कम ऊँचाई वाले पिताओं के बेटों की औसत ऊँचाई इन पिताओं की औसत ऊँचाई से अधिक होने की प्रवृति है। इस घटना को **माध्य की तरफ समाश्रयण होना** कहा गया है। चाहे यह उस समय कुछ अजीब–सा महसूस हुआ हो, लेकिन बाद में यह पाया गया कि इसका कारण वर्ग के उप–वर्गों में प्राकृतिक प्रसरण है। इसी प्रकार की प्रक्रियाएं बहुत–सी समस्याओं तथा आंकड़ों में घटित हुई। इसकी व्याख्या यह है कि कुछ जननिक कारकों के अतिरिक्त, अनियमित प्राकृतिक परिवर्तनों के कारण बहुत से लंबे व्यक्ति औसत ऊँचाई के परिवारों से होते हैं तथा इनके बेटे कुल मिलाकर इनसे कम ऊँचाई के होते हैं। ठीक इसी प्रकार की प्रक्रिया पैमाने के निचले सिरे पर भी लागू होती है।

9.3 रैखिक संबंध : द्विचर स्थिति

X और Y के बीच, सरलतम संबंध शायद एकघाती निर्धारणात्मक फलन के रूप में है और जो है

$$Y_i = a + bX_i \qquad ...(9.1)$$

उपर्युक्त समीकरण में X स्वतंत्र चर या व्याख्यात्मक चर है और Y आश्रित चर या वर्णित चर है। आपको याद होगा कि पादांक i प्रेक्षण संख्या को दर्शाता है और i का दायरा 1 से n तक का है। अतः Y_1 आश्रित चर का पहला प्रेक्षण है और X_5 स्वतंत्र चर का पाँचवा प्रेक्षण है और इसी तरह आगे भी।

समीकरण (9.1) से पता चलता है कि Y पूर्णतया X और a और b प्राचलों से निर्धारित है। मान लीजिए हमारे पास प्राचल मान हैं : $a = 3$ और $b = 0.75$, तब हमारा रैखिक समीकरण होगा, $Y = 3 + 0.75\,X$।

इस समीकरण से हम X के दिए गए मानों के लिए, Y का मान ज्ञात कर सकते हैं। उदाहरण के रूप में, जब $X = 8$ तो हम $Y = 9$ पाते है। अतः यदि हमारे पास X के विभिन्न मान हैं तब (9.1) के आधार पर हम Y के संगत मानों की प्राप्ति करते हैं। यही नहीं, यदि दो प्रेक्षणों के लिए X_i समान है तो Y_i के मान भी, दोनों प्रेक्षणों के लिए एक जैसे ही होंगे। X पर Y वाले आलेख से पता चलता है कि सीधी रेखा से कोई विचलन नहीं है।

यदि (9.1) से प्राप्त निर्धारणात्मक निदर्श की ओर हम ध्यान दें तो हम पाएंगे कि चरों के बीच आर्थिक अंतः संबंध व्यक्त करने में यह उपयुक्त नहीं है। जैसे, मान लीजिए $Y =$ उपभोग व्यय और $X =$ परिवारों की आय है। मान लीजिए आप उत्तरोत्तर महीनों के लिए अपनी आय ओर उपभोग व्यय का रिकार्ड रखते हैं। जिन महीनों में आपकी आमदनी एक जैसी ही रहती है, क्या उन महीनों में उपभोग व्यय भी एक जैसा ही रहता है? यहां हम यह समझाने का प्रयास कर रहे हैं कि आर्थिक संबंध में कुछ निश्चित अनियमितताओं का समावेश रहता है।

इसलिए, हम मान लेते हैं कि Y और X के बीच का संबंध यादृच्छित है और हम (9.1) में एक विभ्रम जोड़ देते हैं। अतः हमारा यादृच्छिक प्रतिमान होगा :

$$Y_i = a + bX_i + e_i \qquad ...(9.2)$$

जहां e_i एक विभ्रम है। जीवन की वास्तविक स्थितियों में e_i मानवीय व्यवहार में होने वाली अनियमितताओं और वर्जित चरों (यदि प्रतिमान में हो तो) को दर्शाता है। याद रखिए कि

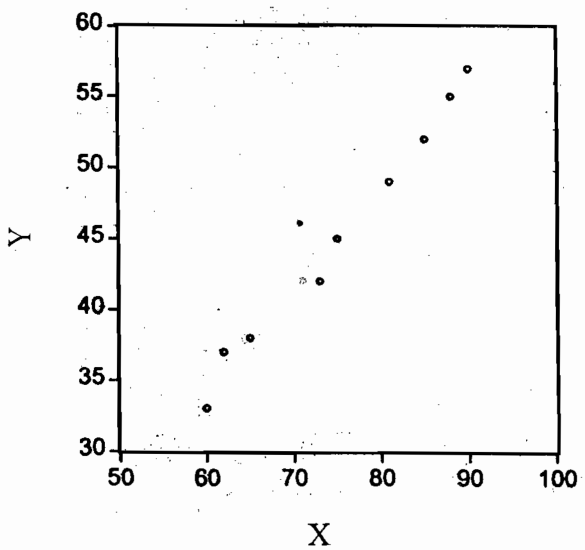
(9.2) के दायें तरफ के दो भाग हैं अर्थात: i) निर्धारणात्मक भाग (जो है, $a + bX_i$) और ii) यादृच्छिक या अनियमित भाग (अर्थात e_i)। समीकरण (9.2) से पता चलता है कि यदि दो प्रेक्षणों के लिए X पहले जैसा ही रहता है तो अलग–अलग e_i के कारण यह आवश्यक नहीं है कि Y भी वैसा ही रहे। अतः हम (9.2) को आरेख पेपर पर खींचते हैं तो प्रेक्षण सीधी रेखा में नहीं रहेंगे।

उदाहरण 9.1

दस वर्षों में हुई वर्षा और कृषि उत्पादन का ब्यौरा, सारणी 9.1 में दिया गया है;

सारणी 9.1 : वर्षा और कृषि उत्पादन

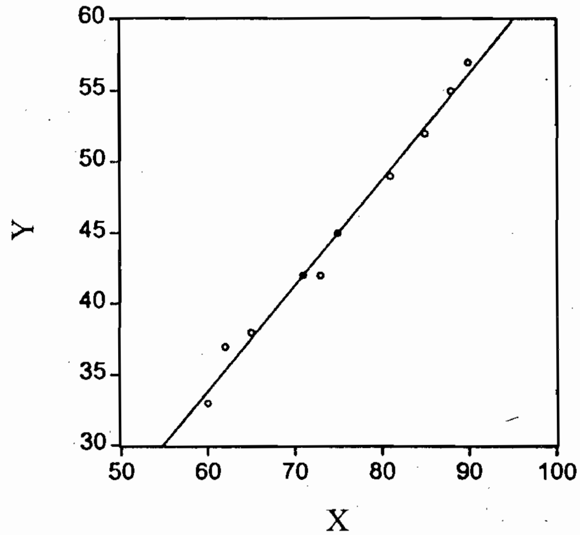
वर्षा, मि. मीटर में	कृषि उत्पादन टनों में
60	33
62	37
65	38
71	42
73	42
75	45
81	49
85	52
88	55
90	57



चित्र 9.1 : प्रकीर्ण आरेख

अब हम इस आंकड़ों का ग्राफ बनाते हैं। प्रकीर्ण आलेख, चित्र 9.1 की भांति नज़र आयेगा। चित्र 9.1 पर ग़ौर करने से हमें पता चलता है कि बिंदु सीधी रेखा में नहीं हैं। लेकिन ऊपर की तरफ बढ़ते हुए वे इस प्रकार प्रवृत है कि उन्हें जोड़ने से सीधी सी रेखा नज़र आयेगी।

आइए अब प्रकीर्ण आलेख के साथ समाश्रयण रेखा भी खींचें।



चित्र 9.2 : समाश्रयण रेखा

समाश्रयण रेखा और प्रेषणों के बीच का ऊर्ध्वाधर अंतर विभ्रम e_i है। समाश्रयण रेखा के संगत मान को प्रागुक्ति मान या प्रत्याशित मान कहते हैं। दूसरी तरफ, स्वतंत्र चर के किसी विशिष्ट मान से संगत करने वाले आश्रित चर के वास्तविक मान को प्रेक्षित मान कहते हैं। अतः विभ्रम से आशय, प्रागुक्ति मान और प्रेक्षित मान के बीच का अंतर है।

अब प्रश्न उठता है कि, 'हम समाश्रयण रेखा की प्राप्ति कैसे करते हैं ? आंकड़ों के लिए सीधी रेखा बनाने की क्रियाविधि इस प्रकार है।

9.4 विभ्रमों का न्यूनतमीकरण

जैसा कि हमने पहले उल्लेख किया था, सीधी रेखा को इस समीकरण द्वारा निरूपित किया जाता है

$$Y_i = a + bX_i$$

जहां b ढाल (slope) है और a, y-अक्ष पर अंतः खंड (intercept) है। सीधी रेखा की अवस्थिति a और b के मानों पर निर्भर करती हैं, जिन्हें प्राचल (parameter) कहते हैं। अतः अब हमें एकत्रित आंकड़ों से इन प्राचलों का **आकलन** (estimation) करना है ('खंड 7: आकलन की संकल्पना' में आप इससे संबंधित अधिक जानकारी प्राप्त करेंगे)। आंकड़ों के अनुरूप सर्वाधिक श्रेष्ठ रेखा प्राप्त करने के लिए हमें a और b के आकलन इस तरीके से प्राप्त करने होंगे ताकि विभ्रम e_i न्यूनतम हो।

जैसा कि हमने पहले माना था, n प्रेक्षणों को (X_i, Y_i) से सूचित किया जाएगा जहां पर $i = 1, 2, \ldots\ldots, n$ है। कृषि उत्पादन और वर्षा से संबंधित उदाहरण 9.1 में $n = 10$ है। मान लीजिए हम X_i पर Y_i के प्रागुक्ति मान को $\hat{Y}_i$ से सूचित करते हैं (संकेत $\hat{Y}_i$ को 'Y_i-cap'

या 'Y_i-hat' के रूप में पढ़ा जाता है)। अतः

$\hat{Y}_i = a + bX_i,\ i = 1, 2, ..., n.$

तब iवें प्रेक्षित बिंदु पर विभ्रम होगा

$$e_i = Y_i - \hat{Y}_i \qquad ...(9.3)$$

सबसे अच्छा तो यह होगा कि हम a और b के ऐसे मान प्राप्त करें जिससे प्रत्येक e_i, $i = 1, 2,......n$, शून्य के बराबर हो जायें। लेकिन, यह तब तक असंभव है, जब तक सभी बिंदु सरल रेखा पर न हों, जिसकी संभावना बहुत कम है। अतः हमें e_i, $i = 1, 2,......n$, के संचय को न्यूनतम करने से ही संतुष्ट होना पड़ेगा। हमारे समक्ष कौन से विकल्प हैं ?

- यह सोचना आकर्षक लगता है कि e_i, $i = 1, 2,......, n$, के कुल योग अर्थात $\sum_{i=1}^{n} e_i$ उचित विकल्प है। लेकिन ऐसा नहीं है, क्योंकि रेखा के ऊपर बिंदुओं के लिए e_i का मान धनात्मक तथा रेखा से नीचे बिंदुओं के लिए e_i का मान ऋणात्मक है। अतः अगर विभ्रम बड़े धनात्मक तथा बड़े ऋणात्मक भी हो तो यह संभव है कि $\sum_{i=1}^{n} e_i$ बहुत छोटा होगा।

- दूसरी संभावना है कि यदि हम $a = \bar{Y}$ (Y_i का समांतर माध्य) और $b = 0$, ले तो $\sum_{i=1}^{n} e_i$ को शून्य के बराबर किया जा सकता है। इस स्थिति में, हमें सभी प्रागुक्तियों के लिए X के मान की आवश्यकता नहीं पड़ती। X के प्रेक्षित मान के बावजूद भी प्रागुक्ति मान समान रहता है। यह प्रामाणिक रूप से ग़लत है।

- तो $\sum_{i=1}^{n} e_i$ निकष में कहाँ गड़बड़ है ? इसमें मुख्य गड़बड़ यह है कि e_i के चिह्न का हिसाब रखता है, जबकि यहां विभ्रम का आकार महत्वपूर्ण है। विभ्रम चाहे धनात्मक हो या ऋणात्मक, वास्तव में कोई महत्व नहीं रखता। अतः $\sum_{i=1}^{n} |e_i|$ न्यूनतम करने में अधिक उपयुक्त निकष है। याद रखें कि $|e_i|$ का अर्थ, e_i के पूर्ण मान (abcolute value) से है। अतः, यदि $e_i = 5$ है तब $|e_i| = 5$ है और यदि $e_i = -5$ है तब भी $|e_i| = 5$ है। लेकिन, इस विकल्प में परिकलन संबंधी कुछ समस्याएं नज़र आ रही हैं।

- सैद्धांतिक और परिकलन संबंधी कारणों से इस निकष की तुलना से न्यूनतम वर्ग निकष को वरीयता दी जाती है। जबकि निरपेक्ष मान निकष में, e_i के चिह्न को इसके निरपेक्ष मान लेकर हटा दिया जाता है। न्यूनतम वर्ग निकष (least squares criterion) में इसके वर्गन से ऐसा किया जाता है। याद रखें कि 5 और –5 दोनों का वर्ग 25 है। इस युक्ति को गणितीय और परिकलन की दृष्टि से अधिक उपयुक्त पाया गया है।

निम्नलिखित अनुभाग में हम न्यूनतम वर्ग विधि की विस्तृत जानकारी देंगे।

9.5 न्यूनतम वर्ग विधि

न्यूनतम वर्ग विधि (method of least squares) में हम विभ्रम मदों के वर्ग के योग अर्थात्, $\sum_{i=1}^{n} e_i^2$ को न्यूनतम बनाते हैं।

(9.1) से, हम पाते हैं $e_i = Y_i - \hat{Y}_i$

जो दर्शाता है $e_i = Y_i - (a + bX_i) = Y_i - a - bX_i$

$$\text{अतः } \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - a - bX_i)^2 \qquad ...(9.4)$$

अगला प्रश्न है : हम (9.4) को कम से कम करने के लिए a और b के मानों की प्राप्ति किस प्रकार करते हैं ?

- आप में से जो विद्यार्थी अवकलन (differentiation) की संकल्पना से अवगत हैं, उन्हें याद होगा कि किसी फलन का मान तब न्यूनतम होता है जब फलन का पहला अवकलज शून्य और दूसरा अवकलज धनात्मक हो। यहां, हमें a और b के मानों का चयन करना है। अतः $\sum_{i=1}^{n} e_i^2$ न्यूनतम होगा जब a और b के सापेक्ष इसके आंशिक अवकलज शून्य हों। $\sum_{i=1}^{n} e_i^2$ के आंशिक अवकलजों की प्राप्ति इस प्रकार की जाती है:

$$\frac{\partial \sum_i e^2}{\partial a} = \frac{\partial \sum_i (Y_i - a - bX_i)^2}{\partial a} = 2.(-1).\sum_i (Y_i - a - bX_i) \qquad ...(9.5)$$

$$\frac{\partial \sum_i e_i^2}{\partial b} = \frac{\partial \sum_i (Y_i - a - bX_i)^2}{\partial b} = 2.(-X_i).\sum_i (Y_i - a - bX_i) \qquad ...(9.6)$$

(9.5) और (9.6) को शून्य के बराबर करने और पदों के हेरफेर से, हम निम्नलिखित दो समीकरणों को प्राप्त करते हैं।

$$\sum_{i=1}^{n} Y_i = na + b\sum_{i=1}^{n} X_i \qquad ...(9.7)$$

$$\sum_{i=1}^{n} X_i Y_i = a\sum_{i=1}^{n} X_i + b\sum_{i=1}^{n} X_i^2 \qquad ...(9.8)$$

(9.7) और (9.8) अर्थात इन दोनों समीकरणों को न्यूनतम वर्ग के प्रसामान्य समीकरण (normal equation) कहते हैं। ये दो अज्ञातों में दो युगपत रैखिक समीकरण हैं। इन्हें a तथा b के मानों की प्राप्ति के लिए हल किया जा सकता है।

- आप में से जो अवकलन की संकल्पना से अवगत नहीं है, उन्हें व्यावहारिक ज्ञान की प्राप्ति करनी चाहिए (हमारा सुझाव है कि आपको अवकलन की संकल्पना सीखनी चाहिए जो अर्थशास्त्र के क्षेत्र में काफी उपयोगी है)। हम कह सकते हैं कि (9.7) और (9.8) में दिए गए प्रसामान्य समीकरणों को प्राप्त करने के लिए, रैखिक समीकरण को a और b के गुणांकों से गुणा कीजिए और सभी प्रेक्षणों के योगफल कीजिए। यहां रैखिक समीकरण $Y_i = a + bX_i$ है। पहला प्रसामान्य समीकरण साधारण रूप से रैखिक समीकरण $Y_i + a + bX_i$ का योग है (क्योंकि a का गुणांक 1 है)

$$\sum Y_i = \sum a + \sum bX_i \text{ या } \sum Y_i = na + b\sum X_i$$

दूसरा प्रसामान्य समीकरण, X_i द्वारा गुणित रैखिक समीकरण का योग है (क्योंकि b का गुणांक X_i)

$$\sum X_iY_i = \sum aX_i + \sum bX_i^2 \text{ या } \sum X_iY_i = a\sum X_i + b\sum X_i^2$$

प्रसामान्य समीकरणों को प्राप्त करने के बाद, हम मौजूद आंकड़ों के समुच्चय से a और b के मानों को परिकलित करते हैं।

उदाहरण 9.2: मान लीजिए कृषि उत्पादन की मात्रा वर्षा की मात्रा पर निर्भर करती है। उदाहरण 9.1 में दिए गए आंकड़ों से रैखिक समाश्रयण बनाइए।

इस मामले में आश्रित चर (Y), कृषि उत्पादन की मात्रा है और स्वतंत्र चर (X), वर्षा की मात्रा है। फिट किये जाने वाले समाश्रयण समीकरण है

$$Y_i = a + bX_i + e_i$$

उपर्युक्त समीकरण के लिए हम न्यूनतम वर्ग विधि द्वारा प्रसामान्य समीकरण ज्ञात करते हैं। ये समीकरण (9.7) और (9.8) में दिए गए हैं। अब निम्नलिखित के अनुसार सारणी बनाइए।

तालिका 9.2 : समाश्रयण रेखा का परिकलन

X_i	Y_i	X_i^2	X_iY_i	$\hat{Y}_i$	e_i
60	33	3600	1980	33.85	-0.85
62	37	3844	2294	35.34	1.66
65	38	4225	2470	37.57	0.43
71	42	5041	2982	42.03	-0.03
73	42	5329	3066	43.51	-1.51
75	45	5625	3375	45.00	0.00
81	49	6561	3969	49.46	-0.46
85	52	7225	4420	52.43	-0.43
88	55	7744	4840	54.66	0.34
90	57	8100	5130	56.15	0.85
$\sum_i X_i=750$	$\sum_i Y_i = 450$	$\sum_i X_i^2=57294$	$\sum_i X_iY_i=34526$	$\sum_i \hat{Y}_i = 450$	$\sum_i e_i=0$

प्रसामान्य समीकरण (9.7) और (9.8) में सारणी (9.2) से मान रखने पर हमें निम्नलिखित की प्राप्ति होती है :

$$450 = 10a + 750b$$

$$34526 = 750a + 57294b$$

इन समीकरणों को हल करने पर हमें $a = -10.73$ और $b = 0.743$ प्राप्त होता है।

इसलिए समाश्रयण रेखा, $\hat{Y}_i = -10.73 + 0.743X_i$ है।

ध्यान दीजिए कि प्रत्याशित समाश्रयण समीकरण के लिए विभ्रम योग $\sum_i e_i$, शून्य है (देखें सारणी 9.2 का अंतिम स्तंभ)

सारणी 9.2 में दिये गये परिकलन में प्रायः बड़ी–बड़ी संख्याएं शामिल होती हैं और इस कारण कठिनाई उत्पन्न हो सकती हैं। अतः प्रसामान्य समीकरणों से a और b के मानों के परिकलन के लिए हम लघुतर विधि का प्रयोग करेंगे।

आइए लें

$x = X - \bar{X}$ और $Y = Y - \bar{Y}$ जहां $\bar{X}$ और $\bar{Y}$, क्रमशः X और Y के समांतर माध्य हैं।

अतः $xy = (X - \bar{X})\,(Y - \bar{Y})$

प्रसामान्य समीकरणों में पदों को पुनः व्यवस्थित करने पर हम पाते हैं कि

$$b = \frac{\sum_{i=1}^{n} xy}{\sum_{i=1}^{n} x^2} \qquad \text{...(9.9)}$$

$$a = \bar{Y} - b\bar{X} \qquad \text{...(9.10)}$$

यदि आप इकाई 8 को ध्यान में लायें तो आपको पता होगा कि सहप्रसरण को इस प्रकार

दर्शाते हैं $\sigma_{xy} = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right) = \frac{1}{n}\sum_{i=1}^{n} X_i Y_i - \bar{X}\bar{Y}$

जबकि X का प्रसरण $\sigma_x^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2 = \sum_{i=1}^{n} x_i^2$ द्वारा दर्शाया जाता है।

चूंकि $b = \frac{\sum_{i=1}^{n} xy}{\sum_{i=1}^{n} x^2}$, हम कह सकते हैं कि $b = \frac{\sigma_{xy}}{\sigma_x^2}$ ...(9.11)

चूंकि इन सूत्रों को प्रसामान्य समीकरण से व्युत्पन्न किया जाता है, इसलिए इस विधि में हम a और b के लिए भी समान मान प्राप्त करते हैं। सारणी 9.1 में दिए गए आंकड़ों के लिए, इस विधि से हम a और b के मान परिकलित करते हैं। इस उद्देश्य के लिए, हम सारणी 9.3 का निर्माण करते हैं।

सारणी 9.3 : समाश्रयण रेखा का परिकलन (लघुतर विधि)

X_i	Y_i	x_i	y_i	x_i^2	$x_i y_i$
60	33	-15	-12	225	180
62	37	-13	-8	169	104
65	38	-10	-7	100	70
71	42	-4	-3	16	12
73	42	-2	-3	4	6
75	45	0	0	0	0
81	49	6	4	36	24
85	52	10	7	100	70
88	55	13	10	169	130
90	57	15	12	225	180
कुल 750	450	0	0	1044	776

सारणी 9.3 के आधार पर हम पाते हैं कि

$$\overline{X} = \frac{750}{10} = 75 \text{ और } \overline{Y} = \frac{450}{10} = 45$$

$$b = \frac{\sum_{i=1}^{n} xy}{\sum_{i=1}^{n} x^2} = \frac{776}{1044} = 0.743$$

$$a = \overline{Y} - b\overline{X} = 45 - 0.743 \times 10 = -10.73$$

अतः इस विधि में भी समाश्रयण रेखा है $\hat{Y}_i = -10.73 + 0.743X_i$ ...(9.12)

(9.12) में गुणांक b, समाश्रयण गुणांक कहलाता है। जब X में यूनिट बढ़ोतरी होती है तो समाश्रयण गुणांक Y में बढ़ने वाली संख्या को प्रभावित करता है। समाश्रयण समीकरण (9.12) में, गुणांक $b = 0.743$ दर्शाता है कि यदि वर्षा की मात्रा में 1 मिमी. से बढ़ोतरी होती है तो कृषि उत्पादन 0.743 हज़ार टन बढ़ जाएगा।

समाश्रयण गुणांक का व्यापक रूप से प्रयोग होता है। यह भी विश्लेषण का एक महत्वपूर्ण उपकरण है। जैसे, यदि Y कुल खपत और X कुल आय है तो b, सीमांत खपत प्रवृत्ति (एम.पी.सी.) को दर्शाता है।

9.6 प्रागुक्ति

समाश्रयण के अध्ययन में मुख्य रुचि, पूर्वानुमान करने की योग्यता में निहित है। पिछले अनुभाग के उदाहरण 9.1 में हमने माना था कि कृषि उत्पादन की मात्रा होने वाली वर्षा की मात्रा पर निर्भर करती है। हमने प्रेक्षित आकड़ों के लिए रैखिक समीकरण फिट किया और इस संबंध की प्राप्ति की

$$\hat{Y}_i = -10.73 + 0.743\,X_i$$

इस समीकरण से हम प्राप्त वर्षा की मात्रा से कृषि उत्पादन की मात्रा का पूर्वानुमान लगा सकते हैं। अब जब वषा 60 मिमी. है तो कृषि उत्पादन है

$(-10.73 + 0.74 \times 60) = 33.85$ हज़ार टन।

यह राशि समाश्रयण समीकरण पर आधारित *प्रागुक्ति मान* (predicted value) है। इसी ढ़ंग से हम X के विभिन्न मानों के लिए Y के प्रागुक्ति मानों की प्राप्ति कर सकते हैं।

प्रागुक्ति मान की प्रेक्षित मान से तुलना कीजिए। सारणी 9.1 से जहां प्रेक्षित मान दिए गए हैं हम पाते हैं कि जब वर्षा की मात्रा 60 मिमी. है तो कृषि उत्पादन 33 हज़ार टन है। असल में X के प्रेक्षित मानों के लिए प्रागुक्ति मान $\hat{Y}_i$, सारणी 9.2 के पांचवे कॉलम में दिए गए हैं। अतः जब वर्षा की मात्रा 60 मिमी. है तो प्रागुक्ति मान 33.85 हज़ार टन है। अतः विभ्रम मान $e_i = -0.85$ हज़ार टन है।

अब प्रश्न उठता है कि, प्रेक्षित और प्रागुक्ति मानों में से हमें किस पर विश्वास करना चाहिए? अन्य शब्दों में भविष्य में कृषि उत्पादन की मात्रा क्या होगी जब वर्षा की मात्रा 60 मिमी. है ? हमारे समाश्रयण रेखा के आधार पर यह 33.85 टन दिया गया है। और ये मान हमें स्वीकृत है, क्योंकि यह समग्र आंकड़ों पर आधारित है। –0.85 के विभ्रम यादृच्छिक अनियमितता है और जिसे दोहराया नहीं जा सकता।

हमारे मस्तिष्क में आने वाला दूसरा प्रश्न है कि, क्या X के किसी भी मान के लिए प्रागुक्ति मान्य है ? उदाहरण के लिए, समाश्रयण समीकरण से हम पाते हैं कि जब वर्षा न के बराबर अर्थात शून्य होगी, कृषि उत्पादन –10.73 हज़ार टन होगा। लेकिन हमारी सामान्य बुद्धि बताती है कि कृषि उत्पादन ऋणात्मक नहीं हो सकता ! क्या हमारे इस समाश्रयण समीकरण में कुछ त्रुटि है ? असल में यहां समाश्रयण समीकरण को 60 से 90 मिमी. की परिसर में होने वाली वर्षा के आंकड़े के आधार पर आकलित किया गया है। अतः X की इस परिसर में प्रागुक्ति मान्य वैध है। हमारा पूर्वानुमान X के दूरस्थ मानों के लिए नहीं होना चाहिए।

यहां उठने वाला तीसरा प्रश्न है कि, 'क्या प्रागुक्ति मान ही सही मान है ?' यह *निर्धारण गुणांक (coefficient of determination) पर* निर्भर करता है। यदि निर्धारण गुणांक 1 के सन्निकट है तो प्रागुक्ति के सही साबित होने की संभावना अधिक है। लेकिन, प्रागुक्ति कुछ ऐसे आकस्मिक अवयवों से बाधित रहती है जो मनुष्य के व्यवहार और कुछ अन्य अप्रत्याशित कारकों से संबंधित है।

9.7 समाश्रयण और सहसंबंध के बीच का संबंध

समाश्रयण विश्लेषण में दो चरों (X,Y) की स्थिति एक दूसरे से भिन्न होती है कि Y ऐसा चर है जिसका पूर्वानुमान करना है और X ऐसा चर है जिससे संबंधित जानकारी को प्रयोग में लाना है। वर्षा–कृषि उत्पादन समस्या में, वर्षा के आधार पर कृषि उत्पादन का पूर्वानुमान लगाना सही है। लेकिन इस बात में दम नहीं कि कृषि उत्पादन के आधार पर होने वाली वर्षा का पूर्वानुमान लगाने का प्रयास किया जाये। लेकिन, अर्थशास्त्र और सांख्यिकी में प्राप्त अंकों के मामले में (देखें इकाई 8 का उदाहरण 8.1), दोनों में से एक X और दूसरा Y होगा। अतः हम दो प्रागुक्ति संबंधी समस्याओं पर विचार करते हैं (i) सांख्यिकी में प्राप्त अंकों (X) के आधार पर अर्थशास्त्र में प्राप्त अंक (Y) का पूर्वानुमान लगाना; और (ii) अर्थशास्त्र में प्राप्त अंकों (Y) से सांख्यिकी में प्राप्त अंकों (X) का पूर्वानुमान लगाना ।

अतः स्वतंत्र और आश्रित चरों के चयन के आधार पर आंकड़ों के दिए गए समुच्चय से हमें दो समाश्रयण गुणांकों की प्राप्ति हो सकती है। ये हैं;

क) X रेखा पर Y, $Y_i = a + bX_i$

ख) Y रेखा पर X, $X_i = \alpha + \beta Y_i$

आप कह सकते हैं, 'कि दो अलग–अलग रेखाओं की प्राप्ति की आवश्यकता क्या है? X रेखा पर Y के पदों को पुनः क्रमबद्ध करने पर हम $X_i = \frac{a}{b} + \frac{1}{B} Y_i$ प्राप्त करते हैं। अतः हमारे पास $\alpha = \frac{a}{b}$ और $\beta = \frac{1}{b}$ होना चाहिए। लेकिन, ये प्रेक्षण सीधी रेखा में नहीं है और X और Y के बीच का संबंध गणितीय नहीं है। आपको याद होगा कि प्राचलों के आकलनों की प्राप्ति, न्यूनतम वर्ग विधि द्वारा की जाती है। अतः समाश्रयण रेखा $\hat{Y}_i = a + bX_i$ की प्राप्ति $\sum_i (Y_i - a - bX_i)^2$ को न्यूनतम करके की जाती है, जबकि समाश्रयण रेखा $\hat{X}_i = \alpha + \beta Y_i$ की प्राप्ति, $\sum_i (X_i - \alpha - \beta Y_i)^2$ को न्यूनतम करके की जाती है।

लेकिन, दो समाश्रयण गुणांकों b और β के बीच एक संबंध है। हमने पहले ध्यान दिया था कि $b = \frac{\sigma_{xy}}{\sigma_x^2}$। इसी सूत्र द्वारा और X और Y की भूमिकाओं में अदला–बदली करके हम $\beta = \frac{\sigma_{xy}}{\sigma_x^2}$ प्राप्त करते हैं। लेकिन यदि परिभाषा की ओर ध्यान दें तो हम देखते हैं कि $\sigma_{xy} = \sigma_{yx}$

अतः $b \times \beta = \frac{\sigma^2_{xy}}{\sigma^2_x \times \sigma^2_y}$ जो r^2 की ही भांति है।

इस r^2 को **निर्धारण गुणांक** कहते हैं। अतः दो समाश्रयण गुणांकों, Y का X पर तथा X का Y पर, का गुणनफल सहसंबंध और समाश्रयण के बीच एक और संबंध है। यहां पर ध्यान दें कि प्रत्येक समाश्रयण का निर्धारण गुणांक वही है अर्थात r^2 के बराबर है, जिसका अर्थ है कि चाहे दो समाश्रयण रेखाएं भिन्न हैं, उनकी प्रागुक्ति की शक्ति समान है। ध्यान दें कि निर्धारण गुणांक r^2 की सीमा 0 और 1 के बीच है अर्थात इसका अधिकतम मान 1 और न्यूनतम मान 0 हो सकता है; लेकिन ऋणात्मक नहीं हो सकता।

पिछली चर्चा से, दो बातें स्पष्ट होकर उभरी हैं :

1) अगर प्रकीर्ण आरेख में बिंदु, सरल रेखा के निकट हैं, तब X तथा Y के बीच एक अच्छा रैखिक संबंध होता है तथा सहसंबंध गुणांक भी उच्च होता है।

2) अगर प्रकीर्ण आरेख में बिंदु सरल रेखा के आसपास स्थित है, तब प्रेक्षित मानों और न्यूनतम वर्ग द्वारा प्रागुक्ति मान बहुत निकट होते हैं तथा प्रागुक्ति विभ्रम $(Y_i - \bar{Y}_i)$ कम होता है।

इस प्रकार, प्रतीत होता है कि न्यूनतम वर्ग द्वारा जनित प्रागुक्ति विभ्रम, सहसंबंध गुणांक से संबंधित हैं। हम यहां इस संबंध की व्याख्या करेंगे। न्यूनतम वर्ग रैखिक समाश्रयण के प्रयोग के कारण विभिन्न बिंदुओं पर विभ्रमों के वर्गों का योग इस प्रकार है;

$$\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

दूसरी तरफ, अगर हमने Y की प्रागुक्ति के लिए X का प्रयोग नहीं किया होता तो प्रागुक्ति, एक स्थिरांक (मान लीजिए), a होती। न्यूनतम वर्ग निकष के अनुसार, a का सर्वोत्तम मान वह होगा जिससे $\sum_{i=1}^{n}(Y_i - a)^2$ न्यूनतम हो, इस a का मान $\bar{Y}$ है। अतः X के प्रयोग के बिना विभिन्न बिंदुओं पर प्रागुक्ति विभ्रमों के वर्ग का योग $\sum_{i=1}^{n}(Y_i - \bar{Y}_i)^2$ है।

इन दोनों का अनुपात $\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \Big/ \sum_{i=1}^{n}(Y_i - \bar{Y})^2$ को एक सूचकांक के रूप में प्रयोग किया जा सकता है जोकि यह व्यक्त करेगा कि X के प्रयोग से कितना लाभ हुआ है। चूंकि इस अनुपात के दोनों अंश और हर, ऋणेतर (non-negative) हैं, इसलिए यह अनुपात शून्य के बराबर या इससे अधिक होगा।

बोध प्रश्न 1

1) निम्नलिखित आंकड़ों से X तथा Y में रैखिक सहसंबंध का गुणांक ज्ञात कीजिए। Y की X पर समाश्रयण रेखा भी ज्ञात कीजिए। जब $X = 12$ हो तो Y का अनुमान भी परिकलित कीजिए।

X :	1	3	4	6	8	9	11	14
Y :	1	2	4	4	5	7	8	9

....

....

....

....

....

....

2) निम्नलिखित आंकड़ों से समाश्रयण की रेखाएं प्राप्त कीजिए :

(X)	1	2	3	4	5	6	7	8	9
(Y)	9	8	10	12	11	13	14	16	15

....

....

....

....

....

....

3) निम्नलिखित आंकड़ों से दो समाश्रयण रेखाएं ज्ञात कीजिए :

पति की आयु (X)	25	22	28	26	35	20	22	40	20	18
पत्नी की आयु (Y)	18	15	20	17	22	14	16	21	15	14

तत्पश्चात (i) पति की आयु का अनुमान ज्ञात कीजिए, जबकि पत्नी की आयु, 19 हो, (ii) पत्नी की आयु का अनुमान ज्ञात कीजिए, जबकि पति की आयु 30 वर्ष हो।

....

....

....

....

....

....

4) निम्नलिखित आंकड़ों से दो समाश्रयण समीकरण प्राप्त कीजिए:

बिक्री :	91	97	108	121	67	124	51	73	111	57
ख़रीद :	71	75	69	97	70	91	39	61	80	47

....

....

....

....

....

....

5) निम्नलिखित सारणी में दिए हुए आंकड़ों से चावल की उपज (Y) की पानी (X) पर समाश्रयण रेखा का समीकरण प्राप्त कीजिए :

पानी (X) (इंचों में)	12	18	24	30	36	42	48
उपज (Y) (टनों में)	5.27	5.68	6.25	7.21	8.02	8.71	8.42

40 इंच पानी के लिए सबसे प्रायिक चावल की उपज का अनुमान कीजिए।

....

....

....

....

....

....

9.8 बहु-समाश्रयण

अभी तक हमने आश्रित चर की स्थिति पर विचार किया जिसे एक स्वतंत्र चर द्वारा स्पष्ट किया जाता है। लेकिन, ऐसे भी बहुत से मामले हैं जहां आश्रित चर की दो या अधिक स्वतंत्र चरों द्वारा व्याख्या की जाती है। जैसे फसलों की उपज (Y) को उर्वरकों (X_1) और सिंचाई जल (X_2) के इस्तेमाल से जोड़ा सकता है। इस प्रकार के निदर्शों को बहु समाश्रयण कहते हैं। यहां, विचाराधीन समीकरण है

$$Y = \alpha + \beta X_1 + \gamma X_2 + e \qquad \text{...... (9.13)}$$

जहां Y आश्रित चर है और X_1 और X_2 स्वतंत्र चर हैं और e विभ्रम पद है। प्रस्तुतीकरण को सरल बनाने के लिए हमने पादांकों को छोड़ दिया है। अनुभाग 9.5 में चर्चित न्यूनतम वर्ग विधि के प्रयोग द्वारा, समाश्रयण समीकरण को (9.13) के लिए फिट किया जा सकता है। यहां भी हम $\sum e^2$ को न्यूनतम बनाते हैं और निम्नलिखित प्रसामान्य समीकरण प्राप्त करते हैं।

$$\sum Y = n\alpha + \beta \sum X_1 + \gamma \sum X_2$$

$$\sum X_1 Y = \alpha \sum X_1 + \beta \sum X_1^2 + \gamma \sum X_1 X_2 \qquad \text{.....(9.14)}$$

$$\sum X_2 Y = \alpha \sum X_2 + \beta \sum X_1 X_2 + \gamma \sum X_2^2$$

उपर्युक्त समीकरणों को हल करने पर हमें α, β और γ के लिए आकलनों की प्राप्ति होती है। हमें प्राप्त होने वाला समाश्रयण समीकरण है :

$$\hat{Y} = \alpha + \beta X_1 + \gamma X_2 \qquad \text{.....(9.15)}$$

याद रखिए X_1 और X_2 के लिए विविध मानों के प्रयोग से, (9.15) के माध्यम से हम Y के प्रागुक्ति या पूर्वानुमानित मान (अर्थात $\hat{Y}$) की प्राप्ति करते हैं।

(Y, X) वाली द्विचर स्थिति में हम, ग्राफ कागज़ पर समाश्रयण रेखा बना सकते हैं। लेकिन ग्राफ कागज़ पर त्रिचर स्थिति (Y, X_1, X_2) बनाना थोड़ा जटिल है, क्योंकि इस स्थिति में तीन आयाम होंगे। लेकिन सहजबोधनीय विचार पहले जैसा ही है और हमें सभी विभ्रमों को न्यूनतम करना है। असल में जब हम सभी विभ्रम पद ($e_1, e_2, ...e_n$) को जोड़ते हैं तो इसका योग शून्य होगा।

बहुत सी स्थितियों में, व्याख्यात्मक चरों की संख्या दो से अधिक हो सकती है। ऐसी स्थिति में, हमें न्यूनतम वर्ग के मूलभूत सिद्धांत का अनुसरण करना होगा अर्थात $\sum e^2$ को न्यूनतम करना होगा। अतः, यदि $Y = a_0 + a_1 X_1 + a_2 X_2 + ... + a_n X_n + e$ तब हमें

$\sum e^2 = \sum (Y - a_0 - a_1 X_1 - a_2 X_2 .... - a_n X_n)^2$ को न्यूनतम करना होगा और प्रसामान्य समीकरण ज्ञात करना होगा।

अब प्रश्न उठता है कि "समाश्रयण समीकरण में कितने चर जोड़े जाने चाहिए ?" इस बात को हम अपनी तार्किक सोच पर छोड़ते हैं कि ऐसे महत्वपूर्ण समझे जाने वाले चर कौन से हैं। क्या सांख्यिकीय परीक्षणों के आधार पर चर की पहचान की जानी चाहिए? इन परीक्षणों की चर्चा खंड 7 में की जाएगी।

अब हम नीचे बहु समाश्रयण का एक उदाहरण प्रस्तुत करते हैं।

उदाहरण 9.2

एक विद्यार्थी विश्वविद्यालय के निकट मकानों के किराये की व्याख्या करना चाहती है। वह मासिक किराया, घर के क्षेत्रफल और विश्वविद्यालय परिसर से घर की दूरी से संबंधित आंकड़े एकत्र करती है और इन्हें, रैखिक समाश्रयण प्रतिमान में फिट करती है।

किराया ('000 रुपयों में) Y	क्षेत्रफल (वर्ग मीटर में) X_1	दूरी (किमी. में) X_2
20	65	5.7
25	66	3.2
26	70	7.5
28	70	6.5
30	75	5
31	76	4
32	72	6
33	75	6.2
35	78	3.5
40	103	2.4

उपर्युक्त उदाहरण में वसूल किया गया किराया (Y) आश्रित चर है, जबकि घर का क्षेत्रफल (X_1) और विश्वविद्यालय परिसर से घर की दूरी (X_2), स्वतंत्र चर है। समाश्रयण रेखा के आकलन में शामिल चरण हैं :

i) अनुमानित समाश्रयण समीकरण को ज्ञात कीजिए। इस स्थिति में, यह $Y = \alpha + \beta X_1 + \gamma X_2 + e$ द्वारा दर्शाया गया है।

ii) अनुमानित समाश्रयण समीकरण के लिए प्रसामान्य समीकरण ज्ञात कीजिए। इस स्थिति में प्रसामान्य समीकरण हैं :

$$\Sigma Y = n\alpha + \beta\Sigma X_1 + \gamma\Sigma X_2$$

$$\Sigma X_1 Y = \alpha\Sigma X_1 + \beta\Sigma X_1^2 + \gamma\Sigma X_1 X_2$$

$$\Sigma X_2 Y = \alpha\Sigma X_2 + \beta\Sigma X_1 X_2 + \gamma\Sigma X_2^2$$

iii) सारणी 9.4 की भांति, सारणी बनाइए।

iv) सारणी से प्राप्त मानों को प्रसामान्य समीकरण में रखें।

v) α, β और γ के आकलनों को हल कीजिए।

सारणी 9.4 : बहु–समाश्रयण का परिकलन

Y	X_1	X_2	X_1Y	X_2Y	X_1^2	X_2^2	X_1X_2	$\hat{Y}$	e_i
20	65	5.7	1300	114	4225	32.49	370.5	25.49	-5.49
25	66	3.2	1650	80	4356	10.24	211.2	25.71	-0.71
26	70	7.5	1820	195	4900	56.25	525	27.94	-1.94
28	70	6.5	1960	182	4900	42.25	455	27.85	0.15
30	75	5	2250	150	5625	25	375	30.00	0.00
31	76	4	2356	124	5776	16	304	30.37	0.63
32	72	6	2304	192	5184	36	432	28.72	3.28
33	75	6.2	2475	204.6	5625	38.44	465	30.11	2.89
35	78	3.5	2730	122.5	6084	12.25	273	31.24	3.76
40	103	2.4	4120	96	10609	5.76	247.2	42.58	-2.58
300	750	50	225000	15000	562500	2500	37500	300	0

उपर्युक्त उल्लिखित चरणों के अनुप्रयोग से, हम इस तरह की अनुमानित समाश्रयण रेखा प्राप्त करते हैं;

$Y = -4.80 + 0.45X_1 + 0.09X_2$

9.9 अरैखिक समाश्रयण

समाश्रयण में प्रयुक्त समीकरण अरैखिक या वक्ररेखी (non-linear) हो सकता है। दरअसल, इसके विविध रूप हो सकते हैं। ऐसा दो चरों वाला सरल रूप, द्विघाती समघात है। समीकरण है;

$$Y = a + bX + cX^2$$

यहां तीन प्राचल अर्थात a, b और c होंगे और प्रसामान्य समीकरण हैं :

$$\sum Y = n\alpha + b\sum X + c\sum X^2$$

$$\sum XY = \alpha\sum X + b\sum X^2 + c\sum X^2$$

$$\sum X^2Y = \alpha\sum X^2 + b\sum X^3 + c\sum X^4$$

इन समीकरणों के लिए हल करने पर हम a, b और c के मान प्राप्त करते हैं।

कुछ निश्चित अरैखिक समीकरणों को, लघुगणक (logarithm) के प्रयोग से रैखिक समीकरणो में परिवर्तित किया जा सकता है। परिवर्तित समीकरणों के प्राचलों के इष्टतम मानों की प्राप्ति करना लगभग वैसा ही है, जैसा कि पिछले अनुभाग में चर्चित प्रक्रिया में हमने बताया था। यहां हम कुछ सामान्य रूप से प्रयुक्त अरैखिक समीकरण और संबंधित परिवर्तित रैखिक समीकरणों पर प्रकाश डालते हैं;

1) $Y = a\,c^{bx}$

प्राकृतिक लघुगणक से, इसे हम इस तरह लिख सकते हैं;

$$\ln Y = \ln a + bX$$

या $Y' = \alpha + \beta X'$

जहां, $Y' = \ln Y$, $\alpha = \ln a$, $X' = X$ और $\beta = b$

2) $Y = a X^b$

सामान्य लघुगणक से समीकरण को इस तरह परिवर्तित किया जा सकता है :

$$\log Y = \log a + b \log X$$

या $Y' = \alpha + \beta X'$

जहां, $Y' = \log Y$, $\alpha = \log a$, $\beta = b$ और $X' = \log X$

3) $Y = \dfrac{1}{a + bX}$

यदि हम $Y' = \dfrac{1}{Y}$ लें, तब

$Y' = a + bX$

4) $Y = a + b\sqrt{X}$

यदि हम $X' = \sqrt{X}$ लें, तब

$Y = a + bX'$

एक बार अरैखिक समीकरण परिवर्तित हो जाएं तो समाश्रयण रेखा को फिट करना, उसी विधि की भांति है जिसकी चर्चा हमने अनुभाग 9.5 में की थी। इससे प्रसामान्य समीकरण की व्युत्पत्ति करते हैं और प्रेक्षित आंकड़ों से परिकलित मानों को उनमें भर देते हैं। रूपांतरित प्राचलों से, प्रतिलोम रूपांतर करके, वास्तविक प्राचलों की प्राप्ति की जा सकती है।

बोध प्रश्न 2

1) इकाई 8 में सारणी 8.1 में दिए सांख्यिकी और अर्थशास्त्र अंकों के आंकड़ों के प्रयोग द्वारा Y का X पर तथा X का Y पर समाश्रयण परिकलित कीजिए और यह जांच कीजिए कि दोनों रेखाएं भिन्न हैं। प्रकीर्ण आरेख पर दोनों समाश्रयण रेखाओं को अंकित कीजिए। यह जांच कीजिए कि समाश्रयण गुणांकों का गुणनफल सहसंबंध गुणांक का वर्ग है।

....

....

....

....

2) मान लीजिए वस्त्रों पर पारिवारिक व्यय (Y रुपए) का वार्षिक पारिवारिक आय (X रुपए) पर रैखिक समाश्रयण $Y = 100 + 0.09X$ प्राप्त किया गया। यहां पर X का परिसर $1000 < X < 1,00,000$ है। इस समांश्रयण रेखा की व्याख्या कीजिए। जब वार्षिक पारिवारिक आय 10,000 रुपये हो तो परिवार का वस्त्र पर व्यय की प्रागुक्ति कीजिए। जिन परिवारों की वार्षिक आय 100 रुपये तथा 10,00,000 रुपये है, उनके बारे में आपकी क्या प्रतिक्रिया है?

....

....

....

....

9.10 सारांश

इस इकाई में हमने एक महत्वपूर्ण सांख्यिकीय उपकरण अर्थात समाश्रयण की चर्चा की। समाश्रयण विश्लेषण में हमारे पास दो प्रकार के चर–आश्रित और स्वतंत्र चर होते हैं। आश्रित चर को स्वतंत्र चर द्वारा स्पष्ट किया जाता है। चरों के बीच का संबंध, गणितीय समीकरण का रूप ले लेता है। हमारी तार्किक सोच, समझ और विश्लेषण के उद्देश्य के आधार पर हम चरों को श्रेणीबद्ध करते हैं और समीकरण स्वरूप की पहचान करते हैं।

समाश्रयण गुणांक, स्वतंत्र चर के दिए गए मानों के अनुरूप आश्रित चर के मान की प्रागुक्ति करता है। लेकिन प्रागुक्ति, विश्लेषण में प्रयुक्त आंकड़ों के परिसर के भीतर

कमोबेश मान्य बनी रहती है। यदि हम स्वतंत्र चरों के दूर के मानों की प्रागुक्ति करने का प्रयास करते हैं तो हमें आश्रित चर के निरर्थक मानों की प्राप्ति होती है।

9.11 शब्दावली

निर्धारण गुणांक (Coefficient of Determination) : इसका मान r^2 अर्थात सहसंबंध गुणांक के वर्ग, के बराबर होता है। आश्रित चर में परिवर्तन की व्याख्या स्वतंत्र चर X के द्वारा की जाती है।

प्रसामान्य समीकरण (Normal Equation) : यह न्यूनतम वर्ग विधि द्वारा व्युत्पन्न युगपत समीकरणों (उदाहरण के लिए समाश्रयण विश्लेषण में समीकरण) का एक समुच्चय होता है। ये प्राचलों के अनुमान के लिए प्रयोग किए जाते हैं।

समाश्रयण (Regression) : यह दो या अधिक चरों के बीच औसत संबंध का सांख्यिकीय परिमाप है, जो आंकड़ों की मूल इकाइयों में व्यक्त किया जाता है।

9.12 कुछ उपयोगी पुस्तकें

Nagar, A.L. and R.K. Das, 1989 : *Basic Statistics,* Oxford University Press, Delhi.

Goon, A.M., M.K. Gupta and B. Dasguprta, 1987 : *Basic Statistics,* The World Press Pvt. Ltd., Calcutta.

9.13 बोध प्रश्नों के उत्तर या संकेत

बोध प्रश्न 1

1) +0.98; $Y = 0.64X + 0.54$; 8.2

2) $X = 0.95Y - 6.4; Y = 0.95X + 7.25$

3) $X = 2.23Y - 12.70$; $Y = 0.39X + 7.33$

   (i) 29.6 (ii) 18.9

4) $Y = 0.613X + 14.81; X = 1.360Y - 5.2$

5) $Y = 3.99 + 0.103X$; 8.11 tons

बोध प्रश्न 2

1) i) $Y = a + bX = 5.856 + 0.676X$

   ii) $X = \alpha + \beta Y = 29.848 + 0.799Y$

   iii) $r = 0.73$

   iv) $0.676 \times 0.799 = 0.54$

2) जब पारिवारिक आय 10,000 रुपये है तो वस्त्रों पर व्यय =1000 रुपये है। जब आय 1000 रुपये से कम या 1,00,000 रुपये से अधिक हो तो समाश्रयण रेखा का लागू होना आवश्यक नहीं है। इन दोनों संख्याओं के बीच एक रुपए के आधार पर आय में वृद्धि होने पर वस्त्र पर व्यय 9 पैसे बढ़ जाता है।

इकाई 10 सूचकांक

इकाई की रूपरेखा

10.0 उद्देश्य

इस इकाई के अध्ययन के बाद आप :

- सूचकांक को परिभाषित करना सीख सकेंगे; और
- उनकी रचना तथा परिकलन कर सकेंगे।

10.1 प्रस्तावना

सामान्य बोध के अनुसार 'सूचक' शब्द का अर्थ संकेतक के अतिरिक्त कुछ और नहीं होता। 'सूचकांक' शब्द इसका बहुवचन रूप होता है, लेकिन इन सभी का एक ही अर्थ होता है।

एक सूचकांक द्वारा, दो या दो से अधिक अवधियों या स्थानों के अंतर्गत बहुत से चरों को एक साथ

लेकर, परिवर्तन के आकार के सामान्य स्तर को व्यक्त किया जाता है। इस परिभाषा में 'चर' शब्द का अर्थ एक संख्यात्मक चर है जोकि मात्रा या राशि में मापा जा सकता है। जैसे वस्तुओं की कीमतें, मात्राएँ आदि। उदाहरण के लिए, हमारी इच्छा 1980 तथा 1990 के बीच या बम्बई तथा कलकत्ता के बीच किसी वस्तु के कीमत स्तर की तुलना करना हो सकती है। मान लीजिए 1985 तथा 1990 में चावल का उत्पादन क्रमश: 50,000 तथा 60,000 टन है। उत्पादन की तुलना के लिए 1985 को आधार वर्ष लिया जाता है, अर्थात् 1985 = 100 लिया जाता है। 1990 के लिए संगत संख्या $\frac{60{,}000}{50{,}000} \times 100 = 120$ होगी। यह सरलतम रूप में एक वस्तु का सूचकांक है, जोकि तुलनात्मक संख्या है। लेकिन व्यवहार में, सूचकांक रचना में, प्राय: बहुत सी वस्तुएँ सम्मिलित होती हैं।

कष्टदायक दशमलवों से बचने के लिए सूचकांक को, जोकि संख्याओं का अनुपात होता है, प्रतिशत के रूप में व्यक्त किया जाता है। अत: यदि किसी वस्तु की 1970 में लागत 45 पैसे हैं तथा 1974 में लागत 1.50 रुपये है तो इसका अनुपात $\frac{150}{45} = 3.33$ होगा। यदि इसके स्थान पर हम इस अनुपात को प्रतिशत में व्यक्त करें तो यह $\frac{150}{45} \times 100 = 333$ होगा। 333 को वर्ष 1970 पर आधारित, जोकि 100 है, 1974 का सूचकांक कहते हैं।

10.2 सूचकांक रचना में चरण

व्यापारिक तथा आर्थिक परिस्थितियों के पूर्वानुमान, सामान्य सूचना प्रदान करना आदि के लिए बहुत सी सरकारी तथा निजी संस्थाएँ सूचकांक के परिकलन में कार्यरत हैं।

बहुत से सामान्य प्रकार के सूचकांकों द्वारा समयावधि में चर के परिवर्तन का माप किया जाता है, लेकिन इसका अर्थ यह नहीं है कि तुलना हमेशा समयावधि में ही की जाती है। इसी प्रकार, सूचकांक की रचना किसी भी चर जैसे बौद्धिक स्तर, अभिरुचि, कार्यकुशलता, उत्पादन आदि में परिवर्तन का अध्ययन करने के लिए भी की जा सकती है, लेकिन, शायद कीमतों की कालश्रेणी के अध्ययन में इसका अधिकतम उपयोग होता है। इसलिए परवर्ती (subsequent) सूचकांक विवेचन में वस्तुओं की कीमतों का विशेष उल्लेख रहेगा। इनकी रचना के सिद्धांतों की सामान्य प्रकृति होने के कारण इनकों अन्य रुचि क्षेत्रों में भी प्रयुक्त किया जा सकता है।

कीमत सूचकांको के बहुत से उपयोग होते हैं। *थोक मूल्य सूचकांक* मुद्रा के मूल्य में हो रही कमी की सूचना देता है। इसके अतिरिक्त, *उपभोक्ता कीमत सूचकांक* या *जीवन निर्वाह व्यय सूचकांक* हमें वास्तविक आय में परिवर्तन के बारे में जानकारी देता है। इसका मुख्य अनुप्रयोग महँगाई भत्ते के परिकलन, जिससे वास्तविक मज़दूरी को कम होने से रोका जा सके या दो भिन्न क्षेत्रों के बीच जीवन स्तर की तुलना, में होता है। इसके द्वारा मुद्रा की क्रय शक्ति में परिवर्तनों का माप भी किया जा सकता है। सामान्य कीमत सूचकांक का व्युत्क्रम, आधारकाल के संदर्भ में, मुद्रा की क्रय शक्ति का सूचक है। उदाहरण के लिए, यदि मूल्य सूचकांक 150 हो गया है तो इसका अर्थ यह है कि मुद्रा की उतनी ही मात्रा अब आधारकाल में खरीदी जाने वाली वस्तुओं की मात्राओं का 100/150 = 0.67 या 67 प्रतिशत ही खरीदा जा सकेगा।

10.2.1 आधार काल का चयन

चूँकि सूचकांक द्वारा तुलनात्मक परिवर्तनों को मापा जाता है, इसलिए इनको एक चयन की गई परिस्थिति (उदाहरण के लिए समय, स्थान आदि) के साथ व्यक्त किया जाता है। इस परिस्थिति

का मान 100 लिया जाता है तथा इसको आधार या सूचकांक श्रेणी का प्रारंभ बिन्दु कहा जाता है। उदाहरण के लिए, हम एक दिनांक को निश्चित करके सभी परिवर्तन उसके आधार पर मापते हैं। आधार कोई भी एक दिन हो सकता है, जैसा कि फुटकर कीमत सूचकांक में होता है, या फिर एक वर्ष का औसत या किसी समयावधि का औसत हो सकता है।

आधारकाल का चयन करते समय निम्नलिखित बातों का ध्यान रखना आवश्यक है :

1) आधार दिनांक सामान्य होना चाहिए ताकि चयन किए गए समंक किसी अनिश्चित या असामान्य परिस्थिति जैसे प्राकृतिक संकट अथवा युद्ध आदि से प्रभावित न हों। अधिक परिशुद्धता के लिए यह आवश्यक है कि विभिन्न तुलनाएँ किसी स्थिर काल से ही की जानी चाहिए।

2) आधारकाल बहुत पहले का नहीं होना चाहिए क्योंकि यदि समयावधि अधिक है तो व्यापार, आयात, उपभोगता अधिमान आदि के प्रतिरूप में अत्याधिक परिवर्तन की संभावना होती है, जिसके कारण की जाने वाली तुलना निरर्थक हो सकती है। दस वर्ष से बीस वर्ष तक का अंतराल एक आधार दिनांक के लिए उपयुक्त हो सकता है जिससे अधिक अंतराल का सूचकांक उत्तरोतर पुराना होता जाता है। ऐसे सूचकांक की तुलना में, जिनमें समयावधि अधिक होती है, अल्पकालीन सूचकांक द्वारा अधिक परिशुद्धता प्राप्त होती है।

3) आर्थिक समंकों से संबंधित सूचकांक में आधार काल का कोई आर्थिक अभिप्राय होना चाहिए।

10.2.2 उपयुक्त माध्य का चयन

एक सूचकांक मूलत: समंकों की श्रेणी का औसत (माध्य) ज्ञात करने का परिणाम होता है (जैसे विभिन्न वस्तुओं के मूल्यानुपातों का माध्य)। लेकिन श्रेणी का माध्य प्राप्त करने की बहुत सी विधियाँ है : माध्य (समांतर माध्य), बहुलक, माध्यिका, गुणोत्तर माध्य, हरात्मक माध्य आदि। यहाँ प्रश्न यह है कि कौन से माध्य का सूचकांक रचना में उपयोग किया जाए।

बहुलक में सरलता का गुण है लेकिन यह अनिश्चित हो सकता है। माध्यिका की भी यही सीमाएँ हैं। इसके अतिरिक्त, इनमें से कोई भी, बंटन के प्रत्येक छोर पर विद्यमान मद के आकार से प्रभावित नहीं होता। हरात्मक माध्य का सूचकांक में व्यवहारिक अनुप्रयोग बहुत ही कम है। परिणामत: बहुलक, माध्यिका तथा हरात्मक माध्य प्राय: सूचकांक के परिकलन में उपयोग नहीं किए जाते। अत: सूचकांक परिकलन में प्राय: समांतर माध्य का उपयोग किया जाता है। परिकलन में थोड़ी कठिनाई के बावजूद कभी-कभी गुणोत्तर माध्य का उपयोग भी किया जाता है।

10.2.3 मदों तथा उनकी संख्याओं का चयन

सूचकांक रचना में सम्मिलित की जाने वाली वस्तुओं की किस्म तथा उनकी संख्या, विचाराधीन प्रश्न-विशेष के लिए, परिकलन की मितव्ययता तथा सुविधा आदि पर निर्भर होती है। सम्मिलित की जाने वाली व्यष्टियों की संख्या तथा किस्म विभिन्न प्रकार के व्यवहारिक विचारों पर आधारित होती है। थोक मूल्य सूचकांक के लिए वस्तुओं की संख्या, जितना संभव हो, अधिक होनी चाहिए। इसके विपरीत, यदि सूचकांक का उद्देश्य समयावधि में हुए परिवर्तनों का सूचक न होकर भविष्य के लिए कीमत परिवर्तन का पूर्वानुमान लगाना है, तो मदों की थोड़ी संख्या भी पर्याप्त हो सकती है। फिर भी, यह सुनिश्चित कर लेना चाहिए कि मदों की

कम संख्या के कारण सूचकांक, सामान्य स्तर का, अप्रतिनिधिक न हो जाए। अत्याधिक दीर्घकाल में वस्तुओं के निश्चित समुच्चय का उपयोग नहीं किया जाना चाहिए, क्योंकि समय के साथ कुछ वस्तुओं का महत्त्व कम हो सकता है तथा कुछ नई वस्तुओं के महत्त्व में वृद्धि हो सकती है। सामान्येत: वस्तुएँ कीमत प्रणाली के विभिन्न तत्त्वों के प्रति संवेदी (sensitive) एवं उनकी प्रतिनिधिक होनी चाहिए।

10.2.4 समंकों का संकलन

क्योंकि विभिन्न बाज़ारों में कीमतें प्राय: भिन्न होती हैं, इसीलिए उनका संकलन प्रतिनिधिक बाज़ार से नियमित अंतराल बाद करते रहना चाहिए। उन दुकानों का, जिनसे उपभोगता प्राय: वस्तुएँ खरीदते हैं, चयन करना वांछनीय होता है। प्रत्येक संघटक व्यष्टि के निवेदित भाव की परिशुद्धता से, सूचकांक की विश्वसनीयता अत्यधिक प्रभावित होती है।

10.3 सूचकांक रचना की विधि

सूचकांक रचना की विभिन्न विधियाँ निम्नलिखित है :

1) सापेक्ष विधियाँ
   a) सापेक्षों का सरल माध्य
   b) सापेक्षों का भारित माध्य
2) समूही विधियाँ
   a) सरल समूही सूत्र
   b) भारित समूही सूत्र
      i) लास्पियर का सूचकांक
      ii) पाशे का सूचकांक
      iii) एजवर्थ-मार्शल का सूचकांक
      iv) फिशर का आदर्श सूचकांक

10.3.1 सापेक्ष विधियाँ

यदि हम बहुत सी वस्तुओं की एक दिए हुए दिनांक पर तथा समरूप वस्तुओं की बाद के एक दिनांक पर कीमतें लिखित करें, तो प्रत्येक वस्तु की कीमत में परिवर्तन, अर्थात् नई कीमत की पुरानी कीमत से तुलना, को प्रतिशत रूप में व्यक्त किया जा सकता है। इसके द्वारा हमें मूल्यानुपात प्राप्त होता है और यदि हमें विभिन्न वस्तुओं के भार ज्ञात हैं तो अगले चरण में इन मूल्यानुपातों को उनके भारों से गुणा कर दिया जाता है। अंतत: इन भारित मूल्यानुपातों को जोड़कर माध्य का परिकलन किया जाय तो हमें सूचकांक प्राप्त होता है।

यह मानना अवास्तविक है कि प्रत्येक वस्तु का उपभोग समान रहा है। इसलिए अधिकतर सूचकांकों में प्रत्येक वस्तु के वास्तविक उपभोग के अनुपात का ध्यान रखा जाता है। इस प्रकार भारित करने की विधि, श्रेणी में प्रत्येक मद के तुलनात्मक महत्त्व को दर्शाती है।

मान लीजिए, k वस्तुओं की आधारवर्ष में कीमतें

$$p_{01}, p_{02}, \cdots p_{0k} \text{ हैं}$$

तथा वर्तमान वर्ष में कीमतें

$p_{n1}, p_{n2}, \dots p_{nk}$ हैं

तब iवीं वस्तु का मूल्यानुपात $\dfrac{p_{ni}}{p_{Oi}}$ होगा, जहाँ पर $i = 1,2,\dots,k$ है तथा अनुलग्न (subscript) 0 आधारवर्ष तथा n वर्तमान वर्ष के संकेतक हैं।

a) **सापेक्षों का सरल माध्य**

मूल्यानुपातों का समांतर माध्य लेने पर हमें निम्नलिखित सूचकांक प्राप्त होता है :

$$\text{सूचकांक} \quad = 100 \times \sum_{i=1}^{k} \frac{\left(p_{ni} / p_{Oi}\right)}{k} \qquad \dots(10.1)$$

सरलता के लिए हम अनुलग्न i को न लिखकर

$$\text{सूचकांक} \quad = 100 \times \sum \left(\frac{p_n / p_O}{k} \right)$$

भी लिख सकते हैं।

b) **सापेक्षों का भारित माध्य**

उपयोग किए जाने वाले भारों में सबसे उपयुक्त भार प्रत्येक वस्तु का मूल्य होती हैं जिनका संकेतन, iवीं वस्तु के लिए, w_i से किया जाता है। हम आधारवर्ष की मात्राओं का, आधारवर्ष की कीमतों पर, मूल्य $(w_{0i} = p_{0i}q_{0i})$ या वर्तमान वर्ष की मात्राओं का, वर्तमान वर्ष की कीमतों पर, मूल्य $(w_{1i} = p_{1i}q_{1i})$ या कोई अन्य मूल्य, भार के रूप में उपयोग कर सकते है। ये भार, किसी विवेकपूर्ण विधि द्वारा प्राप्त स्थिर घटकों का समुच्चय भी हो सकते हैं।

आधारवर्ष के मूल्यों को भार लेने पर मूल्यानुपातों के समांतर माध्य द्वारा निम्नलिखित सूचकांक प्राप्त होता है :

$$\text{सूचकांक} \quad = \frac{\sum \frac{p_n}{p_0} \times w_0}{\sum w_0} \times 100 \qquad \dots(10.2)$$

यहाँ पर सरलता के लिए अनुलग्न i को नहीं लिखा गया है। यहाँ पर यह ध्यान दें कि आधारवर्ष के भारों के उपयोग से निरंतरता तो बनी रहती है लेकिन समय परिवर्तन के साथ आधुनिकता का ह्रास होता है।

उदाहरण 10.1 : निम्नलिखित सारणी में प्रति रेल-यात्रा का औसत भाड़ा प्रस्तुत है। 1948 के औसत को 100 मानकर तथा आधारवर्ष के भारों के उपयोग द्वारा परिकलन किए गए हैं।

टिकट की श्रेणी	1948 में यात्राओं की संख्या (दस लाखों में) (q_0)	भाड़ा (रुपये) 1948 (p_0)	1969 (p_n)	भार $w_0 = p_0 q_0$	मूल्यानुपात $P = (p_n / p_0) \times 100$	$P.w_0$
पूर्ण भाड़ा	23	12	60	276	500	138000
भ्रमण	25	6	30	150	500	75000
उत्सव	20	4	15	80	375	30000
अवधि टिकट	32	5	14	160	280	44800
योग				666		287800

सूत्र (10.2) के उपयोग द्वारा

$$1969 \text{ का सूचकांक } = \frac{287800}{666} = 432.13$$

वर्तमान मानों $(w_n = p_n q_n)$ को भार लेने पर

$$\text{सूचकांक } = \frac{\sum \frac{p_n}{p_0} \times w_n}{\sum w_n} \times 100 \qquad \text{...(10.3)}$$

उदाहरण 10.2 : निम्नलिखित सारणी में प्रति रेलयात्रा का औसत भाड़ा प्रस्तुत हैं। 1948 की औसत को 100 मानकर तथा वर्तमान वर्ष के भारों के उपयोग द्वारा विभिन्न परिकलन किए गए हैं :

टिकट की श्रेणी	1969 में यात्राओं की संख्या (दस लाखों में) (q_n)	भाड़ा (रुपये) 1948 (p_0)	1969 (p_n)	भार $w_n = p_n q_n$	मूल्यानुपात $P = (p_n / p_0) \times 100$	$P.w_n$
पूर्ण भाड़ा	25	12	60	1500	500	750000
भ्रमण	26	6	30	780	500	390000
उत्सव	9	4	15	135	375	50630
अवधि टिकट	27	5	14	378	280	105800
योग				2793		1296430

सूत्र (10.3) के उपयोग द्वारा

$$1969 \text{ का सूचकांक } = \frac{1296430}{2793} = 464.17$$

10.3.2 समूही विधियाँ

इस विधि में, वर्तमान या दिए हुए वर्ष में सभी वस्तुओं के समूह (कुल योग) को आधार वर्ष के इसी प्रकार के समूह के प्रतिशत के रूप में व्यक्त किया जाता है। अत:

सरल समूही सूचकांक

$$\text{सूचकांक} = \frac{\text{वर्तमान वर्ष में कीमतों का योग}}{\text{आधार वर्ष में कीमतों का योग}} \times 100$$

$$= \frac{p_{n1} + p_{n2} + \ldots\ldots + p_{nk}}{p_{01} + p_{02} + \ldots\ldots + p_{0k}} \times 100$$

$$= \frac{\sum p_{ni}}{\sum p_{0i}} \times 100 = \frac{\sum p_n}{\sum p_0} \times 100 \qquad \ldots (10.4)$$

जहाँ पर योग $\left(\sum_{i=1}^{k}\right)$ चुनी हुई k कीमतों पर है।

भारित समूही सूचकांक

$$\text{सामान्य सूचकांक} = \frac{p_{n1}q_1 + p_{n2}q_2 + \ldots\ldots + p_{nk}q_k}{p_{01}q_1 + p_{02}q_2 + \ldots\ldots + p_{0k}q_k} \times 100$$

$$= \frac{\sum p_{ni}q_i}{\sum p_{0i}q_i} \times 100$$

$$\text{या केवल} \quad = \frac{\sum p_n q}{\sum p_0 q} \times 100 \qquad \ldots (10.5)$$

यहाँ पर उपयोग किए गए भार खरीदी या बेची गई वास्तविक मात्राएँ होनी चाहिए और इनमें तब तक परिवर्तन नहीं किया जाना चाहिए जब तक सूचकांक में संशोधन की आवश्यकता न हो।

भारित सामूहिक सूचकांक के बहुत से सूत्र हैं, लेकिन भारों पर आधारित, केवल प्राय: उपयोग किए जाने वाले सूचकांकों का ही विवेचन किया जाएगा।

i) **लास्पियर का सूचकांक**

यदि हम सामान्य भारित समूही सूत्र (10.5) में आधार वर्ष की मात्राओं (q_o) को भार के रूप में उपयोग करें तो हमें लास्पियर का सूत्र (L) प्राप्त होगा।

$$L = \frac{\sum p_n q_0}{\sum p_0 q_0} \times 100 \qquad \ldots (10.6)$$

यहाँ पर ध्यान दिया जाना चाहिए कि इस सूचकांक में स्थिर आधार भारों का उपयोग किया जाता है तथा यह सूत्र (10.2) में दिये मूल्यानुपातों के भारित समांतर माध्य के तुल्य (equivalent) है। अत: हम (10.6) को इस प्रकार भी लिख सकते है :

$$L = \frac{\sum \frac{p_n}{p_0} \times p_0 q_0}{\sum p_0 q_0} \times 100$$

ii) पाशे का सूचकांक

यदि हम सामान्य भारित समूही सूत्र (10.5) में वर्तमान वर्ष की मात्राओं (q_n) को भार के रूप में उपयोग करें, तो हमें पाशे का सूत्र (P) प्राप्त होगा :

$$P = \frac{\sum p_n q_n}{\sum p_0 q_n} \times 100 \qquad \text{... (10.7)}$$

यहाँ पर q_n (जोकि वास्तव में $q_{n1}, q_{n2} ..... q_{nk}$ हैं), वर्तमान वर्ष में खरीदी या बेची गई मात्राएँ हैं।

iii) फिशर का आदर्श सूचकांक

यह सूचकांक लास्पियर तथा पाशे के सूत्रों द्वारा प्राप्त सूचकांकों का गुणोत्तर माध्य (अर्थात् गुणनफल का वर्गमूल) होता है। इस सूचकांक की कुछ विशेषताएँ हैं (जिनका विवेचन बाद में किया जायेगा), तथा इसको फिशर का आदर्श सूचकांक कहते हैं।

$$F = \sqrt{L \times P} = \sqrt{\frac{\sum p_n q_0}{\sum p_0 q_0} \times \frac{\sum p_n q_n}{\sum p_0 q_n}} \times 100 \qquad \text{.. (10.8)}$$

iv) एजवर्थ मार्शल का सूचकांक

इस सूत्र में आधार तथा दिए हुए वर्ष की मात्राओं के माध्य को भार लिया जाता है, अर्थात् $w = \frac{1}{2}(q_0 + q_n)$ है। एजवर्थ-मार्शल का सूत्र इस प्रकार है :

$$I = \frac{\sum p_n (q_0 + q_n)/2}{\sum p_0 (q_0 + q_n)/2} \times 100$$

$$= \frac{\sum p_n (q_0 + q_n)}{\sum p_0 (q_0 + q_n)} \times 100 \qquad \text{... (10.9)}$$

सारणी : 10.1: लास्पियर, पाशे, एजवर्थ-मार्शल तथा फिशर सूचकांको के परिकलन की व्याख्या

मद	आधार वर्ष (1970)		वर्तमान वर्ष (1980)					
	कीमत	मात्रा	कीमत	मात्रा				
	(p_0)	(q_0)	(p_n)	(q_n)	$p_0 q_0$	$p_n q_0$	$p_0 q_n$	$p_n q_n$
A	20	7	25	9	140	175	180	225
B	42	6	40	8	252	240	336	320
C	30	17	25	4	510	425	120	100
D	8	15	14	10	120	210	80	140
E	10	8	13	5	80	104	50	65
योग					1102	1154	766	850

1) लास्पियर का कीमत सूचकांक $= \frac{\sum p_n q_0}{\sum p_0 q_0} \times 100 = \frac{1154}{1102} \times 100 = 104.72 = 105$

2) पाशे का कीमत सूचकांक $= \frac{\sum p_n q_n}{\sum p_0 q_n} \times 100 = \frac{850}{766} \times 100 = 110.97 = 111$

3) एजवर्थ-मार्शल का सूचकांक $= \frac{\sum p_n q_0 + \sum p_n q_n}{\sum p_0 q_0 + \sum p_0 q_n} \times 100$

$$= \frac{1154 + 850}{1102 + 766} \times 100$$

$$= \frac{2004}{1868} \times 100 = 107.28 = 107$$

4) फिशर का आदर्श सूचकांक $= \sqrt{\frac{\sum p_n q_0 \sum p_n q_n}{\sum p_0 q_0 \sum p_0 q_n}} \times 100$

$$= \sqrt{[(L) \times (P)]} = \sqrt{(104.72 \times 110.97)}$$

$$= 107.8 = 108$$

ध्यान दें कि कीमतों में समान परिवर्तन होने पर भी भिन्न सूत्र भिन्न मूल्य प्रदान करते हैं। इसके अलावा लास्पियर कीमत सूचकांक न्यूनतम मूल्य प्रदान करता है जबकि पाशे सूचकांक अधिकतम मूल्य प्रदान करता है। इसलिये यह प्राय: कहा जाता है कि लास्पियर सूचकांक सही कीमत परिवर्तन का एक निम्नानुमान है जबकि पाशे सूचकांक एक ऊर्ध्वानुमान है।

10.3.3 मात्रा या आकार सूचकांक

यदि हम कीमत सूचकांक में p के स्थान पर q तथा q के स्थान पर p का उपयोग करें तो हमें मात्रा या आकार सूचकांक प्राप्त होता है, जोकि वस्तुओं की मात्राओं की तुलना के मापन को व्यक्त करता है।

1) मात्रानुपात (quantity relative) $= \frac{q_n}{q_0} \times 100$

2) मात्रानुपातों का समांतर माध्य $= 100 \sum \left(\frac{q_n}{q_0} \right) / k$

3) मात्रानुपातों के भारित माध्य सूचकांक :

क) आधारवर्ष के भार : $\frac{\sum (q_n / q_0) \times w_0}{\sum w_0} \times 100$ जहाँ पर $w_0 = p_0 q_0$

ख) वर्तमान वर्ष के भार : $\frac{\sum(q_n / q_0) \times w_n}{\sum w_n} \times 100$ जहाँ पर $w_n = p_n q_n$

4) सरल समूही मात्रा-सूचकांक $= \frac{\sum q_n}{\sum q_0} \times 100$

5) लास्पियर का मात्रा सूचकांक $= \frac{\sum q_n p_0}{\sum q_0 p_0} \times 100$

6) पाशे का मात्रा सूचकांक $= \frac{\sum q_n p_n}{\sum q_0 p_n} \times 100$

7) फिशर का आदर्श सूचकांक $= \sqrt{\frac{\sum q_n p_0}{\sum q_0 p_0} \frac{\sum q_n p_n}{\sum q_0 p_n}} \times 100$

8) एजवर्थ-मार्शल का सूचकांक $= \frac{\sum q_n (p_0 + p_n)}{\sum q_0 (p_0 + p_n)} \times 100$

बोध प्रश्न 1

1) सूचकांकों द्वारा क्या मापा जाता है?

....

....

....

....

....

....

....

....

....

....

2) सूचकांक रचना में, विशेष रूप से कीमत सूचकांक के संदर्भ में, आने वाली विभिन्न कठिनाइयों का विवेचन कीजिए।

....

....

....

....

....

....

....

....

....

....

3) 1983 तथा 1984 के लिए 6 भिन्न वस्तुओं की कीमतें निम्नलिखित हैं।

(क) समूही विधि, एवं (ख) मूल्यानुपातों का माध्य विधि; में समांतर माध्य उपयोग करके सूचकांक परिकलित कीजिए।

वस्तुएँ	1983 में कीमत (रुपये)	1984 में कीमत (रुपये)
A	40	50
B	50	60
C	20	30
D	50	70
E	80	80
F	100	110

....

....

....

....

....

....

....

....

....

....

4) निम्नलिखित मदों के समूह द्वारा फिशर का आदर्श सूचकांक परिकलित कीजिए।

मद संख्या	आधार वर्ष		वर्तमान वर्ष	
	कीमत (रुपयों में)	मात्रा (किलो में)	कीमत (रुपयों में)	मात्रा (किलो में)
1	4	1.0	3	4
2	8	1.5	7	5

....

....

....

....

....

....

....

....

....

....

5) निम्नलिखित समंको से लास्पियर तथा पाशे के सूचकांक परिकलित कीजिए :

मद	आधार वर्ष		वर्तमान वर्ष	
	मात्रा	प्रति पाउण्ड कीमत	मात्रा	प्रति पाउण्ड कीमत
रोटी	6.0	40 पैसे	7.0	30 पैसे
मांस	4.0	45 पैसे	5.0	50 पैसे
चाय	0.5	90 पैसे	1.5	40 पैसे

....

....

....

....

....

....

....

....

....

....

10.4 विभिन्न समूही परिमापों के गुण

विभिन्न उद्देश्यों के लिए भिन्न सूचकांकों की रचना की जाती है इसीलिए किसी सूचकांक की उपयुक्तता, उद्देश्य की प्रकृति पर निर्भर होती है।

लास्पियर सूचकांक का परिकलन सरल होता है, क्योंकि इसमें आधारकाल की मात्राओं का भार के रूप में उपयोग किया जाता है जिनको प्राप्त करना कठिन नहीं होता, तथा हर (denominator) को एक बार ही परिकलित करने की आवश्यकता होती है। लेकिन इस सूचकांक में, कीमत परिवर्तन के अनुरूप मात्रा या उत्पादन में परिवर्तन का उपयोग नहीं होता, इसलिए इसमें कीमत वृद्धि को वास्तविकता से अधिक व्यक्त करने की प्रवृत्ति विद्यमान होती है। इसके विपरीत, पाशे के सूचकांक में वर्तमान वर्ष की मात्राएँ भार के रूप में उपयोग होती हैं, जिनमें प्रत्येक वर्ष में परिवर्तन होते रहते हैं। इसके अतिरिक्त, चूँकि वर्तमान वर्ष के भार उपयोग किए जाते हैं, इसमें कीमत वृद्धि को वास्तविक से कम व्यक्त करने की प्रवृत्ति विद्यमान होती है।

चूँकि स्थिर भारों का उपयोग सुविधाजनक होता है, इसलिए संभवत: लास्पियर सूचकांक का उपयोग अधिक होता है। लेकिन समय गुजरने के साथ, भार पुराने हो जाते हैं। उदाहरण के लिए, 1970

में कलकता में टेलिविज़नों की संख्या शून्य थी। 1990 में, टेलिविजनों की संख्या रेफ्रीजरेटरों की संख्या से अधिक हो गई। पाशे के सूचकांक में वर्तमान भारों का उपयोग किया जाता है, जोकि बेहतर होते हैं। लेकिन, चूँकि उत्पादित या उपभोग की गई वस्तुओं के वर्तमान वर्ष के समंक सुविधापूर्वक उपलब्ध नहीं होते, इसलिए लास्पियर सूचकांक अधिक उपयोगी होता हैं।

10.5 सूचकांकों के परीक्षण

एक श्रेष्ठ सूचकांक को, जिसके द्वारा एक अवधि से दूसरी अवधि में किसी तथ्य के बारे में परिवर्तन को मापा जाता है, कुछ परीक्षणों के आधार पर यथेष्ट होना चाहिए। सूचकांक के तीन मुख्य परीक्षण होते है :

i) कालोत्क्रमण परीक्षण,

ii) उपादानोत्क्रमण परीक्षण,

iii) श्रृंखलिक परीक्षण।

10.5.1 कालोत्क्रमण परीक्षण

इस परीक्षण के अनुसार यदि सूचकांक में कीमत (या मात्रा) के समय-अनुलग्नों (जैसे 0 तथा n) को विपरीत कर दिया जाए तो प्राप्त परिणाम, सूचकांक का व्युत्क्रम होना चाहिए।

सांकेतिक रूप में

$$I_{0n} \times I_{n0} = 1$$

जहाँ पर I_{0n} = काल 'n' का सूचकांक जिसका आधार काल '0' है।

I_{n0} = काल '0' का सूचकांक जिसका आधार काल 'n' है।

यदि 1975 से 1982 की अवधि में कीमत परिवर्तन 4 रुपये से 16 रुपये होता है तो 1982 की कीमत, 1975 की कीमत का 400 प्रतिशत है, तथा 1975 की कीमत, 1982 की कीमत का 25 प्रतिशत है। इन दोनों मूल्यानुपातों का गुणनफल 4× 0./25 = 1 है। कालोत्क्रमण परीक्षण इस अनुरूपता पर आधारित है कि जो सिद्धांत एक वस्तु के लिए सत्य है, वही समूह के सूचकांक के लिए भी सत्य होना चाहिए।

सूचकांक रचना की पाँच विधियाँ कालोत्क्रमण परीक्षण को संतुष्ट करती हैं। ये इस प्रकार हैं :

1) मूल्यानुपातों का सरल गुणोत्तर माध्य

2) स्थिर भारों सहित समूही सूचकांक

3) एजवर्थ-मार्शलं सूचकांक

4) मूल्यानुपातों का भारित गुणोत्तर माध्य, जिसमें स्थिर भार उपयोग किए गए हों

5) फिशर का आदर्श सूचकांक

निम्नलिखित में हम यह दर्शाएँगे कि फिशर का आदर्श सूचकांक कालोत्क्रमण परीक्षण को संतुष्ट करता है।

फिशर का आदर्श सूचकांक $F = \sqrt{\frac{\sum p_n q_0}{\sum p_0 q_0} \times \frac{\sum p_n q_n}{\sum p_0 q_n}}$

समय अनुलग्नों को विपरीत करने पर $F' = \sqrt{\frac{\sum p_0 q_n}{\sum p_n q_n} \frac{\sum p_0 q_0}{\sum p_n q_0}}$

क्योंकि $F \times F' = 1$, इसलिए यह परीक्षण संतुष्ट हो जाता है।

10.5.2 उपादानोत्क्रमण परीक्षण

प्राय: उपयोग किए जाने वाले संकेतों की सहायता से "मूल्य सूचकांक" का सूत्र इस प्रकार लिखा जाता है :

$$I_v = \frac{\sum p_n q_n}{\sum p_0 q_0}$$

अब उदाहरण के लिए, लास्पियर के कीमत तथा मात्रा सूचकांक क्रमश:

$$I_p = \frac{\sum p_n q_0}{\sum p_0 q_0}$$

तथा $I_q = \frac{\sum q_n p_0}{\sum q_0 p_0}$ हैं।

उपादानोत्क्रमण परीक्षण के अनुसार $I_p \cdot I_q = I_v$ होना चाहिए।

लेकिन लास्पियर सूचकांक के लिए

$$I_p \cdot I_q = \frac{\sum (p_n q_0)(\sum q_n p_0)}{(\sum p_0 q_0)^2} \neq I_v$$

इसके विपरीत, फिशर का आदर्श सूचकांक इस परीक्षण को संतुष्ट करता है, यह निम्नलिखित में प्रदर्शित किया गया है :

$$I_p = \sqrt{\frac{\sum p_n q_0}{\sum p_0 q_0} \times \frac{\sum p_n q_n}{\sum p_0 q_n}}$$

$$I_q = \sqrt{\frac{\sum q_n p_0}{\sum q_0 p_0} \times \frac{\sum q_n p_n}{\sum q_0 p_n}}$$

$$I_p \cdot I_q = \sqrt{\frac{\sum p_n q_0}{\sum p_0 q_0} \times \frac{\sum p_n q_n}{\sum p_0 q_n} \times \frac{\sum q_n p_0}{\sum q_0 p_0} \times \frac{\sum q_n p_n}{\sum q_0 p_n}}$$

$$= \sqrt{\frac{\sum p_n q_n}{\sum p_0 q_0} \times \frac{\sum p_n q_n}{\sum p_0 q_0}} = \frac{\sum p_n q_n}{\sum p_0 q_0} = I_v$$

इस सिद्धांत को और अच्छी तरह समझने के लिए हम निम्नलिखित उदाहरण पर ध्यान देते हैं।

यदि किसी वस्तु की प्रति इकाई कीमत तथा मात्रा में, 1970 से 1990 की अवधि में, क्रमश: 16 रुपये से 32 रुपये तथा 100 इकाई से 200 इकाई परिवर्तन हुआ हो तब 1990 में कीमत तथा मात्रा, दोनों, 200 प्रतिशत होगी अर्थात् 1970 की कीमत तथा मात्रा का दुगुना होंगी। 1970 में कुल मूल्य (कीमत × मात्रा) 1600 रुपये तथा 1990 में कुल मूल्य 6400 रुपये होगा। इस प्रकार इनका मूल्यानुपात 6400/1600 = 4.00 है। अत: यह जाँच की जा सकती है कि 2.00 × 2.00 = 4.00 है। अत: कीमत अनुपात तथा मात्रा अनुपात का गुणनफल कुल मूल्य अनुपात के बराबर है।

इस परीक्षण को केवल फिशर का आदर्श सूचकांक ही संतुष्ट करता है।

उदाहरण 10.3 : निम्नलिखित समंकों द्वारा यह प्रदर्शित किया गया है कि फिशर का आदर्श सूचकांक उपादानोत्क्रमण परीक्षण को संतुष्ट करता है :

मद	कीमत (रुपयों में)		इकाइयों की संख्या					
	1983 (p_0)	1989 (p_n)	1983 (q_0)	1989 (q_n)	p_0q_0	p_nq_0	p_0q_n	p_nq_n
I	6	10	50	56	300	500	336	560
II	2	2	100	120	200	200	240	240
III	4	6	60	60	240	360	240	360
IV	10	12	30	24	300	360	240	288
V	8	12	40	36	320	480	288	432
योग					1360	1900	1344	1880

कीमत अनुपात : $I_p = \sqrt{\frac{\sum p_nq_0}{\sum p_0q_0} \times \frac{\sum p_nq_n}{\sum p_0q_n}} = \sqrt{\frac{1900}{1360} \times \frac{1880}{1344}}$

मात्रा अनुपात : $I_q = \sqrt{\frac{\sum q_np_0}{\sum q_0p_0} \times \frac{\sum q_np_n}{\sum q_0p_n}} = \sqrt{\frac{1344}{1360} \times \frac{1880}{1900}}$

मूल्य अनुपात : $I_v = \frac{\sum p_nq_n}{\sum p_0q_0} = \frac{1880}{1900}$

$$I_pI_q = \sqrt{\frac{1900}{1360} \times \frac{1880}{1344} \times \frac{1344}{1360} \times \frac{1880}{1900}} = \sqrt{\frac{1880}{1360} \times \frac{1880}{1360}}$$

$$= \frac{1880}{1360}$$

$= I_v$, जोकि यह दर्शाता है कि परीक्षण संतुष्ट हो जाता है।

10.5.3 शृंखला सूचकांक तथा शृंखलिक परीक्षण

सूचकांकों की रचना में दो प्रकार के आधार कालों का उपयोग किया जाता है, जोकि इस प्रकार है : (क) स्थिर आधार, (ख) शृंखला आधार। प्राय: उपयोग किए जाने वाले सूचकांको में स्थिर

आधार का उपयोग किया जाता है। यह विधि, किसी वर्ष में हुई कीमत या मात्रा में परिवर्तनों की अवहेलना करती है। इस विधि में, बाद की किसी तिथि में, महत्त्वपूर्ण हो जाने वाली वस्तुओं को सम्मिलित करना या समय के साथ ह्रासमान महत्त्व वाली वस्तुओं को सूचकांक से निकालना कठिन होता है। शृंखला सूचकांक द्वारा इन कठिनाइयों को दूर किया जा सकता है।

एक उपयुक्त सूचकांक सूत्र (मान लीजिए लास्पियर) के उपयोग द्वारा सर्वप्रथम शृंखलित आपेक्षिक, जोकि निम्नलिखित में परिभाषित है, परिकलित किए जाते हैं।

शृंखलित आपेक्षिक = ऐसा सूचकांक जिसमें पिछला काल आधार के रूप में उपयोग किया गया हो।

विभिन्न शृंखलिक आपेक्षकों को उत्तरोतर गुणा करने पर शृंखला सूचकांक प्राप्त होता है। अत: काल n का शृंखलिक I_{0n}, जिसका आधारकाल 0 है, इस प्रकार प्राप्त किया जाता है :

$$I_{01} = I_{01}$$

$$I_{02} = I_{01} \times I_{12}$$

$$I_{03} = I_{01} \times I_{12} \times I_{23} = I_{02} \times I_{23}$$

....

....

$$I_{0n} = I_{01} \times I_{12} \times \ldots\ldots \times I_{(n-2)(n-1)} \times I_{(n-1)n}$$

उदाहरण 10.4 : निम्नलिखित समंकों के संदर्भ में शृंखला सूचकांकों के परिकलन की व्याख्या की गई है :

वर्ष	शृंखलित आपेक्षिक	शृंखला सूचकांक (आधार वर्ष 1970 = 100)
1970	100	100
1971	$I_{01} = 80$	$100 \times \frac{80}{100} = 80$
1972	$I_{12} = 120$	$80 \times \frac{120}{100} = 96$
1973	$I_{23} = 75$	$96 \times \frac{75}{100} = 72$

अत: 1971 से 1973 तक शृंखला सूचकांक, जिनका आधार वर्ष 1970 है, क्रमश: 80, 96 तथा 72 हैं।

शृंखलिक परीक्षण : यह परीक्षण कालोत्क्रमण परीक्षण का बहुत से वर्षों के लिए विस्तार है। इसके अनुसार, 1973 वर्ष के लिए उपरोक्त परिकलित शृंखला सूचकांक जिसका आधार वर्ष 1970 है, प्रत्यक्ष परिकलित सूचकांक, स्थिर आधार वर्ष 1970, के बराबर होना चाहिए।

संकेतन द्वारा,

$$I_{01} \times I_{12} \times \ldots\ldots \times I_{(n-1)n} \times I_{n0} = 1 \text{ (यहाँ यह ध्यान दीजिए कि } I_{0n} = \frac{1}{I_{n0}}\text{)}$$

एक स्थिर भार के समूह सूचकांक पर ध्यान दीजिए :

$$\frac{\sum p_1 q}{\sum p_0 q}$$

हम परीक्षण की व्याख्या इस प्रकार कर सकते हैं :

आधार वर्ष 0 लेकर हम उपरोक्त सूत्र का 1 से 3 वर्षों के लिए अनुरेखण (trace) कर सकते हैं :

$$\frac{\sum p_1 q}{\sum p_0 q} \times \frac{\sum p_2 q}{\sum p_1 q} \times \frac{\sum p_3 q}{\sum p_2 q} \times \frac{\sum p_0 q}{\sum p_3 q} = 1$$

शृंखलिक सूत्र को संतुष्ट करने वाले सूत्र निम्नलिखित हैं :

1) सरल समूही सूचकांक

2) मूल्यानुपातों का सरल गुणोत्तर माध्य

3) भारित समूही सूचकांक (जैसे स्थिर भार का लास्पियर सूचकांक)

4) मूल्यानुपातों का भारित गुणोत्तर माध्य जिसमें स्थिर भार उपयोग किए गए हों।

इस परीक्षण को फिशर का आदर्श सूचकांक संतुष्ट नहीं करता। यह प्रमाणित किया जा चुका है कि कोई भी सूचकांक दोनों, उपादानोत्क्रमण तथा शृंखलिक, परीक्षणों को एक साथ संतुष्ट नहीं कर सकता।

बोध प्रश्न 2

1) निम्नलिखित सारणी में 1980-84 वर्षों में A, B, तथा C वस्तुओं की औसत थोक बिक्री कीमतें दी हुई हैं। वर्ष 1980 की कीमतों को आधार मानकर, शृंखला सूचकांक का परिकलन कीजिए।

वस्तुएँ	औसत थोक बिक्री कीमतें (रुपयों में)				
	1980	1981	1982	1983	1984
A	20	16	28	35	21
B	25	30	24	36	45
C	20	25	30	24	30

....

....

....

....

....

....

....

....

....

....

....

2) निम्नलिखित समंकों द्वारा फिशर के आदर्श सूचकांक का परिकलन कीजिए तथा यह दर्शाइए कि इसके द्वारा उपादानोत्क्रमण तथा कालोत्क्रमण परीक्षण संतुष्ट होता है।

वस्तुएँ	आधार वर्ष		वर्तमान वर्ष	
	प्रति इकाई कीमत	व्यय (रुपयों में)	प्रति इकाई कीमत	व्यय (रुपयों में)
A	2	40	5	75
B	4	16	8	40
C	1	10	2	24
D	5	25	10	60

....

....

....

....

....

....

....

....

....

....

....

10.6 जीवन निर्वाह व्यय सूचकांक या उपभोक्ता-कीमत सूचकांक

यह व्यक्तियों के एक सजातीय (homogeneous) समूह, जैसे औद्योगिक श्रमिकों के परिवारों, द्वारा जीवन निर्वाह के लिए उपयोग की गई वस्तुओं तथा सेवाओं की कीमतों में परिवर्तन का सूचकांक होता है।

जीवन निर्वाह व्यय सूचकांक की रचना में सम्मिलित की जाने वाली मुख्य उपभोग वस्तुएँ निम्नलिखित होती हैं :

1) खाद्य सामग्री
2) ईंधन तथा प्रकाश
3) वस्त्र
4) मकान का किराया
5) विविध

उपभोग के समंक उस जनसंख्या वर्ग, जिसके लिए सूचकांक की रचना की जानी है, परिवार निर्वाह सर्वेक्षण (family living survey) द्वारा प्राप्त किए जाते हैं। कीमतों सम्बन्धी समंकों का संकलन उन विभिन्न फुटकर बाज़ारों से, जिनसे ये उपभोक्ता वस्तुएँ खरीदते हैं, किया जाता

है। यहाँ यह ध्यान देना चाहिए कि ऊपर लिखे हुए प्रत्येक व्यापक वर्ग में बहुत से छोटे-छोटे वर्ग होते हैं। जैसे खाद्य सामग्री में अनाज, दालें, तेल, मांस, मछली, अंडा, मसाले, सब्जी, फल, शर्बत आदि सम्मिलित होते हैं। इसके अतिरिक्त, विविध वर्ग में चिकित्सा, शिक्षा, परिवहन, मनोरंजन, उपहार, तथा इसी प्रकार की बहुत सी मदें सम्मिलित होती हैं। जब एक ही वस्तु की एक से अधिक निवेदित दरें (price quotations) एकत्रित की गई हों तो इनका सरल माध्य ले लिया जाता है। इन सभी वर्गों के लिए अलग-अलग सूचकांक की रचना वर्ग की कीमतों का भारित माध्य लेकर की जाती है; उपयोग किए जाने वाले भार एक औसत परिवार द्वारा उपभोग की गई वस्तुओं के व्यय के अनुपात में होते हैं। तत्पश्चात् समग्र सूचकांक (जीवन निर्वाह व्यय सूचकांक), इन वर्ग सूचकांकों के भारित माध्य का परिकलन करके प्राप्त किया जाता है।

यहाँ पर भी उपयोग किए जाने वाले भार विभिन्न वर्गों में किए गए व्यय के अनुपात में होते हैं (जैसे खाद्य सामग्री पर 50 प्रतिशत आदि)।

लास्पियर सूत्र के उपयोग द्वारा

जीवन निर्वाह व्यय सूचकांक : $I = \dfrac{\sum w\left(\dfrac{p_n}{p_0} \times 100\right)}{\sum w}$

जहाँ पर $w = \dfrac{p_0 q_0}{\sum p_0 q_0}$, वर्ग सूचकांक का भार है।

जीवन निर्वाह व्यय सूचकांक या उपभोक्ता कीमत सूचकांक के महत्त्वपूर्ण व्यवहारिक तात्पर्य तथा विस्तृत सार्वजनिक उपयोग हैं। इसका सबसे महत्त्वपूर्ण उपयोग मज़दूरी नियमन में होता है। कर्मचारियों का महँगाई भत्ता मुख्यत: इसी सूचकांक के आधार पर निर्धारित किया जाता है। जब मज़दूरी या आय को इसी सूचकांक से भाग किया जाता है तो आय पर कीमत वृद्धि या कमी के प्रभाव का विलोपन हो जाता है। इसको हम अवस्फीति की प्रक्रिया कहते हैं। जोकि वास्तविक मज़दूरी या आय ज्ञात करने में उपयोग की जाती है। जैसा कि पहले जिक्र किया जा चुका है, निर्वाह सूचकांक के व्युत्क्रम द्वारा मुद्रा की क्रय शक्ति को मापा जाता है।

उदाहरण 10.5 : खाद्य सामग्री के सूचकांक की रचना

वस्तुएँ	कीमत		भार		
	p_n	p_0	$P = (p_n + p_0) \times 100$	w	Pw
चावल	50	40	125.0	30	3750.0
गेहूँ	45	30	150.0	20	3000.0
दालें	60	40	150.0	10	1500.0
चीनी	40	20	200.0	5	1000.0
तेल	75	60	125.0	15	1875.0
आलू	60	50	120.0	15	1800.0
मछली	200	150	133.3	5	666.5
योग				100	13591.5

सूचकांक (खाद्य सामग्री) $= \dfrac{\sum w \times (p_n \div p_0)}{\sum w} \times 100$

$= \dfrac{13591.5}{100} = 135.915 = 135.92$

उदाहरण 10.6 : अंतिम जीवन निर्वाह व्यय सूचकांक की रचना

मद	भार (प्रतिशत व्यय)	सूचकांक	भार × सूचकांक
खाद्य सामग्री	45	130	5850
वस्त्र	15	140	2100
आवास	20	170	3400
ईंधन	5	110	550
विविध	15	125	1875
योग	100		13775

जीवन निर्वाह व्यय $= \dfrac{13{,}775}{100} = 137.75 = 138$

बोध प्रश्न 3

1) यदि विभिन्न प्रकार के व्यापार के तुलनात्मक मूल्यों का लेखा-जोखा रखा जाए तो, अक्टूबर 1979 से अक्टूबर 1980 की अवधि में व्यापार के आकार में प्रतिशत परिवर्तन की सूचक एक संख्या का परिकलन कीजिए।

व्यापार की किस्म	टन ('000)		प्राप्तियाँ ('000 रुपये)
	अक्तूबर 1979	अक्तूबर 1980	अक्तूबर 1979
सामान	1246	1206	776
खनिज	1125	981	252
ईंधन	4794	4229	562

....

....

....

....

....

....

....

....

....

....

....

2) निम्नलिखित समंकों से 1980 के लिए, आधार वर्ष 1970 लेकर, पाशे सूचकांक का परिकलन कीजिए।

वस्तु	इकाई	प्रति इकाई कीमत (रुपयों में)		विक्रित मात्रा	
		1970	1980	1970	1980
A	किलो	4	5	95	120
B	किलो	60	70	118	130
C	किलो	35	40	50	70

....

....

....

....

....

....

....

....

....

....

3) निम्नलिखित समंकों से 1980 के लिए, आधार वर्ष 1978 लेकर, लास्पियर सूचकांक का परिकलन कीजिए।

मद	कीमत (रुपयों में)		कुल मूल्य (रुपयों में)
	1978	1980	1978
A	12.50	14.00	112.50
B	10.50	12.00	126.00
C	15.00	14.00	105.00
D	9.40	11.20	47.00

....

....

....

....

....

....

....

....

....

....

....

4) निम्नलिखित समंकों से मार्शल-एजवर्थ सूचकांक का परिकलन कीजिए।

वस्तुएँ	1970		1977	
	कीमत	मात्रा	कीमत	मात्रा
चावल	9.3	100	4.5	90
गेहूँ	6.4	11	3.7	10
ज्वार	5.1	5	2.7	3

....

....

....

....

....

....

....

....

....

....

....

....

10.7 हल किए हुए उदाहरण

सूचकांक विषय के बारे में और जानकारी देने के लिए हम इस भाग में कुछ हल किए हुए उदाहरण देंगे।

उदाहरण 10.7 : कीमत सूचकांक की रचना

मद	इकाई	प्रति इकाई कीमत (रुपयों में)		
		1970 (p_0)	1980 (p_n)	($p_n \div p_0$) × 100
चावल	क्विंटल	100.00	220.00	220
गेहूँ	किलो	1.50	2.40	160
मछली	किलो	15.00	28.00	187
रोटी	पाउंड	0.60	1.35	225
दूध	लीटर	2.50	4.00	160
योग		119.60	255.75	952

i) **समूही विधि**

वर्ष 1980 का सूचकांक (आधार वर्ष 1970 = 100)

$$= \frac{1980 \text{ में प्रति इकाई कीमत}}{1970 \text{ में प्रति इकाई कीमत}} \times 100$$

$$= \frac{\sum p_n / k}{\sum p_0 / k} \times 100 = \frac{255.75}{119.60} \times 100 = 214$$

ii) **मूल्यानुपात विधि**

वर्ष 1980 का सूचकांक (आधार वर्ष 1970 = 100)

$$= \frac{\sum \frac{p_n}{p_0} \times 100}{k}$$

$$= 952/5 = 190$$

उदाहरण 10.8 : निम्नलिखित सूचना द्वारा सूचकांक परिकलित कीजिए :

i) भारित समूही सूत्र के उपयोग द्वारा

ii) मूल्यानुपात के भारित समांतर माध्य द्वारा

वस्तु	इकाई	प्रति इकाई कीमत (रुपयों में)		भार
		आधार वर्ष	वर्तमान वर्ष	
A	क्विंटल	85	115	19
B	किलो	15	15	25
C	दर्जन	45	61	40
D	लीटर	55	100	20
E	पाउंड	17	23	21

सूचकांकों का परिकलन

वस्तु	p_0	p_n	w	p_0w	p_nw	$I = (p_n \div p_0) \times 100$	Iw
A	85	115	19	1615	2185	135.3	2570.7
B	15	20	25	375	500	133.3	3332.5
C	45	61	40	1800	2440	135.6	5424.0
D	55	100	20	1100	2000	181.8	3636.0
E	17	23	21	357	483	135.3	2841.3
योग			125	5247	7608		17804.5

क) भारित समूही सूचकांक $= \dfrac{\sum p_n w}{\sum p_0 w} \times 100 = \dfrac{7608}{5247} \times 100 = 145.0$

ख) मूल्यानुपातों का भारित समांतर माध्य $= \dfrac{\sum Iw}{\sum w} = \dfrac{178045}{125} = 142.4$

उदाहरण 10.9 : निम्नलिखित समंकों में कुछ उपभोग वस्तुओं की कीमतें तथा उनसे सम्बन्धित विभिन्न वस्तुओं के भार दिए हुए हैं। वर्ष 1970 को आधार (= 100) मानकर, मूल्यानुपातों के

i) सरल माध्य के उपयोग से, तथा

ii) भारित माध्य के उपयोग से, 1971 के सूचकांक परिकलित कीजिए :

वस्तुएँ	इकाई	कीमत (रुपयों में)		भार
		1970	1971	
गेहूँ	किलो	0.50	0.75	2
दूध	लीटर	0.60	0.75	5
अण्डा	दर्जन	2.00	2.40	4
चीनी	किलो	1.80	2.10	8
जूते	जोड़ा	8.00	10.00	1

मूल्यानुपातों के सूचकांकों का परिकलन

वस्तुएँ	इकाई	p_0	p_n	$I = (p_n \div p_0) \times 100$	w	Iw
गेहूँ	किलो	0.50	0.75	150	2	300
दूध	लीटर	0.60	0.75	125	5	625
अण्डा	दर्जन	2.00	2.40	120	4	480
चीनी	किलो	1.80	2.10	117	8	936
जूते	जोड़ा	8.00	10.00	125	1	125
योग	--	--	--	637	20	2466

i) मूल्यानुपातों के सरल माध्य का सूचकांक $= \dfrac{\sum \left(p_n / p_0 \right) \times 100}{k} = \dfrac{637}{5} = 127.4$

ii) मूल्यानुपातों के भारित माध्य का सूचकांक $= \dfrac{\sum Iw}{\sum w} = \dfrac{2466}{20} = 123.3$

उदाहरण 10.10 : निम्नलिखित समंकों के आधार पर पाँचों वर्गों का संयुक्त थोक मूल्य सूचकांक परिकलित कीजिए :

वर्ग	भार	27.9.69 को समाप्त होने वाले सप्ताह का सूचकांक (आधार : 1952-53 = 100)
खाद्य सामग्री	50	241
शराब तथा तम्बाकू	2	221
ईंधन, ऊर्जा, प्रकाश तथा स्नेहक (lubricants)	3	204
औद्योगिक कच्चा माल	16	256
निर्मित वस्तुएँ	29	179

हम सामान्य सूचकांक $= \dfrac{\sum Iw}{\sum w}$ परिकलित करते हैं,

जहाँ पर $I =$ वर्ग सूचकांक तथा $w =$ वर्ग भार है।

वर्ग	भार (w)	वर्ग सूचकांक (I)	Iw
खाद्य सामग्री	50	241	12050
शराब तथा तम्बाकू	2	221	442
ईंधन, ऊर्जा, प्रकाश तथा स्नेहक (lubricants)	3	204	612
औद्योगिक कच्चा माल	16	256	4096
निर्मित वस्तुएँ	29	179	5191
योग	100		22391

थोक मूल्य सूचकांक $= \dfrac{22391}{100} = 223.91$

उदाहरण 10.11 : चार वस्तुओं का वार्षिक उत्पादन (10 लाख टन में) निम्नलिखित है :

वस्तुएँ	उत्पादन			भार
	1950	1954	1955	
A	160	200	216	20
B	24	42	45	30
C	50	72	68	13
D	120	168	156	17

वर्ष 1950 को आधार लेकर, दो वर्षों 1954 तथा 1955 के लिए मूल्यानुपातों का (i) सरल समांतर माध्य तथा (ii) भारित समांतर माध्य उपयोग करके, मात्रा सूचकांक परिकलित कीजिए।

हल

1950 को आधार (= 100) लेकर 1954 के लिए मात्रानुपात :

$I = (q_n / q_0) \times 100 = (q_{54} / q_{50}) \times 100$

वस्तु A : $\frac{200}{160} \times 100 = 125$

वस्तु B : $\frac{42}{24} \times 100 = 175$

वस्तु C : $\frac{72}{50} \times 100 = 144$

वस्तु D : $\frac{168}{120} \times 100 = 140$

1950 को आधार (= 100) लेकर 1955 के लिए मात्रानुपात :

$I = (q_{55} / q_{50}) \times 100$

वस्तु A : $\frac{216}{160} \times 100 = 135$

वस्तु B : $\frac{45}{24} \times 100 = 187.5$

वस्तु C : $\frac{68}{50} \times 100 = 136$

वस्तु D : $\frac{156}{120} \times 100 = 130$

वस्तु	मात्रानुपात (I)		भार (w)	Iw	
	1954	1955		1954	1954
A	125	135.0	20	2500	2700
B	175	187.5	30	5250	5625
C	144	136.0	13	1872	1768
D	140	130.0	17	2380	2210
योग	584	588.5	80	12002	12303

i) मात्रानुपातों का सरल समांतर माध्य $= \frac{\sum (q_n / q_0)}{k} \times 100$

(जहाँ पर $k =$ वस्तुओं की संख्या है)

1954 का सूचकांक $= \frac{584}{4} = 146$

1955 का सूचकांक $= \frac{588.5}{4} = 147$

ii) मात्रानुपातों का भारित समांतर माध्य $= \frac{\sum Iw}{\sum w}$

$$1954 \text{ का सूचकांक } = \frac{12002}{80} = 150$$

$$1955 \text{ का सूचकांक } = \frac{12003}{80} = 154$$

उदाहरण 10.12 : निम्नलिखित कीमत (p) तथा मात्रा (q) के समंकों से फिशर का आदर्श सूचकांक परिकलित कीजिए :

वस्तु	1970 (आधार वर्ष)		1978 (वर्तमान वर्ष)	
	कीमत	मात्रा	कीमत	मात्रा
A	12	10	17	10
B	14	9	16	11
C	11	12	13	10

फिशर के आदर्श सूचकांक का परिकलन :

वस्तु	p_0	q_0	p_n	q_n	p_0q_0	p_nq_0	p_0q_n	p_nq_n
A	12	10	17	10	120	170	120	170
B	14	9	16	11	126	144	154	176
C	11	12	13	10	132	156	110	130
योग					378	470	384	476

लास्पियर का कीमत सूचकांक $= \frac{\sum p_n q_0}{\sum p_0 q_0} \times 100 = \frac{470}{378} \times 100 = 124.34 = 124$

पाशे का कीमत सूचकांक $= \frac{\sum p_n q_0}{\sum p_0 q_n} \times 100 = \frac{476}{384} \times 100 = 123.96 = 124$

फिशर का आदर्श सूचकांक $= \sqrt{[L \times P]} = \sqrt{[(124 \times 124)]} = 124$

10.8 सारांश

इस इकाई में आपका परिचय सूचकांक की रचना तथा व्याख्या में उपयोग आने वाली अवधारणाओं तथा विधियों से कराया गया है। आपको यह दर्शाया गया है कि कीमत तथा मात्रा सूचकांक के परिकलन में लास्पियर, पाशे तथा फिशर के सूत्रों का उपयोग किस प्रकार किया जाता है। आपको यह भी ज्ञात हुआ है कि उपभोक्ता कीमत या जीवन निर्वाह व्यय में परिवर्तनों को कैसे मापा जा सकता है।

10.9 शब्दावली

आधार वर्ष	:	यह वर्ष, विचाराधीन चर की दृष्टि से, सामान्य होता है। आधार वर्ष का सूचकांक हमेशा 100 लिया जाता है। वर्तमान वर्ष के सूचकांक को आधार वर्ष के सूचकांक के प्रतिशत के रूप में व्यक्त किया जाता है।
शृंखला सूचकांक	:	वर्तमान वर्ष के सूचकांक को इससे पहले वर्ष के सूचकांक के प्रतिशत के रूप में व्यक्त किया जाता है।
सूचकांक	:	यह केवल एक संख्या है जोकि आधार वर्ष के मूल्य के प्रतिशत के रूप में व्यक्त की जाती है। सूचकांक द्वारा एक समयावधि में, वस्तुओं के एक वर्ग के विचाराधीन चर (कीमत, बिक्री मात्रा या निर्यात आदि) में परिवर्तन को मापा जाता है। यह सम्मिलित की गई वस्तुओं की कीमतों (या कोई और गुण) का एक विशेष भारित माध्य होता हैं।
मूल्यानुपात	:	एक सूचकांक की रचना में किसी वस्तु का मूल्यानुपात वर्तमान वर्ष की तथा आधार वर्ष की कीमतों का अनुपात होता है।
मात्रा सूचकांक	:	इस सूचकांक में विचाराधीन चर वस्तुओं की मात्राएँ होती हैं।

10.10 कुछ उपयोगी पुस्तकें

Nagar, A.L. and R.K. Das, 1989 : *Basic Statistics*, Oxford University Press, Delhi.

Goon, A.M., M.K. Gupta and B. Dasgupta, 1987 : *Basic Statistics*, The World Press Pvt. Ltd., Calcutta.

मेहता बी.सी. ; 1986, प्रारंभिक सांख्यिकी, राजस्थान हिन्दी ग्रन्थ अकादमी : जयपुर अध्याय 12

10.11 बोध प्रश्नों के उत्तर यां संकेत

बोध प्रश्न 1

1) आप स्वयं कीजिए।

2) आप स्वयं कीजिए।

3) सरल समूही विधि सूचकांक = 117.14

   मूल्यानुपात विधि = 122.9

4) 84.2

5) लास्पियर सूचकांक = 86.02

पाशे सूचकांक = 81.25

बोध प्रश्न 2

1) 108.33, 135.41, 160.23, 165.56

2) आप स्वयं कीजिए।

बोध प्रश्न 3

1) हम अक्टूबर 1979 को आधार लेकर अक्टूबर 1980 के लिए मात्रानुपात ज्ञात करते हैं। अपेक्षित सूचकांक को मात्रानुपातों का भारित समांतर माध्य परिकलित करके प्राप्त किया जा सकता है, जिसमें उपयोग किए जाने वाले भार 1979 की प्राप्तियाँ होंगी।

व्यापार की किस्म	q_0	q_n	भार (w)	मात्रानुपात $(q_n + q_0) \times 100$	$(4)\times(5)$
(1)	(2)	(3)	(4)	(5)	(6)
सामान	1246	1206	776	97	75272
खनिज	1125	981	252	87	21924
ईंधन	4794	4229	562	88	49456
योग	--	--	1590	--	146652

$$\text{मात्रा सूचकांक} = \frac{\sum (q_n / q_0) \times 100 \times w}{\sum w} = \frac{146652}{1590} = 92$$

2) पाशे के कीमत सूचकांक का परिकलन

वस्तुएँ	p_0	p_n	q_0	q_n	p_0q_n	p_nq_n
A	4	5	95	120	480	600
B	60	70	118	130	7800	9100
C	35	40	50	70	2450	2800
योग					10730	12500

$$\text{पाशे का कीमत सूचकांक} = \frac{\sum p_nq_n}{\sum p_0a_n} \times 100 = \frac{12500}{10730} \times 100 = 116$$

3) हमें आधार कीमत (p_0), वर्तमान कीमत (p_n) तथा आधार वर्ष में मूल्य (p_0q_0) दिया हुआ है। आधार वर्ष की मात्रा (q_0), ज्ञात करने के लिए निम्नलिखित संबंध का उपयोग कर सकते हैं :

$$q_0 = \frac{p_0q_0}{p_0}$$

p_0, p_n तथा q_0 के उपयोग द्वारा, हम लास्पियर सूचकांक इस सूत्र से ज्ञात कर सकते हैं :

$$L = \frac{\sum p_n q_0}{\sum p_0 q_0} \times 100$$

लास्पियर के कीमत सूचकांक का परिकलन

मद	p_0	p_n	$p_0 q_0$	$q_0 = p_0 q_0 \div p_0$	$p_n q_0$
A	12.50	14.00	112.50	9	126.00
B	10.50	12.00	126.00	12	144.00
C	15.00	14.00	105.00	7	98.00
D	9.40	11.20	47.00	5	56.00
योग	--	--	390.50	--	424.00

लास्पियर का कीमत सूचकांक $= \frac{\sum p_n q_0}{\sum p_0 q_0} \times 100 = \frac{424.00}{390.50} \times 100 = 109$

4) मार्शल-एजवर्थ सूचकांक $= \frac{\sum p_n (q_0 + q_n)}{\sum p_0 (q_0 + q_n)} \times 100$

$$= \frac{\sum p_n q_0 + \sum p_n q_n}{\sum p_0 q_0 + \sum p_0 q_n} \times 100$$

हम 1970 को आधार तथा 1977 को वर्तमान वर्ष ले लेते है :

वस्तु	p_0	q_0	p_n	q_n	$p_0 q_0$	$p_0 q_n$	$p_n q_0$	$p_n q_n$
चावल	9.3	100	4.5	90	930.0	837.0	450.0	405.0
गेहूँ	6.4	11	3.7	10	70.4	64.0	40.7	37.0
ज्वार	5.1	5	2.7	3	25.5	15.3	13.5	8.1
योग					1025.9	916.3	504.2	450.1

अपेक्षित सूचकांक $= \frac{504.2 + 450.1}{1025.9 + 916.3} \times 100 = 49.1$

10.12 पारिभाषिक शब्दावली

आधार काल	:	base period
उपभोक्ता कीमत सूचकांक	:	conusmer price index
उपादानोत्क्रमण परीक्षण	:	factor reversal test
कालश्रेणी	:	time series

कालोत्क्रमण परीक्षण	:	time reversal test
जीवन निर्वाह व्यय सूचकांक	:	cost of living index
मूल्यानुपात	:	price relative
वर्तमान काल	:	current period
समूही विधि	:	aggregative method
सूचकांक	:	index number
शृंखला सूचकांक	:	chain index
शृंखलिक आपेक्षिक	:	link relative
शृंखलिक परीक्षण	:	circular test
मूल्य सूचकांक	:	value index

इकाई 11 निश्चयवादी कालश्रेणी एवं पूर्वानुमान

इकाई की रूपरेखा

11.0 उद्देश्य

इस इकाई का अध्ययन करने के बाद आपको निम्नलिखित के बारे में जानकारी प्राप्त हो सकेगी:

- कालश्रेणी समंकों के लिए एक उपनति रेखा की रचना;
- चल-माध्यों का परिकलन; तथा
- मौसमी विचरण के विभिन्न परिमापों का परिकलन।

11.1 प्रस्तावना

एक कालश्रेणी, उत्तरोतर समय बिन्दुओं पर मापे जाने वाले चर के प्रेक्षणों का एक समुच्चय होती है। प्राय: चर के मान समान समय-अंतरालों, जैसे वार्षिक, त्रैमासिक आदि, पर दर्ज किए जाते है। सामान्यत:, कालश्रेणी आर्थिक समंकों के संदर्भ में होती है लेकिन इसका अनुप्रयोग अन्य क्षेत्रों में, जहाँ मात्रात्मक समंक एकत्रित हों, भी ठीक उसी प्रकार से होता है। राष्ट्रीय आय, कृषि आय, कृषि

उत्पादन की काल श्रेणियाँ, वार्षिक प्रेक्षणों पर आधारित होती है। काल श्रेणी के अन्य उदाहरण विभिन्न वर्षों में एक फसल का उत्पादन, विभिन्न समय बिन्दुओं पर एक देश की जनसंख्या, वर्ष के विभिन्न मौसमों में एक विभागीय भंडार की बिक्री, चाय का त्रैमासिक निर्यात आदि हैं। इस प्रकार के समंकों में 'समय' एक चर होता है, जिसको t से सूचित किया जाता है। तथा दूसरे चर को, जोकि समय पर आश्रित होता है जैसे उत्पादन, जनसंख्या, बिक्री, निर्यात आदि, y_t से सूचित किया जाता है। इस इकाई में विकसित की जाने वाली प्रणाली (methodology) की सहायता से हम इस प्रकार की कुछ काल श्रेणियों का विश्लेषण करेंगे।

11.2 कालश्रेणी समंकों के अध्ययन की समस्याएँ तथा उद्देश्य

कालश्रेणी समंकों द्वारा यह पता चलता है कि सामान्यत: समय परिवर्तन के साथ, आश्रित चर (y_t) के प्रेक्षित मानों में भी परिवर्तन होता है। ये परिवर्तन, बहुत सी शक्तियों जैसे जनसंख्या में वृद्धि, उत्पादन की तकनीक में परिवर्तन, लोगों की रुचि तथा आदतों में परिवर्तन, जलवायु में परिवर्तन आदि, की 'पारस्परिक क्रिया' (interaction) के कारण होते हैं। कालश्रेणी समंकों के अध्ययन का एक मुख्य उद्देश्य विभिन्न घटकों के प्रभावों को पृथक् करना तथा उनका माप करना होता है। इस विश्लेषण द्वारा हमें भूतकाल के व्यवहार तथा भविष्य के लिए पूर्वानुमान प्राप्त करने में सहायता मिलती है। इस प्रकार का पूर्वानुमान एक अर्थशास्त्री या एक व्यापारी, जोकि बिक्री से बहुत पहले अपने उत्पादन की योजना बना सकता है, के लिए बहुत ही महत्त्वपूर्ण होता है।

11.2.1 कालश्रेणी के घटक

कालश्रेणी के समंकों के आलेखी निरूपण द्वारा, समय के साथ परिवर्तनों को, व्यक्त किया जा सकता है। केवल किसी अपवादिक परिस्थिति में ही ऐसा संभव है, कि प्रेक्षण काल में, श्रेणी कोई परिवर्तन प्रदर्शित न करे। ये परिवर्तन पूर्ण रूप से संयोग या यादृच्छिक नहीं होता तथा कम से कम इनके एक अंश की व्याख्या तो की जा सकती है। इनमें से कुछ परिवर्तन आवर्ती (periodic) प्रकृति के होते हैं तथा दूसरे कुछ दीर्घकालीन वर्धन (growth) या पतन को प्रदर्शित करते हैं। इनमें कुछ अननुमेय (unpredictable) परिवर्तन, जोकि यादृच्छिक प्रकृति के होते हैं, भी मिले हुए होते हैं। यहाँ पर यह बात ध्यान में रखनी चाहिए कि सभी श्रेणियों में सभी प्रकार के परिवर्तनों का विद्यमान होना आवश्यक नहीं है। हम यहाँ यह मानते हैं कि एक सामान्य श्रेणी के चार महत्त्वपूर्ण घटक होते हैं :

i) दीर्घकालिक उपनति (T)

ii) मौसमी उच्चावचन (S)

iii) चक्रीय उच्चावचन (C)

iv) अनियमित या यादृच्छिक विचरण (I)

चिरप्रतिष्ठित उपगमन (classical approach) में हम यह मानते हैं कि प्रेक्षित मान y_t ऊपर दिए गए घटकों का, गुणनफल, अर्थात्

$y_t = T \times S \times C \times I$ (गुणात्मक प्रतिरूप)

या योग, अर्थात्

$y_t = T + S + C + I$ (योज्य प्रतिरूप) हो सकता है।

चाहे योज्य प्रतिरूप में परिकलन सुविधाजनक होते हैं, फिर भी काल श्रेणी के विश्लेषण में गुणात्मक प्रतिरूप का अत्यधिक उपयोग किया जाता है।

क) दीर्घकालिक उपनति

दीर्घकालिक उपनति से हमारा आशय श्रेणी के एक समयावधि के प्रेक्षणों में निर्विघ्न, नियमित तथा दीर्घकालिक परिवर्तन से होता है। कुछ श्रेणियाँ समय के साथ ऊर्ध्वमुखी तथा कुछ अधोमुखी उपनति को प्रदर्शित कर सकती है तथा कुछ अन्य लगभग स्थिर रह सकती है। श्रेणी की ऊर्ध्वमुखी उपनति, जनसंख्या वृद्धि, उत्पादन तकनीक में सुधार आदि उपादानों के कारण हो सकती है। उदाहरण के लिए, बहुत से उद्योगों के विकास का प्रतिरूप देश की जनसंख्या वृद्धि का अनुगमन करता है। इसके अतिरिक्त, तकनीकी विकास के कारण बहुत सी आर्थिक श्रेणियों में ऊर्ध्वमुखी विचरण हो सकते हैं। लेकिन सभी काल श्रेणियाँ वृद्धि को प्रदर्शित नहीं करती। कुछ श्रेणियाँ अधोमुखी भी हो सकती हैं तथा कुछ अन्य, उच्चावचनों (fluctuations) को प्रदर्शित कर सकती हैं। संभवतः किसी देश की अशोधित मृत्युदर (crude death rate) की काल श्रेणी अधोमुखी उपनति को प्रदर्शित करेगी।

ख) मौसमी उच्चावचन

अधिकतर काल श्रेणियों के आलेखों से पता चलता है कि दीर्घकालिक उपनति पर बहुत अधिक संख्या में उच्चावचन अध्यारोपित (superimposed) होते हैं। मौसमी विचरणों से अर्थ, श्रेणी के उन आवर्ती विचरणों से है जिनकी अवधि एक वर्ष से अधिक नहीं होती। आवर्ती विचरण एक नियमित समय अंतराल या अवधि के पश्चात् अपने को दोहराता है। उदाहरण के लिए, गर्मियों में शीत पेय की बिक्री में वृद्धि तथा सर्दियों में कमी होती है, वस्त्रों की बिक्री वर्ष के कुछ दिनों में अधिकतम होती है जैसे, मान लीजिए, मई के महीने में या कुछ त्यौहारों के दिनों में, कार्यालय जाने के घंटों के समय बसों में यात्रियों की संख्या अधिकतम होती है, एक सप्ताह के कुछ दिनों में पुस्तकालय से उधार ली गई पुस्तकों की संख्या अधिकतम होती है इत्यादि। इस प्रकार के उच्चावचनों में योगदान देने वाले उपादान विभिन्न मौसमों में जलवायु परिवर्तन, विभिन्न समयों में लोगों के रीति-रिवाजों तथा आदतें आदि होते हैं।

ग) चक्रीय उच्चावचन

चक्रीय उच्चावचन से अर्थ कालश्रेणी की दौलनी (oscillatory) गति से होता है जहाँ पर दोलन की अवधि को चक्र कहते हैं, जोकि एक वर्ष से अधिक होती है। इन उच्चावचनों में वे उपादान सम्मिलित होते हैं जो बारी-बारी से विस्तार एवं संकुचन को जन्म देते हैं। इस प्रकार के लक्षण बहुत सी आर्थिक तथा व्यापारिक श्रेणियों में होते हैं। कभी-कभी ये उच्चावचन अपने आकार, आयाम (amplitude) तथा दिशा की दृष्टि से बहुत ही अनियमित होते हैं। लेकिन इनके द्वारा प्रतिबिंबित परिघटनाएँ (phenomena) – मंदी (depression), समुत्थान (recovery), तेजी (boom) तथा निपात (collapse) – वस्तुतः सभी व्यापारिक तथा आर्थिक समंकों की काल श्रेणियों में देखने को मिलती हैं।

घ) अनियमित विचरण

इस वर्ग में वह सभी उपादान सम्मिलित किए जाते हैं जिनका कहीं और वर्गीकरण नहीं हुआ है। अतः ऐसे उपादान जैसे काम रोको, चुनाव, युद्ध, आग लगना आदि, एक कालश्रेणी को प्रभावित कर सकते हैं। इस वर्ग में वह सभी प्रकार के विचरण सम्मिलित किए जाते हैं जो दीर्घकालिक उपनति, मौसमी तथा चक्रीय उच्चावचनों में सम्मिलित नहीं हुए हैं। दुर्भाग्यवश, इस प्रकार के उपादानों का प्रायः चक्रीय उपादानों से भेद करना कठिन होता हैं। अतः कुछ विवेचनों में चक्रीय तथा अनियमित घटकों को एक साथ मिला दिया जाता है।

11.2.2 कालश्रेणी की रचना-एक उदाहरण

व्याख्या के लिए हम गुणात्मक प्रतिरूप की एक कालश्रेणी तैयार करते है। सारणी 11.1 में एक काल्पनिक कालश्रेणी के उपनति, मौसमी तथा चक्रीय-अनियमित घटक दिखाए गए हैं।

सारणी 11.1 : काल्पनिक कालश्रेणी तथा इसके घटक (त्रैमासिक)

वर्ष	त्रैमास	श्रेणी (y_t)	घटक: उपनति (T)	घटक: मौसमी ($100S$)	घटक: चक्रीय-अनियमित ($100CI$)
1	I	79	80	120	82
	II	58	85	80	85
	III	84	90	92	102
	IV	107	95	108	105
2	I	130	100	120	108
	II	93	105	80	132
	III	121	110	92	120
	IV	161	115	108	130
3	I	216	120	120	150
	II	132	125	80	132
	III	150	130	92	125
	IV	163	135	108	112
4	I	176	140	120	105
	II	112	145	80	97
	III	128	150	92	93
	IV	142	155	108	85
5	I	134	160	120	70
	II	86	165	80	65
	III	94	170	92	60
	IV	104	175	108	55

सारणी 11.1 में श्रेणी एक गुणात्मक प्रतिरूप द्वारा निरूपित है, इसलिए $y_t = T \times S \times C \times I$ होगा।

इस प्रकार पहले वर्ष के पहले त्रैमास के लिए प्रेक्षण $79 = 80 \times \frac{120}{100} \times \frac{82}{100}$

तथा चौथे वर्ष के दूसरे त्रैमास के लिए प्रेक्षण $112 = 145 \times \frac{80}{100} \times \frac{97}{100}$ है।

अत: प्रत्येक त्रैमासिक संख्या (y_t), दीर्घकालिक उपनति (T), मौसमी सूचकांक (S), चक्रीय तथा अनियमित घटक ($C \times I$) का गुणनफल है। इस प्रकार की कृत्रिम रचना, एक वास्तविक कालश्रेणी जैसी लगती है तथा कालश्रेणी समंकों के विश्लेषण के आधार के रूप में इस प्रतिरूप के उपयोग को प्रेरित करती है।

11.3 उपनति के माप

प्राय: हमारी रुचि कालश्रेणी के उपनति चलन की जानकारी प्राप्त करने में होती है। इसके लिए हमें, अन्य घटकों (मौसमी, चक्रीय तथा अनियमित) के श्रेणी पर प्रभाव का विलोपन करना होता है।

उपनति माप की दो महत्त्वपूर्ण विधियाँ चल-माध्य विधि तथा बहुपद समंजन विधि हैं। चल-माध्य विधि में माध्य-कलन-विधि (process of averaging) द्वारा उच्चावचनों का मसृणीकरण (smoothing)

करके दीर्घकालिक उपनति प्राप्त की जाती है। दूसरी विधि में, मूल या रूपांतरित चर के लिए एक उपयुक्त कोटि के बहुपद का चुनाव करके, न्यूनतम वर्ग विधि द्वारा, इसके स्थिरांक ज्ञात किए जाते हैं। बहुपद की कोटि का चुनाव समंकों को आलेख पृष्ठ पर अंकित करके किया जाता है जहाँ पर विभिन्न मापक्रम (scales) जैसे समांतर, अर्ध-लघुगणकीय या दोहरे लघुगणकीय उपयोग किए जा सकते हैं। कालश्रेणी के व्यवहार का अध्ययन तथा भविष्य के पूर्वानुमान के लिए, उपनति का माप आवश्यक होता है।

11.3.1 चल-माध्य विधि

यह उच्चावचनों के मसृणीकरण की सरल विधि है जिसमें श्रेणी के परस्परव्यापी (overlapping) कालों के माध्य परिकलित किए जाते हैं। सर्वप्रथम, चल-माध्य की उपयुक्त अवधि का चयन किया जाता है। यदि चयन की गई अवधि 3 वर्ष है तो 3 लगातार मानों, जोकि श्रेणी के परस्परव्यापी कालों को सम्मिलित करते हों, के माध्यों की श्रेणी परिकलित करके चल-माध्य परिकलित किए जाते हैं। यदि मूल श्रेणी को y_1, y_2, y_3 .... से सूचित किया जाए, तो पहले 3 मानों का माध्य $(y_1 + y_2 + y_3)/3$ होगा तथा यह इन तीन वर्षों की अवधि के मध्य बिन्दु के सम्मुख लिखा जाता है। यह प्रथम चल-माध्य मान होगा। दूसरे चल-माध्य को प्राप्त करने के लिए दूसरे से चौथे काल के मानों का माध्य परिकलित किया जाता है। यह मान $(y_2 + y_3 + y_4)/3$ है तथा इसे दूसरे तथा चौथे वर्ष की अवधि के मध्य के सम्मुख लिखा जाता है। इस प्रकार, इस प्रक्रिया की पुनरावृति की जाती है। यह स्पष्ट हो जाना चाहिए कि इस विधि द्वारा शुरू तथा अंत के कुछ वर्षों के चल-माध्य प्राप्त नहीं किए जा सकते।

यहाँ दो परिस्थितियों में भेद किया जा सकता है, अर्थात् जब चल-माध्य की अवधि विषम हो, तथा जब अवधि सम हो। यदि अवधि विषम है (जैसे तीन वर्ष), तो पहले चल-माध्य को दूसरे वर्ष के सम्मुख, दूसरे चल-माध्य को तीसरे वर्ष के सम्मुख, इत्यादि, लिखा जाता है। यदि अवधि सम है (जैसे चार वर्ष), तो पहला चल-माध्य दूसरे तथा तीसरे वर्ष के बीच लिखा जाएगा, तथा किसी वर्ष की उपनति ज्ञात करने के लिए इनका केन्द्रीयकरण आवश्यक होता है।

व्याख्या के लिए, हम केन्द्रित 4-वर्ष चल-माध्य के परिकलन के विधिवत निरूपण पर विचार करते हैं। यहाँ हम दो विधियाँ प्रस्तुत करेंगे – प्रत्यक्ष विधि (सारणी 11.2) तथा लघुतर विधि (सारणी 11.3)।

सारणी 11.2 : चार वर्षीय केन्द्रित चल-माध्य का परिकलन (प्रत्यक्ष विधि)

वर्ष	y_t	4-वर्षीय चल योग	4-वर्षीय चल-माध्य स्तम्भ 3 ÷ 4	केन्द्रित चल योग	केन्द्रित 4-वर्षीय चल-माध्य
(1)	(2)	(3)	(4)	(5)	(6)
1	y_1			—	—
2	y_2			—	—
		$y_1+y_2+y_3+y_4=T_1$	$T_1/4$		
3	y_3			$(T_1+T_2)/4$	$(T_1+T_2)/8$
		$y_2+y_3+y_4+y_5=T_2$	$T_2/4$		
4	y_4			$(T_2+T_3)/4$	$(T_2+T_3)/8$
		$y_3+y_4+y_5+y_6=T_3$	$T_3/4$		
5	y_5			$(T_3+T_4)/4$	$(T_3+T_4)/8$
		$y_4+y_5+y_6+y_7=T_4$	$T_4/4$		
6	y_6			—	—
7	y_7			—	—

सारणी 11.3 : चार वर्षीय केन्द्रित चल-माध्य का परिकलन (लघुतर विधि)

वर्ष	y_t	4-वर्षीय चल योग (M.T.)	स्तम्भ 3 के दो मदों का चल-योग (केन्द्रित)	केन्द्रित 4-वर्षीय चल माध्य (M.A.), स्तम्भ 4 ÷ 8
(1)	(2)	(3)	(4)	(5)
1	y_1		—	—
2	y_2		—	—
		$y_1+y_2+y_3+y_4=T_1$		
3	y_3		(T_1+T_2)	$(T_1+T_2)/8$
		$y_2+y_3+y_4+y_5=T_2$		
4	y_4		(T_2+T_3)	$(T_2+T_3)/8$
		$y_3+y_4+y_5+y_6=T_3$		
5	y_5		(T_3+T_4)	$(T_3+T_4)/8$
		$y_4+y_5+y_6+y_7=T_4$		
6	y_6		—	—
7	y_7		—	—

ऊपर दिए हुए उदाहरण में चल-माध्य की अवधि 4 वर्ष है। दोनों विधियों, प्रत्यक्ष तथा लघुतर, में स्तम्भ 3, चार वर्षीय चल-योग को प्रदर्शित करता हैं। पहला योग (T_1) दूसरे तथा तीसरे वर्षों में मध्य के सम्मुख, दूसरा योग (T_2) तीसरे तथा चौथे वर्षों के मध्य के सम्मुख लिखा जाता है इत्यादि। तत्पश्चात्, 2 मदों का चल-माध्य परिकलित करके क्रमश: तीसरे, चौथे, ... वर्षों के सम्मुख लिखे जाते हैं (सारणी 11.2)।

लघुतर विधि (सारणी 11.3) में, 4 वर्षीय चल-माध्य परिकलित (जैसा कि सारणी 11.2 के स्तम्भ 4 में दर्शाया गया है) न करके 4 वर्षीय चल-योग परिकलित किए जाते हैं। इसके बाद 2 मद चल-योग (स्तम्भ 4, सारणी 11.3) परिकलित किए जाते हैं। अंत में स्तम्भ 5 में 4-वर्षीय केन्द्रित चल-माध्य परिकलित किए गए हैं। यहाँ पर यह ध्यान दें कि 4-वर्षीय चल-माध्य के लिए केन्द्रीयकरण के कारण श्रेणी के प्रत्येक छोर पर (4/2) = 2 वर्ष के लिए मान प्राप्त नहीं होते।

11.3.2 चल-माध्यों की उपयुक्तता

चल-माध्य विधि का प्रयोग सरल होता है लेकिन इस विधि की सफलता, अवधि के उपयुक्त चुनाव पर निर्भर होती है। एक चल-माध्य, जिसकी अवधि, चक्रीय अवधि के बराबर या इसका गुणज है, चक्रीय घटक का पूर्ण रूप से विलोपन कर देता है तथा उपनति का आकलन (estimate) प्रदान करता है। यह विधि लोचशील है लेकिन श्रेणी के शुरू तथा अंत में कुछ उपनति मान प्राप्त नहीं होते तथा चल माध्य की अवधि में वृद्धि होने पर इनकी संख्या में भी वृद्धि होती है। इसके अतिरिक्त, चूँकि चल-माध्य किसी परिवर्तन के नियम को नहीं अपनाता, इसलिए इस विधि का भविष्य उपनति के पूर्वानुमान में उपयोग नहीं हो सकता।

11.3.3 चल-माध्यों के उदाहरण

उदाहरण 11.3.1: निम्नलिखित समंकों से तीन तथा पाँच वर्षीय चल-माध्यों का परिकलन कीजिए:

वर्ष	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981
उत्पादन ('000 टन)	18	19	20	22	20	19	22	24	25	24	25	26

हल

सारणी 11.3.1 : (I) 3-वर्षीय चल-माध्य, (II) 5-वर्षीय चल-माध्य का परिकलन

वर्ष	उत्पादन	3-वर्षीय चल योग	3-वर्षीय चल माध्य	5-वर्षीय चल योग	5-वर्षीय चल माध्य
(1)	(2)	(3)	(4)	(5)	(6)
1970	18	—	—	—	—
1971	19	57	19.0	—	—
1972	20	61	20.3	99	19.8
1973	22	62	20.7	100	20.0
1974	20	61	20.3	103	20.6
1975	19	61	20.3	107	21.4
1976	22	65	21.7	110	22.0
1977	24	71	23.7	114	22.8
1978	25	73	24.3	120	24.0
1979	24	74	24.7	124	24.8
1980	25	75	25.0	—	—
1981	26	—	—	—	—

परिकलन के चरण

1) स्तम्भ 3 की संख्याएँ, स्तम्भ 2 की लगातार तीन संख्याओं के योग द्वारा प्राप्त की गई हैं। अत: पहला चल-योग $57 = 18 + 19 + 20$ है और इसको 1971 के सम्मुख लिखा गया है। दूसरा चल-योग $61 = 19 + 20 + 22$, 1972 के सम्मुख लिखा गया है।

2) स्तम्भ 4 में दर्शाए गए 3-वर्षीय चल-माध्य, स्तम्भ 3 में दिए हुए चल-योगों को 3 (चल-माध्य की अवधि) से भाग करके प्राप्त किए गए हैं। इस प्रकार $57 \div 3 = 19$, $61 \div 3 = 20.3$ आदि।

3) स्तम्भ 5 में दिए हुए 5-वर्षीय चल-योग, स्तम्भ 2 की लगातार 5 संख्याओं के योग द्वारा, प्राप्त किए गए हैं। इस प्रकार 1972 के सम्मुख प्रथम चल-योग $99 = 18 + 19 + 20 + 22 + 20$ है।

4) स्तम्भ 6 में 5-वर्षीय चल-माध्य, स्तम्भ 5 के चल-योगों को 5 से भाग करके प्राप्त किए गए हैं। इस प्रकार, 1975 का चल-माध्य $107 \div 5 = 21.4$ है। यहाँ यह ध्यान दें कि 3-वर्षीय केन्द्रित चल-माध्य के लिए $\frac{3-1}{2} = 1$ वर्ष तथा 5-वर्षीय केन्द्रित चल-माध्य के लिए $\frac{5-1}{2} = 2$ वर्ष, श्रेणी के क्रमश: आरम्भ तथा अंत में छूट जाते हैं।

उदाहरण 11.3.2 : 4-वर्षीय चल-माध्य के उपयोग द्वारा, निम्नलिखित काल श्रेणी के उपनति मान परिकलित कीजिए।

वर्ष	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988
उत्पादन ('000 टन)	52	54	55	57	58	61	63	66	67	70

सारणी 11.3.2 (क) : 4-वर्षीय चल-माध्य का परिकलन (प्रत्यक्ष विधि)

वर्ष	उत्पादन	4-वर्षीय चल-योग	4-वर्षीय चल-माध्य	स्तंभ 4 की दो मदों का चल योग (केन्द्रित)	केन्द्रित 4-वर्षीय चल-माध्य
(1)	(2)	(3)	(4)	(5)	(6)
1979	52	—	—	—	—
1980	54	—	—	—	—
		218	54.50		
1981	55			110.50	55.250
		224	56.00		
1982	57			113.75	56.875
		231	57.75		
1983	58			117.50	58.750
		239	59.75		
1984	61			121.75	60.875
		248	62.00		
1985	63			126.25	63.125
		257	64.25		
1986	66			130.75	65.375
		266	66.50		
1987	67	—	—	—	—
1988	70	—	—	—	—

सारणी 11.3.2 (ख) : 4-वर्षीय चल-माध्य का परिकलन (लघुतर विधि)

वर्ष	उत्पादन	4-वर्षीय चल योग	स्तंभ 3 के दो मदों का चल-योग (केन्द्रित)	केन्द्रित 4-वर्षीय चल माध्य
(1)	(2)	(3)	(4)	(5)
1979	52	—	—	—
1980	54	—	—	—
		218		
1981	55		442	55.250
		224		
1982	57		455	56.875
		231		
1983	58		470	58.750
		239		
1984	61		487	60.875
		248		
1985	63		505	63.125
		257		
1986	66		523	65.375
		266		
1987	67	—	—	—
1988	70	—	—	—

परिकलन के चरण (प्रत्यक्ष विधि)

1) स्तम्भ 2 की चार लगातार संख्याओं का योग स्तम्भ 3 में लिखा गया है।
इस प्रकार $52 + 54 + 55 + 57 = 218$, $54 + 55 + 57 + 58 = 224$ आदि।

2) स्तम्भ 4 = स्तम्भ 3 ÷ 4, इस प्रकार $218 \div 4 = 54.5$, $224 \div 4 = 56$ आदि।

3) स्तम्भ 4 की लगातार दो संख्याओं का योग स्तम्भ 5 में लिखा गया है।
अत: $54.5 + 56.0 = 110.5$, $56.00 + 57.75 = 113.75$ आदि।

4) स्तम्भ 6 = स्तम्भ 5 ÷ 2, अत: $110.5 \div 2 = 55.25$ आदि।

परिकलन के चरण (लघुतर विधि)

1) स्तम्भ 3 की दो लगातार संख्याओं का योग स्तम्भ 4 में लिखा गया है।
अत: $218 + 224 = 442$, $224 + 231 = 455$ आदि।

2) स्तम्भ 5 = स्तम्भ 4 ÷ 8 है। अत: $442 \div 8 = 55.25$ आदि।

उदाहरण 11.3.3 : निम्नलिखित श्रेणी में 3 वर्षीय भारित चल-माध्य उपयोग करके उपनति मान ज्ञात कीजिए, जिसमें भार 1, 2, 1 हैं।

वर्ष :	1	2	3	4	5	6
मान :	2	3	5	6	8	11

हल

सारणी 11.3.3 : तीन-वर्षीय भारित चल-माध्य का परिकलन

वर्ष	मान	तीन-वर्षीय भारित चल-योग (M.T.)	तीन-वर्षीय भारित चल-माध्य (M.A.)
(1)	(2)	(3)	(4)
1	2	—	—
2	3	13	3.25
3	5	19	4.75
4	6	25	6.25
5	8	33	8.25
6	11	—	—

परिकलन के चरण

1) स्तम्भ 2 की संख्याओं के 3-वर्षीय भारित चल-योग, स्तम्भ 3 में लिखे गए हैं।

अत: $1 \times 2 + 2 \times 3 + 1 \times 5 = 13$

$1 \times 3 + 2 \times 5 + 1 \times 6 = 19$

2) स्तम्भ 4 = स्तम्भ 3 ÷ (भारों का योग)

अत: $13 \div 4 = 3.25$, $19 \div 4 = 4.75$ आदि।

उदाहरण 11.3.4 : निम्नलिखित काल श्रेणी समंकों से 4-वर्षीय केन्द्रित चल-माध्य परिकलित कीजिए।

त्रैमास	वर्ष			
	1980	1981	1982	1983
1	62	66	72	79
2	58	60	67	74
3	72	74	80	88
4	60	64	69	77

हल

वर्ष	त्रैमास	मान	4-त्रैमासिक चल-योग	केन्द्रित चल-योग	4-त्रैमासिक चल-माध्य
(1)	(2)	(3)	(4)	(5)	(6)
1980	1	62	—	—	—
	2	58	—	—	—
			252		
	3	72		508	63.50
			256		
	4	60		514	64.25
			258		
1981	1	66		518	64.75
			260		
	2	60		524	65.50
			264		
	3	74		534	66.75
			270		
	4	64		547	68.38
			277		
1982	1	72		560	70.00
			283		
	2	67		571	71.38
			288		
	3	80		583	72.88
			295		
	4	69		597	74.63
			302		
1983	1	79		612	76.50
			310		
	2	74		628	78.50
			318		
	3	88	—	—	—
	4	77	—	—	—

बोध प्रश्न 1

1) निम्नलिखित समंकों द्वारा 1961 से 1970 तक के औद्योगिक उत्पादन सूचकांक दिए हुए हैं :

वर्ष :									
1961	1962	1963	1964	1965	1966	1967	1968	1969	1970
औद्योगिक उत्पादन सूचकांक :									
109.2	119.8	129.7	140.8	153.8	153.2	152.6	163.0	175.3	184.3

3-वर्षीय चल-माध्य विधि द्वारा उपनति रेखा का आकलन कीजिए।

....

....

....

....

....

....

....

....

....

....

....

2) निम्नलिखित, 1970-71 की कीमतों पर, भारत की शुद्ध राष्ट्रीय उत्पाद की संख्याओं को आलेखित कीजिए तथा 5-वर्षीय चल-माध्य द्वारा परिकलित उपनति से इसको अध्यारोपित (superimpose) कीजिए।

वर्ष	शुद्ध राष्ट्रीय उत्पाद ('00 करोड़ रुपये)	वर्ष	शुद्ध राष्ट्रीय उत्पाद ('00 करोड़ रुपये)	वर्ष	शुद्ध राष्ट्रीय उत्पाद ('00 करोड़ रुपये)
1961	259	1969	335	1977	444
1962	269	1970	356	1978	482
1963	275	1971	378	1979	513
1964	292	1972	386	1980	487
1965	315	1973	382	1981	521
1966	301	1974	397	1982	550
1967	299	1975	400		
1968	324	1976	440		

....

....

....

....

11.4 बहुपद समंजन विधि

बहुपद समंजन विधि शायद उपनति निर्धारण की उत्तम तथा वस्तुनिष्ठ (objective) विधि है। यहाँ उपनति के लिए एक उपयुक्त बहुपद का चयन करके, काल श्रेणी समंकों द्वारा उपनति समीकरण में उपयोग किए जाने वाले स्थिरांकों के मान परिकलित किए जाते है। समंकों के आलेखी निरूपण, जिसमें उपयोग किए जाने वाले समांतर मापक्रम के अतिरिक्त, अर्द्ध-लघुगणकीय या दोहरे-लघुगणकीय मापक्रम भी उपयोग किए जा सकते हैं, के द्वारा उपयुक्त बहुपद के चयन में सहायता मिलती है। यदि साधारण आलेख पृष्ठ पर अंकित समंक लगभग सरल रेखिय प्रवृत्ति दर्शाते हैं तो $Y = a + bx$ (सरल रेखा या प्रथम कोटि बहुपद) समीकरण का उपयोग किया जाता है।

यदि अंकित समंक, अर्द्ध-लघुगणकीय आलेख पृष्ठ पर सरल रेखा प्रदर्शित करते हैं, तो $\log Y = a + bx$ समीकरण उपयोग किया जाता है यह समीकरण $Y = A.B^x$ (घातीय फलन) का लघु (log) लेने पर प्राप्त होती है। यह ध्यान दें कि $a = \log A$ तथा $b = \log B$ है।

कभी-कभी एक द्वि या त्रि-कोटि बहुपद का समंजन भी किया जा सकता है।

$Y = a + bx + cx^2$ (द्वि-कोटि बहुपद या पैराबोला)

$Y = a + bx + cx^2 + dx^3$ (त्रि-कोटि बहुपद)

इन समीकरणों में उपयोग किए गए स्थिरांकों (जैसे $a, b, c, \ldots$) के मान न्यूनतम वर्ग विधि सिद्धांत, जैसा कि समाश्रयण में किया गया था (इकाई 9 देखें), के आधार पर आकलित किए जाते हैं। इस सिद्धांत के अनुसार विभिन्न स्थिरांकों के मान ऐसे होने चाहिए कि विचलनों के वर्ग का योग $\sum (y - Y)^2$ न्यूनतम हो जाय, जहाँ पर y = प्रेक्षित माप तथा Y = प्रत्याशित माप, जोकि $Y = a + bx$ या $Y = a + bx + cx^2$ इत्यादि, उपनति समीकरण से प्राप्त किया गया है। यहाँ पर योग $(\sum)$ सभी प्रेक्षणों पर किया गया है।

जब न्यूनतम वर्ग विधि द्वारा सरल रेखा का समंजन किया जाता है, तब a तथा b स्थिरांकों के मान, निम्नलिखित प्रसामान्य समीकरणों (normal equations) के हल द्वारा प्राप्त किए जाते हैं :

$$\sum y = na + b\sum x \quad \text{तथा}$$

$$\sum xy = a\sum x + b\sum x^2$$

जहाँ पर n वर्षों की संख्या को सूचित करता है।

इसी प्रकार पैराबोला (parabola) या द्वि-कोटि बहुपद के समंजन में a, b, तथा c स्थिरांकों के मान निम्नलिखित तीन प्रसामान्य समीकरणों के हल द्वारा प्राप्त होते हैं।

$$\sum y \quad = na + b\sum x + c\sum x^2$$

$$\sum xy \quad = a\sum x + b\sum x^2 + c\sum x^3$$

$$\sum x^2 y \quad = a\sum x^2 + b\sum x^3 + c\sum x^4$$

प्रसामान्य समीकरण लिखने के नियम

प्रथम सामान्य समीकरण प्राप्त करने के लिए, समीकरण के दोनों पक्षों को a के गुणांक से गुणा करके सभी प्रेक्षणों के लिए योग कर दीजिए।

इस प्रकार सरल रेखा $y = a + bx$ में a का गुणांक 1 है, अत: प्रथम प्रसामान्य समीकरण $\sum y = na + b\sum x$ होगा।

द्वितीय प्रसामान्य समीकरण के लिए, समीकरण के दोनों पक्षों को b के गुणांक से गुणा करके सभी प्रेक्षणों के लिए योग कर दीजिए। सरल रेखा समीकरण में b का गुणांक x है। अत: द्वितीय प्रसामान्य समीकरण $\sum xy = a\sum x + b\sum x^2$ होगा।

अब हम प्रथम-कोटि बहुपद उपनति समंजन पर विचार करते हैं, जिसमें वर्षों की संख्या विषम (सारणी 11.4) तथा सम (सारणी 11.5) है।

स्थिति I : विषम संख्या में वर्ष ($n = 5$)

सारणी 11.4

वर्ष	y	x	x^2	xy
(1)	(2)	(3)	(4)	(5)
1	y_1	−2	4	
2	y_2	−1	1	
3	y_3	0	0	
4	y_4	1	1	
5	y_5	2	4	
योग	Σy	0	10	Σxy

प्रसामान्य समीकरण इस प्रकार हैं :

$$\sum y = 5a + b\sum x = 5a \quad (\text{क्योंकि } \sum x = 0)\,.$$

$$\sum xy = a\sum x + b\sum x^2 = 10b$$

$$\therefore\ a = \frac{\sum y}{5}, b = \frac{\sum xy}{10}$$

जहाँ पर मूल बिन्दू ($x = 0$), 5 वर्ष के अंतराल का मध्य बिन्दु, अर्थात् तृतीय वर्ष है तथा समय की इकाई एक वर्ष है। वास्तविक जीवन परिस्थितियों में y_i के वास्तविक मान लिखित किए जा सकते हैं, अत: Σx तथा Σxy ज्ञात होते हैं।

स्थिति II : सम संख्या में वर्ष ($n = 6$)

सारणी 11.5

वर्ष	y	x	x^2	xy
(1)	(2)	(3)	(4)	(5)
1	y_1	−5	25	
2	y_2	−3	9	
3	y_3	−1	1	
4	y_4	1	1	
5	y_5	3	9	
6	y_6	5	25	
योग	Σy	0	70	Σxy

स्थिरांकों, a तथा b के मान निम्नलिखित समीकरणों से प्राप्त होंगे :

$\Sigma y = 6a$

$\Sigma xy = 70b$

यहाँ पर मूलबिन्दु ($x = 0$), तीसरे तथा चौथे वर्ष के मध्य होगा तथा x की इकाई 6 मास होगी।

11.4.1 न्यूनतम वर्ग विधि की उपयुक्तता

उपनति रेखाओं का उपयोग, काल श्रेणी की संवृद्धि या ह्रास का पूर्वानुमान तथा अर्थव्यवस्था की दीर्घकालिक प्रवृत्तियों के अध्ययन में किया जाता है। बहुपद समंजन विधि में व्यक्तिगत अभिनति (bias) का पूर्ण विलोपन होता है तथा सभी वर्षों के लिए उपनति मान ज्ञात किए जा सकते हैं, जोकि चल-माध्य में सम्भव नहीं है। लेकिन, यहाँ पर बहुपद कि किस्म का चयन स्वेच्छ (arbitrary) होता है तथा दृढ़ रूप से यह जानना सम्भव नहीं है कि कौन सा वक्र (रैखिक या पैराबोला) उपनति का सही प्रकार से निरूपण करेगा। इस प्रकार, उपनति समीकरण का चयन स्वयं अभिनति उत्पन्न कर सकता है। समंकों के प्रकीर्ण आरेख द्वारा उपनति के प्रतिरूप के बारे में कुछ जानकारी प्राप्त की जा सकती है।

11.4.2 न्यूनतम वर्ग विधि के उदाहरण

उदाहरण 11.4.1 : निम्नलिखित काल श्रेणी समंकों के लिए न्यूनतम वर्ग विधि द्वारा सरल रेखा उपनति का समंजन कीजिए।

वर्ष :	1975	1976	1977	1978	1979	1980	1981
उत्पादन :	81	92	100	105	112	120	126

वर्ष 1982 के उत्पादन का आकलन कीजिए।

हल :

यहाँ पर वर्षों की संख्या विषम ($n = 7$) है। मान लीजिए, सरल रेखा उपनति का समीकरण $y = a + bx$ है, जिसका मूल बिन्दू ($x = 0$) 1978 पर है तथा x की इकाई एक वर्ष है। न्यूनतम वर्ग विधि द्वारा प्रसामान्य समीकरण इस प्रकार हैं (इकाई 9 देखें) :

$$\sum y = na + b\sum x$$

$$\sum xy = a\sum x + b\sum x^2$$

सारणी 11.4.1 : सरल रेखा उपनति समंजन

वर्ष	उत्पादन (y)	x	x^2	xy
(1)	(2)	(3)	(4)	(5)
1975	81	– 3	9	– 243
1976	92	– 2	4	– 184
1977	100	– 1	1	– 100
1978	105	0	0	0
1979	112	1	1	112
1980	120	2	4	240
1981	126	3	9	378
योग	736	0	28	203

अतः, इस सारणी से Σy, Σxy, Σx, तथा Σx^2 के मान प्रसामान्य समीकरणों में रखने पर

$7a = 736$, इस प्रकार $a = 105.1$
$28b = 203$, इस प्रकार $b = 7.21$

उपनति समीकरण

$Y = 105.1 + 7.21x$ है,
जिसका मूल बिन्दु 1978 पर तथा x की इकाई एक वर्ष है।

1982 के लिए x का मान 4 होगा। अत: 1982 के लिए उत्पादन का आकलन

$Y = 105.1 + 4 \times 7.21 = 133.94$ होगा।

उदाहरण 11.4.2 :

निम्नलिखित कालश्रेणी समंकों के लिए सरल रेखा का समंजन कीजिए।

वर्ष :	1970	1971	1972	1973	1974	1975
लाभ : (लाख रुपये)	3.1	3.3	3.6	3.2	3.7	3.9

1976 के लाभ का आकलन कीजिए।

हल :

यहाँ पर वर्षों की संख्या सम ($n = 6$) है। मान लीजिए, उपनति समीकरण $y = a + bx$ है जिसका मूल बिन्दु ($x = 0$) 1972 तथा 1973 का मध्य-बिन्दु है तथा x की इकाई 6 मास है। प्रसामान्य समीकरण इस प्रकार है :

$$\sum y = na + b\sum x$$

$$\sum xy = a\sum x + b\sum x^2$$

सारणी 11.4.2 : सरल रेखा उपनति का समंजन

वर्ष	लाभ (y) (लाख रुपये)	x	x^2	xy
(1)	(2)	(3)	(4)	(5)
1970	3.1	– 5	25	–15.5
1971	3.3	– 3	9	– 9.9
1972	3.6	– 1	1	– 3.6
1973	3.2	1	1	0.0
1974	3.7	3	9	11.1
1975	3.9	5	25	19.5
योग	20.8	0	70	4.8

अतः, इस सारणी से Σy, Σxy, Σx, तथा Σx^2 के मान प्रसामान्य समीकरणों में रखने पर

$6a = 20.8$, इस प्रकार $a = 3.47$
$70b = 4.8$, इस प्रकार $b = 0.07$

उपनति समीकरण

$Y = 3.47 + 0.07x$ है,
जिसका मूल बिन्दु 1972 तथा 1973 का मध्य है तथा x की इकाई 6 मास है।

वर्ष 1976 के लिए $x = 7$
इस प्रकार 1976 के लिए आकलन $Y = 3.47 + 0.07 \times 7 = 3.47 + 0.49 = 3.96$
अतः 1976 के लिए आकलित लाभ 3.96 लाख रुपये होगा।

उदाहरण 11.4.3 :

निम्नलिखित सारणी में एक कम्पनी की 1980 से 1986 वर्षों के लिए बिक्री (हजार रुपयों में) दी हुई है। एक घातीय फलन ($Y = A.B^x$) का समंजन कीजिए तथा 1987 वर्ष के लिए बिक्री का आकलन कीजिए।

वर्ष :	1980	1981	1982	1983	1984	1985	1986
बिक्री :	32	47	65	92	132	190	275

हल :

यहाँ वर्षों की संख्या विषम ($n = 7$) है। समीकरण के दोनों पक्षों का लघु लेने पर
$\log Y = \log A + x \log B$

मान लिया, $a = \log A$ तथा $b = \log B$
अतः $\log Y = a + bx$

हम मूल बिन्दु ($x = 0$), 1983 तथा x की इकाई एक वर्ष ले लेते है। न्यूनतम वर्ग प्रसामान्य समीकरण निम्नलिखित हैं :

$$\sum \log y = na + b\sum x$$

$$\sum x \log y = a\sum x + b\sum x^2$$

सारणी 11.4.3 : सरल रेखा उपनति का समंजन

वर्ष	बिक्री (y)	x	x^2	$\log y$	$x.\log y$
1980	32	–3	9	1.5051	–4.5153
1981	47	–2	4	1.6721	–3.3442
1982	65	–1	1	1.8129	–1.8129
1983	92	0	0	1.9638	0
1984	132	1	1	2.1206	2.1206
1985	190	2	4	2.2788	4.5576
1986	275	3	9	2.4398	7.3119
योग	833	0	28	13.7931	4.3237

इस सारणी से $\Sigma \log y$, $\Sigma x.\log y$, Σx, तथा Σx^2 के मान प्रसामान्य समीकरणों में प्रतिस्थापित करने पर

$7a = 13.7931$, या $a = 1.9700$
$28b = 4.3237$, या $b = 0.154$

अत: समंजित फलन $\log Y = a + bx$ या $Y = \text{antilog}\,(a + bx)$ होगा।

वर्ष 1987 के लिए x का मान 4 होगा।
अत:, वर्ष 1987 के लिए आकलित मान

$Y = \text{antilog}\,(1.97 + 0.154 \times 4) = \text{antilog}\,2.586 = 385.48$ है।

जब वर्षों की संख्या सम हो तो, सरल रेखा समंजन विधि (उदाहरण 11.4.2) की तरह विधि अपनाई जाती है।

उदाहरण 11.4.4 :

निम्नलिखित सारणी में 1982 से 1988 तक भारत में सीमेंट उत्पादन दिया गया है। समंकों में द्वि-कोटि बहुपद का समंजन कीजिए।

वर्ष :	1982	1983	1984	1985	1986	1987	1988
उत्पादन (10 लाख टन):	23.7	27.1	30.2	33.1	36.4	39.3	45.0

हल :

यहाँ पर वर्षों की संख्या विषम ($n = 7$) है।

मान लीजिए उपनति समीकरण $y = a + bx + cx^2$ है, जिसका मूल बिन्दु ($x = 0$) 1985 तथा x की इकाई एक वर्ष है। प्रसामान्य समीकरण निम्नलिखित हैं :

$$\sum y = na + b\sum x + c\sum x^2$$

$$\sum xy = a\sum x + b\sum x^2 + c\sum x^3$$

$$\sum x^2 y = a\sum x^2 + b\sum x^3 + c\sum x^4$$

सारणी 11.4.4 : द्वि-कोटि बहुपद का समंजन

वर्ष	y	x	x^2	x^3	x^4	xy	x^2y
1982	23.7	– 3	9	– 27	81	– 71.1	213.3
1983	27.1	– 2	4	– 8	16	– 54.2	108.4
1984	30.2	– 1	1	– 1	1	– 30.2	30.2
1985	33.1	0	0	0	0	0.0	0.0
1986	36.4	1	1	1	1	36.4	36.4
1987	39.3	2	4	8	16	78.6	157.2
1988	45.0	3	9	27	81	135.0	405.0
योग	234.8	0	28	0	196	94.5	950.5

इस सारणी से प्राप्त मानों को प्रसामान्य समीकरणों में प्रतिस्थापित करने पर

$$7a + 28c = 234.8$$
$$28b = 94.5$$
$$28a + 196c = 950$$

इन समीकरणों के हल द्वारा

$$a = 33$$
$$b = 3.37$$
$$c = 0.134$$

अतः द्वि-कोटि बहुपद $Y = 33 + 3.37x + 0.134x^2$ होगा, जिसका मूल बिन्दु $(x = 0)$ 1985 है तथा x की इकाई एक वर्ष है।

उदाहरण 11.4.5 :

निम्नलिखित समंकों के लिए द्वि-कोटि बहुपद का समंजन कीजिए। वर्ष 1982 के लिए उपनति मान आकलित कीजिए।

वर्ष	1976	1977	1978	1979	1980	1981
भारत का वार्षिक आयात (10^8 रुपये)	507	602	681	914	1255	1361

हल :

यहाँ पर वर्षों की संख्या सम $(n = 6)$ है।

मान लीजिए $y = a + bx + cx^2$ उपनति समीकरण है, जिसका मूल बिन्दु $(x = 0)$ 1978 तथा 1979 का मध्य बिन्दु है तथा x की इकाई 6 मास हैं। प्रसामान्य समीकरण निम्नलिखित हैं :

$$\sum y = na + b\sum x + c\sum x^2$$

$$\sum xy = a\sum x + b\sum x^2 + c\sum x^3$$

$$\sum x^2y = a\sum x^2 + b\sum x^3 + c\sum x^4$$

सारणी 11.4.5 : द्वि-कोटि बहुपद का समंजन

वर्ष	y	x	x^2	x^3	x^4	xy	x^2y
1976	507	– 5	– 25	– 125	625	– 2535	12675
1977	602	– 3	9	– 27	81	– 1806	5418
1978	681	– 1	1	– 1	1	– 681	681
1979	914	1	1	1	1	914	914
1980	1255	3	9	27	81	3765	11295
1981	1361	5	25	125	625	6805	34025
योग	5320	0	70	0	1414	6462	65008

इस सारणी से प्राप्त मानों को प्रसामान्य समीकरणों में प्रतिस्थापित करने पर

$$6a + 70c = 5320$$
$$70b = 6462$$
$$70a + 1414c = 65008$$

इस समीकरणों को हल करने पर

$a = 829.2$, $b = 92.31$ तथा $c = 4.924$

अत: द्वि-कोटि बहुपद $Y = 829.2 + 92.31x + 4.924x^2$ होगा,
जिसका मूल बिन्दु ($x = 0$), 1978 तथा 1979 का मध्य बिन्दु तथा x की इकाई 6 मास है।

वर्ष 1982 के लिए x मान 7 होगा।
इसलिए, 1982 का आकलित मान

$$Y = 829.2 + 92.31 \times 7 + 4.924 \times (7)^2$$
$$= 829.2 + 646.17 + 241.28 = 1716.65$$

11.5 वार्षिक समंकों से मासिक या त्रैमासिक उपनति

कालश्रेणी में, वार्षिक समंक विभिन्न रूपों में उपलब्ध हो सकते हैं, जैसे (1) प्रत्येक वर्ष के लिए मासिक या त्रैमासिक औसत; तथा (2) वार्षिक योग।

यदि उपनति का समंजन मासिक या त्रैमासिक समंकों के लिए किया गया है तो मासिक या त्रैमासिक मान प्राप्त करने में कोई कठिनाई नहीं होती। लेकिन, न्यूनतम वर्ग विधि द्वारा मासिक या त्रैमासिक समंकों के लिए उपनति रेखा के समंजन की सलाह नहीं दी जाती। प्राय: उपनति रेखा का समंजन वार्षिक समंकों के लिए किया जाता है तथा इसके उपयोग द्वारा, बाद में विभिन्न मासों या त्रैमासों के लिए उपनति मान प्राप्त किए जाते हैं। मासिक या त्रैमासिक उपनति समीकरण प्राप्त करने की विधि का विवेचन निम्नलिखित हैं।

मान लीजिए वार्षिक उपनति समीकरण $y = a + bx$ है। यदि हम समीकरण के दोनों पक्षों को 12 से भाग करें, तो हमें मासिक औसत समीकरण प्राप्त होगा।

अत: $\frac{y}{12}=\frac{a}{12}+\frac{b}{12}x$ एक मासिक औसत समीकरण है। $y=\frac{y}{12}$, $A=\frac{a}{12}$ तथा $B=\frac{b}{12}$ रखने पर, मासिक औसत समीकरण $Y=A+Bx$ होगा, जहाँ पर $Y=$ मासिक औसत, $B=$ मासिक औसत में परिवर्तन दर (अर्थात् प्रति इकाई x में या एक वर्ष में परिवर्तन होने पर)।

मासिक समीकरण प्राप्त करने के लिए हमें इसके अनुरूप Y में परिवर्तन दर ज्ञात करना पड़ेगा। क्योंकि प्रतिवर्ष मासिक औसत में परिवर्तन B है, इसलिए $\frac{B}{12}$ प्रतिमास औसत परिवर्तन को व्यक्त करेगा। अत: मासिक समीकरण

$Y=A+\frac{B}{12}x$ या $Y=\frac{a}{12}+\frac{b}{144}x$ होगा, जिसमें x मास, न कि वर्ष, को सूचित करता है।

इसी प्रकार

$Y=\frac{a}{4}+\frac{b}{4}x$ एक त्रैमासिक समीकरण होगी, जिसमें x की इकाई एक वर्ष है, तथा $Y=\frac{a}{12}+\frac{b}{16}x$ एक त्रैमासिक समीकरण होगी, जिसमें x की इकाई एक त्रैमास है।

मूल बिन्दु का स्थानांतरण

समीकरण $y=a+bx$ में a का अर्थ, मूल बिन्दु के वर्ष में उपनति का मान होता है। अत: मूल बिन्दु के स्थानांतरण के साथ a के मान में परिवर्तन हो जाता है। मान लिया मूल बिन्दु $(x=0)$ वर्ष 1995 है तथा हम इसको 1998 में स्थानांतरित करना चाहते है। यह ध्यान दें कि 1998 के लिए x का मान 3 है, अत: 1998 की उपनति $=a+3b$ होगी। इस उपनति को स्थिरांक लेने पर उपनति समीकरण

$y=(a+3b)+bx$, एक नया उपनति समीकरण होगा, जिसका मूल बिन्दु 1998 होगा।

उदाहरण 11.5.1 :

कुछ उत्पादन समंकों के लिए उपनति समीकरण

$y=150+24x$ ($y=$ वार्षिक उत्पादन, हजार टनों में तथा $x=$ समय, जिसका मूल बिन्दु 1978 है, x की इकाई एक वर्ष) है।

मई 1983 के लिए उपनति मान का आकलन कीजिए।

हल :

मासिक उपनति समीकरण

$$Y=\frac{150}{12}+\frac{24}{12}x=12.5+12.5+0.167x$$

होगी, जिसमें $y=$ मासिक उत्पादन, x की इकाई एक मास तथा मूल बिन्दु 1978, अर्थात् 30 जून 1978 पर है। मई 1983 की उपनति के आकलन के लिए, हम इस समीकरण में $x=58.5$ प्रतिस्थापित कर देते हैं। इस प्रकार

$Y=12.5+0.167\times 58.5=22.25$ ('000 टन) प्राप्त होगा।

उदाहरण 11.5.2 :

7 वर्षों की त्रैमासिक औसत बिक्री के लिए समंजित उपनति समीकरण $y = 250 + 20x$ (x की इकाई = 1 वर्ष, मूल बिन्दु = 30 जून 1970) है। वर्ष 1973 के प्रथम त्रैमास (जनवरी-मार्च) के लिए उपनति मान आकलित कीजिए।

हल :

यहाँ पर त्रैमासिक औसत का अर्थ, प्रत्येक वर्ष के लिए प्रति त्रैमास औसत से है।

त्रैमासिक उपनति समीकरण $y = 250 + \frac{20}{40}x$ (जिसमें Y = त्रैमासिक बिक्री, x = एक त्रैमास तथा मूल बिन्दु = 30 जून, 1970) होगा।

30 जून, 1970 तथा 1973 के पहले त्रैमास (1 जनवरी, 1973 से 31 मार्च, 1973 का माध्य बिन्दु अर्थात 15 फरवरी, 1973) तक अंतराल 2 वर्ष 7.5 मास या 10.5 त्रैमासों का है। उपनति समीकरण में $x = 10.5$ प्रतिस्थापित करने पर $y = 250 + 5 \times 10.5 = 302.5$ प्राप्त होगा, जोकि 1973 के पहले त्रैमास का आकलित मान है।

बोध प्रश्न 2

1) निम्नलिखित समंकों में एक सरल रेखा का समंजन कीजिए तथा दर्शाइये कि इससे मासिक उपनति मान किस प्रकार प्राप्त किए जा सकते हैं। इस प्रकार के दो मान प्राप्त कीजिए।

वर्ष :	1970	1971	1972	1973	1974
औसत मासिक उत्पादन : ('000 टनों में)	38	40	41	45	47

....

....

....

....

....

....

....

....

....

2) किन्हीं उत्पादन समंकों के लिए उपनति समीकरण $y = 240 + 48x$ (y = वार्षिक उत्पादन, टनों में, x = समय, जिसका मूल-बिन्दु 1985 है, x की इकाई = 1 वर्ष) है।

अक्टूबर, 1991 के लिए उपनति मान आकलित कीजिए।

....

....

....

....

....

....

....

....

....

....

3) त्रैमासिक औसत बिक्री समंकों में समंजित उपनति समीकरण $y = 60 + 8x$ (x की इकाई = 1 वर्ष, मूल बिन्दु = 30 जून, 1988) है। वर्ष 1990 के प्रथम त्रैमास (जनवरी-मार्च) के लिए उपनति मान आकलित कीजिए।

....

....

....

....

....

....

....

....

....

....

11.6 मौसमी विचरणों का माप

काल श्रेणी में मौसमी विचरणों को मापने की बहुत सी विधियाँ हैं, जो इस बात पर निर्भर करती हैं कि अन्य घटक जैसे चक्रीय, उपनति तथा अनियमित विचरण किस प्रकार इसमें विद्यमान हैं। सरलता के लिए, हम केवल मासिक तथा त्रैमासिक समंकों पर ही विचार करेंगे, लेकिन साप्ताहिक तथा दैनिक समंकों की विधि भी समरूप होती है। इन विधियों के उपयोग द्वारा हम त्रैमासिक (या मासिक) समंकों द्वारा 4 (या 12) संख्याओं का परिकलन करते हैं, जिनको मौसमी सूचकांक कहते हैं। इन सूचकांकों को प्राय: प्रतिशत रूप में व्यक्त किया जाता है। एक विशेष त्रैमास (या मास) की संख्या इस बात का सूचक है कि इस त्रैमास (या मास) का मान, एक सामान्य त्रैमास (या मास) के मान से कम है या अधिक। उदाहरण के लिए, यदि एक त्रौमास का मान 80 है तो यह इस बात का सूचक है कि व्यवसाय का निर्यात या बिक्री (मान लीजिये), सामान्य त्रैमास की तुलना में, 20% कम है। मौसमी विचरणों के माप के लिए हम केवल गुणात्मक प्रतिरूप पर ही विचार करेंगे। मौसमी विचरणों के माप की मुख्य विधियाँ निम्नलिखित हैं।

1) सरल माध्य विधि

2) उपनति से अनुपात विधि

3) चल-माध्य से अनुपात विधि

11.6.1 सरल माध्य विधि

यह विधि इस मान्यता पर आधारित है कि काल श्रेणी चर y, दो घटकों, अर्थात् मौसमी (S) तथा अनियमित विचरणों के प्रभावों का परिणाम है। अत: हम $y = S.I$ लिख सकते हैं।

जब हम, सभी वर्षों के लिए, प्रत्येक मास या त्रैमास का माध्य लेते हैं तो अनियमित विचरणों का विलोपन हो जाता है तथा हमारे पास मौसमी घटक शेष रह जाता है। इस विधि की व्याख्या सारणी 11.6 में की गई है।

सारणी 11.6 : सरल माध्य विधि की व्याख्या

वर्ष	त्रैमास			
	I	II	III	IV
1	y_1	y_2	y_3	y_4
2	y_5	y_6	y_7	y_8
3	y_9	y_{10}	y_{11}	y_{12}
4	y_{13}	y_{14}	y_{15}	y_{16}
5	y_{17}	y_{18}	y_{19}	y_{20}
योग	T_1	T_2	T_3	T_4
माध्य	A_1	A_2	A_3	A_4
मौसमी सूचकांक	s_1	s_2	s_3	s_4
मौसमी सूचकांक (समायोजित)	S_1	S_2	S_3	S_4

व्याख्यात्मक विवरण :

i) प्रत्येक वर्ष के प्रथम त्रैमास मानों का योग $T_1 = y_1 + y_5 + y_9 + y_{13} + y_{17}$ है। इसी प्रकार T_2, T_3 तथा T_4, प्रत्येक वर्ष के क्रमशः द्वितीय, तृतीय तथा चतुर्थ त्रैमासों के योग हैं।

ii) i वें त्रैमास का माध्य $A = \dfrac{T_i}{n}$ है जहाँ पर $i = 1, 2, 3, 4$ तथा n वर्षों की संख्या का सूचक है।

iii) $G = \dfrac{\sum A_i}{4}$ को सर्वमाध्य (grand average) परिभाषित किया गया है।

iv) $s_i = \dfrac{A_i}{G} \times 100,\ i = 1,\ 2,\ 3,\ 4$

v) $s = s_1 + s_2 + s_3 + s_4$

vi) S_1, S_2, S_3 तथा S_4 जहाँ पर $S_i = \dfrac{s_i}{s} \times 400,\ i = 1,\ 2,\ 3,\ 4$ है, क्रमशः प्रथम, द्वितीय, तृतीय तथा चतुर्थ त्रैमासों के मौसमी सूचकांक हैं। यह ध्यान दें कि चारों सूचकांकों का योग 400 के बराबर होना अनिवार्य होता है। इसके अतिरिक्त यदि $s = 400$ है तो $S_i = s_i$ होगा।

vii) मासिक समंकों की काल श्रेणी में, 12 मासिक सूचकांकों का येाग 1200 के बराबर होना आवश्यक है।

उदाहरण 11.6.1 :

सरल माध्य विधि द्वारा निम्नलिखित समंकों के लिए मौसमी सूचकांकों का परिकलन कीजिए।

सारणी 11.6.1 : मौसमी सूचकांकों का परिकलन

वर्ष	त्रैमास			
	I	II	III	IV
1972	72	68	80	70
1973	76	70	82	74
1974	74	66	84	80
1975	76	74	84	78
1976	78	74	86	82

हल :

वर्ष	त्रैमास			
	I	II	III	IV
1972	72	68	80	70
1973	76	70	82	74
1974	74	66	84	80
1975	76	74	84	78
1976	78	74	86	82
योग (T)	376	352	416	384
माध्य (A)	75.2	70.4	83.2	76.8
मौसमी सूचकांक	98.43	92.15	108.90	100.52

विवरण :

सर्वमाध्य $G = \dfrac{A_1 + A_2 + A_3 + A_4}{4} = \dfrac{75.2+70.4+83.2+76.8}{4} = 76.4$

प्रथम त्रैमास का मौसमी सूचकांक, $S_1 = 98.43$

द्वितीय त्रैमास का मौसमी सूचकांक, $S_2 = 92.15$

तृतीय त्रैमास का मौसमी सूचकांक, $S_3 = 108.90$

चर्तुथ त्रैमास का मौसमी सूचकांक, $S_4 = 100.52$

क्योंकि इन सूचकांकों का योग 400 के बराबर है, समायोजन की आवश्यकता नहीं है।

11.6.2 उपनति से अनुपात विधि

यदि समंकों में प्रर्याप्त मात्रा मे उपनति विद्यमान है तो सर्वप्रथम हम एक उपयुक्त उपनति समीकरण ज्ञात करके विभिन्न त्रैमासों (या मासों) के उपनति मान ज्ञात करते हैं। प्राय: ये मान त्रैमासिक या मासिक उपनति समीकरण द्वारा प्राप्त किए जाते हैं। तत्पश्चात् मूल y के मानों को उनके अनुरूप उपनति मानों के प्रतिशत के रूप में व्यक्त करके, उपनति का विलोपन किया जाता है। यह विधि इस मान्यता पर आधारित है कि चक्रीय विचरण या तो बहुत कम है या बिल्कुल विद्यमान नहीं है।

संकेतों के उपयोग द्वारा, हम

$$\frac{y}{T}\times 100 = \frac{TSI}{T}\times 100 = SI\times 100$$

लिख सकते हैं।

इसके पश्चात् अनियमित विचरणों का विलोपन सरल माध्य विधि द्वारा किया जाता है।

उदाहरण 11.6.2 :

निम्नलिखित सारणी में एक बड़े विभागीय भंडार की पाँच विभिन्न वर्षों की बिक्री ('000 रुपये में) दी गई है। उपनति से अनुपात विधि द्वारा मौसमी सूचकांक प्राप्त कीजिए।

वर्ष	त्रैमास			
	I	II	III	IV
1980	502	1632	605	362
1981	526	1700	680	390
1982	556	1820	780	422
1983	590	1955	888	464
1984	632	2110	1002	515

हल :

पहले हम त्रैमास औसत मानों के लिए उपनति रेखा का समंजन करते हैं। मान लीजिए, समंजन की जाने वाली उपनति $y = a + bx$ है, जिसमें y = वर्ष की त्रैमासिक औसत तथा x की इकाई एक वर्ष है। दिए हुए समंकों से प्रत्येक वर्ष के चार त्रैमासों का औसत लेकर, निम्नलिखित सारणी की रचना की गई है।

सारणी 11.6.2 (a) : रैखिक उपनति का समंजन

वर्ष	y	x	x^2	xy
1980	775	−2	4	−1550
1981	824	−1	1	−824
1982	894	0	0	0
1983	974	1	1	974
1984	1065	2	4	2130
	4532	0	10	730

प्रसामान्य समीकरण इस प्रकार है :

$$\sum y = na + b\sum x$$

$$\sum xy = a\sum x + b\sum x^2$$

ऊपर दी गई सारणी से, प्रसामान्य समीकरणों में मान प्रतिस्थापित करने पर

$5a = 4532$ or $a = 906.4$

$10b = 730$ or $b = 73$

इस प्रकार औसतन त्रैमासिक उपनति समीकरण $T = 906.4 + 73x$ होगा।
(मूल बिन्दु : 1982, x की इकाई = 1 वर्ष)

यह ध्यान दें कि पहले के विपरित हमने Y के स्थान पर T का उपयोग किया है।

इस समीकरण द्वारा हम निम्नलिखित त्रैमासिक उपनति समीकरण प्राप्त करते हैं।

$$T = 906.4 + \frac{73}{4}.x = 906.4 + 18.25x$$

(मूल बिन्दु : 30 जून 1982, x की इकाई = एक त्रैमास)

मूल बिन्दु का स्थानान्तरण 1982 के तृतीय त्रैमास (इसका मध्य बिन्दु) पर करने पर, त्रैमासिक उपनति समीकरण

$$T = 906.4 + 18.25\,(x + 0.5) = 906.4 + 9.125 + 18.25x$$
$$= 915.125 + 18.25x$$ प्राप्त होता है।

इस समीकरण में x के उपयुक्त मान रखने पर हमें विभिन्न त्रैमासों के उपनति मान (T) प्राप्त हो जाते है। अगले चरण में हम मूल y मानों को उपनति के प्रतिशत के रूप में लिखते हैं अर्थात $(y \div T) \times 100$, जिससे हमें उपनति से अनुपात प्राप्त होते हैं। उपनति मान तथा उपनति से अनुपातों को निम्नलिखित सारणी में दर्शाया गया है।

सारणी 11.6.2 (b) : उपनति से अनुपातों का परिकलन

वर्ष	त्रैमास	x	$T = 915.525 + 18.25x$	y	$(y \div T) \times 100$
1980	I	– 10	733.0	502	68
	II	– 9	751.3	1632	217
	III	– 8	769.5	605	79
	IV	– 7	787.8	362	46
1981	I	– 6	806.0	526	65
	II	– 5	824.3	1700	206
	III	– 4	842.5	680	81
	IV	– 3	860.8	390	45
1982	I	– 2	879.0	556	63
	II	– 1	897.3	1820	203
	III	0	915.5	780	85
	IV	1	933.8	422	45
1983	I	2	952.0	590	62
	II	3	970.3	1955	201
	III	4	988.5	888	90
	IV	5	1006.8	464	46
1984	I	6	1025.0	632	62
	II	7	1043.3	2110	202
	III	8	1061.5	1002	94
	IV	9	1079.8	515	48

इन उपनति से अनुपातों को त्रैमासों के अनुसार व्यवस्थित करके, सरल माध्य विधि द्वारा, मौसमी सूचकांक परिकलित किए जाते हैं।

सारणी 11.6.2 (c) : मौसमी सूचकांकों का परिकलन

वर्ष	त्रैमास			
	I	II	III	IV
1980	68	217	79	46
1981	65	206	81	45
1982	63	203	85	45
1983	62	201	90	46
1984	62	202	94	48
योग	320	1029	429	230
माध्य	64.0	205.8	85.8	46.0
मौसमी सूचकांक	63.74	209.98	85.46	45.82

11.6.3 चल-माध्य से अनुपात विधि

जब काल श्रेणी समंकों में सभी चार घटक विद्यमान हों तो इस विधि का उपयोग किया जाता है। हम यह जानते हैं कि एक चल-माध्य, जिसकी अवधि आवर्ती विचरणों की अवधि के बराबर है, इन विचरणों का पूर्ण विलोपन कर देता है। अत:, यदि हम त्रैमासिक समंको का 4-अवधि (या मासिक समंकों का 12-अवधि) चल-माध्य (M) परिकलित करें, तो y मानों से मौसमी विचरणों और कुछ अनियमित विचरण का विलोपन हो जाता है। तत्पश्चात् y को चल-माध्य के प्रतिशत के रूप में लिखने पर, अर्थात् $(y \div M) \times 100$, हमे जो मान प्राप्त होते हैं उनमें मौसमी तथा अनियमित घटक विद्यमान होते हैं। इसके बाद, पहले की भाँति, सरल माध्य विधि द्वारा, मौसमी घटक को अनियमित घटक से अलग किया जाता हैं।

संकेतों के उपयोग द्वारा हम $\frac{y}{M} \times 100 = \frac{TCSI}{TCI'} = SI$" लिख सकते हैं।

उदाहरण 11.6.3 :

निम्नलिखित समंकों के लिए, चल-माध्य से अनुपात विधि द्वारा, मौसमी सूचकांको का परिकलन कीजिए।

वर्ष	गर्मी	मॉनसून	पतझड़	सर्दी
1989	30	81	62	119
1990	33	104	86	171
1991	42	153	99	221
1992	56	172	129	235
1993	67	201	136	302

हल :

सारणी 11.6.3 : चल-माध्य अनुपात विधि द्वारा मौसमी सूचकांकों का परिकलन

वर्ष	त्रैमास	y	4-अवधि चल योग	केन्द्रित योग	4-अवधि चल माध्य (M)	$(y \div M)\times100$
1989	गर्मी	30	—	—	—	—
	मॉनसून	81	—	—	—	—
			292			
	पतझड़	62		587	73.38	84.50
			295			
	सर्दी	119		613	76.63	155.30
			318			
1990	गर्मी	33		660	82.50	40.00
			342			
	मॉनसून	104		736	92.00	113.04
			394			
	पतझड़	86		797	99.63	86.32
			403			
	सर्दी	171		855	106.88	160.00
			452			
1991	गर्मी	42		917	114.63	36.64
			465			
	मॉनसून	153		980	122.50	124.90
			515			
	पतझड़	99		1044	130.50	75.86
			529			
	सर्दी	221		1077	134.63	164.16
			548			
1992	गर्मी	56		1126	140.75	39.79
			578			
	मॉनसून	172		1170	146.25	117.61
			592			
	पतझड़	129		1195	149.38	86.36
			603			
	सर्दी	235		1235	154.38	152.23
			632			
1993	गर्मी	67		1271	158.88	42.17
			639			
	मॉनसून	201		1345	168.13	119.55
			706			
	पतझड	136	—	—	—	—
	सर्दी	302	—	—	—	—

अब चल-माध्यों से अनुपातों को त्रैमासों के अनुसार व्यवस्थित करके, सरल माध्य विधि द्वारा, मौसमी सूचकांक परिकलित किए जा सकते हैं।

वर्ष	त्रैमास			
	गर्मी	मॉनसून	पतझड़	सर्दी
1989	—	—	84.50	155.30
1990	40.00	113.04	86.32	160.00
1991	36.64	124.90	75.86	164.16
1992	39.79	117.61	86.36	152.23
1993	42.17	119.55	—	—
योग	158.60	475.10	333.04	631.69
माध्य	39.65	118.78	83.26	157.92
मौसमी सूचकांक	36.69	118.89	83.34	158.08

बोध प्रश्न 3

1) निम्नलिखित समंकों में 1972 से 1975 के वर्षों में स्टील के तैयार डिब्बों की संख्या दी हुई है :

स्टील के तैयार डिब्बों का उत्पादन ('000 डिब्बे)

वर्ष	जनवरी	फरवरी	मार्च	अप्रैल	मई	जून	जुलाई	अगस्त	सितंबर	अक्तूबर	नवंबर	दिसंबर
1972	420	414	502	365	368	332	390	396	429	417	422	496
1973	491	466	516	337	342	360	409	402	372	391	394	446
1974	463	465	478	310	325	406	415	437	438	445	430	416
1975	502	487	536	404	418	429	489	492	475	456	476	476

सरल माध्य विधि द्वारा मौसमी सूचकांकों का परिकलन कीजिए।

....

....

....

....

....

....

....

....

....

....

....

....

2) निम्नलिखित सारणी में भारत में, 1972 से 1975 तक विभिन्न त्रैमासों में स्टील उत्पादन ('000 टनों में) दिया हुआ है :

वर्ष	प्रथम त्रैमास	द्वितीय त्रैमास	तृतीय त्रैमास	चतुर्थ त्रैमास
1972	1336	1065	1215	1335
1973	1463	1039	1183	1161
1974	1306	1041	1290	1321
1975	1525	1251	1456	1408

रैखिक उपनति की मान्यता पर, उपनति से अनुपात विधि द्वारा मौसमी सूचकांकों का परिकलन कीजिए।

....

....

....

....

....

....

....

....

....

....

3) 1986 से 1989 वर्षों के लिए त्रैमासिक बिक्री समंक (हजार रुपयों में) निम्नलिखित हैं। चल-माध्य से अनुपात विधि द्वारा मौसमी सूचकांक ज्ञात कीजिए।

वर्ष	त्रैमास			
	I	II	III	IV
1986	290	280	285	310
1987	320	305	310	330
1988	340	321	320	340
1989	370	360	362	380

....

....

....

....

....

....

....

....

....

4) किसी दुकान के एक विशेष प्रकार के वस्त्रों की बिक्री के मौसमी सूचकांक निम्नलिखित हैं।

त्रैमास	मौसमी सूचकांक
जनवरी-मार्च	97
अप्रैल-जून	85
जुलाई-सितम्बर	83
अक्तूबर-दिसम्बर	135

यदि प्रथम त्रैमास में कुल बिक्री का मूल्य 15,000 रुपये हो तो दुकानदार को कितने मूल्य के इस प्रकार के वस्त्र भंडार में रखने चाहिए, जिससे वर्ष के अन्य त्रैमासों की माँग पूरी की जा सके?

....

....

....

....

....

....

....

....

....

....

11.7 सारांश

कालश्रेणी एक चर के प्रेक्षणों का एक समुच्चय होता है जोकि एक समयावधि, प्राय: समान समय अंतराल, में लिखित किए जाते हैं। सम्बन्धित चर में परिवर्तनों को, कुछ सीमा तक, कालश्रेणी के घटकों के आधार पर समझा जा सकता है। ये घटक उपनति, मौसमी विचरण, चक्रीय उच्चावचन तथा यादृच्छिक विचरण होते हैं। चर का प्रेक्षित मान या तो उपरोक्त घटकों की गुणा के रूप में (गुणात्मक प्रतिरूप) या घटकों के योग के रूप में (योज्य प्रतिरूप) निरूपित किया जाता है।

चल-माध्य विधि, या न्यूनतम वर्ग विधि द्वारा समीकरण का समंजन करके उपनति का मापन किया जा सकता है। उपनति का अनुमान लगाने के बाद हम भविष्य के मानों का आकलन कर सकते हैं तथा वार्षिक समंकों से मासिक या त्रैमासिक मान ज्ञात कर सकते हैं।

मौसमी विचरणों के आकलन की तीन विधियाँ हैं : सरल माध्य विधि, उपनति से अनुपात विधि तथा चल-माध्य से अनुपात विधि। सरल माध्य विधि का उपयोग अनियमित घटक के विलोपन के लिए किया जाता है। यदि चक्रीय उच्चावचन विद्यमान न हों तो उपनति से अनुपात विधि तथा सभी चार घटकों के विद्यमान होने पर चल-माध्य से अनुपात विधि का उपयोग किया जाता है।

11.8 शब्दावली

चक्रीय विचरण : कालश्रेणी का दोलनी संचलन जहाँ पर दोलन की अवधि, जिसको चक्र कहते हैं, एक वर्ष से अधिक होती है।

अनियमित विचरण	:	काल श्रेणी का यादृच्छिक संचलन जिसकी अन्य घटकों द्वारा व्याख्या नहीं होती। इस दृष्टि से यह अन्य घटकों का अवशिष्ट होता है।
न्यूनतम वर्ग विधि	:	जब कालश्रेणी में एक बहुपद का समंजन किया जाता है तो न्यूनतम वर्ग विधि के लिए यह आवश्यक है कि फलन के प्राचलों का चयन इस प्रकार किया जाए जिससे वास्तविक प्रेक्षण तथा प्रत्याशित मानों के विचलनों के वर्ग का योग न्यूनतम हो जाय।
चल-माध्य विधि	:	चल-माध्य, एक उपयुक्त अवधि लेकर, मान लीजिए 3 वर्ष, तीन लगातार वर्षों के मानों का औसत परिकलित करके एक श्रेणी के रूप में प्राप्त किए जाते हैं। प्राप्त माध्य, अवधि के मध्य मे लिखा जाता है।
मौसमी विचरण	:	ऐसा आवर्ती संचलन जिसकी अवधि एक वर्ष से अधिक नहीं होती।
दीर्घकालिक उपनति	:	यह कालश्रेणी का एक अवधि में मसृण, नियमित तथा दीर्घकालीन संचलन होता है। किसी अवधि में उपनति उपरिमुखी या बढ़ती हुई, अधोमुखी या घटती हुई, या लगभग स्थिर रह सकती है।

11.9 कुछ उपयोगी पुस्तकें

Nagar, A.L. and R.K. Das, 1989 : *Basic Statistics*, Oxford University Press, Delhi.

Goon, A.M., M.K. Gupta and B. Dasgupta, 1987 : *Basic Statistics*, The World Press Pvt. Ltd., Calcutta.

मेहता , बी.सी. , 1986, *प्रारंभिक सांख्यिकी*, राजस्थान हिन्दी ग्रन्थ अकादमी, जयपुर, अध्याय 11

11.10 बोध प्रश्नों के उत्तर या संकेत

बोध प्रश्न 1

1) उदाहरण 11.3.1 पढ़कर उत्तर दें।

2) उदाहरण 11.3.1 पढ़कर उत्तर दें।

बोध प्रश्न 2

1) $y = 42.2 + 2.3x$, मूल बिन्दु : 1972, x की इकाई = 1 वर्ष (मासिक औसत समीकरण)
$y = 42.3 + 0.19x$, मूल बिन्दु : जुलाई, 1972, x की इकाई = 1 मास (मासिक समीकरण)

मार्च 1971 के लिए आकलन $(x = -16) = 39.23$
सितंबर 1973 के लिए आकलन $(x = 14) = 44.98$

2) 45.17 टन

3) 73

बोध प्रश्न 3

1) 109.57, 107.00, 118.69, 82.71, 84.87, 89.19, 99.47, 100.87, 100.11, 99.82, 100.58, 107.12

2) 112.27, 86.62, 100.21, 100.90

3) 104.2, 97.9, 96.5, 101.4

4) 13,144 रुपये; 12, 835; 20,876.

11.11 पारिभाषिक शब्दावली

अनियमित विचरण	:	irregular variations
उपनति से अनुपात विधि	:	ratio to trend method
चल-माध्य विधि	:	moving average method
गुणात्मक प्रतिरूप	:	multiplicative model
चक्रीय उच्चावचन	:	cyclical fluctuations
दीर्घकालिक उपनति	:	long term trend
न्यूनतम वर्ग विधि	:	method of least squares
बहुपद समंजन विधि	:	method of fitting polynomial
मौसमी विचरण	:	seasonal variations
योज्य प्रतिरूप	:	additive model

इकाई 12 जन्म-मृत्यु सांख्यिकी (Vital Statistics)

इकाई की रूपरेखा

12.0 उद्देश्य

इस इकाई को पढ़ने के बाद, आप :

- जन्म-मृत्यु सांख्यिकी में प्रयुक्त आँकड़ों के स्रोतों का वर्णन कर सकेंगे;
- जन्म-मृत्यु सबंधी विविध दरों की गणना कर सकेंगे;
- वय सारणी के निर्माण की कार्यविधि की व्याख्या कर सकेंगे; और
- वय सारणी के अनुप्रयोगों और सीमाओं के समझ पाएँगे।

12.1 प्रस्तावना

जन्म-मृत्यु सांख्यिकी मुख्य रूप से ऐसे कारकों से संबंद्ध है जिनका जनसंख्या वृद्धि में विशेष योगदान है। ऐसे कुछ कारक हैं; जन्म दर, मृत्यु दर, जीवन की प्रत्याशा (जीवाशा) और प्रवास। इस इकाई को पढ़ने के बाद, आप अर्थशास्त्र में जन्म-मृत्यु सांख्यिकी के महत्त्व और अनुप्रयोग को समझ पाने की स्थिति में होंगे। इस इकाई का मुख्य उद्देश्य आपको जन्म-मृत्यु सांख्यिकी की कुछ बुनियादी संकल्पनाओं और आँकड़ों के स्रोतों से अवगत कराना है और यह बताना है कि विविध अनुपातों को किस प्रकार मापा जाता है और जनसंख्या को प्रक्षेपित करने और जीवन प्रत्याशा की गणना करने में इन अनुपातों को किस प्रकार लागू किया जाता है और बीमाविज्ञान आदि में इनके क्या प्रयोग हैं।

12.2 आँकड़ों के स्रोत

जन्म-मृत्यु सांख्यिकी के लिए आँकड़े आमतौर पर निम्नलिखित चार विधियों से इकट्ठे किए जाते हैं : पंजीकरण, जनगणना, सर्वेक्षण और प्रतिदर्श पंजीकरण पद्धति। आइए अब हम इन विधियों पर चर्चा करें;

i) **पंजीकरण विधि** : इस विधि में जन्म, मृत्यु, विवाह, प्रवास आदि के सतत् और स्थायी रिकार्ड शामिल हैं। भारत समेत बहुत से देशों में कानून के अंतर्गत जन्म और मृत्यु का पंजीकरण कराना अनिवार्य है। जन्म या मृत्यु का पंजीकरण कराने पर पंजीकरण कार्यालय प्रमाणपत्र जारी करता है। यद्यपि पंजीकरण विधि साधारण और प्रभावी है, फिर भी इस विधि से जुड़ी समस्या है कि होने वाले सभी जन्म और मृत्यु पंजीकृत नहीं होते। ऐसा इसलिए है क्योंकि ये कानून, विशेषरूप से, ग्रामीण भारत में इसे सख्ती से लागू नहीं किया जा रहा।

ii) **जनगणना** : विश्व के लगभग सभी देशों में समय-समय पर जनसंख्या की गणना की जाती है। जनगणना से हमें जन्म-मृत्यु सांख्यिकी के साथ-साथ आयु, लिंग, वैवाहिक स्थिति, शिक्षा का स्तर, व्यवसाय, धर्म आदि से जुड़ी जानकारी भी प्राप्त होती है। लेकिन यह जानकारी केवल जनगणना संबंधी वर्षों से ही संबंधित होती है (अर्थात् दस वर्षों में एक बार)। जनगणना के वर्षों को छोड़कर अन्य वर्षों के ये आँकड़ें उपलब्ध नहीं हो पाते।

iii) **सर्वेक्षण** : सर्वेक्षण ऐसे क्षेत्रों में किया जाता है जहाँ पंजीकरण विधि प्रभावी नहीं होती या सुचारू रूप से कार्य नहीं कर पाती। सर्वेक्षण के माध्यम से हम ऐसे क्षेत्रों की जन्म-मृत्यु सांख्यिकी प्राप्त करने के योग्य हो जाते हैं।

iv) **प्रतिदर्श पंजीकरण पद्धति** : यह भारत में बड़े पैमाने का जनसांख्यिकी सर्वेक्षण है ताकि राष्ट्रीय और उप-राष्ट्रीय स्तरों पर जन्म दर, मृत्यु दर और अन्य उर्वरता (fertility) और मृत्यु संख्या सूचकों के विश्वसनीय वार्षिक अनुमान प्राप्त किए जा सकें। आमतौर पर अंशकालिक गणनाकार द्वारा क्षेत्र छान-बीन के अंतर्गत जन्म और मृत्यु संख्या की निरंतर गणना की जाती है। ये व्यक्ति आमतौर पर शिक्षक होते हैं। इसके उपरांत हर छह महीने के बाद किसी सरकारी व्यक्ति द्वारा स्वतंत्र सर्वेक्षण किया जाता है। इन माध्यमों से प्राप्त आँकड़ों का मिलान किया जाता है। इसके बाद बेमेल और आंशिक रूप से मेल खाने वाली घटनाओं में पुन:सत्यापन किया जाता है और तत्पश्चात् जन्म और मृत्यु संबंधी अंतिम आँकड़ों की प्राप्ति होती है। 1964-65 में कुछ गिने-चुने राज्यों में प्रायोगिक स्तर पर भारत के महापंजीयक कार्यालय द्वारा प्रतिदर्श पंजीकरण

पद्धति (एस आर एस) की शुरुआत की गई। 1969-70 के दौरान लगभग 3700 प्रतिदर्श इकाइयों को सम्मिलित करते हुए यह पद्धति पूरी तरह से लागू हो गई। इसके बाद समय-समय पर प्रतिदर्श का फैलाव बढ़ता गया। 1991 जनगणना आँकड़ों के आधार पर हाल ही में इस सूची को नवीतनम रूप दिया गया है।

12.3 जन्म-मृत्यु सांख्यिकी के उपयोग

मानवीय गतिविधियों के विविध क्षेत्रों में जन्म-मृत्यु सांख्यिकी बेहद उपयोगी है। इसके कुछ मुख्य फायदे इस प्रकार हैं :

1) जन्म-मृत्यु सांख्यिकी हमें यह समझने में सहायता करती है कि किसी देश या देश के भीतर किसी क्षेत्र की जनसंख्या की रचना में किस प्रकार के बदलाव आ रहा है। जनसंख्या की रचना से आशय वहाँ के लोगों की आयु, लिंग, धर्म, जन्म दर, मृत्यु दर, प्रवास दर, वैवाहिक स्थिति आदि से सम्बन्धित जानकारी से होता है। देश या क्षेत्र की भावी आबादी की सरंचना का पूर्वानुमान लगाने में ये आँकड़े हमारे लिए उपयोगी सिद्ध होते हैं।

2) जनसंख्या संबंधी प्रवृतियों का आकलन और प्रक्षेपण नीति निर्धारकों और प्रशासकों को बेहतर तरीके से योजना बनाने और आर्थिक और सामाजिक विकास कार्यक्रमों का मूल्यांकन करने में सहायता करता है।

3) मृत्यु दर से जुड़ी सांख्यिकी समुदायों की स्वास्थ्य संबंधी दशाओं को बेहतर बनाने में सहायता करती है। जैसे, संचारी रोगों पर आधारित सांख्यिकी, प्रभावी क्षेत्र की साफ-सफाई संबंधी दशाओं और चिकित्सीय सुविधाओं को बेहतर बनाने में स्वास्थ्य प्राधिकारियों के लिए सहायक सिद्ध होती है।

4) बीमाविज्ञान, जिसमें जीवन बीमा भी शामिल है, जन्म-मृत्यु सांख्यिकी पर आधारित है। इस इकाई के अनुभाग 12.4 में आप इस संदर्भ में और अधिक जानकारी प्राप्त करेंगे।

12.4 जनसंख्या का मापन

किसी भी देश की कुल जनसंख्या को आमतौर पर किसी विशिष्ट समय बिंदु पर अभिव्यक्त किया जाता है। जैसे भारत में नवीनतम जनगणना 2001 में हुई और उस समय देश की कुल जनसंख्या की स्थिति 31.3.2001 को स्पष्ट की गई। जनगणना के समय की गई कुल जनसंख्या की गिनती को ही ठीक-ठीक माना जाता है। जैसाकि आप जानते हैं भारत में हर दस वर्षों में एक बार जनगणना की जाती है। निम्नलिखित विधियों के प्रयोग से हम जनगणना अवधियों के बीच के वर्षों के आँकड़ों का आकलन करते हैं।

12.4.1 रैखिक अंतर्वेशन विधि

विभिन्न आवधिक वर्षों के लिए कुल जनसंख्या का आकलन निम्नलिखित सूत्र से किया जा सकता है :

$$P_t = P_0 + \frac{n}{N}(P_1 - P_0) \qquad \dots(12.1)$$

जहाँ, P_t दिए गए आवधिक (inter-censal) वर्ष 't' में आकलित जनसंख्या है।

P_0 पिछली जनगणना की जनसंख्या है।

P_1 अनुवर्ती जनगणना की जनसंख्या है।

n दिए गए वर्ष और पिछले जनगणना वर्ष के बीच के वर्षों की संख्या है।

N दो जनगणना वर्षों के बीच के वर्षों की संख्या है।

उपर्युक्त विधि बीच के वर्षों के लिए उत्कृष्ट अनुमान प्रदान करती है। इसकी मान्यता यही है कि दोनों जनगणनाओं के बीच जनसंख्या वृद्धि की दर स्थिर रही है।

उदाहरण 12.1: भारत में 1991 की जनगणना के समय कुल आबादी 846 मिलियन थी और 2001 की जनगणना के समय यह 1027 मिलियन थी। 1996 में भारत की कुल आबादी की गणना कीजिए।

यहाँ, $P_0 = 846, P_1 = 1027, N = 10, n = 5$ है

इसलिए $P_{1996} = 846 + \frac{5}{10}(1027 - 846) = 936.5$ मिलियन

लेकिन उपर्युक्त विधि की सीमा है कि इससे हम केवल दो जनगणना वर्षों के बीच के वर्षों की ही आबादी का अनुमान लगा सकते हैं। इससे हम भावी वर्षों के लिए अनुमान नहीं लगा सकते।

12.4.2 चक्रवर्धी वृद्धि दर सूत्र का प्रयोग

आमतौर पर देखा गया है कि जनसंख्या वृद्धि गुणोत्तर श्रेणी में ही होती है। यदि आधार वर्ष जनसंख्या और जनसंख्या चक्रवर्धी वृद्धि दर का (आधार जनगणना वर्ष और अनुवर्ती जनगणना वर्ष के बीच पता हो तो आप निम्नलिखित सूत्र के प्रयोग से दिए गए वर्ष के लिए कुल आबादी का अनुमान लगा सकते हैं।

$$P_t = P_0 (1 + r)^n \qquad ...(12.2)$$

जहाँ, r चक्रवर्धी वृद्धि दर है (आधार जनगणना वर्ष और अनुवर्ती जनगणना वर्ष के बीच की दर)

n आधार वर्ष से वर्षों की संख्या है (आमतौर पर पिछला जनगणना वर्ष)

P_0 आधार वर्ष जनसंख्या है (आमतौर पर पिछला जनगणना वर्ष)

P_t वर्ष t का आकलित जनसंख्या है।

उदाहरण 12.2 : 1991 में किसी, छोटे शहर की जनसंख्या 50,500 थी। 1991 और 2001 के बीच इस शहर की जनसंख्या की संयुक्त वृद्धि दर 0.025 थी। वर्ष 2005 के लिए उस शहर की जनसंख्या का आकलन कीजिए (मान लीजिए 2001 के बाद वृद्धि दर उतनी ही होगी)।

यहाँ, हमें दिया गया है $P_0 = 50500, r = 0.025$

और $n = 14$ (क्योंकि $2005 - 1991 = 14$)

इसलिए, $P_{2005} = 50500 (1 + 0.025)^{14} = 71355$

12.4.3 स्वाभाविक वृद्धि और निबल प्रवास विधि

आप जानते ही हैं कि जनगणना से हमें कुल जनसंख्या का पता चलता है। इसी तरह से जन्म-मृत्यु

और प्रवसन संबंधी सांख्यिकी की कुल संख्या रजिस्ट्रारों से प्राप्त होती है। निम्नलिखित कारकों से किसी भी क्षेत्र की आबादी बढ़ जाती है :

i) स्वाभाविक बढ़ोत्तरी (अर्थात् जन्में शिशुओं की कुल संख्या – मृतकों की कुल संख्या)

ii) निबल प्रवास (अर्थात्, क्षेत्र में आने वाले लोगों की कुल संख्या – उस क्षेत्र को छोड़ कर जाने वाले लोगों की कुल संख्या)

निम्नलिखित सूत्र के प्रयोग से दी गई अवधि के लिए आबादी की गणना की जाती है।

$P_t = P_0 + (B - D) + (I - E)$

जहाँ, P_t आधार वर्ष से दिए गए वर्ष t पर आकलित आबादी है।

P_0 आधार वर्ष है (आमतौर पर पिछला जनगणना वर्ष)

B और D, आधार वर्ष से t वर्ष के दौरान जन्मे शिशुओं और मृतकों की क्रमशः कुल संख्या है।

I और E, आधार वर्ष से t वर्ष तक आप्रवासियों और प्रवासियों की संख्याएँ हैं।

उदाहरण 12.3 : 2001 की जनगणना में किसी छोटे शहर की आबादी 22000 थी। 2001 से 2003 तक जन्मे, मृत, आप्रवासी और प्रवासी लोगों की संख्या क्रमशः 800, 150, 2500 और 1500 है। 2003 में शहर की कुल आबादी ज्ञात कीजिए।

यहाँ, $P_0 = 22000$, B = 800, D = 150, I = 2500, E = 1500 है।

अतः $P_{2003} = 22000 + (800 - 150) + (2500 - 1500) = 23650$

बोध प्रश्न 1

निम्नलिखित सारणी, भारत की मध्य वर्ष कुल जनसंख्या और जनसंख्या की वार्षिक चक्रवर्धी वृद्धि दर की जानकारी हमें देती है।

वर्ष	आबादी (करोड़)	अवधि	चक्रवर्धी वृद्धि दर (%)
1950	36.99	1950-60	1.9
1960	44.59	1960-70	2.2
1970	55.50	1970-80	2.1
1980	68.70	1980-90	2.0
1990	84.17	1990-2000	1.8
2000	100.27	—	—

स्रोत : यू एस सेन्सस ब्यूरो : आईडीबी सॅमरि डैमोग्राफिक डेटा फॉर इंडिया

ध्यान रखें : चक्रवर्धी वृद्धि दर प्रतिशत में होती है। अपेक्षित r की प्राप्ति के लिए इसे 100 से विभाजित करें। जैसे, 1950-60 की समयावधि के लिए संयुक्त वृद्धि दर 1.9% है।

इसलिए $r = 1.9/100 = 0.019$

उपर्युक्त सारणी के आधार पर निम्नलिखित प्रश्नों के उत्तर दीजिए।

1) रैखिक अंतर्वेशन विधि का प्रयोग करते हुए निम्नलिखित वर्षों के लिए मध्य-वर्ष आबादी ज्ञात कीजिए।

वर्ष	मध्य-वर्ष आबादी
1954	
1966	
1973	
1985	
1998	

2) चक्रवर्धी वृद्धि दर विधि का प्रयोग करते हुए निम्नलिखित वर्षों के लिए मध्य-वर्ष आबादी ज्ञात कीजिए।

वर्ष	मध्य-वर्ष आबादी
1954	
1966	
1973	
1985	
1998	

3) उपर्युक्त दोनों विधियों की तुलना कीजिए और निष्कर्ष बताइए।

....

....

....

....

....

....

....

....

....

....

4) जन्म-मृत्यु सांख्यिकी में प्रयुक्त विभिन्न आँकड़ों के स्रोतों का संक्षेप में वर्णन कीजिए।

....

....

....

....

....

....

....

....

....

....

5) जन्म-मृत्यु सांख्यिकी के महत्त्वपूर्ण उपयोग क्या हैं?

....

....

....

....

....

....

....

....

....

....

12.5 जन्म-मृत्यु संबंधी दरें

आमतौर पर, जन्म-मृत्यु सांख्यिकी आधारित आँकड़ें जन्म लेने वालों की सख्या, मृतकों की संख्या आदि के रूप में उपलब्ध होते हैं। इन आँकड़ों की सार्थक उपयोगिता की प्राप्ति के लिए, सामान्यतौर पर हम इन आँकड़ों को जन्म-मृत्यु संबंधी दरों या अनुपातों में परिवर्तित करते हैं। वर्ष में प्रति 100 व्यक्तियों में जन्में या मृतकों की संख्या आमतौर पर निम्न होती है और छोटी भिन्नों के रूप में नज़र आती है। इन अनुपातों में होने वाले परिवर्तन भी पूरी तरह स्पष्ट नहीं होते। इस समस्या को दूर करने के लिए, जन्म-मृत्यु दरों को प्रति हजार व्यक्तियों के आधार पर अभिव्यक्त किया जाता है। इस अनुभाग में आप ऐसी कुछ महत्त्वपूर्ण जन्म-मृत्यु दरों का अध्ययन करेंगे।

12.5.1 अशोधित जन्म दर

अशोधित जन्म दर को, किसी विशिष्ट समुदाय या क्षेत्र में प्रति 1000 की जनसंख्या में जन्मे शिशुओं की संख्या के रूप में परिभाषित किया जाता है। अशोधित जन्म दर की गणना के लिए, हम निम्नलिखित सूत्र का प्रयोग करते हैं।

$$\text{अशोधित जन्म दर} = \frac{\text{जन्में शिशुओं की वर्ष में संख्या (किसी समुदाय या क्षेत्र में)}}{\text{मध्य-वर्ष जनसंख्या (समुदाय या क्षेत्र की)}} \times 100$$

अशोधित जन्म दर से हमें पता चलता है कि क्षेत्र या समुदाय में शिशुओं का जन्म किस दर पर हो रहा है।

उदाहरण 12.4 : 1995 में मध्य प्रदेश में जनजातीय समुदाय में मध्यवर्ष जनसंख्या और जन्मे शिशुओं की संख्या क्रमशः 40,000 और 1200 है। अशोधित जन्म दर ज्ञात कीजिए।

यहाँ, हमारे पास हैं; 1995 की मध्य वर्ष जनसंख्या = 40,000 और 1995 में जन्मे शिशुओं की

संख्या = 1200

$$\text{अशोधित जन्म दर} = \frac{1200}{40000} \times 1000$$

$$= 30 \text{ प्रति } 1000 \text{ व्यक्ति प्रति वर्ष}$$

12.5.2 अशोधित मृत्यु दर

अशोधित मृत्यु दर को किसी विशिष्ट आयु समूह या स्त्री/पुरुष समूह या किसी समुदाय या क्षेत्र में प्रति 1000 आबादी में होने वाली मौतों की संख्या के रूप में परिभाषित किया जाता है। अशोधित मृत्यु दर की गणना के लिए हम निम्नलिखित सूत्र का प्रयोग करते हैं।

$$\text{अशोधित मृत्यु दर} = \frac{\text{मौतों की वार्षिक संख्या (किसी विशिष्ट आयु समूह या स्त्री/पुरुष समूह या समुदाय/क्षेत्र में)}}{\text{मध्य-वर्ष जनसंख्या (उस विशिष्ट आयु समूह या स्त्री/पुरुष समूह या समुदाय/क्षेत्र की)}} \times 1000$$

अशोधित मृत्यु दर से हमें पता चलता है कि किसी आयु समूह, स्त्री/पुरुष समूह या क्षेत्र या समुदाय में किस दर पर मौतें हो रही हैं।

उदाहरण 12.5 : महाराष्ट्र के एक छोटे शहर में 2001 में महिलाओं की पंजीकृत मध्य-वर्ष आबादी और मौतों की संख्या क्रमश: 25000 और 245 है। अशोधित मृत्यु दर बताइए।

यहाँ, 2001 में मध्य वर्ष महिला आबादी है = 25000 और 2001 में महिलाओं की होने वाली मौतों की संख्या है = 245

$$\text{अशोधित मृत्यु दर (महिलाओं की)} = \frac{245}{25000} \times 1000$$

$$= \text{महिलाओं में } 9.8 \text{ प्रति } 1000 \text{ व्यक्ति प्रति वर्ष}$$

12.5.3 स्वाभाविक वृद्धि की अशोधित दर

स्वाभाविक वृद्धि की अशोधित दर को दिए गए वर्ष में ऐसी दर पर परिभाषित किया जाता है जिसमें प्रति 1000 व्यक्तियों के रूप में अभिव्यक्त होने वाली मौतों की तुलना में होने वाले अतिरिक्त जन्मों के कारण, दिए गए वर्ष में आबादी में वृद्धि होती है।

वार्षिक स्वाभाविक वृद्धि को इस प्रकार मापा जाता है :

$$\text{वार्षिक स्वाभाविक वृद्धि} = \text{जन्मों की वार्षिक संख्या} - \text{मौतों की वार्षिक संख्या}$$

स्वाभाविक वृद्धि की अशोधित दर की गणना का सूत्र है :

$$\text{स्वाभाविक वृद्धि की वार्षिक दर} = \frac{\text{स्वाभाविक वार्षिक वृद्धि}}{\text{मध्य वर्ष आबादी}} \times 1000$$

$$= \text{अशोधित जन्म दर} - \text{अशोधित मृत्यु दर}$$

दिए गए वर्ष के लिए स्वाभाविक वृद्धि की अशोधित दर से हमें पता चलता है कि पिछले एक वर्ष से जनसंख्या में स्वाभाविक वृद्धि कितनी हुई है।

उदाहरण 12.6 : भारत में 1997 में अशोधित जन्म दर और अशोधित मृत्यु दर क्रमश: 27.2 और 8.9 है। स्वाभाविक वृद्धि की अशोधित दर बताइए।

स्वाभाविक वृद्धि की अशोधित दर = 27.2 – 8.9 = 18.3 प्रति 1000 प्रति वर्ष

12.5.4 निवल प्रवास की दर

प्रवास को क्षेत्र की किसी विशिष्ट सीमा से पार किसी नये या अर्ध स्थायी रिहायश स्थापित करने के उद्देश्य से जाने वाले व्यक्तियों के रूप में परिभाषित किया जाता है। आप्रवासी वे व्यक्ति हैं जो क्षेत्र में रहने के उद्देश्य से आए हैं और प्रवासी वे व्यक्ति हैं जो क्षेत्र छोड़कर चले गए हैं।

वार्षिक निवल प्रवास को इस प्रकार परिभाषित किया जाता है :

आप्रवासियों की वार्षिक संख्या – प्रवासियों की वार्षिक संख्या

निवल प्रवास की वार्षिक दर की गणना का सूत्र है :

$$\text{निवल प्रवास की वार्षिक दर} = \frac{\text{वार्षिक निवल प्रवास}}{\text{वार्षिक मध्य वर्ष जनसंख्या}} \times 1000$$

निवल प्रवास की वार्षिक दर से हमें पता चलता है कि निर्धारित वर्ष की जनसंख्या में किस दर पर निवल प्रवास ने अपना योगदान दिया है।

उदाहरण 12.7 : किसी क्षेत्र में 2002 के लिए आप्रवासियों, प्रवासियों और मध्य वर्ष आबादी की वार्षिक संख्या क्रमश: 6500, 5200 और 66,700 दी गई है। निवल आप्रवास की वार्षिक दर बताइए।

यहाँ, हमारे पास है, आप्रवासियों की संख्या = 6500
प्रवासियों की संख्या = 5200
मध्य वर्ष जनसंख्या = 66700
वार्षिक निवल आप्रवास = 6500 - 5200 = 1300

$$\text{निवल प्रवास की वार्षिक दर} = \frac{1300}{66700} \times 1000$$

= 19.7 प्रति 1000 प्रति वर्ष

12.5.5 कुल वृद्धि की दर

जनसंख्या की कुल वृद्धि को इस प्रकार मापा जाता है :

वार्षिक स्वाभाविक वृद्धि + वार्षिक निवल आप्रवास

$$\text{कुल वृद्धि की दर} = \frac{\text{वार्षिक कुल वृद्धि}}{\text{वार्षिक मध्य वर्ष जनसंख्या}} \times 1000$$

= स्वाभाविक वृद्धि की अशोधित दर + निवल आप्रवास की दर

दिए गए वर्ष के लिए कुल वृद्धि की दर से हमें पता चलता है कि वर्ष में किस दर से जनसंख्या में वृद्धि हुई है।

उदाहरण 12.8 : किसी क्षेत्र के लिए 1998 में रिकार्ड की गई वार्षिक स्वाभाविक वृद्धि, वार्षिक निवल आप्रवास और वार्षिक मध्य-वर्ष जनसंख्या क्रमशः 1500, 500 और 50,000 है। कुल वृद्धि की दर बताइए।

यहाँ, हमारे पास है :

वार्षिक स्वाभाविक वृद्धि = 1500

वार्षिक निवल आप्रवास = 500

मध्य वर्ष जनसंख्या = 50000

वार्षिक कुल वृद्धि = 1500 + 500 = 2000

$$\text{कुल वृद्धि की दर} = \frac{2000}{50000} \times 1000$$

= 40 प्रति 1000 प्रति वर्ष

12.5.6 शिशु मृत्यु दर

शिशु मृत्यु दर को दिए गए वर्ष में प्रति 1000 जीवित जन्मे (एक वर्ष की आयु से कम) शिशुओं की मौतों की संख्या के रूप में परिभाषित किया जाता है। शिशु मृत्यु दर की गणना का सूत्र इस प्रकार है :

$$\text{शिशु मृत्यु दर} = \frac{\text{वार्षिक शिशु मृत्यु (बाल या बालिका या कुल)}}{\text{वार्षिक जीवित जन्म (वाले बाल या बालिका या कुल)}} \times 1000$$

शिशु मृत्यु दर से हमें पता चलता है कि दिए गए वर्ष के लिए एक वर्ष की आयु से कम वाले शिशुओं में से ऐसे कितने हैं जिनकी जीवित रहने की संभावना नहीं है। बाल या बालिकाओं के लिए अलग से शिशु मृत्यु दर की गणना की जा सकती है।

उदाहरण 12.9 : 1997 में, किसी छोटे शहर में, बालिकाओं के संदर्भ में जीवित जन्मी और इनमें होने वाली मौतें की कुल संख्या क्रमशः 3000 और 25 रिकार्ड की गई है। बालिकाओं में शिशु मृत्यु दर ज्ञात करें।

यहाँ हमारे पास हैं :

वार्षिक जीवित जन्मी बालिकाएँ = 3000

वार्षिक शिशु मृत्यु संख्या = 25

$$\text{शिशु मृत्यु दर} = \frac{25}{3000} \times 1000$$

= 8.33 प्रति 1000 प्रति वर्ष

वर्ष 1997 में भारत में अशोधित जन्म दर, अशोधित मृत्यु दर, स्वाभाविक वृद्धि दर और शिशु मृत्युदर के अनुमान इस प्रकार हैं :

जन्म-मृत्यु दर	कुल	ग्रामीण	शहरी
जन्म दर	27.2	28.9	21.5
मृत्यु दर	8.9	9.6	6.5
स्वाभाविक वृद्धि दर	18.3	19.2	15.0
शिशु मृत्यु दर	71	77	45

स्रोत : **प्रतिदर्श पंजीकरण पद्धति बुलेटिन**, अक्तूबर 1998

ध्यान दें कि शहरी क्षेत्रों की तुलना में ग्रामीण क्षेत्रों में जन्म-मृत्यु की दरें उच्च है। निम्नलिखित में से प्रत्येक के लिए कोई एक सर्वाधिक महत्त्वपूर्ण कारण बताइए :

1) ग्रामीण क्षेत्रों में जन्म दर उच्च होती है क्योंकि

....

....

....

....

....

....

....

....

....

....

2) शहरी क्षेत्रों में मृत्यु दर निम्न है क्योंकि

....

....

....

....

....

....

....

....

....

....

3) ग्रामीण क्षेत्रों में शिशु मृत्यु दर उच्च है क्योंकि

....

....

....

....

....

....

....

....

....

....

12.6 वय सारणी

जीवाशा वर्तमान मृत्युदर आदि के बने रहने पर किसी व्यक्ति के जीवन के शेष वर्षों का अनुमान होती है। वय सारणी उस समय की आयु-विशिष्ट मृत्यु दरों के अनुसार जनसंख्या के प्रत्येक आयु वर्ग या आयु समूह में जीवाशा और मृत्यु की प्रायिकता दर्शाने वाली सारणी है।

वय सारणी से हमें भलीभाँति पता चलता है कि जनसंख्या में कुल कितनी तरह की मौतें हुई हैं। इसे हम एक उदाहरण से स्पष्ट कर सकते हैं। हम 1,00,000 जन्मी बालिका शिशुओं के समूह (आमतौर पर जिसे हम सहगण या cohort कहते हैं) से शुरू करते हैं और ऐसी संख्या का अनुमान लगाते हैं जो हरेक आयु या आयु समूह में जीवित जन्मों से संबंधित है। मान लीजिए कि 1,00,000 जीवित जन्मी बालिकाओं में से पहली 95,000 शिशु बालिकाएँ 15 वर्ष की आयु तक पहुँचेगी, 92,500 बालिकाएँ 25 वर्ष की आयु तक और वह औसत आयु जिस पर अंतत: इन सभी 1,00,000 बालिकाओं की मृत्यु होगी, वह है 72 वर्ष की आयु।

वय सारणी का निर्माण करना एक सरल प्रक्रिया हैं इसमें निम्नलिखित चरण शामिल हैं जो हरेक आयु समूह के लिए दोहराए जाते हैं :

i) **आयु अंतराल (x से $x+n$) :** दो सही (exact) आयु समूहों के बीच की जीवन अवधि। सही आयु (x) 0 से शुरू होती है और प्रत्येक आयु अंतराल की निम्न सीमा को दर्शाती है और धीरे-धीरे 1, 5, 10, 15 और ऐसे ही 100+ तक बढ़ती जाती है (जो एक विवृत अंतराल है)। पहले और दूसरे आयु समूह आमतौर पर '<1' और '1- 4' हैं और अंतिम आयु समूह '100+' है जबकि बाकी के आयु समूह समान आकार के हैं, जैसे '5-9', '10-14', '15-19' ... '95-99'।

ii) **आयु अंतराल (n_x) का विस्तार :** इसका अर्थ आयु अंतराल में वर्षों की संख्या से है (अर्थात् x से $x+n$)। आमतौर पर पहला मान 1 (अंतराल <1), दूसरा 4 (1-4) और अंतिम मान (100 +) के अपवाद के साथ बाकी के मान हैं, 5 (5-9, 10-14, ... 95-99)।

iii) **आयु अंतराल में दर्ज मौतों की संख्या (d_x) :** यह स्तम्भ वय सारणी के उपयुक्त वर्ष के दौरान उस आयु समूह में मरने वाले व्यक्तियों की संख्या को दर्शाता है।

iv) **आयु अंतराल में व्यक्तियों की संख्या (P_x) :** यह स्तम्भ वय सारणी के उपयुक्त वर्ष के दौरान उस आयु अंतराल में व्यक्तियों की संख्या को दर्शाता है।

v) **पार्थक्य गुणक (Speration factor)** $_na_x$: यह ऐसे व्यक्तियों के जीवित वर्षों की औसत संख्या को दर्शाता है जिनकी मृत्यु x और $x + n$ आयु के बीच हुई है। गणना के दौरान, यह गुणक अनिवार्य भूमिका निभाता है, लेकिन सारणी में स्तम्भ के रूप में इस गुणक को दर्शाया नहीं जाता। अंतराल (x से $x+n$) में जीवित रहने वाले प्रत्येक व्यक्ति x पूर्ण वर्षों और अंतराल (x से $x+n$) के कुछ खंड तक जीवित रहा है। पूर्ण वय सारणी में 5 वर्षों की आयु से मान 0.5 (अर्थात् एक वर्ष का आधा) वैध है। आसान गणना के लिए, मान लीजिए, जो वय सारणी के 5 वर्ष आयु अंतरालों में मृत्यु को प्राप्त करते हैं, वे औसतन 2.5 वर्षों तक जीवित रहे हैं। तथापि, याद रखिए कि इस खंड का मान, समूचे अंतराल के मृत्यु संबंधी ढाँचे पर निर्भर करता है न कि किसी एकल वर्ष की मृत्यु दर पर। इसके अलावा चूँकि नवजात शिशुओं के बड़े भाग की मृत्यु जीवन के पहले कुछ हफ्तों में हुई है, इसलिए यह मान <1 और 1 - 4 आयु समूहों में काफी निम्न है।

इसी तरह, पिछले तीन समूहों (अर्थात्, 91-94, 95-99 और 100+) में मृत्यु दर काफी उच्च है। इसलिए 91-94 और 95-99 के आयु समूहों में पार्थक्य गुणक का मान काफी निम्न है। चूँकि मृत्यु निश्चित है इसलिए अंतिम आयु समूह (100+) में हमने पार्थक्य गुणक को 1 के रूप में लिया है।

यदि जन्म और मृत्यु अर्थात् दोनों की तारीखें, उपलब्ध हो तो पार्थक्य गुणक की गणना से तालिका में पार्थक्य गुणक दिया जायेगा। यदि ऐसा न हो तो प्रतिमान वय तालिका, जैसा कि तालिका 12.1 में दर्शाये कोल और डेमने द्वारा सारणीबद्ध मानों को $_1a_0$ और $_4a_1$ के लिए प्रयोग में लाया जा सकता है और बाकी के मानों को आयु अंतराल के लिए 0.5 वर्षों (अर्थात् वर्ष अंतराल में 2.5) के रूप में प्रयोग किया जा सकता है।

तालिका 12.1 : आयु 0 और 1-4 के लिए विच्छेद कारक

	आयु <1 के लिए पार्थक्य गुणक				1-4 आयु समूहों के लिए पार्थक्य गुणक		
	क्षेत्र	पुरुष	महिला	महिला/पुरुष दोनों	पुरुष	महिला	महिला/पुरुष दोनों
शिशु मृत्यु दर >0.100	उत्तर (1)	0.33	0.35	0.3500	1.558	1.570	1.570
	पूर्व (2)	0.29	0.31	0.3100	1.313	1.324	1.3240
	दक्षिण (3)	0.33	0.35	0.3500	1.240	1.239	1.2390
	पश्चिम (4)	0.33	0.35	0.3500	1.352	1.361	1.3610
शिशु मृत्यु दर <0.100	उत्तर (1)	0.0425	0.05	0.0500	1.859	1.733	1.7330
	पूर्व (2)	0.0025	0.01	0.0100	1.614	1.487	1.4870
	दक्षिण (3)	0.0425	0.05	0.0500	1.541	1.402	1.4020
	पश्चिम (4)	0.0425	0.05	0.0500	1.653	1.524	1.5240

स्रोत : कोल, अैन्स्ले जे, डेमने पी (1966) रीजनल मॉडल लाइफ टेबलस एंड स्टेबल पापुलेशनस्, प्रिंसटन यूनिवर्सिटी प्रेस।

नोट : (1) आइसलैंड, नार्वे और स्विज़रलैंड, (2) आस्ट्रिया, चेकोस्लवाकिया, नार्थ-सेंट्रल इटली, पोलैंड और हंगरी; (3) साउथ इटली, पुर्तगाल और स्पेन; (4) विश्व के बाकी के देश।

vi) **केन्द्रीय मर्त्यता (Central Mortality) ($_nM_x$):** यह स्तम्भ आयु अंतराल x से $x+n$ में होने वाली मौतों की संख्या (स्तम्भ d_x) को इस आयु समूह में मौजूद व्यक्तियों की संख्या (स्तम्भ P_x) से विभाजित करने पर प्राप्त परिणाम को दर्शाता है।

$$_nM_x = \frac{d_x}{P_x}$$

vii) **आयु समूह x और $x+n$ ($_nq_x$) के बीच होने वाली मौतों की प्रायिकता** : मरण संबंधी प्रायिकताओं की गणना प्रत्येक आयु समूह के लिए, आयु-विशिष्ट मर्त्यता दरों के आधार पर की जाती है। इस स्तम्भ को, आयु x तक जीवित रहने वाले व्यक्ति के लिए आयु समूहों के बीच मरण संबंधी प्रायिकता के रूप में स्पष्ट किया जा सकता है। सारणी के अंतिम आयु समूह के लिए जहाँ मृत्यु को नकारा नहीं जा सकता, मरण संबंधी प्रायिकता 1 है। अन्य आयु समूहों के लिए गणना अधिक जटिल है। गणना का सूत्र इस प्रकार है :

$$_nq_x = \frac{n_x \times {}_nM_x}{1+(n_x - {}_na_x)\times {}_nM_x}$$

viii) **आयु समूह x और $x+n$ के बीच जीवित रहने की प्रायिकता ($_nP_x$)** : इसे ऐसे व्यक्ति की प्रायिकता के रूप में स्पष्ट किया जा सकता है जो सही आयु $x+n$ तक जीवित पहुँचने के लिए, x आयु तक पहुँचा है। गणना का सूत्र इस प्रकार है :

$$_nP_x = 1 - {}_nq_x$$

चूँकि यह $1 - {}_nq_x$ है, सामान्यतौर पर वय सारणी में इसे हम अलग स्तम्भ के रूप में नहीं दिखाते।

ix) **किसी आयु समूह विशेष, x तक ज़ीवित रहने वाले व्यक्तियों की संख्या ($_nl_x$)** : यह स्तम्भ सहगण जिसे हमने 1,00,000 माना था, में से आयु समूह x से $x+n$ तक जीवित रहने वाले व्यक्तियों की संख्या को दर्शाता है।

x) **आयु समूह विशेष, x और $x+n$ के बीच होने वाली मौतों की संख्या ($_nd_x$)** : निम्नलिखित सूत्र के प्रयोग से इसकी गणना की जाती है।

$$_nd_x = {}_nl_x \times {}_nq_x$$

xi) **अंतराल x से $x+n$ में 1,00,000 नवजातों के समूह द्वारा जीवित रहने वाले वर्षों की संख्या ($_nL_x$)** : सहगण का प्रत्येक सदस्य, जो अंतराल x से $x+n$ तक जीवित रहता है, L के लिए n वर्षों का योगदान देता है। जबकि ऐसा सदस्य जो अंतराल x और $x+n$ में मृत्यु की प्राप्ति करता है, ऐसे व्यक्तियों के जीवित वर्षों की औसतन संख्या में योगदान देता है जो इस अवधि ($_na_x$ मौतों के पार्थक्यगुणक) में मृत्यु की प्राप्ति करते हैं। निम्नलिखित सूत्र के प्रयोग से $_nL_x$ की गणना की जाती है।

$$_nL_x = n_x \times {}_nl_{x+n} + {}_na_x \times {}_nd_x$$

जहाँ $_nl_{x+n} = {}_nl_x \times {}_nP_x$ अथवा

$$_nl_{x+n} = {}_nl_x - {}_nd_x$$

xii) **आयु समूह विशेष x के बाद जीवित रहने के कुल वर्ष ($_nT_x$)** : जीवाशा की गणना में यह संख्या अनिवार्य भूमिका निभाती है। यह वार्षिक x और कुल पीढ़ी समाप्ति के बीच जीवित समूह $_nl_x$ द्वारा जीवित वर्षों की कुल संख्या को दर्शाती है। $_nT_x$ की पहली पंक्ति का मान है, इस समूह के अंतिम घटक की मृत्यु तक समूह द्वारा जीवित वर्षों की कुल संख्या।

$_nT_x = {_nL_x}$ का योग ($_nL_x$ की अंतिम पंक्ति से $_nL_x$ की मौजूदा पंक्ति तक)

xiii) **आयु समूह x में जीवाशा ($_ne_x$)** : वय सारणी द्वारा प्रदत्त सभी सूचकों में से जीवाशा ($_ne_x$) का सर्वाधिक प्रयोग किया जाता है जो दी गई मृत्यु संबंधी दशाओं के अंतर्गत नवजातों की पीढ़ी द्वारा जीवित वर्षों की औसतन संख्या को दर्शाती है।

उदाहरण 12.9 : निम्नलिखित सारणी, वय सारणी के निर्माण के लिए अपेक्षित बुनियादी जानकारी प्रदान करती हैं यह आँकड़े 2000 वर्ष में भारतीय महिलाओं से संबंधित हैं। वय सारणी का निर्माण कीजिए।

आयु (x)	मौतों की संख्या (d_x)	व्यक्तियों की संख्या (P_x)	n_x	पार्थक्य गुणक ($_na_x$)
< 1	788471	11655599	1	0.1
1-4	430704	44728827	4	1.6
5-9	137870	54725561	5	2.5
10-14	69159	52128201	5	2.5
15-19	100055	48475620	5	2.5
20-24	119360	42745630	5	2.5
25-29	116085	39848328	5	2.5
30-34	109226	35983667	5	2.5
35-39	102540	31934500	5	2.5
40-44	124848	27744053	5	2.5
45-49	150315	23125487	5	2.5
50-54	172910	19212249	5	2.5
55-59	226553	16258203	5	2.5
60-64	288036	13715985	5	2.5
65-69	354148	10813430	5	2.5
70-74	368365	7554310	5	2.5
75-79	335430	4615527	5	2.5
80-84	252665	2332329	5	2.5
85-89	130278	817817	5	2.5
90-94	42440	183658	5	2
95-99	8199	24796	5	2
100+	915	1961	1	+
सभी आयु समूह	4428572	488625738		

स्रोत : विश्व स्वास्थ्य संगठन

पहले दिए गए सूत्र के प्रयोग से, निम्नलिखित वय सारणी का निर्माण किया गया है।

आयु	n_x	${}_na_x$	${}_nM_x$	${}_nq_x$	${}_np_x$	${}_nl_x$	${}_nd_x$	${}_nL_x$	${}_nT_x$	${}_ne_x$
< 1	1	0.1	0.06765	0.06377	0.93623	100000	6376.52	94261.1	6268416	62.6842
1 से 4	4	1.6	0.00963	0.03765	0.96235	93623.5	3524.63	366035	6174155	65.9467
5 से 9	5	2.5	0.00252	0.01252	0.98748	90098.8	1127.83	447675	5808120	64.4639
10 से 14	5	2.5	0.00133	0.00661	0.99339	888971	588.245	443384	5360446	60.2493
15 से 19	5	2.5	0.00206	0.01027	0.98973	88382.8	907.437	439645	4917061	55.6337
20 से 24	5	2.5	0.00279	0.1386	0.98614	87475.3	1212.83	434345	4477416	51.1849
25 से 29	5	2.5	0.00291	0.01446	0.98554	86262.5	1247.4	428194	4043071	46.8694
30 से 34	5	2.5	0.00304	0.01506	0.98494	85015.1	1280.57	421874	3614877	42.5204
35 से 39	5	2.5	0.00321	0.01593	0.98407	83734.5	1333.63	415339	3193003	38.1324
40 से 44	5	2.5	0.0045	0.02225	0.97775	82400.9	1833.39	407421	2777665	33.7092
45 से 49	5	2.5	0.0065	0.03198	0.96802	80567.5	2576.57	396396	2370243	29.4193
50 से 54	5	2.5	0.009	0.04401	0.95599	77990.9	3432.36	381374	1973847	25.3087
55 से 59	5	2.5	0.01393	0.06733	0.93267	74558.6	5019.87	360243	1592474	21.3587
60 से 64	5	2.5	0.021	0.09976	0.90024	69538.7	6937.35	330350	1232230	17.7201
65 से 69	5	2.5	0.03275	0.15136	0.84864	62601.3	9475.39	289318	901880	14.4067
70 से 74	5	2.5	0.04876	0.21732	0.78268	53126	11545.3	236767	612562	11.5304
75 से 79	5	2.5	0.07267	0.3075	0.6925	41580.7	12786.2	175938	375795	9.03774
80 से 84	5	2.5	0.10833	0.42622	0.57378	28794.5	12272.9	113290	199857	6.94081
85 से 89	5	2.5	0.1593	0.56964	0.43036	16521.6	9411.37	59079.6	86567	5.23963
90 से 94	5	2	0.23108	0.68236	0.31764	7110.23	4851.74	20996	27487.4	3.8659
95 से 99	5	2	0.33067	0.82999	0.17001	2258.5	1874.53	5668.9	6491.49	2.87425
100 +	1	+	0.46678	1	0	383.969	383.969	822.593	822.593	2.14235

जीवाशा हमेशा सारणी की पहली पंक्ति से अंतिम पंक्ति की ओर घटती जाती है। लेकिन जीवाशा दूसरी पंक्ति में पहली पंक्ति के तुलना में अधिक होती है। ऐसी स्थिति कभी-कभार तीसरी पंक्ति (आयु समूह 5-9) में भी देखने को मिलती है जो उच्च शिशु मृत्यु दर वाले देशों में पहली पंक्ति (आयु समूह/< 1) से उच्च हो सकती है। देखा गया है कि पुरुषों की तुलना में महिलाओं में जीवाशा अधिक होती है और कुल जीवाशा इन दोनों के बीच होनी चाहिए। लेकिन ऐसे देशों में जहाँ मातृ मृत्युदर उच्च है और महिलाओं की सामान्य जीवन दशाएँ काफी खराब हैं, वहाँ पुरुषों की तुलना में महिलाओं की जीवाशा निम्न है।

12.7 वय सारणी के अनुप्रयोग

वय सारणी का प्रयोग जनांकिकी, बीमाविज्ञान, सामाजिक और स्वास्थ्य अध्ययनों में विस्तृत रूप से किया जाता है। वय सारणी का मुख्य उद्देश्य, जन्म के समय और अन्य आयु समूहों में जीवाशा की गणना करना है। वय सारणी से कुछ ऐसे रोचक जनांकिकीय आँकड़े प्राप्त होते हैं जिन्हें विभिन्न तरीकों से प्रयोग किया जाता है। इस अनुभाग में आप वय सारणी के अनुप्रयोगों का अध्ययन करेंगे।

12.7.1 जीवन और मरण संबंधी प्रायिकता की गणना

वय सारणी बनाते समय, आपने सीखा कि ${}_nq_x$ आयु x तक जीवित रहने वाले व्यक्ति के लिए दो आयु समूहों $(x, x+n)$ के बीच मरने की प्रायिकता के बारे में जानकारी देता है। जैसे, तालिका 11.3 में आयु समूह 30-34 वर्ष की तदनुसार पंक्ति पर विचार कीजिए। 30 वर्ष तक जीवित रहने वाली महिलाओं की 30 से 34 वर्ष की आयु समूह के बीच (महिलाओं) की मरने की प्रायिकता है 0.01506 (${}_nq_x$)। इसका अर्थ है, 30 वर्ष की आयु तक जीवित रहने वाली हर 1,00,000 भारतीय महिलाओं में से 1506, (=100,000 × 0.01506) 30 और 34 वर्ष की आयु के बीच मृत्यु की प्राप्ति करेंगी। दूसरा, ${}_nq_x$, आयु x तक जीवित रहने वाले व्यक्तियों के लिए दो आयु समूह अर्थात् (x, 30-34 वर्ष/$x+n$) के बीच जीवित रहने की प्रायिकता के बारे में हमें बताता है। आयु समूह के लिए, जीवित रहने की प्रायिकता है- 1−0.01506 = 0.98494 (${}_nq_x$)। इसका अर्थ है प्रत्येक 1,00,000 भारतीय महिलाओं में से 30 वर्ष की आयु तक जीवित रहने वाली महिलाओं में से 98494, आयु समूह 30-34 वर्ष तक जीवित रहेंगी।

तीसरा, 0 से 4 वर्ष की आयु समूह के बीच में मरने वाले शिशु की जन्म संबंधी प्रायिकता की गणना भी हम कर सकते हैं। यह हमें आयु समूह 0-4 वर्षों के बीच असली जन्म-मरण (${}_nd_x$) की संख्या द्वारा प्राप्त होता है और जिसे असली जन्मों की संख्या (आमतौर पर 1,00,000) से विभाजित किया जाता है। हमारे उदाहरण में ${}_nd_x$ = 1281 और प्रायिकता 0.01281 (=1281/100000) है। यह प्रायिकता हमें बताती है कि भारत में हर 100,000 बालिका जन्मों में से (वर्ष 2000 में आकलित मृत्यु दर को ध्यान में रखकर) औसतन 1281 बालिकाएँ 0 से 4 वर्ष की आयु के बीच मृत्यु की प्राप्ति करती हैं।

12.7.2 बीमा विज्ञान में वय सारणी का उपयोग

बीमा विज्ञान में, विशेषरूप से जीवन बीमा के क्षेत्र में, वय सारणी को विशेष रूप से प्रयोग में लाया जाता है। वय सारणी, जीवन बीमा से जुड़ी विविध अनिवार्य राशियों के संदर्भ में प्रीमियमों की दर निर्धारित करने में आधार प्रदान करती है। वय सारणी बीमाविज्ञान को मज़बूत आधार देती है जिससे बीमा व्यवसाय को मात्र जुआ न समझा जाए बल्कि इससे मनुष्य को अहसास दिलाया जाता है कि आकस्मिक मृत्यु के समय इसके माध्यम से अच्छी खासी गिनी हुई रकम प्राप्त होती है जिससे भविष्य सुरक्षित बनता है। असल में जीवन बीमा के संदर्भ में प्रीमियम राशि तय करने में शामिल गणना काफी जटिल होती है। लेकिन इसके सिद्धांत काफी सरल हैं। आइए ऐसे कुछ उदाहरणों पर विचार करें।

उदाहरण 12.10 : भारत में वर्ष 2000 में मौजूद मृत्यु संबंधी दशाओं के अनुसार, एक भारतीय महिला को कुल 1,00,000 रुपये की जीवन पालिसी पर वार्षिक प्रीमियम कितना अदा करना होगा यदि उसका जीवन बीमा जन्म के समय हो गया हो और बीमा कार्यालय अपनी जमा राशियों पर कोई मुनाफा नहीं कमाता हो।

मान लीजिए प्रतिवर्ष प्रीमियम की राशि रुपये x है। चूँकि माना गया है कि औसतन पूरी आयु में महिला 62.7 वर्षों तक जीवित रहेगी। ऐसी स्थिति में महिला को रुपये $x \times 62.7$ की राशि प्रीमियम के रूप में अदा करनी होगी। यह 1,00,000 रुपये की पॉलिसी के मूल्य के बराबर होगी। इसलिए, रुपये $x \times 62.7 = 10000$ और $x = 100000/62.7$ = रुपये 1594.90 है।

उदाहरण 12.11: उपर्युक्त उदाहरण में यदि 25 वर्ष की आयु में पॉलिसी ली गई हो तो बताइए वार्षिक प्रीमियम कितना होगा।

यदि 25 वर्ष की आयु में पॉलिसी ली गई हो तो 25 वर्ष की आयु में 46.9 वर्षों की जीवाशा के लिए कुल रुपये $x \times 46.9$ प्रीमियम अदा करना होगा। ऐसी स्थिति में वार्षिक प्रीमियम $x = 100000/46.9 = 2132.20$ रुपये होगा।

उदाहरण 12.12 : उदाहरण 12.10 में यदि पॉलिसी-एनडोमेंट पॉलिसी है और जिसे 30 वर्ष की आयु में लिया गया हो और जिसे 50 वर्ष की आयु तक या मृत्यु होने से पहले तक अदा करना है। ऐसी स्थिति में कितना वार्षिक प्रीमियम अदा करना होगा।

यदि पॉलिसी, एनडोमेंट पॉलिसी है और जिसे मान लीजिए 30 वर्ष की आयु में लिया गया और जिसे 50 वर्ष की आयु तक या मृत्यु होने से पहले तक अदा करना है तो हम अलग विधि से इसे हल करेंगे। सारणी 11.3 से हम देखते हैं कि 30 वर्ष की आयु तक 850155 ($_nl_x$) जीवित व्यक्ति, 30 और 50 वर्ष की आयु समूह के बीच 1600529.5 (415338.6 + 407421 + 396396. 1 + 381373.8 ($_nL_x$) वर्षों तक जीवित रहते हैं, जिसके परिणाम स्वरूप औसतन कुल x रुपये $\times$ (1600529.5/85015) प्रीमियम इकट्ठे किए जाएँगे और इसके बाद वार्षिक प्रीमियम 1,00,000 रुपये + 18.83 = 5310.67 रुपये होगा।

12.7.3 वय सारणी के अन्य अनुप्रयोग

बीमा में इसके फायदों के अलावा यह विभिन्न देशों या क्षेत्रों के मर्त्यता संबंधी तुलनात्मक विश्लेषण करने में भी उपयोगी है। आइए वय सारणी के कुछ ऐसे अनुप्रयोगों की चर्चा करें।

i) **किन्हीं विशिष्ट कारणों से मर्त्यता/मृत्यु दर की गणना करना** : जनसंख्या के विविध समूह जैसे लिंग (स्त्री/पुरुष), आयु वितरण (विविध आयु समूह), धर्म, क्षेत्र आदि के लिए वय सारणी की गणना, तुलना संबंधी कार्यों के लिए की जाती हैं। मर्त्यता संबंधी सांख्यिकी हमें जनसंख्या के विविध समूहों में होने वाली विशिष्ट मौतों का पता लगाने के लिए प्रेरित करती है।

ii) **मर्त्यता संबंधी दशाओं की तुलना** : जन्म के समय और अन्य आयु समूहों में जीवाशा का पता लगाना, मर्त्यता संबंधी दशाओं के उत्कृष्ट सूचकांक हैं। ये सूचकांक एक स्थान से दूसरे स्थान पर और समय के साथ-साथ बदलते रहते हैं। अधिकांश देशों में पिछले कई वर्षों से शिशु मृत्यु दर में गिरावट आने की वजह से जीवाशा तेजी से बढ़ी हैं जैसाकि आपने पहले अध्ययन किया है। ऐसे स्थानों को छोड़ कर जहाँ महिला मातृ मृत्यु दर उच्च है, अन्य सभी जगहों पर पुरुष जीवाशा की तुलना में महिला जीवाशा उच्च है। तालिका 12.4 से हमें कुछ चुनिंदा देशों में महिला और पुरुष जीवाशा का ब्यौरा प्राप्त होता है।

तालिका 12.4 : जन्म के समय जीवाशा : 1999 में कुछ चुनिंदा देशों में

(वर्षों में)

	पुरुष	महिला
आस्ट्रेलिया (अ)	76.6	82.0
कनाडा	75.9	81.4
चीन	68.3	72.5
फ्रांस	74.5	82.3
जर्मनी	74.3	80.6
हांगकांग (एस ए आर ऑफ चीन)	76.7	82.2

भारत	62.4	63.3
इंडोनेशिया	63.9	67.7
इटली	75.2	81.6
जापान	77.3	84.1
कोरिया, गणतंत्र	70.9	78.4
नीदरलैंड	75.3	80.7
न्यूजीलैण्ड	74.8	80.1
पापुआ न्यू गुयाना	55.4	57.3
सिंगापुर	75.2	79.6
यूनाइटेड किंगडम	75.0	80.0
अमेरिका	73.9	79.7

(अ) आस्ट्रेलिया के लिए संदर्भ अवधि 1998-2000 है।

स्रोत: डेथस्, आस्ट्रेलिया (3302.0), संयुक्त राष्ट्र विकास कार्यक्रम 2000

iii) **जनसंख्या प्रक्षेपण** : वय सारणी का प्रयोग, आयु और लिंग अर्थात् स्त्री/पुरुष समूह को ध्यान में रख कर जनसंख्या के समग्र अनुमान तैयार करने में भी किया जाता है। इसका अर्थ है कि कुछ भावी तारीखों पर जनसंख्या का आकार क्या होगा, इसका अनुमान भी वय सारणी की सहायता से लगाया जा सकता है।

12.7.4 वय सारणी की सीमाएँ

वय सारणी जनगणना और प्रतिदर्श पंजीकरण पद्धति जैसे स्रोतों से इकट्ठे किए गए आँकड़ों पर आधारित होती है। अत: वय सारणी अनुमानों के कई दोष भी हैं। ये ऐसे दोष हैं जो जनसंख्या संबंधी जनगणनाओं और जन्म-मृत्यु रिकार्डों पर आधारित सांख्यिकीय मापों में पाए जाते हैं। आयु समूहों और मर्त्यता संबंधी पंजीकरण पर आधारित आँकड़े अधूरे या एक तरफा होते हैं। शिशु मृत्यु दर मोटे तौर पर जीवाशा पर आधारित हैं और जिसका अर्थ है कि इस सूचक की अधूरी जानकारी जो बहुत से देशों में आदत सी बन चुकी है। इससे सारणियों के परिणामों पर महत्त्वपूर्ण असर पड़ता हैं। इसी तरह, उच्च मृत्यु दर वाले विशिष्ट आयु समूह और स्त्री और पुरुष समूहों में पाये जाने वाले महत्त्वपूर्ण अंतरों को भी अनदेखा कर दिया जाता है क्योंकि समग्र जीवाशा पर इसका बहुत कम असर पड़ता है।

चूँकि क्षेत्रीय या राष्ट्रीय स्तरों पर प्रवास संबंधी गतिविधियाँ जनसंख्या की संरचना को अधिक प्रभावित करती हैं इसलिए स्थानीय या उप-क्षेत्रीय स्तर पर छोटी जनसंख्या के लिए सामान्यतौर पर वय सारणियाँ बनाने की सिफारिश नहीं की जाती।

ऐसे मामलों में मृत्यु दर की ऐसी बहुत छोटी संख्या प्राप्त होती है जो तालिका के स्तम्भों की गलत गणना दिखा सकती है।

बोध प्रश्न 3

पाठ में तालिका 11.3 में दी गई वय सारणी को पढ़िए। अब निम्नलिखित प्रश्नों के उत्तरों से वय सारणी में मानों को स्पष्ट कीजिए।

1) भारत में 2000 वर्ष में 1 वर्ष की आयु पूरी करने से पहले मरने वाली बालिका शिशु की प्रायिकता क्या होगी?

....

....

....

....

....

2) भारत में 2000 में जन्मी महिला का कितने वर्षों तक जीवित रहने का अनुमान लगाया जाता है?

....

....

....

....

....

3) 15 से 20 वर्षों की आयु में मरने वाली महिला की प्रायिकता क्या होगी?

....

....

....

....

....

4) 15 और 20 वर्ष की आयु के बीच मृत्यु दर क्या होगी?

....

....

....

....

....

5) वह प्रायिकता बताइए, जब महिला 15 वर्ष से 20 वर्ष की आयु में पहुँचती है?

....

....

....

....

....

....

6) भारत में 2000 में 15 और 20 वर्ष की आयु में महिला से कितने अतिरिक्त वर्षों तक जीवित रहने की आशा की जाती है?

....

....

....

....

....

12.8 सारांश

जन्म-मृत्यु सांख्यिकी मुख्य रूप से जन्म और मृत्यु से संबंद्ध है। जन्म-मृत्यु दर की विश्वसनीयता पंजीकरण पद्धति की प्रभाविता पर निर्भर करती है। कानूनी मौजूदगी के बावजूद जन्म और मृत्यु का अधूरा पंजीकरण, जन्म और मृत्यु दरों की सही छवि दिखाने में बाधा डालता है।

वय सारणियाँ, समग्र जनसंख्या की मर्त्यता और जीवन संबंधी अनुभवों को दर्शाती है और विशिष्ट समूहों पर और विशिष्ट संमयावधियों में इसके प्रभाव के मूल्यांकन को कारगर बनाती हैं। यह एक सरल साधन है जिसे नियमित रूप से आकड़ों को इकठ्ठा करके आशानी से बनाया जा सकता है।

यह ध्यान में रखना बेहद ज़रूरी है कि जनगणनाओं और मर्त्यता संबंधी आकड़ों को ध्यान में रख कर वय सारणियाँ निर्मित की जाती हैं। अत: आँकड़ों की गुणवत्ता, वय सारणी की वैधता पर विशेष प्रभाव डालती है।

12.9 शब्दावली

बीमा विज्ञान (Actuarial Science): बीमा विज्ञान, वित्त और बीमा के संदर्भ में गणितीय और सांख्यिकीय विधियों के अनुप्रयोग से संबंधित है। यह विशेषरूप से दीर्घावधि में बीमा आदि के क्षेत्र में जोखिमों के निर्धारण से संबंधित है। बीमा विज्ञान में हम बीमा संबंधी जोखिमों और प्रीमियमों को परिकलित करते हैं।

सहगण (Cohort) ऐसे व्यक्तियों का समूह जो निश्चित समयावधि में एक जैसे सांसारिक/जनांकिकीय अनुभव प्राप्त करते हैं। जैसे 2003 के जन्म सहगण में ऐसे व्यक्तियों का समूह शामिल है जिन सभी का उसी वर्ष में जन्म हुआ था।

शिशु मृत्यु दर (Infant Mortality Rate) : निश्चित समयावधि में प्रति 1000 जीवित जन्मों में से एक वर्ष की आयु पूरी करने से पहले मरने वाले शिशुओं की संख्या।

जीवाशा (Life expectancy) : व्यक्ति के ऐसे अतिरिक्त वर्षों की औसतन संख्या जिसे उस व्यक्ति ने आगे जीना है, बशर्ते मौजूदा मृत्यु दर की प्रवृत्तियाँ उस व्यक्ति के शेष जीवन के लिए निरंतर कार्य रहें। आमतौर पर जन्म के समय जीवाशा का प्रयोग किया जाता है।

मध्य वर्ष जनसंख्या (Mid-year population)	:	यह वर्ष के अंत के आकलनों का औसतन माना जाता है। जैसे 2003 की मध्य वर्ष जनसंख्या 31 दिसम्बर 2002 और 31 दिसम्बर 2003 की जनसंख्या की औसतन संख्या होगी।
प्रवास (Migration)	:	नयी या अर्ध-स्थायी रिहायशों को स्थापित करने के उद्देश्य से व्यक्तियों का किन्हीं निश्चित सीमाओं के आर-पार होना।
स्वाभाविक वृद्धि (Natural Increase)	:	निश्चित समयावधि में जनसंख्या में मौतों की तुलना में अतिरिक्त जन्मों का होना।
स्वाभाविक वृद्धि दर (Rate if Natural increase)	:	होने वाली मौतों की तुलना में अतिरिक्त जन्मों के कारण, दिए गए वर्ष में जिस दर पर जनसंख्या बढ़ती है और जिसे आधार जनसंख्या के प्रति 1000 व्यक्तियों के रूप में अभिव्यक्त किया जाता है। इसमें आप्रवास की दर शामिल नहीं है।

12.10 कुछ उपयोगी पुस्तकें

Agarwal, B.L., (1988), *Basic Statistics*, Wiley Eastern Limited, New Delhi.

Ansari, M.A., Gupta, O.P. and Chaudhary, S.C. (1980), *Applied Statistics*, Kedarnath Ram Nath & Co., Meerut.

Benjamin, B.(1959), *Elements of Vital Statistics, George Allen and Unwin*, London.

Chang, C.J. (1980), *Life Tables and Mortality Analysis*, Geneva: World Health Organisation.

12.11 बोध प्रश्नों के उत्तर या संकेत

बोध प्रश्न 1

1)

वर्ष	मध्य वर्ष जनसंख्या
1950	40.03
1966	51.14
1973	59.46
1985	76.44
1998	97.05

2)

वर्ष	मध्य वर्ष जनसंख्या
1954	39.88
1966	50.81
1973	59.07
1985	75.85
1998	97.08

3) देखें भाग 12.2 और उत्तर दें।

4) देखें भाग 12.3 और उत्तर दें।

बोध प्रश्न 2

1) ग्रामीण क्षेत्रों में जन्म दर उच्च होती है, क्योंकि परिवार नियोजन की विधियों और इसकी आवश्यकता पर वहां के लोगों में जागरूकता का अभाव होता है।

2) छोटे और बड़े शहरों में उन्नत स्वास्थ्य सुविधाओं के कारण, शहरी क्षेत्रों में मृत्यु दर निम्न होती है।

3) ग्रामीण क्षेत्रों में स्वास्थ्य सुविधाओं के अभाव और माताओं में कुपोषण के कारण, ग्रामीण क्षेत्रों में शिशु मृत्यु दर उच्च होती है।

बोध प्रश्न 3

1) भारत में 2000 ($_1q_0$) में 1 वर्ष से कम की आयु में मरने वाली बालिका के लिए प्रायिकता 0.06377 है।

2) भारत में 20000 में जन्मी बालिका के जीवित रहने ($_1e_0$) के अनुमानित 62.68 वर्ष हैं।

3) आयु समूह ($_5q_{15}$) में 15 और 20 वर्षों के बीच मरने वाली महिला की प्रायिकता 0.01027 है।

4) आयु समूह ($_5M_{15}$) में 15 और 20 वर्षों के बीच मृत्यु दर 0.00206 है।

5) 15 – 19 आयु समूह की महिला की 20-24 वर्ष के आयु समूह ($_5q_{15}$) में पहुँचने की प्रायिकता 0.98973 है।

6) भारत में 2000 में 15-20 वर्ष के आयु समूह में महिला की जीवाशा ($_5e_{15}$) 55.63 वर्ष है।

इकाई 13 प्रारंभिक प्रायिकता

इकाई की रूपरेखा

13.0 उद्देश्य

इस इकाई के अध्ययन के पश्चात् आप :

- प्रायिकता की अवधारणा की व्याख्या कर सकेंगे;
- बेज-प्रमेय सहित प्रायिकता नियमों की व्याख्या कर सकेंगे; तथा
- प्रायिकता से संबंधित प्रश्नों के हल कर सकेंगे।

13.1 प्रस्तावना

प्रायः हम निम्न प्रकार के व्याख्यान करते हैं :

कल वर्षा हो सकती है।

टीम अ के प्रतियोगिता जीतने की संभावना अधिक है।

श्रीमान ब के प्रधान बनने की संभावना नहीं है।

शायद श्रीमान क की श्रीमान ख से भेंट हुई।

इन सभी कथनों में हमें कुछ अनिश्चितता की झलक का अनुभव होता है। उदाहरण के लिए प्रथम कथन में हमें इस बात की निश्चितता नहीं है कि कल वर्षा होगी। इसी प्रकार, अंतिम कथन में, हमें यह निश्चित जानकारी नहीं है कि श्रीमान क की श्रीमान ख से भेंट हुई हैं।

कोई कथन, जिसमें घटना के घटित होने के बारे में अनिश्चितता की झलक महसूस हो, एक प्रायिक कथन कहलाता है। इस प्रकार, उपर दिए गए सभी कथन, प्रायिक कथन हैं।

मान लीजिए, प्रथम प्रायिकता कथन के संदर्भ में, हम यह प्रश्न पूछते हैं कि कल वर्षा होने का संभावना कितनी है?

इसके लिए हम यह उत्तर दे सकते हैं कि, कल वर्षा की संभावना 75% है।

यह कथन केवल प्रायिकता कथन ही नहीं है बल्कि इसमें कल वर्षा होने की घटना की निश्चितता (या अंतर्निहित अनिश्चितता) की कोटि का तुलनात्मक परिमाप भी दिया गया है। अतः वर्षा की निश्चितता की कोटि 75%, एक तुलनात्मक माप है तथा साथ ही वर्षा की अंतर्निहित साहचर्य अनिश्चितता 25% है। मान लीजिए निश्चितता की कोटि के माप के लिए हमारे पास एक पैमाना है। इस पैमाने पर निश्चितता की कोटि में 0% से 100% तक परिवर्तन होगा। इसी पैमाने द्वारा अंतर्निहित अनिश्चितता की कोटि को भी मापा जा सकता है। अतः यदि किसी घटना के घटित न होने पर पूर्ण विश्वास है तो इसके घटने की सम्भावना 0% या निश्चित है तथा घटने की असम्भावना 100% या अनिश्चित है। इस प्रकार यदि किसी घटना के घटने में पूर्ण विश्वास है तो इसके घटने की सम्भावना 100% है तथा ना घटने की सम्भावना 0% है। लेकिन, इस बात पर ध्यान दें कि दोनों कथनों में 100% घटित होना (या 0% घटित न होना) परवर्ती प्रायिकता तथा 0% घटित होना (या 100% घटित न होना) में अनिश्चितता की कोई झलक नहीं हैं। अतः पक्के तौर पर इन कथनों को प्रायिकता कथन नहीं कहा जा सकता। किसी घटना के घटित हो सकने की निश्चितता की कोटि के तुलनात्मक माप को घटना की प्रायिकता कहते हैं। परम्परा के अनुसार, इस कोटि का माप इकाई की तुलना में किया जाता है। इस प्रकार किसी घटना के घटित होने की सम्भावना यदि 30% है तो हम यह कहते है कि इस घटना के घटित होने की प्रायिकता 0.3 है। इसी प्रकार, कल वर्षा होने की संभावना 75% है, को हम इस प्रकार लिखते हैं: कल वर्षा होने की प्रायिकता 0.75 है।

इस संदर्भ में एक प्रश्न यह पूछा जा सकता है: हम एक घटना की प्रायिकता कैसे प्राप्त कर सकते हैं? इसका उत्तर अगले भाग में दिया गया है।

13.2 प्रायिकता की परिभाषा

स्पष्टतः किसी घटना के घटित होने की प्रायिकता ज्ञात करने की समस्या वास्तव में घटित होने की निश्चितता की कोटि मापने की समस्या है। एक घटना की निश्चितता की कोटि के बारे में विभिन्न गणितज्ञों की सोच भिन्न हैं, परिणामतः, प्रायिकता की कई परिभाषाएँ दी गई हैं। इन परिभाषाओं के द्वारा हमें किसी घटना की प्रायिकता ज्ञात करने की विधि के बारे में जानकारी प्राप्त होती है। इस इकाई में हम तीन परिभाषाओं का जिक्र करेंगे। ये इस प्रकार हैं :

i) चिर प्रतिष्ठित या गणितीय परिभाषा,

ii) तुलनात्मक बारंबारता या सांख्यिकीय परिभाषा, तथा

iii) आधुनिक भाषा या प्रायिकता की अभिगृहीतीय अभिगम।

13.2.1 चिरप्रतिष्ठित या गणितीय परिभाषा

यह परिभाषा कुछ अवधारणाओं पर आधारित है। पहले हम इनकी जानकारी प्राप्त करेंगे।

क) *सांख्यिकीय प्रयोग*

एक प्रयोग जिसके एक से अधिक परिणाम संभव हों, को सांख्यिकीय प्रयोग कहते हैं। सांख्यिकीय प्रयोग को एक परीक्षण (experiment) भी कहते हैं। इस प्रकार, एक सिक्के के उछाल को, जिसमें यह देखना कि परिणाम चित या पट है, परीक्षण कहते हैं। कुछ कथनों जिनके अर्थ एक से अधिक संभावित परिस्थितियाँ हो सकती है, को भी परीक्षण कहा जा सकता है। उदाहरण के लिए , भाग 13.1 के शुरू में दिए गए चारों कथनों को परीक्षण या सांख्यिकीय प्रयोग कहा जा सकता है।

ख) *घटना*

एक परीक्षण के एक संभावित परिणाम को घटना (event) कहते हैं। एक सिक्के के उछाल में चित प्राप्त होना एक घटना है। इसी प्रकार एक पासे के उछाल में 5 का अंक प्राप्त होना या एक विषम संख्या प्राप्त होना, कुछ संभव घटनाएँ हैं। बाद का उदाहरण इस बात का सूचक है कि एक घटना परीक्षण की एक या अधिक संभव परिणामों से भी बन सकती है। वास्तव में एक विषम संख्या प्राप्त करने की घटना एक पासे के उछाल के तीन परिणामों से बनी है। यहाँ यह ध्यान दें कि परीक्षण के एक ही परिणाम से बनी घटना को प्रायः मूल घटना कहते हैं।

ग) *निःशेष घटनाएँ*

किसी परीक्षण के सभी संभव परिणामों को सम्मिलित करने वाले घटनाओं के समुच्चय को निःशेष घटनाएँ (exhaustive events) कहते हैं। उदाहरण के लिए, एक सिक्के के उछाल में दो घटनाओं, चित या पट के अतिरिक्त और कोई संभावना नहीं होती। अतः समुच्चय (चित, पट) सिक्के के उछाल के साहचर्य निःशेष घटनाओं का एक समुच्चय है। एक और उदाहरण पर विचार कीजिए। हम यह जानते है कि एक पासे के 6 फलक होते हैं। जिन पर 1 से 6 तक बिन्दु अंकित होते हैं। यदि पासे के उछाल में आने वाली संख्या एक घटना मान ली जाय तो समुच्चय (1, 2, 3, 4, 5, 6), निःशेष घटनाओं का समुच्चय कहलाता है। एक निःशेष घटनाओं के समुच्चय में तत्त्वों की संख्या को परीक्षण की घटनाओं की संख्या कहते हैं।

घ) *अनुकूल घटनाएँ*

वह घटनाऐं, जो किसी घटना के घटित होने का समर्थन करती हैं, अनुकूल घटनाएँ (favourable events) कहलाती हैं। मान लिया, एक पासे को उछाल कर यह देखा जाता है कि क्या परिणाम संख्या सम है। इस परीक्षण में वह फलक, जिन पर 2, 4 या 6 बिन्दु हैं, ऐसी घटनाएँ हैं जो पासे पर सम संख्या आने वाली घटना की समर्थक हैं।

ङ) *समप्रायिक घटनाएँ*

यदि किसी परीक्षण में, प्रत्येक संभव घटना के घटित होने की संभावना समान है तो इन घटनाओं को समप्रायिक घटनाएँ (equally likely events) कहते हैं। एक पासे के उछाल में, यदि हमारा विश्वास है कि, सभी 6 फलकों के ऊपर आने की संभावनाएं समान है तो 6 संभव घटनाएँ समप्रायिक कहलाती हैं।

च) परस्पर अपवर्जी घटनाएँ

यदि किसी परीक्षण में, एक घटना के घटित होने पर दूसरी घटना का घटित होना संभव न हो, तो ये दो घटनाएँ परस्पर अपवर्जी घटनाएँ (mutually exclusive events) कहलाती हैं। हम जानते हैं कि एक सिक्के के उछाल में दो घटनाएँ, चित तथा पट, परस्पर अपवर्जी घटनाएँ हैं।

अब हम चिरप्रतिष्ठित परिभाषा को समझने में समर्थ हैं। इस परिभाषा के अनुसार :

यदि एक परीक्षण के n परस्पर अपवर्जी, समप्रायिक तथा निःशेष परिणाम संभव है, जिनमें से m परिणाम एक घटना A के समर्थक हैं, तब A की प्रायिकता जिसको $P(A)$ से सूचित किया जाता है, $P(A) = \frac{m}{n}$ होगी।

स्पष्टतः यदि A एक असंभव घटना है तो n संभव परिणामों में से कोई भी इंसका समर्थन नहीं करेगा, अर्थात $m = 0$ होगा। इस परिस्थिति में $P(A) = \frac{m}{n} = \frac{0}{n} = 0$ होगी।

इसके विपरीत यदि A एक निश्चित घटना है, तब सभी n संभव परिणाम इसका समर्थन करेंगे, अर्थात $m = n$ होगा। इस परिस्थिति में $P(A) = \frac{m}{n} = \frac{n}{n} = 1$ होगी।

अब हम, प्रायिकता परिकलन के लिए चिरप्रतिष्ठित भाषा के कुछ अनुप्रयोगों पर विचार करेंगे।

उदाहरण 13.1

एक निष्पक्ष सिक्के के उछाल में चित प्राप्त करने की प्रायिकता क्या है?

हम यह जानते है कि एक सिक्के के उछाल का परिणाम चित या पट के अतिरिक्त और कुछ नहीं हो सकता। अतः ये दो घटनाएँ, निःशेष घटनाओं का एक समुच्चय है क्योंकि चित तथा पट एक साथ घटित नहीं हो सकते। ये घटनाएँ परस्पर अपवर्जी भी हैं। अंत में, क्योंकि सिक्का निष्पक्ष दिया हुआ है, इसीलिए ये घटनाएँ समप्रायिक हैं। इस प्रकार चिरप्रतिष्ठित परिभाषा की सभी शर्तें संतुष्ट होती है।

निःशेष परिणामों की संख्या $n = 2$ (चित या पट)

अपेक्षित घटना (चित) को समर्थन करने वाले परिणामों की संख्या $m = 1$

अतः चित के घटित होने की प्रायिकता $P(\text{चित}) = \frac{m}{n} = \frac{1}{2}$

आगे दिए हुए उदाहरणों में हम सीधे प्रश्न के हल को प्राप्त करेंगे। लेकिन, आप स्वयं को संतुष्ट कर लें कि चिरप्रतिष्ठित परिभाषा की सभी शर्तें विद्यमान हैं।

उदाहरण 13.2

एक निष्पक्ष पासे को उछाला जाता है। 1 या 6 प्राप्त करने की प्रायिकता क्या है?

एक पासे के 6 फलक होते हैं जिन पर $1, 2, 3, 4, 5$ तथा 6 अंकित होते हैं तथा पासा उछालने पर इनमें से कोई एक फलक उपर आ सकता है। अतः निःशेष परिणामों की संख्या

$n=6$ है। यदि 1 या 6 अंकित फलक उपर आता है तो यह अपेक्षित घटना का समर्थक है। अतः अपेक्षित घटना का समर्थन करने वाले परिणाम $m=2$ हैं।

यदि हम 1 या 6 प्राप्त करने की प्रायिकता को P(1 या 6) से सूचित करें तो

$$P(1 \text{ या } 6) = \frac{m}{n} = \frac{2}{6} = \frac{1}{3}$$

उदाहरण 13.3

एक निष्पक्ष सिक्के को दो बार उछाला जाता है। कम से कम एक बार चित प्राप्त करने की प्रायिकता क्या होगी?

यदि चित को H तथा पट को T से सूचित करें तो चार परिणाम संभव हैं, जोकि इस प्रकार हैं

$$(H, H) \quad (H, T) \quad (T, H) \quad (T, T)$$

इस प्रकार यहाँ पर $n=4$, तथा अपेक्षित घटना का समर्थन करने वाली घटनाओं की संख्या $m=3$ हैं।

अतः $P(\text{कम से कम एक चित}) = \frac{m}{n} = \frac{3}{4}$

चिरप्रतिष्ठित परिभाषा की सीमाएँ

चिरप्रतिष्ठित परिभाषा की कुछ गंभीर सीमाएँ है, जोकि इस प्रकार हैं:

क) इस परिभाषा का उपयोग केवल उन परिस्थितियों में हो सकता है जब परीक्षण के विभिन्न परिणाम समप्रायिक हों। लेकिन वास्तव में, इनका सदैव समप्रायिक होना आवश्यक नहीं होता। उदाहरण के लिए, यदि सिक्का निष्पक्ष नहीं हैं तो इस परिभाषा के आधार पर चित या पट की प्रायिकता ज्ञात करना संभव नहीं है।

ख) चिरप्रतिष्ठित परिभाषा केवल परीक्षण के सीमित संख्या में परिणामों के लिए तर्कसंगत है। परीक्षण के परिणामों की संख्या असीमित होने पर यह विफल हो जाती है। वास्तव में, परिणामों की संख्या सीमित होने पर भी, कुछ परिस्थितियों में, सभी परिणामों की संख्या की गणना करना संभव नहीं होता।

ग) प्रायिकता की परिभाषा के लिए, चिरप्रतिष्ठित परिभाषा, समप्रायिक शब्द का उपयोग करती है, जोकि प्रायिकता की अवधारणा के ज्ञान की पूर्व-मान्यता पर आधारित है। अतः यह परिभाषा वृत्तीय है।

बोध प्रश्न 1

1) एक बक्से में 4 सफेद तथा 6 लाल गैंदे हैं। बक्से में देखे बिना एक गैंद निकाली गई। इसके सफेद होने की क्या प्रायिकता है?

....

....

2) छः फलकों वाले एक पासे को उछाला गया। एक सम संख्या ज्ञात करने की प्रायिकता क्या होगी?

....

....

3) एक सिक्के को दो बार उछाला गया। दोनों चित या दोनों पट प्राप्त करने की क्या प्रायिकता होगी?

....

....

4) 52 ताश के पत्तों में से एक पत्ता निकाला गया। बादशाह प्राप्त न करने प्रायिकता क्या होगी?

....

....

13.2.2 तुलनात्मक बारंबारता या सांख्यिकीय परिभाषा

यह प्रायिकता की एक और परिभाषा है जिसका प्रायः उपयोग किया जाता है। यदि हम एक परीक्षण को बार-बार दोहराएँ तथा घटना के घटित होने का प्रेक्षण करें, तो हम यह पाते हैं कि परीक्षणों की संख्या में वृद्धि होने पर, घटना के घटित होने की संख्या के कुल परीक्षणों की संख्या से अनुपात में एक निश्चित मान पर स्थिर होने की प्रवृति होती है। स्पष्टतः, घटना के घटित होने की संख्या को इसकी बारंबारता कहते हैं, तथा जब इसको कुल परीक्षणों की संख्या से भाग किया जाता है तो हमें घटना की तुलनात्मक बारंबारता प्राप्त होती है। अन्य शब्दों में जब परीक्षणों की संख्या काफी बड़ी हो जाय तो तुलनात्मक बारंबारता में एक सीमा की और अग्रसर होने की प्रवृति होती है। तुलनात्मक बारंबारता की परिभाषा के अनुसार यह, सीमान्त मान, विचाराधीन चर की प्रायिकता होता है। मान लिया हम सिक्के को उछालने का बार-बार परीक्षण करते हैं तथा चित आने की संख्या का प्रेक्षण करते हैं। यदि हम यह पाते है कि, जैसे-जैसे उछालों की संख्या में वृद्धि होती है, अर्थात, 10 से 100 से 1000 से 10000 इत्यादि, तो चित की तुलनात्मक बारंबारता क्रमशः $\frac{1}{2}$ पर स्थिर हो रही है, तब हम कह सकते है कि एक सिक्के के उछाल में चित की प्रायिकता $\frac{1}{2}$ है। गणितीय रूप में यदि कुल परीक्षणों की संख्या n है जिनमें से एक घटना A, m बार घटित हुई, तब A की प्रायिकता $P(A) = \lim_{n \to \infty} \frac{m}{n}$ होगी।

13.2.3 आधुनिक परिभाषा या प्रायिकता की अभिगृहीतीय अभिगम

प्रायिकता की चिरप्रतिष्ठित तथा तुलनात्मक परिभाषाओं की मुख्य सीमा यह है कि इनके द्वारा प्रायिकता के विषय का गणितीय विकास संभव नहीं है। आधुनिक परिभाषा इस सीमा को ध्यान में रखते हुए की गई है। आधुनिक या अभिगृहीतीय अभिगम परिभाषा प्रस्तुत करने से पूर्व निम्नलिखित अवधारणाओं की जानकारी आवश्यक है :

i) **प्रतिदर्श क्षेत्र (sample space)** : यह एक परीक्षण के सभी संभव (या निःशेष) परिणामों का समुच्चय (set) होता है। एक परीक्षण के प्रतिदर्श क्षेत्र को S से सूचित किया जाता है जोकि $S = \{e_1, e_2, ... e_n\}$ है, जहाँ पर $e_1, e_2, ... e_n, n$ मूल घटनाएँ हैं।

   यदि एक सिक्का उछालना एक परीक्षण है तो इसका प्रतिदर्श क्षेत्र $S = \{H, T\}$ होगा। इसी प्रकार, जब एक पासे को उछाला जाए तो प्रतिदर्श क्षेत्र $S = \{1, 2, 3, 4, 5, 6\}$ होगा।

प्रतिदर्श क्षेत्र के तत्त्व क्रमित युग्म (ordered pairs) भी हो सकते हैं। उदाहरण के लिए दो सिक्कों के एक साथ उछालने का प्रतिदर्श क्षेत्र $S = \{H, H\}, \{H, T\}, \{T, H\}, \{T, T\}$ है।

इसके अतिरिक्त एक प्रतिदर्श क्षेत्र, जिसमें विद्यमान तत्त्वों की संख्या सीमित है या असीमित, के अनुसार, सीमित या असीमित हो सकता है।

ii) **घटना** : प्रतिदर्श क्षेत्र के किसी *उपसमुच्चय* (subset) को एक घटना कहते हैं। उदाहरण के लिए यदि पासे पर विषम संख्या प्राप्त करना एक घटना है जोकि A से सूचित की जाती है तब $A = \{1, 3, 5\}$ होगा। इसी प्रकार $B = \{(H, H), (H, T), (T, H)\}$ कम से कम एक चित आने की एक घटना है।

iii) **घटना का घटित होना** : जब भी किसी परीक्षण का परिणाम किसी घटना के समुच्चय का अवयव होता है तो हम यह कहते हैं कि यह घटना घटित हुई है। इस प्रकार, यदि पासे को उछालने पर 1 आता है तो घटना A घटित मानी जाती है। इस अवधारणा के आधार पर हम यह कह सकते हैं कि किसी परीक्षण के प्रतिदर्श क्षेत्र का घटित होना निश्चित होता है।

आधुनिक परिभाषा के अनुसार, एक घटना A की प्रायिकता, जोकि $P(A)$ से सूचित होती है, एक वास्तविक समादृत समुच्चय फलन (real valued set function) होता है जोकि प्रतिदर्श क्षेत्र S के किसी उपसमुच्चय A के लिए एक सहचारी वास्तविक मान, $P(A)$, प्रदान करता है। घटना A की प्रायिकता $P(A)$ होने के लिऐ इसे निम्नलिखित प्रतिबंध संतुष्ट करने आवश्क हैं। इन प्रतिबंधों को प्रायिकता सिद्धांत के अभिगृहीत (postulates) भी कहा जाता है।

1) एक प्रतिदर्श क्षेत्र S में, एक घटना A की प्रायिकता इकाई के बराबर या कम एक अऋणात्मक वास्तविक संख्या होती है, अर्थात, $0 \le P(A) \le 1$ होती है।

2) ऐसी घटना जिसका घटित होना निश्चित है, की प्रायिकता इकाई के बराबर होती है। जैसा कि हम जानते है कि प्रतिदर्श क्षेत्र S का घटना निश्चित होता है, इसीलिए $P(S) = 1$ होती है।

3) एक प्रतिदर्श क्षेत्र S में यदि तीन परस्पर अपवर्जी घटनाएँ A_1, A_2, A_3 हैं, तो

$$P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3)$$

यह संबंध घटनाओं की अधिक संख्याएँ लेकर व्यापक बनाया जा सकता है।

यहाँ पर यह ध्यान दीजिए कि प्रतिदर्श क्षेत्र $S = \{e_1, e_2, ... e_n\}$ में मूल घटनाएँ $e_1, e_2, ... e_n$ परस्पर अपवर्जी हैं। अतः, तृतीय अभिगृहीत के आधार पर हम यह कह सकते है कि

$$P(S) = \sum_{i=1}^{n} P(e_i)$$

अर्थात प्रतिदर्श क्षेत्र की प्रायिकता, इसमें विद्यमान मूल घटनाओं की प्रायिकताओं के योग के बराबर होती है। इसके आधार पर हम यह भी कह सकते हैं *कि किसी घटना की प्रायिकता उसमें विद्यमान मूल घटनाओं की प्रायिकताओं का योग होती है*। अतः, किसी घटना की प्रायिकता ज्ञात करने के लिए हमें इसमें विद्यमान मूल घटनाओं की प्रायिकताओं की जानकारी प्राप्त करना आवश्यक है। यह जानकारी निम्नलिखित तीन विधियों में से किसी एक विधि द्वारा प्राप्त की जा सकती है।

1) विभिन्न मूल घटनाओं के घटित होने के बारे में कोई सूचना न होने पर इनको समप्रायिक मानना युक्ति संगत होता है। अतः हम प्रत्येक मूल घटना की प्रायिकता समान ले लेते हैं।

क्योंकि $P(S) = \sum_{i=1}^{n} P(e_i) = 1$ होता है, इसलिए

हम $P(e_i) = \frac{1}{n}, (i = 1, 2 \ldots n)$ लें लेते है।

यदि किसी घटना A में m मूल घटनाएँ हैं, तो

$$P(A) = \left(\frac{1}{n}\right) + \left(\frac{1}{n}\right) + \ldots + \left(\frac{1}{n}\right), m \text{ बार}$$

$$= \frac{m}{n} = \frac{A \text{ में तत्त्वों की संख्या}}{\text{S में तत्त्वों की संख्या}}$$

जोकि प्रायिकता की चिरप्रतिष्ठित परिभाषा भी है।

2) मूल घटनाओं की प्रायिकता निर्धारित करने की दूसरी विधि में हम एक परीक्षण को बड़ी संख्या में बार-बार दोहराते हैं। यदि परीक्षणों की संख्या, n, काफी बड़ी हो तो मूल घटनाओं की तुलनात्मक बारंबारताएँ उनकी क्रमशः प्रायिकताओं के बराबर होती हैं। यह विधि (पूर्व विवेचित) प्रायिकता की सांख्यिकीय परिभाषा पर आधारित है।

3) परीक्षण करने वाले व्यक्ति द्वारा भी, उसके अनुभव तथा प्रत्याशा के अनुसार, विभिन्न मूल घटनाओं की प्रायिकता निर्धारित की जा सकती है। उदाहरण के लिए, कोई व्यक्ति आपको आज वर्षा की प्रायिकता को निश्चित करने के लिए कह सकता है। यदि वर्षा के दिन हैं तो आप वर्षा की अधिक प्रायिकता, मान लिया 0.8, निर्धारित करना चाहेंगे इत्यादि। यह विधि, विशेष रूप से, प्रबंधकों द्वारा व्यवसायों में लिए जाने वाले विभिन्न निर्णयों के लिए अत्यंत उपयोगी है।

बहुत सी व्यवहारिक परिस्थितियों में घटनाओं के संयोजन होते हैं जिनमें हमें, इन घटनाओं की, प्रायिकताओं का भी संयोजन करना पड़ता है। इस संदर्भ में, हम प्रायिकता के दो महत्वपूर्ण नियमों का विवेचन करेंगे।

13.3 प्रायिंकता के नियम

प्रायिकता के विभिन्न नियमों पर विचार करने से पहले हमें कुछ संकेतनों से परिचित होना आवश्यक है।

क) यदि A तथा B दो घटनाएँ हैं तो $P(A \cup B)$ या $P(A + B)$ का अर्थ A घटित होने या B घटित होने या दोनों का एक साथ घटित होने की प्रायिकता होता है। इसका अर्थ, दो घटनाओं, A तथा B, में से कम से कम एक के घटने की प्रायिकता भी होता है।

ख) या $P(A \cap B)$ या $P(AB)$ का अर्थ दोनों घटनाओं, A तथा B, के एक साथ घटने की प्रायिकता, होता है।

ग) $P(A/B)$ का अर्थ, A के घटित होने की सप्रतिबंध प्रायिकताए जबकि B पहले ही घटित हो चुका है, होता है।

13.3.1 योग नियम

इन नियम के अनुसार दो घटनाओं में से कम से कम एक (अर्थात A या B या दोनों) के घटित होने की प्रायिकता, A की प्रायिकता धन B की प्रायिकता ऋण A तथा B के एक साथ घटने की प्रायिकता, होती है। संकेतनों के उपयोग द्वारा

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad ...(13.1)$$

कुछ विशेष परिस्थितियों में (13.1) में संशोधनों का विवेचन निम्नलिखित है :

क) **परस्पर अपवर्जी घटनाएँ** : मान लिया A तथा B परस्पर अपवर्जी हैं, अर्थात A के घटित होने पर B घटित नहीं होता तथा विलोमतः B के घटित होने पर A घटित नही होता। इस परिस्थिति में $P(A \cap B) = 0$ होता है। अतः

$$P(A \cup B) = P(A) + P(B)$$

ख) **निःशेष घटनाएँ** : यदि, किसी परीक्षण के परिणाम केवल A तथा B घटनाएँ ही है तब A या B या दोनों का घटना निश्चित होता हैं। हम यह जानते हैं कि एक निश्चित घटना के घटने की प्रायिकता 1 होती है। अतः

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 1 \text{ या}$$
$$P(A \cup B) = P(A) + P(B) = 1 \text{ (जब } A \text{ तथा } B \text{ परस्पर अपवर्जी हों)}$$

ग) **पूरक घटनाएँ** (complementary events) : किसी परीक्षण की, यदि A एक घटना है तो स्पष्टतः परीक्षण के दो परिणाम होंगे : A का घटित होना या A का घटित न होना। इस प्रकार A तथा 'A नहीं' दोनों घटनाएँ निःशेष हैं। हम 'A नहीं' घटना $\bar{A}$ को से सूचित करते हैं। अतः

$$P(A) + P(\bar{A}) = 1$$
या $$P(\bar{A}) = 1 - P(A)$$

यहाँ पर A तथा $\bar{A}$ एक दूसरे की पूरक घटनाएँ कहलाती हैं। अतः दो पूरक घटनाओं की प्रायिकताओं का योग इकाई के बराबर होता है।

13.3.2 गुणन नियम

इस नियम के अनुसार दो घटनाओं, A तथा B, के एक साथ घ्टने की प्रायिकता

i) A की प्रायिकता तथा B की सप्रतिबंध प्रायिकता (conditional poobability) जबकि A पहले ही घटित हो चुका है, का गुणनफल होता है।

या

ii) B की प्रायिकता तथा A की सप्रतिबंध प्रायिकता जबकि B पहले ही घटित हो चुका है, का गुणनफल होता है।

संकेतनों के प्रयोग द्वारा उपयोग

$$P(A \cap B) = P(A) \,.\, P(B/A) = P(B) \,.\, P(A/B) \qquad ...(13.2)$$

गुणन नियम के उपयोग द्वारा हम सप्रतिबंध प्रायिकताएँ ज्ञात कर सकते है :

$$P(B/A) = \frac{P(A \cap B)}{P(A)}$$

तथा

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

जब घटनाएँ स्वतंत्र हों तो गुणन नियम में संशोधन की आवश्यकता होती है।

मान लिया B का घटित होना A के घटित होने पर निर्भर नहीं है तब A तथा B घटनाओं को परस्पर स्वतंत्र कहा जाता है। इस परिस्थिति में

$P(B/A) = P(B)$ तथा

$P(A/B) = P(A)$ होता है।

अतः जब A तथा B स्वतंत्र हों तो

$$P(AB) = P(A) \,.\, P(B) = P(B) \,.\, P(A) \qquad ...(13.3)$$

अब हम प्रायिकता नियमों के अनुप्रयोगों के कुछ उदाहरणों पर विचार करेंगे।

13.3.3 प्रायिकता नियमों के अनुप्रयोग

ऊपर दिए गए प्रायिकता नियमों के अनुप्रयोगों को समझने के लिए हम कुछ प्रश्नों को हल करेंगे।

उदाहरण 13.4

एक पासे को उछाला गया। 1 या 6 प्राप्त करने की प्रायिकता क्या है?

स्पष्टतः, एक उछाल में दो घटनाएँ, 1 तथा 6, एक साथ घटित नहीं हो सकती। अतः ये दोनों घटनाएँ परस्पर अपवर्जी हैं।

$$P\,(1 \text{ या } 6) = P(1) + P(6) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

उदाहरण 13.5

52 पतों की गड्डी में से एक ताश का पत्ता निकाला गया। इस पत्ते को प्रतिस्थापित किए बिना एक और पत्ता निकाला गया। दोनों पत्ते हुक्म के होने की प्रायिकता क्या है?

यहाँ पर पहली घटना हुक्म का एक पत्ता प्राप्त करना तथा दूसरी घटना हुक्म का एक और पत्ता प्राप्त करना है। इस प्रकार दूसरी घटना एक सप्रतिबंध घटना है। मान लिया प्रथम घटना की प्रायिकता $P(A)$ है तथा दूसरी घटना की प्रायिकता, जबकि पहली घटित हो चुकी है, $P(B/A)$ है।

$$P(A) = \frac{13}{52} = \frac{1}{4}$$

जब प्रथम पत्ता हुक्म का है, जिसको प्रतिस्थापित नहीं किया गया, तब गड्डी में 51 पत्ते शेष हैं जिनमें से 12 पत्ते हुक्म के हैं। इस प्रकार

$$P(B/A)=\frac{12}{51}$$

अतः अपेक्षित प्रायिकता

$$P(A\cap B)=P(A).P(B/A)=\frac{13}{52}\times\frac{12}{51}=\frac{1}{4}\times\frac{12}{51}=\frac{3}{51}$$

हमें इस बात पर ध्यान देना चाहिए कि जब पहला पत्ता निकाल कर प्रतिस्थापित नहीं किया जाता तो दो घटनाएँ स्वतंत्र नहीं रहती क्योंकि दूसरी घटना के घटित होने की प्रायिकता पहली घटना के घटित होने की प्रायिकता पर निर्भर होती है।

उदाहरण 13.6

ताश के 52 पत्तों में से एक पत्ता निकाला गया। इस पत्ते को गड्डी में प्रतिस्थापित कर दिया गया तथा एक और पत्ता निकाला गया। दोनों पत्ते हुक्म के होने की प्रायिकता क्या है?

इस उदाहरण में पहला पत्ता निकालने के बाद प्रतिस्थापित कर दिया जाता है। इस प्रकार, जब दूसरा पत्ता निकाला जाता है तब भी 52 पत्तों की गड्डी में 13 हुक्म के पत्ते हैं। इसके परिणामस्वरूप, दूसरा पत्ता हुक्म का होने की प्रायिकता पहले पत्ते के परिणाम से स्वतंत्र है। अतः दोनों घटनाएँ स्वतंत्र हैं। यदि दूसरा पत्ता हुक्म का होने की प्रायिकता $P(B)$ है तब

$$P(B/A)=P(B)=\frac{13}{52}=\frac{1}{4}$$

अतः अपेक्षित प्रायिकता

$$P(A\cap B)=P(A).P(B/A)=P(A).P(B)=\frac{13}{52}\times\frac{13}{52}=\frac{1}{4}\times\frac{1}{4}=\frac{1}{16}$$

उदाहरण 13.7

एक पासे को उछाला गया। 5 से कम संख्या या विषम संख्या प्राप्त करने की प्रायिकता क्या है?

मान लिया, 5 से कम संख्या प्राप्त करने की घटना A है तथा विषम संख्या प्राप्त करने की घटना B है। यह ध्यान दें कि ये दोनों घटनाएँ परस्पर अपवर्जी नहीं है क्योंकि उछाल का परिणाम 5 से कम तथा विषम, दोनों, हो सकता है। इस प्रकार अपेक्षित प्रायिकता निम्नलिखित सूत्र के अनुप्रयोग द्वारा ज्ञात की जा सकती है।

$$P(A\cup B)=P(A)+P(B)-P(A\cap B)$$

एक पासे में, 6 संख्याओं में से 4 संख्याएँ (1, 2, 3 तथा 4) 5 से कम होती हैं। अतः

$$P(A)=\frac{4}{6}=\frac{2}{3}$$

इसके अतिरिक्त, 6 संख्याओं में 3 विषम संख्याएँ (1, 3 तथा 5) होती हैं। अतः

$$P(B)=\frac{3}{6}=\frac{1}{2}$$

मान लिया, एक विषम संख्या की प्रायिकता, जबकि वह 5 से कम दी हुई है, $P(B/A)$ है। तब

$$P(B/A)=\frac{2}{4}=\frac{1}{2}$$

इस प्रकार

$$P(A\cap B)=P(A).P(B/A)=\frac{2}{3}\times\frac{1}{2}=\frac{1}{3}$$

अतः, 5 से कम या विषम संख्या प्राप्त करने की प्रायिकता

$$P(A\cup B)=\frac{2}{3}+\frac{1}{2}-\frac{1}{3}=\frac{5}{6}$$

उदाहरण 13.8

A तथा B दो ऐसी घटनाएँ हैं कि $P(A)=\frac{2}{3}, P(\bar{A}\cap B)=\frac{1}{6}$ तथा $P(A\cap B)=\frac{1}{3}$ है। ,

$P(B)$, $P(A\cup B), P(A/B), P(B/A), P(\bar{A}\cup B)$ तथा $P(\bar{A}\cap\bar{B})$ ज्ञात कीजिए।

यह भी जाँच कीजिए कि क्या A तथा B

क) समप्रायिक हैं

ख) निःशेष हैं

ग) परस्पर अपवर्जी हैं

घ) स्वतंत्र हैं

हम यह लिख सकते हैं

$$P(B)=P(\bar{A}\cap B)+(A\cap B)=\frac{1}{6}+\frac{1}{3}=\frac{3}{6}=\frac{1}{2}$$

$$P(A\cup B)=P(A)+P(B)-P(A\cap B)=\frac{2}{3}+\frac{1}{2}-\frac{1}{3}=\frac{5}{6}$$

$$P(A\cup B)=\frac{P(A\cap B)}{P(B)}=\frac{1}{3}\times 2=\frac{2}{3}$$

$$P(B/A)=\frac{P(A\cap B)}{P(A)}=\frac{1}{3}\times\frac{3}{2}=\frac{1}{2}$$

$$P(\bar{A}\cup B)=P(\bar{A})+P(B)-P(\bar{A}\cap B)=\frac{1}{3}+\frac{1}{2}-\frac{1}{6}=\frac{2}{3}\quad\left[\because P(\bar{A})=1-P(A)\right]$$

$P(\bar{A}\cap\bar{B})=-P(A\cup B)=1-\frac{5}{6}=\frac{1}{6}$ (पूरक घटनाओं की अवधरणा के उपयोग द्वारा)

क) क्योंकि $P(A)\neq P(B)$, A तथा B समप्रायिक नहीं हैं।

ख) क्योंकि $P(A \cup B) \neq 1$, A तथा B निःशेष नहीं हैं।

ग) क्योंकि $P(A \cap B) \neq 0$, A तथा B परस्पर अपवर्जी नहीं हैं।

घ) क्योंकि $P(A) \cdot P(B) = P(A \cap B)$, A तथा B स्वतंत्र हैं।

बोध प्रश्न 2

1) एक विद्यार्थी गणित तथा अंग्रेजी की परीक्षाएँ देता है। उसके दोनों परीक्षाओं में उत्तीर्ण होने की स्वतंत्र प्रायिकताएँ क्रमशः $\frac{2}{3}$ तथा $\frac{3}{4}$ हैं।

   क) कम से कम एक परीक्षा में उत्तीर्ण होने की प्रायिकता क्या है?

   ख) दोनों परीक्षाओं में अनुत्तीर्ण होने की प्रायिकता क्या हैं?

....

....

....

....

2) एक ताश की गड्डी में से दो पत्ते निकाले गए।

   क) दोनों के बादशाह होने की प्रायिकताा क्या है, जबकि द्वितीय पत्ता निकालने से पहले प्रथम को प्रतिस्थापित किया जाए?

   ख) दोनों के हुक्म होने की प्रायिकता क्या है, जबकि द्वितीय पत्ता निकालने से पहले प्रथम को प्रतिस्थापित न किया जाय ?

....

....

....

....

3) दो घड़े हैं। प्रथम में 7 सफेद गेंद तथा 3 लाल गैंद हैं। द्वितीय घड़े में 4 सफेद गैंद तथा 6 लाल गैंद हैं। यादृच्छिक विधि से एक घड़े का चयन करके उसमें से एक गैंद निकाली गई। प्रथम घड़ा चयन होने तथा लाल गैंद प्राप्त करने की प्रायिकता क्या है?

....

....

....

....

4) A तथा B के स्वतंत्र रूप से सच बोलने की प्रायिकताएँ क्रमशः $\frac{1}{2}$ तथा $\frac{1}{3}$ हैं। यदि वे एक ही कथन करते हैं तो उनके द्वारा सत्य कथन की प्रायिकता क्या है?

....

....

....

....

13.4 बेज-प्रमेय

मान लिया A_1, A_2, तथा A_3 तीन प्रकार अपवर्जी तथा निःशेष घटनाएँ हैं तथा D एक घटना है जो इनमें से किसी के साथ घटित हो सकती हैं। यदि, वास्तव में D घटित हो जाती है, तब $A_i (i = 1, 2, 3)$ के घटित होने की सप्रतिबंध प्रायिकता, जबकि D घटित हो चुकी है, निम्नलिखित हैं

$$P(A_i / D) = \frac{P(A_i \cap D)}{P(D)} = \frac{P(A_i) \cdot P(D / A_i)}{P(D)}$$

जहाँ पर

$$P(D) = \sum_{i=1}^{3} P(A_i \cap D) = \sum_{i=1}^{3} P(A_i) \cdot P(D / A_i)$$

यह परिणाम कितनी ही परस्पर अपवर्जी तथा निःशेष घटनाओं के लिए व्यापक बनाया जा सकता है।

अब हम बेज-प्रमेय (Bayes theorem) के कुछ व्यवहारिक अनुप्रयोगों पर विचार करते हैं।

उदाहरण 13.9

एक काबले बनाने वाली फैक्ट्री में तीन मशीनें A, B तथा C हैं। ये कुल उत्पादन का क्रमशः 25%, 35% तथा 40% उत्पादन करती हैं। लेकिन इनके उत्पादन के क्रमशः 5%, 4% तथा 2% काबले दोषपूर्ण होते हैं। एक दिन के उत्पादन में से एक काबले का चयन किया गया तथा वह दोषपूर्ण पाया गया। इस काबले के (i) मशीन A, (ii) मशीन B, तथा (iii) मशीन C द्वारा निर्मित होने की प्रायिकता क्या है?

क्योंकि एक दिन के उत्पादन में सभी तीन मशीनों का उत्पादन सम्मिलित है, इसलिए, मशीन A द्वारा इसके निर्माण की प्रायिकता, $P(A) = 0.25$ है। इसी प्रकार $P(B) = 0.35$ तथा $P(C) = 0.40$ होगा। इसके अतिरिक्त, मान लीजिए, काबले के दोषपूर्ण होने की घटना को D से सूचित करते हैं। क्योंकि मशीन A द्वारा 5: दोषपूर्ण काबले निर्मित होते हैं, इसलिए $P(D / A) = 0.05$ है। इसी प्रकार, $P(D / B) = 0.04$ तथा $P(D / C) = 0.02$ होगा।

इस प्रकार

$$P(D) = P(A) \cdot P(D / A) + P(B) \cdot P(D / B) + P(C) \cdot P(D / C)$$
$$= 0.25 \times 0.05 + 0.35 \times 0.04 + 0.40 \times 0.02 = 0.0345$$

मशीन A द्वारा काबला निर्मित होने की प्रायिकता जबकि यह दोषपूर्ण है, अर्थात

$$P(A / D) = \frac{P(A) \cdot P(D / A)}{P(D)} = \frac{0.25 \times 0.05}{0.0345} = \frac{0.0125}{0.0345} = 0.362$$

इसी प्रकार

$$P(B / D) = \frac{P(B) \cdot P(D / B)}{P(D)} = \frac{0.35 \times 0.04}{0.0345} = \frac{0.0140}{0.0345} = 0.406$$

तथा

$$P(C / D) = \frac{P(C) \cdot P(D / C)}{P(D)} = \frac{0.40 \times 0.02}{0.0345} = \frac{0.0080}{0.0345} = 0.232$$

उपरोक्त विधि की वैकल्पिक विधि भी अपनाई जा सकती है। ऊपर दिये गये प्रायिकताओं को निम्नलिखित सारणी की रचना द्वारा भी प्राप्त किया जा सकता है। तीन घटनाओं, A, B तथा C को क्रमशः A_1, A_2, तथा A_3 से सूचित किया गया है।

A_i	A_1	A_2	A_3	योग
$P(A_i)$	0.25	0.35	0.45	1.00
$P(D/A_i)$	0.05	0.04	0.02	
$P(D\cap A_i)$	0.0125	0.014	0.008	$P(D) = 0.0345$
$P(A_i/D) = \frac{P(D\cap A_i)}{P(D)}$	0.362	0.406	0.232	1.00

यह ध्यान दीजिए कि प्रायिकताएँ $P(A_1), P(A_2)$ तथा $P(A_3)$, जिनके मान परीक्षण से पूर्व ज्ञात हैं, को *पूर्ववर्त्ती प्रायिकताएँ* (prior probabilities) कहते हैं। A_1, A_2, तथा A_3 की सप्रतिबंध प्रायिकताएँ, अर्थात $P(A_1/D)$, $P(A_2/D)$ तथा $P(A_3/D)$, परीक्षण का परिणाम ज्ञात होने पर प्राप्त प्रायिकताएँ, *परवर्त्ती प्रायिकताएँ* (posterior probabilities) कहलाती हैं।

किसी समस्या का विश्लेषण करने से पूर्व, एक फर्म का प्रबंधक कुछ घटनाओं की प्रायिकता व्यक्तिपरक (subjective) आधार पर निर्धारित कर सकता है जोकि उसके अनुभव तथा प्रत्याशा पर निर्भर होती है। इन प्रायिकताओं को पूर्ववर्त्ती प्रायिकताएँ कहते हैं। इसके बाद परीक्षण किया जाता है तथा D जैसी घटना के घटित होने के आधार पर पूर्ववर्त्ती प्रायिकताओं का संशोधन किया जाता है। इन संशोधित प्रायिकताओं को परवर्त्ती प्रायिकताएँ कहते है। दूसरे दौर में इन परवर्त्ती प्रायिकताओं को पूर्ववर्त्ती प्रायिकताएँ मान कर तथा उसी प्रकिया को दोहरा कर, परवर्त्ती प्रायिकताएँ प्राप्त की जा सकती हैं। इस प्रकार के कुछ संशोधनों के बाद परवर्त्ती प्रायिकताओं में स्थिर होने की प्रवृति होती है। इस प्रकार व्यक्तिपरक प्रायिकताएँ लगभग वस्तुपरक प्रायिकताएँ बन जाती हैं। व्यवसाय गतिविधियों के विश्लेषण के लिए बेज-प्रमेय बहुत ही उपयोगी है।

उदाहरण 13.10

एक कम्पनी के उत्पाद के सफल होने की प्रायिकता, जबकि सर्वेक्षण का परिणाम अनुकूल है, 0.6 है तथा सफल होने की प्रायिकता, जबकि सर्वेक्षण का परिणाम प्रतिकुल है, 0.3 है। यदि सर्वेक्षण के अनुकूल परिणाम दर्शाने की प्रायिकता 0.7 है, i) उत्पाद सफल होने की प्रायिकता ज्ञात कीजिए, ii) सर्वेक्षण का परिणाम सफल होने की प्रायिकता, जबकि उत्पाद सफल हों, ज्ञात कीजिए, तथा iii) सर्वेक्षण का परिणाम असफल होने की प्रायिकता, जबकि उत्पाद सफल हो, ज्ञात कीजिए।

मान्लिया कम्पनी का उत्पाद सफल होने की घटना को S से तथा सर्वेक्षण का परिणाम अनुकूल होने की घटना को F से सूचित करते हैं। इनकी क्रमशः विपरीत घटनाएँ $\bar{S}$ तथा $\bar{F}$ से सूचित की जाती हैं।

संकेतनों के द्वारा, दी हुई सूचना को हम इस प्रकार लिख सकते हैं।

$P(S/F) = 0.6, P(S/\bar{F}) = 0.3$ तथा $P(F) = 0.7$

इस प्रकार $P(\overline{F}) = 1 - 0.7 = 0.3$

i) उत्पाद के सफल होने की प्रायिकता

$$\begin{aligned} P(S) &= P(S \cap F) + P(S \cap \overline{F}) \\ &= P(F) . P(S/F) + P(\overline{F}) . P(S/\overline{F}) \\ &= 0.7 \times 0.6 + 0.3 \times 0.3 = 0.51 \end{aligned}$$

ii) $P(F/S) = \dfrac{P(F \cap S)}{P(S)} = \dfrac{0.42}{0.51} = 0.824$

iii) $P(\overline{F}/S) = \dfrac{P(\overline{F} \cap S)}{P(S)} = \dfrac{0.09}{0.51} = 0.176$

ध्यान दीजिए कि $P(\overline{F}/S) = 1 - P(F/S)$

बोध प्रश्न 3

1) एक टैल्कम पाउडर बनाने वाली कम्पनी ने एक नए प्रकार का विज्ञापन दिया। कम्पनी द्वारा यह आकलन किया गया कि एक व्यक्ति, जिसने वह विज्ञापन देखा है, द्वारा उनका उत्पाद खरीदने की प्रायिकता 0.7 है तथा जिसने वह विज्ञापन नही देखा, द्वारा उनका उत्पाद खरीदने की प्रायिकता 0.3 है। यदि 1000 व्यक्तियों वाले क्षेत्र में, 70% व्यक्तियों ने उस विज्ञापन को देखा है, तो एक व्यक्ति जिसने उत्पाद खरीदा है (क) के विज्ञापन न देखने, तथा (ख) विज्ञापन देखने की प्रायिकताएं क्या हैं?

....

....

....

....

....

....

....

....

2) एक बीमा कम्पनी ने 2000 स्कूटर चालकों, 4000 कार चालकों तथा 6000 ट्रक चालकों का बीमा किया। क्रमशः वर्ग में दुर्घटना की प्रायिकता $0.01, 0.03$ तथा 0.15 हैं। एक बीमाकृत चालक दुर्घटना ग्रस्त हो जाता है। इसके स्कूटर चालक होने की प्रायिकता क्या होगी?

....

....

....

....

....

....

....

....

....

13.5 सारांश

साधारण भाषा में, प्रायिकता का अर्थ संभावना से होता है। लेकिन, सांख्यिकी में इसका अर्थ कुछ अधिक होता है। यहाँ पर हम, केवल एक घटना के घटित होने की अनिश्चितता पर ही विचार नहीं करते बल्कि इसका संख्यात्मक मान प्रदान करने का प्रयास भी करते हैं। इस प्रकार, प्रायिकता, संभावना का परिमाणात्मक माप है। इस इकाई में हमने प्रायिकता की तीन अभिगमों, अर्थात चिरप्रतिष्ठित, तुलनात्मक बारंबारता तथा अभिगृहीतीय, का अध्ययन किया है। मिश्रित घटनाओं की प्रायिकताएँ, दो नियमों पर आधारित हैं। ये योग नियम तथा गुणन नियम हैं। अंततः बेज-प्रमेय द्वारा, घटनाओं के घटित होने या न घटित होने के आधार पर प्रायिकताओं के संशोधन का ढाँचा, प्रदान किया जाता है। इस प्रकार का संशोधन व्यवसायिक निर्णयों में अत्यंत उपयोगी होता है।

13.6 शब्दावली

प्रायिकता : यह एक घटना के सहचारी निश्चितता (तथा अप्रत्यक्ष रूप से अनिश्चितता) की कोटि का तुलनात्मक परिमाप होता है। एक घटना A के लिए $0 \le P(A) \le 1$ प्रायिकता होती है।

सप्रतिबंध प्रायिकता : यदि A तथा B दो परस्पर अपवर्जी घटनाएँ नहीं हैं तो B की प्रायिकता, जबकि A पहले ही घटित हो चुका है, को B की सप्रतिबंध प्रायिकता कहा जाता है। इस प्रायिकता को $P(B/A)$ से सूचित किया जाता है।

स्वतंत्र घटनाएँ : दो घटनाएँ A तथा B के लिए, यदि B का घटित होना A के घटित होने पर निर्भर नहीं होता तथा इसका विपरीत भी सत्य हो, तो ये परस्पर स्वतंत्र कहलाती हैं।

पूरक घटना : : यदि A एक घटना है तब इसके घटित न होने की घटना, जिसको $\overline{A}$ से सूचित करते हैं, को A की पूरक घटना कहते हैं। एक घटना तथा इसकी पूरक घटना की प्रायिकताओं का योग सदैव इकाई के बराबर होता है।

पूर्ववर्त्ती प्रायिकता : चिरप्रतिष्ठित परिभाषा, सांख्यिकीय परिभाषा या व्यक्तिपरक आधार पर निर्धारित, विभिन्न घटनाओं की प्रायिकताएँ पूर्ववर्त्ती प्रायिकताएँ होती हैं।

परवर्त्ती प्रायिकता : विभिन्न घटनाओं की संशोधित प्रायिकताओं को परवर्त्ती प्रायिकताएँ कहते हैं। यह संशोधन बेज-प्रमेय के उपयोग द्वारा किया जाता है।

13.7 कुछ उपयोगी पुस्तकें

New bold, p., 1991, *Statistics for Business and Economics* (Third Edition) Prentice Hall, New Jersey, Chapter 3.

Nagar, A.L. and Das, R.K., 1989, *Basic Statistics* : Oxford University press, Delhi, Chapter 8.

Anderson, D.R. Seeney, D.J. , and Williams, T.A. , 1993, *Statistics for Business and Economics* (Fifth Edition) : West Publishing Company, Minnaeapolos / St. Pawl, Chapter 4.

13.8 बोध प्रश्नों के उत्तर या संकेत

बोध प्रश्न 1

1) 2/5

2) $\frac{1}{2}$

3) $\frac{1}{2}$

4) $\frac{12}{13}$

बोध प्रश्न 2

1) क) $\frac{11}{12}$, ख) $\frac{1}{12}$

2) क) $\frac{1}{169}$, ख) $\frac{1}{17}$

3) $\frac{3}{20}$

4) $\frac{1}{3}$

बोध प्रश्न 3

1) $\frac{9}{58}, \frac{49}{58}$

2) $\frac{1}{52}$

13.9 पारिभाषिक शब्दावली

अनुमिति	:	inference
अभिगृहीतीय अभिगम	:	axiomatic approach
गणितीय प्रत्याशा	:	mathmetical expectation

चिरप्रतिष्ठित	:	classical
निःशेष घटनाएँ	:	exhaustive events
अनुकूल घटनाएँ	:	favourable events
समप्रायिक घटनाएँ	:	equally likely events
परस्पर अपवर्जी घटनाएँ	:	mutually exclusive events
मूल घटना	:	elementary event
प्रतिदर्श क्षेत्र	:	sample space
पूर्ववर्त्ती प्रायिकता	:	prior probability
परवर्त्ती प्रायिकता	:	posterior probability
निदर्शन बंटन	:	sampling distribution

इकाई 14 प्रायिकता बंटन – I

इकाई की रूपरेखा

14.0 उद्देश्य

इस इकाई को पढ़ने के बाद, आप:

- यादृच्छिक चर के अर्थ को व्यक्त कर सकेंगे;
- प्रायिकता बंटन की संकल्पना की व्याख्या कर सकेंगे;
- असतत् प्रायिकता बंटन और सतत् प्रायिकता बंटन के बीच के अंतर को स्पष्ट कर सकेंगे; और
- द्विपद और पाइसों बंटन की प्रमुख विशेषताओं का वर्णन कर सकेंगे।

14.1 प्रस्तावना

इकाई 13 में हमने किसी घटना के घटित होने की प्रायिकता के बारे में चर्चा की थी। उस इकाई में घटना को संयोगी प्रयोग (chance experiment) के एक या अधिक संभावित परिणामों के समुच्चय के रूप में परिभाषित किया गया था। ऐसे संयोगी प्रयोग के परिणामों को 'यादृच्छिक चर' की संकल्पना से जोड़ा जा सकता है। इस इकाई में हम यादृच्छिक

चर को ध्यान में रखकर प्रायिकता पर विचार करेंगे। इसके बाद हम प्रायिकता बंटन (poobability distribution) की अब धारणा पर चर्चा करेंगे।

इस इकाई में हम असतत् यादृच्छिक चर और सतत् यादृच्छिक चर के बीच के अंतर को स्पष्ट करेंगे। इसके बाद, असतत् यादृच्छिक चरों के संदर्भ में हम दो महत्वपूर्ण असतत् प्रायिकता बंटनों अर्थात द्विपद बंटन और पाइसों बंटन की चर्चा करेंगे। सतत् प्रायिकता बंटन की संकल्पना की चर्चा हम अगली इकाई में करेंगे।

14.2 यादृच्छिक चर

यादृच्छिक चर (random variable) की औपचारिक परिभाषा करने से इस पहले, आइए सहज रूप से इस संकल्पना के अर्थ को समझने का प्रयास करें। हम जानते ही है कि यादृच्छिक चर, संयोगी प्रयोग के परिणामों से संबंधित है। ऐसे संयोगी प्रयोग को यादृच्छिक प्रयोग भी कहते हैं। आइए ऐसे एक उदाहरण पर विचार करें।

मान लीजिए हमने एक सिक्का उछाला इसके दो संभावित परिणाम हैं : शीर्ष (Head) या पुच्छ (Tail)। इससे पहली इकाई में हमने प्रतिदर्श समष्टि (Sample space) की संकल्पना की चर्चा की थी। इस प्रयोग की प्रतिदर्श समष्टि में शीर्ष और पुच्छ परिणाम शामिल हैं। यदि S, प्रतिदर्श समष्टि को व्यक्त करता है, तब

$$S = (H, T)$$

इस प्रयोग में यह सुनिश्चित नहीं है कि परिणाम के रूप में शीर्ष आएगा या पुच्छ। यह संयोगी प्रयोग या यादृच्छिक प्रयोग का एक उदाहरण है। अब मान लीजिए, हम पुच्छ (T) घटित होने को 0 संख्या और शीर्ष (H) के घटित होने को 1 संख्या द्वारा अभिव्यक्त करते हैं। आइए अब X चर को परिभाषित करें जो किसी परिणाम के घटित होने को व्यक्त करता है। तब चर और इसके संभावित मान हैं :

$$X = (0, 1)$$

लेकिन, इस चर और चर संबंधी हमारी सामान्य धारणा के बीच एक महत्वपूर्ण अंतर है (विभिन्न प्रकार के चरों के लिए, इकाई 7 देखें)। यहाँ, वह मान; जिसे चर प्राप्त करेगा, ऐसे संयोग या यादृच्छिक प्रयोग के परिणाम पर निर्भर है जिस पर हम विचार कर रहे हैं। अन्य शब्दों में, हमें पक्का पता नहीं है कि प्रयोग के परिणाम के रूप में क्या चर 0 मान धारण करेगा अथवा इसका मान 1 होगा। हम तो केवल इन मानों से कुछ प्रायिकताओं को जोड़ सकते हैं। ये प्रायिकताएँ, प्रयोग के विविध परिणामों के घटित होने के संयोग पर निर्भर करती हैं। यदि अपने उदाहरण में हम मान लें कि सिक्का अनभिनत हैं। तब पुच्छ आने की प्रायिकता $\frac{1}{2}$ है, और शीर्ष के लिए भी प्रायिकता $\frac{1}{2}$ है; (क्योंकि दोनों परिणामों के आने की संभावना एक जैसी है)। इसी आधार पर, हम $\frac{1}{2}$ की प्रायिकता को 0 और 1 अर्थात दोनों से जोड़ते हैं। चर की रूढ़ धारणा के मामले में, दूसरी तरफ, ऐसी किसी प्रायिकता को चर के किसी मान से नहीं जोड़ा जा सकता।

उपर्युक्त चर्चा के आधार पर हम कह सकते हैं कि यादृच्छिक चर ऐसा चर है जो *कुछ प्रायिकताओं से अलग-अलग मान ले सकता है।* अतः सिक्के को उछालने के संभावित परिणामों को दर्शाने वाला चर X, यादृच्छिक चर का उदाहरण है।

हम निम्नलिखित संकेत का प्रयोग करेंगे : मान लीजिए X यादृच्छिक चर है और इसके $x_1, x, x_3, \ldots x_2$, मान हैं। इस स्थिति में संगत प्रायिकताएं $p_1, p_2, p_3, \ldots\ldots p_n$ होंगी। अत: $p(X = x_1) = P_1$

उदाहरण 14.1 : मान लीजिए किन्हीं दो सिक्कों को एक साथ उछाला जाता है। प्रयोग की प्रतिदर्श समष्टि है :

$$S = \{(H, H), (H, T), (T, H), (T, T)\}$$

यदि हम यादृच्छिक चर X को प्राप्त शीर्षों की संख्या के रूप में परिभाषित करें तब X=2, (H, H) परिणाम के तदनुरूप है; X= 1, (H, T) और (T, H) परिणामों को तदनुरूप है, और अंतत: $X = 0$ परिणाम (T, T) के तदनुरूप है। अत: X के तीन संभावित मान अर्थात 0, 1 और 2 हो सकते हैं।

$$X = (0, 1, 2)$$

उदाहरण 14.2: एक अन्य उदाहरण में हम एक पाँसे को फैंकते हैं। प्रतिदर्श समष्टि है, $S = \{1, 2, 3, 4, 5, 6\}$। यादृच्छिक चर X को इस प्रकार भी परिभाषित कर सकते हैं कि यह 0 के बराबर के मान लें जब पाँसे पर विषम संख्या आती है और 1 लें जब कोई सम संख्या आती है। अत:

$$X = (0, 1)$$

दो सिक्कों को उछालने वाले प्रयोग में, हम यादृच्छिक चर को रुपयों आदि के रूप में भी परिभाषित कर सकते हैं। जैसे, यदि हमें दो शीर्षों की प्राप्ति होती है तो हम खिलाड़ी को 10 रुपए देना तय करते हैं और यदि एक शीर्ष की प्राप्ति होती है तो 5 रुपए और यदि कोई भी शीर्ष नही आता तो खिलाड़ी को (–)8 रुपए अदा करते (अर्थात वह हमें 8 रुपए देगा)। यहाँ X ऐसा यादृच्छिक चर है जो ऐसे भुगतान को व्यक्त करता है जिसे खिलाड़ी को अदा किया जाना है। अत:

$$X = (10, 5, -8)$$

जैसा कि खंड 2 में हमने अध्ययन किया था, चर असतत् या सतत् हो सकता है। इसी तरह से, यादृच्छिक चर भी असतत् या सतत् हो सकता है।

i) **असतत् यादृच्छिक चर** : जब किसी प्रयोग की प्रतिदर्श समष्टि असतत् है तो संगत यादृच्छिक संख्याएं भी असतत् होगी अर्थात निश्चित रूप से उसके कुछ वियुक्त (isolated) मान होंगे। उपर्युक्त उल्लिखित यादृच्छिक चर, असतत् यादृच्छिक चर (discrete random variable) के उदाहरण हैं।

ii) **सतत् यादृच्छिक चर**: जैसा कि हम जानते हैं, सतत् चर का अंतराल में कोई भी मान हो सकता है। इसी तरह से यादृच्छिक चर सतत् होगा जब प्रतिदर्श समष्टि भी सतत् हो। अगली इकाई में हम सामान्य चर की संकल्पना चर्चा करेंगे जो सतत् यादृच्छिक चर (continuous random variable) का उदाहरण है।

14.3 प्रायिकता बंटन

आइए प्रायिकता बंटन (probability distribution) की परिभाषा की बात शुरू करें। इसे, ऐसे प्रकथन (statement) के रूप में परिभाषित किया जाता है जो संबंधित प्रायिकताओं के साथ यादृच्छिक चर के संभावित मानों पर आधारित है।

आइए प्रायिकता बंटन का एक उदाहरण लें। उपर्युक्त उदाहरण 14.1 में हमने दो सिक्कों को उछाला और यादृच्छिक चर X को शीर्षों की संख्या के रूप में परिभाषित किया और इसके आगे X ने तीन मान अर्थात 0, 1, और 2 धारण किए। मान लीजिए कि दोनों सिक्के अनभिनत हैं, तब हम लिख सकते हैं :

$$p(X=0) = \frac{1}{4}$$

$$p(X=1)\ \frac{1}{2}$$ [अर्थात, (H, T) या (T, H) आने की प्रायिकता]

$$P(X=2)\ \frac{1}{4}$$

सारणीबद्ध तरीके से लिखी गई ये प्रायिकताएँ और इनके साथ यादृच्छिक चर के तदनुरूप मान, *यादृच्छिक चर X के प्रायिकता बंटन* की रचना करते हैं जहाँ X शीर्षों की संख्या है। इसे सारणी 14.1 में दर्शाया गया है।

सारणी 14.1: दो अनभिनत सिक्कों को उछालने से प्राप्त शीर्षों की संख्या का प्रायिकता वितरण

शीर्षों की संख्या (x)	प्रायिकता $p(x)$
0	¼
1	½
2	¼

उपर्युक्त उदाहरण में शीर्ष प्राप्त न करना ($X=0$), एक शीर्ष प्राप्त करना ($X=1$) और दो शीर्ष प्राप्त करने ($X=2$) से संबंधित घटनाएं अन्य सभी संभावनाओं को नकार देती है (इसका अर्थ है कि उपर्युक्त तीन के अलावा, और कोई भी संभावित परिणाम नहीं हो सकता)। अतः उपर्युक्त प्रयोग से प्राप्त प्रायिकता बंटन ने यादृच्छिक चर X के सभी संभावित मानों की गणना कर ली है और उन्हें कुछ विशिष्ट प्रायिकताएँ दी हैं। हम देख सकते हैं कि इन प्रायिकताओं का योग 1 के बराबर है।

प्रायिकता बंटन दो प्रकार के हो सकते हैं: असतत् (discrete) प्रायिकता बंटन और सतत् (continuous) प्रायिकता बंटन।

14.3.1 असतत् प्रायिकता बंटन

हम पहले ही देख चुके हैं कि यादृच्छिक चर संबंधी प्रायिकता बंटने हमें बताता है कि किस प्रकार प्रायिकताओं को यादृच्छिक चर के मानों पर बंटित किया जाता है। अब, सतत् यादृच्छिक चर के लिए, प्रायक्ता बंटन को ऐसे फलन द्वारा परिभाषित किया जाता है जिसे *प्रायिकता द्रव्यमान फलन* (probability mass function) कहते हैं और इसे $p(x)$ द्वारा दर्शाया जाता है। यह प्रायिकता द्रव्यमान फलन, असतत् यादृच्छिक चर के प्रत्येक मान के लिए प्रायिकता प्रदान करता है। असल में, दो सिक्कों को उछालने पर आने वाले शीर्षों की संख्या का प्रायिकता बंटन जिसे हमने तालिका 14.1 में दर्शाया था, असतत् प्रायिकता बंटन (discrete probability distribution) का उदाहरण है।

आइए अब असतत् प्रायिकता बंटन के अन्य उदारहण पर विचार करें। मान लीजिए हम अपने इलाके के प्रति परिवार बच्चों की संख्या पर गौर करते हैं, यहाँ, हम बच्चों की संख्या को असतत् यादृच्छिक चर मान सकते हैं। प्रति परिवार में बच्चों की संख्या के लिए असतत् प्रायिकता वितरण का निर्माण, इस यादृच्छिक चर के संभावित मानों के लिए, सापेक्षिक बारंबारता (relative prequency) के अभिकलन द्वारा किया जा सकता है। ऐसे प्रायिकता बंटन को सारणी 14.2 में दर्शाया गया है।

सारणी 14.2: प्रति परिवार के आधार पर बच्चों की संख्या का प्रायिकता बंटन

बच्चों की संख्या (x)	$p(x)$
0	0.10
1	0.15
2	0.23
3	0.25
4	0.14
5	0.13

अतः क्रमबद्ध युग्म (ordered pairs) $[x, p(x)]$ का समुच्चय, असतत् यादृच्छिक चर X या असतत् प्रायिकता बंटन का प्रायिकता बंटन कहलाता है।

चूँकि मान $P(x)$, सभी प्रायिकताओं को दर्शाता है और यादृच्छिक चर का कोई न कोई मान x हमेशा होगा ही, इसलिए प्रायिकता द्रव्यमान फलन इन दो शर्तों को पूरा करेगा :

1) किसी भी घटना की प्रायिकता अर्थात X के किसी मान के लिए प्रायिकता नकारात्मक नहीं हो सकती।

$$P(x) \geq 0$$

2) सभी संभावित परिणामों की प्रायिकताओं का योग 1 के बराबर होता है, अर्थात

$$\sum_{\text{सभी } x} p(x) = 1$$

आइए असतत् प्रायिकता बंटन पर आधारित कुछ समस्याओं पर विचार करें।

उदाहरण 14.3

क्या निम्नलिखित एक मान्य प्रायिकता द्रव्यमान फलन हैं?

$$p(x) = \frac{x^3}{2}, x = -1, 0, 1$$

आइए x के कुछ विशिष्ट मानों ($-1, 0$ और 1) पर हम x की प्रायिकता का आकलन करें।

जब $x = -1$:

$$p(x) = p(-1) = -\frac{1^3}{2} = -\frac{1}{2} < 0$$

लेकिन हम जानते हैं कि किसी भी घटना की प्रायिकता नकारात्मक नहीं हो सकती। इसलिए, यहां प्रायिकता द्रव्यमान फलन की पहली शर्त का उल्लंघन हो रहा है। अतः दिया गया फलन, मान्य प्रायिकता द्रव्यमान फलन नहीं हो सकता।

उदाहरण 14.4

फलन $p(x) = \frac{k}{x}, x = 3, 4, 5$ में k स्थिरांक है। k का मान ज्ञात कीजिए ताकि ये फलन मान्य प्रायिकता द्रव्यमान फलन हो जाए।

इस फलन से, हम देख सकते हैं कि :

$$p(3) = \frac{k}{3}$$

$$p(4) = \frac{k}{4}$$

$$p(5) = \frac{k}{5}$$

प्रायिकता द्रव्यमान फलन की दूसरी शर्त को पूरा करने के लिए, आवश्यक है कि

$$\sum p(x) = \frac{k}{3} + \frac{k}{4} + \frac{k}{5} = 1$$

$$\text{या } k\left(\frac{1}{3} + \frac{1}{4} + \frac{1}{5}\right) = 1$$

$$\text{या } k\frac{47}{60} = 1$$

$$\text{या } k = \frac{60}{47}$$

आइए जाँच करें कि k का उपर्युक्त मान क्या पहली शर्त को पूरा करता है;

$$p(3) = \frac{1}{3} \times \frac{60}{47} = \frac{60}{141} \geq 0$$

$$p(4) = \frac{1}{4} \times \frac{60}{47} = \frac{60}{188} \geq 0$$

$$p(5) = \frac{1}{5} \times \frac{60}{47} = \frac{60}{235} \geq 0$$

अतः, $k = \frac{60}{47}$, प्रायिकता द्रव्यमान फलन के लिए पहली शर्त को भी पूरी करता है।

14.3.2 सतत् प्रायिकता बंटन

सतत् यादृच्छिक चर X की अपने किसी भी मान विशेष को सही मायने में धारण करने की प्रायिकता शून्य होती है। निश्चय ही, यह एक विचित्र सा कथन है। आइए इसे समझने का प्रयास करें। अब हम भार (weight) को यादृच्छिक चर मान कर पर विचार करते हैं। निस्संदेह भार एक सतत् यादृच्छिक चर है, क्योंकि यह निरंतर बदलता रहता है। मान लीजिए हमें किसी एक व्यक्ति का सही भार नहीं पता लेकिन मोटे तौर पर हम जानते हैं कि उसका भार 60 किलो और 61 किलो के बीच है। इन दो सीमाओं के बीच संभावित भारों की अनंत संख्याएं शामिल हैं। इसकी परिभाषा के परिणामस्वरूप व्यक्ति के किसी विशेष भार जैसे (60.3 किग्रा.) के लिए प्रायिकता काफी कम होगी; लगभग शून्य के बराबर। लेकिन हम निश्चित रूप से व्यक्ति के भार जैसे 60 किग्रा. और 61 किग्रा. के बीच के लिए कुछ प्रायिकता तय कर सकते हैं। अतः सतत् यादृच्छिक चर X के लिए किसी अंतराल (न कि किसी विशिष्ट मान) की प्रायिकता तय की जायेगी। यहाँ, हमें फलन $p(x)$ चाहिए जिसे *प्रायिकता घनत्व फलन* (probability derisity function) कहते हैं। इस फलन की सहायता से हम प्रायिकता को अभिकलित कर सकते हैं।

$p(a < x < b)$, यहां a और b अंतराल (a, b) की सीमाएं हैं तथा $a < b$

प्रायिकता घनत्व फलन $p(x)$ को इस ढंग से परिभाषित किया जाता है कि जब x के प्रांत पर अभिकलिन किया जाये तो इसके वक्र के नीचे x-अक्ष से परिबद्ध क्षेत्रफल इकाई के बराबर हो। प्रायिकता घनत्व फलन को वास्तविक संख्याओं R के पूरे समुच्चय पर परिभाषित सतत् यादृच्छिक चर के रूप में मान्य होने के लिए निम्नलिखित तीन शर्तों को पूरा करना होगा :

1) $p(x) > 0$ सभी $x \in R$ के लिए

2) $\int_{-\infty}^{+\infty} p(x)dx = 1$

3) $p(a < x < b) = \int_a^b p(x)dx$

यद्यपि सतत् यादृच्छिक चर के प्रायिकता बंटन को असतत् यादृच्छिक चर की भांति सारणी के रूप में नहीं दर्शाया जा सकता, फिर भी प्रायिकता घनत्व फलन $p(x)$ के विशिष्ट रूप से अभिव्यक्त किया जा सकता है। ऐसे कुछ रूपों (forms) का अध्ययन हम अगली इकाई में करेंगे, जो सतत् यादृच्छिक चरों से संबंधित सैद्धांतिक बंटनों पर आधारित है।

14.3.3 प्रमेयात्मक/अनुमित बंटन

प्रायिकता बंटन, प्रायिकता प्रयोग (experiments) से संबद्ध आनुभाविक प्रेक्षणों पर आधारित होते है। सुसंगत प्रायिकता बंटन की प्राप्ति के लिए, प्रयोग को समान स्थितियों में कई बार दोहराना पड़ता है यह कभी कभार अत्यंत कठिन कार्य सिद्ध होता है। वैकल्पिक रूप से सूत्र के प्रयोग से हम प्रायिकता द्रव्यमान फलन या प्रायिकता घनत्व फलन को स्पष्ट कर सकते हैं। हमें यह ध्यान रखना पड़ता है कि इसमें सभी सैद्धांतिक शर्तों की पूरी हो रही हैं। इस प्रकार के प्रायिकता बंटन को प्रमेयात्मक अथवा अनुमित बंटन कहते हैं। ऐसे बंटन का एक मुख्य फायदा है कि कुछ अनुमित बंटन, जीवन के अनेक घटनाक्रमों को सटीक

रूप से व्यक्त कर सकते हैं। परिणामस्वरूप प्रयोग किए बिना भी हम उन बंटनो के माध्यम से सीधे ही जीवन की सच्ची धटनाओं को समझ सकते हैं। अनुमित बंटन असतत् या सतत् हो सकता है। हम आगे दो महत्वपूर्ण असतत् अनुमित बंटनों की चर्चा कर रहे हैं। इन्हें अक्सर सांख्यिकीय विश्लेषण के लिए प्रयोग में लाया जाता है। अगली इकाई में हम कुछ सतत् अनुमित बंटनों का अध्ययन करेंगे।

14.4 यादृच्छिक चर का माध्य एवं प्रसारण

यादृच्छिक चर के माध्य को इसकी *गणितीय प्रत्याशा* या प्रत्याशित मान के रूप में भी जाना जाता है। इसे यादृच्छिक चर के मूल्यों और संगत प्रायिकताओं के गुणनफलों के योग के रूप में परिभाषित किया जाता है।

अतः यदि X असतत् यादृच्छिक चर है, जिसकी क्रमशः विशिष्ट प्रायिकताएँ $p_1, p_2, p_3 \cdots p_n$ और $x_1, x_2, x_3, \ldots\ldots x_n$ मान हैं, तब X की गणितीय प्रत्याशा है :

$$E(X) = x_1 p_1 + x_2 p_2 + x_3 p_3 + \ldots\ldots + x_n p_n = \sum_{i=1}^{n} x_i p_i$$

मजेदार बात यह है कि यादृच्छिक चर की गणितीय प्रत्याशा, साधारण चर के समांतर माध्य के समान होती है। असतत् यादृच्छिक चर के लिए, इसे आसानी से दर्शाया जा सकता है। हमने प्रायिकता की सापेक्षिक बारंबारता परिभाषा में देखा है कि किसी घटना की प्रायिकता को, उस घटना के घटित होने की सापेक्षिक बारंबारता की सीमा के रूप में व्यक्त किया जा सकता है। जब अभिप्रयोगों की संख्या अनंत की ओर प्रवृत हो, अर्थात :

$$p_i = \frac{f_i}{N}$$

जहाँ f_i, x_i की बारंबारता है और $N = \sum_{i=1}^{n} f_i$ कुल बारंबारता है।

अतः $E(X) = \sum_{i=1}^{n} p_i x_i = \sum_{i=1}^{n} \frac{f_i}{N} x_i$

$$= \frac{1}{N} \sum_{i=1}^{n} f_i x_i = \overline{X}$$

अतः हमने देखा कि यादृच्छिक चर की गणितीय प्रत्याशा उसका समांतर माध्य है।

उदाहरण 14.5

एक अनभिनत सिक्के को उछाला जाता है। यदि शीर्ष आता है तो आप 20 रुपए जीतेंगे और अगर 'पुच्छ' है तो आप 10 रुपए हारेंगे। वह राशि बताइए जिसे हर बार सिक्का उछालने पर आप जीतेंगे या हारेंगे।

चूंकि सिक्का अनभिनत है, इसलिए 'शीर्ष' या 'पुच्छ' प्राप्त करने की प्रायिकता $\frac{1}{2}$ है। मान लीजिए X यादृच्छिक चर है जो ऐसे मानों को लेता है जो हार और जीत वाली राशियों के बराबर है। इसलिए $\frac{1}{2}$ प्रायिकता के साथ $x = 20$ और दुबारा $\frac{1}{2}$ प्रायिकता के साथ $x = -10$ (हानि को नकारात्मक फायदा माना जा सकता है)।

इसलिए, प्रति उछाल पर हारने या जीतने की प्रत्याशित राशि है :

$$\left[20 \times \frac{1}{2} + (-10) \times \frac{1}{2}\right] = 5 \text{ रुपए}$$

सकारात्मक प्रत्याशित फायदे वाले खेल को, खिलाड़ी के पक्ष में एकतरफा कहा जाता है। यदि प्रत्याशित फायदा शून्य है तो खेल को निष्पक्ष कहते हैं। उपर्युक्त खेल को निष्पक्ष बनाया जा सकता है यदि प्रवेश शुल्क (प्रत्याशित मूल्य के बराबर की राशि के रूप में) हम 5 रुपए वसूल करें। यादृच्छिक चर X के संभावित मूल्य अब 15 और –15 और प्रत्याशित मूल्य $E(X) = 0$ है।

सतत् यादृच्छिक चर के लिए, गणितीय प्रत्याशा, निश्चित समाकल (definite integral) का रूप ले लेती है। इसलिए,

$$E(X) = \int_a^b xp(x)dx$$

जहाँ a से b तक के प्रांत में X एक सतत् यादृच्छिक चर है और $p(x)$ इसका प्रायिकता घनत्व है।

14.4.1 गणितीय प्रत्याशा आधारित कुछ प्रमेय

i) स्थिरांक की गणितीय प्रत्याशा स्थिरांक स्वयं होता है। यदि c स्थिरांक है तब,

$E(c) = c$

ii) स्थिरांक और यादृच्छिक चर के गुणनफल की गणितीय प्रत्याशा, यादृच्छिक चर की गणितीय प्रत्याशा और स्थिरांक का गुणनफल है। यदि c स्थिरांक है और X यादृच्छिक चर है, तब

$E(cX) = cE(X)$

iii) यादृच्छिक चर के किसी भी फलन की गणितीय प्रत्याशा, फलन के मानों और यादृच्छिक चर के मानों की तदनुरूप प्रायिकताओं के गुणनफलों का योग है। अतः यदि $f(X)$, यादृच्छिक चर X का फलन है जो विशिष्ट प्रायिकताओं $p_1, p_2, p_3, \ldots\ldots$ p_n के साथ $x_1, x_2, x_3 \ldots x_n$ मानों को लेता है तो $f(X)$ की गणितीय प्रत्याशा है

$$E[f(X)] = \sum_{i=1}^{n} f(x_i)\, p_i$$

हमें यहाँ ध्यान देना है कि उपर्युक्त संकलन सही अर्थों में असतत् यादृच्छिक चर पर ही लागू होता है। लेकिन बिना अपनी सामान्यता खोए यह प्रमेय, सतत् यादृच्छिक चर के लिए भी मान्य है। यहाँ कुछ सीमित मानों के संकलन की बजाय, यादृच्छिक चर के प्रांत का समाकलन (integration) आवश्यक है।

iv) यादृच्छिक चरों की दी गई संख्या के योगफल की गणितीय प्रत्याशा, उसकी प्रत्याशाओं का योगफल है। यदि X और Y दो यादृच्छिक चर हैं, $X + Y$ की गणितीय प्रत्याशा है;

$E(X + Y) = E(X) + E(Y)$

v) *स्वतंत्र यादृच्छिक चरों* की दी गई संख्या के गुणनफल की गणितीय प्रत्याशा, उनकी प्रत्याशाओं का गुणनफल है। यदि X और Y दो स्वतंत्र यादृच्छिक चर हैं, XY की गणितीय प्रत्याशा है;

$E(XY) = E(X).\,E(Y)$

यादृच्छिक चर X का प्रसरण होगा

$$V(X) = E[(X - E(X)]^2 = E(X^2) - [E(X)]^2$$

उदाहरण 14.5 (सिक्का उछालने वाले) में यादृच्छिक चर का प्रसरण, निम्नलिखित तरीके से अभिकलित किया जा सकता है :

पहले हम अभिकलित करेंगे

$$E(X^2) = 20^2.\frac{1}{2} + 10^2.\frac{1}{2} = 200 + 50 = 250$$

अब

$$V(X) = \sigma_x^{\,2} = 250 - 25 = 225$$

X का मानक विचलन भी

$$\sigma_x = \sqrt{225} = 15 \text{ रुपए है।}$$

14.4.2 प्रसरण आधारित कुछ प्रमेय

i) स्थिरांक का प्रसरण शून्य है। यदि c स्थिरांक है, तब

$V(c) = 0$

ii) स्थिरांक और यादृच्छिक चर के गुणनफल का प्रसरण, यादृच्छिक चर का प्रसरण और स्थिरांक के वर्ग का गुणनफल है। यदि c स्थिरांक और X यादृच्छिक चर है, तब

$V(cX) = c^2\, V(X)$

iii) यादृच्छिक चरों की दी गई संख्या के योगफल का प्रसरण उनके प्रसरणों तथा सह प्रसरण का योगफल है। यदि X और Y दो यादृच्छिक चर हैं, तब $X + Y$ का प्रसरण है

$V(X+Y) = V(X) + V(Y) + 2Cov\,(X, Y)$

यहाँ $Cov\,(X, Y)$, X और Y के बीच सहप्रसरण (covariance) कहलाता है। ध्यान रहे कि सहप्रसरण, दो चरों की एक साथ परिवर्तनशीलता का माप है।

सहप्रसरण को इस प्रकार दर्शाया जा सकता है :

$$Cov\,(X, Y) = E\,[\{X - E(X)\}\,\{Y - E\,(Y)\}] = E(XY) - E(X)\,E(Y)$$

लेकिन, यदि X और Y स्वतंत्र हैं तब एक चर में होने वाला परिवर्तन दूसरे चर में परिवर्तन उत्पन्न नहीं कर सकता। जिसके परिणामस्वरूप, $Cov\,(X, Y) = 0$ और

$$V(X+Y) = V(X) + V(Y)$$

यहाँ ध्यान देने योग्य बात है कि उपर्युक्त चर्चित गणितीय प्रत्याशा और प्रसरण पर आधारित सभी मूल प्रमेय सतत् और असतत् अर्थात दोनों प्रकार के यादृच्छिक चरों के लिए मान्य है।

अब हम एक महत्वपूर्ण परिणाम को व्यक्त और सिद्ध कर सकते हैं।

14.4.3 मानक प्रसामान्य विचर

दिए गए माध्य और मानक विचलन वाले किसी भी चर (यादृच्छिक या किसी अन्य) के लिए, जब भी माध्य को चर से कर मानक विचलन से विभाजित किया जाता है तो परिणामी चर का माध्य शून्य के बराबर और मानक विचलन एक के बराबर होता है।

आइए उपर्युक्त कथन को सिद्ध करें

मान लीजिए X एक यादृच्छिक चर है जिसका माध्य (प्रत्याशा) μ और मानक विचलन σ है।

अतः

$$z = \frac{X-\mu}{\sigma}$$

$$E(z) = E\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma} E(X-\mu) = \frac{1}{\sigma} E(X) - E(\mu)] = \frac{1}{\sigma}(\mu-\mu) = 0$$

अब

$$V(z) = V\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2} V(X-\mu) = \frac{1}{\sigma^2}[V(X) + (-1)\, V(\mu)]$$

उपर्युक्त में सहप्रसरण शब्द $Cov(X, \mu)$ लुप्त है क्योंकि X और μ स्वतंत्र हैं।

अतः

$$V(z) = \frac{1}{\sigma^2}[V(X)] = \frac{1}{\sigma^2}\sigma^2 = 1 \; [\because V(\mu) = 0]$$

अब

मानक विचलन $(z) = \sqrt{V(Z)} = \sqrt{1} = 1$

इस तरीके से हम देख सकते हैं

$$E(z) = E\left(\frac{X-\mu}{\sigma}\right) = 0$$

और

मानक विचलन $(z) =$ मानक विचलन $\left(\frac{X-\mu}{\sigma}\right) = 1$

उपर्युक्त विधि से परिभाषित चर z को बहुधा *मानक प्रसामान्य विचर* (standard normal variate) कहते हैं।

अगली इकाई में हम विचार करेंगे कि किस प्रकार प्रसामान्य बंटन के संदर्भ में इस परिणाम को प्रयोग में लाया जाता है।

बोध प्रश्न 1

1) जाँच कीजिए कि निम्नलिखित मान्य प्रायिकता द्रव्यमान फलन है या नहीं?

$$p(x) = \frac{x^2 - x}{16},$$

$x = -2, -1, 0, 1, 2, 3$

....

....

....

....

....

....

2) k का ऐसा एक मान ज्ञात कीजिए कि निम्नलिखित मान्य प्रायिकता द्रव्यमान फलन बन जाए।

$$p(x) = \frac{k}{x^2}, x = 1, 2$$

....

....

....

....

....

....

....

3) 99 रुपए के एक पुरस्कार के लिए A और B एक पासा उछालते हैं जो खिलाड़ी पहले छह उछालेगा, वह पुरस्कार जीतेगा। अगर पासे की पहली उछाल A शुरु करता है तो क्रमशः उनकी प्रत्याशाएं क्या हैं?

....

....

....

....

....

....

....

4) एक ठेकेदार विनिर्माण और बोली लगाने संबंधी खर्च निकालने के बाद, किसी निर्माण परियोजना पर बोली लगाने के लिए 3000 रुपए खर्च करता है और अगर वह बोली जीत लेता है तो इससे उसे 25000 रुपए का मुनाफा होगा। यदि बोली जीतने की संभावना 10 प्रतिशत है तो ठेकेदार के प्रत्याशित मुनाफे को अभिकलित कीजिए और बताइए कि उसे बोली लगानी चाहिए या नहीं।

....

....

....

....

....

5) निम्नलिखित परिणामों को सिद्ध कीजिए।

क) $E(cX) = cE(X)$, जहाँ c स्थिरांक है।

ख) $V(c) = 0$, जहाँ c स्थिरांक है।

ग) $V(cX) = c^2 V(X)$, जहाँ c स्थिरांक है।

....

....

....

....

....

14.5 द्विपद बंटन

द्विपद बंटन (binomial distribution) असतत् प्रायिकता बंटन का एक उदाहरण है। जेम्स बर्नोली ने वर्ष 1700 में इसे प्रस्तुत किया था। 'बायनोमियल' शब्द दो का संकेत कहता है। यह प्रयोग के दो संभावित परिणामों को दर्शाता है, अर्थात, किसी घटना का घटित होना या घटित न होना। एक प्रायिकता प्रयोग को बर्नोली प्रयोग कहा जा सकता है, यदि यह निम्नलिखित शर्तों को पूरा करें,

1) प्रयोग में n पुनरावृत्त अभिप्रयोगों (repeated trials) का अनुक्रम शामिल हो।

2) प्रत्येक अभिप्रयोग का परिणाम ऐसा होना चाहिए जिसे *सफलता* या *असफलता* के रूप में वर्गीकृत किया जा सके।

3) सफलता की प्रायिकता, जिसे p द्वारा दर्शाया जाता है, पूर्व ज्ञात हो और यह प्रत्येक अभिप्रयोग में स्थिर रहती हो, परिणामस्वरूप, $q = 1-p$ द्वारा अभिव्यक्त असफलता की प्रायिकता का भी पता रहता है और यह भी प्रत्येक अभिप्रयोग में स्थिर रहती हो।

4) प्रत्येक अभिप्रयोग एक दूसरे से स्वतंत्र हो।

हम बर्नोली प्रयोग के बारे में कुछ आधार बनाने के लिए एक अनभिनत सिक्का उछालने के उदाहरण पर विचार करते हैं। सिक्का बार-बार उछाला जाता है और आने वाले शीर्षों की संख्या की गिनती की जाती है। मान लीजिए हमने एक अनभिनत सिक्का 5 बार उछाला।

यह स्पष्ट है कि प्रयोग में 5 समान अभिप्रयोगों का अनुक्रम शामिल है। प्रत्येक उछाल के दो संभावित परिणाम हैं, शीर्ष (सफलता) और पुच्छ (असफलता)। शीर्ष (सफलता) की प्रायिकता $\frac{1}{2}$ है और एक उछाल से दूसरे उछाल अर्थात कई बार उछालने पर भी इसमें कोई परिवर्तन नहीं होता। पुच्छ (असफलता) की प्रायिकता भी $\frac{1}{2}$ है और इसके प्रायिक्ता में भी पहले की ही भांति कोई परिवर्तन नहीं होता। अंततः प्रत्येक उछाल दूसरे उछाल से इस तरह बिल्कुल अलग है कि एक उछाल का परिणाम, किसी भी तरह से दूसरे उछाल के परिणाम पर निर्भर नहीं करता। अतः हम पाते हैं कि कुछ निश्चित संख्या में सिक्का उछालने का यह प्रयोग और आने वाले शीर्षों की संख्या पर ध्यान देते हुए यह बर्नोली प्रयोग की सभी शर्तों को भलीभांति पूरा करता है।

बर्नोली प्रयोग में हम सफलताओं की दी गई संख्या की प्रायिकता जानने के इच्छुक हैं, जैसे n अभिप्रयोग में उभरने वाले x (जैसे, पिछले उदाहरण में हम 5 उछाल में 3 शीर्ष प्राप्त करने की प्रायिकता जानने के इच्छुक थे)। अब यह स्पष्ट है कि यादृच्छिक चर X, का मान $0, 1, 2, 3, \ldots n$ में से कोई भी हो सकता है। मान लीजिए हम सफलता को S और असफलता को F से दर्शाते हैं। तब x सफलताएं और $(n-x)$ असफलताएं अलग-अलग अनुक्रमों में उभर सकती हैं। एक संभावित अनुक्रम है कि पहले वाले x अभिप्रयोग, सभी सफलताएं हैं और बाकी के $(n-x)$ अभिप्रयोग सारी असफलताएं हैं। सांकेतिक रूप से, इस अनुक्रम को इस प्रकार दर्शाया जाता है।

$$\frac{SS....S}{x \text{ बार}} \times \frac{FF....F}{(n-x) \text{ बार}}$$

x सफलताओं और $(n-x)$ असफलताओं के उपर्युक्त अनुक्रम की प्रायिकता को, प्रायिकता की गुणन प्रमेय को लागू करके प्राप्त किया जा सकता है। यह प्रायिकता

$$\frac{PP....P}{x \text{ बार}} \times \frac{(1-P)(1-P)\ldots(1-P)}{(n-x) \text{ बार}} = P^x (1-P)^{n-x} \text{ है।}$$

लेकिन, जैसाकि हमने ऊपर बताया था, x सफलताएं और $(n-x)$ असफलताएं, अन्य अनुक्रमों में भी उभर सकती हैं। लेकिन, ऐसा हरेक अनुक्रम जिसमें x सफलताएं और $(n-x)$ असफलताएं उभरेगी $P^x(1-P)^{n-x}$ की प्रायिकता को प्राप्त करेंगे। अतः n अभिप्रयोगों में x सफलताओं की प्रायिकता, किसी भी संभावित अनुक्रमों में x सफलताएं और $(n-x)$ असफलताओं के घटित होने की प्रायिकता हैं। संभावित क्रमों पर प्रायिकता की योगफल प्रमेय को लागू करके, इस प्रायिकता की प्राप्ति की जा सकती है। लेकिन, चूंकि x सफलताओं और $(n-x)$ असफलताओं की प्रायिकता, प्रत्येक संभावित अनुक्रम के लिए एक जैसी है, तो n अभिप्रयोगों में x सफलताओं की अपेक्षित प्रायिकता, संभावित अनुक्रमों की कुल संख्या और अनुक्रम के घटित होने की प्रायिकता का गुणनफल है। अनुक्रमों की कुल संख्या (जिसमें x सफलताएं और $n-x$ असफलताएं, n अभिप्रयोगों में उभर सकती हैं) मूलरूप से समय में x द्वारा n वस्तुओं के संचय की संख्या प्राप्त करने की समस्या है और इसे ${}^nC_x$ द्वारा दर्शाया जाता है। *क्रमचय* और *संचय* (permutation and combination) के गणित से हम जानते ही हैं कि :

$${}^nC_x = \frac{n!}{x!(n-x)!}$$

जहाँ

$$n! = n(n-1)\ (n-2)\ ....2.1$$

$$x! = x(x-1)\ (x-2)\ ....2.1$$

और

$$0! = 1$$

हमें पता होना चाहिए कि चिह्न '!' को *क्रमगुणन* (factorial) कहते हैं। जैसे $4! = 4 \times 3 \times 2 \times 1 = 24$

^nC_x में संकेत C संचय को दर्शाता है। जैसे

$$^5C_3 = \frac{5!}{3!.(5-3).} = \frac{5!}{3!.2!} = \frac{5 \times 4 \times 3 \times 2 \times 1}{3 \times 2 \times 1 \times 2 \times 1} = 10$$

अतः n अभिप्रयोगों में x सफलताओं की प्रायिकता

$p(x) = {}^nC_x p^x\ (1-p)^{n-x}$ है।

जहाँ $x = 0, 1, 2, ...\ n$ मानों को ले सकता है।

उपर्युक्त व्यंजक द्विपद बंटन के लिए प्रायिकता द्रव्यमान फलन है। इस फलन का प्रयोग सारणी 14.3 में n अभिप्रयोगों में $x = 0, 1, 2, ...n$ सफलताओं के द्विपद बंटन को दर्शाने के लिए किया गया है। यहाँ हम देखते हैं कि द्विपद बंटन के दो प्राचल n और p हैं। इसका अर्थ है कि यदि n और p के मानों का पता हो तो बंटन पूरी तरह से स्पष्ट होता है।

सारणी 14.3 : द्विपद बंटन

सफलताओं की संख्या x	प्रायिकता $p(x)$
0	$(1-p)^n$
1	$np(1-p)^{n-1}$
2.	$\frac{n(n-1)}{2.1} p^2\ (1-p)^{n-2}$
:	:
N	p^n
कुल	**1**

आइए अब द्विपद बंटन का माध्य ज्ञात करें।

मान लीजिए अभिप्रयोग में सफलता की प्रायिकता के रूप में द्विपद प्रयोग में p और अभिप्रयोगों की n संख्या है। इसका अर्थ है असफलता की प्रायिकता $1-p$ है।

$$E(X) = \sum_{x=0}^{n} x\ {}^nC_x p^x\ (1-p)^{n-x}$$

उपर्युक्त को सरल करने पर हम पाते हैं कि द्विपद बंटन का माध्य, np है। हम उपर्युक्त का प्रमाण प्रस्तुत नहीं कर रहे क्योंकि यह एक लंबा कार्य है।

इसी तरह से द्विपद बंटन का प्रसरण,

$$V(X) = E[X - E(X)]^2 = E(X^2) - [E(X)]^2 \text{ है।}$$

उपर्युक्त के सरलीकरण से दर्शाया जा सकता है कि द्विपद बंटन का प्रसरण npq होता है [जो $np(1-p)$] के बराबर है।

अतः हम देखते हैं कि द्विपद बंटन का माध्य और प्रसरण, उसके दो प्राचल n और p है। हमने नीचे ऐसे कुछ उदाहरण दिए हैं जो द्विपद बंटन के अनुप्रयोग पर आधारित है।

उदाहरण 14.6

एक मशीन के उत्पादन में आमतौर पर 20% उत्पाद दोषपूर्ण होता है। गुणवत्ता नियंत्रण इंस्पेक्टर यादृच्छिक रूप से 5 इकाइयों का चयन करता है। (i) 1 दोषपूर्ण मद, तथा (ii) कम से कम 3 दोषपूर्ण मदों को प्राप्त करने की प्रायिकता ज्ञात कीजिए।

यह द्विपद बंटन का उदाहरण है जिसमें $p = 0.20$ और $n = 5$ है।

आइए इस प्रश्न को हल करें।

i) हमें पता है n मदों में x दोषपूर्ण मदों (अर्थात, $n-x$ गैर दोषपूर्ण मद) की प्रायिकता ${}^nC_x$ $p^x (1-p)^{n-x}$। यहाँ $n = 5$ और $x = 1$ है।

   अतः 1 दोषपूर्ण मद की प्रायिकता है

   $$p(1) = {}^5C_1\,(0.20).(0.80)^4 = 0.4096$$

ii) कम से कम 3 दोषपूर्ण मदों का अर्थ है कि यहाँ 3 या 4 या 5 दोषपूर्ण मद हो सकते हैं। अतः कम से कम 3 दोषपूर्ण मदों की प्रायिकता, का अर्थ 3 दोषपूर्ण मदों और 4 दोषपूर्ण मदों और 5 दोषपूर्ण मदों की कुल प्रायिकता है।

अब, 3 दोषपूर्ण मदों की प्रायिकता है

$$p(3) = {}^5C_3\,(0.20)^3\,(0.80)^2 = 0.0512$$

4 दोषपूर्ण मदों की प्रायिकता है

$$p(4) = {}^5C_4\,(0.20)^4\,.0.8 = 0.0064$$

5 दोषपूर्ण मदों की प्रायिकता है

$$p(5) = {}^5C_5\,(0.20)^5 = 0.0003$$

इसलिए, कम से कम 3 दोषपूर्ण मदों की प्रायिकता है $= 0.0512 + 0.0064 + 0.0003 = 0.0579$

उदाहरण 14.7

अगर एक अनभिनत सिक्के को 36 बार उछाला जाए तो 6 प्राप्त करने की प्रत्याशित संख्या क्या है? इसका प्रसरण क्या है?

यदि 6 प्राप्त करने की प्रायिकता p है, तब $p = \frac{1}{6}$ और $(1-p) = \frac{5}{6}$

अब, $n = 36$

गणितीय प्रत्याशा

$$np = 36 \times \frac{1}{6} = 6$$

अतः यदि पासे को 36 बार उछाला जाए तो 6 को 6 बार प्राप्त किया जा सकता है।

प्रसरण होगा

$$np(1-p) = 36 \times \frac{1}{6} \times \frac{5}{6} = 5$$

14.6 पाइसों बंटन

पाइसों बंटन (Poisson distribution) एक अन्य असतत् प्रायिकता बंटन है। इसका नामकरण फ्रांसीसी गणितज्ञ साइमन पाइसों के नाम पर हुआ है जिन्होंने 1837 में इस बंटन को निरूपण किया था। यह बंटन असल में द्विपद बंटन का एक विशेष (सीमित) रूप है। जब द्विपद बंटन में सफलता की प्रायिकता p बहुत कम हो और अभिप्रयोगों की संख्या n, इतनी अधिक है कि प्रत्याशा, $\mu = np$ परिमित परिमाण हो। तो द्विपद बंटन, पाइसों बंटन की ओर प्रवृत होता है। यह बंटन विशेषरूप से समय या समष्टि के निर्धारित अंतराल में किसी वस्तु आदि के घटित होने की संख्या से संबंधित है। जैसे, मान लीजिए विचाराधीन यादृच्छिक चर, 1 घंटे में टेलीफोन स्विच बोर्ड पर आने वाली टेलीफोन कॉलों की संख्या है या 100 किलोमीटर की पाइपलाइन में रिसाव की संख्या है या दिल्ली में किसी निर्धारित दिवस पर होने वाली बस दुर्घटनाओं की संख्या है।

पाइसों बंटन के प्रायिकता द्रव्यमान फलन की प्राप्ति के लिए उदाहरण के रूप में हम एक घंटे में टेलीफोन कॉल x की संख्या पर विचार कर सकते हैं और मान लेते हैं कि प्रति घंटा टेलीफोन कॉलों की प्रत्याशित संख्या (अर्थात गणितीय प्रत्याशा) λ है। द्विपद बंटन लागू करने के लिए हम एक घंटे के अंतराल को उप-अंतरालों में विभाजित करते हैं और जो इतने छोटे हैं कि उप-अंतराल में टेलीफोन कॉल करने की प्रायिकता p भी काफी छोटी (कम) है और जो एक कॉल से अधिक कॉलों की प्राप्ति पर लगभग शून्य है। अतः प्रत्येक़ उप-अंतराल को बर्नोली अभिप्रयोग के रूप में माना जा सकता है और जिसके केवल दो संभावित परिणाम हैं अर्थात या तो टेलीफोन काल का आना (सफलता) या टेलीफोन कॉल का न आना (असफलता)। उप-अंतरालों की संख्या को अभिप्रयोगों की कुल संख्या, अर्थात, n के बराबर माना गया है। हम देखते हैं कि टेलीफोन कॉलों की प्रत्याशित संख्या λ वैसी ही रहती है और np के बराबर है (जैसाकि हमने द्विपद बंटन से देखा था)। इसलिए उप-अंतराल में टेलीफोन कॉल की प्रायिकता $\frac{\lambda}{n}$ है। अतः एक घंटे में x टेलीफोन कॉलों की प्रायिकता n अभिप्रयोगों में x सफलताओं की प्रायिकता प्राप्त करने जैसी ही है, (जब n अनंत की ओर प्रवृत हो)। हमने ऊपर तर्क किया था, n अभिप्रयोग, n उप-अन्तरालों के तदनुरूप है जो एक घंटे को पूरा करते हैं। यह प्रत्येक उप-अभिप्रयोग को अत्यंत छोटा बनाने का परिणाम है ताकि अभिप्रयोगों की कुल संख्या, बहुत बड़ी अनंत की ओर प्रवृत हो जाए। यह प्रायिकता, द्विपद बंटन की निम्नलिखित सीमा के रूप में है:

$$\lim_{n \to \infty} {}^nC_x \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

आइए उल्लिखित सीमा ज्ञात करने का प्रयास करें।

पाइसों बंटन के प्रायिकता द्रव्यमान फलन उल्लिखित सीमा से मिल सकता है। यह इस प्रकार हैं:

$$p(x) = \frac{e^{-\lambda}\lambda^x}{x!} \qquad x = 0, 1, 2 \ldots$$

जहाँ X यादृच्छिक चर है जो विशिष्ट समय अंतराल या लंबाई अंतराल में सफलताओं की संख्या को दर्शाता है।

λ = समय या लंबाई आदि के अंतराल में घटित होने वालों का प्रत्याशित मान या औसत संख्या है।

e = एक स्थिरांक है (प्राकृतिक लघुगुणक का आधार) जिसका मान e = 2.7182... है।

सारणी 14.4 : पाइसों बंटन

पाइसों यादृच्छिक चर x का मान	**प्रायिक्ता $p(x)$**
0	$e^{-\lambda}$
1	$\frac{\lambda e^{-\lambda}}{1!}$
2	$\frac{\lambda^2 e^{-\lambda}}{2!}$
:	:
:	:
कुल	1

पाइसों बंटन में, यादृच्छिक चर X (घटनाओं की संख्या की कोई ऊपरी सीमा नहीं है। यह असतत् यादृच्छिक चर है जो ($X = 0, 1, 2, 3 \ldots$) मानों के असीम अनुक्रमों को धारण कर सकता है। इस बंटन का केवल एक प्राचल λ है। तालिका 14.4 पाइसों प्रायिक्ता द्रव्यमान फलन द्वारा जनित पाइसों बंटन को दर्शाती है।

माध्य और प्रसरण

जैसा कि हमने ऊपर देखा, पाइसों बंटन की प्रत्याशा, स्थिरांक λ है। यह दर्शाया जा सकता है कि पाइसों बंटन का प्रसरण भी λ है।

उदाहरण 14.8

एक आंकड़ों का विश्लेषण दर्शाता है कि 15 मिनट के समय में पेट्रोल पंप पर औसत 10 कारें आती हैं। 15 मिनटों में 5 कारें आने की प्रायिकता ज्ञात कीजिए। 3 मिनटों में 1 कार आने की प्रायिकता क्या है?

यहाँ, $\mu = 10$ और $x = 5$, अतः अपेक्षित प्रायिकता है

$$p(5) = \frac{10^5 e^{-10}}{5!} = 0.0378$$

यदि 15 मिनट में कारों के आने की प्रत्याशित संख्या 10 है तब 3 मिनट में कारों के आने की प्रत्याशित संख्या है, $\frac{10}{15} \times 3 = 2$

अतः प्रश्न के दूसरे भाग के लिए, $\mu = 2$ और $x = 1$ है। इसलिए, 3 मिनट में एक कार आने की प्रायिक्ता है

$$p(1) = \frac{2e^{-2}}{1!} = 0.2707$$

बोध प्रश्न 2

1) किसी अस्पताल में मरीजों को लाने और ले जाने के लिए 3 एम्बुलेंस हैं। एम्बुलेंस की उपलब्धता की प्रायिकता 0.75 है। यदि किसी एम्बुलेंस की ज़रूरत पड़े तो प्रायिकता क्या है कि

   क) कोई भी एम्बुलेंस उपलब्ध न हो?

   ख) कम से कम एक एम्बुलेंस उपलब्ध हो?

   ....

   ....

   ....

   ....

2) क्या द्विपद बंटन के लिए, माध्य और प्रसरण क्रमशः 3 और 5 हो सकते हैं?

   ....

   ....

   ....

   ....

3) पिछले अनुभवों से पता चलता है कि किसी सयंत्र में प्रति माह औसतन 4 औद्योगिक दुर्घटनाएं होती हैं। प्रायिकता ज्ञात कीजिए कि दिए गए माह में 4 से कम दुर्घटनाएं होंगी। इसे पाइसों बंटन मान कर हल करें।

   ....

   ....

   ....

   ....

14.7 सारांश

इस इकाई में हमने प्रायिकता बंटन का अर्थ समझने के लिए, प्रायिकता की अवधारणा का प्रयोग किया है। हमने गणितीय प्रत्याशा की संकल्पनाओं और प्रायिकता बंटन के प्रसरण को समझा। हमने असतत् प्रायिकता बंटन और सतत् प्रायिकता बंटन के बीच अंतर को

समझा। इस संदर्भ में, हमने प्रायिकता द्रव्यमान फलन और प्रायिकता घनत्व फलन की अवधारणा को समझा। हमने दो विशिष्ट असतत् प्रायिकता बंटनों का अध्ययन किया और जो है : द्विपद बंटन और पाइसों बंटन। हमने इन बंटनों (विशेषरूप से इनके माध्य और प्रसरण के व्यंजक की विशेषताओं) का अध्ययन किया। हमने विभिन्न स्थितियों में इन दोनों बंटनों के प्रयोग को समझने का प्रयास भी किया है।

14.8 शब्दावली

द्विपद बंटन (Binominal Distribution): यह असतत् प्रायिक्ता बंटन है जो निम्नलिखित शर्तों को पूरा करता है:

1) इसमें समरूप अभिप्रयोगों की पुनरावृत्ति शामिल है।
2) प्रत्येक अभिप्रयोग से ऐसा परिणाम प्राप्त होता है जिसे *सफलता* या *असफलता* के रूप में वर्गीकृत किया जा सकता है।
3) सफलता की प्रायिकता को p से दर्शाते हैं, वह पूर्वज्ञात होती है और प्रत्येक अभिप्रयोग में स्थिर रहती है। जिसके परिणामस्वरूप q द्वारा दर्शायी असफलता की प्रायिक्ता $[q = (1—p)]$ भी पूर्व ज्ञात होती है और प्रत्येक अभिप्रयोग में स्थिर रहती है।
4) एक अभिप्रयोग दूसरे से स्वतंत्र है।
5) द्विपद यादृच्छिक चर x का n अभिप्रयोगों में सफलताओं की कुल संख्या है :

   ${}^{n}C_{x}\, P^{x}(1-P)^{1-x}$ जहाँ $0 \leq x \leq n$

सतत् प्रायिकता बंटन (Continuous Probability Distribution): यह सतत् यादृच्छिक चर के लिए, प्रायिकता बंटन है।

सतत् यादृच्छिक चर (Continuous Random Variable): यह यादृच्छिक चर है जो निश्चित अंतराल में सभी मानों को धारण कर सकता है।

असतत् प्रायिकता बंटन (Discrete Probability Distribution): यह असतत् यादृच्छिक चर के लिए प्रायिकता बंटन है।

सतत् यादृच्छिक चर (Discrete Random Variable): यह ऐसा यादृच्छिक चर है जिसके परिमित संख्या वाले मान या अनंत अनुक्रम (जैसे 1, 2, 3, ...) हो सकते हैं।

गणितीय प्रत्याशा (Mathematical Expectation): यादृच्छिक चर की गणितीय प्रत्याशा या प्रत्याशित मान, यादृच्छिक चर के मूल्यों और संगत प्रायिकताओं के गुणनफलों का योग होता है।

पाइसों बंटन (Poisson Distribution) : यह असतत् प्रायिकता बंटन द्विपद बंटन का एक सीमित रूप है जब कि : द्विपद बंटन में सफलता p की प्रायिकता काफी कम और अभिप्रयोगों की संख्या n इतनी अधिक है कि प्रत्याशा $\mu = np$ परिमित संख्यावाली है। पाइसों बंटन का माध्य एवं प्रसरण समान होता है।

प्रायिकता घनत्व फलन (Probability Denisty Function): यह सतत् यादृच्छिक चर का फलन है, लेकिन प्रायिकता द्रव्यमान फलन की भांति, यह यादृच्छिक चर के विशिष्ट मान के लिए सीधे तौर पर प्रायकिता नहीं दे सकता। यहाँ हम केवल अंतराल के भीतर यादृच्छिक चर की प्रायिकता प्राप्त कर सकते हैं।

प्रायिकता बंटन (Probability Distribution): यह यादृच्छिक चर और उसकी प्रायिकताओं के संभावित मानों का प्रकथन है।

प्रायिकता द्रव्यमान फलन (Probability Mass Function): यह ऐसा फलन है जो असतत् यादृच्छिक चर के विशिष्ट मान की प्रायिकता देता है।

प्रमेयात्मक/अनुमित बंटन (Theoretical Distribution): यह प्रायिकता बंटन है जो यादृच्छिक प्रयोग की विशिष्ट स्थितियों द्वारा जनित होता है। अनुमित प्रायिकता बंटन के कुछ उदाहरण हैं : द्विपद बंटन, पाइसों बंटन, प्रसामान्य बंटन आदि।

प्रसरण (Variance): यदि यादृच्छिक चर X की गणितीय प्रत्याशा $E(X)$ है तो X के प्रसरण को $E[X-E(X)]^2$ के रूप में परिभाषित किया जाता है।

14.9 कुछ उपयोगी पुस्तकें

Anderson, D.R., Sweeney, D.J. and Williams, T.A., 1993. *Statistics for Business and Economics* (Fifth Edition), West Publishing Company, Minneapolis/St. Paul, Chapter 6.

Bhardwaj, R.S. 1999. *Business Statistics* (First Edition), Excel Books, New Delhi, Chapters 19 and 20.

Newbold, P. 1991. *Statistics for Business and Economics* (Third Edition), Prentice Hall, New Jersey, Chapter 5.

Walpole, R.E., Myers, R.H., Myers, S.L. and Ye, K. 2002. *Probability and Statistics for Engineers and Scientists* (Seventh Edition), Pearson Education, India, Chapter 6.

14.10 बोध प्रश्नों के उत्तर या संकेत

बोध प्रश्न 1

1) $p(-2)=\frac{3}{8},\ p(-1)=\frac{1}{8}$

$p(0)=0$

$p(1)=0$

$p(2)=\frac{1}{8}$

$p(3)=\frac{3}{8}$

चूंकि x के सभी मानों के लिए $p(x)\geq 0$, अतः पहली शर्त पूरी होती है।

$\sum p(x)=\frac{3}{8}+\frac{1}{8}+0+0+\frac{1}{8}+\frac{3}{8}=1$; अतः दुसरा शर्त भी पूरी होती है।

इसलिए, दिया गया फलन मान्य प्रायिकता द्रव्यमान फलन है।

2) $k = \frac{4}{5}$

3) A जीत सकता है यदि A पहली बार में छह लाता है या A पहली बार में छह नहीं लाता और B दूसरी बार में छह नहीं लाता और A तीसरी बारे में छह लाता है और इसी तरह आगे। इसलिए, प्रायकिता कि A पहले छह लायेगा :

$$= \frac{1}{6} + \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{1}{6} + \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} = \frac{1}{6} + \ldots = \frac{6}{11}$$

अतः A का प्रत्याशित लाभ है; $99 \times \frac{6}{11} = 54$

B द्वारा पहले '6' का आंकड़ा पाने की प्रायिकता है $= (1-p) = 1 - \frac{6}{11} = \frac{5}{11}$

अतः B का प्रत्याशित लाभ होगा $99 \times \frac{5}{11} = 4$ रुपए

4) 200 रुपए

5) क) $E(cX) = E(c)\,E(X) = cE(X)$ (क्योंकि स्थिरांक की प्रत्याशा स्थिरांक ही है)।

ख) $V(c) = E(c^2) - [E(c)]^2 = c^2 - c^2 = 0$

ग) $V(cX) = E(c^2X^2) - [E(cX)]^2 = c^2\,E(X^2) - c^2\,[E(X)]^2$

$= c^2\,[E(X^2) - [E(X)^2] = c^2V(X)$

बोध प्रश्न 2

1) क) $\frac{1}{64}$ ख) $\frac{63}{64}$

2) माध्य $np = 3$, प्रसरण $np(1-p) = 5$, अब,

$$\frac{np(1-p)}{np} = \frac{5}{3}$$

या $1 - p = \frac{5}{3}$

या $p = 1 - \frac{5}{3} = -\frac{2}{3} < 0$, जो संभव नहीं है।

3) 0.4332

इकाई 15 प्रायिकता बंटन – II

इकाई की रूपरेखा

15.0 उद्देश्य

इस इकाई को पढ़ने के बाद, आप :

- प्रसामान्य बंटन की व्याख्या कर सकेंगे और इसे प्रयोग कर पाएंगे;
- स्वतंत्रता की कोटि की संकल्पना को समझ सकेंगे; और
- काई-स्कवैयर बंटन, स्टूडैन्ट –'टी' बंटन और एफ (F) बंटन के बारे में कुछ बुनियादी विचार समझ पाएंगे।

15.1 प्रस्तावना

पिछली इकाई में हमने असतत् यादृच्छिक चर और सतत् यादृच्छिक चर के बीच अंतर स्पष्ट किया था। उस इकाई में हमने, आपको प्रायिकता बंटन की संकल्पना से अवगत कराया था और हमने देखा था कि यह अनिवार्य रूप से यादृच्छिक चर और उससे संबंद्ध प्रायिकताओं के मानों से संबंधित प्रकथन है। उस इकाई में हमने दो महत्वपूर्ण असतत् प्रायिकता बंटनों अर्थात द्विपद बंटन और पाइसों बंटन का अध्ययन किया था। इस इकाई में हम इसी विषय को आगे बढाएंगे और एक महत्वपूर्ण सतत् प्रायिकता बंटन का अध्ययन करेंगे जिसे हम प्रसामान्य बंटन (normal distribution) कहते हैं। यहां उल्लेख करना आवश्यक है कि सांख्यिकीय अनुमिति और परिकल्पना परीक्षण में प्रसामान्य बंटन की भूमिका अत्यंत महत्वपूर्ण होती है। खंड-7 में हम इसी विषय-वस्तु पर अपनी बातचीत जारी रखेंगे।

वास्तव में हम इकाई 16 में सांख्यिकीय परिकल्पना परीक्षेपणों के आधार स्वरूप प्रतिदर्शी बंटनों के विषय में विचार करने जा रहे हैं। इसलिए प्रसामान्य बंटन के साथ-साथ हम अन्य तीन सतत् प्रायिकता बंटनों के बारे में भी कुछ बुनियादी विचारों की चर्चा से हम प्रतिदर्शी बंटन के गुणों का अधिक स्पष्ट विवेचन कर पाएंगे। इन तीन सतत् प्रायिकता बंटनों से हमारा आशय काई-स्क्वैयर बंटन, स्टूडैन्ट–'टी' (t) बंटन और एफ (F) बंटन से है। इन प्रायिकता बंटनों के बारे में हम इसी इकाई में चर्चा करेंगे

15.2 प्रसामान्य बंटन

प्रसामान्य बंटन शायद सांख्यिकी और संबंधित विषयों में सर्वाधिक व्यापक रूप से प्रयुक्त बंटन है। इसका प्रयोग व्यक्तियों की लंबाई और भार से संबंधित पूछताछ करने, आई.क्यू. (बौदि्धक विकास) स्तर का निर्धारण करने, मापन की त्रुटि और वर्षा आदि के अध्ययनों में किया जाता है। इब्राहिम दि माइर (Ibrahim de Moivre) ने 1733 में प्रसामान्य बंटन का गणितीय समीकरण दिया। कार्ल फ्रेडरिच गॉस ने भी अलग से समान परिमाण की पुनरावर्त मापन त्रुटियों के अध्ययन से इसके समीकरण को व्युत्पन्न किया। इसी आधार पर कभी कभी इसे 'गॉसीय वितरण' के रूप में भी देखा जाता है। गणितीय सांख्यिकी के परवर्ती विकास को इस वितरण ने आधार प्रदान किया है।

हमने पिछली इकाई में देखा था कि सतत् यादृच्छिक चर के लिए प्रायिकता द्रव्यमान फलन का प्रतिरूप, **प्रायिकता घनत्व फलन** है। प्रायिकता घनत्व फलन को हम $p(x)$ द्वारा भी सूचित करते हैं। प्रसामान्य बंटन का अनुगमन करने वाले सतत् यादृच्छिक चर का प्रायिकता घनत्व फलन है :

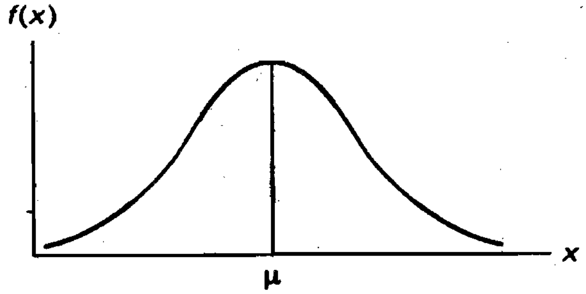
$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{(-)\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

जहाँ $-\infty < x < \infty$, और

$\pi = 3.17141$ (सन्निकटतः)

$e = 2.71828$ (सन्निकटतः)

यह स्पष्ट है कि प्रसामान्य घनत्व फलन पूरी तरह प्राचल μ और σ से निर्धारित किया जाता है। इसका अर्थ है कि यदि μ और σ के मान दिए हैं तो हम x के विभिन्न मानों के लिए $p(x)$ के मानों की प्राप्ति करके प्रसामान्य वक्र का पता लगा सकते हैं। हम यह भी दिखा सकते हैं कि μ और σ क्रमशः प्रसामान्य बंटन के माध्य और मानक विचलन हैं। जब यादृच्छिक चर X माध्य μ और मानक विचलन σ वाले प्रसामान्य बंटन का अनुगमन करता है तो संकेत के रूप में हम इसे $X \sim N(\mu, \sigma)$ के रूप में लिखते हैं और इसे पढ़ते हैं कि चर 'X' माध्य μ और मानक विचलन σ वाले प्रसामान्य बंटन का अनुगामी है। जैसा कि चित्र 15.1 में दर्शाया गया है, प्रसामान्य वक्र संममित घंटाकार वक्र है।



चित्र 15.1: प्रसामान्य वक्र

प्रसामान्य बंटन की महत्वपूर्ण विशेषताएं इस प्रकार हैं :

1) प्रसामान्य वक्र का परिसर $-\infty$ से $+\infty$ के बीच का है। इसका अर्थ है कि प्रसामान्य यादृच्छिक चर (X) के मान $-\infty$ से $+\infty$ के बीच के हैं।

2) वक्र अपने माध्य के लिए सममित है अर्थात $\bar{x} = \mu$. इसका अर्थ है कि $x = \mu + a$ और $x = \mu - a$ के संवात में, $p(x)$ के मान समान रहते हैं (किसी भी मनमाने चयनित 'a' के लिए)।

3) बंटन की माध्यिका (median) और बहुलक (mode) माध्य के साथ संपाती (coincide) है। अतः माध्य = माध्यिका = बहुलक = μ

4) प्रसामान्य वक्र का अधिकतम मान $x = \mu$ पर उभरता है। अतः $p(x)$ अधिकतम है जब $x = \mu$

5) प्रसामान्य वक्र के नतिपरिवर्तन बिंदु (points of inflexion), $x = \mu + \sigma$ और $x = \mu - \sigma$ पर उभरते हैं। नतिपरिर्वतन बिंदु पर, प्रसामान्य वक्र अपना वक्राकार बदल लेता है।

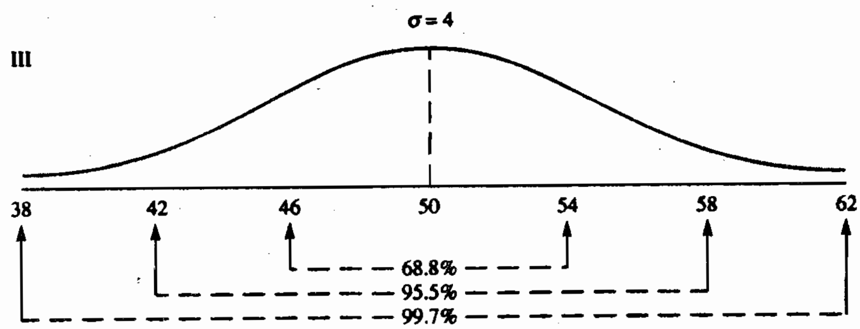
निम्नलिखित गुणधर्म प्रसामान्य बंटन पर मान्य रहते हैं। चित्र 15.2 में हमने माध्य $\mu = 50$ और मानक विचलन $\sigma = 4$ वाला प्रसामान्य वक्र खींचा है;

क) प्रसामान्य वक्र के नीचे 68.8% क्षेत्रफल $\mu - \sigma$ और $\mu + \sigma$ कोटियों के बीच निहित है। अतः चित्र 15.2 में जब x का परिसर 46 और 54 के बीच है तो 68.8% क्षेत्र ढ़का हुआ है।

ख) प्रसामान्य वक्र के नीचे 95.5% क्षेत्रफल $\mu - 2\sigma$ और $\mu + 2\sigma$ कोटियों के बीच निहित है। अतः जब $42 \le x \le 58$, तब चित्र 15.2 में 95.5% क्षेत्र ढका हुआ है।

ग) प्रसामान्य वक्र के नीचे 99.7% क्षेत्रफल (अर्थात बंटन का लगभग पूरा क्षेत्र) $\mu - 3\sigma$ और $\mu + 3\sigma$ कोटियों के बीच निहित है।

यदि हमारे पास μ और σ के अलग अलग मान हैं तो चित्र 15.2 में उल्लिखित x की रेंज बदल जायेगी।



चित्र 15.2 : प्रसामान्य वक्र के नीचे क्षेत्रफल

15.2.1 मानक प्रसामान्य वक्र

हमने पिछले इकाई में देखा कि किसी भी सतत् प्रायिकता बंटन या प्रायिकता घनत्व फलन के लिए वक्र इस प्रकार होना चाहिए कि $x = x_1$ और $x = x_2$ के साथ की दो कोटियों से बद्ध वक्र के नीचे क्षेत्रफल इस प्रकार की प्रायिकता देगा कि यादृच्छिक चर का मान $x = x_1$ और $x = x_2$ के बीच का होगा। अतः प्रसामान्य वक्र के लिए

$$p(x_1 < X < x_2) = \int_{x_1}^{x_2} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

निस्संदेह यह प्रायिकता दो प्राचलों μ और σ के मानों पर निर्भर करती है। लेकिन, प्रसामान्य बंटन के उपर्युक्त उल्लिखित समाकल (integral) को हल करना बेहद कठिन है। इसी कारण विभिन्न अंतरालों में प्रसामान्य चर के मानों की तत्काल प्रायिकताओं की प्राप्ति के लिए प्रसामान्य वक्र क्षेत्रफल की सारणी बनाना ज़रूरी है। लेकिन μ और σ के मानों के प्रत्येक विचारणीय संयोजन के लिए अलग सारणी निर्मित करना पूरी तरह निरर्थक है। सौभाग्यवश इस जटिल कार्य को हल करने में सांख्यिकी में मानक परिणाम का अनुप्रयोग सार्थक सिद्ध होता है। पिछली इकाई में हमने यह देखा और सिद्ध किया था। आइए अब उस परिणाम को संक्षेप में दोहराएं। हमने देखा था कि *दिए गए माध्य और मानक विचलन वाले किसी चर के लिए, जब कभी भी माध्य को चर से घटा कर प्राप्त परिणाम को मानक विचलन से विभाजित किया जाता है; तो परिणामी चर का माध्य शून्य के बराबर और मानक विचलन एक के बराबर होता है।*

अतः यदि X, माध्य (प्रत्याशा) μ और मानक विचलन σ वाला चर है तब $z = \dfrac{X-\mu}{\sigma}$ का माध्य शून्य के बराबर और मानक विचलन एक के बराबर है।

इसका अर्थ है कि μ और σ के विभिन्न संयोजन वाले प्रसामान्य चरों को माध्य 0 और मानक विचलन 1 वाले अनोखे प्रसामान्य चर के रूप में परिवर्तित किया जा सकता है।

अतः, यदि X माध्य (प्रत्याशा) μ और मानक विचलन σ वाला प्रसामान्य चर है तब μ और σ के किसी भी संयोजन के लिए $z = \dfrac{X-\mu}{\sigma}$ सदैव 0 माध्य और 1 मानक विचलन वाला प्रसामान्य चर है।

सांकेतिक रूप से, यदि $X \sim N(\mu, \sigma)$

तो किसी भी μ और σ के लिए

$z = \dfrac{X-\mu}{\sigma} \sim N(0,1)$ होगा।

ऐसे परिवर्तित प्रसामान्य चर को मानक प्रसामान्य विचर (standard normal variate) कहते हैं। *मानक प्रसामान्य विचर* z का प्रायिकता घनत्व फलन है;

$$p(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \quad \text{जहाँ} \quad -\infty < z < \infty$$

एक बार मानक प्रसामान्य विचर प्राप्त कर लेने पर μ और σ के विभिन्न संयोजनों के लिए प्रायिकता क्षेत्र प्राप्त करने का जटिल कार्य बहुत ही सरल बन जायेगा। आइए देखें कि यह कैसे होगा। हमें पता होना चाहिए कि मानक प्रसामान्य विचर का अनूठा माध्य 0 और अनूठा मानक विचलन 1 है। इसका अर्थ है कि हम ऐसे अनूठे मानक प्रसामान्य विचर के प्रायिकता क्षेत्रफलों के लिए सारणी बना सकते हैं और माध्य और मानक विचलन के किसी भी संयोजन वाले सामान्य चर की प्रायिकता प्राप्त करने में उसका प्रयोग किया जा सकता है। यहाँ ध्यान रखने योग्य बात है कि दिए गए प्रसामान्य चर को मानक प्रसामान्य विचर में परिवर्तित किया जाना है।

विभिन्न क्षेत्रफलों (या प्रायिकताओं) के लिए ऐसी सारणी को मानक प्रसामान्य विचर के लिए संकलित किया जा चुका है (देखें सारणी 15.1 : मानक प्रसामान्य वक्र के नीचे क्षेत्रफल) हम सांख्यिकी में इसका कितनी ही दशाओं में प्रयोग करते हैं। माध्य और मानक विचलन वाले किसी प्रसामान्य चर की अपेक्षित प्रायिकता के परिकलन के लिए, दिए गए अंतराल की ऊपरी और निम्न सीमा, (मान लीजिए $x = x_1$ और $x = x_2$) को z-मानों, ($z = z_1$ और $z = z_2$) में परिवर्तित कर दिया जाता है और इस इकाई के अंत में दी गई सारणी 15.1 से सुसंगत क्षेत्रफल की प्राप्ति की जाती है।

ध्यान रहे कि मानक प्रसामान्य वक्र सममित है और इसका क्षेत्रफल 1.0 है। चूंकि z मान का परिसर $-\infty$ और ∞ के बीच का है अतः हम पाते हैं कि $0 < z < \infty$ के बीच क्षेत्रफल 0.5 (अर्थात मानक प्रसामान्य वक्र के नीचे क्षेत्रफल का आधा) है। इसी तरह, $-\infty < z < 0$ के बीच का क्षेत्रफल 0.5 है। चूंकि मानक प्रसामान्य वक्र सममित है, इसलिए हमें इसका एक और भी फायदा है : वक्र के नीचे क्षेत्रफल दोनों तरफ से समान हैं। सारणी 15.1 में z के अलग–अलग मानों का क्षेत्रफल दिया गया है।

यदि हम सारणी 15.1 के स्तंभ 1 में देखें तो हम पायेंगे कि z के मानों का परिसर 0.0 से 3.0 तक है। प्रत्येक मान के तदनुरूप ऐसे 10 स्तंभ हैं जो .00, .01, ...09 के रूप में हैं। ये स्तंभ दशमलव के बाद के दूसरे अंक को दर्शाते हैं। जैसे यदि $z = 0.45$ है तब हम 0.4 के तदनुरूप वाली पंक्ति को देखेंगे। इस पंक्ति में हम दायीं ओर बढेंगे और ऐसे स्तंभ को देखेंगे जो .05 को दर्शाता है। सारणी 15.1 में हम पायेंगे कि जब $z = 0.45$ है तब ढ़का हुआ क्षेत्र 0.1736 है। ध्यान दीजिए जब $z = -0.45$, तब मानक प्रसामान्य वक्र के नीचे क्षेत्रफल दुबारा से 0.1736 है। एक अन्य उदाहरण में, जब $z = 1.31$ है तब मानक प्रसामान्य वक्र के नीचे क्षेत्रफल 0.4049 है। सैद्धांतिक रूप से $-\infty$ और ∞ के बीच z का कोई भी मान हो सकता है। जब $z = 3.00$ तब ढ़का हुआ क्षेत्र 0.4990 है। इसलिए सारणी 15.1 में $z > 3.00$ के लिए क्षेत्रफल नहीं दिए गए।

आइए अब प्रसामान्य क्षेत्रफल सारणी (normal area table) के अनुप्रयोगों के लिए, कुछ उदाहरणों पर विचार करें।

उदाहरण 15.1

मानक प्रसामान्य वक्र के नीचे क्षेत्रफल वाली तालिका का प्रयोग कर निम्नलिखित मामलों में मानक प्रसामान्य वक्र के नीचे क्षेत्रफल ज्ञात कीजिए।

क) $z = 0$ और $z = 1.8$ के बीच

ख) $z = -0.25$ और $z = 0$ के बीच

ग) $z = -0.25$ और $z = 2.25$ के बीच

उत्तर :

क) सारणी 15.1 में आइए z कॉलम के नीचे की तरफ बढ़े जब तक कि हम 1.8 तक न पहुंच जायें। अब दाहिनि ओर 0 के नीचे वाले स्तंभ में देखें। यहाँ हम 0.4641 अंकित पाते हैं। यही अपेक्षित क्षेत्रफल है।

ख) मानक प्रसामान्य वक्र अपने माध्य के संदर्भ में सममित है; तब $z = -0.25$ और $z = 0$ के बीच की आपेक्षित प्रायिकता की प्राप्ति तालिका से $z = 0$ और $z = 0.25$ के बीच का क्षेत्रफल ज्ञात करके की जा सकती है। आइए अब z लिखे कॉलम के नीचे, नीचे की तरफ 0.2 तक पहुंचने तक बढ़ते रहें। अब हम .05 लिखे स्तंभ के दायें ओर

देखते हैं। यहाँ हम 0.0987 के बराबर की प्रविष्टि प्राप्त करते हैं। यही अपेक्षित क्षेत्रफल है।

ग) यह स्पष्ट है कि अपेक्षित क्षेत्र है : ($z = -0.25$ और $z = 0$ के बीच का क्षेत्रफल) + ($z = 0$ और $z = 2.25$ के बीच का क्षेत्रफल)

= ($z = 0$ और $z = 0.25$ के बीच का क्षेत्रफल) (सममिता द्वारा) + ($z = 0$ और $z = 2.25$ के बीच का क्षेत्रफल)

$= 0.1985 + 0.4878 = 0.6863$

उदाहरण 15.2

किसी फैक्टरी में 120 मज़दूरों के प्रतिदर्श में दैनिक मज़दूरी प्राप्त करने का माध्य और मानक विचलन क्रमशः 11.35 रुपए और 3.03 रुपए हैं। फैक्टरी में 9 रु. और 17रु. के बीच की मज़दूरी प्राप्त करने वाले मज़दूरों का प्रतिशत बताइए। मान लीजिए कि मज़दूरी प्रसामान्य बंटित है।

मान लीजिए x मज़दूरी (दर्शाने वाला) यादृच्छिक चर है। तब x एक प्रसामान्य चर है जिसका माध्य $\mu = 11.35$ और मानक विचलन $\sigma = 3.03$ मानक प्रसामान्य विचर है।

$$z = \frac{x - 11.35}{3.03}$$

जब $x = 9, z = \dfrac{9 - 11.35}{3.03} = \dfrac{-2.35}{3.03} = -0.78$

जब $x = 17, z = \dfrac{17 - 11.35}{3.03} = \dfrac{5.65}{3.03} = 1.86$

$\therefore P(9 \le x \le 17) = P(-0.78 \le z \le 1.86)$

$= P(-0.78 \le z \le 0) + P(0 \le z \le 1.86)$

$= P(0 \le z \le 0.78) + P(0 \le z \le 1.86)$

$= 0.2823 + 0.4686$

$= 0.7509 = 75.09\%$

अतः 9 रु. और 17 रु. के बीच मज़दूरी पाने वाले 75.09% मज़दूर हैं।

15.2.2 द्विपद बंटन का प्रसामान्य सन्निकटन

कभी कभार सांख्यिकी में एक बंटन को दूसरे बंटन के सीमित रूप में देखा जाता है। जैसे, इकाई 14 में हमने सीखा था द्विपद बंटन में जब सफलता की प्रायिकता, p काफी कम है और अभिप्रयोगों की संख्या n बहुत अधिक हो तब प्रत्याशा $\mu = np$ परिमित परिमाण हो जाती है। ऐसे मामलों में द्विपद बंटन पाइसो बंटन की ओर प्रवृत है। आपको याद होगा कि द्विपद बंटन और पाइसो बंटन दोनों सतत् बंटन है। लेकिन, आवश्यक नहीं है कि सतत् द्विपद बंटन का सीमित स्वरूप सदा ढंग से सतत् पाइसो बंटन ही हो। जब n काफी अधिक हो और p, 0 या 1 के बहुत ही निकट न हो तब द्विपद बंटन, सतत् प्रसामान्य बंटन का अनुसरण करता है और इसी कारण प्रसामान्य बंटन का अक्सर द्विपद बंटन के सन्निकटन के रूप में प्रयोग किया जाता है। हम पहले ही देख चुके हैं कि मानक प्रसामान्य

सारणी, निश्चित अंतराल के भीतर आने वाले प्रसामान्य चर की प्रायिकता प्राप्त करने में काफी उपयोगी है। अब हम ऐसा परिणाम दे सकते हैं जो यादृच्छिक चर के द्विपद गुणधर्मों को सन्निकट करने के लिए मानक प्रसामान्य वक्र के नीचे वाले क्षेत्रफल के प्रयोग में हमारे लिए सहायक होगा।

यदि X द्विपद चर है जिसका माध्य $\mu = np$ और प्रसरण $\sigma^2 = npq$ तब

$$z = \frac{X - np}{\sqrt{npq}}$$

प्रसामान्य बंटन की ओर प्रवृत है और जिसका माध्य 0 और मानक विचलन 1 है, क्योंकि n अनंत संख्याओं की ओर प्रवृत है।

सांकेतिक रूप से

$$\lim_{n \to \infty} \frac{X - np}{\sqrt{npq}} \sim N(0,1)$$

यह देखा गया है कि प्रसामान्य बंटन द्विपद बंटन के लिए अच्छा सन्निकटन प्रदान करता है चाहे n की संख्या काफी अधिक न हो पर p किसी वजह से 0.5 के निकट हो।

आइए अब एक उदाहरण पर विचार करें।

उदाहरण 15.3

मान लीजिए किसी विशेष प्रकार की मशीन के दोषपूर्ण होने की प्रायिकता 0.4 है। गुणवत्ता नियंत्रण इंस्पैक्टर दोषपूर्ण मशीनों की पहचान करने के लिए ऐसी 15 मशीनों की जांच करता है। प्रायिकता बताइए कि इंस्पैक्टर 4 दोषपूर्ण मशीनों की पहचान कर लेगा।

जैसा कि हम जानते हैं कि इकाई 13 में द्विपद बंटन पर आधारित हमारी चर्चा में, हमने देखा था कि उपलब्ध मशीनों में दोषपूर्ण मशीनों का बंटन द्विपद चर है। यहाँ $n = 15, p = 0.4$ और $q = 1 - p = 0.6$

इसलिए 4 दोषपूर्ण मशीनों की प्रायिकता है

$$^{15}c_4(0.4)^4(0.6)^{11} = \frac{15!}{4!11!}(0.4)^4(0.6)^{11} = 1365 \times 0.0256 \times 0.0036279 = 0.1268$$

अब मान लीजिए, हम प्रसामान्य सन्निकटन द्वारा अपेक्षित प्रायिकता ज्ञात करना चाहते हैं।

तब

$np = 15 \times 0.4 = 6,\ npq = np(1-p) = 15 \times 0.4 \times 0.6 = 3.6$ और $\sqrt{npq} = \sqrt{3.6} = 1.897$

यहाँ है $\quad z = \dfrac{X - np}{\sqrt{npq}}$

जो सन्निकटतः मानक प्रसामान्य विचर है। लेकिन मानक प्रसामान्य विचर सतत् यादृच्छिक चर है। हम जानते हैं कि ऐसे यादृच्छिक चर के लिए कि यह किसी विशिष्ट मान को ले, इसकी प्रायिकता का निर्धारण नहीं किया जा सकता। केवल अंतराल के भीतर वाले यादृच्छिक चर की प्रायिकता की प्राप्ति की जा सकती है। अतः इसकी प्रायिकता कि द्विपद चर का मान 4 हो इसे अपेक्षित असामान्य सन्निकटन के लिए, ऐसे प्रसामान्य चर की

प्रायिकता में परिवर्तित करना होगा जो मान 4 के आसपास के अंतराल में है। चूंकि द्विपद चर शून्य और धनात्मक पूर्णांक है इसलिए 4 से तुरंत पहले का मान और 4 के तुरंत आगे का मान जोकि विचाराधीन द्विपद चर ले सकता है, क्रमशः 3 और 5 है। जिसके फलस्वरूप मान 4 के बराबर का मान लेने वाले द्विपद चर की प्रायिकता को तार्किक रूप से अंतराल (3.5, 4.5) के बीच आने वाले प्रसामान्य चर की प्रायिकता के सन्निकट किया जा सकता है। तब अपेक्षित प्रायिकता $x_1 = 3.5$ और $x_2 = 4.5$ के बीच के प्रसामान्य वक्र के नीचे क्षेत्रफल के बराबर होगी। यदि z-मानों को परिवर्तित किया जाए तो प्राप्त होगा

$$z_1 = \frac{x_1 - np}{\sqrt{npq}} = \frac{3.5-6}{1.897} = 1.32 \text{ और } z_2 = \frac{x_2 - np}{\sqrt{npq}} = \frac{4.5-6}{1.897} = -0.79$$

यदि X द्विपद चर है और z मानक प्रसामान्य विचर, तब

$P(X = 4) = P(-1.32 \le z \le -0.79) = 0.1214$ (मानक प्रसामान्य सारणी से प्राप्त)

हम देख सकते हैं कि 0.1214 का प्रसामान्य सन्निकटन, द्विपद बंटन से प्राप्त 4 दोषपूर्ण मशीनों के लिए 0.1268 की वास्तविक प्रायिकता के काफी निकट है।

बोध प्रश्न 1

1) निम्नलिखित स्थितियों में प्रसामान्य वक्र के नीचे क्षेत्रफल ज्ञात कीजिए।

   क) $z = 1.55$ और $z = 2.55$ के बीच

   ख) $z = -1.5$ के बायें तरफ

   ग) $z = 2.5$ के दायें तरफ

   ....

   ....

   ....

2) 1000 पुरुषों की माध्य ऊँचाई 68 इंच और मानक विचलन 5 इंच है। यदि ऊँचाई प्रसामान्य रूप से बंटित है तो बताइए कि कितने पुरुषों की ऊँचाई 67 इंच और 69 इंच के बीच है?

   ....

   ....

   ....

3) एक कंपनी वर्ष में 5000 बैटरियों की बिक्री करती है और इन बैटरियों की 24 महीनों की गॉरंटी है। बैटरियों की जीवन अवधि 34 महीनों के बराबर के माध्य और 5 महीनों के बराबर के मानक विचलन के संन्निकटतः प्रसामान्य रूप से आकलित है। ऐसी बैटरियों की संख्या बताइए जिन्हें गॉरंटी की समयावधि के दौरान बदला जाना है।

   ....

   ....

   ....

15.3 कुछ अन्य सतत् बंटन

कुछ ऐसे अन्य सतत् प्रायिकता बंटन भी हैं जिनका सांख्यिकी की विविध शाखाओं में विशेष महत्व है। खंड 7 में हम सांख्यिकीय अनुमितियों का अध्ययन करेंगे। इस खंड में हम प्रसामान्य बंटन के अलावा, बहुधा तीन अन्य सतत् बंटनों की संकल्पनाओं का भी प्रयोग करेंगे। जो हैं : काई स्क्वैयर जिसे (काई वर्ग) भी कहते हैं और χ^2 **बंटन** से सूचित करते हैं और 'स्टूडैन्ट' – 'टी' बंटन और एफ (F) बंटन। इस अनुभाग में हम इन बंटनों की संक्षेप में चर्चा करेंगे। सर्वप्रथम हम स्वतंत्रता की कोटि अर्थात इसकी सामान्य संकल्पना पर नज़र डालेंगे क्योंकि उपर्युक्त सभी बंटनों में इसका अनुप्रयोग किया जाता है।

15.3.1 स्वतंत्रता की कोटि

उपर्युक्त उल्लिखित बंटनों के संदर्भ में बहुधा हम *स्वतंत्रता की कोटि* (degrees of freedom) नामक शब्द को पढ़ते हैं। आइए इस शब्द को समझने का प्रयास करें। सामान्य शब्दों में स्वतंत्रता की कोटि से आशय अलग अलग प्रकार की स्वतंत्र सूचना से है जिससे किन्हीं प्रेक्षणों के समुच्चय की किसी विशेषता (जैसे, प्रसरण) को परिकलित करना है। यहाँ हम एक उदाहरण पर विचार करेंगे।

आइए 5 प्रेक्षणों : 4, 7, 12, 3 और 15 पर गौर करें। इनका माध्य $\bar{X}$, 8 है। प्रसरण के परिकलन में माध्य से प्रेक्षणों के वर्गों का योगफल प्रयोग किया जाता है :

$(4-8)^2 + (7-8)^2 + (12-8)^2 = (3-8)^2 + (14-8)^2$

$= (-4)^2 + (-1)^2 + (4)^2 + (-5)^2 + (6)^2$

समांतर माध्य ($\bar{X}$) के गुणधर्मों से हम जानते हैं कि कोष्ठकों के भीतर के मानों का योगफल शून्य के बराबर होना चाहिए अर्थात सामान्य रूप में $\sum_{i=1}^{n}(x_i - \bar{X}) = 0$ इसका अर्थ है कि प्रसरण के परिकलन में, यदि कोष्ठकों के भीतर पहले चार पद, क्रमशः –4, –1, 4 और –5 है तो पाँचवां पद कुछ और नहीं बल्कि 6 ही होगा। अतः इस उदाहरण में, कोष्ठकों के भीतर हमारे पास पांच अलग–अलग सूचनाएं नहीं हैं। कोष्ठक के भीतर पांचवी सूचना अर्थात '6' अपने से संबंधित कोष्ठकों के भीतर की पहली चार सूचनाओं पर आधारित है।

यदि इसका और विस्तार करें तो हम ऐसे व्यक्ति के बारे में सोच सकते हैं जिसे किसी एक प्रेक्षण की जानकारी नहीं है। उसे केवल इतना ही बताया गया है कि कुल 5 प्रेक्षण हैं और पांच प्रेक्षणों के माध्य से प्राप्त मानों के पहले 4 विचलन हैं क्रमशः –4, –1, 4 और –5 अब उसे 5 प्रेक्षणों के प्रसरण को परिकलित करने को कहा जाता है। यदि वह नियम जानती है कि चर के मानों के विचलन का योगफल जो उनके समांतर माध्य से प्राप्त होता है, सदैव शून्य के बराबर होता है तो वह 5वें मान के विचलन को उसके समांतर माध्य से प्राप्त करके 6 बताएंगी। इसके बाद वह इन विचलनों के वर्ग प्राप्त करके उन्हें जोड़ देगी और यही अपेक्षित प्रसरण के परिकलन का अंतिम चरण है। अंतिम चरण में उनके समांतर माध्य के लिए मानों के औसत परिक्षेपण के अनुमान लगाने के लिए इन्हें प्रेक्षणों की संख्या से विभाजित करना होता है। (यहाँ अपेक्षित माप के रूप में हम प्रसरण को लेते हैं) ऐसा सन्निकट मान कौन सा है जिसे, प्रेक्षणों की संख्या के लिए लेना चाहिए? क्या यह 5 है? आइए इस पर गौर करें। हमने देखा कि पांचवां विचलन, पहले चार विचलनों से निर्धारित किया जाता है। और जिसका अर्थ है कि ऐसी पांच विभिन्न सूचनाएं हैं जो उनके समांतर

माध्य के लिए, दिए गए पांच मानों का परिक्षेपण बनाती हैं। इसलिए औसत परिक्षेपण अर्थात प्रसरण का पता लगाने के लिए, पांच विचलनों के वर्गमूल के योग को 5 की बजाए 4 से विभाजित करना चाहिए। अतः इस उदाहरण में स्वतंत्रता की कोटि 4 है। यदि हम उपर्युक्त उदाहरण को n प्रेक्षणों के प्रसरण की प्राप्ति के लिए व्यापक बनाएं तो $\sum_{i=1}^{N}(x_i - \bar{X}) = 0$ प्रतिबंध के कारण यहाँ n–1 स्वतंत्रता की कोटियाँ हैं। जिसके परिणाम स्वरूप, प्रसरण के लिए $\sum_{i=1}^{N}(x_i - \bar{X})^2$ को $n-1$ से विभाजित किया जाता है अर्थात $S^2 = \frac{1}{n-1}\sum_{i=1}^{n=1}(x_i - \bar{X})^2$, जहाँ, S^2 प्रसरण है।

उपर्युक्त चर्चा से हम कह सकते हैं कि किसी भी लक्षण के परिकलन से संबंधित हमें प्राप्त स्वतंत्रता की कोटियाँ, ऐसी संख्या के बराबर है, जो हमें (प्रेक्षणों की संख्या – अपेक्षित अभिलक्षण के परिकलन के लिए लगाई गई प्रतिबंधों की संख्या) प्राप्त करके मिलती है। सांकेतिक रूप से $d.f. = n - r$ जहाँ $d.f.$ स्वतंत्रता की कोटि है, n प्रेक्षणों की संख्या है और r प्रतिबंधों की संख्या है।

हमें यहाँ ध्यान देना चाहिए कि जब प्रसरण या इसके धनात्मक वर्गमूल, अर्थात मानक विचलन के परिकलन के लिए उपलब्ध प्रेक्षणों की संख्या बहुत बड़ी हो तब हम $n-1$ की बजाए, बहुधा n से ही $\sum_{i=1}^{N}(x_1 - \bar{X})^2$ को विभाजित करते हैं। लेकिन, सटीक शब्दों में हमें $n - 1$ से $\sum_{i=1}^{N}(x_1 - \bar{X})^2$ को विभाजित करना चाहिए (विशेषरूप से जब n छोटा हो)।

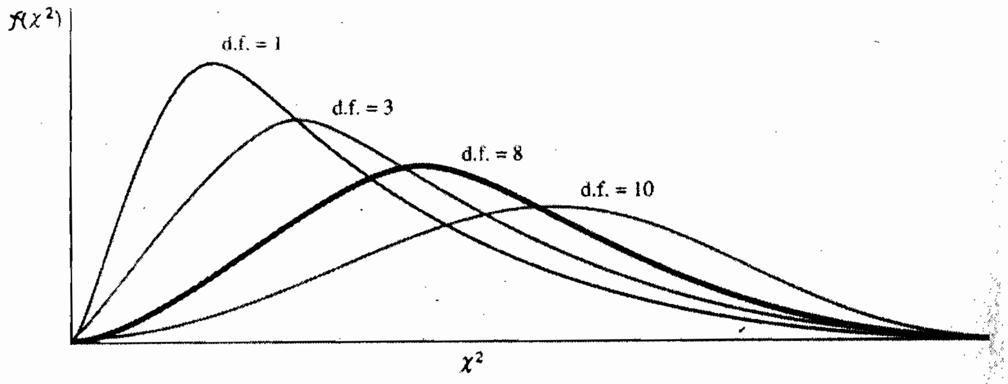
15.3.2 χ^2 (काई-स्क्वैयर) बंटन

मान लीजिए, X प्रसामान्य चर है जिसकी माध्य (प्रत्याशा) μ और मानक विचलन σ है, तब $z = \frac{X-\mu}{\sigma}$ मानक प्रसामान्य विचर अर्थात $z \sim N(0,1)$ है। यदि हम z का वर्ग करें अर्थात

$z^2 = \left(\frac{X-\mu}{\sigma}\right)^2$, तब z^2 ऐसे χ^2 चर के रूप में बंटित है जिसकी स्वतंत्रता की कोटि 1 है और इसे χ_l^2 के रूप में अभिव्यक्त किया जाता है।

यह स्पष्ट है कि चूंकि χ_1^2 वर्गित पद है; अतः $-\infty < z < -\infty$ के लिए χ^2 का मान 0 तथा ∞ के बीच रहेगा। (क्योंकि वर्गित पद के नकारात्मक मान नहीं हो सकते)। आगे, चूंकि z का माध्य, 0 के बराबर है इसलिए z के अधिकतम मान 0 के सन्निकट होंगे। जिसके परिणामस्वरूप χ^2 चर के अधिकाधिक प्रायिकता घनत्व शून्य के निकट हो।

उपर्युक्त उल्लिखित परिणाम को यदि सामान्य रूप में दें तो यदि $z_1, z_2, \ldots, z_k$ स्वतंत्र मानक प्रसामान्य विचर (अर्थात 0 माध्य और 1 प्रसरण वाले प्रसामान्य चर हैं) तब $z = \sum_{i=1}^{k} z_i^2$ ऐसा χ^2 चर है जिसकी स्वतंत्रता की कोटि k है और जिसे χ_k^2 द्वारा सूचित किया जाता है। निम्नलिखित चित्र 15.3, विभिन्न स्वतंत्रता की कोटि वाले χ^2 चरों के लिए प्रायिकता वक्रों को दर्शाता है।



चित्र 15.3: काई स्क्वैयर प्रायिकता वक्र

हमें χ^2 बंटन की निम्नलिखित विशेषताओं को ध्यान में रखना चाहिए।

1) जैसा कि चित्र 15.3 दर्शाता है, χ^2 धनात्मक विषम बंटन है। इसकी विषमता (तिरछेपन) की कोटि, इसकी स्वतंत्रता की कोटियों पर निर्भर करती है। निम्न स्वतंत्रता की कोटियों के लिए बंटन काफी विषम है। स्वतंत्रता की कोटियों की संख्या जैसे–जैसे बढ़ती जाती है, बंटन भी धीरे धीरे सममित होता चला जाता है। असल में, 100 से अधिक स्वतंत्रता की कोटियों के लिए, $\sqrt{2X^2} - \sqrt{(2k-1)}$ चर को मानक प्रसामान्य विचर के रूप में देखा जा सकता है जहाँ k स्वतंत्रता की कोटि है।

2) काई–स्क्वैयर बंटन का माध्य k है और इसका प्रसरण $2k$ है जहाँ k स्वतंत्रता की कोटि है।

3) यदि Z_1 और Z_2 दो स्वतंत्र काई–स्क्वैयर बंटन हैं जिनकी स्वतंत्रता की कोटि क्रमशः k_1 और k_2 है, तो $Z_1 + Z_2$ भी काई–स्क्वैयर चर होगा, जिसकी स्वतंत्रता की कोटि $k_1 + k_2$ के बराबर हैं।

प्रसामान्य बंटन के मामले की भांति χ^2 बंटन के लिए भी इसी तरह की सारणी बनाई गई है (देखें इस इकाई के अंत में दी गई तालिका 15.2)। हमें स्वतंत्रता की विभिन्न कोटि के लिए, χ^2 चर की अपेक्षित प्रायिकता की प्राप्ति के लिए इस तालिका को सिर्फ देखना होता है।

सारणी 15.2 में df स्वतंत्रता की कोटियां को दर्शाता है। जबकि स्तंभ $\chi^2_{0.05}$ और $\chi^2_{0.01}$ क्रमशः 5% और 1% (सार्थकता के लिए) χ^2 मानों को सूचित करता है। सार्थकता के स्तर की संकल्पना का वर्णन हम इकाई 19 में करेंगे।

निम्नलिखित उदाहरण, काई–स्क्वैयर सारणी के प्रयोग को दर्शाता है।

उदाहरण 15.4

34 या अधिक के χ^2 मान प्राप्त करने की प्रायिकता क्या होगी यदि स्वतंत्रता की कोटि 25 है?

सारणी 15.2 से हम देख सकते हैं कि यदि हम 25 के अंक तक पहुंचने के लिए स्वतंत्रता की कोटि कॉलम में नीचे की ओर बढ़ते हैं तो 34 से निकटतम अंक, वहाँ 34.08747 है और 34.08704 की प्रायिकता जैसा कि हम सारणी से देख सकते हैं, 0.10 है। अतः अपेक्षित प्रायिकता 0.10 है।

15.3.3 'स्टूडैन्ट'–'टी' बंटन

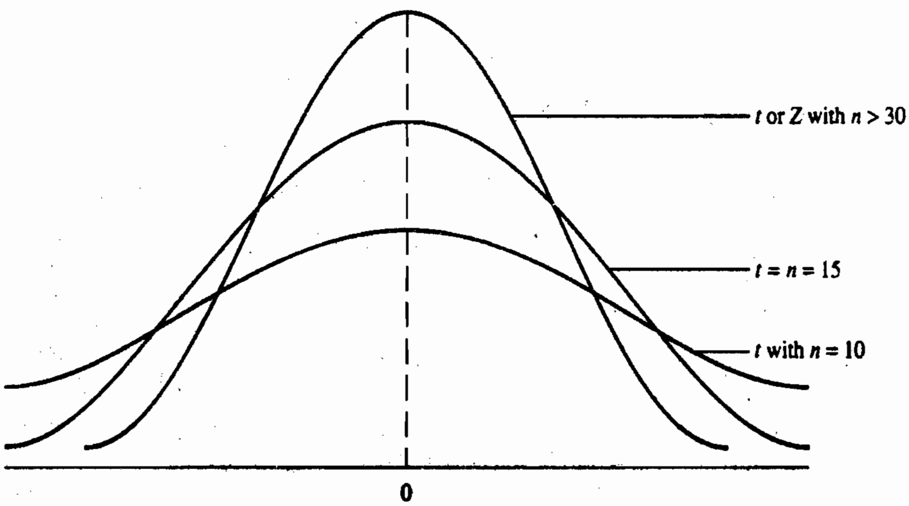
डब्ल्यू. एस. गौसेट ने टी–बंटन की प्रस्तुति की। यह एक रोचक कहानी है कि गौसेट, आयरलैंड में किसी ब्रीवरी में कर्मचारी थे और कंपनी के नियम के अनुसार कोई भी कर्मचारी अपने शोध परिणामों को अपने नाम से प्रकाशित नहीं कर सकता था। तब गौसेट ने उपनाम 'स्टूडैंट' को अपनाया और इस बंटन के बारे में अपने अध्ययन आधारित परिणामों को इसी नाम के अंतर्गत प्रकाशित कराया। तभी से इस बंटन को स्टूडैंट–टी बंटन या सरल रूप से टी (t) बंटन के रूप में जाना जाता है।

यदि z_1 मानक प्रसामान्य विचर अर्थात $z_1 \sim N(0, 1)$ है और z_2 अन्य स्वतंत्र प्रसामान्य विचार है जो ऐसे काई स्क्वैयर बंटन का अनुगमन करता है जिसकी k (स्वतंत्रता की कोटि) है

अर्थात $z_2 \sim \chi_k^2$, तब चर $t = \dfrac{z_1}{\sqrt{(z_2/k)}} = \dfrac{z_1\sqrt{k}}{\sqrt{z_2}}$

k कोटि बंटन वाले स्टूडैंट–टी बंटन का अनुगामी है।

अलग–अलग स्वतंत्रता की कोटि वाले स्टूडैंट–टी बंटन से संबंधित प्रायिकता वक्रों को चित्र 15.4 में दर्शाया गया है।



चित्र 15.4: स्टूडैन्ट–टी प्रायिकता वक्र

इस बंटन की मुख्य विशेषताएं हैं;

1) जैसा कि चित्र 15.4 में हम देख सकते हैं प्रसामान्य बंटन की भांति, स्टूडैन्ट–टी बंटन भी सममित है और इसके विचरण का परिसर भी $-\infty$ से $+\infty$ के बीच का ही है; लेकिन प्रसामान्य बंटन की तुलना में यह अधिक सपाट है। हमें यहाँ ध्यान देना चाहिए कि जैसे जैसे स्वतंत्रता की कोटि में बढ़ोत्तरी होती है, स्टूडैन्ट–'टी' बंटन प्रसामान्य बंटन का अनुगमन करता है।

2) स्टूडैन्ट–'टी' बंटन का माध्य शून्य है और इसका प्रसरण $\dfrac{k}{(k-2)}$ है जहाँ k स्वतंत्रता की कोटि है।

प्रसामान्य बंटन की भांति, स्टूडैन्ट–टी बंटन का भी प्रयोग बहुधा सांख्यिकीय अनुमितियों और परिकल्पना-परीक्षण के लिए किया जाता है। इसकी चर्चा हम खंड–7 में करेंगे। इस

कार्य में इसके धनत्व फलन का समाकलन शामिल है और जो एक लंबा कार्य है। जिसके फलस्वरूप इस मामले में भी, प्रसामान्य बंटन की भांति, संदर्भ तालिकाएं निर्मित की गई है।

इस सारणी का प्रयोग सीखने के लिए अब हम एक उदाहरण पर विचार करेंगे।

उदाहरण 15.5

यदि स्वतंत्रता की कोटि 10 के बराबर है तो (i) 2.764 या अधिक (ii) –2.764 या निम्न के t मान प्राप्त करने की प्रायिकता क्या है?

उत्तर

i) स्टूडेंट–टी बंटन वाली सारणी में, हम स्वतंत्रता की कोटि वाले स्तंभ में नीचे की ओर बढ़ते हैं और 10 के अंक तक पहुंचते हैं और तब साथ वाली पंक्ति में 2.764 के अंक का पता लगाते हैं। 0.01 वाला निम्न प्रायिकता अंक, अपेक्षित प्रायिकता है।

ii) चूंकि स्टूडेंट–टी बंटन सममित है, इसलिए –2.764 या इससे कम का t मान प्राप्त करने की प्रायिकता भी 0.01 है।

15.3.4 एफ (*F*) बंटन

इस इकाई में अंतिम सतत् प्रायिकता बंटन जिस पर अब हम विचार करेंगे, वह F बंटन है।

यदि z_1 और z_2 दो काई–स्क्वैयर चर हैं और जो स्वतंत्रता की कोटि क्रमशः k_1 और k_2 से स्वतंत्र रूप से बंटित हैं, तब चर

$$F = \frac{z_1 / k_1}{z_2 / k_2}$$

F बंटन का अनुगमन करता है जिसकी स्वतंत्रता की कोटि क्रमशः k_1 और k_2 है। चर को F_{k_1,k_2} से सूचित किया जाता है जहाँ पादांक k_1 और k_2 काई–स्क्वैयर चरों से संबंद्ध स्वतंत्रता की कोटि हैं।

यहाँ हम देखते हैं कि k_1 अंश स्वतंत्रता की कोटि है और इसी तरह k_2 हर स्वतंत्रता की कोटि है।

F बंटन के कुछ महत्वपूर्ण गुणधर्मों का उल्लेख इस प्रकार है;

1) काई–स्क्वैयर बंटन की भांति, F बंटन भी दायीं ओर विषम है। लेकिन जैसे–जैसे k_1 और k_2 बढ़ते हैं, F बंटन, प्रसामान्य बंटन का अनुगामी बन जाता है।

2) F बंटन का माध्य $k_1/(k_2-2)$ है जो $k_2 > 2$ के लिए परिभाषित है और इसका प्रसरण $\dfrac{2k_2^2(k_1+k_2-2)}{k_1(k_2-2)^2(k_2-4)}$ है जो $k_2 > 4$ के लिए परिभाषित है।

3) ऐसा F बंटन जिसकी अंश और हर स्वतंत्रता की कोटि क्रमशः 1 और k है, स्टूडेंट–टी बंटन का वर्ग है, जिसकी स्वतंत्रता की कोटि k है सांकेतिक रूप से, $F_{1,k} = t_k^2$

4) बड़े हर वाली स्वतंत्रता की कोटि k_2 के लिए, स्वतंत्रता की कोटि k_1 और F मान का गुणनफल सन्निकटतः ऐसे काई–स्क्वैयर मान के बराबर होगा जिसकी स्वतंत्रता की कोटि k_1 अर्थात $k_1F = \chi^2_{k1}$ है।

जैसा कि अन्य संतत प्रायिकता बंटनों के संदर्भ में हमने उल्लेख किया था, F बंटन का भी सांख्यिकीय अनुमितियों और परिकल्पना परीक्षणों में विस्तृत प्रयोग किया जाता है और ऐसे प्रयोगों के लिए, F प्रायिकता वक्र के नीचे क्षेत्रफल प्राप्त करने की और इसके बाद F घनत्व फलन के समाकलन की आवश्यकता पड़ती है। लेकिन इस स्थिति में भी F सारणी के प्रावधान से हमारा कार्य सुविधाजनक बन जाता है। (इस इकाई के अंत में दी गई सारणी 15.4 देखें)।

15.3.5 प्रसामान्य बंटन से संबंधित बंटन

हम काई-स्क्वैयर, स्टूडैन्ट–टी बंटन और एफ (F) बंटनों की विशेताओं को पहले ही समझ चुके हैं कि अधिक संख्या वाली स्वतंत्रता की कोटि के लिए, ये बंटन, प्रसामान्य बंटन के अनुगामी हैं। जिसके फलस्वरूप, इन बंटनों को प्रसामान्य बंटन से संबंधित बंटनों के रूप में भी देखा जाता है। एक तरफ काई-स्क्वैयर बंटन, स्टूडैंट–टी बंटन और F बंटन के बीच का यह संबंध और दूसरी तरफ प्रसामान्य बंटन अर्थात इनके महत्वपूर्ण व्यावहारिक निहितार्थ हैं। जब स्वतंत्रता की कोटि अधिक होती है तो हम प्रायः काई-स्क्वैयर बंटन या स्टूडैंट–टी बंटन या F बंटन का अलग से प्रयोग करने के बजाए, प्रसामान्य बंटन का प्रयोग कर सकते हैं। इससे हमारा कार्य अपेक्षाकृत सरल हो जाता है।

बोध प्रश्न 2

1) 8 या अधिक वाले χ^2 मान प्राप्त करने की प्रायिकता क्या है, यदि स्वतंत्रता की कोटि 20 है?

....

....

....

....

2) यदि स्वतंत्रता की कोटि 25 के बराबर है तो 1.708 या अधिक का t मान प्राप्त करने की प्रायिकता क्या है?

....

....

....

....

....

3) यदि $k_1 = 10$ और $k_2 = 8$ है तो 5.8 या अधिक का F मान प्राप्त करने की प्रायिकता क्या है?

....

....

....

....

....

15.4 सारांश

इस इकाई में हमने कुछ सतत् प्रायिकता बंटनों का अध्ययन किया। इन बंटनों में से प्रसामान्य बंटन को सर्वाधिक महत्वपूर्ण माना गया है। हमने इसकी विशेषताओं का अध्ययन किया और इसके व्यावहारिक निहितार्थों को देखा। हमने मानक प्रसामान्य विचर की महत्वपूर्ण संकल्पना पर विचार किया। हमने प्रसामान्य बंटन से संबंधित समस्याओं के समाधान के लिए मानक प्रसामान्य वक्र के नीचे क्षेत्रफलों को ज्ञात करने में सारणी प्रयोग की तकनीक भी सीखी। प्रसामान्य बंटन के अतिरिक्त हमने तीन अन्य सतत् प्रायिकता बंटनों पर भी विचार किया और ये हैं, काई–स्क्वैयर बंटन, स्टूडैंट–टी बंटन और एफ बंटन। ये तीनों बंटनों में, स्वतंत्रता की कोटि की धारणा का प्रयोग शामिल है। इसलिए हमने स्वतंत्रता की कोटि की संकल्पना को समझाने का भी प्रयास किया है। हमने इन बंटनों की विशेषताओं का अध्ययन किया और देखा कि इन बंटनों से संबंधित सारणियों के प्रयोग से कैसे विविध स्थितियों में इन बंटनों को लागू किया जा सकता है। अंत में हमने देखा कि जब स्वतंत्रता की कोटि काफी अधिक हो तो ये तीनों बंटन प्रसामान्य बंटन के अनुगामी बन जाते हैं।

15.5 शब्दावली

काई-स्क्वैयर बंटन (Chi-square Distribution): यह एक असममित बंटन है जहाँ यादृच्छिक चर के लिए विचरण का परिसर शून्य से अनंत है। जब स्वतंत्रता की कोटि काफी अधिक हो तो यह प्रसामान्य बंटन का अनुगामी हो जाता है।

काई-स्क्वैयर चर (Chi-square Variable): ऐसा यादृच्छिक चर जो काई-स्क्वैयर बंटन का अनुगामी है।

सतत् प्रायिकता बंटन (Continuous Probability Distribution): यह सतत् यादृच्छिक चर का प्रायिकता बंटन है।

स्वतंत्रता की कोटि (Degrees of Freedom): इससे आशय विभिन्न प्रकार की सूचनाओं से है जिनकी किसी प्रेक्षण समुच्चय के कुछ अभिलक्षणों के परिकलन में आवश्यकता पड़ती है।

असंतत प्रायिकता बंटन (Discrete Probability Distribution): यह असंतत यादृच्छिक चर का प्रायिकता बंटन है।

एफ बंटन (*F* Distribution): यह एक असममित बंटन है जो दायीं ओर विषम है। जब स्वतंत्रता की कोटि काफी अधिक है तो यह भी प्रसामान्य बंटन का अनुगामी बन जाता है।

असतत् यादृच्छिक चर (Discrete Random Variable): ऐसा यादृच्छिक चर जिसके मान अनंत हैं या जिसका अनुक्रम अनंत संख्या वाला है (जैसे, 1,2,3, ...)।

एफ चर (*F* Variable): ऐसा यादृच्छिक चर जो एफ बंटन का अनुगामी है।

गॉउसीय वितरण (Gaussian Distribution): प्रसामान्य बंटन को इस नाम से भी जाना जाता है।

प्रसामान्य बंटन (Normal Distribution): सभी सैद्धांतिक प्रायिकता बंटनों में से सर्वाधिक प्रचलित बंटन। यह घंटाकार सममित प्रायिकता वक्र के रूप में नज़र आती है।

प्रसामान्य चर (Normal Variable): ऐसा यादृच्छिक चर जो प्रसामान्य बंटन का अनुगामी है।

प्रायिकता बंटन (Probability Distribution): यादृच्छिक चर के संभावित मानों और उनकी प्रायिकता से संबंधित प्रकथन।

मानक प्रसामान्य विचर (Standard Normal Variate): प्रसामान्य चर जिसका माध्य 0 और मानक विचलन 1 के बराबर है।

स्टूडैंट–टी बंटन (Student's–t Distribution): यह, 0 मान के गिर्द सममित बंटन है। स्टूडैंट–टी यादृच्छिक चर के लिए विचरण का परिसर $-\infty$ से $+\infty$ के बीच का है। लेकिन यह प्रसामान्य बंटन वक्र की तुलना में अधिक सपाट है। जब स्वतंत्रता की कोटि काफी अधिक हो तो यह प्रसामान्य बंटन का ही अनुगमन करता है।

स्टूडैंट–टी चर (Student's *t*-Variable): ऐसा यादृच्छिक चर जो स्टूडैंट–टी बंटन का अनुगामी है।

15.6 कुछ उपयोगी पुस्तकें

Anderson, D.R., Sweeney, D.J. and Williams, T.A., 1993. *Statistics for Business and Economics* (Fifth Edition), West Publishing Company, Minneapolis/St. Paul, Chapter 6.

Bhardwaj, R.S. 1999. *Business Statistics* (First Edition), Excel Books, New Delhi, Chapters 19 and 20.

Newbold, P. 1991. *Statistics for Business and Economics* (Third Edition), Prentice Hall, New Jersey, Chapter 5.

Walpole, R.E., Myers, R.H., Myers, S.L. and Ye, K. 2002. *Probability and Statistics for Engineers and Scientists* (Seventh Edition), Pearson Education, India, Chapter 6.

15.7 बोध प्रश्नों के उत्तर या संकेत

बोध प्रश्न 1

1) क) 0.0552

   ख) 0.0668

   ग) 0.0062

2) 67 इंच और 69 इंच की बीच की ऊँचाई वाले पुरुषों का समानुपात है 0.1586 इसलिए 67 इंच और 69 इंच के बीच की ऊचाई वाले पुरुषों की संख्या है

   $= 1000 \times 0.1586 = 159$ (लगभग)

3) $z = \dfrac{24-34}{5} = -2$

मानक प्रसामान्य सारणी से, $z = 0$ और $z = 2$ के बीच का क्षेत्रफल 0.4772 है। इसलिए, $z = 0$ और $z = -2$ के बीच का क्षेत्रफल 0.4772 है (क्योंकि मानक प्रसामान्य बंटन सममित है)। जिन बैटरियों को बदला जाना है, उनकी संख्या का पता लगाने के लिए, हमें $z = -2$ के बायें ओर का क्षेत्रफल जो $z = 2$ के दायें ओर का क्षेत्रफल है, पर विचार करना होगा। अब, $z = -2$ के दायें ओर का क्षेत्रफल $0.5 - 0.4772 = 0.0228$ है। इसलिए दोषपूर्ण बैटरी होने की प्रायिकता 0.0228 है। अतः 5000 बैटरियों में से जिन बैटरियों को बदला जाना है, उनकी संख्या होगी $= 0.0228 \times 5000 = 114$

बोध प्रश्न 2

1) 0.99

2) 0.05

3) 0.01

सारणी 15.1 : प्रसामान्य क्षेत्र सारणी

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

सारणी 15.2 : χ^2 बंटन का निर्धारण मूल्य

df\area	0.1	0.05	0.025	0.01	0.005
1	2.706	3.841	5.024	6.635	7.879
2	4.605	5.991	7.378	9.210	10.597
3	6.251	7.815	9.348	11.345	12.838
4	7.779	9.488	11.143	13.277	14.860
5	9.236	11.071	12.833	15.086	16.750
6	10.645	12.592	14.449	16.812	18.548
7	12.017	14.067	16.013	18.475	20.278
8	13.362	15.507	17.535	20.090	21.955
9	14.684	16.919	19.023	21.666	23.589
10	15.987	18.307	20.483	23.209	25.188
11	17.275	19.675	21.920	24.725	26.757
12	18.549	21.026	23.337	26.217	28.300
13	19.812	22.362	24.736	27.688	29.819
14	21.064	23.685	26.119	29.141	31.319
15	22.307	24.996	27.488	30.578	32.801
16	23.542	26.296	28.845	32.000	34.267
17	24.769	27.587	30.191	33.409	35.718
18	25.989	28.869	31.526	34.805	37.156
19	27.204	30.144	32.852	36.191	38.582
20	28.412	31.410	34.170	37.566	39.997
21	29.615	32.671	35.479	38.932	41.401
22	30.813	33.924	36.781	40.289	42.796
23	32.007	35.172	38.076	41.638	44.181
24	33.196	36.415	39.364	42.980	45.559
25	34.382	37.652	40.646	44.314	46.928
26	35.563	38.885	41.923	45.642	48.290
27	36.741	40.113	43.195	46.963	49.645
28	37.916	41.337	44.461	48.278	50.993
29	39.087	42.557	45.722	49.588	52.336
30	40.256	43.773	46.979	50.892	53.672

सारणी 15.3 : t-बंटन का निर्धारण मूल्य

Df\p	0.25	0.10	0.05	0.025	0.01	0.005
1	1.0000	3.0777	6.3138	12.7062	31.8205	63.6567
2	0.8165	1.8856	2.9200	4.3027	6.9646	9.9248
3	0.7649	1.6377	2.3534	3.1825	4.5407	5.8409
4	0.7407	1.5332	2.1318	2.7765	3.7470	4.6041
5	0.7267	1.4759	2.0150	2.5706	3.3649	4.0321
6	0.7176	1.4398	1.9432	2.4469	3.1427	3.7074
7	0.7111	1.4149	1.8946	2.3646	2.9980	3.4995
8	0.7064	1.3968	1.8595	2.3060	2.8965	3.3554
9	0.7027	1.3830	1.8331	2.2622	2.8214	3.2498
10	0.6998	1.3722	1.8125	2.2281	2.7638	3.1693
11	0.6974	1.3634	1.7959	2.2010	2.7181	3.1058
12	0.6955	1.3562	1.7823	2.1788	2.6810	3.0545
13	0.6938	1.3502	1.7709	2.1604	2.6503	3.0123
14	0.6924	1.3450	1.7613	2.1448	2.6245	2.9768
15	0.6912	1.3406	1.7531	2.1315	2.6025	2.9467
16	0.6901	1.3368	1.7459	2.1199	2.5835	2.9208
17	0.6892	1.3334	1.7396	2.1098	2.5669	2.8982
18	0.6884	1.3304	1.7341	2.1009	2.5524	2.8784
19	0.6876	1.3277	1.7291	2.0930	2.5395	2.8609
20	0.6870	1.3253	1.7247	2.0860	2.5280	2.8453
21	0.6864	1.3232	1.7207	2.0796	2.5177	2.8314
22	0.6858	1.3212	1.7171	2.0739	2.5083	2.8188
23	0.6853	1.3195	1.7139	2.0687	2.4999	2.8073
24	0.6849	1.3178	1.7109	2.0639	2.4922	2.7969
25	0.6844	1.3163	1.7081	2.0595	2.4851	2.7874
26	0.6840	1.3150	1.7056	2.0555	2.4786	2.7787
27	0.6837	1.3137	1.7033	2.0518	2.4727	2.7707
28	0.6834	1.3125	1.7011	2.0484	2.4671	2.7633
29	0.6830	1.3114	1.6991	2.0452	2.4620	2.7564
30	0.6828	1.3104	1.6973	2.0423	2.4573	2.7500
inf	0.6745	1.2816	1.6449	1.9600	2.3264	2.5758

सारणी 15.4 : *F*-बंटन का निर्धारण मूल्य (5% महत्व का स्तर)

df2/	1	2	3	4	5	6	7	8	9	10
1	161.448	199.500	215.707	224.583	230.162	233.986	236.768	238.883	240.543	241.882
2	18.513	19.000	19.164	19.247	19.296	19.330	19.353	19.371	19.385	19.396
3	10.128	9.552	9.277	9.117	9.014	8.941	8.887	8.845	8.812	8.786
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964
5	6.608	5.786	5.410	5.192	5.050	4.950	4.876	4.818	4.773	4.735
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637
8	5.318	4.459	4.066	3.838	3.688	3.581	3.501	3.438	3.388	3.347
9	5.117	4.257	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137
10	4.965	4.103	3.708	3.478	3.326	3.217	3.136	3.072	3.020	2.978
11	4.844	3.982	3.587	3.357	3.204	3.095	3.012	2.948	2.896	2.854
12	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849	2.796	2.753
13	4.667	3.806	3.411	3.179	3.025	2.915	2.832	2.767	2.714	2.671
14	4.600	3.739	3.344	3.112	2.958	2.848	2.764	2.699	2.646	2.602
15	4.543	3.682	3.287	3.056	2.901	2.791	2.707	2.641	2.588	2.544
16	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591	2.538	2.494
17	4.451	3.592	3.197	2.965	2.810	2.699	2.614	2.548	2.494	2.450
18	4.414	3.555	3.160	2.928	2.773	2.661	2.577	2.510	2.456	2.412
19	4.381	3.522	3.127	2.895	2.740	2.628	2.544	2.477	2.423	2.378
20	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447	2.393	2.348
21	4.325	3.467	3.073	2.840	2.685	2.573	2.488	2.421	2.366	2.321
22	4.301	3.443	3.049	2.817	2.661	2.549	2.464	2.397	2.342	2.297
23	4.279	3.422	3.028	2.796	2.640	2.528	2.442	2.375	2.320	2.275
24	4.260	3.403	3.009	2.776	2.621	2.508	2.423	2.355	2.300	2.255
25	4.242	3.385	2.991	2.759	2.603	2.490	2.405	2.337	2.282	2.237
26	4.225	3.369	2.975	2.743	2.587	2.474	2.388	2.321	2.266	2.220
27	4.210	3.354	2.960	2.728	2.572	2.459	2.373	2.305	2.250	2.204
28	4.196	3.340	2.947	2.714	2.558	2.445	2.359	2.291	2.236	2.190
29	4.183	3.328	2.934	2.701	2.545	2.432	2.346	2.278	2.223	2.177
30	4.171	3.316	2.922	2.690	2.534	2.421	2.334	2.266	2.211	2.165
40	4.085	3.232	2.839	2.606	2.450	2.336	2.249	2.180	2.124	2.077
60	4.001	3.150	2.758	2.525	2.368	2.254	2.167	2.097	2.040	1.993
120	3.920	3.072	2.680	2.447	2.290	2.175	2.087	2.016	1.959	1.911
inf	3.842	2.996	2.605	2.372	2.214	2.099	2.010	1.938	1.880	1.831

सारणी 15.4 : *F*-बंटन का निर्धारण मूल्य (5% महत्व का स्तर) (क्रमशः)

df2/	12	15	20	24	30	40	60	120	INF
1	243.906	245.950	248.013	249.052	250.095	251.143	252.196	253.253	254.314
2	19.413	19.429	19.446	19.454	19.462	19.471	19.479	19.487	19.496
3	8.745	8.703	8.660	8.639	8.617	8.594	8.572	8.549	8.526
4	5.912	5.858	5.803	5.774	5.746	5.717	5.688	5.658	5.628
5	4.678	4.619	4.558	4.527	4.496	4.464	4.431	4.399	4.365
6	4.000	3.938	3.874	3.842	3.808	3.774	3.740	3.705	3.669
7	3.575	3.511	3.445	3.411	3.376	3.340	3.304	3.267	3.230
8	3.284	3.218	3.150	3.115	3.079	3.043	3.005	2.967	2.928
9	3.073	3.006	2.937	2.901	2.864	2.826	2.787	2.748	2.707
10	2.913	2.845	2.774	2.737	2.700	2.661	2.621	2.580	2.538
11	2.788	2.719	2.646	2.609	2.571	2.531	2.490	2.448	2.405
12	2.687	2.617	2.544	2.506	2.466	2.426	2.384	2.341	2.296
13	2.604	2.533	2.459	2.420	2.380	2.339	2.297	2.252	2.206
14	2.534	2.463	2.388	2.349	2.308	2.266	2.223	2.178	2.131
15	2.475	2.403	2.328	2.288	2.247	2.204	2.160	2.114	2.066
16	2.425	2.352	2.276	2.235	2.194	2.151	2.106	2.059	2.010
17	2.381	2.308	2.230	2.190	2.148	2.104	2.058	2.011	1.960
18	2.342	2.269	2.191	2.150	2.107	2.063	2.017	1.968	1.917
19	2.308	2.234	2.156	2.114	2.071	2.026	1.980	1.930	1.878
20	2.278	2.203	2.124	2.083	2.039	1.994	1.946	1.896	1.843
21	2.250	2.176	2.096	2.054	2.010	1.965	1.917	1.866	1.812
22	2.226	2.151	2.071	2.028	1.984	1.938	1.889	1.838	1.783
23	2.204	2.128	2.048	2.005	1.961	1.914	1.865	1.813	1.757
24	2.183	2.108	2.027	1.984	1.939	1.892	1.842	1.790	1.733
25	2.165	2.089	2.008	1.964	1.919	1.872	1.822	1.768	1.711
26	2.148	2.072	1.990	1.946	1.901	1.853	1.803	1.749	1.691
27	2.132	2.056	1.974	1.930	1.884	1.836	1.785	1.731	1.672
28	2.118	2.041	1.959	1.915	1.869	1.820	1.769	1.714	1.654
29	2.105	2.028	1.945	1.901	1.854	1.806	1.754	1.698	1.638
30	2.092	2.015	1.932	1.887	1.841	1.792	1.740	1.684	1.622
40	2.004	1.925	1.839	1.793	1.744	1.693	1.637	1.577	1.509
60	1.917	1.836	1.748	1.700	1.649	1.594	1.534	1.467	1.389
120	1.834	1.751	1.659	1.608	1.554	1.495	1.429	1.352	1.254
inf	1.752	1.666	1.571	1.517	1.459	1.394	1.318	1.221	1.000

सारणी 15.4 : *F*-बंटन का निर्धारण मूल्य (1% महत्व का स्तर) (क्रमशः)

df2/	1	2	3	4	5	6	7	8	9	10
1	4052.181	4999.500	5403.352	5624.583	5763.650	5858.986	5928.356	5981.070	6022.473	6055.847
2	98.503	99.000	99.166	99.249	99.299	99.333	99.356	99.374	99.388	99.399
3	34.116	30.817	29.457	28.710	28.237	27.911	27.672	27.489	27.345	27.229
4	21.198	18.000	16.694	15.977	15.522	15.207	14.976	14.799	14.659	14.546
5	16.258	13.274	12.060	11.392	10.967	10.672	10.456	10.289	10.158	10.051
6	13.745	10.925	9.780	9.148	8.746	8.466	8.260	8.102	7.976	7.874
7	12.246	9.547	8.451	7.847	7.460	7.191	6.993	6.840	6.719	6.620
8	11.259	8.649	7.591	7.006	6.632	6.371	6.178	6.029	5.911	5.814
9	10.561	8.022	6.992	6.422	6.057	5.802	5.613	5.467	5.351	5.257
10	10.044	7.559	6.552	5.994	5.636	5.386	5.200	5.057	4.942	4.849
11	9.646	7.206	6.217	5.668	5.316	5.069	4.886	4.744	4.632	4.539
12	9.330	6.927	5.953	5.412	5.064	4.821	4.640	4.499	4.388	4.296
13	9.074	6.701	5.739	5.205	4.862	4.620	4.441	4.302	4.191	4.100
14	8.862	6.515	5.564	5.035	4.695	4.456	4.278	4.140	4.030	3.939
15	8.683	6.359	5.417	4.893	4.556	4.318	4.142	4.004	3.895	3.805
16	8.531	6.226	5.292	4.773	4.437	4.202	4.026	3.890	3.780	3.691
17	8.400	6.112	5.185	4.669	4.336	4.102	3.927	3.791	3.682	3.593
18	8.285	6.013	5.092	4.579	4.248	4.015	3.841	3.705	3.597	3.508
19	8.185	5.926	5.010	4.500	4.171	3.939	3.765	3.631	3.523	3.434
20	8.096	5.849	4.938	4.431	4.103	3.871	3.699	3.564	3.457	3.368
21	8.017	5.780	4.874	4.369	4.042	3.812	3.640	3.506	3.398	3.310
22	7.945	5.719	4.817	4.313	3.988	3.758	3.587	3.453	3.346	3.258
23	7.881	5.664	4.765	4.264	3.939	3.710	3.539	3.406	3.299	3.211
24	7.823	5.614	4.718	4.218	3.895	3.667	3.496	3.363	3.256	3.168
25	7.770	5.568	4.675	4.177	3.855	3.627	3.457	3.324	3.217	3.129
26	7.721	5.526	4.637	4.140	3.818	3.591	3.421	3.288	3.182	3.094
27	7.677	5.488	4.601	4.106	3.785	3.558	3.388	3.256	3.149	3.062
28	7.636	5.453	4.568	4.074	3.754	3.528	3.358	3.226	3.120	3.032
29	7.598	5.420	4.538	4.045	3.725	3.499	3.330	3.198	3.092	3.005
30	7.562	5.390	4.510	4.018	3.699	3.473	3.304	3.173	3.067	2.979
40	7.314	5.179	4.313	3.828	3.514	3.291	3.124	2.993	2.888	2.801
60	7.077	4.977	4.126	3.649	3.339	3.119	2.953	2.823	2.718	2.632
120	6.851	4.787	3.949	3.480	3.174	2.956	2.792	2.663	2.559	2.472
inf	6.635	4.605	3.782	3.319	3.017	2.802	2.639	2.511	2.407	2.321

सारणी 15.4 : *F*-बंटन का निर्धारण मूल्य (1% महत्व का स्तर) (क्रमशः)

df2/	12	15	20	24	30	40	60	120	INF
1	6106.321	6157.285	6208.730	6234.631	6260.649	6286.782	6313.030	6339.391	6365.864
2	99.416	99.433	99.449	99.458	99.466	99.474	99.482	99.491	99.499
3	27.052	26.872	26.690	26.598	26.505	26.411	26.316	26.221	26.125
4	14.374	14.198	14.020	13.929	13.838	13.745	13.652	13.558	13.463
5	9.888	9.722	9.553	9.466	9.379	9.291	9.202	9.112	9.020
6	7.718	7.559	7.396	7.313	7.229	7.143	7.057	6.969	6.880
7	6.469	6.314	6.155	6.074	5.992	5.908	5.824	5.737	5.650
8	5.667	5.515	5.359	5.279	5.198	5.116	5.032	4.946	4.859
9	5.111	4.962	4.808	4.729	4.649	4.567	4.483	4.398	4.311
10	4.706	4.558	4.405	4.327	4.247	4.165	4.082	3.996	3.909
11	4.397	4.251	4.099	4.021	3.941	3.860	3.776	3.690	3.602
12	4.155	4.010	3.858	3.780	3.701	3.619	3.535	3.449	3.361
13	3.960	3.815	3.665	3.587	3.507	3.425	3.341	3.255	3.165
14	3.800	3.656	3.505	3.427	3.348	3.266	3.181	3.094	3.004
15	3.666	3.522	3.372	3.294	3.214	3.132	3.047	2.959	2.868
16	3.553	3.409	3.259	3.181	3.101	3.018	2.933	2.845	2.753
17	3.455	3.312	3.162	3.084	3.003	2.920	2.835	2.746	2.653
18	3.371	3.227	3.077	2.999	2.919	2.835	2.749	2.660	2.566
19	3.297	3.153	3.003	2.925	2.844	2.761	2.674	2.584	2.489
20	3.231	3.088	2.938	2.859	2.778	2.695	2.608	2.517	2.421
21	3.173	3.030	2.880	2.801	2.720	2.636	2.548	2.457	2.360
22	3.121	2.978	2.827	2.749	2.667	2.583	2.495	2.403	2.305
23	3.074	2.931	2.781	2.702	2.620	2.535	2.447	2.354	2.256
24	3.032	2.889	2.738	2.659	2.577	2.492	2.403	2.310	2.211
25	2.993	2.850	2.699	2.620	2.538	2.453	2.364	2.270	2.169
26	2.958	2.815	2.664	2.585	2.503	2.417	2.327	2.233	2.131
27	2.926	2.783	2.632	2.552	2.470	2.384	2.294	2.198	2.097
28	2.896	2.753	2.602	2.522	2.440	2.354	2.263	2.167	2.064
29	2.868	2.726	2.574	2.495	2.412	2.325	2.234	2.138	2.034
30	2.843	2.700	2.549	2.469	2.386	2.299	2.208	2.111	2.006
40	2.665	2.522	2.369	2.288	2.203	2.114	2.019	1.917	1.805
60	2.496	2.352	2.198	2.115	2.028	1.936	1.836	1.726	1.601
120	2.336	2.192	2.035	1.950	1.860	1.763	1.656	1.533	1.381
inf	2.185	2.039	1.878	1.791	1.696	1.592	1.473	1.325	1.000

Source: Tables 15.1 to 15.4 are adapted from the website http://www.statsoft.com/textbook/sttable.html (accessed on 23.09.2004)

इकाई 16 प्रतिचयन की बुनियादी संकल्पनाएँ

इकाई की रूपरेखा

16.0 उद्देश्य

इस इकाई को पढ़ने के बाद, आप :

- समष्टि या 'जनसंख्या', प्रतिदर्श, प्राचल, प्रतिदर्शज, आकलक और आकल संबंधी संकल्पनाओं का वर्णन कर सकेंगे;
- जनगणना और प्रतिदर्श सर्वेक्षण के बीच के अंतर को स्पष्ट कर सकेंगे;

- प्रतिदर्श सर्वेक्षण के फायदों का वर्णन कर सकेंगे;
- प्रतिचयन त्रुटि और गैर-प्रतिचयन त्रुटि के बीच के संबंध को स्पष्ट कर सकेंगे;
- प्रतिदर्श बंटन की संकल्पना को व्यक्त कर सकेंगे; और
- मानक त्रुटि की संकल्पना को समझा सकेंगे।

16.1 प्रस्तावना

हमें राष्ट्रीय आय खातों, आगत-निर्गत तालिकाओं, उत्पादन संबंधी विभिन्न सूचकांकों, कीमत सूचकांकों और परिमाण संबंधी अन्य बहुत से सूचकांकों के निर्माण के लिए आंकड़ों की आवश्यकता होती है। यह बात स्पट है कि सुसंगत आंकड़ों के बगैर एक जटिल अर्थव्यवस्था के नीति उद्देश्यों को सूत्रबद्ध करना हमारे लिए संभव नहीं होगा। आधुनिक समाज तेजी से सूचना प्रधान समाज का रूप ले रहा है। ऐसे समाज में विभिन्न आर्थिक एवं सामाजिक प्रक्रियाओं को कुछ विशेष परिमाणात्मक विशेषताओं से दर्शाया जाता है जिसके लिए आंकड़ों के रूप में हमें विभिन्न प्रकार की जानकारी हासिल करनी पड़ती है।

आँकड़ों को एकत्रित करने का कार्य दिन-प्रतिदिन जटिल एवं कठिन होता जा रहा है। अपेक्षित सूचना के लिए कुल मिलाकर कितनी इकाइयों से परामर्श की आवश्यकता है और किनकी छानबीन की जानी है, यह एक भारी काम हो सकता है और हमारे मौजूदा संसाधन अर्थात् धन, समय या व्यक्तिगत प्रयास शायद इस संदर्भ में सीमित होंगे। इसके अलावा बड़े पैमाने पर व्यापक छानबीन से दोषरहित सूचना प्राप्त करना शायद और अधिक चुनौती भरा काम है। इसी कारण हम प्रायः छोटे समूह से सूचना प्राप्त करने का प्रयास करते हैं क्योंकि यह आसानी से मिल जाती है और इसका रखरखाव करना भी आसान होता है। लेकिन, यहाँ सुनिश्चित करना महत्वपूर्ण है कि यह छोटा समूह सुसंगत इकाइयों के समूचे संग्रह का सही प्रतिनिधि हो। प्रतिचयन की विषयवस्तु ऐसे प्रतिनिधि समूह की प्राप्ति के लिए उपयुक्त गणितीय सिद्धांतों से अवगत कराती है।

16.2 जनगणना और प्रतिदर्श सर्वेक्षण

इस अनुभाग में हम आंकड़े एकत्रित करने की जनगणना और प्रतिचयन विधियों के बीच के अंतर को स्पट करेंगे। यहाँ हम जनगणना सर्वेक्षण और प्रतिदर्श सर्वेक्षण की 'व्यापकता' और इसके अर्थ को समझाने का प्रयास करेंगे।

16.2.1 समष्टि और जनगणना

किसी विशेष पूछताछ के लिए हमारे पास सुसंगत इकाइयों का संग्रह है। इस संदर्भ में इकाई एक पहचान है जिसे ध्यान में रखकर और निर्धारित क्रियाविधि का अनुकरण करते हुए हम प्रेक्षण कर सकते हैं। ऐसी इकाइयों के समूचे संग्रह को समष्टि (population) कहते हैं। ऐसी समष्टि मनुष्यों, पशुओं, वृक्षों, कीमतों, उत्पादनों आदि किसी की भी हो सकती है। इस संदर्भ में हमें देखना होगा कि समष्टि परिमित होगी या अपरिमित। यदि इकाइयों की संख्या परिमित है तो यह परिमित समष्टि है और यदि इकाइयों की संख्या अपरिमित है तो यह अपरिमित समष्टि का उदाहरण है। आमतौर पर हम व्यावहारिकता में परिमित समष्टि से ही संबंध रखते हैं।

यदि कोई पूछताछ, समष्टि की सभी इकाइयों से प्राप्त सूचना पर आधारित होती है तो इसे *पूर्ण गणना विधि* (complete enumeration method) या *जनगणना विधि* (census method) कहते हैं।

16.2.2 प्रतिदर्श और प्रतिदर्श सर्वेक्षण

जब हमारे पास समष्टि के किसी भाग/अनुभाग का संग्रह हो तो इसे *प्रतिदर्श* (sample) कहते हैं। जैसा कि हमने पहले देखा था, जनगणना, समष्टि के प्रत्येक सदस्य से प्राप्त सूचना पर आधारित होती है। लेकिन समष्टि के कुछ विशेष लक्षणों के बारे में सूचना प्राप्त करने के लिए, सदैव जनगणना का सहारा लेने की जरूरत नहीं पड़ती। व्यावहारिक तौर पर, समष्टि से लिए गए उपयुक्त प्रतिदर्श के अध्ययन से हम संतोषजनक निष्कर्ष पा सकते हैं। प्रतिदर्श प्राप्ति की प्रक्रिया *प्रतिदर्श सर्वेक्षण* (Sample Survey) कहलाती है। जनगणना विधि में हम समग्र समष्टि की जाँच/परीक्षण करते हैं तो प्रतिदर्श सर्वेक्षण में हम समष्टि के एक प्रतिनिधि भाग पर ही विचार करते हैं और इसी प्रतिदर्श आधारित जानकारी के प्रयोग से, समग्र समष्टि के बारे में निष्कर्ष निकाल लेते हैं।

16.3 कुछ संकल्पनाएँ

प्रतिचयन सिद्धांत में आमतौर पर प्रयुक्त कुछ संकल्पनाएँ इस प्रकार हैं :

16.3.1 प्राचल

आंकड़ों की छानबीन करने में, हमारा ध्यान मुख्य रूप से समष्टि की एक या अधिक विशेाताओं पर केंद्रित रहता है। ऐसी विशेषता के माप को *प्राचल* (parameter) कहते हैं। उदाहरण के रूप में, हम किसी विशेष वर्ष के लिए कुछ क्षेत्रों के व्यक्तियों की माध्य आय जानना चाहते हैं। हम इन व्यक्तियों की आमदनी का मानक विचलन भी जानना चाहेंगे। यहाँ, माध्य और मानक विचलन अर्थात् दोनों प्राचल हैं।

प्राचलों को हम सुविधा के लिए ग्रीक अक्षरों में दर्शाते हैं। जैसे समष्टि माध्य को μ और समष्टि मानक विचलन को σ द्वारा दर्शाया जाता है।

यहाँ सभी समष्टि प्रेक्षणों से प्राचल का मान परिकलित करना अत्यंत महत्वपूर्ण है। अतः प्राचल 'माध्य आय' ऐसे सभी विभिन्न व्यक्तियों की आमदनी संबंधी आंकड़ों से परिकलित की जा सकती है जो समष्टि का गठन करते हैं। इसी तरह, 'ऊँचाई और भार संबंधी सहसंबंध गुणांक' प्राचल की गणना के लिए हमें समष्टि के ऊँचाई एवं भार संबंधी सभी युग्मों के मानों की प्राप्ति की आवश्यकता है। अतः, हम प्राचल को समष्टि मानों के फलन के रूप में परिभाषित कर सकते हैं । यदि θ प्राचल है जिसे हमें $X_1, X_2, \cdots X_N$, समष्टि मानों से प्राप्त करना है, तब

$$\theta = (X_1, X_2, \ldots X_N)$$

16.3.2 प्रतिदर्शज

जनगणना और प्रतिदर्श सर्वेक्षण की चर्चा करते समय, हमने देखा कि विभिन्न मौजूद अवरोधों के कारण, कभी कभार समष्टि के बारे में सूचना एकत्रित करना कठिन होता है। अन्य शब्दों में, समष्टि प्राचल अभिकलित करना सदैव संभव नहीं होता। ऐसी स्थितियों में, प्रतिदर्श से प्राप्त सूचना से हम प्राचल का अनुमान बना सकते हैं। प्रतिदर्श संबंधी यह सूचना, प्रतिदर्शज (statistic) के रूप में संक्षिप्त कर दी जाती है। उदाहरण स्वरूप, प्रतिदर्श माध्य या प्रतिदर्श माध्यिका या प्रतिदर्श बहुलक को प्रतिदर्शज कहते हैं। अतः प्रतिदर्शज की गणना, इकाइयों के ऐसे मानों से की जाती है जो प्रतिदर्श में शामिल किए जाते हैं। इसलिए प्रतिदर्शज को प्रतिदर्श मानों के फलन के रूप में परिभाषित किया जा सकता है। आसानी

से समझने के लिए, प्रतिदर्शज को रोमन वर्ण माला के अक्षरों द्वारा दर्शाया जाता है। प्रतिदर्श माध्य को $\bar{x}$ और प्रतिदर्श मानक विचलन को s से दर्शाया जा सकता है। यदि T ऐसा प्रतिदर्शज है जिसे हमें $x_1, x_2, \cdots x_n$ प्रतिदर्श मानों से प्राप्त करना है, तब $T = f(x_1, x_2, \ldots x_n)$

16.3.3 आकलक एवं आकल

प्रतिदर्शज का मूल उद्देश्य कुछ समष्टि प्राचलों का अनुमान लगाना है। इस संदर्भ में अनुकृत विधि या प्रतिदर्शज परिकलन में प्रयुक्त सूत्र को आकलक (estimator) कहते हैं और परिकलित किया गया प्रतिदर्शज का मान आकल (estimate) कहलाता है।

प्रतिदर्शज परिकलन के लिए हम $\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$ सूत्र का प्रयोग करते हैं। यह सूत्र आकलक है। इसके बाद, यदि हम इस सूत्र का प्रयोग करते हैं और $\bar{x} = 10$, प्राप्त करते हैं तब '10' आकल होगा।

16.4 गैर-प्रतिचयन और प्रतिचयन त्रुटियाँ

जैसा कि हमने ऊपर बताया, प्रतिचयन का मूल उद्देश्य प्रतिदर्श के आधार पर समष्टि के बारे में अनुमिति निकालना है। जैसे, मान लीजिए हमें किसी गाँव की प्रति व्यक्ति आय ज्ञात करनी है। समय, धन और अपेक्षित कार्मिकों के अभाव में हम पूर्ण जनगणना की बजाए, प्रतिदर्श सर्वेक्षण पर ध्यान केंद्रित करते हैं। इस मामले में स्पष्ट है कि प्रतिदर्श से प्राप्त प्रति व्यक्ति आय, गाँव की वास्तविक प्रति व्यक्ति आय के बराबर नहीं होगी। ऐसी त्रुटि होने के दो कारण हो सकते हैं।

हम समष्टि के केवल एक भाग से ही आँकड़े एकत्रित कर रहे हैं (अर्थात् हमारे द्वारा चयनित प्रतिदर्श के आधार पर), प्रतिदर्श माध्य (इस मामले में प्रति व्यक्ति आय) समष्टि माध्य के बराबर नहीं है। यदि संभवतया दोनों बराबर ही हों तो यह अनोखी घटना ही होगी। इसलिए यदि हम प्रतिदर्श माध्य को समष्टि माध्य के रूप में ले रहे हैं, कुछ त्रुटि रह जाती है। इसे प्रतिचयन त्रुटि (sampling error) कहते हैं।

त्रुटि होने का एक अन्य कारण है, आँकड़ों की गलत जानकारी देना या उन्हें रजिस्टर में गलत तरीके से भरना या उनकी गलत तालिका बनाना या आंकड़ों को गलत तरीके से संसाधित करना। इस प्रकार की त्रुटि को गैर-प्रतिचयन त्रुटि (non-sampling error) कहते हैं। ध्यान रखें, गलत-प्रतिचयन त्रुटि, जैसा कि इसके नाम से इंगित है, हमारी प्रतिचयन प्रक्रिया से इसका कोई सरोकार नहीं है। प्रतिदर्श सर्वेक्षण में आंकड़ों की गलत जानकारी प्रस्तुति या इनका गलत संसाधन होने की संभावना बनी रहती है।

इन त्रुटियों के स्रोतों का वर्णन निम्न प्रकार है।

16.4.1 गैर-प्रतिचयन त्रुटि

गैर-प्रतिचयन त्रुटि के विविध स्रोत इस प्रकार हैं :

1) **माप आधार त्रुटि**

यह जाना माना तथ्य है कि किसी भी परिमाण का सही/सटीक माप (measurement) लेना संभव नहीं है। यदि कुछ व्यक्तियों को उदाहरण के रूप में बारी-बारी से कपड़े के किसी एक

टुकड़े की लंबाई मापने को कहा जाए, (मान लीजिए दो दशमलव बिंदुओं तक), तो हम सुनिश्चित तौर पर कह सकते हैं कि सबके उत्तर एक जैसे नहीं होंगे। वास्वत में मापने वाले फीते की परिशुद्धता की कोटि भी एक जैसी नहीं होगी।

किसी छानबीन के मामले में उत्तरदाताओं के प्रतिचयन के संदर्भ में, उदाहरण के रूप में, उनकी आमदनी के बारे में सही आँकड़े प्राप्त नहीं किए जा सकते। यह समस्या ऐसे व्यक्तियों की स्थिति में तो नहीं आएगी जो मजदूरी या वेतन के रूप में निश्चित आय अर्जित करते हैं। लेकिन, स्व-रोजगारी व्यक्ति शायद सही जानकारी देने की स्थिति में नहीं होंगे।

2) गैर-प्रतिक्रिया (non-response) आधारित त्रुटि

कभी कभार उत्तरदाताओं को प्रश्नावली भेजकर अपेक्षित आंकड़ों को एकत्रित किया जाता है। ऐसे व्यक्तियों में से बहुत से अधूरे उत्तर वाली प्रश्नावली वापिस भेजते हैं या प्रश्नावली भेजते ही नहीं। इसका कारण हो सकता है:

क) उत्तरदाता पूछे गए प्रश्नों का उत्तर सावधानी से पढ़ने के बाद नहीं देते।

ख) वे प्रश्नों को समझने की स्थिति में नहीं होते, या

ग) वे अपेक्षित सूचना को बताना नहीं चाहते।

हम ध्यान से देखें तो पाएँगे कि प्रतिक्रिया व्यक्त न करने का कारण, प्रश्नावली का रास्ते में ही खो जाना, भी हो सकता है।

यदि वैयक्तिक साक्षात्कारों के माध्यम से आंकड़े एकत्रित किए जाते हैं तो उपर्युक्त कुछ कारणों में अवश्य कमी आएगी। लेकिन, ऐसे मामले में, कुछेक व्यक्तियों के कारण ऐसी त्रुटि भी उत्पन्न हो सकती है:

क) वे सूचना देना नहीं चाहते, या

ख) बार-बार जाने के बावजूद भी, वे मिलते ही नहीं।

3) आँकड़ों को दर्ज करने में त्रुटि

ऐसी त्रुटि ऐसे चरण पर नज़र आती है जब जाँचकर्ता (investigator) उत्तरों का लिखित रूप दर्ज करता है। ऐसी त्रुटि का मुख्य कारण पूछताछकर्ता का लापरवाही बरतना है।

4) जाँचकर्ता के पक्षपातपूर्ण रवैये पर आधारित त्रुटि

प्रत्येक व्यक्ति निजी पूर्वाग्रहों और पक्षपाती रवैयों से ग्रस्त रहता है। जाँचकर्ताओं को सर्वाधिक संभावित प्रशिक्षण देने के बावजूद भी, उनकी निजी सोच बीच में बाधक बन जाती है जब वे उत्तरदाताओं के प्रश्नों को अपनी समझ के आधार पर समझना शुरू कर देते हैं और इन्हें लिखित स्वरूप भी अपने विचारों के अनुरूप देने लगते हैं।

पूर्ण गणना विधि में गैर-प्रतिचयन त्रुटि होने की संभावना काफी अधिक होती है, क्योंकि आंकड़ा संग्रह प्रक्रिया में बहुत से व्यक्ति शामिल रहते हैं। लेकिन निम्न बातों को अपनाकर हम इस त्रुटि को न्यूनतम कर सकते हैं। ये हैं:

i) सर्वेक्षण की योजना सोच-समझकर बनाना,

ii) जाँचकर्ताओं को उचित प्रशिक्षण देना,

iii) प्रश्नावली को सरल रूप देना।

लेकिन, यहाँ हम इस बात पर जोर देना चाहेंगे कि पूर्ण गणना में काफी अधिक गैर-प्रतिचयन त्रुटियों के होने की संभावना बनी रहती है।

16.4.2 प्रतिचयन त्रुटि

अब तक आप समझ चुके होंगे कि प्रतिचयन विधि में भी गैर-प्रतिचयन त्रुटियाँ उत्पन्न हो सकती हैं। आंकड़ों को पूर्णतया ऐसी त्रुटियों के बिना एकत्रित करना लगभग असंभव होता है। लेकिन, यदि प्रतिदर्श सर्वेक्षण में उत्तरदाताओं की संख्या जनगणना विधि की तुलना में काफी कम हो तो प्रतिचयन विधि में सामान्य तौर पर गैर-प्रतिचयन त्रुटि काफी कम नजर आती है। गैर-प्रतिचयन त्रुटियों के अलावा, प्रतिदर्श सर्वेक्षण में प्रतिचयन त्रुटि भी उत्पन्न होती है। प्रतिचयन त्रुटि, प्राचल और संगत प्रतिदर्शज के बीच का पूर्ण अंतर अर्थात $|T-\theta|$ है।

प्रतिचयन त्रुटि, उत्तरदाता, जाँचकर्ता या कुछ अन्य कारण की वजह से उत्पन्न नहीं होती। ये तो प्रतिचयन क्रियाविधि की प्रकृति से ही उत्पन्न होती है। इनका पूर्ण निवारण संभव नहीं है। लेकिन हमारे पास कुछ ऐसे प्रतिचयन के सुविकसित सिद्धांत मौजूद हैं जिनकी सहायता से ऐसी त्रुटियों को न्यूनतम किया जा सकता है।

16.5 प्रतिदर्श सर्वेक्षण के उपादेयता

पूर्ण गणना या जनगणना विधि की तुलना में, प्रतिदर्श सर्वेक्षण के कुछ महत्वपूर्ण फायदे हैं। ये हैं:

i) **व्यवहार्यता**

कभी-कभी विस्तृत समष्टि के आंकड़ा संग्रह में काम अधिक होने के कारण जनगणना विधि को व्यवहारिक रूप नहीं दिया जा सकता। ऐसी स्थिति में प्रतिदर्श सर्वेक्षण अधिक उपयोगी सिद्ध हो सकता है।

ii) **गति**

जनगणना विधि की तुलना में प्रतिदर्श सर्वेक्षण के माध्यम से आंकड़ों को एकत्रित और इनके संक्षेपण का कार्य तेजी से किया जा सकता है। यह एक महत्वपूर्ण फायदा है। विशेषरूप से जब सूचना की तत्काल आवश्यकता हो।

iii) **परिशुद्धता**

किसी भी सर्वेक्षण, जनगणना या प्रतिदर्श, में प्रश्नावली भरकर अपेक्षित सूचना प्राप्त की जाती है। देखा गया है कि उत्तरदाताओं की तुलना में जब जाँचकर्ता स्वयं प्रश्न पूछकर और उनसे न भरवाकर प्रश्नावली भरने का काम भी स्वयं करते हैं तो अधिक परिशुद्ध परिणाम प्राप्त होते हैं। इसके अलावा, डाक से प्रश्नावली भेजकर और भरने का निवेदन करने की बजाए प्रत्यक्ष साक्षात्कारों से प्राप्त निष्कर्ष अधिक परिशुद्ध सूचना देते हैं। आमतौर पर किसी छानबीन में शामिल जाँचकर्ताओं की संख्या उस काम में शामिल उत्तरदाताओं की संख्या से सीधे तौर पर भिन्न होती है, जिसके परिणामस्वरूप जनगणना की तुलना में प्रतिदर्श सर्वेक्षण अपनाने में निजी साक्षात्कार करना अपेक्षाकृत अधिक सरल होता है। असल में, प्रतिदर्श सर्वेक्षण में अधिक सक्षम और बेहतर ढंग से प्रशिक्षित जाँचकर्ताओं को शामिल करना सहज होता है और ऐसे सर्वेक्षण में जाँचकर्ता प्रत्येक उत्तरदाता को अधिक समय देने की स्थिति में होते हैं। जनगणना की तुलना में प्रतिदर्श सर्वेक्षण भले ही अधिक लोगों तक

पहुँच नहीं पाता हो, लेकिन, फिर भी इसके परिणामों की परिशुद्धता अपेक्षाकृत अधिक होती है।

iv) **कुल खर्च**

निस्संदेह पूर्ण गणना की तुलना में प्रतिदर्श सर्वेक्षण से परिणामों की प्राप्ति कम खर्च पर ही कर ली जाती है क्योंकि ऐसे सर्वेक्षण में समष्टि का केवल सीमित भाग ही शामिल किया जाता है। छानबीन के खर्च संबंधी घटक हैं:

क) सर्वेक्षण संचालित करने वाले संगठन के कुछ बंधे हुए खर्च

ख) आंकड़े एकत्रित करने से संबंधित खर्च

ग) आँकड़ों के संसाधन और तालिका से संबंधित खर्च

घ) सर्वेक्षण के परिणामों के प्रकाशन का खर्च

इस खर्च-वितरण में मद (ख) और (ग) संबंधी खर्च को हम अस्थिर लागत (variable cost) कहते हैं। जबकि (क) और (घ) स्थित लागत मद हैं। जिसके परिणामस्वरूप मद (ख) और (ग) संबंधी खर्च निश्चित रूप से जनगणना की तुलना में प्रतिदर्श सर्वेक्षण में काफी कम होते हैं। ध्यान दीजिए कि उचित प्रतिदर्श सर्वेक्षण की रूपरेखा विकसित करने और उपयुक्त प्रतिदर्श के चयन में अच्छा खासा खर्च हो सकता है। लेकिन, आमतौर पर देखा गया है कि पूर्ण गणना की तुलना में प्रतिदर्श सर्वेक्षण करना सस्ता पड़ता है।

बोध प्रश्न 1

1) निम्नलिखित संकल्पनाओं को परिभाषित कीजिए:

   क) समष्टि

   ख) प्रतिदर्श

   ग) प्राचल

   घ) प्रतिदर्शज

   ....

   ....

   ....

   ....

2) निम्नलिखित के बीच के अंतर को स्पष्ट कीजिए:

   क) आकलक और आकल

   ख) जनगणना और प्रतिदर्श सर्वेक्षण

   ग) प्रतिचयन त्रुटि और गैर-प्रतिचयन त्रुटि

   ....

   ....

   ....

   ....

3) जनगणना की तुलना में प्रतिचयन के फायदे क्या हैं?

....

....

....

....

16.6 प्रतिचयन के प्रकार

समष्टि से प्रतिदर्श चुनने की विधि प्रतिचयन कहलाती है। मुख्यतया प्रतिचयन के दो प्रकार हैं। ये हैं: प्रायिकता प्रतिचयन (probability sampling) और गैर-प्रायिकता प्रतिचयन (non-probability sampling)। प्रायिकता प्रतिचयन में प्रतिचयन इकाइयों का चयन कुछ संयोगी क्रियाविधि (chance mechanism) या चयन की प्रायिकता के आधार पर किया जाता है। दूसरी तरफ, गैर-प्रायिकता प्रतिचयन, चयनकर्ता के विवेक या उसकी इच्छा पर आधारित है। अतः गैर-प्रायिकता प्रतिचयन में सुविधा होने के कारण कुछ निश्चित इकाइयों का चयन हो जाता है या ऐसा इसलिए भी होता है क्योंकि वे उद्देश्य के लिए कारगर सिद्ध होते हैं या शोधकर्ता मानता है कि ये इकाइयाँ समष्टि का प्रतिनिधित्व कर सकती हैं। यहाँ, यादृच्छिक क्रियाविधि के आधार पर इकाइयों का चयन नहीं होता।

16.6.1 प्रायिकता प्रतिचयन

इसे यादृच्छिक प्रतिचयन (random sampling) भी कहते हैं। यह एक ऐसी क्रियाविधि है जिसमें समष्टि के प्रत्येक सदस्य की प्रतिदर्श में चुने जाने की संभावना बनी रहती है। अतः इस संभावना आधारित नजरिए के कारण प्रतिदर्श यादृच्छिक होता है। ''यादृच्छिक'' शब्द का अर्थ यह नहीं है कि बिना किसी नियम का अनुकरण किए संयोग से प्रतिदर्श की प्राप्ति की गई है।

यादृच्छिक प्रतिचयन, प्रायिकता सिद्धांत के सुनिर्मित सिद्धांतों पर आधारित है। इसके कुछ रूपभेद भी हैं। ये हैं: सरल यादृच्छिक प्रतिचयन, क्रमबद्ध यादृच्छिक प्रतिचयन और स्तरित यादृच्छिक प्रतिचयन। इन विविध प्रकारों की चर्चा हम आगे कर रहे हैं :

क) **सरल यादृच्छिक प्रति चयन**

यदि समष्टि के सदस्यों की विशेषताओं में काफी अधिक भिन्नता न हो तो हम सरल यादृच्छिक प्रतिचयन की विधि का अनुसरण कर सकते हैं। इस विधि में हम समूची समष्टि को सजातीय समूह मानकर चलते हैं और प्रतिदर्श के लिए सदस्यों के चयन के लिए यादृच्छिक प्रतिचयन के सिद्धांत का अनुसरण करते हैं।

सरल यादृच्छिक प्रतिचयन के दो रूपभेद हैं : (i) सरल यादृच्छिक प्रतिचयन - प्रतिस्थापन के साथ (SRSWR) और (ii) सरल यादृच्छिक प्रतिचयन - बिना प्रतिस्थापन के (SRSWOR)। इनमें अंतर प्रतिदर्श इकाइयों के चयन की विधि के कारण आता है। प्रतिस्थापन वाली सरल यादृच्छिक प्रतिचयन की क्रियाविधि के अनुसार हम समष्टि से एक इकाई का चयन करते हैं, उसके लक्षणों को नोट करने के बाद, पुनः उसे पूर्ण समष्टि में वापिस डाल देते हैं जिससे इस इकाई के दुबारा भी चुने जाने की संभावना बनी रहती है। इस तरीके से, समष्टि में इकाइयों की कुल संख्या ज्यों की त्यों ही बनी रहती है। अन्य शब्दों में समष्टि की रचना (composition) पहले जैसी ही अर्थात् अपरिवर्तित ही रहती है और समष्टि के

प्रत्येक सदस्य की प्रतिदर्श में चुने जाने की एक जैसी संभावना होती है। यदि समष्टि का आकार N है तो यह प्रायिकता है $\frac{1}{N}$। दूसरी तरफ ना प्रतिस्थापन वाले सरल यादृच्छिक प्रतिचयन के मामले में, इकाई के एक बार चुने जाने के बाद, इसे समष्टि में दुबारा शामिल नहीं किया जाता। अतः दुबारा चयन के लिए यह अयोग्य हो जाती है जिसके परिणामस्वरूप उत्तरोत्तर इकाइयों के चयन में, समष्टि की रचना बदल जाती है। इसलिए, समष्टि से परवर्ती इकाई चयन के लिए, किसी विशिष्ट इकाई के चुने जाने की प्रायिकता भी बदल जाती है। आइए, इसे समझने का प्रयास करें। मान लीजिए, समष्टि का आकार N है और हम SRSWOR के सिद्धांत को लागू करके n आकार के प्रतिदर्श की प्राप्ति करना चाहते हैं। इस विधि में पहली इकाई के चुने जाने से पहले, समष्टि की प्रत्येक इकाई का, प्रतिदर्श में चुने जाने की संभावना एक जैसी (अर्थात $\frac{1}{N}$) होती है। जब प्रतिदर्श के पहले सदस्य का चयन हो जाता है तो समष्टि के बाकी के प्रत्येक $N-1$ सदस्यों की भी प्रतिदर्श में चयन की समान संभावना (अर्थात $\frac{1}{N-1}$) होती है। इसी तरह प्रतिदर्श के nवें सदस्य के चयन से पहले, समष्टि के बाकी के प्रत्येक सदस्य की, प्रतिदर्श में शामिल किए जाने की प्रायिकता $\frac{1}{N-(n+1)}=\frac{1}{N-n-1}$ होती है।

ध्यान दीजिए कि N आकार की समटि से, n आकार के प्रतिदर्शों की संख्या जिनका चयन प्रतिस्थापन के साथ किया जा सकता है N^n है और बिना प्रतिस्थापन के प्रतिदर्शों को निकालने की संख्या ${}^N c_n$ है।

उदाहरण 16.1

मान लीजिए किसी समष्टि में निम्नलिखित 5 इकाइयाँ (4,5,7,9,10) शामिल हैं। 2 आकार के कितने प्रतिदर्शों की प्राप्ति हम इससे कर सकते हैं?

i) यदि हम SRSWR की क्रियाविधि का अनुकरण करते हैं तो प्रतिदर्शों की संख्या जिनका चयन किया जा सकता है, हैं:

$$= N^n = 5^2 = 25$$

संभावित प्रतिदर्श इस प्रकार होंगे:

(4, 4), (4, 5), (4, 7), (4, 9), (4, 10), (5, 4), (5, 5), (5, 7), (5, 9), (5, 10),

(7, 4), (7, 5), (7, 7), (7, 9), (7, 10), (9, 4), (9, 5), (9, 7), (9, 9), (9, 10),

(10, 4), (10, 5), (10, 7), (10, 9), (10, 10).

ध्यान दीजिए कि प्रतिस्थापन वाली प्रतिचयन विधि में, इकाइयों का चयन जिस क्रम में होता है, उसका भी अपना विशेष महत्व है। अतः (4,10) और (10,4) को हम दो अलग-अलग प्रतिदर्शों के रूप में देखेंगे।

ii) यदि हम SRSWOR की क्रियाविधि का अनुसरण करें तो प्रतिदर्शों की संख्या जिनका चयन किया जा सकता है, होंगी:

$$= {}^N c_n = {}^5 c_2 = \frac{5!}{2!(5-2)!} = \frac{5!}{2!\ 3!} = \frac{5\times 4}{2\times 1} = \frac{20}{2} = 10$$

संभावित प्रतिदर्श इस प्रकार होंगे:

(4, 5), (5, 7), (7, 9), (9, 10), (4, 7), (4, 9), (4, 10), (5, 9), (5, 10), (7, 10).

ध्यान दीजिए कि बिना प्रतिस्थापन वाले प्रतिचयन में, यदि किसी सदस्य का एक बार चयन हो जाए तो दुबारा उसका चयन नहीं किया जा सकता। इस तरह (4,4), (5,5) आदि की तरह प्रतिदर्श का चयन नहीं किया जा सकता है। यदि (4,5) जैसे प्रतिदर्श का चयन होता है तो (5,4) जैसे प्रतिदर्श का चयन नहीं किया जा सकता।

ख) क्रमबद्ध यादृच्छिक प्रतिचयन (Systematic Random Sampling)

यादृच्छिक प्रतिचयन के इस रूपभेद में, प्रतिदर्श की केवल पहली इकाई का चयन समष्टि से यादृच्छिक रूप से किया जाता है। इसके बाद की इकाइयों का चयन कुछ निश्चित नियमों का अनुसरण करके किया जाता है। जैसे, मान लीजिए, हमें कृषि के लिए भूखंडों (plots) के प्रतिदर्श का चयन करना है। क्रमबद्ध यादृच्छिक प्रतिचयन में हम सर्वप्रथम यादृच्छिक रूप से केवल एक भूखण्ड का चयन करेंगे और इसके बाद प्रत्येक 10वें भूखण्ड का चयन हम कर सकते हैं।

ग) स्तरित यादृच्छिक प्रतिचयन (Stratified Random Sampling)

यदि विचाराधीन समष्टि में विजातीय इकाइयों का समावेश हो तो स्तरित यादृच्छिक प्रतिचयन उपयुक्त विधि होगी। यहाँ, हम समष्टि को कुछ निश्चित विजातीय समूहों या स्तरानुसार समूहों में विभाजित करेंगे। दूसरा, कुछ इकाइयों का चयन, सरल यादृच्छिक प्रतिचयन द्वारा किया जाता है। तीसरे, प्रत्येक स्तर से इकाइयों के चयन के बाद, इन सभी के मिले-जुले रूप से अंतिम प्रतिदर्श की प्राप्ति की जाती है।

आइए, अब एक उदाहरण पर विचार करें। मान लीजिए हम प्रतिदर्श सर्वेक्षण द्वारा दिल्ली की प्रति व्यक्ति आय का अनुमान लगाना चाहते हैं। जैसा कि हम सभी जानते हैं कि दिल्ली धनी, मध्यम वर्ग और निम्न वर्ग जैसे इलाकों में बंटी हुई है और यह विभाजन इन इलाकों के वासियों की आमदनी के आधार पर है। अब इन इलाकों में से प्रत्येक एक स्तर को गठित कर सकता है जिससे सरल यादृच्छिक प्रतिचयन क्रियाविधि को लागू करके हम कुछ लोगों का चयन कर सकते हैं।

घ) बहु-चरणीय यादृच्छिक प्रतिचयन (Multi-Stage Random Sampling)

आइए अब किसी ऐसी स्थिति पर विचार करें जहाँ हम किसी बड़े शहर, जैसे दिल्ली, में स्थित घरों के प्रतिदर्श से जानकारी प्राप्त करना चाहते हैं। कभी-कभार, सभी घरों की सूची आसानी से प्राप्त नहीं हो सकती जिसकी वजह से घरों के प्रतिदर्श की सीधे तौर पर प्राप्ति करना शायद संभव नहीं होगा। ऐसी स्थिति में हमें विभिन्न चरणों में प्रतिदर्श लेने होंगे। आमतौर पर प्रशासनिक प्रयोजनों की वजह से शहर को कुछ निश्चित भौगोलिक क्षेत्रों में बाँट दिया जाता है। शहरों में ऐसे क्षेत्रों को खंड (ब्लॉक) कहते हैं। इसलिए पहले चरण पर ऐसे कुछ खंडों का चयन यादृच्छिक प्रतिचयन द्वारा किया जाता है। इससे अगले चरण में, पहले चरण के चुनिंदा खंडों में से प्रत्येक से कुछ घरों का चयन दुबारा यादृच्छिक प्रतिचयन के सिद्धांत पर किया जाता है। इस तरीके से बड़े शहर के परिवारों के प्रतिदर्श की अंततः प्राप्ति कर ली जाती है। उपर्युक्त उदाहरण द्वि-चरणी यादृच्छिक प्रतिचयन का मामला है। लेकिन, यदि छानबीन में अपेक्षित हो तो प्रतिचयन की विधि का विस्तार दो चरणों से अधिक भी किया जा सकता है।

16.6.2 गैर-प्रायिकता प्रतिचयन (Non-Probability Sampling)

हमने यादृच्छिक प्रतिचयन की विधि और इसके कुछ रूप भेदों पर विचार किया है। अब हमें यह स्पट हो जाना चाहिए कि यादृच्छिक प्रतिचयन के सिद्धांत का बुनियादी उद्देश्य, समष्टि से प्रतिदर्श के चयन में छानबीनकर्ता के व्यक्तिपरक पक्षपाती रवैये को दूर करना है या अधिकाधिक कम करना है। लेकिन कुछ निश्चित प्रयोजनों के लिए अपने विवेक का प्रयोग करना भी ज़रूरी होता है। जैसे, मान लीजिए किसी अध्यापक को एक प्रतियोगिता के लिए 30 विद्यार्थियों की कक्षा से 4 विद्यार्थियों का चयन करना है। यहाँ अध्यापक कक्षा के ऐसे सर्वोच्च विद्यार्थियों का अपने निजी विवेक के आधार पर चयन कर सकता है। यह प्रयोजनमूलक प्रतिचयन (Purposive Sampling) का उदाहरण है। इस विधि में प्रतिदर्श का प्रयोजन, समष्टि की इकाइयों या कुछ निश्चित सदस्यों के चयन में स्वयं ही मार्गदर्शन करता है।

16.6.3 मिश्रित प्रतिचयन (Mixed Sampling)

मिश्रित प्रतिचयन में, गैर-प्रायिकता प्रतिचयन और यादृच्छिक प्रतिचयन, दोनों के कुछ लक्षण हमें नजर आते हैं। मान लीजिए, किसी संस्थान को गर्मियों की छुट्टियों के दौरान किसी कंपनी में प्रबंधकीय प्रशिक्षण के लिए 5 विद्यार्थियों को भेजना है। सर्वप्रथम, संस्थान को अपने निजी विवेक के प्रयोग से प्रशिक्षण के लिए लगभग ऐसे 20 प्रशिक्षणार्थियों की लघु सूची बनानी होगी जिन्हें वह सर्वाधिक उपयुक्त मानता है। अब इन 20 विद्यार्थियों में से यादृच्छिक प्रतिचयन द्वारा 5 का अंतिम रूप से चयन किया जा सकता है।

16.7 प्रतिदर्शी बंटन

अभी तक आपको स्पष्ट हो चुका होगा कि आमतौर पर मूल समष्टि की तुलना में प्रतिदर्श का आकार अपेक्षाकृत काफी छोटा होता है। जिसके परिणामस्वरूप समान समष्टि से ऐसे बहुत से प्रतिदर्शों का चयन किया जा सकता है जो एक-दूसरे से भिन्न होते हैं। चूंकि प्राचल का आकल, प्रतिदर्श मानों पर निर्भर करता है और ये मान, एक प्रतिदर्श से दूसरे प्रतिदर्श में बदलते रहते हैं। इसलिए समान प्राचल के लिए प्रतिदर्शज के आकल या मान भी अलग-अलग हो सकते हैं। मानों में ऐसे बदलाव को *प्रतिचयन उच्चावचन* कहते हैं। मान लीजिए, हम N आकार की समष्टि से बहुत से प्रतिदर्शों, जिनमें से प्रत्येक n आकार का है, प्राप्ति करते हैं और प्रत्येक प्रतिदर्श के लिए, प्रतिदर्शज का मान परिकलित किया जाता है। यदि प्रतिदर्शों की संख्या बड़ी है तो सापेक्षिक बारंबारता बंटन के रूप में इन मानों को व्यवस्थित किया जा सकता है। जब प्रतिदर्शों की संख्या अनन्त (infinity) की ओर प्रवृत्त हो तो प्रतिदर्शज के मानों की परिणामी सापेक्षिक बारंबारता बंटन, दिए गए प्रतिदर्शज का *प्रतिदर्शी बंटन* (sampling distibution) कहलाएगी।

मान लीजिए, हम समष्टि माध्य को आकलित करने के इच्छुक हैं (जो कि प्राचल है) और जिसे μ द्वारा दर्शाया जाता है। इस समष्टि (N आकार की) से n आकार के यादृच्छिक प्रतिदर्शों का चयन किया जाता है। प्रतिदर्श माध्य $\bar{x} = \frac{1}{n}\Sigma x_i$ प्रतिदर्शज है जो समष्टि माध्य μ. के तदनुरूपी है। ध्यान दीजिए कि $\bar{x}$ यादृच्छिक चर है, इसके मान एक प्रतिदर्श से दूसरे में प्रायिकता के रूप में बदलते रहते हैं।

उदाहरण 16.2

यदि किसी समष्टि में 5 इकाइयाँ, 2,4,6,8 और 10 शामिल हैं तो मान लीजिए इसमें से बिना प्रतिस्थापन वाली सरल यादृच्छिक प्रतिचयन की विधि से 2 आकार वाले प्रतिदशों का चयन हमें करना है। हम प्रतिदर्श माध्य का प्रतिदर्शी बंटन और इसकी मानक त्रुटि की प्राप्ति करना चाहते हैं।

बिना प्रतिस्थापन के चुने जाने वाले प्रतिदर्शों की संख्या

$$= {}^{N}c_n = {}^{5}c_2 = \frac{5!}{2!(5-2)!} = \frac{5!}{2!3!} = \frac{5\times 4}{2\times 1} = \frac{20}{2} = 10$$

संगत प्रतिदर्श माध्य ($\bar{x}$) सहित संभावित प्रतिदर्शों को तालिका 16.1 में दर्शाया गया है।

तालिका 16.1 : संभावित प्रतिदर्श और प्रतिदर्श माध्य

प्रतिदर्श	प्रतिदर्श माध्य ($\bar{x}$)
(2, 4)	3
(2, 6)	4
(2, 8)	5
(2, 10)	6
(4, 6)	5
(4, 8)	6
(4, 10)	7
(6, 8)	7
(6, 10)	8
(8, 10)	9

अब हम प्रतिदर्श माध्य के बारंबारता बंटन की प्राप्ति कर सकते हैं :

सारणी 16.2: प्रतिदर्श माध्यों का बारंबारता बंटन

प्रतिदर्श माध्य ($\bar{x}$)	बारंबारता (f)
3	1
4	1
5	2
6	2
7	2
8	1
9	1

तालिका 16.2 में दिए गए बारंबारता बंटन से, जैसा कि तालिका 16.3 में दर्शाया गया है, हम प्रतिदर्श माध्य के प्रायिकता बंटन को दर्शा सकते हैं।

सारणी 16.3 : प्रतिदर्श माध्यों का प्रतिदर्शी बंटन

प्रतिदर्श माध्य ($\bar{x}$)	प्रायिकता $\left(\frac{f}{\sum f}\right)$
3	$\frac{1}{10}$
4	$\frac{1}{10}$
5	$\frac{2}{10}$
6	$\frac{2}{10}$
7	$\frac{2}{10}$
8	$\frac{1}{10}$
9	$\frac{1}{10}$

ध्यान दीजिए कि $\sum f$ जो पहले दर्शाए गए प्रतिदर्श माध्य का बारंबारता बंटन है, 10 के बराबर है। सारणी 16.3 में, हमने प्रायिकताओं के परिकलन के लिए सापेक्षिक बारंबारता का प्रयोग किया है।

16.8 प्रतिदर्शज की मानक त्रुटि

पिछले अनुभाग में हमने सीखा था कि समष्टि और प्रतिदर्श आकारों के आधार पर हम विविध प्रतिदर्शों की प्राप्ति कर सकते हैं। प्रत्येक प्रतिदर्शज से, हम अपेक्षित प्रतिदर्शज के लिए अलग-अलग मानों की प्राप्ति कर सकते हैं। इन मानों को प्रायिकता बंटन के रूप में व्यवस्थित कर सकते हैं, जिसे संबद्ध प्रतिदर्शज का प्रतिदर्शी बंटन कहते हैं। प्रतिदर्शज भी यादृच्छिक चर की ही भांति होता है। क्योंकि इसके द्वारा प्राप्त प्रत्येक मान से प्रायिकता जुड़ी रहती है। पिछले अनुभाग की सारणी 16.3 में हमने प्रतिदर्शज व उसकी प्रायिकता को दर्शाया है।

इकाई 14 में हमने सीखा था कि यादृच्छिक चर की गणितीय प्रत्याशा, इसके समांतर माध्य के बराबर होती है। आइए, प्रतिदर्शी बंटन के मानक विचलन और गणितीय प्रत्याशा का आकलन करें।

प्रतिदर्शी बंटन के दो महत्वपूर्ण गुणधर्मों पर ध्यान दीजिए।

1) प्रतिदर्शज के बंटन की प्रत्याशा, समष्टि प्राचल के बराबर होती है। यदि हम प्रतिदर्श माध्यों का प्रतिदर्शी बंटन प्राप्त करते हैं तब इसका प्रत्याशित मान, समष्टि माध्य के बराबर होता है। सांकेतिक रूप से $E(\bar{x}) = \mu$

2) प्रतिदर्शी बंटन का मानक विचलन, संबद्ध प्रतिदर्शज की ''मानक त्रुटि'' (standard error) कहलाता है। अतः यदि हमें प्रतिदर्श माध्य का प्रतिदर्शी बंटन प्राप्त है तब इसका मानक विचलन, ''प्रतिदर्श माध्य की मानक त्रुटि'' कहलाएगा। अतः मानक त्रुटि, समष्टि माध्य के गिर्द प्रतिदर्श माध्य के फैलाव को दर्शाता है। खंड 7 में हम देखेंगे कि मानक त्रुटि का प्रयोग परिकल्पना परीक्षण और सांख्यिकीय आकलन के लिए किया जाता है।

उदाहरण 16.3

सारणी 16.3 में दिए गए प्रतिदर्शी बंटन की मानक त्रुटि ज्ञात कीजिए।

हम जानते हैं कि प्रतिदर्श माध्य की मानक त्रुटि, प्रतिदर्शी बंटन का मानक विचलन है। अतः,

$$\sigma_{\bar{x}} = \sqrt{E(\bar{x})^2 - [E(\bar{x})]^2}$$

अब,

$$E(\bar{x}) = 3\times\frac{1}{10} + 4\times\frac{1}{10} + 5\times\frac{2}{10} + 6\times\frac{2}{10} + 7\times\frac{2}{10} + 8\times\frac{1}{10} + 9\times\frac{1}{10} = \frac{60}{10} = 6$$

और

$$E(\bar{x})^2 = 9\times\frac{1}{10} + 16\times\frac{1}{10} + 25\times\frac{2}{10} + 36\times\frac{2}{10} + 49\times\frac{2}{10} + 64\times\frac{1}{10} + 81\times\frac{1}{10} = \frac{390}{10} = 39.$$

$$\sqrt{E(\bar{x})^2 - [E(\bar{x})]^2} == \sqrt{39-36} = \sqrt{3} = 1.73$$

अतः इस मामले में प्रतिदर्श माध्य की मानक त्रुटि 1.73 है।

अब आपके मस्तिष्क में प्रश्न उठ रहा होगा कि, क्या मानक त्रुटि ज्ञात करने के लिए, हमें सभी संभावित प्रतिदर्शों को प्राप्त करना होगा? उपर्युक्त उदाहरण 16.3 में हमने सर्वप्रथम सभी संभावित प्रतिदर्शों को नोट किया, उन्हें आपेक्षिक बारंबारता बंटन के रूप में व्यवस्थित किया और तत्पश्चात् मानक त्रुटि को परिकलित किया। उदाहरण 16.3 में समटि का आकार और प्रतिदर्श आकार काफी छोटा था और इसी वजह से कार्य नियंत्रित दायरे में संपन्न किया जा सकता था। लेकिन, क्या आप कल्पना कर सकते हैं कि जब हमारे पास बड़े आकार की समष्टि और प्रतिदर्श हो तो? क्या होगा यह कार्य न केवल जटिल है बल्कि थकाऊ भी होगा। दरअसल, प्रतिचयन का समग्र लाभ लुप्त हो जाएगा यदि हम सभी संभावित प्रतिदर्शों का चयन शुरू कर देंगे !

दूसरे, क्या प्रतिदर्शी बंटन के लिए (खंड 5 में चर्चित) सैद्धांतिक प्रायिकता बंटन को फिट करना संभव है? दरअसल केंद्रीय सीमा प्रमेय (central limit theorem) के अनुसार, यदि किसी समष्टि से n आकार के प्रतिदर्श लिए जाते हैं तो n के बड़े मानों के लिए प्रतिदर्श माध्य सन्निकटतः प्रसामान्य रूप से बंटित हैं। इसलिए समष्टि का बंटन चाहो कुछ भी हो, $\bar{x}$ का प्रतिदर्शी बंटन, पर्याप्त रूप से बड़े प्रतिदर्शी आकारों के लिए सन्निकटतः रूप से प्रसामान्य होगा। यदि समष्टि प्रसामान्य है तब $\bar{x}$ का प्रतिदर्शी बंटन किसी भी प्रतिदर्श आकार के लिए प्रसामान्य होगा। यदि समष्टि प्रसामान्य रूप से अलग बंटित है तब $\bar{x}$ का

प्रतिदर्शी बंटन, यहाँ तक कि छोटे प्रतिदर्श आकारों के लिए भी लगभग प्रसामान्य होगा। इसके अलावा, यदि समष्टि प्रसामान्य रूप से बंटित नहीं भी है तब भी $\bar{x}$ का प्रतिदर्शी बंटन बड़े प्रतिदर्श आकारों के लिए सन्निकटतः प्रसामान्य होगा।

तीसरे, ऐसी समष्टि जिससे प्रतिदर्श लिए गए हैं, के मानक विचलन (σ) और $\bar{x}$ की मानक त्रुटि के बीच क्या संबंध है? निस्संदेह, समष्टि इकाइयों के प्रसार की तुलना में $\bar{x}$ का प्रसार कम होगा। $\bar{x}$ की मानक त्रुटि है:

$\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$ जहाँ $\sigma_{\bar{x}}$ $\bar{x}$ की मानक त्रुटि है और σ मूल समष्टि का मानक विचलन है।

अतः मानक त्रुटि का मान, समष्टि के मानक विचलन की तुलना में सदैव छोटा होता है क्योंकि मानक त्रुटि, प्रतिदर्श आकार के वर्ग मूल द्वारा विभाजित समष्टि के मानक विचलन के बराबर होती है।

उपर्युक्त कथन प्रतिस्थापन सहित सरल यादृच्छिक प्रतिचयन के मामले में सत्य हैं। जब प्रतिचयन बिना प्रतिस्थापन के हो तो ऐसे मामले में कुछ परिमित समष्टि शुद्धिकरण (finite population corection) करना होगा और मानक त्रुटि होगी $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}} \times \dfrac{N-n}{N-1}$

जब अनुपात $\dfrac{n}{N}$ काफी कम हो तो दोनों क्रियाविधियाँ एक जैसा ही परिणाम देती हैं। लेकिन जब समष्टि के आकार की तुलना में प्रतिदर्श आकार भी काफी बड़ा हो तो संशोधन गुणक को लागू करना ज़रूरी होता है।

मानक त्रुटि को हम कैसे व्यक्त करें? जैसा कि पहले बात हुई है, इससे प्रतिदर्शज के फैलाव का पता चलता है। लेकिन यदि मानक त्रुटि छोटी है तब ये अधिक संभावना होती है कि आकल, संबद्ध प्राचल के सन्निकट है।

उदाहरण 16.4

किसी समष्टि 2,5,8,13 पर विचार कीजिए।

i) समष्टि माध्य और समष्टि मानक विचलन को परिकलित कीजिए।

ii) प्रतिदर्श माध्य के प्रतिदर्शी बंटन का निर्माण कीजिए जब आकार 2 के यादृच्छिक प्रतिदर्शों का चयन समष्टि से किया जाता है

(क) प्रतिस्थापन सहित, (ख) बिना प्रतिस्थापन के

मामलो में बंटन का माध्य और मानक त्रुटि ज्ञात कीजिए।

iii) सत्यापित कीजिए कि प्रतिस्थापन वाले यादृच्छिक प्रतिचयन के मामले में $E(\bar{x}) = \mu$

और $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$ और बिना प्रतिस्थापन वाले यादृच्छिक प्रतिचयन के मामले में

$$E(\bar{x}) = \mu \quad \text{और} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\left(\sqrt{1-\frac{n-1}{N-1}}\right)$$

उत्तर :

प्राप्त समष्टि 2,5,8,13; समष्टि आकार $N = 4$; प्रतिदर्श आकार $n = 2$

i) समष्टि माध्य :

$$\mu = \frac{1}{N}\sum_{k=1}^{N} X_k = \frac{2+5+8+13}{4} = 7.$$

समष्टि मानक विचलनः

$$\sigma = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(X_k-\mu)^2} = \sqrt{\frac{(2-7)^2+(5-7)^2+(8-7)^2+(13-7)^2}{N}}$$

$$= \sqrt{\frac{25+4+1+36}{4}} = \sqrt{\frac{66}{4}} = \sqrt{16.5} = 4.06$$

ii) (क) संभावित प्रतिदर्शों की संख्या (प्रतिस्थापन सहित) =

$N'' = 4^2 = 16.$

प्रतिदर्शः

(2,2), (2,5), (2,8), (2,13),
(5,2), (5,5), (5,8), (5,13),
(8,2), (8,5), (8,8), (8,13),
(13,2), (13,5), (13,8), (13,13)

प्रतिदर्श माध्य :

2, 3.5, 5, 7.5,
3.5, 5, 6.5, 9,
5, 6.5, 8, 10.5,
7.5, 9, 10.5, 13.

प्रतिदर्श माध्य का प्रतिदर्शी बंटन

$\bar{x}$	f	$\frac{f}{N} = P$ (प्रायिकता)
2	1	$\frac{1}{16}$
3.5	2	$\frac{2}{16}$
5	3	$\frac{3}{16}$
6.5	2	$\frac{2}{16}$

7.5	2	$\frac{2}{16}$
8	1	$\frac{1}{16}$
9	2	$\frac{2}{16}$
10.5	2	$\frac{2}{16}$
13	1	$\frac{1}{16}$

प्रतिदर्शी बंटन को माध्य :

$$E\left(\bar{x}\right) = \sum_{i=1}^{n} P_i \bar{x}_i$$ (जहाँ $\bar{x}_i$ iवें प्रतिदर्श का माध्यम हो)

$$= \frac{1}{16}(2+7+15+13+15+8+18+21+13)$$

$$= \frac{1}{16} \times 112 = 7.$$

बंटन की मानक त्रुटि :

$$\sigma_{\bar{x}} = \sqrt{E(\bar{x}^2) - \{E(\bar{x})\}^2}$$

अब,

$$E(\bar{x}) = \sum_{i=1}^{n} P_i \bar{x}_i^2$$

$$= \frac{1}{16}(1\times 2^2 + 2\times 3.5^2 + 3\times 5^2 + 2\times 6.5^2 + 2\times 7.5^2 + 1\times 8^2 + 2\times 9^2 + 2\times 10.5^2 + 1\times 13^2)$$

$$= \frac{1}{16}(1\times 4 + 2\times 12.5 + 3\times 25 + 2\times 42.25 + 2\times 56.25 + 1\times 64 + 2\times 81 + 2\times 110.5 + 1\times 169)$$

$$= \frac{1}{16}(4 + 25 + 75 + 84.5 + 112.5 + 64 + 81 + 221 + 169)$$

$$= \frac{1}{16} \times 748 = 57.31$$

और

$$\{E(\bar{x})\}^2 = 7^2 = 49$$
$$\sigma_{\bar{x}} = \sqrt{57.31-49} = \sqrt{8.31} = 2.83$$

अतः प्रतिस्थापन वाले यादृच्छिक प्रतिचयन के मामले में प्रतिदर्शी बंटन का माध्य और मानक त्रुटि क्रमशः 7 और 7.83 हैं।

(ख) बिना प्रतिस्थापन वाले संभावित प्रतिदर्शों की संख्या है $^NC_n = {}^4C_2 = 6$

प्रतिदर्शः

(2,5), (2,8), (2,13), (5,8), (5,13), (8,13)

प्रतिदर्श माध्यः

3.5, 5, 7.5, 6.5, 9, 10.5

प्रतिदर्श माध्य का प्रतिदर्शी बंटनः

$\bar{x}$	f	$\frac{f}{N} = P$ (प्रायिकता)
3.5	1	$\frac{1}{6}$
5	1	$\frac{1}{6}$
7.5	1	$\frac{1}{6}$
6.5	1	$\frac{1}{6}$
9	1	$\frac{1}{6}$
10.5	1	$\frac{1}{6}$

समष्टि का माध्यः

$$E(\bar{x}) = \sum_{i=1}^{n} P_i \bar{x}_i$$
$$= \frac{1}{6}(3.5+5+7.5+6.5+9+10.5)$$
$$= \frac{1}{6} \times 42 = 7$$

$$\sigma_{\bar{x}} = \sqrt{E(\bar{x}^2) - \{E(\bar{x})\}^2}$$

अब

$$E(\bar{x}^2) = \sum_{i-1} P_i \bar{x}_i^2$$

$$= \frac{1}{6}(3.5^2 + 5^2 + 7.5^2 + 6.5^2 + 7.5^2 + 9^2 + 10.5^2)$$

$$= \frac{1}{6}(12.25 + 25 + 56.25 + 42.25 + 81 + 110.25)$$

$$= \frac{1}{6} \times 327 = 54.5$$

और जैसा कि हमें पता है,

$$\{E(\bar{x})\}^2 = 7^2 = 49$$

$$\sigma_{\bar{x}} = \sqrt{54.5 - 49} = \sqrt{5.5} = 2.35$$

अतः बिना प्रतिस्थापन वाले यादृच्छिक प्रतिचयन के मामले में प्रतिदर्शी बंटन का माध्य एवं मानक त्रुटि क्रमशः 7 और 2.35 हैं।

iii) प्रतिस्थापन वाले यादृच्छिक प्रतिचयन के मामले में,

$$E(\bar{x}) = 7 = \mu$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{4.06}{\sqrt{2}} = \frac{4.06}{1.414} = 2.87 \cong 2.83,$$

जैसा कि हमने प्रतिदर्श माध्य के प्रतिदर्शी बंटन से इसकी प्राप्ति स्वतंत्र रूप से की है।

बिना प्रतिस्थापन वाले यादृच्छिक प्रतिचयन के मामले में,

$$E(\bar{x}) = 7 = \mu$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{\sigma^2}{n} \times \frac{N-n}{N-1}} = \sqrt{\frac{16.48}{2} \times \frac{2}{3}} = \sqrt{8.24 \times 0.67} = \sqrt{5.52} = 2.35,$$

जैसा हमने प्रतिदर्श माध्य के प्रतिदर्शी बंटन से इसकी प्राप्ति स्वतंत्र रूप से की है।

अतः हमारे सभी निष्कर्ष सत्यापित हैं।

16.9 आकलक के वांछनीय गुणधर्म

मान लीजिए, θ ऐसा अज्ञात *प्राचल* है जिसकी हम अपेक्षा करते हैं। हम समष्टि से प्राप्त यादृच्छिक प्रतिदर्श के आधार पर θ को आकलन करना चाहते हैं। इस प्रयोजन के लिए हम प्रतिदर्शज T का प्रयोग करेंगे (जो प्रतिदर्श मानों का फलन है)। यहाँ T, θ का *आकलक* है और T का मान जिसे प्राप्त प्रतिदर्श से हमने प्राप्त किया है, θ का *आकल* है। दरअसल, इस मान को *बिंदु आकल* (point estimate) कहते हैं, क्योंकि यह आकलक का एक विशिष्ट मान है (अधिक जानकारी के लिए इकाई 18 देखें)।

इससे पहले, हमने प्रतिचयन और गैर-प्रतिचयन त्रुटियाँ से संबंधित संकल्पनाओं की चर्चा की थी। यदि हम उन्हें दुबारा ध्यान में लाएँ तो हमें पता है कि प्रतिदर्श प्रतिदर्शज और समष्टि प्राचल के बीच (संकेत को अनदेखा करते हुए) का पूर्ण अंतर अर्थात $|T-\theta|$, प्रतिचयन त्रुटि के विस्तार को मापता है। ध्यान दीजिए कि आकलक अनिवार्यतः, प्राचल के आकल को परिकलित करने का एक सूत्र है। और ऐसे ही बहुत से संभावित आकलक (वैकल्पिक सूत्र) हो सकते हैं जिन्हें इस प्रयोजन के लिए प्रयुक्त किया जा सकता है। अतः ऐसे कुछ वांछनीय गुणधर्म होने चाहिए जिनके आधार पर हम प्राचल का अनुमान लगाने के लिए किसी विशिट आकलक का चयन कर सकते हैं। θ का अच्छा आकलक होने के लिए T और θ के बीच का अंतर $|T-\theta|$ यथासंभव कम होना चाहिए। यह सुनिश्चित करने के लिए विविध नजरियों का सुझाव दिया गया है।

16.9.1 अनभिनता

हमने पहले ही देखा है कि प्रतिचयन परिवर्तनशीलता के कारण प्रतिदर्शज के मान, एक प्रतिदर्श से दूसरे प्रतिदर्श में अलग-अलग होते हैं। यद्यपि प्रतिदर्शज के एकल मान, औसतन, अज्ञात प्राचल से अलग हो सकते हैं लेकिन प्रतिदर्शज का मान, प्राचल के बराबर होना चाहिए। अन्य शब्दों में, T के प्रतिदर्शी बंटन की θ के प्रति केंद्रीय प्रवृत्ति होनी चाहिए। इसे आकलक का अनभिनता (unbiasedness) संबंधी गुण कहते हैं। इसका अर्थ है कि यद्यपि प्राचल के अज्ञात मान की तुलना में दिए गए प्राचल का एकल मान उच्च या निम्न हो सकता है, फिर भी आकलक ऐसे ही मानों की प्राप्ति करने का पक्षधर नहीं होता जो अज्ञात प्राचल की तुलना में सदैव बड़े या छोटे होते हैं। यदि हम मान लें कि माध्य (यहाँ प्रत्याशा) केंद्रीय प्रवृत्ति के लिए सही मान है तब θ के लिए *T अनभिनत आकलक* है, यदि $E(T) = \theta$ है।

16.9.2 न्यूनतम प्रसरण

यह भी वांछनीय है कि समटि प्राचल के इर्द-गिर्द किसी अनभिनत आकलक के सभी संभावित मानों का औसतन प्रसार यथासंभव कम होना चाहिए। इससे प्राचल से आकल के दूर होने की संभावना कम हो जाएगी। यदि हम मान लें कि प्रकीर्णन के लिए प्रसरण उचित माप है तब हम चाहेंगे कि सभी अनभिनत आकलकों में से T का न्यनूतम प्रसरण होना चाहिए: सांकेतिक रूप से $V(T) \leq V(T')$ जहाँ V प्रसरण और T' कोई अन्य अनभिनत आकलक है।

कोई आकलक T, जो अनभिनत है और जिसका सभी अनभिनत आकलकों में से न्यूनतम प्रसरण हो, ''न्यूनतम प्रसरण अनभिनत आकलक'' (Minimum Variance Unbiased Estimator) कहलाता है। आइए किसी उदाहरण पर विचार करें। मान लीजिए N आकार की प्राप्त समष्टि से हमारे पास n आकार के यादृच्छिक प्रतिदर्श हैं। इस मामले में प्रतिदर्श माध्य है। $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, जहां x_i प्रतिदर्श का i वां सदस्य है। यह सिद्ध किया जा सकता है कि यह समष्टि माध्य μ का अनभिनत आकलक है। सांकेतित रूप से

$$E(\bar{x}) = \mu$$

तथापि, यह दर्शाया जा सकता है कि

प्रतिदर्श प्रसरण $s^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$ का प्रत्याशा σ^2 के बराबर नहीं है।

$$E\left(s^2\right) \neq \sigma^2$$

इसके विपरीत यदि हम $s'^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$, के रूप में प्रतिदर्श प्रसरण को परिभाषित करें, तब $\sigma'^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2$ का s'^2 अनभिनत आकलक है।

मान लीजिए, इसके बाद प्रतिदर्श मान न केवल यादृच्छिक है बल्कि स्वतंत्र (प्रतिस्थापन सहित यादृच्छिक प्रतिदर्श) भी है और निहित समटि प्रसामान्य है। यह भी दर्शाया जा सकता है कि प्रतिदर्श माध्य $\bar{x}$ समष्टि माध्य μ का न केवल अनभिनत आकलक है बल्कि μ के सभी अनभिनत आकलकों में से इसका प्रसरण न्यूनतम भी है।.

16.9.3 संगति और दक्षता

एक अन्य दृष्टिकोण है कि आकलक T को प्रतिदर्श आकार n के बढ़ने के साथ-साथ अज्ञात समष्टि प्राचल θ के सन्निकट पहुंचना चाहिए। यहां T स्वयं भी यादृच्छिक चर है। इस बात को हम प्रायिकतात्मक या प्रसंभाव्यता के रूप में अभिव्यक्त कर सकते हैं क्योंकि प्रतिदर्शज T को प्रसंभाव्यता प्राचल θ के प्रति अभिसरित होना चाहिए जब $n \to \infty$. इस गुणधर्म वाला प्रतिदर्शज T, θ का *संगत* (consistent) आकलक कहलाता है।

असल जीवन में समान प्राचल θ के संगत आकलकों की बड़ी संख्या बहुधा देखने को मिलती है। ऐसी स्थिति में निस्संदेह ऐसे संगत आकलकों से कुछ अतिरिक्त निकष (कसोटी) का चयन आवश्यक होता है। ऐसा एक निकष यह माँग सकता है कि न केवल T प्रसंभाव्य रूप से अभिसरित हो बल्कि इस कार्य को यह यथाशीघ्र भी करे। अधिक गहन अध्ययन न करते हुए हम सिर्फ इतना उल्लेख करना चाहेंगे कि जब किसी प्रतिदर्श का आकार n लगातार बढ़ता है तो आकलक, प्रसामान्य बंटन का रूप धारण कर लेता है। ऐसे आकलक उपगामी रूप से प्रसामान्य (asymptotically normal) कहलाते हैं। यदि हम ऐसे संगत आकलकों पर ध्यान केंद्रित करें जो उपगामी प्रसामान्य हैं तो उनके अभिसर की तीव्रता उनके संबद्ध *उपगामी प्रसरण* (Asymptotic Variance) द्वारा दर्शायी जाती है। दरअसल, न्यूनतम उपगामी प्रसरण आकलक के लिए अभिसरण सबसे तीव्र रूप से होता है। समष्टि प्राचल के ऐसे सभी उपगामी प्रसामान्य संगत आकलकों मे से इस प्रकार के आकलक को 'दक्ष आकलक' (efficient estimator) कहते हैं।

बोध प्रश्न 2

1) निम्नलिखित संकल्पनाओं को परिभाषित कीजिए:

क) सरल यादृच्छिक प्रतिदर्श

ख) प्रतिदर्शी बंटन

ग) मानक त्रुटि

....

....

....

....

....

....

2) निम्नलिखित के बीच का अंतर स्पष्ट कीजिए:

   घ) सरल यादृच्छिक प्रतिचयन (प्रतिस्थापन सहित) और सरल यादृच्छिक प्रतिचयन (प्रतिस्थापन रहित)

   च) सरल यादृच्छिक प्रतिचयन और स्तरित यादृच्छिक प्रतिचयन

....

....

....

....

....

....

3) यदि प्राप्त समष्टि 1,2,5,6 है तो आकार 2 के सभी संभावित प्रतिदर्शों को दर्शाइए:

   i) प्रतिस्थापन सहित और

   ii) प्रतिस्थापन रहित

....

....

....

....

....

....

4) समष्टि 2, 4,6 है । मान लीजिए बिना प्रतिस्थापन वाली यादृच्छिक प्रतिचयन विधि का प्रयोग करते हुए आकार 2 के प्रतिदर्श का चयन हमें इस समष्टि से करना है।

   क) प्रतिदर्श माध्य का प्रतिदर्शी बंटन प्रस्तुत कीजिए।

   ख) मानक त्रुटि परिकलित कीजिए।

....

....

....

....

....

16.10 सारांश

इस इकाई में हमने सांख्यिकीय छानबीन करने की जनगणना विधि और प्रतिदर्श विधि के बीच के अंतर को स्पष्ट किया। हमने देखा कि विविध संसाधन अवरोधों के कारण, हमेशा

जनगणना विधि को अपनाया नहीं जा सकता। इसके अलावा इसमें काफी अधिक कार्य विस्तार होने कारण इसमें गैर-प्रतिचयन संबंधी त्रुटियाँ उत्पन्न होने की संभावना बहुत अधिक होती है। दूसरी तरफ, प्रतिदर्श सर्वेक्षण के कुछ निश्चित फायदे हैं। यदि सही ढंग से प्रतिदर्श सर्वेक्षण किया जाए तो आमतौर पर त्रुटि होने की संभावना काफी कम हो जाती है। प्रतिदर्श सर्वेक्षण का एक ठोस वैज्ञानिक आधार है। जिसके परिणामस्वरूप (यादृच्छिक प्रतिदर्शों से प्राप्त) संबद्ध प्रतिदर्शज का प्रतिदर्शी बंटन, प्राचल के लिए मूल्यांकन का विषयपरक आधार बनाता है।

16.11 शब्दावली

आकल (Estimate) : आकलक द्वारा प्राप्त किए जाने वाला विशिष्ट मान।

आकलक (Estimator) : प्रतिदर्शज या इसके परिकलन में शामिल सूत्र का विशिट कार्यात्मक स्वरूप। आमतौर पर प्रतिदर्शज और आकलक का अर्थ समान ही होता है।

प्राचल (Parameter) : समष्टि की कुछ विशेाताओं का माप।

समष्टि (Population) : किसी निश्चित समय और स्थान पर विशिष्ट प्रकार की इकाइयों का समूचा संग्रह।

यादृच्छिक प्रतिचयन (Random Sampling) : ऐसी प्रक्रिया जिसमें समष्टि के सभी सदस्यों की प्रतिदर्श में चुने जाने की निश्चित संभावना होती है। इसे प्रायिकता प्रतिचयन भी कहते हैं।

प्रतिदर्श (Sample) : समष्टि का उप-समुच्चय। अतः यह समष्टि की कुछ इकाइयों का समूह है।

प्रतिदर्शी बंटन (Sampling Distribution) : प्रतिदर्शज का प्रायिकता बंटन

प्रतिचयन त्रुटि (Sampling Error) : प्राचल और संबद्ध प्रतिदर्शज के बीच का पूर्ण अंतर।

प्रतिचयन उच्चावचन (Sampling Fluctuation) : विभिन्न प्रतिदर्शों से परिकलित प्रतिदर्शज के मानों में पाया जाने वाला अन्तर।

सरल यादृच्छिक प्रतिचयन (Simple Random Sampling) : ऐसी प्रतिचयन संबंधी प्रक्रिया जिसमें समष्टि के प्रत्येक सदस्य की प्रतिदर्श में चुने जाने की संभावना एक जैसी होती है।

मानक त्रुटि (Standard Error) : प्रतिदर्श में शामिल की जाने वाली इकाइयों के मानों का फलन। इसका मूल प्रयोजन कुछ प्राचलों का आकल करना है।

प्रतिदर्शज (Statistic) : प्रतिदर्श में शामिल की जाने वाली इकाइयों के मानों का फलन। इसका मूल प्रयोजन कुछ प्राचलों का आकलन करना है।

सांख्यिकीय अनुमिति (Statistical Inference) समष्टि से प्राप्त ज्ञात प्रतिदर्श के आधार पर अज्ञात समष्टि विशषताओं के बारे में निष्कर्ष निकालने की प्रक्रिया।

16.12 कुछ उपयोगी पुस्तकें

Bhardwaj, R.S., 1999, *Business Statistics* (First Edition), Excel Books, New Delhi, Chapter 20.

Nagar, A.L. and Das, R.K., 1988, *Basic Statistics*, Oxford University Press, Delhi, Chapter 9.

Goon, A.M., Gupta, M.K. and Dasgupta, B., 1971, *Fundamentals of Statistics*, volume 1 (Fourth Edition), The World Press Private Limited, Calcutta, Chapters 14, 15 and 16.

16.13 बोध प्रश्नों के उत्तर या संकेत

बोध प्रश्न 1

1) इकाई पढ़ें और प्रत्येक (पारिभाषिक) शब्द को एक या दो वाक्यों में व्यक्त करें।
2) इकाई पढ़ें और अंतर स्पष्ट करें।
3) इकाई पढ़ें और कुछ वाक्यों में उत्तर स्पष्ट करें।

बोध प्रश्न 2

1) इकाई पढ़ें और प्रत्येक (पारिभाषिक) शब्द को एक या दो वाक्यों में व्यक्त करें।
2) इकाई पढ़ें और अंतर स्पष्ट करें।
3) इकाई में दिए उदाहरण 16.2 को पढ़ें और हल करें।
4) 0.82

इकाई 17 प्रतिचयन की क्रियाविधि

इकाई की रूपरेखा

17.0 उद्देश्य

इस इकाई का अध्ययन करने के बाद, आप :

- प्रतिदर्श की विविध विधियों का वर्णन कर सकेंगे;
- प्रतिदर्श के लिए यादृच्छिक संख्या सारणियों का प्रयोग कर सकेंगे; और
- प्रतिदर्श आमाप का निर्धारण कर सकेंगे।

17.1 प्रस्तावना

इकाई 2 में आपने प्राथमिक और द्वितीयक आंकड़ों के बारे में अध्ययन किया। उस इकाई में हमने प्राथमिक आंकड़ें इकट्ठा करते समय प्रत्यक्ष साक्षात्कार, टेलीफ़ोन सर्वेक्षण, डाक संबंधी सर्वेक्षण जैसी विविध सर्वेक्षण तकनीकों के प्रयोग के बारे में भी चर्चा की थी। इकाई 16 में आपने प्रतिचयन (sampling) के अर्थ, प्रतिचयन के लाभ और प्रतिचयन त्रुटि जैसी बातों की भी जानकारी प्राप्त की थी।

सांख्यिकी में, अक्सर समष्टि या बड़े समुच्चय के बारे में निष्कर्ष निकालने के लिए हम प्रतिदर्श (sample) जोकि बड़े समुच्चय के उपसमुच्चय है, पर निर्भर करते हैं। जैसे, आगामी चुनावों में आप दिल्ली वासियों के मतदान संबंधी व्यवहार के बारे में जानना चाहते हैं। इस बारे में, आप किससे पूछताछ करेंगे? निस्संदेह, दिल्ली के हरेकवासी को पूछना कि वह किसे वोट देगा/देगी, संभव नहीं है। बजाय इसके, आप दिल्ली के मतदाताओं के छोटे से समूह से इस बारे में बातचीत करके, उनकी प्रतिक्रिया को ध्यान में रखकर समूची दिल्ली के बारे में अपना निष्कर्ष प्राप्त कर सकते हैं। इस मामले में दिल्ली के कुल मतदाता ऐसी समष्टि है और प्रतिदर्श के रूप में हमने उन मतदाताओं को रखा है, जिनसे असल में हमने बातचीत की थी।

आदर्श रूप से प्रतिदर्श में उस समष्टि की विशेषताएं प्रतिबिंबित होनी चाहिए, जिसमें से उसे लिया गया है। ऐसे मामलों में, प्रतिदर्श से प्राप्त निष्कर्ष, उससे संबंधित पूरे समष्टि पर लागू किए जा सकते हैं।

इस इकाई में आप सीखेंगे कि समष्टि की विभिन्न विशेषताओं के अंतर्गत प्रतिदर्श किस प्रकार सृजित किए जाते हैं और प्रतिदर्श आमाप (size) का निर्धारण किस प्रकार किया जाता है।

17.2 प्रतिचयन प्रक्रिया

प्रतिदर्श सर्वेक्षण करते समय, प्रतिचयन प्रक्रिया निर्धारित करती है कि सर्वेक्षण में प्रतिचयन संबंधी कौन सी इकाइयों को शामिल किया जाएगा। संग्रहित नमूने आंकड़ों को अधिक प्रबन्ध और लागत धार्य बनाते हैं। ये प्रतिदर्श के आधार पर समष्टि की विशेषताओं को इस योग्य बनाते हैं कि इससे प्राप्त निष्कर्षों में कम से कम त्रुटियाँ नज़र आएं। प्रतिचयन

प्रक्रिया में प्रतिदर्शदायक समष्टि की परिभाषा, प्रतिचयन समूह की पहचान, प्रतिचयन विधि का चयन और प्रतिचयन इकाइयों का चयन शामिल है।

1) **सर्वेक्षण का उद्देश्य:** प्रतिदर्श सर्वेक्षण में सर्वप्रथम विशिष्ट उद्देश्यों पर ध्यान केंद्रित होता है। हमें सर्वेक्षण के उद्देश्यों का स्पष्ट पता होना चाहिए, क्योंकि बाकी के सभी चरणों की रूपरेखा जैसे लक्ष्य समष्टि (target population), प्रतिचयन समूह (sampling frame), प्रतिचयन की क्रियाविधि, सर्वेक्षण संबंधी उद्देश्यों को ध्यान में रख कर बनाई जाती है।

2) **प्रश्नावली निर्माण:** सर्वेक्षण के उद्देश्यों को ध्यान में रख कर प्रश्नावली की रूपरेखा विकसित की जाती है। इस पाठ्यक्रम की इकाई 2 में हम पहले ही प्रश्नावली के निर्माण में शामिल मुख्य चरणों का अध्ययन कर चुके हैं। प्रश्नावली के साथ-साथ हमें जाँचकर्त्ताओं के लिए प्रशिक्षण संबंधी दस्तावेज़ विकसित करने की भी आवश्यकता है। [विशेषरूप से तब, जब प्रतिदर्श सर्वेक्षण बड़े पैमाने पर किया जाता है और जिसमें बड़ी संख्या में जाँचकर्त्ता शामिल होते हैं।]

3) **लक्ष्य समष्टि को परिभाषित करना:** समष्टि से प्रतिदर्श एकत्र करने के लिए हमें ऐसे लक्ष्य समष्टि की पूरी ज़ानकारी होनी चाहिए, जिसके बारे में प्रतिदर्श के मद्देनज़र निष्कर्ष निकाले जाने हैं। लक्ष्य समष्टि को सार्वभौमिक रूप में भी माना जाता है। यह ऐसे व्यक्तियों का समूह है जिनके लिए हम प्रतिदर्श के माध्यम से निष्कर्ष निकालते हैं या जिसे हम सामान्य रूप देना चाहते हैं। जैसे, आप दिल्ली में योग्य दंपत्तियों द्वारा प्रयुक्त परिवार नियोजन की विधियों पर प्रतिदर्श सर्वेक्षण करना चाहते हैं। इस संदर्भ में जनन आयु समूह में शामिल दिल्ली के ऐसे सभी दंपत्ति लक्ष्य समष्टि की रचना करते हैं।

4) **प्रतिचयन समूह (Sampling Frame) की पहचान करना:** प्रतिचयन समूह लक्ष्य समष्टि से प्राप्त इकाइयों की सूची है। यह असल में लक्ष्य समष्टि की व्यावहारिक परिभाषा है। हमारे पहले उदाहरण में, जो परिवार नियोजन विधियाँ प्रयोग करने वाले दिल्ली के योग्य दंपत्तियों से संबंधित था, उस उदाहरण में जनन आयु समूह में शामिल सभी व्यक्ति प्रतिचयन–समूह का निर्माण करते हैं। कई बार किसी न किसी वज़ह से हम लक्ष्य समष्टि की सभी इकाइयों को सूचीबद्ध करने की स्थिति में नहीं होते। जैसे, हम टेलीफ़ोन डायरेक्टरी पर आधारित सर्वेक्षण करना चाहते हैं। ऐसे मामले में निश्चित रूप से ऐसे लोगों को सूची में शामिल नहीं किया जाएगा, जिनके टेलीफ़ोन नम्बर डायरेक्टरी में नहीं हैं। इस प्रकार की त्रुटि को प्रतिचयन समूह त्रुटि कहते हैं।

5) **प्रतिचयन की क्रियाविधि का चयन:** प्रतिचयन समूह की एक बार पहचान हो जाने पर, हम सर्वेक्षण संबंधी प्रतिदर्श के चयन के लिए, प्रतिचयन की उपयुक्त क्रियाविधि का चयन कर सकते हैं। इस इकाई के अगले अनुभाग में हम प्रतिचयन की विविध क्रियाविधियों की विस्तृत चर्चा करेंगे।

6) **प्रतिचयन इकाइयों का चयन:** प्रतिचयन इकाइयाँ प्रतिचयन समूह की ऐसी इकाइयां हैं, जिन्हें प्रतिचयन की उपयुक्त क्रियाविधि के प्रयोग से प्रतिदर्श में शामिल किया जाता है। अनिवार्यतः एक 'प्रतिचयन इकाई' ऐसी इकाई होती है जिससे जुड़े आंकड़ों को एकत्र किया जाता है। जैसे, आप प्रतिचयन समूह से 1000 प्रतिचयन इकाइयाँ (दिल्ली के जनन आयु समूह वाले सभी व्यक्ति) अपने प्रतिदर्श सर्वेक्षण के लिए लेना चाहते हैं।

7) **सर्वेक्षण आँकड़ा प्रक्रिया:** प्रतिचयन इकाइयों के चयन के बाद, अगला चरण आँकड़ों को इकट्ठा करना और इन्हें संसाधित करना है। अब हमें अधूरी प्रश्नावलियों की जाँच और शंकाग्रस्त प्रतिक्रियाओं का निवारण करने की आवश्यकता है। इसके बाद आँकड़ा प्रविष्टि और सारणी बनाने का कार्य शुरू होता है।

8) **आँकड़ों का विश्लेषण:** श्रृंखला का अगला चरण आँकड़ों का विश्लेषण करना है। अपनी आवश्यकताओं को ध्यान में रखकर हम विविध सांख्यिकीय साधनों के प्रयोग से आँकड़ों का विश्लेषण करते हैं।

9) **परिणामों का प्रकाशन और प्रचार:** विश्लेषित आँकड़ों के आधार पर हम तकनीकी और शोध आधारित रिपोर्ट तैयार करते हैं। अंत में, संगोष्ठियों में सर्वेक्षण के सामाजिक-आर्थिक परिणामों और उनके प्रभावों की चर्चा की जाती है।

17.3 प्रतिचयन का प्रकार

प्रतिदर्श चयन में शामिल महत्वपूर्ण चरण है प्रतिदर्श चयन की विधि का निर्धारण करना। मोटेतौर पर प्रतिचयन की दो मुख्य क्रियाविधियाँ हैं – (क) प्रायिकता (या यादृच्छिक) प्रतिचयन, (ख) गैर-प्रायिकता (या गैर-यादृच्छिक) प्रतिचयन।

17.3.1 प्रायिकता प्रतिचयन

प्रायिकता प्रतिचयन में, समष्टि की सभी इकाइयों को प्रतिदर्श में शामिल किए जाने की संभावना बनी रहती है। प्रायिकता प्रतिचयन को यादृच्छिक प्रतिचयन भी कहते हैं। प्रायिकता प्रतिचयन विषयनिष्ठ क्रियाविधि है जिसमें प्रतिचयन इकाई के चयन की प्रायिकता का पहले से पता होता है। यहाँ प्रतिदर्श में हर इकाई के शामिल किए जाने की प्रायिकता गैर-शून्य है।

प्रायिकता प्रतिचयन क्रियाविधि के प्रयोग करने का फ़ायदा है कि इसमें हमारी कोई भूमिका नहीं है कि प्रतिदर्श के लिए कौन-सी विशिष्ट समष्टि इकाइयों को चुना गया है। प्रायिकता प्रतिचयन की क्रियाविधि व्यक्तिगत पक्षपात से मुक्त है और प्रतिदर्श के चयन के लिए वस्तुनिष्ठ योजना पर ध्यान केंद्रित करती है। इस कारण से प्रायिकता प्रतिचयन की क्रियाविधियाँ यह सुनिश्चित करने में अच्छा-ख़ासा समय लेती हैं कि प्रतिदर्श में हर इकाई को शामिल किए जाने की गैर-शून्य प्रायिकता है। प्रायिकता प्रतिचयन क्रियाविधि प्रयोग करने का अन्य फ़ायदा है कि प्रतिचयन त्रुटि के विस्तार को समझ पाना संभव है और अनुमानित त्रुटि का आकलन, बहुत-सी स्थितियों में प्राप्त निष्कर्ष में क्षमता सृजित करने में सहायता करता है। प्रायिकता प्रतिचयन क्रियाविधि को पूरी तरह से लागू किया जाता है तो इसमें पक्षपात की संभावना ख़त्म हो जाती है और इसलिए अब यह सापेक्षिक रूप से समष्टि का प्रतिनिधित्व कर सकती है। व्यावहारिक रूप से हम कभी भी पूरी तरह निश्चित नहीं हो सकते कि प्रतिदर्श से प्राप्त परिणाम, समष्टि के लिए भी सही साबित होंगे। इसी वज़ह से बहुधा हमारे लिए जानना पर्याप्त है कि समष्टि से विचलित होने का जोख़िम 1 प्रतिशत या 5 प्रतिशत है जैसा कि विश्वास्यता अंतरालों की गणना करते समय हम अध्ययन करेंगे (इकाई 18 देखें।)

17.3.2 गैर-प्रायिकता प्रतिचयन

गैर-प्रायिकता प्रतिचयन की क्रियाविधि, क्रियाविधि की विषयनिष्ठता की तुलना में मनुष्य के विवेक/व्यक्तिगत निर्णय पर अधिक निर्भर करती है। गैर-प्रायिकता प्रतिचयन

क्रियाविधि को बहुधा ग़ैर-यादृच्छिक प्रतिचयन क्रियाविधि के रूप में भी देखा जाता है। शामिल की जाने वाली प्रतिचयन इकाइयों के निर्धारण में मनुष्य का व्यक्तिगत निर्णय विशेष भूमिका निभाता है। ग़ैर-प्रायिकता प्रतिचयन में हरेक समष्टि इकाई के लिए चयन की प्रायिकता पहले से पता नहीं होती। इसलिए हम पूरी तरह सुनिश्चित नहीं हो सकते कि प्रतिदर्श समष्टि का प्रतिनिधि होगा। इसके साथ-साथ हम चयन की प्रायिकता का निर्धारण नहीं कर सकते अनुभव के आधार पर शोधकर्त्ता की विषय विशेषज्ञता और निष्पादन या प्रतिदर्श के रूप में शामिल की जाने वाली इकाइयों की पहचान की जाती है। कभी-कभार इसमें व्यक्तिगत पक्षपात भी शामिल होता है।

17.4 सरल यादृच्छिक प्रतिदर्श का चयन

सरल यादृच्छिक प्रतिचयन (simple randam sampling), प्रतिचयन की बुनियादी क्रियाविधि है जहाँ हम समष्टि से प्रतिदर्श का चयन करते हैं। इस क्रियाविधि में हरेक इकाई का चयन समग्र रूप से संयोग तंत्र (chance mechanism) के माध्यम से होता है और इसमें समष्टि की हरेक इकाई की प्रतिदर्श के रूप में शामिल होने की एक जैसी संभावना होती है। दिए गए आमाप के हरेक संभव प्रतिदर्श के चुने-जाने की समान संभावना होती है अर्थात समष्टि की हर इकाई की, प्रतिचयन प्रक्रिया में किसी भी चरण पर चुने जाने की संभावना होती है और किसी एक इकाई के चयन का समष्टि में किसी दूसरी इकाई के चयन पर प्रभाव नहीं होता।

सरल यादृच्छिक प्रतिचयन का प्रयोग समरूप (homogeneous) समष्टि के लिए किया जाना चाहिए अर्थात समष्टि की सभी इकाइयों में ऐसा समान गुण होना चाहिए जिसे हम मापना चाहते हैं। समरूप विशेषताओं में आयु, लिंग, आय, सामाजिक स्थिति, भौगोलिक क्षेत्र आदि शामिल किए जा सकते हैं।

ऐसी दो सर्वाधिक सामान्य विधियाँ हैं जिनका प्रयोग सरल यादृच्छिक प्रतिदर्श के सार की प्राप्ति के लिए किया जाता है। पहली है–लॉटरी विधि और दूसरी है–यादृच्छिक संख्या चयन विधि। हम कौन-सी विधि का प्रयोग करेंगे, इस ओर ध्यान दिए बिना, प्रतिचयन समूह में शामिल प्रत्येक इकाई को अंकित संख्या दी जानी चाहिए।

17.4.1 लॉटरी/प्रचय विधि

यादृच्छिक प्रतिदर्श चयन की सरलतम विधि लॉटरी निकालना है। इस विधि में, एक डिब्बे में इकाई अंकित नंबरों को रखकर, अच्छे से हिला दिया जाता है। आख़िर में यादृच्छिक तरीके से डिब्बे से तब तक नंबर निकाले जाते हैं जब तक कि वांछनीय प्रतिदर्श आमाप (sample size) की प्राप्ति न हो जाएं। मान लीजिए, हम N समष्टि इकाइयों से n प्रतिदर्श इकाइयाँ चुनना चाहते हैं। इसके लिए हमारे पास 1 से N तक की संख्याएं हैं। हम प्रत्येक समष्टि इकाई को एक संख्या देते हैं; और इन संख्याओं को N पर्चियों पर लिखते हैं। यथासंभव, इन पर्चियों को आकृति, आकार, रंग आदि में एक जैसा होना चाहिए। अब इन पर्चियों को डिब्बे में डाल कर, अच्छी तरह हिलाया जाता है। आख़िर में, एक-एक करके हम n पर्चियों को निकालते हैं। निकाली गयीं पर्चियों की संख्याओं के तदनुरूप वाली n इकाइयाँ, यादृच्छिक प्रतिदर्श को गठित करती है।

उदाहरण 17.1: मान लीजिए आप किसी बैंक की शाखा से संबंधित कुछ शोधकार्य कर रहे हैं और बैंक की इस शाखा में प्रदत्त सेवा की गुणवत्ता पर ग्राहकों के विचारों का मूल्यांकन करना चाहते हैं। आपको लॉटरी विधि से सरल यादृच्छिक प्रतिचयन क्रियाविधि के

प्रयोग से प्रतिदर्श (नमूने) के रूप में 100 ग्राहकों का चयन करना है। इसके लिए, सर्वप्रथम आपको प्रतिचयन-समूह को क्रमबद्ध करना होगा। इसके लिए आपको बैंक के रिकार्ड से खाता धारकों की पहचान करनी होगी। अब, आपको इन सभी खाताधारकों को क्रमसंख्या देने की आवश्यकता है। मान लीजिए, खाताधारकों की संख्या 1000 है और आपको 100 को चुनना है तो आप 100 / 1000 = 10% खाताधारकों को प्रतिदर्श के रूप में चुनेंगे। आप चाहें तो क्रम संख्या को लिख सकते हैं, उन्हें फाड़ कर अलग पर्चियों का रूप दे कर, उन्हें डिब्बे में रख सकते हैं। अब इन्हें अच्छे से हिलाइए, आँखें बंद कीजिए और पहली 100 पर्चियों को बाहर निकाल लीजिए।

लॉटरी विधि का दोष यह है कि विधि अनावश्यक लंबी है और प्राप्त प्रतिदर्श की गुणवत्ता इस बात पर निर्भर करती है कि हमने इन पर्चियों को किस प्रकार मिलाया और किस प्रकार बेतरतीब तरीके से इन्हें उठाया। इसके अलावा, समष्टि आमाप बढ़ने के साथ-साथ, लॉटरी विधि के प्रयोग से प्रतिदर्श निकालने में कठिनाई भी अधिकाधिक बढ़ने लगती है।

17.4.2 यादृच्छिक संख्या चयन विधि

यादृच्छिक संख्याएं ऐसे अंकों का संग्रहण हैं जिन्हें प्रायिकता तंत्र के माध्यम से जनित किया जाता है। यादृच्छिक संख्याओं के निम्नलिखित गुणधर्म हैं:

क) इसकी प्रायिकता कि प्रत्येक अंक (0,1,2,3,4,5,6,7,8 या 9) किसी जगह उभर सकता है, एक जैसी है अर्थात 1 / 10 है।

ख) किन्हीं दो अंकों का किन्हीं दो स्थानों पर उभरना, एक दूसरे से स्वतंत्र है।

इस विधि में समष्टि की प्रत्येक इकाई को क्रमबद्ध रूप में विशेष संख्या दी जाती है। प्रतिदर्श निकालने के लिए हम यादृच्छिक संख्याओं की सारणी (Random Number Table) को फिशर एंड येटस् (1963): स्टैटिसिटिकल टेबलस् फॉर बायोलॉजिकल, एग्रीकल्चरल एंड मेडिकल रिसर्च में अलग-अलग जगह देख सकते हैं। यादृच्छिक संख्या सारणी का एक उदाहरण सारणी 17.1 में दर्शाया गया है।

सारणी 17.1 यादृच्छिक संख्या सारणी

39634	62349	74088	65564	16379	19713	39153	69459	17986	24537
14595	35050	40469	27478	44526	67331	93365	54526	22356	93208
30734	71571	83722	79712	25775	65178	07763	82928	31131	30196
64628	89126	91254	24090	25752	03091	39411	73146	06089	15630
42831	95113	43511	42082	15140	34733	68076	18292	69486	80468
80583	70361	41047	26792	78466	03395	17635	09697	82447	31405
00209	90404	99457	72570	42194	49043	24330	14939	09865	45906
05409	20830	01911	60767	55248	79253	12317	84120	77772	50103
95836	22530	91785	80210	34361	52228	33869	94332	83868	61672
65358	70469	87149	89509	72176	18103	55169	79954	72002	20582
72249	04037	36192	40221	14918	53437	60571	40995	55006	10694
41692	40581	93050	48734	34652	41577	04631	49184	39295	81776
61885	50796	96822	82002	07973	52925	75467	86013	98072	91942
48917	48129	48624	48248	91465	54898	61220	18721	67387	66575
88378	84299	12193	03785	49314	39761	99132	28775	45276	91816
77800	25734	09801	92087	02955	12872	89848	48579	06028	13827

24028	03405	01178	06316	81916	40170	53665	87202	88638	47121
86558	84750	43994	01760	96205	27937	45416	71964	52261	30781
78545	49201	05329	14182	10971	90472	44682	39304	19819	55799
14969	64623	82780	35686	30941	14622	04126	25498	95452	63937
58697	31973	06303	94202	62287	56164	79157	98375	24558	99241
38449	46438	91579	01907	72146	05764	22400	94490	49833	09258
62134	87244	73348	80114	78490	64735	31010	66975	28652	36166
72749	13347	65030	26128	49067	27904	49953	74674	94617	13317
81638	36566	42709	33717	59943	12027	46547	61303	46699	76243
46574	79670	10342	89543	75030	23428	29541	32501	89422	87474
11873	57196	32209	67663	07990	12288	59245	83638	23642	61715
13862	72778	09949	23096	01791	19472	14634	31690	36602	62943
08312	27886	82321	28666	72998	22514	51054	22940	31842	54245
11071	44430	94664	91294	35163	05494	32882	23904	41340	61185
82509	11842	86963	50307	07510	32545	90717	46856	86079	13769
07426	67341	80314	58910	93948	85738	69444	09370	58194	28207
57696	25592	91221	95386	15857	84645	89659	80535	93233	82798
08074	89810	48521	90740	02687	83117	74920	25954	99629	78978
20128	53721	01518	40699	20849	04710	38989	91322	56057	58573
00190	27157	83208	79446	92987	61357	38752	55424	94518	45205
23798	55425	32454	34611	39605	39981	74691	40836	30812	38563
85306	57995	68222	39055	43890	36956	84861	63624	04961	55439
99719	36036	74274	53901	34643	06157	89500	57514	93977	42403
95970	81452	48873	00784	58347	40269	11880	43395	28249	38743
56651	91460	92462	98566	72062	18556	55052	47614	80044	60015
71499	80220	35750	67337	47556	55272	55249	79100	34014	17037
66660	78443	47545	70736	65419	77489	70831	73237	14970	23129
35483	84563	79956	88618	54619	24853	59783	47537	88822	47227
09262	25041	57862	19203	86103	02800	23198	70639	43757	52064

स्रोत : यादृच्छिक संख्याओं की सारणी से http://www.wrs.edu/sungurea/in trostat// public//instruction/ranbox/randomnumbers//.html के माध्यम से उद्धृत।

उपर्युक्त यादृच्छिक संख्या सारणी में 5 अंक वाली 450 यादृच्छिक संख्याएं हैं।

17.4.3 यादृष्छिक संख्या सारणी के प्रयोग संबंधी चरण

यादृष्छिक संख्या सारणी के प्रयोग से सरल यादृष्छिक प्रतिदर्श का चयन करते समय हमें निम्नलिखित चरणों का अनुसरण करना होगा।

1) समष्टि आमाप (N) का निर्धारण
2) प्रतिदर्श आमाप (n) का निर्धारण
3) समष्टि की सभी इकाइयों को सूचीबद्ध करना। संख्याओं को क्रमबद्ध तरीके से लगाना। मान लीजिए समष्टि में 100 इकाइयां हैं, ऐसी स्थिति में 00 से 99 तक को क्रम संख्या में लगाएं।
4) यादृच्छिक संख्या सारणी के पृष्ठ से प्रतिदर्श के चयन के आरंभिक बिंदु का निर्धारण करें और इसके लिए पृष्ठ पर आँखें बन्द करके किसी भी संख्या पर अपनी अंगुली रख दें।

5) दिशा का चयन करें जिसमें आप संख्याओं को पढ़ना चाहते हैं (मान लीजिए बायें से दायें या दायें से बायें और ऊपर से नीचे या नीचे से ऊपर)

6) मान लीजिए आप दो अंकों वाली (00 से 99) संख्याओं को ढूंढ रहे हैं, लेकिन इन संख्याओं को आप सारणी से सीधेतौर पर नहीं पढ़ सकते, क्योंकि ये 5 अंकों वाली संख्या अर्थात 54245 है (अर्थात दी गई सारणी 17.1 में यादृच्छिक संख्या सारणी की 29वीं पंक्ति और 10वें स्तम्भ की संख्या) का चयन आपने किया है। इस स्थिति में दो अंकों वाली संख्या 45 है, यदि आपने संख्या के अंतिम दो अंकों को चुना है।

7) प्रत्येक समष्टि इकाई की दी गई संख्याओं पर ही ध्यान दें । यदि संख्या समष्टि की किसी एक इकाई को दर्शाती है तो वह प्रतिदर्श का भाग बन जाती है। मान लीजिए आप 10 प्रतिदर्श इकाइयों का चयन करना चाहते हैं, तो जिन अन्य संख्याओं का आप चयन करेंगे, वे हैं; 71 (11071), 30 (44430), 64 (94664), 94 (91294), 63 (35163); 82 (32882), 04 (23904), 40 (41340), 85(61185)। गौर कीजिए आपने 94(05494) को छोड़ दिया है क्योंकि इस संख्या का चयन आप पहले ही कर चुके हैं।

8) संख्या का एक बार चयन करने पर, उसे दुबारा न चुनें।

9) अपेक्षित प्रतिदर्श प्राप्त करने से पहले, यदि आप सारणी के अंतिम बिंदु तक पहुँच जाते हैं तो यादृच्छिक संख्या सारणी में अन्य आरंभिक बिंदु को लें और प्रतिदर्श के लिए बाकी की इकाइयों का चयन करें।

उदाहरण 17.2: मान लीजिए यादृच्छिक संख्या सारणी के प्रयोग से कुल 1000 खाताधारकों में से प्रतिदर्श के रूप में आपको 100 खाताधारकों को चुनना है। यहाँ, सर्वप्रथम आप प्रत्येक खाताधारक को 000 से 999 में से कोई एक संख्या देंगे। 100 खाताधारकों का प्रतिदर्श बनाने के लिए आपको 000 से 999 में से तीन अंकों वाली 100 संख्याओं को लेना होगा। दी गई सारणी 17.1 में से किसी भी पंक्ति या स्तम्भ में यादृच्छिक संख्याओं को चुनें। मान लीजिए प्रतिदर्श बनाने के लिए, आरंभिक बिंदु के रूप में आपने चौथी पंक्ति और पहले कॉलम का चयन किया है। पहली अंक संख्या 628 (64628) है, यदि आपने अपने उद्देश्य के लिए, संख्या के रूप में अंतिम 3 अंकों को चुना है। यहाँ, आप संख्या के अंतिम 3 अंकों को पढ़ें। यदि संख्या 000 से 999 के बीच है तो संख्या को प्रतिदर्श में शामिल करें। अन्यथा संख्या को छोड़ दें और चुनी गई दिशा में अगले अंक को पढ़ें। यदि इस संख्या का चयन आप पहले से कर चुके हैं तो इसे छोड़ दीजिए। इस उदाहरण में चूंकि आपने चौथी पंक्ति और पहले स्तम्भ से शुरू किया था और आप बायें से दायें की दिशा में पढ़ रहे थे तो प्रतिदर्श की निम्नलिखित 100 संख्याओं का चयन किया जाएगा।

628	126	254	090	752	091	411	146	089	630
831	113	511	082	140	733	076	292	486	468
583	361	047	792	466	395	635	697	447	405
209	404	457	570	194	043	330	939	865	906
409	830	911	767	248	253	317	120	772	103
836	530	785	210	228	869	332	868	672	358
469	149	509	176	169	954	002	582	249	037
192	221	918	437	571	995	006	694	692	581
050	734	652	577	631	184	295	776	885	796
822	973	822	467	013	072	942	917	129	624

यदि समष्टि में इकाइयों की संख्या काफ़ी अधिक है तो उपर्युक्त दोनों विधियाँ कारगर सिद्ध नहीं होंगी। आजकल कंप्यूटर के प्रयोग से हम सरलता से यादृच्छिक प्रतिदर्श का चयन कर सकते हैं। ऐसे बहुत से कंप्यूटर कार्यक्रम (program) हैं जो यादृच्छिक संख्याओं की पूरी श्रृंखला जनित कर सकते हैं, लेकिन इसके लिए कंप्यूटर में समष्टि की सूचीबद्ध इकाइयों का होना आवश्यक है।

कंप्यूटर जनित यादृच्छिक संख्याओं के प्रयोग से प्रतिदर्श चयन के एक अन्य तरीके का वर्णन, अब हम करेंगे। इस उदाहरण में, मान लीजिए हम एक्सेल स्प्रेडशीट (Excel spreadsheet) में स्तम्भों में खाताधारकों की सूची को उतार सकते हैं। तब, इससे दायें स्तम्भ में हम फलन = RANDO को पेस्ट करते हैं, जिसका अर्थ है कोष्ठों (cells) में 0 और 1 के बीच यादृच्छिक संख्या रखने का एक्सेल का तरीका। इसके बाद, संग्रहित सूची में हम पहले 100 नामों को लेंगे। यह पूरी प्रक्रिया एक मिनट का समय लेगी यदि कंप्यूटर में एक्सेल (EXCEL) कार्यक्रम प्रयोग करने से हम अवगत हों तो।

17.4.4 सरल यादृच्छिक प्रतिचयन के उपादेयता

1) सरल यादृच्छिक प्रतिदर्श में हमें समष्टि इकाइयों के एक जैसे होने (सजातीय) का भलीभांति पता होता है और इसलिए समष्टि की विशेषताओं पर अतिरिक्त जानकारी प्राप्त करने की आवश्यकता नहीं होती।

2) सरल यादृच्छिक प्रतिचयन के प्रयोग से हम निष्पक्ष प्रतिदर्श का चयन कर सकते हैं। ऐसा इसलिए है, क्योंकि इस प्रतिदर्श में समष्टि की प्रत्येक इकाई के शामिल होने की संभावना एक जैसी होती है। इसमें मनुष्य द्वारा पक्षपात करने की संभावना पूरी तरह कम हो जाती है।

3) प्रतिचयन त्रुटि आकलन के माध्यम से हम परिणामों की परिशुद्धता का निर्धारण कर सकते हैं।

4) यदि समष्टि का आकार अधिक बड़ा नहीं है तो सरल यादृच्छिक प्रतिचयन सरल और आसान विधि है जिसे प्रतिचयन क्रियाविधि के रूप में अपनाया जा सकता है।

17.4.5 सरल यादृच्छिक प्रतिदर्श की सीमाएं

1) सरल यादृच्छिक प्रतिचयन की सबसे बड़ी सीमा है कि यदि समष्टि आमाप काफ़ी अधिक हो तो समष्टि की इकाइयों को सूचीबद्ध करने में हमें काफ़ी अधिक समय लगेगा।

2) यदि समष्टि सजातीय हो तो सरल यादृच्छिक क्रियाविधि अति कारगर सिद्ध होगी। मान लीजिए लिंग, आयु, सामाजिक स्थिति जैसी विशेषताओं वाली समष्टि का हमें अध्ययन करना है। तब हमें समष्टि इकाइयों की उन सभी विशेषताओं वाले प्रतिनिधि प्रतिदर्श (Representative Sample) को रखने के लिए बड़े प्रतिदर्श आमाप की आवश्यकता पड़ेगी। इस मुद्दे से निपटने का बेहतर तरीका स्तरित प्रतिचयन की क्रियाविधि को प्रयोग में लाना है और इसकी जानकारी इकाई में आगे दी गई है।

17.5 क्रमबद्ध यादृच्छिक प्रतिदर्श का चयन

क्रमबद्ध यादृच्छिक प्रतिचयन (systematic random sampling) की क्रियाविधि ऐसी है जो न्यूनाधिक सरल यादृच्छिक प्रतिचयन की क्रियाविधि के जैसी है। इस प्रतिचयन की क्रियाविधि में, हमें यादृच्छिक रूप से आरंभिक बिंदु का चयन करना है और इसके बाद

विशिष्ट प्रतिचयन अंतरालों पर समष्टि इकाइयों से क्रमबद्ध तरीक़े से प्रतिदर्श इकाइयों का चयन करना है।

आरंभिक बिंदु और प्रतिचयन अंतराल, अपेक्षित प्रतिदर्श आमाप पर निर्भर करते हैं। प्रतिचयन अंतराल को K के रूप में दर्शाया जाएगा। क्रमबद्ध यादृच्छिक प्रतिचयन की क्रियाविधि के प्रयोग से प्रतिदर्श का चयन करना बेहद आसान है। मान लीजिए, समष्टि में N इकाइयाँ हैं और आपने क्रमबद्ध यादृच्छिक प्रतिचयन की क्रियाविधि के प्रयोग से n इकाइयों के प्रतिदर्श का चयन करने का निर्णय लिया है, तो हमें निम्नलिखित चरणों का अनुसरण करना होगा।

1) 1 से N तक की समष्टि में इकाइयों की संख्या (मान लीजिए आपके पास 1000 इकाइयाँ हैं)

2) आवश्यकतानुसार, प्रतिदर्श आमाप n का चयन करें (मान लीजिए आप 100 इकाइयों का चयन करना चाहते हैं)।

3) प्रतिदर्श आमाप से समष्टि का विभाजन करके प्रतिचयन अंतराल का निर्धारण करें अर्थात $K = N/n =$ अंतराल आमाप (यहाँ K= 1000/100=10)

4) पहले K इकाइयों (1 से K) से यादृच्छिक तरीक़े से इकाई का चयन करें (मान लीजिए पहली प्रतिदर्श इकाई के रूप में आपने इकाई संख्या 5 का चयन किया है)।

5) इसके बाद पिछली इकाई में K को जोड़ते हुए बाद की प्रतिदर्श इकाइयों का चयन करें। बाद के प्रतिदर्श होंगे : $15\,(=5+10), 25\,(=15+10), \ldots, 995\,(=985+10)$.

उदाहरण 17.3: 500 इकाइयों वाली समष्टि से, क्रमबद्ध यादृच्छिक प्रतिचयन की क्रियाविधि के प्रयोग से 60 इकाइयों का प्रतिदर्श बनाइए।

क्रमबद्ध यादृच्छिक प्रतिचयन विधि का प्रयोग करने के लिए, सर्वप्रथम 1 से 500 तक संख्याओं को देते हुए यादृच्छिक रूप में समष्टि इकाइयों को सूचीबद्ध कीजिए। प्रतिचयन अंतराल है K = 500/60=8.3 या 8 (मान लीजिए)। अब पहली 8 समष्टि इकाइयों से यादृच्छिक रूप से पहली प्रतिदर्श इकाई का चयन कीजिए। इसके बाद की चयनित प्रतिदर्श इकाइयाँ हैं: 13, 21, 29, 37..........477। इसलिए, प्रतिदर्श में चयनित समष्टि इकाइयाँ इस प्रकार हैं :

5	13	21	29	37	45	53	61	69	77
85	93	101	109	117	125	133	141	149	157
165	173	181	189	197	205	213	221	229	237
245	253	261	269	277	285	293	301	309	317
325	333	341	349	357	365	373	381	389	397
405	413	429	429	437	445	453	461	469	477

अतः क्रमबद्ध यादृच्छिक प्रतिचयन की क्रियाविधि में, पहली प्रतिदर्श इकाई का चयन यादृच्छिक रूप से किया जाता है और यह प्रतिदर्श इकाई आगे चल कर चुनी जाने वाली बाद की प्रतिदर्श इकाइयों का निर्धारण करती है। लेकिन, समष्टि में इकाइयों का यादृच्छिक अवस्था में होना अनिवार्य है। कुछ मामलों में, हम सरल यादृच्छिक प्रतिचयन

क्रियाविधि की तुलना में हम क्रमबद्ध यादृच्छिक प्रतिचयन की क्रियाविधि को प्रयोग में लाना बेहतर समझते हैं, क्योंकि प्रतिदर्श इकाइयों के चयन में यह अधिक सरल है। जैसे, यदि कहीं आप नारियल के पेड़ों की फ़सल का पता लगाना चाहते हैं तो क्रमबद्ध यादृच्छिक रूप से पेड़ को यादृच्छिक तौर पर लें, अन्य पेड़ों का, प्रतिचयन अंतराल के बराबर की दूरी पर स्वतः चयन हो जाएगा।

17.5.1 क्रमबद्ध यादृच्छिक प्रतिचयन के उपादेयता

क) इस विधि के प्रयोग का मुख्य फ़ायदा है कि सरल यादृच्छिक प्रतिचयन क्रियाविधि की तुलना में इस विधि के प्रयोग से समय कम लगता है और मेहनत भी कम करनी पड़ती है। चुनावों में वोटरों की प्रवृत्ति और विपणन शोधकार्य में उपभोक्ताओं की राय और विचार लेने के लिए इस विधि का भरपूर प्रयोग किया जाता है।

ख) इस क्रियाविधि के प्रयोग का अन्य फ़ायदा है कि इस विधि का प्रयोग समष्टि इकाइयों की कोई औपचारिक सूची उपलब्ध न हो, तब भी किया जा सकता है। जैसे, मान लीजिए हम बैंक द्वारा प्रदत्त सेवाओं को बेहतर बनाने पर उपभोक्ताओं की राय जानना चाहते हैं, तब हम उस बैंक शाखा में जाने वाले हर K खाताधारक का आसानी से चयन कर सकते हैं बशर्ते हमें उस बैंक के खाताधारकों की कुल संख्या का पता हो; (1) जैसे, समष्टि में 2000 खाताधारक हैं और प्रतिदर्श आमाप के रूप में हमें 200 खाताधारकों की आवश्यकता है। तब, $K = 2000/200 = 10$। हम बैंक जाने वाले प्रत्येक 10वें खाताधारक का चयन करेंगे।

17.5.2 क्रमबद्ध यादृच्छिक प्रतिचयन की सीमाएं

क) क्रमबद्ध यादृच्छिक प्रतिचयन क्रियाविधि का मुख्य दोष है कि समष्टि की इकाइयाँ यदि अलग-अलग समय में उभरती हैं तो क्रमबद्ध प्रतिचयन क्रियाविधि का प्रयोग अत्यंत प्रतिनिधित्वहीन प्रतिदर्श प्रदान करता है। जैसे, मान लीजिए आप अपने इलाके के एक स्टोर पर आने वाले उपभोक्ताओं के विचार/राय जानने के इच्छुक हैं। आप उस स्टोर पर आने वाले सभी उपभोक्ताओं को उस स्टोर पर आने की तारीख़ के अनुसार क्रमबद्ध कीजिए और हर महीने की पहली तारीख़ को उस स्टोर पर आने वाले उपभोक्ताओं को प्रतिदर्श के रूप में चुनें। आप जानते हैं कि हरेक माह का पहला दिन, पूरे महीने का प्रतिनिधित्व नहीं कर सकता।

ख) क्रमबद्ध यादृच्छिक प्रतिचयन क्रियाविधि का अन्य दोष है कि समष्टि की प्रत्येक इकाई की चुने जाने की एक जैसी संभावना नहीं होती। इसके बजाए, प्रतिदर्श में समष्टि इकाइयों का चयन, पहली चुनिंदा इकाई पर निर्भर करता है। इस ओर ध्यान दिए बिना कि प्रतिदर्श की पहली इकाई का चयन हम किस प्रकार करते हैं, बाद की इकाइयों का निर्धारण स्वतः हो जाता है। इस विधि से सभी कुछ क्रमबद्ध ही रहता है।

17.6 स्तरित यादृच्छिक प्रतिदर्श का चयन

कुछ मामलों में समष्टि समरूप नहीं होती, अर्थात् जिसका सर्वेक्षण हम करना चाहते हैं, उसकी इकाइयाँ की विशेषताएं एक जैसी नहीं होती। इसमें देखने योग्य बातें, समष्टि की इकाइयों से जुड़ी हो सकती हैं जैसे, महिला/पुरुष, ग्रामीण/शहरी, साक्षर/निरक्षर, उच्च

उच्च आय/निम्न आय समूह आदि। ऐसी स्थितियों में जहाँ इकाइयों में भारी विविधता है, वहाँ सरल यादृच्छिक प्रतिचयन की क्रियाविधि या क्रमबद्ध यादृच्छिक प्रतिचयन की क्रियाविधि हमें प्रतिनिधि प्रतिदर्श प्रदान नहीं करेंगी। ऐसी स्थिति में हम स्तरित यादृच्छिक प्रतिचयन (stratified random sampling) के प्रयोग से ही प्रतिनिधि प्रतिदर्श प्राप्त कर सकते हैं।

स्तरित प्रतिचयन में, हम समष्टि को ऐसे तरीके से विभिन्न स्तरों में बांट देते हैं कि प्रत्येक स्तर की इकाइयाँ समरूप हों। ये स्तर एक-दूसरे से अलग होंगे। मान लीजिए स्त्री/पुरुष वितरण के आधार पर हम समष्टि को स्तरों में बाँटना चाहते हैं। ऐसी स्थिति में हम स्त्री या पुरुष के अनुसार अलग अलग रूप से समष्टि इकाइयों को सूचीबद्ध करेंगे। इसके बाद, हम प्रतिदर्श आमाप तय करेंगे जिसके लिए प्रत्येक स्तर से प्रतिदर्श हम प्राप्त करेंगे। प्रत्येक स्तर से प्रतिदर्श आमाप तय करने के, दो उपागम हैं; (क) समानुपातिक (proportional) स्तरित प्रतिदर्श, और (ख) असमानुपातिक (disproportional) स्तरित प्रतिदर्श। इन दोनों क्रियाविधियों की चर्चा हम आगे कर रहे हैं।

17.6.1 समानुपातिक स्तरित प्रतिचयन

जब हम विविध स्तरों वाली समष्टि से प्रतिदर्श लेते हैं तो प्रत्येक स्तर से हमें प्रतिदर्श लेने की आवश्यकता पड़ती है। ऐसे प्रतिदर्श कुल समष्टि आमाप से स्तरित समष्टि आमाप के समानुपात में हो सकते हैं। मान लीजिए हम समष्टि (N) को K गैर अतिव्याप्त स्तरों N_1, N_2, N_3......... अर्थात $N_1 + N_2 + N_3 + ... + N_k = N$ में बांट देते हैं। हम n माप का प्रतिदर्श बनाना चाहते हैं। तब विविध स्तरों के प्रतिदर्श समानुपातों को निम्नलिखित द्वारा दर्शाया जाता है:

$$\frac{n_1}{N_1} = \frac{n_2}{N_2} = \frac{n_3}{N_3} = ... \frac{n_k}{N_k}$$

उदाहरण 17.4 : मान लीजिए 1000 इकाइयों वाली समष्टि से 200 इकाइयों का प्रतिदर्श आप बनाना चाहते हैं। समष्टि में उच्च/ निम्न आय और ग्रामीण/शहरी जैसी प्रकृति वाले विभिन्न समूह शामिल हैं। ऐसे स्तरित समष्टि आमाप इस प्रकार है :

उच्च आय–शहरी	=	200
निम्न आय–शहरी	=	400
उच्च आय–ग्रामीण	=	100
निम्न आय–ग्रामीण	=	300

प्रतिनिधि प्रतिदर्श की प्राप्ति के लिए, प्रतिदर्श के प्रत्येक स्तर को समष्टि में तदनुरूप स्तर को दर्शाना होगा। इसके लिए हमें स्तर आमाप के आधार पर प्रत्येक स्तर से विभिन्न प्रतिदर्श आमाप लेने चाहिए। प्रत्येक स्तर में निर्णायक गुणक उतना ही है जितना कि समष्टि के कुल प्रतिदर्श का समानुपात। हमारे उदाहरण में, 200 इकाइयों के प्रतिदर्श की प्राप्ति में, प्रत्येक स्तर में समष्टि के लिए प्रतिदर्श का समानुपात है;

$$\frac{n}{N} = \frac{200}{1000} = 0.2$$

ध्यान दीजिए कि प्रत्येक स्तर के लिए हम एक जैसे ही अनुपात पर विचार कर रहे हैं। तब, प्रत्येक स्तर से प्राप्त प्रतिदर्श इस प्रकार होगा;

स्तर श्रेणी (1)	स्तर समष्टि आमाप (N_i) (2)	प्रतिदर्श (समष्टि अनुपात संबंधी) (3)	स्तर प्रतिदर्श (4) = (2)×(3)
उच्च आय–शहरी	200	0.2	40
निम्न आय–शहरी	400	0.2	80
उच्च आय–ग्रामीण	100	0.2	20
निम्न आय–ग्रामीण	300	0.2	60
कुल	1000	0.2	200

सरल यादृच्छिक प्रतिचयन की तुलना में स्तरित प्रतिचयन के कई अधिक फ़ायदे हैं। स्तरित प्रतिचयन न केवल समग्र समष्टि बल्कि प्रत्येक स्तर के प्रतिदर्श प्रतिनिधि को भी सुनिश्चित करता है। जहां स्तर का आकार छोटा है, वहां इस बात का विशेष महत्व है। इसके अलावा स्तरित प्रतिचयन के सामान्यतौर पर सरल यादृच्छिक प्रतिचयन की तुलना में, जो आंकड़ें प्राप्त होते हैं वे अधिक सार्थक होते हैं।

17.6.2 असमानुपातिक स्तरित प्रतिचयन

समानुपातिक स्तरित प्रतिचयन में, हमने माना था कि समष्टि का प्रत्येक स्तर एक जैसा है। जिसके परिणामस्वरूप हम अपेक्षा करते हैं कि समग्र रूप से समष्टि संबंधी परिवर्तनशीलता से स्तर के भीतर की परिवर्तनशीलता निम्न है। दूसरी तरफ़, यदि प्रत्येक स्तर के भीतर परिवर्तनशीलता कम नहीं हो तो हम असमानुपातिक स्तरित प्रतिचयन क्रियाविधि का प्रयोग करते हैं। ऐसी क्रियाविधि में स्तर आबंटन, आकार और परिवर्तनशीलता (अर्थात विचाराधीन विशेषताओं के मानक विचलन) पर आधारित होता है। इस क्रियाविधि को कभी-कभार द्विशः तोलन योजना (double weighing scheme) भी कहते हैं और दिए गए प्रतिदर्श आमाप के लिए यह सर्वाधिक उपयुक्त प्रतिदर्श और विश्वसनीय आकलन भी प्रदान करती है। इसके लिए, हां हमें केवल प्रत्येक स्तर के भीतर विचाराधीन विशेषता के मानक विचलन का ज्ञान/आकलन प्राप्य होना चाहिए।

असमानुपातिक स्तरित प्रतिचयन के प्रयोग के लिए, निम्नलिखित चरणों का अनुसरण करें।

1) अपने द्वारा चयनित किसी विशेषता/गुण को ध्यान में रखकर, समष्टि स्तरों का विभाजन करें (जैसे ग्रामीण/शहरी, महिला/पुरुष आदि)

2) प्रत्येक स्तर से ली गई इकाइयों की संख्या सीधे तौरपर स्तर n_i के सापेक्षिक आकार और विचाराधीन विशेषता के मानक विचलन s_i के प्रतिलोम में समानुपातिक है। मान लीजिए s_1, s_2, s_3............s_k, k स्तर के मानक विचलन है और $P_1, P_2, P_3, \ldots\ldots P_k$ कुल समष्टि के स्तरित अनुपात हैं और $n\,(= n_1 + n_2 + \ldots\ldots\ldots + n_2)$ अपेक्षित प्रतिदर्श आमाप है। तब असमानुपातिक स्तरित प्रतिचयन की क्रियाविधि का प्रयोग करते हुए स्तरित प्रतिदर्श आमाप होगा।

$$n_i = \frac{P_i \times s_i \times n}{\sum P_i s_i}$$

3) सरल यादृच्छिक प्रतिचयन या क्रमबद्ध यादृच्छिक प्रतिचयन अर्थात दोनों में से किसी एक के प्रयोग से प्रत्येक स्तर से प्रतिदर्श का चयन करें।

आइए उदाहरण 17.4 पर दुबारा से नज़र डालें, जहां हमने समष्टि को 4 स्तरों में विभाजित किया है। हम देखते हैं कि उच्च आय स्तर में कई परिवार हैं और निम्न आय स्तर में परिवारों की बड़ी संख्या शामिल है। मान लीजिए कि उच्च आय समूहों में आय परिवर्तनशीलता निम्न आय समूहों की परिवर्तनशीलता से उच्च है। इसलिए प्रतिदर्श में उच्च आय समूहों के अवप्रतिनिधित्व (under-representative) को अनदेखा करने के लिए, प्रत्येक स्तर में एक असमानुपातिक प्रतिदर्श को शामिल किया गया है। इसका अर्थ है, यदि स्तर के भीतर परिवर्तनशीलता उच्च है तो आकलनों की सटीकता को बढ़ाने के लिए, हमारे पास उस स्तर का बड़ा प्रतिदर्श आमाप अवश्य होना चाहिए। इसी तरह, यदि स्तर के भीतर परिवर्तनशीलता निम्न है तो उस स्तर का छोटा प्रतिदर्श आमाप ही पर्याप्त रहेगा। अर्थात, उच्च स्तर प्रसरण के लिए उच्च स्तर प्रतिदर्श आमाप और निम्न स्तर प्रसरण के लिए निम्न प्रतिदर्श आमाप। यह इस तथ्य के अलावा है कि बड़े स्तरित आमाप के लिए बड़ा प्रतिदर्श आमाप चाहिए।

उदाहरण 17.5: उदाहरण 17.4 पर दुबारा विचार कीजिए। मान लीजिए स्तर प्रसरण (Stratum Variances) इस प्रकार हैं, जैसा कि नीचे दिए गए हैं:

स्तर	**प्रसरण** (s^2)
उच्च आय शहरी	6.5
निम्न आय शहरी	2.5
उच्च आय ग्रामीण	4.5
निम्न आय ग्रामीण	2.0

असमानुपातिक स्तरित प्रतिचयन की क्रियाविधि का प्रयोग करें और प्रत्येक स्तर से प्रतिदर्श आमाप का चयन करें। कुल प्रतिदर्श आमाप 200 है।

इस उदाहरण के लिए, प्रत्येक स्तर के लिए असमानुपातिक प्रतिदर्श आमाप इस प्रकार है;

स्तर	**स्तर समष्टि**	**स्तर समष्टि अनुपात (P_i)**	**स्तर प्रसरण (s_i^2)**	**स्तर मानक विचलन ($s_i = \sqrt{s_i^2}$)**	$P_i \times s_i$	**प्रतिदर्श आमाप** $\frac{P_i \times s_i \times n}{\Sigma P_i s_i}$
उच्च आय–शहरी	200	0.20	6.5	2.5	0.50	56
निम्न आय–शहरी	400	0.40	2.5	1.6	0.64	72
उच्च आय–ग्रामीण	100	0.10	4.5	2.1	0.21	24
निम्न आय–ग्रामीण	300	0.30	2.0	1.4	0.42	47
कुल	**1000**				**1.77**	**200**

17.6.3 स्तरित प्रतिचयन के उपादेयता

क) स्तरित यादृच्छिक प्रतिचयन में, समष्टि के प्रत्येक स्तर से प्रतिदर्श निकाले जाते हैं। इसलिए स्तरित यादृच्छिक प्रतिचयन की क्रियाविधि अधिक दर्शाने योग्य है।

ख) सरल यादृच्छिक प्रतिचयन की तुलना में स्तरित यादृच्छिक प्रतिचयन की क्रियाविधि अधिक सटीक है। इसलिए बड़ी सीमा तक, यह क्रियाविधि प्रतिदर्श चयन संबंधी पक्षपात को दूर करती है।

ग) जैसाकि हमने सरल यादृच्छिक प्रतिचयन और क्रमबद्ध यादृच्छिक प्रतिचयन क्रियाविधियों में देखा, जहां समष्टि में विविधता है, वहां उपयुक्त प्रतिदर्श की प्राप्ति के लिए हमें बड़े प्रतिदर्श आमाप की आवश्यकता है। लेकिन, स्तरित यादृच्छिक प्रतिचयन से हम छोटे आकार के प्रतिदर्श प्राप्त कर सकते हैं। इससे आंकड़ों को इकठ्ठा करने में समय, धन और अन्य संसाधनों की बचत होती है।

17.6.4 स्तरित प्रतिचयन की सीमाएं

क) स्तरित यादृच्छिक प्रतिचयन की क्रियाविधि का मुख्य दोष है कि हमें समष्टि में विशेषताओं के वितरण की विस्तृत जानकारी प्राप्त करने की आवश्यकता होती। यदि हम सही तरीके से सजातीय समूहों की पहचान नहीं कर सकते तो बेहतर होगा कि हम सरल यादृच्छिक प्रतिदर्श का प्रयोग करें, क्योंकि गलत स्तरीकरण से गंभीर त्रुटियां उत्पन्न हो सकती है।

ख) स्तरित यादृच्छिक प्रतिचयन का अन्य दोष है कि हमें प्रत्येक स्तर के लिए, अलग से, समष्टि संबंधी इकाइयों की सूची बनाने की आवश्यकता है। प्रत्येक विशेषता के लिए यदि समष्टि की इकाइयों की सूची तत्काल रूप से उपलब्ध न हो तो सूची बनाना और भी अधिक कठिन कार्य बन जाता है।

17.7 गुच्छ प्रतिदर्श का चयन

अक्सर समष्टि इकाइयों का फैलाव, विशाल भौगोलिक क्षेत्र तक होता है। ऐसे मामले में, सरल यादृच्छिक प्रतिचयन के माध्यम से आंकड़ों को इकट्ठा करने में, समय, धन और जनशक्ति की अधिक ज़रूरत पड़ती है, क्योंकि चुनिंदा इकाइयों पर आंकड़ों को इकठ्ठा करने के लिए समग्र भौगोलिक क्षेत्र को तय करना पड़ता है। मान लीजिए आपको व्यक्तिगत इंटरव्यू लेने के लिए समूचे उत्तर प्रदेश में व्याप्त उत्तरदाताओं का प्रतिदर्श लेना है। चूंकि उत्तरदाताओं का फैलाव पूरे राज्य में है, इसलिए सरल यादृच्छिक प्रतिदर्श के प्रयोग में, आपको उत्तरदाताओं से मिलने के लिए यात्रा करनी पड़ेगी और इसमें खर्चा भी काफी होगा। ऐसी स्थिति में गुच्छ प्रतिचयन (cluster sampling) की क्रियाविधि अधिक उपयोगी सिद्ध होगी।

गुच्छ प्रतिचयन के बुनियादी सिद्धांत इस प्रकार हैं:

i) गुच्छ के भीतर मौजूद अंतर या परिवर्तनशीलता (variability) यथासंभव अधिक होनी चाहिए। जितना संभव हो, प्रत्येक गुच्छ के भीतर परिवर्तनशीलता उतनी होनी चाहिए जितनी सकल समष्टि में है।

ii) गुच्छों के बीच की परिवर्तनशीलता यथासंभव कम होनी चाहिए। गुच्छों का चयन एक बार हो जाए तो चयनित गुच्छों की सभी इकाइयों को आकड़ों की प्राप्ति के लिए, प्रतिदर्श में शामिल किया जाता है।

गुच्छ प्रतिचयन में हम समष्टि को समूहों में बांट देते हैं, जिन्हें गुच्छ (cluster) कहते हैं। तब हम सरल यादृच्छिक प्रतिचयन विधि के प्रयोग से प्रतिदर्श गुच्छ का चयन करते हैं। प्रत्येक गुच्छ की 'समष्टि' इकाइयों को उतना ही विषमतापूर्ण माना गया है जितनी विषमता सकल समष्टि में पाई जाती है। इसका अर्थ है प्रत्येक गुच्छ स्वयं समष्टि का प्रतिनिधि है।

17.7.1 गुच्छ प्रतिचयन के चरण

गुच्छ प्रतिचयन में, हम निम्नलिखित चरणों का अनुसरण करते हैं:

1) समष्टि को अलग-अलग गुच्छों में बांटें।
2) अपने प्रतिदर्श की आवश्यकतानुसार गुच्छों की संख्या का निर्धारण करें।
3) गुच्छों के प्रतिदर्श का चयन यादृच्छिक रूप से करें।
4) प्रतिदर्श गुच्छों के भीतर सभी इकाइयों का सर्वेक्षण करें।

मानलीजिए, गुच्छों का बंटवारा, समष्टि की भौगोलिक सीमाओं पर आधारित है, ऐसी स्थिति में इसे *क्षेत्र प्रतिचयन* (area sampling) कहते हैं। आपने देखा होगा कि गुच्छ प्रतिचयन के मामले में, यादृच्छिक प्रतिचयन विधि के प्रयोग से, गुच्छों का चयन किया जाता है। तत्पश्चात् प्रत्येक प्रतिदर्श गुच्छ के भीतर की सभी समष्टि इकाइयों को प्रतिदर्श में शामिल किया जाता है। मान लीजिए प्रत्येक चुनिंदा गुच्छ की सभी इकाइयों को शामिल करने के बजाए, आपने प्रत्येक गुच्छ के भीतर केवल कुछ इकाइयों के प्रतिदर्श को शामिल करने का निर्णय लिया तब आप भलीभांति समझ सकते हैं कि यहां दो चरण हैं।

पहले चरण में आप गुच्छों को प्रतिदर्श का रूप देते हैं और दूसरे चरण में आप प्रत्येक प्रतिदर्श गुच्छ के भीतर प्रतिदर्श इकाई का चयन करते हैं, इस प्रतिचयन की क्रियाविधि को *द्वि-चरणी प्रतिचयन* कहते हैं। यहां, गुच्छों को प्राथमिक इकाइयां और प्रतिदर्श इकाइयों के भीतर की इकाइयां 'द्वितीयक इकाइयां' कहलाती हैं।

उदाहरण 17.6: मान लीजिए आप उत्तर प्रदेश राज्य के किसी बैंक के एटीएम प्रयोगकर्ताओं की राय जानना चाहते हैं। हम राज्य को 30 गुच्छों (अर्थात हम जिले को इकाई के रूप में मान कर, एक या दो जिलों को एक गुच्छ) में परिवर्तित कर सकते हैं। यहां, मान लीजिए, इन गुच्छों में से प्रत्येक समग्र रूप से उत्तर प्रदेश के एटीएम प्रयोगकर्ताओं की राय को दर्शाएंगें। इसके बाद हम गुच्छों के प्रतिदर्श का चयन करेंगे और प्रत्येक गुच्छ में से सभी एटीएम प्रयोगकर्ताओं की राय प्राप्त करेंगे।

17.7.2 गुच्छ प्रतिचयन के उपादेयता

क) गुच्छ प्रतिचयन का मुख्य फायदा है कि इससे यात्रा में लगने वाला समय कम लगेगा और इस संदर्भ में आंकड़ों को इकठ्ठा करने में कम खर्चा होगा।

ख) शोधकर्त्ता को सभी गुच्छों पर प्रकाश डालने की आवश्यकता नहीं रहती, इसलिए केवल प्रतिदर्श गुच्छ पर ही ध्यान केंद्रित किया जाएगा। यह अधिक व्यावहारिक विधि है जो क्षेत्रकार्य को अधिक सरल बनाती है।

17.7.3 गुच्छ प्रतिचयन की सीमाएं

क) गुच्छ प्रतिचयन में हम मान लेते हैं कि प्रत्येक गुच्छ, सभी गुच्छों की समष्टि इकाइयों की विसमता को दर्शाती है। लेकिन यह कल्पना कई मामलों में सही साबित नहीं होती क्योंकि बहुधा प्रवृत्ति है कि समग्र समष्टि की इकाइयों की तुलना में गुच्छों की

इकाइयां अधिक समरूप होती हैं। इसका अर्थ है कि विसम गुच्छों को गठित करना कठिन कार्य है।

ख) यादृच्छिक प्रतिचयन और स्तरित प्रतिचयन की तुलना में, दिए गए प्रतिदर्श आमाप के लिए, गुच्छ प्रतिचयन की प्रतिचयन क्षमता निम्न होती है। यह केवल कम खर्चीली विधि है, लेकिन सांख्यिकीय दृष्टिकोण से अधिक कुशल विधि नहीं है।

17.8 बहुचरणी प्रतिचयन

गुच्छ प्रतिचयन में हमने देखा कि जब हम प्रत्येक गुच्छ से सभी इकाइयों को लेने के बजाए, प्रतिदर्श का चयन करते हैं तो हम इसे द्वि-चरणी प्रतिचयन कहते हैं। बहुचरणी प्रतिचयन (multistage sampling), द्वि-चरणी प्रतिचयन का विस्तार है।

अभी तक हमने चार विधियों का अध्ययन किया, जो हैं : (क) सरल यादृच्छिक प्रतिचयन, (ख) क्रमबद्ध यादृच्छिक प्रतिचयन, (ग) स्तरित प्रतिचयन, और (घ) गुच्छ प्रतिचयन। ये सरलतम प्रायिकता (या यादृच्छिक) प्रतिचयन की क्रियाविधियां हैं। लेकिन, असल–जीवन या सामाजिक विज्ञान शोध में हम जिन प्रतिचयन की विधियों का प्रयोग करते हैं, वे उपर्युक्त चार विधियों की तुलना में अधिक जटिल हैं। बहुचरणी प्रतिचयन का बुनियादी सिद्धांत है कि अपनी प्रतिचयन संबंधी आवश्यकताओं को पूरा करने के लिए, विविध उपयोगी तरीकों से हम इन सरल विधियों को एक दूसरे से जोड़ सकते हैं। जब हम उपर्युक्त प्रतिचयन की दो या अधिक विधियों को एक दूसरे से जोड़ देते हैं तो ऐसे प्रतिचयन को हम बहुचरणी प्रतिचयन कहते हैं।

उदाहरण 17.7: मान लीजिए आप हरियाणा में किसी स्कूल के बच्चों से बातचीत करना चाहते हैं ताकि अभिभावकों की सामाजिक–अर्थिक पृष्ठभूमि के आधार पर स्कूलों को श्रेणीबद्ध किया जा सके। इसके लिए, आपको पहले चरण में, गुच्छ प्रतिचयन विधि लागू करनी होगी। हमें हरियाणा को विविध गुच्छ या ज़िलों में बांटना होगा। इसके बाद, सरल यादृच्छिक प्रतिचयन विधि का प्रयोग करते हुए हम ज़िलों (गुच्छों) के प्रतिदर्श का चयन करेंगे। दूसरे चरण में हम स्तरित प्रतिचयन विधि का प्रयोग करते हुए स्कूलों को बांटेंगे। इस संदर्भ में स्तर से आशय है; सरकारी स्कूल, सरकारी सहायता प्राप्त स्कूल, केंद्रीय विद्यालय और पब्लिक स्कूल। हम सरल यादृच्छिक प्रतिचयन या क्रमबद्ध यादृच्छिक प्रतिचयन अर्थात दोनों में से कोई भी एक विधि का प्रयोग करते हुए प्रत्येक स्तर में स्कूलों के प्रतिदर्श का चयन करेंगे। तीसरे चरण में, हम दुबारा सरल यादृच्छिक प्रतिचयन या क्रमबद्ध यादृच्छिक प्रतिचयन विधि का प्रयोग करते हुए प्रत्येक प्रतिदर्शी कक्षा से प्रतिदर्श-विद्यार्थियों का चयन करेंगे।

बहुचरणी प्रतिचयन में प्रतिनिधि प्रतिदर्श की प्राप्ति के लिए, विविध चरणों के प्रयोग पर विचार करना संभव है। प्रत्येक चरण में प्रतिचयन की उपयोगी विधि का प्रयोग किया जाता है। प्रत्येक चरण के अंत में प्रतिदर्श का आकार कम हो जाता है।

उपादेयता

क) बहुचरणी प्रतिचयन की क्रियाविधि, आंकड़ा संग्रहण संबंधी लागतों में कटौति द्वारा खर्चा बचाती हैं।

ख) बहुचरणी प्रतिचयन विधि अधिक लचीली हैं और प्रतिचयन के विविध चरणों में प्रतिचयन की विविध क्रियाविधियों के प्रयोग की अनुमति प्रदान करती है।

ग) यदि समष्टि का फैलाव, विस्तृत भौगोलिक क्षेत्र में है तो बहुचरणी प्रतिचयन, सर्वाधिक उपयुक्त प्रतिचयन की विधि है।

सीमाएं

यदि विविध चरणों पर चयनित प्रतिचयन की इकाइयां प्रतिनिधि नहीं है तो बहुचरणी प्रतिचयन विधि अपेक्षाकृत कम सटीक और कम कारगर सिद्ध होती है।

बोध प्रश्न 1

1) निम्नलिखित में से कौन सी समष्टि प्रतिदर्श चयन करने की क्रियाविधि है?

   क) यादृच्छिक प्रतिचयन

   ख) गैर-यादृच्छिक प्रतिचयन

   ग) स्तरित प्रतिचयन

   घ) उपर्युक्त सभी

2) मान लीजिए आप समष्टि के लिए स्तरित यादृच्छिक प्रतिचयन की क्रियाविधि लागू कर रहे हैं, तो आप प्रतिदर्श का चयन किस प्रकार करेंगे?

   क) प्रत्येक स्तर से यादृच्छिक रूप से इकाइयों की समान संख्या का चयन

   ख) प्रत्येक स्तर से समान संख्या में इकाइयों को लेना और निष्कर्षों की तुलना करना।

   ग) समष्टि के अनुपात में, प्रत्येक स्तर से यादृच्छिक रूप से प्रतिदर्श का चयन

   घ) ख और ग

   च) क और ग

3) बताइए, निम्नलिखित कथनों में से कौन से सही हैं और कौन से गलत;

   क) प्रतिचयन की ऐसी क्रियाविधि जो एकसमान अंतरालों पर समष्टि से इकाइयों का चयन करती है, सरल यादृच्छिक प्रतिचयन कहलाती है।

   ख) प्रतिचयन की ऐसी क्रियाविधि जो समष्टि को ऐसे सुपरिभाषित समूहों में विभाजित करती है जिससे यादृच्छिक प्रतिदर्शों की प्राप्ति होती है, स्तरित प्रतिचयन कहलाती है।

4) किसी समष्टि में ऐसे विभिन्न समूह हैं जहां समूह एक दूसरे से काफी भिन्न हैं लेकिन प्रत्येक समूह के भीतर अधिक भिन्नता नहीं है। ऐसी स्थिति में प्रयोग हेतु, सर्वाधिक उपयुक्त प्रतिचयन की क्रियाविधि है;

   क) गुच्छ प्रतिचयन

   ख) क्रमबद्ध प्रतिचयन

   ग) स्तरित प्रतिचयन

   घ) बहुचरणी प्रतिचयन

17.9 गैर-प्रायिकता प्रतिचयन की क्रियाविधियां

गैर-प्रायिकता प्रतिचयन के विविध प्रकार हैं, जैसे;

1) सुविधाजनक प्रतिचयन (convenience sampling)
2) ऐच्छिक प्रतिचयन (judgement sampling)
3) कोटा/नियतमात्रात्मक प्रतिचयन (quota sampling)
4) तुषारपिंडीय प्रतिचयन (snowball sampling)

आइए गैर-प्रायिकता प्रतिचयन प्राप्ति की क्रियाविधि की चर्चा करें।

17.9.1 सुविधाजनक प्रतिचयन

यह ग़ैर-प्रायिकता प्रतिचयन की सर्वाधिक सामान्य रूप से प्रयुक्त विधियों में से एक है। इस विधि में शोधकर्ता की सहूलियत, प्रतिदर्श के चयन का आधार बनती है। विशेषरूप से सर्वेक्षण संबंधी शोध के लिए, आंकड़े इकठ्ठे करने पर विशेष ज़ोर दिया जाता है। ऐसी स्थितियों में प्रतिचयन संबंधी इकाइयों के चयन का ज़िम्मा साक्षात्कर्ता पर होता है। समष्टि इकाइयों को सहजता से प्रतिदर्श में शामिल किया जाता है क्योंकि वे सही समय में सही जगह पर मिल जाती हैं। बहुधा इस विधि का प्रयोग प्रारंभिक शोध-प्रयासों के दौरान किया जाता है ताकि प्रतिदर्श चयन के लिए बिना समय और धन की बर्बादी के परिणामों का कुल अनुमान प्राप्त किया जा सकें। उदाहरण के तौर पर बजट सत्र के दौरान या जब किसी वस्तु की क़ीमत बढ़ जाती है या जब नयी सरकार बनती है, तो आम जनता की राय जानने के लिए शोधकर्ता/पत्रकार सुविधाजनक प्रतिदर्शों का प्रयोग करते हैं। विपणन संबंधी शोध कार्यों में इनका विस्तृत प्रयोग किया जाता है।

सुविधाजनक प्रतिचयन का लाभ है कि यह कम खर्चीली विधि है और इसमें समय भी कम लगता है। इस विधि की सीमाएं हैं: क) इसमें प्रतिदर्श चयन एक तरफा हो सकता है, और ख) इससे समष्टि का प्रतिनिधि प्रतिदर्श प्राप्त नहीं होता और इसलिए हम परिणामों को व्यावहारिक रूप नहीं दे सकते।

17.9.2 ऐच्छिक प्रतिचयन

ग़ैर-प्रायिकता प्रतिचयन की यह ऐसी क्रियाविधि है जो सामान्य रूप से प्रयोग में लाई जाती है। इस क्रियाविधि को अक्सर **उद्देश्यपूर्ण प्रतिचयन** के रूप में जाना जाता है। इस क्रियाविधि में शोधकर्ता अपने निर्णय/विवेक के आधार पर प्रतिदर्श का चयन करता है। शोधकर्ता का मानना है कि चुनिंदा प्रतिदर्श अवयव, समष्टि के प्रतिनिधि हैं। जैसे, उपभोक्ता कीमत सूचकांक की गणना, ऐच्छिक प्रतिचयन पर आधारित होती है। इस विधि में प्रतिदर्श में उपभोक्ता वस्तुओं की टोकरी और अन्य वस्तुओं और सेवाएं शामिल हैं जो प्रतिनिधि प्रतिदर्श को प्रतिबिंबित कर सकती हैं। इन वस्तुओं की क़ीमत, ऐसे चुनिंदा शहरों से ली जाती है जिन्हें ऐसे विशिष्ट शहरों का दर्जा दिया गया है जहां की जनसांख्यिकीय रूपरेखा राष्ट्रीय रूपरेखा से मेल खाती है।

ऐच्छिक प्रतिचयन का लाभ है कि यह सस्ती, सरल और आसानी से अपनाई जाने वाली विधि है। इसका दोष है कि यह समष्टि के प्रत्यक्ष सामान्यीकरण की अनुमति नहीं देती। प्रतिदर्श की गुणवत्ता शोधकर्ता के विवेक पर निर्भर करती है।

17.9.3 कोटा/नियतमात्रात्मक प्रतिचयन

इस क्रियाविधि में, समष्टि को लिंग, आयु शिक्षा, धर्म, आय-समूह आदि जैसी कुछ विशेषताओं के आधार पर समूहों में बांट दिया जाता है। प्रत्येक समूह से इकाइयों के कोटे का निर्धारण किया जाता है। यह कोटा समानुपातिक या गैर-समानुपातिक हो सकता है। समानुपातिक कोटा प्रतिचयन, समष्टि की प्रत्येक विशेषता के समानुपात पर आधारित होता है क्योंकि प्रतिदर्श समानुपात, समष्टि समानुपात को दर्शाता है। जैसे, यदि आप जानते हैं कि शहर में ऐसे 80 प्रतिशत परिवार हैं जिनकी आय प्रतिवर्ष 1,00,000 रुपए से कम है और 20 प्रतिशत परिवार हैं जिनकी प्रतिवर्ष आय 10,00,000 से ऊपर है। यहां हमारा उद्देश्य, समष्टि की प्रत्येक विशेषता के आधार पर प्रतिचयन के समानुपातिक कोटे की पूर्ति करना है।

गैर-समानुपातिक कोटा प्रतिचयन विधि अपेक्षाकृत कम प्रतिबंधी है। इस क्रियाविधि में आप प्रत्येक समूह से प्रतिदर्शी इकाइयों की न्यनूतम संख्या का विशेष रूप से उल्लेख कर सकते हैं। इसमें आपका समष्टि से संबंधित अनुपातों से कोई सरोकार नहीं है। जैसे, उपर्युक्त उदाहरण में, आप 80 प्रतिशत और 20 प्रतिशत की बजाए, प्रत्येक आय समूह से केवल 50 परिवारों का इंटरव्यू ले सकते हैं। साक्षात्कर्ता को अपने विवेक या निर्णय के आधार पर प्रत्येक समूह के लिए कोटा पूरा करने की हिदायत दी जाती है। कोटा प्रतिचयन का मुख्य उद्देश्य, समष्टि में विविध समूहों को उस सीमा तक दर्शाना है, जहां तक जांचकर्ता दर्शाना चाहता है।

कोटा प्रतिचयन को स्तरित प्रतिचयन (जिसका अध्ययन आपने पहले किया था) की भांति न समझें। स्तरित प्रतिचयन में आप प्रत्येक स्तर या समूह से यादृच्छिक प्रतिदर्शों का चयन करते हैं जबकि कोटा प्रतिचयन में साक्षात्कर्ता का नियत कोटा होता है। जैसे, किसी शहर में पांच विपणन केंद्र हैं। कंपनी अपने नये उत्पाद के लिए मांग का निर्धारण करना चाहती है और प्रत्येक केंद्र से 50 संभावित ग्राहकों से प्रश्न पूछने के लिए अपने 5 जांचकर्ताओं को भेजती है जो इन लोगों से बातचीत करके मांग का निर्धारण करेंगे। यदि उत्पाद, महिलाओं से संबंधित हैं तो आप घरेलू, नौकरीपेशा, युवा या वृद्ध महिलाओं जैसे महिला ग्राहकों के विविध समूहों से पूरी जानकारी प्राप्त नहीं कर सकते। प्रतिचयन की इस विधि में आपको प्रत्येक जाँचकर्ता के लिए कोटा निर्धारित करना पड़ेगा।

यदि आपके विचारानुसार प्रतिदर्श, समष्टि की विशेषताओं के अनुरूप है तो अन्य विधियों की तुलना में, कोटा प्रतिचयन के अधिक फायदे हैं। इसके अलावा आंकड़ें इकठ्ठे करने में इससे धन और समय भी कम लगता है। लेकिन, इसके साथ-साथ इस विधि के कुछ दोष भी हैं। कोटा प्रतिचयन में, यादृच्छिक प्रतिदर्श चयन की बजाए, प्रतिदर्शों को जांचकर्ता की सहूलियत के आधार पर चुना जाता है। इसलिए चुनिंदा प्रतिदर्श एकं तरफा भी हो सकते हैं। यदि ऐसी विशेषताओं की संख्या अधिक है जिनके आधार पर कोटा तय किया जाता है तो प्रत्येक समूह/उप-समूह के लिए कोटा/उप-कोटा तय करना बेहद कठिन होता है। इसके अलावा, जांचकर्ता भी आमतौर पर ऐसे लोगों से ही सूचना इकठ्ठा करना चाहते हैं जो इसके इच्छुक होते हैं। वे स्वयं भी अनिच्छुक इकाइयों को अनदेखा करते हैं।

17.9.4 तुषारपिंडीय प्रतिचयन

तुषारपिंडीय प्रतिचयन में हम सर्वप्रथम ऐसे व्यक्ति की पहचान करते हैं जो हमारे अध्ययन के अनुरूप है और हमारे मानदंड़ों को पूरा करता है। ऐसे व्यक्ति को चुनने के बाद हम उससे ऐसे अन्य व्यक्तियों की सिफारिश करने को कहते हैं जो सभी हमारे मानदंड़ों को

पूरा कर सकते हों। यद्यपि इस विधि से प्रतिनिधि प्रतिदर्श की प्राप्ति बेहद मुश्किल है लेकिन कभी-कभी यह श्रेष्ठ उपलब्ध विधि भी हो सकती है। यह विशेषरूप से उपयोगी होती है जब आप ऐसी समष्टि तक पहुंच स्थापित करना चाहते हैं जिस तक पहुंचा नहीं जा सकता या जिन्हें ढूंढ़ना मुश्किल है। जैसे, मान लीजिए आप बेघर लोगों का अध्ययन कर रहे हैं। लेकिन किसी विशिष्ट भौगोलिक क्षेत्र में आप उन्हें ढूंढ़ नहीं पा रहे। लेकिन उस क्षेत्र में यदि आप एक या दो ऐसे लोगों की पहचान कर लेते हैं तो वे अपने क्षेत्र के ऐसे अन्य लोगों को कैसे और कहां ढूंढ़ना है, इसकी जानकारी आपको स्वयं दे देंगे।

17.10 प्रतिदर्श आमाप का निर्धारण

प्रतिनिधि प्रतिदर्श की प्राप्ति के लिए, उपयुक्त प्रतिचयन की क्रियाविधि को प्रयोग में लाना बेहद ज़रूरी है। लेकिन यह शर्त पर्याप्त नहीं है। उपर्युक्त के अलावा, हमें प्रतिदर्श आमाप का निर्धारण भी करना चाहिए। प्रतिदर्श का आकार कितना हो, यह एक कठिन प्रश्न है। विविध विचारों द्वारा प्रतिदर्श आकार आमाप का निर्धारण किया जा सकता है। प्रतिदर्श आमाप निर्धारित करने में, ऐसे कुछ विचार निम्नलिखित हैं:

क) प्रतिचयन त्रुटि (sampling error)

ख) संभावित तुलनाओं की संख्या

ग) प्रतिक्रिया दरें (response rates)

घ) उपलब्ध निधियां

क) **प्रतिचयन त्रुटि**: इकाई 15 में आपने सीखा कि बड़े प्रतिदर्श स्थापित करने की तुलना में छोटे प्रतिदर्श बनाते समय प्रतिचयन संबंधी त्रुटियां अधिक होने की संभावना होती है। दूसरी तरफ, छोटे प्रतिदर्शों की तुलना में बड़े प्रतिदर्शों की गैर-प्रतिचयन त्रुटियां अधिक होती हैं। प्रतिचयन त्रुटि ऐसी संख्या है जो प्रतिदर्श के अनुमान की परिशुद्धता को व्यक्त करती है। इसे आमतौर पर सांख्यिकीय विश्वस्यता स्तर (level of confidence) से संबद्ध त्रुटि उपांत (margin of error) के रूप में अभिव्यक्त किया जाता है। जैसे, आप कह सकते है कि प्रधान मंत्री अधिमानी निर्वाचन में पदासीन को 95 प्रतिशत विश्वास्यता स्तर पर 5 प्रतिशत (–) या (+) बिंदुओं के त्रुटि उपांत (परिशुद्धता) से 65 प्रतिशत लोगों का समर्थन मिल रहा है। इसका अर्थ है कि यदि ऐसे ही सर्वेक्षण, मतदाताओं के 100 विविध प्रतिदर्शों पर किए जाएं तो ऐसे 95 सर्वेक्षण दिखाएंगें कि पदाधिकारी को 60 प्रतिशत और 70 प्रतिशत के मतदाताओं के बीच (65% $\pm$ 5%) के किसी प्रतिशत का समर्थन मिला है। ध्यान रखें अपने परिणामों के परिशुद्धता स्तर को जितना बढ़ाएंगें, उतने ही बड़े प्रतिदर्श की आपको आवश्यकता होगी।

ख) **संभावित तुलनाओं की संख्या**: कभी कभार हम प्रतिदर्श में दो या अधिक समूहों (स्तरों) की तुलना करना चाहते हैं। जैसे, हम महिला और पुरुष उत्तरदाताओं या ग्रामीण और शहरी उत्तरदाताओं के बीच तुलना करना चाहते हैं। या हम देश के 4 भौगोलिक क्षेत्र जैसे उत्तर, दक्षिण, पूर्व और पश्चिम के लिए परिणामों की तुलना करना चाहते हैं। इसके लिए हमें समष्टि के प्रत्येक क्षेत्र या स्तर में पर्याप्त आकार के प्रतिदर्श की आवश्यकता होगी। इसलिए, समष्टि, विशेषताओं की विषमता प्रतिदर्श आमाप तय करने में महत्वपूर्ण भूमिका निभाती है।

ग) **प्रतिक्रिया दर** : डाक संबंधी सर्वेक्षणों में, हमें पता है कि उत्तरदाताओं को भेजी जाने वाली सभी प्रश्नावलियां, भरी हुई वापस नहीं आती। डाक सर्वेक्षण से प्राप्त अनुभवों के आधार पर, प्रतिक्रिया दरें 10% से 50% तक के बीच की होती हैं। तब, यदि आप 20% प्रतिक्रिया दर की आशा करते हैं तो आपको अपेक्षित प्रतिदर्श आमाप की संख्या से 5 गुना अधिक प्रश्नावलियां डाक से भेजनी होगी।

घ) **उपलब्ध निधियां:** उपलब्ध निधियां, प्रतिदर्श आमाप को प्रभावित कर सकती हैं। यदि अध्ययन के लिए उपलब्ध निधियां सीमित हैं तो आंकड़ों को इकट्ठा करने के लिए आप उपलब्ध कुल धन की निश्चित राशि से अधिक खर्च करने की स्थिति में नहीं होंगे।

जब आप गैर-प्रायिकता प्रतिचयन की क्रियाविधियों का प्रयोग करते हैं तो प्रतिदर्श का आकार, तय करना और भी कठिन हो जाता है। ऐसा इसलिए है कि गैर-प्रायिकता प्रतिचयन की क्रियाविधियां के ऐसे कुछ सुनिश्चित नियम नहीं हैं। जिनका अनुसरण करना, हमारे लिए आवश्यक है। यह सब कुछ विशिष्ट बातों पर निर्भर करता है जैसे; आप क्या जानना चाहते हैं, पूछताछ का उद्देश्य क्या है, क्या उपयोगी होगा, इसकी विश्वसनीयता क्या है और उपलब्ध समय और संसाधनों से हम क्या कर सकते हैं। उद्देश्यपूर्ण प्रतिचयन में, प्रतिदर्श का निर्णय उद्देश्य के आधार पर लिया जाता है। गैर-प्रायिकता प्रतिचयन की क्रियाविधियों में प्रतिदर्श आमाप के बजाए वैधता (validity), सार्थकता (meaningfulness) और पूरी जानकारी का चुनिंदा प्रतिदर्श इकाइयों की सूचना-उत्कृष्टता से अधिक गूढ़ संबंध होता है।

प्रतिदर्श आमाप निर्धारण के कुछ सूत्र

तकनीकी विशेषज्ञों का मानना है कि अपेक्षित प्रतिदर्श आमाप, ऐसे आकलनों की परिशुद्धता का फल है जिसकी प्राप्ति आप करना चाहते है और समष्टि के प्रसरण और विश्वस्यता का ऐसा स्तर है जिसका प्रयोग आप करना चाहते हैं। यदि आपको परिशुद्धता और विश्वस्यता का स्तर अधिक बड़ा चाहिए तो प्रतिदर्श का आकार भी आपको बड़ा करना होगा। सर्वाधिक प्रयुक्त विश्वस्यता का स्तर 95% और 99% है और अधिकाधिक प्रयुक्त परिशुद्धता के स्तर (precision level) 5% और 1% है। उपर्युक्त चर्चित विविध विचारों के मद्देनज़र, ऐसे बहुत से सूत्र हैं जिनका प्रयोग प्रतिदर्श के आकार का निर्धारण करने में किया जाता है। इस अनुभाग में हम इनमें से तीन की चर्चा करेंगे।

i) यदि हम जवाबी (responding) प्रतिदर्श के प्रतिशत (समानुपातों) के रूप में परिणामों को दर्शाना चाहते हैं तो हम निम्नलिखित सूत्र का प्रयोग करेंगे:

$$n_i = \frac{P_i(1-P_i)}{\dfrac{\alpha^2}{z^2} + \dfrac{P_i(1-P_i)}{N_i}}$$

जहां n_i = अपेक्षित iवें गुण का प्रतिदर्श आमाप है

P_i = विषय (interest) के i वे गुण वाली समष्टि का आकलित समानुपात है जैसे, महिला, पुरुष, शहरी, ग्रामीण आदि के समानुपात।

α = अपेक्षित परिशुद्धता (0.01, 0.05 आदि)

z = विश्वास्यता स्तर दर्शाने वाला का मानकित मूल्य (95% विश्वास्यता स्तर पर $z = 1.96$ और 99% विश्वास्यता स्तर पर $z = 2.58$)

N_i = i वे गुण की समष्टि का आमाप (निर्धारित या आकलित)

उदाहरण 17.8 : किसी समष्टि में 80% ग्रामीण और 20% शहरी व्यक्ति शामिल हैं। यदि उस समष्टि की संख्या 50,000 है अपेक्षित प्रतिदर्श आमाप का निर्धारण कीजिए। मान लीजिए वांछनीय परिशुद्धता और विश्वस्यता का स्तर क्रमशः 1% और 99% है।

इस उदाहरण में,

P_1 = ग्रामीण व्यक्तियों का समानुपात = 0.80 शहरी

P_2 = शहरी व्यक्तियों का समानुपात = 0.20

N_1 = ग्रामीण समष्टि आमाप $50000 \times 0.80 = 40000$

N_2 = शहरी समष्टि आमाप $50000 \times 0.20 = 10000$

$\alpha = 0.01$

$z = 2.58$ (99% विश्वस्यता स्तर पर)

अपेक्षित समष्टि आमाप है

$$n_1 = \text{ग्रामीण प्रतिदर्श} = \frac{P_1(1-P_1)}{\dfrac{\alpha^2}{z^2} + \dfrac{P_1(1-P_1)}{N_1}}$$

$$= \frac{0.80\,(1-0.80)}{\dfrac{0.01^2}{2.58^2} + \dfrac{0.80\,(1-0.80)}{40000}}$$

$$= \frac{0.80\,(.20)}{\dfrac{0.0001}{6.6564} + \dfrac{0.80\,(0.20)}{40000}}$$

$$= \frac{0.16}{0.000019 + \dfrac{0.16}{40000}}$$

$$= \frac{0.16}{0.000019 + 0.000004} = 8410.8$$

अथवा आप लिख सकते है, 8411

$$n_2 = \text{शहरी प्रतिदर्श} = \frac{P_2(1-P_2)}{\dfrac{\alpha^2}{z^2} + \dfrac{P_2(1-P_2)}{N_2}}$$

$$= \frac{0.20\,(1 - 0.20)}{\frac{0.01^2}{2.58^2} + \frac{0.20\,(1 - 0.20)}{10000}}$$

$$= \frac{0.20\,(0.80)}{\frac{0.0001}{6.6564} + \frac{0.20\,(0.80)}{10000}}$$

$$= \frac{0.16}{0.000019 + \frac{0.16}{10000}}$$

$$= \frac{0.16}{0.000019 + 0.000016}$$

$$= \frac{0.16}{0.000035} = 4568.4$$ अथवा आप लिख सकते है, 4568

अतः हमें 8411 + 4568 = 12979 इकाइयों के आमाप का प्रतिदर्श चाहिए।

ii) यदि हम जवाबी (responding) प्रतिदर्श के माध्य या औसत के रूप में परिणाम देना चाहते हैं, तो हमें निम्नलिखित सूत्र का प्रयोग करना होगा :

$$n_i = \frac{P_i^2}{\frac{\alpha_i}{Z^2} + \frac{P_i^2}{N_i}}$$

जहां n_i = अपेक्षित i वे गुण का अतिदर्श आमाप है

P_i = i वे गुण का आकलित मानक विचलन है (जैसे उच्च आय समूह, निम्न आय समूह आदि की औसतन आय)

α = अपेक्षित परिशुद्धता (0.01 या 0.05 मामलानुसार)

z = विश्वस्यता स्तर (95% विश्वस्यता स्तर पर z = 1.96 और 99% विश्वास्यता स्तर पर z = 2.58) दर्शाने का मानकित मूल्य)

N_i = i वे गुण की समष्टि का आकार (निश्चित या आकलित)

उदाहरण 17.9 : किन्हीं परिवारों की औसतन आय जानने के लिए, कोई अध्ययन करने की योजना बनाई जाती है। यदि परिवारों का मानक विचलन 2.5 और समष्टि संख्या 10,000 हो तो अपेक्षित प्रतिदर्श आमाप का निर्धारण कीजिए। मान लीजिए इष्ट परिशुद्धता और विश्वस्यता के स्तर क्रमशः 5% और 95% है।

इस उदाहरण में,

P_1 = आय का मानक विचलन हैं = 2.5

N_1 = परिवारों की संख्या हैं = 10,000

α = 0.05

z = 1.96 (95% विश्वस्यता के स्तर पर)

अपेक्षित प्रतिदर्श आमाप है

$$n_1 = \frac{P_1^2}{\frac{\alpha^2}{z^2} + \frac{P_1^2}{N_1}}$$

$$= \frac{2.5^2}{\frac{0.05^2}{1.96^2} + \frac{2.5^2}{10000}}$$

$$= \frac{6.25}{\frac{0.0025}{3.8416} + \frac{6.25}{10000}}$$

$$= \frac{6.25}{0.000651 + 0.000625}$$

$$= \frac{6.25}{0.001276} = 4898$$

iii) यदि हम परिणामों को अलग–अलग तरीकों से देना चाहते हैं या विषय के गुण के मानक विचलन या समानुपात के आकलन में हमें कठिनाई है तो हम निम्नलिखित सूत्र का प्रयोग करेंगे :

$$n = \frac{0.25}{\frac{\alpha^2}{z^2} + \frac{0.25}{N}}$$

जहां, n = अपेक्षित प्रतिदर्श आमाप है

α = अपेक्षित परिशुद्धता (0.01 या 0.05 यथा मामला)

z = विश्वस्यता स्तर दर्शान का मानकित मूल्य

(95% विश्वास्यता स्तर पर z = 1. 96 और 99% विश्वास्यता स्तर पर z = 2.58)

N = समष्टि आमाप (निश्चित या आकलित)

उदाहरण 17.10: किसी समष्टि में 10,000 व्यक्ति हैं। अपेक्षित प्रतिदर्श आमाप का निर्धारण कीजिए जब इष्ट परिशुद्धता और विश्वास्यता के स्तर क्रमशः 5% और 99% हैं।

इस उदाहरण में,

N = 10,000

α = 0.05

I = 2.58 (99% विश्वस्यता स्तर पर)

अपेक्षित प्रतिदर्श आमाप है

$$n = \frac{0.25}{\dfrac{0.05^2}{2.58^2} + \dfrac{0.25}{10000}}$$

$$n = \frac{0.25}{\dfrac{0.0025}{6.6564} + \dfrac{0.25}{10000}}$$

$$n = \frac{0.25}{0.0003756 + 0.000025} = \frac{0.25}{0.000401} = 624$$

बोध प्रश्न 2

1) बताइए कि निम्नलिखित कथनों में से कौन से सही या ग़लत हैं।

   क) जब इकाइयों को जांचकर्त्ता की जांच के आधार पर प्रतिदर्श में शामिल किया जाता है तो ऐसे प्रतिचयन को यादृच्छिक कहते हैं।

   ख) जैसे–जैसे प्रतिदर्श का आकार (आमाप) बढ़ता जाता है, प्रतिचयन त्रुटियां कम होती जाती हैं।

   ग) सुविधाजनक प्रतिचयन का दोष है कि यह प्रतिनिधि प्रतिदर्श नहीं हो सकती।

2) ऐच्छिक प्रतिचयन का मुख्य दोष है कि

   क) यह क्रियाविधि लंबी और जटिल है।

   ख) प्रतिदर्श–चयन, जांचकर्त्ता की व्यक्तिगत जांच पर निर्भर करता है।

   ग) इससे लघु प्रतिदर्श आकार (आमाप) की प्राप्ति होती है।

   घ) यह अत्यंत ख़र्चीली है।

17.11 सारांश

सरल यादृच्छिक प्रतिचयन ऐसी क्रियाविधि है जिसका प्रयोग सर्वाधिक किया जाता है, क्योंकि यह सभी समष्टि इकाइयों को प्रतिदर्श में शामिल किए जाने की अनुमति देती है। यादृच्छिक संख्या सारणी के प्रयोग से प्रतिदर्श इकाइयों का चयन किया जाता है। क्रमबद्ध यादृच्छिक प्रतिदर्श विधि में आरंभिक बिंदु के रूप में यादृच्छिक रूप से पहली प्रतिदर्श इकाई का प्रयोग किया जाता है और इसके बाद की प्रतिदर्श इकाइयों का चयन स्वतः हो जाता है। स्तरित प्रतिदर्श, प्रत्येक स्तर से इकाइयों के समावेश को सुनिश्चित करता है। गुच्छ प्रतिदर्श में यादृच्छिक रूप से चुने एक या अधिक गुच्छों की पूर्ण गणना शामिल है।

ग़ैर–प्रायिकता प्रतिचयन की क्रियाविधियों में सुविधाजनक प्रतिचयन, ऐच्छिक प्रतिचयन, कोटा प्रतिचयन और तुषारपिंडीय प्रतिचयन शामिल हैं। ये प्रतिचयन की क्रियाविधियां प्रतिचयन संबंधी पक्षपात से मुक्त नहीं है, लेकिन कुछ स्थितियों विशेष रूप से विपणन संबंधी शोधकार्यों में अभी भी, ये प्रचलित हैं।

प्रतिदर्श आमाप का निर्णय लेने मे बहुत से कारक शामिल हैं। ये, समष्टि में शामिल विविध समूह, समष्टि विषमता और उपलब्ध निधियां और समय जैसे कारक हो सकते हैं।

शोधकार्यों में प्रतिदर्श के प्रयोग से समय, धन और जनशक्ति की बचत होती है, यदि इकाइयों के चयन में उपयोगी प्रतिचयन क्रियाविधि का प्रयोग किया जाए तो उपयुक्त प्रतिदर्श आमाप का चयन किया जाता है और प्रतिचयन त्रुटियों को कम करने की अनिवार्य सावधानियां बरती जाती है। इससे समष्टि के बारे में चुनिंदा प्रतिदर्श से हमें वैध और विश्वसनीय जानकारी ही प्राप्त होगी।

17.12 शब्दावली

गुच्छ प्रतिचयन (Cluster Sampling)	:	प्रतिचयन की ऐसी क्रियाविधि जहां समग्र समष्टि को समूहों या गुच्छों में बांट दिया जाता है और इसके बाद यादृच्छिक तरीके से गुच्छों का चयन किया जाता है। चुनिंदा गुच्छों के सभी प्रेक्षण, प्रतिचयन में शामिल किए जाते हैं।
सुविधाजनक प्रतिचयन Convenience Sampling)	:	इसका अर्थ, प्रतिदर्श प्राप्त करने की ऐसी विधि से है जो शोधकर्त्ता को सर्वाधिक आसान तरीके से उपलब्ध हो।
ऐच्छिक प्रतिचयन (Judgement Sampling)	:	प्रतिचयन की ऐसी क्रियाविधि जिसमें प्रतिदर्श इकाइयों की अपेक्षित उपयुक्त विशेषताओं के बारे में शोधकर्त्ता के व्यक्तिगत निर्णय के आधार पर प्रतिदर्श का चयन किया जाता है।
बहुचरणी प्रतिचयन (Multi-Stage Sampling)	:	इस विधि में प्रतिदर्श का चयन विभिन्न चरणों के माध्यम से किया जाता है।
कोटा/नियतमात्रात्मक प्रतिचयन (Quota Sampling)	:	प्रतिचयन की इस क्रियाविधि में, कुछ मानकों जैसे; आयु, लिंग, भौगोलिक क्षेत्र, शिक्षा, आय, धर्म आदि के आधार पर प्रतिदर्शों का चयन किया जाता है।
यादृच्छिक प्रतिचयन (Random Sampling)	:	यह प्रतिचयन की तकनीक है जहां हम समष्टि से प्रतिदर्श का चयन करते हैं। यहां, समष्टि की प्रत्येक इकाई की प्रतिदर्श में शामिल होने की संभावना होती है।
सरल यादृच्छिक प्रतिचयन (Simple Rendom Sampling)	:	यह प्रतिचयन की बुनियादी क्रियाविधि है, जब हम लॉटरी विधि या यादृच्छिक संख्या सारणियों का प्रयोग करते हुए प्रतिदर्श का चयन करते हैं।
तुषारपिंडीय प्रतिचयन (Snowball Sampling)	:	तुषारपिंडीय प्रतिचयन; अतिरिक्त प्रतिचयन की इकाइयों को सृजित करने के लिए प्रारंभिक प्रतिचयन इकाइयों से प्राप्त परामर्श पर निर्भर करती हैं।
स्तरित प्रतिचयन (Stratified Sampling)	:	इस प्रतिचयन की क्रियाविधि में समष्टि को समूहों (स्तरों) में बांटा जाता है और इसके बाद यादृच्छिक प्रतिचयन विधि के प्रयोग से प्रत्येक स्तर से प्रतिदर्शों का चयन किया जाता है।

क्रमबद्ध प्रतिचयन (Systematic Sampling)	:	प्रतिचयन की ऐसी क्रियाविधि जिसमें समष्टि से इकाइयों का चयन एक समान अंतराल पर होता है और जिसे समय, नियम या स्थान में मापा जा सकता है।

17.13 कुछ उपयोगी पुस्तकें

Kothari, C.R. (1985) *Research Methodology: Methods and Techniques,* Wiley Eastern, New Delhi.

Levin, R.I. and D.S. Rubin. (1999) *Statistics for Management,* Prentice-Hall of India, New Delhi.

Mustafi, C.K. (1981) *Statistical Methods in Managerial Decisions,* Macmillan, New Delhi.

Plane, D.R. and E.B. Oppermann. (1986) *Business and Economic Statistics,* Business Publications, Inc: Plano.

Zikmund, William G. (1988) *Business Research Methods,* The Dryden Press, New York.

17.14 बोध प्रश्नों के उत्तर या संकेत

बोध प्रश्न 1

1) घ

2) च

3) क) सही   3) ख) सही

4) ग

बोध प्रश्न 2

1) क) ग़लत   1) ख) सही   1) ग) सही

2) ख

इकाई 18 सांख्यिकीय आकलन

इकाई की रूपरेखा

18.0 उद्देश्य

इस इकाई को पढ़ने के बाद, आपः

- आकलन की संकल्पना को व्यक्त कर सकेंगे;
- बिंदु आकलन और अंतराल आकलन के बीच अंतर को स्पष्ट कर सकेंगे;
- प्राचल के लिए विश्वस्यता-अंतराल का आकलन कर सकेंगे; और
- विश्वस्यता-स्तर की संकल्पना की व्याख्या कर सकेंगे।

18.1 प्रस्तावना

कई बार कुछ सीमाओं, जैसे अपर्याप्त निधि या श्रम शक्ति का अभाव या समय की कमी के कारण हम समष्टि की सभी इकाइयों का सर्वेक्षण नहीं कर पाते। ऐसी स्थिति में हम प्रतिचयन का सहारा लेते हैं अर्थात हम समष्टि के किन्हीं अंशों का ही सर्वेक्षण करते हैं। प्रतिदर्श में शामिल सूचना के आधार पर हम समष्टि के बारे में निष्कर्ष निकालते हैं। यह प्रक्रिया सांख्यिकीय अनुमिति कहलाती है। हमारा यकीन है कि सांख्यिकीय अनुमिति का अर्थशास्त्र के साथ-साथ अन्य अनेक क्षेत्रों, जैसे समाजशास्त्र, मनोविज्ञान, राजनीति विज्ञान, आयुर्विज्ञान आदि में भी बड़े पैमाने पर इस्तेमाल किया जाता है। उदाहरण के लिए, चुनाव की प्रक्रिया शुरू होने से पहले या चुनाव परिणामों की घोषणा होने से ठीक पहले अनेक समाचार पत्र और दूरदर्शन के चैनल चुनाव सर्वेक्षण या एक्जिट पोल का संचालन करते हैं। इनका उद्देश्य वास्तविक परिणाम घोषित होने से पहले चुनावी परिणामों के बारे में पूर्वानुमान लगाना है। उस स्थिति में सर्वेक्षणकर्ताओं के लिए सभी मतदाताओं से उनके पसंद के प्रत्याशी के बारे में जान पाना संभव नहीं होता। यह समयावधि बहुत छोटी होती

है, संसाधन बहुत सीमित होते हैं, पर्याप्त संख्या में इस काम के लिए व्यक्ति उपलब्ध नहीं होते हैं और चुनाव से पहले पूर्ण रूप से सर्वेक्षण करने से चुनाव का मूल उद्देश्य ही समाप्त हो जाता है।

उपर्युक्त उदाहरण में सर्वेक्षणकर्ता वास्तव में उस परिणाम के बारे में नहीं जानता है, जो मतदाता के मतदान के फलस्वरूप दिखाई देगा। यहाँ पर सभी मतदाताओं की कुल संख्या से समष्टि का निर्माण होता है। इस संदर्भ में सर्वेक्षणकर्ता समष्टि के 'प्रतिनिधि' प्रतिदर्श से आँकड़े इकट्ठा करता है, सभी मतदाताओं से नहीं। प्रतिदर्श से प्राप्त सूचना के आधार पर सर्वेक्षणकर्ता समग्र समष्टि के बारे में अपना पूर्वानुमान व्यक्त करता है।

इस इकाई में हम सांख्यिकीय अनुमिति की संकल्पना और सांख्यिकीय आकलन की विधियों के बारे में अध्ययन करेंगे। प्रांचल, जैसा कि आप जानते हैं, समष्टि इकाइयों का एक फलन है जबकि प्रतिदर्शज प्रतिदर्शी इकाइयों का फलन है। प्राचल और संगत प्रतिदर्शजों की संख्या काफी हो सकती है। लेकिन, अपनी प्रस्तुति को आसान बनाने के लिए हम इसे केवल समांतर माध्य तक ही सीमित रखेंगे।

18.2 सांख्यिकीय पृष्ठभूमि

पिछले दो खंडों में हमने दो महत्वपूर्ण पहलुओं की चर्चा की है: सैद्धांतिक प्रायिकता बंटन और प्रतिदर्श तकनीक। ये दो पहलू बुनियादी सांख्यिकीय अनुमिति का निर्माण करते हैं।

खंड 5 की इकाई 14 में हमने यादृच्छिक चर की संकल्पना का वर्णन किया था। हमने सीखा कि X एक यादृच्छिक चर है, यदि इसके मान $x_1, x_2, \dots\dots\dots, x_n$ और संगत प्रायिकताएँ $p_1, p_2, \dots\dots, p_n$ हैं। यहाँ पर x_1 के उभरने की प्रायिकता p_1 है तो x_2 की प्रायिकता p_2 है। यदि मान $x_1, x_2, \dots\dots\dots, x_n$ असंतत हैं तो x के वियुक्त मानों (isolated values) की प्रायिकता ज्ञात हो सकती है। दूसरी तरफ X यदि सतत् यादृच्छिक चर हो तो निश्चित परिसर के भीतर X की प्रायिकता ज्ञात की जाएगी जैसा कि $P(a \le X \le b) = p_1$

खंड 5 की इकाई 14 और 15 में हमने सैद्धांतिक असतत् प्रायिकता बंटन (जैसे कि द्विपद और पाइसो) और सतत प्रायिकता बंटन (जैसेकि प्रसामान्य और t) की चर्चा की। हमने सीखा कि यदि X की प्रसार अनंत हो, तब ये प्रायिकता बंटन, प्रसामान्य बंटन का अनुसरण करते हैं। अतः प्रसामान्य बंटन इन प्रायिकता बंटनों की परिसीमित स्थिति है और इसे एक आदर्श प्रायिकता बंटन माना जा सकता है।

प्रसामान्य बंटन को दो प्राचलों द्वारा परिभाषित किया जाता है। ये हैं : माध्य (μ) और मानक विचलन (σ)। यदि यादृच्छिक चर से संबंद्ध प्रायिकताएँ, प्रसामान्य बंटन के आधार पर बंटित हों (इसका अर्थ है कि यदि X, प्रसामान्य बंटन का अनुसरण करता है) तब इसके प्रायिकता बंटन फलन के लिए समीकरण का प्रयोग करते हुए हम $p(a \le X \le b) = p_1$ प्रायिकता ज्ञात कर सकते हैं।

यहाँ हमारे समक्ष एक समस्या है कि μ और σ कोई भी मान ग्रहण कर सकते हैं और संगत प्रायिकता ज्ञात करने में काफी समय लगता है। इस समस्या का समाधान है कि प्रसामान्य चर से μ को घटाना और इसे σ से विभाजित करना। इस तरह हम मानक प्रसामान्य विचर, $z = \dfrac{x-\mu}{\sigma}$ प्राप्त करते हैं, जिसका माध्य = 0 और मानक विचलन = 1 है। आलेख पर z के विभिन्न मानों के लिए प्रायिकताओं के आलेखन से हम 'मानक

प्रसामान्य वक्र' (standard normal curve) को प्राप्त करते हैं, जो सममित है और वक्र के नीचे का क्षेत्र = 1 है। ध्यान रखें कि मानक प्रसामान्य वक्र के मामले में हम x-अक्ष पर $z = \frac{x-\mu}{\sigma}$ को मापते हैं और y-अक्ष पर z के उभरने की प्रायिकता अर्थात $p(z)$ है। अतः मान लीजिए यदि हम प्रसामान्य वक्र के किसी विशेष भाग (जैसे z के दो मानों से बद्ध, मान लीजिए, z_1 और z_2)पर विचार करें तो वक्र के नीचे का क्षेत्र इसकी प्रायिकता हमें देगा। ध्यान रखें कि इस पाठ्यक्रम के खंड 1 में वर्णित बारम्बारता वक्र से प्रसामान्य वक्र भिन्न होता है। प्रसामान्य वक्र के नीचे के क्षेत्रफल से बारम्बारता प्राप्त नहीं होती। इससे हम प्रायिकताओं की प्राप्ति करते हैं।

खंड 6 की इकाई 16 में हमने अध्ययन किया था कि अक्सर संपूर्ण समष्टि का अध्ययन करना संभव नहीं होता । इसलिए प्रतिदर्श सर्वेक्षण का सहारा लेना पड़ता है। यदि प्रत्येक समष्टि की इकाई से संबद्ध उपयुक्त प्रायिकता के माध्यम से यादृच्छिक तौर पर प्रतिदर्श लिया जाता है और प्रतिदर्श का आकार बहुत छोटा भी नहीं है तो ऐसा प्रतिदर्श, समष्टि के इस भाग का प्रतिनिधित्व कर सकता है। याद रखें कि हम प्राप्त समष्टि से अनेक प्रतिदर्श प्राप्त कर सकते हैं और ऐसे प्रत्येक प्रतिदर्श से हमें प्रतिदर्श माध्य ($\bar{x}$) की प्राप्ति होती है। अतः प्रतिदर्श माध्यों के बारंबारता बंटन को 'प्रतिदर्शी बंटन' कहते हैं।

जैसा कि खंड 6 की इकाई 16 में हमने अध्ययन किया था कि प्रतिदर्श माध्य ($\bar{x}$) के अलग-अलग मान हो सकते हैं और ऐसे प्रत्येक मान की प्रायिकता भी होती है। अतः प्रतिदर्श माध्य को यादृच्छिक चर के रूप में देखा जा सकता है। वास्तविक जीवन में हमारे पास परिसीमित समष्टि और अनेक प्रतिदर्श होते हैं (और इसलिए प्रतिदर्श माध्यों की संख्या) परिसीमित होती है। ऐसे मामले में $\bar{x}$ असतत् यादृच्छिक चर है लेकिन जब प्रतिदर्शों की संख्या अनंत हो तो $\bar{x}$ सतत् यादृच्छिक चर हो सकता है।

आइए, अब हम इकाई 16 में चर्चित एक अन्य महत्वपूर्ण संकल्पना अर्थात केंद्रीय सीमा प्रमेय पर विचार करें। इसके अनुसार $\bar{x}$ का प्रतिदर्शी बंटन प्रसामान्य है यदि वह समष्टि जिससे प्रतिदर्श लिया गया है भी प्रसामान्य हो। लेकिन, $\bar{x}$ का प्रतिदर्शी बंटन सन्निकटतः प्रसामान्य होगा, यदि प्रतिदर्श आकार (n) बड़ा हो और फिर चाहे मूल समष्टि प्रसामान्य नहीं भी हो। यदि मूल समष्टि प्रसामान्य प्रायः हो तो छोटे आकार के प्रतिदर्शों के माध्य का प्रतिदर्शी बंटन भी प्रसामान्य प्रायः होगा।

हमें पता है कि प्रतिदर्श माध्य का प्रकीर्णन, मूल समष्टि अर्थात जिससे प्रतिदर्श लिया गया है, के प्रकीर्णन से छोटा होता है। स्मरण करें कि प्रतिदर्शी बंटन का मानक विचलन, मानक त्रुटि कहलाता है। इसलिए यदि समष्टि का मानक विचलन σ है तब प्रतिदर्श माध्य की मानक त्रुटि $\frac{\sigma}{\sqrt{n}}$ है।

उपर्युक्त बातों से हमने सीखा कि प्रतिदर्श माध्य को यादृच्छिक चर माना जा सकता है और जब प्रतिदर्श का आकार बड़ा होता है, तब यह प्रसामान्य बंटन के सन्निकट होती है। आमतौर पर यदि $n > 30$ हो तो हम प्रतिदर्श को आकार में बड़ा कहते हैं। छोटे प्रतिदर्श ($n \leq 30$) के लिए प्रतिदर्श माध्य का प्रतिदर्शी बंटन, स्टूडेन्ट टी बंटन के समान होता है। ध्यान रखें कि t बंटन के मामले में प्रायिकता वक्र का आकार, इसकी स्वतंत्रता की कोटियों के मुताबिक बदलता रहता है।

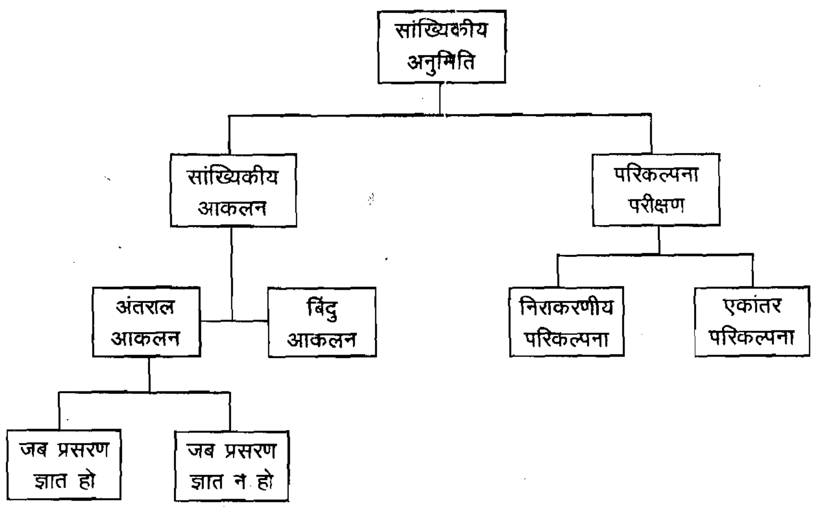
18.3 सांख्यिकीय अनुमिति की संकल्पना

जैसा कि पहले उल्लेख किया गया है, सांख्यिकीय अनुमिति, समष्टि से प्राप्त प्रतिदर्श में शामिल जानकारी के आधार पर समष्टि विशेषताओं के बारे में निष्कर्ष निकालने की विधियों से संबंधित है। ध्यान रखें क़ि हमें समष्टि माध्य का पता नहीं है लेकिन प्रतिदर्श माध्य की जानकारी हमें है। सांख्यिकीय अनुमिति में हम दो प्रकार के प्रश्नों के उत्तर जानना चाहेंगे। पहला, समष्टि माध्य का मान क्या होगा? हमारा उत्तर, समष्टि माध्य के बारे में सोचे-समझे अनुमान में निहित है। सांख्यिकीय अनुमिति का यह पहलू 'आकलन' (estimation) कहलाता है। हमारा दूसरा प्रश्न समष्टि माध्य के बारे में हमारे कुछ विशेष दावे से संबंधित है। मान लीजिए, बिजली के बल्ब बनाने वाले किसी उत्पादक का दावा है कि बिजली के बल्बों का माध्य जीवन 2000 घंटों के बराबर है। प्रतिदर्श सूचना के आधार पर क्या हम कह सकते हैं कि यह दावा सही नहीं है? सांख्यिकीय अनुमिति का यह पक्ष परिकल्पना परीक्षण (hypothesis testing) कहलाता है।

अतः सांख्यिकीय अनुमिति के दो पहलू हैं: आकलन और परिकल्पना परीक्षण। इस इकाई में हम सांख्यिकीय आकलन के बारे में चर्चा करेंगे और परिकल्पना परीक्षण की चर्चा हम अगली इकाई में करेंगे। नीचे चित्र 18.1 में सांख्यिकीय अनुमिति के विविध पहलुओं को संक्षेप में हमने दर्शाया है। यहाँ ध्यान देने योग्य महत्वपूर्ण कारक है कि क्या हमें समष्टि प्रसरण का पता है या नहीं। निस्संदेह जब हमें समष्टि माध्य का पता न हो तो हम समष्टि प्रसरण का पता कैसे लगा सकते हैं? हम ऐसे मामले से शुरू करते हैं जहाँ हमें समष्टि प्रसरण का पता हो, क्योंकि संकल्पनाओं की व्याख्या में यह कारगर सिद्ध होगा। तत्पश्चात हम अज्ञात समष्टि प्रसरण के अपेक्षाकृत अधिक यथार्थिक मामलों पर ध्यान केंद्रित करेंगे।

आकलन भी दो तरह का होता है: बिंदु आकलन और अंतराल आकलन। बिंदु आकलन में हम एकल बिंदु के रूप में समष्टि प्राचल के मान का आकलन करते हैं। जबकि दूसरी तरफ, अंतराल आकलन के मामले में हम प्रतिदर्श माध्य के इर्द-गिर्द निम्न एवं उच्च परिबंधों का आकलन करते हैं जिनके भीतर समष्टि माध्य रहता है।

समष्टि के बारे में हमारा दावा, निराकरणीय परिकल्पना (null hypothsis) और इसके प्रतिपक्ष एकांतर परिकल्पना (alternative hypothesis) के रूप में होगा। परिकल्पना के परीक्षण की विधियों और इसकी संकल्पनाओं का वर्णन हमारी अगली इकाई में है।



चित्र 18.1 : सांख्यिकीय अनुमिति

बोध प्रश्न 1

1) निम्नलिखित संकल्पनाओं को स्पष्ट कीजिए :

   क) मानक प्रसामान्य विचर

   ख) यादृच्छिक चर

   ग) प्रतिदर्शी बंटन

   घ) केंद्रीय सीमा प्रमेय

   ....

   ....

   ....

   ....

   ....

2) बताइए कि निम्नलिखित कथन सही हैं या गलत।

   क) प्रसामान्य बंटन, द्विपद बंटन का परिसीमित रूप है।

   ख) प्रतिदर्शज के प्रतिदर्शी बंटन का मानक विचलन, मानक त्रुटि कहलाता है।

   ग) पाइसो बंटन, सतत् बंटन का एक उदाहरण है।

   घ) सांख्यिकीय आकलन, सांख्यिकीय अनुमिति का एक भाग है।

   ....

   ....

   ....

   ....

   ....

18.4 बिंदु आकलन

जैसा कि हमने पहले बताया था, हमें प्राचल मान का पता नहीं है और प्रतिदर्श सूचना के प्रयोग से हम इसका अनुमान लगाना चाहते हैं। निस्संदेह, प्रतिदर्शज का मान, श्रेष्ठ अनुमान होगा। उदाहरण के तौर पर हमें समष्टि माध्य का पता नहीं है तो इस स्थिति में प्रतिदर्श माध्य, श्रेष्ठ अनुमान है। यहाँ इस मामले में हम प्राचल के 'अनुमान' के रूप में एकल मान या बिंदु का प्रयोग करेंगे।

इकाई 16 में हमने आकल (estimate) और आकलक (estimator) की संकल्पनाओं को व्यक्त किया था। स्मरण कीजिए कि आकलक एक सूत्र है और आकल, इस सूत्र के प्रयोग से प्राप्त विशिष्ट मान। जैसे, समष्टि माध्य के आकलन के लिए हम प्रतिदर्श माध्य का प्रयोग करते हैं तब $\frac{1}{n}\sum_i x_i$ आकलक है। मान लीजिए किसी प्रतिदर्श पर आंकड़ों को एकत्रित किया जाता है और इस सूत्र में प्रतिदर्शी इकाइयों को रखकर, प्रतिदर्श माध्य के लिए, ऐसे किसी विशिष्ट मान की प्राप्ति की जाती है; मान लीजिए यह 120 है। तब ऐसी स्थिति में 120 समष्टि माध्य का आकल है। यह संभव है कि आप समान समष्टि से एक अन्य प्रतिदर्श प्राप्त कर लें, प्रतिदर्श माध्य के लिए सूत्र $\frac{1}{n}\sum_i x_i$ का प्रयोग करें, और एक अलग मान

जैसे 123 की प्राप्ति करें। यहाँ 120 और 123 अर्थात दोनों समष्टि माध्य के आकल हैं। लेकिन इन दोनों मामलों में आकलक एक ही है अर्थात $\frac{1}{n}\sum_i x_i$ । याद रखें कि पारिभाषिक शब्द प्रतिदर्शज, जो प्रतिदर्श मानों के फलन के संदर्भ में प्रयुक्त है, आकलक शब्द का समानार्थक है।

ऐसी स्थितियाँ भी हो सकती हैं जब आप प्राचल के लिए एक से अधिक संभावी (potential) आकलकों (एकांतर सूत्र) की प्राप्ति करेंगे। इन प्राचलों में से श्रेष्ठ के चयन के लिए हमें कुछ निर्धारित मानदंडों का अनुसरण करना होगा। इन मानदंडों के आधार पर आकलक को कुछ निश्चित वांछनीय गुणधर्मों को पूरा करना होगा। वैसे तो आकलक के लिए वांछनीय गुणधर्म गिने-चुने हैं लेकिन सर्वाधिक महत्वपूर्ण है इसकी अनभिनता (unbiasedness)।

अनभिनता का अर्थ है कि आकल, प्राचल के अज्ञात मान से उच्च या निम्न हो सकता है। लेकिन आकल का प्रत्याशित मान प्राचल के बराबर होना चाहिए। जैसे, प्रतिदर्श माध्य एक प्रतिदर्श से दूसरे प्रतिदर्शज में अलग-अलग होता है लेकिन-औसतन यह समष्टि माध्य के बराबर होगा। अन्य शब्दों में $E(\bar{x}) = \mu$

लेकिन, $\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ समष्टि प्रसरण $\sigma^2 = \frac{1}{N}\Sigma(X_i - \bar{X})^2$ का अनभिनत आकलक नहीं है। दरअसल यदि हम $s^2 = \frac{1}{n-1}\Sigma(x_i - \bar{x})^2$ को परिभाषित करें तब s^2, σ^2 का भी अनभिनत आकलक है। अतः, प्रतिदर्श मानक विचलन s की, समष्टि मानक विचलन σ से कम होने की प्रवृत्ति है। इस शर्त को संशोधित करने के लिए हम n की बजाए कृत्रिम रूप से किसी छोटी संख्या $(n-1)$ से s को उच्च करने के लिए विभाजित करते हैं।

परिकल्पना के परीक्षण के लिए बिंदु आकलन का विशेष महत्व है और इसका अध्ययन हम इकाई 19 में करेंगे।

18.5 ज्ञात प्रसरण संबंधी विश्वस्यता अंतराल

जैसा कि हमने ऊपर बिंदु आकलन में देखा, आमतौर पर हम एकल मान अर्थात संगत प्रतिदर्शज द्वारा प्राचल को आकलित करते हैं। इस तरीके से बिंदु आकलन वास्तविकता पूर्ण होना जरूरी नहीं होता, क्योंकि प्रायल का मान इससे भिन्न भी हो सकता है। इसकी एक वैकल्पिक प्रक्रिया है, अंतराल का निर्धारण जो प्राचल को कुछ निश्चित प्रायिकता तक बाँधे रखे। यहाँ हम निम्न सीमा एवं उच्च सीमा पर विशेष ध्यान देते हैं जिसके भीतर प्राचल का मान बना रहेगा। इसके अलावा हम प्राचल की प्रायिकता पर भी विशेष ध्यान डालते हैं जो उस अंतराल में बनी हुई है। इस संदर्भ में **अंतराल** को हम 'विश्वश्यता अंतराल' और इस अंतराल में प्राचल की प्रायिकता को 'विश्वस्यता स्तर' या 'विश्वस्यता गुणांक' कहते हैं।

आइए अब एक उदाहरण लें। मान लीजिए आपको छत्तीसगढ़ राज्य के रायगढ़ जिले के लोगों की औसतन आमदनी का अनुमान लगाना है। आपने 500 परिवारों के प्रतिदर्श से आंकड़े एकत्रित किए और पाया कि औसतन अम्मदनी (मान लीजिए $\bar{x}$) प्रति वर्ष 18,250 रु. है। प्रतिदर्शी त्रुटि के कारण छत्तीसगढ़ के रायगढ़ ज़िले की वास्तविक औसतन आमदनी (μ) के संदर्भ में यह प्रतिदर्श शायद सही परिणाम नहीं दे रहा होगा। इसलिए हम निश्चित रूप से नहीं कह सकते कि जिले की औसतन आमदनी 18,250 रुपए है या नहीं। दूसरी तरफ, यह कहना बेहतर होगा कि छत्तीसगढ़ के रायगढ़ जिले की औसतन आमदनी प्रति

वर्ष 17,900 रुपए और 18,600 रुपए के बीच है। इसके अलावा हमें यह भी स्पष्ट करना होगा कि औसतन आमदनी इन सीमाओं में ही रहेगी, इसका प्रायिकता 95% है। अतः इस मामले में हमारा विश्वस्यता अंतराल 17,900 - 18,600 रुपए है और विश्वस्यता स्तर या विश्वस्यता गुणांक 95% है।

इस संदर्भ में हमारे मस्तिष्क में उठने वाला प्रश्न होगा कि हम विश्वस्यता अंतराल और विश्वस्यता गुणांक की प्राप्ति कैसे करते हैं? आइए, विश्वस्यता गुणांक से शुरुआत करें। हमें पता है कि बड़े प्रतिदर्शों के लिए $\bar{x}$ का प्रतिदर्शी बंटन प्रसामान्य रूप से बंटित होता है और इसका माध्य μ और मानक त्रुटि $\frac{\sigma}{\sqrt{n}}$ है और जहाँ n प्रतिदर्श का आकार है। प्रतिदर्शी बंटन को परिवर्तित करने पर हमें मानक प्रसामान्य विचर ($z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$) की प्राप्ति होती है जिसका माध्य शून्य और प्रसरण 1 है। मानक प्रसामान्य वक्र सममित है और इसलिए, $0 \le z \le \infty$ के लिए वक्र के नीचे का क्षेत्रफल 0.5 है और जिसे हमने खंड 5 की इकाई 15 की सारणी 15.1 में दर्शाया है। आइए अब मान लें कि हमारा विश्वस्यता गुणांक 95% (अर्थात 0.95) है। इस संदर्भ में हमें z के लिए परिसर ज्ञात करना होगा जो मानक प्रसामान्य वक्र का 0.95 क्षेत्र ढक लेगा। चूँकि z का बंटन सममित है, इसलिए $z = 0$ के दाये ओर 0.475 क्षेत्र और बायें ओर 0.475 क्षेत्र रहना चाहिए। यदि हम प्रसामान्य क्षेत्र तालिका (तालिका 15.1) देखें तो पाते हैं कि 0.475 क्षेत्र आच्छादित होगा जब $z = 1.96$ होगा। अतः इसकी प्रायिकता कि z का परिसर -1.96 और 1.96 के बीच हो, 0.95 होगी। इस सूचना के आधार पर आइए पिछले प्रश्न पर दुबारा ध्यान केंद्रित करें और ऐसा परिसर ज्ञात करें जिसके भीतर μ बना रहेगा।

हम पाते हैं कि

$$p(-1.96 \le z \le 1.96) = 0.95 \qquad \ldots(18.1)$$

$$or\ p\left(-1.96 \le \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \le 1.96\right) = 0.95$$

$$or\ p\left(-1.96\frac{\sigma}{\sqrt{n}} \le \bar{x} - \mu \le 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$$

$$or\ p\left(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95 \qquad \ldots(18.2)$$

आइए, उपर्युक्त की व्याख्या करें। स्मरण कीजिए कि हरेक प्रतिदर्श हमें $\bar{x}$ का अलग मान देगा। इसी आधार पर, विश्वस्यता अंतराल भी अलग होगा। प्रत्येक मामले में विश्वस्यता अंतराल में अज्ञात प्राचल शामिल हो सकता है या नहीं भी । समीकरण (18.2) से आशय है कि यदि यादृच्छिक प्रतिदर्शों की संख्या अधिक है तो प्राप्त समष्टि से n आकार का हरेक प्रतिदर्श लिया जाता है और यदि ऐसे हरेक प्रतिदर्श के लिए $\left(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}, \le \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right)$

अंतराल का निर्धारण हो जाता है तब लगभग 95 फीसदी मामलों में, अंतराल में समष्टि माध्य μ शामिल होगा।

विश्वस्यता गुणांक को $(1-\alpha)$ से दर्शाया जाता है जहाँ α सार्थकता का स्तर है (इकाई 19 में हम 'सार्थकता के स्तर' की संकल्पना की चर्चा करेंगे)। विश्वस्यता गुणांक किसी भी मान को ले सकता है। हम अपनी निष्कर्षों की सार्थकता का अनुमान लगाने के लिए, मान लीजिए 81% या 97% को विश्वस्यता का स्तर मान सकते हैं। सुविधा के नज़रिए से अक्सर दो विश्वस्यता स्तर अर्थात 95% और 99% ही व्यापक रूप से प्रयोग में लाए जाते हैं। हाँ, कभी-कभार 90% विश्वस्यता स्तर भी प्रयोग में हम लाते हैं। आइए, विश्वस्यता अंतराल ज्ञात करें जब विश्वस्यता गुणांक $(1-\alpha)=0.99$ हो। इस मामले में 0.495 मानक प्रसामान्य वक्र के दायें/बायें अर्थात दोनों में से किसी एक तरफ होना चाहिए। यदि हम प्रसामान्य क्षेत्र तालिका (तालिक 15.1) पर नज़र डालें तो हम पाते हैं कि 0.495 क्षेत्र आच्छादित होगा जब $z = 2.58$ होगा।

अतः $P(-2.58 \le z \le 2.58) = 0.99$ ...(18.3)

उपर्युक्त में मदों को पुनःव्यवस्थित करने पर हम पाते हैं कि

$$p\left(\bar{x} - 2.58\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + 2.586\frac{\sigma}{\sqrt{n}} \right) = 0.99 \qquad \text{...(18.4)}$$

समीकरण (18.4) से पता चलता है कि μ के लिए 99% विश्वस्यता अंतराल $\bar{x} \pm 2.58\frac{\sigma}{\sqrt{n}}$ द्वारा दिखाया गया है।

प्रसामान्य क्षेत्र तालिका में नज़र डालें तो 0.90 के विश्वस्यता गुणांक के लिए अब हम विश्वस्यता अंतराल निकाल सकते हैं और पता लगा सकते हैं कि

$$p\left(\bar{x} - 1.65\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + 1.65\frac{\sigma}{\sqrt{n}} \right) = 0.90 \qquad \text{...(18.5)}$$

हमने उपर्युक्त समीकरण (18.2), (18.4) और (18.5) पर गौर किया कि जैसे-जैसे विश्वस्यता अंतराल विस्तृत होते जाते हैं, प्राचल (इस मामले में μ) को अंतराल में रखने की संभावना बढ़ जाती है।

विश्वस्यता अंतराल की दो सीमाएँ **विश्वस्यता सीमाएँ** कहलाती हैं। जैसे, 95% विश्वस्यता के लिए, निम्न विश्वस्यता सीमा है $\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}$ और उच्च विश्वस्यता सीमा है $\bar{x} + 1.96\frac{\sigma}{\sqrt{n}}$ 1 μ की अंतराल में ही सहीं ढंग से बनाए रखने के लिए इन दो सीमाओं में जो विश्वास हम कायम कर लेते हैं, उसे विश्वस्यता गुणांक के रूप में व्यक्त किया जा सकता है।

उदाहरण 18.1

कोई पेपर कंपनी अनुमान लगाना चाहती है कि नयी मशीन पर एक रीम पेपर बनाने में औसतन कितना समय लगता है। 36 रिमों का चादृच्छिक प्रतिदर्श दर्शाता है कि एक रिम

पेपर औसतन 1.5 मिनट में बनते हैं। समष्टि मानक विचलन 0.30 मिनट है। 95% विश्वस्यता स्तर पर अंतराल आकल निर्मित कीजिए।

प्राप्त जानकारी के आधार पर

$\bar{x} = 1.5,\ \sigma = 0.30$ और $n = 36.$

चूँकि $n = 36\ (> 30)$, यहाँ प्रतिदर्श को बड़ा प्रतिदर्श माना गया है और इसी आधार पर $\bar{x}$ प्रसामान्य रूप से बंटित है। और इसका माध्य μ और मानक त्रुटि $\frac{\sigma}{\sqrt{n}}$ है

$$\frac{\sigma}{\sqrt{n}} = \frac{0.30}{\sqrt{30}} = 0.05$$

95% विश्वस्यता अंतराल इस प्रकार होगा,

$$\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}$$

अथवा $1.5 - 1.96 \times 0.05 \le \mu \le 1.5 + 196 \times 0.05$

अर्थात $1.402 \le \mu \le 1.598$

अतः यदि 95% विश्वस्यता स्तर है तो हम कह सकते हैं कि नई मशीन के लिए औसतन उत्पादन समय 1.402 मिनट और 1.598 मिनट के बीच होगा। यहाँ 1.402 निम्न विश्वस्यता सीमा है और 1.598 उच्च विश्वस्यता सीमा है।

18.6 अज्ञात प्रसरण संबंधी विश्वस्यता अंतराल

पिछले अनुभाग में हमने समष्टि माध्य के लिए विश्वस्यता अंतराल आकलित किया था और इस संदर्भ में हमने माना था कि समष्टि प्रसरण हमें ज्ञात है। यह थोड़ा सा अविश्वसनीय प्रतीत होता है कि समष्टि माध्य का हमें पता नहीं (हम इसे आकलन करना चाहते हैं) और हमें समष्टि प्रसरण का पता है। इस संदर्भ में अधिक उपयुक्त धारणा तो यह होती है कि समष्टि माध्य और प्रसरण दोनों अज्ञात हैं। प्रतिदर्श माध्य और प्रसरण के आधार पर हम समष्टि माध्य के लिए विश्वस्यता स्तर ज्ञात करना चाहते हैं।

चूँकि समष्टि मानक विचलन (σ) अज्ञात है, इसलिए इसके बदले हम प्रतिदर्श मानक विचलन (s) का प्रयोग करते हैं। लेकिन ऐसे मामले में $\bar{x}$ का प्रतिदर्शी बंटन प्रसामान्य नहीं है और इसके बदले यह स्टूडेंट t बंटन का अनुसरण करता है। प्रतिदर्श माध्य की मानक त्रुटि $\frac{s}{\sqrt{n}}$ होगी।

मानक प्रसामान्य विचर की भांति, t बंटन का माध्य शून्य है और माध्य के लिए यह सममित है और इसका परिसर $-\infty$ से ∞ के बीच है। लेकिन इसका प्रसरण 1 से अधिक है। असल में इसका प्रसरण, स्वतंत्रता की कोटि के आधार पर बदलता है। लेकिन जब $n > 30$ हो तब t-बंटन का प्रसरण 1 के काफी निकट होता है और तब z-बंटन नज़र आता है।

z-प्रतिदर्शज की भांति t-प्रतिदर्शज इस प्रकार परिकलित किया जाता है।

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

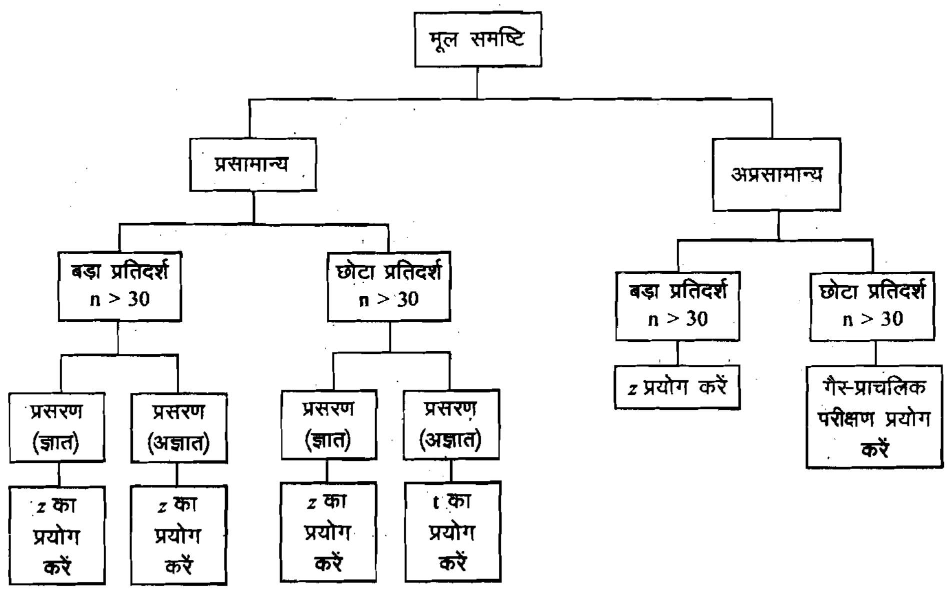
t-बंटन की क्षेत्रफल सारणी (इकाई 15 की सारणी 15.3) पर यदि निगाह डालें तो अपेक्षित विश्वस्यता स्तर के प्रायिकता मान हम प्राप्त करते हैं। अतः विश्वस्यता अंतराल होगा,

$$\bar{x} - t \cdot \frac{s}{\sqrt{n}} \le \mu \le \bar{x} + t \cdot \frac{s}{\sqrt{n}} \qquad ...(18.6)$$

उदाहरण 18.2

20 बालकों का माध्य वजन (किग्रा में) 15 और मानक विचलन 4 है। उपर्युक्त जानकारी के आधार पर ऐसी समष्टि के माध्य वजन का 95% विश्वस्यता अंतराल आकलित कीजिए जिससे हमने प्रतिदर्श लिए हैं। मान लीजिए कि समष्टि प्रसामान्य रूप से बंटित है।

चूँकि समष्टि प्रसामान्य है और प्रतिदर्श आकार छोटा है, इसलिए विश्वस्यता अंतराल के आकलन के लिए हम t-बंटन लागू करते हैं। चूँकि $n = 20$ है इसलिए स्वतंत्रता की कोटियाँ = 19 हैं। अब हम सारणी 15.3 के पहले स्तंभ से नीचे ऐसी पंक्ति की ओर बढ़ते हैं जो 19 के संगत में है। चूँकि हमें 95% विश्वस्यता अंतराल चाहिए इसलिए $t = 0$ के दोनों तरफ हमें 0.025 क्षेत्र छोड़ना होगा और पिछले अनुभाग में भी हमने ऐसा ही किया था। स्वतंत्रता की कोटियाँ = 19 और α = 0.025 के लिए हम पाते हैं कि t का मान 2.093 है।



चित्र 18.2 : उपयुक्त परीक्षण प्रतिदर्शज का चयन

बोध प्रश्न 2

1) किसी प्रतिदर्श में शामिल 50 कर्मचारियों से घर से दफ्तर तक तय की जाने वाली दूरी के बारे में पूछा जाता है तो हम पाते हैं कि समष्टि माध्य 4.5 कि.मी. है। समष्टि के लिए 95% विश्वस्यता अंतराल ज्ञात कीजिए। मान लीजिए कि समष्टि प्रसामान्य रूप से बंटित है और इसका प्रसरण 0.36 है।

....

....

....

....

....

2) किसी प्रतिदर्श में शामिल स्कूल के 25 विद्यार्थियों की माध्य ऊँचाई 95 से.मी. है और मानक विचलन 4 से.मी. है। 99% विश्वस्यता अंतराल ज्ञात कीजिए।

....

....

....

....

....

3) बताइए कि निम्नलिखित कथन सही हैं या गलत।

क) जब मूल समष्टि अप्रसामान्य नहीं हो और प्रतिदर्श का आकार छोटा हो तो विश्वस्यता अंतराल आकलन करने के लिए हम t-बंटन का प्रयोग करते हैं।

ख) t-बंटन का परिसर 0 से अनंत होता है।

ग) जब विश्वस्यता स्तर 90% हो तो सार्थकता का स्तर 10% होगा।

....

....

....

....

....

18.7 सारांश

प्रतिदर्श सूचना के आधार पर समष्टि के बारे में निष्कर्ष निकालने को सांख्यिकीय अनुमिति कहते हैं। इस संदर्भ में हमें मूलतः दो कार्य करने हैं : आकलन और परिकल्पना परीक्षण। इस इकाई में हमने मुख्यतया आकलन पर ध्यान केंद्रित किया था जबकि परिकल्पना परीक्षण की चर्चा हम आगे की इकाइयों में करेंगे।

अज्ञात प्राचल का आकल बिंदु या अंतराल, अर्थात दोनों में से कोई एक, हो सकता है। दूसरी तरफ, अंतराल आकलन में हम प्रतिदर्श माध्य के इर्द-गिर्द दो सीमाएँ (उच्च एवं निम्न) का निर्माण करते हैं। विश्वस्यता के निर्धारित स्तर को ध्यान में रखकर हम कह सकते हैं कि समष्टि माध्य, जिसका फिलहाल हमें पता नहीं है, विश्वस्यता अंतराल में बना रहेगा। विश्वस्यता अंतराल का निर्मित करने के लिए हमें समष्टि प्रसरण या इसके आकल का पता होना ज़रूरी है। जब हमें समष्टि प्रसरण का पता है तो विश्वस्यता अंतराल निर्मित करने के लिए हम प्रसामान्य बंटन लागू करते हैं। ऐसे मामले में जहाँ समष्टि प्रसरण अज्ञात हो, उपर्युकत उद्देश्य के लिए हम स्टूडेंट t का प्रयोग करते हैं। स्मरण रहे कि जब प्रतिदर्श का आकार बड़ा ($n > 30$) होता है तो t-बंटन, प्रसामान्य बंटन के सन्निकटतः होता है। अतः बडे प्रतिदर्शों के लिए यदि समष्टि प्रसरण अज्ञात होता है तो हम प्रतिदर्श माध्य और प्रतिदर्श

प्रसरण के आधार पर विश्वस्यता अंतराल ज्ञात करने के लिए प्रसामान्य बंटन का प्रयोग कर सकते हैं।

18.8 शब्दावली

विश्वस्यता स्तर (Confidence level) : प्रतिदर्शों की प्रतिशत (प्रायिकता) जहाँ समष्टि माध्य, प्रतिदर्श माध्य के इर्दगिर्द विश्वस्यता अंतराल में निहित रहता है। ों यदि α सार्थकता का स्तर (level of significance) है तो विश्वस्यता स्तर ($1-\alpha$) है।

आकलन (Estimation) : प्रतिदर्शजों के आधार पर प्राचल मानों का पूर्वानुमान लगाने की विधि।

आकलक (Estimator) : आकलन प्रमेय में प्रतिदर्शज का अन्य नाम।

प्राचल (Parameter) : समष्टि की कुछ विशेषताओं का माप।

समष्टि (Population) : किसी स्थान और समय विशेष पर किंही विशिष्ट इकाइयों का संपूर्ण संग्रहण।

यादृच्छिक प्रतिदर्श (Random Sampling) : ऐसी प्रक्रिया जिसमें समष्टि के प्रत्येक सदस्य के प्रतिदर्श के लिए चुने जाने की निश्चित संभावना और प्रायिकता होती है। इसे प्रायिकता प्रतिदर्श भी कहा जाता है।

प्रतिदर्श (Sample) : समष्टि का उप-समुच्चय। इसे प्रायिकता के नियमों को लागू करके वैज्ञानिक तरीके से समष्टि से प्राप्त किया जा सकता है जिससे व्यक्तिगत पक्षपात दूर किया जा सके। किसी समष्टि से अनेक प्रतिदर्श लिए जा सकते हैं और प्रतिदर्श ज्ञात करने की विधियाँ अनेक होती हैं।

प्रतिदर्श त्रुटि (Sampling Error) : क्ष्प्रतिचयन विधि में प्राप्त समष्टि के कुछ लक्षणों को इससे प्राप्त प्रतिदर्श के सन्निकट करने का प्रयास करते हैं। अब चूँकि प्रतिदर्श में समष्टि के सभी सदस्यों को शामिल नहीं किया जाता है, इसलिए निकटतम ही सन्निकट होता है। यह समष्टि के अपेक्षित लक्षण के समरूप नहीं होती और इसी वजह से त्रुटि हो जाती है। इस त्रुटि को प्रतिदर्शी त्रुटि कहते हैं।

सार्थकता का स्तर (Level of significance) : कुछेक प्रतिदर्श ऐसे हो सकते हैं, जिनमें समष्टि माध्य, प्रतिदर्श माध्य के इर्दगिर्द विश्वस्यता अंतराल में नहीं रहता। इस प्रकार के मामलों की प्रतिशत (प्रायिकता) सार्थकता का स्तर कहलाती है। इसे आमतौर पर α द्वारा दर्शाया जाता है। जब $\alpha = .05$ (अर्थात 5 प्रतिशत) हम कह सकते हैं कि 5% मामलों में हम गलत निर्णय पर पहुँचते हैं या प्रथम कोटि की त्रुटि करते हैं। सार्थकता का स्तर किसी भी स्तर पर हो सकता है लेकिन आमतौर पर इसे 5% या 1% स्तर माना जाता है।

प्रतिदर्शज (Statistic) : इकाइयों के मानों का ऐसा फलन जिन्हें प्रतिदर्श में शामिल किया जाता है। प्रतिदर्शज का मुख्य उद्देश्य कुछ समष्टि प्राचलों का आकलन करना है।

प्रतिदर्शी बंटन (Sampling Distribution) : प्रतिदर्शज के मानों की सापेक्षिक बारंबारता या प्रायिकता बंटन जब प्रतिदर्शों की संख्या अनंत की ओर प्रवृत्त होती है।

मानक त्रुटि (Standard Error) : प्रतिदर्शज के प्रतिदर्शी बंटन का मानक विचलन।

सांख्यिकीय अनुमिति (Statistical Inference) : किसी अज्ञात समष्टि से ज्ञात प्रतिदर्श लेकर, इनके बारे में निष्कर्ष निकालने की प्रक्रिया।

आकलन की समस्या (Problem of Estimation) : समष्टि की किसी ऐसी विशेषता को जानना जिसके बारे में पूरी तरह से कुछ नहीं जानते और समष्टि से लिए यादृच्छिक प्रतिदर्श के आधार पर हम कोई विशेष अनुमान लगाना चाहते हैं। सांख्यिकीय अनुमिति की यह समस्या आकलन की समस्या कहलाती है।

18.9 कुछ उपयोगी पुस्तकें

Nagar, A. L. and Das, R. K., 1989, *Basic Statistics:* Oxford University Press, Delhi, Chapter 9.

Newbold, P., 1991. *Statistics for Business and Economics* (Third Edition): Prentice Hall, New Jersey, Chapters 6, 7, 8 and 9.

Keller, G, and B. Warrack, 1991, *Essentials of Business Statistics*, Wordsworth Publishing Co., California, Chapters 7 and 8.

18.10 बोध प्रश्नों के उत्तर

बोध प्रश्न 1

1) अनुभाग 18.2 का भलीभांति अध्ययन करें और उत्तर दें।

2) क) सही ख) सही ग) सही घ) सही

बोध प्रश्न 2

1) चूँकि यह बड़ा प्रतिदर्श है इसलिए हम z-प्रतिदर्शज का प्रयोग करते हैं। विश्वस्यता अंतराल है $4.40 \leq \mu \leq 4.60$

2) चूँकि यह छोटा प्रतिदर्श है और समष्टि प्रसरण नहीं दिया गया है इसलिए हम t प्रतिदर्शज का प्रयोग करते हैं ($t = 2.49$)। विश्वस्यता अंतराल है $93.01 \leq \mu \leq 96.99$ ।

3) क) गलत ख) गलत ग) सही

इकाई 19 परिकल्पना परीक्षण

इकाई की रूपरेखा

19.0 उद्देश्य

इस इकाई को पढ़ने के बाद आप आपः

- निराकरणीय परिकल्पना और वैकल्पिक परिकल्पना की अवधारणाओं का वर्णन कर सकेंगे;
- सार्थकता के स्तर के आधार पर निर्णायक क्षेत्र की पहचान कर सकेंगे;
- प्रथम कोटि और द्वितीय कोटि की त्रुटियों के बीच के अंतर को स्पट कर सकेंगे;
- एकल प्रतिदर्श के आधार पर समष्टि माध्य से संबंधित परिकल्पना-परीक्षण कर सकेंगे; और
- दो प्रतिदर्शों से प्राप्त प्रतिदर्श माध्यों के बीच के अंतर के लिए परीक्षण कर सकेंगे।

19.1 प्रस्तावना

पिछली इकाई में हमने प्रतिदर्श आँकड़ों के आधार पर समष्टि माध्य के लिए विश्वस्यंता अंतराल का आकलन करना सीखा था। इस इकाई में हम सांख्यिकीय अनुमिति के दूसरे

पहलू अर्थात परिकल्पना परीक्षण पर ध्यान केंद्रित करेंगे। परिकल्पना प्राचल के बारे में किया जाने वाला एक दावा है। जैसे मान लीजिए कि हमें पता चलता है कि छत्तीसगढ़ राज्य की प्रति व्यक्ति आय प्रति वर्ष 20,000 रुपए है। यदि हमें इस राज्य के समग्र परिवारों की संपूर्ण जनगणना का पता है तो हम उपर्युक्त कथन को सही मान सकते हैं। इसे पता चलता है कि हमने छत्तीसगढ़ राज्य के सभी परिवारों की आमदनी पर आँकड़े इकट्ठे किए हैं और राज्य की प्रति व्यक्ति आय को परिकलित किया है। लेकिन, समय, धन और जनशक्ति जैसे अवरोध, प्रतिदर्श सर्वेक्षण में बाधा डालते हैं और इसी से प्रतिदर्श सूचना के आधार पर हम ऐसे कथन के बारे में निष्कर्ष निकालते हैं। उपर्युक्त अध्ययन की प्रक्रिया, परिकल्पना-परीक्षण की विषयवस्तु है।

परिकल्पना परीक्षण का विविध क्षेत्रों और विविध स्थितियों में विस्तृत प्रयोग किया जाता है। जैसे, मान लीजिए हमें क्षय रोग के निवारण में प्रयुक्त नई दवा की प्रभाविता की जाँच करनी है। इस संदर्भ में ज़रूरी नहीं है कि इस दवा की प्रभाविता देखने के लिए क्षय रोग से पीड़ित सभी रोगियों को नयी दवा दी जाए। यहाँ हमारा काम है प्रतिनिधि प्रतिदर्श की प्राप्त करना और परीक्षण करना कि क्या नई दवा वर्तमान दवाइयों से अधिक कारगर है। आइए, एक अन्य उदाहरण लें : योजनाकार का मानना है कि बिहार और राजस्थान में अशोधित जन्म दर एक समान है। इस मामले में पिछले वर्ष के दौरान बिहार और राजस्थान में सभी जन्म लेने वाले शिशुओं की गणना करके, अशोधित जन्म दर परिकलित करना शायद संभव नहीं होगा। इस स्थिति में भी प्रतिदर्श सर्वेक्षण करके, योजनाकार द्वारा प्राप्त निष्कर्ष का परीक्षण किया जाता है।

परिकल्पना परीक्षण में हम निम्नलिखित प्रकार के प्रश्नों के उत्तर देने का प्रयास करते हैं: क्या विचाराधीन प्रतिदर्श किसी विशिट समष्टि से निकाला गया है? क्या दो प्रतिदर्शों के बीच का अंतर इतना पर्याप्त रूप से महत्वपूर्ण है कि वे एक ही समष्टि से संबंधित नहीं हो सकते?

19.2 परिकल्पना का निरूपण

समष्टि के अभिलक्षण के बारे में परिकल्पना एक अस्थायी कथन है। जैसे हाल ही के वर्षों के सरकारी आँकड़े दर्शाते हैं कि उड़ीसा में महिला साक्षरता 51 प्रतिशत है। यहाँ महिला साक्षरता की दर के बारे में एक कथन या दावा प्रस्तुत किया गया है। अतः इसे हम एक परिकल्पना मान सकते हैं।

परिकल्पना परीक्षण में चार महत्वपूर्ण घटक शामिल हैं : i) निराकरणीय परिकल्पना, ii) वैकल्पिक परिकल्पना, iii) परीक्षण प्रतिदर्श, और iv) निष्कर्षों की व्याख्या। हम इन सभी के विषय में चर्चा करेंगे।

आमतौर पर सांख्यिकीय परिकल्पना को H वर्ण से दर्शाया जाता है। परिकल्पना दो तरह की होती है: निराकरणीय परिकल्पना और वैकल्पिक परिकल्पना। निराकरणीय परिकल्पना ऐसा कथन है जिसे हम समष्टि के संदर्भ में सत्य मानते हैं और परीक्षण प्रतिदर्शज द्वारा इसका परीक्षण करते हैं। इस निराकरणीय परिकल्पना को H_0 द्वारा दर्शाया जाता है। उड़ीसा में महिला साक्षरता पर आधारित हमारे उदाहरण में, हमारी निराकरणीय परिकल्पना है :

$$H_0 : \mu = 0.51 \qquad ...(19.1)$$

जहाँ μ प्राचल है, इस मामले में उड़ीसा में महिला साक्षरता दर संभावना है कि निराकरणीय परिकल्पना जिसका परीक्षण हम करना चाहते हैं, सही नहीं है और महिला

साक्षरता 51 प्रतिशत के बराबर नहीं है। अतः ऐसी वैकल्पिक परिकल्पना की रचना की आवश्यकता है जो निराकरणीय परिकल्पना के असत्य होने पर खरी उतरे। वैकल्पिक परिकल्पनाओं को हम संकेत से दर्शाते हैं। और इसे इस तरह सूत्रबद्ध करते हैं :

$$H_A : \mu \neq 0.51 \qquad ...(19.2)$$

हमें ध्यान में रखना होगा कि निराकरणीय परिकल्पना और वैकल्पिक परिकल्पना परस्पर अपवर्जी हैं अर्थात दोनों एक-साथ सत्य नहीं हो सकतीं। दूसरा, H_0 और H_A में कुल मिलाकर प्राचल के संदर्भ में सभी संभावित विकल्प समाहित हैं, अर्थात, इस संदर्भ में तीसरी संभावना नहीं हो सकती। जैसे, उड़ीसा में महिला साक्षरता के मामले में दो संभावनाएँ हैं: साक्षरता का दर 51 प्रतिशत होना या 51 प्रतिशत न होना अर्थात यहाँ कोई तीसरी संभावना नहीं है।

यह एक विरल संयोग है कि प्रतिदर्श माध्य ($\bar{x}$), समष्टि माध्य (μ) के बराबर हो। अधिकांश मामलों में हम $\bar{x}$ और μ के बीच अंतर पाते हैं। क्या यह अंतर प्रतिचयन उच्चावचन की वजह से है या सही मायने में प्रतिदर्श और समष्टि के बीच कोई अंतर है। इस प्रश्न का उत्तर देने के लिए हमें दोनों के बीच के अंतर के परीक्षण के लिए परीक्षण प्रतिदर्शज की आवश्यकता है। परीक्षण प्रतिदर्शज के प्रयोग से हमें जो परिणाम प्राप्त होगा, उसे समझाने की आवश्यक है और निराकरणीय परिकल्पना को अस्वीकार करना है या निराकृत करना है, इस संदर्भ में यह निर्णय लेने की भी आवश्यकता है।

परिकल्पना परीक्षण के लिए परीक्षण प्रतिदर्शज के संवर्धन और परिणामों की व्याख्या को खुलकर समझाने की आवश्यकता है। इन दो चरणों पर आगे चर्चा करने से पहले हम एक अन्य अवधारणा — निर्णायक क्षेत्र पर प्रकाश डालते हैं।

19.3 अस्वीकृति क्षेत्र एवं त्रुटियों के प्रकार

परिकल्पना परीक्षण और (पिछली इकाई में चर्चित) अंतराल आकलन एक जैसे ही विचार सूत्र पर आधारित हैं। इकाई 18 को ध्यान में लाइए जहाँ हमने अध्ययन किया था कि कुछ निश्चित विश्वस्यता स्तर के साथ प्रतिदर्श माध्य के इर्द-गिर्द विश्वस्यता अंतराल निर्मित किया जाता है। 95 प्रतिशत के विश्वस्यता स्तर का अर्थ है कि 95 प्रतिशत मामलों में प्रतिदर्श माध्य से आकलित विश्वस्यता अंतराल में समष्टि माध्य रहेगा। यहाँ स्पष्ट नहीं होगा कि 5 प्रतिशत मामलों में समष्टि माध्य विश्वस्यता अंतराल के भीतर बना रहेगा या नहीं। ध्यान दीजिए कि जब समष्टि माध्य विश्वस्यता अंतराल के भीतर नहीं रहता तो हमें निराकरणीय परिकल्पना को अस्वीकृत कर देना चाहिए।

19.3.1 बड़े प्रतिदर्शों के लिए अस्वीकृति क्षेत्र

आइए, बड़े प्रतिदर्शों के लिए निर्णायक क्षेत्र (critical region) की अवधारणा का वर्णन करें। इसके बाद हम छोटे प्रतिदर्शों तक अपनी संकल्पना का विस्तार करेंगे।

जैसा कि पिछली इकाइयों में हम पहले ही अध्ययन कर चुके हैं, प्रतिदर्श माध्य ($\bar{x}$) का प्रतिदर्शी बंटन प्रसामान्य बंटन का अनुसरण करता है और जिसका माध्य μ और मानक विचलन $\frac{\sigma}{\sqrt{n}}$ है। अतः $\bar{x}$ को मानक प्रसामान्य विचर, z में परिवर्तित किया जा सकता है। वह प्रसामान्य बंटन का अनुसरण करेगा और जिसका माध्य $= 0$ और मानक विचलन $= 1$ हो।

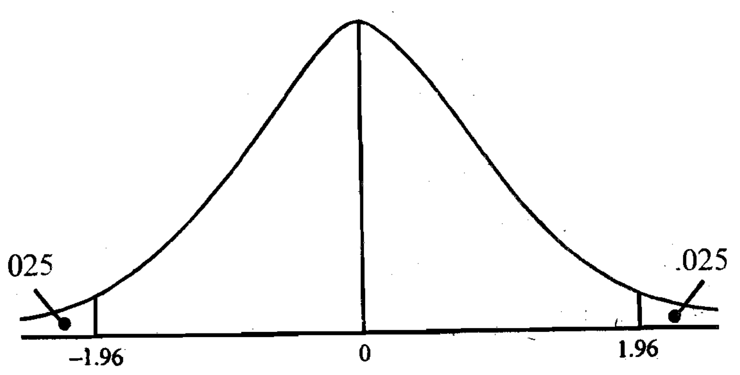
सांकेतिक रूप से $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ और $z \sim N(0,1)$। इकाई 15 में हमने सीखा है कि मानक प्रसामान्य वक्र के नीचे का क्षेत्र, z द्वारा माने गए विविध प्रकार के मानों के लिए प्रायिकता देता है। इन प्रायिकताओं को सारणी के रूप में भी प्रस्तुत किया जा सकता है (देखें खंड 5 की इकाई 15 की सारणी 15.1)।

आइए, चित्र 19.1 में प्रस्तुत मानक प्रसामान्य चर पर ध्यान केंद्रित करें, जहाँ x-अक्ष, z चर को और y-अक्ष, z की प्रायिकता अर्थात $p(z)$ को दर्शाते हैं। हमें निम्नलिखित बातों को ध्यान में रखना चाहिए।

- जब प्रतिदर्श माध्य समष्टि माध्य के बराबर हो (अर्थात $\bar{x} = \mu$), तब $z = 0$ होगा। जब $\bar{x} > \mu$ तब z सकारात्मक होगा। दूसरी तरफ जब $\bar{x} < \mu$ तब z नकारात्मक होगा।
- यहां हम $\bar{x}$ और μ के बीच के अंतर पर ध्यान केंद्रित कर रहे हैं। इसलिए z का संकेत नकारात्मक हो या धनात्मक हो, इससे कोई फर्क नहीं पड़ता।
- $\bar{x}$ और μ के बीच का अंतर जितना उच्च होगा, z का निरपेक्ष मान (absolute value) भी उतना ही उच्च होगा। अतः z-मान, $\bar{x}$ और μ के बीच के अन्तर का पता लगाता है और इसलिए परीक्षण प्रतिदर्शज के रूप में इसे प्रयुक्त किया जा सकता है।
- हमें z के निर्णायक मान (critical value) का पता होना चाहिए क्योंकि इसके परे $\bar{x}$ और μ के बीच का अंतर विशेष महत्व रखता है।
- यदि z का निरपेक्ष मान, निर्णायक मान से निम्न है, तो हमें निराकरणीय परिकल्पना को अस्वीकृत नहीं करना चाहिए।

अतः बड़े प्रतिदर्शों के मामले में z के निरपेक्ष मान को परिकल्पना परीक्षण के लिए परीक्षण प्रतिदर्शज के रूप में लिया जा सकता है, जैसे कि

$$z = \frac{|\bar{x} - \mu|}{\sigma/\sqrt{n}} \qquad ...(19.3)$$



चित्र 19.1 : अस्वीकृति क्षेत्र

आइए, चित्र 19.1 में दिए गए मानक प्रसामान्य वक्र के माध्यम से अस्वीकृति क्षेत्र (rejection region) की संकल्पना का वर्णन करें। जब हमारे पास 95% का विश्वस्यता गुणांक हो तो मानक प्रसामान्य वक्र के नीचे का आच्छादित क्षेत्र 95% है। अतः वक्र के नीचे का 95% क्षेत्र $-1.96 \le z \le 1.96$ से बद्ध है। बाकी का 5% क्षेत्र $z \le -1.96$ और $z \ge 1.96$ से आच्छादित है। अतः मानक प्रसामान्य वक्र के दोनों तरफ के 2.5% क्षेत्रांश अस्वीकृति क्षेत्र को रचना करते है। इस क्षेत्र को चित्र 19.1 में दर्शाया गया है। यदि प्रतिदर्श माध्य अस्वीकृति क्षेत्र में आ जाता है तो हम निराकरणीय परिकल्पना को अस्वीकृत करते हैं।

19.3.2 एकपुच्छ एवं द्विपुच्छ परीक्षण

चित्र 19.1 में हमने मानक प्रसामान्य वक्र के दोनों तरफ अस्वीकृति क्षेत्र को दर्शाया है। हालांकि, बहुत से मामलों में हम मानक प्रसामान्य वक्र के (बायें या दायें) अर्थात किसी भी एक तरफ अस्वीकृति क्षेत्र बना सकते हैं।

याद रखिए यदि α सार्थकता का स्तर है, तब द्विपुच्छ परीक्षण (two-tail test) में $\frac{\alpha}{2}$ क्षेत्र के लिए, इसे मानक प्रसामान्य वक्र के दोनों तरफ रखा जाता है। लेकिन यदि यह एकपुच्छ परीक्षण (one-tail test) है, तब α क्षेत्र को मानक प्रसामान्य वक्र के एक-तरफ ही रखा जाता है। अतः एकपुच्छ और द्विपुच्छ परीक्षण के लिए निर्णायक मान एक-दूसरे से अलग होते हैं।

एकपुच्छ या द्विपुच्छ परीक्षण का चयन वैकल्पिक परिकल्पना के रचनासूत्र पर निर्भर करता है। जब वैकल्पिक परिकल्पना $H_A: \bar{x} \ne \mu$ प्रकार की है तो हम द्विपुच्छ परीक्षण करते हैं क्योंकि $\bar{x}, \mu$ से बड़ा या छोटा हो सकता है। दूसरी तरफ यदि वैकल्पिक परिकल्पना, $H_A: \bar{x} < \mu$ प्रकार की है तो समूचा अस्वीकृति क्षेत्र, मानक प्रसामान्य वक्र के बायें तरफ होगा। इसी तरह यदि वैकल्पिक परिकल्पना $H_A: \bar{x} > \mu$ प्रकार की है तो समूचा अस्वीकृति क्षेत्र, मानक प्रसामान्य वक्र के दायें तरफ होगा। यहाँ एकपुच्छ परीक्षण का प्रयोग होगा।

z के निर्णायक मान, सार्थकता के स्तर पर निर्भर करते हैं। सारणी 19.1 में सार्थकता (α) के निर्धारित स्तरों के लिए, इन निर्णायक मानों को प्रसामान्य बंटन की अभिधारणा के अंतर्गत किए जाने वाले परीक्षणों के लिए, दिया गया है। ये मान द्वि-पुच्छ और एकपुच्छ परीक्षणों के लिए दिए गए हैं।

सारणी 19.1 : z-प्रतिदर्शज संबंधी निर्णायक मान

सार्थकता का स्तर (α)	0.10	0.05	0.01	0.005
द्वि-पुच्छ परीक्षण	1.65	1.96	2.58	2.81
एक-पुच्छ परीक्षण	1.28	1.65	2.33	2.58

नोट : यह सारणी, सारणी 15.2 से व्युत्पन्न है।

19.3.3 प्रथम कोटि और द्वितीय कोटि की त्रुटियां

परिकल्पना परीक्षण में हम विश्वस्यता की निश्चित कोटि पर परिकल्पना को अस्वीकृत करते हैं या नहीं करते। जैसा कि आपको पता है 0.95 के विश्वस्यता गुणांक का अर्थ है कि 100 प्रतिदर्शों में से 95% भाग में प्राचल, स्वीकृति क्षेत्र के भीतर रहता है जबकि बाकी के 5% भाग में प्राचल, अस्वीकृति क्षेत्र में रहता है। अतः इन 5% मामलों में, प्रतिदर्श की प्राप्ति भले ही समष्टि से की जाती है किन्तु प्रतिदर्श माध्य, समष्टि माध्य से बहुत दूर रहता है। ऐसे मामलों में प्रतिदर्श समष्टि से संबद्ध होता है, लेकिन हमारी परीक्षण प्रक्रिया इसे अस्वीकार

कर देती है। अर्थात H_0 सही है लेकिन इसे अस्वीकृत कर दिया जाता है। इसे 'प्रथम कोटि की त्रुटि' (type I errer) कहते हैं। इसी तरह ऐसी भी स्थितियाँ हो सकती हैं जब H_0 सत्य नहीं होता लेकिन प्रतिदर्श सूचना के आधार पर हम उसे अस्वीकार नहीं कर पाते। निर्णय लेने में ऐसी त्रुटि को 'द्वितीय कोटि की त्रुटि' (type II errer) कहते हैं (देखें सारणी 19.2)।

ध्यान दीजिए कि प्रथम कोटि त्रुटि से पता चलता है कि किस सीमा तक गलती सहनीय होती है। प्रथम कोटि त्रुटि, सार्थकता के स्तर के बराबर होती है और इसे α से दर्शाया जाता है। याद रखिए कि विश्वस्यता गुणांक $1-\alpha$ के बराबर है।

सारणी 19.2 : त्रुटियों के प्रकार

	H_0 सही	H_0 सही नहीं
H_0 अस्वीकृत	प्रथम कोटि की त्रुटि	सही निर्णय
H_0 स्वीकृत	सही निर्णय	द्वितीय कोटि की त्रुटि

19.3.4 छोटे प्रतिदर्शों के लिए अस्वीकति क्षेत्र

आइए, इकाई 18 के चित्र 18.2 पर दुबारा निगाह डालें जहाँ अंतराल आकलन में उचित परीक्षण प्रतिदर्श के प्रयोग के लिए हमने कुछ मानदंड दिए हैं। जैसा कि छोटे प्रतिदर्शों के मामले में ($n \le 30$), यदि समष्टि मानक विचलन ज्ञात हो तो परिकल्पना परीक्षण के लिए हम z-प्रतिदर्शज लागू करते हैं। दूसरी ओर यदि समष्टि मानक विचलन अज्ञात हो तो हम t-प्रतिदर्शज का प्रयोग करते हैं। परिकल्पना परीक्षण में भी हम इन्हीं मानदंडों का अनुसरण हैं।

छोटे प्रतिदर्शों के मामले में, यदि समष्टि मानक विचलन ज्ञात हो तो परीक्षण प्रतिदर्शज है:

$$z = \frac{|\bar{x}-\mu|}{\sigma/\sqrt{n}} \qquad \text{...(19.4)}$$

दूसरी तरफ, यदि समष्टि मानक विचलन अज्ञात हो तो तो परीक्षण प्रतिदर्शज है:

$$t = \frac{|\bar{x}-\mu|}{s/\sqrt{n}} \qquad \text{...(19.5)}$$

टी-बंटन के मामले में, हालांकि वक्र के नीचे का क्षेत्र (जिसे प्रायिकता कहते हैं) स्वतंत्रता की कोटियों के अनुसार बदल जाता है। अतः t का निर्णायक मान ज्ञात करते समय हमें स्वतंत्रता की कोटियों को भी ध्यान में रखना चाहिए। जब प्रतिदर्श आकार n है तो स्वतंत्रता की कोटि $n-1$ होगी। अतः t का क्रांतिक मान ज्ञात करते समय हमें दो बातें अवश्य ध्यान में रखनी चाहिए, (i) सार्थकता का स्तर और (ii) स्वतंत्रता की कोटि।

बोध प्रश्न 1

1) निम्नलिखित के बीच के अंतर को स्पष्ट कीजिए:

   क) निराकरणीय परिकल्पना और वैकल्पिक परिकल्पना

   ख) एक-पुच्छ और द्वि-पुच्छ परीक्षण

ग) विश्वस्यता का स्तर और और सार्थकता का स्तर

घ) प्रथम कोटि और द्वितीय कोटि की त्रुटियाँ

....

....

....

....

....

2) मान लीजिए 100 विद्यार्थियों के प्रतिदर्श की माध्य आयु 12.5 वर्ष की है। 5% के सार्थकता स्तर पर अस्वीकृति क्षेत्र के आरेख को दर्शाते हुए परिकल्पना परीक्षण कीजिए कि प्रतिदर्श की माध्य आयु, समष्टि आयु से अधिक है। मान लीजिए कि समष्टि आयु और मानक विचलन क्रमशः 10 वर्ष और 2 वर्ष है।

....

....

....

....

....

....

19.4 एकल प्रतिदर्श संबंधी परिकल्पना परीक्षण

बहुत सी स्थितियों में हमें पता लगाना होता है कि क्या प्रतिदर्श मुख्यतया प्राप्त समष्टि से महत्त्वपूर्ण रूप से भिन्न है या नहीं। जैसे, मान लीजिए हमने छत्तीसगढ़ के रायगढ़ जिले में 400 परिवारों के प्रतिदर्श का सर्वेक्षण किया और इन परिवारों की प्रति व्यक्ति आय को परिकलित किया। जिसके फलस्वरूप हमारा कार्य परिकल्पना परीक्षण करना है कि क्या प्रतिदर्श से परिकलित प्रति व्यक्ति आय, जिले की प्रति व्यक्ति आय से अलग तो नहीं है।

उपर्युक्त उदाहरण में हमारे सम्मुख दो विभिन्न स्थितियाँ हैं: (i) समष्टि (इस मामले में जिले के सभी परिवार) प्रसरण ज्ञात है, (i) हमें समष्टि प्रसरण ज्ञात नहीं है। हम प्रत्येक मामले में शामिल चरणों का वर्णन निम्न प्रकार से करते हैं ।

19.4.1 समष्टि प्रसरण ज्ञात है

आइए, एक ऐसे मामले पर विचार करें जहाँ हमें छत्तीसगढ़ के रायगढ़ जिले की प्रति व्यक्ति आय और इसके प्रसरण का पता है। मान लीजिए कि सरकारी रिकार्डों में उपलब्ध आंकड़े दर्शाते हैं कि रायगढ़ जिले की प्रति व्यक्ति आय 10,000 रु. और प्रति व्यक्ति आय का मानक विचलन 1500 रु. है। हालांकि हमने 400 परिवारों का प्रतिदर्श सर्वेक्षण किया था और पाया था कि इनकी प्रति व्यक्ति आय 10,500 रु. है। क्या सरकारी रिकार्डों में प्रदत्त आंकड़े हमें स्वीकृत होने चाहिए?

इस मामले में μ = रु. 10000

σ = रु. 1500

$$\bar{x} = \text{रु. } 10500$$
$$n = 400$$

केंद्रीय सीमा प्रमेय से हमें पता चलता है कि जब प्रतिदर्श आकार बड़ा होगा तो प्रतिदर्श माध्य सन्निकटतःप्रसामान्य रूप से बंटित होगा। यह बात ऐसे मामलों में तो सत्य होती है जहाँ मूल समष्टि प्रसामान्य रूप से बंटित है। अतः यह उदाहरण, प्रसामान्य बंटन के अनुप्रयोग के उपयुक्त है।

इस मामले में हमारी निराकरणीय परिकल्पना है

$$H_0 : \bar{x} = \mu$$

निराकरणीय परिकल्पना से पता चलता है कि प्रतिदर्श माध्य, समष्टि माध्य के बराबर है। अन्य शब्दों में प्रतिदर्श से प्राप्त प्रति व्यक्ति आय वही है जैसा कि सरकारी रिकर्डों में वर्णित है। हमारी वैकल्पिक परिकल्पना है:

$$H_A : \bar{x} \neq \mu$$

मान लीजिए हमारे पास बताने के लिए कोई कारण नहीं है कि प्रतिदर्श ($\bar{x}$) से प्राप्त **प्रति** व्यक्ति आय, सरकारी आंकड़ों में उपलब्ध प्रति व्यक्ति आय से अधिक है या कम। अतः हमारी वैकल्पिक परिकल्पना है कि $\bar{x}$, μ के किसी भी तरफ हो सकता है। इसलिए, हमें द्वि-पुच्छ परीक्षण करना होगा ताकि अस्वीकृति क्षेत्र, मानक प्रसामान्य वक्र के दोनों तरफ हो और परीक्षण प्रतिदर्शज है :

$$z = \frac{|\bar{x} - \mu|}{\sigma / \sqrt{n}} \qquad \ldots(19.5)$$

उपर्युक्त में मानों के प्रतिस्थापन पर, हम पाते हैं:

$$z = \frac{|10500 - 10000|}{15000/\sqrt{400}} = \frac{500}{500/20} = \frac{500}{75} = 6.67$$

मानक प्रसामान्य वक्र और z के विविध मानों से संबंधित क्षेत्र को ध्यान में लाइए (तालिका 15 में सारणी 15.1 देखें)। हम देखते हैं कि जब $z = 1.96$ हो तो मानक प्रसामान्य वक्र के नीचे का आच्छादित क्षेत्र 0.4750 होगा। **इसलिए सार्थकता का स्तर 5 प्रतिशत है। इसी तरह** जब $z = 2.58$ **तब मानक प्रसामान्य वक्र के नीचे का आच्छादित क्षेत्र 0.495 है, इसलिए सार्थकता का स्तर 1 प्रतिशत** है।

उपर्युक्त मामले में चूँकि $z = 6.67$ है, इसलिए प्रतिदर्श, अस्वीकृति क्षेत्र में निहित है और हम परिकल्पना को अस्वीकृत कर देते हैं। अतः प्रतिदर्श से प्राप्त प्रति व्यक्ति आय सरकारी रिकॉर्डों में प्रदत्त **प्रति** व्यक्ति आय से महत्वपूर्ण रूप से भिन्न है। अनुसरण संबंधी चरण हैं:

1) निराकरणीय परिकल्पना का विशेष रूप से उल्लेख कीजिए।

2) पता लगाइए कि क्या इसमें एक-पुच्छ या द्वि-पुच्छ परीक्षण की आवश्यकता है। इसी आधार पर अस्वीकृति क्षेत्र की पहचान कीजिए। वैकल्पिक परिकल्पना के विस्तृत ब्यौरे में यह सहायक होगा।

3) (19.5) में z-प्रतिदर्श के लिए प्रतिदर्श मानों का प्रयोग कीजिए।

4) सार्थकता के स्तर के अनुरूप z-सारणी से निर्णायक मान का पता लगाइए।

5) यदि निर्णायक मान से निम्न मान की प्राप्ति होती है तो निराकरणीय परिकल्पना को अस्वीकृत नहीं कीजिए।

6) यदि निर्णायक मान से बड़े मान की प्राप्ति होती है तो निराकरणीय परिकल्पना को अस्वीकृत करें और वैकल्पिक परिकल्पना को स्वीकार करें।

उदाहरण 19.1

मान लीजिए किसी विशेष ब्रांड की बैटरी से जनित वोल्टेज प्रसामान्य रूप से बंटित है। ऐसी 100 बैटरियों के यादृच्छिक प्रतिदर्श का परीक्षण किया गया और 1.4 वोल्ट की माध्य वोल्टेज पाई गई। 0.01 के सार्थकता के स्तर पर क्या इससे पता चलता है कि इन बैटरियों की औसत वोल्टेज 1.5 वोल्ट से भिन्न है? मान लीजिए समष्टि मानक विचलन 0.21 वोल्ट है।

यहाँ $H_0 : \mu = 1.5$

चूँकि प्रतिदर्श की औसत वोल्टेज, समष्टि की औसत वोल्टेज से भिन्न हो सकती है– यह 1.5 वोल्ट से कम हो या अधिक। हमारा अस्वीकृति क्षेत्र, प्रसामान्य वक्र के दोनों तरफ है। अतः यह द्वि-पुच्छ परीक्षण का मामला है और वैकल्पिक परिकल्पना है।

$H_A : \mu \neq 1.5$

चूंकि समष्टि मानक विचलन σ ज्ञात है और परीक्षण प्रतिदर्शज है:

$$z = \frac{|\bar{x}-\mu|}{\frac{\sigma}{\sqrt{n}}} = \frac{|1.4-1.5|}{\frac{0.21}{\sqrt{100}}} = 4.8$$

मानक प्रसामान्य वक्र के नीचे के क्षेत्र से संबंधित सारणी से हम पाते हैं कि 1% सार्थकता के स्तर पर निर्णायक मान 2.58 है। चूँकि z का प्रेक्षित मान, 2.58 से बड़ा है, इसलिए हम 1% स्तर पर निराकरणीय परिकल्पना को अस्वीकृत करते हैं और वैकल्पिक परिकल्पना को स्वीकार करते हैं कि बैटरियों की औसत वोल्टेज 1.5 वोल्ट से भिन्न है।

19.4.2 समष्टि प्रसरण अज्ञात हों

यह अभिधारणा है कि समष्टि मानक विचलन (σ) हमें ज्ञात है, अवास्तविक नज़र आती है। क्योंकि हमें स्वयं समष्टि माध्य का पता नहीं है। जब σ अज्ञात हो तो हमें प्रतिदर्श मानक विचलन (s) द्वारा इनका अनुमान लगाना पड़ता है। ऐसी स्थितियों में प्रतिदर्श आकार के आधार पर दो संभावनाएँ नज़र आती हैं। यदि प्रतिदर्श आकार बड़ा है $(n > 30)$, तब हम z- प्रतिदर्श लागू कर सकते हैं अर्थात

$$z = \frac{|\bar{x}-\mu|}{s/\sqrt{n}} \qquad \ldots(19.6)$$

यदि प्रतिदर्श आकार छोटा है $(n \leq 30)$ तो हम t प्रतिदर्शज लागू करते हैं जहाँ स्वतंत्रता की कोटि $n-1$ है। परीक्षण प्रतिदर्शज है:

$$t = \frac{|\bar{x}-\mu|}{s/\sqrt{n}} \qquad \ldots(19.7)$$

अनुरक्षण करने योग्य चरण हैं:

1) निराकरणीय परिकल्पना का विशेष रूप से उल्लेख कीजिए।

2) पता लगाइए कि क्या एक-पुच्छ या द्वि-पुच्छ परीक्षण करना ज़रूरी है। इसी आधार पर मानक प्रसामान्य वक्र में अस्वीकृति क्षेत्र की पहचान कीजिए। वैकल्पिक परिकल्पना को स्पष्ट करने में यह सहायक होगा।

3) जाँच कीजिए कि प्रतिदर्श आकार बड़ा $(n > 30)$ है या छोटा $(n \leq 30)$।

4) यदि $n > 30$ है, तो (19.6) को ध्यान में रखकर z-प्रतिदर्शज लागू कीजिए।

5) सार्थकता के स्तर (α) के आधार पर z-सारणी से निर्णायक मान को ज्ञात कीजिए।

6) यदि $n < 30$ है, तो (19.7) को ध्यान में रखकर t- प्रतिदर्शज लागू कीजिए।

7) $n-1$ (स्वतंत्रता की कोटि) और सार्थकता के स्तर (α) के लिए t-सारणी से निर्णायक मान का पता लगाइए।

8) यदि निर्णायक मान से छोटा मान प्राप्त होता है तो निराकरणीय परिकल्पना को अस्वीकृत न कीजिए।

9) यदि निर्णायक मान से बड़ा मान प्राप्त होता है तो निराकरणीय परिकल्पना को अस्वीकृत कीजिए और वैकल्पिक परिकल्पना को अपनाइए।

उदाहरण 19.2

मान लीजिए दवा की गोली में एस्प्रीन की औसतन 10 मि.ग्रा. मात्रा शामिल है। 100 गोलियों का यादृच्छिक प्रतिदर्श 10.2 मि.ग्रा. मात्रा का माध्य एस्प्रीन दर्शाता है और मानक विचलन 1.4 मि.ग्रा. है। क्या 0.05 सार्थकता के स्तर पर आप निष्कर्ष निकाल सकते हैं कि माध्य एस्प्रीन की मात्रा वास्तव में 10 मि.ग्रा. ही है?

यहाँ निराकरणीय परिकल्पना है: $H_0: \mu = 10$

अस्वीकृति क्षेत्र, 10 मि.ग्रा. के दोनों तरफ है। अतः यहाँ द्वि-पुच्छ परीक्षण करना होगा और $H_A: \mu \neq 10$

इसके अलावा प्रतिदर्श माध्य $\bar{x} = 10.2$ है और प्रतिदर्श आकार $n = 100$। चूँकि समष्टि मानक विचलन अज्ञात है, इसलिए इसका अनुमान हम प्रतिदर्श मानक विचलन s से लगाते हैं और हमारा परीक्षण प्रतिदर्शज है : $z = \dfrac{|\bar{x} - \mu|}{s/\sqrt{n}}$ प्रतिदर्श से प्रासंगिक मानों को लागू करके, हम प्राप्त करते हैं:

$$z = \frac{|\bar{x} - \mu|}{\dfrac{s}{\sqrt{n}}} = \frac{|10.2 - 10|}{\dfrac{1.4}{\sqrt{100}}} = 1.43$$

5% के सार्थकता के स्तर पर z का निर्णायक मान 1.96 है। z का मान, जिसकी प्राप्ति हमने की है, 1.96 से कम है। इसलिए हम निराकरणीय परिकल्पना को अस्वीकृत नहीं कर सकते। अतः एस्प्रीन का माध्य स्तर 10 मि.ग्रा. है।

उदाहरण 19.3

हरिपुरा जिले की समष्टि की माध्य जीवन प्रत्याशा 60 र्वा है। जिले में स्वास्थ्य देखभाल संबंधी कुछ विशेष उपायों को अपनाया गया जिसके फलस्वरूप 25 व्यक्तियों का यादृच्छिक प्रतिदिर्श 60.5 वर्षों की औसतन जीवन प्रत्याशा को दर्शाता है और इसका मानक विचलन 2 वर्ष है। क्या हम 0.05 प्रतिशत सार्थकता के स्तर पर निष्कर्ष निकाल सकते हैं कि जिले की औसतन जीवन प्रत्याशा वास्तव में बढ़ गई है?

यहाँ $H_0 : \mu = 60$

जीवन प्रत्याशा में बढ़ोत्तरी के लिए हमें परीक्षण करना होगा। अतः यह एक-पुच्छ परीक्षण का मामला है और अस्वीकृति क्षेत्र, मानक प्रसामान्य वक्र के दायें हाथ (पुच्छ) पर होगा।

अतः हमारी वैकल्पिक परिकल्पना है:

$H_A : \mu = 60$

यहाँ समष्टि मानक विचलन σ अज्ञात है और हम प्रतिदर्श मानक विचलन s द्वारा इसका आकलन कर सकते हैं। यहाँ प्रतिदर्श का आकार छोटा है। अतः (19.7) को ध्यान में रखकर हमें t-प्रतिदर्शज को लागू करना होगा।

$$t = \frac{|\bar{x} - \mu|}{\frac{s}{\sqrt{n}}} = \frac{|60.5 - 60|}{\frac{2}{\sqrt{25}}} = 1.25$$

चूँकि प्रतिदर्श का आकार 25 है ओर स्वतंत्रता की कोटियाँ $25 - 1 = 24$ है। t-सारणी से हम पाते हैं कि 24 स्वतंत्रता की कोटियाँ, 5 प्रतिशत सार्थकता का स्तर और एक-पुच्छ परीक्षण के लिए t-का मान 1.71 है।

t-का प्रेक्षित मान निर्णायक मान से कम है। इसलिए हम निराकरणीय परिकल्पना को अस्वीकृत नहीं करते। अतः स्वास्थ्य देखभाल संबंधी उपायों को अपनाने के बावजूद भी जिले की जीवन प्रत्याशा में कोई बदलाव नहीं आया।

बोध प्रश्न 2

1) एक रिपोर्ट का मानना है कि किसी स्कूल की परीक्षा में गणित में औसतन 78 अंक प्राप्त किए गए, मानक विचलन 16 था। हालांकि, 37 विद्यार्थियों का यादृच्छिक प्रतिदर्श गणित में औसतन 84 अंक दर्शाता है। इस प्रमाण को ध्यान में रखकर क्या हम निष्कर्ष निकाल सकते हैं कि गणित में औसतन 84 अंक प्राप्त किए गए थे। 0.05 की सार्थकता के स्तर का प्रयोग करें।

   ....

   ....

   ....

   ....

2) यात्री कार बनाने वाली किसी कंपनी का दावा है कि कारों की औसत इंधन दक्षता 35 किमी. प्रति लीटर पैट्रोल है। 50 कारों के यादृच्छिक प्रतिदर्श से ऐसी क्षमता औसतन 32 किमी. प्रति लीटर पायी गई है और मानक विचलन 1.2 किमी. रहा है। क्या यह

प्रमाण 0.01 सार्थकता के स्तर पर कंपनी के दावे को गलत साबित करता है?

....

....

....

....

3) नारियल के तेल के 200 टिनों के यादृच्छिक प्रतिदर्श से औसतन प्रति टिन 4.95 किग्रा. के भार का पता चलता है और जिसका मानक विचलन 0.21 किग्रा. है। क्या आप 0.01 के सार्थकता के स्तर पर प्रति टिन 5 किग्रा. के नेट भार की परिकल्पना को स्वीकार करते हैं?

....

....

....

....

....

4) किसी रिपोर्ट के अनुसार, हाल ही के वर्ष के दौरान सरकारी कर्मचारियों की राष्ट्रीय औसतन वार्षिक आय 24,632 रुपए थी और जिसका मानक विचलन 1827 रुपए था। इसी वर्ष के दौरान 49 सरकारी कर्मचारियों के यादृच्छिक प्रतिदर्श से 25,415 रुपए की औसतन वार्षिक आमदनी का पता चलता है। इस प्रतिदर्श को ध्यान में रखकर, 0.05 के सार्थकता के स्तर पर क्या हम निष्कर्ष निकाल सकते हैं कि सरकारी कर्मचारियों की राष्ट्रीय औसतन वार्षिक आय सही मायने में 24,632 रुपए ही थी।

....

....

....

....

....

19.5 दो प्रतिदर्शों के बीच के अंतर से संबंधित परीक्षण

कई बार हमें दो प्रतिदर्शों के बीच के अंतर के लिए परीक्षण करना पड़ता है। इसका उद्देश्य पता लगाना होता है कि क्या दोनों प्रतिदर्श समान समष्टि से लिए गए हैं या इससे जाँच की जाती है कि क्या दोनों समष्टियों में कोई एक सामान्य लक्षण मौजूद है। जैसे, हम एक परिकल्पना करते हैं कि किसी प्लांट A में प्रति कामगार उत्पादन उतना ही है जितना कि प्लांट B में प्रति कामगार है। ऐसी परिकल्पना के परीक्षण की प्रक्रिया की चर्चा हम आगे कर रहे हैं।

यहाँ हमारे सम्मुख दो अलग-अलग स्थितियाँ हैं : क्या दोनों समष्टियों का प्रसरण हमें ज्ञात है। दूसरी कि प्रतिदर्श की आकार छोटा है या बड़ा।

निराकरणीय परिकल्पना है कि दोनों समष्टियों का समष्टि माध्य समान है। सांकेतिक रूप से

$$H_0 : \mu_1 = \mu_2 \qquad \ldots(19.8)$$

वैकल्पिक परिकल्पना होगी कि दोनों समष्टियों माध्य अलग-अलग हैं। सांकेतिक रूप से

$$H_A : \mu_1 \neq \mu_2 \qquad \ldots(19.9)$$

19.5.1 समष्टि प्रसरण ज्ञात हों

जब दोनों समष्टियों के मानक विचलन (प्रसरण के सकारात्मक वर्गमूल) ज्ञात हो तो हम z प्रतिदर्शज को लागू करेंगे।

$$z = \frac{\left|(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)\right|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \qquad \ldots(19.10)$$

उपर्युक्त (19.10) में, पादांक 1 से आशय पहले प्रतिदर्श से है और पादांक 2 से आशय दूसरे प्रतिदर्श से है। (19.10) में उपयुक्त आंकड़ों को प्रयोग करके हम z के प्रेक्षित मान की प्राप्ति करते हैं। और सार्थकता के विशिष्ट स्तर के लिए इसकी तुलना निर्णायक मान से करते हैं।

उदाहरण 19.4

कोई बैंक दिल्ली और कोलकाता के अपने ग्राहकों की औसत बचत का पता लगाना चाहता है। दिल्ली के 250 खातों के प्रतिदर्श से 22,500 रुपए की औसत बचत का पता चलता है जबकि कोलकाता के 200 खातों के प्रतिदर्श से 21,500 रुपए की औसत बचत का पता चलता है। यह ज्ञात है कि दिल्ली में बचत का मानक विचलन 150 रु. है और कोलकाता में यह 200रुपए है। क्या 1 प्रतिशत के सार्थकता स्तर पर हम निष्कर्ष निकाल सकते हैं कि दिल्ली और कोलकाता में ग्राहकों का बैंकिंग व्यवहार एक जैसा है?

इस मामले में निराकरणीय परिकल्पना है $H_0 : \mu_1 = \mu_2$

और वैकल्पिक परिकल्पना है $H_A : \mu_1 \neq \mu_2$

सुलभ जानकारी के आधार पर

$\bar{x}_1 = 22500$ रुपए $\qquad \sigma_1 = 150$ रुपए

$\bar{x}_2 = 22400$ रुपए $\qquad \sigma_2 = 200$ रुपए

$n_1 = 250$ रुपए $\qquad n_2 = 200$

चूँकि σ_1 और σ_2 हमें ज्ञात हैं, इसलिए हम z-परीक्षण लागू करते हैं

परीक्षण प्रतिदर्शज है $z = \dfrac{\left|(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)\right|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$

उपर्युक्त जानकारी प्रयोग करके, हमें प्राप्त होता है :

$$z = \frac{|22500 - 22400|}{\sqrt{\frac{150^2}{250} + \frac{200^2}{200}}} = \frac{100}{90 + 200} = 5.87$$

हम पाते हैं कि 1 प्रतिशत के सार्थकता के स्तर पर सारणी 19.3 से प्राप्त निर्णायक मान 2.58 है।

चूँकि t का प्रेक्षित मान, t के निर्णाचक मान से अधिक है, इसलिए निराकरणीय परिकल्पना अस्वीकृत की जाती है और वैकल्पिक परिकल्पना स्वीकार की जाती है। अतः दिल्ली और कोलकाता में ग्राहकों का बैंकिंग व्यवहार अलग-अलग है।

19.5.2 समष्टि प्रसरण अज्ञात हो

जब समष्टि मानक प्रसरण (σ^2) अज्ञात हो तो हम प्रतिदर्श मानक प्रसरण (s^2) से इसका अनुमान लगाते हैं। यदि दोनों प्रतिदर्श आकार में बड़े हैं अर्थात (n>30) तब हम z प्रतिदर्शज को इस प्रकार लागू करेंगे:

$$z = \frac{\left|(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)\right|}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} \qquad \ldots(19.11)$$

दूसरी तरफ, यदि प्रतिदर्श का आकार छोटा है अर्थात ($n \leq 30$) तब निम्नलिखित की भांति हम t-प्रतिदर्शज को लागू **करेंगे**:

$$t = \frac{\left|(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)\right|}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} \qquad \ldots(19.12)$$

t-परीक्षण में स्वतंत्रता की कोटियाँ हैं $= (n_1 - 1) + (n_2 - 1) = n_1 + n_2 - 2$

उदाहरण 19.5

गणित की कोई अध्यापिका, कक्षा X के दो अनुभागों की कार्य प्रगति की तुलन करना चाहती है। वह अनुभाग A में 25 विद्यार्थियों को एक जैसा प्रश्न पत्र हल करने के लिए देती है। उसने पाया कि अनुभाग A विद्यार्थियों के अंक 78 हैं। उनका मानक विचलन 4 है। अनुभाग B के विद्यार्थियों के अंक 75 हैं और मानक विचलन 5 है। क्या दोनों अनुभागों में विद्यार्थियों की प्रगति 1 प्रतिशत के सार्थकता के स्तर पर अलग-अलग है?

इस मामले में निराकरणीय परिकल्पना है $H_0 : \mu_1 = \mu_2$

और वैकल्पिक परिकल्पना है $H_A : \mu_1 \neq \mu_2$

उपलब्ध जानकारी है :

$\bar{x}_1 = 78$ $\quad\quad$ $s_1 = 4$

$\bar{x}_2 = 75$ $\quad\quad$ $s_2 = 5$

$n_1 = 25$ $\quad\quad$ $n_2 = 20$

चूँकि σ_1 और σ_2 अज्ञात हैं और प्रतिदर्शों का आकार छोटा है, इसलिए हम t-परीक्षण लागू करते हैं

$$t = \frac{\left|(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)\right|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{|78-75|}{\sqrt{\frac{4^2}{25} + \frac{5^2}{20}}} = \frac{3}{1.37} = 2.18$$

इस मामले में स्वतंत्रता की कोटियाँ हैं: $25 + 20 - 2 = 43$.

सारणी 15.3 से पता चलता है कि 43 स्वतंत्रता की कोटियों के लिए 1 प्रतिशत के सार्थकता के स्तर पर t का मान 2.69 है।

चूँकि t का निर्णायक मान, t के प्रेक्षित मान से कम है, इसलिए हम निराकरणीय परिकल्पना को अस्वीकृत करते हैं और वैकल्पिक परिकल्पना को स्वीकृत करते हैं। इसलिए गणित में कार्य-प्रगति के संदर्भ में अनुभाग A के विद्यार्थी अनुभाग B से भिन्न हैं।

बोध प्रश्न 3

1) निम्नलिखित परिकल्पना का परीक्षण कीजिए :

$n_1 = 50$ $\quad$ $n_2 = 50$

$\bar{x}_1 = 52.3$ $\quad$ $\bar{x}_2 = 52.3$

$\sigma_1 = 6.1$ $\quad$ $\sigma_2 = 6.1$

परीक्षण कीजिए कि :

$H_o : \mu_1 = \mu_2$

$H_A : \mu_1 \neq \mu_2$

$\alpha = 0.05$

....

....

....

....

2) दो प्रसामान्य समष्टियों से दो प्रतिदर्श निकाले जाते हैं और उनसे यह सूचना प्राप्त होती है :

$n_1 = 15$ $\quad$ $n_2 = 10$

$\bar{x}_1 = 140$ $\quad$ $\bar{x}_2 = 150$

$s_1 = 10$ $\quad$ $s_2 = 15$

1 प्रतिशत के सार्थकता के स्तर पर परिकल्पना का परीक्षण कीजिए कि दोनों समष्टियों के बीच कोई अंतर नहीं है।

....

....

....

....

3) मान लीजिए दो प्रसामान्य समष्टियों से $n_1 = 20$ और $n_2 = 15$ आकार के प्रतिदर्श प्राप्त किए जाते हैं। प्रतिदर्शज इस प्रकार हैं :

$\bar{x}_1 = 225 \qquad \bar{x}_2 = 125$

$s_1^2 = 225 \qquad s_2^2 = 150$

क्या हम 5 प्रतिशत के सार्थकता के स्तर पर निष्कर्ष निकाल सकते हैं कि $\mu_1 < \mu_2$?

....

....

....

....

19.6 सारांश

इस इकाई में हमने परिकल्पना परीक्षण और समष्टि के बारे में निष्कर्ष निकालने की विधियों के बारे में चर्चा की। परिकल्पना एक ऐसा कथन है जो प्राचल से संबंधित है। परिकल्पना परीक्षण के लिए उपलब्ध जानकारी के आधार पर हम परीक्षण प्रतिदर्शज को सूत्रबद्ध करते हैं। इस इकाई में हम इन स्थितियों पर विचार किया करते हैं : i) एकल प्रतिदर्श की व्याख्या, और ii) दो प्रतिदर्शों के बीच की तुलना।

परीक्षण प्रतिदर्शज का निर्माण, समष्टि प्रसरण और प्रतिदर्श आकार से संबंधित जानकारी पर निर्भर करता है। जब समष्टि प्रसरण हमें ज्ञात हो या प्रतिदर्श का आकार बड़ा हो तो हम प्रसामान्य बंटन लागू करते हैं और परिकल्पना परीक्षण के लिए z का प्रयोग करते हैं। दूसरी तरफ जब हमें समष्टि प्रसरण का पता नहीं होता और प्रतिदर्श आकार छोटा हो तो t-बंटन के आधार पर हम परीक्षण प्रतिदर्शज का निर्माण करते हैं। याद रखिए कि बड़े प्रतिदर्शों के लिए बंटन सन्निकटतः प्रसामान्य होता है और इसलिए हम z प्रतिदर्शज का प्रयोग कर सकते हैं।

19.7 शब्दावली

आकलक (Estimator) : आकलन सिद्धांत में प्रतिदर्शज को दिया जाने वाला अन्य नाम।

प्राचल (Parameter) : समष्टि के कुछ अभिलक्षणों का माप।

परिकल्पना परीक्षण की समस्या (Problem of Hypothesis Testing) : कई बार समष्टि के कुछ विशिट अभिलक्षणों के बारे में हमारे पास कुछ जानकारी होती है और हम जाँच करना चाहते हैं कि क्या समष्टि से प्राप्त यादृच्छिक प्रतिदर्श के आधार पर ऐसी जानकारी मान्य है या नहीं। सांख्यिकीय अनुमिति की समस्या को परिकल्पना परीक्षण की समस्या कहते हैं।

प्रतिचयन त्रुटि (Sampling Error) : प्रतिचयन विधि में हम प्राप्त समष्टि से प्रतिदर्श प्राप्त करके ऐसी समष्टि के कुछ लक्षणों का अनुमान करने का प्रयास करते हैं। चूँकि प्रतिदर्श में समष्टि के सभी सदस्य शामिल नहीं किए जाते, इसलिए इस पर आधारित आकलन अनुमान ही रहते हैं। ये अपेक्षित जानकारी के लक्षण के समरूप नहीं होता और इसी वजह से कुछ त्रुटि हो जाती है। ऐसी त्रुटि को प्रतिचयन त्रुटि कहते हैं।

प्रतिचयन उच्चावयन (Sampling Fluctuation) : परिकलित प्रतिदर्शज के मानों में पाई जाने वाली विविधता।

प्रतिदर्शज (Statistic) : ऐसी इकाइयों के मानों का फलन जिन्हें प्रतिदर्श में शामिल किया जाता है। इसका बुनियादी उद्देश्य कुछ समष्टि प्राचलों का अनुमान लगाना है।

मानक त्रुटि (Standard Error) : प्रतिदर्शज के प्रतिदर्शी बंटन का मानक विचलन।

18.8 कुछ उपयोगी पुस्तकें

Nagar, A. L. and Das, R. K., 1989, *Basic Statistics:* Oxford University Press, Delhi, Chapter 9.

Newbold, P., 1991. *Statistics for Business and Economics* (Third Edition): Prentice Hall, New Jersey, Chapters 6, 7, 8 and 9.

Keller, G, and B. Warrack, 1991, *Essentials of Business Statistics*, Wordsworth Publishing Co., California, Chapters 7 and 8.

18.9 बोध प्रश्नों के उत्तर

बोध प्रश्न 1

1) अनुभाग 18.2 और 18.3 को भलीभाँति पढ़कर उत्तर दें।

2) यह बड़ा प्रतिदर्श है और प्रसरण अज्ञात है। इसके लिए एक-पुच्छ परीक्षण करना होगा। अतः अस्वीकृति क्षेत्र, मानक प्रसामान्य वक्र की दायें तरफ की पुच्छ है। इसी आधार पर आरेख बनाइए।

बोध प्रश्न 2

1) चूँकि यह ज्ञात प्रसरण वाला, बड़ा प्रतिदर्श है, इसलिए हम z-प्रतिदर्श लागू कर सकते हैं। चूँकि वैकल्पिक परिकल्पना $\mu > 78$ है, इसलिए हम एक-पुच्छ परीक्षण करेंगे। z का प्रेक्षित मान 2.28 है। 5 प्रतिशत के सार्थकता के स्तर पर z का निर्णायक मान 1.65 है। चूँकि निर्णायक मान की तुलना में प्रेक्षित मान बड़ा है, इसलिए हम निराकरणीय परिकल्पना को अस्वीकार करते हैं। अतः निष्कर्ष है कि औसतन अंक 78 से अधिक थे।

2) यह एक बड़ा प्रतिदर्श है और जिसका प्रसरण अज्ञात है। इसके लिए द्वि-पुच्छ परीक्षण करना होगा। z का प्रेक्षित मान 17.68 है और 1 प्रतिशत के सार्थकता के स्तर पर z का निर्णायक मान 2.58 है। चूँकि प्रेक्षित मान, निर्णायक मान से अधिक है, इसलिए निराकरणीय परिकल्पना अस्वीकृत की जाती है।

3) अज्ञात प्रसरण वाला यह एक बड़ा प्रतिदर्श है। इसके लिए z-प्रतिदर्शज वाला द्वि-पुच्छ परीक्षण करना होगा। z का प्रेक्षित मान 3.37 है। निराकरणीय परिकल्पना अस्वीकृत की जाती है।

4) चूँकि यह ज्ञात मानक विचलन वाला बड़ा प्रतिदर्श है। इसके लिए द्वि-पुच्छ परीक्षण करना होगा। z का प्रेक्षित मान 3.00 है और 5 प्रतिशत के सार्थकता के स्तर पर z का निर्णायक मान 2.58 है। निराकरणीय परिकल्पना अस्वीकृत की जाती है। इसलिए, सरकारी कर्मचारियों की राष्ट्रीय औसतन आय 24,632रुपए से भिन्न है।

बोध प्रश्न 3

1) प्रतिदर्श बड़े आकार के है और समष्टि मानक विचलन ज्ञात है। इसलिए हम z-प्रतिदशज को लागू करते हैं और z का प्रेक्षित मान 2.58 है। चूँकि वैकल्पिक परिकल्पना $\mu_1 \neq \mu_2$ है। अतः $\sigma = 0.05$ पर द्वि-पुच्छ परीक्षण करना होगा और z का निर्णायक मान 1.96 है। निराकरणीय परिकल्पना अस्वीकृत की जाती है।

2) प्रतिदर्श का आकार छोटा है और σ अज्ञात है। इसलिए हम t-प्रतिदर्शज लागू करते हैं और t का प्रेक्षित मान 0.61 है। H_A: $\mu_1 \neq \mu_2$ की तुलना में परिकल्पना H_o: $\mu_1 = \mu_2$ है। अतः इसके लिए द्वि-पुच्छ परीक्षण करना होगा। 1 प्रतिशत के सार्थकता के स्तर पर 23 स्वतंत्रता की कोटियों के लिए t का निर्णायक मान 2.50 है। अतः निराकरणीय परिकल्पना अस्वीकृत नहीं होती।

3) प्रतिदर्श आकार में छोटा है और σ अज्ञात है, t-प्रतिदर्शज लागू किया जाता है। t का प्रेक्षित मान 0.72 है। H_o $\mu_1 = \mu_2$ है और H_A $\mu_1 < \mu_2$ है। अतः एक पुच्छ परीक्षण करना होगा। 33 स्वतंत्रता की कोटियों पर 5 प्रतिशत सार्थकता के स्तर के लिए, t का निर्णायक मान 2.00 है। H_o अस्वीकृत नहीं है।

इकाई 20 नामीय आंकड़ों से संबंधित काई-वर्ग परीक्षण

इकाई की रूपरेखा

20.0 उद्देश्य

इस इकाई को पढ़ने के बाद आप :

- आसंग सारणी के रूप में नामीय आंकड़ों को प्रस्तुत कर सकेंगे;
- काई-वर्ग परीक्षण प्रतिदर्शज को व्यक्त कर सकेंगे; और
- आसंग सारणियों के संदर्भ में काई-वर्ग परीक्षण का प्रयोग कर सकेंगे।

20.1 प्रस्तावना

पिछली दो इकाइयों में हमने प्रतिदर्श सूचना के आधार पर समष्टि प्राचलों के बारे में निष्कर्ष निकालने की प्रक्रिया की चर्चा की थी। बहुत से मामलों में, विशेष रूप से नामीय चरों के लिए हमें प्राचल प्राप्त नहीं हैं। इस संदर्भ में एक चर (या गुण) ऐसा मान बन जाता है, जो परिमित संख्या की श्रेणियों से संबंधित होता है और हम प्रत्येक श्रेणी में प्रेक्षणों की संख्या की गिनती कर सकते हैं। नामीय आंकड़ों के मामले में निष्कर्ष निकालना ही इस इकाई की विषयवस्तु है।

पिछली इकाई में हमने परिकल्पना परीक्षण की विधि की चर्चा की थी और जिसमें समष्टि से संबंधित कुछ निश्चित अवधारणाओं को शामिल करना ज़रूरी होता है और जिससे हम प्रतिदर्श की प्राप्ति करत हैं। जैसे, छोटे प्रतिदर्शों के लिए टी-परीक्षण लागू करने के लिए ज़रूरी है कि मूल समष्टि प्रसामान्य रूप से बंटित हो। इसी तरह प्राचल के लिए कोई विशिष्ट मान निर्धारित करके परिकल्पना को सूत्रबद्ध किया जाता है। अतः ऐसे परीक्षणों को प्राचलिक परीक्षण (parametric test) कहते हैं।

यदि प्राचल के मान के बारे में कोई भी अभिधारणा कायम करना संभव नहीं है तो पिछली इकाई में वर्णित परीक्षण प्रक्रिया असफल रहेगी। ऐसी स्थितियाँ, जहाँ समष्टि प्रसामान्य बंटन का अनुसरण नहीं करती, या जिन स्थितियों में प्राचल मान निर्धारित करना संभव नहीं होता, वहाँ हम गैर-प्राचलिक परीक्षणों (non-parametric tests) का प्रयोग करते हैं।

20.2 आसंग सारणी

आसंग सारणी (contingency table) एक आयताकार सारणी है जिसमें समष्टि से प्राप्त प्रेक्षणों को दो अभिलक्षणों के आधार पर वर्गीकृत किया जाता है। इसे द्विधा सारणी (two-way table) भी कहते हैं और इकाई 7 में हमने इसकी चर्चा भी की थी। ऐसे गुणात्मक आंकड़ों को याद कीजिए जिन्हें श्रेणियों में वर्गीकृत किया जा सकता है और जिन्हें हमने द्विधा सारणी में प्रस्तुत किया था।

काई-वर्ग परीक्षण के अनुप्रयोग को समझाने के लिए, आइए, एक ठोस उदाहरण लें। इकाई 7 में हमने पिता के व्यवसाय और बच्चों की संख्या संबंधी उदाहरण प्रस्तुत किया था। हमने व्यवसाय को पाँच श्रेणियों में विभाजित किया : i) बेरोजगार, ii) अकुशल श्रमिक, iii) कुशल श्रमिक, iv) स्व-रोजगार प्राप्त, और v) पेशेवर। इसी तरह, बच्चों की संख्या के आधार पर हमने परिवारों को पाँच श्रेणियों में विभाजित किया था : i) निःसंतान, ii) एक संतान, iii) दो संतान, iv) तीन संतान, और v) तीन से अधिक संतान वाले परिवार। 650 परिवारों के प्रतिदर्श के लिए प्राप्त आँकड़ों को सारणी 20.1 में प्रस्तुत किया गया है।

तालिका 20.1 : व्यवसाय और बच्चों की संख्या पर आधारित प्रेक्षित बारंबारता

बच्चों की संख्या	बेरोजगार	अकुशल श्रमिक	व्यवसाय कुशल श्रमिक	स्व-रोजगार	पेशेवर	कुल
	(1)	(2)	(3)	(4)	(5)	
0	10	15	10	12	11	58
1	35	25	17	18	25	120
2	22	33	45	40	43	183
3	11	40	48	58	30	187
≥ 4	11	33	30	19	9	102
कुल	89	146	150	147	118	650

सारणी 20.1, आसंग सारणी है क्योंकि हम पता करने का प्रयास कर रहे हैं कि क्या बच्चों की संख्या, पिता के व्यवसाय पर आश्रित है।

हमारा उद्देश्य पिता के व्यवसाय और बच्चों की संख्या के बीच के संभावित संबंध का परीक्षण करना है। अतः निराकरणीय परिकल्पना इस प्रकार होगी:

H_0 : वैकल्पिक परिकल्पना के संदर्भ में पिता का व्यवसाय और बच्चों की संख्या एक-दूसरे से अलग है।

H_A : बच्चों की संख्या और पिता का व्यवसाय एक-दूसरे पर निर्भर है।

सारणी 20.1 में हमने प्रत्येक कोष्ठिका (cell) के लिए प्रेक्षित बारंबारता (observed frequency) को सारणी में प्रस्तुत किया है। जब विचाराधीन चरों के बीच कोई संबंध नहीं है तो प्रत्याशित बारंबारता (expected frequency) क्या होनी चाहिए। इसका उत्तर हम नीचे देंगे।

20.3 प्रत्याशित बारंबारताएँ

जैसा कि हमने ऊपर बताया, प्रत्याशित बारंबारता को ऐसी अभिधारणा के अंतर्गत परिकलित किया जाता है जब बच्चों की संख्या और पिता के व्यवसाय के बीच कोई संबंध नहीं होता। तालिका 20.1 में प्रत्येक कोष्ठिका के लिए प्रत्याशित बारंबारता की प्राप्ति, कोष्ठिका प्रायिकता द्वारा n आकार के प्रतिदर्शों को गुणा करके की जाती है। कोष्ठिका प्रायिकता परिकलित करने के लिए हमें सर्वप्रथम प्रत्येक पंक्ति और स्तंभ के लिए सीमांत बारंबारता को ज्ञात करना होगा। जैसा कि इकाई 7 में हमने बताया था, पंक्ति सीमांत (row marginal) पंक्ति के कुल योग (row total) के समान और इसी तरह 'स्तम्भ सीमांत' स्तम्भ के कुल योग (column totals) द्वारा प्राप्त होते हैं। पंक्तियों के लिए, हम 'सीमांत पंक्ति प्रायिकता' ज्ञात कर सकते हैं। पहली पंक्ति के लिए, सीमांत पंक्ति प्रायिकता $p(r_1)$ होगी।

$$p(r_2) = \frac{120}{650} = 0.18 \qquad \ldots(20.1)$$

अन्य पंक्तियों के लिए सीमांत पंक्ति प्रायिकताएँ हैं:

$$p(r_2) = \frac{120}{650} = 0.18 \qquad p(r_3) = \frac{183}{650} = 0.28$$

$$p(r_4) = \frac{187}{650} = 0.29 \qquad p(r_5) = \frac{102}{650} = 0.16$$

इसी तरह स्तम्भ 1 के लिए सीमांत स्तम्भ प्रायिकता $p(c_1)$ होगी :

$$p(c_1) = \frac{89}{650} = 0.14 \qquad \ldots(20.2)$$

अन्य स्तम्भों के लिए सीमांत स्तम्भ प्रायिकताएँ हैं:

$$p(c_2) = \frac{146}{650} = 0.22 \qquad p(c_3) = \frac{150}{650} = 0.23$$

$$p(c_4) = \frac{147}{650} = 0.23 \qquad p(c_5) = \frac{118}{650} = 0.18$$

इकाई 13 को ध्यान में लाइए जहाँ हमने अध्ययन किया था कि यदि A और B अलग-अलग घटनाएँ हैं, तो A और B के संयुक्त रूप से उभरने की प्रायिकता होगी

$$p(A \cap B) = p(A).p(B)$$

इसलिए, यदि हम निराकरणीय प्रायिकता को सत्य मान लें तब पहली कोष्ठिका (c_1, r_1) के लिए कोष्ठिका प्रायिकता होगी:

$$p(r_1 \cap c_1) = p(r_1).p(c_1) = \frac{58}{650} \times \frac{89}{650} = 0.0892 \times 0.1369 = 0.0122. \qquad \ldots(20.3)$$

अतः पहली कोष्टिका के लिए प्रत्याशित बारंबारता होगी:

$$E_{11} = n.p(r_1 \cap c_1) = 650 \times 0.0122 = 7.94 \qquad \ldots(20.4)$$

सामान्य भाषा में हम कह सकते हैं कि ij कोष्ठिका की प्रत्याशित बारंबारता है

$$E_{ij} = \frac{(\text{पंक्ति } i \text{ का कुल योग})(\text{स्तम्भ } j \text{ का कुल योग})}{\text{प्रतिदर्श आकार}} \qquad \ldots(20.5)$$

(20.5) को लागू कर हम प्रत्येक कोष्ठिका के लिए प्रत्याशित कोष्ठिका बारंबारता को परिकलित कर सकते हैं और सारणी 20.2 जैसी सारणी तैयार कर सकते हैं।

तालिका 20.2 : प्रत्येक कोष्ठिका के लिए प्रत्याशित बारंबारता का परिकलन

बच्चों की संख्या		व्यवसाय					कुल
		बेरोजगार	अकुशल श्रमिक	कुशल श्रमिक	स्व-रोजगार	पेशेवर	
		c_1	c_2	c_3	c_4	c_5	
0	r_1	7.94	13.03	13.38	13.12	10.53	58.00
1	r_2	16.43	26.95	27.69	27.14	21.78	120.00
2	r_3	25.06	41.10	42.23	41.39	33.22	183.00
3	r_4	25.60	42.00	43.15	42.29	33.95	187.00
≥4	r_5	13.97	22.91	23.54	23.07	18.52	102.00
कुल		89.00	146.00	150.00	147.00	118.00	650.00

हमारा अगला चरण प्रत्याशित बारंबारता से प्रेक्षित बारंबारता की तुलना करना है।

20.4 काई-वर्ग परीक्षण प्रतिदर्शज

प्रेक्षित बारंबारता से प्रत्याशित बारंबारता की तुलना करने के लिए हम काई-वर्ग प्रतिदर्शज का निर्माण करेंगे, जो है:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} \qquad \ldots(20.6)$$

जहाँ O प्रेक्षित बारंबारता है और E प्रत्याशित बारंबारता।

काई-वर्ग प्रतिदर्शज की स्वतंत्रता की कोटियाँ $(r-1)(c-1)$ हैं। जैसे, यदि 3 पंक्तियाँ और 4 स्तम्भ हैं तो स्वतंत्रता की कोटि है $(3-1)(4-1)=6$

आइए काई-वर्ग परीक्षण का अनुसरण करने में शामिल चरणों को संक्षेप में समझें। ये हैं:

1) निराकरणीय और वैकल्पिक परिकल्पनाओं को स्पष्ट कीजिए।

2) (20.5) का प्रयोग कर प्रत्येक कोष्ठिका के लिए प्रत्याशित बारंबारता परिकलित कीजिए।

3) (20.6) का प्रयोग कर χ^2 प्रतिदर्शज के प्रेक्षित मान को परिकलित कीजिए।

4) सूत्र $(r-1)(c-1)$ के आधार पर स्वतंत्रता की कोटि का निर्धारण कीजिए।

5) अपेक्षित सार्थकता के स्तर (α) की जाँच कीजिए।

6) खंड 5 की इकाई 15 में दी गई सारणी 15.3 से α और प्रासंगिक स्वतंत्रता की कोटि के लिए χ^2 के निर्णायक मान ज्ञात कीजिए।

7) χ^2 के प्रेक्षित मान की तुलना χ^2 के निर्णायक मान से कीजिए।

8) यदि प्रेक्षित मान, निर्णायक मान से कम है तो H_0 को स्वीकृत कीजिए।

9) यदि प्रेक्षित मान, निर्णायक मान से बड़ा है तब H_0 को स्वीकृत न कीजिए और H_A को स्वीकृत कीजिए।

सारणी 20.1 में प्रदत्त आंकड़ों के लिए, आइए χ^2 के प्रेक्षित मान को ज्ञात करें।

सारणी 20.3 : हरेक कोष्ठिका के लिए $\frac{(O_i - E_i)^2}{E_i}$

बच्चों की संख्या	व्यवसाय						कुल
		बेरोजगार	अकुशल श्रमिक	कुशल श्रमिक	स्व-रोजगार	पेशेवर	
		c_1	c_2		c_4	c_5	
0	r_1	0.53	0.30	0.86	0.10	0.02	1.80
1	r_2	20.99	0.14	4.13	3.08	0.47	28.81
2	r_3	0.37	1.60	0.18	0.05	2.88	5.08
3	r_4	8.33	0.10	0.54	5.84	0.46	15.26
≥ 4	r_5	0.63	4.44	1.77	0.72	4.89	12.46
कुल		30.85	6.58	7.48	9.77	8.72	63.41

चूंकि 5 पंक्तियाँ और 5 स्तम्भ हैं, इसलिए स्वतंत्रता की कोटियाँ (5 - 1) (5 - 1) = 16 हैं।

16 स्वतंत्रता की कोटि के लिए, 5% के सार्थकता के स्तर पर χ^2 का निर्णायक मान 26.30 होगा (सारणी 15.3 देखें)। सारणी 20.3 से हम पाते हैं कि χ^2 का प्रेक्षित मान 63.41 है। चूँकि प्रेक्षित मान, निर्णायक मान से बड़ा है, इसलिए हम निराकरणीय परिकल्पना को अस्वीकृत करते हैं और वैकल्पिक परिकल्पना को अपना लेते हैं। अतः निष्कर्ष निकलता है कि बच्चों की संख्या और पिता का व्यवसाय एक-दूसरे पर निर्भर हैं।

बोध प्रश्न 1

1. निम्नलिखित अवधारणाओं का वर्णन कीजिए:

   क) सीमांत बारंबारता

   ख) कोष्ठिका प्रायिकता

   ग) प्रत्याशित बारंबारता

घ) χ^2 के निर्णायक मान

....

....

....

....

....

....

2) किसी कंपनी द्वारा तीन तरह के (संतरा, कोला और नींबू) पेय पदार्थ तैयार किए जाते हैं। दो राज्यों (पहला उत्तर का पंजाब और दूसरा दक्षिण का तमिलनाडु) में 160 व्यक्तियों के सर्वेक्षण से निम्नलिखित जानकारी प्राप्त होती है।

	संतरा	कोला	नींबू
पंजाब	33	26	31
तमिलनाडु	17	24	29

परिकल्पना परीक्षण कीजिए कि दोनों राज्यों में किसी एक पेय पदार्थ को विशेष रूप से अधिक पसंद नहीं किया जा रहा है ($\alpha = 0.05$).

....

....

....

....

....

20.5 सारांश

गुणात्मक आंकड़ों के मामले में हम प्राचलिक मान प्राप्त नहीं कर सकते। इसलिए, z-प्रतिदर्शज या t- प्रतिदर्शज के आधार पर परिकल्पना परीक्षण नहीं किया जा सकता । ऐसी स्थितियों में काई-वर्ग परीक्षण लागू किया जाता है। काई-वर्ग परीक्षण गैर-प्राचलिक परीक्षण है जहाँ समष्टि के बारे में कोई अभिधारणा कायम करना जरूरी नहीं होता । काई-वर्ग परीक्षण के अलावा कुछ अन्य प्रकार के गैर-प्राचलिक परीक्षण भी होते हैं। हालांकि आसंग सारणी के अलावा बहुत सी स्थितियों में काई-वर्ग परीक्षण लागू किया जा सकता है।

आसंग सारणी में हमने निराकरणीय परिकल्पना का परीक्षण किया है । वैकल्पिक परिकल्पना की तुलना में जहाँ चर एक-दूसरे से संबद्ध होते हैं, वहीं विचाराधीन चर एक-दूसरे से अलग होते हैं। यहाँ हम प्रत्याशित बारंबारता की तुलना प्रेक्षित बारंबारता से करते हैं और काई-वर्ग प्रतिदर्शज का निर्माण करते हैं। यदि काई वर्ग का प्रेक्षित मान, काई-वर्ग के प्रत्याशित मान से बड़ा होता है तो हम निराकरणीय परिकल्पना का खंडन करते हैं।

20.5 शब्दावली

नामीय चर (Nominal Variable) : ऐसे चर के गुणात्मक मान होते हैं और इनमें कोई क्रमबद्ध संबंध नहीं होता। जैसे लिंग में दो मान अर्थात स्त्री और पुरुष शामिल होते हैं लेकिन इसमें स्त्री/पुरुष की कोई क्रमबद्ध स्थिति का समावेश नहीं होता। नामीय चर को गुणवाची चर (attribute) भी कहते हैं।

आसंग सारणी (Contingency Table) : द्विचर आंकड़ों को प्रस्तुत करने की द्विधा सारणी। इसे आसंग सारणी भी कहते हैं क्योंकि हम पता लगाने का प्रयास करते हैं कि क्या एक चर, दूसरे चर पर आश्रित है या नहीं।

प्रत्याशित बारंबारता (Expected Frequency) : इस अभिधारणा के अंतर्गत प्रत्याशित कोष्ठिका बारंबारता कि दोनों चर एक-दूसरे से भिन्न हैं।

20.7 : कुछ उपयोगी पुस्तकें

Kiess, H.O., 1989, *Statistical Concepts for the Behavioral Science*, Allyn and Bacon, Boston.

Keller, G. and B. Warrack, 1991, *Essentials of Business Statistics,* Wordsworth Publishing Co., California.

20.8 बोध प्रश्नों के उत्तर / संकेत

बोध प्रश्न 1

1) इकाई का भलीभांति अध्ययन कर आवधारणाओं को स्पष्ट करें।

2) प्रत्याशित बारंबारता हैः

	संतरा	कोला	नींबू
पंजाब	28.13	28.13	33.75
तमिलनाडु	21.88	21.88	26.25

काई-वर्ग प्रतिदर्शज का प्रेक्षित मान 2.98 है और स्वतंत्रता की कोटि 2 है। 2 स्वतंत्रता की कोटियों पर 5 प्रतिशत के सार्थकता स्तर पर, काई-वर्ग का निर्णायक मान 5.99 है। अतः निराकरणीय परिकल्पना का खंडन नहीं किया जा सकता । क्षेत्रों में पेय पदार्थ का उपभोग अलग-अलग है।