इकाई–1

प्रस्तावना एवम् अवलोकन (Introduction and Qverview)

संरचना (Structure)

1.1 उद्देश्य (Objectives)

इस इकाई के अध्ययन के पश्चात् विद्यार्थी योग्य होंगे–

- प्रतिदर्श तथा प्रांचल की अवधारणा को समझने में।
- समष्टि और प्रतिदर्श को जानने में।

1.2 प्रस्तावना (Introduction)

किसी भी अनुसंधान का मुख्य उद्देश्य कुछ ऐसे तथ्यों, सिद्धान्तों अथवा उपयोगों का पता लगाना होता है जो कुछ नवीन ज्ञान प्रदान करके मानव जाति की समस्याओं का समाधान करने में सहायता प्रदान कर सकें। अनुसंधान सूचनाओं का संकलन मात्र नहीं हैं, वरन् यह एक ऐसी प्रक्रिया है जिसमें समंकों के संकलन तथा विश्लेषण के द्वारा नवीन तथ्यों व सिद्धान्तों को प्रतिपादित करके ज्ञान भण्डार में वृद्धि करने के लिए वैज्ञानिक विधि का प्रयोग किया जाता है। व्यावहारिक अनुसंधान का उद्देश्य विभिन्न प्रकार की समस्याओं का वैज्ञानिक ढंग से अध्ययन करके ऐसे समाधान खोजना है जो सर्वत्र (Universally) प्रयुक्त किये जाने योग्य हो क्योंकि लगभग समस्त सामाजिक विज्ञानों जैसे शिक्षा शास्त्र मनोविज्ञान, समाजशास्त्र आदि की समस्याएं प्राय: व्यक्तियों से सम्बन्धित होती हैं तथा व्यक्तियों से प्राप्त सूचनाओं के आधार पर ही उन समस्याओं का समाधान करना सम्भव होता है इसलिए अनुसंधानकर्ता को अपने अध्ययन के लिए किसी क्षेत्र या समूह का निर्धारण करना होता है तथा वह उस क्षेत्र या

समूह के व्यक्तियों का अध्ययन करके ही किन्हीं तथ्यों या सिद्धान्तों को प्रतिपादित करता है। अनुसंधानकर्ता के द्वारा चयनित समूह के छोटा होने पर सर्वत्र-प्रयुक्त-योग्य (Universally Applicable) तथ्यों या सिद्धान्तों का प्रतिपादन करना सम्भव नहीं होता है इसलिए चयनित क्षेत्र या समूह प्राय: काफी बड़ा होता है। ऐसी स्थिति में अनुसंधानकर्ता के लिए यह अत्यन्त कठिन हो जाता है कि वह चयनित समूह के प्रत्येक व्यक्ति से सूचनाऐं प्राप्त करके उस समूह के सम्बन्ध में कोई निष्कर्ष प्राप्त करे।

उदाहरण के लिए यदि कोई अनुसंधानकर्ता किसी प्रदेश के माध्यमिक विद्यालयों की कक्षा दस में पढ़ने वाले समस्त छात्रों जिनकी संख्या लाखों में होगी, के लिए किसी तथ्य जैसे औसत लम्बाई अथवा औसत निष्पत्ति अथवा औसत बुद्धि आदि का ज्ञान करना चाहता है तो उसको इन लाखों छात्रों से सूचनाएं संकलित करनी होगी। यह कार्य न केवल अत्यन्त कठिन बल्कि लगभग असम्भव होगा। इस कार्य को सम्पादित करने के लिए धन, समय तथा श्रम की एक विशाल मात्रा की आवश्यकता होगी जिसे कर पाना प्राय: सम्भव नहीं होता है। ऐसी परिस्थिति में अनुसंधानकर्ता प्राय: उस बड़े समूह, जिसका वह अध्ययन करना चाहता है से एक छोटा प्रतिनिधित्व समूह छाँट लेता है तथा इस छोटे समूह का अध्ययन करके बड़े समूह की विशेषताओं का अनुमान लगाने का प्रयास करता है बड़े समूह को अध्ययन की जनसंख्या अथवा समष्टि (Population) अथवा समग्र (Universe) कहते हैं। जबकि छांटे गए छोटे समूह को प्रतिदर्श अथवा न्यादर्श (Sample) कहते हैं। आधुनिक युग में अधिकांश अनुसंधान कार्य प्रतिदर्श विधि (Method of Samples) के द्वारा किये जाते हैं। प्रतिदर्श विधि के प्रयोग का तर्क है कि (*i*) यदि अनुसंधानकर्ता के द्वारा छाँटा गया प्रतिदर्श समष्टि का सही ढंग से प्रतिनिधित्व करता है तब प्रतिदर्श से संकलित सूचनाओं की सहायता से समष्टि समग्र के सम्बन्ध में ठीक-ठीक अनुमान लगाया जा सकता है तथा (*ii*) वैज्ञानिक ढंग से छाँटे गये प्रतिदर्श में वे सभी विशेषताएं होने की सम्भावना होती है जो समष्टि में विद्यमान है तथा ऐसा प्रतिदर्श समष्टि का सही ढंग से प्रतिनिधित्व करने वाला माना जा सकता है। प्रतिदर्श से प्राप्त सूचनाओं की सहायता से समष्टि के सम्बन्ध में प्राप्त सूचनाओं को सांख्यिकीय अनुमान (Statistical Inferences) कहते हैं तथा इन्हें ज्ञात करने के लिए जिन सांख्यिकीय विधियों का प्रयोग किया जाता है उन्हें अनुमानात्मक सांख्यिकीय विधियाँ (Inferential Statistical Methods) कहा जाता है।

अनुमानात्मक सांख्यिकीय के दो प्रमुख कार्य–(*i*) प्रतिदर्शजों (Statistics) के ज्ञात होने पर प्राचलों (Parameters) का अनुमान लगाना तथा (*ii*) परिकल्पनाओं (Hypotheses) का परीक्षण करना है। प्राचलों का अनुमान लगाने तथा परिकल्पनाओं का परीक्षण इन दोनों ही कार्यों में प्रतिचयन वितरणों (Sampling Distributions) की आवश्यकता होती है। प्रस्तुत अध्याय में प्रतिदर्शज व प्राचल, समष्टि व प्रतिदर्श, प्रतिचयन विधियाँ तथा प्रतिचयन वितरणों की चर्चा की गई है।

1.3 प्रतिदर्शज तथा प्राचल (Statistics and Parameters)

अनुसंधानकर्ता किसी समूह की विशेषताओं को मध्यमान, मध्यांक, मानक विचलन, सहसम्बन्ध गुणांक आदि के द्वारा व्यक्त करता है। क्योंकि समूह दो प्रकार के हो सकते हैं (*i*) समष्टि समूह अर्थात ऐसे बड़े समूह जिनके सम्बन्ध में अनुसंधानकर्ता वास्तव में अध्ययन करना चाहता है तथा (*ii*) प्रतिदर्श समूह अर्थात अपेक्षाकृत रूप से ऐसे छोटे समूह जिनका अनुसंधानकर्ता वास्तव में अध्ययन करता है व जिनसे प्राप्त सूचनाओं के आधार पर समष्टि के सम्बन्ध में अनुमान लगाया जाता है। इसलिए समष्टि तथा प्रतिदर्श दोनों के लिए ही विभिन्न विशेषताओं अर्थात मध्यमान, मध्यांक, मानक विचलन, सहसम्बन्ध गुणांक आदि की चर्चा की जा सकती है। वास्तव में व्यवहार में केवल प्रतिदर्श के लिए ही विभिन्न वर्णनात्मक मापों की गणना की जाती है क्योंकि समष्टि तो इतनी बड़ी होती है कि उसके सभी सदस्यों से सूचना प्राप्त करके किसी माप की गणना करना व्यावहारिक रूप में (धन, समय व श्रम की दृष्टि से) सम्भव नहीं होता है तथा यदि समष्टि का अध्ययन किया जा सकता है तब प्रतिदर्श का चयन करने की आवश्यकता ही समाप्त हो जाती है परन्तु जिस तरह से प्रतिदर्श की वर्णनात्मक मापों की गणना की जा सकती है ठीक उसी प्रकार से समष्टि की वर्णनात्मक मापों की भी गणना की जा सकती है। यही कारण है कि समष्टि की वर्णनात्मक मापों की व्यवहार में गणना न करने के बावूजद भी सैद्धान्तिक दृष्टि से समष्टि के सभी सदस्यों से

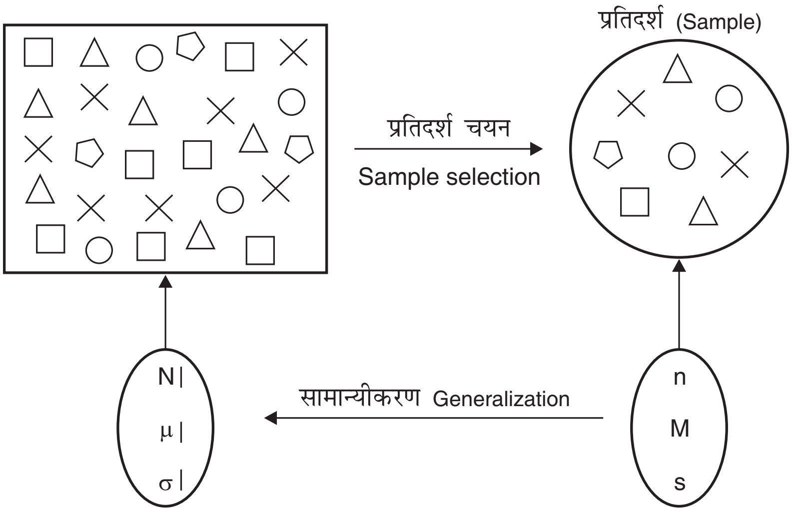
प्राप्त सूचना के आधार पर ज्ञात की जा सकने वाली वर्णनात्मक मापें विद्यमान रहती हैं। प्रतिदर्श के लिए ज्ञात की जाने वाली वर्णनात्मक मापों को **प्रतिदर्शज** (Statistics एक वचन में Statistic) कहते हैं जबकि समष्टि के लिए विद्यमान वर्णनात्मक मापों को प्राचल (Parameters) कहते हैं। उदाहरण के लिए यदि कक्षा दस के एक लाख छात्रों की किसी समष्टि से 500 छात्रों का प्रतिदर्श छाँटा गया हो, तो इन 500 छात्रों के लिए किसी चर (जैसे शैक्षिक उपलब्धि, ऊँचाई, भार आदि) पर प्राप्त अंकों के मध्यमान (अथवा मानक विचलन आदि) का मान प्रतिदर्शज कहलायेगा जबकि समष्टि के सभी एक लाख छात्रों से प्राप्त अंकों के मध्यमान (अथवा मानक विचलन आदि) का मान प्राचल कहलायेगा। प्रतिदर्शजों को प्राय: रोमन अक्षरों (Roman Letters) से व्यक्त करते हैं जबकि प्राचलों को प्राय: ग्रीक अक्षरों (Greek Letters) अथवा रोमन अक्षरों के नीचे Pop लिखकर व्यक्त करते हैं। सारणी 1 में कुछ प्रमुख वर्णनात्मक मापांकों के लिए प्रतिदर्शजों व उनके सापेक्ष प्राचलों हेतु प्रयुक्त किये जाने वाले संकेताक्षरों को प्रस्तुत किया गया है।

सारणी 1

विभिन्न वर्णनात्मक मापांकों के लिए प्रतिदर्शज व प्राचल संकेताक्षर

प्रतिदर्श संकेताक्षर	वर्णनात्मक मापांक	प्राचल संकेताक्षर
n	**आकार**	N
M	**मध्यमान**	M_{pop} या μ (Mu - म्यू)
s	**मानक विचलन**	σ (sigma - सिगमा)
s^2	**प्रसरण**	σ^2 (sigma square - सिगमा वर्ग)
r	**सहसम्बन्ध गुणांक**	r_{pop} या ρ (Rho - रौ)

क्योंकि अनुसंधानकर्ता व्यवहार में प्रतिदर्शजों की गणना ही कर पाता है इसलिए ज्ञात किये गये प्रतिदर्शजों की सहायता से प्राचलों का अनुमान लगाना पड़ता है। प्रतिदर्शजों की सहायता से प्राचलों का अनुमान लगाना ही सांख्यिकीय अनुमान (Statistical Inference) कहा जाता है तथा जिन सांख्यिकीय विधियों का प्रयोग करके प्राचलों के सांख्यिकीय अनुमान लगाये जाते हैं उन्हें अनुमानात्मक सांख्किीय विधियाँ (Inferential Statistical Methods) कहते हैं। सांख्यिकीय का वह भाग जो इन विधियों से सम्बन्धित होता है उसे अनुमानात्मक सांख्यिकीय (Inferential Statistics) कहते हैं।



चित्र 1. प्रतिदर्शज तथा प्राचल का परस्पर सम्बन्ध

1.4 समष्टि एवं प्रतिदर्श (Population and Sample)

पीछे समष्टि तथा प्रतिदर्श शब्दों का प्रयोग किया जा चुका है। इस खण्ड में समष्टि तथा प्रतिदर्श शब्दों को कुछ अधिक यथार्थ ढंग से परिभाषित करने एवं प्रतिदर्श छाँटने की विधि का संक्षिप्त वर्णन करने का प्रयास किया गया है। इस खण्ड में तथा इसके उपरान्त इस पुस्तक में अन्यत्र सभी जगह समष्टि व जनसंख्या को समानार्थक शब्दों के रूप में स्वीकार करते हुए इन दोनों शब्दों के लिए अंग्रेजी पर्याय Population तथा समग्र शब्द के लिए अंग्रेजी पर्याय Universe का प्रयोग किया जायेगा।

अनुसंधान में समष्टि अथवा जनसंख्या (Population) से तात्पर्य उन सभी इकाइयों के समुच्चय (Set of Units) से होता है जिनमें कुछ सामान्य विशेषताएं होती हैं तथा जिनके सम्बन्ध में अनुसंधानकर्ता कुछ निष्कर्ष ज्ञात करना चाहता है। जैसे यदि अनुसंधानकर्ता उत्तर प्रदेश में अध्ययनरत कक्षा दस के छात्रों के सम्बन्ध से कोई निष्कर्ष ज्ञात करना चाहता है तो उसके अध्ययन की समष्टि उ. प्र. में स्थित माध्यमिक विद्यालयों की कक्षा दस के समस्त छात्रों का समुच्च होगा। इसी प्रकार से प्राथमिक विद्यालय अध्यापकों की, चतुर्थ श्रेणी कर्मचारियों की, शिक्षित गृहणियों की, किसी रोग विशेष से पीड़ित व्यक्तियों की समष्टियां हो सकती है। समष्टियाँ सीमित (Finite) भी हो सकती हैं तथा असीमित (Infinite) भी हो सकती हैं। **सीमित समष्टि** (Finite Population) वे समुच्चय होते हैं जिनमें इकाइयों की संख्या सुनिश्चित होती है। जैसे भारत में माध्यमिक अध्यापकों की संख्या, किसी विश्वविद्यालय में अध्यापकों की संख्या, किसी महानगर में परिवारों की संख्या, किसी बड़े चिकित्सालय में रोगियों की संख्या आदि सीमित समष्टि के उदाहरण हैं। इसके विपरीत **असीमित समष्टि** (Infinite Population) वे जनसंख्या हैं जिनमें इकाइयों की संख्या अनिश्चित अथवा अनन्त होती है। जैसे आकाश में तारों की संख्या, किसी भाषा में बोले जाने वाले शब्दों की संख्या आदि असीमित समष्टि के उदाहरण हैं। यद्यपि वास्तव में कोई भी समष्टि असीमित नहीं होती है परन्तु कभी-कभी समष्टि में इकाइयों की संख्या इतनी अधिक होती है कि उसे गिनना व्यवहार में लगभग असम्भव होता है जिसके कारण ऐसी समष्टियों को व्यावहारिक दृष्टि से असीमित मानना तर्कसंगत प्रतीत होता है। जैसे आकाश में तारों की कोई-न-कोई संख्या अवश्य होगी परन्तु इसे ज्ञात करना सम्भव न होने के कारण इसे असीमित समष्टि माना जा सकता है। समष्टियों को परिकल्पित समष्टि तथा यथार्थ समष्टि में भी विभक्त किया जा सकता है। जो समष्टि मूर्त रूप में विद्यमान होती है उसे **यथार्थ समष्टि** (Existent Population) कहते हैं। दूसरे शब्दों में कहा जा सकता है कि यथार्थ समष्टि की समस्त इकाइयां वास्तव में विद्यमान होती हैं। जैसे छात्रों की समष्टि, पुस्तकों की समष्टि अथवा मनोरोगियों की समष्टि आदि यथार्थ समष्टियों के उदाहरण हैं। इसके विपरीत जो समष्टि मूर्त रूप से विद्यमान नहीं होती हैं उन्हें **परिकल्पित समष्टि** (Hypothetical Population) कहते हैं स्पष्ट है कि परिकल्पित समष्टि की इकाइयों की केवल कल्पना ही की जा सकती है। वे सभी वास्तव में विद्यमान नहीं होती हैं। जैसे सिक्के को उछालने पर चित्त (Head) व पट (Tail) आने की सभी सम्भावित आवृत्तियों की समष्टि एक परिकल्पित समष्टि है। सिक्के को उछाले बिना ही ऐसी समष्टि की परिकल्पना की जा सकती है

प्रतिदर्श को न्यादर्श भी कहते हैं। प्रतिदर्श (Sample) समष्टि का वह छोटा भाग है जिसे अनुसंधानकर्ता के द्वारा वास्तविक अध्ययन के लिए चयनित किया जाता है। प्रतिदर्श समष्टि का प्रतिनिधित्व सूक्ष्म-रूप (Representative Miniature) होता है तथा इससे प्राप्त सूचनाओं का सामान्यीकरण करके समष्टि के बारे में अनुमान लगाया जाता है। स्पष्ट है कि प्रतिदर्श समष्टि की विशेषताओं का प्रतिबिम्ब होता है। प्रतिदर्श के अंग्रेजी पर्याय Sample का उद्भव लैटिन के Exemplum शब्द से हुआ जिसका अर्थ है Exmple अर्थात उदाहरण। इस शाब्दिक अर्थ से भी संकेत मिलता है कि प्रतिदर्श समष्टि की कुछ ऐसी इकाइयों का संकलन है जिन्हें समष्टि की विशेषताओं को स्पष्ट करने के लिए उदाहरण स्वरूप चुना जाता है। अत: कहा जा सकता है कि प्रतिदर्श वास्तव में समष्टि का अध्ययन हेतु चयनित किया गया वह अंश है जिसका अनुसंधानकर्ता वास्तव में अध्ययन करता है तथा जिससे प्राप्त प्रतिदर्श की विशेषताओं (प्रतिदर्शजों) का सामान्यीकरण करके समष्टि की विशेषताओं (प्राचलों) का अनुमान लगाया जाता है। समष्टि की इकाइयों को प्रतिचयन इकाइयाँ (Sampling Units) कहते हैं तथा किसी

समष्टि की सभी प्रतिचयन इकाइयों के समूह को प्रतिचयन ढाँचा (Sampling Frame) कहते हैं। प्रतिदर्श में चुनी गई इकाइयों को प्रतिदर्श इकाइयाँ (Sample Units) कहते हैं। प्रतिदर्श व समष्टियाँ व्यक्तियों अथवा वस्तुओं के समूह भी हो सकते हैं तथा इनका प्रयोग व्यक्तियों या वस्तुओं के प्राप्तांकों के समूहों के लिए भी किया जा सकता है।

प्रतिदर्श के आधार पर समष्टि के विषय में सांख्यिकीय अनुमान सदैव एक अनिवार्य मान्यता के साथ प्रारम्भ होता है। इस अनिवार्य मान्यता के अनुसार प्रतिदर्श समष्टि का सही ढंग से प्रतिनिधित्व करता है अर्थात प्रतिदर्श में ठीक वही गुण या विशेषताएं हैं जो समष्टि में हैं। यदि प्रतिदर्श समष्टि का ठीक ढंग से प्रतिनिधित्व नहीं कर पाता है अर्थात यदि प्रतिदर्श में समष्टि की वास्तविक विशेषता परिलक्षित या प्रतिबिम्बित नहीं होती है तब अनुमानात्मक सांख्यिकीय विधियों से प्राप्त अनुमान त्रुटिपूर्ण हो सकते हैं। दूसरे शब्दों में तब अनुमानित प्राचल वास्तविक प्राचलों से भिन्न हो सकते हैं। जैसे यदि उत्तर प्रदेश में कक्षा दस में पढ़ने वाले छात्रों की औसत ऊँचाई ज्ञात करने के लिए केवल शहरों में स्थित माध्यमिक विद्यालयों की कक्षा दस में अध्ययनरत छात्रों का प्रतिदर्श ले लिया जाये तब यह प्रतिदर्श उत्तर प्रदेश में कक्षा दस में पढ़ने वाले छात्रों की समष्टि का सही प्रतिनिधित्व नहीं कर सकेगा क्योंकि ग्रामीण क्षेत्रों में स्थित माध्यमिक विद्यालयों के छात्रों को इसमें सम्मिलित नहीं किया गया है। जब तक सभी प्रकार के छात्रों को प्रतिदर्श में नहीं चुना जायेगा तब तक प्रतिदर्श अपनी समष्टि की विशेषताओं को ठीक ढंग से परिलक्षित नहीं कर सकेगा तथा इस प्रकार से छाँटे गये अप्रतिनिधित्व प्रतिदर्श (Unrepresentative Sample) से प्राप्त परिणामों को सामान्यीकरण समष्टि के लिए करना त्रुटिपूर्ण होगा। अत: यह आवश्यक है कि अनुसंधानकर्ता समष्टि से किसी प्रतिनिधित्व प्रतिदर्श का चयन करें जिससे प्रतिदर्श से प्राप्त परिणामों का तर्क संगत ढंग से समष्टि के लिए सामान्यीकरण किया जा सके। प्रतिदर्श के प्रतिनिधित्व होने अथवा नहीं होने का प्रश्न उठने का प्रमुख कारण सामाजिक विज्ञानों (Social Sciences) में प्रयुक्त समष्टियों की विभिन्न इकाइयों में वैभिन्नता का होना है। भौतिक विज्ञानों (Physical Sciences) जैसे रसायन शास्त्र, चिकित्सा, कृषि आदि में भी प्रतिदर्शों का उपयोग किया जाता है परन्तु वहाँ पर प्रतिदर्श की प्रतिनिधित्वता पर शंका नहीं की जाती है क्योंकि वहाँ प्रयुक्त समष्टियों में विद्यमान सभी इकाइयाँ प्राय: पूर्णत: एक जैसी होती हैं। किसी तत्व, यौगिक या मिश्रण के सभी कण अथवा किसी व्यक्ति के खून की सभी बूंदें अथवा किसी खेत में उत्पादित गेहूँ के ढेर के सभी दाने लगभग एक समान होते हैं। जिसके कारण इनकी केवल एक ही इकाई समष्टि का उचित प्रतिनिधित्व करने में समर्थ होती है। ऐसी समष्टियों को **समजातीय समष्टि** (Homogenious Populations) कहा जाता है परन्तु सामाजिक विज्ञानों जैसे मनोविज्ञान, शिक्षा शास्त्र, समाजशास्त्र आदि में प्रयुक्त समष्टियाँ प्राय: जीवधारियों (मानव या पशु आदि) के समूह होते हैं जिनकी विभिन्न इकाइयाँ परस्पर भिन्न-भिन्न होती हैं। जैसे किसी समूह के छात्रों की बुद्धि या व्यक्तित्व या रुचियाँ आदि एक-दूसरे से पर्याप्त भिन्न होती हैं ऐसी समष्टियों को **विषमजातीय समष्टि** (Hetrogeneous Population) कहते हैं। इस प्रकार की समष्टियों से ऐसे प्रतिदर्श का चयन करना अपेक्षाकृत कठिन हो जाता है जो अपनी समष्टि का उचित ढंग से प्रतिनिधित्व कर सके। कोई प्रतिदर्श अपनी समष्टि का प्रतिनिधित्व कर रहा है अथवा नहीं इसका निर्धारण करना लगभग असम्भव सा कार्य है। ऐसा करने के लिए प्रतिदर्श तथा समष्टि दोनों की ही विशेषताओं का पूर्ण ज्ञान आवश्यक है क्योंकि तब ही दोनों से प्राप्त परिणामों की तुलना की जा सकती है। यदि समष्टि की विशेषताएं सीधे-सीधे ज्ञात की जा सकती हैं तब प्रतिदर्श के चयन व उससे प्राप्त परिणामों का सामान्यीकरण करने का प्रश्न ही नहीं उठता है। वास्तव में सांख्यिकीय में सदैव ही प्रतिदर्श के अपने समष्टि का प्रतिनिधित्व करने की प्रायिकता देखी जाती है न कि उसका निर्धारण किया जाता है। यही कारण है कि अनुमानात्मक सांख्यिकीय विधियों से प्राप्त समस्त परिणाम प्रायिकता के रूप में ही व्यक्त किये जाते हैं। क्योंकि सांख्यिकीय दृष्टि से किसी रेन्डम प्रतिदर्श (Random Sample) के द्वारा अपनी समष्टि का प्रतिनिधित्व करने की प्रायिकता सर्वाधिक होती है, इसलिए रैन्डम प्रतिदर्श को प्रतिनिधित्व प्रतिदर्श स्वीकार करके ही सामान्यीकरण किया जाता है। वास्तव में अनुमानात्मक सांख्यिकीय की अधिकांश बहुप्रचलित विधियाँ रैन्डम प्रतिदर्श की मूलभूत मान्यता (Fundamental Assumption) पर आधारित होती है। आगे के पृष्ठों पर प्रतिदर्श चयन की विभिन्न विधियों का संक्षेप में वर्णन किया जा रहा है।

सरल रैन्डम प्रतिचयन

(Simple Random Sampling)

किसी समष्टि से सर्वाधिक सम्भाव्य प्रतिनिधित्व वाला प्रतिदर्श छाँटने की एक सरल तथा वस्तुनिष्ठ विधि रैन्डम प्रतिदर्श विधि है। जिसे सरल रैन्डम विधि (Simple Random Method) भी कहा जाता है। रैन्डम प्रतिदर्श (Random Sample) शब्द युग्म में रैन्डम (Random) शब्द से तात्पर्य अस्तव्यस्त (Haphazard), अनियोजित (Unplanned) अथवा लापरवाही (Hit and Miss) से नहीं है वरन सांख्यिकीय मानदण्डों की दृष्टि से सुमान्य प्रतिचयन विधि को इंगित करने के लिए रैन्डम शब्द का प्रयोग किया जाता है। रैन्डम प्रतिदर्श विधि की प्रमुख विशेषता है कि इसमें समष्टि की प्रत्येक इकाई के प्रतिदर्श में चुने जाने की समान प्रायिकता होती है। अत: यदि किसी प्रतिदर्श का चयन इस ढंग से किया गया है कि सभी इकाइयों के चयन की प्रायिकता समान थी तब ऐसे प्रतिदर्श को रैन्डम प्रतिदर्श (Random Sample) कहा जाता है।

यहाँ पर सरल रैन्डम प्रतिदर्श तथा रैन्डम प्रतिदर्श के अन्तर व सम्बन्ध को स्पष्ट करना उचित ही होगा। सरल रैन्डम प्रतिदर्श (Simple Random Sample) में वास्तव में समष्टि की विभिन्न इकाइयों के चयन की प्रायिकता स्थिर (Constant) रहती है अर्थात विभिन्न इकाइयों का चयन परस्पर स्वतन्त्र होता है। तथा कुछ इकाइयों के चयन से शेष इकाइयों के चयन की प्रायिकता प्रभावित नहीं होती है। उदाहरण के लिए 10 इकाइयों की किसी समष्टि में से 2 इकाइयों का एक प्रतिदर्श छाँटना है। यदि 10 इकाइयों में से 2 इकाइयों को एक-एक करके छाँटा जायेगा तब प्रथम बार सभी इकाइयों के चयन की प्रायिकता 1/10 होगी जबकि दूसरी बार शेष 9 इकाइयों के चयन की प्रायिकता परिवर्तित हो कर 1/9 हो जायेगी। यह रैन्डम प्रतिदर्श चयन तो है परन्तु सरल रैन्डम प्रतिदर्श चयन नहीं है परन्तु यदि पहली बार छांटी गई इकाई को पुन: शेष इकाइयों में मिला दिया तथा फिर दूसरी इकाई छाँटी जाये तो प्रतिदर्श चयन की विधि सरल रैन्डम हो जायेगी। प्रतिदर्श चयन की इन विधियों को **प्रतिस्थापन रहित प्रतिचयन** (Sampling Without Replacement) तथा **प्रतिस्थापन सहित प्रतिचयन** (Sampling with Replacement) भी कहा जाता है। पाठकों को यह स्पष्ट हो गया होगा कि असीमित समष्टियों (Infinite Populations) से छाँटे गये रैन्डम प्रतिदर्श सदैव ही सरल रैन्डम प्रतिदर्श भी होंगे क्योंकि असीमित समष्टि की स्थिति में कुछ इकाइयों का चयन शेष इकाइयों के चयन की प्रायिकता को प्रभावित नहीं कर पाता है परन्तु सीमित समष्टियों (Finite Populations) की स्थिति में प्रतिचयन प्रतिस्थापन रहित तथा सहित दोनों ही प्रकार का हो सकता है।

प्रतिस्थापन रहित प्रतिचयन की स्थिति में यह स्पष्ट कहना उचित ही होगा कि कुछ इकाइयों के चयन के उपरान्त शेष इकाइयों का चयन वास्तव में प्रतिबन्धात्मक (Conditional) होता है अर्थात अवशेष इकाइयों में से किसी/किन्हीं इकाई का चयन तब ही सम्भव है जब वे इकाई इससे पूर्व चयनित न की गई हों। इकाइयों के चयन की कुल प्रतिबन्धात्मक प्रायिकता (Conditional Probability) का मान सभी इकाइयों के लिए स्थिर रहता है जिसे प्रायिकता के गुणनफल नियम (Multiplication Law) से ज्ञात किया जा सकता है। जैसे उपरोक्त उदाहरण में दूसरी बार में किसी इकाई के चयन की प्रतिबन्धित प्रायिकता का मान प्रथम बार में उसके न चुने जाने को प्रायिकता व दूसरी बार में चुने जाने की प्रायिकता के गुणनफल के बराबर अर्थात $(9/10) \times (1/9) = 1/10$ ही होगा। इसी प्रकार से तीसरी बार में किसी इकाई के चुने जाने के कुल प्रतिबन्धित प्रायिकता का मान उसके प्रथम बार में न चुने जाने, दूसरी बार में न चुने जाने तथा तीसरी बार में चुने जाने की प्रायिकताओं के गुणनफल के बराबर अर्थात $(9/10) \times (8/9) \times (1/8) = 1/10$ होगा।

रैन्डम प्रतिदर्श को समष्टि से सम्भाव्य समस्त प्रतिदर्शों में से रैन्डम विधि के द्वारा छांटे गये प्रतिदर्श के रूप में भी परिभाषित किया जा सकता है। यदि किसी समष्टि से n आकार के समस्त सम्भाव्य प्रतिदर्शों में से किसी एक को इस प्रकार से छांटा गया है कि सभी प्रतिदर्शों के चयन की प्रायिकता समान हो तो इस प्रकार से छांटे गये प्रतिदर्श को रैन्डम प्रतिदर्श कहा जा सकता है। जैसे यदि पाँच इकाइयों (अ, ब, स, द, य) वाली समष्टि से 2 ईकाई वाले प्रतिदर्श का चयन करना हो तब कुल दस सम्भाव्य प्रतिदर्श– (अ, ब), (अ, स), (अ, द), (अ, य), (ब, स), (ब, द), (ब, य), (स, द), (स, य) तथा (द, य) होंगे। यदि इन दस प्रतिदर्शों में से किसी एक का

चयन इस प्रकार से किया जाता है कि प्रत्येक के चयन की प्रायिकता समान हो तो चयनित प्रतिदर्श को रैन्डम प्रतिदर्श कहा जायेगा। रैन्डम प्रतिदर्श का यह केवल सैद्धान्तिक पक्ष है क्योंकि इस प्रकार से सभी सम्भाव्य प्रतिदर्श बनाकर उनमें से किसी एक का चयन करना व्यावहारिक दृष्टि से सम्भव नहीं है। नि:सन्देह यह प्रतिस्थापन रहित प्रतिचयन हैं। यद्यपि सैद्धान्तिक रूप से अनुमानात्मक सांख्यिकीय विधियों के लिए प्रतिदर्श का रैन्डम होना एक अनिवार्य आवश्यकता है तथापि रैन्डम विधियों से न चुने गये प्रतिदर्शों से भी समष्टि के बारे में अनुमान लगाये जा सकते हैं। परन्तु ऐसा करते समय यह सुनिश्चित करना उचित ही होता है कि प्रतिदर्श का चयन किसी अवाँछित कारण से प्रभावित नहीं हो रहा है। किसी समष्टि से रैन्डम प्रतिदर्श चुनने की कई विधियाँ हैं। इनमें से दो विधियाँ–लाटरी विधि (Lottery Method) तथा रैन्डम अंक सारणी विधि (Table of Random Numbers Method) सर्वाधिक सरल, लोकप्रिय तथा उपयुक्त है। इन दोनों ही विधियों में समष्टि की समस्त ईकाइयों का ज्ञान होना आवश्यक है जिसके कारण इन्हें केवल सीमित समष्टियों से प्रतिदर्श छांटने के लिए उपयोग मे लाया जा सकता है। लाटरी विधि में समष्टि की सभी प्रतिचयन इकाइयों (Sampling Units) के नाम अथवा क्रम संख्या अलग-अलग कागज के छोटे टुकड़ों पर लिखकर उन्हें मोड़कर पर्ची तैयार कर ली जाती हैं। इस प्रकार से समष्टि की इकाइयों की संख्या के बराबर कागज की पर्चियाँ तैयार हो जाती हैं जिन्हें किसी थैले में रखकर व अच्छी तरह से मिलाकर बिना देखे एक-एक करके ठीक उतनी ही पर्चियाँ निकाल लेते हैं जितनी इकाइयाँ प्रतिदर्श के लिए छांटनी होती हैं। इस प्रकार से छांटी गई पर्चियों पर लिखे नाम अथवा क्रमसंख्या वाली ईकाइयाँ ही प्रतिदर्श की रचना करती हैं। जैसे यदि 500 छात्रों की किसी समष्टि से 20 छात्रों का एक रैन्डम प्रतिदर्श छांटना हो तो पहले समष्टि के सभी 500 छात्रों के नामों को अलग-अलग कागज की पर्चियों पर लिखकर इन पर्चियों पर लिखकर इन पर्चियों को एक ही प्रकार से मोड़कर किसी थैले में डालकर अच्छी तरह से हिला लेंगे। तत्पश्चात इस थैले से एक-एक करके 20 पर्ची निकाल लेंगे। जिन 20 छात्रों के नाम इन पर्चियों पर लिखें होंगे वे 20 छात्र ही प्रतिदर्श की संरचना करेंगे तथा इन 20 छात्रों के प्रतिदर्श पर अध्ययन करके सम्पूर्ण समष्टि (अर्थात 500 छात्रों का समूह) के लिए अनुमान ज्ञात किये जायेंगे। लाटरी द्वारा रैन्डम प्रतिदर्श छाँटने के लिए ढोल घुमाकर (Rotating Drum) या यान्त्रिक विधि (Mechanical Method) से अंक चक्र यन्त्र चलाकर (Moving Number Wheel) इकाइयों के चयन करने की विधि का प्रयोग भी किया जाता है।

रैन्डम अंक सारणी अधिक वस्तुनिष्ठ व व्यवस्थित ढंग से रैन्डम प्रतिदर्श प्राप्त करने का एक साधन है। रैन्डम अंक सारणी वास्तव में 0 से 9 तक के दस अंकों की कुछ पंक्तियों व स्तम्भों में इस प्रकार से बारम्बार प्रस्तुति है कि विभिन्न पंक्तियों तथा स्तम्भों में सभी दस अंकों की आवृत्ति लगभग समान होती है एवं किसी भी अंक के दायें, बायें, ऊपर अथवा नीचे कौन सा अंक आयेगा इसको अनुमान लगाना असम्भव होता है। दूसरे शब्दों में रैन्डम अंक सारणी में से किसी अंक को बिना देखे चयनित करने पर उसके 0, 1, 2, 3, 4, 5, 6, 7 ,8 , या 9 होने की सम्भावना समान होती है। रैन्डम अंक सारणी में पंक्तियों व स्तम्भों की संख्या कितनी भी हो सकती है। वास्तव में रैन्डम अंक सारणी का मनचाहा विस्तार किया जा सकता है। सुविधा के लिए रैन्डम अंक सारणी के स्तम्भों को (तथा कभी कभी पंक्तियों को भी) चार या पांच के समूहों में प्रस्तुत करते हैं। तब रैन्डम अंक सारणी के विभिन्न स्तम्भ चार या पांच अंकों वाली रैन्डम संख्याओं को इंगित करते हैं। सबसे पहले एल. एच. सी. टिपेट (L. H. C. Tippet) ने जनगणना के आँकड़ों से 41600 अंकों को छांटकर उनकी सहायता से चार-चार अंकों वाली 10400 रैन्डम संख्याएं तैयार की थीं। बाद में कैन्डल (M.C. Kendall) व स्मिथ (B.B. Smith) तथा फिशर (R.A Fisher) व येटस (F. Yates) आदि ने भी रैन्डम अंकों/संख्याओं की सारणियाँ तैयार कीं। कम्प्यूटर के आविष्कार से पूर्व रैन्डम अंक संख्या सारणी बनाना एक कठिन कार्य था परन्तु कम्प्यूटर ने इस कार्य को सरल कर दिया है। आज कम्प्यूटर की सहायता से रैन्डम अंक प्राप्त करने की अनेक विधियां जैसे केन्द्रीय वर्गीकरण विधि (Central Squaring Method), घात अवशेष विधि (Power Residue Method) आदि उपलब्ध हैं परन्तु अनुसंधानकर्ता को रैन्डम अंक/संख्या सारणी बनाने की आवश्यकता नहीं है क्योंकि अनेक रैन्डम अंक/संख्या सारणी उपलब्ध है। अनुसंधान विधियों तथा सांख्यिकी की लगभग सभी पुस्तकों में भी रैन्डम अंक/संख्या सारणी दी गई होती है तथा अनुसंधानकर्ता इनका प्रयोग करके अपने प्रतिदर्श का चयन कर सकता है। कुछ सांख्यिकीय कम्प्यूटर के द्वारा तैयार

किये गये रैन्डम अंकों/संख्याओं को छदम रैन्डम अंक/संख्या (Pseudo Random Digits Numbers) मानते हैं क्योंकि ये पूर्णतया ज्ञात गणना (Completely Determined Calculation) विधियों से तैयार (Generate) किये जाते हैं तथा गणना विधि व प्रारम्भ बिन्दु के निर्धारण के तुरन्त बाद सभी अंक/संख्या निश्चित हो जाती हैं।

नोट

रैन्डम अंक/संख्या सारणी की सहायता से प्रतिदर्श का चयन करने के लिए सबसे पहले समष्टि की सभी इकाइयों की सूची बनाकर प्रत्येक इकाई को एक-एक क्रम संख्या दे दी जाती है। सबसे बड़ी क्रम संख्या में जितने अंक (Digits) होते हैं ठीक उतने ही अंकों की कोई संख्या आँख बन्द करके रैन्डम अंक संख्या/सारणी मे कहीं पर चयनित कर लेते हैं तथा फिर उस संख्या के ऊपर या नीचे या दायें या बायें ओर चलकर उतने ही अंकों की संख्याएं लेते जाते हैं। इस प्रकार से प्राप्त संख्याओं में से वे सभी संख्याएं, जिनका मान समष्टि की सबसे बड़ी क्रम संख्या के बराबर अथवा कम होता है, प्रतिदर्श में सम्मिलित की जाने वाली इकाइयों की क्रम संख्या होती है। यदि कोई संख्या एक से अधिक बार आती है तब केवल पहली बार ही उसे लेते हैं तथा बाद में छोड़ते जाते हैं। इस प्रकार से प्रतिदर्श के आकार के अनुरूप वांछित संख्या में इकाइयों का चयन कर लेते हैं जो प्रतिदर्श का निर्माण करती हैं। दो प्रश्न उठते हैं जिनका निराकरण वस्तुनिष्ठ ढंग से रैन्डम सारणी के प्रयोग के लिए आवश्यक होगा-प्रथम, सारणी में प्रथम संख्या का चयन कैसे किया जाये तथा द्वितीय, सारणी में किस दिशा में चला जाये। रैन्डम अंक/संख्या सारणी में प्रथम संख्या का चयन करने के लिए पहले सारणी के किसी पृष्ठ का चयन करना होगा। इसके लिए या तो लाटरी डाली जा सकती है अथवा प्रतिदर्श चयन की दिनांक, घन्टा, मिनट में से किसी एक का प्रयोग किया जा सकता है। यदि दिनांक (अथवा घन्टा अथवा मिनट जिसका भी प्रयोग करना हो) पृष्ठों की संख्या से अधिक हो तो दिनांक के अंकों को जोड़कर प्राप्त संख्या वाले पृष्ठ से प्रारम्भ किया जा सकता है। नौ अथवा इससे अधिक पृष्ठों वाली सारणी के लिए इस विधि को सरलता से प्रयोग में ला सकते हैं। रैन्डम सारणी के प्रारम्भ पृष्ठ के चयन के उपरान्त प्रवेश बिन्दु (Point of Entry) का चयन करना होगा। इसके लिए आँखें बन्द करके पेन्सिल को सारणी में कहीं पर टिका सकते हैं तथा जिस अंक पर पेन्सिल टिकती है उसके अगल-बगल के अंकों से मिलकर बनने वाली संख्या प्रवेश संख्या (Entry Number) हो सकती है। दायें ओर के अंकों को लेकर प्रवेश संख्या बनानी है अथवा बाईं और के अंकों को लेकर, इसके लिए घड़ी के सैकन्ड वाली सुई (Second Hand of clock) की दिशा का प्रयोग किया जा सकता है। रैन्डम सारणी के पृष्ठ पर पेन्सिल टिकाने के तुरन्त बाद सैकन्ड की सुई को देखें। यदि यह घड़ी के दायें भाग में 12 से 6 के बीच हो तो दाईं ओर के अंकों को लेकर प्रथम संख्या बनाई जा सकती है। यदि सैकन्ड की सुई बायें भाग में 12 से 6 के बीच हो तो बाईं और के अंकों को लेकर प्रथम संख्या बनाई जा सकती है। प्रवेश संख्या के निर्धारण के उपरान्त यह निर्धारित करना होगा कि किस दिशा में बढ़ा जाये। इसके लिए भी घड़ी की सेकेन्ड की सुई के 11, 12, या 1 पर होने पर ऊपर की और, सेकेन्ड की सुई के 2, 3 या 4 पर होने पर दाईं और, सेकेन्ड की सुई के 5, 6 या 7 पर होने पर नीचे की ओर तथा सेकेन्ड की सुई के 8, 9, या 10 पर होने पर बाईं ओर चलने का निश्चय वस्तुनिष्ठ निर्णय हो सकता है। रैन्डम सारणी की सहायता से प्रतिदर्श का चयन करने की विधि निम्न उदाहरण से स्पष्ट हो सकेगी।

उदाहरण के लिए माना कि किसी विद्यालय में अध्ययनरत 500 छात्रों की समष्टि ($N = 500$) से 20 छात्रों का एक रैन्डम प्रतिदर्श ($n = 20$) छाँटना है। सबसे पहले समष्टि के 500 छात्रों को क्रमबद्ध करके 001 से 500 तक की क्रम संख्याएं प्रदान की जायेंगी। क्योंकि सबसे बड़ी क्रम संख्या 500 तीन अंकों की है, इसलिए तीन अंकों वाली रैन्डम संख्याओं का चयन करना होगा। रैन्डम अंक सारणी के दूसरे पृष्ठ पर दी 54वीं पंक्ति के स्तम्भ छह, सात व आठ पर स्थित अंकों से प्रारम्भ करके नीचे की ओर चलकर रैन्डम संख्याएं छाँटने का निश्चय किया जाये तब निम्न संख्याएं मिलेंगी–

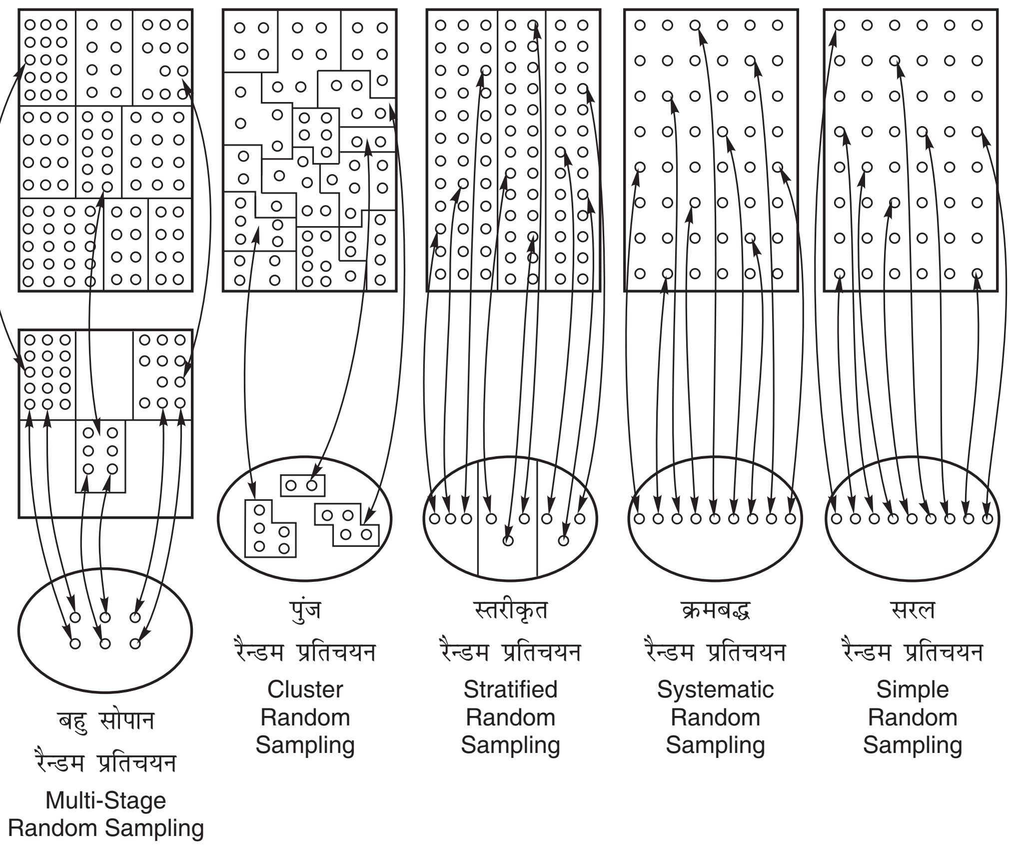
940	746	169*	767	897	299*
561	455*	332*	532	054*	585
959	571	630	279*	345*	862
400*	007*	477	442*	911	929
805	224*	764	938	461*	729

477*	946	041*	179*	520	219*
804	399*	202*	909	895	
566	967*	854	416*	626	

उपरोक्त सूची में अन्तिम संख्या 219 है। अब अन्य संख्याओं की आवश्यकता नहीं है क्योंकि इस संख्या तक बीस ऐसी संख्याएं मिल गई हैं जो प्रतिदर्श के लिए आवश्यक है। तारांकित संख्याएं (Starred Numbers) समष्टि की सूची की वे क्रम संख्याएं हैं जिन्हें प्रतिदर्श में सम्मिलित करना है। स्पष्ट है कि केवल 001 से 500 तक की क्रम संख्याओं को तारांकित किया गया है तथा दोहराई गई संख्याओं को केवल एक बार चुना गया है। अत: रैन्डम प्रतिदर्श के लिए चयनित क्रम संख्याएं निम्नवत लिखी जा सकती हैं–

400,	477,	455,	007,	224,	399,	367,	169,	332,	041,
202,	279,	442,	179,	416,	054,	345,	461,	299,	219,

यदि सारणी की अन्तिम पंक्ति तक पहुँचने पर भी आवश्यक क्रम संख्याएं नहीं मिल पाती हैं तब अन्तिम पंक्ति में दायें या बायें घूमकर फिर उपर की ओर से संख्याएं छांटी जा सकती हैं। दायें या बायें घूमने के निर्णय के लिए भी घड़ी की सेकेन्ड वाली सुई की दिशा का प्रयोग किया जा सकता है। समष्टि में 100,1000 या 10,000 आदि इकाइयाँ होने पर क्रम देना क्रमश: 00,000 या 0000 से प्रारम्भ करना उचित रहता है क्योंकि तब सभी क्रम संख्याएं क्रमश: 2, 3 या 4 अंकों की होगी जिससे संख्याएं छाँटने का कार्य सरल हो जाता है।



चित्र 2. प्रतिचयन विधियाँ

प्रतिदर्श का चयन करने की उपरोक्त वर्णित विधियों के अतिरिक्त कुछ अन्य ऐसी विधियाँ भी हैं जिनके द्वारा प्रतिनिधित्व प्रतिदर्श (Representative Sample) का चयन किया जा सकता है। क्रमबद्ध रैन्डम प्रतिदर्श

(Systematic Random Sample), स्तरीकृत रैन्डम प्रतिदर्श (Stratified Random Sample), बहु-सोपान रैन्डम प्रतिदर्श (Multi-Stage Random Sample) तथा पुन्ज रैन्डम प्रतिदर्श (Cluster Random Sample) सांख्यिकीय दृष्टि से काफी सीमा तक रैन्डम प्रतिदर्श के समतुल्य होते हैं तथा इनका प्रयोग अनुमानात्मक सांख्यिकीय विधियों के लिए किया जा सकता है।

क्रमबद्ध रैन्डम प्रतिचयन

(Systematic Random Sampling)

क्रमबद्ध रैन्डम प्रतिचयन विधि (Systematic Random Sampling Method) में भी सबसे पहले समष्टि की सभी इकाइयों की सूची बनाकर इकाइयों को क्रम संख्याएं दे दी जाती हैं। तत्पश्चात समष्टि व प्रतिदर्श के आकारों का अनुपात ($N/n = k$) ज्ञात करके देख लिया जात है कि समष्टि की प्रत्येक कितनी इकाइयों में से प्रतिदर्श के लिए एक इकाई चयनित करनी है। दूसरे शब्दों में k बताता है कि प्रतिदर्श की प्रत्येक इकाई समष्टि की कितनी इकाइयों का प्रतिनिधित्व करती है। अब समष्टि की प्रथम k इकाइयों में से एक इकाई को लाटरी अथवा रैन्डम अंक सारणी की सहायता से चुन लेते हैं तथा फिर उस इकाई के बाद प्रत्येक kवें अन्तराल पर स्थित इकाइयों का चयन करते जाते हैं। इस प्रकार से चुनी गई इकाइयां ही प्रतिदर्श का निर्माण करती हैं। जैसे 1 से 500 तक क्रमबद्ध सूची के 500 व्यक्तियों की समष्टि से 20 व्यक्तियों का क्रमबद्ध रैन्डम प्रतिदर्श चुनते समय समष्टि व प्रतिदर्श के आकारों का अनुपात (k) का मान 25 होगा अर्थात समष्टि के प्रत्येक 25 व्यक्तियों के लिए एक व्यक्ति का चयन करना है। अब समष्टि सूची के प्रथम 25 व्यक्तियों (क्रम संख्या 1 से 25 तक) में से एक का चयन रैन्डम विधि से करना होगा। माना यह 12वीं क्रम संख्या है। अब 12 के उपरान्त प्रत्येक 25 वीं क्रम संख्या अर्थात 37, 62, 87 आदि को चुनते जायेंगे। इस प्रकार से चयनित बीस संख्याएं अर्थात 12, 37, 62, 87, ------ 462 व 487 प्रतिदर्श में सम्मिलित किये जाने वाले व्यक्तियों की क्रम संख्याएं हैं। क्योंकि इस विधि में केवल प्रथम क्रमसंख्या को रैन्डम ढंग से चुनते हैं तथा बाद की सभी क्रम संख्याऐं क्रमबद्ध ढंग से चुनी जाती हैं, इसलिए इस विधि को क्रमबद्ध रैन्डम प्रतिचयन विधि कहते हैं तथा इस विधि से चयनित प्रतिदर्श को क्रमबद्ध रैन्डम प्रतिदर्श कहते हैं। यह विधि रैन्डम विधि से सरल तथा अल्प समयव्ययी है।

स्तरीकृत रैन्डम प्रतिचयन

(Stratified Random Sampling)

कभी-कभी समष्टि को कुछ सुस्पष्ट उपसमूहों अथवा स्तरों (Strata) में बाँटा जा सकता है तथा अनुसंधानकर्ता उन सभी उपसमूहों अथवा स्तरों से इकाइयों का चयन सुनिश्चित करना चाहता है। तब प्राय: स्तरीकृत रैन्डम प्रतिचयन विधि का प्रयोग किया जाता है। जैसे उच्च माध्यमिक स्तर के छात्रों को कक्षा 6, कक्षा 7 व कक्षा 8 के छात्रों के तीन उपसमूहों में अथवा ग्रामीण व शहरी छात्रों के दो स्तरों में अथवा लड़के व लड़कियों के दो स्तरों में अथवा उच्च, औसत व निम्न बुद्धि छात्रों के तीन वर्गों में बाँटा जा सकता है। इसी प्रकार से परिवारों को उनके सामाजिक व आर्थिक स्तर के आधार पर अथवा शैक्षिक स्तर के आधार पर स्तरीकृत किया जा सकता है तथा नागरिकों का भौगोलिक क्षेत्र, भाषा, धर्म या जाति के आधार पर स्तरीकरण किया जा सकता है। स्पष्ट है कि समष्टि को विभिन्न स्तरों में बाँटने का परिस्थितिनुसार आधार जैसे लिंगभेद, कक्षा-भेद, आर्थिक स्तर, पारिवारिक स्तर, भौगोलिक स्थिति, बुद्धि स्तर, सम्प्राप्ति स्तर, राजनैतिक सम्बद्धता आदि कुछ भी हो सकता है। ऐसी स्थिति में प्रतिदर्श चयन के लिए स्तरीकृत रैन्डम प्रतिचयन अच्छा माना जाता है। स्तरीकृत रैन्डम प्रतिचयन विधि में समष्टि को कुछ स्तरों में बाँटकर प्रत्येक स्तर से इकाइयों का चयन रैन्डम ढंग से किया जाता है जो मिलकर प्रतिदर्श की रचना करती है। किस स्तर से कितनी इकाइयों का चयन करना चाहिए इसके निर्धारण के चार विकल्प– (*i*) समान आबँटन (Equal Allocation), (*ii*) अनुपातिक आबँटन (Proportionate Allocation), (*iii*) अ-अनुपातिक आबंटन (Disproportionate Allocation) तथा (*iv*) अनुकूलतम आबंटन (Optimum Allocation) हो सकते हैं।

सारणी 2

स्तरीकृत प्रतिचयन में चार प्रकार के आबंटन के उदाहरण

स्तर Stratum	समष्टि		प्रतिदर्श के आकार			
	आकार (Size)	प्रसरण (Variance)	समान आबंटन (Equal Allocation)	अनुपातिक आबंटन (Proportionate Allocation)	अ-अनुपातिक आबंटन (Disproporationate Allocation)	अनुकूलतम आबंटन (Optimum Allocation)
उच्च बुद्धि	200	25	33	20	30	28
औसत बुद्धि	500	20	34	50	40	55
निम्न बुद्धि	300	10	33	30	30	17
कुल	**1000**		**100**	**100**	**100**	**100**

नोट

जब समष्टि के सभी स्तरों से समान संख्या में इकाइयाँ चयनित की जाती हैं तब इसे समान आबंटन की स्थिति कहते हैं। जब समष्टि के विभिन्न स्तरों से उनके आकार के अनुपात में इकाइयाँ चयनित की जाती हैं तब इसे अनुपातिक आवंटन कहते हैं। जब समष्टि के विभिन्न स्तरों से चयनित की जाने वाली इकाइयों का निर्धारण अनुसंधानकर्ता अपनी स्वेच्छा से करता है तब इसे अ-अनुपातिक आबँटन कहते हैं। जब समष्टि के विभिन्न स्तरों से चयनित इकाइयों के निर्धारण में स्तरों के आकार व विचलनशीलता दोनों का ध्यान इस प्रकार से रखा जाता है कि बड़े आकार व अधिक विचलनशीलता वाले (विजातीय) स्तरों से अधिक इकाइयाँ छाँटी जाये जबकि छोटे आकार वक कम विचलनशीलता वाले (सजातीय) स्तरों से कम इकाइयाँ छाँटी जायें तो इसे अनुकूलतम आबँटन कहते हैं। अनुकूलतम आबंटन में समष्टि के स्तरों के आकार (N) तथा प्रसरण (σ^2) के गुणनफलों के अनुपात में प्रतिदर्श की इकाइयाँ निर्धारित की जाती है। स्पष्ट है कि अनुकूलतम आबँटन केवल तब ही प्रयोग में लाया जा सकता है जबकि समष्टि को स्तरों में विभाजित करने वाले चर को कम से कम अन्तरित पैमाने (Interval Scale) पर मापा गया हो। अनुसंधानकर्ता अपने अनुसंधान की परिस्थिति व आवश्यकता के अनुरूप इनमें से किसी एक आबँटन का प्रयोग कर सकता है। विभिन्न स्तरों से इकाइयों के चयन के लिए लाटरी अथवा रैन्डम अंक/संख्या सारणी का प्रयोग किया जा सकता है।

पुंज रैन्डम प्रतिचयन

(Cluster Random Sampling)

अभी तक वर्णित प्रतिचयन की तीनों विधियों अर्थात सरल रैन्डम विधि, क्रमबद्ध रैन्डम विधि तथा स्तरीकृत रैन्डम विधि में समष्टि की सभी इकाइयों को सूचीबद्ध करना होता है। परन्तु कभी-कभी समष्टि की सभी इकाइयों की सूची प्राप्त करना अत्यन्त कठिन या लगभग असम्भव होता है। जैसे उत्तर प्रदेश में कक्षा दस में पढ़ने वाले सभी छात्रों की सूची प्राप्त करना एक असम्भव सा कार्य है। अपने अधिकतम प्रयास से अनुसंधानकर्ता अधिक से अधिक सभी हाईस्कूलों की सूची प्राप्त कर सकता है। इसके अतिरिक्त उपरोक्त तीनों ही विधियों से प्राप्त प्रतिदर्श में चयनित इकाइयाँ सम्पूर्ण समष्टि में यत्र-तत्र सर्वत्र फैली हुई होती हैं जिसके कारण ऐसे प्रतिदर्श से समंक संकलित करना धन, समय व श्रम की दृष्टि से अत्यन्त दुरूह हो जाता है। ऐसी स्थिति में उपरोक्त वर्णित तीनों विधियाँ अव्यावहारिक हो जाती हैं। इस प्रकार की परिस्थितियों में पुंज रैन्डम प्रतिचयन विधि (Cluster Random Sampling Method) का प्रयोग किया जा सकता है। पुंज रैन्डम प्रतिचयन विधि में इकाइयों के स्थान पर इकाइयों के पुंजों अर्थात झुंडों (Clusters) को छांटा जाता है। जैसे छात्रों का चयन करने के लिए विभिन्न स्कूलों में पढ़ने वाले छात्रों

को पुंज या झुंड मानकर छांटा जा सकता है। अत: पुंज प्रतिचयन विधि से कक्षा दस के छात्रों का प्रतिदर्श छाँटने के लिए समस्त हाईस्कूलों में से कुछ हाईस्कूलों को छांटकर उनकी कक्षा 10 में पढ़ने वाले सभी छात्रों से प्रतिदर्श बनाया जा सकता है। इस विधि में पुंजों का चयन करने के लिए पीछे वर्णित किसी भी विधि का प्रयोग किया जा सकता है तथा तदनुसार प्रतिदर्श को पुंज सरल रैन्डम प्रतिदर्श अथवा पुंज क्रमबद्ध रैन्डम प्रतिदर्श अथवा पुंज स्तरीकृत रैन्डम प्रतिदर्श कहा जा सकता है।

बहु-सोपान प्रतिचयन

(Multi Stage Sampling)

कभी-कभी अनुसंधानकर्ता को प्रतिदर्श का चयन करने के लिए एक से अधिक बार प्रतिचयन करना पड़ता है। जैसे उ० प्र० कक्षा दस में अध्ययनरत छात्रों के किसी प्रतिदर्श का चयन करने के लिए पहले कुछ जनपद, फिर छांटे गये जनपदों में से प्रत्येक जनपद से कुछ हाईस्कूल तथा फिर छांटे गये हाईस्कूलों में से प्रत्येक हाईस्कूल से कुछ छात्र छाँटे जा सकते हैं। स्पष्ट है कि अन्तिम रूप में छाँटे गये छात्र ही वास्तविक प्रतिदर्श का निर्माण करेंगे। इस विधि में अनुसंधानकर्ता को कुल तीन बार प्रतिचयन करना होगा। एक से अधिक बार प्रतिचयन करने के कारण प्रतिदर्श छाँटनें की इस विधि को बहु-सोपान रैन्डम प्रतिचयन (Multi Stage Random Sampling) तथा इस विधि से छांटे गए प्रतिदर्श को बहु-सोपान रैन्डम प्रतिदर्श कहते हैं। इस विधि के विभिन्न सोपानों पर उपरोक्त वर्णित चारों विधियों का प्रयोग किया जा सकता है।

उपरोक्त के अतिरिक्त सोद्देश्य प्रतिदर्श (Purposive Sample), निर्णीत प्रतिदर्श (Judgement Sample), आकस्मिक प्रतिदर्श (Incidental Sample), उपलब्ध प्रतिदर्श (Available Sample) तथा नियतांश प्रतिदर्श (Quota Sample) की चर्चा भी की जाती है। इन सभी प्रकार के प्रतिदर्शों का तात्पर्य इनके नाम से स्व-स्पष्ट है। जब अनुसंधानकर्ता अपने अनुसंधान उद्देश्य को ध्यान में रखकर स्वविवेक से प्रतिचयन का निर्धारण करता है तब इसे सोद्देश्य प्रतिदर्श अथवा निर्णीत प्रतिदर्श कहते हैं। इस प्रकार के प्रतिदर्श प्राय: प्रारम्भिक अध्ययनों (Pilot Studies) के लिए छांटे जाते हैं। जब अनुसंधानकर्ता अपने सम्मुख अचानक पड़ जाने वाली इकाइयों से प्रतिदर्श की रचना कर लेता है, तब ऐसे प्रतिदर्श को आकस्मिक प्रतिदर्श कहते हैं। जब सुगमता से उपलब्ध इकाइयों को प्रतिदर्श के रूप में स्वीकार कर लिया जाता है तो ऐसे प्रतिदर्श को उपलब्ध प्रतिदर्श कहते हैं। जब समष्टि के विभिन्न उपसमूहों से सम्मिलित की जाने वाली इकाइयों की संख्या का निर्धारण करके उन्हें किसी भी ढंग से छाँट लेते हैं तब ऐसे प्रतिदर्श को नियतांश प्रतिदर्श कहते हैं। परन्तु इन पाँचों प्रकार के प्रतिदर्शों से प्राप्त सूचनाओं के आधार पर समष्टि के सम्बन्ध में विश्वसनीय व वैध अनुमान प्राप्त करने प्राय: कठिन होते हैं। यही कारण है कि अनुमानात्मक सांख्यिकी में इन सभी प्रतिदर्शों की उपादेयता लगभग नहीं के बराबर है।

छात्र क्रियाकलाप (Student Activity)

1. सांख्यिकीय में प्रतिदर्शज औ प्राचल से क्या तात्पर्य है?

2. सीमित समष्टि और असीमित समष्टि का अर्थ स्पष्ट कीजिए।

3. सरल रैन्डम प्रतिचयन से आप क्या समझते हैं?

1.5 सारांश (Summary)

1. किसी भी अनुसंधान का मुख्य उद्देश्य कुछ ऐसे तथ्यों, सिद्धान्तों अथवा उपयोगों का पता लगाना होता है जो कुछ नवीन ज्ञान प्रदान करके मानव जाति की समस्याओं का समाधान करने में सहायता प्रदान कर सकें।
2. प्रतिदर्श के लिए ज्ञात की जाने वाली वर्णनात्मक मापों को **प्रतिदर्शज** (Statistics एक वचन में Statistic) कहते हैं जबकि समष्टि के लिए विद्यमान वर्णनात्मक मापों को प्राचल (Parameters) कहते हैं।
3. अनुसंधान में समष्टि अथवा जनसंख्या (Population) से तात्पर्य उन सभी इकाइयों के समुच्चय (Set of Units) से होता है जिनमें कुछ सामान्य विशेषताएं होती हैं तथा जिनके सम्बन्ध में अनुसंधानकर्ता कुछ निष्कर्ष ज्ञात करना चाहता है।
4. रैन्डम प्रतिदर्श (Simple Random Sample) में वास्तव में समष्टि की विभिन्न इकाइयों के चयन की प्रायिकता स्थिर (Constant) रहती है अर्थात विभिन्न इकाइयों का चयन परस्पर स्वतन्त्र होता है। तथा कुछ इकाइयों के चयन से शेष इकाइयों के चयन की प्रायिकता प्रभावित नहीं होती है।

अभ्यास-प्रश्न (Exercise Questions)

1. प्रतिचयन की विभिन्न विधियों का उल्लेख करते हुए क्रमबद्ध प्रतिचयन का वर्णन कीजिए।
2. बहु-सोपान (Multi-stage) रैन्डम प्रतिचयन विधि की विवेचना करो।
3. प्रतिस्थापन रहित चयन और प्रतिस्थापन सहित चयन से क्या तात्पर्य है?

सन्दर्भ ग्रन्थ (Reference Books)

इकाई–2

प्रायिकता
(Probability)

संरचना (Structure)

2.1 उद्देश्य (Objectives)

इस इकाई के अध्ययन के पश्चात् विद्यार्थी योग्य होंगे–

- प्रायिकता का अर्थ तथा परिभाषा जानने में।
- प्रायिकता का माप जानने में।
- प्रायिकता प्रमेय तथा बेस प्रमेय से अवगत होंगे।

2.2 प्रस्तावना (Introduction)

वास्तव में प्रायिकता सिद्धान्त (Theory of Probability) के अध्ययन की प्रेरणा सर्व-प्रथम जुआरियों से मिली थी। भाग्य देवी (Goddess Fortune) अथवा अन्ध विश्वास (super-stitions) पर आस्था रखने वाले सत्रहवीं शताब्दी के कुछ जुआरी जुए में जीतने के अवसरों से सम्बन्धित समस्याओं का समाधान कराने के लिए तत्कालीन प्रसिद्ध गणितज्ञों के पास जाते थे। उस शताब्दी में इस सिद्धान्त का जन्म हुआ जबकि फ्रांसिसी रईस एन्टोयने गोमबाल्ड (Antoine Gombauld) जिसे चेवलेयर डी मेरे (Chevalier de Mere) के नाम से जाना

जाता है, ने अवसर के खेल के बारे में कुछ प्रश्न उठाये। विशेषकर एक जोड़ी पासे को 24 बार फेंकने में दो-छः आने की सम्भावना उसके लिए एक पहेली बनी हुई थी। डी मेरे ने यह प्रश्न एक फ्रान्सिसी युवा गणितज्ञ ब्लेसे पास्कल (Blaise Pascal) के सामने उठाया, जिसने उस पहेली का हल निकाला। बाद में पास्कल ने इसे तथा अन्य पहेलियों को, जिसे डी मेरे उठाता था, प्रसिद्ध फ्रान्सिसी गणितज्ञ पैरे डी फरमेट (Pierre de Fermat) के साथ विवेचन किया। इस प्रकार प्रायिकता सिद्धान्त की नींव पड़ी। उन्नीसवीं शताब्दी के प्रारम्भ में फ्रांस के प्रसिद्ध गणितज्ञ लॉप्लेस (Laplace) तथा जर्मन गणितज्ञ गॉस (Gauss) ने इस सिद्धान्त के ज्ञान में महत्त्वपूर्ण वृद्धि की। राष्ट्रीय अर्थ-व्यवस्था के विस्तार के साथ-साथ अनेक महान गणितज्ञों, जैसे डी मॉवेर (De Moivre-1718), जैम्स बर्नोली (Games Bernoulli-1713), बेस (Bayes-1768) आदि भी प्रायिकता सिद्धान्त के विकास तथा अनेक क्षेत्रों में उसके प्रयोग के लिए प्रेरित हुए। बाद के वर्षों में फिशर (R. A. Fisher), पीयर्सन (Pearson), जे. नेमेन (J. Neyman) आदि ने प्रायिकता सिद्धान्त पर आधारित प्रतिदर्श-सिद्धान्त (Sampling Theory) का विकास किया।

2.3 अर्थ तथा परिभाषा (Meaning and Definition)

प्रायिकता सिद्धान्त के विकास का एक प्रमुख कारण जीवन के लगभग प्रत्येक पहलू में दैव-तत्व (Random phenomenon) की उपस्थिति रहना है। एक स्थिति दैव होती है यदि उसके परिणाम अवसर पर निर्भर करते हैं। सभी सम्बन्ध परिणाम पूर्व ही ज्ञात हो सकते हैं, परन्तु किसी प्रयोग में एक घटना का क्या परिणाम होगा, पूर्व-निर्धारित न किया जा सकता हो। फिर भी प्रक्रिया में कुछ नियमितता पायी जाती हैं, जिससे प्रत्येक सम्भव परिणाम को प्रायिकता अंश दिया जा सकता है। दैव तत्व की उपस्थिति का सरलतम उदाहरण एक सिक्के के उछालने से प्राप्त परिणामों को दिया जा सकता है। यद्यपि सभी सम्भव परिणाम ज्ञात होते हैं कि एक सिक्के के उछालने पर वह या तो चित्त गिरेगा या पट्ट गिरेगा परन्तु एक उछाल में वह चित्त गिरेगा या पट्ट गिरेगा यह केवल अवसर-कारक ही निर्धारित करते हैं। यहाँ पर कोई भी निर्धारणात्मक नियमितता नहीं है। इसी प्रकार एक पासे को फेंकने पर हम यह पूर्व-निर्धारण नहीं कर सकते कि पासे का कौन-सा ओर ऊपर आयेगा, यद्यपि सभी सम्भावित परिणाम पूर्व ज्ञात होते हैं। दैव-तत्व इतने व्यापक एवं विस्तृत क्षेत्रों में विद्यमान होता है कि सामाजिक एवं प्राकृतिक विज्ञानों के विद्यार्थियों को प्रायिकता सिद्धान्त का ज्ञान आवश्यक हो जाता है।

सांख्यिकी में प्रायिकता का विशेष महत्व है क्योंकि अनेक सिद्धान्त एवं रीतियाँ इसी विचार पर आधारित हैं। वास्तव में प्रायिकता हमारे जीवन में विशेष महत्त्व रखती है तथा इसका प्रयोग दिन-प्रतिदिन की बोलचाल में होता है। हम प्रायः इस प्रकार के कथन सुनते रहते हैं : ''अच्छा हो कि तुम छाता ले लो, क्योंकि पानी बरसने की सम्भावना है'', ''अमुक व्यक्ति के जीतने की सम्भावना काफी कम है'', ''यह सम्भावना है कि आज शाम तक पानी बरसे'', ''सम्भवतः तुम्हारी बात सही हो'' उस व्यक्ति के परीक्षा में पास होने की सम्भावना पचास-पचास है'', आदि। इन समस्त कथनों में अनिश्चितता की भावना को व्यक्त किया गया है। गोथे (Goethe) ने कहा है कि, ''अज्ञानता में कार्य करने से अधिक डर की बात और कोई नहीं है। सम्भावनाओं के रूप में तर्क करने से अनिश्चितता अथवा अज्ञान को कम करने का प्रयत्न किया जाता है। सांख्यिकी विज्ञान में 'सम्भावना' अथवा प्रायिकता शब्द का अर्थ अधिक सूक्ष्म है। सांख्यिकी में प्रायिकता किसी एक घटना के घटने की निश्चितता नापने का एक संख्यात्मक माप है

2.4 प्रायिकता की माप (Measurement of Probability)

प्रायिकता की माप की मुख्यतः तीन विधियाँ हैं। ये विभिन्न वैचारिक दृष्टिकोण का प्रतिनिधित्व करती हैं। वे हैं–

तर्कपूर्ण या स्वयंसिद्ध प्रायिकता (Priori Probability)–प्रायिकता सिद्धान्त का विकास जुए के खेलों के कारण हुआ था, अत: प्रायिकता के मापने की विधि जुए की दशाओं के अनुरूप ही विकसित हुई है। प्रायिकता मापने की इस विधि को प्राचीन (Classical) अथवा स्वयंसिद्ध (Priori) प्रायिकता विचार कहते हैं। ऐसी दशा में, किसी एक घटना के घटने की प्रायिकता उस घटना के अनुकूल परिणामों की संख्या का सभी सम्भव परिणामों की संख्या से अनुपात होता है, जबकि उनमें से प्रत्येक परिणाम घट सकता है (The probability of an event is simply the ratio of number of favourable observes of an event to the number of possible observes, where each outcome is equally likely to occur)। अन्य शब्दों में, यदि एक घटना अनेक प्रकार से घट सकती हो तो उनमें से किसी एक घटना घटने की प्रायिकता उस घटना के अनुकूल घट जाने की संख्या का समस्त सम्भावित घटनाओं की संख्या के अनुपात के रूप में होगी। यदि एक घटना 'A' के अनुकूल घट जाने के 'a' तथा अनुकूल न घट जाने के 'b' सम्भव परिणाम हैं तथा ये सभी सम्भव परिणाम समान रूप से घट सकते हैं तथा वे अपवर्जी हैं तो A घटना के घटने की प्रायिकता $\{P(A)\}$ होगी–

$$\text{P(A)} = \frac{a}{a+b} = \frac{\text{Number of outcomes favourable to occurence of A}}{\text{Total number of possible outcomes}}$$

तथा घटना A न घटने की प्रायिकता $\{q(A)\}$ होगी–

$$q(\text{A}) = \frac{b}{a+b} = \frac{\text{Number of outcomes not favourable to occurence of A}}{\text{Total number of possible outcomes}}$$

उदाहरणस्वरूप, यदि एक सिक्के को उछाला (toss) जाय, तो चित्त (Head) गिरने की प्रायिकता $P = \frac{\text{H}}{\text{H + T}}$ या $\frac{1}{2}$, क्योंकि प्रतिकूल घटना की संख्या 1 है तथा सभी सम्भव परिणामों की संख्या 2 है। सिक्के के पट्ट (Tail) गिरने की प्रायिकता $q = \frac{\text{T}}{\text{H + T}} = \frac{1}{2}$ होगी।

किसी घटना के घटने (P) और न घटने (q) का योग 1 के बराबर होता है।

आधारभूत धारणाएँ

(i) **दैव प्रयोग** (Random Experiment)–प्रायिकता घटनाओं के परिणामों जिन्हें 'दैव प्रयोग' कहते हैं, के लिए ज्ञात की जाती है। 'प्रयोग' शब्द का अर्थ रसायनशास्त्र अथवा भौतिक शास्त्र की अपेक्षा सांख्यिकी में अधिक विस्तृत है। उदाहरणस्वरूप, सिक्का उछालना एक सांख्यिकी प्रयोग है। प्रयोगों के दो गुण होते हैं–(अ) प्रत्येक प्रयोग के कई सम्भव परिणाम होते हैं, जिन्हें पहले से ही स्पष्ट किया जा सकता है, (ब) प्रत्येक प्रयोग के परिणाम के बारे में हम अनिश्चित होते हैं। सिक्का उछालने में, यह स्पष्ट किया जा सकता है कि सिक्का चित्त या पट्ट गिरेगा, परन्तु हम यह निश्चित रूप से नहीं कह सकते कि सिक्के की अमुक उछाल में यह चित्त गिरेगा या पट्ट। हम प्रयोग शब्द इसलिए उपयोग में लाते हैं क्योंकि परिणाम तो अभी निश्चित किया जाना है तथा 'दैव' यह स्पष्ट करता है कि कोई विशेष परिणाम अनिश्चित होता है।

(ii) **समान रूप से घटने वाली घटनाएँ** (Equally Likely Events)–प्रायिकता सिद्धांत का विवरण देने में 'समान रूप से घटने वाली घटनाएँ' वाक्यांश का प्रयोग किया जाता है। ये वे घटनाएँ हैं जिनमें से कोई एक घटना घट सकती है। एक सिक्का उछालने में या तो चित्त आयेगा या पट्ट, अत: समान रूप से घटने वाली घटनाएँ दो हैं।

(iii) **प्रतिदर्श-समूह** (Sample Space)–किसी प्रयोग का प्रतिदर्श-समूह सभी सम्भव परिणामों का समूह (set or collection) होता है। एक सिक्का उछालने में केवल दो ही परिणाम हो सकते हैं–चित्त या पट्ट, अत: प्रतिदर्श समूह = (Head, Tail), एक पासा फेंकने का प्रतिदर्श समूह होगा (1, 2, 3, 4, 5, 6)। प्रतिदर्श समूह में दिया गया प्रत्येक परिणाम अवयव या प्रतिदर्श बिन्दु (Element of Sample Point) कहलाता है।

यदि दो सिक्कों को एक साथ एक बार उछाया जाय तो प्रयोग के सम्भव परिणाम इस प्रकार होंगे–

Coin 1 \ Coin 2	H	T
H	HH	HT
T	TH	TT

इस प्रयोग का प्रतिदर्श समूह = (HH, HT, TH, TT).

यदि दो पाँसों को एक साथ फेंका जाय तो सम्भव परिणाम 36 होंगे–

Outcome of First die	*Outcome of Second die*						
	1	2	3	4	5	6	
1	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)	The
2	(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)	sample
3	(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)	space
4	(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)	of the
5	(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)	experi-
6	(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)	ment

(*iv*) **घटना** (Event)–एक घटना प्रतिदर्श-समूह (sample space) का एक अवयव (element) होता है। एक घटना सरल घटना (simple event) कहलाती है यदि उसमें केवल एक ही अवयव होता है। यदि उसमें एक से अधिक अवयव होता है तो उसे संयुक्त या मिश्रित घटना (compound or complex event) कहते हैं। अन्य शब्दों में, संयुक्त घटना में दो या अधिक सरल घटनाओं का समावेश होता है।

(*v*) **पारस्परिक अपवर्जी घटनाएँ** (Mutually Exclusive Events)–पारस्परिक अपवर्जी घटनाएँ वे घटनाएँ होती हैं जिनके घटने पर अन्य घटनाओं के घटने की सम्भावना समाप्त हो जाती है तथा इसके विपरीत अन्य घटना के घट जाने पर पूर्व घटना के घटने की सम्भावना समाप्त हो जाती है। अन्य शब्दों में, पारस्परिक अपवर्जी घटनाएँ एक साथ नहीं घट सकतीं। उदाहरणार्थ, यदि एक सिक्के को उछाला जाय तो चित्त गिर जाने पर पट्ट गिरने अथवा पट्ट गिर जाने पर चित्त गिर जाने की सम्भावना समाप्त हो जाती है। इसी प्रकार यदि पाँसे को फेंका जाय तो उसकी कोई भी साइड ऊपर आ जाने पर अन्य साइडों के ऊपर आने की सम्भावना समाप्त हो जाती है। पाँसे की छः साइड होती हैं परन्तु एक बार में एक ही साइड ऊपर रह सकती है अतः यह पारस्परिक अपवर्जी घटनाएँ हैं।

(*vi*) **स्वतन्त्र घटनाएँ** (Independent Events)–यदि किसी घटना के घट जाने से आगे किसी घटना के घटने पर प्रभाव नहीं पड़ता तो ऐसी घटनाएँ स्वतंत्र होती हैं। जब दो घटनाओं का प्रभाव एक दूसरे पर नहीं पड़ता, तो वे घटनाएँ स्वतन्त्र होती हैं। उदाहरणार्थ, एक बार सिक्का उछालने के परिणाम का प्रभाव दुबारा सिक्का उछालने के परिणाम पर नहीं पड़ता।

(*vii*) **आश्रित घटनाएँ** (Dependent Events)–यदि एक घटना के घट जाने का प्रभाव दूसरी घटना के घटने पर पड़ता है तो उन घटनाओं को आश्रित घटनाएँ (dependent events) कहा जाता है। उदाहरणार्थ ताश की एक गड्डी में से बादशाह निकालने की सम्भावना $\frac{4}{52}$ या $\frac{1}{13}$ है, यदि पहली बार ताश निकालने में बादशाह निकल आता है जो दूसरी बार बादशाह निकलने की सम्भावना $\frac{3}{51}$ होगी। अतः पहले वाली घटना का प्रभाव दूसरी घटना पर पड़ा, ये आश्रित घटनाएँ हैं।

प्रायिकता की मूल परिभाषा से यह स्पष्ट होता है कि किसी घटना के घटने की प्रायिकता सदैव '0' से लेकर '1' के बीच में होगी। यह इसलिए कि प्रायिकता भिन्न में अंश (numerator) न तो कभी ऋणात्मक हो

सकता है और न हर (denominator) से अधिक। इससे दो बातें स्पष्ट होती हैं–प्रथम, एक घटना जो निश्चित रूप से घटित होगी उसके अंश हर समान होंगे। ऐसी दशा में प्रायिकता '1' होगी। दूसरे, किसी असम्भव घटना के घटने की प्रायिकता '0' होगी क्योंकि प्रायिकता अंश में अंश (numerator) शून्य होगा।

अत: P (निश्चित घटना) = 1, P (असम्भव घटना) = 0

प्रायिकता को व्यक्त करने के ढंग–किसी घटना के घटने की प्रायिकता को अनुपात (ratio), अंश (fraction) या प्रतिशत (percentage) के रूप में व्यक्त किया जा सकता है। उदाहरणार्थ, एक सिक्के की उछाल में उसके चित्त गिरने की सम्भावना को $\frac{1}{2}$, 0.5 या 50% के रूप में व्यक्त किया जा सकता है।

उदाहरण (Illustration)

There are 50 balls, each ball having two colours, one black or white and the other red, orange or green as shown in the following table :

	Red	*Orange*	*Green*	*Total*
Black	3	12	15	30
White	7	3	10	20
Total	10	15	25	50

(It means there are 3 balls which are black and red, 12 balls are black and orange and so on.)

If out of these balls, one ball is selected at random, find the probability of each type of ball being drawn up.

समाधान (Solution)

प्रायिकता-तालिका (*Probability Table*)

	Red	*Orange*	*Green*	*Total*
Black	0.06	.24	.30	0.60
White	0.14	.06	.20	0.40
Total	0.20	.30	.50	1.00

2.5 प्रायिकता-प्रमेय (Probability Theorems)

प्रायिकता का प्रयोग करने वाली अनेक समस्याओं के हल के लिए कुछ आधारभूत नियमों, जिनसे प्रायिकता-व्यवसाय प्रभावित होता है, को समझना आवश्यक होता है। इन्हें सामान्यत: 'प्रायिकता-प्रमेय' की संज्ञा दी जाती है। इनका यहाँ विवेचन किया जा रहा है।

(1) **योग-प्रमेय** (Addition Theorem)–इस प्रमेय के अनुसार, "यदि दो घटनाएँ पारस्परिक अपवर्जी (mutually exclusive) हों तथा एक घटना के घटने की प्रायिकता p_1 तथा दूसरी घटना के घटने की प्रायिकता p_2 हो तो दोनों में से किसी भी घटना के घट जाने की प्रायिकता $p_1 + p_2$ होगी।" यह प्रमेय सरल व स्वयंसिद्ध है।

उदाहरणार्थ, एक पाँसा फेंकने में 1 नम्बर ऊपर आने की प्रायिकता $\frac{1}{6}$ होती है, 3 नम्बर ऊपर आने की प्रायिकता भी $\frac{1}{6}$ होती है तथा 5 नम्बर ऊपर आने की प्रायिकता भी $\frac{1}{6}$ होती है। एक बार पासा फेंकने में किसी विषम (odd) संख्या (1, 3 तथा 5) आने की प्रायिकता ज्ञात करने के लिए तीनों नम्बरों की अलग-अलग प्रायिकताओं को जोड़ना चाहिए। अर्थात् एक बार पाँसा फेंकने में विषम संख्या आने की प्रायिकता $\frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6}$ or $\frac{1}{2}$ होगी।

योग-प्रमेय उसी दशा में सत्य होगा जबकि–

(अ) घटनाएँ पारस्परिक अपवर्जी (mutually exclusive) हों,

(ब) पारस्परिक अपवर्जी घटनाएँ एक ही प्रकार की हों।

उदाहरण (Illustration)

What is the probability of drawing a black card or a king from a pack of ordinary playing cards?

समाधान (Solution)

Number of black cards		number of kings		number of black kings
26	+	4	–	2

$$\text{Hence } (p) = \frac{26}{52} + \frac{4}{52} - \frac{2}{52} = \frac{28}{52}$$

घटनाओं का पारस्परिक अपवर्जी (mutually exclusive) होने के साथ-साथ उनका एक ही प्रकार (belong to same set) का होना आवश्यक है। वॉन मिसेस (Von Mises) के निम्न उदाहरण से इसकी आवश्यकता स्पष्ट हो जाती है–

"यदि यह कल्पना की जाय कि एक मनुष्य के 40-41 वें वर्ष में मरने की प्रायिकता 0.011 तथा 41-42वें वर्ष में विवाह करने की प्रायिकता 0.009 है। ये दोनों घटनाएँ पारस्परिक अपवर्जी हैं, परन्तु यह नहीं कहा जा सकता कि मनुष्य के 40-41 वें वर्ष में मर जाने तथा 41-42वें वर्ष में विवाह करने की प्रायिकता .011 + .009 = .02 है, क्योंकि ये दोनों घटनाएँ एक ही प्रकार की नहीं हैं।"

(2) **गुणन-प्रमेय** (Multiplication Theorem)–"यदि दो घटनाएँ पारस्परिक रूप से स्वतन्त्र (mutually independent) हों तथा एक घटना के घटने की प्रायिकता p_1 तथा दूसरी घटना के घटने की प्रायिकता p_2 हो तो दोनों घटनाओं के एक साथ घटने की प्रायिकता $p_1 \times p_2$ होगी।" उदाहरणार्थ, एक सिक्का उछालने में उसके चित्त गिरने की प्रायिकता $\frac{1}{2}$ है, एक पाँसा फेंकने में 4 संख्या आने की प्रायिकता $\frac{1}{6}$ है। यदि सिक्के तथा पाँसे दोनों को एक साथ उछाला तथा फेंका जाय तो सिक्के के चित्त गिरने तथा पाँसे में 4 संख्या आने की प्रायिकता $\frac{1}{2} \times \frac{1}{6} = \frac{1}{12}$ होगी।

उदाहरण (Illustration)

An aircraft is equipped with three engines that operate independently. The probability of an engine failure is .01. What is the probability of a successful flight if only one engine is needed for the successful operation of the aircraft ?

समाधान (Solution)

Since the flight is unsuccessful only when all the three engines fail, then the probability of unsuccessful flight is :

$$.01 \times .01 \times .01 = .000001.$$

The probability of successful flight = 1 – .000001 = .999999.

प्रायिकता के गुणन प्रमेय के प्रयोग में भी यह आवश्यक है कि सभी घटनाएँ एक ही प्रकार (same set) की होनी चाहिए। इस आवश्यकता का महत्त्व स्पष्ट करने के लिए मोरोने (Moroney) ने अपनी पुस्तक '*Facts from Figures*' में एक बड़ा ही रोचक उदाहरण दिया है जो इस प्रकार है–

"उस व्यक्ति के बारे में सोचिए जो अपनी भावी पत्नी में अनेक असम्बन्धित प्रकृति के गुणों की माँग करता है। कल्पना कीजिए वह चाहता है कि उसकी भावी पत्नी की ग्रीकवासियों जैसी नाक हो, सुनहले बाल हों, आँखों के रंग अलग-अलग हों–एक आँख नीली तथा एक आँख बादामी हो तथा उसे सांख्यिकी विज्ञान का उच्चस्तीय ज्ञान हो। ऐसी स्त्री मिलने की क्या प्रायिकता होगी?"

इस उदाहरण में प्रायिकता का गुणन प्रमेय प्रयोग नहीं किया जा सकता क्योंकि सभी घटनाएँ अलग-अलग प्रकार की हैं।

(3) **शर्तपूर्ण प्रायिकताओं का प्रमेय** (Theorem of Conditional Probabilities)–यदि उप-घटनाएँ (sub-events) स्वतन्त्र न हों और आश्रितता (dependence) की प्रकृति ज्ञात हो तो शर्तपूर्ण प्रायिकताओं के प्रमेय का प्रयोग होता है। यह प्रमेय गुणक प्रमेय का ही उपप्रमेय (corollary) है। इस प्रमेय के अनुसार "दो आश्रित घटनाओं के घटने की प्रायिकता, प्रथम घटना के घटने की प्रायिकता तथा प्रथम घटना के घट जाने पर दूसरी घटना के घटने की प्रायिकता के गुणनफल के बराबर होगी।" संकेतानुसार, $p(\text{A and B}) = p(\text{A}) \times p\ (\text{B/A})$, यहाँ $p(\text{B/A})$ 'B' घटना की आश्रित प्रायिकता है, अर्थात् 'B' के घटने की उस दशा की प्रायिकता है जबकि 'A' पहले ही घटित हो चुकी है।

उदाहरणार्थ, ताश की एक गड्डी में से एक ताश निकाले जाने पर उसके बादशाह होने की प्रायिकता $\frac{4}{52}$ है, यदि पहला ताश निकाले जाने पर बादशाह निकलता है तो पुन: एक ताश निकाले जाने पर बादशाह निकलने की प्रायिकता $\frac{3}{51}$ होगी। अत: एक साथ दो बार बादशाह निकलने की प्रायिकता $\frac{4}{52} \times \frac{3}{51} = \frac{1}{26} \times \frac{2}{52}$ होगी। शर्तपूर्ण प्रायिकताओं को एक प्रतिदर्श के आंशिक खाली होने की प्रायिकता (Probability due to partial exhaustion of a sample) भी कहते हैं।

2.6 बेस का प्रमेय (Bayes's Theorem)

दैव प्रयोग (random experiment) से सम्बन्धित नयी जानकारी मिलने पर प्रायिकताओं में संशोधन किया जा सकता है। प्रायिकताओं में संशोधन करने का विचार हम सभी के लिए जाना-माना है, उन व्यक्तियों के लिए भी जिन्हें प्रायिकता ज्ञात करने का पूर्व अनुभव नहीं है–जो ऐसे वातावरण में रहते हैं जहां अवसर या मनमौजी होने का बोलबाला रहता है और जो अनजाने में प्रायिकता-निर्णय करते हैं। कुछ तथ्यों का अवलोकन करके, हम आत्म-बोध से प्रायिकताओं में संशोधन करते हैं और उसके अनुरूप व्यवहार करते हैं। प्रायिकताओं में संशोधन करने का उद्देश्य प्रयोगात्मक जानकारी का श्रेष्ठतर उपयोग करना है। इसको आदरणीय थॉमस बेस (Reverend Thomes Bayes) के नाम से बेस प्रमेय (Bayes Theorem) कहा जाता है क्योंकि उन्होंने 18वीं शताब्दी में प्रयोगात्मक निष्कर्षों के आधार पर प्रायिकताओं में संशोधन करने का प्रस्ताव किया था।

प्राय: व्यापारी किसी विशेष घटना या समस्या पर अतिरिक्त जानकारी रखता है। वह ज्ञान या तो व्यक्तिगत विश्वास के आधार पर हो सकता है अथवा घटना के पूर्व-व्यवहार के आधार पर। प्रयोग के परिणाम देखे बिना व्यक्तिगत अनुभव के आधार पर प्रायिकता निश्चित करने को तर्क-पूर्ण या स्वयंसिद्ध प्रायिकता (Prior Probability) कहते हैं। उदाहरणस्वरूप, भूतकालीन बिक्री की प्रायिकता निर्धारण करना या मशीन द्वारा निर्मित दोषपूर्ण उत्पादों की प्रायिकता निर्धारण करना, स्वयंसिद्ध प्रायिकता के उदाहरण हैं। बेस के नियम के आधार पर जब प्रायिकताएँ संशोधित की जाती हैं तो उन्हें अनुभवसिद्ध प्रायिकताएँ (Posterior Probabilities) कहते हैं। अतिरिक्त जानकारी के आधार पर व्यवहारिक व्यापारिक समस्याओं के समाधान में बेस प्रमेय अत्यन्त ही उपयोगी रहती है।

उदाहरणार्थ, एक दैव-प्रयोग (random experiment) के $E_1, E_2 \cdots\cdots$आदि अनेक पारस्परिक अपवर्जी घटनाएँ हैं जिनमें से प्रत्येक घटना की प्रायिकता $P(E_1), P(E_2) \cdots\cdots$ज्ञात की गयी हैं। इन प्रायिकताओं को स्वयंसिद्ध (Prior) प्रायिकताएँ कहा जाता है, क्योंकि ये वास्तविक अनुसन्धान के परिणामों को जानने से पहले ही घटनाओं के घटने के अवसरों का प्रतिनिधित्व करती हैं। अनुसन्धान के अनेक सम्भव परिणाम हो सकते हैं, जो संख्यात्मक रूप से E_s पर निर्भर करते हैं। प्रत्येक परिणाम के लिए (जिसे R से संकेताक्षर दिया गया है) सशर्त प्रायिकताएँ $P(R/E_1), P(R/E_2) \cdots\cdots$उपलब्ध होंगी। परिणामों से घटनाओं की प्रायिकताओं को घटा-बढ़ा कर संशोधित किया जा सकता है। संशोधित प्रायिकताएँ अनुभवसिद्ध प्रायिकताएँ (Posterior Probabilities) कहलाती हैं, क्योंकि वे परिणामों की जानकारी हो जाने के पश्चात् लागू होती हैं। अनुभवसिद्ध प्रायिकताएँ (posterior probabilities) मान वास्तव में $P(R/E_1), P(R/E_2)$ के स्वरूप की सशर्त प्रायिकताएँ (Conditional

Probabilities) जिन्हें बेस प्रमेय के अनुसार ज्ञात किया गया है। किसी घटना (E_j) की अनुभवसिद्ध प्रायिकता (posterior probability), प्रयोगात्मक अनुसन्धान के परिणाम (R) के लिए इस प्रकार ज्ञात की जायेगी–

$$P(E_j/R) = \frac{P(E_j)\,P(R/E_j)}{P(E_1)\,P(R/E_1) + P(E_2)\,P(R/E_2)\ldots\ldots\ldots}.$$

Illustration. *A* can hit a target 3 times in 5 shots, *B* 2 times in 5 shots, *C* 3 times in 4 shots. They fire a volley. What is the probability that 2 shots hit ?

समाधान (Solution). 'Fire a volly' means that *A*, *B*, and *C* all try to hit the target simultaneously. Two shots hit the target in one of the following ways :

(*a*) *A* and *B* hit and *C* fails to hit.

(*b*) *A* and *C* hit and *B* fails to hit.

(*c*) *B* and *C* hit and *A* fails to hit.

The chance of hitting by A = $\frac{3}{5}$ and of not hitting by him = $1 - \frac{3}{5} = \frac{2}{5}$

The chance of hitting by B = $\frac{2}{5}$ and of not hitting by him = $1 - \frac{2}{5} = \frac{3}{5}$

The chance of hitting by C = $\frac{3}{4}$ and of not hitting by him = $1 - \frac{3}{4} = \frac{1}{4}$

The Probability of (*a*) = $\frac{3}{5} \times \frac{2}{5} \times \frac{1}{4} = \frac{6}{100}$

The Probability of (*b*) = $\frac{3}{5} \times \frac{3}{4} \times \frac{2}{5} = \frac{27}{100}$

The Probability of (*c*) = $\frac{2}{5} \times \frac{3}{4} \times \frac{2}{5} = \frac{12}{100}$

Since (*a*), (*b*) and (*c*) are mutually exclusive events, the probability that two shots hit

$$\frac{6}{100} + \frac{27}{100} + \frac{12}{100} = \frac{45}{100} = \frac{9}{20} \text{ or } \frac{9}{20} \times 100 = 45\%$$

तर्कपूर्ण या स्वयंसिद्ध (priori) प्रायिकता माप में दो दिलचस्प विशेषताएँ होती हैं। प्रथम, जिस घटना की प्रायिकता ज्ञात की जाती है उसका गैर-विभ्रमीय (unbiased) होना आवश्यक है। अन्य शब्दों में, सिक्का, पाँसा या ताशों की गड्डी सच्ची (fair) होने चाहिए। दूसरे, इस प्रकार की घटनाओं की प्रायिकताएँ ज्ञात करने के लिए प्रयोग करने की आवश्यकता नहीं है। अर्थात् सिक्का उछाल कर, पाँसा फेंक कर या तासों की गड्डी में से ताश खींचकर प्रायिकताएँ जानना जरूरी नहीं है। प्रायिकताएँ ज्ञात करने के लिए किसी प्रयोगात्मक-समंकों का संकलन करने की आवश्यकता नहीं होती है, प्रायिकताओं की गणना पूर्णत: तर्क पर आधारित रहती है।

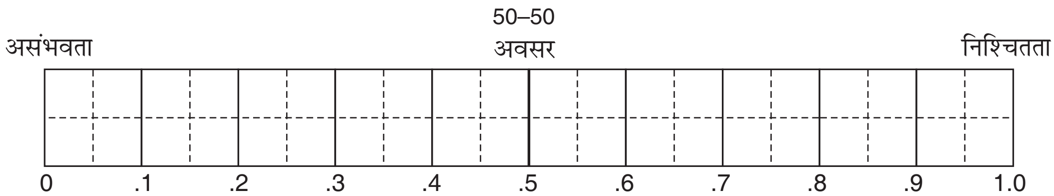
यह सम्भव हो सकता है कि अल्प-संख्या में प्रयोग करने से परिणाम स्वयंसिद्ध प्रायिकता से पर्याप्त भिन्न हों, परन्तु प्रयोगों की संख्या जितनी बढ़ती जायेगी, परिणाम स्वयंसिद्ध प्रायिकता के उतने ही निकट आते जायेंगे। सम्भव हो सकता है कि 10 बार सिक्का उछालने में 8 बार चित्त गिरे तथा 2 बार पट्ट गिरे जबकि प्रायिकता 5 बार चित्त तथा 5 बार पट्ट गिरने की होती है, परन्तु 500 या 1000 बार सिक्का उछालने में चित्त या पट्ट गिरने की संख्या $\frac{1}{2}$ के अत्यन्त ही निकट होगी।

अनुभवसिद्ध या सापेक्ष-आवृत्ति प्रायिकता (Relative Frequency of Occurrence or Empirical Probability)–स्वयंसिद्ध प्रायिकता की धारणा अवसर के खेलों से सम्बन्धित समस्याएँ सुलझाने में उपयोगी रहती हैं परन्तु उसमें कुछ कमियाँ पायी जाती हैं जिनके कारण अन्य प्रकार की अनेक समस्याओं का हल उसमें नहीं पाया जा सकता है। उदाहरणार्थ, एक उत्पादन-प्रक्रिया में निर्मित दोषपूर्ण इकाई की प्रायिकता स्वयंसिद्ध प्रायिकता धारणा के आधार पर नहीं ज्ञात की जा सकती है। इस प्रकार की प्रायिकता निश्चित करने के लिए अनुभव-सिद्ध समंक

या आवृत्ति को संदर्भित करना होता है। किसी घटना के घटने की प्रायिकता भूतकालीन प्रलेखों (Past record) या आवृत्ति-वितरण के आधार पर निर्धारित की जा सकती है। उदाहरणार्थ, यदि एक रेलगाड़ी प्रतिदिन आती-जाती है। पिछले वर्ष (365 दिनों) में वह गाड़ी 13 दिन देर से आयी तो उसकी किसी दिन देर से आने की प्रायिकता : $P = \frac{13}{365}$ होगी।

आत्म-चेतन प्रायिकता (Subjective Probability)–प्रायिकता निर्धारण की आत्म चेतन या व्यक्तिगत (personalistic) धारणा का विकास नया ही है। इस धारणा के अनुसार, किसी घटना के घटने की प्रायिकता उस घटना के घटने के बारे में किसी व्यक्ति का उपलब्ध जानकारी के आधार पर विश्वास की मात्रा से निर्धारित की जाती है। निर्णयकर्त्ता को अक्सर अनिश्चितता के मध्य निर्णय लेना पड़ता है। ऐसी दशा में आत्म-चेतना से किसी घटना के घटने की प्रायिकता ज्ञात की जाती है तथा उसके आधार पर निर्णय लिये जाते हैं। माँग का अनुमान लगाना, मूल्य-स्तर क्या होगा आदि के निर्णय प्राय: व्यक्तिगत अनुमानित प्रायिकता से लिये जाते हैं। असम्भवता (प्रायिकता = 0) से लेकर निश्चितता (प्रायिकता = 1) के मध्य प्रायिकता का अनुमान लगाया जाता है। प्रायिकता चित्र इस तथ्य को स्पष्ट करता है–

सम्भावना-चित्र



छात्र क्रियाकलाप

1. दैव प्रयोग (Rendum Experiment) से आप क्या समझते हैं?

2. 'शर्तपूर्ण प्रायिकताओं का प्रमेय' का क्या अर्थ है?

3. साक्षेप-आवृत्ति प्रायिकता को उदाहरण सहित स्पष्ट कीजिए।

2.7 सारांश (Summary)

1. वास्तव में प्रायिकता सिद्धान्त (Theory of Probability) के अध्ययन की प्रेरणा सर्व-प्रथम जुआरियों से मिली थी। भाग्य देवी (Goddess Fortune) अथवा अन्ध विश्वास (super-stitions) पर आस्था रखने वाले सत्रहवीं शताब्दी के कुछ जुआरी जुए में जीतने के अवसरों से सम्बन्धित समस्याओं का समाधान कराने के लिए तत्कालीन प्रसिद्ध गणितज्ञों के पास जाते थे।

2. प्रायिकता सिद्धान्त के विकास का एक प्रमुख कारण जीवन के लगभग प्रत्येक पहलू में दैव-तत्व (Random phenomenon) की उपस्थिति रहना है। एक स्थिति दैव होती है यदि उसके परिणाम अवसर पर निर्भर करते हैं। सभी सम्बन्ध परिणाम पूर्व ही ज्ञात हो सकते हैं, परन्तु किसी प्रयोग में एक घटना का क्या परिणाम होगा, पूर्व-निर्धारित न किया जा सकता हो।

3. दैव प्रयोग (random experiment) से सम्बन्धित नयी जानकारी मिलने पर प्रायिकताओं में संशोधन किया जा सकता है। प्रायिकताओं में संशोधन करने का विचार हम सभी के लिए जाना-माना है, उन व्यक्तियों के लिए भी जिन्हें प्रायिकता ज्ञात करने का पूर्व अनुभव नहीं है—जो ऐसे वातावरण में रहते हैं जहां अवसर या मनमौजी होने का बोलबाला रहता है और जो अनजाने में प्रायिकता-निर्णय करते हैं।

अभ्यास-प्रश्न (Exercise Questions)

1. प्रायिकता का अर्थ स्पष्ट करते हुए इसके विकास क्रम का उल्लेख कीजिए।
2. प्रायिकता के विभिन्न मापों का वर्णन करो।
3. बेस का प्रमेय (Base's Theorem) की विवेचना कीजिए।

सन्दर्भ ग्रन्थ (Reference Books)

इकाई–3

अनियमित चर एवम् प्रायिकता वितरण (Random variables and probability distribution)

संरचना (Structure)

3.1 उद्देश्य (Objectives)

इस इकाई के अध्ययन के पश्चात् विद्यार्थी योग्य होंगे–

- सैद्धान्तिक वितरण की उपयोगिता को समझने में।
- सामान्य प्रायिकता वक्र को जानने में।
- किसी वितरण को सामान्यीकरण करने में।

3.2 प्रस्तावना (Introduction)

आवृत्ति-वितरण या आवृत्ति-तालिका में समंकों को वर्गों में विभक्त किया जाता है तथा प्रत्येक वर्ग में आने वाले पदों की संख्या को अभिलिखित किया जाता है। आवृत्ति-वितरण की रचना दो प्रकार से की जा सकती है–

(1) वास्तविक या अवलोकित आवृत्ति वितरण (Empirical or Observed Frequency Distribution)

(2) सैद्धान्तिक अथवा प्रत्याशित आवृत्ति वितरण (Theoretical or Expected Frequency Distribution)

वास्तविक आवृत्ति-वितरण की रचना सांख्यिकीय अनुसन्धानों से उपलब्ध समंकों या अवलोकित समंकों के आधार पर की जाती है। उदाहरणार्थ, किसी कारखाने में 100 श्रमिकों की साप्ताहिक मजदूरी के आधार पर बनी आवृत्ति तालिका, वास्तविक आवृत्ति वितरण कहलायेगी। इस प्रकार का वितरण निम्न प्रकार का होगा–

साप्ताहिक मजदूरी (₹ में)	:	0–20	20–40	40–60	60–80	80–100	योग
श्रमिकों की संख्या	:	10	15	30	25	20	100

वास्तविक अवलोकनों के आधार पर बने आवृत्ति-वितरण के विपरीत, प्रायिकता सिद्धान्त (Theory of Probability) के आधार पर प्रत्याशित आवृत्तियाँ ज्ञात करके भी आवृत्ति-वितरण की रचना की जा सकती है। इस प्रकार संरचित आवृत्ति वितरण को सैद्धान्तिक अथवा प्रायिकता आवृत्ति वितरण की संज्ञा दी जाती है। "एक दैव-चर के प्रत्येक सम्भव मूल्य की प्रायिकताओं की सूची प्रायिकता वितरण कहलाता है।" एक दैव चर (random variable) के प्रत्येक सम्भव मूल्य की प्रायिकता ज्ञात की जाती है, अतः सभी मूल्य की प्रायिकताओं का योग 1 होना चाहिए। उदाहरणार्थ, यदि तीन सिक्कों को एक साथ 128 बार उछाला जाय तो ये तीन सिक्के अग्र 8 प्रकार से गिरेंगे–

	1	2	3	4	5	6	7	8
प्रथम सिक्का :	H	T	H	H	T	T	H	T
द्वितीय सिक्का :	H	H	T	H	T	H	T	T
तृतीय सिक्का :	H	H	H	T	H	T	T	T

यहाँ पर H = Head (चित्त), T = Tail (पट्ट)

सैद्धान्तिक अथवा प्रायिकता वितरण इस प्रकार होगा–

चित्त सिक्कों की संख्या (*Number of Heads*)	प्रायिकता (*Probability*)	प्रत्याशित आवृत्ति (*Expected Frequency*)
0	$\frac{1}{8}$	16
1	$\frac{3}{8}$	48
2	$\frac{3}{8}$	48
3	$\frac{1}{8}$	16
योग (Total)	1	128

3.3 सैद्धान्तिक वितरण की उपयोगिका (Utility of Theoretical Distribution)

सैद्धान्तिक आवृत्ति वितरण आधुनिक सांख्यिकी में महत्त्वपूर्ण आधार प्रदान करते हैं तथा प्रमुख आवृत्ति वितरण अनेक परिस्थितियों में निर्णय लेने में सहायक होते हैं। सैद्धान्तिक आवृत्ति वितरण की उपयोगिता निम्न दशाओं में होती है।–

(1) निश्चित मान्यताओं तथा दशाओं के अन्तर्गत आवृत्ति वितरण की प्रवृत्ति क्या होगी। इससे जोखिम तथा अनिश्चितता का विश्लेषण किया जा सकता है।

(2) प्रत्याशित आवृत्ति वितरण अनिश्चितता के मध्य निर्णय लेने में योग देता है।

(3) प्रत्याशित समंकों की सहायता से पूर्वानुमान किये जा सकते हैं।

(4) जब वास्तविक अनुसंधानों से अधिक व्यय हो तो प्रत्याशित आवृत्ति वितरण का स्थानापन्न हो सकता है।

(5) वास्तविक आवृत्ति वितरण को सैद्धान्तिक आवृत्ति वितरण से तुलना करके यह पता लगाया जा सकता है कि दोनों में अन्तर प्रतिदर्शन के उच्चावचनों (fluctuations of sampling) के कारण है या अन्य किन्हीं कारणों से।

(6) अनुसन्धान करने से पूर्व यह आभास हो जाता है कि आवृत्ति वितरण की प्रकृति क्या होगी।

(7) अनेक प्रकार की व्यापारिक व अन्य समस्याओं का समाधान इन वितरणों के आधार पर किया जा सकता है। उदाहरणार्थ, बने-बनाये कपड़े बनाने वाला प्रसामान्य आवृत्ति बंटन (Normal Distribution) के आधार पर विभिन्न आकारों में बनाये जाने वाले कपड़ों की मात्रा निश्चित कर सकता है। गुण-नियन्त्रण, रेलवे बुकिंग आदि जहाँ '*Q*' लगते हैं, प्वॉयसन वितरण के आधार पर महत्त्वपूर्ण निर्णय ले सकते हैं।

सैद्धान्तिक वितरण के प्रकार

सांख्यिकीय अध्ययनों में अनेकों सैद्धान्तिक आवृत्ति-वितरणों जैसे, प्रसामान्य वितरण, द्वि-पद वितरण, बहुपद वितरण (multinomial distribution), हायपरजॉमेट्रिक वितरण (Hypergeometric Distribution), काई-वर्ग वितरण आदि का प्रयोग किया जाता है। सांख्यिकीय विश्लेषण में निम्नलिखित तीन प्रकार के आवृत्ति वितरण अधिक महत्त्वपूर्ण हैं–

(*i*) द्विपद-वितरण (Binomial Distribution)

(*ii*) प्वॉयसन वितरण (Poisson Distribution)

(*iii*) प्रसामान्य वितरण (Normal Distribution)

मनुष्य को जीवन के विभिन्न क्षेत्रों में भावी अनिश्चितताओं (Uncertainities) के होते हुए भी उनसे सम्बन्धित निर्णय लेने होते हैं। उसके निर्णय इन अनिश्चित घटनाओं के सम्बन्ध में उसकी प्रत्याशा (Expectations) पर आधारित होते हैं। वास्तव में जीवन का सुख अनिश्चितताओं तथा प्रत्याशाओं के बीच निहित होता है। किसी अनिश्चित घटना के प्रति तर्कसंगत अनुमान को सम्भाविता या प्रायिकता (Probability) के द्वारा व्यक्त किया जा सकता है। जैसे कहा जाता है किसी सिक्के को उछालने पर उसके चित्त (Head) गिरने की सम्भावना 1/2 है, या छः पहलू वाले पासे (Six-Faced Die) के फेंके जाने पर पाँच के आने के सम्भावना 1/6 है, या उपग्रह इन्सैट 2 बी के सफल होने की 99 प्रतिशत सम्भावना है। ये सभी कथन प्रायिकता कथन (Probability Statements) हैं जो पूर्ववर्ती घटित घटनाओं के आधार पर भविष्य की घटनाओं के लिए लगाये अनुमानों को व्यक्त करते हैं। ध्यान रहें कि ये सभी तथा अन्य प्रायिकता कथन किसी भी घटना के होने अथवा नहीं होने के बारे में कोई निश्चित (Definite) कथन प्रस्तुत नहीं करते हैं वरन केवल अनुमान व्यक्त करते हैं जो गलत भी सिद्ध हो सकता है। गणितज्ञों को प्रायिकता सिद्धान्त का विधिवत अध्ययन करने की प्रेरणा सत्रहवीं शताब्दी में उस समय मिली जब यूरोप के जुआरियों तथा सटोरियों ने भाग्यलक्ष्मी (Goddess of Fortune) की चपलता से निराश होकर तत्कालीन गणितज्ञों से अपनी द्यूत क्रीड़ा सम्बन्धी समस्याओं के समाधान के लिए परामर्श लेना प्रारम्भ किया। तब गणितज्ञों ने भावी घटनाओं के घटित होने का विधिवत अध्ययन किया जिसके फलस्वरूप प्रायिकता सिद्धान्त का प्रादुर्भाव हुआ। किन्तु आज प्रायिकता सिद्धान्त का उपयोग जुआरियों तथा सटोरियों तक सीमित नहीं है। यह सिद्धान्त वास्तव में उन सभी क्षेत्रों में अत्यन्त महत्त्वपूर्ण तथा उपयोगी है जहाँ घटनाएँ अनिश्चित होती हैं तथा उनके भविष्य में घटित होने के सम्बन्ध में अनुमान लगाने की आवश्यकता होती है।

घटनाएँ कई प्रकार की हो सकती हैं। जब एक समय में केवल एक ही घटना होने की बात की जाती है तब इसे **सरल घटना** (Simple Event) कहते हैं किन्तु जब एक साथ कई घटनाओं के घटित होने की चर्चा की जाती है तब ऐसी घटनाओं को **संयुक्त घटना** (Compound Event) कहा जाता है। जैसे एक सिक्के को उछालने या एक पाँसे को फेंकने की घटनायें सरल घटनायें हैं जबकि एक साथ कई सिक्कों को उछालने या कई पाँसों को फेंकने की घटनाएँ संयुक्त घटना का उदाहरण है। संयुक्त घटनाएँ परस्पर स्वतंत्र भी हो सकती हैं तथा परस्पर परतन्त्र या आश्रित भी हो सकती हैं। जब एक साथ घटित होने वाली घटनाएँ एक दूसरे को प्रभावित नहीं करती हैं तब उन्हें **स्वतंत्र घटनायें** (Independent Events) कहते हैं। इसके विपरीत जब एक घटना का दूसरी घटनाओं पर प्रभाव

पड़ता है तब घटनाओं को **आश्रित घटनायें** (Dependent Events) कहते हैं। जैसे दो सिक्कों को एक साथ उछालना स्वतंत्र घटनायें हैं क्योंकि एक सिक्के के चित्त या पट गिरने का दूसरे सिक्के के गिरने पर कोई प्रभाव नहीं पड़ेगा। परन्तु ताश की गड्डी से एक पत्ता खींचने के बाद उसे बिना रखे दूसरा पत्ता खींचने की घटना परस्पर आश्रित घटनायें हैं क्योंकि पहले खींचा गया पत्ता पुन: नहीं खींचा जा सकता है। आश्रित घटनाओं की प्रायिकता प्रतिबंधित (Conditioned) होती है। जो घटनायें एक साथ घटित नहीं हो सकती हैं उन्हें **परस्पर अपवर्जी घटनाएँ** (Mutually Exclusive Events) कहते हैं। जैसे सिक्के के उछालने पर चित्त व पट का आना परस्पर अपवर्जी घटनायें हैं क्योंकि एक का आना दूसरे को रोक देता है जिसके कारण दोनों एक साथ नहीं आ सकते हैं। इसी तरह से पासे को उछालने पर 1, 2, 3, 4, 5 व 6 अंकों का आना परस्पर अपवर्जी है क्योंकि पांसे को उछालने पर ये अंक एक साथ कभी नहीं आ सकते हैं।

प्रायिकता किसी घटना के घटित होने अथवा घटित न होने के सम्बन्ध में प्रत्याशा की द्योतक होती है। दूसरे शब्दों में कहा जा सकता है कि समान रूप से सम्भाव्य अनेक घटनाओं में किसी एक घटना के घटित होने की प्रायिकता वांछित घटना के अनुकूल परिस्थितियों की संख्या के साथ कुल सम्भाव्य घटनाओं की संख्या का अनुपात होता है। अत: प्रायिकता या प्रायिकता अनुपात (Probability Ratio) को अग्रांकित सूत्र से ज्ञात किया जा सकता है–

$$\textbf{प्रायिकता अनुपात } (\text{P.R.}) = \frac{\text{वांछित घटनाओं की संख्या}}{\text{कुल सम्भाव्य घटनाओं की संख्या}}$$

किसी सिक्के को उछालने पर दो सम्भाव्य परिणाम हो सकते हैं–या तो सिक्का चित्त (Head) गिरेगा अथवा सिक्का पट (Tail) गिरेगा। अत: चित्त आने की प्रायिकता 1/2 है। इसी प्रकार से पट आने की प्रायिकता भी 1/2 है। छ: पहलू वाले पासे (Six Faced Die) के फेंकने पर छ: भिन्न-भिन्न सम्भावना प्राप्त हो सकती हैं अर्थात् 1, 2, 3, 4, 5 व 6 में से कोई भी पहलू ऊपर आ सकता है। इसलिए प्रत्येक अंक के आने की प्रायिकता 1/6 होगी। इसी प्रकार से 52 ताश के पत्तों की गड्डी (Pack of 52 Playing Cards) से यों ही एक पत्ता खींचने पर उसका हुक्म का बादशाह (King of Spade) होने की प्रायिकता 1/52 है जबकि उसका चारों में से कोई भी बादशाह (King either of Heart, Diamond, Spade or Club) होने की प्रायिकता 4/52 अर्थात् 1/13 है तथा उसका हुक्म को कोई भी पत्ता होने की प्रायिकता 13/52 अर्थात् 1/4 है।

दो सिक्कों को एक साथ उछालने पर चार सम्भावनाएँ हो सकती हैं–

(*i*) पहला तथा दूसरा दोनों ही सिक्के चित्त गिरे,

(*ii*) पहला सिक्का चित्त तथा दूसरा सिक्का पट गिरे,

(*iii*) पहला सिक्का पट तथा दूसरा सिक्का चित्त गिरे,

(*iv*) दोनों ही सिक्के पट गिरे।

इन चारों सम्भावनाओं को संक्षेप में HH, HT, TH व TT से व्यक्त कर सकते हैं जिनमें पहला अक्षर पहले सिक्के के तथा दूसरा अक्षर दूसरे सिक्के के गिरने की स्थिति को बताता है। परन्तु HT व TH एक समान है, अत: चार सम्भावनाओं में से एक HH की, दो HT की तथा एक TT की है। स्पष्ट है कि दोनों चित्त या दोनों पट की प्रायिकता 1/4 होगी जबकि एक चित्त व एक पट की प्रायिकता 2/4 अर्थात् 1/2 होगी। इसी प्रकार से तीन सिक्कों के एक साथ उछालने पर कुल आठ सम्भावनाएँ निम्नवत होंगी–

HHH,	HHT,	HTH,	HTT,
THH,	THT,	TTH,	TTT

क्योंकि HHT = HTH = THH तथा TTH = THT = HTT, इसलिए आठ सम्भावनाओं में HHH व TTT के एक-एक बार तथा HHT व TTH के तीन-तीन बार आने की सम्भावनाएँ हैं। अत: तीनों चित्त या तीनों पट की प्रायिकता 1/8 है जबकि दो चित्त एक पट या दो पट एक चित्त की प्रायिकता 3/8 है।

इस प्रकार से दो छ: मुखी पांसों (Six Faced Die) को एक साथ फेंकने पर कुल 36 सम्भावनाएँ हो सकती हैं। इन्हें सारणी 64 में प्रस्तुत किया गया है।

सारणी 3.1. दो पांसों को एक साथ उछालने पर होने वाली 36 सम्भावनाएँ तथा उनकी प्रायिकताएँ

पांसे		कुल अंक	पांसे		कुल अंक	पांसे		कुल अंक	पांसे		कुल अंक	पांसे		कुल अंक	पांसे		कुल अंक
I	II		I	II		I	II		I	II		I	II		I	II	
1	1	2	2	1	3	3	1	4	4	1	5	5	1	6	6	1	7
1	2	3	2	2	4	3	2	5	4	2	6	5	2	7	6	2	8
1	3	4	2	3	5	3	3	6	4	3	7	5	3	8	6	3	9
1	4	5	2	4	6	3	4	7	4	4	8	5	4	9	6	4	10
1	5	6	2	5	7	3	5	8	4	5	9	5	5	10	6	5	11
1	6	7	2	6	8	3	6	9	4	6	10	5	6	11	6	6	12

सारणी 2.1 में प्रस्तुत की गयी दो पांसों को एक साथ फेंकने पर होने वाली 36 संभावनाओं को सारांश रूप में निम्नवत ढंग से प्रस्तुत किया जा सकता है–

सारणी 3.2

कुल अंक	2	3	4	5	6	7	8	9	10	11	12
सम्भावनाओं की संख्या	1	2	3	4	5	6	5	4	3	2	1
प्रायिकता	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

सारणी 2.2 के अवलोकन से यह स्पष्ट है कि दो पांसों को एक साथ उछालने पर कुल अंकों के 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 व 12 होने की प्रायिकता क्रमश: 1/36, 2/36, 3/36, 4/36, 5/36, 6/36, 5/36, 4/36, 3/36, 2/36 व 1/36 हैं।

अत: यदि किसी घटना के होने तथा न होने के सभी ढंग सम-प्रायिक (Equally Likely) हो एवं वह घटना m बार हो सकती हो तथा n बार नहीं हो सकती हो, तब उस घटना के घटित होने की प्रायिकता $m/(m + n)$ होगी तथा उसके घटित न होने की प्रायिकता $n/(m + n)$ होगी। घटना के घटित होने की प्रायिकता को p संकेताक्षर से तथा घटित न होने की प्रायिकता को q संकेताक्षर से लिखते हैं। अत: किसी समप्रायिक घटना के

घटित होने की प्रायिकता, $p = \dfrac{m}{m+n}$

घटित न होने की प्रायिकता, $q = \dfrac{n}{m+n}$

स्पष्ट है कि p तथा q का योग सदैव ही एक के बराबर होगा। अत: p तथा q में से एक के ज्ञात होने पर दूसरे को प्राप्त किया जा सकता है। जैसे ताश के 52 पत्तों में से यादृच्छिक रूप से (Randomly) एक पत्ता निकालने पर उसके कोई भी इक्का (Any Ace) होने की प्रायिकता 4/52 अर्थात् 1/13 तथा इक्का न होने की प्रायिकता 48/52 अर्थात् 12/13 है। इन दोनों प्रायिकताओं का योग एक के बराबर है। प्रायिकता अनुपात का मान शून्य से एक के बीच होता है। शून्य प्रायिकता का अर्थ है घटना के घटित होने की कोई सम्भावना नहीं है जबकि प्रायिकता के एक होने का अर्थ है कि घटना के घटने की पूर्ण सम्भावना अर्थात निश्चितता है। किन्तु प्रायिकता के

ये दोनों मान (Extreme Value) कभी नहीं होते हैं क्योंकि प्रायिकता का शून्य या एक होने का अर्थ है घटना के घटित होने या न होने के सम्बन्ध में निश्चित अनुमान, जो कि प्रायिकता के क्षेत्र से बाहर है। प्रायिकता को भिन्न या दशमलव या प्रतिशत के रूप में प्रस्तुत किया जा सकता है।

ऊपर प्रस्तुत किया गया विवेचन प्रायिकता का गणितीय, सैद्धान्तिक या तार्किक प्रत्यय प्रस्तुत करता है। परन्तु केवल तर्क पर आधारित प्रायिकता का ही प्रयोग नहीं होता है। सांख्यिकी में भूतकालीन अनुभवों तथा प्रयोग से प्राप्त सूचनाओं के आधार पर भी प्रायिकता का अनुमान लगाया जाता है। जैसे मरण तालिकाओं (Mortality Tables) के आधार पर बीमा कम्पनी विभिन्न आयु वाले व्यक्तियों के जीवित रहने की प्रायिकता ज्ञात करती है। या जन्म समंकों (Birth Data) से अनुमान लगाया जाता है कि जन्म लेने वाले शिशु के जीवित रहने की प्रायिकता क्या है? इस प्रकार की प्रायिकता को **आनुभाविक प्रायिकता** (Empirical Probability) या **सांख्यिकीय प्रायिकता** (Statistical Probability) कहते हैं। यदि समान परिस्थितियों में घटी n सम्भाव्य घटनाओं में से कोई वांछित घटना कुल r बार घटित हुई है तब भविष्य में इस वांछित घटना के घटित होने की प्रायिकता को n के अनन्त की ओर प्रवृत्त होने पर प्राप्त r/n की सीमा से व्यक्त किया जा सकता है। अत: सूत्र रूप में लिख सकते हैं कि–

$$p = \operatorname*{Limit}_{n\to\infty} \frac{r}{n}$$

उदाहरण के लिए यदि पिछले जन्म समंकों से ज्ञात होता है कि पिछले एक लाख प्रसवों में से 45 हजार प्रसवों में कन्या का तथा 55 हजार प्रसवों में पुत्र का जन्म हुआ, तब कहा जा सकता है कि कन्या के जन्म की प्रायिकता 45 तथा पुत्र के जन्म की प्रायिकता .55 है। यादृच्छिक चरों (Random Variables) के लिए n के अनन्त की ओर प्रवृत्त होने पर आनुभाविक प्रायिकता का मान तार्किक प्रायिकता के निकट आता जाता है।

प्रायिकता अनुपात ज्ञात करने के लिए वांछित घटना के घटित होने की संख्या तथा घटनाओं की कुल संख्या की ठीक-ठीक ज्ञान होना आवश्यक है। कुल घटनाओं तथा वांछित घटनाओं की संख्या सरलता से ज्ञात करने के लिए क्रमचय (Permutation) तथा संचय (Combination) के प्रत्ययों व सूत्रों का प्रयोग किया जा सकता है। अत: यहाँ पर क्रमचय तथा संचय की प्रारम्भिक जानकारी की चर्चा करना उचित ही होगा।

क्रमचय (Permutation) से तात्पर्य उन सभी क्रमों (Orders) या विन्यासों (Arrangements) से है जिनमें दी गई वस्तुओं या उनमें से ली गई कुछ वस्तुओं को विन्यासित (Arrange) किया जा सकता है। जैसे तीन पुस्तकें A, B व C को छ: भिन्न-भिन्न विन्यासों ABC, ACB, BAC, BCA, CAB तथा CBA में व्यवस्थित किया जा सकता है। अत: इन तीन पुस्तकों के लिए कुल छ: क्रमचय होंगे। परन्तु यदि चार पुस्तकें A, B, C व D दी हो तथा इनमें से दो-दो पुस्तकें लेकर क्रम बनाने हों तब कुल 12 विन्यास AB, AC, AD, BA, BC, BD, CA, CB, CD, DA, DB तथा DC बनेंगे। अत: 4 पुस्तकों में से दो-दो के कुल 12 क्रमचय बनेंगे। क्रमचय में क्रम (order) का विशेष महत्त्व है तथा एक ही वस्तुओं को भिन्न-भिन्न क्रम में रखने पर भिन्न-भिन्न क्रमचय बनते हैं। यही कारण है कि AB तथा BA दो भिन्न-भिन्न क्रमचय हैं जबकि दोनों में वस्तुएँ एक ही है परन्तु क्रम भिन्न होने के कारण क्रमचय अलग-अलग हैं। n असमान वस्तुओं में से r वस्तुओं को एक साथ लेने पर बनने वाले क्रमचयों की संख्या निम्न सूत्र से ज्ञात की जा सकती है–

$$\textbf{क्रमचय}\text{ (Permutations)} = {}^{n}P_{r} = \frac{n\,!}{(n-r)\,!}$$

जहाँ $n!$ को क्रमगुणित एन (Factorial n) कहकर पढ़ा जाता है तथा इसका मान 1 से n तक की सभी संख्याओं के गुणनफल के बराबर होता है। अत:

$$n! = 1 \times 2 \times 3 \times \ldots\ldots \times n$$

परन्तु सुविधा के लिए $n\,!$ के अंक विपरीत क्रम में लिखे जाते हैं अर्थात्

$$n! = n \times (n-1) \times (n-2) \times \ldots\ldots \times 1$$

अत: $4! = 4 \times 3 \times 2 \times 1 = 24$

$$5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$$

क्रमगुणित शून्य (Factorial Zero) तथा क्रमगुणित एक (Factorial One) का मान एक के बराबर होता है। अत: यदि 8 वस्तुओं में 5–5 वस्तुओं को लेकर क्रमचय बनाये जाये तब कुल क्रमचयों की संख्या 6720 होगी, क्योंकि

$$^8P_5 = \frac{8!}{(8-5)!} = 6720$$

क्रम को ध्यान में रखते हुए दी गई वस्तुओं में से कुछ वस्तुओं को लेकर बनने वाले विन्यासों को **संचय** (Combination) कहते हैं। जैसे यदि चार पुस्तकों A, B, C व D में से दो-दो पुस्तकों को लेकर ऐसे विन्यास बनाये जायें जिनमें क्रम पर ध्यान न दिया जाये तब कुल छ: विन्यास AB, AC, AD, BC, BD व CD बनेंगे। स्पष्ट है कि यहाँ संचयों की संख्या क्रमचयों की संख्या से आधी है। ऐसा इसलिए है क्योंकि प्रत्येक संचय के लिए दो क्रमचय बन सकते हैं। जैसे AB संचय से AB तथा BA दो क्रमचय बनेंगे। n असमान वस्तुओं में से r वस्तुओं को एक साथ लेकर बनने वाले संचयों की संख्या निम्न सूत्र से ज्ञात की जा सकती है–

$$\textbf{संचय} \text{ (Combinations)} = {}^nC_r = \frac{n!}{(n-r)!\,r!}$$

अत: यदि 8 वस्तुओं में से 5–5 वस्तुओं को लेकर संचय बनाये जायें तब कुल संचयों की संख्या 56 होगी क्योंकि

$$^8C_5 = \frac{8!}{(8-5)!\,5!}$$

क्रमचय एवं संचय के सूत्रों के अवलोकन से स्पष्ट है कि संचयों की संख्या सदैव ही क्रमचयों की संख्या से कम होती है।

दो या दो से अधिक परस्पर अपवर्जी घटनाओं (Mutually Exclusive Events) में से किसी भी एक घटना (Either of the Events) के घटित होने की प्रायिकता, उनके अलग-अलग घटित होने की प्रायिकताओं का योग होती है। अत: यदि A तथा B परस्पर अपवर्जी घटनाएँ हैं तब A अथवा B के घटित होने की प्रायिकता

PR (A अथवा B) = PR (A) + PR (B)

जैसे पांसे को फेंकने पर प्रत्येक अंक के आने की प्रायिकता 1/6 होती है। अत: 2 व 4 में से किसी के भी आने की प्रायिकता 1/6 + 1/6 = 2/6 = 1/3 होगी। इसी तरह से 2, 4 व 6 में से किसी के भी आने की प्रायिकता 1/2 होगी।

दो या दो से अधिक स्वतंत्र घटनाओं के एक साथ घटित होने की प्रायिकता उनके अलग-अलग घटित होने की प्रायिकताओं के गुणनफल के बराबर होती है। अत: यदि A तथा B स्वतंत्र घटनाएँ हैं तब A तथा B के साथ घटित होने की प्रायिकता

PR (A तथा B) = PR (A) . PR (B)

जैसे पांसें को दो बार फेंकने पर पहली बार 4 व दूसरी बार 5 आने की प्रायिकता 1/6 × 1/6 = 1/36 होगी। इसी तरह से पहली बार सम (Even) अंक तथा दूसरी बार विषम (Odd) अंक आने की प्रायिकता 1/2 × 1/2 = 1/4 होगी।

प्रायिकता के उपरोक्त वर्णित सिद्धान्त का प्रयोग करके वैज्ञानिक ढंग से घटनाओं के घटित होने के सम्बन्ध में अनुमान लगाया जा सकता है। इसके लिए वैज्ञानिक सबसे पहले किसी सुनिश्चित पूर्व-कल्पना (Presumption) अथवा मान्यता (Assumption) के आधार पर विभिन्न घटनाओं के घटित होने से सम्बन्धित गणितीय निदर्श

(Mathematical Model) का प्रतिपादन करते हैं तथा फिर इस निदर्श (Model) के आधार पर भावी घटनाओं या अज्ञात घटनाओं का अनुमान लगाते हैं। वैज्ञानिकों के द्वारा प्रतिपादित ये निदर्श पूर्णत: परिकल्पित (Hypothetical) होते हैं। इन निदर्शों से प्राप्त प्रायिकता अनुपातों को प्रयासों (Trials) की संख्या (N) से गुणा करके विभिन्न घटनाओं के घटित होने की सम्भावित आवृत्तियाँ भी ज्ञात की जा सकती हैं। जैसे किसी सुडौल सिक्के को 50 बार उछालने पर चित्त आने की सम्भावित आवृत्ति 25 होगी तथा पट आने की सम्भावित आवृत्ति भी 25 ही होगी। क्योंकि PR (H) = 1/2 जिसे 50 से गुणा करने पर 25 प्राप्त होगा तथा PR(T) = 1/2 जिसे 50 से गुणा करने पर 25 प्राप्त होगा। स्पष्ट है कि 50 बार की उछाल में चित्त (H) तथा पट (T) आने की आवृत्ति 25-25 ही होगी। इसी प्रकार से दो सिक्कों तथा तीन सिक्कों को एक साथ अनेक बार उछालने पर प्राप्त होने वाले परिणामों की सम्भावित आवृत्तियाँ ज्ञात की जा सकती हैं। सारणी 2.3 में दो सिक्कों को तथा सारणी 2.4 में तीन सिक्कों को 200 बार उछालने पर प्राप्त होने वाले गणितीय निदर्श प्रस्तुत किये गये हैं–

सारणी 3.3

दो सिक्कों की 200 बार उछाल का गणितीय निदर्श

सिक्कों की स्थिति		प्रायिकता	आवृत्ति
दोनों चित्त	HH	1/4	50
एक चित्त तथा एक पट	HT या TH	2/4	100
दोनों पट	TT	1/4	50
कुल		1.00	200

सारणी 3.4

तीन सिक्कों की 200 बार उछाल का गणितीय निदर्श

सिक्कों की स्थिति		प्रायिकता	आवृत्ति
तीनों चित्त	HHH	1/8	25
दो चित्त तथा एक पट	HHT HTH THH	3/8	75
एक चित्त तथा दो पट	HTT TTT	3/8	75
तीनों पट	TTT	1/8	25
	कुल	1.00	200

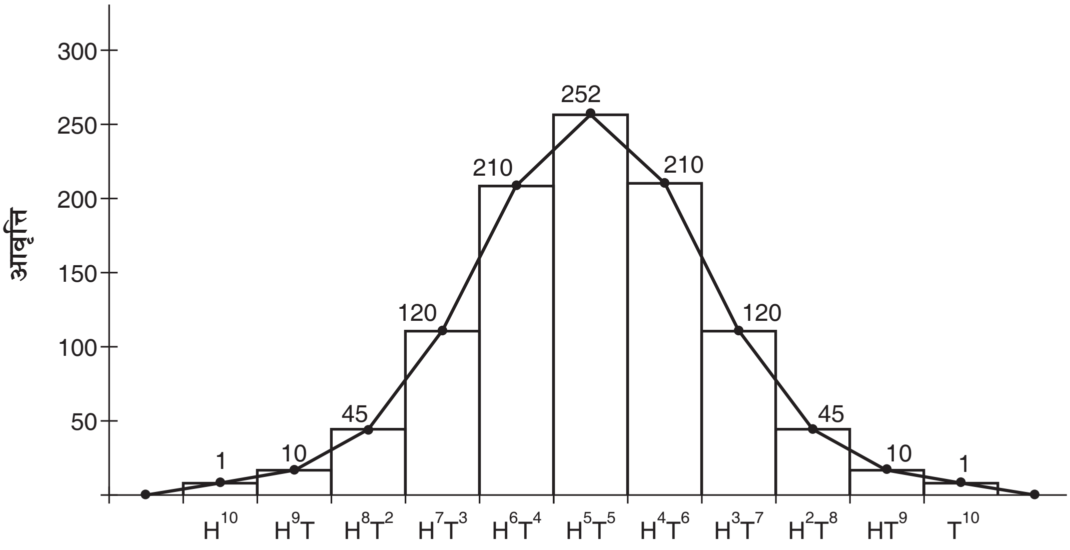
इसी तरह से यदि दस सिक्कों को एक साथ अनेक बार (सुविधा के लिए माना कि 1024 बार) उछाला जाये तब प्रायिकता सिद्धान्त के आधार पर प्राप्त होने वाले सम्भावित आवृत्ति वितरण को सारणी 2.5 में दिया गया है इस सम्भावित वितरण को आवृत्ति बहुभुज के रूप में चित्र 1 में प्रस्तुत किया गया है। गणितीय निदर्श (Mathematical Model) के आधार पर तैयार किये गये।

सारणी 3.5. दस सिक्कों को 1024 बार उछालने पर सम्भावित आवृत्ति वितरण

सिक्कों की स्थिति		प्रायिकता	आवृत्ति
दसों सिक्के चित्त	H^{10}	1/1024	1
नौ सिक्के चित्त तथा एक सिक्का पट	H^9T	10/1024	10
आठ सिक्के चित्त तथा दो सिक्के पट	H^8T^2	45/1024	45
सात सिक्के चित्त तथा तीन सिक्के पट	H^7T^3	120/1024	120
छ: सिक्के चित्त तथा चार सिक्के पट	H^6T^4	210/1024	210
पाँच सिक्के चित्त तथा पाँच सिक्के पट	H^5T^5	252/1024	252
चार सिक्के चित्त तथा छह सिक्के पट	H^4T^6	210/1024	210

नोट

तीन सिक्के चित्त तथा सात सिक्के पट	H^3T^7	120/1024	120
दो सिक्के चित्त तथा आठ सिक्के पट	H^2T^8	45/1024	45
एक सिक्का चित्त तथा नौ सिक्के पट	HT^9	10/1024	10
दसों सिक्के पट	T^{10}	1/1024	1
	कुल	1.00	1024



चित्र 1. सारणी में प्रस्तुत आवृत्ति वितरण का रेखाचित्रीय निरूपण

इस प्रकार के आवृत्ति वितरण को **सैद्धान्तिक आवृत्ति वितरण** (Theoritical Frequency Distribution) या **प्रत्याशित आवृत्ति वितरण** (Expected Frequency Distribution) या **आदर्श आवृत्ति वितरण** (Ideal Frequency Distribution) कहते हैं। अत: सैद्धान्तिक आवृत्ति वितरण वह वितरण है जो किन्हीं पूर्व निश्चित मान्यताओं (Assumptions) के आधार पर गणितीय ढंग से तैयार किया जाता है। स्पष्ट है कि सैद्धान्तिक वितरण वास्तविक अवलोकनों (Actual Observations) पर आधारित नहीं होते हैं। यहाँ यह कहना उचित ही होगा कि सिक्कों की उछाल के लिए बने सभी सैद्धान्तिक वितरण सममित (Symmetrical) हैं। नि:सन्देह सैद्धान्तिक वितरण गणितीय प्रायिकता सिद्धान्त पर आधारित होता है परन्तु यदि वास्तव में सिक्कों को उछाला जायेगा तब भी लगभग ऐसे ही परिणाम प्राप्त होंगे। अनेक प्रयोगों ने सिद्ध कर दिया है कि उछालों की संख्या के बढ़ते जाने पर वास्तविक या अवलोकित वितरण (Real or Observed Distribution) प्रायिकता पर आधारित सैद्धान्तिक वितरण के अनुरूप होता जाता है। सैद्धान्तिक आवृत्ति वितरण वास्तव में आधुनिक सांख्यिकी के प्राण हैं। अनेक परिस्थितियों में इनका प्रयोग अपरिहार्य होता है। सैद्धान्तिक वितरणों के प्रमुख उपयोग निम्नवत हैं–

(*i*) इससे ज्ञात हो जाता है कि दी गई निश्चित मान्यताओं के अन्तर्गत समंकों के वितरण की प्रवृत्ति क्या होगी।

(*ii*) ये विवेकपूर्ण निर्णय लेने का आधार बन सकते हैं।

(*iii*) ये भावी अनुमान लगाने में सहायक होते हैं।

(*iv*) वास्तविक वितरणों की प्राप्ति के अत्यधिक व्ययसाध्य या असम्भव होने पर सैद्धान्तिक वितरण अवलोकित वितरणों के स्थानापन्न (Substitute) होते हैं।

(*v*) अवलोकित वितरणों की तुलना सैद्धान्तिक वितरणों से करके प्रतिचयन उच्चावचनों (Sampling Fluctuations) की सार्थकता ज्ञात की जा सकती है।

(*vi*) सैद्धान्तिक वितरण वास्तविक अनुसंधान की सार्थकता तथा व्यय का अनुमान लगाने में सहायक होते हैं।

(*vii*) सैद्धान्तिक वितरण दैनिक जीवन की समस्याओं के समाधान में सहायक होते हैं।

अनेक सिक्कों को एक साथ उछालने पर आने वाले विभिन्न सम्भावित संचयों की संख्या ज्ञात करना पाठकों को सम्भवत: कुछ जटिल सा प्रतीत हो रहा होगा। परन्तु इसके लिए विभिन्न संचयों को बनाने की आवश्यकता नहीं है। इस प्रकार की स्थिति में प्रायिकता ज्ञात करने के लिए द्विपदीय विस्तार (Binomial Expansion) की गणितीय समीकरण का प्रयोग किया जा सकता है। दो सिक्कों की स्थिति में यह समीकरण होगी।

$$\left(\frac{1}{2}+\frac{1}{2}\right)^2 = \left(\frac{1}{2}\right)^2 + 2\left(\frac{1}{2}\times\frac{1}{2}\right) + \left(\frac{1}{2}\right)^2$$

$$= \frac{1}{4}+\frac{2}{4}+\frac{1}{4}$$

इस समीकरण के दाँये पक्ष के तीनों पद क्रमश: दो चित्त (Two H), एक चित्त (One H) तथा कोई चित्त नहीं (No H) के आने की प्रायिकता बताते हैं। इसी तरह से तीन सिक्कों को उछालने की स्थिति में समीकरण होगी–

$$\left(\frac{1}{2}+\frac{1}{2}\right)^3 = \left(\frac{1}{2}\right)^3 + 3\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right) + 3\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^3$$

$$= \frac{1}{8}+\frac{3}{8}+\frac{3}{8}+\frac{1}{8}$$

इसी प्रकार से चार सिक्कों को एक साथ उछालने के लिए द्विपदीय समीकरण (Binomial Equation) निम्नवत होगी–

$$\left(\frac{1}{2}+\frac{1}{2}\right)^4 = \left(\frac{1}{2}\right)^4 + 4\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right) + 6\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2 + 4\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^3 + \left(\frac{1}{2}\right)^4$$

$$= \frac{1}{16}+\frac{4}{16}+\frac{6}{16}+\frac{4}{16}+\frac{1}{16}$$

उपरोक्त समीकरणों में प्रत्येक पद का हर (Denominator) कुल सम्भाव्य घटनाओं (outcomes) की संख्या को तथा अंश (Numerator) वांछित घटनाओं की सम्भाव्य संख्या को व्यक्त कर रहा है। चित्त (Head) आने की प्रायिकता को p से तथा पट (Tail) आने की प्रायिकता को q से लिखें, तब n सिक्कों को उछालने पर विभिन्न संचयों में चित्त (Head) तथा पट (Tail) आने की प्रायिकता को $(p + q)^n$ के विस्तार से निम्नवत ढंग से व्यक्त किया जा सकता है–

एक सिक्के के लिए, $(p + q)^1 = p + q$

दो सिक्कों के लिए, $(p + q)^2 = p^2 + 2pq + q^2$

तीन सिक्कों के लिए, $(p + q)^3 = p^3 + 3p^2q + 3pq^2 + q^3$

चार सिक्कों के लिए, $(p + q)^4 = p^4 + 4p^3q + 6p^2q^2 + 4pq^3 + q^4$

n सिक्कों के लिए, $(p + q)^n = p^n + np^{n-1}q^1 + \frac{n(n-1)}{n}p^{n-2}q^2 + \ldots q^n$

उपरोक्त समीकरणों को एक सामान्य समीकरण (General Equation) के रूप में निम्नवत ढंग से लिखा जा सकता है–

$$= {}^nC_0p^nq^0 + {}^nC_1p^{n-1}q^1 + {}^nC_2p^{n-2}q^2 + \ldots {}^nC_np^0q^n$$

इस सामान्य समीकरण की सहायता से स्पष्ट है कि $n = 5$ के लिए द्विपदीय विस्तार (Binomial Expansion) निम्नवत होगा।

$$(p + q)^5 = {}^5C_0p^5q^0 + {}^5C_1p^4q^1 + {}^5C_2p^2q^3 + {}^5C_4p^1q^4 + {}^5C_5p^0q^5$$

$$= 1p_5 + 5p^4q + 10p^3q^2 + 10p^2q^3 + 5pq^4 + 1q^5$$

$(p + q)^n$ के लिए उपरोक्त वर्णित सामान्य विस्तार को **द्विपद वितरण** (Binomial Distribution) कहते हैं। इसका प्रतिपादन स्विस गणितज्ञ जेम्स बर्नोली (James Bernoulli) ने किया था जिनके नाम पर इसे बर्नोली वितरण भी कहते हैं। द्विपद का अर्थ है दो पद, एक पद घटना की सफलता को तथा दूसरा विफलता को इंगित करता है। जैसे पीछे p को चित्त आने (Head) के लिए तथा q को चित्त न आने (Non-Head) के लिए प्रयुक्त किया गया है। उपरोक्त से स्पष्ट है कि द्विपद वितरण में $(n + 1)$ पद होते हैं। द्विपद वितरण का प्रत्येक पद संख्यात्मक गुणांक (Numerical Coefficient), p के घातांक (Power of p) तथा q के घातांक (Power of q) की गुणा से बना होता है। प्रथम तथा अन्तिम पद का संख्यात्मक गुणांक सदैव एक होता है इसलिए इसे लिखने की आवश्यकता नहीं होती। प्रथम पद में p की घात n तथा q की घात शून्य होती है जिसे p^n से लिखते हैं। आगामी पदों में p की घात क्रमश: एक-एक घटती जाती है तथा q की घात बढ़ती जाती है। अन्तिम पद में p की घात शून्य होती है तथा q की घात n हो जाती है जिसे q^n से लिखते हैं। स्पष्ट है कि प्रत्येक पद में p तथा p की घातों का योग n के बराबर होता है। द्विपद वितरण के गुणांक सदैव सममितीय (Symmetrical) होते हैं तथा इनका योग 2^n के बराबर होता है जो कुल सम्भावनाओं की संख्या को बताता है। द्विपद वितरण में p तथा q का मान समान होने की कोई शर्त नहीं है परन्तु p तथा q का योग सदैव एक के बराबर होता है। द्विपद वितरण में p तथा q का मान समान होने पर द्विपद विस्तार पूर्णतया सममित होता है परन्तु यदि p तथा q असमान होते हैं तब द्विपद विस्तार के विभिन्न पदों का वितरण असममित होता है। इस स्थिति में n के मान में वृद्धि के साथ असममितता कम होती जाती है। द्विपद वितरण के विभिन्न पदों के संख्यात्मक गुणांकों (Numerical Coefficients) को **पास्कल त्रिभुज** (Pascal Triangle) की सहायता से सरलता से ज्ञात किया जा सकता है।

सारणी 3.6. $(p + q)^n$ के विस्तार के लिए संख्यात्मक गुणांकों का पास्कल त्रिभुज

घात (n)	गुणांक	योग
1	1 1	2
2	1 2 1	4
3	1 3 3 1	8
4	1 4 6 4 1	16
5	1 5 10 10 5 1	32
6	1 6 15 20 15 6 1	64
7	1 7 21 35 35 21 7 1	128
8	1 8 28 56 70 56 28 8 1	256
9	1 9 36 84 126 126 84 36 9 1	512
10	1 10 45 120 210 252 210 120 45 10 1	1024

पास्कल त्रिभुज के अवलोकन से स्पष्ट होगा कि प्रत्येक पंक्ति के दोनों किनारों के गुणांक एक के बराबर है जबकि अन्य सभी गुणांक उससे ऊपर की पंक्ति के दोनों ओर के दो गुणांकों का योग है। जैसे चौथी पंक्ति में 6 = 3 + 3, छठी पंक्ति में 15 = 5 + 10 तथा सातवीं पंक्ति में 21 = 6 + 15 आदि। पास्कल त्रिभुज से स्पष्ट है कि $n = 5$ के लिए गुणांकों का मान क्रमश: 1, 5, 10, 10, 5 व 1 है जो पूर्ववत ज्ञात गुणांकों के समान ही हैं।

द्विपद वितरण निदर्श (Binomial Distribution Model) का व्यावहारिक विज्ञानों में अत्यन्त उपयोगी एवं महत्त्वपूर्ण स्थान है। जिन क्षेत्रों में घटनाओं की सफलता-असफलता से आधार पर उनका द्वि-विभाजन (Dichotomous Classification) किया जा सकता है वहां द्विपद वितरण का प्रयोग उचित होता है। इस निदर्श का व्यावहारिक उपयोग एक सरल उदाहरण से स्पष्ट हो सकेगा। जैसे यदि किसी छात्र को किसी विषय पर 10 सत्यासत्य प्रश्नों वाला परीक्षण हल करने के लिए दिया गया। हमारी मान्यता है कि वह छात्र उस विषय में कुछ

भी नहीं जानता है तथा वह प्रश्नों का उत्तर यादृच्छिक ढंग से (Randomly) देगा। स्पष्टत: उस विषय में कुछ जानने पर छात्र का झुकाव सही उत्तरों की ओर अधिक होगा जबकि न जानने पर सही तथा गलत दोनों प्रकार के उत्तरों की ओर समान रूप से होगा। यह देखने के लिए कि क्या वास्तव में छात्र के द्वारा दिये गये उत्तर यादृच्छिक हैं। अथवा उसका झुकाव सही उत्तरों की ओर हैं, उसके वास्तविक प्रत्युत्तरों की तुलना यादृच्छिक प्रत्युत्तरों से करनी होगी। अत: द्विपद वितरण $(p + q)^{10}$ के रूप में यादृच्छिक परिकल्पना बनानी होगी जो यादृच्छिक ढंग से उत्तर देने की स्थिति में सही उत्तरों के विभिन्न संचयों की प्रायिकता को बता सकें। नि:सन्देह $n = 10$ की स्थिति में, जैसा कि पास्कल त्रिभुज से स्पष्ट होगा, 5 प्रश्नों को सही करने की प्रायिकता सर्वाधिक (252/1024) होगी, अत: यदि छात्र 5 से अधिक प्रश्न सही हल करता है तब कहा जा सकता है कि उसके उत्तर देने की प्रवृत्ति यादृच्छिक न होकर सही उत्तरों के प्रति अधिक है। दूसरे शब्दों में कहा जा सकता है कि इस बात की सम्भावना है कि वह छात्र सम्बन्धित विषय में कुछ-न-कुछ अवश्य जानता है। द्विपद वितरण निदर्श को बहुविकल्प प्रश्नों के लिये भी प्रयुक्त किया जा सकता है। जैसे चार विकल्प वाले दो प्रश्नों के परीक्षण पर $p = 1/4$ तथा $q = 3/4$ होगा। तब द्विपद वितरण निम्नवत होगा–

$$\left(\frac{1}{4}+\frac{3}{4}\right)^2 = \left(\frac{1}{16}\right) + 2\left(\frac{3}{16}\right) + \left(\frac{9}{16}\right)$$

अत: यादृच्छिक ढंग से दोनों प्रश्नों को सही हल करने की सम्भावना 1/16 होगी, एक प्रश्न सही हल करने की सम्भावना 6/16 होगी तथा दोनों प्रश्न गलत हल करने की सम्भावना 9/16 होगी। ये दोनों उदाहरण द्विपद वितरण की व्यावहारिक उपयोगिता की पुष्टि के लिए पर्याप्त हैं। जनसंख्या में पुरुष-महिला का होना, कारखाने में निर्मित वस्तुओं का दोषपूर्ण व दोषरहित होना जैसी द्वन्द्वात्मक घटनाओं से सम्बन्धित व्यवहारिक समस्याओं के अध्ययन में भी इस वितरण का उपयोग हो सकता है।

(I) **द्विपद-वितरण** (Binomial Distribution)

द्विपद-वितरण के साथ स्विस गणितज्ञ जेम्स बर्नोली (James Bernoulli, 1654-1705) का नाम जुड़ा हुआ है तथा इसे बर्नोली वितरण (Bernaulli Distribution, Bernaulli Trial or Bernaulli Process) भी कहा करते हैं। इस सैद्धान्तिक वितरण का प्रतिपादन बर्नोली ने ही किया था और इसका प्रकाशन उनकी मृत्यु के 8 वर्ष पश्चात् 1713 में हुआ था। द्विपद (Binomial) का अर्थ 'दो नाम' (Two names) है, अत: आवृत्ति वितरण दो वर्गों के अन्तर्गत आता है। द्विपद-वितरण एक खण्डित आवृत्ति वितरण है जो द्वन्द्वात्मक-विकल्पों–एक समंक समूह की सफलता व असफलता के रूप में सम्भावनाओं पर आधारित होता है। द्विपद-वितरण की विशेषताएँ निम्नलिखित हैं–

(1) प्रयोगों (trials) की संख्या निश्चित होती है।

(2) प्रत्येक प्रयोग के पारस्परिक अपवर्जी (mutually exclusive) परिणाम दो होते हैं।

(3) प्रयोग की सफलता का संकेताक्षर 'p' होता है तथा असफलता का 'q' । 'q', $1 - p$, के बराबर होता है।

(4) प्रयोग (trials) स्वतन्त्र (independent) होते हैं, अर्थात् किसी एक प्रयोग के परिणाम का प्रभाव आगे किये जाने वाले प्रयोगों के परिणामों पर नहीं पड़ता है।

सामान्य रूप से द्विपद वितरण की विशेषताओं के कारण इसका प्रयोग अनेक दशाओं में सम्भव होता है। यह वितरण किसी भी दशा में व्यवहार में लाया जा सकता है जहाँ किसी दैव-प्रयोग के 'n' स्वतन्त्र प्रयोग निहित हों, प्रत्येक प्रयोग के दो सम्भव परिणाम हों, सफलता (success), जिसकी प्रायिकता 'p' हो तथा असफलता (failure) जिसकी प्रायिकता 'q' हो जो '$1 - p$' के बराबर होती है। यहाँ पर इस बात पर बल दिया जा सकता है कि प्रयोग स्वतन्त्र होना चाहिए, तथा सफलता की प्रायिकता (अत: असफलता की प्रायिकता भी) विभिन्न प्रयोगों में नहीं बदलनी चाहिए। 'सफलता' तथा 'असफलता' को दो सम्भव परिणामों के रूप में माना जाना चाहिए न कि उनके शाब्दिक अर्थ के रूप में।

सिक्का उछालने के प्रयोगों के उदाहरण से द्विपद वितरण की व्युत्पत्ति को भली-भाँति दर्शाया जा सकता है। एक सच्चा (fair) सिक्का उछालने में पारस्परिक अपवर्जी दो परिणाम चित्त गिरना (Head up) या पट्ट गिरना (Tail up)—हो सकते हैं। सिक्के के चित्त गिरने की प्रायिकता : $p = \frac{1}{2}$ है तथा उसके पट्ट गिरने की प्रायिकता : $q = 1 - p$ या $1 - \frac{1}{2} = \frac{1}{2}$ है। अत: $p + q = 1$ इसे द्विपद सूत्र को इस रूप $(p + q)^n$ में व्यक्त किया जा सकता है।

इसी प्रकार यदि दो सिक्के (A तथा B) एक साथ उछाले जायँ तो सम्भावित परिणाम निम्न प्रकार होंगे :

A	B	
H	H	
T	H	
H	T	(H = Head, T = Tail)
T	T	

दोनों सिक्कों को एक साथ उछालने पर उनके चित्त गिरने (H) की सम्भावना $\frac{1}{4}$ है तथा पट्ट गिरने (T) की सम्भावना भी $\frac{1}{4}$ है। सम्भावना प्रमेय से भी यही परिणाम ज्ञात होता है :

प्रथम सिक्के को उछालने में चित्त (H) गिरने की सम्भावना = $\frac{1}{2}$

द्वितीय सिक्के को उछालने में चित्त (H) गिरने की सम्भावना = $\frac{1}{2}$

दोनों सिक्कों की एक साथ उछाल में दोनों के चित्त (H) गिरने की सम्भावना = $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$ है। पट्ट गिरने (T) की सम्भावना भी $\frac{1}{4}$ है।

यह चित्त गिरने की घटना को सफलता माना जाय और उसकी सम्भावना का संकेत 'p' रखा जाय तथा पट्ट गिरने की घटना को असफलता माना जाय और उसकी सम्भावना का संकेत 'q' रखा जाय तो उपर्युक्त परिणामों को निम्न प्रकार व्यक्त किया जा सकता है—

$$\underbrace{\underset{pp}{\text{HH}}}_{\text{or } p^2} \qquad \underbrace{\underset{pq}{\text{HT}} \qquad \underset{qp}{\text{TH}}}_{2pq} \qquad \underbrace{\underset{qq}{\text{TT}}}_{q^2}$$

$p^2 + 2\,pq + q^2$, $(p + q)^2$ का विस्तार है। अत: दो स्वतन्त्र घटनाओं की सामूहिक सम्भावना को सरल द्विपदी सूत्र $(p + q)^2$ है।

सिक्के के उछालने में उसके चित्त गिरने की सम्भावना $(p) = \frac{1}{2}$

सिक्के के उछालने में उसके पट्ट गिरने की सम्भावना $(q) = \frac{1}{2}$

अत: $(p + q)^2 = p^2 + 2\,pq + q^2$ or $\left(\frac{1}{2} + \frac{1}{2}\right)^2$ or $\frac{1}{4} + \frac{1}{2} + \frac{1}{4}$

इसी प्रकार तीन सिक्के—A, B, C उछालने पर 8 सम्भावित परिणाम हो सकते हैं—

A	B	C	
H	H	H	$= p^3$
H	H	T	
H	T	H	$= 3p^2q$
T	H	H	
T	H	T	
T	T	H	
H	T	T	$= 3\,pq^2$
T	T	T	$= q^3$

एक सिक्के की उछाल में उसके चित्त गिरने (H) की सम्भावना $= \frac{1}{2}$

तीन सिक्कों की उछाल में उनके चित्त गिरने (H) की सम्भावना $= \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$

तीन सिक्कों की उछाल में उनके पट्ट गिरने (T) की सम्भावना $= \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$

तीन सिक्कों की उछाल में उसमें 2 के चित्त तथा 1 के पट्ट गिरने की सम्भावना $= \left(\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}\right)$

तीन सिक्कों को तीन-तीन बार उछालने में 2 के पट्ट तथा 1 के चित्त गिरने की सम्भावना भी $\frac{3}{8}$ है। यदि $p = \frac{1}{2}, q = \frac{1}{2}$:

$$(p + q)^3 = \left(\frac{1}{2} + \frac{1}{2}\right)^3 = p^3 + 3\,p^2q + 3\,pq^2 + q^3 = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8}$$

यह विस्तार बीजगणित के निम्न सूत्रों पर आधारित हैं–

$$(a + b)^2 = a^2 + b^2 + 2ab$$

$$(a + b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$$

इस विस्तार-सूत्र को $(p + q)^n$ के रूप में व्यक्त किया जा सकता है। यदि विभिन्न परिणामों की सम्भावित आवृत्तियाँ ज्ञात करना चाहते हैं तो $N\,(p + q)^n$ सूत्र की सहायता से उन्हें ज्ञात किया जा सकता है। इस सूत्र में N योग (number of trials) व n स्वतन्त्र घटनाओं की संख्या का प्रतिनिधित्व करता है। यदि आवृत्तियों का कुल योग 100 हो तथा स्वतन्त्र घटनाएँ दो हों तो सम्भावित आवृत्ति विवरण इस प्रकार होगा–

$N\,(p + q)^n$ or $\quad 100\,(p + q)^2$

$100\,(p^2 + 2pq + q^2)$ if $p = \dfrac{1}{2}, q = \dfrac{1}{2}$

Then $\quad = 100\left(\dfrac{1}{4} + \dfrac{1}{2} + \dfrac{1}{4}\right) = 100\left(\dfrac{1}{4}\right) + 100\left(\dfrac{1}{2}\right) + 100\left(\dfrac{1}{4}\right)$

$\quad = 25 + 50 + 25 = 100$

अन्य रूप में $= 25$ for two successes
50 for one success and one failure
25 for no successes

यदि $N = 100$, स्वतन्त्र घटनाओं की संख्या 3 हो तो द्विपद विस्तार इस प्रकार होगा–

$$N(p + q)^n = 100\left(\frac{1}{2} + \frac{1}{2}\right)^3$$

$$100(p^3 + 3p^2q + 3pq^2 + q^3)$$

$$= 100\left(\frac{1}{8}\right) + 100\left(\frac{3}{8}\right) + 100\left(\frac{3}{8}\right) + 100\left(\frac{1}{8}\right) = 12.5 + 37.5 + 37.5 + 12.5 = 100$$

द्विपद-वितरण 4, 5, 6, 7, 8.........n घटनाओं के लिए $(p + q)^n$ सूत्र के विस्तार से ज्ञात किया जा सकता है–

$$(p + q)^4 = p^4 + 4p^3q + 6p^2q^2 + 4pq^3 + q^4$$
$$(p + q)^5 = p^5 + 5p^4q + 10p^3q^2 + 10p^2q^3 + 5pq^4 + q^5$$
$$(p + q)^6 = p^6 + 6p^5q + 15p^4q^2 + 20p^3q^3 + 15p^2q^4 + 6pq^5 + q^6$$
$$(p + q)^7 = p^7 + 7p^6q + 21p^5q^2 + 35p^4q^3 + 35p^3q^4 + 21p^2q^5 + 7pq^6 + q^7$$

विस्तार का सूत्र इस प्रकार है–

$$(p+q)^n = p^n + np^{n-1}q^1 + \frac{n(n-1)}{1\times 2}p^{n-2}q^2 + \frac{n(n-1)(n-2)}{1\times 2\times 3}p^{n-3}q + \frac{n(n-1)(n-2)(n-3)}{1\times 2\times 3\times 4}p^{n-4}q^4 \ldots + q^n$$

यदि $n = 5$, तब $(p+q)^n = (p+q)^5$

$$(p+q)^5 = p^5 + 5p^{5-1}q^1 + \frac{5(5-1)}{2}p^{5-2}q^2 + \frac{5(5-1)(5-2)}{3\times 2}p^{5-3}q^3 + \frac{5(5-1)(5-2)(5-3)}{4\times 3\times 2}p^{5-4}q^4 + \frac{5(5-1)(5-2)(5-3)(5-4)}{5\times 4\times 3\times 2}p^{5-5}q^5$$

$$= p^5 + 5p^4q + 10p^3q^2 + 10p^2q^3 + 5pq^4 + q^5$$

संयोग (Combination) के नियमों के आधार पर भी सूत्र का विस्तार निम्न प्रकार से किया जा सकता है–

$$\begin{aligned}
(p+q)^5 &= {}^nC_5p^5q^0 + {}^nC_4p^4q^1 + {}^nC_3p^3q^2 + {}^nC_2p^2q^3 + {}^nC_1p^1q^4 + {}^nC_0p^0q^5 \\
&= {}^5C_5p^5q^0 + {}^5C_4p^4q^1 + {}^5C_3p^3q^2 + {}^5C_2p^2q^3 + {}^5C_1p^1q^4 + {}^5C_0p^0q^5 \\
&= 1p^5 + 5p^4q + 10p^3q^2 + 10p^2q^3 + 5pq^4 + 1q^5 \\
(q+p)^n &= {}^nC_0q^np^0 + {}^nC_1q^{n-1}p^1 + {}^nC_2q^{n-2}p^2 + {}^nC_3q^{n-3}p^3 + \ldots + {}^nC_nq^0p^n \\
&= q^n + {}^nC_1q^{n-1}p^1 + {}^nC_2q^{n-2}p^2 + {}^nC_3q^{n-3}p^3 + \ldots + p^n
\end{aligned}$$

यदि किसी घटना की $p = \frac{1}{2}, q = \frac{1}{2}$ तथा $n = 8$, तो सूत्र का विस्तार इस प्रकार होगा–

$$\left(\frac{1}{2}+\frac{1}{2}\right)^8 = \left(\frac{1}{2}\right)^8 + \frac{8}{1}\left(\frac{1}{2}\right)^7\left(\frac{1}{2}\right) + \frac{8(7)}{(1)(2)}\left(\frac{1}{2}\right)^6\left(\frac{1}{2}\right)^2 + \frac{8(7)(6)}{(1)(2)(3)}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^3 + \frac{8(7)(6)(5)}{(1)(2)(3)(4)}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^4 + \frac{8(7)(6)(5)(4)}{(1)(2)(3)(4)(5)}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^5 + \frac{8(7)(6)(5)(4)(3)}{(1)(2)(3)(4)(5)(6)}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^6 + \frac{8(7)(6)(5)(4)(3)(2)}{(1)(2)(3)(4)(5)(6)(7)}\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^7 + \left(\frac{1}{2}\right)^8$$

सरल करने पर–

$$\left(\frac{1}{2}+\frac{1}{2}\right)^8 = \frac{1}{256} + \frac{8}{256} + \frac{28}{256} + \frac{56}{256} + \frac{70}{256} + \frac{56}{256} + \frac{28}{256} + \frac{8}{256} + \frac{1}{256}$$

यदि 8 सिक्कों को एक साथ उछाला जाय तो विभिन्न परिणामों की प्रायिकता इस प्रकार होगी

H	T	p	p
8	0	$\frac{1}{256}$	.004
7	1	$\frac{8}{256}$	.031
6	2	$\frac{28}{256}$	.109

5	3	$\frac{56}{256}$	.219
4	4	$\frac{70}{256}$	.273
3	5	$\frac{56}{256}$	.219
2	6	$\frac{28}{256}$	.109
1	7	$\frac{8}{256}$	.031
0	1	$\frac{1}{256}$	.004
Total		$\frac{256}{256}$	1.000

(II) प्वॉयसन वितरण (Poisson Distribution)

अपने समय के प्रसिद्ध फ्रेंच गणितज्ञ साइमन डेनिस प्वॉयसन (Simon Denis Poission) (1781-1840) ने 1837 में एक आवृत्ति वितरण का प्रतिपादन किया जिसे प्वॉयसन वितरण की संज्ञा दी गयी है। यह एक महत्त्वपूर्ण प्रायिकता आवृत्ति-वितरण है। प्वॉयसन ने इस आवृत्ति-वितरण का विचार उस समय खोजा जबकि वे फौज के डिवीजनों में, जहाँ खच्चरों का उपयोग सिपाहियों व सामान ढोने के लिए किया जाता था, खच्चरों की लातों से मरने वाले व्यक्तियों का अध्ययन कर रहे थे। त्यॉयसन वितरण भी उन्हीं मान्यताओं पर आधारित है जिन पर द्विपद वितरण। इससे तात्पर्य यह है कि प्वॉयसन प्रयोग में सफलता तथा असफलता से व्यवहार किया जाता है, सफलताएँ एक-दूसरे से स्वतन्त्र होती हैं, और इस प्रकार सम्पूर्ण प्रयोग में सफलता की प्रायिकता समान रहती है। प्वॉयसन को द्विपद वितरण की सीमा के एक स्वरूप के रूप में देखा जा सकता है जबकि n अनन्त की ओर जा रहा हो ($n \to \infty$) तथा p शून्य की ओर जा रहा हो ($p \to 0$), इस प्रकार कि उनका गुणनफल एक निश्चित संख्या (m) हो अर्थात् वह स्थिर रहे। अन्य शब्दों में, त्यायसन वितरण का व्यवहार उस दशा में होता है जहाँ पर अनेक दैव-दशाएँ हों, जहाँ एक सफलता की प्रायिकता बहुत ही कम हो तथा प्रयोगों की संख्या काफी अधिक हो। इस प्रकार के प्रायिकता वितरण का व्यवहार बीमा के दावों, मशीनों का रुक जाना, प्रति पृष्ठ टंकण-त्रुटियाँ, ग्राहकों का आगमन आदि में विशेष रूप से किया जाता है। प्वॉयसन विधि से सृजित घटनाओं में एक निश्चित अन्तराल (समय, दूरी या स्थान का) होना चाहिए जिसे छोटे-छोटे अन्तरालों में बाँटा जा सकता हो, तथा प्रत्येक को एक प्रयोग माना जाए। उप-अन्तराल (sub-interval) काफी छोटे होने चाहिए जिससे एक उप-अन्तराल में एक से अधिक सफलता न घट सके। उप-अन्तरालों की संख्या ही घटनाओं की संख्या होगी। यह मान्यता होती है कि घटनाएँ स्वतन्त्र होंगी।

प्वॉयसन वितरण का स्वरूप (Form of Poission Distribution)

द्विपद वितरण की भाँति ही त्यॉयसन वितरण के चर-मूल्य (variate) खण्डित (discrete) होते हैं, अर्थात् वे पूर्णांक (integer) होते हैं। 0, 1, 2, 3·········सफलताओं की प्रायिकताएँ निम्न सूत्र के विस्तार से ज्ञात की जा सकती हैं–

$$p = e^{-m}\left[1 + \frac{m}{1!} + \frac{m^2}{2!} + \frac{m^3}{3!} + \frac{m^4}{4!} \cdots\cdots\cdots \frac{m^r}{r!}\right]$$

इसको इस प्रकार भी लिखा जा सकता है–

No. of successes :	0	1	2	3	4	......	r
Probabilities (p) :	e^{-m}	me^{-m}	m^2e^{-m}	m^3e^{-m}	m^4e^{-m}	......	m^re^{-m}

गणन-क्रिया

प्वॉयसन वितरण के लिए 0, 1, 2, 3.........सफलताओं की प्रायिकताएँ निम्न प्रकार ज्ञात की जाती हैं–

(i) समंकों का माध्य (m) निकाला जाता है।

(ii) e^{-m} का मान ज्ञात किया जाता है।

e (प्राकृतिक लघुगणक का आधार) का मान 2.7183 होता है।

$$e^{-m} = \frac{1}{e^m} = \frac{1}{(2.7183)^m} = \frac{1}{\text{Antilog (Log } 2.7183 \times m)}$$

$$= \frac{1}{\text{Antilog } (.4343 \times m)} = \text{Reciprocal of [Antilog } (.4343 \times m)]$$

$$\therefore \quad e^{-m} = \text{Reciprocal [Antilog } (.4343 \times m)]$$

(iii) निम्न सूत्र के प्रयोग से प्वॉयसन वितरण में 0, 1, 2, 3, 4.........सफलताओं की प्रायिकताएँ ज्ञात की जायेंगी–

$$P(r) = e^{-m}\frac{m^r}{r!}$$

प्वॉयसन वितरण की विशेषताएँ
(Characteristics of Poission Distribution)

प्वॉवसन वितरण में निम्नलिखित विशेषताएँ पायी जाती हैं–

(1) **खण्डित वितरण**–द्विपद वितरण की भाँति प्वॉयसन वितरण भी खण्डित वितरण (discrete distribution) होता है, अर्थात् उसका सम्बन्ध उन घटनाओं से होता है जिनका खण्डित दैव-चर से वितरण दिया जा सकता है। ये दैव-चर सफलताओं की संख्या 0, 1, 2, 3...से इंगित किये जाते हैं। यहाँ पर यह बताना उचित होगा कि प्वॉयसन दैव-चर 0, 1, 2,...से लेकर अनन्त संख्या तक हो सकते हैं जबकि द्विपद चरों की सीमा होती है–0, 1, 2,...n।

(2) **'p' तथा 'q' का मान**–इस वितरण का व्यवहार उन दशाओं में होता है जहाँ किसी घटना के घटने की प्रायिकता (p) अत्यन्त ही कम होती है तथा उस घटना के न घटने की प्रायिकता (q) काफी अधिक (1 के लगभग) होती है। इसमें 'n' भी अधिक होता है।

(3) **मुख्य मापांक**–प्वॉयसन वितरण के मुख्य मापांक (Parameter) माध्य ($m = np$) होता है। यदि m ज्ञात हो तो अन्य मापांकों के मूल्य ज्ञात किये जा सकते हैं।

(4) **अचरों के मूल्य**–प्वॉयसन वितरण में अचरों (Constants) के मूल्य इस प्रकार होते हैं–

$$\overline{X} \text{ or } m = np, \sigma = \sqrt{m} = \sqrt{np},$$

$$\mu_1 = 0,\ \mu_2 = m, \mu_3 = m, \mu_4 = m(3m + 1).$$

(5) **वितरण का स्वरूप**–प्वॉयसन वितरण विषम वितरण होता है। m का मान बढ़ने के साथ-साथ वितरण दायीं ओर बढ़ता जाता है तथा विषमता कम होती जाती है।

(6) **मान्यताएँ**–प्वॉयसन वितरण निम्नलिखित मान्यताओं पर आधारित होता है–(i) किसी घटना का घटना या न घटना अन्य के घटने या न घटने को प्रभावित नहीं करता है। (ii) एक छोटे समय, क्षेत्र या स्थान का अन्तराल समय, क्षेत्र या स्थान की लम्बाई के अनुपात में है। (iii) एक से अधिक घटना के लघु-अन्तराल में घटने की प्रायिकता नगण्य होती है।

(7) **उपयोगिता**–प्वॉयसन वितरण ऐसे दैव-चरों से सम्बन्धित प्रायिकताएँ निर्धारित करने में अधिक उत्तम निदर्श (model) रहता है जहाँ 'p' अत्यन्त ही कम हो तथा n का परिमाण अधिक हो, जैसे, टेलीफोन स्विचबोर्ड पर टेलीफोन कॉल आने की संख्या, निर्मित पार्ट में दोषों की संख्या, दुर्घटनाओं की संख्या, किसी सेवा-सुविधा पर आने वाले ग्राहकों की संख्या आदि।

3.4 सामान्य प्रायिकता वक्र (Normal Probability Curve)

ऊपर वर्णित द्विपद वितरण एक खण्डित वितरण (Discrete Distribution) है जिसका प्रयोग केवल उन्हीं परिस्थितियों में किया जा सकता है जब किसी घटना के केवल दो ही सम्भावित परिणाम सफलता व असफलता होते हैं परन्तु व्यवहार में अनेक ऐसी घटनाऐं होती हैं जिनके परिणाम सतत् (Continuous) रूप में प्राप्त होते हैं। जैसे किसी परीक्षण पर छात्रों के प्राप्तांक का अथवा व्यक्तियों के भार व ऊँचाई मापांक आदि का वितरण सतत होता है। ऐसी घटनाओं के सम्बन्ध में अनुमान लगाने के लिए सामान्य प्रायिकता वितरण निदर्श (Normal Probability Distribution Model) का प्रयोग किया जाता है। द्विपद वितरण में जब p व q का मान समान होता है तथा घातांक (Exponent) का मान अनन्त (Infinity) की ओर अग्रसर होता है तब सामान्य प्रायिकता वितरण प्राप्त होता है। यह एक अखण्डित, सतत व सममित वितरण है। आधुनिक सांख्यिकी में सामान्य प्रायिकता वितरण का केन्द्रीय स्थान है। सामान्य प्रायिकता वितरण भी एक सैद्धान्तिक वितरण है। n के अनन्त की ओर अग्रसित होने के कारण इस वितरण का रेखाचित्रीय रूप सतत वक्र जैसा होता है। इसलिए इसे प्राय: **सामान्य प्रायिकता वक्र** (Normal Probability Curve) अथवा एन. पी. सी. (N.P.C.) के नाम से सम्बोधित किया जाता है।

सामान्य प्रायिकता वक्र का इतिहास
(History of Normal Probability Curve)

सामान्य प्रायिकता वक्र (Normal Probability Curve) वास्तव में एक सैद्धान्तिक (Theoritical), आदर्श (Ideal) तथा गणितीय (Mathematical) वक्र है। इसे सामान्य वक्र (Normal Curve) या एन. पी. सी. (N.P.C.) के नाम से भी पुकारा जाता है। सामान्य प्रायिकता वक्र को ज्ञात करने का श्रेय लन्दन में रह रहे एक फ्रांसीसी शरणार्थी अब्राहम डी मोइवर (Abraham De Moivre) को जाता है जिन्होंने सन् 1733 में इस वक्र की गणितीय समीकरण प्रस्तुत की थी। परन्तु इस वक्र का व्यावहारिक उपयोग फ्रांसीसी गणितज्ञ पीयरे साइमन (Pierre Simon), जो मारकूस डी लाप्लास (Marquis De laplace) भी कहलाता था, तथा जर्मन खगोलविज्ञ कार्ल फ्रेडरिच गॉस (Carl Freidrich Gauss) ने उन्नीसवीं शताब्दी के प्रारम्भ में किया। गॉस ने पाया कि खगोलविदों के द्वारा की जाने वाली अंकन त्रुटियों (Recording Errors) का वितरण सामान्य प्रायिकता वक्र के अनुरूप होता है तथा सामान्य प्रायिकता वक्र को द्विपद वितरण (Binomial Distribution) में घातांक n को अनन्त (Infinity) तक बढ़ाकर प्राप्त किया जा सकता है। उसके इस कार्य को इतनी ख्याति मिली कि सामान्य प्रायिकता वक्र को **गॉसियन वक्र** (Gaussian Curve) अथवा **त्रुटियों का सामान्य वक्र** (Normal Curve of Errors) भी कहा जाने लगा। सामान्य प्रायिकता वक्र के व्यावहारिक उपयोग को बेल्जियम के एडोल्फी क्यूटलेट (Adolphe Quetlet) ने उन्नीसवीं शताब्दी के मध्य भाग में अत्यधिक विस्तार दिया। उसका विचार था कि सामान्य प्रायिकता वक्र का उपयोग मानवशास्त्र तथा मानवीय घटनाओं से सम्बन्धित समस्याओं के अध्ययन में माडल (Model) के रूप में किया जा सकता है। उसका विश्वास था कि मापे जाने पर विभिन्न मानसिक तथा नैतिक गुण (Mental and Moral Traits) सामान्य प्रायिकता वक्र के रूप में वितरित पाये जायेंगे। उसका यह विश्वास कालान्तर में सत्य सिद्ध हुआ। कालान्तर में जब विभिन्न मानसिक व नैतिक गुणों का मापन सम्भव हुआ तब उनका वितरण सामान्य प्रायिकता वक्र के अनुरूप ही पाया गया। उन्नीसवीं शताब्दी के अन्तिम वर्षों में सर फ्रान्सिस गाल्टन (Sir Fransis Galton) ने व्यक्तिगत विभिन्नताओं से सम्बन्धित अपने प्रयोगों में पाया कि अनेक मानसिक तथा शारीरिक विशेषताओं (Mental and Physical traits) के वितरण सामान्य प्रायिकता वक्र के समान होते हैं।

वे प्राकृतिक घटनाओं (Natural Phenomena) में सामान्य प्रायिकता वक्र के उपयोग की सम्भावना से भी बहुत प्रभावित हुए।

गणितीय दृष्टि से सामान्य प्रायिकता वक्र एक ऐसा द्विपदीय वक्र है जिसमें घातांक n का मान अनन्त (Infinity) तथा p व q के मान समान अर्थात् .5 होता है। क्योंकि यह वक्र प्रायिकता सिद्धातों पर आधारित है इसलिए इसे सामान्य प्रायिकता वक्र का नाम दिया गया। डी मोइवर के द्वारा प्रतिपादित होने के कारण इसे **डी मोइवर वक्र** (De Moivre Curve) भी कहा जाता है। आधुनिक सांख्यिकी में सामान्य प्रायिकता वक्र का एक अत्यन्त महत्त्वपूर्ण स्थान है। विभिन्न चरों तथा त्रुटियों का एक ही प्रकार के सैद्धान्तिक वक्र (सामान्य प्रायिकता वक्र) के रूप में वितरित होने का सांख्यिकीविदों ने लाभ उठाया एवं इस वक्र को माडल मान कर अनेक साँख्यिकीय विधियों का विकास कर लिया। वास्तव में सामान्य प्रायिकता वक्र निष्कर्षात्मक सांख्यिकी (Inferential Statistics) का प्राण है। यदि सामान्य प्रायिकता वक्र के प्रत्यय को निकाल दिया जाये, तो सम्पूर्ण निष्कर्षात्मक सांख्यिकी का ज्ञान कुछ पृष्ठों मात्र में सीमित होकर रह जायेगा।

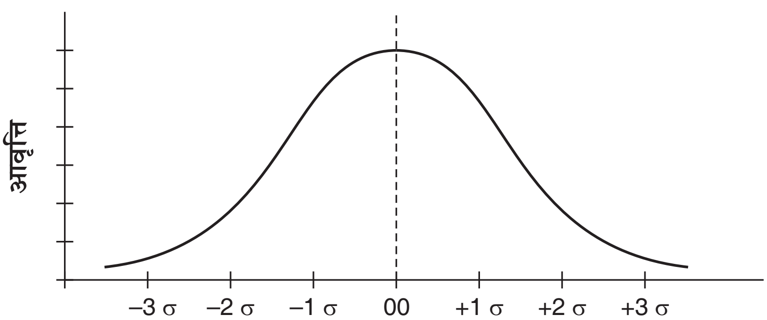
यद्यपि व्यवहार में सामान्य प्रायिकता वक्र की प्राप्ति संभव नहीं है क्योंकि कोई भी चर पूर्ण रूप से सामान्य प्रायिकता वक्र के रूप में वितरित नहीं होता है तथापि अवलोकित वक्रों की प्रवृत्ति सामान्य प्रायिकता वक्र के आकार को प्राप्त करने की होती है। जैसे-जैसे N बढ़ता जाता है वैसे-वैसे अवलोकित आवृत्ति वक्र का आकार सामान्य प्रायिकता वक्र के अनुरूप होता जाता है। अवलोकित वक्रों की इस प्रवृत्ति के कारण N के बड़ा होने पर व्यावहारिक समस्याओं के अध्ययन में विभिन्न चरों को सामान्य प्रायिकता वक्र के अनुरूप वितरित माना जा सकता है। परिणामत: सामान्य वक्र की विशेषताओं की सहायता से इस प्रकार की समस्याओं को हल किया जा सकता है। स्पष्ट है कि सामान्य प्रायिकता वक्र एक ऐसा सैद्धान्तिक, गणितीय तथा आदर्श वक्र है, जिसकी व्यवहार में पूर्ण प्राप्ति लगभग असम्भव ही है परन्तु जिसका व्यावहारिक उपयोग अत्यन्त अधिक है।

सामान्य प्रायिकता वक्र की विशेषताएँ

(Characteristics of Normal Probability Curve)

प्रयुक्त सांख्यिकी में सामान्य प्रायिकता वक्र (NPC) के उपयोग को ध्यान में रखकर यह उचित ही होगा कि इस वक्र का गहन अध्ययन किया जाये। अत: सामान्य प्रायिकता वक्र की कुछ प्रमुख विशेषताएँ अग्रांकित प्रस्तुत की गई हैं–

1. आकृति (Shape)–सामान्य प्रायिकता वक्र (NPC) वास्तव में **घंटाघर** (Bell Shaped), **पूर्ण सममित** (Symmetrical), **एकल बहुलांकी** (Unimodal), **सामान्य वक्रता वाला** (Mosokurtic) वक्र है। इस वक्र में दोनों सिरों पर आवृत्तियाँ कम होती है तथा केन्द्र की ओर अधिक आवृत्तियाँ होती हैं। अधिकतम आवृत्तियाँ केन्द्र में स्थित होती है जहाँ वक्र का शीर्ष बिन्दु स्थित होता है।



चित्र 2

2. गणितीय समीकरण (Mathematical Equation)–सामान्य प्रायिकता वक्र (NPC) की गणितीय समीकरण (Mathematical Equation) निम्नवत् होती है–

$$y = \frac{N}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}$$

जहाँ y = X बिन्दु पर स्थिति कोटि की ऊँचाई

x = मध्यमान से प्राप्तांक X की दूरी अर्थात् $x = X - M$

σ = प्राप्तांकों का मानक विचलन

N = कुल आवृत्ति

π = स्थिरांक, जिसका मान 3.1416 होता है।

e = स्थिरांक, जिसका मान 2.71828 होता है।

3. केन्द्रीय मानों की समानता (Equality of Measures of Central Tendencey)–सामान्य प्रायिकता वक्र (NPC) में केन्द्रीय प्रवृत्ति के तीनों मान अर्थात् मध्यमान, मध्यांक व बहुलाँक समान होते हैं तथा वक्र के मध्य बिन्दु पर स्थित होते हैं। अत: सामान्य प्रायिकता वक्र में

$$\mathbf{M = Md = M_0}$$

4. विषमता गुणांक (Coefficient of Skewness)–सामान्य प्रायिकता वक्र (NPC) कोटि अक्ष के सापेक्ष सममित (Symmetrical) होता है जिसके कारण इस का विषमता गुणांक (Coefficient of Skewness) का मान शून्य होता है अर्थात्

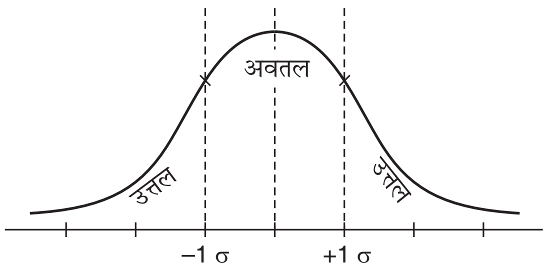
$$\mathbf{S_k = 0}$$

5. वक्रता गुणांक (Coefficient of Kurtosis)–सामान्य प्रायिकता वक्र (NPC) न तो बहुत चपटा और न ही बहुत नुकीला होता है। यह वक्र औसत ऊँचाई वाला वक्र होता है तथा इसके लिए वक्रता गुणांक (Coefficient of Kurtosis) का मान .263 होता है अर्थात्

Ku = .263

6. अनन्तस्पर्शी (Agymptotic)–सामान्य प्रायिकता वक्र (NPC) दोनों दिशाओं में अनन्त (Infinity) की ओर अग्रसर होता है। इसके दोनों सिरे आधार रेखा के निकट तो आते रहते हैं परन्तु वे आधार रेखा को कभी भी स्पर्श नहीं करते हैं। दूसरे शब्दों में कह सकते हैं कि सामान्य प्रायिकता वक्र आधार रेखा को अनन्त पर स्पर्श करता है।

7. वक्रता दिशा परिवर्तन बिन्दु (Point of Inflection)–वक्रता दिशा परिवर्तन बिन्दु वे बिन्दु होते हैं जहाँ वक्र अपनी वक्रता दिशा (Curvature) में परिवर्तन करता है। सामान्य प्रायिकता वक्र (NPC) मध्यमान से एक मानक विचलन ऊपर व नीचे अर्थात् $\pm 1\sigma$ पर अपनी दिशा परिवर्तित करता है। यह वक्र $+1\sigma$ से -1σ के बीच आधार रेखा की ओर अवतल (Concave) होता है जबकि $+1\sigma$ के ऊपर व -1σ के नीचे अर्थात् दोनों सिरों (Tails) पर आधार रेखा की ओर उत्तल (Convex) होता है।



चित्र 3

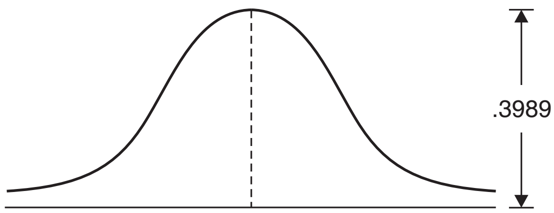
8. चतुर्थांकों में समान अन्तर (Equal Difference of Quartiles from Median)–सामान्य प्रायिकता वक्र (NPC) में प्रथम व तृतीय चतुर्थांकों का मध्यांक से अन्तर समान होता है। यह अन्तर चतुर्थांक विचलन के बराबर होता है जिसे संभाव्य त्रुटि (Probable Error–PE) भी कहते हैं। अर्थात्

$$Q_3 - Md = Md - Q_1 = Q \text{ या } PE$$

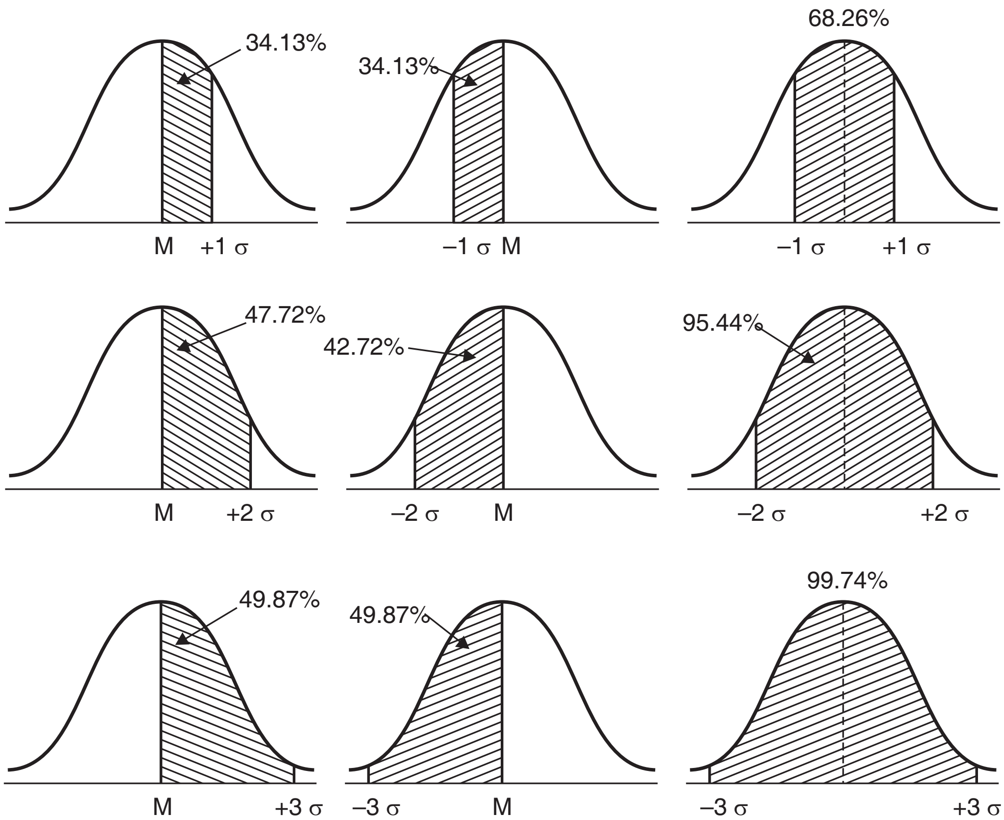
9. चतुर्थांक विचलन तथा मानक विचलन में सम्बन्ध (Relationship between Q.D. and S.D.)–सामान्य प्रायिकता वक्र (NPC) में चतुर्थांक विचलन का मान मानक विचलन के मान का लगभग 2/3 होता है। सूत्र के रूप में लिख सकते हैं कि–

$$Q = .6745\ \sigma \text{ तथा } \sigma = 1.482\ Q$$

10. सर्वोच्च कोटि (Highest Ordinate)–सामान्य प्रायिकता वक्र (NPC) में आधार रेखा के बिल्कुल मध्य में अर्थात् मध्यमान बिन्दु पर स्थित कोटि की ऊँचाई अधिकतम होती है तथा यह कुल आवृत्तियों अर्थात् N की .3989 होती है। इस कोटि को सर्वोच्च कोटि कहते हैं।



चित्र 4



चित्र 5. *सामान्य प्रायिकता वक्र में क्षेत्रफल* (***Area Under Normal Probability Curve***)

11. क्षेत्रफल सम्बन्ध (Area under N.P.C.)–सामान्य प्रायिकता वक्र (NPC) तथा आधार रेखा के बीच का क्षेत्रफल सामान्य प्रायिकता वक्र का क्षेत्रफल कहलाता है तथा यह कुल आवृत्तियों को प्रकट करता है। सामान्य प्रायिकता वक्र (NPC) की किन्हीं भी दो कोटियों के बीच का क्षेत्रफल उन कोटियों के सापेक्ष प्राप्तांकों के बीच अंक पाने वाले छात्रों की संख्या को प्रदर्शित करता है तथा यह कुल क्षेत्रफल का एक निश्चित प्रतिशत होता है। मध्यमान से एक मानक विचलन दूरी पर स्थित कोटि व मध्यमान पर स्थित कोटि के बीच कुल प्राप्तांकों के 34.13% प्राप्तांक होते हैं। मध्यमान से दो मानक विचलन दूरी पर स्थित कोटि व मध्यमान पर स्थित कोटि के बीच कुल प्राप्तांकों के 47.72% प्राप्तांक होते हैं मध्यमान से तीन मानक विचलन पर स्थित कोटि व मध्यमान पर स्थित कोटि के बीच कुल प्राप्तांकों के 49.87% प्राप्तांक होते हैं। मध्यमान के दोनों तरफ 50%-50% प्राप्तांक होते हैं।

स्पष्ट है कि सामान्य वक्र में–

(*i*) **मध्यमान से ± 1σ के बीच 68.26% प्राप्तांक होते हैं।**

(*ii*) **मध्यमान से ± 2σ के बीच 95.44% प्राप्तांक होते हैं।**

(*iii*) **मध्यमान से ± 3σ के बीच 99.74% प्राप्तांक होते हैं।**

क्योंकि ± 3σ के बीच 99.74% अर्थात् लगभग सभी प्राप्तांक आ जाते हैं। अत: व्यावहारिक समस्याओं के समाधान के लिए मान लिया जाता है। कि सामान्य प्रायिकता वक्र में ±3σ के बीच सभी (अर्थात् 100%) प्राप्तांक आ जाते हैं।

मानक सामान्य प्रायिकता वक्र
(Standard Normal Probability Curve)

सामान्य प्रायिकता वक्र (NPC) की समीकरण से स्पष्ट है कि सामान्य प्रायिकता वक्र के निर्धारण के लिए केवल समूह के आकार (N), मध्यमान (M) तथा मानक विचलन (σ) की आवश्यकता होती है, क्योंकि समीकरण में प्रयुक्त अन्य पद स्थिरांक (Constants) है। इन तीनों के ज्ञात होने पर सामान्य प्रायिकता वक्र की रचना की जा सकती है। विभिन्न वितरणों के लिए समूह का आकार, मध्यमान तथा मानक विचलन प्राय: भिन्न-भिन्न होते हैं, अत: उनके सापेक्ष सामान्य प्रायिकता वक्र भी भिन्न-भिन्न होंगे। यदि ऐसे वितरणों को सामान्य रूप से वितरित मानकर कुछ गणनाएँ करनी हों, तब इनके सापेक्ष भिन्न-भिन्न सामान्य प्रायिकता वक्रों का प्रयोग करना होगा। स्पष्ट है कि ऐसे वक्रों की विभिन्न कोटियों (Ordinates) के बीच के क्षेत्रफल या प्राप्तांकों की संख्या को ज्ञात करने के लिए भिन्न-भिन्न सारणियों की आवश्यकता होगी, जो न केवल एक कठिन बल्कि लगभग असम्भव सा कार्य होगा। सामान्य प्रायिकता वक्र की क्षेत्रफल सम्बन्धी विशेषताओं में स्पष्ट किया जा चुका है कि मध्यमान पर स्थित कोटि तथा मध्यमान से मानक विचलन की इकाइयों में दूर स्थित कोटियों के बीच का क्षेत्रफल कुल क्षेत्रफल का एक निश्चित प्रतिशत होता है। दूसरे शब्दों में कहा जा सकता है कि यदि प्राप्तांकों को मानक जेड प्राप्तांकों (Z-Scores) के रूप में व्यक्त करे तब विभिन्न कोटियों के बीच का क्षेत्रफल कुल क्षेत्रफल का एक निश्चित प्रतिशत अथवा अनुपात होता है। जेड मानक प्राप्तांकों के रूप में प्रस्तुत सामान्य प्रायिकता वक्र को मानक सामान्य प्रायिकता वक्र (Standard Normal Probability Curve) कहते हैं। अत: मानक सामान्य प्रायिकता वक्र एक ऐसा सामान्य प्रायिकता वक्र है जिसके लिए मध्यमान का मान शून्य के बराबर तथा मानक विचलन का मान एक के बराबर होता है। तब मानक सामान्य प्रायिकता वक्र की समीकरण निम्न हो जायेगी–

$$Y = \frac{N}{\sqrt{2\pi}} e^{-z^2/2}$$

जब N = 1 होता है तब मानक सामान्य प्रायिकता वक्र को इकाई सामान्य प्रायिकता वक्र (Unit Normal Probability Curve) कहा जाता है। स्पष्ट है कि इकाई सामान्य प्रायिकता वक्र का कुल क्षेत्रफल एक इकाई के बराबर होता है। इकाई सामान्य प्रायिकता वक्र की समीकरण निम्न होती है–

$$y = \frac{\mathbf{1}}{\sqrt{\mathbf{2\pi}}} \mathbf{e}^{-\mathbf{z^2/2}}$$

क्योंकि इकाई सामान्य प्रायिकता वक्र, मध्यमान, मानक विचलन या समूह के आकार पर आधारित नहीं होता है इसलिए इसके सापेक्ष एक ऐसी सारणी बनाई जा सकती है जिसका प्रयोग विभिन्न व्यावहारिक समस्याओं के समाधान में किया जा सके। सांख्यिकी की पुस्तकों में दी गई सामान्य प्रायिकता वक्र की क्षेत्रफल सम्बन्धी सारणियाँ प्राय: इकाई वक्र के लिए होती है जो किन्हीं दो कोटियों के बीच के क्षेत्रफल या आवृत्तियों को अनुपात (Proportion) के रूप में व्यक्त करती है। मानक सामान्य प्रायिकता वक्र की सारणी की सहायता से किसी भी सामान्य प्रायिकता वितरण की किन्हीं भी दो कोटियों के बीच के क्षेत्रफल/आवृत्तियों को ज्ञात किया जा सकता है। इसके लिए दिये गये मध्यमान तथा मानक विचलन वाले वितरण को ऐसे वितरण में परिवर्तित कर लेते हैं जिसका मध्यमान शून्य तथा मानक विचलन एक हो, दूसरे शब्दों में कह सकते हैं कि मानक सामान्य प्रायिकता वक्र सारणी मानक प्राप्तांकों के लिए ही प्रयुक्त की जा सकती है। अत: मानक सामान्य प्रायिकता वक्र की सारणी का उपयोग करने के लिए पहले विभिन्न प्राप्तांकों को मानक प्राप्तांकों (Standard Scores), जिन्हें **जेड प्राप्तांक** (Z-Scores) अथवा **सिगमा मान** (σ Values) भी कहते हैं, में बदल लेते हैं। तत्पश्चात इन जैड प्राप्तांकों या सिगमा मानों के सापेक्ष विभिन्न कोटियों के बीच का क्षेत्रफल सारणी से ज्ञात कर लेते हैं। क्योंकि मध्यमान सामान्य प्रायिकता वक्र की आधार रेखा के मध्य बिन्दु पर स्थित होता है तथा मध्यमान के लिए जैड प्राप्तांक या सिगमा मान सदैव शून्य होता है इसलिए कहा जा सकता है कि मानक सामान्य प्रायिकता वक्र में आधे जैड प्राप्तांक (सिगमा मान) शून्य से कम होते हैं तथा आधे जैड प्राप्तांक (सिगमा मान) शून्य से अधिक होते हैं। यदि जैड प्राप्तांक या सिगमा मान धनात्मक होता है तो यह मध्यमान बिन्दु या शून्य जैड प्राप्तांक या शून्य सिगमा मान के दाँयी तरफ (Right Side) में स्थित होता है जबकि ऋणात्मक जैड प्राप्तांक या सिगमा मान बाँयी तरफ (Left Side) में स्थित रहता है। क्योंकि किसी मूल प्राप्तांक को जैड प्राप्तांक में बदलने के लिए उसमें से मध्यमान को घटाकर मानक विचलन से भाग देते हैं ($Z = (X - M)/\sigma$), इसलिए प्राप्तांक के मध्यमान से बड़ा होने पर धनात्मक जैड प्राप्तांक या सिगमा मान प्राप्त होता है जबकि प्राप्तांक के मध्यमान से छोटा होने पर ऋणात्मक जैड प्राप्तांक या सिगमा मान प्राप्त होता है। किसी जैड प्राप्तांक या सिगमा मान को मध्यमान व मानक विचलन के ज्ञात होने पर मूल प्राप्तांक में समीकरण ($X = M + Z\sigma$) की सहायता से बदला जा सकता है।

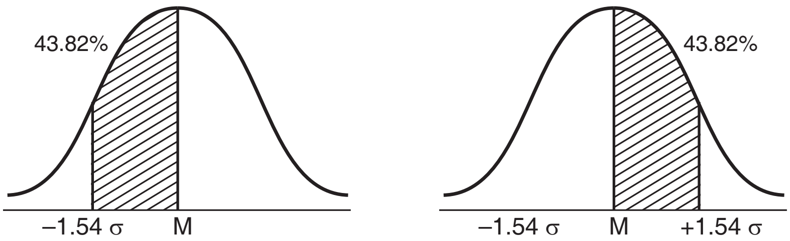
सामान्य प्रायिकता वक्र सारणी

(Table of NPC)

परिशिष्ट–1 में दी गई सामान्य प्रायिकता वक्र की क्षेत्रफल सम्बन्धी सारणी के प्रथम स्तम्भ (Column) में Z प्राप्तांक या σ मान के विभिन्न मान दशमलव के एक स्थान तक लिखे गये हैं तथा प्रथम पंक्ति (Row) में इसके लिए दशमलव का दूसरा स्थान लिखा गया है। प्रथम स्तम्भ में .0 से 5.0 तक के अंक तथा प्रथम पंक्ति में .00 से .09 तक के दस अंक लिखे हुए हैं। सारणी के शेष भाग में मध्यमान कोटि व अन्य विभिन्न Z प्राप्तांकों (या σ मानों) पर स्थित कोटियों के बीच व्यक्तियों के प्राप्तांकों का अनुपात दिया हुआ है। यहाँ यह बात ध्यान देने की है कि कुछ सारणियों में प्रतिशत दिया होता है। अत: किसी प्रकार की शंका होने पर सारणी के साथ की गई टिप्पणी का अवलोकन करना चाहिये। इस सारणी की सहायता से मध्यमान पर स्थित कोटि तथा किसी दिये गये Z प्राप्तांक (या σ मान) पर स्थित कोटि के बीच आने वाले प्राप्तांकों का अनुपात सरलता से ज्ञात किया जा सकता है। जैसे यदि किसी प्राप्तांक का σ मान 1.54 है तथा मध्यमान व 1.54 σ पर स्थित कोटियों के बीच प्राप्तांकों का अनुपात/प्रतिशत ज्ञात करना हो तो प्रथम स्तम्भ में 1.5 तथा प्रथम पंक्ति में .04 देखेंगे। तब 1.5 वाली पंक्ति व .04 वाले स्तम्भ के कटाव पर लिखी संख्या पढ़ लेंगे। सारणी से स्पष्ट है कि यह संख्या .4382 है। इसका अर्थ है कि मध्यमान कोटि व 1.54 σ मान वाली कोटि के बीच प्राप्तांकों का अनुपात .4382% प्राप्तांक है। दूसरे शब्दों में मध्यमान व 1.54 σ मान के सापेक्ष प्राप्तांक के बीच अंक प्राप्त करने वाले छात्रों की संख्या 43.82 प्रतिशत है। क्योंकि सामान्य वक्र एक सममित वक्र है इसलिए मध्यमान से दांयी ओर व बांयी ओर बराबर दूरी पर स्थित कोटि अक्षों तथा मध्यमान पर स्थित कोटि के बीच प्राप्तांकों की संख्या बराबर होती है। अत: यदि σ मान ऋणात्मक होगा तो सामान्य प्रायिकता वक्र में क्षेत्रफल मध्यमान के बांयी ओर होगा, तथा यदि σ मान धनात्मक होगा तो क्षेत्रफल मध्यमान के दांयी ओर होगा, किन्तु वे दोनों क्षेत्रफल बराबर होंगे।

इसके अतिरिक्त परिशिष्ट-2 में विभिन्न Z मानों (σ मानों) के सापेक्ष सामान्य प्रायिकता वक्र में कोटियों की ऊँचाई एवं परिशिष्ट-3 में विभिन्न क्षेत्रफल विभाजन के सापेक्ष कोटियों की ऊँचाई तथा प्रमापीकृत प्राप्तांकों (Z) के मान दिये गये हैं।

नोट



चित्र 6

3.5 सामान्य प्रायिकता वक्र के प्रत्यय का उपयोग (Use of the Concept of Normal Probability Curve)

जैसा कि स्पष्ट किया जा चुका है कि शिक्षा, मनोविज्ञान, समाज शास्त्र आदि के अधिकांश चर सामान्य प्रायिकता वक्र के रूप में वितरित होते हैं। मापन व मूल्यांकन के क्षेत्र में सामान्य प्रायिकता वक्र के प्रत्यय का अत्यन्त महत्त्व है। अध्यापकगण सामान्य प्रायिकता वक्र का उपयोग अपनी व्यावहारिक समस्याओं के समाधान के लिए कर सकते हैं। यदि किसी चर पर प्राप्तांकों के वितरण को सामान्य प्रायिकता वितरण के रूप में स्वीकार किया जा सके तो सामान्य प्रायिकता वक्र की सहायता से निम्न प्रकार की समस्याओं का समाधान किया जा सकता है :

1. किसी समूह में किसी दिये गये प्राप्तांक से अधिक या कम अंक पाने वाले छात्रों की संख्या ज्ञात करना।
2. किसी समूह में दिये गये किन्हीं दो प्राप्तांकों के बीच अंक प्राप्त करने वाले छात्रों की संख्या ज्ञात करना।
3. किसी समूह में किसी विशेष स्थिति वाले छात्रों के लिए प्राप्तांक सीमाएँ ज्ञात करना।
4. किसी परीक्षण के विभिन्न प्रश्नों का सापेक्षिक कठिनाई स्तर ज्ञात करना।
5. किसी समूह को इस प्रकार से कुछ उपसमूहों (Subgroups) में विभाजित करना कि प्रत्येक उपसमूह में योग्यता का प्रसार (Range of ability) बराबर रहें।

इन सभी प्रकार की समस्याओं के समाधान में सामान्य प्रायिकता वक्र के उपयोग को कुछ उदाहरणों की सहायता से स्पष्ट किया गया है। इसके अतिरिक्त प्रतिदर्शजों के प्रतिचयन वितरणों के सामान्य प्रायिकता वक्र के अनुरूप होने का लाभ उठाकर सांख्यिकीयविद सार्थकता परीक्षणों में भी सामान्य प्रायिकता वक्र के प्रत्यय का उपयोग करते हैं। इसकी चर्चा आगे के अध्यायों में की गई है।

1. किसी प्राप्तांक से अधिक या कम अंक प्राप्त करने वाले व्यक्तियों की संख्या ज्ञात करना (Determination of the Number of Individuals getting Scores above or below the given Score)

कभी-कभी किसी समूह के लिए किसी चर पर मध्यमान व मानक विचलन ज्ञात होता है तथा किसी प्राप्तांक से अधिक या कम अंक प्राप्त करने वाले छात्रों की संख्या ज्ञात करनी होती है। यदि उस पर चर प्राप्तांकों के वितरण को सामान्य वितरण स्वीकार किया जा सके तो सामान्य प्रायिकता वक्र सारणी की सहायता से इस प्रकार की समस्याओं को हल कर सकते हैं। इसके लिए सबसे पहले प्राप्तांक की, आवश्यकतानुसार, उच्च या निम्न सीमा को σ मान में परिवर्तित कर लेते हैं तथा इस मान पर स्थित कोटि व मध्यमान कोटि के बीच के प्राप्तांकों का प्रतिशत ज्ञात कर लेते हैं। तत्पश्चात σ मान नीचे या ऊपर के प्राप्तांकों की प्रतिशत संख्या ज्ञात कर लेते हैं।

क्योंकि सामान्य प्रायिकता वक्र वस्तुत: किसी सतत् चर (Continuous Variable) का वक्र है, इसलिए सामान्य प्रायिकता वक्र की सहायता से समस्या हल करते समय विभिन्न प्राप्तांकों की उच्च या निम्न सीमा पर (जो भी उपयुक्त हो) स्थित कोटि के दाएं या बाएं ओर स्थित छात्रों की प्रतिशत संख्या ज्ञात करते हैं न कि उस प्राप्तांक पर स्थित कोटि के दाएं या बाएं ओर स्थित छात्रों की संख्या। जैसे यदि 29 से कम अंक प्राप्त करने वाले छात्रों की संख्या ज्ञात करनी हों तो सामान्य प्रायिकता वक्र में 28.5 के सापेक्ष σ मूल्य पर स्थित कोटि के बाँयी ओर स्थित छात्रों को देखेंगे। सतत चर होने के कारण वास्तव में उन सभी छात्रों को 29 अंक दिये जाते हैं जिनसे प्राप्तांक 28.5 से लेकर 29.5 तक होते हैं। अत: जब 29 से कम अंक वाले छात्रों की चर्चा करते हैं तो इसका अर्थ है कि 28.5 से कम वाले छात्र। इसी प्रकार से जब 29 से अधिक अंक वाले छात्रों की चर्चा की जाती है तो इसका अर्थ है कि 29.5 से अधिक वाले छात्र। अत: जब किसी प्राप्तांक से कम अंक प्राप्त करने वाले छात्रों की संख्या ज्ञात करनी होती है तो उस प्राप्तांक की निम्न सीमा पर विचार करते हैं तथा जब किसी प्राप्तांक से अधिक अंक प्राप्त करने वाले छात्रों की संख्या ज्ञात करनी होती है तो उस प्राप्तांक की उच्च सीमा पर विचार करते हैं।

उदाहरण—200 छात्रों के एक समूह को हिन्दी परीक्षण दिया गया, जिस पर मध्यमान 50 तथा मानक विचलन 12 था। हिन्दी परीक्षण पर प्राप्तांकों के वितरण को सामान्य रूप से वितरित मानकर ज्ञात करिये कि

(*i*) कितने छात्रों ने 60 से अधिक अंक प्राप्त किये होंगे?

(*ii*) कितने छात्रों ने 35 से अधिक अंक प्राप्त किये होंगे?

(*iii*) कितने छात्रों ने 55 से कम अंक प्राप्त किये होंगे?

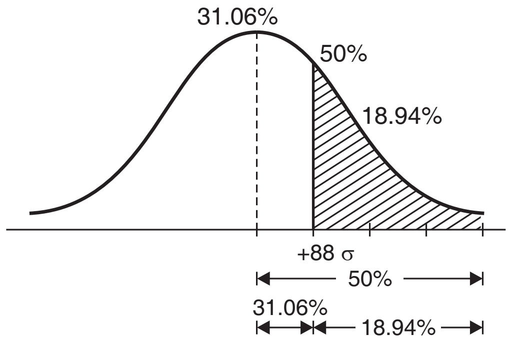
(*iv*) कितने छात्रों ने 30 से कम अंक प्राप्त किये होंगे?

हल—स्पष्ट है यहाँ पर N = 200, M = 50 σ = 12

(*i*) क्योंकि 60 से अधिक अंक प्राप्त करने वाले छात्रों की संख्या ज्ञात करनी है, इसलिए सामान्य प्रायिकता वक्र में 60 की उच्च सीमा अर्थात् 60.5 के σ मान पर स्थित कोटि के दांयी ओर का क्षेत्रफल उन छात्रों की संख्या को व्यक्त करेगा जिनके प्राप्तांक 60 से अधिक हैं। 60.5 को σ के मान में परिवर्तित करने पर

$$\textbf{X का } \sigma \textbf{ मान} = \frac{\mathbf{X - M}}{\sigma}$$

$$\textbf{60.5 का } \sigma \textbf{ मान} = \frac{\mathbf{60.5 - 50}}{\mathbf{12}} = .88\ \sigma$$



चित्र 7

अत: सामान्य प्रायिकता वक्र में, +.88 σ पर स्थित कोटि के दांयी ओर स्थित छात्र ही वांछित छात्र होंगे, जिन्हें आड़ी रेखाओं से प्रदर्शित किया गया है। सामान्य प्रायिकता वक्र सारणी से स्पष्ट है कि

मध्यमान या 0 σ के दांयी ओर छात्र = 50%

मध्यमान या 0 σ से .88 σ तक छात्र = 31.06%

अत: .84 σ के दाईं ओर कुल छात्र 50 – 31.06 = 18.94%

अत: कहा जा सकता है कि 18.94% छात्रों ने 60 से अधिक अंक प्राप्त किये होंगे।

क्योंकि 100 छात्रों में से 18.94 छात्र 60 से अधिक अंक प्राप्त करते हैं।

अत: 1 छात्रों में से $\frac{18.94}{100}$ छात्र 60 से अधिक अंक प्राप्त करेंगे।

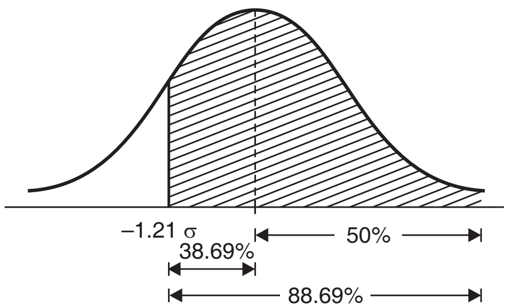
अत: 200 छात्रों में से $\frac{18.94}{100} \times 200 = 37.88$ छात्र 60 से अधिक अंक प्राप्त करेंगे।

क्योंकि छात्र संख्या पूर्णांक में ही हो सकती है इसलिए कहेंगे कि 200 छात्रों में से 38 छात्रों ने 60 से अधिक अंक प्राप्त किये होंगे।

(*ii*) क्योंकि 35 से अधिक अंक प्राप्त करने वाले छात्रों की संख्या ज्ञात करनी है। इसलिए सामान्य प्रायिकता वक्र में 35 की उच्च सीमा अर्थात् 35.5 के σ मान पर स्थित कोटि के दाईं ओर का क्षेत्रफल इन छात्रों की संख्या को प्रदर्शित करेगा जिनके प्राप्तांक 35 से अधिक हैं। 35.5 को σ मान में परिवर्तित करने पर

$$\textbf{35.5 का } \sigma \textbf{ मान} = \frac{35.5 - 50}{12} - 1.21\,\sigma$$

अत: सामान्य प्रायिकता वक्र में – 1.21 σ पर स्थित कोटि के दाएं ओर का क्षेत्रफल वांछित छात्रों को प्रदर्शित करेगा। स्पष्ट है कि चित्र में –1.21 σ मध्यमान के बाईं ओर स्थित छात्रों को तिरछी रेखाओं से प्रदर्शित किया गया है। सामान्य प्रायिकता वक्र सारणी से स्पष्ट है कि–



चित्र 8

मध्यमान या 0 σ से दाईं ओर छात्र = 50%

– 1.21 σ से 0 σ तक छात्र = 38.69%

– 1.21 σ से दाईं ओर कुल छात्र 50 + 38.69 = 88.69%

अत: 88.69% छात्रों ने 35 से अधिक अंक प्राप्त किये होंगे।

क्योंकि 100 छात्रों में से 88.69 छात्र 35 से अधिक अंक प्राप्त करते हैं।

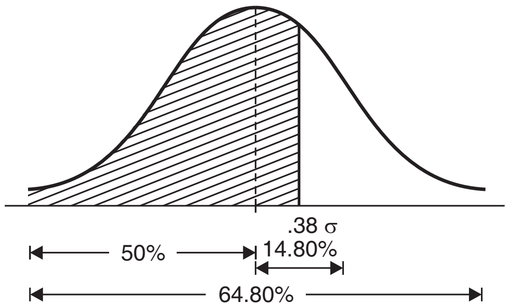
अत: 1 छात्र में से $\frac{88.69}{100}$ छात्र 35 से अधिक अंक प्राप्त करेंगे।

200 छात्रों में से $\frac{88.69}{100} \times 200 = 177.38$ छात्र 35 से अधिक अंक प्राप्त करेंगे।

अत: कहा जा सकता है कि 200 छात्रों में से 177 छात्रों के अंक 35 से अधिक होंगे।

(*iii*) क्योंकि 55 से कम अंक प्राप्त करने वाले छात्रों की संख्या ज्ञात करनी है इसलिए सामान्य प्रायिकता वक्र में 55 की निम्न सीमा अर्थात् 54.5 के σ मान पर स्थित कोटि के बाईं ओर का क्षेत्रफल इन छात्रों की संख्या को बतायेगा जो 55 से कम अंक प्राप्त करते हैं। 54.5 को σ मान में बदलने पर

$$54.5 \text{ का } \sigma \text{ मान} = \frac{54.5 - 50}{12} = .38\,\sigma$$



चित्र 9

अत: सामान्य प्रायिकता वक्र में .38 पर स्थित कोटि के बांयी ओर के छात्र ही वांछित छात्र होंगे जिन्हें तिरछी रेखा से व्यक्त किया गया है। सामान्य प्रायिकता वक्र सारणी से स्पष्ट है कि

मध्यमान या 0 σ से बाईं ओर छात्र = 50%

मध्यमान या 0 σ से .38 तक छात्र = 14.80%

अत: .37 σ से बांयी ओर कुल छात्र 50 + 14.80 = 64.80%

अत: 64.80% छात्रों ने 55 से कम अंक प्राप्त किये होंगे।

क्योंकि 100 छात्रों में से 64.80 छात्रों के प्राप्तांक 55 से कम है।

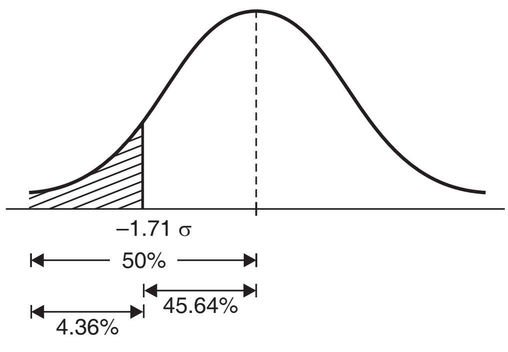
अत: 1 छात्र में से $\frac{64.80}{100}$ छात्रों के प्राप्तांक 55 से कम होंगे।

इसलिए 200 छात्र में से $\frac{64.80}{100} \times 200 = 129.60$ छात्रों के प्राप्तांक 55 से कम होंगे।

अत: कहा जा सकता है कि 200 छात्रों में से 129 छात्रों के प्राप्तांक 55 से कम होंगे।

(*iv*) क्योंकि 30 से कम अंक प्राप्त करने वाले छात्रों की संख्या ज्ञात करनी है, इसलिए सामान्य प्रायिकता वक्र में 30 की निम्न सीमा अर्थात् 29.5 के σ मान पर स्थित कोटि के बाईं ओर का क्षेत्रफल उन छात्रों की संख्या बतायेगा जो 30 से कम अंक प्राप्त करते हैं। प्राप्तांक 29.5 को σ मान में बदलने पर

$$29.5 \text{ का } \sigma \text{ मान} = \frac{29.5 - 50}{12} = -1.71\,\sigma$$



चित्र 10

अत: सामान्य प्रायिकता वक्र में $-1.71\,\sigma$ पर स्थित कोटि के बाईं ओर के छात्र ही वांछित छात्र होंगे, जिन्हें तिरछी रेखा से प्रकट किया गया है। सामान्य प्रायिकता वक्र सारणी से

मध्यमान या $0\,\sigma$ के बाईं ओर छात्र = 50%

मध्यमान या $0\,\sigma$ से $-1.71\,\sigma$ तक छात्र = 45.64%

अत: $-1.71\,\sigma$ के बाईं ओर कुल छात्र = 50 − 45.64 = 4.36%

अत: 4.36% छात्रों के प्राप्तांक 30 के कम होंगे।

क्योंकि 100 छात्रों में से 4.36 छात्रों के प्राप्तांक 30 से कम हैं।

अत: 1 छात्र में से $\frac{4.36}{100}$ छात्रों के प्राप्तांक 30 से कम होंगे।

इसलिए 200 छात्रों में से $\frac{4.36}{100} \times 200 = 8.72$ छात्रों के प्राप्तांक 30 से कम होंगे।

अत: कहा जा सकता है कि 200 छात्रों में से 9 छात्रों के प्राप्तांक 30 से कम होंगे।

2. दो प्राप्तांकों के बीच अंक प्राप्त करने वाले व्यक्तियों की संख्या ज्ञात करना

(Determination of the Number of Individuals getting Scores within given limits of Scores)

किसी समूह का मध्यमान व मानक विचलन ज्ञात होने पर किन्हीं भी दो प्राप्तांकों के बीच अंक प्राप्त करने वाले छात्रों की प्रतिशत संख्या को सामान्य प्रायिकता वक्र की सहायता से ज्ञात किया जा सकता है। इसके लिए छोटे प्राप्तांक की निम्न सीमा व बड़े प्राप्तांक की उच्च सीमा को σ मानों में परिवर्तित करके इन दोनों σ मानों पर स्थित कोटियों के बीच का क्षेत्रफल ज्ञात कर लेते हैं। यह क्षेत्रफल ही इन दोनों प्राप्तांकों के बीच अंक प्राप्त करने वाले छात्रों की प्रतिशत संख्या को व्यक्त करेगा। छोटे प्राप्तांकों की निम्न सीमा व बड़े प्राप्तांक की उच्च सीमा इसलिए लेते हैं क्योंकि इन दोनों प्राप्तांकों के बराबर (वास्तव में 0.5 से कम या अधिक) प्राप्तांक वाले छात्र भी दोनों प्राप्तांकों के बीच अंक प्राप्त करने वाले छात्रों के समूह में आयेंगे। यहाँ यह इंगित करना उचित ही होगा कि गैरट व गिलफोर्ड आदि अनेक विद्वानों ने दो प्राप्तांकों के बीच के छात्रों की संख्या ज्ञात करते समय प्राप्तांक को ही सिगमा नाम में परिवर्तित करने का सुझाव दिया है परन्तु लेखक इस बात से सहमत नहीं है। लेखक का तर्क है कि यदि किसी समूह के छात्रों को 40 से नीचे, 40 व 50 के बीच तथा 50 से अधिक अंक पाने वाले छात्रों को तीन वर्गों में बाँटते समय यदि सामान्य प्रायिकता वक्र में ये उपसमूह 39.5 से नीचे, 40 व 50 के बीच तथा 50.5 से ऊपर के बनाये जायेंगे तब 39.5 से 40 के बीच तथा 50 से 50.5 के बीच वाले छात्र कहाँ जायेंगे।

उदाहरण—एक बुद्धि परीक्षण 400 छात्रों के समूह को दिया गया जिसका मध्यमान 100 तथा मानक विचलन 15 पाया गया। छात्रों के बुद्धिलब्धांक को सामान्य रूप से वितरित मानकर ज्ञात करो कि

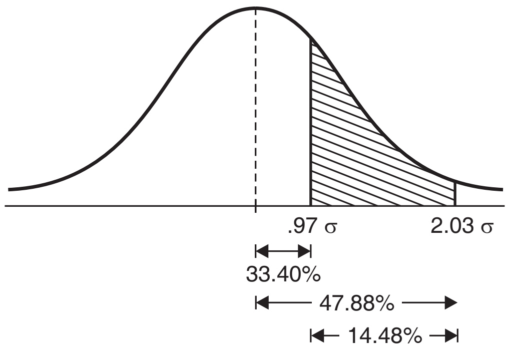
(*i*) कितने छात्र के बुद्धिलब्धांक 115 व 130 के बीच होंगे?

(*ii*) कितने प्रतिशत छात्रों के बुद्धिलब्धांक 80 व 112 के बीच होंगे?

हल—स्पष्ट है कि यहाँ पर $N = 400$ $M = 100$ $\sigma = 15$

(*i*) क्योंकि 115 व 130 के बीच बुद्धिलब्धांक वाले छात्रों की संख्या ज्ञात करनी है इसलिए सामान्य प्रायिकता वक्र में 115 की निम्न सीमा अर्थात् 114.5 तथा 130 की उच्च सीमा अर्थात् 130.5 पर स्थित कोटियों के बीच का क्षेत्रफल इन छात्रों को प्रदर्शित करेगा जिनके प्राप्तांक 115 व 130 के बीच होंगे। अत: 114.5 व 130.5 को σ मूल्यों में बदलने पर

नोट



चित्र 11

$$\textbf{114.5 का } \sigma \textbf{ मान} = \frac{114.5 - 100}{15} = .97\,\sigma$$

$$\textbf{130.5 का } \sigma \textbf{ मान} = \frac{130.5 - 100}{15} = 2.03\,\sigma$$

अत: सामान्य प्रायिकता वक्र में .96 σ व 2.03 σ पर स्थित कोटियों के बीच का क्षेत्रफल वांछित छात्रों को प्रदर्शित करेगा, जिन्हें तिरछी रेखाओं से व्यक्त किया गया है। सामान्य वक्र सारणी से स्पष्ट है कि

0 σ से 2.03 σ तक छात्र = 47.88%

0 σ से .97 तक छात्र = 33.40%

अत: .97 σ से 2.03 σ तक छात्र = 47.88 – 33.40 = 14.48%

अत: 14.48% छात्रों से बुद्धिलब्धिांक 115 व 130 के बीच होंगे।

क्योंकि 100 छात्रों में से 14.48 छात्रों के प्राप्तांक 115 व 130 के बीच हैं।

अत: 1 छात्रों में से $\frac{14.48}{100}$ छात्रों के प्राप्तांक 115 व 130 के बीच होंगे।

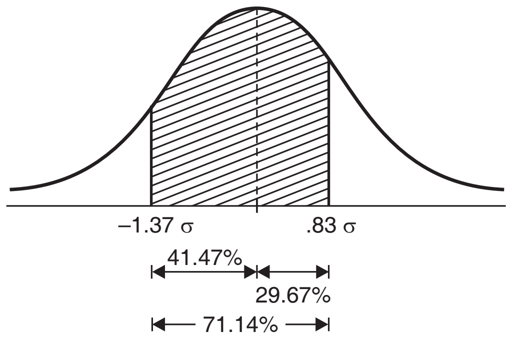
इसलिए 400 छात्रों में से $\frac{14.48}{100} \times 400 = 57.92$ छात्रों के प्राप्तांक 115 व 130 के बीच होंगे।

अत: सामान्य प्रायिकता वक्र के आधार पर कहा जा सकता है कि 400 में से केवल 58 छात्रों की बुद्धिलब्धि 115 व 130 के बीच होंगी।

(*ii*) क्योंकि 80 व 112 के बीच अंक प्राप्त करने वाले छात्रों की प्रतिशत संख्या ज्ञात करनी है, इसलिए सामान्य प्रायिकता वक्र में 80 की निम्न सीमा अर्थात् 79.5 तथा 112 की उच्च सीमा अर्थात् 112.5 के σ मानों पर स्थित कोटियों के बीच का क्षेत्रफल उन छात्रों को व्यक्त करेगा जिनके प्राप्तांक 80 व 112 के बीच हैं। अत: 79.5 व 112 σ को σ मानों में बदलने पर।

$$\textbf{79.5 का } \sigma \textbf{ मान} = \frac{79.5 - 100}{15} = -\,1.37\,\sigma$$

$$\textbf{112.5 का } \sigma \textbf{ मान} = \frac{112.5 - 100}{15} = +\,.83\,\sigma$$



चित्र 12

अत: सामान्य प्रायिकता वक्र में $-1.37\,\sigma$ व $.83\,\sigma$ पर स्थित कोटियों के बीच का क्षेत्रफल वांछित छात्रों को प्रदर्शित करेगा, जिन्हें तिरछी रेखाओं से व्यक्त किया गया है। अत: सामान्य वक्र सारणी देखने से

$0\,\sigma$ से -1.37 तक छात्र $= 41.47\,\sigma\,\%$

0 से $+.83\,\sigma$ तक छात्र $= 29.67\%$

अत: सामान्य प्रायिकता वक्र के प्रत्यय के आधार पर कहा जा सकता है कि $-1.37\,\sigma$ से $+.83\,\sigma$ तक कुल छात्र $= 71.14\%$

अत: 71.14 छात्रों के प्राप्तांक 80 व 112 के बीच होंगे।

3. दिये गये प्रतिशत या संख्या वाले व्यक्तियों के प्राप्तांकों की उच्च व निम्न सीमाएँ ज्ञात करना

(Determination of the Limits of Scores which include a given Percentage or Number of Individuals)

किसी समूह का मध्यमान व मानक विचलन ज्ञात होने पर सामान्य प्रायिकता वक्र के प्रत्यय की सहायता से वे प्राप्तांक ज्ञात किये जा सकते हैं जिनके बीच किसी दिये गये प्रतिशत या संख्या के छात्र अंक प्राप्त करते हैं। इस प्रकार की समस्याओं में सबसे पहले सामान्य प्रायिकता वक्र में छात्रों का वांछित प्रतिशत दो कोटियों के बीच के क्षेत्रफल के रूप में व्यक्त कर लिया जाता है तथा फिर इन कोटियों के सापेक्ष σ मूल्यों को सामान्य प्रायिकता वक्र सारणी से ज्ञात करके मूल प्राप्तांकों में बदल लेते हैं। ये प्राप्तांक ही वांछित छात्रों के प्राप्तांकों की सीमाएँ प्रकट करते हैं। किसी σ मान को मूल प्राप्तांक में बदलने के लिए निम्न सूत्र का प्रयोग किया जाता है–

मूल प्राप्तांक, X = M + (σ मान x S.D.)

दूसरे शब्दों में, किसी σ मान को मूल प्राप्तांक में बदलने के लिए उसे मानक विचलन से गुणा करके मध्यमान में जोड़ देते हैं।

उदाहरण–छात्रों के एक समूह में 500 छात्र थे, जिनका सामान्य ज्ञान परीक्षण पर मध्यमान 40 तथा मानक विचलन 12 प्राप्त हुआ। ज्ञात करिये कि–

(*i*) सामान्य ज्ञान परीक्षण पर 20% श्रेष्ठतम् छात्र किस प्राप्तांक से अधिक अंक प्राप्त करेंगे?

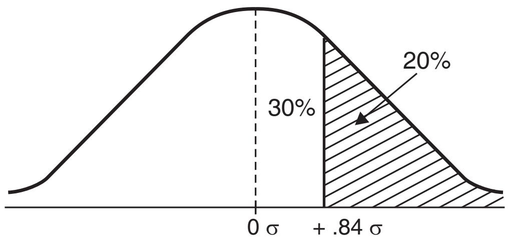
(*ii*) सामान्य ज्ञान परीक्षण पर मध्य के 50% छात्र किस-किस प्राप्तांकों के बीच अंक प्राप्त करेंगे?

(*iii*) सामान्य ज्ञान परीक्षण पर नीचे के 350 छात्रों के प्राप्तांक किस प्राप्तांक से कम होंगे?

हल–स्पष्ट है कि $N = 500 \quad M = 40\ \sigma = 12$

(*i*) क्योंकि 20% श्रेष्ठ छात्र ही वांछित छात्र है, इसलिए ये सामान्य प्रायिकता वक्र में दांयी ओर स्थित होंगे तथा वक्र के सबसे दांयी ओर का 20% क्षेत्रफल इन छात्रों को प्रकट करेगा। सामान्य प्रायिकता वक्र में इस क्षेत्रफल को तिरछी रेखाओं से व्यक्त किया गया है। स्पष्ट है कि जो कोटि इस 20% क्षेत्रफल को शेष 80% क्षेत्रफल से

अलग कर रही है उस कोटि के सापेक्ष मूल प्राप्तांक ही 20% श्रेष्ठ छात्रों के प्राप्तांकों की निम्न सीमा होगी। अत: इस कोटि के सापेक्ष मानक प्राप्तांक सामान्य प्रायिकता वक्र सारणी की सहायता से ज्ञात करके उसे मूल प्राप्तांक में बदलेंगे। परन्तु सामान्य प्रायिकता वक्र सारणी में 0 σ पर स्थित कोटि व 0 σ से किसी दूरी पर स्थित कोटि के बीच स्थित प्राप्तांकों की प्रतिशत संख्या दी गई होती हैं, इसलिए पहले वांछित कोटि व 0 σ पर स्थित कोटि के बीच छात्रों का प्रतिशत ज्ञात करेंगे। क्योंकि 0 σ के दांयी ओर 50% छात्र होते हैं तथा वांछित कोटि के दांयी ओर 20% छात्र है, इसलिए 0 σ व वांछित कोटि के बीच 30% छात्र होंगे। अब सामान्य वक्र सारणी की सहायता से देखा जायेगा कि इन 30% (या इसके सर्वाधिक निकट प्रतिशत) के लिए σ मान क्या है। सारणी में 29.95% के लिए सिगमा मान .84 σ है। क्योंकि वांछित कोटि 0 σ के दाँयी ओर स्थित है इसलिए इसका σ मान धनात्मक होगा। अत: .0 σ व +.84 σ के बीच लगभग 30% छात्र होंगे तथा +.84 σ के ऊपर लगभग 20% छात्र होंगे। अब +.84 σ को मूल प्राप्तांक में बदलना होगा–



चित्र 13

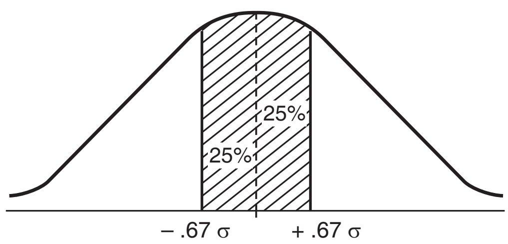
$$\text{मूल प्राप्तांक}, X = M + (\sigma \text{ मान} \times S.D.)$$

$$\text{अत: .84 के लिए मूल प्राप्तांक}, X = 40 + (.84 \times 12)$$

$$= 50.08$$

अत: कहा जा सकता है कि ऊपर के 20% छात्रों के प्राप्तांक 50.08 से अधिक होंगे। दूसरे शब्दों में, श्रेष्ठ 20% छात्र कम से कम 50 या इससे अधिक अंक प्राप्त करेंगे।

(*ii*) क्योंकि बीच के 50% छात्र ही वाँछित छात्र हैं, अत: इसके आधे अर्थात् 25% मध्यमान या 0 σ से नीचे होंगे तथा आधे अर्थात 25% मध्यमान या 0 σ से ऊपर होंगे। सामान्य प्रायिकता वक्र में इन छात्रों को तिरछी रेखाओं से प्रदर्शित किया गया है। स्पष्ट है कि इन छात्रों को प्रदर्शित करने वाले क्षेत्रफल को शेष क्षेत्रफल से अलग करने वाली कोटियों के सापेक्ष प्राप्तांक ही वाँछित प्राप्तांक होंगे। 0 σ व वाँछित कोटियों के बीच 25% छात्र है, अत: सामान्य वक्र सारणी में 25% के लिए σ मान देखा जायेगा, जो कि .67 σ है। स्पष्ट है कि 0 σ के दांयी ओर की कोटि के लिए यह + .67 σ होगा तथा बांयी ओर की कोटि के लिए यह – .67 σ होगा।



चित्र 14

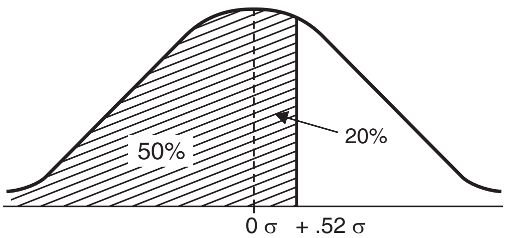
अतः – .67 σ तथा + .67 σ को मूल प्राप्तांकों में बदलने पर

– .67 σ के लिए मूल प्राप्तांक = 40 + (– .67 × 12) = 31.96

+ .67 σ के लिए मूल प्राप्तांक = 40 + (.67 × 12) = 48.04

क्योंकि – .67 σ व + .67 σ के बीच मध्य के 50% छात्र होंगे, अतः – .67 σ का मूल प्राप्तांक अर्थात् 31.96 बीच के 50 छात्रों की निम्न सीमा होगी तथा +.67 σ का मूल प्राप्तांक अर्थात् 48.04 बीच के 50% छात्रों की उच्च सीमा होगी। दूसरे शब्दों में कहा जा सकता है कि बीच के 50% छात्र 32 तथा 48 के बीच अंक प्राप्त करेंगे।

(*iii*) क्योंकि नीचे के 350 छात्र ही वांछित छात्र है इसका अर्थ है कि नीचे के (350/500) × 100 = 70% छात्र ही वांछित छात्र हैं तथा ये सामान्य प्रायिकता वक्र में बांयी ओर स्थिर होंगे अतः वक्र के बांयी ओर का 70% क्षेत्रफल इन छात्रों को प्रदर्शित करेगा। क्योंकि सामान्य प्रायिकता वक्र में 0 σ से नीचे भी तथा ऊपर भी 50% छात्र होते हैं, इसलिए वांछित छात्रों में से 50% छात्र 0 σ के बाईं ओर तथा 20% छात्र 0 σ के दाईं ओर होंगे। इन छात्रों को तिरछी रेखाओं से प्रकट किया गया है। स्पष्ट है कि नीचे के 70% छात्रों को शेष 30% छात्रों से अलग करने वाली कोटि के सापेक्ष प्राप्तांक ही नीचे के 70% छात्रों के प्राप्तांकों की उच्च सीमा होगी। क्योंकि 0 σ पर स्थित कोटि व वांछित कोटि के बीच 20% छात्र है अतः सामान्य प्रायिकता वक्र सारणी में 20% के लिए σ मान देखेगें। स्पष्ट है कि 19.85% के लिए σ मान .52 σ है। क्योंकि वांछित कोटि 0 σ के दाईं ओर स्थित है अतः वांछित कोटि का σ मान + .52 σ होगा।



चित्र 15

अब + .52 σ को मूल प्राप्तांक में बदलने पर।

+ .52 σ के लिए मूल प्राप्तांक = 40 + .52 × 12 = 46.24

अतः सामान्य प्रायिकता वक्र के आधार पर कहा जा सकता है कि नीचे के 350 छात्र 46 या इससे कम अंक प्राप्त करेंगे।

4. परीक्षण के प्रश्नों का सापेक्षिक कठिनाई स्तर ज्ञात करना
(Determination of the Relative Difficulty of Test Items)

सामान्य प्रायिकता वक्र के प्रत्यय की सहायता से किसी परीक्षण के विभिन्न प्रश्नों का सापेक्षिक कठिनाई स्तर ज्ञात करके उनकी तुलना भी की जा सकती है। इसके लिए प्रत्येक प्रश्न के लिए उस प्रश्न को सही हल करने वाले तथा सही हल न कर पाने वाले छात्रों को एक दूसरे से अलग करने वाली कोटि ज्ञात की जाती है। इस कोटि के सापेक्ष σ मान को उस प्रश्न का कठिनाई स्तर कहा जाता है। यह कठिनाई स्तर अर्थात् σ मान जितना अधिक होता है, प्रश्न उतना ही अधिक कठिन होता है।

उदाहरण—किसी परीक्षण के तीन प्रश्नों को सही हल करने वाले छात्रों की संख्या निम्नवत् थी। प्रश्नों के सापेक्षिक कठिनाई स्तर की तुलना करो।

प्रश्न	सही हल करने वाले छात्रों का प्रतिशत
1	50%
2	71%
3	40%

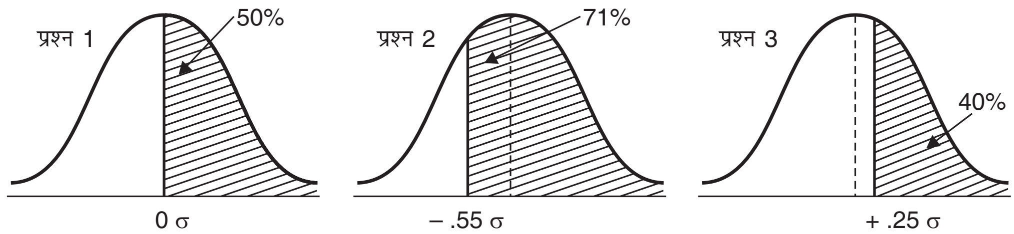
हल–प्रश्न 1 को 50% छात्र हल करते हैं। स्पष्ट है कि ये 50% छात्र श्रेष्ठ छात्र होंगे तथा सामान्य वक्र में दाईं ओर स्थित होंगे। इसी प्रकार से प्रश्न 2 व 3 को सही हल करने वाले 71% व 40% छात्र श्रेष्ठ छात्र होने के कारण सामान्य वक्र में दाईं ओर स्थित होंगे। सामान्य वक्र सारणी से उन कोटियों का, जो सही हल करने वाले छात्रों को शेष छात्रों से अलग करती है, σ मान ज्ञात करने पर प्रश्नों का कठिनाई स्तर ज्ञात हो जायेगा। अत:

प्रश्न 1 का कठिनाई स्तर $= 0\ \sigma$

प्रश्न 2 का कठिनाई स्तर $= -\ .55\ \sigma$

प्रश्न 3 का कठिनाई स्तर $= +\ .25\ \sigma$

स्पष्ट है कि इन तीनों प्रश्नों में प्रश्न 2 सबसे सरल तथा प्रश्न 3 सबसे कठिन है। प्रश्न 1 प्रश्न 2 की अपेक्षा $.55\ \sigma$ अधिक कठिन है तथा प्रश्न 3 प्रश्न 1 की अपेक्षा $.25\ \sigma$ अधिक कठिन है। जबकि प्रश्न 3 प्रश्न 2 की अपेक्षा $.80\ \sigma$ अधिक कठिन है।



चित्र 16

5. किसी समूह को योग्यता के आधार पर कुछ भागों में विभक्त करना (Division of a Group into Sub-Groups on the basis of Ability)

सामान्य प्रायिकता वक्र के प्रत्यय का उपयोग छात्रों के बड़े समूहों को ऐसे क्रमबद्ध उपसमूहों में विभाजित करने के लिए भी किया जा सकता है कि प्रत्येक उपसमूह में योग्यता का विस्तार अर्थात् छात्रों के प्राप्तांकों का प्रसार बराबर-बराबर हो। ऐसा करने के लिए सामान्य प्रायिकता वक्र की क्षैतिज दूरी को उतने ही बराबर भागों में बाँट लेते हैं जितने कि उपसमूह बनाने होते हैं तथा इन भागों से सम्बन्धित कोटियों के सापेक्ष σ मान ज्ञात करके उन्हें मूल प्राप्तांकों में बदल लेते हैं। ये मूल प्राप्तांक ही विभिन्न उपसमूहों के छात्रों की उच्च व निम्न सीमा बताते हैं। यदि विभिन्न उपसमूहों में छात्रों की संख्या ज्ञात करनी होती है तो विभिन्न मानों के लिए सामान्य प्रायिकता वक्र सारणी से छात्रों का प्रतिशत ज्ञात करके विभिन्न उपसमूहों में छात्रों की प्रतिशत संख्या ज्ञात कर लेते हैं। क्योंकि सामान्य प्रायिकता वक्र में $-3\ \sigma$ से $+3\ \sigma$ के बीच अर्थात् $6\ \sigma$ की दूरी में 99.74% छात्र आ जाते हैं, इसलिए व्यावहारिक समस्याओं में सामान्य प्रायिकता वक्र की क्षैतिज दूरी $6\ \sigma$ ही मानी जाती है। दूसरे शब्दों में मान लिया जाता है कि $-3\ \sigma$ से $+3\ \sigma$ के बीच सभी 100% छात्र आ जाते हैं।

उदाहरण–छात्रों के एक समूह का भाषा परीक्षण पर मध्यमान 60 व मानक विचलन 10 था। अध्यापक छात्रों को छ: ग्रेड A, B, C, D, E व F इस प्रकार से देना चाहता है कि प्रत्येक ग्रेड को प्राप्त करने वाले छात्रों के प्राप्तांकों का विस्तार समान हो। प्राप्तांकों के वितरण को सामान्य प्रायिकता वक्र के अनुरूप मानते हुए ज्ञात करिये कि

(*i*) प्रत्येक ग्रेड को कितने-कितने प्रतिशत छात्र प्राप्त करें?

(*ii*) प्रत्येक ग्रेड में किन-किन प्राप्तांकों वाले छात्र आयेंगे?

हल–ज्ञात है कि $M = 60$ तथा $\sigma = 10$

क्योंकि व्यावहारिक परिस्थितियों में माना जा सकता है कि सम्पूर्ण समूह -3σ से $+3\sigma$ के बीच अर्थात् 6σ दूरी में वितरित होता है। अत: $-3\ \sigma$ से $+3\ \sigma$ तक की $6\ \sigma$ दूरी को विभिन्न उपसमूहों में बराबर-बराबर बांटा

जायेगा। क्योंकि समस्त छात्रों को 6 उपसमूहों में बांटना है अत: प्रत्येक उपसमूह में 1 σ दूरी आयेगी। अत: सामान्य प्रायिकता वक्र में विभिन्न उपसमूहों की स्थिति चित्र 58 के अनुसार होगी। चित्र से स्पष्ट है कि

उपसमूह A में + 2 σ से ऊपर वाले छात्र होंगे।

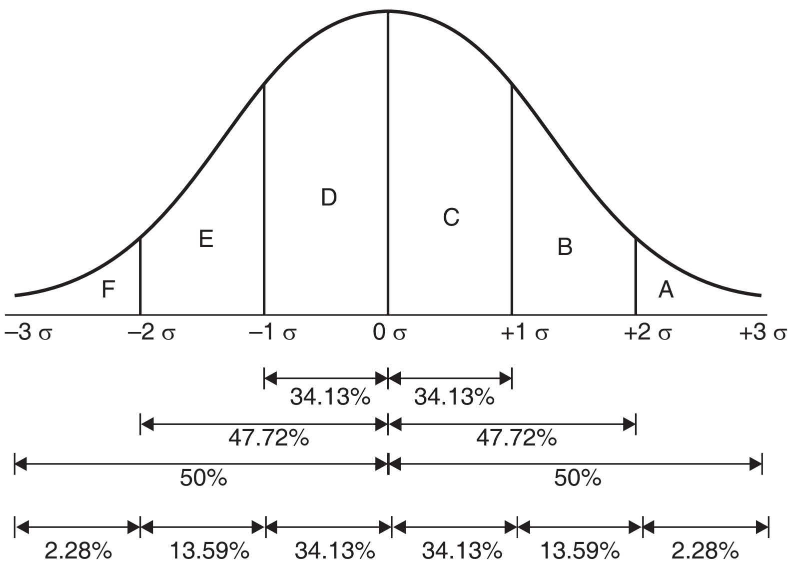
उपसमूह B में + 1 σ से + 2 σ तक के छात्र होंगे।

उपसमूह C में 0 σ से + 1 σ तक के छात्र होंगे।

उपसमूह D में – 1 σ से 0 σ तक के छात्र होंगे।

उपसमूह E में – 2 σ से – 1 σ तक के छात्र होंगे।

उपसमूह F में – 2 σ से नीचे वाले छात्र होंगे।



चित्र 17

(*i*) क्योंकि प्रत्येक ग्रेड को प्राप्त करने वाले छात्रों का प्रतिशत ज्ञात करना है इसलिए विभिन्न σ मानों के सापेक्ष छात्रों की संख्या सामान्य प्रायिकता वक्र सारणी से देखेगें, जो कि चित्र 58 में प्रस्तुत की गई है। स्पष्ट है कि ग्रेड A को 2.28%, ग्रेड B को 13.59%, ग्रेड C को 34.13%, ग्रेड D को 34.13%, ग्रेड E को 13.59 व ग्रेड F को 2.28% छात्र प्राप्त करेंगे।

(*ii*) क्योंकि विभिन्न ग्रेडों को प्राप्त करने वाले छात्रों के प्राप्तांकों की उच्च व निम्न सीमा भी ज्ञात करनी हैं इसलिए विभिन्न σ मानों को मूल प्राप्तांक में बदलेंगे–

+ 2 σ का मूल प्राप्तांक = 60 + (2 × 10) = 80

+ 1 σ का मूल प्राप्तांक = 60 + (1 × 10) = 70

0 σ का मूल प्राप्तांक = 60 + (0 × 10) = 60

– 1 σ का मूल प्राप्तांक = 60 – (1 × 10) = 50

– 2 σ का मूल प्राप्तांक = 60 – (2 × 10) = 40

अत: कहा जा सकता है कि

80 से अधिक प्राप्तांक पाने वाले छात्र A ग्रेड पायेंगे।

70 से 80 के बीच अंक पाने वाले छात्र B ग्रेड पायेंगे।

60 से 70 के बीच अंक पाने वाले छात्र C ग्रेड पायेंगे।

50 से 60 के बीच अंक पाने वाले छात्र D ग्रेड पायेंगे।

40 से 50 के बीच अंक पाने वाले छात्र E ग्रेड पायेंगे।

40 से कम प्राप्तांक पाने वाले छात्र F ग्रेड पायेंगे।

उदाहरण–400 छात्रों के एक समूह को सामूहिक बुद्धि परीक्षण दिया गया। छात्रों के इस समूह को पाँच भागों – तीव्र बुद्धि, अधिक बुद्धि, सामान्य बुद्धि, अल्प बुद्धि तथा क्षीण बुद्धि में इस प्रकार से बाँटना है कि विभिन्न भागों में बुद्धि प्राप्तांकों का विस्तार समान रहे। सामान्य प्रायिकता वक्र के अनुरूप वितरण की परिकल्पना मानते हुए ज्ञात करिये कि विभिन्न भागों में कितने-कितने छात्र आयेंगे?

हल–सामान्य प्रायिकता वक्र की व्यावहारिक दूरी 6 σ को पाँच बराबर भागों में बाँटने पर प्रत्येक भाग में 1.2 σ की दूरी आयेगी। अत: सामान्य प्रायिकता वक्र में विभिन्न भागों की स्थिति निम्नवत होंगी–

+ 1.8 σ से ऊपर के छात्र तीव्र बुद्धि वाले समूह में आयेंगे।

+ .6 σ से + 1.8 तक के छात्र अधिक बुद्धि वाले समूह में आयेंगे।

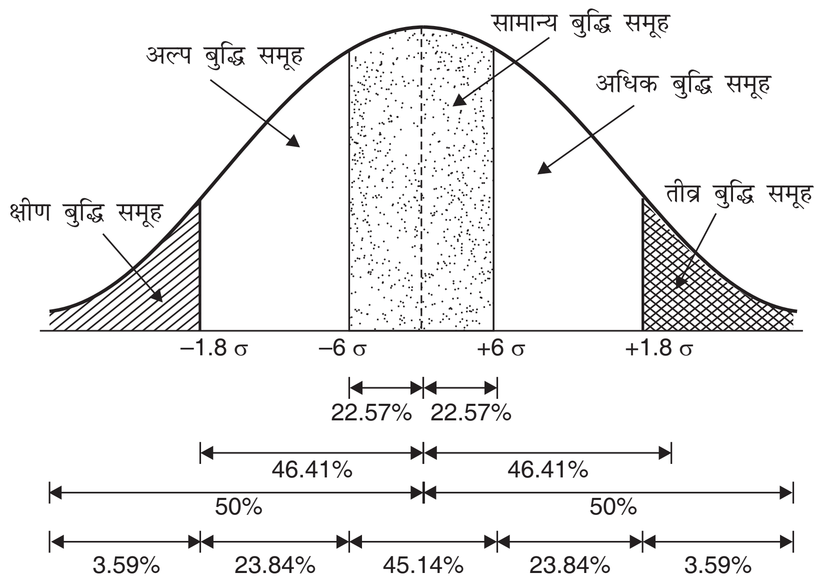
– .6 σ से + .6 σ तक के छात्र सामान्य बुद्धि वाले समूह में आयेंगे।

– 1.8 σ से – .6 σ तक के छात्र अल्प बुद्धि वाले समूह में आयेंगे।

– 1.8 σ से नीचे को छात्र क्षीण बुद्धि वाले समूह में आयेंगे।

विभिन्न समूहों को चित्र 18 में प्रदर्शित किया गया है। सामान्य वक्र सारणी के अवलोकन से स्पष्ट है कि 0 σ व .6 σ के बीच 22.57% छात्र तथा 0 σ व 1.8 σ के बीच 46.41% छात्र होंगे। अत: कहा जा सकता है कि

तीव्र बुद्धि वाले समूह में	50 – 46.41	= 3.59% छात्र होंगे।
अधिक बुद्धि वाले समूह में	46.41 – 22.57	= 23.84% छात्र होंगे।
सामान्य बुद्धि वाले समूह में	22.57 + 22.57	= 45.14% छात्र होंगे।
अल्प बुद्धि वाले समूह में	46.41 – 22.57	= 23.84% छात्र होंगे।
क्षीण बुद्धि वाले समूह में	50 – 46.41	= 3.59% छात्र होंगे।



चित्र 18

क्योंकि कुल 400 छात्र हैं अत:

तीव्र बुद्धि वाले समूह में $\frac{3.59}{100} \times 400$ = 14.36 अर्थात् 14 छात्र होंगे।

अधिक बुद्धि वाले समूह में $\frac{23.84}{100} \times 400$ = 95.36 अर्थात् 95 छात्र होंगे।

सामान्य बुद्धि वाले समूह में $\frac{45.14}{100} \times 400$ = 180.56 अर्थात् 181 छात्र होंगे।

अल्प बुद्धि वाले समूह में $\frac{23.84}{100} \times 400$ = 95.36 अर्थात् 95 छात्र होंगे।

क्षीण बुद्धि वाले समूह में $\frac{3.59}{100} \times 400$ = 14.36 अर्थात् 14 छात्र होंगे।

3.6 प्राप्तांकों के किसी वितरण को सामान्यीकृत करना (Normalizing a Distribution of Scores)

जैसा कि स्पष्ट किया जा चुका है सामान्य प्रायिकता वक्र (N.P.C.) के प्रत्यय का उपयोग केवल उन्हीं परिस्थितियों में किया जा सकता है जब समंकों का वितरण सामान्य प्रायिकता वक्र के अनुरूप होता है। परन्तु कभी-कभी ऐसी परिस्थितियाँ उत्पन्न होती हैं कि संकलित समंकों का वितरण सामान्य प्रायिकता वक्र से काफी भिन्न होता है जिसके कारण उन्हें सामान्य रूप से वितरित मानना सम्भव नहीं हो पाता है। ऐसे समंकों के साथ सामान्य प्रायिकता वक्र के प्रत्यय तथा उस पर आधारित सांख्यिकीय गणनाओं का प्रयोग करना किसी भी दृष्टि से उचित नहीं हैं। किसी वितरण की सामान्यता का निश्चय करने के लिए यह आवश्यक हो जाता है कि उस वितरण के सापेक्ष सामान्य प्रायिकता वितरण तैयार किया जाये। वितरणों की सामान्यता (Normalcy) का परीक्षण किये जाने वाले विभिन्न प्रकार के सांख्यिकीय परीक्षणों (जैसे कोई-वर्ग परीक्षण, के. एस. परीक्षण आदि) में अवलोकित वितरण की उसके सापेक्षिक सामान्य प्रायिकता वितरण से तुलना करके देखा जाता है कि क्या दृष्टिगत अन्तर को संयोगवश स्वीकार किया जा सकता है। इसके अतिरिक्त कभी-कभी अनुसंधानकर्ता अपनी अनुसंधान समस्या का समाधान प्राप्त करने के लिए अथवा अनुसंधान परिकल्पनाओं का परीक्षण करने के लिए किन्हीं ऐसी सांख्यिकीय प्राविधियों का प्रयोग करने के लिए कटिबद्ध होता है जिनमें समंकों का वितरण सामान्य होना आवश्यक होता है। समंकों के असामान्य रूप से वितरित होने की स्थिति में ऐसी प्राविधियों के प्रयोग से प्राप्त परिणामों को वैध स्वीकार नहीं किया जा सकता है। ऐसी परिस्थिति में यदि अनुसंधानकर्ता तर्कपूर्ण ढंग से मान सकता है कि समष्टि में समंकों का वितरण वस्तुत: सामान्य है तथा किसी कारणवश प्रतिदर्श के लिए समंकों का वितरण असामान्य आ गया है तब अवलोकित वितरण के लिए सामान्य वितरण का आसंजन (Fitting of Normal Distribution) किया जा सकता है तथा इस प्रकार से प्राप्त आसंजित सामान्य वितरण पर वांछित सांख्यिकीय प्राविधियों का प्रयोग वैध ढंग से किया जा सकता है। उपरोक्त कारणों से कभी-कभी प्राप्य या अवलोकित वितरण को सामान्य प्रायिकता वितरण के रूप में बदलना अथवा आसंजित करना आवश्यक तथा उपयोगी होता है। समंकों के किसी वितरण के लिए परिवर्तित अथवा आसंजित सामान्य वितरण वह वितरण है जो सामान्य प्रायिकता वक्र के अनुरूप होता है तथा जिसका मध्यमान, मानक विचलन तथा व्यक्ति संख्या (N) अपने मूल वितरण के समतुल्य होती हैं। असामान्य रूप से वितरित प्राप्तांकों को सामान्यीकृत करने पर होने वाले प्रभाव को चित्र 19 में दर्शाया गया है। चित्र से स्पष्ट होगा कि सामान्यीकरण के उपरान्त प्राप्तांकों का वितरण लगभग सामान्य प्रायिकता वक्र के अनुरूप ही है।

किसी प्राप्य असामान्य आवृत्ति वितरण (Observed Non-Normal Frequency Distribution) का सामान्यीकरण (Normalizing) या सामान्य वितरण आसंजन (Fitting of Normal Distribution) दो विधियों से किया जा सकता है–

(*i*) **कोटिअक्ष विधि** (Method of Ordinates)

(*ii*) **क्षेत्रफल विधि** (Method of Areas)

दोनों ही विधियों में प्रयुक्त किये जाने वाले विभिन्न सोपानों को आगे स्पष्ट किया गया है परन्तु यहाँ यह बात ध्यान देने की है कि एक ही वितरण के लिए कोटि अक्ष विधि तथा क्षेत्रफल विधियों से प्राप्त सामान्यीकृत वितरणों की आवृत्तियों में मामूली सा अन्तर हो सकता है।

कोटिअक्ष विधि

(Method of Ordinates)

कोटिअक्ष विधि के द्वारा किसी दिये गये असामान्य वितरण के लिए सामान्य वितरण का आसंजन करने के लिए निम्न सोपानों का विचलन अनुसरण किया जाता है–

(*i*) सर्वप्रथम दिये गये वितरण के लिए मध्यमान (M) तथा मानक वितरण (S.D.) की गणना की जाती है।

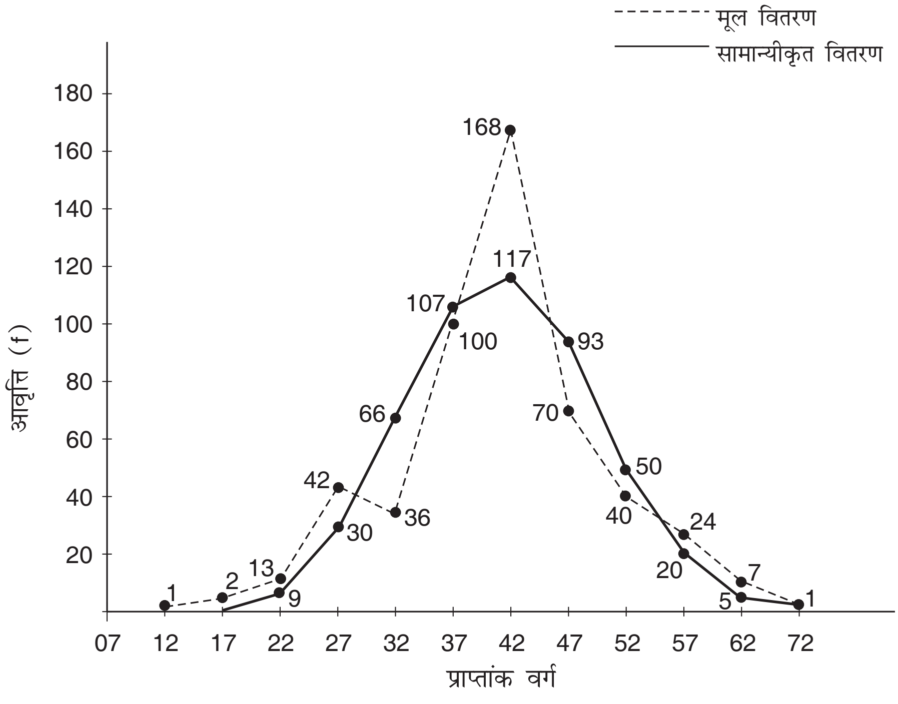
(*ii*) विभिन्न वर्गों के मध्यबिन्दु (X) ज्ञात किये जाते हैं।

(*iii*) वर्गों के मध्यबिन्दु (X) में से मध्यमान (M) को घटा दिया जाता है जिससे विचलित प्राप्तांक (x) प्राप्त हो जाते हैं।

(*iv*) विचलित प्राप्तांकों (x) को मानक विचलन (S.D.) से भाग देकर जेड प्राप्तांक (Z) ज्ञात कर लेते हैं।

(*v*) परिशिष्ट में दी गई सामान्य प्रायिकता वक्र की कोटियों की सारणी (Table of Ordinates of Normal Probability Curve) से विभिन्न Z प्राप्तांकों के सापेक्ष कोटियों की ऊँचाइयाँ (Hights of Ordinates) अर्थात् y ज्ञात कर ली जाती हैं।

वर्ग	15–19	20–24	25–29	30–34	35–39	40–44	45–49	50–54	55–59	60–64	65–69
अवलोकित आवृत्ति	0	13	42	36	100	168	70	40	24	7	0
सामान्यीकृत आवृत्ति	2	9	30	66	107	117	93	50	20	5	1



चित्र 19. ***सामान्यीकरण का प्रभाव (Effect of Normalizing)*** (*vi*) विभिन्न कोटियों की ऊँचाइयों (y) को वर्ग अन्तराल (i) व N से गुणा करके तथा मानक विचलन से भाग देकर विभिन्न वर्गों की अपेक्षित आवृत्तियाँ (f_N) ज्ञात कर लेते हैं।

अग्रांकित उदाहरण से किसी दिये गये आवृत्ति वितरण के सामान्य वितरण आसंजन की कोटि अक्ष विधि स्पष्ट हो सकेगी।

उदाहरण–निम्न वितरण के लिए सामान्य वितरण का आसंजन कोटि अक्ष विधि से कीजिए।

सारणी 2.7

आवृत्ति वितरण

वर्ग	आवृत्ति f
55–59	1
50–54	4
45–49	12
40–44	25
35–39	41
30–34	22
25–29	10
20–24	3
15–19	2
कुल	**120**

हल–सबसे पहले मध्यमान तथा मानक विचलन की गणना की जायेगी। गणना करने पर मध्यमान का मान 37.25 तथा मानक विचलन का मान 7.15 प्राप्त हुआ। अब सारणी बनाकर X, x, Z, y तथा f_N ज्ञात किये जायेंगे। इन्हें सभी को सारणी 2.8 में प्रस्तुत किया गया है।

सारणी 2.8

कोटिअक्ष विधि से किसी दिये गये वितरण को सामान्यीकृत वितरण में बदलना

वर्ग	आवृत्ति f	मध्य बिन्दु X	विचलित प्राप्तांक x	जेड प्राप्तांक Z	कोटि अक्ष y	सामान्यीकृत आवृत्ति $f_N = \frac{i \times N \times y}{S.D.}$
55–59	1	57	19.75	2.76	.0088	.74 = 1
50–54	4	52	14.75	2.06	.0478	4.01 = 4
45–49	12	47	9.75	1.36	.1582	13.28 = 13
40–44	25	42	4.75	.66	.3209	26.93 = 27
35–39	41	37	−.25	−.03	.3988	33.47 = 33
30–34	22	32	−5.25	−.73	.3056	25.64 = 26
25–29	10	27	−10.25	−1.43	.1435	12.04 = 12
20–24	3	22	−15.25	−2.13	.0413	3.47 = 3
15–19	2	17	−20.25	−2.83	.0073	.61 = 1
i **= 5**	**N = 120**		**M = 37.25**	**S.D. = 7.15**		**N = 120**

नोट

सारणी 2.8 में सामान्यीकृत आवृत्ति वाले स्तम्भ के अवलोकन से स्पष्ट है कि अब आवृत्तियों का वितरण सामान्य प्रायिकता वक्र के समान है। सामान्यीकृत आवृत्ति वितरण के लिए M = 37.25 तथा S.D. = 7.18 है जोकि मूल वितरण के लगभग समान ही है। इस स्थान पर यह बात ध्यान रखने की है कि यदि सबसे ऊपर के वर्ग के लिए Z का मान $+3\sigma$ से कम होता है तब ऊपर की ओर अतिरिक्त वर्ग बनाकर उनकी प्राप्य आवृत्ति शून्य मान लेते हैं। इसी प्रकार से सबसे नीचे के वर्ग के लिए Z मान – 3.00 से कम होने पर अतिरिक्त वर्ग बनाकर उनकी आवृत्ति शून्य मान ली जाती है।

क्षेत्रफल विधि

(Method of Areas)

क्षेत्रफल विधि से किसी दिये गये असामान्य वितरण के लिए सामान्य वितरण का आसंजन करते समय निम्नांकित सोपानों का अनुगमन किया जाता है–

(*i*) सर्वप्रथम दिये गये वितरण के लिए मध्यमान (M) तथा मानक विचलन (S.D.) की गणना की जाती है।

(*ii*) विभिन्न वर्गों की उच्च सीमाएँ (Upper Limits) ज्ञात की जाती हैं।

(*iii*) वर्गों की उच्च सीमाओं में से मध्यमान घटाकर उसके लिए विचलित प्राप्तांक x ज्ञात किये जाते हैं।

(*iv*) प्राप्त x मानों को मानक विचलन से भाग देकर जेड मान (Z) में बदल दिया जाता है।

(*v*) सामान्य प्रायिकता वक्र की क्षेत्रफल सारणी (Table of Areas under Normal Probability Curve) से विभिन्न Z मानों के लिए प्रायिकता सामान्य प्रायिकता वक्र के अधीनस्थ प्रतिशत क्षेत्रफल (A) देख लेते हैं।

(*vi*) विभिन्न वर्गों के अधीनस्थ प्रतिशत क्षेत्रफल (Ac) ज्ञात कर लेते हैं।

(*vii*) विभिन्न वर्गों के अधीनस्थ क्षेत्रफल (Ac) को N से गुणा करके तथा 100 से भाग करके सामान्यीकृत आवृत्ति (f_N) ज्ञात कर लेते हैं।

किसी दिये गये वितरण को सामान्यीकृत वितरण में परिवर्तित करने की क्षेत्रफल विधि (Method of Areas) अग्रांकित उदाहरण से स्पष्ट हो सकेगी।

उदाहरण–निम्न सारणी में दिये जा रहे आवृत्ति वितरण को क्षेत्रफल विधि से सामान्यीकृत आवृत्ति वितरण में परिवर्तित करो।

सारणी 3.9

वर्ग	**15–17**	**18–20**	**21–23**	**24–26**	**27–29**	**30–32**	**33–35**	**36–38**	**39–41**	**कुल N**
आवृत्ति f	2	4	16	25	43	30	17	8	5	150

हल–सर्वप्रथम मध्यमान तथा मानक विचलन की गणना करनी होगी। गणना करने पर मध्यमान 28.62 तथा मानक विचलन 4.90 प्राप्त होता है। अब सारणी बनाकर विभिन्न गणना की जायेगी जिन्हें सारणी 2.10 में प्रस्तुत किया गया है।

सारणी 3.10

क्षेत्रफल विधि से दिये गये वितरण को सामान्यीकृत वितरण में बदलना

वर्ग	आवृत्ति f	वर्ग की उच्च सीमा U_L	$x = U_L - M$	$Z = \frac{U_L - M}{S.D.}$	मध्य कोटि से Z के सापेक्ष कोटि क्षेत्रफल का अनुपात A	वर्ग के अधीनस्थ क्षेत्रफल का अनुपात Ac	सामान्यीकृत आवृत्ति $f_N = A_c \times \frac{N}{100}$
39–41	5	41.5	12.88	2.63	.5000	.0217	3.26 = 3
36–38	8	38.5	9.88	2.02	.4783	.0591	8.87 = 9
33–35	17	35.5	6.88	1.40	.4192	.0134	20.10 = 20
30–32	30	32.5	3.88	.79	.2852	.2138	32.07 = 32
27–29	43	29.5	.88	.18	.0714	.2378	35.67 = 36
24–26	25	26.5	–2.12	–.43	.1664	.1844	27.66 = 28
21–23	16	23.5	– 5.12	–1.04	.3508	.1007	15.11 = 15
18–20	4	20.5	–8.12	–1.66	.4515	.0369	5.54 = 5
15–17	2	17.5	–11.12	–2.27	.4884	.0116	1.74 = 2
12–14	0	14.5	–14.12	–2.88	.5000		
i = 3	**N = 150**	**M = 28.62**		**S.D. = 4.90**			**N = 150**

सारणी 2.10 में सामान्यीकृत आवृत्ति वाले स्तम्भ के अवलोकन से स्पष्ट है कि आवृत्तियों का यह वितरण सामान्य वितरण के समान है। इस सामान्यीकरण वितरण मध्यमान का मान 28.60 तथा मानक विचलन 4.90 है जो कि मूल वितरण के लगभग समान है।

मूल प्राप्तांकों को सामान्यीकृत प्राप्तांकों में परिवर्तित करना
(Converting Raw Scores into Normalized Scores)

पीछे समंकों के किसी असामान्य वितरण के लिए सामान्य प्रायिकता वक्र के आसंजन की आवश्यकता तथा विधि का वर्णन किया जा चुका है। यद्यपि अनेक परिस्थितियों में इस प्रकार का आसंजन उपयोगी होता है फिर भी कुछ परिस्थितियों में इस प्रकार का आसंजन पर्याप्त नहीं होता है। पीछे वर्णित आसंजन विधि में अवलोकित वितरण के मध्यमान तथा मानक विचलन को स्थिर रखते हुए सामान्यीकृत वितरण की स्थिति में विभिन्न वर्गों की आवृत्तियाँ ज्ञात की जाती हैं। परन्तु मूल प्राप्तांकों के सापेक्ष सामान्यीकृत वितरण के प्राप्तांकों को ज्ञात करने का कोई प्रयास नहीं किया गया है। कभी-कभी असामान्य रूप से वितरित मूल प्राप्तांकों को इस प्रकार से सामान्यीकृत प्राप्तांकों में परिवर्तित करने की आवश्यकता होती है कि परिवर्तित सामान्यीकृत प्राप्तांकों का मध्यमान तथा मानक विचलन मूल प्राप्तांकों के मध्यमान तथा मानक विचलन के बराबर हों। प्राप्तांकों की व्याख्या करने के लिए प्राय: मापन के क्षेत्र में इस प्रकार की आवश्यकता होती है जैसे मूल प्राप्तांकों को सामान्यीकृत टी-प्राप्तांकों में बदलते समय मापनकर्ता को प्राप्तांकों को सामान्यीकृत करना होता है। असामान्य रूप से वितरित चरों के लिए गुणनफल आघूर्ण सहसम्बन्ध ज्ञात करते समय भी इस प्रकार की परिस्थिति उत्पन्न हो सकती है। प्राप्तांकों के किसी समुच्चय को सामान्यीकृत प्राप्तांकों के रूप में बदलना कुछ जटिल तथा समयसाध्य है फिर भी इसकी उपयोगिता व आवश्यकता को देखते हुए प्राप्तांकों को सामान्यीकृत प्राप्तांकों में परिवर्तित करने की विधि का वर्णन किया जा रहा है। किन्हीं दिये गये प्राप्तांकों को सामान्यीकृत प्राप्तांकों में परिवर्तित करने के लिए अग्रांकित सोपानों का अनुगमन किया जाता है–

(*i*) सर्वप्रथम दिये गये प्राप्तांकों के लिए मध्यमान (M) तथा मानक विचलन (S.D.) की गणना की जाती है।

(*ii*) इसके उपरान्त प्राप्तांकों को क्रमबद्ध करके प्रत्येक प्राप्तांक की आवृत्ति (*f*) तथा संचयी आवृत्ति (*cf*) ज्ञात की जाती है।

नोट

(*iii*) तत्पश्चात प्रत्येक प्राप्तांक का शतांशीय क्रम (PR) ज्ञात किया जाता है। इसके लिए प्राप्तांक की संचयी आवृत्ति (या क्रम) में से 5 घटाकर 100/N से गुणा कर देते हैं। इस प्रकार से प्राप्त शतांशीय क्रम बताता है कि उस प्राप्तांक के नीचे आवृत्तियों का कुल कितना प्रतिशत स्थित है।

(*iv*) शतांशीय क्रम (PR) में से 50 को घटाकर मूल प्राप्तांक की मध्यबिन्दु से स्थिति ज्ञात कर लेते हैं। स्पष्ट है कि PR-50 का धनात्मक चिन्ह बतायेगा कि सामान्य प्रायिकता वितरण की स्थिति में मूल प्राप्तांक मध्यमान से ऊपर स्थित होता है जबकि ऋणात्मक चिन्ह बतायेगा कि मूल प्राप्तांक मध्यमान से नीचे स्थित होता।

(*v*) सामान्य प्रायिकता वक्र की क्षेत्रफल सारणी की सहायता से (PR-50)% भाग के सापेक्ष Z मान ज्ञात कर लिये जाते हैं। ये Z मान ही मूल प्राप्तांकों के सापेक्ष सामान्यीकृत मानक प्राप्तांक (Normalized Z-scores) होंगे।

(*vi*) जब इन Z मानों को मानक विचलन (S.D.) से गुणा करके मध्यमान में जोड़ देते हैं जिससे सामान्यीकृत (X) प्राप्त हो जाते हैं।

आगे प्रस्तुत किये गये उदाहरण से प्राप्तांकों को सामान्यीकृत प्राप्तांकों में परिवर्तित करने की विधि के उपरोक्त वर्णित सोपान स्पष्ट हो सकेंगे।

उदाहरण—निम्न प्राप्तांकों को सामान्यीकृत प्राप्तांकों में परिवर्तित कीजिए—

66	69	65	72	65	66	69	65	67	68	68	66	64	68	64
70	45	67	68	64	67	62	68	70	64	64	71	50	65	69
64	68	67	61	66	72	62	69	58	66	67	65	67	60	65
64	69	60	71	67	65	67	68	64	68	69	68	53	64	68
67	64	70	64	71	61	69	64	72	53	67	61	67	71	68
70	55	68	59	69	68	66	70	54	69	66	69	45	68	66
65	69	56	69	65	61	71	66	60	70	62	72	65	70	67
68	50	67	68	67	69	68	62	68	59	68	53	60	62	69
61	69	61	71	66	58	70	47	69	62	70	67	68		71
52	68	65	61	68	46	66	52	70	59	66	66	62	69	62

हल—सर्वप्रथम प्राप्तांकों को क्रमबद्ध करके उनका मध्यमान तथा मानक विचलन ज्ञात किया जायेगा जो क्रमश: 64.9267 तथा 5.65587 प्राप्त हुआ है। सामान्यीकृत आसंजन का गणना कार्य सारणी 2.11 में प्रस्तुत किया गया है। सारणी के अवलोकन से स्पष्ट है कि सामान्यीकृत आसंजन की स्थिति से 72 अंक वाले छात्र को 80 जबकि 45 अंक पाने वाले छात्र को 52 अंक दिये जायेंगे।

नोट

सारणी 3.11

दिये गये मूल प्राप्तांकों को सामान्यीकृत प्राप्तांकों में परिवर्तित करना

मूल प्राप्तांक X	आवृत्ति *f*	संचयी आवृत्ति *cf*	शतांकीय क्रम PR	मध्य बिन्दु से स्थिति PR-50	सामान्यीकृत मानक प्राप्तांक (सारणी से) Z	सामान्यीकृत प्राप्तांक $X' = M + Z\sigma$
72	5	150	99.67	49.67	2.72	80.31 = 80
71	7	145	96.33	46.33	1.79	75.06 = 75
70	10	138	91.67	41.67	1.38	72.73 = 73
69	17	128	85.00	35.00	1.04	70.81 = 71
68	22	111	73.67	23.67	0.63	68.49 = 68
67	15	89	59.00	9.00	0.23	66.23 = 66
66	13	74	49.00	– 1.00	– 0.03	64.76 = 65
65	11	61	40.33	– 9.67	– 0.24	63.57 = 64
64	12	50	33.00	– 17.00	– 0.44	62.44 = 62
62	8	38	25.00	– 25.00	– 0.67	61.14 = 61
61	7	30	19.67	– 30.33	– 0.85	60.12 = 60
60	4	23	15.00	– 35.00	– 1.04	59.05 = 59
59	3	19	12.33	– 37.67	– 1.16	58.37 = 58
58	2	16	10.33	– 39.67	– 1.26	57.80 = 58
56	1	14	9.00	– 41.00	– 1.34	57.35 = 57
55	1	13	8.33	– 41.67	– 1.38	57.12 = 57
54	1	12	7.67	– 42.33	– 1.43	56.84 = 57
53	3	11	7.00	– 43.00	– 1.48	56.56 = 57
52	2	8	5.00	– 45.00	– 1.65	55.59 = 56
50	2	6	3.67	– 46.33	– 1.79	54.80 = 55
47	1	4	2.33	– 47.67	– 1.99	53.67 = 54
46	1	3	1.67	– 48.33	– 2.13	52.88 = 53
45	2	2	1.00	– 49.00	– 2.33	51.75 = 52
			M = 65.6667			**S.D. = 5.9907**

अन्य प्राप्तांकों के लिए दिये जाने वाले सामान्यीकृत प्राप्तांक सारणी में दिये गये हैं। सामान्यीकृत प्राप्तांकों के लिए मध्यमान का मान 65.6667 तथा मानक विचलन का मान 5.9907 है जो काफी हद तक मूल प्राप्तांकों के मध्यमान तथा मानक विचलन के बराबर है। मूल प्राप्तांकों तथा सामान्यीकृत प्राप्तांकों के मध्यमानों व मानक विचलनों के बीच दृष्टिगोचित अन्तर मुख्यत: अनेक बार होने वाली सन्निकटीकरण त्रुटियों (Rounding Errors) के कारण है।

छात्र क्रियाकलाप (Student Activity)

1. द्विपद-वितरण से आप क्या समझते हैं?

2. प्वॉयसन प्रायिकता आवृत्ति वितरण का परिचय दीजिए।

3. सामान्य प्रायिकता वक्र के इतिहास का संक्षिप्त उल्लेख कीजिए।

4. प्राप्तांक से कम या ज्यादा अंक प्राप्त करने वाले व्यक्तियों की संख्या कैसे ज्ञात करते हैं?

3.7 सारांश (Summary)

1. सैद्धान्तिक आवृत्ति वितरण आधुनिक सांख्यिकी में महत्त्वपूर्ण आधार प्रदान करते हैं तथा प्रमुख आवृत्ति वितरण अनेक परिस्थितियों में निर्णय लेने में सहायक होते हैं।
2. घटनाएँ कई प्रकार की हो सकती हैं। जब एक समय में केवल एक ही घटना होने की बात की जाती है तब इसे **सरल घटना** (Simple Event) कहते हैं किन्तु जब एक साथ कई घटनाओं के घटित होने की चर्चा की जाती है तब ऐसी घटनाओं को **संयुक्त घटना** (Compound Event) कहा जाता है।
3. द्विपद-वितरण के साथ स्विस गणितज्ञ जेम्स बर्नोली (James Bernoulli, 1654-1705) का नाम जुड़ा हुआ है तथा इसे बर्नोली वितरण (Bernaulli Distribution, Bernaulli Trial or Bernaulli

Process) भी कहा करते हैं। इस सैद्धान्तिक वितरण का प्रतिपादन बर्नोली ने ही किया था और इसका प्रकाशन उनकी मृत्यु के 8 वर्ष पश्चात् 1713 में हुआ था।

4. सामान्य प्रायिकता वक्र (Normal Probability Curve) वास्तव में एक सैद्धान्तिक (Theoritical), आदर्श (Ideal) तथा गणितीय (Mathematical) वक्र है। इसे सामान्य वक्र (Normal Curve) या एन. पी. सी. (N.P.C.) के नाम से भी पुकारा जाता है।
5. किसी समूह का मध्यमान व मानक विचलन ज्ञात होने पर किन्हीं भी दो प्राप्तांकों के बीच अंक प्राप्त करने वाले छात्रों की प्रतिशत संख्या को सामान्य प्रायिकता वक्र की सहायता से ज्ञात किया जा सकता है।
6. किसी समूह का मध्यमान व मानक विचलन ज्ञात होने पर सामान्य प्रायिकता वक्र के प्रत्यय की सहायता से वे प्राप्तांक ज्ञात किये जा सकते हैं जिनके बीच किसी दिये गये प्रतिशत या संख्या के छात्र अंक प्राप्त करते हैं।
7. सामान्य प्रायिकता वक्र (N.P.C.) के प्रत्यय का उपयोग केवल उन्हीं परिस्थितियों में किया जा सकता है जब समंकों का वितरण सामान्य प्रायिकता वक्र के अनुरूप होता है।

अभ्यास-प्रश्न (Exercise Questions)

1. सामान्य प्रायिकता वक्र के प्रयोग से जिन समस्याओं का हल ढूंढ़ा जा सकता है, उनका वर्णन करें।
2. 'सैद्धान्तिक आवृत्ति वितरण आधुनिक सांख्यिकी में महत्त्वपूर्ण आधार प्रदान करते हैं।' विवेचन कीजिए।
3. प्वॉयसन वितरण की विशेषताओं का उल्लेख कीजिए।
4. सामान्य प्रायिकता वक्र क्या है? इसका ऐतिहासिक परिचय दीजिए।
5. प्राप्तांकों के किसी वितरण को किस प्रकार सामान्यीकृत किया जाता है?

सन्दर्भ ग्रन्थ (Reference Books)

नोट

इकाई–4

सहसंबंध एवं प्रतीपगमन (Correlation and Regression)

संरचना (Structure)

4.1 उद्देश्य (Objectives)

इस इकाई के अध्ययन के पश्चात् विद्यार्थी योग्य होंगे–

- आंशिक सहसम्बन्ध गुणांक और बहु-सहसम्बन्ध गुणांक।
- बहु-प्रतीपगमन समीकरण को समझने में।
- बीटा गुणाकों की सार्थकता को जानने में।

4.2 प्रस्तावना (Introduction)

सहसम्बन्ध एक विवरणात्मक सांख्यिकीय माप (descriptive statistics) है जो दो चरों के मध्य पाये जाने वाले सम्बन्ध की मात्रा का विवरण देती है। कुछ प्रकार के चरों में पारस्परिक सम्बन्ध पाया जाता है जैसे मूल्य तथा माँग, उत्पादन तथा रोजगार, मजदूरी तथा मूल्य निर्देशांक आदि चरों में सम्बन्ध होता है। दिन-प्रतिदिन के अनुभव से यह स्पष्ट होता है कि किस प्रकार विभिन्न तथ्य पारस्परिक रूप से सम्बन्धित होते हैं। उदाहरणार्थ, पति-पत्नियों

की आयुओं में, पिता-पुत्र की ऊँचाइयों में, युवा व्यक्तियों की ऊँचाई एवं भार में, विनियोजित पूँजी एवं अर्जित लाभ में तथा अन्य ऐसे ही तथ्यों में निकट का सम्बन्ध पाया जाता है। सहसम्बन्ध दो चरों के मध्य अन्तरनिर्भरता (interdependence) इंगित करता है। मुख्य रूप से सहसम्बन्ध दो चरों के मध्य सम्बन्ध की माप है।

सहसम्बन्ध की परिभाषा

(Definiton of Correlation)

जब दो तथ्यों में एक साथ एक ही दिशा में अथवा विपरीत दिशाओं में परिवर्तन हो तथा एक तथ्य में परिवर्तन दूसरे तथ्य में परिवर्तन का कारण हो तो यह कहा जाता है कि उन दोनों तथ्यों में सहसम्बन्ध है। डेवनपोर्ट (E .Davenport) के मतानुसार, "सहसम्बन्ध या सम्पूर्ण विषय पृथक् विशेषताओं के मध्य पाये जाने वाले उस पारस्परिक सम्बन्ध की ओर संकेत करता है जिसके अनुसार वे कुछ सीमा तक साथ-साथ परिवर्तन होने की प्रवृत्ति रखते हैं।" अत: सहसम्बन्ध दो चरों के मध्य अन्तरनिर्भरता (interdependence) अथवा सम्बन्ध का अध्ययन करता है। यह एक सांख्यिकीय तकनीक है जो दो या अधिक चरों के मध्य सम्बन्ध की प्रकृति व मात्रा मापती है।

सांख्यिकी-विश्लेषण में सहसम्बन्ध सिद्धान्त एवं तकनीक का बहुत महत्त्व है। इस सिद्धान्त के मूल तत्वों का प्रतिपादन सर्वप्रथम फ्रांस के खगोलशास्त्री ब्रावेश (Bravais) ने किया था, परन्तु बिन्दु-रेखीय रूप से सहसम्बन्ध तकनीक का अन्वेषण सर्वप्रथम सर फ्रांसिस गाल्टन (Sir Francis Galton) ने किया था। 1896 में प्रसिद्ध संख्याशास्त्री कार्ल पियर्सन (Karl Pearson) ने सहसम्बन्ध गुणांक (Coefficient of Correlation) द्वारा सहसम्बन्ध ज्ञात करने की गणितीय विधि का प्रतिपादन किया। इन दोनों संख्याशास्त्रियों (गाल्टन तथा पियर्सन) ने इस तकनीक की सहायता से प्राणिशास्त्र (Biology) तथा जनन-विद्या (Genetics) की अनेक समस्याओं का विवेचन किया। अर्थशास्त्र में भी इस तकनीक का विशेष महत्त्व है। अर्थशास्त्र में सहसम्बन्ध के उपयोग के बारे में नीसवेंजर (Neiswanger) लिखते हैं, "सहसम्बन्ध-विश्लेषण आर्थिक व्यवहार को समझने में योग देता है, विशेष महत्त्वपूर्ण चरों, जिन पर अन्य चर निर्भर करते हैं, को खोजने में सहायता देता है; अर्थशास्त्री को उन सम्बन्धों को स्पष्ट करता है जिनसे गड़बड़ी फैलती है तथा उसे उन उपायों का सुझाव देता है जिनके द्वारा स्थिरता लाने वाली शक्तियाँ प्रभावी हो सकती हैं।" प्रतीपगमन (regression) तथा विचरण-अनुपात (ratio of variation) के विचार सहसम्बन्ध की माप पर ही आधारित है। सहसम्बन्ध की माप यह भी आश्वस्त करती है कि सम्बन्धित चरों में आन्तरगणन अथवा बाह्यगणन एवं पूर्वानुमान विश्वसनीय होगा।

सहसम्बन्ध के प्रकार

(Types of Correlation)

सम्बद्ध चरों के मध्य परिवर्तनों की दिशा, अनुपात आदि के आधार पर सहसम्बन्ध निम्नलिखित प्रकारों का हो सकता है–

(1) धनात्मक अथवा ऋणात्मक सहसम्बन्ध (Positive or Negative Correlation)

(2) सरल, आंशिक अथवा बहुगणी सहसम्बन्ध (Simple, Partial or Multiple Correlation)

(3) रेखीय अथवा अ-रेखीय सहसम्बन्ध (Linear or Non-linear Correlation)

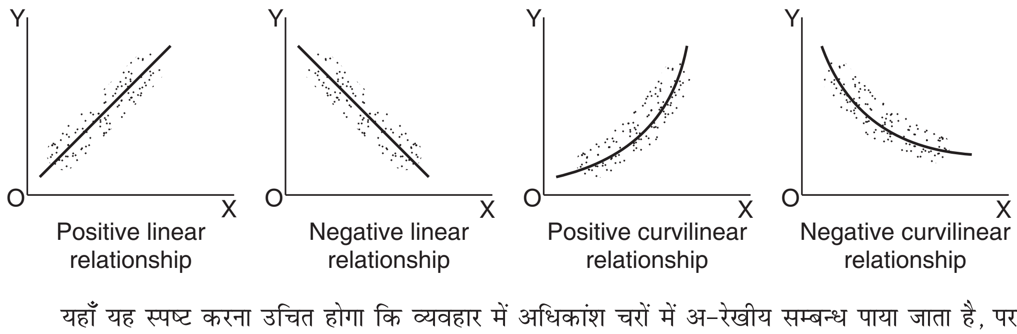
(1) **धनात्मक अथवा ऋणात्मक सहसम्बन्ध** (Positive or Negative Correalation)–दो चरों में यदि एक ही दिशा अथवा विपरीत दिशाओं में परिवर्तन होते हों तो उनके मध्य सहसम्बन्ध होता है। यदि एक चर-मूल्य घटने पर दूसरा चर-मूल्य भी घटे अथवा एक चर-मूल्य के बढ़ने पर दूसरा चर-मूल्य भी बढ़े तो ऐसा सहसम्बन्ध धनात्मक (Positive) होता है। मूल्य एवं पूर्ति में इस प्रकार का सम्बन्ध पाया जाता है। यदि किसी वस्तु का मूल्य बढ़ता है तो उसकी पूर्ति (supply) भी बढ़ जाती है और वस्तु का मूल्य घटने पर उसकी पूर्ति भी घट जाती है। ऋणात्मक सहसम्बन्ध उस दशा में होता है जब एक चर-मूल्य के घटने पर दूसरा चर-मूल्य बढ़ता हो तथा उस चर-मूल्य के बढ़ने पर दूसरे चर-मूल्य में कमी होती हो। इस प्रकार के सहसम्बन्ध को विलोम (inverse) सहसम्बन्ध भी कहते हैं। इस प्रकार का सहसम्बन्ध मूल्य एवं माँग में पाया जाता है। किसी वस्तु का मूल्य बढ़ने पर उसकी माँग (demand) कम हो जाती है और मूल्य कम हो जाने पर माँग बढ़ जाती है।

(2) **सरल, आंशिक अथवा बहुगुणी सहसम्बन्ध** (Simple, Partial or Multiple Correlation)– स्वतन्त्र एवं आश्रित चर-मूल्यों (independent and dependent variables) की संख्या तथा सहसम्बन्ध ज्ञात करने में निहित तकनीक के आधार पर सहसम्बन्ध सरल आंशिक अथवा बहुगुणी प्रकार का हो सकता है। दो चर-मूल्यों के सहसम्बन्ध को सरल सहसम्बन्ध (simple correlation) कहते हैं। इनमें एक श्रेणी, जिसे आधार श्रेणी (subject series) कहते हैं, के चर-मूल्य स्वतन्त्र (independent variables) होते हैं तथा दूसरी श्रेणी, जिसे सम्बद्ध श्रेणी (relative series) कहते हैं, के चर-मूल्य आश्रित (dependent variables) होते हैं। आंशिक सहसम्बन्ध (Partial Correlation) में दो मूल्यों में एक अन्य स्वतन्त्र चर-मूल्य का समावेश करके, सहसम्बन्ध ज्ञात किया जाता है। बहुगुणी सहसम्बन्ध (Multiple Correlation) में तीन या अधिक चर-मूल्यों के मध्य पारस्परिक सहसम्बन्ध का अध्ययन किया जाता है। उदाहरणार्थ, कृषि उपज, वर्षा तथा प्रयुक्त खाद के परिणामों के मध्य सहसम्बन्ध बहुगुणी सहसम्बन्ध कहलाता है।

(3) **रेखीय अथवा अ-रेखीय सहसम्बन्ध** (Linear or Non-Linear Correlation)–रेखीय अथवा अ-रेखीय सहसम्बन्ध के मध्य अन्तर का आधार विचारगत चर-मूल्यों के मध्य परिवर्तन-अनुपात की नियमितता होती है। यदि दो चर-मूल्यों के मध्य परिवर्तन का अनुपात समान होता है तो उनमें रेखीय सहसम्बन्ध होगा। उनका सम्बन्ध एक सरल रेखा से व्यक्त किया जा सकता है। उनके सम्बन्ध में अंकगणितीय वृद्धि (arithmetic progression) झलकती है जो धनात्मक अथवा ऋणात्मक हो सकती है। इन चर-मूल्यों को यदि बिन्दु-रेखीय पत्र पर प्रांकित किया जाय तो बिन्दु एक सीधी रेखा के रूप में होंगे। अ-रेखीय सहसम्बन्ध, जिसे वक्र-रेखीय सहसम्बन्ध (Curvilinear Correlation) भी कहते हैं, में एक चर-मूल्य के परिवर्तनों की मात्रा व दूसरे चर-मूल्य के परिवर्तनों की मात्रा एक अनुपात में नहीं होगी। इन चर-मूल्यों को बिन्दुरेखा पर प्रांकित करने पर वक्र बन जाता है। निम्न तालिका इस प्रवृत्ति को इंगित करती है–

रेखीय सहसम्बन्ध		अ-रेखीय सहसम्बन्ध	
X	**Y**	**X**	**Y**
20	50	50	10
40	100	55	12
60	150	60	22
80	200	90	34
100	250	98	45
120	300	120	56

रेखीय तथा अ-रेखीय सम्बन्ध रखने वाले चरों का विक्षेप चित्र अथवा बिन्दुरेखीय चित्र का स्वरूप इस प्रकार होगा–



यहाँ यह स्पष्ट करना उचित होगा कि व्यवहार में अधिकांश चरों में अ-रेखीय सम्बन्ध पाया जाता है, परन्तु अ-रेखीय सहसम्बन्ध तकनीक अत्यधिक जटिल होने के कारण, हमारी यह मान्यता रहती है कि विचारगत चरों के मध्य रेखीय सम्बन्ध है।

अनेक अनुसंधान कार्यों में दो से अधिक चरों (More than two variables) के लिए किसी प्रतिदर्श से समंक संकलित किये जाते हैं। इस प्रकार के समंकों को बहु-चरीय समंक (Multi-Variate Data) कहते हैं। अनुसंधानकर्ता को कभी-कभी इस प्रकार के बहु-चरीय समंकों का एक साथ विश्लेषण करके परिणाम ज्ञात करने होते हैं क्योंकि ऐसी स्थिति में समंक विश्लेषण के कार्य में एक साथ दो से अधिक चर सम्मिलित होते हैं इसलिए इस प्रकार के समंक विश्लेषण को बहु-चरीय समंक विश्लेषण (Multi-Variate Data Analysis) कहते हैं। विभिन्न प्रकार की परिस्थितियों के लिए विभिन्न बहु-चरीय समंक विश्लेषण विधियों का प्रतिपादन सांख्यिकीविदों के द्वारा किया जा चुका है। यहाँ बहु-चरीय समंकों के सहसम्बन्धात्मक विश्लेषण (Correlational Analysis) के लिए प्रयुक्त की जाने वाली प्रमुख बहु-चरीय सहसम्बन्धात्मक विधियों (Multivariate Corrlational Methods) अर्थात आंशिक सहसम्बन्ध (Partial Correlation) तथा बहु-सहसम्बन्ध (Multiple Correlation) की चर्चा की जा रही है। जब किसी प्रतिदर्श के लिए कम से कम तीन अथवा अधिक चरों के लिए प्राप्तांक उपलब्ध होते हैं अथवा उन चरों के सभी सम्भव दो-दो युग्मों के बीच सहसम्बन्ध गुणांक ज्ञात होते हैं तब आंशिक सहसम्बन्ध तथा बहु-सहसम्बन्ध गुणांक ज्ञात किये जा सकते हैं।

4.3 आंशिक सहसम्बन्ध गुणांक (Partial Correlation Coefficient)

दो चरों के मध्य सहसम्बन्ध गुणांक एक चर के घटने या बढ़ने पर दूसरे चर के घटने-बढ़ने की प्रवृत्ति को बताता है परन्तु दो चरों के मध्य परिगणित सहसम्बन्ध गुणांक का मान भ्रामक (Misleading) भी हो सकता है। कभी-कभी दोनों चरों के मध्य वास्तव में सहसम्बन्ध नगण्य अथवा शून्य होता है परन्तु दोनों चरों के किसी तीसरे चर पर निर्भर (Depend) करने के कारण उन दोनों चरों के बीच परिमित या उच्च सहसम्बन्ध गुणांक प्राप्त हो जाता है। जैसे ऊँचाई (Height), भार (Weight,) शारीरिक क्षमता (Physical Strength), शब्द भण्डार (Vocabulary), मानसिक योग्यता (Mental Ability), सामान्य ज्ञान (General Knowledge), जूते का आकार (Size of Shoe), पोशाक पर किया व्यय (Money Spent on Dresses), विषम लिंगी आकर्षण (Interest in Opposite Sex) आदि कुछ ऐसे चर हैं जो दस वर्ष से लेकर बीस वर्ष की आयु तक आयु (Age) के साथ-साथ बढ़ते हैं यदि आयु की दृष्टि से पर्याप्त प्रसार (Variation) वाले किसी प्रतिदर्श के लिए इनमें से किन्हीं भी दो चरों के मध्य सहसम्बन्ध गुणांक की गणना की जायेगी तब सम्भवत: उच्च धनात्मक सहसम्बन्ध (High Positive Correlation) गुणांक ही प्राप्त होगा परन्तु यदि लगभग समान आयु वाले प्रयोज्यों से युक्त प्रतिदर्श के लिए ऐसे दो चरों के मध्य सहसम्बन्ध गुणांक ज्ञात किया जायेगा तब सम्भवत: लगभग शून्य (Almost Zero) या नगण्य (Negligible) सहसम्बन्ध गुणांक ही प्राप्त होगा। प्रथम स्थिति में दो चरों जैसे ऊँचाई तथा भार के बीच प्राप्त उच्च धनात्मक सहसम्बन्ध वास्तव में आयु के बढ़ने पर ऊँचाई तथा भार दोनों के बढ़ने के कारण प्राप्त हो रहा है। यद्यपि ऊँचाई तथा भार में किसी विशेष सम्बन्ध होने का औचित्य नहीं है परन्तु प्रतिदर्श में सम्मिलित प्रयोज्यों की आयु में पर्याप्त विभिन्नता होने के कारण ऊँचाई तथा भार के बीच उच्च धनात्मक सहसम्बन्ध प्राप्त हो जाता है जो दूसरी स्थिति में आयु को स्थिर (Constant) करने पर लुप्त (Eliminate) हो जाता है। अत: यदि इस प्रकार के दो चरों के बीच वास्तविक सहसम्बन्ध (Real Correlation) को ज्ञात करना है तब आयु चर (Age Variable) के प्रभाव को समाप्त अथवा नियन्त्रित करना होगा। किन्हीं दो चरों के बीच के सहसम्बन्ध को एक साथ कई अवांछित चर (Undesirable Variables) प्रभावित कर सकते हैं इसलिए दो चरों के बीच वास्तविक सहसम्बन्ध का अध्ययन करने के लिए ऐसे अवांछित चरों के प्रभाव को समाप्त करना आवश्यक हो जाता है। यहाँ पर प्रयुक्त अवांछित चर से तात्पर्य किसी ऐसे चर से नहीं है जो वस्तुत: अवांछित हो वरन् केवल इतना ही तात्पर्य है कि यह चर अध्ययन किये जा रहे दो चरों के बीच के सहसम्बन्ध को प्रभावित कर रहा है तथा सहसम्बन्ध को समझने के लिए इसके प्रभाव को समाप्त करना आवश्यक है। वस्तुत: जब तक दो चरों के बीच का सहसम्बन्ध गुणांक अन्य अवांछित चरों (Undesirable Variables) के प्रभाव से मुक्त नहीं होगा तब तक उसकी सही ढंग से व्याख्या करना सम्भव नहीं हो सकेगा।

अन्य चरों के प्रभाव से युक्त सहसम्बन्ध गुणांक भ्रामक सहसम्बन्ध को वास्तविक सहसम्बन्ध बताते हुए त्रुटिपूर्ण निष्कर्ष प्रदान कर सकता है। जिन चरों के बीच वास्तविक सहसम्बन्ध गुणांक ज्ञात करना होता है उन्हें प्राथमिक चर (Primary Variables) कहते हैं तथा जिन अवांछित चरों के प्रभाव को समाप्त या नियन्त्रित करना होता है उन्हें नियन्त्रित चर (Controlled Variables) कहते हैं। बहु-चरीय सहसम्बन्ध विश्लेषण में दो प्राथमिक चरों के मध्य वास्तविक सहसम्बन्ध गुणांक को अवांछित चरों के प्रभाव से मुक्त रखने के लिए अवाँछित चरों को दो ढंग से प्रथम, प्रयोगात्मक ढंग से (Experimentally) तथा द्वितीय, सांख्यिकीय ढंग से (Statistically) नियन्त्रित किया जा सकता है। किसी चर को प्रयोगात्मक ढंग से नियन्त्रित करने के लिए उस चर को स्थिर (Constant) रखा जाता है अर्थात प्रतिदर्श के प्रयोज्यों को इस प्रकार से छांटते हैं कि नियन्त्रित किये जाने वाले चर पर सभी प्रयोज्यों के लगभग समान प्राप्तांक हों। किसी चर को सांख्यिकीय ढंग से नियन्त्रित करने के लिए आंशिक सहसम्बन्ध विधि का प्रयोग किया जाता है। क्योंकि प्रथम विधि में समान प्राप्तांक वाले प्रयोज्यों का चयन करना कठिन होता है इसलिए व्यावहारिक दृष्टि से प्रथम विधि की तुलना में दूसरी विधि को अधिक पसन्द किया जाता है। सांख्यिकीय ढंग से नियन्त्रित करने के लिए चर को स्थिर रखकर जब दो चरों के मध्य सहसम्बन्ध ज्ञात किया जाता है तब इसे आंशिक सहसम्बन्ध गुणांक (Partial Correlation Cofficient) कहते हैं। अत: दो चरों के बीच आंशिक सह-सम्बन्ध गुणांक को किसी तीसरे अथवा अधिक चरों के प्रभाव को समाप्त करने के उपरान्त उन दोनों चरों के बीच दृष्टिगोचित सहसम्बन्ध गुणांक के रूप में परिभाषित किया जा सकता है। जैसे यदि आयु के प्रभाव को समाप्त करने के उपरान्त बालकों की ऊँचाई तथा भार के बीच सहसम्बन्ध गुणांक ज्ञात किया जाये तब इसे आयु को अप्रभावी करने पर प्राप्त ऊँचाई तथा भार के बीच प्राप्त आंशिक सहसम्बन्ध गुणांक कहेंगे। इस आंशिक सहसम्बन्ध गुणांक की व्याख्या आयु की दृष्टि से समजातीय (Homogeneous in Age) अनेक प्रतिदर्शों के लिए ऊँचाई तथा भार के बीच प्राप्त सहसम्बन्ध गुणांकों के औसत के रूप में भी की जा सकती है।

आंशिक सहसम्बन्ध गुणांक केवल एक अथवा अनेक चरों के प्रभाव को समाप्त करके ज्ञात किया जा सकता है। दो चरों के बीच परिगणित किसी आंशिक सहसम्बन्ध गुणांक को जितने अन्य चरों के प्रभावों से मुक्त रखा जाता है वह संख्या उस आंशिक सहसम्बन्ध गुणांक का स्तर (Order) कहलाती है, अत: जब केवल एक चर के प्रभाव को समाप्त किया जाता है तब प्राप्त आंशिक सहसम्बन्ध गुणांक को प्रथम स्तर का आंशिक सहसम्बन्ध गुणांक (First Order Partial Correlation Cofficient) कहते हैं। जब दो चरों के प्रभाव को समाप्त किया जाता है तब दो स्तर का आंशिक सहसम्बन्ध गुणांक (Second Order Partial Correlation Coefficient) प्राप्त होता है। स्पष्ट है कि जितने चरों के प्रभाव से सहसम्बन्ध को मुक्त रखा जाता है उतने ही स्तर (Order) का सहसम्बन्ध गुणांक प्राप्त होता है। सहसम्बन्ध गुणांकों के स्तर (Order) के सम्बन्ध में यह स्पष्ट करना उचित ही होगा कि साधारण सहसम्बन्ध गुणांक (Simple Correlation Coefficient) को शून्य स्तर का सहसम्बन्ध गुणांक (Zero Order Correlation Coefficient) भी कहते हैं। यह तथ्य इस बात से भी स्पष्ट है कि साधारण सहसम्बन्ध में किसी भी अन्य चर के प्रभाव को समाप्त नहीं करते हैं।

आंशिक सहसम्बन्ध गुणांक वास्तव में दो चरों (X_1 तथा X_2) पर प्राप्त अंकों में से तीसरे चर (X_3) जिसके प्रभाव को विलग करना है, के द्वारा स्पष्ट किये जा सकने वाले अंशों को निकालकर बचे अवशिष्ट प्राप्तांकों (Residual Scores) के बीच परिगणित सहसम्बन्ध होता है। अवशिष्ट प्राप्तांकों को निम्न ढंग से प्राप्त किया जा सकता है–

(*i*) चर X_1 के लिए अवशिष्ट प्राप्तांक

$$d_1 = X_1 - X'_1$$

जहाँ $X'_1 =$ चर X_3 की सहायता से चर X_1 का पूर्वकथित मान, जिसे निम्न समीकरण से ज्ञात किया जा सकता है–

$$X'_1 = M_1 + r_{13} \frac{\sigma_1}{\sigma_3} (X_3 - M_3)$$

(*ii*) चर X_2 के लिए अवशिष्ट प्राप्तांक

$$d_2 = X_2 - X'_2$$

जहाँ X_2 = चर X_3 की सहायता से चर X_2 का पूर्वकथित मान, जिसे निम्न समीकरण से ज्ञात किया जा सकता है–

$$X'_2 = M_2 + r_{23} \frac{\sigma_2}{\sigma_3}(X_3 - M_3)$$

परन्तु अवशिष्ट प्राप्तांक (Residual Scores) प्राप्त करके उनके बीच सहसम्बन्ध गुणांक की गणना करके आंशिक सहसम्बन्ध गुणांक ज्ञात करना एक जटिल कार्य है। व्यवहार में ऐसा करने की आवश्यकता नहीं है। वस्तुत: आंशिक सहसम्बन्ध गुणांक की गणना विभिन्न चरों के बीच शून्य स्तरीय साधारण सहसम्बन्ध गुणांकों के ज्ञात होने पर सरलता से की जा सकती है। प्रथम स्तर के आंशिक सहसम्बन्ध गुणांक (First Order Partial Correlation) की गणना करने का सूत्र निम्नवत है–

$$r_{12.3} = \frac{r_{12} - r_{13}\, r_{23}}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}}$$

जहाँ $r_{12.3}$ = चर 3 के प्रभाव को समाप्त करने पर चर 1 तथा चर 2 के बीच अवशिष्ट आंशिक सहसम्बन्ध गुणांक

r_{12} = चर 1 तथा चर 2 के बीच साधारण सहसम्बन्ध गुणांक

r_{13} = चर 1 तथा चर 3 के बीच साधारण सहसम्बन्ध गुणांक

r_{23} = चर 2 तथा चर 3 के बीच साधारण सहसम्बन्ध गुणांक

इसी प्रकार से $r_{13.2}$ तथा $r_{23.1}$ के सूत्र अग्रांकित होंगे–

$$r_{13.2} = \frac{r_{13} - r_{12}\, r_{32}}{\sqrt{(1-r_{12}^2)(1-r_{32}^2)}}$$

$$r_{23.1} = \frac{r_{23} - r_{21}\, r_{31}}{\sqrt{(1-r_{21}^2)(1-r_{31}^2)}}$$

यदि चरों को 1, 2, 3 के स्थान पर x, y तथा z से प्रदर्शित करें तब प्रथम स्तरीय आंशिक सहसम्बन्ध गुणांक के सूत्र होंगे–

$$r_{xy.z} = \frac{r_{xy} - r_{xz}\, r_{yz}}{\sqrt{(1-r_{xz}^2)(1-r_{yz}^2)}}$$

$$r_{xz.y} = \frac{r_{xz} - r_{xy}\, r_{zy}}{\sqrt{(1-r_{xy}^2)(1-r_{zy}^2)}}$$

$$r_{yz.x} = \frac{r_{yz} - r_{yx}\, r_{zx}}{\sqrt{(1-r_{yx}^2)(1-r_{zx}^2)}}$$

द्वि स्तरीय आंशिक सहसम्बन्ध गुणांक (Second Order Partial Correlation Cofficient) को निम्न सूत्र से ज्ञात किया जा सकता है–

$$r_{12.34} = \frac{r_{12.3} - r_{14.3}\, r_{24.3}}{\sqrt{(1-r_{14.3}^2)(1-r_{24.3}^2)}}$$

जहाँ $r_{12.34}$ = चर 3 तथा चर 4 के प्रभाव को समाप्त करने पर चर 1 तथा चर 2 के बीच अवशिष्ट आंशिक सहसम्बन्ध गुणांक

नोट

$r_{12.3}$ = चर 3 के प्रभाव को समाप्त करने पर चर 1 तथा चर 2 के बीच अवशिष्ट आंशिक सहसम्बन्ध गुणांक

$r_{14.3}$ = चर 3 के प्रभाव को समाप्त करने पर चर 1 तथा चर 4 के बीच अवशिष्ट आंशिक सहसम्बन्ध गुणांक

$r_{24.3}$ = चर 3 के प्रभाव को समाप्त करने पर चर 2 तथा 4 के बीच अवशिष्ट आंशिक सहसम्बन्ध गुणांक

$r_{12.3}$, $r_{14.3}$ तथा $r_{24.3}$ की गणना पूर्ववत प्रस्तुत प्रथम स्तरीय आंशिक सहसम्बन्ध गुणांक के सूत्रों की सहायता से की जा सकती है।

इसी प्रकार से किसी भी स्तर (Order) के आंशिक सहसम्बन्ध गुणांक को उससे एक कम स्तर के आंशिक सहसम्बन्ध गुणांकों (Next lower Order Partial Correlation Coffcients) की सहायता से ज्ञात किया जा सकता है। अत: यदि $n = (k-2)$ तब nवें स्तर (nth order) के आंशिक सहसम्बन्ध गुणांक की गणना के लिए प्रयुक्त किया जाने वाला सूत्र होगा–

$$r_{12.34......k} = \frac{r_{12.345......(k-1)} - r_{1k.345......(k-1)}\ r_{2k.345......(k-1)}}{\sqrt{\left(1-r^2_{1k.3345......(k-1)}\right)\left(1-r^2{}_{2k.345......(k-1)}\right)}}$$

आंशिक सहसम्बन्ध गुणांक को इंगित करने वाले संकेताक्षरों के अवलोकन से यह स्पष्ट हो गया होगा कि आंशिक सहसम्बन्ध गुणांक को r के नीचे चरों को इंगित करने वाले अंक या अक्षर प्रत्यय (Suffix) की तरह से लगाकर लिखते हैं। बिन्दु (Point) से पूर्व के दो अंक या अक्षर उन चरों को इंगित करते हैं जिनके मध्य अवशिष्ट सहसम्बन्ध ज्ञात करना है जबकि बाद के अंक या अक्षर उन चरों को इंगित करते हैं जिनके प्रभाव को समाप्त किया गया है। स्पष्ट है कि बिन्दु के बाद लिखे अंकों अथवा अक्षरों की संख्या आंशिक सहसम्बन्ध गुणांक के स्तर (Order) की द्योतक होती है। बिन्दु के पूर्व तथा बाद में लिखे चरों के क्रम का कोई विशेष महत्त्व नहीं होता है जैसे द्वि-स्तरीय आंशिक सहसम्बन्ध की स्थिति में

$$r_{12.34} = r_{12.43} = r_{21.34} = r_{21.43}$$

जैस-जैसे आंशिक सहसम्बन्ध गुणांक का स्तर (Order) बढ़ता जाता है वैसे-वैसे उसकी व्याख्या करना जटिल होता जाता है। यही कारण है कि प्राय: केवल प्रथम अथवा द्वितीय स्तरीय आंशिक सहसम्बन्ध गुणांक की गणना की जाती है। प्रथम तथा द्वितीय स्तरीय आंशिक सहसम्बन्ध गुणांकों की गणना विधि को आगे दिये उदाहरणों से स्पष्ट किया गया है। उच्च स्तरीय आंशिक सहसम्बन्ध गुणांक (High Order Partial Correlation Coefficient) की गणना में रुचि रखने वाले पाठक सूत्रों का प्रयोग करके गणना कार्य स्वयं कर सकेंगे, ऐसी आशा है।

उदाहरण–यदि $r_{12} = .45, r_{13} = -.40$ तथा $r_{23} = .62$ हो तो $r_{12\cdot3}$ की गणना कीजिये।

हल–ज्ञात है–

$$r_{12} = +.45$$
$$r_{13} = -.40$$
$$r_{23} = +.62$$

क्योंकि $r_{12\cdot3}$ की गणना का सूत्र है–

$$r_{12.3} = \frac{r_{12} - r_{13}\, r_{23}}{\sqrt{\left(1-r_{13}^2\right)\left(1-r_{23}^2\right)}}$$

$$r_{12.3} = \frac{.45 - (-.40)\times(.62)}{\sqrt{[1-(-.40)^2]\,[1-(.62)^2]}}$$

$$= \frac{.698}{.7191} = .97$$

उदाहरण–200 छात्रों के एक प्रतिदर्श के लिए ऊँचाई, भार तथा आयु के लिए साधारण सहसम्बन्ध गुणांक निम्नवत थे। आयु के प्रभाव को निकाल देने पर ऊँचाई तथा भार के बीच अवशिष्ट सहसम्बन्ध गुणांक क्या होगा?

ऊँचाई तथा भार के बीच सहसम्बन्ध गुणांक = .68

ऊँचाई तथा आयु के बीच सहसम्बन्ध गुणांक = .80

भार तथा आयु के बीच सहसम्बन्ध गुणांक = .75

हल–यदि ऊँचाई को X संकेताक्षर से, भार को Y संकेताक्षर से तथा आयु को Z संकेताक्षर से इंगित करें तब

$r_{xy} = .68, \quad r_{xz} = .80, \quad r_{yz} = .75$

क्योंकि आयु (Z) के प्रभाव से मुक्त ऊँचाई (X) तथा भार (Y) के बीच सहसम्बन्ध को ज्ञात करना है इसलिए $r_{xy.z}$ की गणना करनी होगी अत:

$$r_{xy.z} = \frac{r_{xy} - r_{xz}\, r_{yz}}{\sqrt{\left(1-r_{xz}^2\right)\left(1-r_{yz}^2\right)}}$$

$$= \frac{.68 - .80 \times .75}{\sqrt{(1-(.80)^2)\,(1-(.75)^2)}}$$

$$= \frac{.08}{.3969} = .202$$

अत: आयु के प्रभाव को समाप्त करने पर ऊँचाई तथा भार के मध्य आंशिक सहसम्बन्ध गुणांक का मान .202 होगा।

संगतता आवश्यकता

(Consistency Requirement)

दो चरों के बीच साधारण सहसम्बन्ध गुणांक तथा आंशिक सहसम्बन्ध गुणांक में परस्पर सम्बन्ध अनिश्चित रहता है। यद्यपि प्राय: आंशिक सहसम्बन्ध गुणांक का मान साधारण सहसम्बन्ध गुणांक से कम प्राप्त होता है परन्तु आंशिक सहसम्बन्ध गुणांक का मान साधारण सहसम्बन्ध गुणांक के मान से अधिक भी प्राप्त हो सकता है। कभी-कभी तो दो चरों के बीच साधारण सहसम्बन्ध गुणांक तथा उन्हीं दो चरों के बीच आंशिक सहसम्बन्ध गुणांक के मान विपरीत चिन्हों (Opposite Signs) वाले भी प्राप्त हो जाते हैं। इन तीनों ही स्थितियों को एक उदाहरण से स्पष्ट किया जा सकता है। यदि $r_{12} = .90$, $r_{13} = .60$ तथा $r_{23} = .30$ हो तब $r_{12.3}$ का मान .791, $r_{13.2}$ का मान .794 तथा $r_{23.1}$ का मान –.688 प्राप्त होगा। स्पष्ट है कि $r_{12.3}$ का मान r_{12} के मान से कम है, तथा $r_{13.2}$ का मान r_{13} के मान से अधिक है जबकि $r_{23.1}$ तथा r_{23} के चिन्ह परस्पर विपरीत हैं। आंशिक सहसम्बन्ध गुणांक के इस प्रकार से प्राप्त होने पर उसकी व्याख्या करने में कठिनाई आती है। परन्तु कभी-कभी तो आंशिक सहसम्बन्ध गुणांक का मान 1.00 से अधिक प्राप्त हो जाता है जो कि सहसम्बन्ध गुणांक के प्रत्यय के विपरीत होता है। जैसे यदि $r_{12} = .80$, $r_{13} = .70$, तथा $r_{23} = –.20$ हो तब $r_{12.3}$ का मान 1.343 प्राप्त होगा जो कि सहसम्बन्ध के प्रत्यय के अनुरूप नहीं है। वस्तुत: प्रथम स्तरीय सहसम्बन्ध गुणांक (First Order Partial Correlation) की गणना के लिए यह आवश्यक है कि तीनों साधारण सहसम्बन्ध गुणांकों में परस्पर इस प्रकार का सम्बन्ध होना चाहिए कि कोई भी आंशिक सहसम्बन्ध गुणांक का मान 1.00 से अधिक न आ सकें। तीन सहसम्बन्ध गुणांकों में इस प्रकार के सम्बन्ध को संगतता सम्बन्ध (Consistency Relationship) कहते हैं। इस संगतता सम्बन्ध के लिए निम्न शर्त पूरी होनी चाहिए–

$$1 - r_{12}^2 - r_{13}^2 - r_{23}^2 + 2\, r_{12}\, r_{13}\, r_{23} > 0$$

यदि तीन सहसम्बन्ध गुणांक इस शर्त को पूरा करते हैं तब ऐसे तीन सहसम्बन्ध गुणांकों को संगत (Consistent) कहते हैं। परन्तु यदि तीनों सहसम्बन्ध गुणांक इस शर्त को पूरा नहीं करते हैं तब इन्हें असंगत (Inconsistent) कहते हैं। पीछे वर्णित उदाहरण में $r_{12} = .80$, $r_{13} = .70$, तथा $r_{23} = -.20$ के लिए–

नोट

$$1 - r_{12}^2 - r_{13}^2 - r_{23}^2 + 2\, r_{12}\, r_{13}\, r_{23}$$
$$= 1 - .80^2 - .70^2 - (-.20)^2 + 2 \times .80 \times .70 \times (-.20)$$
$$= -.394$$

क्योंकि – .394 का मान शून्य से अधिक नहीं है इसलिए ये तीनों सहसम्बन्ध गुणांक संगतता की शर्त को पूरा नहीं कर पाने के कारण असंगत (Inconsistent) हैं एवं इनके लिए आंशिक सहसम्बन्ध गुणांकों की गणना करना सम्भव नहीं है।

आंशिक सहसम्बन्ध गुणांक की सीमाएँ
(Limitations of Partial Correlation Coefficient)

किन्हीं भी दो चरों के बीच आंशिक सहसम्बन्ध गुणांक (Partial Correlation Coefficient) निम्नलिखित सीमाओं से युक्त होता है–

(*i*) केवल रेखीय सहसम्बन्ध के कारण आ रहे प्रभाव को समाप्त किया जा सकता है। वक्रीय सहसम्बन्ध के प्रभाव को समाप्त करना सम्भव नहीं होता है इसलिए यह आवश्यक है कि आंशिक सहसम्बन्ध गुणांक की गणना में प्रयुक्त सभी शून्य स्तरीय (zero order) सहसम्बन्ध गुणांकों की गणना के लिए प्रयुक्त समंक रेखीय ढंग से वितरित हो।

(*ii*) आंशिक सहसम्बन्ध गुणांक को पूर्णरुपेण स्पष्ट करना प्राय: सम्भव नहीं हो पाता है।

(*iii*) साधारण सहसम्बन्ध गुणांक के समान आंशिक सहसम्बन्ध गुणांक भी कार्य-कारण सम्बन्ध (Cause-Effect Relationship) प्रस्तुत नहीं करता है।

(*iv*) सभी चरों के प्रभाव को समाप्त करना सम्भव नहीं हो पाता है।

(*v*) संगतता शर्त (Consistency Condition) के पूरा न होने पर आंशिक सहसम्बन्ध गुणांक का मान 1.00 से अधिक आ सकता है।

आंशिक सहसम्बन्ध गुणांक की सार्थकता
(Significane of Partial Correlation Coefficient)

साधारण सहसम्बन्ध गुणांक की सार्थकता के लिए प्रयुक्त की जाने वाली सभी विधियों को आंशिक सहसम्बन्ध गुणांक की सार्थकता की जाँच के लिए प्रयुक्त किया जा सकता है। दोनों में अन्तर केवल मुक्तांशों (*df*) के निर्धारण का है। आंशिक सहसम्बन्ध गुणांक की सार्थकता की जाँच के समय नियन्त्रित किये गये प्रत्येक चर के लिए एक-एक मुक्तांश (*df*) और कम होता जाता है। क्योंकि साधारण सहसम्बन्ध गुणांक की मानक त्रुटि $(1 - r^2)/\sqrt{(n-2)}$ होती है। अत: प्रथम स्तरीय (First Order) आंशिक सहसम्बन्ध की मानक त्रुटि ज्ञात करने का सूत्र निम्नवत होगा–

$$\sigma_{r_{12.3}} = \frac{1 - r_{12.3}^2}{\sqrt{N - 3}}$$

तथा प्रथम स्तर के आंशिक सहसम्बन्ध गुणांक ($r_{12.3}$) की सार्थकता के लिए प्रयुक्त टी-परीक्षण के लिए टी अनुपात का मान ज्ञात करने का सूत्र होगा–

$$t = \frac{r_{12.3}\sqrt{N - 3}}{\sqrt{1 - r_{12.3}^2}}$$

इस टी-अनुपात की सार्थकता $df = N - 3$ पर देखी जायेगी इसी प्रकार से $(K - 2)$ वें स्तर के आंशिक सहसम्बन्ध गुणांक की मानक त्रुटि तथा टी-अनुपात के सूत्र होंगे–

$$\sigma_{r_{12.34......k}} = \frac{1 - r^2_{12.3......k}}{\sqrt{N - K}}$$

तथा
$$t = \frac{r_{12.34......k}\sqrt{N - K}}{\sqrt{1 - r^2_{12.34......K}}} \qquad (df = N - K)$$

परन्तु साधारण सहसम्बन्ध गुणांक की सार्थकता के निर्धारण के समान आंशिक सहसम्बन्ध गुणांक की सार्थकता का अधिक सही ढंग से निर्धारण फिशर के जेड फलन प्रत्यावर्तन (Fisher's Z Conversion) के द्वारा किया जा सकता है। क्योंकि साधारण सहसम्बन्ध गुणांक के लिए प्रत्यावर्तित जेड की मानक त्रुटि $1/(N - 3)$ होती है इसलिए प्रथम स्तरीय आंशिक सहसम्बन्ध के अनुरूप प्रत्यावर्तित जेड की मानक त्रुटि का सूत्र निम्नवत होगा–

$$\sigma_Z = \frac{1}{\sqrt{N - 4}}$$

इसी प्रकार से k वें स्तर के आंशिक सहसम्बन्ध (Partial Correlation Coefficient) के लिए प्रत्यावर्तित जेड की मानक त्रुटि ज्ञात करने का सूत्र होगा–

$$\sigma_Z = \frac{1}{\sqrt{N - K - 3}}$$

आंशिक सहसम्बन्ध गुणांक की सार्थकता की विधि निम्न उदाहरणों से स्पष्ट हो सकेगी।

उदाहरण–100 प्रयोज्यों के एक प्रतिदर्श के लिए $r_{12.3} = .30$ प्राप्त हुआ। क्या यह आंशिक सहसम्बन्ध गुणांक सार्थक है?

हल–ज्ञात है $N = 100$ तथा $r_{12.3} = .30$

प्रथम स्तरीय (First Order) आंशिक सहसम्बन्ध की सार्थकता की जाँच के लिए प्रयुक्त टी परीक्षण में टी अनुपात की गणना निम्न सूत्र से की जायेगी–

$$t = \frac{r_{12.3}\sqrt{N - 3}}{\sqrt{1 - r^2_{12.3}}}$$

$$= \frac{.30\sqrt{100 - 3}}{\sqrt{(1 - .30^2)}}$$

$$= \frac{2.9547}{.9539}$$

$$= 3.097$$

परिगणित $t = 3.097$ के लिए $df = 97$ है। अत: $df = 97$ पर टी-सारणी देखने पर स्पष्ट है कि $t_{.05} = 1.98$ तथा $t_{.01} = 2.63$ है। क्योंकि परिगणित $t = 3.097$ इन दोनों ही सारणी मानों से अधिक है। अत: यह .01 स्तर पर सार्थक है। इसलिए .01 स्तर पर सार्थक इस टी-अनुपात के आधार पर कहा जा सकता है कि $r_{12.3} = .30$ का मान .01 स्तर पर सार्थक है।

अर्द्ध आंशिक सहसम्बन्ध गुणांक
(Semi-Partial Correlation Coefficient)

आंशिक सहसम्बन्ध गुणांक के सैद्धान्तिक विवेचन में स्पष्ट किया जा चुका है कि आंशिक सहसम्बन्ध ज्ञात करते समय दोनों ही चरों से किसी तीसरे चर (अथवा कई चरों) के प्रभाव को विलग किया जाता है। दूसरे शब्दों

में कहा जा सकता है कि दोनों ही चरों के प्राप्तांकों में से तीसरे चर के प्रभाव वाले अंश को घटाने पर प्राप्त अवशिष्ट प्राप्तांकों के बीच जो सहसम्बन्ध गुणांक प्राप्त होता है, उसे ही आंशिक सहसम्बन्ध गुणांक कहते हैं। यदि केवल एक ही चर से तीसरे चर के प्रभाव को हटाये तथा दूसरे चर को यथावत रहने दे, तब प्राप्त सहसम्बन्ध गुणांक को अर्द्ध-आंशिक सहसम्बन्ध गुणांक (Semi-Partial Correlation Coefficient) कहते हैं। कुछ विद्वान इसे अंश सहसम्बन्ध गुणांक (Part Correlation Coefficient) के नाम से भी सम्बोधित करते हैं। अर्द्ध-आंशिक सहसम्बन्ध गुणांक को प्राय: $r_{1(2.3)}$ संकेताक्षर से लिखते हैं जो प्रथम चर का द्वितीय चर के साथ अर्द्ध-आंशिक सहसम्बन्ध गुणांक को इंगित करता है। तृतीय चर के प्रभाव को केवल द्वितीय चर से विलग किया गया है। स्पष्ट है कि कोष्ठक में लिखी प्रथम संख्या वाले चर से उससे आगे की संख्या वाले चर (चरों) के प्रभाव को विलग किया गया है। आंशिक सहसम्बन्ध वाले खण्ड में प्रयुक्त अवशिष्ट प्राप्तांकों का यदि सन्दर्भ लिया जाये तब कहा जा सकता है कि $r_{1(2.3)}$ वास्तव में X_1 तथा d_2 के बीच प्राप्त सहसम्बन्ध गुणांक है। अर्द्ध-आंशिक सहसम्बन्ध गुणांक की गणना निम्न सूत्र से की जा सकती है–

$$r_{1(2.3)} = \frac{r_{12} - r_{13} \cdot r_{23}}{\sqrt{1 - r_{23}^2}}$$

यदि $r_{1(2.3)}$ के सूत्र की तुलना $r_{12.3}$ के सूत्र से की जाये तब स्पष्ट हो सकेगा कि अर्द्ध-आंशिक सहसम्बन्ध गुणांक $r_{1(2.3)}$ के सूत्र के हर में एक पद $\sqrt{1 - r_{13}^2}$ नहीं है, शेष सभी पद यथावत हैं। क्योंकि $\sqrt{1 - r_{13}^2}$ का मान सदैव ही एक से कम प्राप्त होगा इसलिए कहा जा सकता है कि $r_{1(2.3)}$ का मान $r_{12.3}$ के मान से सदैव ही कम प्राप्त होता है। इसी प्रकार से अधिक स्तर वाले अर्द्ध आंशिक सहसम्बन्ध गुणांकों की गणना की जा सकती है। अर्द्ध-आंशिक सहसम्बन्ध गुणांक का बहुप्रतीपगमन में विशेष महत्त्व है। चार स्वतन्त्र चरों के लिए किसी बहुसहसम्बन्ध गुणांक के वर्ग को अर्द्ध-आंशिक सहसम्बन्ध गुणांकों के रूप में निम्न ढंग से लिखा जा सकता है–

$$R_{1.2345}^2 = r_{12}^2 + r_{1(3.2)}^2 + r_{1(4.23)}^2 + r_{1(5.234)}^2$$

जहाँ r_{12}^2 = चर 1 तथा 2 के बीच उभयनिष्ठ प्रसरण

$r_{1(3.2)}^2$ = चर 2 के साथ सहभागी प्रसरण को घटा कर, चर 1 तथा 3 के बीच अवशिष्ट उभयनिष्ट प्रसरण

$r_{1(4.23)}^2$ = चर 2 तथा 3 के साथ सहभागी प्रसरण को घटाकर चर 1 तथा 4 अवशिष्ट उभयनिष्ठ प्रसरण

$r_{1(5.234)}^2$ = चर 2, 3 तथा 4 के साथ सहभागी प्रसरण को घटाकर चर 1 तथा 5 के बीच अवशिष्ट उभयनिष्ठ प्रसरण

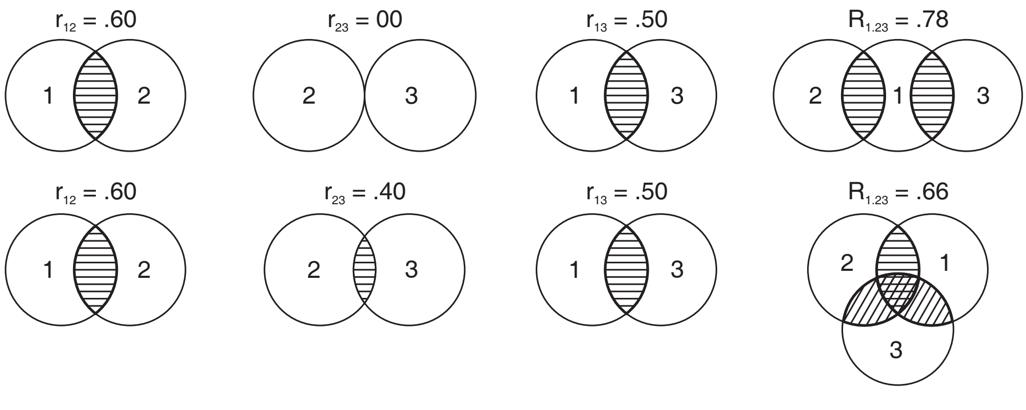
बहु-सहसम्बन्ध गुणांक की चर्चा आगे के पृष्ठों पर विस्तार से की जा रही है।

4.4 बहु-सहसम्बन्ध गुणांक (Multiple Correlation Coefficient)

जिस चर के मान का पूर्वकथन किया जाता है। उसे आश्रित चर (Dependent Variable) अथवा निकष चर (Criterion Variable) कहते हैं तथा जिस चर का मान ज्ञात होता है उसे स्वतन्त्र चर (Independent Variable) अथवा पूर्वकथक (Predictor) के नाम से सम्बोधित किया जाता है। स्वतन्त्र चर के आधार पर आश्रित चर के पूर्वकथन की परिमार्जितता (Precision) दोनों चरों के मध्य सहसम्बन्ध गुणांक (r) के मान पर निर्भर करती है। वस्तुत: पूर्वकथन की परिमार्जितता दोनों चरों के मध्य उभयनिष्ठ प्रसरण (Common Variance)

पर निर्भर करती है जो कि उनके बीच के सहसम्बन्ध गुणांक (r) के वर्ग के बराबर होता है। जैसे यदि बुद्धि तथा शैक्षिक निष्पत्ति के मध्य .60 का सहसम्बन्ध हो तब बुद्धि तथा शैक्षिक निष्पत्ति प्राप्तांकों के बीच 36% प्रसरण उभयनिष्ठ होगा जो बुद्धि के ज्ञात होने पर शैक्षिक निष्पत्ति के पूर्वकथन की परिमार्जितता का भी द्योतक होगा। यही कारण है कि दोनों चरों के मध्य सहसम्बन्ध गुणांक (r) का मान (चाहे धनात्मक हो अथवा ऋणात्मक हो) जितना अधिक होता है पूर्वकथन की परिमार्जितता उतनी ही अधिक होती है।

पूर्वकथन को अधिक परिमार्जितता बनाने के लिए कभी-कभी एक से अधिक स्वतन्त्र चरों अथवा पूर्वकथनों का प्रयोग किया जाता है क्योंकि अभिप्रेरणा तथा शैक्षिक निष्पत्ति भी परस्पर सह सम्बन्धित होते हैं इसलिए शैक्षिक निष्पत्ति के पूर्वकथन के लिए बुद्धि के साथ-साथ अभिप्रेरणा चर का भी प्रयोग किया जा सकता है उदाहरणार्थ यदि शैक्षिक निष्पत्ति व बुद्धि के बीच .60 का तथा शैक्षिक निष्पत्ति व अभिप्रेरणा के मध्य .50 का सहसम्बन्ध हो तब शैक्षिक निष्पत्ति व बुद्धि चरों के बीच 36% प्रसरण उभयनिष्ठ होगा जबकि शैक्षिक निष्पत्ति व अभिप्रेरणा चरों के बीच 25% प्रसरण उभयनिष्ठ होगा। यदि बुद्धि तथा अभिप्रेरणा के बीच कोई सहसम्बन्ध न हो तब इन दोनों पूर्व-कथकों तथा शैक्षिक निष्पत्ति के बीच 36% + 25% = 61% प्रसरण उभयनिष्ठ होगा। इस 61% उभयनिष्ठ प्रसरण के आधार पर कहा जा सकता है कि शैक्षिक निष्पत्ति का बुद्धि व अभिप्रेरणा के साथ संयुक्त सहसम्बन्ध (Combined Correlation) का मान $\sqrt{.61} = .78$ है। किसी एक चर का दो या दो से अधिक चरों के साथ इस प्रकार से संयुक्त सहसम्बन्ध को ही बहु-सहसम्बन्ध (Multiple Correlation) कहा जाता है। अत: बहु-सहसम्बन्ध गुणांक किसी आश्रित चर (Dependent Variable) तथा दो या दो से अधिक स्वतन्त्र चरों (Independent Variables) या निकष चर (Criterion Variable) या पूर्वकथकों (Predectors) के बीच सहसम्बन्ध की मात्रा को बताता है। परन्तु व्यवहार में बहु-सहसम्बन्ध गुणांक की गणना इतनी सरल नहीं है क्योंकि किन्ही भी दो स्वतन्त्र चरों या पूर्वकथकों के बीच सहसम्बन्ध गुणांक का मान शून्य नहीं होता है। दो स्वतन्त्र चरों के बीच सहसम्बन्ध के शून्य न होने पर आश्रित चर तथा दोनों स्वतन्त्र चरों के बीच के उभयनिष्ठ प्रसरणों में कुछ अतिच्छादन (Overlapping) होती हैं जिसके कारण आश्रित चर तथा दोनों स्वतन्त्र चरों के बीच उभयनिष्ठ प्रसरणों का योग करके आश्रित चर तथा स्वतन्त्र चरों के बीच कुल उभयनिष्ठ प्रसरण को ज्ञात कर पाना सम्भव नहीं हो पाता है। परिणामत: बहु-सहसम्बन्ध गुणांक की गणना कठिन हो जाती है। अतिच्छादन (Overlapping) के कारण होने वाली इस कठिनाई को चित्र 96 के अवलोकन से अधिक अच्छी तरह से समझा जा सकता है।



चित्र
विभिन्न चरों के बीच उभनिष्ठ प्रसरण का रेखाचित्रीय प्रदर्शन

स्वतन्त्र चरों के बीच परस्पर सहसम्बन्ध होने पर आश्रित चर का कई स्वतन्त्र चरों से संयुक्त सहसम्बन्ध ज्ञात करने के लिए बहु-सहसम्बन्ध गुणांक की गणना के लिए विशिष्ट सूत्रों का प्रतिपादन किया गया है जिनकी सहायता से बहु-सहसम्बन्ध गुणांक की गणना सरलता से की जा सकती है। बहु-सहसम्बन्ध समस्या में कुल तीन

चर अर्थात् एक आश्रित चर (X) तथा दो स्वतन्त्र चर (Y तथा Z) होने पर बहु-सहसम्बन्ध गुणांक की गणना निम्न सूत्र से की जा सकती है–

नोट

$$R_{X.YZ} = \sqrt{\frac{r_{XY}^2 + r_{XZ}^2 \; 2r_{XY} \; r_{XZ} \; r_{YZ}}{1 - r_{YZ}^2}}$$

जहाँ $R_{X.YZ}$ = आश्रित चर (X) का स्वतन्त्र चर (Y) तथा स्वतन्त्र चर (Z) के साथ बहु-सहसम्बन्ध गुणांक

r_{XY} = चर X तथा चर Y के बीच साधारण सहसम्बन्ध गुणांक

r_{XZ} = चर X तथा चर Z के बीच साधारण सहसम्बन्ध गुणांक

r_{YZ} = चर Y तथा चर Z के बीच साधारण सहसम्बन्ध गुणांक

क्योंकि स्वतन्त्र चरों की संख्या दो से अधिक भी हो सकती है इसलिए बहु-सहसम्बन्ध की समस्याओं में आश्रित चर को 1 से तथा स्वतन्त्र चरों को 2, 3, 4....आदि अंकों से इंगित करना अधिक अच्छा समझा जाता है। तब बहु-सहसम्बन्ध गुणांक का उपरोक्त सूत्र निम्नवत हो जायेगा–

$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12} \; r_{13} \; r_{23}}{1 - r_{23}^2}}$$

कभी-कभी आश्रित चर को Y से तथा स्वतन्त्र चरों को X_1, X_2....आदि से इंगित करते हैं तब उपरोक्त सूत्र होगा–

$$R_{Y.X_1X_2} = \sqrt{\frac{r_{yx_1}^2 + r_{yx_2}^2 - 2r_{yx_1} \; r_{yx_2} \; r_{x_1x_2}}{1 - r_{x_1x_2}^2}}$$

उपरोक्त विवेचन से स्पष्ट है कि बहु-सहसम्बन्ध गुणांक (Multiple Correlation Coefficient) को R संकेताक्षर से लिखते हैं। संकेताक्षर R के साथ प्रयुक्त प्रत्यय (Suffix) में बिन्दु से पूर्व का अंक या अक्षर आश्रित चर को इंगित करता है जबकि बाद के अंक या अक्षर स्वतन्त्र चरों को इंगित करते हैं। जैसे $R_{1.2345}$ में आश्रित चर को 1 से तथा स्वतन्त्र चरों को 2, 3, 4 व 5 से इंगित किया गया है। कभी-कभी आश्रित चर को कोष्ठक में बन्द करके इंगित करते हैं। जैसे $R_{(x)yz}$ में आश्रित चर X है जबकि $R_{(1)23}$ में आश्रित चर को अंक 1 से इंगित किया गया है। बहु-सहसम्बन्ध गुणांक की गणना विधि आगे दिये गये उदाहरणों से स्पष्ट हो सकेगी।

उदाहरण–यदि $r_{12} = .58$, $r_{13} = .42$ तथा $r_{23} = .35$ हो, तो $R_{1.23}$ का मान ज्ञात कीजिये।

हल–ज्ञात है

$r_{12} = .58$, $r_{13} = .42$ $r_{23} = .35$

चर एक का चर दो व चर तीन से बहु-सहसम्बन्ध गुणांक ज्ञात करने का सूत्र है–

$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12} \; r_{13} \; r_{23}}{1 - r_{23}^2}}$$

$$= \sqrt{\frac{.58^2 + .42^2 - 2 \times .58 \times .42 \times .35}{1 - .35^2}}$$

$$= \sqrt{\frac{.3364 + .1764 - .17052}{1 - .1225}}$$

$$= \sqrt{\frac{.34228}{.8785}}$$

$$= \sqrt{.390063}$$

$$= .625$$

बहु-सहसम्बन्ध गुणांक की व्याख्या
(Interpretation of Multiple Correlation Coefficient)

नोट

बहु-सहसम्बन्ध गुणांक की गणना के उपरान्त उसकी व्याख्या का प्रश्न उठता है। वस्तुत: बहु-सहसम्बध गुणांक की व्याख्या ठीक उसी प्रकार से की जाती है जिस प्रकार से साधारण सह-सम्बन्ध गुणांक की व्याख्या की जाती है परन्तु बहु-सहसम्बन्ध गुणांक के सम्बन्ध में यह बात ध्यान देने योग्य है कि बहु-सहसम्बन्ध की दिशा की ओर कोई संकेत नहीं करता है वरन केवल आश्रित चर तथा स्वतन्त्र चरों के बीच संयुक्त सहसम्बन्ध (Joint Relationship) के मान को बताता है। यह बात बहु-सहसम्बन्ध ज्ञात करने के सूत्र से भी स्पष्ट होती है। क्योंकि किसी धनात्मक संख्या का वर्गमूल धनात्मक तथा ऋणात्मक दोनों ही हो सकता है इसलिए बहु-सहसम्बन्ध गुणांक को '+' अथवा '–' चिन्ह देना सम्भव नहीं हो पाता है। बहु-सहसम्बन्ध गुणांक की व्याख्या R^2 के रूप में की जा सकती है। बहु-सहसम्बन्ध गुणांक के वर्ग अर्थात R^2 को **बहु-निर्धारक गुणांक** (Coefficient of Multiple Determination) कहते हैं तथा यह आश्रित चर व स्वतन्त्र चरों के बीच उभयनिष्ठ प्रसरण (Common variance) को अनुपात के रूप में बताता है जिसे 100 से गुणा करने पर उभयनिष्ठ प्रसरण का मान प्रतिशत में प्राप्त हो जाता है। अत: यदि $R = .75$ तब $R^2 = .5625$ होगा। तब कह सकते हैं कि आश्रित चर तथा स्वतन्त्र चरों के बीच उभयनिष्ठ प्रसरण का मान 56.25 प्रतिशत है। अवशिष्ट प्रसरण (Remaining Variance) अर्थात् $1 - R^2$ को **बहु-अनिर्धारक गुणांक** (Coefficient of Multiple Non-determination) कहते हैं तथा इसे K से प्रदर्शित करते हैं। बहु-अनिर्धारक गुणांक (K) बताता है कि आश्रित चर का कितना प्रसरण अनुपात स्वतन्त्र चरों से अभिव्यक्त नहीं हो पा रहा है। बहु-अनिर्धारक गुणांक वास्तव में साधारण सहसम्बन्ध गुणांक के अलगाव गुणांक (Coefficient of Alienation) के समकक्ष है तथा जिस प्रकार से वहाँ पर $r^2 + k^2$ का मान 1.0 के बराबर होता है, ठीक उसी प्रकार से $R^2 + K^2$ का मान 1.0 होता है।

किसी आश्रित चर तथा अनेक स्वतन्त्र चरों के मध्य बहु-सहसम्बन्ध गुणांक का मान सदैव ही आश्रित चर तथा स्वतन्त्र चरों के बीच के साधारण सहसम्बन्ध गुणांकों के मान (+ अथवा – पर ध्यान दिये बिना) से अधिक होता है। अत: $R_{1.23}$ का मान सदैव ही r_{12} तथा r_{13} दोनों के मान से अधिक होगा। बहु-सहसम्बन्ध गुणांक के सूत्र से स्पष्ट होगा कि बहु-सहसम्बन्ध गुणांक का अधिकतम मान तब प्राप्त होता है जब $r_{23} = 0$ होता है। वस्तुत: $r_{23} = 0$ होने पर बहु-सहसम्बन्ध गुणांक का सूत्र $R_{1.23} = \sqrt{r_{12}^2 + r_{13}^2}$ हो जाता है जो किन्हीं भी दिये गये r_{12} तथा r_{13} के लिए $R_{1.23}$ का अधिकतम मान प्रदान करेगा।

बहु-सहसम्बन्ध गुणांक के सम्बन्ध में दो प्रमुख नियम हैं कि (*i*) स्वतन्त्र चरों के मध्य परस्पर सहसम्बन्ध (r_{23}) के बढ़ने पर बहु-सहसम्बन्ध गुणांक का मान घटता जाता है जबकि स्वतन्त्र चरों के मध्य परस्पर सहसम्बन्ध (r_{23}) के घटने पर बहु-सहसम्बन्ध गुणांक का मान बढ़ जाता है तथा (*ii*) आश्रित चर तथा स्वतन्त्र चरों के मध्य सहसम्बन्ध गुणांकों के मानों के बढ़ने पर बहु-सहसम्बन्ध गुणांक का मान बढ़ता है। अत: बहु-सहसम्बन्ध गुणांक का मान तब अधिकतम होता है जब आश्रित चर का स्वतन्त्र चरों से सहसम्बन्ध अधिक पुष्ट (High) हो जबकि स्वतन्त्र चरों से सहसम्बन्ध अत्यन्त क्षीण (Low) हो परन्तु कुछ अपवादात्मक स्थितियाँ (Exceptional Cases) ऐसी हो सकती हैं जो उपरोक्त वर्णित नियमों के विपरीत परिणाम प्रस्तुत कर देती हैं। ये दोनों नियम एवं इनकी अपवादात्मक स्थितियाँ सारणी 1 में प्रस्तुत परिकल्पित समंकों से स्पष्ट हो सकेंगी।

सारणी 1

विभिन्न साधारण सहसम्बन्ध गुणांकों का बहु-सहसम्बन्ध गुणांक पर प्रभाव

क्रम-संख्या	r_{12}	r_{13}	r_{23}	$R^2_{1.23}$	$R_{1.23}$
1	.5	0	0	.2500	.50
2	.5	0	.4	.2976	.55
3	.5	0	.6	.3906	.63
4	.5	0	.8	.6944	.83

नोट

5	.5	.3	0	.3400	.58
6	.5	.3	.4	.2619	.51
7	.5	.3	.6	.2500	.50
8	.5	.3	.8	.2778	.53
9	.5	.6	0	.6100	.78
10	.5	.6	.4	.4405	.66
11	.5	.6	.6	.3906	.63
12	.5	.6	.8	.3611	.60
13	.5	.3	– .4	.5476	.74
14	.5	– .3	.4	.5476	.74
15	.5	– .3	– .4	.2619	.51
16	– .5	– .3	– .4	.5476	.74

उपरोक्त विवेचन एवं सारणी के अवलोकन से बहु-सहसम्बन्ध गुणांक के सम्बन्ध में निम्नांकित महत्त्वपूर्ण बातें स्पष्ट होती हैं–

(*i*) बहु-सहसम्बन्ध गुणांक सहसम्बन्ध की दिशा को इंगित नहीं करता है।

(*ii*) आश्रित चर तथा स्वतन्त्र चरों के मध्य सहसम्बन्ध गुणांकों के बढ़ने पर बहु-सहसम्बन्ध गुणांक का मान भी बढ़ता है।

(*iii*) स्वतन्त्र चरों के मध्य परस्पर सहसम्बन्ध गुणांकों के बढ़ने पर बहु-सहसम्बन्ध गुणांक का मान घटता है।

(*iv*) बहु-सहसम्बन्ध गुणांक का मान आश्रित चर तथा स्वतन्त्र चरों के मध्य साधारण सहसम्बन्ध गुणांकों के मानों से अधिक या बराबर होता है।

4.5 बहु-प्रतीपगमन समीकरण (Multiple Regression Equation)

बहु-सहसम्बन्ध गुणांक को दो से अधिक स्वतन्त्र चरों के लिए भी विस्तृत किया जा सकता है, परन्तु इसकी चर्चा के पूर्व बहु-प्रतीपगमन को समझना आवश्यक होगा। जिस प्रकार से दो-चर समस्या (Two Variable Problem) में एक स्वतन्त्र चर की सहायता से आश्रित चर का पूर्वकथन करने के लिए बहु-प्रतीपगमन समीकरण विकसित की गई थी, ठीक उसी प्रकार से बहु-चर समस्या में कई स्वतन्त्र चरों के ज्ञात होने पर आश्रित चर का पूर्वकथन करने के लिए प्रतीपगमन समीकरण को विकसित किया जा सकता है। किसी आश्रित चर का दो स्वतन्त्र चरों की सहायता से पूर्वकथन करने के लिए प्रयुक्त किए जाने वाली प्रतीपगमन समीकरण का सामान्य रूप (General Form) निम्नवत होता है–

$$\overline{X}_1 = b_{12.3}\,X_2 + b_{13.2}\,X_3 + K$$

जहाँ X_1 = आश्रित चर X_1 पर पूर्वकथित मान

X_2 = स्वतन्त्र चर X_2 पर ज्ञात मान

X_3 = स्वतन्त्र चर X_3 पर ज्ञात मान

$b_{12.3}$ = X_1 के पूर्वकथन के लिए X_2 को दिया गया भार

$b_{13.2}$ = X_1 के पूर्वकथन के लिए X_3 को दिया गया भार

K = स्थिरांक

किसी प्रयोज्य के स्वतन्त्र चरों X_2 तथा X_3 पर प्राप्तांक ज्ञात होने पर उसके आश्रित चर X_1 के प्राप्तांक को उपरोक्त बहु-प्रतीपगमन समीकरण की सहायता से ज्ञात किया जा सकता है। स्थिरांक (Constant) K का मान

इस प्रकार से चयनित किया जाता है कि पूर्वकथित X_1 मानों का मध्यमान वास्तविक X_1 मानों के मध्यमान के बराबर हो। उपरोक्त समीकरण में $b_{12.3}$ तथा $b_{13.2}$ गुणांक (Coefficients) वस्तुत: X_2 तथा X_3 की सहायता से X_1 के पूर्वकथन के समय X_2 तथा X_3 चरों के मानों को दिये जाने वाला भार (Weights) है। अत: गुणांक $b_{12.3}$ का मान बताता है कि X_2 में एक इकाई (Unit) की वृद्धि होने पर X_1 के मान में कितनी इकाईयों की वृद्धि होगी जबकि X_3 के प्रभाव को समाप्त कर दिया गया हो। इसी प्रकार से गुणांक $b_{13.2}$ का मान बताता है कि X_3 में एक इकाई की वृद्धि होने पर X_1 के मान में कितनी इकाइयों की वृद्धि होगी जबकि X_2 के प्रभाव को समाप्त कर दिया गया है। इन बी गुणांकों को आंशिक प्रतीपगमन गुणांक (Partial Regression Coefficients) कहते हैं। उपरोक्त बहु-प्रतीगमन समीकरण को विचलित प्राप्तांकों (Deviated Scores) तथा मानक प्राप्तांकों (Standard Scores) के रूप में निम्नवत लिखा जा सकता है–

$$\bar{x}_1 = b_{12.3}\, x_2 + b_{13.2}\, x_3$$

तथा

$$\bar{z}_1 = \beta_{12.3}\, Z_2 + \beta_{13.2}\, Z_3$$

उपरोक्त समीकरण में प्रयुक्त बीटा (β) गुणांकों को मानक आंशिक प्रतीपगमन गुणांक (Standard Partial Regression Weights) कहते हैं। बीटा गुणांकों को 'मानक' इसलिए कहते हैं क्योंकि ये मानक प्राप्तांकों के लिए प्रयुक्त किये जाते हैं। विभिन्न बी-गुणांकों (*b*-Coefficients) की गणना बीटा-गुणांकों (β-coefficients) की सहायता से की जा सकती है। विभिन्न बी-गुणांकों, बीटा-गुणांकों तथा स्थिरांक की गणना निम्न सूत्रों से की जा सकती है–

$$b_{12.3} = \frac{s_1}{s_2}\, \beta_{12.3}$$

$$b_{13.2} = \frac{s_1}{s_3}\, \beta_{13.2}$$

$$\beta_{12.3} = \frac{r_{12} - r_{13}\, r_{23}}{1 - r_{23}^2}$$

$$\beta_{13.2} = \frac{r_{13} - r_{12}\, r_{23}}{1 - r_{23}^2}$$

$$K = M_1 - (b_{12.3}\, M_2 + b_{13.2}\, M_3)$$

उपरोक्त सूत्रों में प्रयुक्त संकेताक्षर स्वस्पष्ट है। बीटा-गुणांकों तथा साधारण सहसम्बन्ध गुणांकों की सहायता से बहु-सहसम्बन्ध गुणांक (R) की गणना निम्न ढंग से की जा सकती है–

$$R_{1.23} = \sqrt{\beta_{12.3}\, r_{12} + \beta_{13.2}\, r_{13}}$$

सुविधा के लिए $b_{12.3}$ व $\beta_{12.3}$ को क्रमश: b_2 व β_2 संकेताक्षर से तथा $b_{13.2}$ व $\beta_{13.2}$ को क्रमश: b_3 व β_3 संकेताक्षर से भी लिखा जाता है। तब प्रतीपगमन समीकरण तथा विभिन्न सूत्रों का रूप निम्नवत हो जायेगा–

$$\bar{X}_1 = b_2\, X_2 + b_3\, X_3 + K$$

$$\bar{x}_1 = b_2\, x_2 + b_3\, x_3$$

$$\bar{z}_1 = \beta_2 Z_2 + \beta_3\, Z_3$$

$$b_2 = \frac{s_1}{s_2}\, \beta_2 \quad \text{तथा} \quad b_3 = \frac{s_1}{s_3}\, \beta_3$$

$$\beta_2 = \frac{r_{12} - r_{13}\, r_{23}}{1 - r_{23}^2} \quad \text{तथा} \quad \beta_3 = \frac{r_{13} - r_{12}\, r_{23}}{1 - r_{23}^2}$$

$$K = M_1 - (b_2\, M_2 + b_3 M_3)$$

$$R_{1.23} = \sqrt{\beta_2\, r_{12} + \beta_3\, r_{13}}$$

यदि आश्रित चर को Y से तथा स्वतन्त्र चरों को X_1 व X_2 से प्रदर्शित किया जाये, तब विभिन्न प्रतीपगमन समीकरणों तथा सूत्रों का रूप निम्नवत हो जायेगा

नोट

$$Y = b_1X_1 + b_2X_2 + K$$

$$\bar{y} = b_1x_1 + b_2x_2$$

$$\bar{z}_y = \beta_1 z_1 + \beta_2 z_2$$

$$b_1 = \frac{s_y}{s_1}\beta_1 \quad \text{तथा} \quad b_2 = \frac{s_y}{s_2}\beta_2$$

$$\beta_1 = \frac{r_{yx_1} - r_{yx_2}\, r_{x_1x_2}}{1 - r^2_{x_1x_2}} \quad \text{तथा} \quad \beta_2 = \frac{r_{yx_2} - r_{yx_1}\, r_{x_1x_2}}{1 - r^2_{x_1x_2}}$$

$$K = M_y - (b_1M_1 + b_2M_2)$$

$$R_{y.x_1x_2} = \sqrt{\beta_1 r_{yx_1} + \beta_2 r_{yx_2}}$$

उपरोक्त वर्णित प्रतीपगमन समीकरणों, बी गुणांकों तथा बीटा गुणांकों के अवलोकन से पाठकों को स्पष्ट हो गया होगा कि प्रतीपगमन समीकरण तथा सूत्रों के ये सभी रूप वस्तुत: एक समान हैं, अन्तर केवल संकेताक्षरों के प्रयोग का है। इन सभी रूपों को प्रस्तुत करना इसलिए उपयुक्त समझा गया है क्योंकि विभिन्न पुस्तकों तथा अनुसंधान कार्यों में ये भिन्न-भिन्न रूप प्रयुक्त किये जाते हैं। पाठकगण किसी भ्रम में पड़े बिना इनमें से किसी भी एक रूप का प्रयोग कर सकते हैं। प्रस्तुत पुस्तक में आगे प्रथम रूप का ही प्रयोग किया जायेगा। यहाँ यह भी स्पष्ट करना उचित होगा कि प्रतीपगमन समीकरण में विभिन्न स्वतन्त्र चरों के बीटा तथा बी-गुणांकों के मान प्रतीपगमन समीकरण में उनके क्रम पर निर्भर करते हैं। यदि पहले X_3 को तथा बाद में X_2 को सम्मिलित किया जायेगा तब उनके गुणांक भी परिवर्तित हो सकते हैं। बहुप्रतीपगमन की प्रक्रिया आगे दिये गये उदाहरण से स्पष्ट हो सकेगी।

उदाहरण—यदि तीन चरों X_1, X_2 तथा X_3 के लिए मध्यमान व मानक विचलन क्रमश: 25.6 व 5.8, 47.2 व 7.3 तथा 16.9 व 3.4 हो एवं $r_{12} = .47, r_{13} = .38$ तथा $r_{23} = .24$ हो, तब चर 2 व 3 की सहायता से चर 1 के पूर्व कथन के लिए प्रतीपगमन समीकरण ज्ञात करें तथा बहु-सहसम्बन्ध गुणांक की गणना करे। जिस प्रयोज्य के X_2 पर 34 व X_3 पर 16 अंक हों, उसके X_1 चर के प्राप्तांकों का पूर्वकथन भी करें।

हल—ज्ञात है

$r_{12} = .47$	$M_1 = 25.6$	$\sigma_1 = 5.8$
$r_{13} = .38$	$M_2 = 47.2$	$\sigma_2 = 7.3$
$r_{23} = .24$	$M_3 = 16.9$	$\sigma_3 = 3.4$

चर संख्या 2 व चर संख्या 3 की सहायता से चर संख्या 1 के पूर्वकथन के लिए प्रतीपगमन समीकरण निम्नवत होगी—

$$\overline{X}_1 = b_2\, X_2 + b_3\, X_3 + K$$

बी गुणांकों तथा स्थिरांक की गणना निम्न सूत्रों से की जायेगी—

$$b_2 = \frac{s_1}{s_2}\,\beta_2 \quad \text{जहाँ} \quad \beta_2 = \frac{r_{12} - r_{13}\, r_{23}}{1 - r^2_{23}}$$

$$b_3 = \frac{s_1}{s_3}\,\beta_3 \quad \text{जहाँ} \quad \beta_3 = \frac{r_{13} - r_{12}\, r_{23}}{1 - r^2_{23}}$$

तथा $$K = M_1 - (b_2\, M_2 + b_3\, M_3)$$

अत: $\beta_2 = \frac{.47 - .47 \times .24}{1 - .24 \times .24} = \frac{.3572}{.9424} = .379$

$b_2 = \frac{5.8}{7.3} \times .379 = .30$

$\beta_3 = \frac{.38 - .38 \times .24}{1 - .24 \times .24} = \frac{.2888}{.9424} = .306$

$b_2 = \frac{5.8}{3.4} \times .306 = .52$

$$\begin{aligned} K &= 25.6 - (.30 \times 47.2 + .52 \times 16.9) \\ &= 25.6 - (14.16 + 8.79) \\ &= 25.6 - 22.95 \\ &= 2.65 \end{aligned}$$

अत: X_1 के लिए प्रतीपगमन समीकरण होगी–

$$\overline{X}_1 = .30\,X_2 + .52\,X_3 + 2.65$$

क्योंकि बहु-सहसम्बन्ध गुणांक का सूत्र है–

$$R_{1.23} = \sqrt{\beta_2 r_{12} + \beta_3 r_{13}}$$

अत: $R_{1.23} = \sqrt{.379 \times .47 + .306 \times .38}$

$$\begin{aligned} &= \sqrt{.17813 + .11628} \\ &= \sqrt{.29441} \\ &= .543 \end{aligned}$$

$X_2 = 34$ तथा $X_3 = 16$ वाले प्रयोज्य के लिए X_1 का पूर्वकथित मान ज्ञात करने के लिए प्रतीपगमन समीकरण में $X_2 = 34$ तथा $X_3 = 16$ रखने पर–

$$\begin{aligned} \overline{X}_1 &= (.30 \times 34) + (.52 \times 16) + 2.65 \\ &= 10.2 + 8.32 + 2.65 \\ &= 21.17 \end{aligned}$$

अत: $X_2 = 34$ तथा $X_3 = 16$ वाले प्रयोज्य के X_1 चर पर 21 अंक होने की सम्भावना की जा सकती है।

4.6 दो से अधिक स्वतन्त्र चरों के लिए बहु-प्रतीपगमन (Multiple Regression for More than Two Independent Variables)

बहु-सहसम्बन्ध की समस्या को दो से अधिक स्वतन्त्र चरों के लिए भी विस्तारित (Extend) किया जा सकता है। दो से अधिक स्वतन्त्र चरों के होने पर बहु-सहसम्बन्ध गुणांक की गणना कुछ जटिल हो जाती है तब प्रतीपगमन समीकरण के बीटा गुणांकों अथवा बी-गुणांकों की सहायता से ही बहु-सहसम्बन्ध की गणना की जाती है। दो से अधिक स्वतन्त्र चरों की स्थिति में प्रतीपगमन समीकरण के गुणांकों को ज्ञात करने के लिए डूलिटिल विधि (Doolittle Method) अथवा ऐटकिन विधि (Aitkin's Method) का प्रयोग किया जाता है। इन दोनों ही विधियों में प्रयुक्त सोपान (Steps) काफी जटिल व श्रमसाध्य हैं। आजकल कम्प्यूटर के द्वारा दो से अधिक स्वतन्त्र चरों के लिए प्रतीपगमन समीकरण के गुणांकों को ज्ञात करने की सुविधा उपलब्ध है, इसलिए इन विधियों की

विस्तार से चर्चा करने की आवश्यकता नहीं है। फिर भी पाठकों को बहु-प्रतीपगमन विधि से, परिचित कराने के लिए डूलिटिल विधि का आगे वर्णन किया गया है। दो से अधिक $(m-1)$ स्वतन्त्र चरों के होने पर बहु-प्रतीपगमन समीकरण का सामान्य रूप अग्रांकित होता है।

$$\overline{X}_1 = b_2X_2 + b_3X_3 + \ldots\ldots\ldots + b_mX_m + K$$

या

$$\overline{X}_1 = \beta_2\frac{s_1}{s_2}X_2 + \beta_3\frac{s_1}{s_3}X_3 + \ldots\ldots\ldots + \beta_m\frac{s_1}{s_m}X_m + K$$

इस सूत्र में β_2 का पूर्ण रूप $\beta_{12.34\ldots\ldots\ldots m}$ है, β_3 का पूर्ण रूप $\beta_{13.245\ldots\ldots\ldots m}$ तथा β_m का पूर्ण रूप $\beta_{1m.2345\ldots\ldots\ldots m-1}$ है इस स्थिति में बहु-सहसम्बन्ध गुणांक (R) की गणना निम्न सूत्र से की जा सकती है–

$$R_{1.23\ldots\ldots\ldots i} = \sqrt{\beta_2r_{12} + \beta_3r_{13} + \ldots\ldots\beta_mr_{1m}}$$

नि:सन्देह स्वतन्त्र चरों की संख्या के बढ़ाने पर बहु-सहसम्बन्ध गुणांक (R) तथा पूर्वकथन की परिमार्जितता (Accuracy) बढ़ जाती है, परन्तु प्रतीपगमन के लिए कितने स्वतन्त्र चरों का चयन किया जाये, यह एक विचारणीय प्रश्न है। यदि कोई चर अथवा अनेक स्वतन्त्र चरों को प्रतीपगमन समीकरण जोड़ने पर बहु-सहसम्बन्ध गुणांक तथा परिमार्जितता में नगण्य अथवा अत्यन्त कम वृद्धि होती है तब उनको जोड़ने का कोई लाभ नहीं है। ऐसी स्थिति में प्रश्न उठता है कि क्या किसी स्वतन्त्र चर को प्रतीपगमन विश्लेषण में सम्मिलित करने पर बहु-सहसम्बन्ध गुणांक में सार्थक वृद्धि हो रही है। इस प्रश्न का उत्तर पाने की तीन प्रमुख विधियाँ अग्रगामी समाधान (Forward solution), पूर्वगामी समाधान (Backward Solution) तथा सोपानीकृत समाधान (Stepwise Solution) हैं। अग्रगामी समाधान में एक-एक करके चरों को प्रतीपगमन विश्लेषण सम्मिलित किया जाता है जबकि पूर्वगामी समाधान में एक-एक करके चरों को विश्लेषण से हटाया जाता है। सोपानीकृत प्रतीपगमन विश्लेषण (Stepwise Regression Analysis) वस्तुत: अग्रगामी समाधान का ही एक प्रकार है जिसमें पूर्व के सोपानों में सम्मिलित उन चरों को हटा दिया जाता है जिनकी उपयोगिता नहीं रही है। अग्रगामी समाधान में प्रत्येक सोपान में एक-एक करके स्वतन्त्र चरों को प्रतीपगमन समीकरण में इस प्रकार से सम्मिलित किया जाता है कि प्रत्येक सोपान में अधिकतम बहु-सहसम्बन्ध गुणांक प्राप्त हो सके। इस प्रकार से सोपानीकृत प्रतीपगमन विश्लेषण में प्रथम सोपान में एक सर्वोत्तम पूर्वकथक (Best Predicter) को, दूसरे सोपान में दो सर्वोत्तम पूर्वकथकों को, ...तथा इसी प्रकार से mवें सोपान से m सर्वोत्तम पूर्वकथकों को सम्मिलित करके प्रतीपगमन समीकरण तैयार की जाती है। प्रत्येक सोपान पर प्राप्त बहु-सहसम्बन्ध गुणांक में उससे पूर्व के सोपान पर प्राप्त बहु-सहसम्बन्ध गुणांक की तुलना में हुई वृद्धि की सार्थकता का परीक्षण प्रविष्टि के लिए एफ-परीक्षण (F to enter Test) के द्वारा किया जा सकता है। यदि बहु-सहसम्बन्ध गुणांक में सार्थक वृद्धि नहीं होती है तब प्रतीपगमन विश्लेषण में अवशिष्ट स्वतन्त्र चरों को सम्मिलित करना बन्द कर देते हैं, इसके विपरीत पूर्वगामी समाधान में प्रथम सोपान में सभी स्वतन्त्र चरों को सम्मिलित कर लेते हैं तब फिर एक-एक करके उन चरों को हटाते जाते हैं जिनका योगदान सबसे कम होता है। सोपानीकृत समाधान में प्रत्येक सोपान पर सभी चरों के योगदान की सार्थकता देखते हैं तथा आवश्यकता पड़ने पर असार्थक योगदान वाले चरों को हटा देते हैं। सोपानीकृत प्रतीपगमन विश्लेषण में सहनशीलता (Tolerance) का प्रत्यय अत्यन्त महत्त्वपूर्ण होता है। सहनशीलता (Tolerance) वास्तव में गणना कार्य में परिशुद्धता (Computational Accuracy) बनाये रखने का एक उपाय है। किसी स्वतन्त्र चर की सहनशीलता से तात्पर्य उस चर तथा उससे पूर्व प्रतीपगमन विश्लेषण में सम्मिलित चरों (Variable entered) के मध्य बहु सहसम्बन्ध गुणांक के वर्ग के पूरक (Complement) अर्थात् $1 - r^2$ से है। अत्यन्त अल्प अथवा नगण्य सहनशीलता (जैसे .001) वाले चर को प्रतीपगमन विश्लेषण में सम्मिलित करने पर सन्निकटीकरण त्रुटियों (Rounding Errors) के कारण प्रतीपगमन विश्लेषण के परिणामों पर दुष्प्रभाव पड़ने की अधिक सम्भावना रहती है। वैसे भी किसी चर की सहनशीलता के नगण्य अथवा शून्य होने का अर्थ है कि वह चर पूर्व में सम्मिलित (Extered) चरों का रेखीय संयोजन (Linear Combination) है तथा ऐसे चर को प्रतीपगमन विश्लेषण में सम्मिलित करने से कोई

अतिरिक्त सूचना (Additional Information) नहीं मिल सकती है। सोपानीकृत प्रतीपगमन विश्लेषण का गणना कार्य कम्प्यूटर की सहायता से कराना अधिक सुविधाजनक होता है।

बहु-प्रतीपगमन की डूलिटिल विधि
(Doolittle Method of Multiple Regression)

दो से अधिक स्वतन्त्र चरों की स्थिति में बहु-प्रतीपगमन समीकरण तथा बहु-सहसम्बन्ध गुणांक ज्ञात करने की एक सरल तथा उपयोगी विधि को एम.एच. डलिटिल (M. H. Doolittle) ने सन 1878 में प्रस्तुत किया था। पीछे इंगित किया जा चुका है कि

$$\overline{X}_1 = \beta_2 \frac{s_1}{s_2} X_2 + \beta_3 \frac{s_1}{s_3} X_3 + \ldots + \beta_j \frac{s_1}{s_j} X_j + K$$

अथवा $$\overline{X}_1 = b_2 x_2 + b_3 x_3 + \ldots B_j X_j + K$$

प्रतीपगमन विश्लेषण के द्वारा वस्तुतः β मानों को इस प्रकार से चयनित करना होता है कि प्रयोज्यों के X_1 तथा $\overline{X}_1$ मानों के बीच अन्तर न्यूनतम हो। गणित में निपुण पाठक सरलता से समझ सकेंगे कि न्यूनतम वर्ग विधि (Method of Least Squares) का प्रयोग करने पर निम्न रेखीय युगपत समीकरण (Linear Simultaneous Equations) प्राप्त होंगी

$$\begin{aligned}
&\beta_2 + r_{23}\beta_3 + r_{24}\beta_4 + \ldots + r_{2j}\beta_j = r_{12} \\
&r_{23}\beta_2 + \beta_3 + r_{34}\beta_4 + \ldots + r_{3j}\beta_j = r_{13} \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
&r_{2j}\beta_2 + r_{3j}\beta_3 + r_{4j}\beta_4 + \ldots + \beta_j = r_{1j}
\end{aligned}$$

इन समीकरणों में β मान अज्ञात है। जितने β मान होते हैं ठीक उतनी ही समीकरणें होती हैं। इन समीकरणों को सामान्य समीकरण (Normal Equation) कहते हैं परन्तु सामान्य समीकरण पद में प्रयुक्त सामान्य शब्द का सामान्य प्रायिकता वक्र से कोई सम्बन्ध नहीं है। चार स्वतन्त्र चरों की स्थिति में प्रतीपगमन समीकरण निम्नवत होगी–

$$\overline{X}_1 = \beta_2 \frac{s_1}{s_2} X_2 + \beta_3 \frac{s_1}{s_3} X_3 + \beta_4 \frac{s_1}{s_4} X_4 + \beta_5 \frac{s_1}{s_5} X_5 + \beta_j \frac{s_1}{s_j} X_j + K$$

अथवा $$\overline{X}_1 = b_2 X_2 + b_3 X_3 + b_4 X_4 + b_5 X_5 + k$$

इस स्थिति में सरल समीकरणें (Normal Equations) निम्नवत होंगी–

$$\begin{aligned}
\beta_2 + r_{23}\beta_3 + r_{24}\beta_4 + r_{25}\beta_5 &= r_{12} \\
r_{23}\beta_2 + \beta_3 + r_{34}\beta_4 + r_{35}\beta_5 &= r_{13} \\
r_{24}\beta_2 + r_{34}\beta_3 + \beta_4 + r_{45}\beta_5 &= r_{14} \\
r_{25}\beta_2 + r_{35}\beta_3 + r_{45}\beta_4 + \beta_5 &= r_{15}
\end{aligned}$$

यदि इन समीकरणों में प्रयुक्त सभी सहसम्बन्ध गुणांकों का मान ज्ञात हो तब इन समीकरणों को सरल करके β मान ज्ञात किये जा सकते हैं। जब केवल दो ही स्वतन्त्र चर होते हैं तब β मानों के लिए सूत्र सरलता से प्राप्त किये जा सकते हैं। दो स्वतन्त्र चरों के लिए पिछले खण्ड में वर्णित β_2 तथा β_3 के सूत्र इसी प्रकार की समीकरणों से प्राप्त किये गये हैं। परन्तु अधिक स्वतन्त्र चरों के होने पर β मानों को ज्ञात करने के लिए इन समीकरणों में सहसम्बन्ध गुणांकों का मान रखकर ही उनका समाधान करना ही श्रेयस्कर रहता है। इस प्रकार की प्रथम घातीय समीकरणों (First Degree Equations) का समाधान करने की कई विधियाँ हैं, परन्तु डूलिटिल के द्वारा प्रतिपादित विधि कुछ अधिक सुविधाजनक है। यह विधि सहसम्बन्ध गुणांकों में सममितता (Symmetry) की

विशेषता अर्थात् $r_{12} = r_{21}, r_{13} = r_{31}, ..., r_{ij} = r_{ji}$ का लाभ उठाकर गणना कार्य को सरल बना देती है। उदाहरणार्थ किसी आश्रित चर (X_1) तथा चार स्वतन्त्र चर (X_2, X_3, X_4 व X_5) के लिए परिकल्पित मध्यमानों, मानक विचलनों तथा सहसम्बन्ध गुणांकों को सारणी 2 में प्रस्तुत किया गया है।

सारणी 2

पाँच चरों के लिए समंक (N = 500)

मध्यमान M	मानक विचलन s	चर Variable	सहसम्बन्ध गुणांक (r)				
			X_1	X_2	X_3	X_4	X_5
120.30	24.39	X_1		.61	.42	.24	.53
46.20	6.56	X_2	.61		.39	.23	.57
32.50	7.34	X_3	.42	.39		.36	.55
10.40	2.60	X_4	.24	.23	.36		.38
72.50	16.47	X_5	.53	.57	.55	.38	

सामान्य समीकरणों में स्वतन्त्र एवं आश्रित चरों के बीच प्राप्त विभिन्न सहसम्बन्ध गुणांकों (r's) के मान रखने पर निम्न समीकरणें प्राप्त होंगी–

$$\beta_2 + .39\,\beta_3 + .23\,\beta_4 + .57\,\beta_5 = .61 \qquad \textbf{(प्रथम समीकरण)}$$

$$.39\,\beta_2 + \beta_3 + .36\,\beta_4 + .55\,\beta_5 = .42 \qquad \textbf{(द्वीतीय समीकरण)}$$

$$.23\,\beta_2 + .36\,\beta_3 + \beta_4 + .38\,\beta_5 = .24 \qquad \textbf{(तृतीय समीकरण)}$$

$$.57\,\beta_2 + .55\,\beta_3 + .38\,\beta_4 + \beta_6 = .53 \qquad \textbf{(चतुर्थ समीकरण)}$$

इन समीकरणों को हल करने की विधि ठीक वही है जो हाईस्कूल कक्षाओं में बीजगणित के अन्तर्गत पढ़ाई जाती है। डूलिटिल विधि में वर्णन से पूर्व सारणी 3 में इन समीकरणों को बीजगणितीय ढंग से हल किया गया है जिससे पाठकगण डूलिटिल विधि में निहित तर्क को समझ सकें। सारणी 3 के अवलोकन से स्पष्ट है कि समीकरणों का यह समाधान चक्रीय है। प्रत्येक चक्र (Cycle) में एक-एक अज्ञात राशि (β) को हटाया गया है। अज्ञात राशियों को हटाने का क्रम तब तक चलता है जब तक अन्तिम अज्ञात राशि का मान प्राप्त नहीं हो जाता है। फिर इस अज्ञात राशि का मान इससे पूर्व के चक्र में प्राप्त समीकरण में रखकर दूसरी अज्ञात राशि का मान ज्ञात कर लेते हैं तथा इसी प्रकार से सभी अज्ञात राशियों (β's) के मान ज्ञात कर लेते हैं। सहसम्बन्ध सारणी में सहसम्बन्ध गुणांकों की सममितता की विशेषता का उपयोग करके डूलिटिल ने सारणी 3 में प्रयुक्त विधि का एक सरलीकृत रूप प्रस्तुत किया। इस विधि में प्रयुक्त सोपानों को सारणी 4 में प्रस्तुत किया जा रहा है। सारणी 4 में विभिन्न पंक्तियों को A, B, C, ...आदि अक्षरों से तथा स्तम्भों को X_1, X_2 ... आदि संकेताक्षरों से इंगित किया गया है। अन्तिम स्तम्भ, जिसे Σ चिन्ह से इंगित किया गया है, गणना की आन्तरिक जाँच के लिए बनाया गया है। किसी पंक्ति तथा स्तम्भ के कटाव पर बने खाने के मान को उस पंक्ति तथा उस स्तम्भ के संकेताक्षरों को एक साथ बड़े कोष्ठक में रखकर इंगित किया गया है। जैसे [AX_2] से तात्पर्य पंक्ति A तथा स्तम्भ X_2 के कटाव वाले मान से है। सारणी 5 में बीटा गुणांकों के गणना कार्य को तथा सारणी 6 में प्रतीपगमन गुणांकों व स्थिरांक की गणना को प्रस्तुत किया गया है। वस्तुत: डूलिटिल विधि से बहु-प्रतीपगमन विश्लेषण करने का समस्त गणना कार्य सारणी 4, 5 तथा 6 से स्व स्पष्ट है फिर भी अब बोध की दृष्टि से इस विधि में प्रयुक्त सोपानों को आगे प्रस्तुत किया गया है।

सारणी 3
सामान्य समीकरणों का बीजगणितीय समाधान

पंक्ति (Row)	संक्रिया (Operation)	समाधान (Solution)
1	प्रथम समीकरण लिखो	$\beta_2 + .39\,\beta_3 + .23\,\beta_4 + .57\,\beta_5 = .61$
2	द्वितीय समीकरण लिखो	$.39\,\beta_2 + \beta_3 + .36\beta_4 + .55\,\beta_5 = .42$
3	पंक्ति 1 को – .39 से गुणा करो	$-.39\beta_2 - .1521\,\beta_3 - .0897\beta_4 - .2223\,\beta_5 = -.2379$
4	पंक्ति 2 तथा पंक्ति 3 को जोड़ो	$0 + .8479\,\beta_3 + .2703\,\beta_4 + .3277\,\beta_5 = .1821$
5	पंक्ति 4 को .8479 से भाग करो	$\beta_3 + .3188\,\beta_4 + .3865\,\beta_5 = .2148$
6	तृतीय समीकरण लिखो	$.23\,\beta_2 + .36\,\beta_3 + \beta_4 + .38\,\beta_5 = .24$
7	पंक्ति 1 को – .23 से गुणा करो	$-.23\beta_2 - .0897\beta_3 - .0529\beta_4 - .1311\,\beta_5 = -.1403$
8	पंक्ति 5 को – .2703 से गुणा करो	$-.2703\,\beta_3 - .0862\,\beta_4 - .1045\,\beta_5 = -.0581$
9	पंक्ति 6, 7 तथा 8 को जोड़ो	$0 + 0 - .8609\,\beta_4 - .1444\,\beta_5 = -.0416$
10	पंक्ति 9 को .8609 से भाग करो	$\beta_4 + .1677\,\beta_5 = .0483$
11	चतुर्थ समीकरण लिखो	$.57\,\beta_2 + .55\,\beta_3 + .38\,\beta_4 + \beta_5 = .53$
12	पंक्ति 1 को – .57 से गुणा करो	$-.57\beta_2 - .2223\,\beta_3 - .1311\,\beta_4 - .3249\beta_5 = -.3477$
13	पंक्ति 5 को – .3277 से गुणा करो	$-.3277\,\beta_3 - .1045\,\beta_4 - .1267\,\beta_5 = -.0704$
14	पंक्ति 10 को – .1444 से गुणा करो	$-.1444\,\beta_4 - .0242\,\beta_5 = -.0070$
15	पंक्ति 11, 12, 13 तथा 14 को जोड़ो	$0 + 0 + 0 + .5242\,\beta_5 = .1049$
16	पंक्ति 15 को .5242 से भाग करो	$\beta_5 = .2001$
17	पंक्ति 10 में $b_5 = .2001$ रखो	$\beta_4 + (.1677 \times .2001) = .0483$ अतः $\beta_4 = .0147$
18	पंक्ति 5 में $\beta_4 = .0147$ तथा $\beta_5 = .2001$ रखो	$\beta_3 + (.3188 \times .0147) + (.3865 \times .2001) = .2148$ अतः $\beta_3 = .1328$
19	पंक्ति 1 में $\beta_3 = .1328$, $\beta_4 = .0147$ तथा $\beta_5 = .2001$ रखो	$\beta_2 + (.39 \times .1328) + (.23 \times .0147) + (.57 \times .2001) = .61$, अतः $\beta_2 = .4408$

सारणी 4
डूलिटिल विधि के द्वारा बहु-प्रतीपगमन विश्लेषण
(Multiple-Regression by the Doolittle Method)

पंक्ति (Row)	संक्रिया (Operation)	स्तम्भ (Column) X_2	X_3	X_4	X_5	X_1	Σ
A	चर X_2 के साथ सहसम्बन्ध गुणांक	1.00	.39	.23	.57	.61	2.80
B	पंक्ति A को – $[AX_2] = -1.00$ से भाग	**– 1.00**	**– .39**	**– .23**	**– .57**	**– .61**	**– 2.80**
C	चर X_3 के साथ सहसम्बन्ध गुणांक		1.0000	.3600	.5500	.4200	+ 2.720
D	पंक्ति A की $[B\,X_3] = -.39$ से गुणा		– .1521	– 0.897	– .2223	– .2379	– 1.002

नोट

E	पंक्ति C तथा पंक्ति D का योग	.8479	.2703	.3277	.1821	+ 1.628
F	पंक्ति E को – [E X_3] = – .8479 से भाग	**– 1.0000**	**– .3188**	**– .3865**	**– .2148**	**– 1.920**
G	चर X_4 के साथ सहसम्बन्ध गुणांक		1.0000	.3800	.2400	+ 2.210
H	पंक्ति A की [B X_4] = – .23 से गुणा		– .0529	– 1311	– .1403	– .6440
I	पंक्ति E की [F X_4] = .3188 से गुणा		– .0862	– .1045	– .0581	– .5190
J	पंक्ति G, पंक्ति H तथा पंक्ति I का योग		.8609	.1444	.0416	+1.0470
K	पंक्ति J को – [J X_4] से भाग		**–1.0000**	**– .1677**	**– .0483**	**–1.2162**
L	चर X_5 के साथ सहसम्बन्ध गुणांक			1.0000	.5300	+3.0300
M	पंक्ति A की [BX_5] = – .57 से गुणा			– .3249	– .3477	– 1.5960
N	पंक्ति E की [FX_5] = – .3865 से गुणा			– .1266	– .0704	– .6292
O	पंक्ति J की [KX_5] = – .1677 से गुणा			– .0242	– .0070	– .1756
P	पंक्ति L, पंक्ति M, पंक्ति N तथा O का योग			.5243	.1049	+ .6292
Q	पंक्ति P को – [PX_5] = – .5243 से भाग			**– 1.000**	**– .2001**	**–1.2001**

सारणी 5
बीटा गुणांकों की गणना
(Solution of the Beta Coefficients)

बीटा गुणांक	गणना कार्य
β_5	– [QX_1] = .2001
β_4	– [KX_1] + β_5 [KX_5] = .0483 + .2001 (– .1677) = .0147
β_3	– [F X_1] + β_5 [F X_5] + β_4 [F X_4] = .2148 + .2001 (– .3865) + .0147 (– .3188) = .1328
β_2	– [B X_1] + β_5 [B X_5] + β_4 [B X_4] + β_3 [B X_3] = .6100 + .2001 (– .57) + .0147 (– .23) + .1328 (– .39) = .4408

सारणी 6
बहु-सहसम्बन्ध गुणांक, प्रतीपगमन गुणांकों तथा स्थिरांक की गणना
(Solution of the multiple Correlation, the Regression Coefficients and the Constant)

चर	β_i	r_{1i}	$\beta_i\, r_{1i}$	s_i	s_1/s_i	b_i	M_i	$M_i\, b_i$
X_2	.4408	.61	.2689	6.56	3.7180	1.6389	46.20	75.7172
X_3	.1328	.42	.0558	7.34	3.3229	.4413	32.50	14.3423
X_4	.0147	.24	.0035	2.60	9.3808	.1379	10.40	1.4342
X_5	.2001	.53	.1061	16.47	1.4809	.2963	72.50	21.4818
			$\Sigma\, b_i\, r_{1i}$ = .4043					$\Sigma M_i b_i$ = 112.9755

M_1 = 120.30 $\quad R^2_{1.2345} = \Sigma\, \beta_i\, r_{1i}$ = .4043 $\quad K = M_1 - \Sigma\, M_i\, b_i$

s_1 = 24.39 $\quad R_{1.2345} = \sqrt{.4043}$ = .6358 $\quad$ = 120.30 – 112.9755 = 7.3245

प्रतीपगमन समीकरण (Regression Equation)

$$\overline{X}_1 = b_2 X_2 + b_3 X_3 + b_4 X_4 + b_5 X_5 + K$$

$$X_1 = 1.6389\, X_2 + .4413\, X_3 + .1379\, X_4 + .2963\, X_5 + 7.3245$$

सोपान (Steps)

डूलिटिल विधि (Doolittle Method) से बहुप्रतीपगमन करते समय अनुसरण किये जाने वाले विभिन्न सोपानों को निम्नवत लिखा जा सकता है–

(*i*) पंक्ति A में चर X_2 के साथ अन्य चरों के सहसम्बन्ध गुणांक लिखो। इस पंक्ति में X_2 वाले स्तम्भ अर्थात $[AX_2]$ में 1.00 लिखो। पंक्ति में लिखे मानों का योग करके Σ स्तम्भ में लिखो।

(*ii*) पंक्ति B में A के मानों को $-[AX_2]$ अर्थात -1.00 से भाग करके लिखो।

(*iii*) पंक्ति C में चर X_3 के साथ अवशिष्ट चरों (Remaining Variables) के सहसम्बन्ध गुणांक लिखो। क्योंकि r_{23} का मान पहले सोपान में लिखा जा चुका है इसलिए यहाँ पर अवशिष्ट चरों से तात्पर्य X_2 के अतिरिक्त अन्य चरों से है। इस पंक्ति के X_3 वाले स्तम्भ अर्थात $[CX_3]$ में 1.00 लिखना होगा। इस पंक्ति में लिखे सभी मानों एवं X_3 के साथ अन्य चरों के सहसम्बन्ध गुणांकों का योग करके Σ स्तम्भ में लिखो।

(*iv*) पंक्ति D में पंक्ति A के मानों को $[BX_3]$ अर्थात $-.39$ से गुणा करके लिखो।

(*v*) पंक्ति E में पंक्ति C तथा पंक्ति D का योग करके लिखो।

(*vi*) पंक्ति F में पंक्ति E के मानों को $-[EX_3]$ अर्थात $-.8479$ से भाग करके लिखो।

जाँच–देखो कि पंक्ति F के मानों का योग (अन्तिम स्तम्भ के मान को छोड़कर) अन्तिम स्तम्भ के मान अर्थात [F] के बराबर है।

(*vii*) पंक्ति G में चर X_4 के साथ अवशिष्ट चरों के सहसम्बन्ध गुणांक लिखो। क्योंकि r_{24} तथा r_{34} के मान पूर्व में लिखे जा चुके हैं इसलिए यहाँ पर अवशिष्ट चरों से तात्पर्य X_2 तथा X_3 के अतिरिक्त अन्य चरों से हैं। इस पंक्ति के X_4 वाले स्तम्भ अर्थात $[GX_4]$ में 1.00 लिखो। इस पंक्ति में लिखे सभी मानों एवं X_4 के साथ अन्य चरों के सहसम्बन्ध गुणांकों का योग करके Σ स्तम्भ में लिखो।

(*viii*) पंक्ति H में पंक्ति A के मानों को $[BX_4]$ अर्थात $-.23$ से गुणा करके लिखो।

(*ix*) पंक्ति I में पंक्ति E के मानों को $[FX_4]$ अर्थात $-.3188$ से गुणा करके लिखो।

(*x*) पंक्ति J में पंक्ति G, पंक्ति H तथा पंक्ति I का योग करके लिखो।

(*xi*) पंक्ति K में पंक्ति J के मानों को $-[JX_4]$ अर्थात $-.8609$ से भाग करके लिखो।

जाँच–देखो कि पंक्ति K के मानों का योग अन्तिम स्तम्भ के मान अर्थात [KΣ] के बराबर है।

(*xii*) पंक्ति L में चर X_5 के साथ अवशिष्ट चरों के सहसम्बन्ध गुणांक लिखो। क्योंकि r_{25}, r_{35} तथा r_{45} के मान पूर्व में लिखे जा चुके हैं इसलिए यहाँ पर अवशिष्ट चरों से तात्पर्य X_2, X_3 तथा X_4 चरों के अतिरिक्त अन्य चरों से है। इस पंक्ति के X_5 वाले स्तम्भ अर्थात $[LX_5]$ में 1.00 लिखो। इस पंक्ति में लिखे सभी मानों एवं X_5 के साथ अन्य चरों के सहसम्बन्ध गुणांकों का योग Σ स्तम्भ में लिखो।

(*xiii*) पंक्ति M में पंक्ति A के मानों को $[BX_5]$ अर्थात $-.57$ से गुणा करके लिखो।

(*xiv*) पंक्ति N में पंक्ति E के मानों को $[FX_5]$ अर्थात $-.3865$ से गुणा करके लिखो।

(*xv*) पंक्ति O में पंक्ति J के मानों को $[KX_5]$ अर्थात् – .1677 से गुणा करके लिखो।

(*xvi*) पंक्ति P में पंक्ति L, पंक्ति M, पंक्ति N तथा पंक्ति O का योग करके लिखो।

(*xvii*) पंक्ति Q में पंक्ति P के मानों को $-[PX_5]$ अर्थात – .5243 से भाग करके लिखो।

नोट

जाँच–देखो कि पंक्ति Q के मानों का योग अन्तिम स्तम्भ के मान अर्थात [QΣ] के बराबर है।

(*xviii*) निम्न सूत्रों से बीटा गुणांकों की गणना करो–

$\beta_5 = -[QX_1]$

$\beta_4 = -[KX_1] + \beta_5 [KX_5]$

$\beta_3 = -[FX_1] + \beta_5 [FX_5] + \beta_4 [FX_4]$

$\beta_2 = -[BX_1] + b_5 [bX_5] + \beta_4 [BX_4] + \beta_3 [BX_3]$

जाँच–किसी भी एक सामान्य समीकरण में बीटा गुणांकों के मान रखकर देखो कि दोनों पक्षों का मान समान है।

(*xix*) निम्न सूत्रों से R^2 तथा R का मान ज्ञात करो–

$$R^2_{1.2345} = \beta_2 r_{12} + \beta_3 r_{13} + \beta_4 r_{14} + \beta_5 r_{15}$$

$$R_{1.2345} = \sqrt{R^2_{1.2345}}$$

(*xx*) निम्न सूत्रों में बीटा गुणांकों तथा मानक विचलनों के मान रखकर बी गुणांकों के मान ज्ञात करो–

$$b_2 = \beta_2 \frac{s_1}{s_2} \qquad b_3 = \beta_3 \frac{s_1}{s_3} \qquad b_4 = \beta_4 \frac{s_1}{s_4} \qquad b_2 = \beta_5 \frac{s_1}{s_5}$$

(*xxi*) निम्न समीकरण की सहायता से स्थिरांक K का मान ज्ञात करो–

$K = M_1 - [b_2 M_2 + b_3 M_3 + b_4 M_4 + b_5 M_5]$

उपरोक्त ढंग से प्रतीपगमन गुणांकों तथा स्थिरांक का मान ज्ञात करने के उपरान्त प्रतीपगमन समीकरण (Regremion Equation) सरलता से लिखी जा सकती है। स्पष्ट है कि पीछे वर्णित उदाहरण के लिए प्रतीपगमन समीकरण निम्नवत प्राप्त होगी–

$$X_1 = 1.6389\, X_2 + .4413\, X_3 + .1379\, X_4 + .2963\, X_5 + 7.3245$$

इस समीकरण की सहायता से किसी भी ऐसे प्रयोज्य के X_1 चर पर प्राप्तांकों का पूर्वकथन किया जा सकता है जिसके X_2, X_3, X_4 तथा X_5 चरों पर प्राप्तांक ज्ञात हों।

प्रतीपगमन विश्लेषण के उपरोक्त वर्णित सोपान पाँच चरों (एक आश्रित चर तथा चार स्वतन्त्र चरों) के लिए हैं। इनको पाँच से अधिक चरों के लिए भी विस्तृत (Extend) किया जा सकता है। चरों की संख्या अधिक होने पर पंक्तियों तथा स्तम्भों की संख्या बढ़ जायेगी। नि:सन्देह सम्पूर्ण गणना कार्य चक्रों (cycles) में विभक्त है। गणना चक्रों की संख्या स्वतन्त्र चरों की संख्या के बराबर होती है। प्रत्येक चक्र नये चर के साथ सहसम्बन्ध गुणांकों की प्रविष्टि से प्रारम्भ होता है तथा उस चक्र की अन्तिम पंक्ति के प्रथम स्तम्भ में – 1.00 के प्राप्त होने तक चलता है। पीछे वर्णित उदाहरण में X_1 को आश्रित चर तथा X_2, X_3, X_4 व X_5 को स्वतन्त्र चरों के रूप में लिया गया था। किसी भी चर को आश्रित चर लेकर प्रतीपगमन विश्लेषण किया जा सकता है, परन्तु आश्रित चर की गणना सारणी में सदैव ही स्तम्भ से ठीक पूर्व वाले स्तम्भ में रखा जाता है।

बहु-सहसम्बन्ध गुणांक की सार्थकता

(Significance of Multiple Correlation)

बहु-सहसम्बन्ध गुणांक (R) की मानक त्रुटि (Standard Error) निम्न सूत्र से ज्ञात की जा सकती है–

$$\sigma_R = \frac{1-R^2}{\sqrt{N-m-1}}$$

जहाँ R = बहु-सहसम्बन्ध गुणांक

N = प्रतिदर्श का आकार

m = स्वतन्त्र चरों की संख्या

मानक त्रुटि की सहायता से बहु-सहसम्बन्ध गुणांक के लिए विश्वस्तता अन्तराल प्राप्त किये जा सकते हैं। किसी बहु-सहसम्बन्ध गुणांक की सार्थकता के लिए एफ-अनुपात (F-Ratio) की गणना निम्न सूत्र से की जा सकती है–

$$F = \frac{R^2}{1-R^2} \cdot \frac{N-m-1}{m}$$

जहाँ R = बहु-सहसम्बन्ध गुणांक का मान

N = प्रतिदर्श का आकार

m = स्वतन्त्र चरों की संख्या

इस एफ-अनुपात की सार्थकता df = m तथा $N - m - 1$ पर देखी जाती है। यह एफ-अनुपात बताता है कि परिगणित बहु-सहसम्बन्ध गुणांक (R) का मान शून्य से सार्थक रूप से भिन्न है अथवा नहीं। जैसे N = 200 तथा $m = 2$ के लिए यदि R = .45 हो तब इसकी सार्थकता के लिए एफ-अनुपात

$$F = \frac{.45^2}{1-.45^2} \cdot \frac{200-2-1}{2}$$

$$= 25.01$$

क्योंकि $df = 2$ व 197 के लिए .05 तथा .01 स्तरों पर एफ का सारणी मान 3.04 तथा 4.71 है अत: परिगणित $f = 25.01$ का मान .01 स्तर पर सार्थक है। अत: R = .45 का मान .01 स्तर पर सार्थक है।

बहु-सहसम्बन्ध गुणांक (R) की सार्थकता के निर्धारण के लिए परिशिष्ट 7 में दी गई सार्थक बहु-सहसम्बन्ध गुणांकों की सारणी का भी उपयोग किया जा सकता है। इस सारणी में N तथा K के विभिन्न मानों के लिए .05 तथा .01 स्तरों पर सार्थकता के लिए आवश्यक बहु-सहसम्बन्ध गुणांक के न्यूनतम मान दिये गये हैं। जैसे N = 200 तथा $m = 2$ के लिए R का मान कम से कम .172 व .212 होने पर वह क्रमश: .05 व .01 स्तरों पर सार्थक होगा।

4.7 बीटा गुणांकों की सार्थकता (Significance of Beta Coefficient)

प्रमापीकृत प्रतीपगमन गुणांकों (B's) की सार्थकता के लिए टी-परीक्षण का उपयोग किया जा सकता है। बीटा गुणांक की सार्थकता की जाँच के लिए टी-अनुपात (t-Ratio) की गणना निम्न लिखित सूत्र से की जाती है–

$$t = \frac{\beta_{12.34......m}}{\sigma_{\beta_{12.34......m}}}$$

जहाँ बीटा गुणांक की मानक त्रुटि (Standard Error of Beta) की गणना निम्न सूत्र से की जाती है।

$$\sigma_{\beta_{12.34......m}} = \sqrt{\frac{1 - R^2_{1.2345......m}}{\left(1 - R^2_{2.34......m}\right)}}$$

उपरोक्त सूत्र से परिगणित टी-अनुपात की सार्थकता $df = (N-m)$ पर देखी जाती है।

दो बहु-सहसम्बन्ध गुणांकों की तुलना

(Comparison of Two Multiple Correlation Coefficients)

कभी-कभी अधिक स्वतंत्र चरों के लिए परिगणित बहु-सहसम्बन्ध गुणांक की तुलना अपेक्षाकृत कम स्वतंत्र चरों के लिए परिगणित बहु-सहसम्बन्ध गुणांक से करनी होती है। ऐसी स्थिति में एफ-परीक्षण का प्रयोग दोनों बहु-सहसम्बन्ध गुणांकों की तुलना के लिए किया जाता है। तब एफ-अनुपात की गणना निम्न सूत्र से की जाती है–

$$F = \frac{(R_1^2 - R_2^2)(N - m_1 - 1)}{(1 - R_1^2)(m_1 - m_2)}$$

जहाँ R_1 = अधिक स्वतंत्र चरों के लिए बहु-सहसम्बन्ध गुणांक

R_2 = कम स्वतन्त्र चरों के लिए बहु-सहसम्बन्ध गुणांक

m_1 = अधिक स्वतंत्र चरों की संख्या

m_2 = कम स्वतंत्र चरों की संख्या

इस एफ-अनुपात की सार्थकता का निर्धारण $(m_1 - m_2)$ तथा $(N - m_1 - 1)$ मुक्तांशों (df) पर किया जाता है।

आंशिक तथा बहु-सहसम्बन्ध गुणांकों में सम्बन्ध

(Relation between Partial and Multiple Correlation Coefficients)

किन्हीं दिये गये चरों के लिए परिगणित आंशिक तथा बहु-सहसम्बन्ध गुणांकों के मध्य सम्बन्ध को निम्न समीकरण से प्रस्तुत किया जा सकता है–

$$1 - R^2_{1.23} = (1 - r^2_{12})(1 - r^2_{13.2})$$

दो से अधिक स्वतंत्र चरों जैसे 4 स्वतंत्र चरों के होने पर

$$1 - R^2_{1.2345} = (1 - r^2_{12})(1 - r^2_{13.2})(1 - r^2_{14.23})(1 - r^2_{15.234})$$

आंशिक सहसम्बन्ध गुणांक की गणना बीटा अथवा बी गुणांकों की सहायता से निम्न सूत्र के द्वारा की जा सकती है–

$$r^2_{12.345......m} = \beta_{12.345......m} \cdot \beta_{21.345......m}$$

अथवा $$r^2_{12.345......m} = b_{12.345......m} \cdot b_{21.345......m}$$

आंशिक तथा बहु-सहसम्बन्ध गुणांक की मान्यताएँ

(Assumptions of Partial and Multiple Correlation Coefficients)

क्योंकि आंशिक तथा बहु-सहसम्बन्ध गुणांकों की गणना साधारण सहसम्बन्ध गुणांकों के द्वारा जी जाती है, इसलिए आंशिक तथा बहु-सहसम्बन्ध के लिए यह आवश्यक है कि वे सभी साधारण सहसम्बन्ध गुणांक आवश्यक मान्यताओं की पूर्ति करते हों। अत: आंशिक तथा बहु-सहसम्बन्ध गुणांकों की सही व्याख्या के लिए यह अत्यन्त महत्त्वपूर्ण है कि समंक सभी चरों के लिए गुणनफल-आघूर्ण सहसम्बन्ध गुणांक की मान्यताओं को पूरा कर रहे हों।

गुणनफल-आघूर्ण सहसम्बन्ध गुणांक की तरह से आंशिक तथा बहु-सहसम्बन्ध गुणांक की मान्यताओं को निम्न ढंग से लिखा जा सकता है–

(*i*) सभी चरों के लिए एक चरीय वितरण (Univariate Distributions) सामान्य प्रायिकता वक्र (N.P.C.) के अनुरूप है।

(*ii*) सभी द्वि-चरीय वितरणों (Bi-Variate Distributions) के लिए समप्रसरणता (Homoscdasticity) की शर्त पूरी हो रही है।

(*iii*) सभी द्वि-चक्रीय वितरणों (Bi-Variate Distributions) के लिए प्रतीपगमन (Regression) की प्रकृति रेखीय (Linear) है।

आंशिक तथा बहु-सहसम्बन्ध की चर्चा के अन्त में कैनोनिकल सहसम्बन्ध (Canonical Correlation) का उल्लेख करना उचित ही होगा। एक साथ अनेक आश्रित चरों (Dependent Variables) तथा स्वतंत्र चरों (Independent Variable) के मध्य सहसम्बन्ध ज्ञात करने के लिए कैनोनिकल सहसम्बन्ध विश्लेषण (Canonical Correlation Analysis) का प्रयोग किया जाता है। यहाँ यह इंगित करना उचित ही होगा कि बहु-सहसम्बन्ध विश्लेषण वस्तुत: कैनोनिकल सहसम्बन्ध विश्लेषण का ही एक विशिष्ट प्रकार है। कैनोनिकल सहसम्बन्ध विश्लेषण का गणना कार्य अत्यन्त जटिल है इसके लिए कम्प्यूटर का ही प्रयोग किया जाता है।

छात्र क्रियाकलाप (Student Activity)

1. सहसम्बन्ध के कितने प्रकार होते हैं?

2. अर्थ आंशिक सहसम्बन्ध गुणांक का अर्थ स्पष्ट कीजिए।

3. दो बहु-सहसम्बन्ध गुणांकों की तुलना को परिभाषित करो?

3.8 सारांश (Summary)

1. सहसम्बन्ध एक विवरणात्मक सांख्यिकीय माप (descriptive statistics) है जो दो चरों के मध्य पाये जाने वाले सम्बन्ध की मात्रा का विवरण देती है। कुछ प्रकार के चरों में पारस्परिक सम्बन्ध पाया जाता है जैसे मूल्य तथा माँग, उत्पादन तथा रोजगार, मजदूरी तथा मूल्य निर्देशांक आदि चरों में सम्बन्ध होता है।
2. जब दो तथ्यों में एक साथ एक ही दिशा में अथवा विपरीत दिशाओं में परिवर्तन हो तथा एक तथ्य में परिवर्तन दूसरे तथ्य में परिवर्तन का कारण हो तो यह कहा जाता है कि उन दोनों तथ्यों में सहसम्बन्ध है।
3. दो चरों के मध्य सहसम्बन्ध गुणांक एक चर के घटने या बढ़ने पर दूसरे चर के घटने-बढ़ने की प्रवृत्ति को बताता है परन्तु दो चरों के मध्य परिगणित सहसम्बन्ध गुणांक का मान भ्रामक (Misleading) भी हो सकता है।
4. दो चरों के बीच साधारण सहसम्बन्ध गुणांक तथा आंशिक सहसम्बन्ध गुणांक में परस्पर सम्बन्ध अनिश्चित रहता है। यद्यपि प्राय: आंशिक सहसम्बन्ध गुणांक का मान साधारण सहसम्बन्ध गुणांक से कम प्राप्त होता है परन्तु आंशिक सहसम्बन्ध गुणांक का मान साधारण सहसम्बन्ध गुणांक के मान से अधिक भी प्राप्त हो सकता है।
5. किसी आश्रित चर तथा अनेक स्वतन्त्र चरों के मध्य बहु-सहसम्बन्ध गुणांक का मान सदैव ही आश्रित चर तथा स्वतन्त्र चरों के बीच के साधारण सहसम्बन्ध गुणांकों के मान (+ अथवा – पर ध्यान दिये बिना) से अधिक होता है।
6. बहु-सहसम्बन्ध की समस्या को दो से अधिक स्वतन्त्र चरों के लिए भी विस्तारित (Extend) किया जा सकता है। दो से अधिक स्वतन्त्र चरों के होने पर बहु-सहसम्बन्ध गुणांक की गणना कुछ जटिल हो जाती है तब प्रतीपगमन समीकरण के बीटा गुणांकों अथवा बी-गुणांकों की सहायता से ही बहु-सहसम्बन्ध की गणना की जाती है। दो से अधिक स्वतन्त्र चरों की स्थिति में प्रतीपगमन समीकरण के गुणांकों को ज्ञात करने के लिए डूलिटिल विधि (Doolittle Method) अथवा ऐटकिन विधि (Aitkin's Method) का प्रयोग किया जाता है।
7. आंशिक तथा बहु-सहसम्बन्ध गुणांकों की गणना साधारण सहसम्बन्ध गुणांकों के द्वारा जी जाती है, इसलिए आंशिक तथा बहु-सहसम्बन्ध के लिए यह आवश्यक है कि वे सभी साधारण सहसम्बन्ध गुणांक आवश्यक मान्यताओं की पूर्ति करते हों।

अभ्यास-प्रश्न (Exercise Questions)

1. सहसम्बन्ध की परिभाषा देते हुए इसके विभिन्न प्रकारों का विस्तार से वर्णन कीजिए।
2. आंशिक सहसम्बन्ध गुणांक की विवेचना करो।
3. बहु-सहसम्बन्ध गुणांक की व्याख्या कीजिए।
4. बहु-प्रतीगमन से आपका क्या तात्पर्य है?

   $\overline{X}_1 = b_{12.3} X_2 + b_{13.2} X_3 + K$

   में X, b और K को स्पष्ट करो
5. बहु प्रतीपगमन की डूलिटिल विधि का विश्लेषण कीजिए।

सन्दर्भ ग्रन्थ (Reference Books)

इकाई–5

प्रतिचयन वितरण
(Sampling Distribution)

संरचना (Structure)

5.1 उद्देश्य (Objectives)

इस इकाई के अध्ययन के पश्चात विद्यार्थी योग्य होंगे–

- परिमित समष्टि की अवधारणा समझने में।
- t-वितरण और X_2-वितरण का जानने में।
- काई वर्ग परिक्षण तथा वितरण के अन्य प्रकारों को समझने में।

5.2 प्रस्तावना (Introduction)

व्यवहार में समष्टि से केवल एक ही प्रतिदर्श लिया जाता है तथा इस प्रतिदर्श से प्राप्त विभिन्न प्रतिदर्शजों की सहायता से समष्टि के प्राचलों का अनुमान लगाया जाता है। इसके लिए प्रतिदर्शजों के वितरण (Distributions of Statistics), जिन्हें प्रतिचयन वितरण (Sampling Distributions) भी कहते हैं, का

उपयोग किया जाता है। किसी प्रतिदर्शज का प्रतिचयन वितरण वास्तव में समष्टि से एक ही आकार के अनेक रैन्डम प्रतिदर्शों से प्राप्त प्रतिदर्शज मानों का वितरण हैं। अत: यदि किसी समष्टि से एक ही आकार के अनेक भिन्न-भिन्न रैन्डम प्रतिदर्शों का चयन किया जाये तो उनमें से प्रत्येक प्रतिदर्श के लिए किसी चर पर कोई वांछित प्रतिदर्शज मान (जैसे मध्यमान) ज्ञात किये जा सकते हैं। स्पष्ट: उतने ही प्रतिदर्शज मान (जैसे मध्यमान) प्राप्त होंगे जितने प्रतिदर्श छाँटे गये हैं। इस प्रकार से प्राप्त प्रतिदर्शज मान परस्पर कुछ भिन्न हो सकते हैं तथा इन्हें आवृत्ति वितरण के रूप में व्यवस्थित किया जा सकता है। प्रतिदर्शज मानों के इस वितरण को ही प्रतिचयन वितरण (Sampling Distribution) कहते हैं।

प्रतिचयन वितरण के सैद्धान्तिक प्रत्यय को किसी स्थूल उदाहरण से अधिक अच्छे ढंग से स्पष्ट किया जा सकता है। माना कि किसी समष्टि में केवल नौ इकाइयां हैं जिनके किसी चर पर प्राप्तांक क्रमश: 1, 2, 3, 4, 5, 6, 7, 8, व 9 हैं। इस समष्टि का मध्यमान 5.0 तथा मानक विचलन 2.58 है। नौ प्राप्तांकों वाली इस समष्टि (1, 2, 3, 4, 5, 6, 7, 8 व 9) से यदि दो-दो प्राप्तांकों वाले प्रतिदर्श लिये जायें तब 36 विभिन्न प्रतिदर्श (प्रतिस्थापन रहित) सम्भव है। यह 36 प्रतिदर्श तथा उनके लिए माध्यमान निम्नवत् होंगे–

सारणी–1

विभिन्न प्रतिदर्श एवं उनके लिए मध्यमान

प्रतिदर्श प्राप्तांक	मध्यमान	प्रतिदर्श प्राप्तांक	मध्यमान	प्रतिदर्श प्राप्तांक	मध्यमान	प्रतिदर्श प्राप्तांक	मध्यमान
1, 2	1.5	2, 4	3.0	3, 7	5.0	5, 7	6.0
1, 3	2.0	2, 5	3.5	3, 8	5.5	5, 8	6.5
1, 4	2.5	2, 6	4.0	3, 9	6.0	5, 9	7.0
1, 5	3.0	2, 7	4.5	4, 5	4.5	6, 7	6.5
1, 6	3.5	2, 8	5.0	4, 6	5.0	6, 8	7.0
1, 7	4.0	2, 9	5.5	4, 7	5.5	6, 9	7.5
1, 8	4.5	3, 4	3.5	4, 8	6.0	7, 8	7.5
1, 9	5.0	3, 5	4.0	4, 9	6.5	7, 9	8.0
2, 3	2.5	3, 6	4.5	5, 6	5.5	8, 9	8.5

यदि सारणी 1 में दिखाये गये इन छत्तीस प्रतिदर्शों से प्राप्त मध्यमानों का आवृत्ति वितरण तैयार किया जाये तो निम्नानुसार वितरण प्राप्त होगा।

सारणी–2

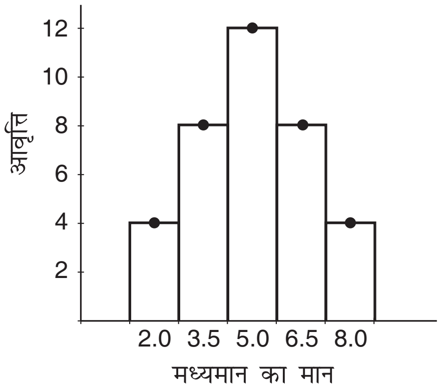
मध्यमान	आवृत्ति	मध्यमान	आवृत्ति	मध्यमान	आवृत्ति
8.5	1	6.0	3	3.5	3
8.0	1	5.5	4	3.0	2
7.5	2	5.0	4	2.5	2
7.0	2	4.5	4	2.0	1
6.5	3	4.0	3	1.5	1

इस वितरण को ही मध्यमान का प्रतिचयन वितरण (Sampling Distribution of Mean) कहेंगे। इस वितरण को वर्गीकृत करके सारणी 3 में प्रस्तुत किया गया है तथा इसके सापेक्ष आवृत्ति दण्डाकृति को चित्र 1 में दर्शाया है।

सारणी–3

मध्यमान का प्रतिचयन वितरण (Sampling Distribution of Mean)

वर्ग	आवृत्ति
7.5–8.5	4
6.0–7.0	8
4.5–5.5	12
3.0–4.0	8
1.5–2.5	4
	K = 36

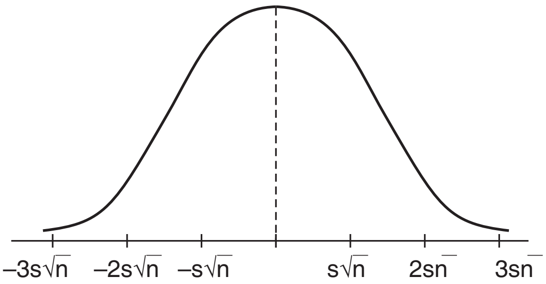


चित्र 1

मध्यमान के उपरोक्त प्रतिचयन वितरण से दो बातें स्पष्ट हैं–(*i*) यह एक सममित या समाकृत वितरण (Symmetrical Distribution) है जिसके केन्द्र में आवृत्ति अधिक हैं व किनारों की और आवृत्तियाँ कम होती जा रही हैं, तथा (*ii*) इस वितरण का मध्यमान 5.0 है जोकि समष्टि के मध्यमान के बराबर है। इसी प्रकार से अन्य प्रतिदर्शजों जैसे मध्यांक, मानक विचलन, सहसम्बन्ध गुणांक अदि के लिए भी प्रतिचयन वितरण तैयार किये जा सकते हैं। अनुमानात्मक सांख्यिकीय विधियों के द्वारा प्राचलों का अनुमान लगाने में प्रतिदर्शज वितरणों का विशेष उपयोग है। वास्तव में प्रतिदर्शज वितरणों की विशेषताओं (Characteristics) के आधार पर ही प्राचलों का अन्तराल अनुमान लगाना सम्भव है। अधिकांश प्रतिदर्शज वितरणों में दो विशेषताऐं पाई जाती है (*i*) प्रतिचयन वितरण का मध्यमान मूल समष्टि के मध्यमान के बराबर होता है तथा (*ii*) प्रतिदर्शज के आकार (n) के बड़ा होने पर प्रतिचयन वितरण का रूप सामान्य प्रायिकता वितरण (NPC) के अनुरूप होता जाता है। प्रतिचयन वितरणों की इन विशेषताओं को केन्द्रीय सीमा प्रमेय (Central Limit Theorem) के रूप में प्रस्तुत किया जाता है, जोकि प्रायिकता सिद्धान्त की एक अत्यन्त महत्त्वपूर्ण व अनोखी विशेषता है। मध्यमान के लिए केन्द्रीय सीमा प्रमेय को निम्न ढँग से लिखा जा सकता है।

"यदि किसी समष्टि का मध्यमान μ तथा मानक विचलन σ हो तो इस समष्टि से समान आकार (n) के छाँटे गये रैन्डम प्रतिदर्शों का वितरण, n के मान के पर्याप्त बड़ा होने पर, ऐसे सामान्य प्रायिकता वितरण के अनुरूप होता है जिसका मध्यमान μ तथा मानक विचलन $\sigma / \sqrt{n}$ के बराबर होता है।"

नोट



चित्र 2. प्रतिदर्श मध्यमानों का वितरण

किसी प्रतिदर्शज के प्रतिचयन वितरण के मानक विचलन को उस प्रतिदर्शज की मानक त्रुटि (Standard Error) कहते हैं। अत: मध्यमान को मानक त्रुटि $\sigma/\sqrt{n}$ के बराबर होती है। मानक त्रुटि को SE अथवा चिह्न के बाद सम्बन्धित प्रतिदर्शज का संकेताक्षर प्रत्यय (Suffix) के रूप में लिखकर इंगित किया जाता है। अत: मध्यमान की मानक त्रुटि

$$\sigma E_M \text{ या } \sigma_M = \frac{\sigma}{\sqrt{n}}$$

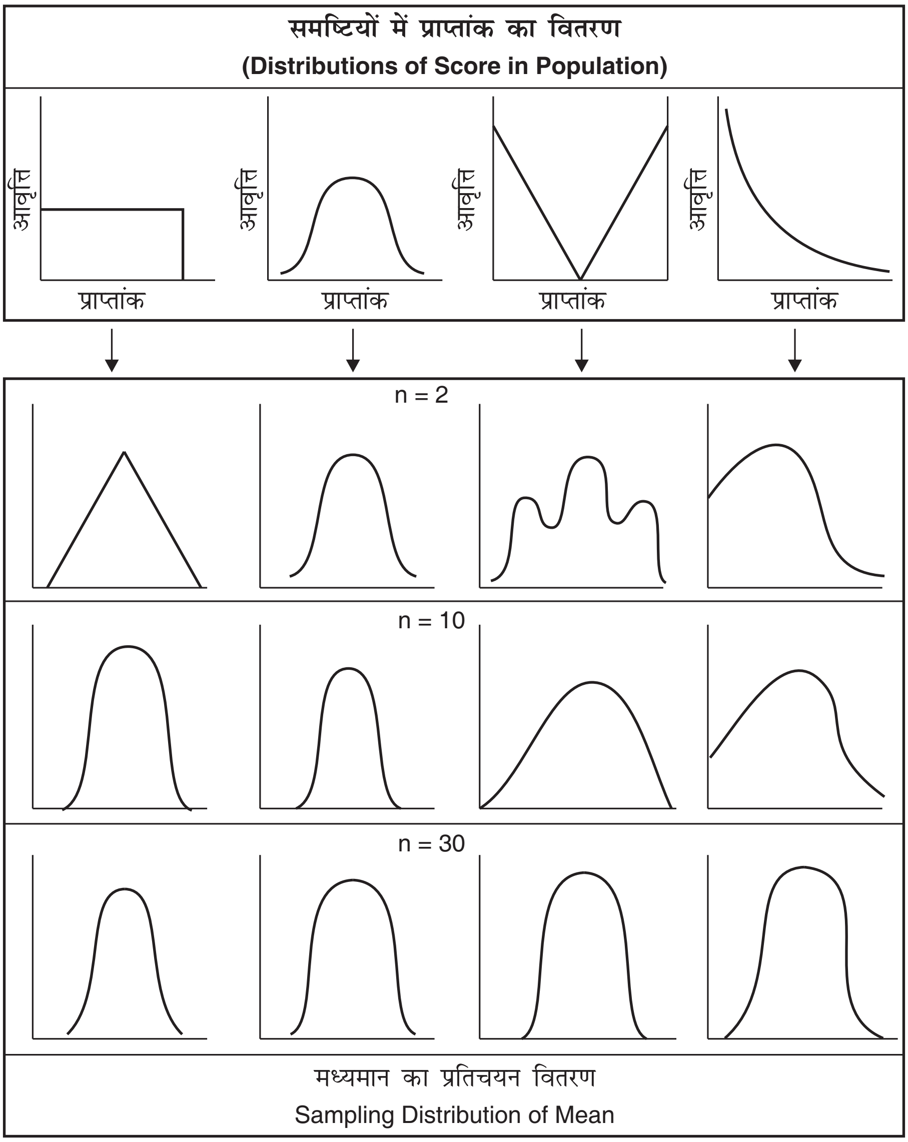
केन्द्रीय सीमा प्रमेय के सम्बन्ध में दो बातें महत्त्वपूर्ण हैं। प्रथम, प्रतिचयन वितरण प्रतिदर्श के आकार n के बड़ा होने पर समष्टि के वितरण की प्रकृति से प्रभावित नहीं होता है। समष्टि में प्राप्तांकों का वितरण किसी भी प्रकार का क्यों न हो प्रतिदर्श के आकार को बढ़ाने पर प्रतिचयन वितरण सामान्य प्रायिकता वक्र के अनुरूप होता जाता है। यह तथ्य चित्र 3 से स्पष्ट हो सकेगा। द्वितीय, प्रतिचयन वितरण के मानक विचलन अर्थात सम्बन्धित प्रतिदर्शज की मानक त्रुटि के सूत्र में प्रयुक्त σ समष्टि का मानक विचलन है न कि प्रतिदर्श का मानक विचलन। समष्टि के मानक विचलन को ज्ञात करना लगभग असम्भव होने के कारण σ का अनुमान प्रतिदर्श के मानक विचलन (s), जोकि ज्ञात होता है, से लगाते हैं क्योंकि प्रतिदर्श समष्टि का एक अंश (Part) होता है। इसलिए सम्भावना यही है कि किसी रैन्डम प्रतिदर्श की विचलनशीलता अपनी समष्टि से कम ही होगी जिसके परिणामस्वरूप प्रतिदर्श के मानक विचलन के सदैव ही समष्टि के मानक विचलन से कम होने की ही सम्भावना होती है। दूसरे शब्दों में कहा जा सकता है प्रतिदर्श का मानक विचलन अपनी समष्टि के मानक विचलन को कम करके आँकता (Underestimate) है। प्रतिदर्श के मानक विचलन (s) से असीमित समष्टि के मानक विचलन (σ) का सर्वोत्तम अनुमान (Best Estimate) निम्न सूत्र की सहायता से लगाया जा सकता है–

$$\sigma = s \cdot \sqrt{\frac{n}{n-1}}$$

जब प्रतिदर्श का आकार (n) पर्याप्त बड़ा (प्राय: 30 से अधिक) होता है तब $\sqrt{n/(n-1)}$ का मान लगभग एक के बराबर ही होता है। जैसे के 30 होने पर यह मान 1.017, n के 50 होने पर 1.010, n के 100 होने पर 1.005, n के 500 होने पर 1.001 तथा n के 1000 होने पर 1.0005 होगा। इसलिए n के बड़ा होने पर समष्टि के मानक विचलन के स्थान पर प्रतिदर्श के मानक विचलन का बिना किसी विशेष त्रुटि के प्रयोग किया जा सकता है परन्तु जब प्रतिदर्श का आकार n छोटा होता है तब σ के स्थान पर s का प्रयोग त्रुटिपूर्ण होगा क्योंकि तब $\sqrt{n/(n-1)}$ का मान एक से काफी भिन्न होगा जो σ के अनुमान को काफी प्रभावित करेगा। अत: n के छोटा होने पर या तो मानक त्रुटि के सूत्र में समष्टि के मानक विचलन का मान रखना चाहिए अथवा सूत्र को प्रतिदर्श के मानक विचलन के अनुरूप संशोधित कर लेना चाहिए। समष्टि के मानक विचलन का प्रतिदर्श प्राप्तांकों से सीधे अनुमान निम्न सूत्र के प्रयोग से लगाया जा सकता है–

$$\sigma = \sqrt{\frac{\Sigma x^2}{n-1}} \quad \text{अथवा} \quad \sigma = \sqrt{\frac{\Sigma(X - M)}{n-1}}$$

नोट



चित्र 3. विभिन्न प्रकार की समष्टियों से लिये प्रतिदर्शों के लिए मध्यमान का प्रतिचयन वितरण

जहाँ X प्रतिदर्श के प्राप्तांक, M प्रतिदर्श का मध्यमान तथा n प्रतिदर्श का आकार है। समष्टि के मानक विचलन (σ) के स्थान पर प्रतिदर्श के मानक विचलन (s) का प्रयोग करने पर मध्यमान की मानक त्रुटि के सूत्र को निम्नवत लिखा जा सकता है–

नोट

सारणी–4

विभिन्न प्रतिदर्शजों की मानक त्रुटियाँ (Standard Errors of Various Statistics)

प्रतिदर्शज Statistic	मानक त्रुटि के लिए संकेताक्षर Symbol for SE	मानक त्रुटि सूत्र Standard Error Formula
1. मध्यमान	SE_M या σ_m	$\frac{\sigma}{\sqrt{n}}$
2. मध्यांक	SE_{MD} या σ_{md}	$\frac{1.253\sigma}{\sqrt{n}}$
3. मानक विचलन	SE_σ या σ_σ	$\frac{\sigma}{\sqrt{2n}}$ या $\frac{.707\sigma}{\sqrt{n}}$
4. चतुर्थांक विचलन	SE_Q या σ_Q	$\frac{.7867\sigma}{\sqrt{n}}$
5. आवृत्ति	SE_f या σ_f	$\sqrt{npq}$
6. प्रतिशत	$SE_{\%}$ या $\sigma_{\%}$	$\sqrt{\frac{PQ}{N}}$
7. अनुपात	SE_p या σ_p	$\sqrt{\frac{pq}{n}}$
8. श्रेणीक्रम सहसम्बन्ध गुणांक	SE_p या σ_p	$\frac{1}{\sqrt{n-1}}$
9. गुणनफल-आघूर्ण सहसम्बन्ध गुणांक	SE_r या σ_r	$\frac{1-r^2}{\sqrt{n-1}}$

विशेष–प्रतिदर्श के छोटा होने अर्थात n के 30 से कम होने पर मानक त्रुटि के प्रथम चार सूत्रों में प्रयुक्त हर (Denominaters) में n के स्थान पर (n – 1) रखना होगा जिससे समष्टि के मानक विचलन σ के स्थान पर प्रतिदर्श के मानक विचलन (s) का प्रयोग बिना किसी त्रुटि के किया जा सकें।

माध्यमान के प्रतिचयन वितरण की तरह से मध्यांक, मानक विचलन, चतुर्थांक विचलन, प्रतिशत, अनुपात आदि के लिए केन्द्रीय सीमा प्रमेय का विस्तार किया जा सकता है परन्तु इन सभी के प्रतिचयन वितरणों का मानक विचलन अर्थात इनकी मानक त्रुटि ज्ञात करने के सूत्र भिन्न-भिन्न होंगे जिन्हें सारणी 4 में प्रस्तुत किया गया है। यहाँ यह स्मरणीय है कि मानक विचलन किसी चर पर प्राप्तांकों के वितरण की विचलनशीलता का द्योतक है जबकि मानक त्रुटि किसी प्रतिदर्शज जैसे मध्यमान, मध्यांक, मानक विचलन आदि के विभिन्न मानों के वितरण की विचलनशीलता का परिचायक है। मानक त्रुटियों के सूत्रों से स्पष्ट है कि मानक त्रुटियाँ समष्टि की विचलनशीलता अर्थात मानक विचलन के समानुपाती तथा प्रतिदर्श के आकार के विलोमानुमाती होता है। दूसरे शब्दों में कहा जा सकता है कि मानक विचलन के बढ़ने तथा प्रतिदर्श के आकार के घटने पर मानक त्रुटि बढ़ जाती है।

5.3 परिमित समष्टि (Finite Population)

मानक त्रुटियों के लिए सारणी 4 में वर्णित सभी सूत्र अपरिमित समष्टियों (Infinite Populations) से लिए गये प्रतिदर्शों के लिए हैं। जब समष्टि की तुलना में प्रतिदर्श काफी छोटा होता है तब भी ये सूत्र मानक त्रुटि

का सही अनुमान देते हैं परन्तु जब प्रतिदर्श समष्टि की तुलना में काफी छोटा नहीं होता है तब इन सूत्रों का प्रयोग उचित नहीं है। यद्यपि सामाजिक विज्ञानों में प्रयुक्त प्रतिदर्श प्राय: समष्टि की तुलना में बहुत ही छोटे होते हैं तथा कभी-कभी तो विशेषकर प्रयोगात्मक अनुसंधानों में समष्टि पूर्ण रूपेण विद्यमान (Exist) भी नहीं रहती है। जैसे बालकों के किसी प्रतिदर्श से प्राप्त परिणामों को उन बालकों पर भी लागू किया जा सकता है जिनका अभी जन्म भी नहीं हुआ है। इसलिए अधिकांश स्थितियों में सारणी 4 में प्रस्तुत किये गये मानक त्रुटि के सूत्रों का प्रयोग बिना किसी कठिनाई के किया जा सकता है परन्तु जब समष्टि परिमित हो तथा प्रतिदर्श व समष्टि का अनुपात पर्याप्त हो तब सांख्यिकीयविद् को सजग रहने की आवश्यकता होती है। वस्तुत: परिमित समष्टियों से छाँटे गये प्रतिदर्शों के लिए मानक त्रुटि के सूत्रों में कुछ आवश्यक संशोधन करने की आवश्यकता होती है। परिमित समष्टियों की स्थिति में मानक त्रुटि कुछ कम होती है। ऐसी स्थिति में मानक त्रुटि न केवल प्रतिदर्श के आकार बल्कि समष्टि के आकार अथवा इन दोनों के अनुपात से भी प्रभावित होती है। आकार N की सीमित से आकार n के प्रतिदर्श के लिए मध्यमान की मानक त्रुटि निम्न सूत्र से ज्ञात की जा सकती है–

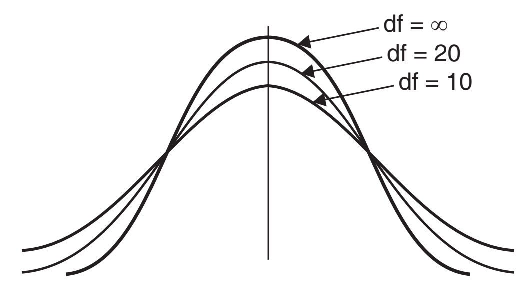
$$\sigma_M = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

जहाँ σ समष्टि के लिए मानक विचलन है

उपरोक्त सूत्र से स्पष्ट है कि जब $N = n$ अर्थात सम्पूर्ण समष्टि को प्रतिदर्श के रूप में ले लेते है तब σ_M का मान शून्य हो जाता है। जब N की तुलना में n बहुत छोटा होता है तब $\sqrt{(N-n)/(N-1)}$ का मान लगभग 1 के बराबर हो जाता है जिसे सूत्र से हटाया जा सकता है। परिमाणत: यह सूत्र वही हो जाता है जो अपरिमित समष्टियों से लिए प्रतिदर्शों के लिए होता है।

5.4 टी-वितरण (t-Distribution)

यद्यपि केन्द्रीय सीमा प्रमेय (Central Limit Theorem) के द्वारा बड़े प्रतिदर्शों के प्रतिचयन वितरणों की प्रकृति को स्पष्ट किया जा सकता है परन्तु छोटे प्रतिदर्शों के लिए केन्द्रीय सीमा प्रमेय के अनुसार प्रतिचयन



कतिपय मुक्तांशों (df n) के लिए टी वितरण

चित्र 4

वितरण का सामान्य प्रायिकता वक्र के रूप में स्वीकार करना बहुधा उचित नहीं होता है। सन् 1908 में डब्लू. एस. गॉसेट (William Sealy Gossett) ने पाया कि यद्यपि बड़े प्रतिदर्शों के माध्यमानों का वितरण सामान्य रूप से वितरित होता है किन्तु जैसे-जैसे प्रतिदर्श का आकार छोटा करते जाते हैं वैसे-वैसे प्रतिचयन वितरण सामान्य वितरण से दूर हटता जाता है तथा यह वितरण अधिक नुकीला (Peaked) होता जाता है। दूसरे शब्दों में प्रतिचयन वितरण की वक्रता (Kurtosis) बढ़ती जाती है जिससे वह अधिक वक्रीय (Leptokurtic) होता जाता है। गॉसेट ने $(M - \mu)/s\sqrt{n}$ का वितरण तैयार किया जिसे टी वितरण (t-Distribution) कहते हैं। यहाँ यह ज्ञातव्य है कि

गॉसेट ब्रिटेन के डब्लिन (Dublin) शहर में स्थित गुइनेज शराब फैक्ट्री (Guiness Brewery) में गणितज्ञ के रूप में कार्यरत था परन्तु फैक्ट्री की शर्त के अनुसार फैक्ट्री का कोई भी कर्मचारी अपने नाम से कोई गणितीय लेख प्रकाशित नहीं कर सकता था। इसलिए गॉसेट ने स्टूडैन्ट (Student) के छद्म नाम (Pseudonym) से प्रकाशित अपने शोध पत्र (The Probable Error of a Mean) में अपने विचार प्रस्तुत किये थे। बाद में सन् 1926 में आर. ए. फिशर (R. A. Fisher) ने गॉसेट के द्वारा सुझाया t वितरण प्रस्तुत किया। स्टूडैन्ट के नाम पर प्रस्तुत किये गये t-वितरण को स्टूडैन्ट का t वितरण (Student's t-Distribution) के नाम से भी पुकारा जाता है।

टी वितरण एक सतत् वितरण (Continuous Distribution) है जो सैद्वान्तिक रूप से $-\infty$ से $+\infty$ तक फैला होता है। यह $t = 0$ के सापेक्ष सममित (Symmetrical) होता है तथा कुछ अधिक फैला होने के अतिरिक्त काफी हद तक सामान्य प्रायिकता वक्र के जैसा होता है। टी वितरण का आकार प्रतिदर्श के आकार n पर निर्भर करता है इसीलिए विभिन्न आकारों के प्रतिदर्शों के लिए t–वितरण भिन्न-भिन्न होते हैं। प्रतिदर्श के आकार n के बढ़ते जाने पर t–वितरण का रूप सामान्य प्रायिकता वक्र के नजदीक आता जाता है तथा n के बहुत बड़ा होने पर यह सामान्य वक्र जैसा हो जाता है। विभिन्न आकारों के लिए टी–वितरण भिन्न-भिन्न होने के कारण सभी टी–वितरणों को प्रस्तुत करना लगभग असम्भव कार्य है। परिशिष्ट 5 में दी गई टी–वितरण सारणी में कुछ चुने गये सार्थकता स्तरों के लिए विभिन्न मुक्तांश (df) पर t के मान दिये गये है। मुक्तांश (df) एक तकनीकी पद है जिसकी चर्चा आगे की गई है। यहाँ पर केवल इतना जानना पर्याप्त है कि मध्यमानों के सन्दर्भ में df का मान $(n-1)$ होता है।

गॉसेट के द्वारा टी–वितरण के प्रस्तुत किये जाने पर इसके अनेक उपयोगों की खोज की गई। आधुनिक समय में छोटे प्रतिदर्शों के लिए मध्यमान तथा अन्य प्रतिदर्शजों के प्रतिचयन वितरणों के रूप में सामान्य प्रायिकता वितरण के स्थान पर टी–वितरण का उपयोग अधिक उपयुक्त माना जाता है।

5.5 X_2-वितरण और इसका उपयोग (X_2-distribution and their uses)

जब 'मध्यमान शब्द का प्रयोग किया जा रहा है तब उपलब्ध समंक (Available Data) मात्रात्मक प्रकृति (Quantitativ Nature) का होता है जिसे कम से कम (At least) मापन के अन्तरिक स्तर (Interval Level) पर मापा गया होता है। परन्तु कभी-कभी अनुसंधान कार्य में गुणात्मक प्रकृति (Qualitative Nature) के समंक उपलब्ध होते हैं। ऐसी स्थिति में चरों को नामित स्तर (Nominal Level) अथवा क्रमित स्तर (Ordinal Level) पर मापा गया होता है। कभी-कभी किन्हीं परिस्थितिजन्य कारणों से अन्तरित स्तर पर मापे चरों को क्रमित स्तर पर परिवर्तित करके समंक विश्लेषण करना पड़ता है। यहाँ यह ज्ञातव्य है कि नामित या क्रमित स्तरों पर प्रयोज्यों (Subjects) को उनमें विद्यमान किसी विशिष्ट गुण के आधार पर कुछ वर्गों (Categories) में विभक्त किया जाता है तथा विभिन्न वर्गों में आये प्रायोज्यों की संख्या की गणना करके आवृत्तियाँ ज्ञात की जाती हैं। क्रमित स्तर पर प्रयोज्यों को जिन वर्गों में बाँटा जाता है उनमें श्रेष्ठता या हीनता का एक अन्तर्निहित क्रम (Inherent Order) भी होता है। दूसरे शब्दों में कहा जा सकता है कि नामित तथा क्रमित स्तरों पर उपलब्ध समंक कम से कम दो या अधिक वर्गों (At least two or more than two Categories) की आवृत्तियों के रूप में होते हैं। जैसे छात्रों को लिंगभेद, अथवा सामाजिक-आर्थिक स्तर, अथवा बौद्धिक स्तर, अथवा परीक्षा परिणाम, अथवा अधिगम शैली, अथवा व्यक्तित्व प्रकार आदि के आधार पर कुछ वर्गों में विभक्त करके गिना (count) जा सकता है। इसी प्रकार से नागरिकों को राजनैतिक दलों के प्रति प्रतिबद्धता, अथवा किसी सामाजिक परिवर्तन के सम्बन्ध में राय, अथवा शैक्षिक स्तर अथवा व्यवसाय अथवा धर्म, अथवा जाति आदि के आधार पर विभिन्न वर्गों में बाँटा जा सकता है। विभिन्न वर्गों में सम्मिलित व्यक्तियों की संख्याओं को उन वर्गों की आवृत्तियाँ (Frequencies) कहते हैं तथा f संकेताक्षर से निर्दिष्ट करते हैं। आवृत्तियों के रूप में उपलब्ध इस प्रकार के समंकों का विश्लेषण करने के लिए पूर्व वर्णित सांख्यिकीय परीक्षणों का प्रयोग नहीं किया जा सकता है ऐसी स्थिति में प्राय: काई–वर्ग परीक्षण (Chi-Square Test) का प्रयोग किया जाता है। काई–परीक्षण एक अत्यन्त महत्त्वपूर्ण अप्राचलिक सांख्यिकीय

प्रविधि (Non-Parametric Statistical Technique) है। सम्भवत: असतत वर्गों (Discrete Categories) तथा आवृत्तियों (Frequencies) के रूप में उपलब्ध समंकों के लिए सर्वाधिक उपयुक्त (Most Appropriate) तथा बहुतायत से प्रयुक्त (Widely Used) परीक्षण यही है। सामाजिक विज्ञानों के अनुसंधान कार्यों में काई-वर्ग परीक्षण का महत्त्व तथा उपयोग सर्वाधिक प्रयुक्त होने वाले कुछ गिन-चुने सांख्यिकीय प्रत्ययों जैसे सामान्य प्रायिकता वक्र, टी-परीक्षण एवं एफ-परीक्षण के समतुल्य स्वीकार किया जाता है।

काई (χ) वास्तव में ग्रीक भाषा (Greek Language) का एक अक्षर (Latter) है। काई-वर्ग वितरण (Chi-Square Distribution) की खोज सन् 1875 में हेलमर्ट (Helmert) ने की थी तथा बाद में सन् 1900 में कार्ल पियर्सन (Karl Pearson) ने पुन: इसका प्रतिपादन स्वतन्त्र रूप से किया एवं इसके आधार पर काई-वर्ग परीक्षण (Chi-Square Test) को विकसित किया। काई-वर्ग के गणितीय विवेचन के कुछ जटिल होने तथा व्यावहारिक उपयोग में इस गणितीय विवेचन की कोई विशेष आवश्यकता न होने के कारण, प्रस्तुत अध्याय में, काई-वर्ग के गणितीय विवेचन का वर्णन नहीं किया जा रहा हैं। काई-वर्ग परीक्षण वस्तुत: आवृत्तियों के किसी अवलोकित वितरण (Observed Distribution) की तुलना किसी पूर्व निश्चित परिकल्पना के आधार पर तैयार किये गये परिकल्पित वितरण (Hypothetical Distribution) जिसे सैद्धान्तिक वितरण (Theoritical Distribution) अथवा प्रत्याशित वितरण (Expected Distribution) भी कहते हैं, से करता है। इस तुलना में देखा जाता है कि अवलोकित वितरण तथा परिकल्पित वितरण के विभिन्न वर्गों में स्थित आवृत्तियों में सहमति (Agreement) अथवा विभिन्नता (Divergence) किस सीमा तक है। अवलोकित वितरण तथा परिकल्पित वितरण की आवृत्तियों के मध्य दृष्टिगोचित अन्तर की सार्थकता के लिए काई-वर्ग की गणना की जाती है। अत: काई-वर्ग परीक्षण के द्वारा ज्ञात किया जाता है कि क्या कोई दिया गया अवलोकित वितरण किसी परिकल्पना के आधार पर तैयार किये गये प्रत्याशित वितरण से सार्थक रूप से भिन्न है अथवा नहीं। इस प्रश्न का उत्तर प्रकारान्तर में उस परिकल्पना की स्वीकृति अथवा अस्वीकृति को इंगित करता है जिसके आधार पर प्रत्याशित वितरण तैयार किया गया है। उदाहरण के लिए जैसे यदि किसी सामाजिक परिवर्तन के प्रति रुझान को जानने के लिए 100 प्रयोज्यों के किसी प्रतिदर्श में 25 ने पूर्ण सहमति, 30 ने सहमति, 12 ने अनिश्चित, 15 ने असहमति तथा 18 ने पूर्ण असहमति व्यक्त की। यदि परिकल्पना यह है कि समष्टि में प्रयोज्यों की प्रतिक्रियाएँ पाँचों वर्गों में समान रूप से वितरित हैं अर्थात पाँचों प्रतिक्रिया वर्गों के लिए आवृत्तियाँ समान हैं। इस परिकल्पना के आधार पर पाँचों वर्गों में 20-20 प्रयोज्य होने चाहिए थे। स्पष्ट है कि यहाँ पर अवलोकित वितरण की आवृत्तियाँ क्रमश: 25, 30, 12, 15 व 18 हैं जबकि प्रत्याशित वितरण की आवृत्तियाँ क्रमश: 20, 20, 20, 20 व 20 हैं। प्रश्न यह है कि क्या अवलोकित वितरण की आवृत्तियों तथा प्रत्याशित वितरण की आवृत्तियों में अन्तर प्रतिचयन त्रुटि (Sampling Error) के कारण हैं अथवा वास्तव में अन्तर है। इसी प्रकार से यदि अनुसंधानकर्ता जानना चाहता है कि क्या किसी चर पर प्रतिदर्श से प्राप्त समंक रूप से वितरित हैं, तब भी उसे अवलोकित आवृत्ति वितरण की तुलना सामान्य वितरण की परिकल्पना पर आधारित प्रत्याशित वितरण से करके देखना होगा कि दोनों वितरणों में दृष्टिगोचित अन्तर प्रतिचयन त्रुटि के कारण प्रायिक हो सकता है अथवा वास्तव में अन्तर है। ऐसी परिस्थितियों में काई-वर्ग परीक्षण का प्रयोग किया जा सकता है। स्पष्ट है कि काई-वर्ग परीक्षण आवृत्तियों के मध्य अन्तर की सार्थकता का परीक्षण करता है। इसमें काई-वर्ग (Chi-Square) नाम के गुणांक की गणना की जाती है, जिसका सूत्र निम्नवत है—

$$\textbf{काई-वर्ग},\ \chi^2 = \Sigma \frac{(f_0 - f_e)^2}{f_e}$$

जहाँ f_0 किसी वर्ग की अवलोकित आवृत्ति (Observed Frequency) है तथा f_e उस वर्ग की प्रत्याशित आवृत्ति (Expected Frequency) है एवं Σ का चिह्न सभी वर्गों के लिए $(f_0 - f_e)^2/f_e$ पदों के योग को इंगित करता है। काई-वर्ग के सूत्र से स्पष्ट है कि इसका मान अवलोकित तथा प्रत्याशित आवृत्तियों के अन्तर के वर्गों का सम्बन्धित प्रत्याशित आवृत्तियों से अनुपातों का योग होता है। अवलोकित तथा प्रत्याशित आवृत्तियों के अन्तरों के शून्य होने पर अर्थात् अवलोकित तथा प्रत्याशित आवृत्तियों के समान होने पर काई-वर्ग का मान शून्य हो जाता है जबकि

अन्तरों के अधिक होने पर काई-वर्ग का मान बढ़ता जाता है। काई-वर्ग के सूत्र से यह भी स्पष्ट है कि काई-वर्ग का मान सदैव ही धनात्मक (+ve) होता हैं। अत: काई-वर्ग का मान शून्य अथवा शून्य से अधिक ही हो सकता है। क्योंकि काई-वर्ग की गणना में अवलोकित तथा प्रत्याशित आवृत्तियों के अन्तर के वर्गों $(f_0 - f_e)^2$ का प्रयोग होता है इसलिए यह अन्तर की दिशा (Direction of Difference) पर कोई ध्यान नहीं देता है।

काई-वर्ग परीक्षण में परिगणित काई-वर्ग की सार्थकता का निर्धारण काई-वर्ग के प्रतिचयन वितरण (Sampling Distribution of Chi-Square) के आधार पर प्रायिकता के रूप में (In Terms of Probability) किया जाता है। काई-वर्ग का प्रतिचयन वितरण केवल अवलोकित वितरण के वर्गों के लिए मुक्तांशों (dfs) पर निर्भर करता है। दूसरे शब्दों में काई-वर्ग के प्रतिचयन वितरण पर प्रतिदर्श के आकार का कोई प्रभाव नहीं पड़ता है। मुक्तांशों के बहुत कम होने पर काई-वर्ग का प्रतिचयन वितरण धनात्मक रूप से विषम (Positively Skewed) होता है, परन्तु जैसे-जैसे मुक्तांशों (df) का मान बढ़ता जाता है, यह सामान्य वितरण की ओर प्रवृत (Tend) होने लगता है। विभिन्न मुक्तांशों (dfs) पर काई-वर्ग प्रतिचयन वितरणों के लिए कुछ चुने हुए सार्थकता स्तरों पर काई-वर्ग के मान परिशिष्ट 8 में दिए गये हैं। यदि परिगणित काई-वर्ग (χ^2) का मान सम्बन्धित मुक्तांशों (dfs) तथा वांछित सार्थकता स्तर पर सारणी मान से कम होता है तब प्राप्त काई-वर्ग को असार्थक कहा जाता है। असार्थक काई-वर्ग का अर्थ है कि अवलोकित वितरण तथा प्रत्याशित वितरण के बीच की भिन्नता संयोगवश हुई प्रतिचयन त्रुटि के कारण माना जा सकता है एवं अवलोकित वितरण तथा प्रत्याशित वितरण के मध्य अन्तर न होने के सांख्यिकीय शून्य परिकल्पना (H_o) को स्वीकार किया जा सकता है। परिणामत: समष्टि में वितरण को प्रत्याशित वितरण के अनुरूप माना जा सकता है तथा उस परिकल्पना को, जिसके आधार पर प्रत्याशित वितरण तैयार किया गया था, स्वीकार किया जा सकता है। इसके विपरीत यदि परिमाणित काई-वर्ग (χ^2) का मान सम्बन्धित मुक्तांशों (dfs) तथा वांछित सार्थकता स्तर पर सारणी मान से अधिक होता है तब प्राप्त काई-वर्ग को सार्थक कहा जाता है। सार्थक काई-वर्ग का अर्थ है कि अवलोकित तथा प्रत्याशित वितरण के बीच दृष्टिगोचित भिन्नता को संयोगवश हुई प्रतिचयन त्रुटि के कारण स्वीकार नहीं किया जा सकता है एवं अवलोकित तथा प्रत्याशित वितरण के बीच अन्तर न होने की सांख्यिकीय शून्य परिकल्पना (H_o) को निरस्त किया जा सकता है। परिणामत: समष्टि में वितरण को प्रत्याशित वितरण के अनुरूप नहीं माना जा सकता है तथा उस परिकल्पना को, जिसके आधार पर प्रत्याशित वितरण तैयार किया गया था, अस्वीकार (Reject) किया जा सकता है।

वैसे तो प्रत्याशित वितरण किसी भी परिकल्पना के आधार पर तैयार किया जा सकता है तथा उसकी तुलना अवलोकित वितरण से की जा सकती है परन्तु साधारणत: तीन प्रकार की परिकल्पनाओं के आधार पर प्रत्याशित आवृत्ति विवरण तैयार किये जाते हैं। ये तीन परिकल्पनाएँ हैं–

(*i*) **समान रूप से वितरित होने की परिकल्पना**

(*ii*) **सामान्य रूप से वितरित होने की परिकल्पना**

(*iii*) **स्वतन्त्र रूप से वितरित होने की परिकल्पना**

समान रूप से तथा सामान्य रूप से वितरित होने की दोनों परिकल्पनाएँ एकमार्गी अथवा एक चर वितरण के लिए हैं। इस प्रकार की परिकल्पनाओं का अभीष्ट यह देखना होता है कि क्या अवलोकित वितरण को सम्बन्धित परिकल्पना के अनुरूप वितरित माना जा सकता है अथवा नहीं। इस प्रकार की परिस्थिति में काई-वर्ग परीक्षण वास्तव में **समरूपता आँकलन परीक्षण** या **समानुकता परीक्षण** (Test of Goodness of Fit) अथवा **सहमति परीक्षण** (Agreement Test) है जो अवलोकित वितरण/आवृत्तियों (f_o's) तथा प्रत्याशित वितरण/आवृत्तियों (f_o's) के परस्पर समान (fit) होने की मात्रा (Goodness) को ज्ञात करता है। स्वतन्त्र रूप से वितरित होने की तीसरी परिकल्पना द्वि-मार्गी वितरण के लिए होती हैं। जब किसी समूह को एक साथ दो-नामित, क्रमित, अन्तरित अथवा अनुपातिक चरों पर बाँटा जाता है तब प्राप्त आवृत्ति वितरण को द्वि-मार्गी अथवा द्वि-चर आवृत्ति वितरण कहते हैं। अब यदि दोनों चरों के परस्पर स्वतन्त्र (Independent) होने अथवा परस्पर निर्भर (Dependent) होने की परिकल्पना की जाँच करनी होती है तब प्रत्याशित आवृत्ति वितरण दोनों चरों के स्वतन्त्र रूप से आवृत्तियों के वितरित

होने के आधार पर तैयार किया जाता है। इस प्रकार की परिकल्पनाओं का अभीष्ट यह देखना होता है कि क्या अवलोकित द्वि-मार्गी वितरण दोनों चरों के परस्पर स्वतन्त्र होने पर प्राप्त होने वाले प्रत्याशित वितरण के अनुरूप माना जा सकता है अथवा नहीं। इस प्रकार की परिस्थिति में काई-वर्ग परीक्षण वास्तव में **स्वतन्त्रता परीक्षण** (Test of Independence) होता है जो दोनों चरों के मध्य सम्बन्ध (Association) की सार्थकता को ज्ञात करता है। दो समूहों के प्राप्त आवृत्ति वितरणों की तुलना के लिए भी काई-वर्ग परीक्षण का उपयोग किया जा सकता है। वस्तुत: काई-वर्ग का यह उपयोग प्रकारान्तर से स्वतन्त्र रूप से वितरित होने की परिकल्पना का ही एक विस्तार (Extension) है जिसमें एक चर के लिए दोनों समूहों के लिए आवृत्ति वितरण ज्ञात किया जाता है तथा जिस आधार पर दोनों समूहों में विभेद होता है, उसे दूसरा चर मान लिया जाता है। तत्पश्चात् दोनों चरों के परस्पर स्वतन्त्र होने की परिकल्पना की जाँच काई-वर्ग परीक्षण से की जाती है। जैसे छात्र तथा छात्राओं के द्वारा कुछ चुने गये खेलों के प्रति पसन्द के लिए बने आवृत्ति वितरणों की तुलना के लिए काई-वर्ग परीक्षण का प्रयोग किया जा सकता है। इस स्थिति में लिंगभेद (Sex Difference) को एक चर मानकर, लिंगभेद तथा खेलों के प्रति पसन्द के लिए द्वि-मार्गी आवृत्ति वितरण तैयार करके स्वतन्त्र वितरण की परिकल्पना की जाँच की जा सकती है। दो समूहों की तुलना के लिए परिगणित काई-वर्ग के सार्थक होने का तात्पर्य होता है कि दोनों समूहों के वितरणों में सार्थक अन्तर है जबकि परिगणित काई-वर्ग के असार्थक होने से संकेत मिलता है कि दोनों समूहों के वितरणों में अन्तर सार्थक नहीं है।

5.6 काई-वर्ग परीक्षण में मुक्तांश (Degrees of Freedom in chi-Square Test)

आवृत्ति वितरण के रूप में प्रस्तुत समंकों के लिए मुक्तांशों की संख्या उन वर्गों (Categories) की संख्या के बराबर होती है जिनमें पंक्ति व स्तम्भों के योगों के रूप में आरोपित वाह्य बन्धनों को सन्तुष्ट करते हुए, इच्छानुसार (Arbitrary) आवृत्तियाँ रखी जा सकती हैं। जैसा कि पीछे स्पष्ट किया जा चुका है, आवृत्ति वितरण एक-मार्गी (One-way) तथा द्वि-मार्गी (Two-way) हो सकते हैं। एकमार्गी आवृत्ति वितरण में केवल एक ही पंक्ति होती है तथा वाह्य रूप से आरोपित बंधन मात्र विभिन्न वर्गों (Categories) में सम्मिलित आवृत्तियों का योग अर्थात् कुल आवृत्ति (N) होता है। यदि कुल आवृत्तियों (N) को K वर्गों में बाँटकर आवृत्ति वितरण बनाया गया होता है तब कुल योग (N) को स्थिर रखते हुए केवल (K – 1) वर्गों में इच्छानुसार आवृत्तियाँ आवंटित की जा सकती हैं परन्तु अन्तिम वर्ग की आवृत्ति को इस प्रकार से समायोजित करना होगा कि सभी K वर्गों की आवृत्तियों का योग N के बराबर ही हो। अत: इस प्रकार की स्थिति में मुक्तांशों (df) की संख्या (K – 1) के बराबर होती है, जहाँ K वर्गों की संख्या है। जैसे यदि किसी अवलोकित वितरण में 400 छात्र संकाय के आधार पर पाँच वर्गों में सारणी 1 के अनुसार वितरित हैं तब यहाँ पर मुक्तांशों (df) का मान $5 - 1 = 4$ होगा क्योंकि अनुसंधानकर्ता को 400 का योग बनाये रखते हुए केवल चार वर्गों में मनचाही आवृत्ति ही रखने की छूट हो सकती है। पाँचवें वर्ग की आवृत्ति को कुल योग के अनुरूप रखना होगा। यहाँ पर यह ध्यान रखना होगा कि योग वाले वर्ग को मुक्तांश (df) ज्ञात करते समय वर्गों की संख्या में नहीं गिना जाता है।

सारणी–5

400 छात्रों का संकायाधारित आवृत्ति वितरण

कला संकाय	विज्ञान संकाय	वाणिज्य संकाय	शिक्षा संकाय	विधि संकाय	योग
100	80	90	60	70	400

द्विमार्गी अथवा द्विचर आवृत्ति वितरण में आवृत्तियाँ कुछ पंक्तियों तथा कुछ स्तम्भों में वितरित होती हैं। इस प्रकार के वितरण में न केवल आवृत्तियों का कुल योग वरन विभिन्न स्तम्भों व पंक्तियों के योग भी वाह्य रूप से आरोपित बंधन होते हैं। यदि आवृत्ति वितरण में पंक्तियों की संख्या (r) तथा स्तम्भों की संख्या (c) होती है तब विभिन्न पंक्तियों तथा स्तम्भों की आवृत्तियों के योगों को स्थिर रखते हुए केवल $(r - 1)(c - 1)$ प्रकोष्ठों को (Cells)

में इच्छानुसार आवृत्तियाँ आबंटित की जा सकती हैं जबकि शेष $(r + c - 1)$ प्रकोष्ठों की आवृत्तियों को इस प्रकार से समायोजित करना होगा कि विभिन्न पंक्तियों तथा स्तम्भों की आवृत्तियों का योग अवलोकित आवृत्ति वितरण के अनुरूप हो सकें। अत: इस प्रकार की स्थिति में मुक्तांशों (df) की संख्या $(r-1)(c-1)$ के बराबर होती हैं, जहाँ r पंक्तियों की संख्या व c स्तम्भों की संख्या है। जैसे यदि किसी द्वि-मार्गी अवलोकित वितरण में बौद्धिक स्तर के लिए पाँच वर्गों तथा सामाजिक-आर्थिक स्तर के लिए तीन वर्गों से बने द्वि-मार्गी वितरण के विभिन्न प्रकोष्ठों में कुल 500 व्यक्ति सारणी 164 के अनुरूप वितरित हों, तब ऐसे द्विमार्गी वितरण के लिए मुक्तांशों (df) की संख्या $(3-1)(5-1) = 8$ होगी,

सारणी–6

व्यक्तित्व-प्रकार तथा सामाजिक-आर्थिक स्तर के लिए 500 व्यक्तियों का द्वि-मार्गी आवृत्ति वितरण

सामाजिक आर्थिक स्तर	बौद्धिक स्तर					
	अत्यन्त निम्न	निम्न	सामान्य	उच्च	अत्यन्त उच्च	योग
उच्च	30	25	20	15	30	120
मध्यम	40	45	30	25	40	180
निम्न	30	50	40	40	40	200
योग	**100**	**120**	**90**	**80**	**110**	**500**

क्योंकि अनुसंधानकर्ता विभिन्न पंक्तियों तथा स्तम्भों का योग बनाये रखते हुए 15 प्रकोष्ठों में से केवल 8 प्रकोष्ठों में मनचाही आवृत्ति रख सकता है, शेष 7 प्रकोष्ठों की आवृत्तियों को पंक्ति व स्तम्भ योगों के अनुरूप रखना होगा। यहाँ पर भी यह ध्यान रखना होगा कि योग वाली पंक्ति तथा योग वाले स्तम्भ को मुक्तांश ज्ञात करते समय गिना नहीं जाता है।

काई-वर्ग परीक्षण के सोपान (Steps of Chi-Square Test)

काई-वर्ग परीक्षण की (Chi-Square Test) सम्पूर्ण प्रक्रिया (Process) को निम्न सोपानों (Steps) में बाँटा जा सकता है–

(*i*) आवृत्ति वितरण के सम्बन्ध में परिकल्पना का निर्माण करना।

(*ii*) आवृत्ति वितरण सम्बन्धी परिकल्पना के आधार पर प्रत्याशित आवृत्तियों (f_e) को ज्ञात करके प्रत्याशित आवृत्ति वितरण तैयार करना।

(*iii*) सांख्यिकीय शून्य परिकल्पना (H_0), कि अवलोकित वितरण तथा प्रत्याशित वितरण में अन्तर नहीं है, का निर्माण करना।

(*iv*) सार्थकता स्तर का चयन करना।

(*v*) मुक्तांशों (df) को ज्ञात करना।

(*vi*) वांछित सार्थकता स्तर तथा सम्बन्धित मुक्तांशों (df) के लिए काई-वर्ग का सारणी मान ज्ञात करना।

(*vii*) अग्रांकित सूत्र से काई-वर्ग की गणना करना–

$$\chi^2 = \Sigma \frac{(f_0 - f_e)^2}{f_e}$$

(*viii*) परिगणित काई-वर्ग (χ^2) की सार्थकता देखकर, शून्य परिकल्पना की स्वीकृति या अस्वीकृति के सम्बन्धों में निर्णय लेना।

(*ix*) शून्य परिकल्पना की स्वीकृति/अस्वीकृति के आधार पर आवृत्ति वितरण के सम्बन्धों में बनाई गई परिकल्पना को स्वीकार अथवा निरस्त करना।

काई-वर्ग परीक्षण के ये विभिन्न सोपान आगे दिये गये उदाहरणों से स्पष्ट हो सकेंगे। यहाँ यह इंगित करना उचित ही होगा कि उपरोक्त नौ सोपानों में से कुछ सोपानों को विशेषकर शून्य परिकल्पना के निर्माण के सोपान को प्राय: औपचारिक रूप से लिखा नहीं जाता है। काई-वर्ग परीक्षण में वितरण की परिकल्पना तथा शून्य परिकल्पना एक ही बात को दो अलग-अलग ढंग से व्यक्त करती हैं।

5.7 समान वितरण की परिकल्पना (Hypothesis of Equal Distribution)

कभी-कभी अनुसंधानकर्ता यह ज्ञात करने का इच्छुक होता है कि क्या विभिन्न वर्गों में अवलोकित आवृत्तियों को समान रूप से वितरित स्वीकार किया जा सकता है अथवा वे परस्पर सार्थक रूप से भिन्न-भिन्न हैं। इस प्रश्न का उत्तर समान वितरण की परिकल्पना के आधार पर तैयार किये गये प्रत्याशित वितरण (Expected Distribution) की तुलना अवलोकित वितरण (Observed Distribution) से काई-वर्ग परीक्षण के द्वारा करके प्राप्त किया जा सकता है। समान वितरण की परिकल्पना को समान प्रायिकता की परिकल्पना (Hypothesis of Equal Probability) भी कहते हैं। इस परिकल्पना के अनुसार समष्टि के अन्दर विभिन्न वर्गों (Categories) में आवृत्तियाँ समान हैं। दूसरे शब्दों में कह सकते हैं कि जब समष्टि में आवृत्तियाँ विभिन्न वर्गों में समान रूप से वितरित होने की परिकल्पना की जाती है तब इसे समान वितरण की परिकल्पना कहते हैं। स्पष्ट: यह परिकल्पना इंगित करती है कि प्रतिदर्श में भी आवृत्तियाँ समान रूप से वितरित होनी चाहिए, परन्तु संयोगवश हुई प्रतिचयन त्रुटि के कारण प्राय: ऐसा नहीं होता है, तब अनुसंधानकर्ता को अवलोकित आवृत्ति वितरण के समान रूप से वितरित होने की परिकल्पना का परीक्षण करने की आवश्यकता होती है। ऐसी स्थिति में काई-वर्ग परीक्षण के द्वारा अवलोकित तथा प्रत्याशित वितरणों के अन्तर की सार्थकता की जाँच की जाती है। क्योंकि समष्टि में आवृत्तियों के समान रूप से वितरित होने की परिकल्पना की जाँच करनी है, इसलिए प्रतिदर्श के प्रत्याशित वितरण के विभिन्न वर्गों (Categories) की प्रत्याशित आवृत्तियाँ (f_e) समान होनी चाहिए। विभिन्न वर्गों की प्रत्याशित आवृत्ति (f_e) ज्ञात करने के लिए कुल आवृत्तियों के योग (N) को वर्गों की संख्या (K) से भाग देते हैं। अत: समान वितरण की परिकल्पना में प्रत्याशित वितरण के सभी वर्गों की प्रत्याशित आवृत्तियों (f_e) का मान N/K के बराबर होता है। प्रत्याशित वितरण (Expected Distribution) को ज्ञात करने के उपरान्त विभिन्न वर्गों के लिए f_0 व f_e का अन्तर अर्थात् ($f_0 - f_e$) ज्ञात करके, इसका वर्ग कर लेते हैं तथा f_e से भाग दे देते हैं। सभी वर्गों के लिए इस प्रकार से ज्ञात किये गये $(f_0 - f_e)^2/f_e$ पदों का योग ही काई-वर्ग (χ^2) का मान होता है जिसकी सार्थकता $df = \text{K} - 1$ पर काई-वर्ग के सार्थक मानों की सारणी से वांछित सार्थकता स्तर पर निर्धारित की जा सकती है। समान वितरण की परिकल्पना के परीक्षण की सम्पूर्ण प्रक्रिया आगे उदाहरणों की सहायता से स्पष्ट की जा रही है।

उदाहरण–500 छात्रों के एक रैन्डम प्रतिदर्श में 300 लड़के व 200 लड़कियाँ थीं। क्या समष्टि में लड़के-लड़कियों के वितरण को समान रूप से वितरित माना जा सकता है?

हल–क्योंकि समान वितरण की परिकल्पना है, इसलिए प्रत्याशित वितरण में लड़के व लड़कियों की आवृत्तियाँ समान होनी चाहिए। कुल 500 छात्र हैं, इसलिए दोनों ही वर्गों की प्रत्याशित आवृत्ति 250-250 होगी। अवलोकित तथा प्रत्याशित वितरण के अन्तर की सार्थकता का परीक्षण काई-वर्ग परीक्षण (Chi-Square Test) से किया जा सकता है। यहाँ पर शून्य परिकल्पना होगी कि दोनों वितरणों में अन्तर नहीं है। अत: शून्य परिकल्पना (Null Hypothesis) को संकेताक्षरों में निम्न ढंग से लिख सकते हैं–

$$\text{H}_{0\text{-}} : f_{0_i} - f_{e_i}$$

सार्थकता स्तर = .01

मुक्तांश (df) = K – 1 = 2 – 1 = 1

सारणी मान, $df = 1$ के लिए .01 स्तर पर $\chi^2 = 6.64$

अब अवलोकित तथा प्रत्याशित आवृत्ति वितरणों की तुलना के लिए काई-वर्ग की गणना की जायेगी। काई-वर्ग की गणना कार्य सारणी 3 में प्रस्तुत है।

सारणी–7

काई-वर्ग की गणना

वर्ग	अवलोकित आवृत्ति f_0	प्रत्याशित आवृत्ति f_e	$f_0 - f_e$	$(f_0 - f_e)^2$	$\frac{(f_0 - f_e)^2}{f_e}$
लड़के	300	250	50	2500	10
लड़कियाँ	200	250	50	2500	10
	N = 500	क्योंकि $\chi^2 = \Sigma \frac{(f_0 - f_e)^2}{f_e}$		अत:	$\chi^2 = 20$

क्योंकि परिगणित $\chi^2 = 20$ का मान $df = 1$ के लिए .01 स्तर पर काई-वर्ग के सारणी मान 6.64 से अधिक है, अत: यह .01 स्तर पर सार्थक है। अत: शून्य परिकल्पना, कि अवलोकित तथा प्रत्याशित वितरणों में अन्तर नहीं है, को .01 स्तर पर निरस्त किया जा सकता है। परिणामत: समान वितरण की परिकल्पना को भी निरस्त किया जाता है। अत: कहा जा सकता है कि समष्टि में लड़के तथा लड़कियों की आवृत्तियाँ समान स्वीकार नहीं की जा सकती हैं।

5.8 सामान्य वितरण की परिकल्पना (Hypothesis of Normal Distribution)

अनुसंधानों में प्राय: अनुसंधानकर्ता को यह ज्ञात करने की आवश्यकता होती है कि क्या किसी चर पर प्राप्त समंकों को सामान्य प्रायिकता वक्र (N.P.C) के अनुरूप वितरित स्वीकार किया जा सकता है अथवा वे सामान्य वितरण से सार्थक रूप से भिन्न हैं। इस प्रश्न का उत्तर सामान्य प्रायिकता वक्र के अनुरूप वितरण की परिकल्पना के आधार पर तैयार किये गये प्रत्याशित वितरण (Expected Distribution) की तुलना में अवलोकित वितरण (Observed Distribution) से काई-वर्ग परीक्षण के द्वारा करके दिया जा सकता है। यदि परिगणित काई-वर्ग का मान सार्थक होता है तब शून्य परिकल्पना, कि अवलोकित वितरण तथा सामान्य वितरण की परिकल्पना के आधार पर तैयार प्रत्याशित वितरण में अन्तर नहीं है, को निरस्त कर दिया जाता है जो प्रकारान्तर में इंगित करता है कि अवलोकित वितरण को सामान्य वितरण के समरूप नहीं माना जा सकता है। इसके विपरीत यदि परिगणित काई-वर्ग का मान असार्थक होता है तब उपरोक्त शून्य परिकल्पना को स्वीकार कर लिया जाता है जो प्रकारान्तर में इंगित करता है कि अवलोकित वितरण को सामान्य वितरण के अनुरूप माना जा सकता है। सामान्य वितरण की परिकल्पना को सामान्य प्रायिकता की परिकल्पना (Hypothesis of Normal Probability) भी कहते हैं। इस से तात्पर्य है कि समष्टि मे प्राप्तांक सामान्य प्रायिकता वक्र की तरह से वितरित हैं। सामान्य वितरण की परिकल्पना तब ही सम्भव हो सकती है जब समंक कम से कम क्रमित स्तर (Ordinal Level) अथवा अन्तरित स्तर (Interval Level) पर संकलित किये गये हों। इस प्रकार की परिकल्पना दो परिस्थितियों में सम्भव हो सकती हैं–(*i*) जब समंक वर्गान्तरों (Class Intervals) की आवृत्तियों के रूप में व्यवस्थित होते हैं अथवा (*ii*) जब समंक अन्तरित स्तर पर बने क्रमित वर्गों (Categories) की आवृत्तियों के रूप में दिये होते हैं। पहली प्रकार की परिस्थिति का उदाहरण परीक्षण प्राप्तांकों के लिए बना आवृत्ति वितरण है जबकि दूसरी प्रकार की परिस्थिति का उदाहरण छात्रों को अध्यापकों के द्वारा दिये विभिन्न ग्रेडों (Grades) के लिए बना आवृत्ति वितरण है। इन दोनों ही परिस्थितियों में काई-वर्ग परीक्षण का प्रयोग वितरण की समान्यता की जाँच के लिए किया जाता है। किसी अवलोकित वितरण की

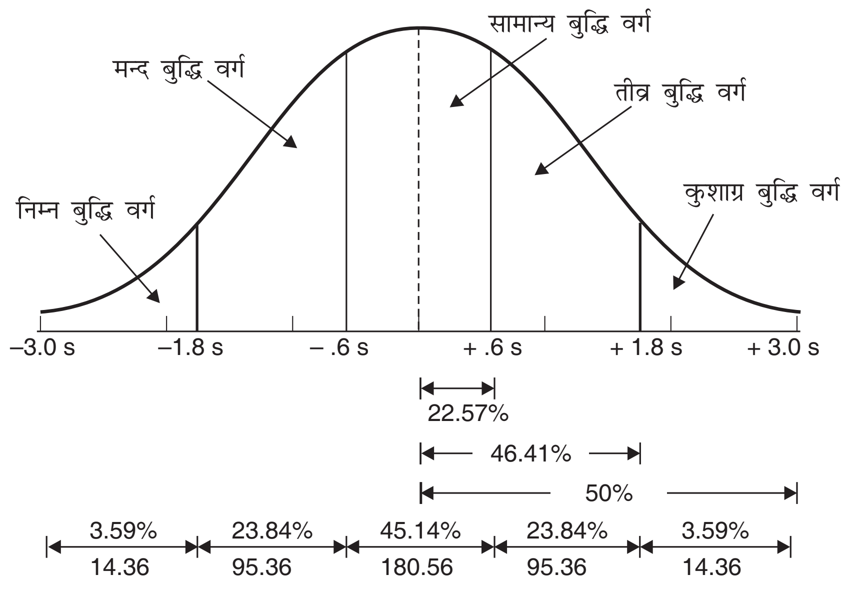
तुलना उसके सापेक्षिक बने सामान्य प्रायिकता वितरण से करने के लिए सर्वप्रथम सामान्य वक्र सारणी की सहायता से प्रत्याशित वितरण तैयार करना होता है। यहाँ यह स्पष्ट कर देना उचित ही होगा कि तैयार किये गये प्रत्याशित आवृत्ति वितरण के लिए मध्यमान तथा मानक विचलन का मान अवलोकित आवृत्ति वितरण के मध्यमान तथा मानक विचलन के बराबर होना चाहिए तथा वितरण की प्रकृति सामान्य प्रायिकता वक्र के अनुरूप होनी चाहिए।

सामान्य प्रायिकता वितरण की परिकल्पना के अनुरूप प्रत्याशित आवृत्ति वितरण ज्ञात करने के उपरान्त पूर्ववत् ढंग से काई-वर्ग की गणना करके उसकी सार्थकता की जाँच की जाती है। सार्थक काई-वर्ग मान शून्य परिकल्पना तथा सामान्य प्रायिकता वितरण की परिकल्पना की अस्वीकृति का द्योतक होता है जबकि असार्थक काई-वर्ग मान शून्य परिकल्पना तथा सामान्य वितरण की परिकल्पना की स्वीकृति का द्योतक होता है। काई वर्ग परीक्षण के द्वारा सामान्य वितरण की परिकल्पना के परीक्षण की सम्पूर्ण प्रक्रिया आगे उदाहरणों के द्वारा स्पष्ट की गई है।

उदाहरण–400 छात्रों के एक प्रतिदर्श को उनकी बुद्धि के आधार पर पाँच क्रमित वर्गों में निम्नवत् ढंग से बाँटा गया। क्या छात्रों का यह वितरण सामान्य प्रायिकता वितरण के अनुरूप माना जा सकता है?

वर्ग	**कुशाग्र बुद्धि**	**तीव्र बुद्धि**	**सामान्य बुद्धि**	**मन्द बुद्धि**	**निम्न बुद्धि**
आवृत्ति	20	85	200	80	15

हल–(i) यहाँ पर अवलोकित वितरण की तुलना सामान्य वितरण की परिकल्पना के आधार पर बने प्रत्याशित वितरण से करनी है, इसलिए सर्वप्रथम 400 छात्रों के इस प्रतिदर्श को सामान्य प्रायिकता वक्र के अनुरूप पाँच वर्गों में विभक्त करके प्रत्याशित वितरण तैयार करना होगा। क्योंकि व्यावहारिक समस्याओं के लिए सामान्य प्रायिकता वक्र (NPC) के विस्तार को $-3Z$ से $+3Z$ तक विस्तृत माना जा सकता है, इसलिए $6Z$ की इस दूरी को पाँच बराबर भागों में बाँटने पर प्रत्येक भाग में $1.2Z$ दूरी होगी। इन पाँचों वर्गों का विभाजन तथा सामान्य प्रायिकता वक्र सारणी से विभिन्न वर्गों की प्रत्याशित आवृत्तियाँ (f_e) निम्न चित्र के अनुसार होंगी।



चित्र 5. 400 छात्रों का सामान्य प्रायिकता वक्र के अनुसार पाँच भागों में विभाजन

(*ii*) स्पष्ट: काई-वर्ग परीक्षण के द्वारा जाँच की जाने वाली शून्य परिकल्पना होगी कि अवलोकित वितरण तथा सामान्य रूप से वितरित प्रत्याशित वितरण की आवृत्तियों में अन्तर नहीं है अर्थात्

$$H_{ai} : a_i = f_{ei}$$

(*iii*) सार्थकता स्तर = .05 तथा .01 एवं $df = K - 1 = 5 - 1 = 4$, अत: $df = 4$ के लिए .05 तथा .01 स्तरों पर काई-वर्ग सारणी देखने पर,

$$\chi^2 = .05\ (df = 4) = 9.488 \text{ तथा } \chi^2\ .01\ (df = 4) = 13.277$$

(*iv*) अवलोकित तथा प्रकाशित आवृत्ति वितरणों की तुलना हेतु काई-वर्ग सारणी बनाने पर

सारणी–8

काई-वर्ग की गणना

सोपान \ वर्ग	निम्न बुद्धि	मन्द बुद्धि	सामान्य बुद्धि	तीव्र बुद्धि	कुशाग्र बुद्धि	योग
f_0	15.00	80.00	200.00	85.00	20.00	400
f_e	14.36	95.36	180.56	95.36	14.36	400
$f_0 - f_e$	0.64	15.36	19.44	10.36	5.64	
$(f_0 - f_e)^2$	0.41	235.93	377.91	107.33	31.81	
$\frac{(f_0 - f_e)^2}{f_e}$	$\frac{.41}{14.36} = .03$	$\frac{235.93}{95.36} = 2.47$	$\frac{377.91}{180.56} = 2.09$	$\frac{107.33}{95.36} = 1.13$	$\frac{31.81}{14.36} = 2.22$	**7.94**

$$\chi^2 = \Sigma \frac{(f_0 - f_e)^2}{f_e}$$
$$= .03 + 2.47 + 2.09 + 1.13 + 2.22$$
$$= 7.94$$

(*v*) क्योंकि परिगणित $\chi^2 = 7.94$ का मान .05 व .01 स्तरों पर सार्थकता के लिए आवश्यक न्यूनतम सारणी मान 9.488 व 13.277 से कम है, अत: यह सार्थक नहीं है। प्राप्त असार्थक काई-वर्ग के आधार पर शून्य परिकल्पना को स्वीकार करते हुए कहा जा सकता है कि क्योंकि अवलोकित वितरण तथा प्रत्याशित वितरण के बीच दृष्टिगोचित अन्तर संयोगवश आ सकता है, इसलिए छात्रों के अवलोकित वितरण को सामान्य प्रायिकता वितरण के अनुरूप माना जा सकता है।

5.9 स्वतन्त्र वितरण की परिकल्पना (Hypothesis of Independent Distribution)

स्वतन्त्र वितरण की परिकल्पना को स्वतन्त्रता की परिकल्पना (Hypothesis of Independence) भी कहते हैं। कभी-कभी अनुसांधनकर्ता यह जानना चाहता है कि क्या आवृत्तियों के रूप में मापे गये दो चर परस्पर स्वतन्त्र (Independent) या असम्बन्धित (Unrelated) है अथवा नहीं। इस प्रश्न का उत्तर काई-वर्ग परीक्षण के द्वारा दिया जा सकता है। इस प्रकार की परिस्थिति में आवृत्ति वितरण द्वि-मार्गी होता है अर्थात् दो चरों के लिए एक साथ बने प्रकोष्ठों (Cells) में अवलोकित आवृत्तियाँ स्थित होती हैं। इन अवलोकित आवृत्तियों की तुलना स्वतन्त्र वितरण या स्वतन्त्रता की परिकल्पना के आधार पर प्रत्याशित आवृत्तियों से करने के लिए काई-वर्ग परीक्षण का प्रयोग किया जाता है। काई-वर्ग परीक्षण बताता है कि क्या अवलोकित वितरण तथा प्रत्याशित वितरण का

आवृत्तियों में अन्तर संयोगवश हुई प्रतिचयन त्रुटि के कारण है अथवा नहीं। स्वतन्त्रता की परिकल्पना के आधार पर द्विचर वितरण के विभिन्न प्रकोष्ठों (Cells) के लिए प्रत्याशित आवृत्तियाँ ज्ञात करने के लिए उस प्रकोष्ठ के स्तम्भ आवृत्ति योग तथा पंक्ति आवृत्ति योग की गुणा करके प्राप्त गुणनफल को कुल आवृत्ति योग (N) से भाग दे देते हैं। अत: सूत्र रूप में कह सकते हैं कि iवीं पंक्ति तथा jवें स्तम्भ से बने प्रकोष्ठ के लिए प्रत्याशित आवृत्ति–

$$f_e = \frac{f_r \times c_j}{\mathrm{N}}$$

जहाँ $r_i = i$वीं पंक्ति की आवृत्तियों का योग

$c_j = j$वें स्तम्भ की आवृत्तियों का योग

N = कुल आवृत्तियों का योग

इस सूत्र से सभी प्रकोष्ठों के लिए प्रत्याशित आवृत्तियों की गणना कर लेते हैं। द्विचर आवृत्ति वितरण में प्रत्याशित आवृत्तियों (f_e's) को प्राय: कोष्ठकों (Brackets) मे लिखते हैं।

सभी प्रकोष्ठों के लिए प्रत्याशित आवृत्तियों (f_e's) को ज्ञात करने के उपरान्त काई-वर्ग के सूत्र का प्रयोग करके काई-वर्ग की गणना कर लेते हैं। यदि द्विचर आवृत्ति वितरण में पंक्तियों की संख्या r तथा स्तम्भों की संख्या c होती है तब प्राप्त काई वर्ग के लिए मुक्तांशों की संख्या, $df = (r-1)(c-1)$ होती है। अत: प्राप्त काई-वर्ग की सार्थकता वांछित स्तर पर $(r-1)(c-1)$ मुक्तांशों के लिए देखी जाती है। सार्थक काई-वर्ग मान शून्य परिकल्पना तथा स्वतन्त्र वितरण की परिकल्पना की अस्वीकृति का जबकि असार्थक काई-वर्ग मान स्वीकृति का परिचायक होता है। दो या अधिक आवृत्ति वितरणों की तुलना के लिए काई-वर्ग का प्रयोग भी इसी ढंग से किया जाता है। तब इसे समानुपातिक आवृत्तियों की परिकल्पना (Hypothesis of Proportional Frequencies) भी कहा जाता है। इस स्थिति में काई-वर्ग परीक्षण से देखा जाता है कि क्या विभिन्न समूहों के लिए आवृत्ति वितरण एक ही अनुपात में है अथवा नहीं। समानुपातिकता की परिकल्पना के परीक्षण के लिए प्रत्याशित वितरण ठीक उसी ढंग से बनाते हैं जिस ढंग से स्वतन्त्र वितरण की परिकल्पना के लिए बनाते हैं। इसीलिए समानुपातिकता की परिकल्पना को प्राय: अलग से नहीं लिखा जाता है। स्वतन्त्र वितरण की परिकल्पना के परीक्षण की सम्पूर्ण विधि आगे उदाहरणों से स्पष्ट की जा रही है।

उदाहरण–एक सौ छात्रों के एक प्रतिदर्श के लिए उनके वार्षिक परीक्षा परिणाम तथा सामाजिक-आर्थिक स्तर चरों के लिए आवृत्ति वितरण निम्नानुसार था। क्या ये समंक परीक्षा-परिणाम के सामाजिक-आर्थिक स्तर से स्वतन्त्र होने की और संकेत करते हैं?

परीक्षा परिणाम सामाजिक-आर्थिक स्तर	**प्रथम श्रेणी**	**द्वितीय श्रेणी**	**तृतीय श्रेणी**	**अनुत्तीर्ण**	**योग**
उच्च सा॰ आ॰ स्तर	6	15	5	4	**30**
औसत सा॰ आ॰ स्तर	5	10	15	10	**40**
निम्न सा॰ आ॰ स्तर	4	10	10	6	**30**
योग	**15**	**35**	**30**	**20**	**100**

हल–(i) क्योंकि अवलोकित वितरण की तुलना स्वतन्त्रता की परिकल्पना (Hypothesis of Independence) के आधार पर बने प्रत्याशित वितरण से करनी है, इसलिए सबसे पहले प्रत्याशित वितरण तैयार करना होगा। प्रत्याशित आवृत्तियों की गणना सारणी 4 में की गई है।

नोट

सारणी–9

प्रत्याशित आवृत्तियों (f_e's) की गणना

परीक्षा परिणाम सामाजिक-आर्थिक स्तर	प्रथम श्रेणी	द्वितीय श्रेणी	तृतीय श्रेणी	अनुत्तीर्ण	योग
उच्च सा॰ आ॰ स्तर	$\frac{30\times15}{100} = 4.5$	$\frac{30\times35}{100} = 10.5$	$\frac{30\times30}{100} = 9.0$	$\frac{30\times20}{100} = 6.0$	30
औसत सा॰ आ॰ स्तर	$\frac{40\times15}{100} = 6.0$	$\frac{40\times35}{100} = 14.0$	$\frac{40\times30}{100} = 12.0$	$\frac{40\times20}{100} = 8.0$	40
निम्न सा॰ आ॰ स्तर	$\frac{30\times15}{100} = 4.5$	$\frac{30\times35}{100} = 10.5$	$\frac{30\times30}{100} = 9.0$	$\frac{30\times20}{100} = 6.0$	30
योग	15	35	30	20	100

(*ii*) काई-वर्ग परीक्षण के द्वारा जाँच की जाने वाली शून्य परिकल्पना है कि "अवलोकित वितरण तथा स्वतन्त्र वितरण की परिकल्पना के आधार पर बनाये गये प्रत्याशित वितरण की आवृत्तियों में अन्तर नहीं है" अर्थात्

$$H_0 : f_{0_i} = f_{e_i} = 0$$

(*iii*) सार्थकता स्तर = .05 तथा .01 एवं $df = (r-1)(c-1) = (3-1)(4-1) = 6$, अतः $df = 6$ पर काई-वर्ग सारणी देखने पर–

$$\chi^2 = .05\ (df = 6) = 12.592 \quad \chi^2 = .01\ (df = 6) = 16.812$$

(*vi*) अवलोकित वितरण तथा प्रत्याशित वितरण की तुलना हेतु काई-वर्ग की गणना के लिए सारणी बनाने पर–

सारणी–10

काई-वर्ग की गणना

पंक्ति	स्तम्भ	अवलोकित आवृत्ति f_0	प्रत्याशित आवृत्ति f_e	$f_0 - f_e$	$(f_0 - f_e)^2$	$\frac{(f_0 - f_e)^2}{f_e}$
1	1	6	4.5	1.5	2.25	0.50
1	2	15	10.5	4.5	20.25	1.93
1	3	5	9.0	4.0	16.00	1.78
1	4	4	6.0	2.0	4.00	0.67
2	1	5	6.0	1.0	1.00	0.17
2	2	10	14.0	4.0	16.00	1.14
2	3	15	12.0	3.0	9.00	0.75
2	4	10	8.0	2.0	4.00	0.50
3	1	4	4.5	0.5	0.25	0.06
3	2	10	10.5	0.5	0.25	0.02
3	3	10	9.0	1.0	1.00	0.11
3	4	6	6.0	0.0	0.00	0.00
						$\chi^2 = \Sigma \frac{(f_0 - f_e)^2}{f_e} = 7.63$

(v) परिगणित काई-वर्ग का मान 7.63 है जो .05 व .01 स्तरों पर सार्थक काई-वर्ग मानों (12.592 व 16.812) से कम है, अत: यह सार्थक नहीं है। इस असार्थक काई-वर्ग के आधार पर शून्य परिकल्पना को स्वीकार करते हुए स्वतन्त्र वितरण की परिकल्पना को स्वीकार किया जा सकता है। अत: कहा जा सकता है कि परीक्षा-परिणाम तथा सामाजिक-आर्थिक स्तर एक-दूसरे से स्वतन्त्र हैं।

छात्र क्रियाकलाप (Student Activity)

1. काई-वर्ग परीक्षण में मुक्तांश से आप क्या समझते हैं? उदाहरण सहित बताएं।

2. X_2-वितरण और इसकी उपयोगिता का संक्षेप में वर्णन कीजिए।

3. स्वतन्त्र वितरण की परिकल्पना को स्पष्ट कीजिए।

5.10 सारांश (Summary)

1. व्यवहार में समष्टि से केवल एक ही प्रतिदर्श लिया जाता है तथा इस प्रतिदर्श से प्राप्त विभिन्न प्रतिदर्शजों की सहायता से समष्टि के प्राचलों का अनुमान लगाया जाता है। इसके लिए प्रतिदर्शजों के वितरण (Distributions of Statistics), जिन्हें प्रतिचयन वितरण (Sampling Distributions) भी कहते हैं, का उपयोग किया जाता है।
2. जब 'मध्यमान शब्द का प्रयोग किया जा रहा है तब उपलब्ध समंक (Available Data) मात्रात्मक प्रकृति (Quantitativ Nature) का होता है जिसे कम से कम (At least) मापन के अन्तरिक स्तर

(Interval Level) पर मापा गया होता है। परन्तु कभी-कभी अनुसंधान कार्य में गुणात्मक प्रकृति (Qualitative Nature) के समंक उपलब्ध होते हैं। ऐसी स्थिति में चरों को नामित स्तर (Nominal Level) अथवा क्रमित स्तर (Ordinal Level) पर मापा गया होता है।

नोट

3. काई (χ) वास्तव में ग्रीक भाषा (Greek Language) का एक अक्षर (Latter) है। काई-वर्ग वितरण (Chi-Square Distribution) की खोज सन् 1875 में हेलमर्ट (Helmert) ने की थी तथा बाद में सन् 1900 में कार्ल पियर्सन (Karl Pearson) ने पुन: इसका प्रतिपादन स्वतन्त्र रूप से किया एवं इसके आधार पर काई-वर्ग परीक्षण (Chi-Square Test) को विकसित किया।
4. काई-वर्ग परीक्षण में परिगणित काई-वर्ग की सार्थकता का निर्धारण काई-वर्ग के प्रतिचयन वितरण (Sampling Distribution of Chi-Square) के आधार पर प्रायिकता के रूप में (In Terms of Probability) किया जाता है। काई-वर्ग का प्रतिचयन वितरण केवल अवलोकित वितरण के वर्गों के लिए मुक्तांशों (dfs) पर निर्भर करता है।
5. कभी-कभी अनुसंधानकर्ता यह ज्ञात करने का इच्छुक होता है कि क्या विभिन्न वर्गों में अवलोकित आवृत्तियों को समान रूप से वितरित स्वीकार किया जा सकता है अथवा वे परस्पर सार्थक रूप से भिन्न-भिन्न हैं। इस प्रश्न का उत्तर समान वितरण की परिकल्पना के आधार पर तैयार किये गये प्रत्याशित वितरण (Expected Distribution) की तुलना अवलोकित वितरण (Observed Distribution) से काई-वर्ग परीक्षण के द्वारा करके प्राप्त किया जा सकता है। समान वितरण की परिकल्पना को समान प्रायिकता की परिकल्पना (Hypothesis of Equal Probability) भी कहते हैं।
6. सामान्य वितरण की परिकल्पना को सामान्य प्रायिकता की परिकल्पना (Hypothesis of Normal Probability) भी कहते हैं। इस से तात्पर्य है कि समष्टि मे प्राप्तांक सामान्य प्रायिकता वक्र की तरह से वितरित हैं। सामान्य वितरण की परिकल्पना तब ही सम्भव हो सकती है जब समंक कम से कम क्रमित स्तर (Ordinal Level) अथवा अन्तरित स्तर (Interval Level) पर संकलित किये गये हों।

अभ्यास-प्रश्न (Exercise Questions)

1. काई-वर्ग परीक्षण के अन्तर्गत आने वाले विभिन्न सोपानों का उल्लेख करें।
2. समान वितरण की परिकल्पना का संक्षिप्त में परिचय दीजिए।
3. परिमित समष्टि से आप क्या समझते हैं?
4. किसी होटल में आए कुल मेहमानों में 200 पुरुष थे तथा 100 महिलाएं। क्या समष्टि के अन्तर्गत पुरुष-महिलाओं के वितरण को समान रूप से वितरित माना जा सकता है।

संदर्भ ग्रंथ (Reference Books)

1. क
2. क
3. क
4. क