हिन्दी कार्यशाला

Hindi Workshop

कृषि एवं प्रतिदर्श आँकड़ों के विश्लेषण हेतु सांख्यिकीय सॉफ़्टवेयर का अनुप्रयोग

Application of Statistical Software for Analysis of Agricultural and Survey Data

6-13 सितम्बर, 2023

6-13 September, 2023

सन्दर्भ पुस्तिका

Reference Manual

पाठ्यक्रम समन्वयक : डॉ. पंकज दास

पाठ्यक्रम सह-समन्वयक : डॉ. राहुल बनर्जी

डॉ. भारती

प्रतिदर्श सर्वेक्षण प्रभाग

भा.कृ.अनु.प.-भारतीय कृषि सांख्यिकी अनुसंधान संस्थान

लाइब्रेरी एवेन्यू, पूसा, नई दिल्ली-110012

DIVISION OF SAMPLE SURVEYS

ICAR-INDIAN AGRICULTURAL STATISTICS RESEARCH INSTITUTE

LIBRARY AVENUE, PUSA, NEW DELHI-110012

2023

प्राक्कथन

भा.कृ.अनु.प.- भारतीय कृषि सांख्यिकी अनुसंधान संस्थान, सांख्यिकीय विज्ञान (सांख्यिकी, संगणक अनुप्रयोग एवं जैव सूचना विज्ञान) में प्रासंगिक कार्यों में संलग्न एक प्रमुख संस्थान है और कृषि अनुसंधान की गुणवत्ता को समृद्ध करने और नीतिगत निर्णय लेने के लिए कृषि विज्ञान में इनके विवेकपूर्ण संलयन में इसका प्रमुख योगदान है। 1930 में अपनी स्थापना के बाद से, तत्कालीन इंपीरियल काउंसिल ऑफ एग्रीकल्चरल रिसर्च (ICAR) के एक छोटे सांख्यिकीय खंड के रूप से, संस्थान का कद ऊंचा उठा और राष्ट्रीय और अंतर्राष्ट्रीय स्तर पर अपनी उपस्थिति दर्ज कराने में सक्षम हुआ। संस्थान बहुत सक्रिय रूप से परामर्शदात्री सेवा प्रदान कर रहा है जिसने संस्थान को राष्ट्रीय कृषि अनुसंधान और शिक्षा प्रणाली (NARES) एवं राष्ट्रीय कृषि सांख्यिकी प्रणाली (NASS) दोनों में अपनी उपस्थिति दर्ज कराने में सफल हुआ है। संस्थान में NARES में एक उच्च स्तरीय सांख्यिकीय संगणना के लिए उपयुक्त वातावरण बनाने में अग्रणी भूमिका निभाई है।

आँकड़ें एवं उसका विश्लेषण एक उभरता हुआ एवं अत्यंत महत्वपूर्ण क्षेत्र है जो हर क्षेत्र में आधुनिक तकनीकी प्रगति के साथ जुड़ा हुआ है। विभिन्न क्षेत्रों में आँकड़ों को समझने, विश्लेषण करने एवं समस्याओं का समाधान करने के लिए विशेषज्ञों को आँकड़ों के विश्लेषण के साथ साथ संबंधित सॉफ़्टवेयर के उपयोग कीजानकारी होना आवश्यक है।समष्टि के आँकड़े एकत्रित करने में सर्वेक्षण एक महत्वपूर्ण भूमिका निभाता है। प्रतिदर्श सर्वेक्षण तकनीक अनुसन्धान में समय एवं धन की बचत करता है एवं एकत्रित आँकड़ों से उद्देश्यपूर्ण परिणाम निकालने में मदद करता है।

"कृषि एवं प्रतिदर्श आँकड़ों के विश्लेषण हेतु सांख्यिकीय सॉफ़्टवेयर का अनुप्रयोग " नामक यह हिन्दी कार्यशाला कार्यक्रम 6-13 सितम्बर, 2023 के दौरान आयोजित किया जा रहा है। सॉफ़्टवेयर की मदद सेविभिन्न विश्लेषण तकनीकों को आसानी से लागू किया जा सकता हैएवं उनसे अधिक बेहतर परिणाम प्राप्त किये जा सकते हैं।इस कार्यशाला के माध्यम सेआँकड़ों के विश्लेषण के महत्व को समझाने, विभिन्न सॉफ़्टवेयर टूल्स के बारे में अवगत कराने एवं उनके उपयोग का महत्व समझाने का प्रयास किया जाएगा।

इस कार्यशाला के संकाय में प्रतिदर्श सर्वेक्षण एवं संबंधित क्षेत्रों में गहन अनुभव रखने वाले वैज्ञानिक एवं प्रख्यात सांख्यिकीविद शामिल हैं।कार्यशाला के लिए ई-संदर्भ संहिता तैयार की गई है एवं इस ई-संदर्भ संहिता में कार्यशाला में दिए गए व्याख्यान का विस्तृत विवरण किया गया है। मुझे आशा है कि यह संदर्भ संहिता सहभागियों के लिए उपयोगीसाबित होगी। डॉ पंकज दास (पाठ्यक्रमसमन्वयक), डॉ राहुल बनर्जी एवं डॉ भारती (पाठ्यक्रमसह-समन्वयक) एवं अन्य सहयोगी इस संदर्भ संहिता को तैयार करने के लिए सराहना के पात्र हैं।

05 सितम्बर 2023
नई दिल्ली

(राजेन्द्र प्रसाद)
निदेशक,
भा.कृ.अनु.प.–भा.कृ.सां.अ.सं.

आमुख

भारतीय कृषि अनुसंधान परिषद्-भारतीय कृषि सांख्यिकी अनुसंधान संस्थान देश में कृषि सांख्यिकी, संगणक अनुप्रयोग एवं जैव सूचना विज्ञान में अग्रणी संस्थान है। यह संस्थान भारतीय कृषि अनुसन्धान परिषद् के कृषि शिक्षा प्रभाग के मानव संसाधन विकास कार्यक्रम के तत्वावधान में कृषि सांख्यिकी एवं संगणक अनुप्रयोग में उच्च संकाय प्रशिक्षण केंद्र (जो कि पहले उच्च अध्ययन केंद्र के नाम से जाना जाता था) के रूप में कार्य कर रहा है। वर्तमान हिंदी कार्यशाला संस्थान की हिंदी गतिविधियों को सुचारु संचालन हेतु हिंदी एकक के निदेशानुसार आयोजित की गई है।

प्रतिदर्श सर्वेक्षण प्रभाग द्वारा कृषि, बागवानी, पशुपालन, मात्स्यिकी इत्यादि में महत्वपूर्ण प्राचल के आकलन के लिए विभिन्न प्रतिदर्श पद्धतियों का विकास किया गया है। इन विकसित प्रतिदर्श पद्धतियों का उपयोग राष्ट्रीय एवं अंतराष्ट्रीय स्तर पर किया जा रहा है। प्रतिदर्श सर्वेक्षण प्रभाग के वैज्ञानिक प्रतिदर्श सर्वेक्षण के विभिन्न पहलुओं जैसे जटिल सर्वेक्षणों की अभिकल्पना एवं विश्लेषण, विचरण आकलन तकनीकें, लघु क्षेत्र आकलन, सर्वेक्षण आंकड़ों के विश्लेषण के लिए सांख्यिकीय सॉफ्टवेयर विकसित करना, भौगोलिक सूचना प्रणाली एवं सुदूर संवेदन तकनीकों आदि पर अनुसंधान कर रहे हैं।

वर्तमान में, आँकड़ों के विश्लेषण के लिए कई सॉफ़्टवेयर उपलब्ध हैं जिनकी सहायता से विभिन्न विश्लेषण तकनीकों को आसानी से लागू किया जा सकता है एवं उनसे अधिक बेहतर परिणाम प्राप्त किये जा सकते हैं। आँकड़ें एवं उसका विश्लेषण एक उभरता हुआ एवं अत्यंत महत्वपूर्ण क्षेत्र है जो हर क्षेत्र में आधुनिक तकनीकी प्रगति के साथ जुड़ा हुआ है। विभिन्न क्षेत्रों में आँकड़ों को समझने, विश्लेषण करने एवं समस्याओं का समाधान करने के लिए विशेषज्ञों को आँकड़ों के विश्लेषण के साथ साथ संबंधित सॉफ़्टवेयर के उपयोग की जानकारी होना आवश्यक है। इस कार्यशाला "कृषि एवं प्रतिदर्श आँकड़ों के विश्लेषण हेतु सांख्यिकीय सॉफ़्टवेयर का अनुप्रयोग" के माध्यम से आँकड़ों के विश्लेषण के महत्व को समझाने, विभिन्न सॉफ़्टवेयर टूल्स के बारे में अवगत कराने एवं उनके उपयोग का महत्व समझाने का प्रयास किया जाएगा। यह संदर्भ पुस्तिका प्रतिभागियों के उपयोग के लिए सरलीकृत रूप से प्रस्तुत की गयी है।

इस कार्यशाला को सार्थक एवं सफल बनाने में अपना समय एवं प्रयासों को समर्पित करने के लिए हम सभी संकाय सदस्यों के बहुत आभारी हैं। हम इस हिंदी कार्यशाला में प्रतिभागियों को नामित करने के लिए भारतीय कृषि अनुसंधान परिषद् के विभिन्न संस्थानों एवं राज्य कृषि विश्वविद्यालयों के आभारी हैं। हम डॉ राजेन्द्र प्रसाद, निदेशक, भारतीय कृषि सांख्यिकी अनुसंधान संस्थान के आभारी है जिन्होंने, हमें इस कार्यशाला का संचालन करने की अनुमति दी एवं आवश्यक सुविधाएं उपलबध करवाई। हम इस हिंदी कार्यशाला के आयोजन के लिए निरंतर सहायता प्रदान करने के लिए डॉ तौकीर अहमद, प्रभागाध्यक्ष, प्रतिदर्श सर्वेक्षण प्रभाग के बहुत आभारी हैं। इस कार्यशाला के सुचारु रूप से संचालन के लिए हम हिंदी एकक के भी आभारी हैं। अंत में, हम उन सभी का आभार प्रकट करते हैं जिन्होंने, इस सन्दर्भ पुस्तिका को तैयार करने में सहयोग दिया है।

लेखकगण

विषय सूची

संं.क्र.	विषय	प्रवक्ता	पृष्ठ संख्या
1.	वैज्ञानिक एवं तकनिकी कार्यों में हिन्दी टाइपिंग एवं यूनिकोड फ़ॉन्ट का इस्तेमाल	श्री ए.के. जगदीशन, श्री उदय वीर सिंह	1-5
2.	MS Excel का आँकड़ों के विश्लेषण में अनुप्रयोग	डॉ. भारती	6-14
3.	आँकड़ों के विश्लेषण का अवलोकन	डॉ. राहुल बनर्जी	15-28
4.	सांख्यिकीय विश्लेषण में R सॉफ्टवेयर का उपयोग	डॉ. पंकज दास	29-39
5.	सांख्यिकीय विश्लेषण में SPSS सॉफ्टवेयर का उपयोग	डॉ. राजू कुमार	40-50
6.	प्रतिदर्श सर्वेक्षणों में प्रारंभिक अवधारणाएँ	डॉ. अंकुर विश्वास	51-67
7.	भारत में कृषि सांख्यिकी प्रणाली का परिचय	डॉ. तौक़ीर अहमद	68-80
8.	eLISS ऐप एवं eLISS पोर्टल द्वारा पशुधन आँकड़ों का उत्पादन	डॉ. प्राची मिश्रा साहू	81-88
9.	QGIS का परिचय	डॉ. भारती	89-98
10.	उत्तर-पूर्वी पर्वतीय क्षेत्र में सुदूर संवेदन, भौगोलिक सूचना तंत्र एवं भूमि सर्वेक्षण आँकड़ों के उपयोग द्वारा फसल क्षेत्र आकलन	डॉ. प्राची मिश्रा साहू	99-105
11.	फ़सल उपज आकलन हेतु फ़सल कटाई प्रयोग तकनीक	डॉ. मान सिंह	106-136
12.	आँकड़ों का विज़ुअलाइज़ेशन एवं अन्वेषणात्मक विश्लेषण	डॉ. दीपक सिंह	137-141
13.	सहसंबंध एवं प्रतिगमन विश्लेषण तकनीकें	डॉ. राहुल बनर्जी	142-152
14.	लघु क्षेत्र आकलन विधि-एक अवलोकन	डॉ. कौस्तव आदित्य	153-163
15.	मशीन लर्निंग के अनुप्रयोग का परिचय	डॉ. पंकज दास	164-174

अध्याय 01

वैज्ञानिक एवं तकनिकी कार्यों में हिन्दी टाइपिंग एवं यूनिकोड फ़ॉन्ट का इस्तेमाल

ए.के . जगदीशन[1], उदय वीर सिंह[2]
[1]भा.कृ.अनु.प.-भारतीय बागवानी अनुसंधान संस्थान, बेंगलूरू
[2]भा.कृ.अनु.प.-भारतीय कृषि साँख्यिकी अनुसंधान संस्थान, लाइब्रेरी एवेन्यू, नई दिल्ली-110012

1. प्रस्तावना (Introduction):

यूनिकोड क्या है (What is Unicode?): सर्वप्रथम यह समझना आवश्यक है कि यूनिकोड क्या है? क्या यूनिकोड कोई फोंट है? क्या यूनिकोड कोई टंकण टूल है? कम्प्युटर में टाइप करने के लिए अनेक फोंट्स उपलब्ध होते हैं। सभी फोंट्स को दो भागों में बांटा गया है: 1. नॉन-यूनिकोड 2. यूनिकोड

हिन्दी में टाइप करने के लिए प्रायः हम लोग क्रुतिदेव अथवा मंगल फोंट्स का प्रयोग करते हैं। इनमें क्रुतिदेव फॉन्ट नॉन-यूनिकोड के अंतर्गत आता है जबकि मंगल फॉन्ट यूनिकोड के अंतर्गत आता है। यूनिकोड एक वैज्ञानिक तकनीक है, एक व्यवस्था है, जो प्रचलित प्रत्येक लिपि के वर्णमाला के अक्षर को चार अंकों का विशेष कोड या नम्बर प्रदान करता है इन्टरनेट पर जो हम हिन्दी में लिखे हुए लेख देखते हैं, यह सब यूनिकोड तकनीक का इस्तेमाल करके ही लिखे जाते हैं।

नॉन-यूनिकोड और यूनिकोड में अंतर (Difference between non-Unicode and Unicode):

नॉन-यूनिकोड (non-Unicode): जब हम नॉन-यूनिकोड फोंट्स अर्थात क्रुतिदेव फॉन्ट का उपयोग करके हिन्दी में टाइप करते हैं एवं टाइप किये गये पत्र, लेख अथवा संदेश को इंटरनेट या किसी और माध्यम से एक कम्प्युटर से दूसरे कम्प्युटर पर अथवा एक उपयोगकर्ता से दूसरे उपयोगकर्ता को भेजते हैं तो अक्सर यह होता है कि यदि दूसरे उपयोगकर्ता के पास कम्प्युटर में क्रुतिदेव फॉन्ट उपलब्ध नहीं है तो उस कम्प्युटर पर भेजा हुआ पत्र या लेख पढ़ने में नहीं आता है, उसका स्वरूप बदल जाता है वह अपाठ्य हो जाता है, चाहे दूसरे उपयोगकर्ता के पास क्रुतिदेव फॉन्ट उपलब्ध हो भी, तब भी हिन्दी वर्णमाला के कुछ अक्षरों मे अपभ्रंश हो जाता है, उनका स्वरूप ही बदल जाता है अर्थात उस वाक्य का अर्थ ही बदल जाता है।

हम सभी इंग्लिश में तो टाइप करने के अभ्यस्त हैं। क्योंकि हमें कम्प्युटर में इंग्लिश के कीबोर्ड पर इंग्लिश के अक्षरों की तो जानकारी होती है, लेकिन इंग्लिश के कीबोर्ड पर हिन्दी भाषा/ लिपि की वर्णमाला के अक्षरों की जानकारी नहीं होती है, कि हिन्दी का कौन सा अक्षर कहाँ पर अत: नॉन-यूनिकोड फॉन्ट का उपयोग केवल हिन्दी टाइपिस्ट अथवा टाइप करने का अभ्यास करने वाले ही टाइप कर सकते हैं।

यूनिकोड (Unicode): जब हम यूनिकोड फॉन्ट का उपयोग करके टाइप किये गये पाठ, सामग्री, पत्र, लेख अथवा संदेश को इंटरनेट या किसी और माध्यम से एक कम्प्युटर से दूसरे कम्प्युटर पर अथवा एक उपयोगकर्ता से दूसरे उपयोगकर्ता को भेजते हैं तो उसका स्वरूप बिलकुल भी नहीं बदलता है, वह पूरी दुनिया में कही भी पढ़ा जा सकता है, इसके लिये किसी विशेष फॉन्ट की आवश्यकता नहीं होती है। यूनिकोड फॉन्ट का उपयोग कम्प्युटर में अँग्रेजी के कीबोर्ड पर अँग्रेजी के अक्षरों की जानकारी रखने वाले भी हिन्दी में बहुत अच्छी तरह से टाइप कर सकते है।

2. यूनिकोड के लाभ (Importance of Unicode):

इसका लाभ यह है कि चाहे किसी भी सॉफ्टवेयर में या किसी भी भाषा में यूनिकोड का प्रयोग किया जा सकता है। यूनिकोड में टाइप किये गये पाठ या सामग्री को कहीं भी ले जाने पर उसका स्वरूप नहीं बदलता है, वह पूरी दुनिया में कही भी पढी जा सकती है, इसके लिये किसी विशेष फान्ट की आवश्यकता नहीं होती है।

यूनिकोड विश्व की ज्यादातर भाषाओं में बदला जा सकता है। यूनिकोड मानक को एपल, एच.पी., आई.बी.एम., माइक्रोसौफ्ट, औरेकल जैसी प्रमुख कम्पनियो ने अपनाया है । यूनिकोड में हिन्दी तथा अन्य भाषाओ में कम्प्युटर पर अँग्रेजी की तरह आसानी से 100% कार्य किया जा सकता है। जैसे- वर्ड प्रोसेसिंग, डाटा प्रोसेसिंग, ई-मेल, वैबसाइट निर्माण आदि किए जा सकते हैं। यूनिकोड तीन प्रकार का होता है:

1. यूटीएफ़-8 2. यूटीएफ़-16 3. यूटीएफ़-32

भारतीय भाषाओ के लिए यूनिकोड एङ्कोडिंग के लिए यूटीएफ़-8 का प्रयोग किया जाता है।

यूनिकोड के अन्य लाभ:

1. आप बिना हिन्दी टाइप जाने हिन्दी में टाइप कर सकते हैं ।
2. आप गूगल सर्च में हिन्दी में सर्च कर सकते हैं ।
3. हिन्दी में ई-मेल भेज सकते हो।
4. कम्प्यूटर में विभिन्न फाइल और फोल्डरों के नाम हिन्दी में रख सकते हैं ।
5. हिन्दी में चैट कर सकते हैं ।
6. हिन्दी में वेब साइट या ब्लाग बना सकते हैं ।
7. वर्ड और एक्सेल में बिना हिन्दी फॉन्ट डाउनलोड किये हिन्दी में टाइपिंग की जा सकती है।

8. फेसबुक और ट्विटर जैसे सोशल नेटवर्किंग साइट पर आसानी से हिन्दी में लिखा जा सकता है।

9. यूनिकोड में लिखी किसी भी सामग्री को आसानी से दूसरी भाषा में परिवर्तित किया जा सकता है।

सरकारी कार्यालयों में यूनिकोड का प्रयोग (Use of Unicode in Government Offices): भारत सरकार द्वारा यूनिकोड को अपने सभी कार्यालयों में अनिवार्य कर दिया गया है, भारत सरकार की सभी साइटों पर यूनिकोड का प्रयोग किया जा रहा है। यहाँ तक कि नई भर्तियों के लिये अभ्यर्थियों को यूनिकोड फॉन्ट में टाइपिंग टेस्ट भी अनिवार्य कर दिया गया है।

3. यूनिकोड के अंतर्गत मंगल फॉन्ट का उपयोग कर टाइप करने के कुछ उदाहरण:

Bhartiya Krishi Sankhyiki Anusandhan Sansthan
भारतीय कृषि सांख्यिकी अनुसंधान संस्थान

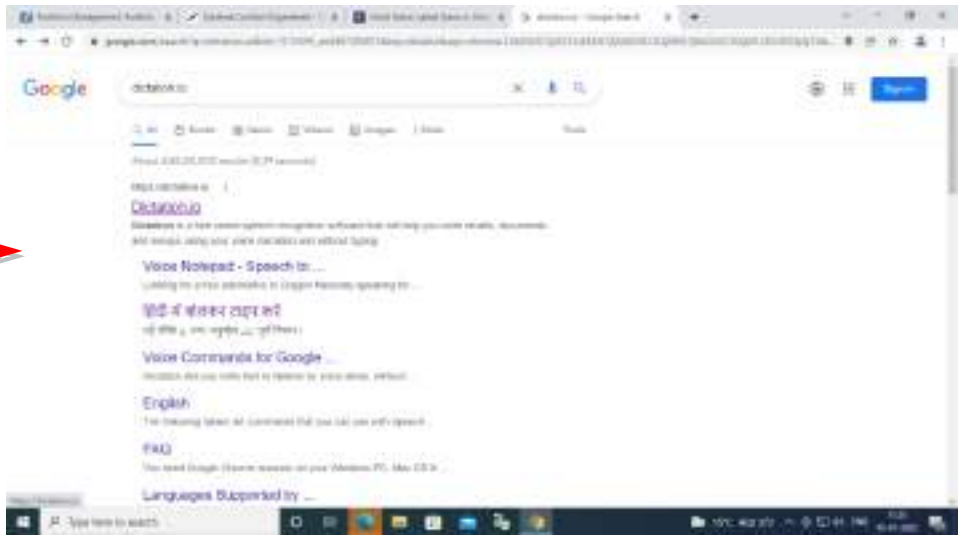
Hindi Karyshala
हिन्दी कार्यशाला

Krishi mein pratidarsh taknikon ka anuprayog evam aankdon ka vishleshan
कृषि में प्रतिदर्श तकनीकों का अनुप्रयोग एवं आंकड़ों का विश्लेषण

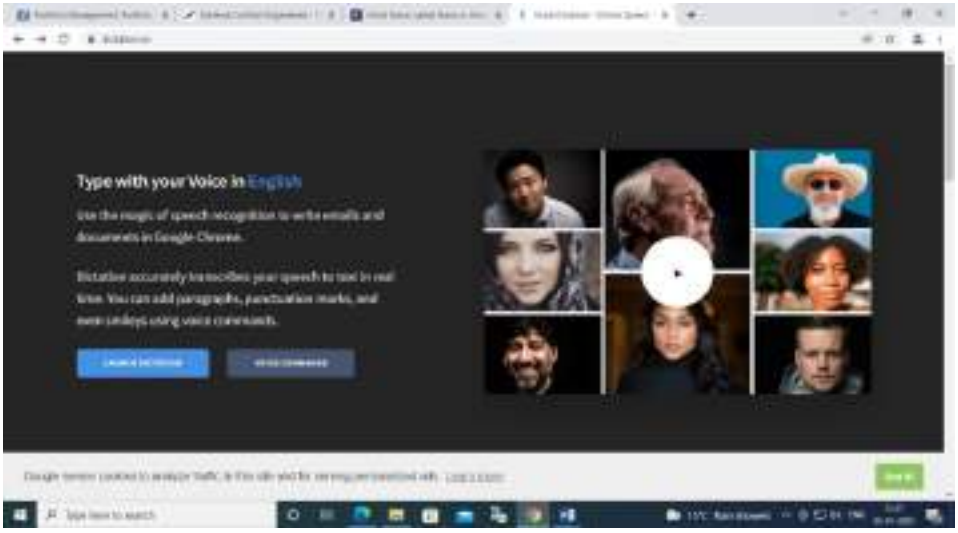
Pramukh Vaigyanik	**Takniki Adhikari**	**Prashsnik Adhikari**
प्रमुख वैज्ञानिक	तकनीकी अधिकारी	प्रशासनिक अधिकारी

Ek Kisaan ki Mera Gaon mera gaurav pariyojna ke antargat krishi Vagyanik se Vartalap
एक किसान की मेरा गाँव मेरा गौरव परियोजना के अंतर्गत कृषि वैज्ञानिक से वार्तालाप

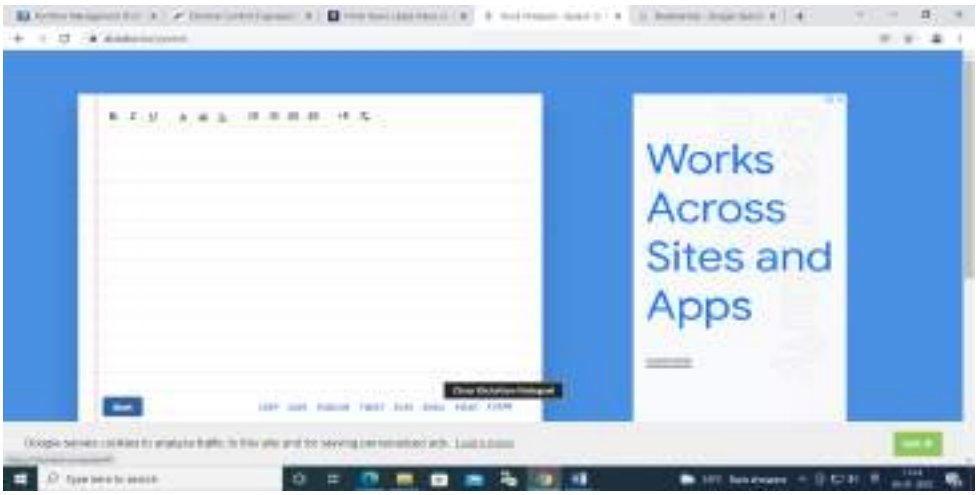
यूनिकोड के अतिरिक्त कम्प्युटर पर ऑनलाइन आवाज/ बोलने से टाइप/ टंकण करने की सुविधा



गूगल में dictation.io सर्च करके web site https://dictation.io पर क्लिक करें।



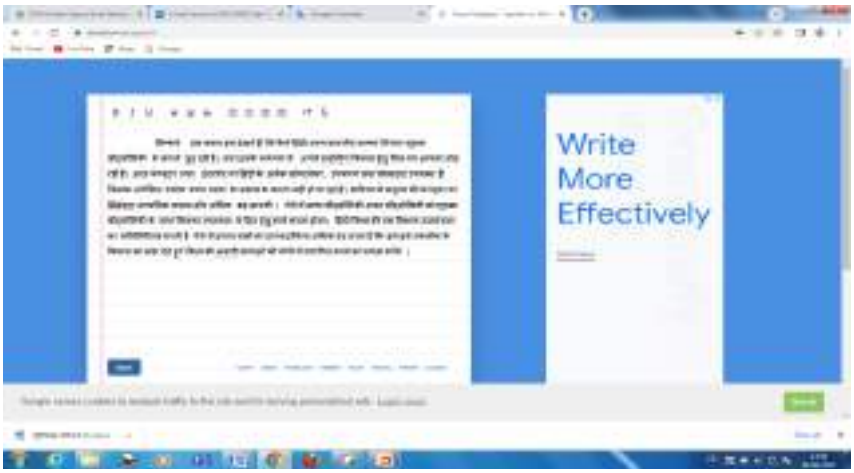
Launch Dictation पर क्लिक करें।



रूलर को स्क्रोल कर के नीचे लाएँ ।

स्क्रोल नीचे करने के बाद भाषा चुनें और माइक्रोफोन अथवा स्टार्ट पर क्लिक करें

विभिन्न चिन्हों को टाइप करने के लिए उक्त शब्दों का प्रयोग करें ।



ऑन लाइन टाइप करने के बाद एम.एस. वर्ड में कॉपी, पेस्ट और एडिट किया हुआ लेख

4. निष्कर्ष (Conclusion):

इस प्रकार हम देखते हैं कि कैसे हिंदी एवं अन्य भारतीय भाषाएं निरंतर सूचना प्रौद्योगिकी के क्षेत्र में जुड़ रही है। तथा इसके माध्यम से अपने सर्वांगीण विकास हेतु नित नए आयाम जोड़ रही है । आज कंप्यूटर तथा इंटरनेट पर हिंदी के अनेक सॉफ्टवेयर, उपकरण तथा वेबसाइट उपलब्ध है जिनका अपेक्षित प्रयोग प्रचार प्रसार के अभाव के कारण नहीं हो पा रहा है । भविष्य में मनुष्य की कंप्यूटर पर निर्भरता और अधिक बढ़ जाएगी । ऐसे में भाषा प्रौद्योगिकी को सूचना प्रौद्योगिकी के साथ मिलकर मानवता के हित हेतु कार्य करना होगा। हिंदी विश्व की एक विशाल जनसंख्या का प्रतिनिधित्व करती है। ऐसे में हम सभी का उत्तरदायित्व और अधिक बढ़ जाता है कि हम इसे तकनीक के विकास का लाभ देते हुए विश्व की अग्रणी भाषाओं की श्रेणी में स्थापित करने का प्रयास करें ।

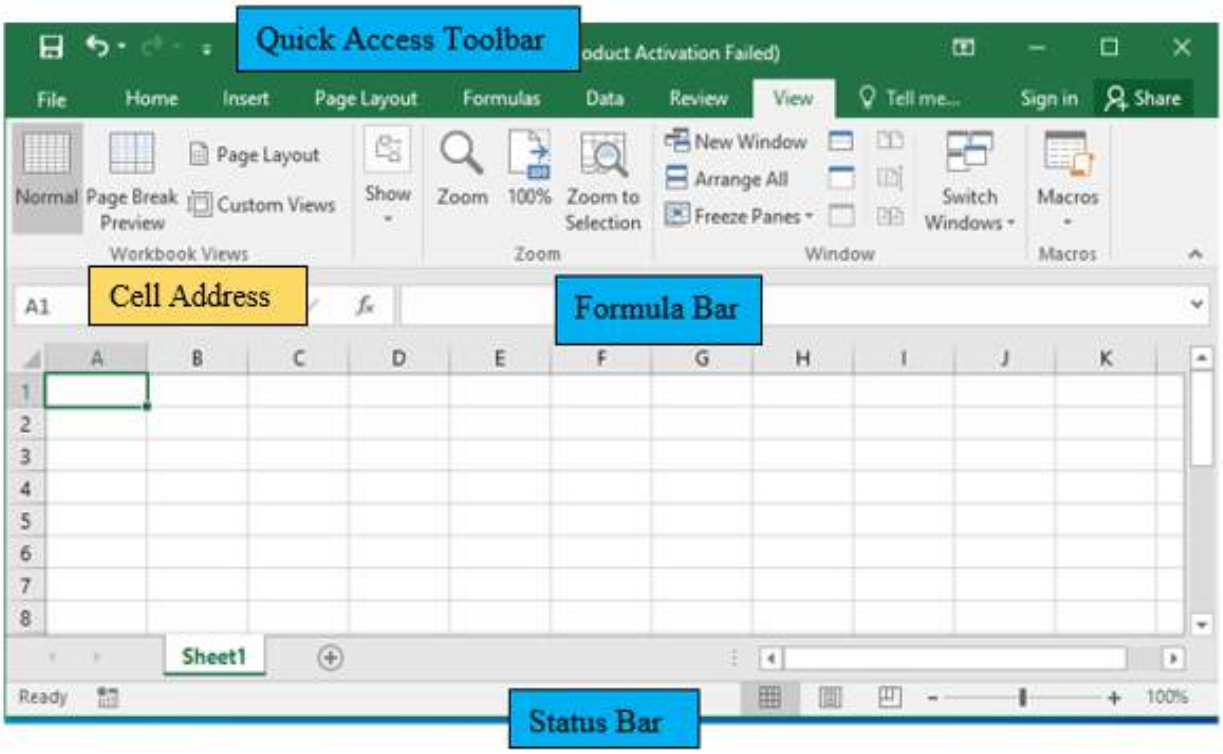
अध्याय 02

MS Excel का आँकड़ों के विश्लेषण में अनुप्रयोग

भारती
वैज्ञानिक, प्रतिदर्श सर्वेक्षण प्रभाग
भा.कृ.अनु.प.-भारतीय कृषि साँख्यिकी अनुसंधान संस्थान, लाइब्रेरी एवेन्यू, नई दिल्ली-110012

1. प्रस्तावना (Introduction):

MS Excel एक शक्तिशाली और प्रसिद्ध स्प्रेडशीट एप्लिकेशन है जिसका उपयोग डेटा को संग्रहित, संचित, विश्लेषित और प्रस्तुत करने के लिए किया जाता है। यह माइक्रोसॉफ्ट ऑफ़िस सुइट का हिस्सा है और व्यावसायिक और शिक्षात्मक उद्देश्यों के लिए व्यापक उपयोग किया जाता है।



चित्र 1: **MS Excel** स्क्रीन

MS Excel द्वारा विभिन्न प्रकार के आँकड़ों को तालिकाओं में संग्रहित किया जा सकता है जिन्हें विश्लेषण और अन्य गणनाओं के लिए प्रयोग किया जा सकता है। इसके साथ ही, चार्ट्स और ग्राफ्स की मदद से आँकड़ों को ग्राफिकल रूप में प्रस्तुत किया जा सकता है, जो आँकड़ों के पैटर्न को समझने में मदद करता है। **MS Excel** में फ़ॉर्मूलों और फ़ंक्शन्स की मदद से अंकगणितिक, सांख्यिकीय और लॉजिकल गणनाओं को सरलता से किया जा सकता है। इसे बचत और निवेश की गणनाओं, वित्तीय प्रबंधन, इंवेंटरी कंट्रोल, व्यक्तिगत बजट तैयारी, ग्रेड बुक कार्य, और अन्य कई कार्यों के लिए भी किया जा सकता हैं।

2. MS Excel स्क्रीन **(MS Excel Screen)**:

- **टाइटल बार** (**Title Bar):** यह वर्कशीट के शीर्ष में स्थित होता है और वर्तमान फ़ाइल का नाम दिखाता है।
- **मेनू बार** (**Menu Bar):** यह मेनू विकल्पों का संग्रहण स्थल होता है जैसे कि होम, इन्सर्ट, पेज लेआउट इत्यादि।
- **स्टेटस बार** (**Status Bar):** यह वर्कबुक में सबसे नीचे स्थित होता है और सेल की विभिन्न गणनाओं की जानकारी को प्रदर्शित करता है, जैसे कि सेल की स्थिति, सेल में डेटा का योग इत्यादि।
- **सूत्र बार** (**Formula Bar):** सेल की सूचना और फ़ॉर्मूला प्रदर्शित करता है और जब डेटा टाइप करते हैं तो सूत्र बार पर दिखाई देता है।
- **स्क्रॉल बार** (**Scroll Bars):** यह वर्कशीट में ऊपर और नीचे, बाएँ और दाएँ स्थानों की ओर स्क्रॉल करने में मदद करता है।

3. MS Excel में आँकड़ों को दर्ज करना (Entering Data in MS Excel):

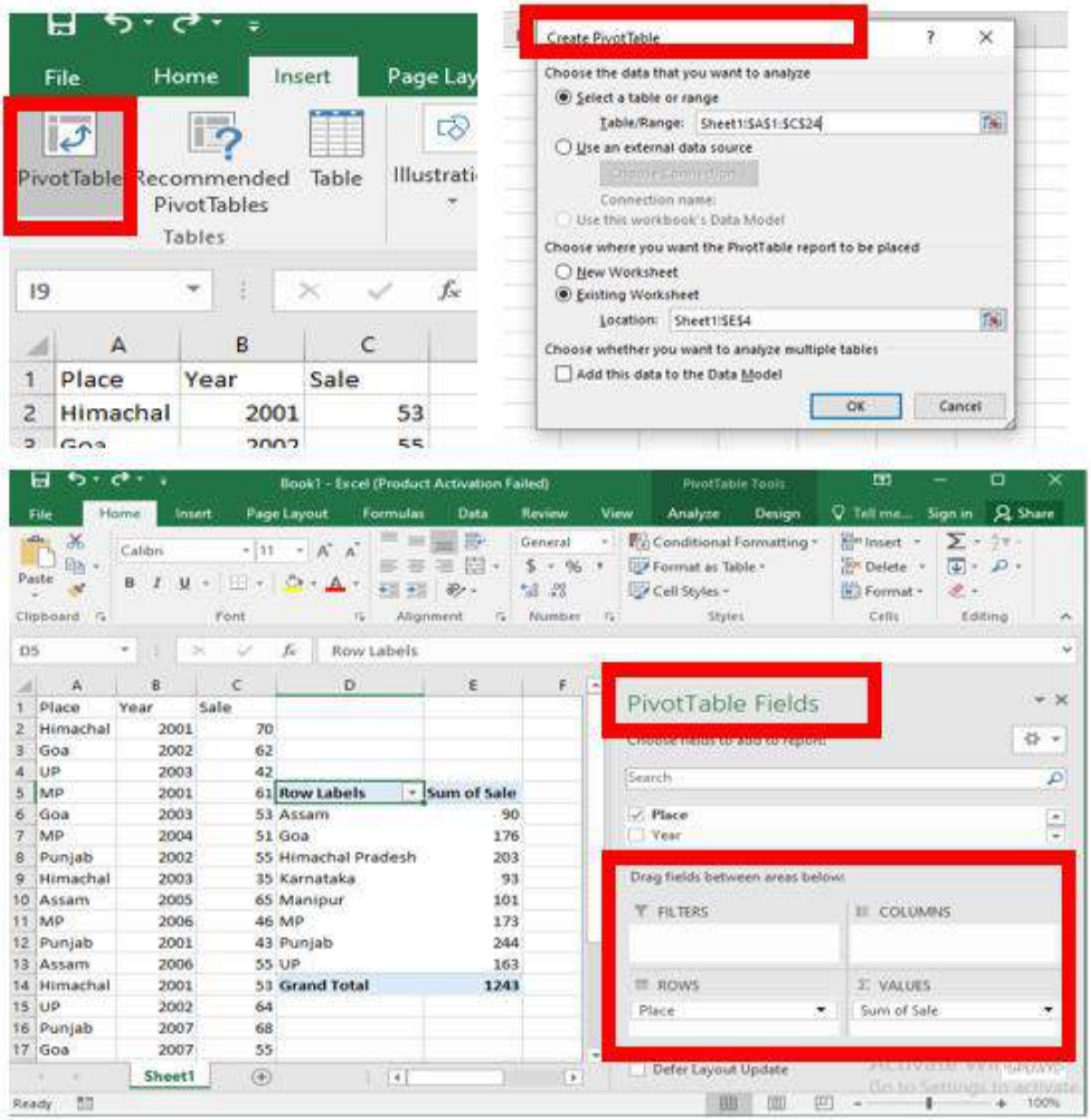
MS Excel में एक वर्कशीट में **256** स्तंभ और **65,536** पंक्तियाँ होती हैं । **MS Excel** वर्कशीट पंक्तियों और स्तंभों की ग्रिड होती है। पंक्तियाँ संख्याओं से और स्तंभ अक्षरों से चिह्नित होती हैं। प्रत्येक पंक्ति और स्तंभ के प्रतिच्छेदन को सेल कहते है। प्रत्येक सेल का पता स्तंभ अक्षर और पंक्ति संख्या होता है। चित्र **1** में वर्कशीट पर दाहिने ओर **A1** सेल पता की ओर इशारा करता है, जिसे वर्तमान में हाइलाइट किया गया है, इससे स्पष्ट होता है कि यह एक सक्रिय सेल है। किसी सेल में आँकड़ों को दर्ज करने के लिए सेल सक्रिय होना आवश्यक है।

सेल में डेटा दर्ज करने के लिए, सेल को चुनें और टाइपिंग शुरू करें। सेल में टाइपिंग करते समय जानकारी सूत्र बार में भी प्रदर्शित होती है। सूत्र बार में भी जानकारी दर्ज की जा सकती हैं, और यह जानकारी चयनित सेल में प्रदर्शित होती है। नीचे सेल की ओर जाने के लिए "एंटर" दबाएं और दाईं ओर जाने के लिए "टैब" दबाएं।

4. **सेल में डेटा दर्ज करने के बाद विभिन्न तरीकों से आँकड़ों का विश्लेषण किया जा सकता है।**

4.1 **डेटा समीकरण** (**Data Summarization): MS Excel** का उपयोग बड़े डेटा सेट्स को संक्षिप्त और सारगर्भित रूप में प्रस्तुत करने के लिए किया जा सकता है। **PivotTables** और **PivotCharts** का उपयोग करके डेटा के विभिन्न पहलुओं को समझ सकते हैं और सार्वजनिक रूप से प्रस्तुत कर सकते हैं। **Pivot Table** इंटरैक्टिव टेबल हैं, जो रिपोर्टिंग और एनालिसिस के लिए बड़ी मात्रा के डेटा को छोटे और टेबुलर फॉर्मेट में तैयार करने की सुविधा देती है। ये डेटा को सॉर्ट, काउंट और टोटल कर सकती है, और इसके साथ ही ये कई तरह के स्प्रेडशीट प्रोग्राम में भी उपलब्ध है।

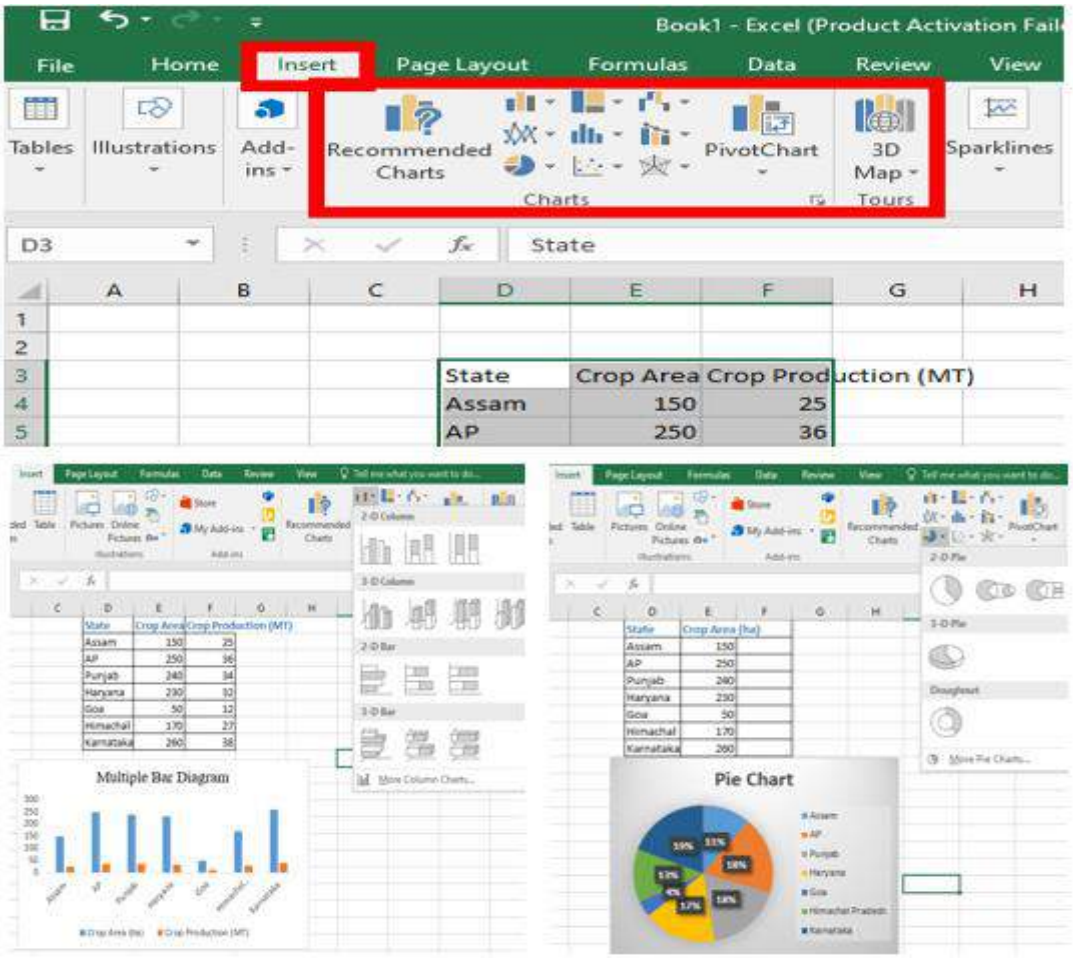
Pivot Table बनाने के लिए "**Insert**" पर क्लिक करें और फिर "**Pivot table**" पर क्लिक करें।



चित्र 2: **MS Excel Pivot Table**

- **" Pivot table Field List**" से "**Column Labels", "Row Labels"** और "Σ **Values"** के तहत संबंधित क्षेत्र **(fields)** को ड्रैग एंड ड्रॉप करने के लिए चुनें।
- "Σ **Values"** में डाउन एरो पर क्लिक करके " **Value Field Settings** " को चुनें और उपयुक्त विकल्प का चयन करें और फिर "**Ok**" पर क्लिक करें।

4.2 **ग्राफिकल विश्लेषण** (**Graphical Visualization): Excel** में विभिन्न ग्राफिकल चार्ट्स और ग्राफ्स का उपयोग करके डेटा के पैटर्न और रिलेशनशिप को समझ सकते हैं। चार्ट्स के माध्यम से आँकड़ों की विभिन्न पहलुओं की स्पष्टता से प्रदर्शन के लिए किया जा सकता है। चार्ट्स और ग्राफ्स बनाने के लिए **"Insert"** मेनू पर क्लिक करे और फिर **"Chart"** पर क्लिक करे जैसा की चित्र **3** में दर्शाया गया है।



चित्र 3: **MS Excel** चार्ट्स और ग्राफ्स

4.3 **क्रमबद्धीकरण (Sorting) और फ़िल्टरिंग (Filtering):** क्रमबद्धीकरण (**Sorting**) और फ़िल्टरिंग (**Filtering**) के द्वारा **MS Excel** में डेटा को संगठित और संवादनशील तरीके से प्रदर्शित किया जा सकता हैं।

क्रमबद्धीकरण (**Sorting**)**:** क्रमबद्धीकरण का उपयोग डेटा को विशिष्ट कॉलम के माध्यम से वर्गीकृत करने के लिए किया जाता है। क्रमबद्धीकरण करने के लिए:

- डेटा रेंज का चयन करें जिसे आप क्रमबद्ध करना चाहते हैं।
- "**Data**" टैब पर "**Sort & Filter**" ग्रुप में से "**Sort**" विकल्प का चयन करें।
- क्रमबद्धीकरण के विभिन्न विकल्प हो सकते हैं जैसे एक आरोही क्रम, एक अवरोही क्रम, या किसी विशिष्ट कॉलम के आधार पर क्रमबद्धीकरण ।

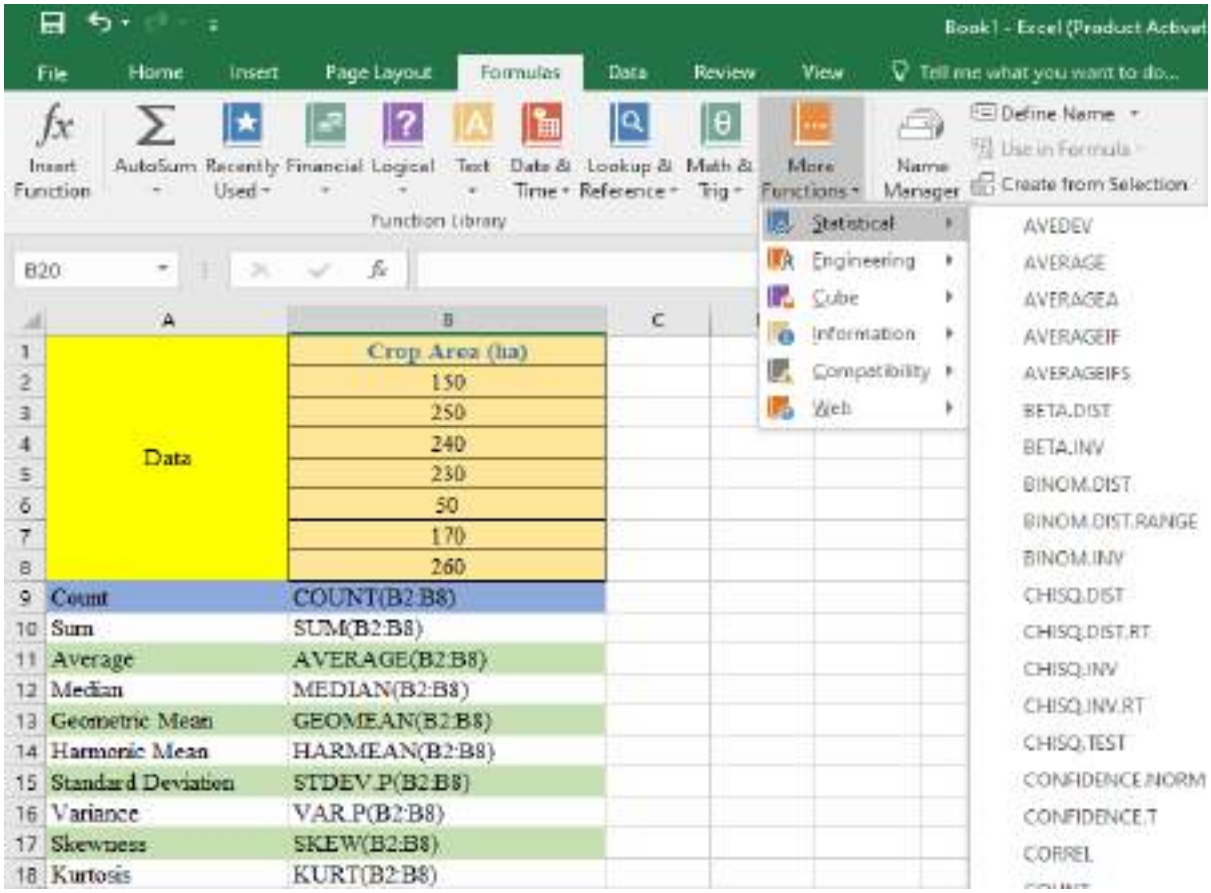
फ़िल्टरिंग (**Filtering**)**:** फ़िल्टरिंग का उपयोग डेटा को चुनें हुए कॉलमों के आधार पर दिखाने के लिए किया जाता है ताकि सिर्फ़ विशिष्ट मानों के साथ डेटा देख सकें। फ़िल्टरिंग करने के लिए:

- डेटा रेंज का चयन करें जिसे आप फ़िल्टर करना चाहते हैं।

- "**Data**" टैब पर "**Sort & Filter**" ग्रुप में से "**Filter**" विकल्प का चयन करें।
- यह ड्रॉपडाउन फ़िल्टर मेनू प्रदान करेगा जिससे कॉलम मानों के आधार पर फ़िल्टरिंग की जा सकती हैं।

ये दोनों कार्यविधियाँ डेटा को आसानी से संवादित करने में मदद कर सकती हैं और डेटा के विभिन्न दृश्य प्रदर्शित करने में सहायक होती हैं।

5. **सांख्यिकीय विश्लेषण** (**Statistical Analysis): MS Excel** में कई संख्यात्मक फ़ंक्शंस होते हैं जिनका उपयोग विभिन्न आँकड़ों के सांख्यिकीय विश्लेषण के लिए किया जा सकता है। उनमें से कुछ प्रमुख फ़ंक्शंस शामिल हैं - **AVERAGE (**औसत**), SUM (**योग**), COUNT (**गणना**), MIN (**न्यूनतम**), MAX (**अधिकतम) आदि। सामान्यत: सांख्यिकीय फ़ंक्शन आमतौर पर एकल संख्यात्मक मानों के स्थान पर सूचियों को सांख्यिकीय विश्लेषण के लिए लेते हैं। एक सूची एकाधिक संख्याओं की एक समूह हो सकती है जो अल्पविराम द्वारा अलग किए गए हो, जैसे **(9, 5, 14, 25, 16, 13, 54)** या निर्दिष्ट सेलों का एक विशिष्ट सीमा, जैसे (**B7:B11),** जो कि सूची (**B7, B8, B9, B10, B11)** को टाइप करने के समकक्ष है।



चित्र 4: **MS Excel** सांख्यिकीय विश्लेषण **(**संख्यात्मक फ़ंक्शंस**)**

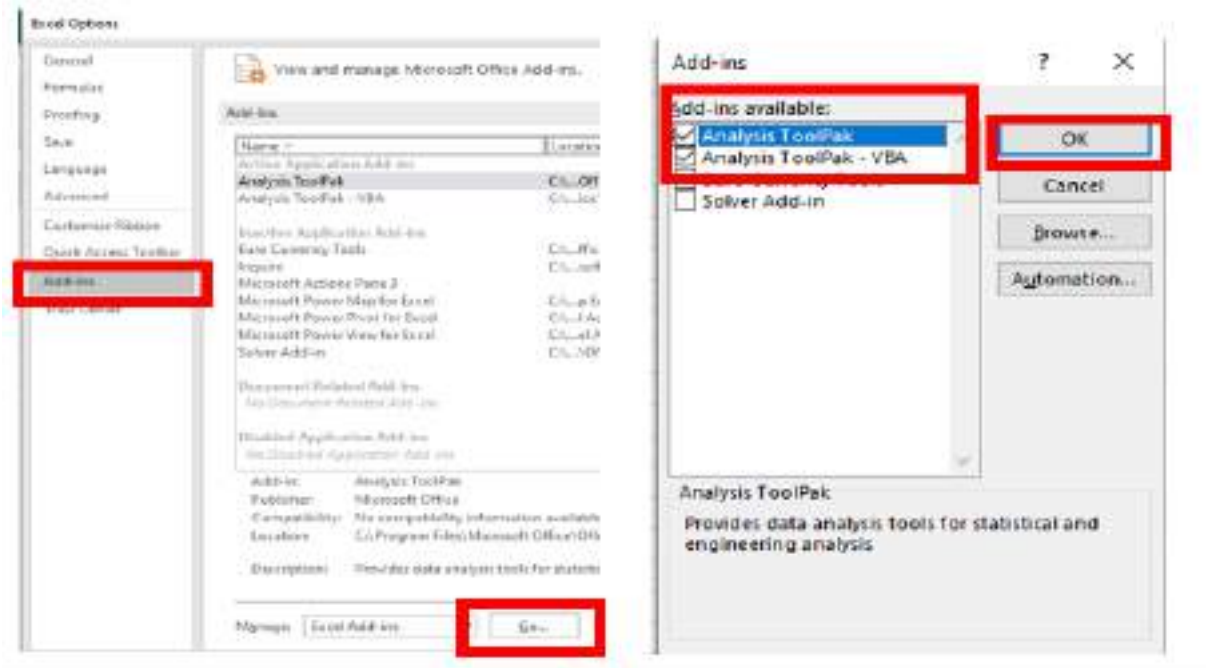
सूत्र (**Formulas)** वर्कशीट में मूल्यों पर गणना करने वाले समीकरण होते हैं। एक सूत्र एक समान चिह्न (=) से शुरू होता है। उदाहरण के लिए, **=7+6*4** सूत्र **6** को **4** से गुणा करके और फिर **7** को परिणाम में जोड़ता है। सूत्र (**Formula)** के लिए निम्नलिखित चरणों का पालन करें:

- उस सेल पर क्लिक करें जिसमें आप सूत्र दर्ज करना चाहते हैं।
- **=** (एक समान चिह्न) टाइप करें।
- सूत्र दर्ज करें।
- एंटर दबाएं

इसके बाद, आपका सूत्र तैयार हो जाएगा और विभिन्न मानों के साथ गणनाएँ कर सकते हैं।

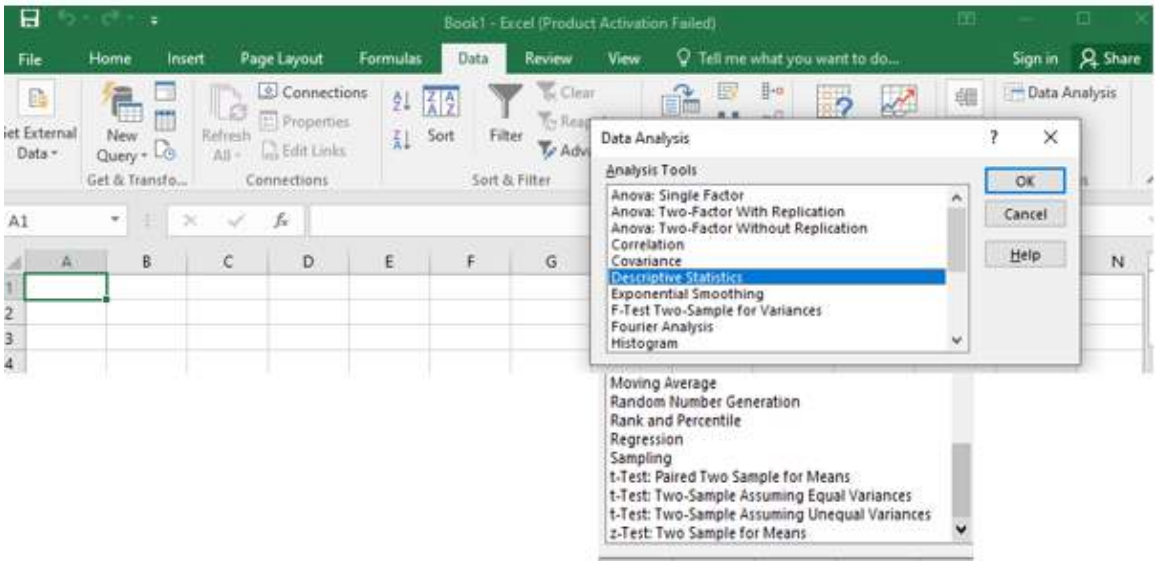
6. स्टैटिस्टिकल एनालिसिस टूल्स (STATISTICAL ANALYSIS TOOLS):

Ms Excel में **"Analysis ToolPak"** का उपयोग करके विभिन्न सांख्यिकीय विश्लेषणों को किया जा सकता हैं । इस टूल में विश्लेषण के लिए डेटा और पैरामीटर्स प्रदान करके। यह टूल उपयुक्त सांख्यिक या इंजीनियरिंग मैक्रो फ़ंक्शन का उपयोग करता है और फिर परिणामों को तालिका में प्रदर्शित करता है। कुछ उपकरण आउटपुट तालिकाओं के साथ-साथ चार्ट भी उत्पन्न होते हैं। डेटा विश्लेषण उपकरण तक पहुँचने के लिए "**Data**" मेनू पर क्लिक करें, फिर "**Data Analysis**" पर क्लिक करें और उपयुक्त विश्लेषण विकल्प का चयन करें। अगर "**Data Analysis**" कमांड उपलब्ध नहीं है, तो "**Add-Ins**" से "**Analysis ToolPak**" को "**Select and OK**" करें।



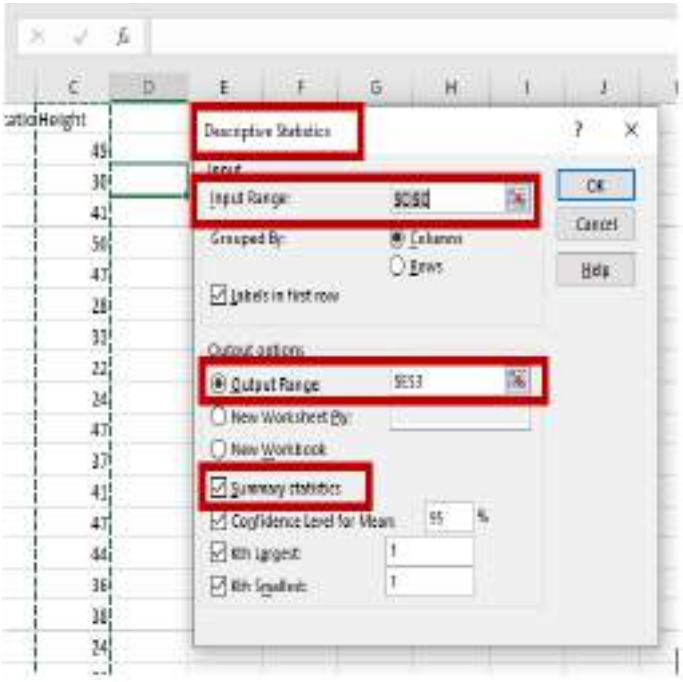
चित्र 5: **MS Excel Analysis ToolPak**

MS Excel में " **Data Analysis**" के अंतर्गत विभिन्न सांख्यिकीय विश्लेषण किये जा सकते हैं जैसे कि **"ANOVA", "Correlation", "Covariance", "Descriptive Statistics", " Exponential Smoothing", "F-test", "t-test"** इत्यादि जैसा कि चित्र **6** में दर्शाया गया है।



चित्र 6: **MS Excel Data Analysis**

Data – Data Analysis – Descriptive Statistics



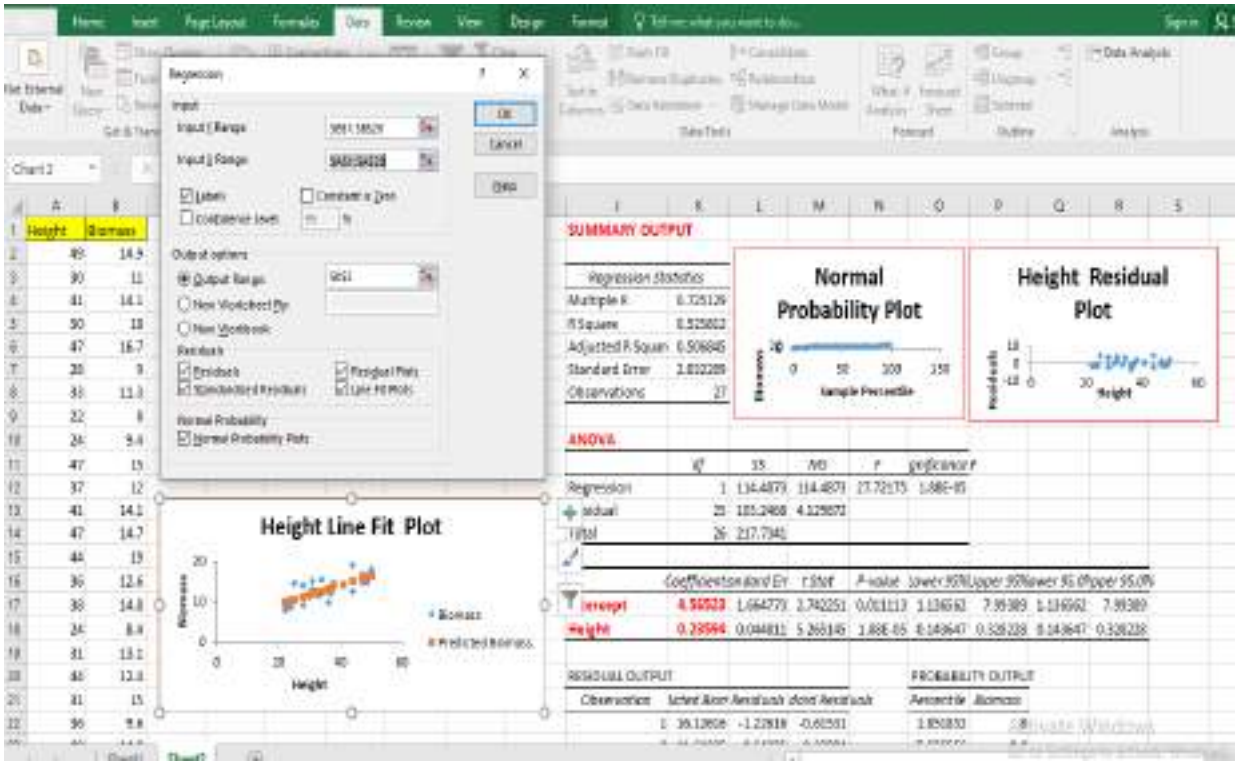
Height	
Mean	36.11111111
Standard Error	1.711643829
Median	36
Mode	47
Standard Deviation	8.893962227
Sample Variance	79.1025641
Kurtosis	-1.251194015
Skewness	0.000437352
Range	28
Minimum	22
Maximum	50
Sum	975
Count	27
Largest(1)	50
Smallest(1)	22
Confidence Level(9	3.518334278

चित्र 7: **MS Excel "Data Analysis" में "Descriptive Statistics"**

Data – Data Analysis – Correlation

चित्र 7: **MS Excel "Data Analysis" में "Correlation"**

Data – Data Analysis – Regression



चित्र 7: **MS Excel "Data Analysis" में "Regression"**

MS Excel पर अभ्यास (Practical Exercise on MS Excel)

1. निम्नलिखित तालिका में मिट्टी के संग्रहित नमूनों में पाए गए **pH** और जैविक कार्बन सामग्री की मूल्यांकन दी गई है। आपको इस डेटा की औसत, माध्य, मानक विचलन, दायित्व और वाक्रता (**skewness)** की गणना करनी है।

संग्रहित नमूनों	**pH**	जैविक कार्बन
1	5.6	2.60
2	6.2	3.26
3	6.1	1.95
4	5.5	1.42
5	5.8	2.65
6	5.2	1.98
7	6.1	1.75
8	5.2	1.54
9	5.8	2.63
10	5.9	3.65

2. निम्नलिखित डेटा के लिए ग्राफ बनाएं:

	S1	**S2**
T1	15	11
T2	10	11
T3	13	12
T4	14	13
T5	13	13
T6	12	10
T7	12	13

3. निम्नलिखित तालिका में चावल के पौधों की दो किस्मों की ऊँचाई दी गई है। पहले, **F**-परीक्षण का उपयोग करके देखें कि क्या ये दोनों नमूने ऐसे समष्टि से आते हैं जिनका विचलन समान है। फिर, उपयुक्त **t**-परीक्षण का उपयोग करके देखें कि क्या दो समूहों की औसत ऊँचाई समान है।

प्रकार **A**	**15**	**12**	**16**	**18**	**19**	**20**	**24**	**26**	**24**	**23**	**22**	**25**
प्रकार **B**	**25**	**24**	**26**	**28**	**32**	**35**	**36**	**32**	**40**	**25**	**36**	**34**

4. निम्नलिखित तालिका में दिए गए पौधों की ऊँचाई और बायोमास के बीच सहसंबंध और समाश्रयण गुणांक का मान ज्ञात करे:

ऊँचाई	**25**	**22**	**19**	**18**	**20**	**25**	**24**	**26**	**24**	**23**	**21**	**23**
बायोमास	**65**	**62**	**58**	**54**	**62**	**67**	**65**	**68**	**68**	**62**	**61**	**62**

अध्याय 03

आँकड़ों के विश्लेषण का अवलोकन

राहुल बनर्जी
वैज्ञानिक, प्रतिदर्श सर्वेक्षण प्रभाग
भा.कृ.अनु.प.-भारतीय कृषि साँख्यिकी अनुसंधान संस्थान, लाइब्रेरी एवेन्यू, नई दिल्ली-110012

1. प्रस्तावना (Introduction):

केंद्रीय प्रवृत्ति के माप (Measures of Central Tendency) आँकड़ों को संक्षेप में समझाने की एक संख्यात्मक विधि है। आप प्रतिदिन की ज़िन्दगी में बड़े सेट ऑफ आँकड़ों को संक्षेपित करने के उदाहरण देख सकते हैं, जैसे कि किसी टेस्ट में छात्रों द्वारा प्राप्त औसत अंक, किसी क्षेत्र में प्राप्त औसत वर्षा, किसी कारख़ाने में प्राप्त औसत उत्पादन, किसी परिस्थितिकी या किसी कंपनी में काम करने वाले व्यक्तियों की प्राप्त औसत आय, आदि। बैजू एक किसान है। वह बिहार के बक्सर ज़िले के बालापुर गांव में अपनी ज़मीन में खाद्यान्न उगाता है। गांव में कुल 50 छोटे किसान हैं। बैजू के पास 1 एकड़ ज़मीन है। आप बालापुर के छोटे किसानों की आर्थिक स्थिति को जानने में रुचि रखते हैं। आप बालापुर गांव में बैजू की आर्थिक स्थिति की तुलना करना चाहते हैं। इसके लिए, आपको उनकी ज़मीन की आकार मापनीय बनानी होगी, अन्य बालापुर के किसानों की ज़मीन की आकार की तुलना करके।

आप देख सकते हैं कि बैजू द्वारा स्वामित्व में रखी गई ज़मीन –

- सामान्य मान में औसत से ऊपर है (अंकगणितीय माध्य)
- उसकी ज़मीन उसके पास है जो कि गांव के किसानों के आधे से ज़्यादा ज़मीन रखते हैं (माध्यांक)
- उसकी ज़मीन उसके पास है जो कि गांव के किसानों की अधिकांशिक ज़मीन रखते हैं (मोड/ बहुलक)

तो, आप इन मापों की मदद से बैजू की ज़मीन की स्थिति का आकलन कर सकते हैं।

बैजू की सांख्यिकीय आर्थिक स्थिति का मूल्यांकन करने के लिए, आपको बालापुर के किसानों के ज़मीन के स्वामित्व के सम्पूर्ण डेटा को संक्षेपित करना होगा। इसे केंद्रीय प्रवृत्ति का प्रयोग करके किया जा सकता है, जो डेटा को एक ही मूल्य में संक्षेपित करती है, इस प्रकार कि यह एक ही मूल्य पूरे डेटा को प्रतिनिधित्व कर सकता है। केंद्रीय प्रवृत्ति की मापनी होती है, जिससे डेटा को एक प्रकार के प्रतिनिधित्वात्मक मूल्य में संक्षेपित किया जा सकता है।

केंद्रीय प्रवृत्ति या "औसतों" के कई सांख्यिकीय माप होते हैं। तीन सर्वाधिक प्रचलित औसत निम्नलिखिय हैं:

- समान्तर माध्य/अंकगणितीय माध्य (Arithmetic Mean): सभी आंकड़ों का योग उनकी कुल संख्या से विभाजित करके प्राप्त होता है। यह सबसे प्रसिद्ध औसत है और सामान्यत: "औसत" के रूप में प्रयुक्त होता है।
- मध्यिका/माध्यांक (Median): डेटा को आर्डर में लागू किया जाता है और चोटीले और निचले 50% के बीच मध्य स्थिति का मान माध्यांक होता है। यह आवश्यक नहीं है कि डेटा सभी आंकड़ों का एक सामान्य वितरण हो, और यह आउटलायर्स के प्रति प्रभावित नहीं होता है।
- बहुलक/मोड (Mode): जो मूल्य डेटा में सबसे बार दिखता है, वह मोड होता है। अगर कोई मूल्य सबसे ज्यादा बार दिखता है, तो डेटा का मोड मात्रात्मक होता है।

ये तीनों माप डेटा की प्रतिनिधित्व करने के लिए प्रयुक्त होते हैं और उन्हें केंद्रीय प्रवृत्ति के रूप में संक्षिप्त किया जा सकता है। आपको यह भी ध्यान रखना चाहिए कि दो अन्य प्रकार के औसत और भी हैं, जैसे ज्यामितीय माध्य तथा इरात्मक माध्य, जो विशिष्ट परिस्थितियों में उपयुक्त होते हैं।

2. समान्तर माध्य/अंकगणितीय माध्य (Arithmetic Mean):

मान लीजिए कि छह परिवारों की मासिक आय (रुपयों में) निम्नलिखित है:

1600, 1500, 1400, 1525, 1625, 1630। माध्य परिवारिक आय प्राप्त की जाती है जिसे आयों को जोड़कर और परिवारों की संख्या से विभाजित करके प्राप्त किया जाता है।

माध्य = (1600 + 1500 + 1400 + 1525 + 1625 + 1630) / 6

= 9280 / 6

= 1546.67 रुपये (लगभग)

इसलिए, माध्य परिवारिक आय लगभग 1546.67 रुपये है।

समान्तर माध्य सेंट्रल टेंडेंसी का सबसे आमतौर पर प्रयुक्त माप है। इसे सभी अवलोकनों के मूल्यों का योग अवलोकनों की संख्या से विभाजित करके परिभाषित किया जाता है और आमतौर पर $\bar{X}$ द्वारा प्रकट किया जाता है। आमतौर पर, अगर N अवलोकन होते हैं जैसे X_1, X_2, X_3,..., X_N, तो समान्तर माध्य को इस प्रकार प्राप्त किया जायेगा:

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_N}{N} = \frac{1}{N}\sum_{i=1}^{N} X_i$$

यहाँ, "i" एक सूचक है जो क्रमिक मान 1, 2, 3,...,N लेता है।

समान्तर माध्य का परिकलन कैसे किया जाता है (Computation of Arithmetic Mean):

समान्तर माध्य की गणना को दो व्यापक श्रेणियों में अध्ययन किया जा सकता है:

- अनुगणित आँकड़ों का समान्तर माध्य।
- समूहित आँकड़ों का समान्तर माध्य।

अनुगणित आँकड़ों की श्रृंखला के लिए समान्तर माध्य (Arithemtic Mean for Ungrouped Data):

प्रत्यक्ष विधि:

प्रत्यक्ष विधि के द्वारा समान्तर माध्य निकालने के लिए किसी श्रृंखला में सभी प्रेक्षणों के योग को प्रेक्षणों की कुल संख्याओं से विभाजित किया जाता है।

उदाहरण 1: किसी कक्षा के अर्थशास्त्र की इम्तिहान में प्राप्तांक प्रदर्शित करने वाले आँकड़ों से समान्तर माध्य का परिकलन करें। 40, 50, 55, 78, 58.

दिए गए श्रृंखला में छात्रों के अर्थशास्त्र परीक्षा में अंक दिखाने से समान्तर माध्य (औसत) की गणना करने के लिए आप निम्नलिखित सूत्र का प्रयोग कर सकते हैं:

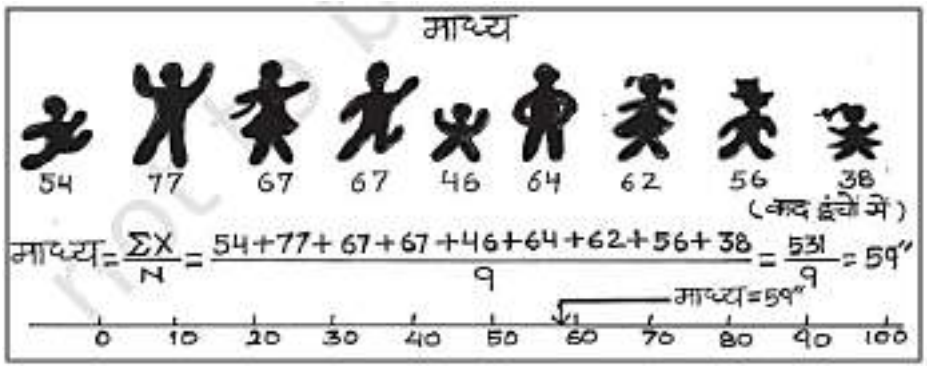
समान्तर माध्य = (सभी अवलोकनों का योग) / (कुल अवलोकनों की संख्या)

समान्तर माध्य = (40 + 50 + 55 + 78 + 58) / 5

समान्तर माध्य = 281 / 5

समान्तर माध्य = 56.2

अर्थशास्त्र की परीक्षा में छात्रों के औसत अंक 56.2 हैं।



(चित्र का स्रोत: NCERT 11 वी कक्षा की क़िताब अर्थशास्त्र में साँख्यिकी)

समूहित आँकड़ों के लिए समान्तर माध्य की गणना (Arithemtic Mean for Grouped Data):

विविक्त श्रृंखला (Discrete Series):

अगर श्रृंखला विविक्त है, तो प्रत्येक प्रेक्षण की बारंबारता (फ्रीक्वेंसी) को प्रेक्षण के मान के द्वारा गुणा किया जाता है। इससे जो मान प्राप्त होते हैं, उन्हें जोड़ा जाता है और बारंबारताओं की कुल सँख्या के द्वारा बिभाजित किया जाता है। प्रतीक के रूप में:

$$\bar{X} = \frac{\sum fx}{\sum f}$$

जहाँ पर, $\sum fx$ चरों के उत्पाद तथा बारंबारताओं का योग।

$\sum f$ = बारंबारताओं का योग

उदाहरण 2: एक आवासीय कॉलोनी में भूखण्ड सिर्फ तीन आकारों में आते हैं: 100 वर्ग मीटर, 200 वर्ग मीटर और 300 वर्ग मीटर और भूखण्डों की संख्या क्रमशः 200, 50 और 10 है।

भूखण्ड का आकर (वर्ग मीटर) 'x'	भूखण्डों की संख्या 'f'	fx
100	200	20000
200	50	10000
300	10	3000
कुल	260	33000

प्रत्यक्ष विधि के प्रयोग द्वारा समांतर माध्य,

$$\bar{X} = \frac{\sum fx}{\sum f} = \frac{33000}{260} = 126.92 \text{ वर्ग मीटर}$$

अतः आवासीय कॉलोनी में औसत भूखण्ड का आकार 126.92 वर्ग मीटर है।

संतत श्रृंखला (Continuous Series):

यहाँ, वर्ग अंतराल दिए गए हैं। संतत श्रृंखला के मामले में समान्तर माध्य की गणना की प्रक्रिया विविक्त श्रृंखला के मामले की तरह ही होती है। एकमात्र अंतर है कि विभिन्न वर्ग अंतरालों के बीच के मध्यबिंदु लिए जाते हैं। हम पहले से जानते हैं कि वर्ग अंतराल अपवर्जी या समावेशी या असमान आकार के हो

सकते हैं। अपवर्जी वर्ग अंतराल का उदाहरण, मान लीजिए, 0–10, 10–20 आदि है। समावेशी वर्ग अंतराल का उदाहरण, मान लीजिए, 0–9, 10–19 आदि है। असमान वर्ग अंतराल का उदाहरण, मान लीजिए, 0–20, 20–50 आदि है। इन सभी मामलों में, अंकगणितीय माध्य की गणना एक ही तरीके से की जाती है।

समूहित आँकड़ा (संतत श्रृंखला) के समान्तर माध्य की गणना करने के लिए निम्नलिखित चरणों का पालन करें:

- प्रत्येक समूह का मध्यांक ज्ञात करें। समूह के बाएं सीमा और दाएं सीमा का योग करके 2 से विभाजित करें। यह मध्यांक होगा।
- प्रत्येक समूह के मध्यांक को उस समूह के फ़्रीक्वेंसी से गुणा करें।
- प्रत्येक समूह के गुणित माध्यांकों का योग करें।
- उपरोक्त योग को उस समूहों की कुल फ़्रीक्वेंसी के साथ विभाजित करें। यह आपका समूहित डेटा का समान्तर माध्य होगा।

उदाहरण 3: यहाँ एक उदाहरण देखते हैं:

वर्ग	फ़्रीक्वेंसी	मध्यांक = (बाएं सीमा + दाएं सीमा) / 2	मध्यांक * फ़्रीक्वेंसी
10-20	5	15	75
20-30	8	25	200
30-40	12	35	420
40-50	10	45	450
50-60	15	55	825
योग	50		1970

कुल फ़्रीक्वेंसी = 5 + 8 + 12 + 10 + 15 = 50

समूहित डेटा (संतत श्रृंखला) का समान्तर माध्य = योग / कुल फ़्रीक्वेंसी = 1970/ 50 = 39.40

इस प्रकार, आप समूहित डेटा (संतत श्रृंखला) के समान्तर माध्य की गणना कर सकते हैं।

समान्तर माध्य की दो रोचक विशेषताएँ (Characteristics of Arithmetic Mean):

- समान्तर माध्य से मदों के विचलन का योग हमेशा शून्य के बराबर होता है।

- औसत माध्य चरम मूल्यों द्वारा प्रभावित होता है। कोई भी चरम मूल्य, किसी भी तरफ, औसत माध्य को ऊपर या नीचे धकेल सकता है।

3. मध्यिका/माध्यांक (Median):

मध्यिका वह स्थानिक मान होता है जो वितरण को दो बराबर भागों में विभाजित करता है, एक भाग में माध्यम मान से अधिक या उसके समान सभी मूल्य शामिल होते हैं और दूसरे भाग में उससे कम या उसके समान सभी मूल्य शामिल होते हैं। मध्यिका वह "मध्य" तत्व होता है जब डेटा सेट को विचार के क्रम में व्यवस्थित किया जाता है। क्योंकि माध्य विभिन्न मूल्यों के स्थान के द्वारा निर्धारित होता है, यह असर नहीं पड़ता है, यदि कहें, सबसे बड़े मूल्य का आकार बढ़ जाता है।

मध्यिका की गणना (Computation of Median):

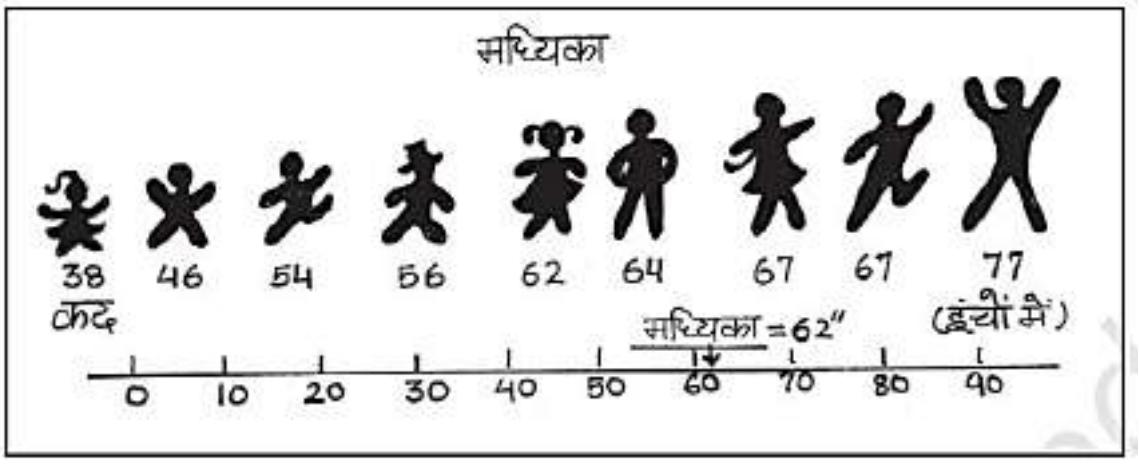
मध्यिका की गणना आसानी से की जा सकती है जिसमें आँकड़ों को सबसे छोटे से बड़े तक क्रमित करके मध्य मूल्य की पहचान की जाती है।

उदाहरण 4: मान लीजिए हमारे पास एक आंकडा समुच्च में निम्नलिखित अवलोकन हैं: 5, 7, 6, 1, 8, 10, 12, 4, और 3। आँकड़ों को आरोही क्रम में व्यवस्थित करते हुए, आपके पास निम्नलिखित क्रम होगा:

1, 3, 4, 5, 6, 7, 8, 10, 12

यहाँ पर "मध्य अंक" 6 है, इसलिए मध्यिका 6 हैं । आधे अंक 6 से बड़े हैं और आधे अंक उससे छोटे हैं।

यदि आंकड़ों में सम संख्याऐं होती हैं, तो मध्य में दो प्रेक्षण होंगे। ऐसी स्थिति में मध्यिका उन दो मध्य मूल्यों के समान्तर माध्य के रूप में गणना किया जाता है।



(चित्र का स्रोत: NCERT 11 वी कक्षा की क़िताब अर्थशास्त्र में साँख्यिकी)

4. बहुलक/मोड (Mode):

कभी-कभी, आपको एक श्रृंखला के सबसे प्रतिष्ठित मूल्य या ऐसे मूल्य की जानकारी हो सकती है जिसके चारों ओर आइटमों की अधिकांश संघटित होती है। उदाहरण के लिए, एक निर्माता को उस साइज की जानकारी चाहिए होती है जिसमें मैक्सिमम मांग होती है या वो शर्ट की स्टाइल जिसकी अधिकांश मांग होती है। यहाँ, मोड सबसे उपयुक्त माप है। "Mode" शब्द का उत्थान फ्रेंच शब्द "la Mode" से हुआ है, जिसका अर्थ होता है वितरण के सबसे फैशनेबल मूल्य, क्योंकि यह श्रृंखला में सबसे अधिक बार दोहराया जाता है। मोड सबसे अधिक देखे जाने वाले डेटा मूल्य होता है। इसे M_o से दर्शाया जाता है।

बहुलक का अभिकलन (Computation of Mode):

विविक्त श्रृंखला (Discrete Series):

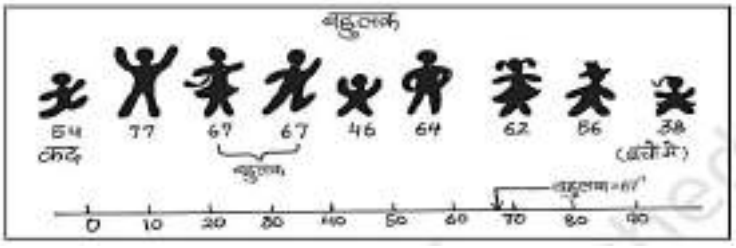
उदाहरण 5: आँकड़ा समुच्च 1, 2, 3, 4, 4, 5 पर विचार करें।

इस आँकड़ा समुच्च के लिए बहुलक 4 है क्योंकि आंकडा समुच्च में 4 सबसे अधिक बार (दो बार) होता है।

निम्नलिखित विविक्त श्रृंखला को देखियें:

चर	10	20	30	40	50
बारंबारता (फ्रीक्वेंसी)	2	8	20	10	5

यहाँ, जैसा कि आप देख सकते हैं, अधिकतम फ्रीक्वेंसी 20 है, मोड की मान 30 है। इस मामले में, जैसा कि एक अद्वितीय मोड की मान है, डेटा एकमोदाल है। लेकिन मोड आवश्यक रूप से अद्वितीय नहीं होता है, यह अंकगणितीय माध्य और माध्य में की तरह। आपके पास दो मोड वाले डेटा (द्विमोदल) या दो से अधिक मोड वाले डेटा (बहुमोदल) हो सकते हैं। ऐसा हो सकता है कि वितरण में कोई मूल्य किसी अन्य मूल्य से अधिक बार नहीं प्रकट होता है तो कोई मोड नहीं हो सकता है। उदाहरण के लिए, एक श्रृंखला 1, 1, 2, 2, 3, 3, 4, 4 में कोई मोड नहीं है।



(चित्र का स्रोत: NCERT 11 वी कक्षा की क़िताब अर्थशास्त्र में साँख्यिकी)

5. परिक्षेपण के माप (Measures of Dispersion):

पिछले यूनिट में, आपने सीखा है कि कैसे आँकड़ों को एक एकल प्रतिनिधित्वक मूल्य में संक्षेपित किया जाता है। हालांकि, वह मूल्य उस विविधता को प्रकट नहीं करता जो आँकड़ों में मौजूद होती है। इस यूनिट में आप वे मापें सीखेंगे, जो आंकड़ों की विविधता को मापने का प्रयास करती हैं। तीन दोस्त, राम, रहीम और मारिया, एक कप चाय के साथ बातचीत कर रहे हैं। उनकी बातचीत के कोर्स के दौरान, उन्होंने अपने परिवार के आय के बारे में बात करना शुरू किया। राम उन्हें बताते हैं कि उनके परिवार में चार सदस्य हैं और प्रति सदस्य औसत आय 15,000 रुपये है। रहीम कहते हैं कि उनके परिवार में आय की औसत समान है, हालांकि सदस्यों की संख्या छह है। मारिया कहती है कि उनके परिवार में पांच सदस्य हैं, जिनमें से एक काम नहीं कर रहा है। उन्होंने गणना की है कि उनके परिवार में भी औसत आय 15,000 रुपये है। वे थोड़े आश्चर्यित हैं क्योंकि उन्हें पता है कि मारिया के पिता की बहुत अधिक तनख़्वाह है। उन्होंने विस्तार से पता किया एवं निम्नलिखित आंकड़ों को इकठ्ठा किया :

पारिवारिक आय (रुपयों में)

क्र.स.	राम	रहीम	मारिया
1.	12,000	7,000	0
2.	14,000	10,000	7,000
3.	16,000	14,000	8,000
4.	18,000	17,000	10,000
5.	.....	20,000	50,000
6.	.....	22,000	.....
कुल आय	60,000	90,000	75,000
औसत आय	*15,000*	*15,000*	*15,000*

क्या आपने ध्यान दिया कि हालांकि औसत समान है, व्यक्तिगत आयों में काफी भिन्नताएँ हैं?

यह काफी स्पष्ट है कि औसतें केवल एक पहलू को दर्शाने की कोशिश करती हैं, जो की मूल्यों की एक प्रतिष्ठानक आकार होती है। इसे और बेहतर समझने के लिए, आपको मूल्यों की प्रसार को भी जानने की आवश्यकता है। आप देख सकते हैं कि राम के परिवार में आय में अंतर संज्ञानीय रूप से कम है। रहीम के परिवार में अंतर अधिक है और मारिया के परिवार में अंतर सबसे अधिक है। केवल औसत का ज्ञान पर्याप्त नहीं है।

उदाहरण के लिए, प्रति व्यक्ति आय केवल औसत आय प्रदान करता है। एक विचलन का माप आपको आय की असमानताओं के बारे में बता सकता है, इसके आधार पर सामाजिक वर्गों के अलग-अलग परिप्रेक्ष्य में आनंदित जीवन मानकों की समझ को सुधार सकता है।

विचलन वह दिशानिर्देश है जिसमें वितरण में मूल्यों की विभिन्नता औसत से कितनी अलग होती है।

विभिन्नता की मात्रा को मापने के लिए कुछ निम्नलिखित माप होते हैं:

- परास (Range)
- चतुर्थक विचलन (Quartile Deviation)
- माध्य विचलन (Mean Deviation)
- मानक विचलन (Standard Deviation)

परास एवं चतुर्थक विचलन परिक्षेपण की माप उस प्रसरण के परिकलन द्वारा करते हैं, जिसमें ये मान निहित होते हैं। माध्य विचलन तथा मानक विचलन औसत से मानों के अंतर की मात्रा को मापते हैं।

मानों के प्रसरण पर आधरित माप (Measures based upon Spread of Values):

6. परास (Range):

परास किसी वितरण में अधिकतम L एवं न्यूनतम S मानों के बीच का अंतर है। अतः R = L-S परास का अधिक मान अधिक परिक्षेपण दर्शाता है और, इसके विपरीत कम मान निम्न परिक्षेपण को दर्शाता है।

परास: टिप्पणी

- परास चरम मान के द्वारा अनुचित रूप से प्रभावित होता है। यह सभी मानों पर आधरित नहीं है। जब तक न्यूनतम एवं अधिकतम मान अपरिवर्तित रहते हैं, तब तक दूसरे मानों में कोई भी बदलाव परास को प्रभावित नहीं करता। इसे मुक्तांत बारंबारता वितरण में परिकलित नहीं किया जा सकता है।
- कुछ सीमाओं वेफ होते हुए भी परास अपनी सरलता के कारण आसानी से समझा एवं बहुधा प्रयुक्त किया जाता है। उदाहरण वेफ लिए, हम लोग दूरदर्शन पर विभिन्न शहरों का दैनिक अधिकतम एवं न्यूनतम तापमान देखते रहते हैं और तापमान विविध्ता के आधर पर उनके बारे में राय बनाते हैं।

7. चतुर्थक विचलन (Quartile Deviation):

किसी वितरण में उच्च या निम्न किसी भी चरम मान की उपस्थिति परिक्षेपण के माप के रूप में परास की उपयोगिता को घटा सकती है। इसलिए, आपको एक ऐसे माप की ज़रूरत हो सकती है, जो कि बाह्यमूल्यों से अनुचित रूप से प्रभावित न हो। ऐसी स्थिति में, यदि संपूर्ण आँकड़ों को चार बराबर भागों में विभाजित

किया जाए, तो प्रत्येक में मानों का 25% भाग समाहित होगा, जिससे हमें चतुर्थकों एवं मध्यिका का मान प्राप्त होता है ।

उच्च एवं निम्न चतुर्थक (क्रमशः Q_3 एवं Q_1) का प्रयोग अंतर-चतुर्थक परास के परिकलन में किया जाता है, जो Q_3-Q_1 हैं। अंतर-चतुर्थक परास, किसी वितरण में माध्य के 50% मानों पर आधरित होता है। अतः वह चरम मान के द्वारा प्रभावित नहीं होता है। अंतर-चतुर्थक परास के आधे को चतुर्थक-विचलन (QD) कहा जाता है। अतः

$$QD = \frac{Q_3 - Q_1}{2}$$

चतुर्थक विचलन को अर्ध-अंतर-चतुर्थक परास भी कहा जाता है।

असमूहित आँकड़ों के लिए परास और चतुर्थक विचलन का परिकलन (Computation of Quartile Deviation for Ungrouped Data):

उदाहरण 6: निम्नलिखित प्रेक्षणों का परास और चतुर्थक विचलन परिकलित कीजिए:

20, 25, 29, 30, 35, 39, 41, 48, 51, 60 और 70

स्पष्टतः परास 70-20 = 50 है।

चतुर्थक विचलन के लिए हमें उच्च Q_3 एवं निम्न Q_1 के मानों को परिकलित करने की आवश्यकता होती है।

Q_1मान $\frac{n+1}{4}$ वें मद का आकार है।

चूँकि n 11 है, Q_1 तीसरे मद का आकार है। क्योंकि मानों को पहले ही आरोही क्रम में व्यवस्थित किया हुआ है, यह देखा जा सकता है कि Q_1 तीसरा मान 29 है। ठीक इसी तरह से, Q_3 मान $\frac{3(n+1)}{4}$ वें मद का आकार है। अर्थात 9वें मद का मान, 51 है। अतः Q_3 = 51

$$QD = \frac{Q_3 - Q_1}{2} = \frac{51 - 29}{2} = 11$$

क्या आपने ध्यान दिया है कि QD मध्यिका से चतुर्थकों का औसत अंतर है।

औसत से परिक्षेपण के माप (Measures of Dispersion from Average):

आपको याद होगा कि परिक्षेपण हमें यह बतलाता है कि किसी वितरण की विभिन्न मदों का मान वितरण के औसत मान से किस सीमा तक भिन्न है। परास और चतुर्थक विचलन माप में उपयोगी नहीं हैं कि मान अपने औसत से कितनी दूर हैं, फिर भी मानों के प्रसरण के परिकलन द्वारा वे परिक्षेपण के बारे में एक अच्छा अनुमान दे देते हैं। दो माप, जोकि मानों के अपने औसत से विचलन पर आधरित होते हैं वे हैं, माध्य विचलन और मानक विचलन।

चूँकि औसत एक केन्द्रीय मान है, कुछ विचलन धनात्मक और कुछ ऋणात्मक होते हैं। अगर उन्हें ऐसे ही जोड़ दिया जाए, तो जोड़ से कोई परिणाम नहीं निकलेगा। वास्तव में समांतर माध्य से विचलनों का योग सदैव शून्य होता है। मानों के निम्न दो समुच्चयों को देखेंः

समुच्चय अः 5, 9, 16

समुच्चय बः 1, 9, 20

आप देख सकते हैं कि समुच्चय ब में मान अपने औसत से अधिक दूर है और इसलिए समुच्चय अ के मानों की अपेक्षा अधिक प्रसरित है। यहाँ समांतर माध्य से विचलनों को परिकलित कीजिए और फिर उन्हें जोड़ दीजिए। आपने क्या देखा? अब यही क्रिया मध्यिका के साथ दोहराइए। क्या आप परिकलित मानों से विचरण की मात्रा पर टिप्पणी कर सकते हैं? माध्य विचलन, विचलनों के संकेतों की उपेक्षा करके इस समस्या का समाधन करने की कोशिश करता है, अर्थात् यह सभी विचलनों को धनात्मक मानता है। मानक विचलन के लिए, पहले विचलनों के वर्गों का परिकलन करके उनका औसत निकाला जाता है। इसके बाद औसत का वर्गमूल निकाला जाता है।

8. माध्य विचलन (Mean Deviation):

मान लीजिए पाँच कस्बों A, B, C, D और E के लिए एक कॅालेज प्रस्तावित किया जाता है। ये कस्बे एक सड़क के किनारे इसी क्रम से स्थित हैं। A कस्बे से दूसरे कस्बों की दूरी (किलोमीटर में) तथा छात्रों की संख्या नीचे दी जा रही है।

कस्बे	कस्बा A से दूरी	छात्रों की संख्या
A	0	90
B	2	150
C	6	100
D	14	200
E	18	80
		620

अब, यदि कॉलेज कस्बा A में स्थित होता है तो कस्बा B के 150 छात्रा 2 किमी प्रति छात्रा के हिसाब से (कुल 300 किमी) यात्रा करके कॉलेज पहुँचेंगे। उद्देश्य यह है कि ऐसी जगह पता करें, जिससे छात्रों को कम से कम औसत दूरी की यात्रा करनी पड़े। आप देख सकते हैं कि यदि कॉलेज A या E कस्बे में स्थित होता है तो छात्रों को औसतन अधिक यात्रा करनी होगी और यदि कॉलेज किसी मध्यवर्ती जगह पर स्थित होता है, तो उन्हें अपेक्षाकृत कम यात्रा करनी पड़ेगी। माध्य विचलन औसत से अंतरों का समांतर माध्य है। यहाँ प्रयुक्त उपयुक्त सांख्यिकी उपकरण है जिससे छात्रों द्वारा तय की गई औसत दूरी का आकलन किया जा सकता है। माध्य विचलन औसत या तो समांतर माध्य है या मध्यिका। चूँकि बहुलक एक स्थिर औसत नहीं है अतः माध्य विचलन के परिकलन हेतु इसका प्रयोग नहीं किया जाता है।

असमूहित आँकड़ों के लिए समांतर माध्य से माध्य विचलन का परिकलन:

इस विधि के निम्नलिखित चरण हैं:

- मानों का समांतर माध्य परिकलित किया जाता है।
- प्रत्येक मान और समांतर माध्य के दरमियानी अंतर का परिकलन किया जाता है। ये सभी अंतर धनात्मक माने जाते हैं। इन्हें $|d|$ द्वारा दर्शाया जाता है।
- इन अंतरों का समांतर माध्य, माध्य विचलन है। अर्थात्

$$MD = \frac{\sum |d|}{n}$$

उदाहरण 7: निम्नलिखित मानों का माध्य विचलन परिकलित कीजिए: 2, 4, 7, 8 एवं 9.

$$\text{समांतर माध्य (A.M.)} = \frac{\Sigma X}{n} = \frac{30}{5} = 6$$

X	$\|d\|$
2	4
4	2
7	1
8	2
9	3
	12

$$MD = \frac{\sum |d|}{n} = \frac{12}{5} = 2.4$$

9. मानक विचलन (Standard Deviation):

मानक विचलन माध्य से विचलनों के वर्गों के माध्य का धनात्मक वर्गमूल है। इसलिए यदि पाँच मान X_1, X_2, X_3, X_4 एवं X_5 हैं, तो सबसे पहले इनका माध्य परिकलित किया जाता है। इसके बाद माध्य से मानों के विचलन परिकलित किए जाते हैं। फिर इन विचलनों का वर्ग किया जाता है। इन वर्ग विचलनों का माध्य प्रसरण कहलाता है। प्रसरण का धनात्मक वर्गमूल मानक विचलन होता है। (यह ध्यान दें कि मानक विचलन का परिकलन सिर्फ़ माध्य के आधार पर होता है)

असमूहित आँकड़ों के लिए मानक विचलन का परिकलन (Standard Deviation for Ungrouped data):

उदाहरण 8: मान लीजिए, आपको निम्नलिखित मानों का मानक विचलन परिकलित करना हैः

5, 10, 25, 30, 50

इसकी गणना के लिए प्रथम चरण यह होगा:

$$\bar{X} = \frac{5 + 10 + 25 + 30 + 50}{5} = \frac{120}{5} = 24$$

X	$d = (X - \bar{X})$	d^2
5	-19	361
10	-14	196
25	+1	1
30	+6	36
50	+26	676
	0	1270

अब निम्नलिखित सूत्रा प्रयुक्त होगाः

$$\sigma = \sqrt{\frac{\sum d^2}{n}}$$

$$\sigma = \sqrt{\frac{\sum (X - \bar{X})^2}{n}}$$

$$\sigma = \sqrt{\frac{1270}{5}} = \sqrt{254} = 15.937$$

10.निष्कर्ष (Conclusion):

केन्द्रीय प्रवृत्ति की माप या औसतों का प्रयोग आँकड़ों के संक्षेपण के लिए किया जाता है। यह आँकड़ा-समुच्च का वर्णन करने के लिए एकल प्रतिनिधि मान को दर्शाता है। समांतर माध्य सर्वाधिक प्रयोग किया जाने वाला औसत है। यह परिकलन में सरल एवं सभी प्रेक्षणों पर आधरित होता है। लेकिन यह चरम मदों की उपस्थिति से अनुचित रूप से प्रभावित होता है। इस प्रकार के आँकड़ों के लिए मध्यिका अच्छा संक्षेपण है। बहुलक का प्रयोग सामान्यतः गुणात्मक आँकड़ों की व्याख्या में किया जाता है। मध्यिका एवं बहुलक को आलेखी तौर पर आसानी से अभिकलित किया जा सकता है। मुक्तांत वितरणों के लिए भी इनका अभिकलन सरलता से किया जा सकता है। इसलिए यह महत्वपूर्ण है कि हम विश्लेषण के उद्देश्य तथा वितरण की प्रकृति को देखते हुए उपयुक्त औसत का चुनाव करें। यद्यपि परास समझने तथा परिकलन के लिए सबसे सरल है, लेकिन यह चरम मानों से अनुचित रूप से प्रभावित होता है। चतुर्थक विचलन चरम मानों से प्रभावित नहीं होता है, क्योंकि वह आँकड़ों के केवल मध्यवर्ती 50» आँकड़ों पर आधरित होता है। परंतु, चतुर्थक विचलन का निर्वचन बहुत कठिन होता है। माध्य विचलन और मानक विचलन दोनों ही मानों के अपने औसत से विचलनों पर आधरित होते हैं। माध्य विचलन औसत से विचलनों के औसत को परिकलित करता है, लेकिन विचलन के चिन्हों की (धनात्मक तथा ऋणात्मक) अपेक्षा करता है।

अध्याय 04

सांख्यिकीय विश्लेषण में R सॉफ्टवेयर का उपयोग

पंकज दास
वैज्ञानिक, प्रतिदर्श सर्वेक्षण प्रभाग
भा.कृ.अनु.प.-भारतीय कृषि साँख्यिकी अनुसंधान संस्थान, लाइब्रेरी एवेन्यू, नई दिल्ली-110012

1. प्रस्तावना (Introduction):

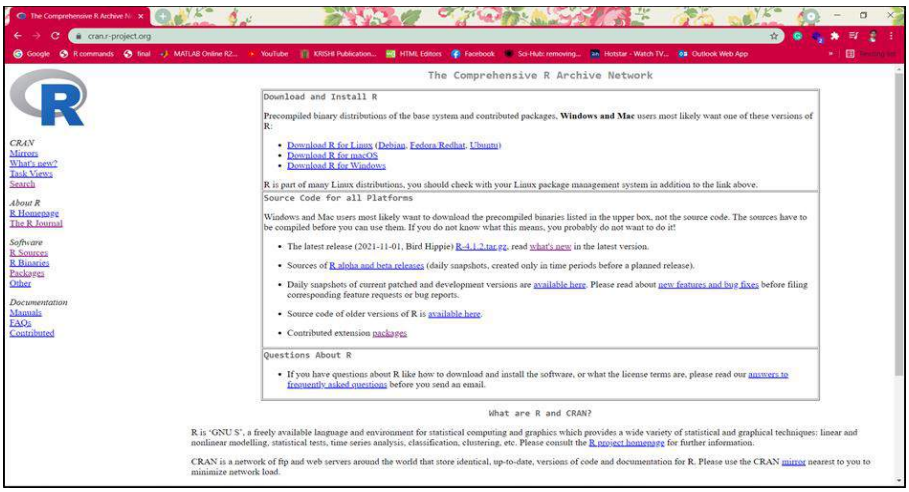
R एक प्रोग्रामिंग भाषा और वातावरण है जो आमतौर पर सांख्यिकीय कंप्यूटिंग, डेटा विश्लेषण और वैज्ञानिक अनुसंधान में उपयोग किया जाता है। यह सांख्यिकीविदों, डेटा विश्लेषकों, शोधकर्ताओं और विपणक द्वारा डेटा को पुनः प्राप्त करने, साफ करने, विश्लेषण करने, कल्पना करने और प्रस्तुत करने के लिए उपयोग की जाने वाली सबसे लोकप्रिय भाषाओं में से एक है। R को न्यूजीलैंड के ऑकलैंड विश्वविद्यालय में रॉस इहाका और रॉबर्ट जेंटलमैन द्वारा बनाया गया था, और वर्तमान में R डेवलपमेंट कोर टीम द्वारा विकसित किया गया है।

R की विशेषताएं (Features of R):

- R एक अच्छी तरह से विकसित, सरल और प्रभावी प्रोग्रामिंग भाषा है जिसमें सशर्त, लूप, उपयोगकर्ता परिभाषित पुनरावर्ती कार्य और इनपुट और आउटपुट सुविधाएं शामिल हैं।
- R के पास प्रभावी डेटा प्रबंधन और भंडारण की सुविधा है।
- R सरणियों, सूचियों, वैक्टर और मैट्रिक्स पर गणना के लिए ऑपरेटरों का एक सूट प्रदान करता है।
- R डेटा विश्लेषण के लिए उपकरणों का एक बड़ा, सुसंगत और एकीकृत संग्रह प्रदान करता है।
- R डेटा विश्लेषण और सीधे प्रदर्शित करने के लिए चित्रमय सुविधाएं प्रदान करता है।

2. R सॉफ्टवेयर को डाउनलोड करना (To Download R-Softwsare)

- सॉफ्टवेयर को डाउनलोड करना (**http://cran.r-project.org**).



- सी आर ए एन डाऊनलोड करें।
- अपने मिरर को सेट करे : भारत या कोई अन्य देश।
- दायी और आप विण्डोज के लिए R डाऊनलोड करें यह देखेंगे। उस पर क्लिक करें।
- बेस पर क्लिक करें।
- **R-4.1.2-win.exe** पर क्लिक करें और इसे अपने हार्डडिस्क में सेव करें।
- यह सॉफ्टवेयर का नवीनतम संस्करण है। यह एक प्रबन्धक (executive) फाइल '.exe' है इसे हार्डडिस्क पर सेव किया जा सकता है।

इस फाइल के नाम पर दो बार क्लिक करने से R अपने आप संस्थापित हो जाता है। इसे संस्थापित करने के लिए इन सभी प्रक्रियाओं को अपनाने की जरूरत है।

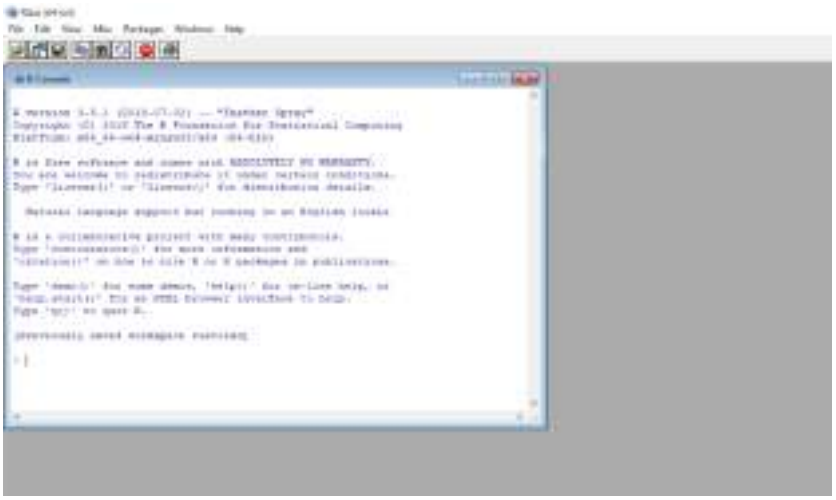
सॉफ्टवेयर को खोलना (Opening R Software)

स्थापना प्रक्रिया R के लिए एक स्वतः शॉर्टकट आईकॅन बनाती है। R परिवेश में खोलने के लिए इस आईकॅन को दो बार क्लिक करें। R के क्रियान्वयन का आभास एक मानक विण्डो के साथ खुलते हुए होगा। R एक व्याख्यात्मक भाषा है जो पंक्तिबद्ध तरीके से प्रक्रिया सम्पन्न करती है।

अतः R कोड में एक लाइन टंकित (अथवा पेस्ट करने) करने के पश्चात एन्टर को दबाना आवश्यक है ताकि R द्वारा इसे क्रियान्वित किया जा सके।

3. R प्रोग्राम कोड चलाना (To run R programme code)

R वातावरण के अन्तर्गत **R** कन्सोल (R console) एक सक्रिय विण्डो है। यह एक संयुक्त विण्डो में एक पंक्ति एडिटर और आउटपुट दिखाने वाला प्रोग्राम है। यहाँ पर कमांण्ड प्रोम्पट (command prompt) (चिन्ह >) में हम R कमांण्ड को एन्टर कर सकते हैं जो वाहक (carriage) कि रिटर्न–की दबाने पर तुरन्त चलाती (रन) हैं। (चिन्ह >) को प्रोम्पट कहते हैं क्योंकि यह उपयोगकर्ता को तुरन्त कुछ लिखने के लिए कहता है। कृपया देखें



हम कोड के खण्ड को भी चला सकते हैं जिसे हमनें दूसरे स्रोतों से बफर में चिपकाया है। इस सत्र में हम प्रदर्शन के लिए विण्डो आपूर्ति एडिटर नोट पेड का प्रयोग करेंगे और अपने R प्रोग्राम कोड को सम्पादित करेंगे। यदि हमें कोड मे कुछ और लिखना है तो साधारणतया इसे एडिटर से कापी करते है और इसे R कन्सोल में पेस्ट कर देते हैं, इससे कोड सही समय पर चलता है।

4. एडिटर को खोलना (Opening editor):

यहाँ हम प्रदर्शित करने के लिए विण्डो आपूर्ति एडिटर नोट पैड का प्रयोग करते है और अपने R–कार्यक्रम को एडिट करते है। हालांकि कोई भी व्यापक प्रयोजन के लिये एडिटर पर्याप्त होगा। नोट पैड को खोलने के लिए स्टार्ट बटन पर जाकर उसे क्लिक करें।

start > all programme > Accessories> Note> Pad आरम्भ > सम्पूर्ण कार्यक्रम > सहायक सामग्री > नोट पैड

नोट पैड को खुला रखकर, फाइल को खोलें उदाहरणतया **Intro_to_R.txt** विकल्प सूची (Menu) से निम्नलिखित विकल्प का चयन करते हुए फाइल को खोले (कार्यक्रम कोड युक्त, माना कि यह सी ड्राइव में कापी है)

File > open फाईल > खोले

"ओपन" नामक संवाद बॉक्स के ऊपर 'डाऊन–एरो' पर क्लिक करें तथा सिलेक्शन को 'लुक इन' c:/ में परिवर्तित करें। अब आप फाइलों की सूची में फाइल नाम **Intro_to_R.txt** देखेंगे। इसे खोलने के लिए फाइल के नाम पर दो बार क्लिक करें।

कुछ अन्य उपयोगी बातें :

कृपया ध्यान रखे कि R प्रकरण–संवेदनशील (case sensitive) है इसलिए अपने लक्ष्य की पूर्ति के लिए हमें छोटे और बड़े दोनो अक्षरों के प्रयोग में सावधानी बरतने की आवश्यकता है। जब कार्यक्रम समाप्त होता है तब हमें R कन्सोल विण्डो में लाल कमाण्ड प्रोम्पट (command promp) (>) दिखाई देना चाहिए। इससे यह अकित होता है कि कन्ट्रोल प्रयोगकर्ता के पास वापस आ जाता है ताकि यदि आप चाहे R कमाण्ड में और टंकित कर सकते हैं। R कोड में टिप्पणी हैश संकेत (**#**) से शुरू होती है।

सहायता (Help)

मैन्यू के माध्यस से, हेल्प का उपयोग किया जा सकता है। वेवपेज से आप मैन्यूल्स, बार–बार पूछे जाने वाले प्रश्न, व्यक्तिगत पैकेज के लिये सहायता का संदर्भ, और सबसे अधिक महत्वपूर्ण खोज इंजन

(search engine) को एक्सिस कर सकते हैं। नये फंक्शन्स विषय में पता लगाने के लिए 'सहायता' (help) एक महत्वपूर्ण साधन है इसे उपयोग करने के सबंध में इसका विवरण प्राप्त करें।

निम्न कमाण्ड की कोशिश करें।

```
> help(Syntax)
> ?Syntax
```

5. R के साथ शुरू करना (Getting started with R)

कमाण्ड प्रोम्पट में **R** की कमाण्ड्स उपलब्ध होती है। साधारण गणना, वैक्टर और ग्राफिक्स (Simple calculations, vectors and graphics)

शुरू के साथ R हम को केलकुलेटर की तरह उपयोग करना बताँएगे। निम्न कमाण्ड की कोशिश करें।

```
2 + 7
2/(3+5)
sqrt(9) $ 5^2
sin(π/2)-log(exp(1))
```

विशेष कमाण्ड लिखने की सहायता के लिये किसी कमाण्ड से पहले प्रश्नवाचक चिन्ह लगाकर कर सकते हैं उदाहरणतया

```
?log
```

एक विकल्प के रूप में, सहायता का उपयोग कर सकते हैं। इस मामले में, सहायता (लॉग) सहायता फाईल एक विशेष स्रोत है और आप बार बार प्रयोग करने पर इससे अवगत हो जायेंगे । कमाण्ड को निम्न प्रकार हैश संकेत (रु) का प्रयोग करते हुए लिख सकते हैं।

```
2+3 # उत्तर 5 होना चाहिए
```

- **सदिश और आव्यूह (vectors and matrix) :**

सदिश और आव्यूह अनेक संख्यात्मक समस्याओं के समाधान में बहुत महत्वपूर्ण है उलकंजं नाम के सदिश बनाने में इसको तथा 7–2, 5 मूल्य देने के लिये, हम इसे निम्न प्रकार लिखते है :

```
mydata <- c(7,-2,5)
```

संकेत <- (या विकल्प के प्रयोग =) को एशाइन (assigns) चाहिये। कमाण्ड c की व्याख्या कॉलम अथवा संयुक्त रुप मे कर सकते है सदिश के दूसरे तत्व को इस कमाण्ड के द्वारा निर्दिष्ट (Referred) करते हैं।

```
mydata[2]
```

और इसे 2 ओर 3 तत्व के मध्य व्यत्त करते है

```
mydata[2:3]
```

सदिश में हेर फेर भी किया जा सकता है। उदाहरण के लिए, सभी तत्वों को निम्न प्रकार से लगातार जोड़ने के द्वारा,

```
myconst <-100; mydata+ myconst
```

सेमिकोलन ";" (semicolon) का उपयोग हमें एक लाइन में एक से अधिक कमाण्ड लिखने की अनुमति देता है।

एक सदिश X, 1 से 10 पूर्णाकों के मध्य बना होता है ; 1, 2, ·····.10 ; को निम्न प्रकार से बनाकर लिख सकते हैं :

```
x <- c(1:10)
```

सदिश अनुक्रम संख्याओ के पास विशेष वृद्धि के साथ seq कमाण्ड से सृजित किया जा सकता है।

mydata1<- seq(0,10,2) # 0 और 10 पूर्णाकों के मध्य, 2 की वृद्धि के साथ rbind और cbind को पंक्ति या स्तम्भ सदिश का आव्यूह विलय के लिए प्रयोग किया जा सकता है निम्न लिखित प्रयास करें

```
X< - c (1, 2, 3)
Y< - c (4, 5, 6)
A = cbind (x, y)
B = rbind (x, y)
C = t (B)
```

अन्तिम कमाण्ड B को आव्यूह परिवर्तन देता है। अब A, B अथवा C लिखकर देखें कि ये भिन्न आव्यूह कैसी देखती है।

- **सरल ग्राफिक्स (Simple graphics)**

फलन का रेखांकन उपयुक्त सदिश निर्माण और आदेश कथानक के उपयोग द्वारा किया जा सकता है।

```
x <- seq(0,10,0.2); y <- sqrt(x); plot(x,y); grid()
```

जैसा कि एक अनुमान हो सकता है कि अंतिम कमाण्ड ग्रिड को कथानक से जोडता है।

- **आँकड़ा प्रबन्धन (Handling data)**

आँकड़ा ढाँचा तैयार करना (Creating data frames) विभिन्न प्रकार के आँकड़ें को संगठित करने तथा उक्त आँकडो के उपसमूह (subsets) प्राप्त करने के लिय कमाण्ड आँकड़ा का उपयोग किया जा सकता है। मान लीजिये कि हमारे पास तीन व्यक्तियों से सम्बन्धित आँकड़े हैं जिसे हम निम्न प्रकार से संग्रह करते है :

```
Length<-c(180, 175, 190)
Weight<-c(75, 82, 88)
Name<-c("Anil", "Ankit" "Sunil")
Friends<-data.frame (name,length, weight)
```

Friends अब एक आँकड़ा फ्रेम है जिसमें तीन व्यक्तियों से सम्बन्धित आँकड़े है। आँकड़े सरलता से प्राप्त किये जा सकते है :

```
my.name<-friends$name
length1<-friends$length [1]
```

6. फाइल से आँकड़ो को पढ़ना (Reading data from file):

यह सामान्य सी बात है कि आँकड़े को टेक्स्ट फाइल मे संग्रहित किया जाता है तथा हम इस आँकड़े को R में लेना चाहते हैं। हम दो मामलों का अध्ययन करेंगे, एक पूर्णतया संख्यात्मक आँकड़ों के साथ तथा दूसरा हैडर (Header) सहित संख्यात्मक आँकड़ो के साथ। coins.dat फाइल में विभिन्न अवधि के 27 चाँदी के सिक्कों में चाँदी की माञा के बारे में आँकडे संग्रहित हैं । जब हम निम्न कमाण्ड का उपयोग करते है :

```
mynt<-read.table("coins.dat")
```

(अवधि) नामक दो कॉलम है। दसवे प्रेक्षणो की विशेषताए देखने के लिये हम निम्न प्रकार से लिख सकते है :

```
mynt1 [10,] # [10,] ''सब कुछ दसवीं पंक्ति में''
```

यदि 10वें सिक्के मे केवल चांदी की मात्रा देखना चाहते हैं तब हम निम्न प्रकार लिख सकते हैं :

```
mynt1$ V1[10] # vectorV1( the first vector), row 10
```

अथवा

```
mynt1 [10,1] # 10वीं पंक्ति, स्तम्भ 1
```

यह जानना महत्वपूर्ण है कि स्तम्भ में किस प्रकार के आँकड़े संग्रहित हैं। यह स्प" ट करने के लिये आँकड़े लेने के समय स्तम्भ के ऊपर शीर्षक लिख देना चाहिए।

```
mynt2 <- read.table ("coin.dat",col.names=c ("silver","Epoch'))
```

इस प्रकार V1 और V2 के स्थान पर चांदी और अवधि का उपयोग करेंगे। इस आँकड़ा समुचय के अध्ययन में हम आँकड़ों को अलग प्रकार से संग्रहित करते हैं। फाइल coin.txt में यही आँकड़े संग्रहित हैं परन्तु फाइल की पहली पंक्ति में शीर्षक संग्रहित है। शीर्षक सहित आँकड़ों को लाने हेतु हम इस प्रकार लिख सकते हैं :

```
mynt2<- read.table ("coins.txt, header= TRUE)
```

पूर्व की तरह विशेष प्रेषणों (particular observatives) की सूचना प्राप्त करने का प्रयास करैं। आँकड़े प्राप्त करने में फलन (functions) है उदाहरणतयाः डेटाबेस अथवा एक्सिल स्प्रेड–शीट, परन्तु ये अत्यन्त उच्च श्रेणी के हैं और इस व्याख्यान में शामिल नहीं हैं। प्रायः कोई भी डाटाबेस एक्सिल स्प्रेड–शीट से टेक्स फाइल में कापी कर सकता है इसलिए read.table कमाण्ड का प्रयोग करते हैं।

7. R सत्र से बाहर निकलें (Exit the R session):

R सत्र से बाहर निकलने के लिए, R प्रॉम्प्ट में quit() टाइप करें, और कार्यक्षेत्र छवि को सहेजने के लिए 'n' (नहीं) कहें। इसका मतलब है, हम वर्तमान सत्र में हमारे द्वारा टाइप की गई सभी कमांड की मेमोरी को सहेजना नहीं चाहते हैं:

```
> quit()
Save workspace image? [y/n/c]: n
```

8. R सत्र सहेजा जा रहा है (Saving the R session):

ध्यान दें कि वर्तमान सत्र को न सहेजकर, हम वर्तमान सत्र कमांड की सभी मेमोरी और आर प्रॉम्प्ट से बाहर निकलने पर बनाए गए चर और वस्तुओं को खो देते हैं। जब हम R में काम करते हैं, तो हमारे द्वारा बनाए और लोड किए गए R ऑब्जेक्ट्स को वर्कस्पेस नामक मेमोरी भाग में संग्रहीत किया जाता है। जब हम कार्यस्थान को सहेजने के

लिए 'नहीं' कहते हैं, तो हम इन सभी वस्तुओं को कार्यस्थान स्मृति से मिटा दिया जाता है। यदि हम 'हाँ' कहते हैं, तो वे ".**RData**" नामक फ़ाइल में सहेजे जाते हैं जो वर्तमान कार्यशील निर्देशिका को लिखा जाता है।

9. डाटासेट का विश्लेषण (Analysis of a data set):

हम 1970 के प्रारम्भिक दशक की विभिन्न कारो के आँकड़ों के बारे में डाटासेट का अध्ययन करेंगे। डाटासेट को निम्न प्रकार से लिखकर लोड करे :

```
data (mtcars)
```

हेल्प फाइल को देखकर आँकड़ों के बारे में अधिक जानकारी प्राप्त कर सकते हैं।

```
? mtcars
```

mtcars {datasets} R Documentation

Motor Trend Car Road Tests

Description

The data was extracted from the 1974 *Motor Trend* US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973–74 models).

Usage

```
mtcars
```

Format

A data frame with 32 observations on 11 (numeric) variables.

[, 1] mpg Miles/(US) gallon
[, 2] cyl Number of cylinders
[, 3] disp Displacement (cu.in.)
[, 4] hp Gross horsepower
[, 5] drat Rear axle ratio
[, 6] wt Weight (1000 lbs)
[, 7] qsec 1/4 mile time
[, 8] vs Engine (0 = V-shaped, 1 = straight)
[, 9] am Transmission (0 = automatic, 1 = manual)
[,10] gear Number of forward gears

[,11] carb Number of carburetors

Source

Henderson and Velleman (1981), Building multiple regression models interactively. *Biometrics*, **37**, 391–411.

Examples

```
require(graphics)
pairs(mtcars, main = "mtcars data", gap = 1/4)
coplot(mpg ~ disp | as.factor(cyl), data = mtcars,
       panel = panel.smooth, rows = 1)
## possibly more meaningful, e.g., for summary() or bivariate plots:
mtcars2 <- within(mtcars, {
   vs <- factor(vs, labels = c("V", "S"))
   am <- factor(am, labels = c("automatic", "manual"))
   cyl  <- ordered(cyl)
   gear <- ordered(gear)
   carb <- ordered(carb)
})
summary(mtcars2)
```

अभ्यास : फाइल के उपयोग की सहायता से निम्न लिखित प्रश्नों के उत्तर दीजिए :
1. डाटा सेट में कितनी कार शामिल है ?

2. मॉडल किस वर्ण का है ?
3. mpg value क्या वर्णन करता है ?

सम्पूर्ण डाटा को देखकर, केवल लिखे

mtcars

अभ्यास : डाटासेट को जानने के लिए, निम्नलिखित गैर सांख्यिकीय प्रश्नों के उत्तर दीजिए :
1. क्या ऐसी कोई कार है जिसका वजन 5000 lb/1000 से ज्यादा है ?

2. Volvo 142 E की मोटर कार में कितने सिलेन्डर है ?
3. क्या 5 फारवर्ड (forward) गियर वाली कोई कार है ? क्या उनका ट्रॉंसमिसन स्वचालित या हस्त चालित है ?

विवरणात्मक सांख्यिकी :
आँकड़ों को माध्य, माध्यिका, मानक विचलन, अधिकतम और न्यूनतम जैसे सरल मापों का उपयोग करके सार प्रस्तुत कर सकते है। mtcars data set के लिए कुछ इसी तरह के मापों के सारांश निम्न प्रकार लिखकर प्राप्त किये जा सकते हैं।

- **summary(mtcars)**

माप का एक साथ भी अध्ययन किया जा सकता है

```
mean(mtcars$hp);median(mtcars$hp);qua
ntile(mtcars$wt);max(mtcars$mpg)
sd(mtcars$mpg) # standard deviation मानक विचलन

var(mtcars$mpg) # variance
```

विचलन / प्रसरण

```
sd (mtcars$mpg) ^ 2 # sd* sd=var?
```

जब सम्बन्धित कमाण्ड (Attach command) डाटा फ्रेम के साथ काम करती है तो बहुत उपयोगी होती है। ऊपर लिखे संदर्भ लिखने के बजाय attach (mtcars) लिखने से mtcars में संदर्भित चरों (variables) को छाँट सकते हैं।

```
mean(hp); median(hp); quantile(wt); max(mpg)
```

- **आयत चित्र (Histogram)**

सतत आँकड़ों के विवरण के अध्ययन के लिए आयत चित्र का प्रयोग कर सकते हैं। यदि वे स्पष्ट रूप में नहीं है जब ीपेज कमाण्ड का उपयोग करते हैं तो त क्लास की संख्या तथा क्लास की चौड़ाई चुनता है। नीचे दिये गये कमाण्ड, क्रम के अनुसार एक विण्डो में दो आयत चित्र बनाता है पहला आयत चित्र उचह के लिये और दूसरा आयत चित्र ूज के लिए।

```
par(mfrow = c(1, 2) ; hist(mtcars$mpg);
hist(mtcars$wt)
Par(mfrow = c(a,b)) प्रत्येक पंक्ति कें b भूखंडों के साथ ं पंक्ति देता है।
```

प्राचल बारम्बारताओं (parameters frequencies) का उपयोग करते हुए जब hist कमाण्ड का उपयोग करते हैं तो हम बारम्बारता की बजाय सम्बन्धित बारम्बारता वाला आयत चित्र बना सकते हैं। इस प्रकार के आयत चित्र आँकड़ों के घनत्व फलन (density function) के रूप में देखे जा सकते हैं हैल्प फांइल में hist(mtcars$mpg freq=FALSE) के बारे में पढे और कमाण्ड टाईप करके स्वय देखें।

- **बॉक्सप्लॉटस (Boxplots)**

बॉक्सप्लॉटस आँकड़ों के अध्ययन के लिए उपयोगी साधन है। यह माध्यिका, चतुर्थक (Quartiles) और
आउटलायरी को दर्शाता है । R कमाण्ड बॉक्सप्लॉटस है जिसे हम समरूप चरों (same variable) के आयत चित्र मे उपयोग कर सकते हैं ।

```
boxplot(mtcars$mpg); x11( ); boxplot(mtcars$wt)
```

x11 कमाण्ड एक नई विण्डो खोलता है जो इसमे इससे अगली आकृति बनायेगा ।

स्केटर प्लॉट्रस (Scatter plots)

स्केटर प्लॉट्रस चरों के बीच निर्भरताओं का अध्ययन करने के लिए उपयोगी है। लिखने का प्रयास करें :

plot(mtcars$wt,mtcars$mpg)

क्या गुच्छ ;बसनेजमतद्ध का स्लोप (जो रेंखिक स्वतन्त्रता को मापता है) उचित प्रतीत होता है ? कमाण्ड बवत (हैल्प फाइल का उपयोग करें) के प्रयोग द्वारा सहसंबंध की गणना कर सकते हैं इस मामले में क्या सहसंबंध है ? क्या यह इस स्लोप से सहमत है ?

```
cor(mtcars$wt,mtcars$mpg)
```

- **रैखीय समाश्रयण (Linear regression)**

```
lm(mtcars$wt~mtcars$mpg)
```

देखने मे मदद लेने की कोशिश करो (IM) R को कमाण्ड q( ) से बन्द किया जा सकता है। बन्द की कमाण्ड देने के बाद, **R** पूछता है कि वर्कस्पेस को सैव किया जाये या नहीं। आमतौर पर वर्क्रस्पेस को सैव करने का एक अच्छा विचार है। चूँकि यह एक विशेष फाइल बनाता है जिससे सीधे **R** में पढ़ा जा सकता है तथा उस ही डाटा सिट तथा पहले से प्राप्त परिणाम पर कार्य आरम्भ किये जा सकता हैं।

फाइल में सैव की गयी वर्कस्पेस **R data** कहते हैं। तथा उस ही सत्र के दौरान दी गयीं सभी कमाण्ड एक फाइल में सैव रहती हैं जिसे **R** हिस्ट्री कहते हैं। **R** में पुनः वर्कस्पेस लोड करने के लिये **R** कंजं फाइल को दो बार किलिक करे इससे **R** स्वतः खुल जाएगा जिससे डाटा एवं परिणाम लोड हुए है। यह नोट करे कि लाइब्रेरी स्वतः लोड नहीं होती, इन्हे कार्य शुरू करने से पहले इसे लोड करना चाहिये यदि आवश्यक है।

संदर्भ

R विकास कोर टीम (2010): सांख्यिकीय गणना के लिए एक भाषा एवं वातावरण, संख्यिकीय गणना के लिए **R** फाउण्डेशन, वियाना, आस्ट्रिया, URL: http://www.R-project.org

अध्याय 05

सांख्यिकीय विश्लेषण में SPSS सॉफ्टवेयर का उपयोग

राजू कुमार
वैज्ञानिक, प्रतिदर्श सर्वेक्षण प्रभाग
भा.कृ.अनु.प.-भारतीय कृषि साँख्यिकी अनुसंधान संस्थान, लाइब्रेरी एवेन्यू, नई दिल्ली-110012

1. प्रस्तावना (Introduction):

स्टेटिसिटिकल पैकेज फार सोशल साइंसेज (एस.पी.एस.एस.) समाजिक विज्ञान के ऑकड़ों के विश्लेषण के लिए सर्वाधिक उपयोग में लाए जाने वाले साफ्टवेयर्स में से एक है। यह पहले मूल रूप से समाजिक विज्ञान के सांख्यिकीय पैकेज का संक्षिप्त रूप था लेकिन अब यह सांख्यिकीय उत्पाद एवं सेवा प्रदान के लिए उपयोग में लाया जाता है । इसमें डाटा–एन्ट्री, डाटा प्रबन्धन, सांख्यिकीय विश्लेषण व प्रस्तुतीकरण के लिए साफ्टवेयर का एक सैट उपलब्ध है। एस.पी.एस.एस. से जटिल आंकड़े व फाईल प्रबन्धन तथा सांख्यिकीय विश्लेषण व रिपोर्टिंग फलनों ;निदबजपवदेद्ध को इंटीग्रेट किया जा सकता है । इसकी सहायता से लगभग सभी प्रकार की फाइलों से आंकड़ों को लेकर उनसे सारणीकृत रिपोर्ट, चार्ट्स, वितरण व प्रवृति ;ज्तमदकेद्ध के प्लाट, विस्तृत सांख्यकी व जटिल सांख्यिकीय विश्लेषण प्राप्त किये जा सकते हैं । मूल SPSS मैनुअल (Nieet et al, 1970) को ''समाजशास्त्र के सबसे प्रभावशाली पुस्तकों'' में से एक के रूप में वर्णित किया गया हैं जो किसी को भी अपने स्वयं के सांख्यिकीय विश्लेषण करने में सक्षम कराता है। मूल रूप से यह ''सामाजिक विज्ञान के लिए सांख्यिकीय पैकेज (SPSS)'' नाम से परिचित था, लेकिन यह अब ''सांख्यिकीय उत्पाद और सेवा समाधान (SPSS)'' नाम से परिचित है। SPSS Inc. द्वारा लम्बे समय से स्थापित SPSS सॉफ्टवेयर को सन 2009 में आईबीएम (IBM) ने अधिगृहित किया। सन 2009 से 2010 के दौरान यह PASW (Predictive Analysis Software) के नाम से बुलाया जाता था। सन 2010 का वर्तमान संस्करण आधिकारिक तौर पर''आईबीएम SPSS सांख्यिकी (IBM SPSS statistics)'' से नामित किया गया हैं।

यह सबसे लोकप्रिय सांख्यिकीय पैकेज होने के साथ अत्यधिक जटिल आंकड़ों का अत्यधिक साधारण तरीके से विश्लेषण कर सकता हैं। यह पैकेज व्यक्तिगत कंप्युटर के साथ मेनफ्रेम कंप्यूटर के लिए भी उपलब्ध है। SPSS पैकेज ऑकड़ा प्रविष्टि ऑकड़ा पैकेज प्रबंधन, सांख्यिकीय वि लेशण एवं प्रस्तुति के लिए साफ्टवेयर सामग्री का एक संग्रह हैं। SPSS लगभग किसी भी प्रकार के डाटा फाइल की उपयोग कर विभिन्न वितरण एवं प्रवृतियों, वर्णनात्मक ऑकड़े और जटिल सांख्यिकीय विश्लेषण को सारणीबद्ध करके रिपोर्ट, चार्ट, और प्लाट उत्पन्न कर सकता है।

हाल के वर्षों में जारी SPSS के संस्करणों निम्न प्रकार हैं

- SPSS सांख्यिकी 17.0.1 – दिसंबर 2008

- PASW सांख्यिकी 17.0.3 – सितम्बर 2009
- PASW सांख्यिकी 18.0, 18.0.1, 18.0.2, 18.0.3
- आईबीएम SPSS सांख्यिकी 19.0 – अगस्त 2010
- आईबीएम SPSS सांख्यिकी 19.0.1, 20.0, 20.0.1, 21.0, 23.0, 24.0, 25.0, 26.0, 27.0,

SPSS Inc. परिवार के अन्य साथी उत्पादों विभिन्न अन्य महत्वपूर्ण विश्लेषण जैसे नमूना सर्वेक्षण (आईबीएम SPSS डेटा संग्रह), आँकड़ा खनन (आईबीएम SPSS मॉडलर), टेक्स्ट एनालिटिक्स, और सहभागिता (बैच और स्वचालित स्कोरिंग सेवाएं) के लिए उपयोग किया जाता है। यह लेख का विषय SPSS की बुनियादी सुविधाओं को पेश करना एवं कुछ बुनियादी सर्वेक्षण आँकड़ों का सांख्यिकीय विश्लेषण करने में इस सॉफ्टवेयर के उपयोग पर है।

आईबीएम SPSS सांख्यिकी की प्रमुख विशेषता में से कुछ निम्न प्रकार हैं :

- यह सीखने में आसान है।
- यह साफ्टवेयर अपने कुछ प्रमुख सुविधायें जैसे पुल–डाउन मेनु की सहायता से बहुत उपयोगी साबित हुआ है।
- यह डेटा प्रबंधन की विभिन्न प्रणाली की एक पूरी श्रृंखला में शामिल होता है।
- यह गहरी सांख्यिकीय विश्लेषण में उपयोगी है।
- यह प्लाटिगं, रिपोर्टिग एवं प्रस्तुति सुविधाओं की व्यापक उपयोगिता प्रदान करता है।

सांख्यिकीय विश्लेषण के अलावा, डेटा प्रबंधन और डेटा प्रलेखन मूल सॉफ्टवेयर की अन्य विशेषताएं हैं। मूल सॉफ्टवेयर में शामिल कुछ महत्वपूर्ण सांख्यिकी विश्लेषण किस्म निम्न प्रकार हैं :

वर्णनात्मक संख्यिकी जैसे बारंबारता–बंटन, वर्णनात्मक आदि, द्विचर आँकड़े जैसे सहसंबंध एवं समाश्रयण, अप्राचलिक परीक्षण आदि, संख्यात्मक परिणामों के लिए भविश्यवाणी जैसे रेखीय प्रतिगमन, एकाधिक समाश्रयण आदि, पहचान समूहों के लिए भविश्यवाणी जैसे कारक विश्लेषण, क्लस्टर/ गुच्छविश्लेषण, विभेदक विश्लेषण आदि।

2. सर्वेक्षण आँकड़े विश्लेषण में SPSS का उपयोगिता (Use of SPSS in analysis of Survey Data)

SPSS सॉफ्टवेयर में उपलब्ध विश्लेषण मेनू के जटिल नमूने उप–मेनू के तहत जटिल डिजाइन से नमूने चयन के लिए प्रक्रियाऐं प्रदान करता हैं और सांख्यिकी विश्लेशण में डिजाइन विर्निदे ाों को भाामिल करके, सुनिि चत परिणाम निकालने में सहायता करता

हैं। परंपरागत सॉफ्टवेयर समूह सरल यादृच्छिक नमूने के द्वारा प्रतिचयन किया करता हैं। एक जटिल नमूना कई मायनों में एक सरल यादृच्छिक नमूना चयन से अलग हो सकते हैं। एक सरल यादृच्छिक नमूना में, नमूने इकाइयों को बराबर संभावना से पूरी आबादी से सीधे यादृच्छिक चयन किया जाता है जब एक जटिल नमूना चयन में निम्न सुविधाऐं भी हो सकते है।

स्तरीकरण (Stratification): स्तरित प्रतिचयन में, जनसंख्या के उपसमूहों से स्वतंत्र रूप से नमूनों का चयन किया जाता हैं, इन उपसमूहों को स्तर (stratum) कहा जाता हैं। उदाहरणतः, अर्थ–सामाजिक समूह, नौकरी श्रेणियाँ, आयु समूह, जातीय समूह, राज्यो में समस्त जिले आदि को स्तर के रूप में चयन किया जा सकता है।

समूहीकरण (Clustering): गुच्छ या समूह प्रतिचयन में समूहों का चयन किया जाता है। यहाँ समूह को नमूना इकाई के तौर पर लिया जाता है। उदाहरणतः, स्कूलों, अस्पतालों या भौगोलिक क्षेत्रों को समुह के तौर पर लिया जा सकता हैं।

बहु–चरण (Multi-Stage): बहुचरणी यादृच्छिक प्रतिचयन के पहले चरण में नमूने को समूह के आधार पर चयन करते है और दूसरे चरण में चुने हुई समूह से उप–नमूना का चयन किया जाता हैं। फिर चुने हुए उप–नमूने में से फिर नमूना का चयन किया जा सकता है। उदाहरणतः, पहले चरण में एक जिले से तहसीलों का चयन, दूसरे चरण में . प्रत्येक चयनित तहसील से गावों का चयन और फिर तीसरे चरण में प्रत्येक गाँव से घरों का चयन कर सकते हैं।

असमान–संभावना प्रतिचयन (Varying Probability Sampling): जब नमूना इकाई असमान आकार के हो तो इकाईयोंका चयन के लिए आकार के अनुपातिक संभावना प्रतिचयन (PPS) का उपयोग कर सकते हैं।

अप्रतिबंधित प्रतिचयन (Unrestricted Sampling): अप्रतिबंधित प्रतिचयन यादृच्छिक रूप से प्रतिस्थापन के बिना इकाईयों का चयन किया जाता है। इस लिए एक इकाइ एक से अधिक बार नमूना में चयनित हो सकती है।

नमूना भार (Sampling Weight): किसी जटिल नमूने के चयन करते समय नमूना भार कि आकलन स्वतः हो जाता है जो किसी नमूना इकाई का लक्षित जनसंख्या में आवृत्ति बताता है। जटिल नमूने वि लेशण प्रक्रिया के लिए नमूना भार अति आव यक हैं।

जटिल नमूने प्रक्रिया का उपयोग दो वि ोश कार्य सम्पादन के लिए किया जाता हैं–

क) किसी जटिल प्रतिचयन डिजाइन की योजना बनाना और सर्वेक्षण कराना जिसके लिए **नमूना विजार्ड** (Sampling Wizard) का उपयोग करते हैं।

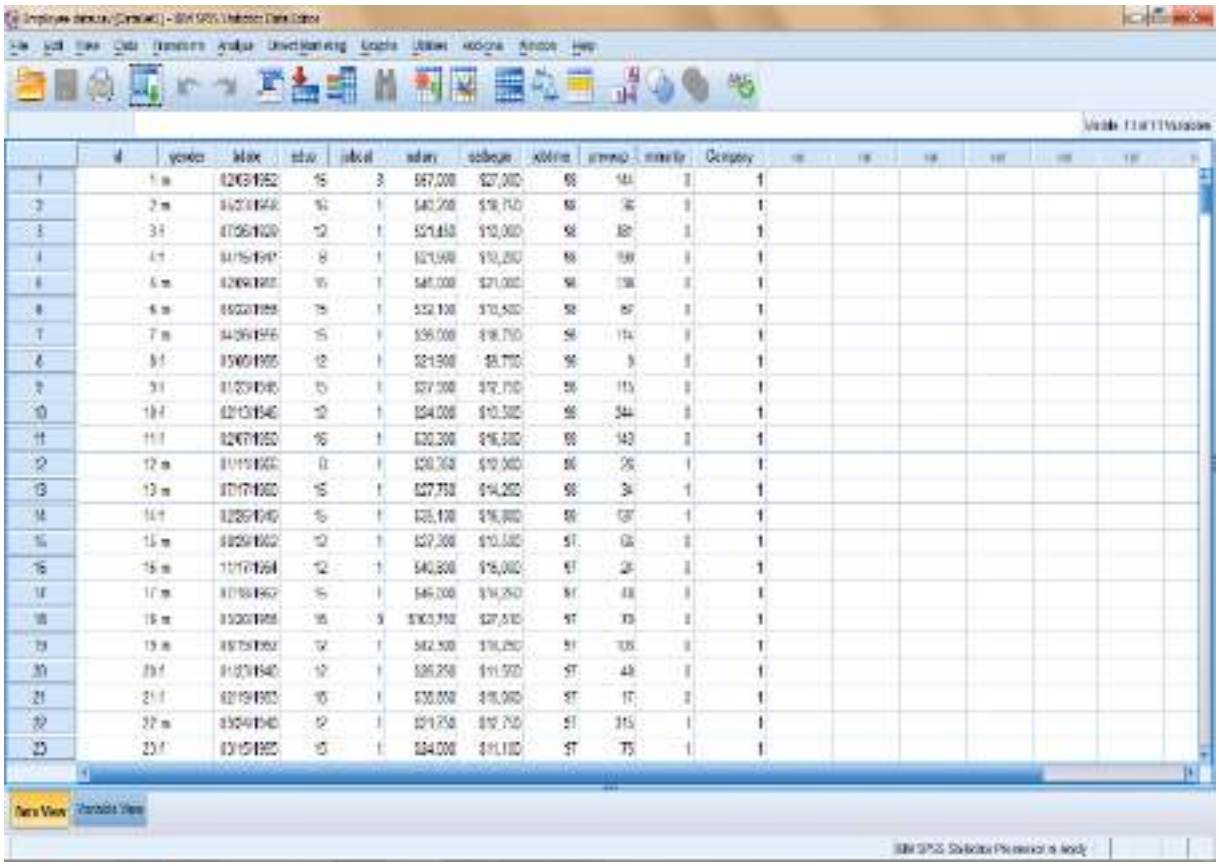
ख) जटिल प्रतिचयन डिजाइन द्वारा प्राप्त नमूना आकड़ों का सांख्यिकी वि लेशण करना जिसके लिए **विलेशण निर्माण विजार्ड** (Analysis Preparation Wizard) उपयोगी हैं।

जटिल प्रतिचयन डिजाइन के समस्त विर्निदे ाों के पुनः उपयोग के लिए इन्हें एक **योजना फाईल** में संग्रहित करना अति आव यक हैं। योजना फाईल दो प्रकार के होते हैं–

नमूना योजना (Sampling plan): जटिल नमुनों के चयन में उपयोगी विर्निदे ाों को नमुना योजना फाइल में भाामिल किया जाता हैं।

विश्लेषण योजना (Analysis plan): जटिल नमूनो के विचरण के सटीक आकलन हेतु उपयोगी जटिल नमूने विलेशण प्रक्रिया के विर्निदेाों को विलेशण योजना में भाामिल किया जाता हैं।

इस लेख में एक उदाहरण की सहायता से सर्वेक्षण आँकड़े विश्लेषण के लिए IBM SPSS की उपयोगीता को विस्त्रित रूप से वर्णित किया गया है। उदाहरण के लिए निजी कम्पयूटर के C ड्राइव में प्रोग्राम फाइल फोल्डर में स्थित IBM SPSS की नमूने फोल्डर में उपलब्ध "Employee Data" का उपयोग किया जा रहा हैं। इसके अलावा इस डाटासेट में ''कंपनी'' (Company) नाम से एक नया चर जोड़ा गया है जो कर्मचारियों के कम्पनी के कोड बताता है। अतः यह डाटासेट 400 कर्मचारीयों के विभिन्न कार्मिक विशेषताओं को दर्शाता है जिनको कंपनी चर के अनुसार 40 कर्मचारी के 10 समूह मे बाँटा जा सकता है। पुनः इस डाटासेट के "Employee Category" चर में तीन प्रकार है जिसे स्तर के रूप में चयन कर सकते हैं। ।



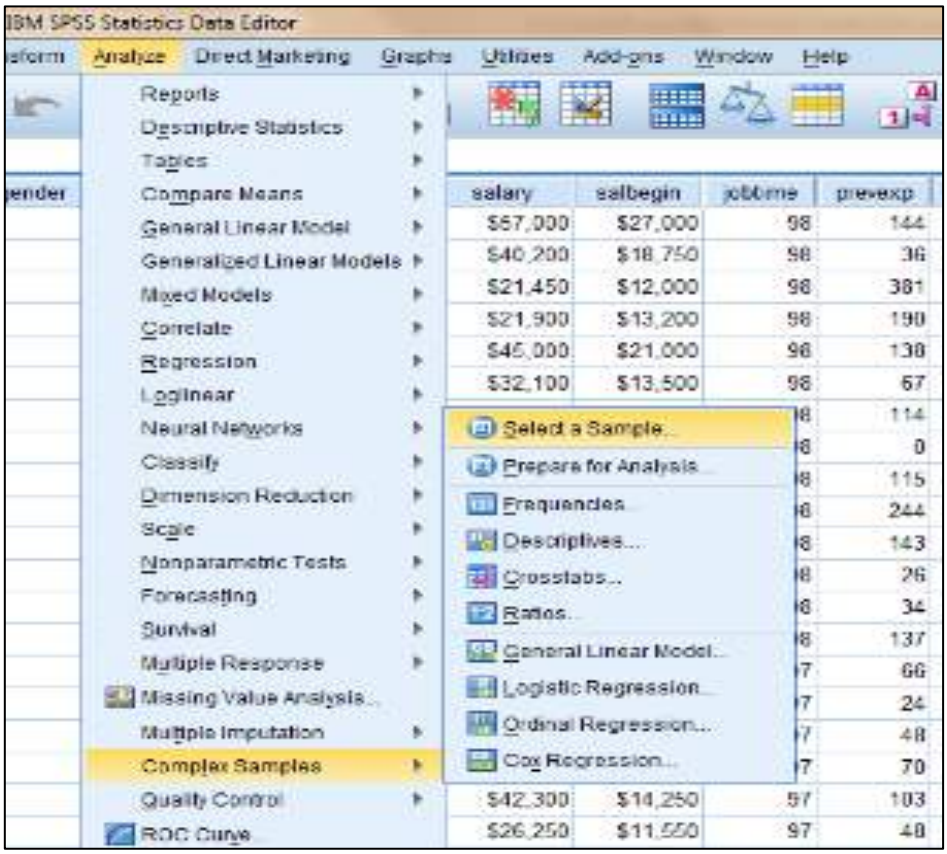
सर्वेक्षण आकड़े के सांख्यिकी विश्लेषण के लिए SPSS सॉफ्टवेयर में उपलब्ध नमूना विजार्ड बहुत कार्यकारी है। किसी भी नमूना प्रतिचयन के लिए नयी नमूना योजना फाइल को

बनाने में, संशोधित करने में या क्रियान्वित करने के प्रक्रिया में नमूना विज़ार्ड पथ प्रदर्शक के रूप में कार्यकारी है। विज़ार्ड का उपयोग करने से पहलेएक लक्ष्य जनसंख्या (target population),नमूने इकाइयों की एक सूची और एक उपयुक्त नमूना डिजाइन परिभाशित करना आवश्यक है। इस लेख में इस डाटासेट को विभिन्न तरह की नमूना चयन डिजाइन के उदाहरण के लिए लक्ष्य जनसंख्या की तौर में व्यवहार किया जाएगा। SPSS सॉफ्टवेयर में उपलब्ध जटिल नमूने विकल्प के द्वारा, एक जटिल नमूना चयन डिजाइन के अनुसार एक नमूना का चयन और डिजाइन विनिदेशों को शामिल करते हुऐ नमूना डेटा का सांख्यिकी विश्लेषण किया जाता हैं।

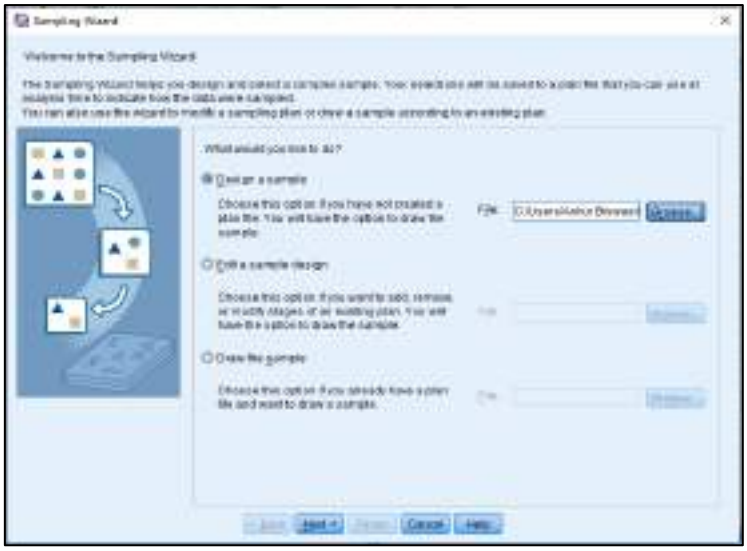
3. एक नया नमूना योजना बनाने कि पद्धति (Method to create a new sample plan) :

1. मेनू में जाकर विकल्प सूची से चुनें

विश्लेषण (Analyze)→ जटिल नमूने (Complex samples)→ एक नमूना का चयन (Select a sample) ....



2. नमूने के डिजाइन का चयन करते हुए नमूना योजना फाइल के नाम का चयन करें और Next क्लिक करें।



3. **''डिजाइन चर''** चरण में, स्तर (Strata), समूहों (Cluster), और इनपुट नमूना के डिजाइन weight को परिभाषित कर सकते हैं जो जटिल नमूना चयन पदधतिओं में बहुत कार्यकारी हैं।

हमारे वर्तमान उदाहरण में नमूनों को चयन के लिए विभिन्न नमूना चयन डिजाइन जैसे सरल यादृच्छिक प्रतिचयन (SRS), स्तरित यादृच्छिक प्रतिचयन, गुच्छ प्रतिचयन, बहुचरणी यादृच्छिक प्रतिचयन डिजाइन के द्वारा करना चाहते हैं। ये नमूना प्रतिचयन डिजाइन सर्वेक्षण में सबसे अधिक लोकप्रिय हैं।

सरल यादृच्छिक प्रतिचयन के लिए स्तर और समूहों के चयन की जरूरत नहीं है। स्तरित यादृच्छिक प्रतिचयन के लिए, ''Employee Category'' चर को स्तर के रूप में चयन कर सकते हैं। गुच्छ प्रतिचयन के लिए समूह के रूप में ''कंपनी'' चर का चयन करें और फिर Next क्लिक करें।

4. **नमूने विधि (Sampling method)** चरण में एक इकाई का प्रतिचयन करने के लिए कोई एक विधि का चयन कर सकते है। सरल यादृच्छिक प्रतिचयन

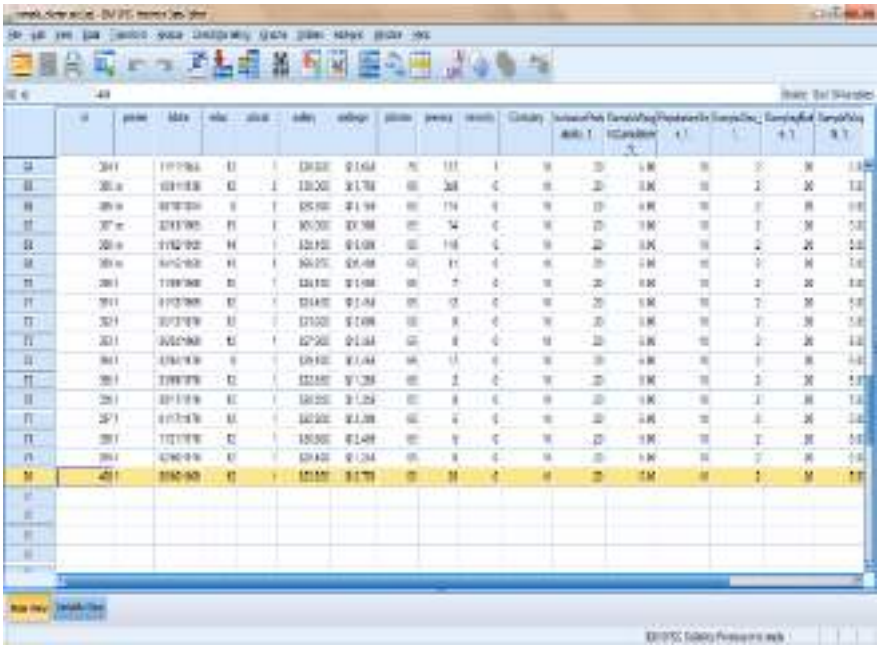
- सामान्य रूप से सरल यादृच्छिक प्रतिचयन का चयन करें।

- परिवर्ती प्रायिकता प्रतिचयन के लिए पीपीएस Brewer या पीपीएस मूर्ति का चयन करें।

Next क्लिक करें।

5. **नमूना आकार (Sample size)** चरण में, नमूने की संख्या या नमूने के लिए इकाइयों के अनुपात को निर्दिष्ट करें।

उदाहरण के तौर पर, सरल यादृच्छिक प्रतिचयन के लिए नमूने की संख्या 80 कर सकते हैं जिस से 80 इकाइयां चयनित होगा। स्तरित यादृच्छिक प्रतिचयन के लिए नमूने की संख्या 20 कर सकते हैं जिस से हर एक स्तर में 20 इकाइ चयनित होगा। गुच्छ प्रतिचयन के लिए नमूने की संख्या 2 कर सकते हैं जिससे 2 समूह हर एक में 40 इकाई चयनित होगा।

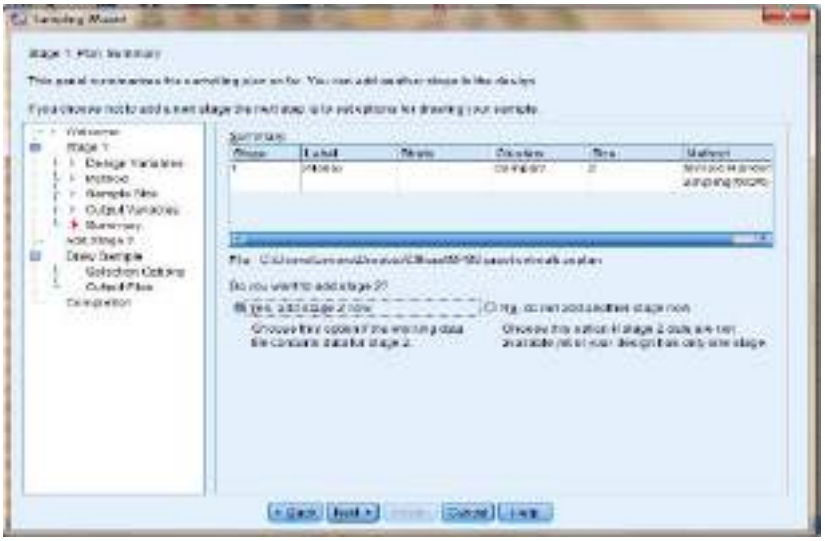


6. वैकल्पिक रूप से, अगले चरणों में कोई निम्नलिखित का प्रयोग कर सकते हैं :

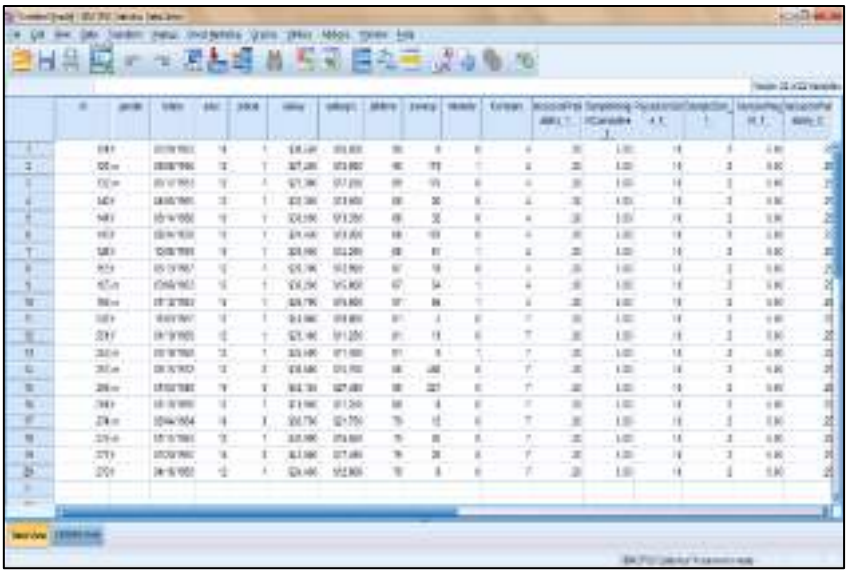
- सुरक्षित करने के लिए आउटपुट चर को चुनें।

- विभिन्न चयन विकल्पों को निर्दिष्ट करें जैसे कि Stage से नमूने का चयन करना है, यादृच्छिक संख्या बीज और यदि User missing मान डिजाइन चर के वैध्य मान के रूप में लिया गया है।
- आउटपुट डेटा को कहाँ सुरक्षित करना है इसका चयन करें।
- डिजाइन में दूसरा या तीसरा Stage जोड़ें।

  बहुचरणी यादृच्छिक प्रतिचयन डिजाइन के द्वारा नमूने के चयन के लिए किसी को ''क्या दूसरा चरण चुनना चाहते है ?'' प्रश्न के जवाब में ''हां दूसरा चरण चुनें'' को चुनना चाहिए।



अब स्टेज–2 के लिए ''**नमूना आकार**'' को परिभाषित करें तथा जहॉ फाइल को सूरक्षित करना है पथ प्रर्दर्शित करें। एक आऊट पुट फाइल को नीचे दिया गया है जिस में, पहले, दो समूहो को सरल यादृच्छिक प्रतिचयन के द्वारा तथा चयनित समूहो के अंदर 10 इकाईयों को सरल यादृच्छिक प्रतिचयन के द्वारा चुना गया है।



7. अब नमूने को चयनित करने के लिए Finish क्लिक करें।

विकसित नमूना योजना को आगे भी यादृच्छिक नमूना चयन के लिए इस्तेमाल किया जा सकता जो कि इस प्रकार है

विश्लेषण (Analyze)→ जटिल नमूने (Complex samples)→नमूना ड्रा (Draw sample)...

नमूना का चयन करने के बाद अगले चरण में चयनित नमूने को विश्लेषण के लिए तैयार करने कि आवश्यकता हैं। विभिन्न जटिल नमूने विश्लेषण (Complex sample analysis) प्रक्रिया के प्रयोग से विश्लेषण निर्माण विज़ार्ड विश्लेषण योजना को विकास तथा संशोधण प्रक्रिया का पथ प्रदर्शक है। विज़ार्ड का उपयोग करने से पहले, एक जटिल डिजाइन के अनुसार चयनित नमूना तैयार रखना चाहिए।

4. एक नई विश्लेषण योजना का निर्माण (Formulation of New Analysis Plan)

1. मेनू से चुनें :

   विश्लेषण (Analyze)→ जटिल नमूने (Complex samples)→ विश्लेषण के लिए तैयार करें (Prepare for Analysis)...

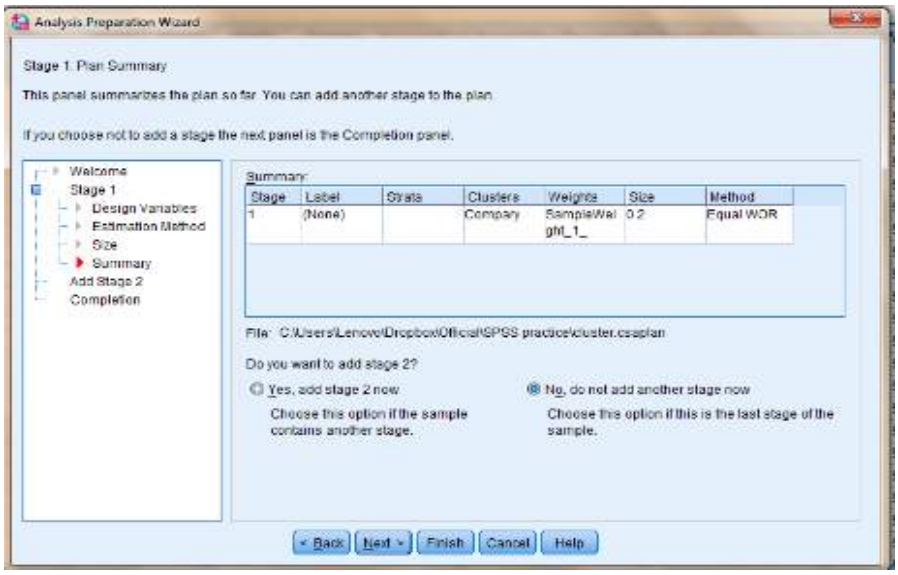
2. योजना फाइल बनायें (Create a plan file) चुनें। विश्लेषण योजना फाइल को सुरक्षित करने के लिए फाइल कानाम चुने और Next बटन को क्लिक करें।

3. **''डिजाइन चर''** चरण में, नमूना वजन युक्त चर निर्दिष्ट करें, वैकल्पिक तौर पर स्तर और समूहों को परिभाषित करें।

4. अगले चरण में निम्नलिखत क्रिया को वैकल्पिक तौर में व्यवहार किया जा सकता है।

- **आकलन विधि** चरण में मानक त्रुटियों का आकलन करने के लिए एक विधि का चयन करें।
- नमूना प्रतिदर्शित इकाईयों की संख्या को निर्दिष्ट करें या प्रति इकाई को सम्मिलत किये जाने की प्रायिकता का उल्लेख करें।
- दूसरे और तीसरे चरण को डिजाइन में शामिल करें।

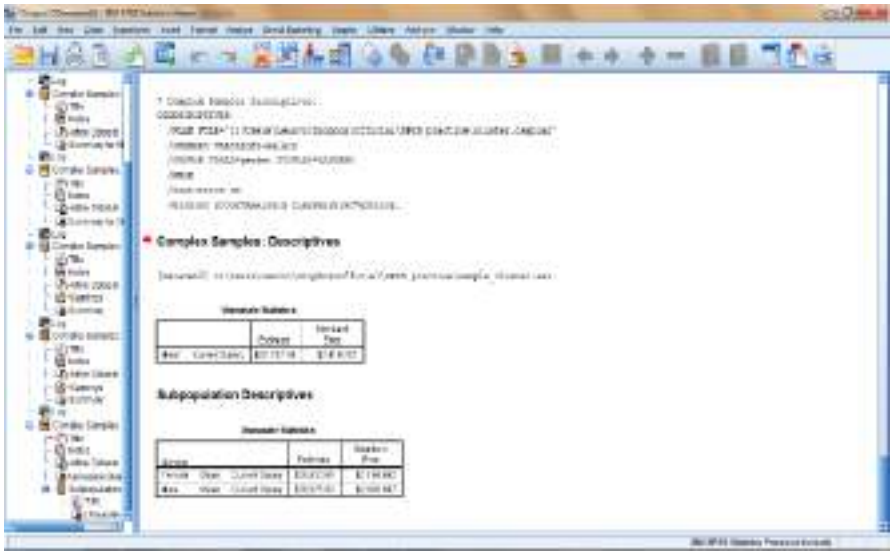
5. अब योजना को सुरक्षित करने के लिए Finish बटन को क्लिक करें।



अब विश्लेषण योजना फाइल को उपयोग कर हम बहुत से आऊटपुट को जटिल नमूने (Complex samples) विक्लप से प्राप्त कर सकते है।

- आवृत्तियो
- वर्णनात्मक
- क्रासटेबस
- अनुपात
- सामान्य रेखीय मॉडल
- वृद्धिघात समाश्रयण
- क्रमवार समाश्रयण
- कॉक्स समाश्रयण

वर्णनात्मक विकल्प से ''वर्तमान वेतन'' का निम्न वर्णनात्मक परिणाम प्राप्त किया गया



संदर्भ (References)

- मॉर्गन, जी.ए., बैरेट, के.सी., लीच,एन.एल. और ग्लॉकनर, जी. (2019) इंट्रोडक्टरी सांख्यिकी के लिए IBM SPSS : प्रयोग और व्याख्या। छठा संस्करण, रूटलेज।
- नील, एन.एच., बेंट, डी.एच. और हल, सी.एच. (1970)। SPSS : सामाजिक विज्ञान के लिए सांख्यिकीय पैकेज। न्यूयॉर्क : मैकग्रा–हिल।

अध्याय 06

प्रतिदर्श सर्वेक्षणों में प्रारंभिक अवधारणाएँ

अंकुर विश्वास
वरिष्ठ वैज्ञानिक, प्रतिदर्श सर्वेक्षण प्रभाग
भा.कृ.अनु.प.-भारतीय कृषि साँख्यिकी अनुसंधान संस्थान, लाइब्रेरी एवेन्यू, नई दिल्ली-110012

1. प्रस्तावना (Introduction):

सांख्यिकीय सर्वेक्षण का उद्देश्य, समष्टि के बारे में सूचना प्राप्त करना है। 'समष्टि' से तात्पर्य है – सर्वेक्षण के उद्देश्यों के अनुसार परिभाषित इकाइयों का एक समूह। अतः समष्टि के अंतर्गत, क्षेत्र और उपज संबंधी सर्वेक्षणों में विनिर्धारित फसल के अंतर्गत सभी खेत, या कृषि सर्वेक्षणों में विनिर्धारित आकार से अधिक के सभी कृषि जोत–क्षेत्र आते हैं। निश्चित रूप से समष्टि अर्थात् जनसंख्या का तात्पर्य व्यक्तियों की गणना से भी होता है, चाहे वह किसी देश की पूरी जनसंख्या – हो या उसके किसी विशेष क्षेत्र की जनसंख्या । जनसंख्या सम्बन्धित सूचना ज्यादातर इकाइयों की कुल संख्या, विभिन्न विशेषताओं के समग्र मूल्य, विशेषताओं के औसत (प्रति इकाई) विनिर्धारित गुणों वाली इकाईयों के अनुपात आदि के बारे में होती है। ऑकड़ों का एकत्रीकरण मूलतः दो पद्धतियों द्वारा किया जाता है – (i) सम्पूर्ण परिगणना, जिसमें समष्टि की प्रत्येक इकाई के लिए सर्वेक्षण संबंधी मदों पर ऑकड़ों का एकत्रीकरण किया जाता है। इस प्रक्रिया का प्रयोग जनगणना, कृषि, पशुधन, खुदरा भंडारों, औद्योगिक प्रतिष्ठानों की गणना आदि के लिए किया जाता है। (ii) प्रतिचयन पद्धति, जिसमें जनसंख्या की चयनित इकाइयों पर ही सर्वेक्षण मदों के लिए ऑकड़ों का एकत्रीकरण किया जाता है। प्रतिचयन पद्धति, जनसंख्या से इकाइयॉ चुनने की एक वैज्ञानिक और विषय–पूरक प्रक्रिया है और ऐसा प्रतिदर्श उपलब्ध कराती है जो पूरी जनसंख्या का निरूपण करने में सक्षम होता है। सम्पूर्ण परिगणना पद्धति, प्रतिचयन पद्धति का एक विशेष रूप ही है, अतः प्रतिचयन पद्धति का महत्व अधिक बढ़ जाता है। प्रतिचयन पद्धति उन परिणामों के आकलन की प्रक्रिया भी उपलब्ध कराती है, जो समष्टि में सभी इकाइयों पर तुलनात्मक सर्वेक्षण किए जाने पर प्राप्त हो सकते हैं। दूसरे शब्दों में, प्रतिचयन पद्धति सर्वेक्षण प्राचलनों के आकार को कम करने के साथ–साथ समष्टि के कुल योगों, औसतों, अनुपातों का आकलन भी उपलब्ध कराती है।

2. प्रतिचयन त्रुटियॉ (Sampling Errors)

उत्कृष्ट प्रतिदर्श सर्वेक्षणों के लिए प्रतिचयन पद्धतियों के प्रयोग पर आधारित सर्वेक्षणों की प्रमुख विशेषताएँ हैं। इस विशेषता के द्वारा प्रतिदर्श आकलनों एवं समष्टि मूल्यों के बीच विसंगति (discrepancy) का पता चलता है जो प्रतिदर्श परिगणना के तरीके से ही समष्टि की सभी इकाइयों की परिगणना करने पर प्राप्त होती है। यह विसंगतियॉ अपरिहार्य (unavoidable) हैं क्योंकि प्रतिदर्श आकलक, केवल प्रतिदर्श इकाइयों के ऑकड़ों पर ही आधारित होते हैं। तथापि, प्रतिचयन पद्धतियों के द्वारा इन विसंगतियों के औसत परिमाण का आकलन किया जा सकता है। प्रतिचयन पद्धतियॉ, पहले से ही सर्वेक्षण डिजाइन के ब्योरे सम्बन्धित सूचना, जैसे प्रतिदर्श का आकार आदि इस ढंग से उपलब्ध कराती हैं कि प्रतिचयन त्रुटियों की औसत मात्रा, आबंटन–पूर्व संभाव्यता की अनुमानित मात्रा से अधिक न हो। दूसरे शब्दों में, प्रतिचयन पद्धतियॉ पूर्व निर्धारित सीमा में, प्रतिदर्श आकलनों की परिशुद्धता (precision) को नियंत्रित करने में सक्षम बनाती हैं।

प्रतिचयन पद्धतियॉ जोखिम के सिद्धांत (laws of chance) एवं प्रायिकता के सिद्धांत पर आधारित हैं। प्रतिचयन की अन्य पद्धतियॅा उद्देश्यात्मक प्रतिचयन (purposive sampling), अपना निर्णयात्मक प्रतिचयन (judgement sampling) के रूप में उपलब्ध हैं। इन पद्धतियों में, प्रतिदर्श में इकाइयों का चयन, समष्टि में इन इकाइयों की विशिष्टता के आधार पर, विषय–वस्तु में विशेषज्ञों के निर्णय के अनुसार किया जाता है। इस प्रकार की प्रतिचयन पद्धतियों के प्रयोग से किये गए प्रतिदर्श, विषय–वस्तु में विशेषज्ञों के व्यक्तिगत निर्णय से प्रभावित होते हैं। यह प्रक्रिया वस्तुपूरक (objective) नहीं है और न ही यह प्रायिकता के सिद्धांतों पर आधारित है। परिणामस्वरूप यह प्रक्रियाएं, प्रतिचयन त्रुटियों के परिमाण का आकलन करने तथा उन्हें नियंत्रित करने की संभावना भी उपलब्ध नहीं करती हैं। प्रस्तुत लेख केवल प्रायिकता प्रतिचयन (probability sampling) से सम्बन्धित है। प्रायिकता प्रतिदर्श प्राप्त करने का एक सरल तरीका है – समष्टि की प्रत्येक इकाई को प्रथम एवं तदंतर (subsequent) प्रत्येक ड्रा (draw) पर चयन की ज्ञात प्रायिकता देते हुए एक के बाद एक इकाई लेते हुए प्रतिदर्श प्राप्त करना।

तदंनतर ड्रा, पूर्व प्राप्त इकाइयों को प्रस्थापित करके या प्रतिस्थापित किए बिना लिए जा सकते हैं। इकाइयों को प्रतिस्थापित करके प्राप्त होने वाले प्रतिदर्श को प्रस्थापन सहित प्रतिचयन (sampling with replacement) पद्धति कहते हैं तथा दूसरी प्रायिकता प्रतिचयन को पद्धति लागू करने में कल्पना की जाती है कि समष्टि को बिना प्रस्थापन के प्रतिचयन (sampling without replacement) पद्धति कहते हैं। सुस्पष्ट (distant)

एवं अभिज्ञेय (identifiable) सीमित इकाइयों में उप–विभाजित किया जा सकता है–जिन्हें प्रतिचयन इकाइयॉ कहा जाता है। प्रतिचयन प्रक्रिया के लिए यह असंगत है कि प्रतिचयन इकाइयॉ कौन सी हैं। वे स्वाभाविक (natural) इकाइयॉ मानव जनसंख्या में व्यक्ति या फसल सर्वेक्षण में खेत या ऐसी इकाइयों का स्वाभाविक समूह जैसे परिवार या गांव या अस्वाभाविक (artificial) इकाइयॉ जैसे एकल पौधा, पौधों की एक पंक्ति, विनिर्धारित आकार का एक प्लाट हो सकती हैं। प्रतिचयन के लिए समष्टि की सभी प्रतिचयन इकाइयों की सूची तैयार करना अत्यावश्यक है। ऐसी सूची को ढांचा (frame) कहा जाता है और यह प्रतिदर्श इकाइयों का चयन एवं पहचान करने के लिए एक आधार प्रदान करता है। उदाहरणार्थ कृषि के खेतों की सूची, गांव में परिवारों की सूची, किसी जिले में गांवों की सूची, किसी राज्य में जिलों की सूची, भारत में राज्यों की सूची आदि हैं। गांव, प्रतिचयन इकाई माना जाए तो उसके परिवार, खेत एवं प्लाट तदंनतर चयन करने के लिए इकाइयॉ उपलब्ध कराते हैं। ढांचे में ज्यादातर समष्टि के आकार और संरचना के बारे में सूचना होती है। प्रतिदर्श सर्वेक्षणों में इन सूचनाओं का प्रयोग विभिन्न तरीकों से किया जाता है, जो निम्न प्रकार है :

3. सांख्यिकीय ऑकड़ों की आवश्यकता (Need of Statistical Data):

मानव क्रियाकलाप के लगभग प्रत्येक क्षेत्र में सूचना एकत्र करने की आवश्यकता होती है। सामान्य बातचीत वाले एवं विवादास्पद प्रश्नों को हल करने के लिए संख्यात्मक ऑकडों की आवश्यकता होती है। विभिन्न सिद्धान्तों (theory) एवं प्राक्कल्पनाओं (hypothesis) की जॉच करने के लिए भौतिक, रासायनिक और जैविक परीक्षणों से प्राप्त ऑकड़ों का प्रयोग किया जाता है। सुसंगत ऑकड़ों के प्रयोग एवं विश्लेषण के द्वारा विभिन्न सामाजिक एवं आर्थिक अन्वेषण किए जाते हैं। एकत्रित किए गए वस्तुपूरक विश्लेषित एवं उचित रूप से प्रस्तुत आंकड़े दैनिक जीवन के विभिन्न क्षेत्रों में नीतिगत निर्णय लेने के लिए प्रयोग में लाए जाते हैं।

सांख्यिकीय ऑकड़ों के महत्वपूर्ण प्रयोगकर्ताओं में सरकार, उद्योग, व्यवसाय, अनुसंधान संस्थान, सरकारी संगठन और अंतर्राष्ट्रीय संगठन शामिल हैं। अपनी विभिन्न जिम्मेदारियों का निर्वाहन करने के लिए सरकार को, अर्थव्यवस्था, व्यापार, औद्योगिक उत्पादन, स्वास्थ्य एवं मृत्यु–दर, जनसंख्या, पशुधन, कृषि, वानिकी, पर्यावरण, मौसम विज्ञान तथा उपलब्ध संसाधनों के विभिन्न सेक्टरों से सम्बन्धित अनेक प्रकार की सूचनाओं की आवश्यकता पड़ती है। ऑकड़ों से निकाले गए निष्कर्ष, राष्ट्र की भावी आवश्यकताऍ निर्धारित करने और लोगों की सामाजिक और आर्थिक समस्याऍ हल करने में भी सहायक होते हैं। उदाहरण के लिए, वेतन और मूल्य स्तरों से सम्बन्धित नीतियॉ बनाने में देश के विभिन्न भागों में रह रहे लोगों की विभिन्न श्रेणियों के

निर्वाह–व्यय (cost of living) सम्बन्धित सूचना का विशेष महत्व है। जनसंख्या वृद्धि पर रोक लगाने के लिए नीतियॉ बनाने हेतु स्वास्थ्य, मृत्यु–दर और जनसंख्या सम्बन्धित ऑकड़ों का प्रयोग किया जा सकता है। इसी प्रकार स्वच्छ एवं स्वस्थ जीवन के लिए नीति योजना बनाने हेतु वानिकी और पर्यावरण सम्बन्धित सूचना की आवश्यकता पड़ती है। राष्ट्र की पोषण सम्बन्धित योजना बनाने हेतु, कृषि उत्पादन सम्बन्धित ऑकड़ों का अत्यधिक प्रयोग किया जाता है। उद्योग और व्यवसाय में उत्पादन स्तरों और बिक्री अभियानों की उचित योजना बनाने के लिए श्रम, उत्पादन की लागत और गुणवत्ता पर सूचना एकत्रित की जाती है।

पिछले निष्कर्षों का सत्यापन करने तथा नए अनुमान लगाने के लिए अनुसंधान संस्थानों को ऑकड़ों की आवश्यकता होती है। सरकारी नीतियों का आकलन करने के लिए सरकारी संगठनों द्वारा ऑकड़ों का प्रयोग किया जाता है और यदि ये लोगों की प्रत्याशाओं पर खरे नहीं उतरते तो प्रशासन को बताया जाता है। अंतर्राष्ट्रीय संगठन विभिन्न देशों की तुलनात्मक स्थितियॉ प्रस्तुत करने के लिए अर्थव्यवस्था, शिक्षा, स्वास्थ्य, संस्कृति आदि से सम्बन्धित ऑकड़े एकत्र करते हैं। इसके अतिरिक्त जन कल्याण के लिए अंतरराष्ट्रीय स्तर पर अपनी नीतियॉ बनाने हेतु भी इसका प्रयोग करते हैं।

4. ऑकड़ों के प्रकार (Type of Data):

अपेक्षित सूचना का एकत्रीकरण, एक ओर अध्ययन के स्वरूप, उद्देश्य एवं विषय–क्षेत्र पर निर्भर करता है और दूसरी ओर वित्तीय संसाधनों, समय और जन शक्ति की उपलब्धता पर निर्भर है। सांख्यिकीय ऑकड़े दो प्रकार के होते हैं: (1) प्राथमिक (primary) ऑकड़े, एवं (2) गौण (secondary) ऑकड़े।

परिभाषा 4.1: अनुसंधानकर्ता द्वारा मूल स्रोत से ऑकड़ों के एकत्रीकरण को प्राथमिक ऑकड़े कहा जाता है।

परिभाषा 4.2: यदि अपेक्षित ऑकड़ों का एकत्रीकरण किन्हीं संस्थाओं या व्यक्तियों द्वारा पहले से ही कर लिया गया है एवं वह ऑकड़े प्रकाशित या अप्रकाशित रिकार्ड के रूप में उपलब्ध हैं तो इन्हें गौण ऑकड़ों के रूप में जाना जाता है।

इस प्रकार जब प्राथमिक ऑकड़ों का प्रयोग किसी अन्य अनुसंधानकर्ता/संस्था द्वारा किया जाता है तो वे गौण ऑकड़े हो जाते हैं। बहुत से प्रकाशनों में गौण ऑकड़े प्रकाशित किये जाते हैं जिनमें से कुछ महत्वपूर्ण प्रकाशन निम्न प्रकार हैं।

- संघीय, राज्य सरकारों एवं स्थानीय सरकारों के अधिकारिक प्रकाशन,
- समितियों एवं आयोगों की रिपोर्ट,

- व्यवसायिक संगठनों, व्यापारिक संघों एवं वाणिज्य मंडलों की रिपोर्ट एवं प्रकाशन,
- पत्रिकाओं, जर्नलों एवं समाचारपत्रों द्वारा प्रकाशित किए गए ऑकड़े,
- संयुक्त राष्ट्र संगठन, विश्व बैंक, अंतरराष्ट्रीय वित्तीय निधि, संयुक्त राष्ट्र व्यापार एवं विकास सम्मेलन, अंतरराष्ट्रीय श्रम संगठन, खाद्य एवं कृषि संगठन, इत्यादि विभिन्न अंतरराष्ट्रीय संगठनों के प्रकाशन।

गौण ऑकडों का प्रयोग करते समय सावधानी अत्यावश्यक है क्योंकि उनमें प्राथमिक स्रोत से प्रतिलेखन करने में त्रुटियॉ हो सकती हैं।

5. कुछ तकनीकी शब्दावली (Some Technical Terms):

परिभाषा 5.1: घटक (element) एक ऐसी इकाई है, जिसके लिए सूचना मांगी जाती है।

परिभाषा 5.2: समष्टि (population), घटकों का एक ऐसा समूह है, जिसके सम्बन्ध में अनुमान लगाया जाना होता है। समष्टि का सीमित या असीमित होना इसमें शामिल इकाइयों की संख्या पर निर्भर करता है।

परिभाषा 5.3: प्रतिचयन इकाइयॉ, समष्टि के घटकों की व्यापन बिना (non-overlapping) संग्रहण हैं।

परिभाषा 5.4: समष्टि की सभी इकाइयों की सूची जिसमें से प्रतिचयन करना है, ढांचा या प्रतिचयन ढांचा कहलाता है।

परिभाषा 5.5: समष्टि के घटकों का विश्लेषण करने के लिए समष्टि के ढांचे में से उसका एक 'सब–सेट' चुना जाता है जिसे प्रतिदर्श कहते हैं।

व्यवहारिक तौर पर, प्रतिदर्श के लिए चुनी गई इकाइयों की संख्या समष्टि में दर्शाई गई संख्या से बहुत कम होती है। प्रतिदर्श के लिए चुनी गई इकाईयों के अध्ययन चर पर की गई टिप्पणियों के आधार पर पूरी समष्टि का अनुमान लगाया जाता है।

6. प्रतिदर्श का प्रयोजन (Purpose of Sampling):

प्रतिदर्श का प्रयोजन समष्टि की समस्त इकाइयों पर किसी चर विशेष (characteristics of interest) पर सूचना के एकत्रीकरण को संपूर्ण परिगणना या जनगणना के रूप में जाना जाता है। जनगणना करने के लिए अधिकतर ज्यादा अपेक्षित धनराशि और समय की आवश्यकता होती है और ऐसी अनेक परिस्थितियॉ होती हैं जहॉ सीमित संसाधनों के साथ संपूर्ण परिगणना संभव नहीं होती। बहुत से ऐसे

भी उदाहरण हैं जहॉ समष्टि की सभी इकाइयॉ अप्राप्य होती हैं जिस कारण सभी इकाइयों की परिगणना करना व्यवहार्य नहीं हो पाता ऐसे सभी मामलों में अनुसंधानकर्ता के पास प्रतिदर्श सर्वेक्षण के अलावा अन्य कोई विकल्प नहीं रह जाता है ।

प्रतिदर्श के लिए चुनी गई इकाइयों की संख्या (आवश्यक नहीं है कि वे भिन्न हों) को प्रतिदर्श आकार के रूप में जाना जाता है और इस n द्वारा निर्दिष्ट किया जाता है, जबकि समष्टि की सभी इकाइयों की संख्या को समष्टि आकार कहा जाता है और 'N' द्वारा निर्दिष्ट करते हैं। प्रतिदर्श आकार एवं समष्टि आकार के अनुपात अर्थात n/N को प्रतिचयनानुपात (sampling fraction) कहा जाता है। सम्पूर्ण परिगणना के मुकाबले प्रतिदर्श सर्वेक्षण के कुछ प्रयोगात्मक लाभ निम्नलिखित हैं:

6.1 अपेक्षाकृत अधिक गति (greater speed)

प्रतिदर्श के लिए ऑकड़ों के एकत्रीकरण और विश्लेषण में लगा समय, संपूर्ण परिगणना में लगने वाले समय की तुलना में काफी कम होता है। व्यवहार्यतया हमारे सामने ऐसी स्थितियॉ आती हैं, जहॉ सूचना का एकत्रीकरण एक विनिर्धारित समय में ही करना होता है। ऐसे मामलों में जहॉ उपलब्ध समय कम होता है या समष्टि अत्यधिक बड़ी होती है, वहॉ प्रतिचयन ही केवल एक विकल्प रह जाता है।

6.2 अपेक्षाकृत अधिक परिशुद्धता (more accuracy)

ज्यादातर संपूर्ण परिगणना में कोई बड़ा एवं अव्यहवस्तिक (unwidely) संगठन शामिल होता है अतः अनेक प्रकार की त्रुटियॉ (error) हो सकती हैं। कभी–कभी इन त्रुटियों को पर्याप्त ढंग से नियंत्रित करना भी संभव नहीं हो पाता है । प्रतिदर्श सर्वेक्षण में, कार्य की मात्रा काफी कम हो जाती है। अतः प्रशिक्षित और दक्ष स्टाफ की सेवाऍ बिना किसी विशेष कठिनाई के उपलब्ध हो जाती हैं। जोकि संपूर्ण परिगणना की तुलना में अधिक परिशुद्ध परिणाम उपलब्ध करने में सहायक सिद्ध होते हैं।

6.3 अधिक विस्तृत सूचना (detailed information)

चूंकि किसी प्रतिदर्श में इकाइयों की संख्या, संपूर्ण परिगणना की इकाइयों की संख्या की तुलना में काफी कम होती है, इसलिए इसमें अधिक चरों पर सूचना प्राप्त की जा सकती है। तथापि, संपूर्ण परिगणना में इस प्रकार का प्रयास अपेक्षाकृत अधिक कठिन हो जाता है।

6.4 कम लागत (reduced cost)

समष्टि की तुलना में प्रतिदर्श इकाइयों की संख्या अपेक्षाकृत कम होने के कारण, समष्टि की तुलना में प्रतिदर्श इकाइयों के प्रेक्षण में समय, धन और ऊर्जा की ज्यादा बचत हो जाती है।

निष्कर्षतया कह सकते हैं कि प्रतिदर्श सर्वेक्षण अधिक किफायती है तथा अधिक परिशुद्ध सूचना उपलब्ध कराता है और संपूर्ण परिगणना की तुलना में प्रतिदर्श सर्वेक्षण विषय विशेष पर अधिक सूचना प्राप्त कराने में सक्षम है। तथापि, यहां ध्यान देने योग्य बात यह भी है कि प्रतिदर्श सर्वेक्षणों के परिणामों में प्रतिचयन त्रुटियॉ मौजूद होती हैं क्योंकि सम्पूर्ण परिगणना के केवल एक भाग का सर्वेक्षण किया जाता है। दूसरी ओर प्रतिदर्श सर्वेक्षण की तुलना में, सम्पूर्ण परिगणना अध्ययन में अधिक गैर प्रतिचयन त्रुटियॉ (non- sampling errors) होने की संभावना होती है।

7. प्रतिचयन प्रकियाएँ (sampling techniques)

परिभाषा 7.1: यदि प्रतिदर्श इकाइयों का चयन, किसी प्रायिकता प्रक्रियातंत्र का प्रयोग करके, किया जाता है तो इस पकार की प्रक्रिया को प्रायिकता प्रतिचयन (probability sampling) कहा जाता है।

इस प्रकार के सर्वेक्षण में, समष्टि की प्रत्येक इकाई को प्रतिदर्श में चयन किए जाने की एक निश्चित प्रायिकता प्रदान की जाती है। विकल्पतः यह हमें भिन्न–भिन्न प्रतिदर्शों के सेट परिभाषित करने में समर्थ बनाता है, जिसका चयन करने में प्रक्रिया सक्षम हो यदि इसे विशिष्ट समष्टि पर लागू किया जाए। प्रतिचयन प्रक्रिया, प्रत्येक संभावित प्रतिदर्श को चयन किए जाने की एक ज्ञात प्रायिकता उपलब्ध करती है। कोई भी व्यक्ति, प्रायिकता प्रतिदर्श के लिए समष्टि के विभिन्न चरों के लिए उपयुक्त आकलक बना सकता है। इस प्रकार की किसी प्रतिचयन प्रक्रिया के लिए कोइ भी व्यक्ति प्रायिकता उपकरणों का प्रयोग करके सिद्धांत बनाने में सक्षम हो पाता है। यह भी संभव है कि एक विशेष समष्टि पर बार–बार इसी प्रक्रिया का प्रयोग करके आकलक मूल्यों का बारंबारता वितरण (frequency distribution) भी प्राप्त किया जा सकता है। इस प्रकार की प्रक्रियाओं के लिए प्रतिचयन प्रसारण के माप तथा वास्तविक मूल्यों (true value) के निश्चित अंतराल में आकलनों का अनुपात प्राप्त किया जा सकता है। इस व्याख्यान में केवल इस प्रकार की प्रक्रियाओं पर ही विचार किया जाएगा।

परिभाषा 7.2: कोई प्रायिकता प्रक्रियातंत्र इस्तेमाल किए बिना किसी प्रतिदर्श का चयन करने की प्रक्रिया को गैर–प्रायिकता प्रतिचयन (non-probability sampling) कहा जाता है।

जब समष्टि के प्रतिनिधिनात्मक सब–सेट बनाने हेतु विशेष कौशल की आवश्यकता होती है तब उद्देश्यात्मक प्रतिचयन (जिसे निर्णयात्मक प्रतिचयन भी कहा जाता है) विशेषतया उपयोगी सिद्ध होता है। उदाहरण के लिए, लेखा–परीक्षक, यह पता लगाने के लिए कि क्या मदों की संपूर्ण लेखा–परीक्षा आवश्यक है, अध्ययन के लिए मदों का चयन करने हेतु निर्णयात्मक प्रतिचयन का प्रयोग करते हैं। कभी–कभी किए जा रहे अध्ययन से सुसंगत बातों के आधार पर समष्टि की विभिन्न श्रेणियों के लिए कोटा निश्चित किए जाते हैं और श्रेणियों के अंदर किए जाने वाले चयन व्यक्तिगत निर्णय पर आधारित होते हैं। इस प्रकार की प्रतिचयन प्रक्रिया को कोटा प्रतिचयन भी कहा जाता है।

स्पष्टतः इन पद्धतियों के प्रयोग में मानवीय अभिनति (bias) भी होती है। उपयुक्त स्थितियों में, ये पद्धतियॉ लाभप्रद परिणाम उपलब्ध करा सकती हैं। तथापि, वे प्रासंगिक सिद्धांत और सांख्यिकीय विश्लेषण के विकास हेतु उत्तरदायी नहीं हैं। ऐसी पद्धतियों में, प्रतिचयन त्रुटियों को निश्चित नहीं किया जा सकता, इसलिए इनकी तुलना उपलब्ध प्रायिकता प्रतिचयन नमूनाकरण पद्धतियों से नहीं की जा सकती है।

परिभाषा 7.3: प्रतिस्थापन प्रतिचयन में क्रमिक ड्रा निष्पादित करने से पहले, किसी विशेष ड्रा में चुनी गई इकाईयों को प्रतिस्थापित करते हुए समष्टि से एक–एक करके इकाइयों का चयन किया जाता है।

चूंकि प्रत्येक ड्रा में समष्टि का संघटन (constitution) समान रहता है, इसलिए प्रतिस्थापन प्रतिचयन कुछ इकाइयां एक से अधिक बार चुनी जा सकती हैं। यह प्रक्रिया N^n संभावित प्रतिदर्श उपलब्ध कराती है यदि प्रतिदर्श में इकाइयों के चयन के क्रम को ध्यान में रखा जाए, जहां N और n क्रमशः समष्टि एवं प्रतिदर्श के आकार को निर्दिष्ट करता है।

उदाहरण 7.1: एक प्रशिक्षण कार्यक्रम के 4 प्रतिभागियों के वजन (किलोग्राम में) निम्न प्रकार से हैं:

प्रतिभागी :	क	ख	ग	घ
वजनः	55	80	65	70

आकार 2 के सभी संभावित प्रतिस्थापन प्रतिदर्श की परिगणना करें । प्रतिदर्श इकाइयों के लिए अध्ययन चरों (वजन) के मान भी लिखें।

हलः यहॉ $N = 4$ और $n = 2$ है। इसलिए $4^2 = 16$ संभावित प्रतिदर्श होंगे। प्रतिदर्श इकाइयों के वजनमान के साथ इन प्रतिदर्श की परिगणना नीचे दी गई है।

सारणी 7.1: चर मानों के साथ संभावित प्रतिदर्श

प्रतिदर्श संख्या	प्रतिदर्शी प्रतिभागी	प्रतिदर्शी प्रतिभागियों के वजन–मान	प्रतिदर्श संख्या	प्रतिदर्शी प्रतिभागी	प्रतिदर्शी प्रतिभागियों के वजन–मान
1	ए, ए	55 , 55	9	सी, ए	65 , 55
2	ए, बी	55 , 80	10	सी, बी	65 , 80
3	ए, सी	55 , 65	11	सी, सी	65 , 65
4	ए, डी	55 , 70	12	सी, डी	65 , 70
5	बी, ए	80 , 55	13	डी, ए	70 , 55
6	बी, बी	80 , 80	14	डी, बी	70 , 80
7	बी, सी	80 , 65	15	डी,सी	70 , 65
8	बी, डी	80 , 70	16	डी, डी	70 , 80

परिभाषा 7.4: बिना प्रतिस्थापन वाले प्रतिदर्श में समष्टि में से इकाइयों का चयन एक–एक करके किया जाता है और किसी ड्रा विशेष में चुनी गई इकाई को, अगले ड्रा में किसी इकाई का चयन करने से पहले समष्टि में वापस प्रतिस्थापित नहीं किया जाता।

स्पष्टतः बिना प्रतिस्थापन वाले प्रतिदर्श में किसी भी इकाई का चयन एक से अधिक बार नहीं किया जाता। यदि नमूने में इकाइयों के चयन के क्रम पर ध्यान न दिया जाए तब इस चयन प्रक्रिया में $\left(\frac{N}{n}\right)$ संभावित प्रतिदर्श होंगे।

उदाहरण 7.2:

उदाहरण 7.1 के ऑकड़ों का प्रयोग करके, आकार 2 के बिना प्रतिस्थापन वाले सभी संभावित प्रतिदर्शों की परिगणना करें और संबंधित प्रतिदर्श इकाइयों के वजन मान की सूची भी बनाएँ।

हलः यहॉ संभाविक प्रतिदर्श की संख्या $({}^{4}_{2})$ = 6 होगी । इनकी परिगणना नीचे की गई है। नोट करें कि ए,ए या बी,बी जैसे कोई भी प्रतिदर्श, संभावित प्रतिदर्शों की सूची में नहीं है। ए, बी तथा बी, ए क्रम संख्या वाले प्रतिदर्शों को एक प्रतिदर्श के रूप में लिया जाता है।

सारणी 7.2: चर मानों के साथ संभावित प्रतिदर्श

प्रतिदर्श संख्या	प्रतिदर्शी प्रतिभागी	प्रतिदर्शी प्रतिभागियों के वजन–मान	प्रतिदर्श संख्या	प्रतिदर्शी प्रतिभागी	प्रतिदर्शी प्रतिभागियों के वजन–मान
1	ए, बी	55 , 80	4	बी, सी	80 , 65
2	ए, सी	55 , 65	5	बी, डी	80 , 70
3	ए, डी	55 , 70	6	सी, डी	65 , 70

8. प्रतिचयन की आवश्यकता (Need of Sampling):

व्यवहारिकतया आवश्यक ऑकड़ों में त्रुटि की कुछ सीमा की अनुमति है, सम्पूर्ण परिगणना सर्वेक्षण का एक प्रभावी विकल्प, प्रतिदर्श सर्वेक्षण हो सकता है, जहां समष्टि में से उपयुक्त पद्धति से चुनी गई केवल कुछ इकाइयों का सर्वेक्षण किया जाता है और उन इकाइयों के आधार पर समष्टि के बारे में अनुमान लगाया जाता है। यह आसानी से देखा जा सकता है कि प्रतिदर्श सर्वेक्षण की तुलना में सम्पूर्ण परिगणना सर्वेक्षण में अधिक समय लगता है, अतिव्ययी है, सीमित विषय समावेश करने की दृष्टि से इसमें कम गुंजाइश है और इसमें समावेश (coverage) प्रेक्षणात्मक और सारणीकरण संबंधी त्रुटियॉ होने की अधिक संभावना है। कुछ जांच–पड़तालों में यह आवश्यक है कि ऑकड़ों के एकत्रीकरण के लिए विशेष प्रकार के उपकरण तथा अत्यधिक प्रशिक्षित फील्ड स्टाफ की जरूरत होती है जिसमें प्रतिचयन पद्धति के अलावा अन्य किसी पद्धति से सर्वेक्षण करना लगभग असंभव होता है। इसके अलावा, विवरणात्मक सर्वेक्षणों में, संपूर्ण परिगणना सर्वेक्षण करना असंभव एवं अव्यवहारिक है। अतः यदि किसी समूह में बिजली के बल्बों का औसत जीवनकाल प्राप्त करना हो तो प्रेक्षणों को समष्टि या समष्टि के एक भाग (या प्रतिदर्श) तक सीमित रहते हुए प्रतिदर्शी प्रेक्षणों के आधार पर संपूर्ण समष्टि के बारे में अनुमान लगाना होगा। तथापि, चूंकि प्रतिदर्श सर्वेक्षण में समष्टि के एक भाग के आधार पर संपूर्ण समष्टि के बारे में अनुमान लगाया जाता है, इसलिए प्रतिदर्श सर्वेक्षण से प्राप्त परिणाम, समष्टि मानों से भिन्न पाये जाते हैं और यह

अंतर चुने गए भागों या प्रतिदर्श पर निर्भर करता है। इस प्रकार किसी प्रतिदर्श द्वारा उपलब्ध कराई गई सूचना में ऐसी त्रुटि हो सकती है जिसे प्रतिचयन त्रुटि के रूप में जाना जाता है। दूसरी ओर, चूंकि समष्टि के केवल एक भाग का सर्वेक्षण ही किया जाना है, इसलिए किसी सम्पूर्ण परिगणना सर्वेक्षण की तुलना में प्रतिदर्श सर्वेक्षण में उचित नियंत्रणों द्वारा और प्रशिक्षित स्टाफ नियोजित करके अभिनिश्चयन या प्रेक्षणात्मक त्रुटियों को समाप्त करने की अधिक गुंजाइश है। यदि प्रतिदर्श सर्वेक्षण, कुछ विनिर्धारित सांख्यिकीय सिद्धांतों के अनुसार किया जाए तो ना केवल संपूर्ण समष्टि के लक्षणों का आकलन प्राप्त किया जा सकेगा बल्कि आकलन की प्रतिचयनात्मक त्रृटियों का वैध आकलन प्राप्त करना भी संभव होगा। प्रतिदर्श सर्वेक्षण की योजना बनाने और उसके कार्यान्वयन में विभिन्न चरण होते हैं। प्रतिदर्श सर्वेक्षण में एक प्रधान चरण, प्राथमिक ऑकड़ों के एकत्रीकरण की पद्धतियों से संबंधित है।

9 विभिन्न अवधारणाएँ और परिभाषाएँ (Assumptions and Definitions):

(क) जनसंख्या/समष्टि (Population)

किसी समय या अवधि विशेष पर किसी क्षेत्र विशेष में किसी विशिष्ट प्रकार की सभी इकाइयों के संग्रहण को जनसंख्या या समष्टि कहा जाता है। इस प्रकार, अपेक्षित ऑकड़ों के स्वरूप पर निर्भर करते हुए, हम किसी क्षेत्र में व्यक्तियों, परिवारों, फार्मों, मवेशियों की जनसंख्या या किसी वन में पेड़ों या पत्तियों की संख्या या किसी तालाब में मछलियों की संख्या पर विचार कर सकते हैं।

(ख) प्रतिचयन इकाई (Sampling Unit):

प्रारंभिक इकाइयॉ या ऐसी इकाइयों का समूह, जो स्पष्ट रूप से परिभाषित किए जाने के साथ–साथ अभिज्ञेय और अवलोकनीय हैं, तथा प्रतिदर्श के उद्देश्य के लिए सुविधाजनक हैं उन्हें प्रतिचयन इकाइयॉ कहा जाता है। उदाहरण के लिए एक पारिवारिक बजट पूछताछ में, आमतौर पर किसी परिवार को प्रतिचयन इकाई माना जाता है क्योंकि प्रतिचयन के लिए और अपेक्षित सूचना का पता लगाने के लिए इसे सुविधाजनक पाया गया है। किसी फसल सर्वेक्षण में किसी परिवार के स्वामित्व वाले या उसके द्वारा चलाए जाने वाले किसी फार्म या फार्मों के समूह को प्रतिचयन इकाई माना जाता है।

(ग) प्रतिचयन ढांचा (Sampling Frame):

अध्ययन की जाने वाली समष्टि की सभी प्रतिदर्शी इकाइयों (पहचान ब्योरा एवं नक्शे की सीमा दर्शाते हुए) की सूची को प्रतिचयन ढांचा कहा जाता है। फार्मों की सूची और भारत में गांवों या संयुक्त राज्य के देशों में उपयुक्त क्षेत्र खण्डों की सूची किसी ढांचे

के उदाहरण हैं। यह ढांचा अद्यतन होना चाहिए तथा प्रतिचयन इकाइयॉ अनुलिपिकरण (duplication) एवं लुप्त (ommision) त्रुटियों से मुक्त होनी चाहिए।

(घ) यादृच्छिक प्रतिदर्श (Random Sample):

किसी निर्धारित प्रक्रिया के अनुसार समष्टि से चुनी गई एक या एक से अधिक प्रतिचयन इकाइयों को प्रतिदर्श कहा जाता है। प्रतिदर्श को यादृच्छिक (random) या प्रायिकता (probability) प्रतिदर्श माना जाएगा यदि इसका चयन, अवसर के निश्चित नियमों द्वारा नियंत्रित हो। दूसरे शब्दों में, कोई यादृच्छिक या प्रायिकता प्रतिदर्श ऐसे तरीके से लिया गया प्रतिदर्श होता है जिसमें समष्टि से चयन की पूर्व–निर्धारित संभाव्यता हो। उदाहरण के लिए यदि किसी समष्टि में N प्रतिचयन इकाइयॉ $U_1, U_2, \ldots, U_N$ हों तो हम, पिछले ड्रा में चुनी गई प्रतिचयन इकाइयों को प्रतिस्थापित करके या प्रतिस्थापित किए बिना, प्रत्येक इकाई के लिए समान प्रायिकता के साथ इकाई चुन कर n इकाइयों के प्रतिदर्श का चयन कर सकते हैं।

(ड.) गैर–यादृच्छिक प्रतिदर्श (Non random Sample):

किसी गैर–यादृच्छिक प्रक्रिया द्वारा चुने गए प्रतिदर्श को गैर–यादृच्छिक (non-random) प्रतिदर्श कहा जाता है। गैर–यादृच्छिक प्रतिदर्श जो निरूपक प्रतिदर्श प्राप्त करने की दृष्टि से किसी अनुमान का प्रयोग करके लिया गया है, को अनुमान या उद्देश्यात्मक प्रतिदर्श कहा जाता है। उद्देश्यात्मक प्रतिदर्श में इकाइयों का चयन, प्रतिदर्श में समष्टि का परिवर्तन सुनिश्चित करने की दृष्टि से उपलब्ध सहायक सूचना पर विचार करके किया जाता है। इस प्रकार के प्रतिचयन प्रयोग कदाचित बड़े पैमाने पर किए जाने वाले सर्वेक्षणों में मुख्यतः इसलिए नहीं किया जाता है क्योंकि आमतौर पर यह संभव नहीं होता कि विषयात्मक चयन में पक्षपात के खतरे और इकाइयों के चयन की संभाव्यताओं पर सूचना के अभाव के कारण, विचाराधीन समष्टि मापदंडों और उनकी प्रतिचयनात्मक त्रुटियों का पूरी तरह वैध आकलन प्राप्त किया जा सके।

(च) समष्टि प्राचल (parameter)

मान लो एक सीमित समष्टि की N इकाइयॉ $U_1, U_2, \ldots, U_N$ हैं और Y_i, $i=1,2\ldots..N$ चर Y का i-th इकाई Ui के लिए मान है । उदाहरण के लिए इकाई एक फार्म हो सकता है और अध्ययनाधीन चर किसी फसल विशेष के अंतर्गत क्षेत्र हो सकता है। सभी समष्टि इकाइयों (समष्टि सूचकों) के माप का फलन (function) या केवल प्राचल के रूप में जाना जाता है। सर्वेक्षणों में आकलन किए जाने वाले अपेक्षित

कुछ महत्वपूर्ण प्राचल समष्टि योग $Y = \sum_{i=1}^{N} Y_i$ और समष्टि माध्य (औसत) $\overline{Y} = \sum_{i=1}^{N} Y_i / N$ हैं।

(छ) प्रतिदर्शज, आकलक एवं आकलन (Sample, Estimator and Estimation):

मान लो प्रायिकता पद्धति के अनुसार N इकाइयों की समष्टि से n इकाइयों के प्रतिदर्श का चयन किया गया है और मान लो कि प्रतिदर्श प्रेक्षणों को $y_1, y_2, \ldots, y_n$ द्वारा निर्दिष्ट किया गया है। इन मानों के फलन को, जो अज्ञात समष्टि प्राचल से मुक्त है, प्रतिदर्शज कहते हैं।

आकलक एक प्रतिदर्शज है, जो किसी समष्टि प्राचल के आकलन करने के लिए विनिर्धारित प्रक्रिया द्वारा किया जाता है। आकलक एक यादृच्छिक चर है और इसका मान अलग–अलग प्रतिदर्शों में भिन्न–भिन्न होता है और प्रतिदर्शों का चयन विनिर्धारित प्रायिकता के साथ किया जाता है। किसी निश्चित प्रतिदर्श के लिए आकलक द्वारा प्रप्त विशेष मान को आकलन के रूप में जाना जाता है।

(ज) प्रतिदर्श अभिकल्पना (Sampling Design):

किसी एक तरह के सभी संमाव्य प्रतिदर्शों के (प्रायिकता सहित) एक स्पष्ट विनिर्देशन को प्रतिदर्श अभिकल्पना कहा जाता है। उदाहरण के लिए, मान लो हम प्रतिस्थापन के साथ समान प्रायिकता वाली n इकाइयों के एक प्रतिदर्श का चयन करते हैं तो प्रतिदर्श अभिकल्पना में N^n संभावित प्रतिदर्श (चयन के क्रम और प्रतिदर्श में इकाइयों की पुनरावृत्ति को ध्यान में रखते हुए) होंगे जिनमें प्रत्येक की चयन प्रायिकता $1/N^n$ होगी क्योंकि n ड्रा में से प्रत्येक में, N इकाइयों में से किसी भी इकाई का चयन हो सकता है। इसी प्रकार प्रतिस्थापन के बिना समान संभाव्यता वाली n इकाइयों के प्रतिचयन में संभावित प्रतिदर्शों की संख्या (इकाइयों के चयन के क्रमों की उपेक्षा करके) (N/n) है जिनमें प्रत्येक की चयन प्रायिकता $1/(N/n)$ होगी ।

(झ) अनभिनत आकलन (unbiased)

मान लो i-th प्रतिदर्श प्राप्त करने की प्रायिकता p_i तथा आकलन t_i है अर्थात इस प्रतिदर्श पर आधारित समष्टि प्राचल θ के आकलक t का मान $M_o (i=1, 2, \ldots, M_o)$ विनिर्धारित प्रायिकता पद्धति के लिए संभावित प्रतिदर्शों की कुल संख्या है। आकलक t का अपेक्षित मान या ओसत $E(t) = \sum_{i=1}^{M_o} t_i \, p_i$

आकलक t को समष्टि प्राचल θ का एक अनभिनत आकलक कहा जायेगा यदि इसका अपेक्षितमान θ के बराबर होगा चाहे y–मान कितना भी हो। यदि आकलक का अपेक्षित मान समष्टि प्राचल के बराबर नहीं है तो आकलक t को θ का अभिनत (bias) आकलक कहा जाता है। अकलक t को, समष्टि प्राचल के लिए सकारात्मक अभिनत या नकारात्मक अभिनत कहा जायेगा यदि अभिनत मान क्रमशः सकारात्मक या नकारात्मक होगा।

(ञ) त्रुटि के माप (Measure of error):

प्रतिदर्श अभिकल्पना के द्वारा जयादातर अलग–अलग प्रतिदर्श प्राप्त होते हैं, पर इसलिए प्रतिदर्श प्रेक्षणों पर आधारित आकलन, सामान्यतः प्रत्येक प्रतिदर्श में अलग–अलग होते हैं और विचाराधीन प्राचलों के माप से अलग–अलग होंगे i-th प्रतिदर्श पर आधारित आकलन t_i एवं प्राचलों के मान के बीच के अंतर अर्थात $(t_i - \theta)$ को आकलन की त्रुटि कहा जा सकता है और यह त्रुटि अलग–अलग प्रतिदर्शों में भिन्न भिन्न होती है। सही मान (true value) से विभिन्न आकलनों के (expected value) संभावित मान विचलन का औसत परिमाण, का मान त्रुटि–वर्ग के संभावित मान द्वारा निकाला जाता है।
अर्थात

M(t) $E(t-\theta)^2 = \sum_{i=1}^{M_o}(t_i - \theta)^2 \rho_i$ है और इसे आकलक की औसत वर्ग त्रुटि (एम0एस0ई0) के रूप में जाना जाता है। एमएसई को परिशुद्धता का मापक माना जा सकता है जिससे आकलक t के द्वारा प्राचल का आकलन किया जाता है।

संभावित मान से आकलक के अन्तर के वर्ग को प्रतिचयन विचलन कहा जाता है। यह संभावित मान से आकलक के विचलन का मापक है और इसे $V(t) = \sigma^2 t = E\{t - E(t)\}^2 = E\{t\}^2 - \{E(t)\}^2$ द्वारा दर्शाया जाता है। इसे आकलक t की परिशुद्धता (precision) के नाम से भी जाना जाता है। t के एमएसई को प्रतिचयन विचलन और त्रुटि वर्ग के योग के रूप अभिव्यक्त किया जा सकता है। अभिनत आकलक के लिए एमएसई और प्रतिचयन विचलन एक ही होते हैं। प्रतिचयन विचलन के वर्गमूल $\sigma(t)$ को आकलक t की मानक त्रुटि (एसई) कहा जाता है व्यवहार में अधिकतर $\sigma(t)$ का वास्तविक मान ज्ञात नहीं होता और इसलिए इसका आकलन स्वंय प्रतिदर्श से ही किया जाता है।

(ट) विश्वास्यता अंतरा (confidence internal)

प्रतिदर्श आकलन पर आधारित आकलक t के मान के अनुसार प्रतिदर्शों के बारंबारता बंटन को आकलक t का प्रतिदर्शी बंटन (sampling distribution) कहा जाता है। यह उल्लेख करना अनिवार्य है कि हालांकि समष्टि बंटन संभवतः प्रसामान्य (normal) न हो परंतु आकलक t का प्रतिदर्शी बंटन अधिकतर प्रसामान्य के निकट होता है जबकि प्रतिदर्श का आकार प्रर्याप्त रूप से बड़ा हो। यदि आकलक t अनभिनत है और प्रसामान्य बंटन है तो अंतराल $\{t \pm kSE(t)\}$ में घटनाओं के P% में प्राचल θ शामिल होने की संभावना होती है जहां P, मानक प्रसामान्य वेरियेट के बंटन के क्षेत्र – K और + K के बीच का अनुपात हो। माने गए अंतराल को, विश्वास्यता सीमा t-K SE (t) और t+K SE (t) वाले P% के विश्वास्यता गुणांक के साथ प्राचल θ का विश्वास्यता अंतराल कहा जाता है। उदाहरण के लिए, यदि किसी बड़े कारखाने में दैनिक प्रयोग की बैटरियों के रिकार्ड का कोई यादृच्छिक प्रतिदर्श, मानक त्रुटि SE(t) = 4.6 दिन के साथ औसत जीवनकाल t = 394 दिन दर्शाता है तो इस बात के 100 में से 99 अवसर हैं कि बैटरियों की संख्या में औसत जलवनकाल होगाः

$t_L = 394 - (2.58)(4.6) = 382$ दिन
$t_u = 394 + (2.58)(4.6) = 406$ दिन

382 दिन और 406 दिन सीमाओं को, t के लिए 99 प्रतिशत विश्वास्यता अंतराल की निम्न और उच्च सीमाएँ कहा जाता है। किसी एकल सर्वेक्षण से किये गये एकल आकलन के लिये यह कथन निश्चित रूप से सही नहीं है कि θ का मान 382 और 406 दिन के बीच में है। 99 प्रतिशत विश्वास्यता का तात्पर्य यह है कि यदि किसी समष्टि में अनेक बार एक ही प्रतिचयन पद्धति का प्रयोग किया गया है तो इन कथनों के 99 प्रतिशत कथन सही होंगे और 1 प्रतिशत गलत होंगे।

(ठ) प्रतिचयन एवं गैर प्रतिचयन त्रुटि (Sampling and Non Sampling Error):

प्रेक्षणों के आधार पर प्रतिदर्शों द्वारा समष्टि के बारे में अनुमान लगाने के कारण हुई त्रुटि को प्रतिचयन त्रुटि कहा जाता है। सम्पूर्ण परिगणना सर्वेक्षण में प्रतिचयन त्रुटि मौजूद नहीं होती क्योंकि सम्पूर्ण इकाइयों का सर्वेक्षण किया जाता है । प्रतिचयन त्रुटियों के अलावा अन्य त्रुटियों जैसे उत्तर न देने, अधूरी और उत्तर के अयथार्थ होने से उत्पन्न त्रुटियों को गैर प्रतिचयन त्रुटियॉ कहा जाता है और किसी प्रतिदर्श सर्वेक्षण

की तुलना में किसी सम्पूर्ण परिगणना सर्वेक्षण में उनके अधिक व्यापक और महत्वपूर्ण होने की संभावना होती है। गैर प्रतिचयन त्रुटियॉ आरंभिक चरण, जब सर्वेक्षण की योजना एवं अभिकल्पना तैयार की जाती है, से लेकर अंतिम चरण जब आंकड़ों पर कार्यवाही और उनका विश्लेषण किया जाता है, तक विभिन्न कारणों की वजह से उत्पन्न होती हैं।

प्रतिचयन त्रुटियां, प्रतिदर्श आकार – (प्रतिदर्श के लिए चुनी गई इकाइयों की संख्या) में वृद्धि के साथ ज्यादातर घटती हैं जबकि गैर प्रतिचयन त्रुटि के प्रतिदर्श आकार में वृद्धि के साथ बढ़ने की संभावना होती है।

जहॉ तक गैर प्रतिचयन त्रुटि का संबंध है, प्रतिदर्श सर्वेक्षण की तुलना में सम्पूर्ण परिगणना सर्वेक्षण में इसके अधिक होने की संभावना होती है क्योंकि पहले मामले में बाद वाले मामले की अपेक्षा फील्ड और सारणीकरण के चरणों में बेहतर संगठन और उपयुक्त ढंग से प्रशिक्षित कार्मिकों का प्रयोग करके गैर प्रतिचयन त्रुटि को कम करना संभव है।

प्रतिदर्श अभिकल्पना के मौलिक सिद्धांत (Sampling Design and Fundamental Principle):

प्रत्येक प्रतिदर्श अभिकल्पना के साथ सर्वेक्षण की लागत और बनाये गए आकलनों की परिशुद्धता (प्रतिचयन भिन्नता के रूप में मापी जाती है) जुड़ी होती है। अभिकल्पना व्यावहारिक होनी चाहिए ताकि इन्हें अपेक्षित विनिर्देशनों के अनुसार बनाया जाना संभव हो। इन सभी अभिकल्पनाओं में से एक ऐसी अभिकल्पना को वरीयता दी जानी चाहिए जो सर्वेक्षण की एक निश्चित लागत पर उच्चतम परिशुद्धता या परिशुद्धता के एक विनिर्धारित स्तर के लिए न्यूनतम लागत पर हो। यह प्रतिदर्श अभिकल्पना का मार्गदर्शी सिद्धांत है।

संदर्भ (References):

1. सुखात्मे, पी.वी. और सुखात्मे, बी.वी (1970): सैम्पलिंग थ्योरी ऑफ सर्वेज विद एप्लीकेशन, द्वितीय संस्करण, आयोवा स्टेट यूनीवर्सिटी प्रेस, यू.एस.ए.।
2. सुखात्मे, पी.वी.; सुखात्मे, बी.वी .; सुखात्मे, एस. और अशोक, सी. (1984): सैम्पलिंग थ्योरी ऑफ सर्वेज विद एप्लीकेशन, तृतीय संस्करण, आयोवा स्टेट यूनीवर्सिटी प्रेस, यू.एस.ए.।
3. कोकरन विलियम जी. (1977): सैम्पलिंग टेकनीक्स, तृतीय संस्करण, जोह्न विले एण्ड संस।

4. मूर्ती, एम.एन. (1977): सैम्पलिंग थ्योरी एण्ड मैथड्स, स्टेटस्टीकल पब्लिशिंग सोसाइटी, कलकत्ता।
5. सिंह, दरोगा और चौधरी, एफ.एस. (1986): थ्योरी एण्ड एनालिसिज ऑफ सेम्पल सर्वे डिज़ाइन्स, विले ईस्टर्न लिमिटेड।
6. सिंह, दरोगा; सिंह, पद्म और प्राणेश कुमार (1978): हैंडबुक ऑफ सैम्पलिंग मैथड्स। आई.ए.एस.आर.आई., नई दिल्ली।
7. देशराज (1968): सैम्पलिंग थ्योरी, टाटा मेकग्राहिल पब्लिशिंग कंपनी लिमिटेड।
8. देशराज और प्रमोद चंडोक (1998): सैम्पल सर्वे थ्योरी, नरोसा पब्लिश्शिग हाउस।
9. सिंह रविन्द्र और मंगत, एन.एस. (1996): एलीमेंट्स ऑफ सर्वे सैम्पलिंग, क्लूवेर अकेडेमिक पब्लिशर्स।
10. राउत के.सी. और सिंह डी. (1976): मेथड्स ऑफ कलेक्शन ऑफ एग्रीकल्चरल स्टेटस्टिक्स इन इन्डिया, आई.ए.एस.आर.आई., नई दिल्ली।

अध्याय 07

भारत में कृषि साँख्यिकी प्रणाली का परिचय

तौक़ीर अहमद
प्रभाग-अध्यक्ष एवं प्रमुख वैज्ञानिक, प्रतिदर्श सर्वेक्षण प्रभाग
भा.कृ.अनु.प.-भारतीय कृषि साँख्यिकी अनुसंधान संस्थान, लाइब्रेरी एवेन्यू, नई दिल्ली-110012

1. परिचय (Introduction):

भारत मुख्य रूप से कृषि आधारित देश है और इसकी अर्थव्यवस्था काफी हद तक कृषि पर निर्भर करती है। वर्तमान में, कृषि का योगदान राष्ट्रीय सकल घरेलू उत्पाद का लगभग एक तिहाई है और संबद्ध गतिविधियों में 70 प्रतिशत से अधिक भारतीय आबादी को रोजगार प्रदान करता है। इसलिए हमारे देश का विकास काफी हद तक कृषि के विकास पर निर्भर करता है। कृषि के विभिन्न क्षेत्रों के लिए संसाधनों के नियोजन और आबंटन के लिए कृषि उत्पादन की जानकारी बहुत महत्वपूर्ण है। भारत में कृषि सांख्यिकी की एक लंबी परंपरा है। कौटिल्य के अर्थशास्त्र में उनके संग्रह का उल्लेख प्रशासनिक व्यवस्था के एक भाग के रूप में किया गया है। मुगल काल के दौरान भी राजस्व प्रशासन की जरूरतों को पूरा करने के लिए कुछ बुनियादी कृषि आँकड़े एकत्र किए गए थे। आइन–ए–अकबरी सबसे महत्वपूर्ण दस्तावेज़ है जो मुगल काल के दौरान आंकड़े एकत्र करने के तरीके पर बहुत प्रकाश डालता है। मुगल काल के बाद ब्रिटिश शासन ने रैयतवारी प्रणाली शुरू की, जिसे 18वीं शताब्दी के दौरान भू–राजस्व एकत्र करने के लिए शुरू किया गया था। 1866 में, ब्रिटिश सरकार ने मुख्य रूप से राजस्व प्रशासन के उप–उत्पाद के रूप में कृषि सांख्यिकी का संग्रह शुरू किया ओर ये भू–राजस्व के संग्रह में सरकार के तत्कालीन प्राथमिक हित को दर्शाते थे। इसके बाद, मुख्य रूप से ब्रिटिश व्यापार हितों की सेवा के लिए तैयार किए गए फसल पूर्वानुमानों पर जोर दिया गया। लिवरपूल की एक प्रमुख फर्म, गेहूँ के व्यापार द्वारा, 1884 में गेहूँ के पूर्वानुमान की तैयारी की गई और 1884 से देश में भूमि उपयोग के आँकड़े उपलब्ध हैं। 1900 तक, तिलहन, चावल, कपास, जूट नील और गन्ने को भी पूर्वानुमानित फसलों की सूची में जोड़ा गया था।

कृषि सांख्यिकी प्रणाली बड़ी संख्या में मापदंडों पर मूल्यवान आँकड़े उत्पन्न करती है। कुछ बहुत ही महत्वपूर्ण आँकड़े भूमि उपयोग के आँकड़े हैं और समय पर रिपोर्टिंग योजना (TRS) के माध्यम से प्रमुख फसलों के तहत क्षेत्र और साथ ही पूर्ण गणना के आधार पर, सामान्य फसल अनुमान सर्वेक्षण (GCES), खेती अध्ययन की लागत की योजना के माध्यम से उपज का अनुमान है। सी.सी.एस. (CCS) फसल सांख्यिकी में सुधार (आई सी एस) की योजना के माध्यम से टीआरएस की विश्वसनीयता और उपज अनुमानों की जांच करता है, उत्पादन अनुमानों की लागत, कृषि मजदूरी, सिंचाई के आंकड़े, कृषि जनगणना के संचालन और पंचवर्षीय आधार पर पशुधन की गणना करता है। कृषि सांख्यिकी प्रणाली एकीकृत नमूना सर्वेक्षण (आई एस एस) की योजना के माध्यम से पशुधन उत्पादों पर डेटा भी उत्पन्न करती है, थोक और

खुदरा मूल्य एकत्र करती है, बाजार की खुफिया जानकारी का संचालन करती है और वर्षा और मौसम की स्थिति का निरीक्षण करती है।

2.फसल क्षेत्र सांख्यिकी (Crop Area Statistics):

क्षेत्र के आंकड़ों के संग्रह के संबंध में देश को चार व्यापक श्रेणियों में विभाजित किया जा सकता है, अर्थात् (i) अस्थायी रूप से बसे हुए राज्य, (ii) स्थायी रूप से बसे हुए राज्य और (iii) अन्य क्षेत्र।

(i) अस्थायी रूप से बसे हुए राज्य

हमारे देश में अस्थायी रूप से बंदोबस्त की व्यवस्था 1892 में एक अवधि के लिए भू–राजस्व तय करने की दृष्टि से शुरू की गई थी, जो अगले बंदोबस्त के समय परिवर्तन के अधीन थी। सामान्यतः दो बस्तियों के बीच का अन्तराल 25 से 30 वर्ष का होता था। भू–राजस्व का निर्धारण करने के लिए उत्पादन पूर्वानुमान लगाने के लिए भू–राजस्व, भूमि मूल्य आदि के बारे में विस्तृत आँकड़े एकत्र किए जाने हैं। अस्थायी रूप से बसे हुए क्षेत्रों में फसल क्षेत्र के आंकड़ों की जानकारी ग्राम लेखाकार या पटवारी द्वारा एकत्र की जाती है और एक रजिस्टर में दर्ज की जाती है जिसे उत्तरी भारत में खसरा के नाम से जाना जाता है। ग्रामीण लेखाकार को देश के विभिन्न भागों में अलग–अलग नामों से पुकारा जाता है जैसे दक्षिण में कर्णम, महाराष्ट्र में तेलत्ती, बिहार में कर्मचारी, उत्तर प्रदेश में लेखपाल आदि। यह श्रेणी देश के कुल सूचित क्षेत्र का लगभग 86 प्रतिशत है।

ग्राम लेखाकार द्वारा एकत्रित फसल क्षेत्र के आँकड़े पूरी गणना के आधार पर गिरदावरी कहलाते हैं। ग्राम लेखापाल को प्रत्येक फसल के मौसम में गाँव के प्रत्येक खेत का दौरा करना होता है और विभिन्न फसलों/भूमि उपयोग श्रेणियों के तहत क्षेत्र और खसरा रजिस्टर नामक रूपों में इसकी स्थिति जैसी जानकरी दर्ज करनी होती है। ग्राम लेखापाल का कार्य उत्तर भारत में कानुंगो के नाम से जाने जाने वाले तत्काल वरिष्ठ अधिकारी द्वारा पर्यवेक्षण किया जाता है। अस्थायी रूप से बसे राज्यों के अधिकांश भौगोलिक क्षेत्रों का भूकर सर्वेक्षण किया जाता है और विस्तृत नक्शे तहसील ओर जिला मुख्यालयों में उपलब्ध है। विभिन्न ग्राम लेखाकारों द्वारा प्राप्त आँकड़ों को उच्च प्रशासनिक इकाइयों जैसे ब्लॉक, तहसील, जिला, राज्यों आदि में फसल क्षेत्र के आँकड़े प्रापत करने के लिए एकत्र किया जाता है। डेटा संग्रह की इस प्रणाली का पालन 18 राज्यों अर्थात् आंध्र प्रदेश, असम (पहाड़ी जिलों को छोड़कर) बिहार, छत्तीसगढ़, गोवा, गुजरात, हरियाणा, हिमाचल प्रदेश, जम्मू और कश्मीर, झारखंड, कर्नाटक, मध्य प्रदेश, महाराष्ट्र, पंजाब, में किया जा रहा है। राजस्थान, तमिलनाडु, उत्तरांचल और उत्तर प्रदेश और 5 केंद्र शासित प्रदेश चंडीगढ़, दादरा और नगर हवेली, दमन और डयू, दिल्ली और पांडिचेरी।

(ii) स्थायी रूप से बसे हुए राज्य

केरल, उड़ीसा और पश्चिम बंगाल नाम के तीन राज्य स्थायी रूप से बसे राज्यों की श्रेणी में आते हैं। इन राज्यों के मामले में भू–राजस्त स्थायी रूप से तय किया गया था और संशोधन का सवाल ही नहीं उठता था। इन राज्यों में क्षेत्र के आँकड़ों के विवरण दर्ज करने की कोई व्यवस्था नहीं है

क्योंकि गाँव के लेखाकार जैसे गाँव के लिए कोई स्थायी राजस्व कर्मचारी नहीं है जैसा कि अस्थायी रूप से बसे हुए क्षेत्र के मामले में होता है। प्रारंभ में इन क्षेत्रों में क्षेत्र के आंकड़े एकत्र करने की कोई समान प्रणाली नहीं थी। पुलिस चौकीदार या ग्राम प्रधान आमतौर पर अनुमान के आधार पर आंकड़े उपलब्ध कराते थे जो काफी अविश्वसनीय थे। स्थायी रूप से बसे राज्यों में इन आंकड़ों की गुणवत्ता में सुधार करने के लिए, वर्तमान में, इन राज्यों में क्षेत्र के आंकड़े "कृषि सांख्यिकी रिर्पोटिंग एजेंसी (ईएआरएएस) की स्थापना" नामक योजना के तहत विशेष रूप से नियुक्त फील्ड स्टाफ द्वारा एकत्र किए जाते हैं, जिसे 1968–69 मे शुरू किया गया था। ईएआरएएस के अंतर्गत आने वाले राज्यों में, सभी क्षेत्रों (सर्वेक्षण संख्या) अर्थात् गिरदावरी की संपूर्ण गणना राज्यों के 20 प्रतिशत गाँवों के यादृच्छिक नमूने में प्रत्येक वर्ष आयोजित की जाती है, जिनका चयन इस प्रकार किया जाता है कि 5 वर्ष की अवधि के दौरान, पूरे राज्य को कवर किया गया है। यह श्रेणी देश के कुल रिर्पोट किए गए क्षेत्र का लगभग 9 प्रतिशत है।

(iii) अन्य क्षेत्र

उत्तर पूर्वी क्षेत्रों में शेष आठ राज्यों अर्थात् अरूणाचल प्रदेश, मणिपुर, मेघालय, मिज़ोरम, नागालैंड, सिक्किम और त्रिपुरा और दो अन्य केंद्र शासित प्रदेशों जैसे अंडमान और निकोबार द्वीप समूह और लक्षद्वीप में उचित रिपोर्टिंग प्रणाली नहीं है, हालांकि त्रिपुरा और स्किम राज्यों (कुछ छोटे क्षेत्रों को छोड़कर) का भूकर सर्वेक्षण किया जाता है। इन क्षेत्रों में क्षेत्र के आँकड़ों का संकलन पारंपरिक तरीकों पर आधारित होता है जिसमें गाँव के चौकीदारों द्वारा व्यक्तिगत मूल्यांकन के आधार पर अनुमानों की सूचना दी जाती है। यह श्रेणी देश के कुल रिर्पोट किए गए क्षेत्र का लगभग 5 प्रतिशत है।

3. समय पर रिपोर्टिंग योजना (टीआरएस) (Timely Reporting System):

बुआई और क्षेत्र के अनुमानों की उपलब्धता और फसलों की कटाई और उत्पादन के अनुमानों की उपलब्धता के बीच समय अंतराल को कम करने के लिए, मुख्य फसलों के क्षेत्र और उत्पादन(टीआरएस) के अनुमानों की समय पर रिपोर्टिंग के लिए एक केंद्र प्रोयाजित योजना वर्ष 1968–70 में कृषि और सिंचाई मंत्रालय द्वारा शुरू की गई थी। टी आर एस का मूल उद्देश्य स्थायी रूप से बसे राज्यों में फसल काटने के प्रयोगों के लिए फसल उगाने वाले क्षेत्रों के चयन के नमूना फ्रेम प्रदान करने के अलावा प्रमुख फसलों के क्षेत्र के आंकड़े उपलब्ध कराने के लिए समय अंतराल को कम करना है। इस योजना के तहत, प्रत्येक स्तर (तहसील/राजस्व निरीक्षक सर्कल/पटवारी सर्कल आदि) के गांवों को 5 स्वतंत्र गैर–अतिव्यापी सेटों में विभाजित किया गया है, जिनमें ये प्रत्येक में गांवों की कुल संख्या का पांचवां हिस्सा शामिल है। बेतरतीब ढंग से चुने गए गाँव के एक सेट को प्रत्येक मौसम में बुवाई के तुरंत बाद प्राथमिकता के आधार पर फसल निरीक्षण के लिए चुना

जाता है, लेकिन इस तरह के फसल निरीक्षण के लिए भूमि अभिलेख नियमावली में निर्धारित अवधि से पहले प्रमुख फसलों के लिए अग्रिम रूप से फसल क्षेत्र के आँकड़ों का अनुमान लगाने के लिए गाँव के फसल क्षेत्र के विवरण उच्च अधिकारियों को निर्धारित तिथि में प्रस्तुत किए जाते हैं। इन अनुमानों का उपयोग आगे फसल पूर्वानुमान उद्देश्यों के लिए किया जाता है। टीआरएस के तहत सैंपल किए गए गांवों को अस्थायी क्षेत्र से इस तरह से चुना जाता है कि देश के पूरे अस्थायी रूप से बसे हुए हिस्से पांच साल की अवधि में कवर हो जाते हैं।

टीआरएस चयनित गांवों में सिंचाई के तहत क्षेत्र के साथ–साथ उच्च उपज देने वाली किस्मों के तहत क्षेत्र को रिकॉर्ड करने का प्रावधन करता है। फसलों के तहत क्षेत्र की गणना की सटीकता और समयबद्धता सुनिश्चित करने के अलावा, योजना के तहत सांख्यिकीय कर्मचारियों को फसल काटने के प्रयोगों के फील्डवर्क का निरीक्षण करने और रिटर्न का समय पर प्रेषण सुनिश्यित करने की आवश्यकता है। यह योजना उत्तर प्रदेश और महाराष्ट्र से शुरू होकर विभिन्न राज्यों में चरणबद्ध तरीके से शुरू की गई है।

4. कृषि सांख्यिकी योजना की रिपोर्टिंग के लिए एक एजंसी की स्थापना (Establishment of an Agency for Reporting Agriculture Statistics):

केरल, उड़ीसा और पश्चिम बंगाल राज्यों में केंद्र और राज्यों दोनों द्वारा उपयोग के लिए 20 प्रतिशत नमूने के आधार पर क्षेत्र अनुमान प्राप्त करने समान उद्देश्यों के साथ टीआरएस के समान एक योजना शुरू की गई थी। यहां भी, यह परिकल्पना की गई थी कि टीआरएस के मामले में पांच साल की अवधि में सभी गांवों के लिए क्षेत्र के आंकड़ों के लिए खेतों की पूरी गणना उपलब्ध होगी।

5. क्षेत्र अनुमान के लिए डेटा संग्रह की प्रणाली (Data Collection System for Area Estimation):

उन राज्यों में जहां भूमि रिकॉर्ड बनाए रखा जाता है (अस्थायी रूप से बसा हुआ) ग्राम लेखाकार एक कृषि वर्ष के लिए प्रत्येक फसल मौसम में फसल क्षेत्र और भूमि उपयोग के आँकड़ों को रिकार्ड करने के लिए एक गांव या गांवों के समूह का प्रभारी होता है। वह खसरा रजिस्टर में क्षेत्र और भूमि उपयोग से संबंधित फसल विवरण दर्ज करने वाला है। गांव के प्रत्येक सर्वेक्षण संख्या के लिए प्रविष्टियां पूरी होने के बाद, विभिन्न फसलों के तहत बोए गए क्षेत्र का एक सार "जिनेश्वर विवरण" तैयार किया जाता है और राजस्व पदानुक्रम में अगले उच्च अधिकारी को भेजा जाता है। प्रत्येक कृषि वर्ष के अंत में भूमि उपयेग क्षेत्र के आँकड़े संकलित किए जाते हैं और सार संबंधित उच्च अधिकारी भेजा जाता है। विभिन्न गाँवों से प्राप्त फसल–वार और भूमि उपयोग–वार क्षेत्र के आँकड़े राजस्व मंडल, तहसील और जिला स्तर पर एकत्रित किए जाते हैं। जिलेवार क्षेत्र के आंकड़े राजय कृषि सांख्यिकी प्राधिकरण (एसएसएसएस) को भेजे जाते हैं, जो आमतौर पर सांख्यिकी ब्यूरो के निदेशक या कृषि निदेशक या भूमि अभिलेख निदेशक होते हैं। राज्य

स्तरीय एकत्रीकरण एसएएसए द्वारा किया जाता है और अर्थशास्त्र और सांख्यिकी निदेशालय(डीईएस), कृषि और सहकारिता मंत्रालय, भारत सरकार को, जो राज्य स्तर और अखिल भारतीय स्तर के अनुमानों को जारी करने के लिए नोडल एजेंसी है, को अग्रेषित किया जाता है।

कृषि सांख्यिकी की समयबद्धता और गुणवत्ता में सुधार के लिए, कृषि और सहकारिता मंत्रालय, भारत सरकार ने टीआरएस और ईएआरएएस की शुरूआत की। टीआरएस के तहत क्षेत्र की गणना एक राज्य में प्रत्येक फसल के मौसम के दौरान 20 प्रतिशत गांवों के यादृच्छिक नमूने में प्राथमिकता के आधार पर पूरी की जानी है। ईएआरएएस को गैर–भूमि रिकॉर्ड राज्यों केरल, उड़ीसा और पश्चिम बंगाल में टीआरएस की अगली कड़ी के रूप में पेश किया गया था। यह योजना हर साल 20 प्रतिशत गांवों को कवर करने के लिए पूर्णकालिक एजेंसी स्थापित करने का प्रावधान करती है ताकि एक राज्य के सभी गांवों को 5 वर्षो में कवर किया जा सके। इस योजना के तहत नमूना गांवों में फसल क्षेत्र की पूरी गणना के आधार पर रिपोर्ट की जानी है।

6. उपज अनुमान के लिए डेटा संग्रह की प्रणाली (Data Collection System for Yield Estimation):

आमतौर पर फसल आकलन सर्वेक्षण के लिए अपनाया गया नमूना डिज़ाइन स्तरीकृत बहु–चरण यादृच्छिक नमूनाकरण में से एक है जिसमें तहसीलों/तालुकों/राजस्व निरीक्षक मंडलों/ब्लॉकों/आंचलों आदि के साथ स्तर और राजस्व गांव के रूप में नमूनाकरण के पहले चरण इकाई के रूप में एक स्तर के भीतर होता है, दूसरे चरण में नमूना इकाई के रूप में प्रत्येक चयनित गांव के भीतर सर्वेक्षण संख्या/क्षेत्र और नमूनाकरण की अंतिम इकाई के रूप में एक निर्दिष्ट आकार और आकार प्रयोगात्मक भूखंड। प्रत्येक चयनित प्राथमिक इकाई में आम तौर पर फसल काटने के प्रयोगों के संचालन के लिए प्रयोगात्मक फसल उगाने वाले दो सर्वेक्षण संख्या/क्षेत्र चुने जाते हैं। हालांकि दादरा और नगर हवेली में दो के बजाय तीन क्षेत्रों का चयन किया गया है।

आम तौर पर, एक प्रमुख जिले में फसल के लिए 80–120 प्रयोग किए जाते हैं जहां एक जिले को किसी फसल के लिए प्रमुख माना जाता है यदि जिले में फसल के तहत क्षेत्र 80,000 हेक्टेयर से अधिक है या 40,000 और 80,000 हेक्टेयर के बीच है लेकिन राज्य के प्रति जिले औसत क्षेत्र से अधिक है। अन्यथा, दी गई फसल के लिए जिले को अवयस्क माना जाता है। छोटे जिलों में प्रयोग इतने समायोजित हैं कि अनुमानों की सटीकता काफी अधिक है और फील्ड स्टाफ पर काम का बोझ प्रबंधनीय है। एक छोटे से जिले में औसतन लगभग 44 या 46 प्रयोगों की योजना है। किसी जिले को आबंटित किए गए प्रयोगों की संख्या को जिले के भीतर स्तरों के बीच मोटे तौर पर उस परत में फसल के तहत क्षेत्र के अनुपात में वितरित किया जाता है। आम तौर पर, अधिकांश राज्यों में अधिकांश फसलों के लिए $5m \times 5m$ आकार के भूखंड में फसल की कटाई की जाती है। हालांकि, यू पी में भूखंड का आकार 10 मीटर आकार के एक

समबाहु त्रिभुज का है और पश्चिम बंगाल में 1.745 त्रिज्या का एक गोलाकार भूखंड फसल काटने के लिए लिया जाता है। औसत उपज कटाई, थ्रेसिंग, तौल और चयनित भूखंडों से उपज के वनज को रिकॉर्ड करने के बाद प्राप्त की जाती है। प्रयोगों के एक उप–नमूने में फसल की प्रकृति के आधार पर सूखें अनाज या उपज के विपणन योग्य अनाज की प्रतिशत वसूली निर्धारित करने के लिए कटाई की गई उपज की आगे की प्रक्रिया की जाती है। तीन गैर–भूमि रिकॉर्ड राज्यों यानी केरल, उड़ीसा और पश्चिम बंगाल के मामले में प्रति प्रतिदर्श सर्वेक्षण के आधार पर क्षेत्र और उपज दोनों का अनुमान लगाया जाता है। क्षेत्र की गणना के उद्देश्य से चयनित प्राथमिक इकाइयों के उप–नमूने में फसल काटने के प्रयोगों की योजना बनाई गई है। प्रतिदर्श इकाइयों के चयन की सामान्य प्रक्रिया अन्य राज्यों की तरह विभिन्न चरणों में समान रहती है।

केरल ब्लॉक/नगर निगम या नगर पालिकाओं में 10 वर्ग किमी के क्षेत्र और इससे ऊपर को अलग स्तर के रूप में माना जाता है। वर्ग किमी से कम क्षेत्रफल वाली नगरपालिओं को निकटवर्ती ब्लॉकों के साथ विलय कर दिया जाता है और एक ही स्तर के रूप में माना जाता है। इन ब्लॉकों को एक ब्लॉक के क्षेत्र, भूमि की प्रकृति आदि के आधार पर कई अनवेषक क्षेत्रों में विभाजित किया गया है। नगर निगम क्षेत्र को तीन अन्वेषक क्षेत्रों में विभाजित किया गया है। 10 वर्ग किमी से अधिक क्षेत्रफल वाली प्रत्येक नगरपालिका को एकल अन्वेषक के रूप में माना जाता है। प्रत्येक अन्वेषक क्षेत्र में किए गए फसल काटने के प्रयोगों की संख्या एक कृषि वर्ष में धान के लिए छह, नारियल और केले के लिए तीन–तीन और टैपिओका, सुपारी, काजू, काली मिर्च, केला और कटहल के लिए दो–दो है। एक अलग अन्वेषक क्षेत्र वाले नगरपालिका क्षेत्र में धान के संबंध में प्रति सीज़न 10 फसल काटने के प्रयोग और नारियल के लिए 5 प्रयोग प्रति वर्ष किए जाते हैं। नगर निगम क्षेत्रों के लिए एक अन्वेषक क्षेत्र में प्रति सीज़न धान के लिए छह और नारियल के लिए पांच प्रयोग प्रति वर्ष किए जाते हैं।

7. फसल पूर्वानुमान के लिए प्रणाली (System for Crop Forecasting):

फसल क्षेत्र और उत्पादन के अग्रिम अनुमान प्रमुख खाद्य और गैर–खाद्य फसलों (खाद्य अनाज, तिलहन, गन्ना, फाइबर आदि) के संबंध में जारी किए जाते हैं, जो लगभग 87 प्रतिशत कृषि उत्पादन को कवर करते हैं। चार पूर्वानुमान जारी किए जाते हैं, पहला सितंबर के मध्य में, दूसरा जनवरी में, तीसरी मार्च में और चौथा मई के अंत तक।

सितम्बर में जारी किए गए अग्रिम खरीफ फसलों से संबंधित हैं, जो ज्यादातर क्षेत्र के अधिकारियों के दृष्य अवलोकन के आधार पर राज्यों द्वारा प्रस्तुत रिपोर्ट पर आधारित है। दूसरा पूर्वानुमान जो खरीफ और रबी दोनों को कवर करता है और जनवरी में कृषि इनपुट, कीटों और बीमारियों की घटनाओं, क्षेत्र कवरेज, खड़ी फसलों की स्थिति आदि के संबंध में राज्य सरकार से साप्ताहिक रिपोर्ट सहित विभिन्न स्त्रोतों से प्राप्त अतिरिक्त जानकारी को ध्यान में रखते हुए जारी किया गया था। वर्तमान में अनुमान इस स्तर पर रिमोट सेंसिंग के माध्यम से प्राप्त किए गए कार्यों पर भी विचार किया जाता है। तीसरा पूर्वानुमान, जो मार्च में

किया जाता है, खरीफ और रबी मौसम के अनुमानों को बाजार खुफिया इकाइयों, मौसम विभाग और फसल मौसम निगरानी समूह (सीडब्ल्यूडब्ल्यूजी) जैसे स्त्रोतों से प्राप्त जानकारी के आधार पर संशोधित किया जाता है। मई के अंत तक किया गया पूर्वानुमान राज्य कृषि सांख्यिकी प्राधिकरणों (एसएएसए) द्वारा जीसीईएस के माध्यम से प्राप्त उपज अनुमानों का उपयोग करके आपूर्ति किए गए वास्तविक आंकड़ों पर आधारित है। इन चार पूर्वानुमानों के अलावा, डीईएस, एमओए दिसंबर में अंतिम अनुमान प्रदान करता है। पूरी तरह से संशोधित अनुमान अगले फसल वर्ष में दिसंबर में प्राप्त किए जाते हैं जिसमें सभी विलंबित सूचनाओं को शामिल किया जाता है और अखिल भारतीय फसल आंकड़े जारी किए जाते हैं। राष्ट्रीय फसल पूर्वानुमान केंद्र (एनसीएफसी) की स्थापना कृषि मंत्रालय द्वारा पूर्वानुमान के मौजूदा तंत्र की जांच करने और अधिक वस्तुनिष्ठ तकनीक विकसित करने के उद्देश्य से की गई थी। हालांकि, एनसीएफसी को ठोस सांख्यिकीय तकनीकों पर आधारित अधित वस्तुनिष्ठ तकनीकों और मॉडलों को शामिल करके देश की फसल पूर्वानुमान प्रणाली को मजबूत करने की आवश्यकता है।

8. डेटा संग्रह का समन्वय (Organization of Data Collection):

एनएसएसओ के फील्ड ऑपरेशन डिवाज़न के पास राज्यों में फसल अनुमान सर्वेक्षण (सीईएस) में विश्वसनीय और समय पर अनुमान प्राप्त करने, तकनीकी मार्गदर्शन प्रदान करने ओर समान अवधारणाओं, परिभाषाओं और प्रक्रियाओं को अपनाने के लिए उपयुक्त तकनीक विकसित करने में राज्यों की सहायता करने की समग्र जिम्मेदारी है। यह डिज़ाइन, योजना, कार्यान्वयन के विवरण और सर्वेक्षण के परिणामों की समीक्षा करता है और राज्यों के फील्ड स्टाफ के लिए प्रशिक्षण शिविरों का आयोजन एवं संचालन में खुद को संबद्ध करता है और तकनीकी पर्यवेक्षण के प्राथमिक क्षेत्र के काम में भाग लेता है।

9. डेटा संग्रह का पर्यवेक्षण (Supervision of Data Collection):

फील्डवर्क का पर्यवेक्षण एकत्र किए गए किसी भी बड़े पैमाने पर प्रतिदर्श सर्वेक्षण डेटा की गुणवत्ता सुनिश्चित करने के लिए अनिवार्य हिस्सा है। फसल आकलन सर्वेक्षण के तहत नियोजित फसल कटाई प्रयोगों के पर्यवेक्षण के लिए राज्यों में त्रिस्तरीश दृष्टिकोण अपनाया गया है। जिसमे शामिल है:

1. राज्य कृषि सांख्यिकी प्राधिकरणों (एसएएसए) के सांख्यिकीय कर्मचारियों द्वारा पर्यवेक्षण,
2. विभागीय कर्मचारियों द्वारा पर्यवेक्षण अर्थात् विभागों के पर्यवेक्षी अधिकारियों द्वारा जिनके कार्यकर्ता खेत में फसल काटने के प्रयोगों के संचालन के लिए जिम्मेदार हैं और
3. राष्ट्रीय प्रतिदर्श सर्वेक्षण कार्यालय के एफओडी के तकनीकी कर्मियों द्वारा पर्यवेक्षण।

गोवा, उड़ीसा, पश्चिम बंगाल और केंद्र शासित प्रदेश पांडिचेरी में जहां फील्ड वर्क केवल सांख्यिकी विभाग के कर्मचारियों द्वारा किया जाता था, पर्यवेक्षण केवल सांख्यिकीय कर्मचारियों द्वारा किया जाता था जबकि बिहार, हिमाचल प्रदेश, केंद्र शासित प्रदेशों के मामले में दादरा और नगर हवेली और दमन और दीव के लिए हालांकि यहां अन्य प्राथमिक क्षेत्र की एजेंसिंयां हैं, पर्यवेक्षण केवल राज्य सांख्यिकीय कर्मचारियों द्वारा

किया गया था। यद्यपि वर्ष 1973–74 (रबी) से सांख्यिकी विंग की स्थापना के बाद से विभिन्न राज्यों में फसल कटाई प्रयोगों के संचालन का पर्यवेक्षण प्रचलन में था।, एनएसएसओ के कर्मचारी फसल काटने के प्रयोगों पर सैंपल जांच कर पर्यवेक्षण में भाग ले रहे हैं। हालांकि विभिन्न राज्यों में फसल कटाई प्रयोगों के संचालन की निगरानी 20 राज्यों और 2 केंद्र शासित प्रदेशों में फसल सांख्यिकी सुधार योजना (आईसीएस) के तहत फसल कटाई के बाद के चरणों में फसल काटने के प्रयोगों की जांच पूर्व–निर्धारित सैंपल में की जाती है। इस योजना के तहत, राज्य के सांख्यिकीय कर्मचारी भी मिलान के आधार पर समान सैंपल जांच करते हैं।

10. रिमोट सेंसिंग और जीआईएस प्रौद्योगिकी के अनुप्रयोग (Application of Remote Sensing and GIS Technology):

भारत में, भारतीय कृषि अनुसंधान परिषद (आईसीएआर) और भारतीय अंतरिक्ष अनुसंधान संगठन (इसरो) ने संयुक्त रूप से 1969 में नारियल में ज़ड़–विल्ट रोग की पहचान के लिए पहला बहु–वर्णक्रमीय वायुजनित अध्ययन किया।

रिमोट सेंसिंग प्रौद्योगिकियों के अनुप्रयोगों से संबंधित देश स्तरीय अध्ययन आईआरएस–आईए उपग्रह के प्रक्षेपण के बाद शुरू किए गए थे। फसल क्षेत्र और उत्पादन अनुमान (सीएपीई) इस दिशा में गेहूं, चावल, कपास, मूंगफली, ज्वार और सरसों के तहत फसल क्षेत्र के आकलन के लिए महत्वपूर्ण परियोजनाओं में से एक था। इस राष्ट्रीय स्तर की परियोजनाओं के अलावा, अंतरिक्ष विभाग द्वारा कई छोटे अध्ययन कृषि और ग्रामीण विकास के विभिन्न क्षेत्रों में उपग्रह डेटा के अनुप्रयोग के लिए कार्यप्रणाली विकसित करने के लिए किए गए हैं। इनमें से कुछ अध्ययन डधवाल एट अल द्वारा किए गए हैं (1985, 1991), आदि...भारतीय कृषि सांख्यिकी अनुसंधान संस्थान (आईएएसआरआई), नई दिल्ली में फसल क्षेत्र और उत्पादन के आकलन से संबंधित कइ पद्धतिगत अध्ययन किए गए हैं। सिंह एट अल (1992) ने सामान्य फसल आकलन सर्वेक्षण के लिए फसल क्षेत्र के स्तरीकरण के लिए उपग्रह डेटा का इस्तेमाल किया और फसल उपज का अधिक स्टीक अनुमानक प्राप्त किया। सिंह एट अल (1999) ने फसल उपज के छोटे क्षेत्र के अनुमानक को भी विकसित किया। सिंह एट अल (2002) एक विश्वसनीय फसल उपज मॉडल विकसित करने के लिए उपग्रह डेटा और किसानां की आंखों के अनुमान का इस्तेमाल किया। राय एट अल (2004) द्वारा स्थानिक मॉडल का उपयोग करके भूमि उपयोग के आंकड़ों के आकलन के लिए रिमोट सेंसिंग और जीआईएस प्रौद्योगिकी के अनुप्रयोग का पता लगाया गया है। अब, समय पर राष्ट्रव्यापी और बहु–फसल की विश्वसनीय पूर्वाभास की आवश्यकताओं को पूरा करने के लिए कृषि मंत्रालय के राष्ट्रीय फसल पूर्वानुमान केंद्र (एनसीएफसी) के तहत अंतरिक्ष, कृषि–मेट्रोलॉजी और भूमि आधारित अवलोकन (एफएएसएएल) का उपयोग करके कृषि उत्पादन का पूर्वानुमान लगाने की एक परियोजना शुरू की गई है। एक परियोजना

आईएएसआरआई, नई दिल्ली, अंतरिक्ष अनुप्रयोग केंद्र (एसएसी) अहमदाबाद और उत्तर–पूर्वी अंतरिक्ष अनुप्रयोग केंद (एनईएसएसी) शिलांग के द्वारा संयुक्त रूप से मेघालय राज्य के आर्थिक और सांख्यिकी निदेशालय के सहयोग से रिमोट सेंसिंग प्रौद्योगिकी, जीआईएस और क्षेत्र सर्वेक्षण के एकीकरण द्वारा क्षेत्र की फसलों के उप्तादन और क्षेत्र के अनुमान की संभावना का पता लगाने के लिए शुरू की गई है। इन सभी अध्ययनों के परिणाम बहुत उत्साहजनक हैं और यह इंगित करते हैं कि भविष्य में रिमोट सेंसिंग और जीआईएस के पास देश के क्षेत्र और उत्पादन आंकड़ों की गुणवत्ता में सुधार करने के लिए एक महान संभावित क्षमता है।

11. फसल सांख्यिकी में सुधार (आईसीएस) (Improved Crop Statistics):

राज्य सरकार द्वारा विकसित फसल क्षेत्र के अनुमानों के अलावा, राष्ट्रीय प्रतिदर्श सर्वेक्षण (एनएसएस) अपने नियमित दौर के सर्वेक्षणों के दौरान राष्ट्रीय प्रतिदर्श सर्वेक्षणों के आधार पर क्षेत्र अनुमान विकसित करने के लिए उपयोग करता है। पूरे देश के लिए और कुछ जनसंख्या क्षेत्रों के लिए भी अनुमान प्राप्त किए गए थे। फसल क्षेत्र के आंकड़ों पर डेटा की दो श्रृंखलाओं के बीच महत्वपूर्ण अंतर हुआ करता था। इन उच्च अंतरों की जांच के लिए 1963 में फसल सांख्यिकी पर एक तकनीकी समिति का गठन किया गया था। समिति ने अन्य बातों के साथ–साथ पूर्ण गणना के आधार पर अनुमानों का समर्थन किया। परिणामस्वरूप एनएसएस ने 1970–71 में घरेलू सर्वेक्षणों के तहत अपने भूमि उपयोग सर्वेक्षण और फसल काटने के प्रयोगों को भी बंद कर दिया। इसके बाद, एनएसएसओ ने 1973–74 में केंद्र और राज्य प्राधिकरणों के संयुक्त प्रयासों के माध्यम से आंकड़ों की गुणवत्ता में सुधार लाने के उद्देश्य से आईसीएस योजना शुरू की। वर्तमान में यह योजना 20 राज्यों और दो केंद्र शासित प्रदेशों दिल्ली और पांडिचेरी में लागू है। इस योजना में एक स्वतंत्र एजेंसी (एनएसएसओ) टीआरएस सैंपल गांवों में से प्रत्येक में पांच सर्वेक्षण संख्या के चार समूहों के उप–नमूना में गिरदावरी का पर्यवेक्षण और भौतिक सत्यापन करती है। प्रतिदर्श में चयनित सर्वेक्षण संख्याओं में से प्रत्येक के लिए ग्राम लेखाकार द्वारा पूरी की गई पर्यवेक्षक और गिरदावरी की प्रविष्टियों के बीच विसंगतियों की सीमा के लिए एक आकलन किया जाता है। पर्यवेक्षक गाँव के फसल सार की जाँच एकत्रीकरण की संभावित त्रुटियों की जांच के लिए करते हैं, जो पटवारी द्वारा तैयार किया जाता है। स्थायी रूप से बसें राज्यों को भी इस योजना के तहत कवर किया जाता है जहां अस्थायी रूप से बसे हुए क्षेत्र के लिए अपनाई गई पद्धति के अनुसार ईएआरएएस सैंपल गांवों (सर्वेक्षण संख्या) के उप–नमूने की जांच की जाती है। आम तौर पर, कुल 10,000 नमूना गांव आईसीएस द्वारा कवर किए जाते हैं, जिनमें से 8,500 अस्थायी रूप से बसे राज्यों में और 1500 स्थायी रूप से बसे राज्यों में हैं। राष्ट्रीय प्रतिदर्श सर्वेक्षण कार्यालय (एनएसएसओ) पर्यवेक्षण के लिए पूर्णकालिक फील्ड स्टाफ को नियुक्त करके आईसीएस की

योजना और संचालन के लिए मुख्य रूप से जिम्मेदार है। क्षेत्र पर्यवेक्षण की जिम्मेदारी नामित राज्य एजेंसियों द्वारा साझा की जाती है जो लगभग आधे सैंपल वाले गांवों में क्षेत्र पर्यवेक्षण करने के लिए जिम्मेदार है।

11.1 आईसीएस योजना से उभरने वाले प्रमुख मुद्दे (Main Issues emerged in ICS Plan)

1. पटवारी द्वारा प्रस्तुत फसल विवरण कई बार अधूरी गिरदावरी पर आधारित होते हैं।
2. गांव के फसल विवरण समय पर जमा नहीं किए जाते हैं और बड़ी संख्या में गैर–प्रतिक्रिया होती है।
3. .गिरदावरी में प्रविष्टियां कम से कम एक तिहाई सर्वेक्षण संख्या के लिए सही नहीं है।
4. मिश्रित फसल के तहत रिकॉड़िंग क्षेत्र त्रुटियों का एक प्रमुख स्रोत है क्योंकि यह सभी राज्यों में एक समान नहीं है।
5. कभी–कभी फसल के अंतर्गत बोए गए क्षेत्र या कटाई के क्षेत्र के रूप में रिकॉड़िंग क्षेत्र की अनिश्चितता होती हैं। इससे क्षेत्र का गलत अनुमान होता है, यदि बोए गए क्षेत्र को फसल के तहत क्षेत्र के रूप में दर्ज किया जाता है और उम्मीद के मुताबिक अंकुरण नहीं होता है।
6. एक से अधिक बार बोया गया क्षेत्र विभिन्न फसलों के अंतर्गत क्षेत्र के आंकड़ों के बारे में कुछ भ्रम के लिए भी जिम्मेदार है।
7. माप में मेड़, मेड़ को शामिल (Inclusion of field ridges, bunds in measurements) करने से भी सटीकता प्राप्त होती है, जो कुछ मामलों में अधिक हो सकती है।
8. नई तकनीक/किस्मों की शुरूआत के कारण कम अवधि की फसले उगाई जाती हैं और साथ ही, मूल्य वर्धित फसलों की ओर फसल पैटर्न में बदलाव होता है जो गिरदावरी में ठीक से परिलक्षित नहीं होती है।
9. यह देखा गया है कि राज्य सरकार द्वारा अनुमोदित फील्ड कर्मचारी निर्धारित प्रक्रियाओं का कड़ाई से पालन नहीं करते हैं और इस प्रकार सर्वेक्षण अनुमान विभिन्न प्रकार की गैर–नमूना त्रुटियों के अधीन हैं।
10. त्रुटियां मुख्य रूप से क्षेत्रों के गलत चयन और चयनित प्रयोगात्मक भूखंडों की अवधि के कारण पेश की जाती हैं। उचित तौल मशीन जैसे दोषपूर्ण उपकरणों का उपयोग माप त्रुटियों की काफी मात्रा का परिचय देता है।
11. राज्य के राजस्व ओर कृषि विभाग, जो सर्वेक्षण करने के लिए जिम्मेदार हैं, इन कार्यक्रमों को कम प्राथमिकता पर रखते हैं और क्षेत्र संचालन के पर्यवेक्षण और नियंत्रण अपर्याप्त स्तर का है।राज्यों

में "कृषि सांख्यिकी पर उच्च स्तरीय समन्वय समिति" का डेटा की गुणवत्ता में सुधार पर बहुत कम प्रभाव पड़ता है।

12. ब्लॉक/ग्राम पंचायत स्तर पर विशेष रूप से फसल बीमा उद्देश्यों के लिए अनुमान प्राप्त करने की आवश्यकताओं को पूरा करने के लिए कुछ राज्यों ने फसल काटने के प्रयोगों की संख्या में काफी वृद्धि की है।यह फील्ड एजेंसी पर भारी बोझ डालता है, गैर–नमूनाकरण त्रुटियों में काफी वृद्धि करता है, जिसके परिणामस्वरूप जीसीईएस (GCES) के माध्यम से एकत्र किए गए डेटा की गुणवत्ता में गिरावट आती है। बीमाकृत किसानों के स्थानीय दबाव के कारण फसल बीमा के मामले में उपज दरों के कम अनुमान की संभावना रहती है, जहां फसल की उपज को कम करने में रूचि रखते हैं।

13. यह अतीत में एक बड़ी बहस का विषय रहा है क्योंकि विभिन्न स्रोतों/एजेंसियों द्वारा प्राप्त उत्पादन आंकड़ें काफी अलग हैं।

14. फसल काटने के प्रयोगों के संचालन के लिए फील्ड स्टाफ को अपर्याप्त प्रशिक्षण प्रदान किया जाता है।

15. एक अन्य महत्वपूर्ण कारक, जो उत्पादन के आंकड़ों की गुणवत्ता को प्रभावित करता है, कुछ राज्यों में खरीफ मौसम में कुछ फसलों के लिए देर से समय सारणी निर्धारित की जाती है। इस मामले में जल्दी कटाई के कारण मौसम खत्म होने से पहले फसल काटने के प्रयोग किए जाने चाहिए।

16. गुजरात, हरियाणा, कर्नाटक और मध्य प्रदेश जैसे राज्यों में खरीफ फसलों जैसे मक्का, ज्वार, बाजरा, मूगफली, कपस, सोयाबीन आदि के संबंध में ऐसी स्थिति उत्पन्न हो रही है। इन फसलों की जल्दी कटाई के कारण, फसल के तहत क्षेत्र आमतौर पर कम रिपोर्ट किया जाता है और इसलिए उत्पादन भी होता है। उपयुक्त पूर्वानुमान मॉडल विकसित करने की सख्त आवश्यकता है जो फसल उप्तादन से संबंधित मापदंडों जैसे फसल की स्थिति, कृषि मौसम विज्ञान, पानी की उपलब्धता आदि पर विभिन्न स्रोतों से जानकारी को एकत्रित करता है।

17. कोई बहु–आयामी मॉडल मौजूद नहीं है जिसमें विभिन्न स्रोतों से प्राप्त जानकारी को एकीकृत किया जा सके।

18. विभिन्न उप्तादन एजेंसियों से सूचना का प्रवाह समयबद्ध और उपयुक्त नहीं है।

19. डीईएस, एमओए उप्तादन के आंकड़ों में लगातार बदलाव के कारण उपयोगकर्ता समूह का विश्वास खो रहा है, विशेष रूप से पूर्वानुमानित अनुमानों में ज्यादातर अंतर बहुत बड़ा होता है। ये अंतर उपयोगकर्ता के बीच भ्रम और संदेह पैदा करता है।

20. वर्तमान तकनीक ज्यादातर व्यक्तिपरक है और ध्वनि सांख्यिकीय तकनीक पर आधारित नहीं है।

संदर्भ (References):

1. बसु, डी. (1969). रोल आफ सफीशियन्सी एण्ड लाईकलीहुड प्रिंसिपलस इन सैंपल सर्वे थ्योरी. संख्या, 31, 441–454.
2. बसु, डी. (1971). एन ऐस्से ऑन द लॉजिक्ल फाउंडेशन ऑफ सर्वे सैंपलिंग, एपार्ट आई. इन फाउंडेशन ऑफ स्टैटिस्टिकल इंफ्रेंस, होल्ट, राइनहास्ट और विंस्टन, टोरंटो, 203–242.
3. बाउले, ए.एल. (1906). एड्रेस टू द इकॉनोमिक साइंस एण्ड स्टेटिस्टिकस सेक्शन ऑफ द ब्रिटिश एसोसिऐशन फॉर द एडवांसमेंट ऑफ साइंस. जे. रॉय. स्टैटिस्टि. सोसायटी., 69, 548–557.
4. बाउले, ए.एल. (1906). मैजरमेंट आफ द प्रीसिज़न अटेंड इन सैम्पलिंग. बुल.इंट. स्टैटिस्टि.इंस्ट., 22 लीवर आई.
5. ब्रेवर, के.आर. डब्लयू. (1963). ए मॉडल आफ सिस्टेमेटिक सैंपलिंग विद अनइक्वल प्रोबेबिलिटिज़, ऑस्ट्रल.जे. स्टैटिस्टि., 5, 3–5.
6. कैसल, सी.एम., सरंडल, सी.ई., एण्ड रिटमैन, जे.एच. (1977). फाउंडेशनस ऑफ इंफ्रेंस इन सर्वे सैंपलिंग. न्यूयॉर्कः विली.
7. कोचरन, डब्ल्यू.जी. (1942). सैंपलिंग थ्योरी वैन द सैंपलिंग यूनिटस आर ऑफ अनइक्वल साईज़िज.' जे.आमेर. स्टैटिस्टि. एसो., 37, 199–212.
8. कोचरन, डब्ल्यू.जी. (1953). सैंपलिंग टेकनीकस, 1 एडी., 2 एडी., 3 एडी., (1977). न्यूयॉर्कः विली.
9. कोचरन, डब्ल्यू.जी. एण्ड वाटसन, डी.जे. (1936). एन एक्सपेरीमेंट आन ऑब्ज़रवरस बायस इन द सलेक्शन आफ शूट–हाईटस. एम्पायर जे. एक्सप. एग्रीकलचर,4, 69–76.
10. डेटनियस, टी. (1962). रिसेंट एडवांसिस इन सैंपल सर्वे थ्योरी एण्ड मैथ्डस. एन.मैथ. स्टैटिस्टि. 33, 325–349.
11. डेमिंग, डब्ल्यू. ई., (1950). सम थ्योरी आफ सैंपलिंग. न्यूयॉर्क. विली.
12. गोडाम्बे, वी.पी. (1955). ए यूनिफाईड थ्योरी आफ सैंपलिंग फ्रोम फिनिट पोपुलेशनस. जे.रॉय. स्टैटिस्टि. सोसा. बी, 17, 269–278.एए
13. हैनसेन, एम.एच., हर्विट्ज़, डब्ल्यू.एन. एण्ड मैडो, डब्ल्यू.जी. (1953). सैंपल सर्वे मैथ्डस एण्ड थ्योरी, जॉन विली एण्ड सन्स, न्यूयॉर्क, वोल्यू. I एण्ड II.
14. होर्विट्ज़, डी.जी. एण्ड थॉम्ससन, डी.जे. (1952). ए जनरलाईजेशन आफ सैंपलिंग विदाउट रिपलेस्मेंट फ्रोम ए फिनिट यूनिवर्स. जे. आमेर. स्टैटिस्टि. एसोसि., 47, 663–685.
15. जेसन, आर.जे. (1942). स्टैटिस्टिकल इनवैस्टिगेशन आफ ए सैंपल सर्वे फार ओब्टेनिंग फार्म फ़ेक्टस. आयोवा एग्रीक्लचरल एक्सपेरिमेंटल स्टेशन रिसर्च बुलेटिन न. 304.
16. कियार, ए.एन. (1895–6). ऑब्ज़र्वेशनस आफ एक्सपीरियंसस कंर्सनेंट डीस डिनोमब्रीमेंटस रिप्रसेंटेटिफस, बुल. इंट. स्टैटिस्टि. इंस्ट., 9, लिव. 2, 176–183.
17. मैकार्थी, पी.जे.(1966). स्यूडो–रेपलिकेशनः एन अप्रोच टू द एनालिसिस आफ डेटा फ्रोम कमप्लेक्स सर्वेस. वाशिंगटनः एनसीएचएस सीरीज़ 2, न.14.
18. नारायण, आर.डी. (1951). ऑन सैंपलिंग विदाउट रिप्लेसमेंट विद वेरिंग प्रोबेबिलिटिज़, इंड. सोस. एग्रील.स्टैटिस्टि., 3, 169–174.
19. नेमन, जे. (1951). ऑन द टू डिफरेंट आस्पेक्टस आफ द रिप्रेजेनटेटिव मैथ्डः द मैथ्ड आफ स्ट्रेटिफाईड सैंपलिंग एण्ड द मैथ्ड आफ परपसिव सलेक्शन. जे.रॉय.स्टैटिस्टि. सोस., 97, 555–606.
20. नेमन, जे. (1938). कंट्रीब्यूशन टू द थ्योरी आफ सैंपलिंग ह्युमन पोपुलेशनस. जे. आमेर.स्टैटिस्टि.एसोस., 33, 101–116.

21. ओ'मुइरचेर्टेघ, सी एण्ड वोंग, टी. (1981). द इमपेक्ट आफ सैंपलिंग थ्योरी आन सर्वे सैंपलिंग प्रैक्टिस. बुल.इंट. स्टैटिस्टि. इंस्ट. ILIX बुक 1, 465–493.
22. पैटरसन, एच.डी. (1950). सैंपलिंग आन सक्सेसिव ओकेज़नस विद पार्शियल रिप्लेसमेंट आफ यूनिटस. जे. रॉय स्टैटिस्टि. सोस., बी, 312–255.
23. राव, सी.आर, (1985). इवेल्यूएशन आफ डेटा कलेक्शप सेंससिस. सैंपल सर्वेस एण्ड डिज़ाइन आफ एक्सपेरिमेंट.
24. राव, जे.एन.के. (1963). आन टू सिस्टमस आफ अनइक्वल प्रोबेबिलिटि सैंपलिंग विदाउट रिप्लेसमेंट. एन. इंस्ट. स्टैटिस्टि मैथ्स. 15, 67–72.
25. राव, जे.एन.के. एण्ड स्कॉट, ऐ.जे. (1981). द एनालिसिस आफ कैटेगरिकल डेटा फ्रॉम कम्प्लेकस सैंपल सर्वेसः ची–स्क्वायर टेस्टस फार गुडनेस आफ फिट एण्ड इंडिपेंडेंस इन टू–वे टेबलस. जे. आमेर. स्टैटिस्टि एसोस., 76, 221–230.
26. रॉयल, आर.एम. (1968). एन ओल्ड एप्रोच टू फिनिट पोपुलेशन सैंपलिंग थ्योरी. जे. आमेर. स्टैटिस्टि एसोस., 63, 1269–1279.
27. रॉयल, आर.एम. (1970). आन फिनिट पोपुलेशन सैंपलिंग थ्योरी अंडर सर्टेन लिनियर रिग्रेशन मॉडल्स. बायोमेट्रिका, 57, 377–387.
28. वर्मा, वी., स्कॉट, सी., एण्ड ओ'मुइरचेर्टेघ, सी. (1980). सैंपल डिज़ाइन एण्ड सैंपलिंग एररस फार द वर्ल्ड फर्टिलिटि सर्वे, जे.रॉय. स्टैटिस्टि. सोस.,ए143, 431–473.
29. सुखात्मे, पी.वी. (1959). मेजर डेवलपमेंटस इन द थ्योरी एण्ड एप्लीकेशन आफ सैंपलिंग डयूरिंग द लास्ट टवेंटी फाइव इयरस, एस्टाडिसटिका, 17, 62–679.

अध्याय 08

eLISS ऐप एवं eLISS पोर्टल द्वारा पशुधन आँकड़ों का उत्पादन

प्राची मिश्रा साहू
प्रमुख वैज्ञानिक, प्रतिदर्श सर्वेक्षण प्रभाग
भा.कृ.अनु.प.-भारतीय कृषि साँख्यिकी अनुसंधान संस्थान, लाइब्रेरी एवेन्यू, नई दिल्ली-110012

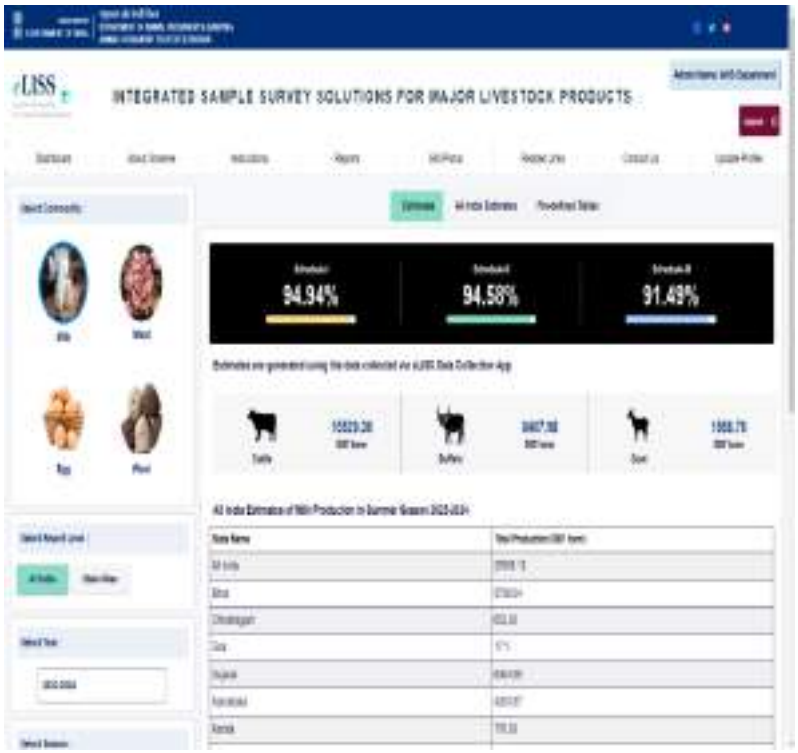
1.0 परिचय (Introduction):

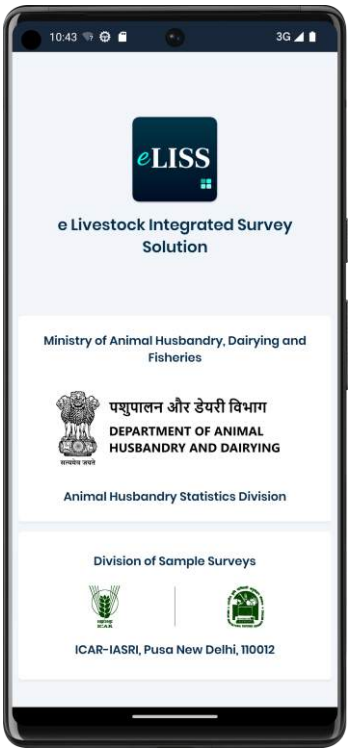
पशुपालन और डेयरी विभाग (DAHD) के पशुपालन सांख्यिकी (AHS) प्रभाग को "पशुपालन संगणना और एकीकृत नमूना सर्वेक्षण" के माध्यम से पशुपालन सांख्यिकी का उत्पन्न करने का काम सौंपा गया है। यह योजना पशुपालन और दुग्ध विभाग द्वारा राज्य पशुपालन विभागों के माध्यम से कार्यान्वित की जाती है।

इस योजना के तहत, चार प्रमुख पशुओं के उत्पाद (MLPs) जैसे कि दूध, अंडा, मांस और ऊन की मात्रा का अनुमान राष्ट्रीय और राज्य दोनों स्तरों पर वार्षिक नमूना सर्वेक्षण नामतः "एकीकृत नमूना सर्वेक्षण" के आधार पर लगाया जाता है। कुल अर्थव्यवस्था में पशुधन क्षेत्र और इसके उत्पादों के योगदान को मापने के लिए यह एक आवश्यक सर्वेक्षण रहा है। ग्रामीण और शहरी दोनों क्षेत्रों में 36 राज्यों/केंद्र शासित प्रदेशों को कवर करते हुए पूरे देश में सर्वेक्षण किया जाता है। सर्वेक्षण की अवधि मार्च से फरवरी तक होती है और एक वर्ष की पूरी अवधि को तीन मौसमों में विभाजित किया जाता है, गर्मी, बरसात और सर्दी

प्रत्येक में चार महीने की अवधि होती है। यह समुच्चय के प्रगतिशील अनुमानों को लाने में मदद करता है और अध्ययन के तहत वर्णों में यदि कोई हो तो मौसमी का भी ध्यान रखता है।

सभी राज्य और केंद्र शासित प्रदेश इस योजना को लागू करते हैं ताकि पशुओं की संख्या और प्रमुख MLPs के उत्पादन के लिए सांख्यिकी उत्पन्न की जा सकें। केंद्र सभी राज्यों और केंद्र शासित प्रदेशों से प्राप्त डेटा को समेकित करता है, ताकि पैन इंडिया आंकड़े जारी किए जा सकें। क्षेत्र स्तर पर जानकारी भरने के लिए कागजी अनुसूचियों का परंपरागत दृष्टिकोण कंप्यूटर सहायित व्यक्तिगत साक्षात्करण (Computer Aided Personal Interviewing, CAPI) विधि द्वारा बदल दिया गया है। डेटा को फ़ील्ड स्तर से सीधे एनुमरेटर द्वारा टैबलेट/हैंडसेट में दर्ज किया जाता है, जिसके लिए भा.कृ.अनु.प.-भा.कृ.सां.अ.सं. ने "इंटीग्रेटेड सैंपल सर्वे फॉर मेजर लिवेस्टॉक प्रोडक्ट्स" परियोजना के तहत eLISS ऐप एवं eLISS पोर्टल का विकास किया।





इसके अतिरिक्त, राज्यों द्वारा एकत्रित यूनिट स्तर का पूर्ण डेटा उपलब्ध नहीं था और उपलब्ध डेटा को डिजिटलाइज़ नहीं किया गया था, जो मौजूदा डेटा प्रबंधन प्रक्रिया में सुधार करने की आवश्यक है। इसलिए, "इंटीग्रेटेड सैंपल सर्वे फॉर मेजर लिवेस्टॉक प्रोडक्ट्स" शीर्षक द्वारा एक परियोजना को भा.कृ.अनु.प.-भा.कृ.सां.अ.सं., नई दिल्ली को सौंपा गया था ताकि जिलों/राज्यों/केंद्र शासित प्रदेशों और

सम्पूर्ण भारत स्तर पर पशुधन संख्या और पशुधन उत्पादन के आकलनों को उत्पन्न किया जा सके और डिजिटल ऑनलाइन एंड-टू-एंड समाधान विकसित किया जा सके, जिसमें नमूना चयन, आवंटन, फ़ील्ड डेटा संग्रहण, फ़ील्ड डेटा पर्यवेक्षण, डेटा विश्लेषण आदि शामिल हैं। इस अध्ययन में, एक eLISS वेब पोर्टल और एक एंड्रॉयड-आधारित eLISS डेटा संग्रहण एप्लिकेशन विकसित किए गए थे जो ISS योजना में किए जाने वाले कार्यों को मॉनिटर करने के लिए उपयोगी है ।

वर्तमान में eLISS वेब पोर्टल https://iss.icar/gov.in पर लाइव है। इस पोर्टल में नमूना चयन मॉड्यूल, नमूना आवंटन मॉड्यूल और डेटा विश्लेषण मॉड्यूल शामिल हैं। पोर्टल के माध्यम से उत्पन्न आकलन को उपयोगकर्ता द्वारा विभिन्न प्रारूपों में और विभिन्न स्तरों पर देखा और डाउनलोड किया जा सकता है। एक एंड्रॉयड-आधारित eLISS डेटा संग्रहण ऐप विकसित किया गया है और गूगल प्ले स्टोर पर उपलब्ध है ताकि क्षेत्र से डेटा को एकत्र किया जा सके और पशुओं के उत्पाद जैसे कि दूध, मांस, अंडा और ऊन उत्पादन का आकलन करने के लिए एक अंत से अंत समाधान प्रदान किया जा सके। इंटीग्रेटेड सैंपल सर्वे (ISS) योजना के सभी आठ अनुसूचियों को ऐप द्वारा विवरणीत किया गया है। eLISS वेब पोर्टल और eLISS डेटा संग्रहण ऐप देश के सभी राज्यों/केंद्र शासित प्रदेशों में डेटा संग्रहण और आकलन के लिए संचालन में हैं।

2.0 पद्धति (Methodology):

सर्वेक्षण राज्यों/केंद्र शासित प्रदेशों क्षेत्रों के संपूर्ण ग्रामीण और शहरी क्षेत्रों में किया जाता है। सर्वेक्षण सभी घरेलू उद्यमों, गैर-घरेलू उद्यम संस्थानों जैसे सभी फार्म हाउस, बूचड़खानों, कसाई दुकानों को शामिल करते हुए चयनित नमूना गांवों/शहरी वार्डों में किया जाता है।

2.1 प्रतिचयन डिजाइन (Sampling Design):

ISS में नमूना डिजाइन एक स्तरीकृत तीन चरण डिजाइन है जिसमें जिला को स्तर के रूप में रखा गया है। नवीनतम पशुधन गणना की सूची में से (जो प्रतिचयन फ्रेम का गठन करती है), पूर्ण गणना और विस्तृत सर्वेक्षण के लिए राज्य में गांवों/शहरी वार्डों के 15% (प्रत्येक मौसम के लिए 5%) का चयन किया जाता है। प्रत्येक जिले (यानी स्ट्रेटम) को उसकी पशुधन आबादी के अनुपात में गांवों/शहरी वार्डों की संख्या आवंटित की जाती है, जिसे निम्नानुसार वर्णित किया गया है (शहरी वार्डों के मामले में समान):

नवीनतम पशुधन जनगणना गांवों की सूची प्रतिचयन फ्रेम का गठन करती है। पशुधन की संख्या के आकलन के लिए, पशुधन और पोल्ट्री आबादी (प्रत्येक मौसम में 5%) की पूर्ण गणना के लिए राज्य में 15% (प्रत्येक मौसम का 5%) गांवों / शहरी वार्डों का चयन दो स्वतंत्र उप-नमूनों के रूप में प्रतिस्थापन के बिना सरल यादृच्छिक नमूनाकरण (एसआरएसडब्ल्यूओआर) का उपयोग करके किया जाता है। प्रत्येक जिले (यानी स्ट्रेटम) में गांवों की संख्या का चयन उसकी पशुधन आबादी के अनुपात के अनुसार आवंटित किया जायेगा । गांवों के इन दो उप-नमूनों की सूची में से, जिला स्तर पर उपज पर जानकारी एकत्र करने के लिए विस्तृत अध्ययन के लिए 4 गांवों और एक-एक शहरी वार्ड के प्रतिनिधि नमूने का चयन किया जाएगा। उपलब्ध स्टाफ संसाधन के आधार पर छोटे राज्यों/केंद्र शासित प्रदेशों में नमूना आकार बढ़ाया जा सकता है जहां गांवों की संख्या के साथ-साथ पशुधन की संख्या काफी कम है।

3.0 eLISS एंड टू एंड सोल्युशन (eLISS End-to-End Solution):

eLISS वेब पोर्टल और eLISS डेटा कलेक्शन ऐप भारत में प्रमुख पशुधन उत्पादों के लिए एक व्यापक एंड-टू-एंड समाधान प्रदान करता है। eLISS एंड-टू-एंड समाधान का उपयोग करते हुए, 48,000 गांवों/शहरी वार्डों में 71 लाख से अधिक परिवारों/उद्यमों का सर्वेक्षण किया गया है, पूरे देश में लगभग 730 बूचड़खानों का सर्वेक्षण किया गया है और 33,000 से अधिक वाणिज्यिक पोल्ट्री फार्मों से डेटा एकत्र किया गया है। वर्तमान में, 22,000 से अधिक प्रगणक, 7,500 पर्यवेक्षक और 730 DNOs eLISS एंड-टू-एंड समाधान पर सक्रिय हैं।

4.0 eLISS वेब पोर्टल (eLISS Web Portal):

eLISS वेब पोर्टल में प्रत्येक जिले में 15% गांवों/शहरी वार्डों का चयन करने के लिए नमूना चयन मॉड्यूल शामिल है (प्रत्येक मौसम यानी गर्मी, सर्दी और बारिश के लिए 5%)। गांवों/शहरी वार्डों के चयन के बाद, चयनित गांवों/शहरी वार्डों को आंकड़ा संग्रहण के लिए प्रगणकों (enumerators) को तथा नमूना आवंटन माड्यूल का उपयोग करके डेटा संवीक्षा के लिए पर्यवेक्षकों/जिला नोडल अधिकारी (supervisors/District Nodal Officer) को आबंटित किया जाता है। eLISS वेब पोर्टल चयनित गांवों/शहरी वार्डों के लिए मुख्यालय/राज्य/संघ राज्य क्षेत्र/जिले को 10-स्तरीय रियल टाइम सर्वेक्षण आँकड़े प्रदान करने में सक्षम है।

इसके अलावा, पोर्टल में डेटा विश्लेषण मॉड्यूल शामिल है जिसका उपयोग करके जिला नोडल अधिकारी द्वारा जिला स्तर के अनुमानों को कैप्चर किया जाता है। पोर्टल द्वारा कैप्चर किए गए जिला स्तर के अनुमानों को इस योजना में अपनाई जा रही आकलन प्रक्रिया के अनुसार राज्य/केंद्र शासित प्रदेशों स्तर पर एकत्र किया जाता हैं।

पोर्टल के माध्यम से उत्पन्न अंतिम अनुमानों को उपयोगकर्ताओं द्वारा विभिन्न प्रारूपों में देखा और डाउनलोड किया जा सकता है। इसके अलावा, अनुमानों को राज्य-वार, जिला-वार, वर्ष-वार, पशु-वार (मवेशी, भैंस, बकरी, भेड़, सुअर, कुक्कुट आदि), समूह-वार (विदेशी/संकर, स्वदेशी/मवेशियों में गैर-डिस्क्रिप्ट आदि), क्षेत्र-प्रकार-वार (ग्रामीण/शहरी) और आउटपुट-चर-वार (उत्पादन, संख्या, औसत उपज) जैसे विभिन्न स्तरों पर भी देखा और डाउनलोड किया जा सकता है। दूध, मांस, अंडा और ऊन से संबंधित जानवरों के उत्पादन और संख्या के आंकड़ों को विषयगत मानचित्रों के रूप में भी दर्शाया गया है।

5.0 eLISS डेटा संग्रहण ऐप (eLISS Data Collection App):

डेटा संग्रह के लिए एक एंड्रॉइड आधारित एप्लिकेशन विकसित किया गया है जो क्षेत्र से डेटा कैप्चर करने के लिए उपयोग किया जाता है जिसे पहले प्रगणकों द्वारा पेपर-आधारित शेड्यूल का उपयोग करके

मैन्युअल रूप से एकत्र किया जाता था। डेटा संग्रह ऐप में ISS योजना की सभी आठ अनुसूचियों को सभी फील्ड और प्रविष्टियों के साथ डिजाइन किया गया है। डेटा संग्रह ऐप में एकत्र किए गए डेटा की बेहतर डेटा गुणवत्ता के लिए कई सत्यापन जांच शामिल हैं। यह डेटा संग्रह ऐप प्रतिचयन फ्रेम के रूप में अनुसूची-II में कैप्चर किए गए परिवारों / उद्यमों की सूची का उपयोग करके दूसरे चरण का नमूना भी तैयार करता है। इस ऐप के माध्यम से कैप्चर किए गए डेटा को प्रगणक (enumerator) द्वारा सर्वर से सिंक किया जा सकता है। प्रगणक द्वारा एकत्र किए गए आंकड़ों को पर्यवेक्षक (supervisor) और जिला नोडल अधिकारी स्तर पर सत्यापित किया जा सकता है जिसे राज्यों/केंद्र शासित प्रदेशों द्वारा देखा जा सकता है। इसके अलावा, पर्यवेक्षक और जिला नोडल अधिकारी किसी गांव/शहरी वार्ड के शेड्यूल के लिए प्रश्न उठा सकते हैं, यदि कुछ सुधार की आवश्यकता होती है।

5.1 डेटा संग्रह (Data Collection):

eLISS डेटा संग्रह ऐप में डेटा संग्रह की शुरुआत प्रगणक (Enumerator) द्वारा आवंटित गांवों / शहरी वार्डों के लिए अनुसूचियों को भर कर की जाती है। निम्नलिखित आठ अनुसूचियों का विवरण है जिनके उपयोग से डेटा एकत्र किया जाता है:

अनुसूची I (Schedule I): सामान्य जानकारी (General information)

अनुसूची II (Schedule II): पूर्ण गणना (Complete enumeration)

अनुसूची III (Schedule III): चयनित परिवारों/उद्यमों में दूध की पैदावार का विवरण (Details of milk yield in selected households/enterprise)

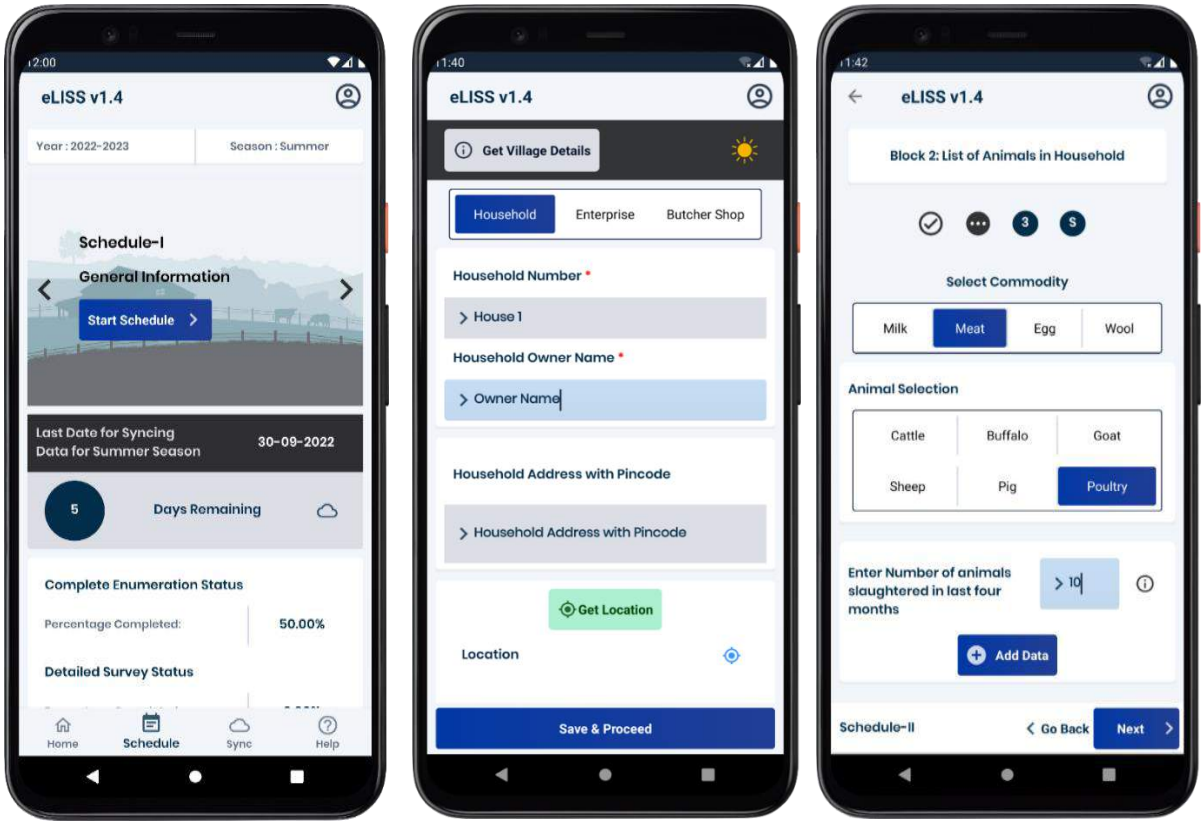
अनुसूची IV (Schedule IV): चयनित परिवारों/उद्यमों में अंडा उत्पादन का विवरण (Details of egg production in selected households/enterprises)

अनुसूची V (Schedule V): वाणिज्यिक पोल्ट्री फार्मों में अंडे के उत्पादन का विवरण (Details of Egg production in the commercial poultry farms)

अनुसूची VI (Schedule VI): नमूना घरों / उद्यमों में ऊन उत्पादन का विवरण (Details of wool production in sample households/enterprise)

अनुसूची VII (Schedule VII): मान्यता प्राप्त बूचड़खानों से मांस की उपज के बारे में जानकारी (Information on yield of meat from recognized slaughter houses)

अनुसूची VIII (Schedule VIII): वाणिज्यिक पोल्ट्री फार्मों में ब्रायलर और परतों के मांस उत्पादन का विवरण (Details of Broilers and Layers' Meat production in the commercial poultry farms)



6.0 अनुमानों का सृजन (Generation of Estimates):

जिला/राज्य और सम्पूर्ण भारतीय स्तर पर पशुओं की संख्या प्राप्त करने और दूध, मांस, अंडा और ऊन के अनुमान उत्पादन के लिए डेटा विश्लेषण मॉड्यूल विकसित किया गया है। यह डेटा विश्लेषण मॉड्यूल पिछले वर्षों के साथ परिणामों की तुलना करने की सुविधा भी प्रदान करता है। इसके अलावा, यह डेटा विश्लेषण मॉड्यूल जानवरों की वर्षवार संख्या और दूध, मांस, अंडा और ऊन के उत्पादन से संबंधित विभिन्न चार्ट और ग्राफ तैयार भी करता है।

7.0 निष्कर्ष (Conclusion):

इस अध्ययन में, भारत में पशुधन के आंकड़े उत्पन्न करने के लिए एक एंड-टू-एंड समाधान विकसित किया गया है जिसमें eLISS पोर्टल और eLISS डेटा संग्रह ऐप शामिल हैं। यह पोर्टल समय पर सटीक और विश्वसनीय अनुमानों का निर्माण सुनिश्चित करता है। यह जिला/राज्य और राष्ट्रीय स्तर पर एकल रिपॉजिटरी में जानवरों के विभिन्न श्रेणियों का विशाल डेटाबेस भी रखता है। घरेलू स्तर पर यूनिट स्तर का डेटा भी डेटाबेस में संग्रहीत किया जाता है और इसे एक क्लिक के साथ एक्सेस किया जा सकता है। पशुओं की संख्या और दूध, मांस, अंडा और ऊन के उत्पादन के लिए अनुमान तैयार करने के अलावा, यह प्रणाली ISS अनुसूची के विभिन्न अन्य मापदंडों का उपयोग करने में सक्षम है। डेटा का यह दोहन विभिन्न व्युत्पन्न आउटपुट उत्पन्न करता है जिसका उपयोग पशुधन क्षेत्रों के सतत विकास के लिए महत्वपूर्ण नीतिगत निर्णय लेने के लिए किया जा सकता है।.

- **eLISS वेब पोर्टल लाइव:** https://iss.icar/gov.in

अध्याय 09

QGIS का परिचय

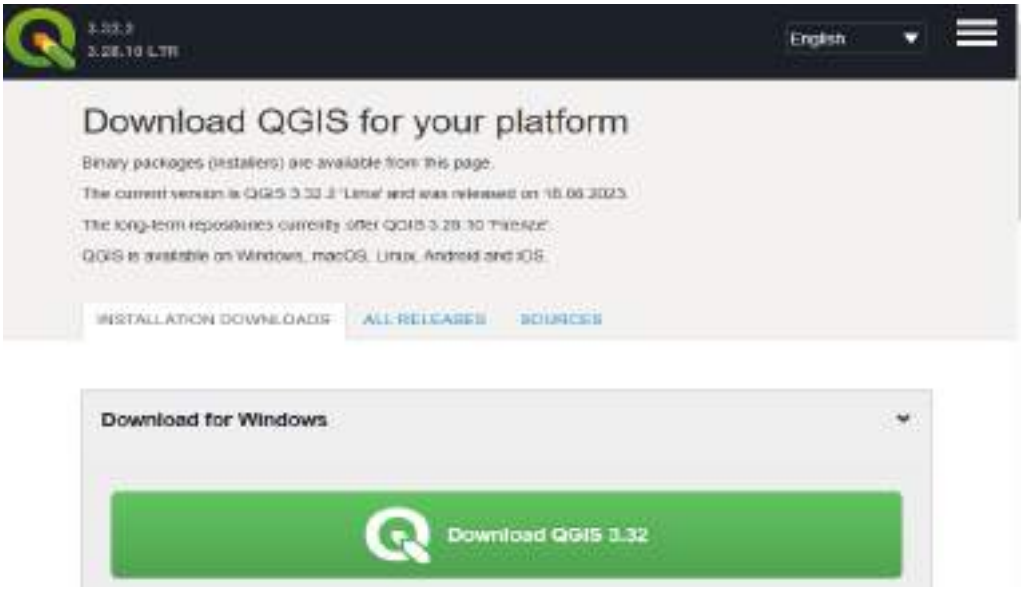
भारती
वैज्ञानिक, प्रतिदर्श सर्वेक्षण प्रभाग
भा.कृ.अनु.प.-भारतीय कृषि साँख्यिकी अनुसंधान संस्थान, लाइब्रेरी एवेन्यू, नई दिल्ली-110012

1. प्रस्तावना (Introduction):

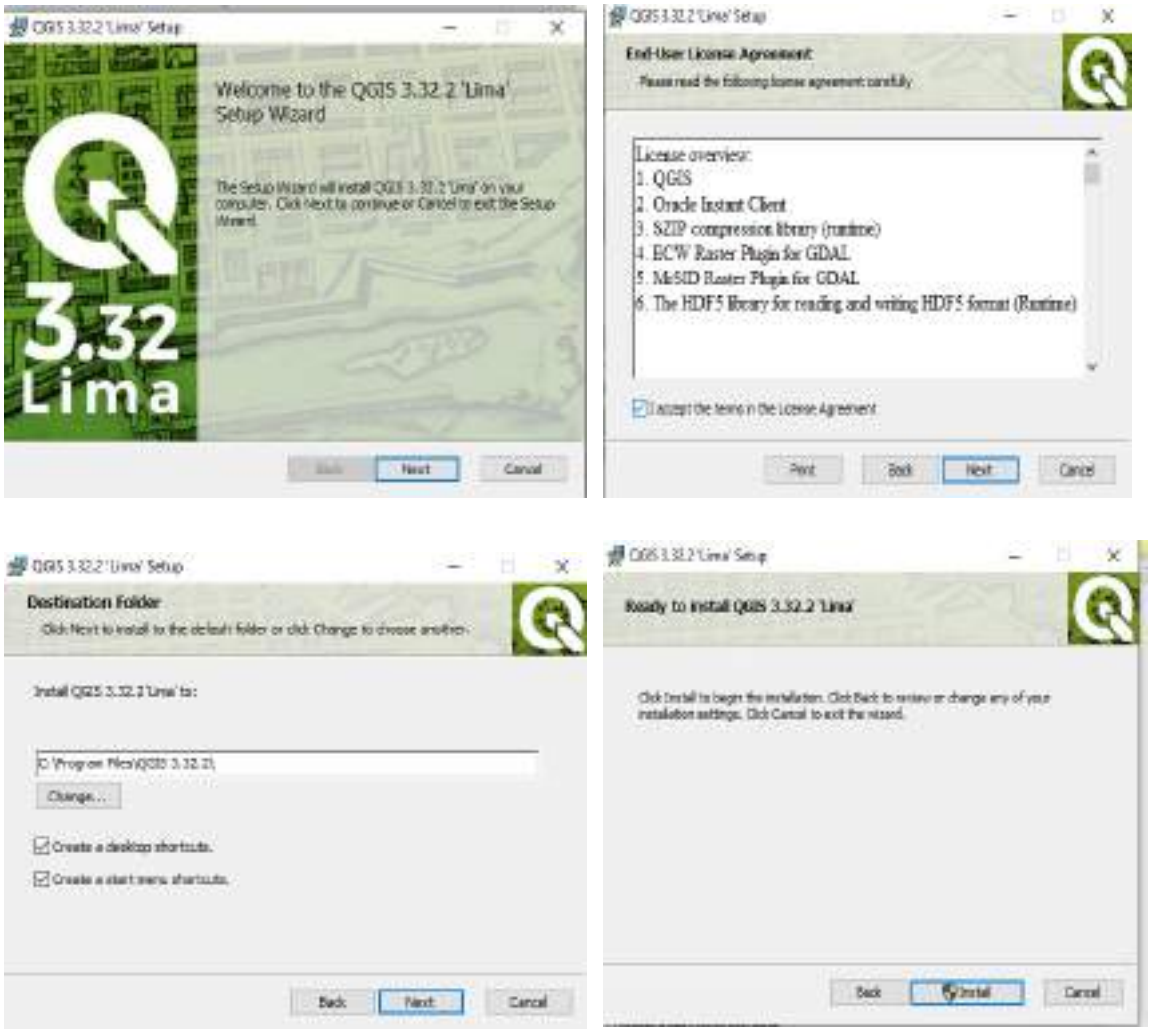
QGIS (Quantum GIS) एक मुफ्त और ओपन रिसोर्स जीओस्पेशियल सॉफ़्टवेयर है जिसका उपयोग जोग्राफिक डेटा को विश्लेषण और प्रबंधन करने के लिए किया जाता है। यह एक GIS (भूगोलीय सूचना प्रणाली) सॉफ़्टवेयर है जिसका मतलब है कि यह स्थानीय और ग्लोबल स्थानीय डेटा के साथ काम करने में मदद करने के लिए डिज़ाइन किया गया है। QGIS वेक्टर और रास्टर डेटा को समर्थित करता है, जिससे ग्राफिकल और स्थानीय डेटा का विश्लेषण किया जा सकता हैं। QGIS में विभिन्न प्लगइन्स उपलब्ध होते हैं जिनकी मदद से आवश्यकताओं के अनुसार सॉफ़्टवेयर की विशेषताओं को बढ़ावा दे सकते हैं। QGIS विभिन्न कोऑर्डिनेट सिस्टम्स का समर्थन करता है, जिससे विभिन्न स्थानों के डेटा को सही रूप से प्रदर्शित कर सकते हैं। QGIS की मदद से इंटरएक्टिव मैप्स बनाये जा सकते हैं । QGIS में मेटाडेटा को प्रबंधित करने के लिए टूल्स का उपयोग कर सकते हैं, जिससे डेटा की गुणवत्ता और प्रामुखता को सुनिश्चित कर सकते हैं।

QGIS को डाउनलोड करना का तरीका (Downloading QGIS):

- QGIS की आधिकारिक वेबसाइट पर जाएं: https://qgis.org/



- "Download Now" या समर्थित संस्करण का चयन करें (Windows, macOS, Linux आदि)।
- डाउनलोड पृष्ठ पर, विभिन्न विकल्प दिए जाएंगे। आपके ऑपरेटिंग सिस्टम के अनुसार चयन करें और डाउनलोड बटन पर क्लिक करें।
- डाउनलोड पूरा होने के बाद, डाउनलोड की गई फ़ाइल को खोलें और स्थापना प्रक्रिया शुरू करें।

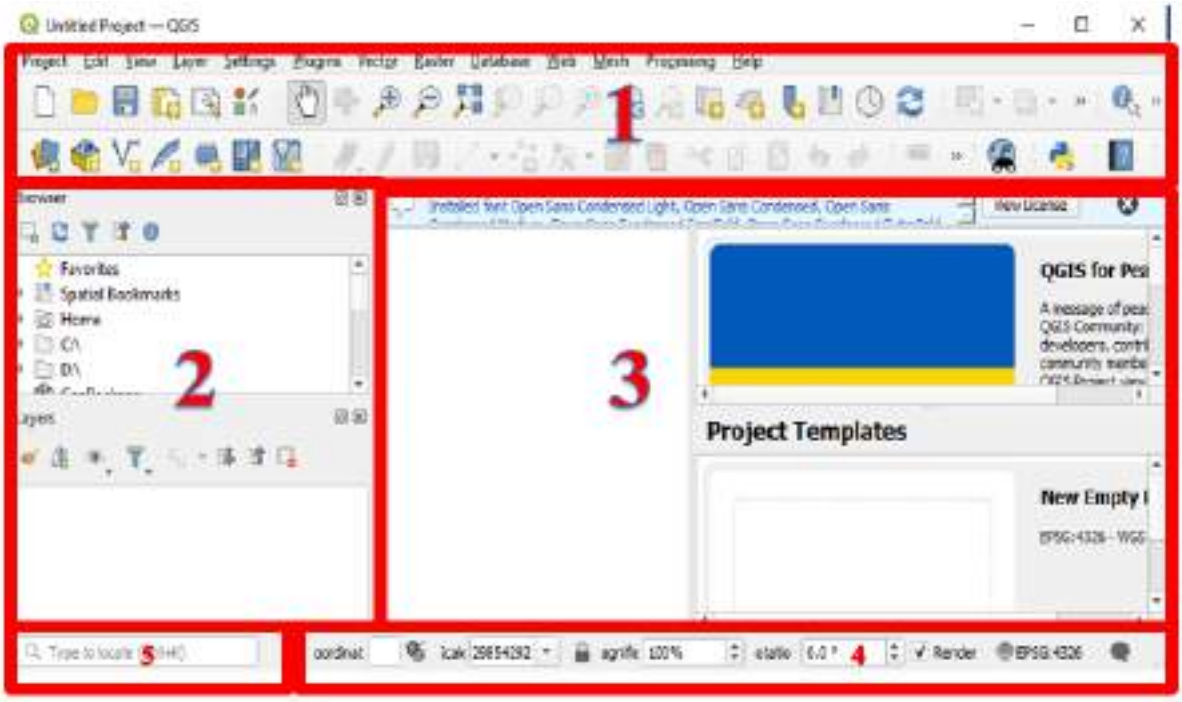


2. QGIS के विभिन्न संस्करण (QGIS Versions):

QGIS ने कई संस्करणों को जारी किया है, प्रत्येक में नई और बेहतर फीचर्स जोड़े गए हैं और सिस्टम की सुधार की गई है। यह कुछ महत्वपूर्ण QGIS संस्करण हैं:

- QGIS 1.x सीरीज़
- QGIS 2.x सीरीज़
- QGIS 3.x सीरीज़
- QGIS 3.16 'Hannover'
- QGIS 3.18 'Zürich'

3. QGIS इंटरफ़ेस (QGIS Interface):

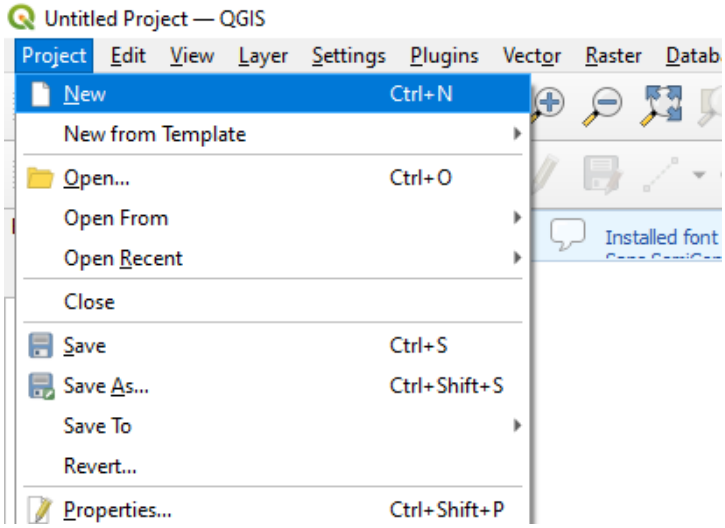


1. मेनू बार और टूलबार (Menu Bar and Toolbars); 2. लेयर पैनल/ ब्राउज़र पैनल (Layers List / Browser Panel); 3. मानचित्र व्यू (Map canvas); 4. स्टेटस बार (Status bar); 5. लोकेटर बार (Locator bar)

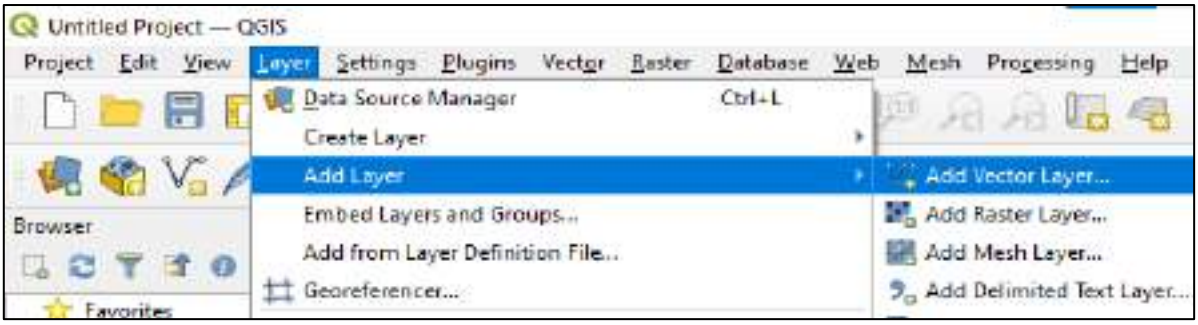
4. QGIS में डेटा इम्पोर्ट और एक्सपोर्ट (Data Import and Export in QGIS)

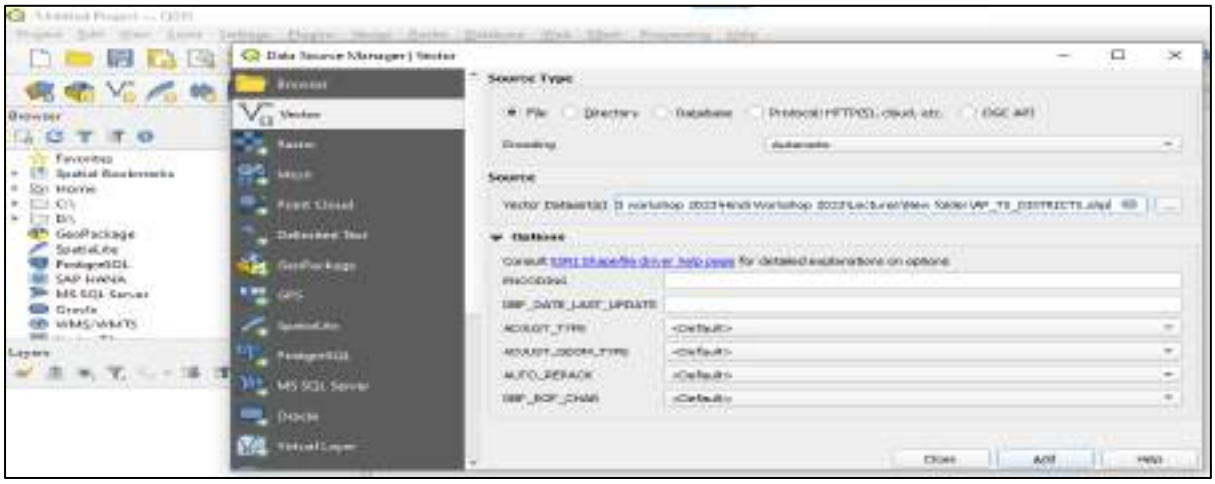
QGIS में पहली लेयर्स को जोड़ने के लिए निम्नलिखित कदम अनुसरण करें:

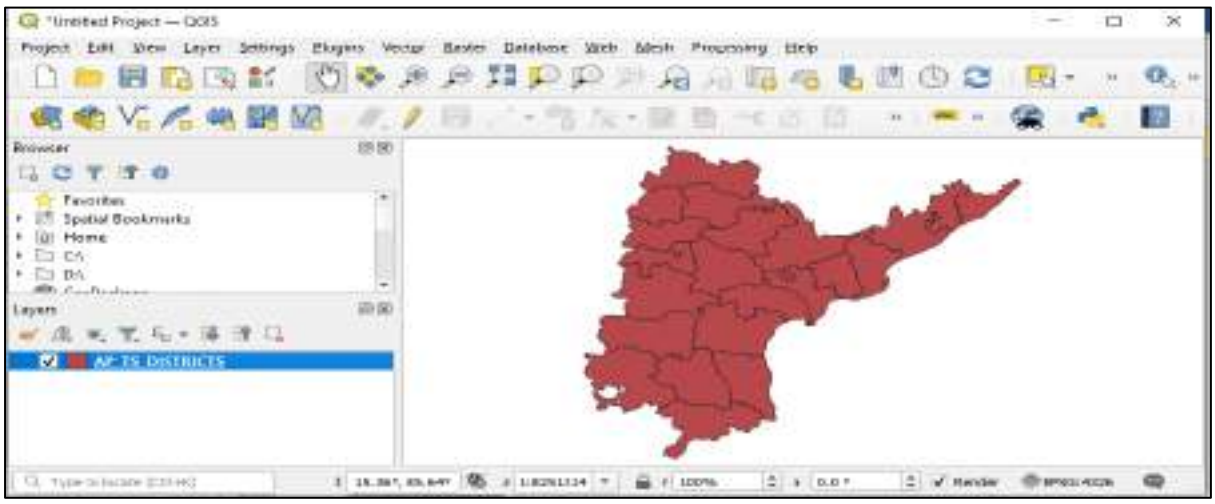
- सबसे पहले, QGIS सॉफ़्टवेयर खोलें।
- प्रोजेक्ट खोलने के बाद, "**Project**" मेनू में जाएं और "**New**" का चयन करें।



- नया प्रोजेक्ट बनाने के बाद, "**Layor**" मेनू में जाएं और "**Add Layor**" का चयन करें। लेयर जोड़ने के लिए, आपको किस प्रकार के डेटा को जोड़ना है उसका चयन करें। उदाहरण स्वरूप में, आप शेपफ़ाइल (Shapefile), रास्टर (Raster), डेटाबेस आदि को चुन सकते हैं।

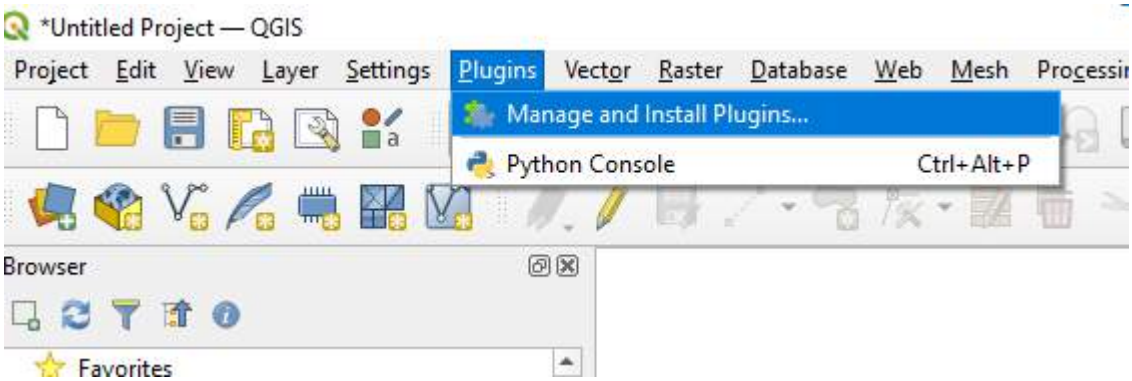




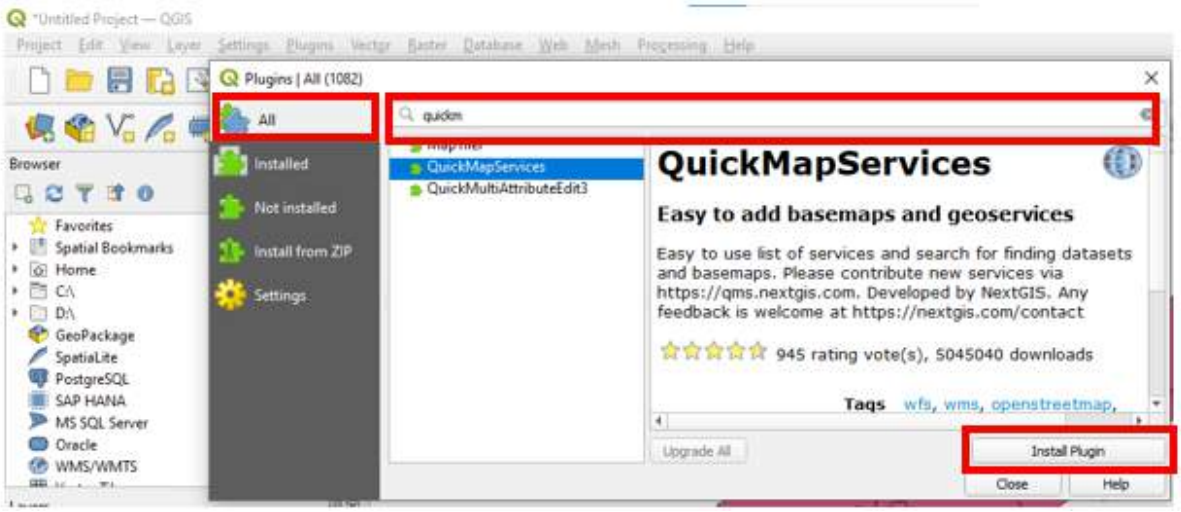


5. **QGIS में बाहरी प्लगइन डाउनलोड करना (Downloading Plugin in QGIS)**:

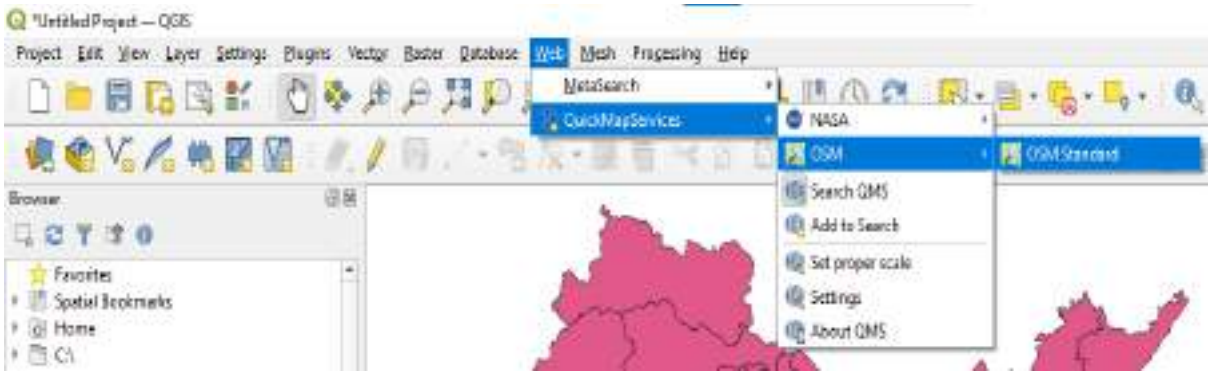
- QGIS के मुख्य मेनू से '**Plugins**' पर जाएं और वहाँ से '**Manage and Install Plugins**' का चयन करें।



- प्लगइन मैनेजर खुलने पर, '**All**' टैब का चयन करें।
- खोज बॉक्स में बाहरी प्लगइन का नाम या विवरण लिखें । आपकी खोज के आधार पर, सूची में उपलब्ध प्लगइन दिखाई जाएंगे। आप उनमें से जिस प्लगइन को डाउनलोड करना चाहते हैं, उस पर क्लिक करें।
- प्लगइन का विवरण खुलने पर, '**Install Plugin**' बटन पर क्लिक करें।

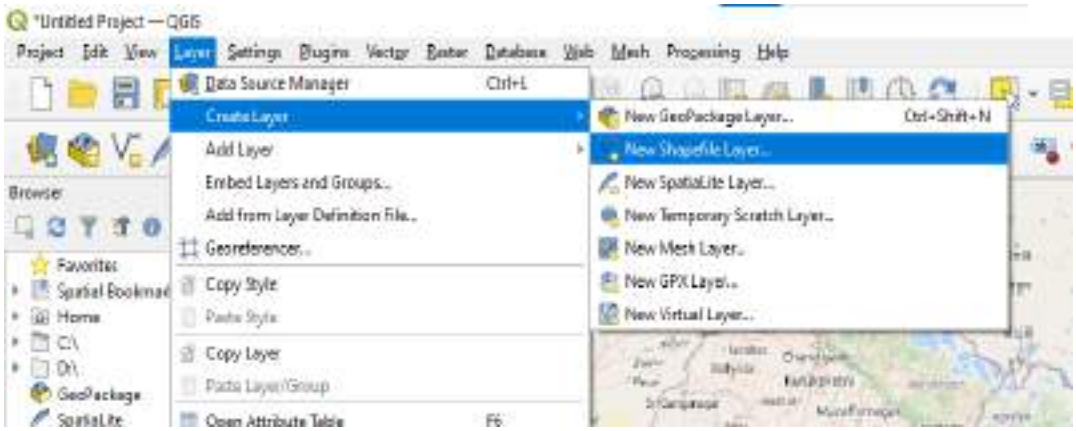


- प्लगइन सफलतापूर्वक स्थापित होने के बाद, उसे सक्रिय करने की आवश्यकता हो सकती है। इसके लिए '**Plugins**' मेनू से '**Manage and Install Plugins**' चुनें और फिर प्लगइन के नाम पर क्लिक करके उसे सक्रिय करें।

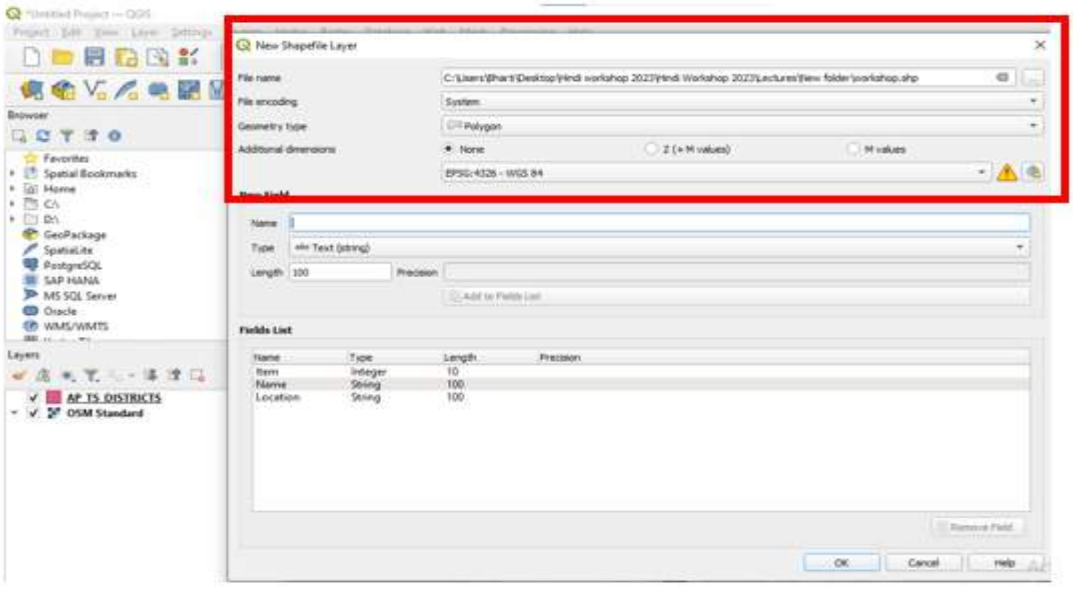


6. QGIS में एक नया लेयर बनाना (Create Layer in QGIS):

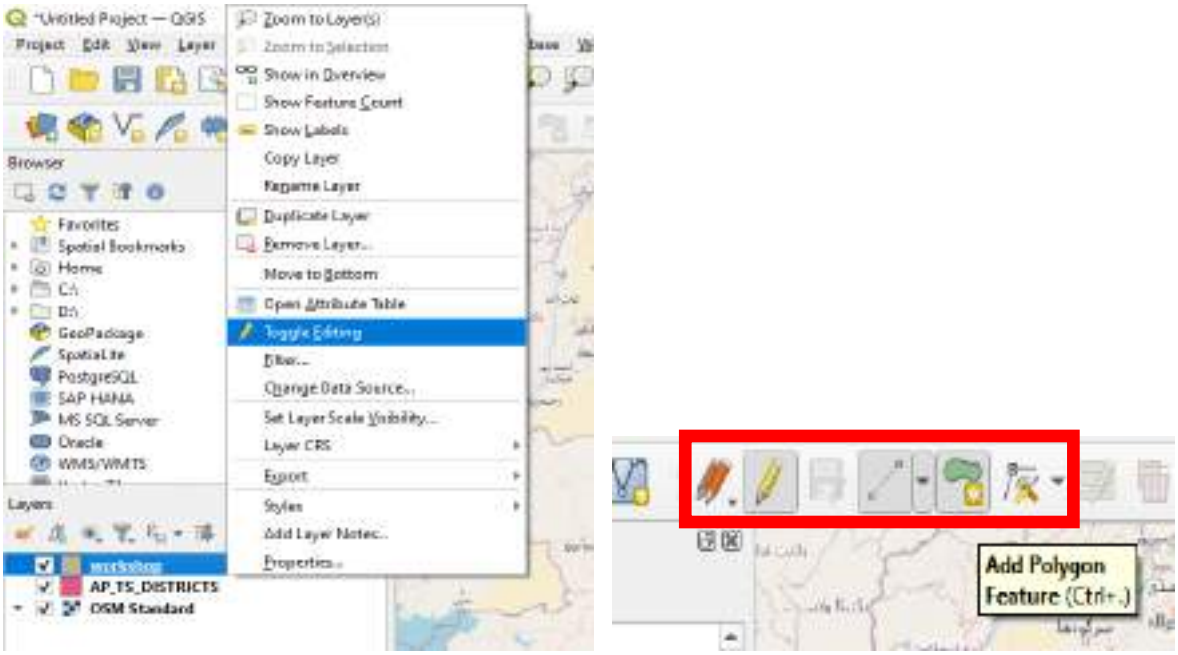
- QGIS के मुख्य मेनू से '**Layer**' पर जाएं और फिर '**Create Layer**' का चयन करें।



- एक नई लेयर में लेयर के प्रकार का चयन करने के लिए विकल्प दिखाए जाएंगे, जैसे कि पॉइंट, लाइन, और पॉलिगॉन। आपके आवश्यकताओं के आधार पर उपयुक्त प्रकार का चयन करें।



- लेयर प्रकार का चयन करने के बाद, लेयर के कोआर्डिनेट सिस्टम का चयन करे। आप अपने प्रोजेक्ट की आवश्यकतानुसार सही कोआर्डिनेट सिस्टम का चयन करें। लेयर कोन्फ़िगरेशन के लिए आपको विभिन्न विकल्प दिखाए जा सकते हैं, जैसे कि लेयर का नाम, फ़ील्ड की संरचना, आदि। आपको लेयर की विशिष्टताओं को कॉन्फ़िगर करने के लिए उपयुक्त विकल्प भरने की आवश्यकता होगी। सभी आवश्यक विशिष्टताएँ भरने के बाद, '**OK**' बटन पर क्लिक करें।
- नई लेयर अब QGIS में दिखनी चाहिए। आप उसे मानचित्र में दिखाने और संशोधित करने के लिए उपयोग कर सकते हैं।

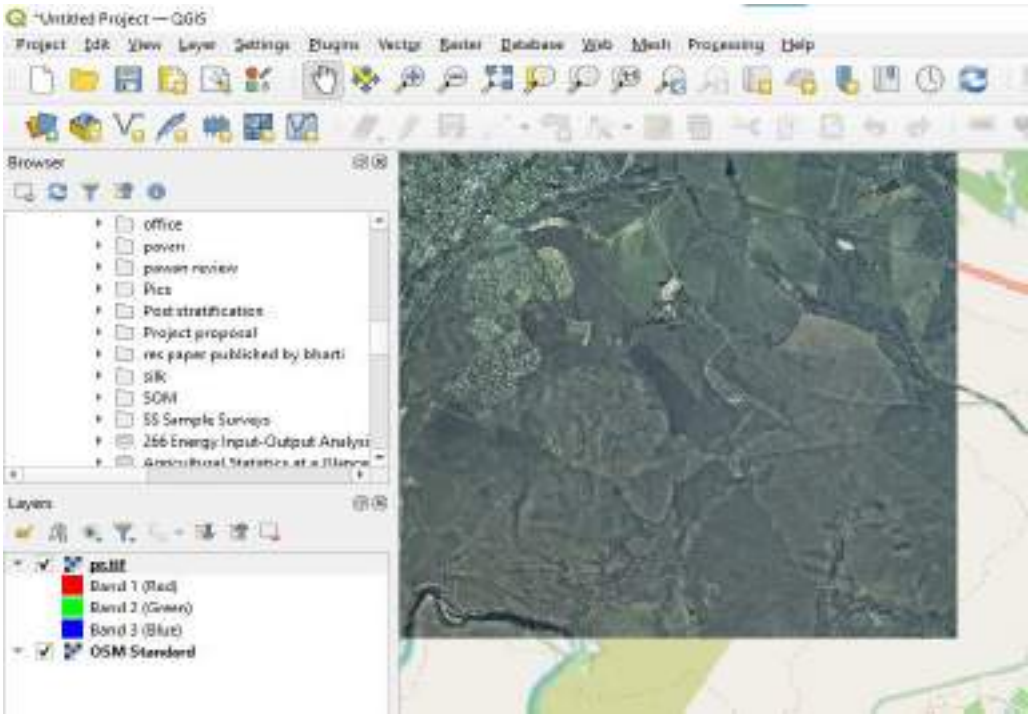


7. **QGIS में रास्ट्र लेयर जोड़ना (Add Raster Layer in QGIS):**

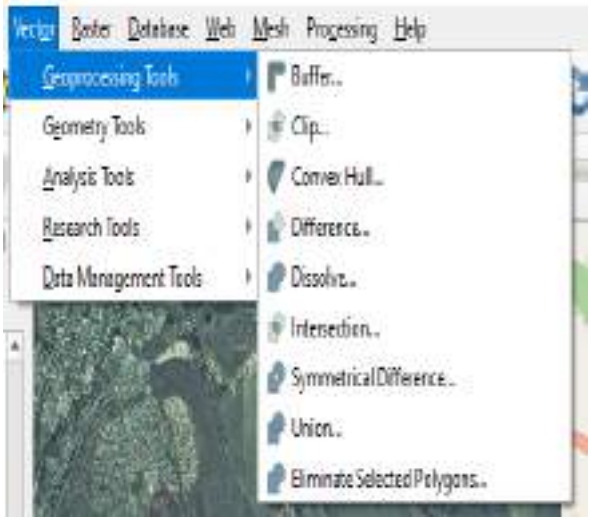
- QGIS को खोलने के बाद, मुख्य मेनू में "**Layer**" विकल्प पर क्लिक करें। "**Add Raster Later**" चुनें ।

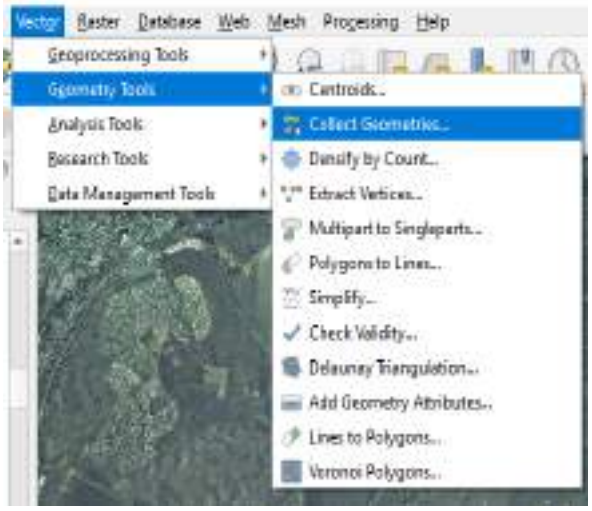
- "**Add Raster Layer**" क्लिक करने के बाद, आपको आपके सिस्टम पर संग्रहित रास्टर फ़ाइल का चयन करने के लिए एक डायलॉग बॉक्स दिखाई देगा। आपको उस फ़ाइल का पथ ब्राउज़ करके चयन करना होगा। जब आप अपने रास्टर फ़ाइल को चुन लेते हैं, तो "**Add**" बटन पर क्लिक करें।

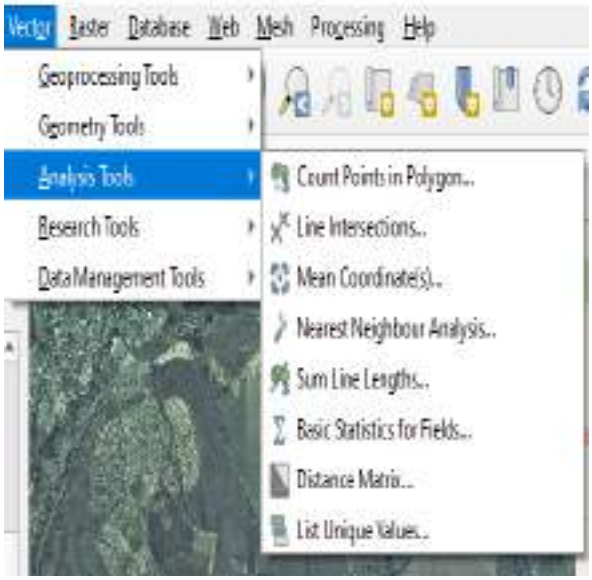
- चयनित रास्टर फ़ाइल के आधार पर, एक नई रास्टर लेयर आपके QGIS प्रोजेक्ट में जोड़ी जाएगी और आप इसे प्रदर्शित कर सकेंगे।

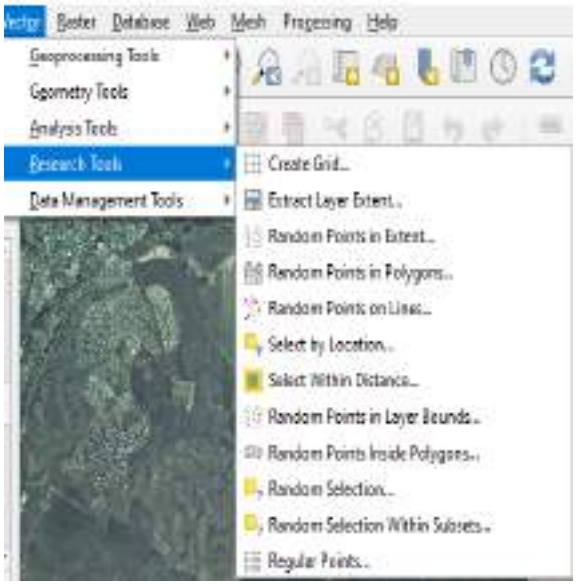


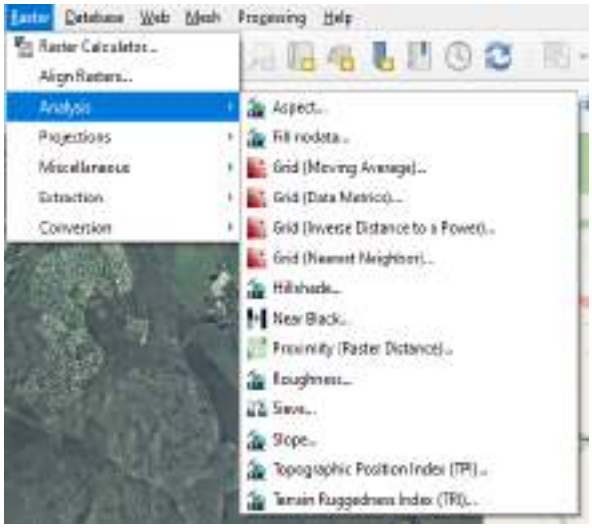
8. QGIS में "Analysis Tools":

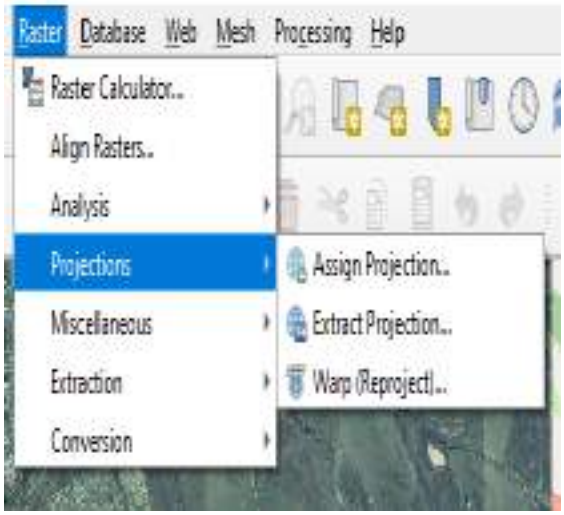












9. कृषि में QGIS का अनुप्रयोग (Application of QGIS in Agriculture):

- **स्थानिक विश्लेषण**: QGIS में मिट्टी के प्रकार, भूमि उपयोग पैटर्न जैसे स्थानिक डेटा का विश्लेषण किया जा सकता है । इस जानकारी का उपयोग फसल का चयन, बोने जाने वाले पैटर्न, और सिंचाई रणनीतियों के बारे में सूचित निर्णयों के लिए किया जा सकता है।
- **प्रेसिजन एग्रिकल्चर**: QGIS सैटेलाइट डेटा जैसे विभिन्न स्रोतों से डेटा को एकत्रित करके क्षेत्र की स्थितियों के सटीक नक्शे बना सकता है। किसान इन नक्शों का उपयोग पौष्टिकता, आर्द्रता या कीटों के विभिन्न स्तरों वाले क्षेत्रों की पहचान के लिए कर सकते हैं, जिससे उर्वरक, कीटनाशक और पानी का उचित इस्तेमाल किया जा सकता है।
- **उपज की पूर्वानुमान**: मौसम के पैटर्न, मिट्टी की गुणवत्ता और बोने जाने वाले तरीकों के साथ ऐतिहासिक उपज डेटा का विश्लेषण करके QGIS का उपयोग फसलों की उपज के लिए पूर्वानुमानी मॉडल बनाने में किया जा सकता है। यह जानकारी किसानों को बेहतर उत्पादन और विपणन के निर्णयों में मदद करती है।
- **जल प्रबंधन**: QGIS जल स्रोतों, सिंचाई प्रणालियों और मौसम पैटर्न से संबंधित डेटा को एकत्रित कर सकता है। इससे किसान सुगंधित सिंचाई अनुसूचियों को बना सकते हैं और जल संसाधनों का प्रबंधन प्रभावी तरीके से कर सकते हैं।
- **पर्यावरणीय प्रभाव मूल्यांकन**: नए कृषि परियोजनाओं की योजना बनाते समय, QGIS का उपयोग संभावित पर्यावरणीय प्रभावों का मूल्यांकन करने के लिए किया जा सकता है। यह महत्वपूर्ण क्षेत्रों, आवासों और जल स्रोतों की पहचान करने में मदद करता है जिन्हें संरक्षित किया जाना चाहिए।

अध्याय 10

उत्तर-पूर्वी पर्वतीय क्षेत्र में सुदूर संवेदन, भौगोलिक सूचना तंत्र एवं भूमि सर्वेक्षण आँकड़ों के उपयोग द्वारा फसल क्षेत्र आकलन

प्राची मिश्रा साहू
प्रमुख वैज्ञानिक, प्रतिदर्श सर्वेक्षण प्रभाग
भा.कृ.अनु.प.-भारतीय कृषि साँख्यिकी अनुसंधान संस्थान, लाइब्रेरी एवेन्यू, नई दिल्ली-110012

1. प्रस्तावना (Introduction):

भारत प्रमुख रुप से रोजगार एवं राष्ट्रीय आय में अपने योगदान की वजह से एक कृषि प्रधान अर्थ व्यवस्था है। विश्वसनीय एवं सामयिक कृषि आँकड़ों की उपलब्धता नियोजकों, प्रशासकों, नीति निर्माताओं एवं अनुसंधानकर्ताओं के लिए अति महत्वपूर्ण हैं क्योंकि उत्पादन, कीमत निर्धारण, प्रोसेसिंग, प्रापण, भण्डारण, परिवहन, विपणन, आयात/निर्यात, सार्वजनिक वितरण एवं निवेश नियोजन सहित अन्य सम्बद्ध मामलों पर नीति सम्बन्धी निर्णय लेने के लिए सरकार इन आँकड़ों पर निर्भर रहती है। विश्वसनीय एवं सामयिक कृषि आँकड़ो की उत्पत्ति के लिए इस तंत्र (mechanism) का निरंतर मूल्यांकन विशेष महत्व रखता है।

कृषि और सहकारिता विभाग (D.A.C.) में कृषि सम्बन्धी प्रमुख आँकड़ें जैसे – क्षेत्रफल, प्रमुख फसलों के उत्पादन आकलक इत्यादि के एकत्रीकरण, संकलन एवं प्रकाशन के लिए अर्थशास्त्र एवं सांख्यिकी निदेशालय (D.E.C.) एक नोडल अधिकारिक एजेन्सी है।

क्षेत्रफल सम्बन्धी आँकड़ों के एकत्रीकरण के लिए देश के राज्यों को तीन वृहत श्रेणियों में बांटा गया हैं : पहली श्रेणी में आने वाले राज्य और संघ शासित क्षेत्र भूकर सर्वेक्षित (cadastrally surveyed) हैं। इन क्षेत्रों में भूमि उपयोग (land use) के आँकड़े राजस्व एजेन्सी द्वारा बनाये गये भूमि अभिलेख (land record) का भाग है जो अस्थायी रूप से बसे राज्य (temporarily settled states) हैं। तीन राज्यों (केरल, उड़ीसा और पश्चिम बंगाल) और चार संघ शासित प्रदेश (चंडीगढ़, दिल्ली, दादर एवं नागर हवेली और पांडिचेरी) के अलावा सभी मुख्य राज्यों में भूमि अभिलेख (land record) का पालन किया जा रहा है । ये राज्य/संघ शासित प्रदेश देश के रिर्पोटिंग क्षेत्रफल के लगभग 86 प्रतिशत हैं। इन राज्यों में क्षेत्र आंकड़े पूर्ण गणना के आधार पर एकत्र होते हैं।

दूसरी श्रेणी जो आमतौर पर गैर भूमि अभिलेख (non-land record) राज्य या स्थायी रूप से बसे राज्य (permanently settled states) कहलाते हैं उन राज्यों में क्षेत्र आंकड़े प्रतिदर्श सर्वेक्षण के आधार पर एकत्र होते है। यह तीन राज्य केरल, उड़ीसा और पश्चिम बंगाल हैं। इन राज्यों में 'कृषि सांख्यिकी रिर्पोटिंग के लिए एजेन्सी की स्थापना' (EARAS) नामक एक योजना शुरू की गई है जो अन्य बातों के साथ–साथ

20 प्रतिशत गाँव/अनुमान क्षेत्र जैसे बड़े प्रतिदर्श क्षेत्र में पूर्ण गणना या प्रतिदर्श सर्वेक्षण करते हैं। यह राज्य रिर्पोटिंग क्षेत्र के लगभग 9 प्रतिशत हैं ।

असम राज्य के पहाड़ी क्षेत्र, उत्तर–पूर्वी क्षेत्र में शेष राज्य, सिक्किम, गोआ एवं अण्डमान द्वीप समूह, दमन व दीव और लक्षद्वीप जैसे संघ शासित क्षेत्र जहाँ पर कोई रिपोर्टिंग एजेन्सी कार्य नहीं कर रही थी वहाँ आंकड़ों के संग्रह का कार्य गाँव के मुखिया को सौंपा गया। यह तीसरी श्रेणी के राज्य हैं जो रिपोर्टिंग क्षेत्रफल के लगभग 5 प्रतिशत हैं। इन क्षेत्रों में क्षेत्र आंकड़े गाँव के मुखिया से इम्प्रेशनिस्टिक अप्रोच के अनुसार तदर्थ विधि के आधार पर एकत्रित किए जाते हैं। यहां विभिन्न फसलों का क्षेत्र आकलन केवल अनुमान से किया जाता है। अत: यह आंकड़े पूर्णतः अवैज्ञानिक एवं गैर–सांख्यिकीय होते हैं। इसलिए प्रभावी सांख्यिकीय पृष्ठभूमि द्वारा एक ऐसी वैज्ञानिक पद्धति विकसित करने की आवश्यकता महसूस की गयी जो इन क्षेत्रों में विभिन्न फ़सलों के अन्तर्गत क्षेत्रफल के विश्वसनीय आकलक उपलब्ध कराने में सक्षम हो।

सत्तर के दशक में सुदूर संवेदन (Remote Sensing) तकनीक कृषि के क्षेत्र में विशेष योगदान होने के कारण दुनिया भर में कृषि प्रणाली में सुधार के लिए एक नयी पद्धति के रूप में विकसित हुई । सुदूर संवेदन तकनीक द्वारा भूमि उपयोग आंकड़ों को प्राप्त करने और भौगोलिक क्षेत्र के व्यापक कवरेज के कारण इसको लोकप्रियता हासिल हुई है। इसलिए नियमित आधार पर इन राज्यों में कृषि आंकड़े एकत्रित करने के लिए उपग्रह सुदूर संवेदन के उपयोग की आवश्यकता है। इसके अलावा भौगोलिक तकनीकी क्षेत्र में भौगोलिक सूचना तंत्र (Geographical Information System) के कारण भौगोलिक सत्यता के अध्ययन में सांख्यिकी दृष्टिकोण से बदलाव आया है। भौगोलिक सूचना तंत्र, भौगोलिक निर्देशांक के विभिन्न स्रोतों जैसे जनगणना, सर्वेक्षण और सुदूर संवेदन से प्राप्त भौगोलिक आंकड़ें संभालने में सक्षम है, इसलिए भौगोलिक सूचना तंत्र और उपग्रह डिजिटल आंकड़े (digital data) का एकीकरण, फसल आंकड़ों के विश्वसनीय आकलन देने में सक्षम है।

उत्तर पूर्वी राज्यों में विद्यमान समस्याओं के कारण विभिन्न फसलों के क्षेत्रफल आकलन के लिये कोई तार्किक पद्धति नहीं है। इन राज्यों में मुख्यतः घने जंगलों वाले पहाड़ हैं। अधिकतर क्षेत्र ऊबड़–खाबड़ हैं और वहां पहुंच पाना बहुत कठिन है। फसलों के क्षेत्र के सापेक्ष कुल भौगोलिक क्षेत्र (total geographical area) का प्रतिशत बहुत कम है, सामान्यत: यह 10 प्रतिशत ही है। इन क्षेत्रों में ज्यादातर सीढ़ीदार और झूम खेती होती रही है। इन क्षेत्रों में विशेषतः मेघालय में अधिकतर बादल छाये रहते है इसलिये यहां बादल मुक्त उपग्रह के चित्र मिलना बहुत कठिन है। अत: केवल सुदूर संवेदी उपग्रह आंकड़ों का उपयोग फसल क्षेत्रफल आकलन के लिये पूर्णतः विश्वसनीय जानकारी नहीं दे सकता है। इन क्षेत्रों के लिए कोई भूकर नक्शे और गांव की सीमा के नक्शे भी उपलब्ध नहीं है। इसके अलावा एक गांव में किसानों की

कुल संख्या, प्रत्येक किसानों के पास कुल खेतों का क्षेत्रफल और किसानों के द्वारा उगायी जाने वाली फसलें सही तरह से ज्ञात नहीं है। इसलिये इन क्षेत्रों में फसल क्षेत्र आकलन की पारंपरिक पद्धति लागू करना भी कठिन है।

इन तथ्यों के आधार पर यह माना गया कि भौगोलिक सूचना तन्त्र वातावरण में भूमि सर्वेक्षण आंकड़ों के साथ उपग्रह आंकड़ों का उपयोग फसलों के अंतर्गत क्षेत्र के लिए विश्वसनीय आकलन प्राप्त करने के लिए उपयोगी हो सकता है। इस स्थिति में संतोषजनक तकनीक के अभाव में भारतीय कृषि सांख्यिकी अनुसंधान संस्थान (I.A.S.R.I.) नई दिल्ली ने अन्तरिक्ष अनुप्रयोग केन्द्र (S.A.C.) अहमदाबाद एवं उत्तर–पूर्वी अन्तरिक्ष अनुप्रयोग केन्द्र (N.E.S.A.C.), शिलाँग के सहयोग से एक परियोजना ''मेघालय में कृषि सांख्यिकी के एकत्रीकरण हेतु सुदूर संवेदन पद्धति का विकास'' शुरू की। इस परियोजना का मुख्य उद्देश्य मेघालय के एक जिले में धान के फसल के क्षेत्र आकलन के लिए एक उपयुक्त सर्वेक्षण पद्धति विकसित करना था। इसके लिए रीबोई जिला जो मेघालय का राइस बाउल माना जाता है, चयनित किया गया। पायलट अध्ययन के दौरान विकसित की गई पद्धति को मान्य (validate) करने के लिए यह पद्धति उसी जिले (रीबोई) में दोहरायी गयी एवं राज्य के एक अन्य जिले जयंतिया हिल्स में धान के अर्न्तगत क्षेत्र का आकलन करने के लिए भी लागू की गयी।

2. सर्वेक्षण आंकडो के साथ सुदूर संवेदन और भौगोलिक सूचना तन्त्र का उपयोग करने हेतु विकसित पद्धति (Methodology developed by using Remote Sensing and Geographic Information System with Survey data):

फसल क्षेत्र आकलन की पारंपरिक पद्धति, चाहे यह पूर्ण गणना से हो या प्रतिदर्श सर्वेक्षण से हो, जो देश के अन्य भागों में आयोजित की जाती है मेघालय में लागू नहीं है क्योंकि इस राज्य में कोई भूमि अभिलेख प्रणाली (land record system) नहीं है यहां कोई भूकर नक्शा (cadastral map) नहीं है और न ही गांव की सीमा का नक्शा (village boundary map) उपलब्ध है । सभी जिलों/ब्लाक के गांव की संख्या ठीक तरह से मालूम नहीं है। इसके अलावा एक गांव में किसानों की कुल संख्या, प्रत्येक किसानों के पास खेतों की संख्या और किसानों के द्वारा उगायी जाने वाली फसलें भी उपलब्ध नहीं हैं। वैकल्पिक रूप से यदि कोई फसल क्षेत्र आकलन के लिये सुदूर संवेदन उपकरण लगाने के बारे में सोचता है जो मैदानों में बहुत सफल रहा है तो कुछ अन्य समस्याओं जैसे घने जंगल, ऊबड़–खाबड़ क्षेत्र, अधिकतर क्षेत्र में पहुंचने में कठिनाई, सीढ़ीदार और झूम खेती के कारण उपग्रह आंकड़ो का सीधा उपयोग फसल क्षेत्र आकलन के लिये कठिन हो जाता है।

इसके अलावा मेघालय में अधिकतर बादल छाये रहते हैं जिससे इन क्षेत्रों में बादल मुक्त छवि मिलना बहुत मुश्किल है। अतः केवल सुदूर संवेदी उपग्रह आंकड़ों का उपयोग विभिन्न फसलों के अन्तर्गत क्षेत्र की विश्वसनीय जानकारी देने में सक्षम नहीं है। यह देखते हुये भौगोलिक सूचना तन्त्र (जी.आई.एस.) वातावरण में भूमि सर्वेक्षण द्वारा उपग्रह आंकड़ों का उपयोग फसलों के अन्तर्गत क्षेत्र के विश्वसनीय आकलन के लिये उपयोगी हो सकता है।

अतः धान को फसल के अन्तर्गत क्षेत्र आकलन के लिये एक एकीकृत दृष्टिकोण सुदूर संवेदन, व्यापक भूमि सर्वेक्षण और भौगोलिक सूचना तन्त्र के उपयोग द्वारा विकसित किया गया है। क्षेत्र की समस्याओं को देखते हुये इसको तीन प्रमुख श्रेणियों में बांटा गया है। इन समस्याओं का हल ढूढ़ने का प्रयास किया जा रहा है।

सुदूर संवेदन आंकड़ों के उपयोग द्वारा फसल क्षेत्र का आकलन करने के लिये निम्न समस्याओं को तीन समूह में बांटा गया हैः

1) ऊबड़ खाबड़ क्षेत्र, गलत वर्गीकरण और संवेदन (sensor) कोण होने के कारण छवि में फसल के अन्तर्गत क्षेत्र और जमीन पर फसल के अर्न्तगत वास्तविक क्षेत्र में महत्वपूर्ण अन्तर हो सकता है जो वर्गीकरण त्रुटियों का विशेष कारण हो सकता है।
2) धान के फसल के अन्तर्गत क्षेत्र जो पहाड़ या घाटी से ढके हैं वहां क्षेत्र उपग्रह सेंसर से नहीं दिखते हैं क्योंकि उपग्रह सेंसर सूर्य तुल्यकालिक (sun-synchronous) है। इसके अतिरिक्त छोटे धान का खेत भी छवि में नहीं आ पाता क्योंकि LISS III सेंसर का स्थानिक रिजल्यूशन कम होता है।
3) उपग्रह छवि में बादल का आवरण होने के कारण

प्रत्येक समस्या का समाधान निकाल कर इस पद्धति का विकास किया गया। ऊबड़ खाबड़ क्षेत्र और गलत वर्गीकरण त्रुटियों के कारण धान की फसल के अन्तर्गत क्षेत्रों को सुधारने के लिये वर्गीकृत छवि में धान की फसल के अन्तर्गत क्षेत्र और जमीन पर धान की फसल के अन्तर्गत क्षेत्र में एक सम्बन्ध स्थापित किया गया है।

पहाड़ की छाया और सेंसर के सीमित स्थानिक रिजल्यूशन के कारण जो धान के फसल के अन्तर्गत क्षेत्र उपग्रह सेंसर द्वारा नहीं लिये जा सकते है। वह भौगोलिक सूचना तन्त्र वातावरण के उपयोग द्वारा चयनित सड़कों में बफर जोन बना कर एक उपयुक्त प्रतिदर्श सर्वेक्षण के द्वारा लिया गया। इस बफर जोन में धान के फसल के अन्तर्गत क्षेत्र का आकलन करने के लिये एक उपयुक्त आकलन विकसित किया गया है। इस बफर का वेक्टर लेयर वर्गीकृत छवि पर आवरणित किया गया। बादल हटाने के लिये शुरू में संयुक्त

(कम्पोजिट) तकनीक लागू किया गया। इसके अलावा बादल या बादल छाया के अन्तर्गत धान के क्षेत्र के आकलक भी विकसित किये गये। इन आकलकों के उपयोग द्वारा पूरे जिले में धान के अन्तर्गत क्षेत्र के आकलक को विकसित किया गया।

इस परियोजना के अन्तर्गत भूमि सर्वेक्षण, सुदूर संवेदन एवं भौगोलिक सूचना तन्त्र के आधार पर धान की फ़सल के क्षेत्रफल आकलन के लिए एक पद्धति विकसित कर वैधीकृत की गयी तथा पूरे राज्य में क्रियान्वयन हेतु संतोषजनक पायी गयी । इसके पश्चात् राज्य में बहु–फ़सलों के अन्तर्गत क्षेत्रफल आकलन के लिए पद्धति विकसित करने की आवश्यकता महसूस की गयी ।

अतः एक दूसरी परियोजना 'उत्तर पूर्वी पर्वतीय क्षेत्र में कृषि सांख्यिकी के एकत्रीकरण हेतु सुदूर संवेदन पद्धति का विकास' शुरू की गयी जिसका उद्देश्य सुदूर संवेदन, भौगोलिक सूचना तंत्र एवं भूमि सर्वेक्षण का उपयोग करके बहु फसलों के क्षेत्र आकलन लिए एक एकीकृत पद्धति विकसित कर सके, था।

बहु फसलों में अनाज फ़सलें जैसे – धान एवं मक्का, फल फ़सलें जैसे अनानास तथा शाक फ़सलें जैसे आलू एवं इस क्षेत्र में उगायी जाने वाली अन्य सब्जियां (आलू के संवर्ग के अतिरिक्त) और वृक्षारोपण फ़सलें जैसे – काजू सम्मिलित है।

इस परियोजना के संचालन के लिए मेघालय को प्रतिनिधित्व राज्य के रूप में चुना गया । इस राज्य के चार जिले रिबोई, पूर्वी खासी हिल्स, पूर्वी गारो हिल्स एवं पश्चिमी गारो हिल्स नामित किए गये। जहाँ तक फ़सलीय पद्धति में विचरण एवं कृषि क्रियाओं का सम्बन्ध है, ये जिले समस्त राज्य की फ़सल दशाओं को सम्मिलित करते हैं ।

यह पाया गया कि मेघालय राज्य में विभिन्न ऊंचाइयों पर विभिन्न फ़सलें उगायी जाती हैं जैसे – धान की खेती घाटी में होती है, अनानास पहाड़ी ढलान में उगाया जाता है, आलू एवं अदरक तुलनात्मक रूप में पहाड़ी के ऊपर समतल स्थल पर उगाये जाते हैं। अतः इन क्षेत्रों में फ़सल उगाने के लिए ऊंचाई की महत्वपूर्ण भूमिका है । इसके अतिरिक्त राज्य में फ़सल के अन्तर्गत क्षेत्रफल बहुत कम (केवल 10 प्रतिशत) है और यह पूरे राज्य में फैला हुआ है। अतः प्रशासनिक सीमाओं पर आधारित स्तरीकरण दक्ष स्ट्राटा उपलब्ध नहीं करा पाता। चूँकि एक फ़सल के तहत क्षेत्रफल को प्रभावित करने के लिए ऊंचाई एवं खेती के विस्तार दो मुख्य कारक हैं, अतः एक ऐसे स्तरीकरण कसौटी की आवश्यकता थी जो इन दोनों कारकों को शामिल कर सके। इसलिए विभिन्न ऊंचाइयों पर उगायी जा रही सभी फ़सलों को शामिल करने के उद्देश्य से खेती की ऊंचाई एवं विस्तार पर आधारित स्थानिक स्तरीकरण की पद्धति विकसित की गयी। एलीवेशन, एन.आर.एस.ए. (NRSA) से प्राप्त 90 मीटर रिजल्यूशन सहित डिज़िटल एलीवेशन मॉडल (DEM) से प्राप्त किया गया । डिज़िटल एलीवेशन मॉडल लेयरों को न्यूनतम (25वें परसेन्टाइल तक), मध्यम (25वें एवं 75वें परसेन्टाइल के बीच) एवं उच्चतम (75वें परसेन्टाइल से अधिक) के रूप में वर्गीकृत

किया गया है। इस भूमि उपयोग/भूमि कवर (land use/ land cover) मैप से खेती के अर्न्तगत क्षेत्र की लेयर प्राप्त की गयी। खेती के अन्तर्गत क्षेत्र की लेयर को न्यूनतम (खेती के अन्तर्गत 05 प्रतिशत क्षेत्र से कम), मध्यम (खेती के अन्तर्गत 5 से 10 प्रतिशत के बीच क्षेत्र) एवं उच्चतम (खेती के अन्तर्गत 10 प्रतिशत क्षेत्र से अधिक) के रूप में वर्गीकृत किया गया है। भौगोलिक सूचना तंत्र वातावरण में दोनों लेयरों को आच्छादित किया गया तथा इण्टरसेक्शन आपरेशन का उपयोग करते हुए 09 स्ट्राटा प्राप्त किये गये जिनके नाम उच्च खेती उच्च एलीवेशन, उच्च खेती मध्यम एलीवेशन, उच्च खेती न्यून एलीवेशन, उच्च खेती उच्च एलीवेशन, उच्च खेती मध्यम एलीवेशन, उच्च खेती न्यून एलीवेशन, मध्यम खेती उच्च एलीवेशन, मध्यम खेती मध्यम एलीवेशन, मध्यम खेती न्यून एलीवेशन, न्यून खेती उच्च एलीवेशन, न्यून खेती मध्यम एलीवेशन एवं न्यून खेती न्यून एलीवेशन हैं। स्थानिक स्तरीकरण के पश्चात्, प्रत्येक जिले में प्रत्येक स्ट्रैटम के खेती के अन्तर्गत क्षेत्र के अनुपात में 60 गाँवों का यादृच्छिक रूप से चयन किया गया । इन प्रत्येक चयनित गाँवों में से 05 कृषकों का यादृच्छिक रूप से चयनित किया गया । इन चयनित कृषकों से विभिन्न चरों पर आँकडे एकत्रित किये गये जैसे गाँव की सामान्य सूचना, चालू मौसम के दौरान उगायी गयी फ़सलें, औसत बीज दर, बुआई का महीना, कटाई का महीना, औसत उपज एवं प्रत्याशित उत्पादन, प्रत्येक खेत के क्षेत्रफल के मौखिक एवं माप किये गये आकलन । इसके अतिरिक्त, मौखिक आकलक एवं वास्तविक क्षेत्रफल के बीच सम्बन्ध देखने के उद्देश्य से एक खेत के क्षेत्रफल को जी.पी.एस. (GPS) द्वारा मापा गया । प्राथमिक आँकड़ों के एकत्रीकरण के लिए अनुदेश मैनुअलों के साथ–साथ अनुसूचियाँ तैयार कर संशोधित की गयीं तथा उन्हें अन्तिम रूप दिया गया। विभिन्न फ़सलों के अन्तर्गत क्षेत्र के आकलक प्राप्त करने के लिए एक उपयुक्त आकलन प्रक्रिया विकसित की गयी।

यह पाया गया कि सभी चार जिलों में धान के फ़सल के आकलक 5 प्रतिशत तक की बहुत कम प्रतिशत मानक त्रुटियों के साथ आकलित किये गये क्योंकि धान इस राज्य की प्रमुख फसल है। अदरक की फ़सल में 10 से 12 प्रतिशत के बीच मानक त्रुटि आकलित की गयी जो काफी ठीक है। पूर्वी खासी हिल्स को छोड़कर अधिकांश जिले में आलू लघु फ़सल है, अतः अन्य जिलों में इस फ़सल के आकलन प्राप्त नहीं किये जा सके। पूर्वी खासी हिल्स में आलू की खेती के अन्तर्गत क्षेत्र 9.95 प्रतिशत मानक त्रुटि के साथ आँका गया जो काफी उचित है। मक्का रबी मौसम की फ़सल है अतः समस्त जिलों से मक्का के क्षेत्रफल की उचित सूचना प्राप्त नहीं हो सकी। अनानास रिबोई जिले की मुख्य फ़सल है परन्तु 300 चयनित किसानों में से केवल 18 किसानों के लघु प्रतिदर्श आकार होने के कारण अनानास के अन्तर्गत क्षेत्र बहुत कम आंका गया। अन्य जिलों के लिए यह लघु फ़सल है अतः अन्य जिलों में अनानास के अन्तर्गत कोई क्षेत्र नहीं पाया गया। काजू मुख्यतः पश्चिमी गारो हिल्स में उगाया जाता है इसलिए रिबोई एवं पूर्वी खासी हिल्स से इस फ़सल के अन्तर्गत क्षेत्र के कोई आँकड़े प्राप्त नहीं हुए तथा पूर्वी गारो हिल्स में बहुत ही छोटा क्षेत्र था जो लगभग नगण्य था। इसके अतिरिक्त, काजू पश्चिमी गारो हिल्स जिले की मुख्य फ़सल है जो 8.16 प्रतिशत मानक त्रुटि के साथ आंकी गयी। सब्जियों के सम्बन्ध में प्रतिशत मानक

त्रुटियाँ अपेक्षाकृत अधिक पायी गयीं क्योंकि बहु फ़सलों वाले एकीकृत सर्वेक्षणों में सब्जियों से सम्बन्धित प्रतिदर्श के आकार बहुत छोटे हैं। सब्जियों के मामले में उच्च मानक त्रुटियाँ स्वीकार्य हैं क्योंकि सर्वेक्षण व्यवहारिक तौर पर कठिन हो जाते हैं। कुछ सब्जियों की अनेक पिकिंग्स होती हैं तथा सम्पूर्ण आकलन प्राप्त करने के लिए सभी पिकिंग्स से आंकड़े एकत्रित करने की आवश्यकता है। सब्जियों की खेती के अन्तर्गत क्षेत्र के आकलन की परम्परागत विधि में भी न्यूनतम प्रतिशत मानक त्रुटि अधिक ही देखी गयी है।

यह पहला अध्ययन है जो मुख्य फ़सलों के क्षेत्रफल आकलन के लिए एक प्रभावी, दक्ष एवं सम्भाव्य पद्धति विकसित करने के उद्देश्य से आरम्भ की गयी तथा जो भविष्य में देश के उत्तर–पूर्वी भाग में अर्थशास्त्र एवं सांख्यिकी निदेशालय (D.E.S.) द्वारा अपनायी जा सकती है। इस अध्ययन में बहु फ़सलों के क्षेत्रफल आकलन के साथ–साथ अनाज, फलों एवं सब्जी की फ़सलों के लिए स्थानिक स्तरीकरण पर आधारित एक वैज्ञानिक पद्धति का सुझाव दिया गया। यह पद्धति वस्तुपरक, वैज्ञानिक एवं सत्यापन योग्य है परन्तु इसके वास्तविक क्रियान्वयन में कुछ समस्यायें पायी गयीं जिनकी विस्तार से चर्चा की गयी है। यदि इस पद्धति को भविष्य में अपनाया जाता है तो इसके लिए कुछ सिफ़ारिशें भी की गयी हैं।

प्रस्तावित पद्धति को लागू करने से पूर्व इसकी जाँच करने की आवश्यकता है। यह देखा गया है कि कुछ फ़सलें कुछ विशेष ब्लॉकों तथा पाकेटों तक ही केन्द्रित थीं। जैसे रिबोई जिले के एक विशेष ब्लॉक में अनानास बहुतायत में उगाया जाता है। यदि इस विशेष ब्लॉक, पॉकेट से कम गाँवों का चयन किया जाता है तो इससे फ़सल के अन्तर्गत क्षेत्र के अवआकलक प्राप्त होंगे । प्रतिचयन अभिकल्पना में सुझाया गया एक संशोधन यह है कि एक वृहत मास्टर प्रतिदर्श लिया जाना चाहिए जिससे सभी फ़सलों के आकलन प्राप्त किये जा सकते हैं और इसके अतिरिक्त, क्षेत्र की एकल फ़सल से सम्बन्धित लघु उप–प्रतिदर्श लिया जाना चाहिए जहाँ यह सबसे अधिक केन्द्रित है तथा संयुक्त आकलन विकसित किये जाने चाहिए। इससे आकलनों की दक्षता में सुधार होगा। सभी फ़सलों के अन्तर्गत क्षेत्र का वर्णन करने के लिए यदि मल्टीडेट हाई रिजल्यूशन डाटा प्रयुक्त किया जाता है तो क्षेत्रफल के आकलनों में सुधार होगा । इस अध्ययन में डी.ई.एम (DEM) के लिए लिया गया एस.आर.टी.एम. (SRTM) डाटा कोर्सर रिजल्यूशन (90 m) का था और यदि फाइनर रिजल्यूशन DEM का प्रयोग होता है तो इसमें बहुत सुधार हो सकता है। इसके अतिरिक्त, LISS III पर आधारित आकलकों में सुधार के लिए प्रतिदर्शक (sampler) के रूप में LISS IV का प्रयोग उचित है। भविष्य में CARTOSAT अथवा RADARSAT आँकड़ों के प्रयोग से आकलकों की दक्षता में वृद्वि हो सकती है।

अध्याय 11

फसल उपज आकलन हेतु फसल कटाई प्रयोग तकनीकें

मान सिंह
पूर्व क्षेत्रीय अधिकारी, प्रतिदर्श सर्वेक्षण प्रभाग
भा.कृ.अनु.प-भारतीय कृषि साँख्यिकी अनुसंधान संस्थान, लाइब्रेरी एवेन्यू, नई दिल्ली-110012

1. प्रस्तावना (Introduction)

फसल उत्पादन का अनुमान, फसल के अंतर्गत क्षेत्रफल और प्रति हेक्टेयर उपज के आधार पर किया जाता है। भारतवर्ष में फसल क्षेत्रफल के आँकड़े पूर्ण गणना और फसल उपज दर के अनुमान भारतीय सांख्यिकीविदों द्वारा विकसित वैज्ञानिक तकनीक **"फसल कटाई प्रयोग"** द्वारा **"प्रतिदर्श सर्वेक्षण विधि"** के माध्यम से प्राप्त किये जाते हैं। फसल कटाई प्रयोग की वैज्ञानिक पद्धति का ढाँचा विकसित करने में भारतीय कृषि सांख्यिकी अनुसंधान संस्थान, नई दिल्ली ने एक महत्वपूर्ण भूमिका निभाई है। एशिया प्रशांत और अफ्रीका के कुछ देषों में भी भारत में विकसित फसल कटाई प्रयोग तकनीक का प्रयोग किया जा रहा है।

अस्थाई व्यवस्थित (Temporarilly Settled) अथवा भू–अभिलेख (Land Records) राज्यों में कुल फसल क्षेत्रफल का 86% **पूर्ण गणना विधि** (Complete enumeration method) द्वारा, **स्थाई व्यवस्थित (Permanantly Settled)** राज्यों में **"कृषि सांख्यिकी को अभिलेखित करने के लिए एक ऐजेन्सी की स्थापना** (Establishment of an agency for reporting of Agriculture Statistics)" नामक योजना के अर्न्तगत प्रतिदर्श सर्वेक्षण विधि द्वारा 9% तथा **गैर प्रतिवेदन** (Non-Reporting) **क्षेत्रों** में व्यक्तिगत ज्ञान के आधार पर 5% फसल क्षेत्रफल के आँकडें एकत्र किये जाते है।

संपूर्ण भारतवर्ष में भारत सरकार राज्य सरकारों के माध्यम से कृषि वर्ष के तीनों मौसमों (खरीफ, रबी एवं जायद) में उपज दर के विश्वसनीय अनुमान प्राप्त करने के लिए फसल कटाई प्रयोगों का आयोजन किया जाता है। फसलों की बुबाई के तरीकों के आधार पर (पंक्तियों अथवा बिना पंक्तियों) अलग–अलग फसल कटाई प्रयोग विधि की सिफारिश की गई है। भारतवर्ष के राज्यों में 5 मीटर x 5 मीटर, 10 मीटर x 10 मीटर आकार के वर्गाकार अथवा 10 मीटर x 5 मीटर आकार के आयताकार फसल कटाई प्रयोग के प्रयोगात्मक प्लॉट तैयार किये जाते हैं। उत्तर प्रदेश में गन्ना, आलू, प्याज, तम्बाकू के लिए 5 मीटर भुजा के वर्गाकार प्रयोगात्मक प्लॉट के अलावा अधिकांष फसलों के लिए समबाहू त्रिभुज (त्रिभुज की प्रत्येक भुजा की लम्बाई 10 मीटर और क्षेत्रफल 43.3 वर्ग मीटर) की आकृति तथा पश्चिम

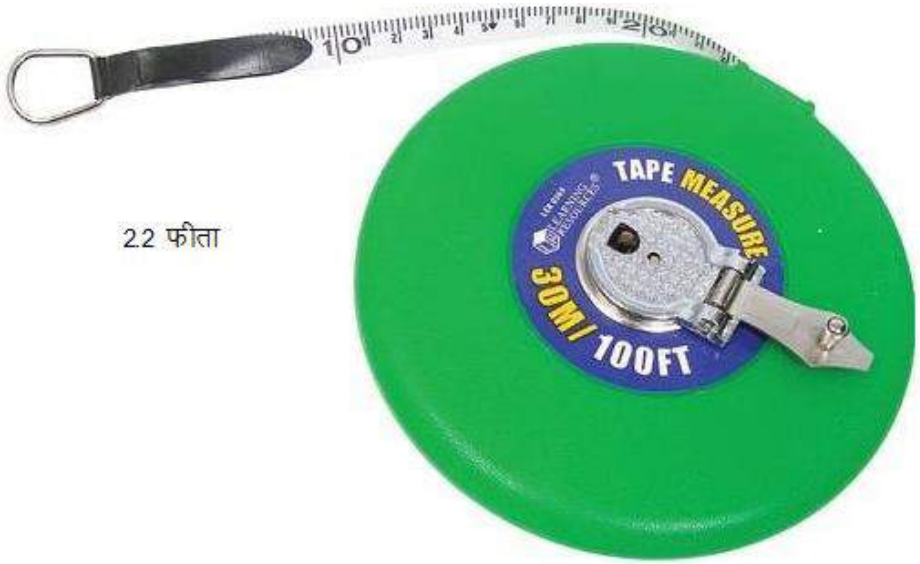
बंगाल में 2 फीट, 4 फीट एवं 1 फीट 7.5 इंच त्रिज्या के तीन संकेन्द्रित (concentric) वृत्तीय आकृति (त्रिज्या 5 फीट 7.5 इंच /1.7145 मीटर, क्षेत्रफल 9.284 वर्ग मीटर) के फसल कटाई प्रयोग के प्रयोगात्मक प्लॉट तैयार किये जाते हैं।

फसल उपज दर का अनुमान प्रतिदर्ष सर्वेक्षण विधि द्वारा अर्थ एवं सांख्यिकी निदेषालय, कृषि एवं किसान कल्याण मंत्रालय, भारत सरकार की ''सामान्य फसल आकलन सर्वेक्षण'' (General crop estimation surveys) नामक योजना के अर्न्तगत वैज्ञानिक तरीके से विकसित फसल कटाई प्रयोग विधि (सुखात्मे और पान्से (1951)) के आधार पर प्राप्त किये जाते है।

2. उपकरण एवं सामग्री

फसल कटाई प्रयोग के आयोजन में वर्गाकार प्रयोगात्मक प्लॉट तैयार करने सम्बन्धी उपकरणों की विशेष भूमिका है। फसल कटाई प्रयोग के आयोजन में प्रयोगात्मक प्लॉट तैयार करने में निम्नलिखित उपकरणों की आवश्यकता होती है

2.1 खूँटें



2.2 फीता

2.3 रस्सी

कमानीदार तराजू

तराजू

2.4 तराजू

2.5 बाटों का सेट

2.6 टाट–पट्टी

2.7 कपड़े की थैली

उपरोक्त के अतिरिक्त नमूना हेतु कपड़े की थैली, दो मजबूत जलरोधी (Water proof) थैले (फसल कटाई प्रयोग के उपकरणों एवं दस्तावेज रखने हेतु), खाली अनुसूचीयाँ, अनुदेश पुस्तिका, यादृच्छिक संख्या तालिकाऐं, लेखन सामग्री एवं फसल कटाई प्रयोग संबंधित अन्य दस्तावेजों की आवश्यकता होती है। फसल काटने के लिए दरांती, गाहाई, ओसाई एवं सफाई सम्बन्धी उपकरण/ सामग्री की भी आवश्यकता होती है। जो किसान के पास मिल जाते है।

3. फसल कटाई प्रयोग

फसल कटाई प्रयोग की प्रक्रिया विभिन्न चरणों द्वारा सम्पन्न होती है। इस प्रक्रिया में गाँवों का चयन, चयनित गाँवों में फसल कटाई प्रयोग के लिए अध्ययनाधीन फसल के खेतों की सूची तैयार करना, फसल कटाई प्रयोग के लिए खेतों का चयन, चयनित खेत में संस्तुति आकार का प्रयोगात्मक प्लॉट तैयार करना, प्रयोगात्मक प्लॉट से अध्ययनाधीन फसल की कटाई, गहाई, औसाई/सफाई के तुरन्त बाद उत्पाद का वजन एवं सूखे उत्पाद का पुनः वजन करना शामिल हैं।

3.1 गाँवों का चयन

प्रत्येक राज्य के लिए अध्ययनाधीन फसल के फसल कटाई प्रयोगों की संख्या का निर्धारण राज्य स्तर पर किया जाता है। अध्ययनाधीन फसल के क्षेत्रफल के आधार पर राज्य के जिलों को दो समूहों **"प्रमुख"** एवं **"गौण"** में विभाजित किया जाता हैं। यदि जिले में अध्ययनाधीन फसल का क्षेत्रफल 80000 हेक्टेयर से अधिक या 40000 से 80000 हेक्टेयर के बीच या राज्य के अन्य जिलों के औसत से अधिक है तो अध्ययनाधीन फसल के क्षेत्रफल के आधार पर उसे जिले को **प्रमुख** जिला माना जाता है। प्रमुख जिले में एक अध्ययनाधीन फसल के सामान्यतः 80 से 120 फसल कटाई प्रयोग आयोजित किये जाते हैं।

इसके अतिरिक्त अन्य जिलों को अध्ययनाधीन फसल के लिए गौण जिला माना जाता है। गौण जिले में फसल कटाई प्रयोगों की संख्या इस प्रकार समायोजित की जाता है कि अनुमानों की उच्च परिशुद्धता और क्षेत्रीय कार्यकर्ताओं का कार्यभार प्रबंधनीय हो सके। एक गौण जिले में अध्ययनाधीन फसल के औसतन 44 से 46 फसल कटाई प्रयोग आयोजित किये जाते हैं।

जिले के लिए आवंटित अध्ययनाधीन फसल के फसल कटाई प्रयोगों को जिले के सामुदायिक विकास खंड/तहसील (स्तरों) में अध्ययनाधीन फसल के क्षेत्रफल के अनुपात में वितरित किया जाता हैं। सामान्यतः सामुदायिक विकास खंड/तहसील स्तर पर अध्ययनाधीन फसल के अधिकतम 16 फसल कटाई प्रयोग किये जाते है। एक गाँव में दो फसल कटाई प्रयोग किये जाने की सिफारिश की गई है अतः गाँवों की संख्या, फसल कटाई प्रयोगों की संख्या से आधी होती है। राज्य स्तर पर जिले के सामुदायिक विकास खंड/तहसील वार चयनित गाँवों की सूची तैयार करके जिला के सम्बन्धित अधिकारियों को भेज दिया जाता है जो इन गाँवों को सामुदायिक विकास खंड/तहसील के चयनित गाँवों में कार्यरत प्राथमिक कार्यकर्ताओं में वितरित कर देते हैं।

3.2 खेत का चयन

3.2.1 खेत

फसल कटाई प्रयोग सम्पन्न करने के लिए खेत को निम्न प्रकार परिभाशित किया गया है।

''भूमि का एक ऐसा सुस्पष्ट टुक्ड़ा, जो चारों तरफ से मेंडों, दूसरी फसलों या अकृषित खेतों से स्पष्ट रूप से सिमांकित हो और जिसमें अध्ययनाधीन फसल उगायी गयी हो तथा खेत में संस्तुति आकार (25, 50 अथवा 100 वर्ग मीटर) का फसल कटाई प्रयोग प्लॉट बन सके''।

3.2.2 सर्वेक्षण संख्या का चयन

प्रत्येक चयनित गांव में दो फसल कटाई प्रयोग अलग–अलग दो खेतों में आयोजित किये जाते हैं। प्रत्येक चयनित गांव में दो खेतों का चुनाव करने के लिए दो अलग–2 यादृच्छिक संख्या आवंटित की जाती हैं। प्राथमिक कार्यकर्ता चयनित गांव में इन दो आवंटित यादृच्छिक संख्याओं का प्रयोग कर दो खेतों का चुनाव करता हैं।

गांव में दो सर्वेक्षण अथवा खसरा अथवा खेत संख्या का चुनाव यादृच्छिक विधि द्वारा किया जाता है। यदि आवंटित यादृच्छिक संख्या, गांव में कुल सर्वेक्षण अथवा खसरा अथवा खेतों की संख्या से छोटी है तो आवंटित यादृच्छिक संख्या के अनुरुप सर्वेक्षण अथवा खसरा अथवा खेत संख्या का चुनाव किया जाता है और यदि यादृच्छिक संख्या, गांव में कुल सर्वेक्षण अथवा खसरा अथवा खेतों की संख्या से बड़ी है तो आवंटित यादृच्छिक संख्या को खेतों की अधिकतम संख्या से भाग (Divide) करने पर प्राप्त शेषफल के अनुरुप सर्वेक्षण संख्या और शेषफल शून्य होने पर सबसे बड़ी सर्वेक्षण संख्या का चयन किया जाता है। अध्ययनाधीन

फसल यदि चयनित खेत में नहीं उगाई गयी हो तो अगली सर्वेक्षण संख्या का चयन करना चाहिए।

यदि गांव का चकबन्दी का कार्य पूर्ण नहीं हुआ है तो पेमाइस अथवा पट्टा संख्या का चयन करना चाहिए। जिस गांव में पेमाइस अथवा पट्टा संख्या भी उपलब्ध न हो सके तो काश्तकारों के नाम क्रमानुसार अथवा वर्णानुक्रम में व्यवस्थित कर चयन करना चाहिए।

3.2.3 उप–प्रभाग का चयन

यदि चयनित सर्वेक्षण संख्या, उप–प्रभागों (Sub-divisions) में विभाजित है। तो केवल एक उप–प्रभाग का चयन यादृच्छिक विधि द्वारा किया जाना चाहिए। यदि चयनित सर्वेक्षण संख्या कुल उप–प्रभागों की संख्या से छोटी है तो सर्वेक्षण संख्या के अनुरुप उप–प्रभाग का चयन करना चाहिए। यदि चयनित सर्वेक्षण संख्या उप–प्रभागो की कुल संख्या से बड़ी है तो चयनित सर्वेक्षण संख्या को उप–प्रभाग की संख्या से भाग देने से प्राप्त शेषफल के अनुरुप उप–प्रभाग का चयन करना चाहिए।

मानाकि चयनित सर्वेक्षण संख्या 116 है और सर्वेक्षण संख्या 116 में उप–प्रभागों की संख्या 6 (116/1, ..., 116/6) है। सर्वेक्षण संख्या 116, उप–प्रभागों की संख्या 6 से बड़ी है अतः 116 को 6 से भाग देने पर शेषफल 2 के अनुरुप उप–प्रभाग संख्या 116/2 का चयन होगा। और यदि चयनित सर्वेक्षण संख्या 2 है और सर्वेक्षण संख्या 2 में उप–प्रभागो की संख्या 3 (2/1, 2/2, 2/3) है। सर्वेक्षण संख्या 2, उप–प्रभागों की संख्या 3 से छोटी है अतः उप–प्रभाग संख्या 2/2 का चयन होगा।

3.2.4 सर्वेक्षण/उप–प्रभाग में खेत का चयन

चयनित सर्वेक्षण/उप–प्रभाग संख्या में अध्ययनाधीन फसल के बीज की प्रजाति, बुवाई के समय अथवा अन्य कारणों से एक से अधिक भागों (खेतों) में बोया गया हो, तो सर्वेक्षण/उप–प्रभाग संख्या के दक्षिण–पष्चिम कोने के पास वाले भाग (खेत) को फसल कटाई प्रयोग के लिए चयनित करना चाहिए है। यदि अध्ययनाधीन फसल के दो अथवा दो से अधिक खेत दक्षिण–पष्चिम कोने से समान दूरी पर है तो दक्षिण कोने के पास वाले भाग (खेत) को फसल कटाई प्रयोग के लिए चयनित करना चाहिए। यदि दक्षिण कोने के पास वाले खेत में फसल कटाई प्रयोग समायोजित न हो सके तो इसके सीधे हाथ वाले खेत को चुनना चाहिए। यह प्रक्रिया घड़ी की उल्टी दिशा में खेत का क्षेत्रफल फसल कटाई प्रयोग प्लॉट के क्षेत्रफल से बड़ा होने तक दोहरानी चाहिए है। यदि इस सर्वेक्षण/उप–प्रभाग का कोई भी खेत फसल कटाई प्रयोग समायोजित होने लायक नहीं है तो अगला सर्वेक्षण/उप–प्रभाग का चुनाव फसल कटाई प्रयोग के लिए करना चाहिए।

3.2.5 सर्वेक्षण संख्या/उप–प्रभाग का विलय

यदि चयनित सर्वेक्षण/उप–प्रभाग संख्या और पास के अन्य सर्वेक्षण /उप–प्रभाग संख्या के बीच कोई स्पष्ट सीमांकन अथवा मेंड़ नहीं है और इन सभी सर्वेक्षण/उप–प्रभाग में प्रयोगाधीन फसल एक साथ बोई गई है तो साथ वाले सर्वेक्षण/उप–प्रभाग को भी चयनित सर्वेक्षण/उप–प्रभाग संख्या में शामिल कर लेना चाहिए। गांव के खसरा रजिस्टर में साथ वाले सर्वेक्षण/उप–प्रभाग का विलय दर्ज होना आवश्यक नहीं है। लेकिन खसरा रजिस्टर में साथ वाले सर्वेक्षण/उप–प्रभाग में प्रयोगात्मक फसल का क्षेत्रफल दर्ज होना चाहिए।

3.2.6 चयन प्रक्रिया में अन्य आवश्यक शर्तें

- प्रयोगाधीन फसल चयनित खेत में बोयी गई हो और बोया गया क्षेत्रफल खसरा रजिस्टर में चयनित खेत (सर्वेक्षण/उप–प्रभाग) में दर्ज किया हो।
- चयनित खेत का क्षेत्रफल फसल कटाई प्रयोगात्मक प्लॉट के क्षेत्रफल से अधिक होना चाहिए।
- चयनित खेत में एक से अधिक फसलें एक साथ (मिश्रण) बोयी गई है, तो प्रयोगाधीन फसल का चयनित खेत के क्षेत्रफल का कम से कम 10 प्रतिशत होना चाहिए।
- फसल पुरस्कार, प्रतियोगिता, बीज उत्पादन, प्रदर्शन या चारे के उद्देश्य से नही उगायी गयी हो।
- चयनित खेत के किसान को उसके खेत में फसल कटाई प्रयोग आयोजित करने की सूचना देनी चाहिए।
- प्राथमिक कार्यकर्ता को प्रयोगाधीन फसल कटाई की सामान्य तारीख से लगभग एक माह पूर्व चयनित खेत में प्रयोगात्मक प्लॉट का दक्षिण–पष्चिम कोना चिन्हित करना चाहिए और किसान से मिलकर चयनित खेत में बोई प्रयोगाधीन फसल कटाई की अनुमानित तारीख पूछना चाहिए तथा प्रयोगाधीन फसल कटाई की तारीख के लिए लगाकर किसान के सम्पर्क में रहना चाहिए।
- प्राथमिक कार्यकर्ता को किसान से निवेदन करना चाहिए कि चयनित खेत की प्रयोगाधीन फसल की कटाई, फसल कटाई प्रयोग सम्पन्न होने के बाद शुरु करे तथा प्रयोगात्मक प्लॉट की फसल कटाई उसकी उपस्थिति में ही होगी।
- चयनित खेत की प्रयोगाधीन फसल का क्षेत्रफल यदि खसरा रजिस्टर में दर्ज है और प्रयोगाधीन फसल पशुओं, जंगली जानवरों अथवा कीटों/रोगों/अधिक वर्षा/अपर्याप्त वर्षा/ओला वृष्टि ने पूर्ण रुप से क्षतिग्रस्त हो तो प्रयोगाधीन फसल के फसल कटाई प्रयोगात्मक प्लॉट के उत्पाद की मात्रा को शून्य दर्ज किया जाना चाहिए।

- चयनित खेत अथवा प्रयोगात्मक प्लॉट में फसल अच्छी न होने के कारण खेत को रद्द नहीं करना चाहिए।
- किसान यदि चयनित खेत की प्रयोगाधीन फसल को आंशिक या पूर्ण रुप से काट लेता तो, प्राथमिक कार्यकर्ता को इस खेत के स्थान पर अन्य खेत चुनने की अनुमति नहीं है। ऐसी दशा में इस चयनित खेत के फसल कटाई प्रयोग को हानि के रुप में माना जायेगा।

4. फसल कटाई प्रयोग विधि

फसलों को मुख्य रुप से छिटकवां अथवा कम दूरी पर पंक्तियों में, एक दिशा में स्पष्ट पंक्तियों में एवं दोनों दिशाओं में स्पष्ट पंक्तियों में बोया जाता है। फसल बुवाई की तीनों दशाओं के अनुसार उपज दर अनुमान हेतु चयनित खेत की प्रयोगाधीन फसल के प्रत्येक पौधा को फसल कटाई प्रयोग की समष्टि के चयन में समान अवसर प्रदान हो सके इसलिए फसल कटाई प्रयोग की प्रक्रिया अलग–अलग है।

प्राथमिक कार्यकर्ता को प्रयोगाधीन फसल कटाई के दिन कटाई शुरु होने से पहले चयनित खेत पर पहुँचकर सर्व प्रथम प्रयोगात्मक प्लॉट तैयार करना चाहिए। इसके बाद प्रयोगात्मक प्लॉट की प्रयोगाधीन फसल की कटाई, उसके बंड़ल बनाना, उन्हें चिन्हित करना एवं खलिहान में सुरक्षित स्थान पर रखना चाहिए। इसके बाद ही शेष खेत की फसल कटाई शुरु होनी चाहिए। प्रयोगात्मक प्लॉट तैयार करने में निम्नलिखित पहलु शामिल हैं।

4.1 बिना पंक्तियों (छिटकवां) अथवा कम दूरी पर पंक्तियों में बोई गई फसलों की फसल कटाई प्रयोग विधि

गेहूं, जौ, चना, सरसों, मसूर, मक्का, बाजरा, धान, ज्वार, उर्द, मूंग इत्यादि फसलों बुवाई छिटकवां अथवा कम दूरी पर पंक्तियों में बोया जाता है। यदि बुवाई पंक्तियों में बोई गई है तो पंक्ति से पंक्ति की दूरी कम और पंक्ति में पौधा से पौधा की दूरी निश्चित नहीं होती है।

4.1.1 चयनित खेत का दक्षिण–पश्चिम कोना

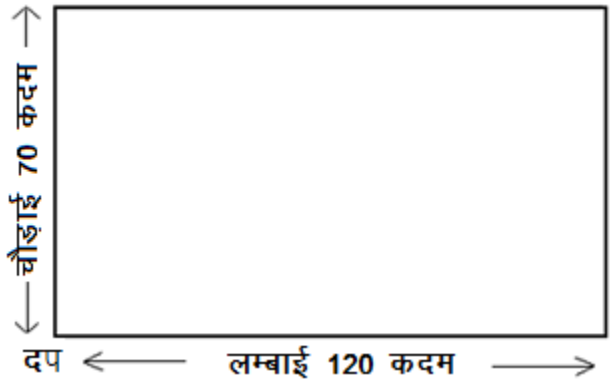
चयनित खेत में प्रयोगात्मक प्लॉट के निर्धारण में समानता रखने हेतू चयनित खेत का दक्षिण–पश्चिम कोना चिन्हित करना अनिवार्य किया गया है। सर्वप्रथम चयनित खेत के दक्षिण–पश्चिम कोने की पहचान की जाती है। यदि आप चयनित खेत के दक्षिण–पष्चिम कोने पर उत्तर की ओर मुँह करके खड़े हैं तो चयनित खेत आपके सामने और आपके दाहिने हाथ की ओर होना चाहिए। चयनित खेत, यदि ठीक उत्तर–दक्षिण और पूर्व–पश्चिम दिशा में नहीं है तो वह कोना जो लगभग दक्षिण–पश्चिम दिशा में है, दक्षिण–पश्चिम कोने के रूप में लिया जा सकता है।

दक्षिण–पश्चिम

4.1.2 चयनित खेत की लम्बाई एवं चौड़ाई

4.1.2.1 नियमित आकृति

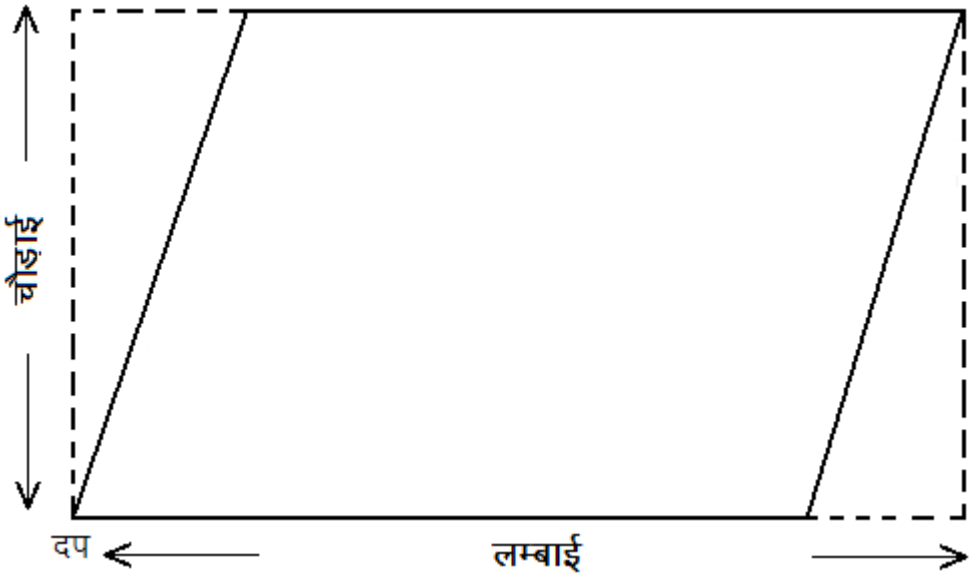
चयनित खेत की आकृति यदि वर्गाकार अथवा आयताकार है तो चयनित खेत की लम्बाई एवं चौड़ाई को सामान्य कदमों में नापते है। चयनित खेत के दक्षिण–पश्चिम कोने से लम्बी भुजा को लम्बाई एवं चौड़ी भुजा को चौड़ाई के रुप में नापते है (आकृति–4.1.2.1)।



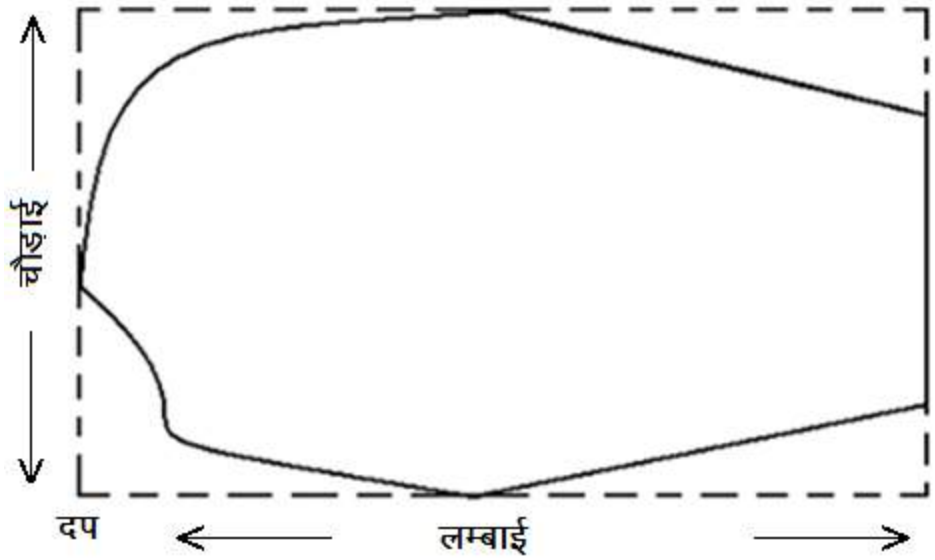
आकृति–4.1.2.1: नियमित आकृति

4.1.2.2 अनियमित आकृति

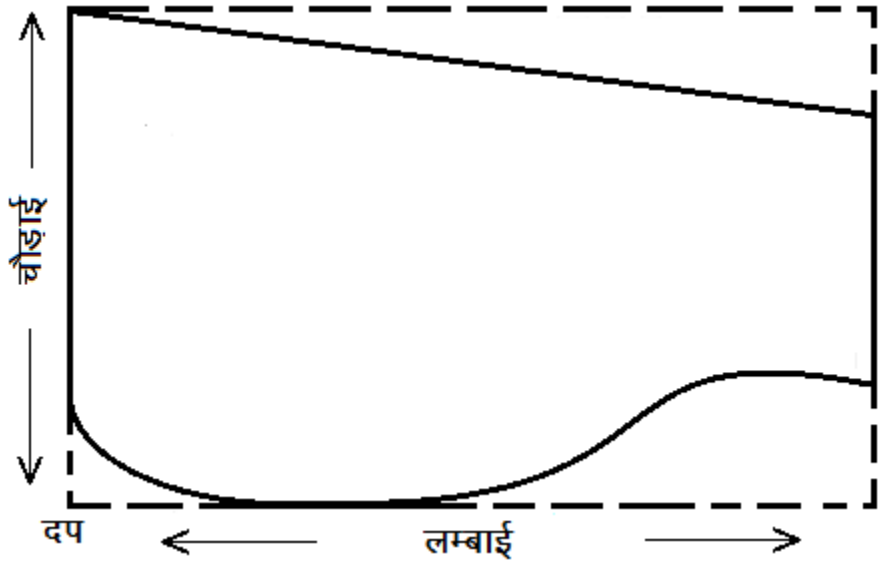
चयनित खेत की आकृति यदि वर्गाकार अथवा आयताकार नहीं है तो सर्वप्रथम चयनित खेत को वर्गाकार अथवा आयताकार आकृति प्रदान की जाती है। अनियमित खेत के बाहरी किनारों को छूते हुए चारों ओर से वर्गाकार अथवा आयताकार आकृति में सीमांकित करते है। सीमांकित वर्गाकार अथवा आयताकार आकृति के दक्षिण–पश्चिम कोने को अनियमित खेत का दक्षिण–पश्चिम कोना मानते है। सीमांकित वर्गाकार अथवा आयताकार आकृति के दक्षिण–पश्चिम कोने के आधार पर ही अनियमित खेत में प्रयोगात्मक प्लॉट का दक्षिण–पश्चिम कोना निर्धारित किया जाता है। सीमांकित वर्गाकार अथवा आयताकार आकृति के दक्षिण–पश्चिम कोने से लम्बी भुजा को लम्बाई एवं चौड़ी भुजा को चौड़ाई के रुप में कदमों से नापते है (आकृति–4.1.2.2)।



आकृति–4.1.2.2: अनियमित आकृति



आकृति–4.1.2.2: अनियमित आकृति



आकृति–4.1.2.2: अनियमित आकृति

4.1.3 यादृच्छिक संख्या का चयन

चयनित खेत में प्रयोगात्मक प्लॉट का दक्षिण–पश्चिम कोना निर्धारित करने हेतु, दो यादृच्छिक संख्याओं का चयन किया जाता है। पहली यादृच्छिक संख्या लम्बाई एवं दूसरी चौड़ाई के लिए होती है। फसल कटाई प्रयोगात्मक प्लॉट संपूर्ण रुप से चयनित खेत में समायोजित हो सके इसलिए प्रयोगात्मक प्लॉट की कदमों में लम्बाई एवं चौड़ाई को क्रमशः चयनित खेत की कदमों में लम्बाई एवं चौड़ाई से घटाते है। शेष लम्बाई एवं चौड़ाई के अनुसार यादृच्छिक संख्याओं का चयन किया जाता है। मान लीजिए, प्रयोगात्मक प्लॉट की लम्बाई एवं चौड़ाई 5 मीटर है और 5 मीटर 7 कदमों के बराबर है तो यादृच्छिक संख्याओं के चयन हेतु शेष लम्बाई एवं चौड़ाई निम्न प्रकार होगी।

उदाहरणः 4.1.3

चयनित खेत की लम्बाई	=	120 कदम
प्रयोगात्मक प्लॉट की लम्बाई	=	7 कदम
चयनित खेत की लम्बाई से प्रयोगात्मक प्लॉट की लम्बाई घटाने पर शेष लम्बाई	=	113 कदम
चयनित खेत की चौड़ाई	=	70 कदम
प्रयोगात्मक प्लॉट चौड़ाई	=	7 कदम
चयनित खेत की चौड़ाई से प्रयोगात्मक प्लॉट की चौड़ाई घटाने पर शेष चौड़ाई	=	63 कदम

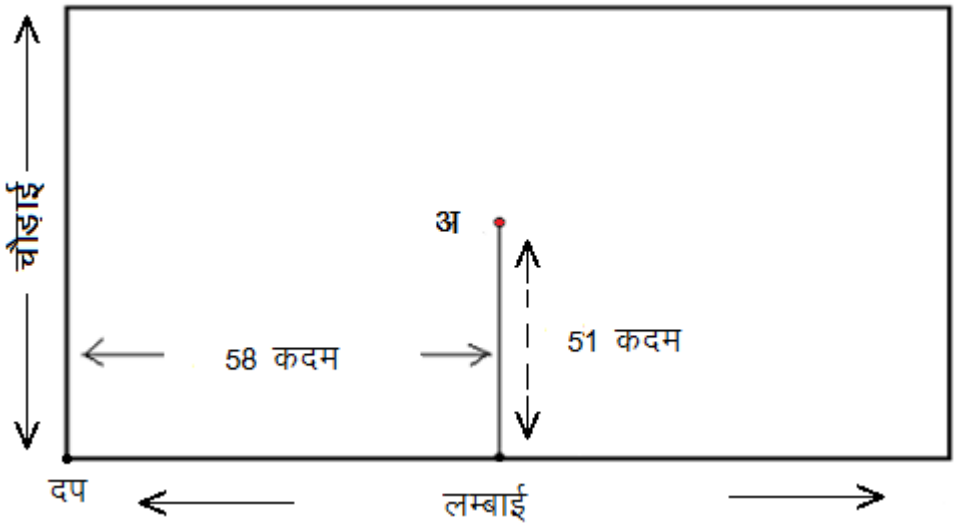
यादृच्छिक संख्या का चयन यादृच्छिक संख्या तालिका की मद्द से किया जाता है। चयन के लिए प्रत्येक प्राथमिक कार्यकर्ता को यादृच्छिक संख्या तालिका का अलग स्तंभ निर्दिष्ट किया जाता है। मानाकि, किसी एक प्राथमिक कार्यकर्ता को यादृच्छिक संख्या तालिका का स्तंभ **एक** निर्दिष्ट किया गया है। लम्बाई के लिए यादृच्छिक संख्या का चयन हेतु चयनित खेत की शेष लम्बाई 113 के अनुसार तीन अंकों की यादृच्छिक संख्या तालिका के स्तंभ **एक** को संदर्भित करते हुए 113 या इससे कम जो भी यादृच्छिक संख्या सबसे पहले आती है उसके चारों ओर घेरा बनाकर लिख लिया जाता है। इसी प्रकार चौड़ाई हेतु चयनित खेत की शेष चौड़ाई 63 के अनुसार दो अंकों की यादृच्छिक संख्या तालिका की स्तंभ **एक** में 63 या इससे कम जो भी यादृच्छिक संख्या सबसे पहले आती है उसके चारों ओर घेरा बनाकर उसे भी लिख लिया जाता है। माना लम्बाई एवं चौड़ाई हेतु जो चयनित यादृच्छिक पहले आती है वे क्रमशः 58 एवं 51 है। चयनित यादृच्छिक संख्याओं की सहायता से चयनित खेत में प्रयोगात्मक प्लॉट का दक्षिण–पश्चिम कोना निर्धारित करते है।

यदि यादृच्छिक संख्या के चयन की प्रक्रिया के दौरान निर्दिष्ट स्तंभ समाप्त हो जाता है तो दायें वाले स्तंभ को संदर्भित करना चाहिए। चयनित खेत की अनियमित आकृति के कारण, यदि प्रयोगात्मक प्लॉट आंशिक अथवा संपूर्ण, चयनित खेत की सीमा के बाहर निकल जाता है, तो इस चयनित यादृच्छिक संख्या जोड़ा रद्द करना चाहिए और नया यादृच्छिक संख्या जोड़ा प्रयोगात्मक प्लॉट संपूर्ण रुप से चयनित खेत की सीमा में आने तक चुनना चाहिए।

4.1.4 प्रयोगात्मक प्लॉट तैयार करना

4.1.4.1 प्रयोगात्मक प्लॉट का दक्षिण–पश्चिम कोना

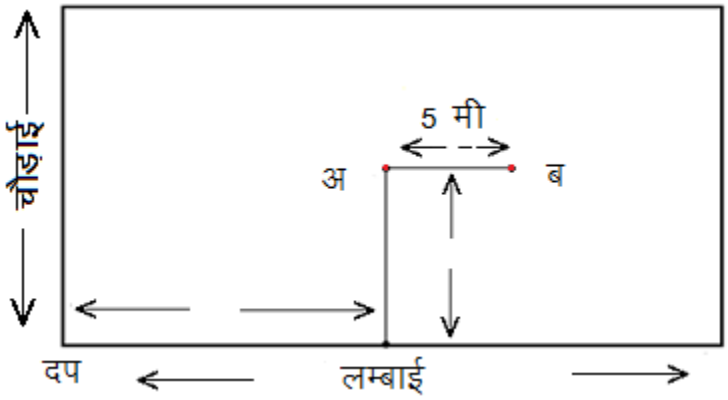
चयनित खेत के दक्षिण–पश्चिम कोने से उसकी लम्बाई की दिशा में 58 कदम चलने पर जिस बिन्दु पर पहुँचते है वहाँ से चयनित खेत की चौड़ाई की दिशा में 51 कदम चलने पर जो बिन्दु आता है वह प्रयोगात्मक प्लॉट का दक्षिण–पश्चिम कोना कहलाता है। इस बिन्दु को प्रयोगात्मक प्लॉट का मुख्य–बिन्दु (Key-Point) और प्रथम कोना भी कहते है। इस मुख्य–बिन्दु "**अ**" पर पहली खूँटी गाड़ देते है (आकृति–4.1.4.1)।



आकृति–4.1.4.1: प्रयोगात्मक प्लॉट का दक्षिण–पश्चिम कोना

4.1.4.2 प्रयोगात्मक प्लॉट का दूसरा कोना

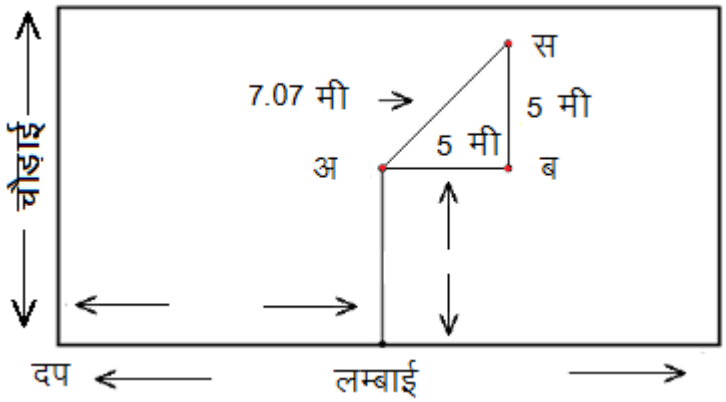
प्रयोगात्मक प्लॉट के मुख्य–बिन्दु "**अ**" से चयनित खेत की लम्बाई की ओर 5 मीटर की दूरी पर प्रयोगात्मक प्लॉट का दूसरा कोना "**ब**" कहलाता है। "**ब**" बिन्दु पर दूसरी खूँटी गाड़ देते है (आकृति–4.1.4.2)। "**अ**" "**ब**" को मिलाने वाली रेखा को प्रयोगात्मक प्लॉट की आधार रेखा भी कहते है।



आकृति–4.1.4.2: प्रयोगात्मक प्लॉट का दूसरा कोना

4.1.4.3 प्रयोगात्मक प्लॉट का तीसरा कोना

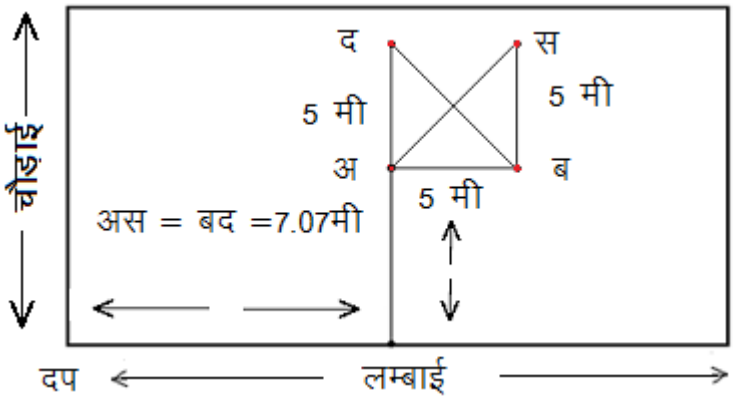
प्रयोगात्मक प्लॉट का तीसरा कोना चिन्हित करने हेतु समकोण त्रिभुज बनाने की विधि का उपयोग किया जाता है। "**ब**" बिन्दु पर 90^0 का कोण बनाने के लिए पहला व्यक्ति "**अ**" बिन्दु पर मापक फीता को शून्य तथा दूसरा व्यक्ति "**ब**" बिन्दु पर 12.07 (7.07+5.00) मीटर के निशान पर पकड़कर खड़ा होता है और तीसरा व्यक्ति मापक फीता को "**अ**" बिन्दु से 7.07 {वर्ग मूल (5^2+5^2)} एवं "**ब**" बिन्दु से 5.00 मीटर की दूरी पर पकड़कर चयनित खेत की चौड़ाई की ओर खींचते हुए जिस बिन्दु पर पहुँचता है उसे प्रयोगात्मक प्लॉट का तीसरा कोना "**स**" कहते है। "**स**" बिन्दु पर तीसरी खूँटी गाड़ देते है (आकृति–4.1.4.3)।



आकृति–4.1.4.3: प्रयोगात्मक प्लॉट का तीसरा कोना

4.1.4.4 प्रयोगात्मक प्लॉट का चौथा कोना

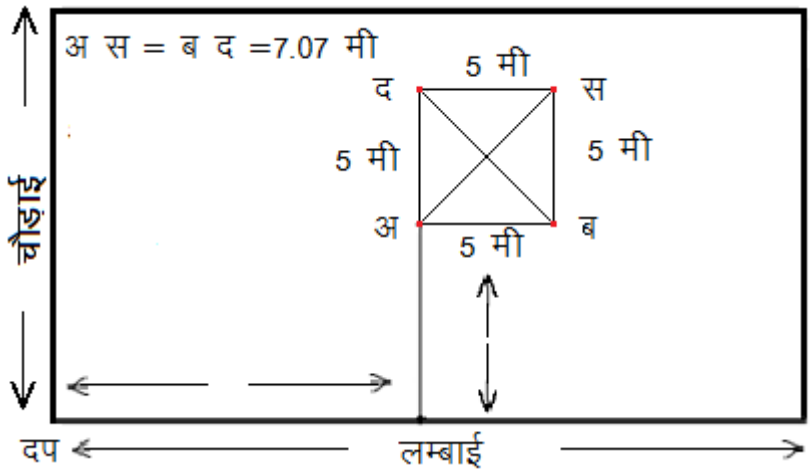
तीसरे कोने के चिह्नांकन के अनुसार ही चौथा कोना चिन्हित करते है। "**अ**" बिन्दु पर 90^0 का कोण बनाने हेतु "**अ**" तथा "**ब**" बिन्दु पर खड़ा व्यक्ति पहले की तरह मापक फीता को क्रमशः शून्य तथा 12.07 (7.07+5.00) मीटर के निशान पर पकड़ते हैं और तीसरा व्यक्ति मापक फीता को "**अ**" बिन्दु से 5.00 एवं "**ब**" बिन्दु से 7.07 {वर्ग मूल (5^2+5^2)} मीटर की दूरी पर पकड़कर चयनित खेत की चौड़ाई की दिशा में खींचते हुए जिस बिन्दु पर पहुँचता है उसे प्रयोगात्मक प्लॉट का चौथा कोना **"द"** कहते है। **"द"** बिन्दु पर चौथी खूँटी गाड़ देते है (आकृति–4.1.4.4)।



आकृति–4.1.4.4: प्रयोगात्मक प्लॉट का चौथा कोना

4.1.4.5 प्रयोगात्मक प्लॉट

प्रयोगात्मक प्लॉट के चारों कोनों ''अ ब स द'' पर खूँटों को मजबूती से भूमि में सीधा गाड़ते है और प्रयोगात्मक प्लॉट की चारों भुजाओं तथा दोनों कर्णों की दूरी की ठीक से जाँच करनी चाहिए। प्रत्येक भुजा 5 और कर्ण 7.07 मीटर का होना चाहिए (आकृति–4.1.4.5)।



आकृति–4.1.4.5: प्रयोगात्मक प्लॉट

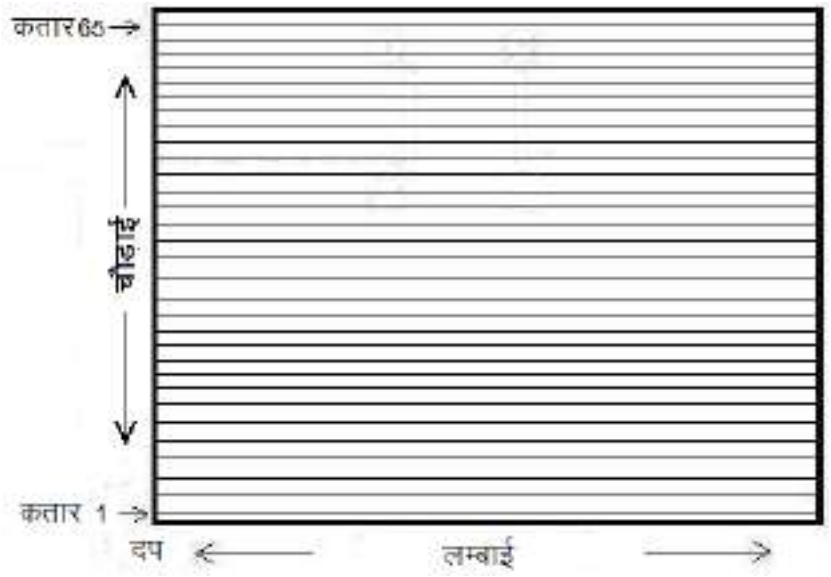
प्रयोगात्मक प्लॉट

4.2 एक दिशा में स्पष्ट पंक्तियों में बोई गई फसलों की फसल कटाई प्रयोग विधि

आलू, अरहर, कपास, गन्ना, अरड़ी जैसी फसलें एक दिशा में स्पष्ट पंक्तियों में बोई जाती हैं। इन फसलों की फसल कटाई प्रयोग विधि, बिना पंक्तियों एवं दोनों दिशाओं में पंक्तियों में बोई गई फसलों की फसल कटाई प्रयोग विधि से भिन्न हैं। खेत के चयन से लेकर चयनित खेत के दक्षिण–पष्चिम कोने के चिह्नांकन तक 3.2, 3.2.1, 3.2.2, 3.2.3, 3.2.4, 3.2.5, 4.1.1 में किये गये उल्लेख के समान ही है।

4.2.1 कतारों की गणना

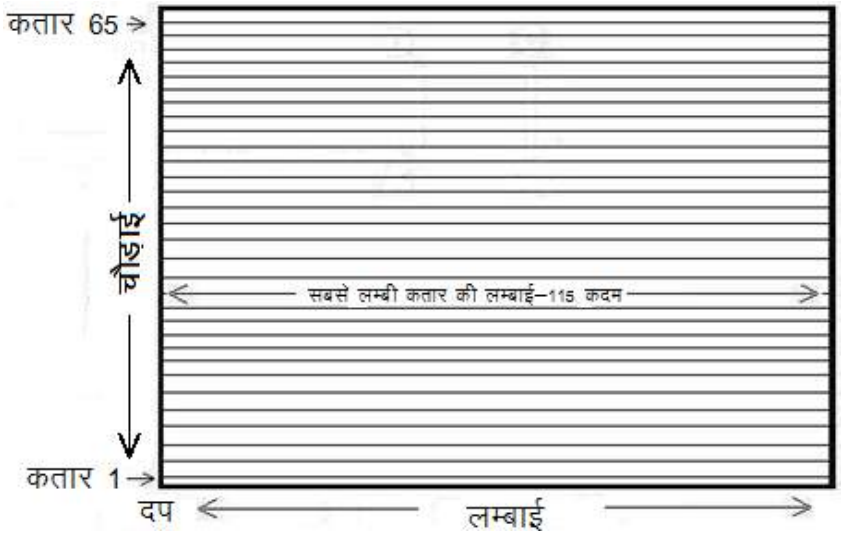
चयनित खेत के दक्षिण–पष्चिम कोने से कतारों की दिशा का निर्णय किया जाता है। जिस भुजा से कतारों में बुवाई की गई है उसे चयनित खेत की चौड़ाई और कतार की लम्बाई को चयनित खेत की लम्बाई माना जाता है। चयनित खेत के दक्षिण–पष्चिम कोने से कतारों की गणना करते हैं (आकृति–4.2.1)।



आकृति–4.2.1: कतारों की गणना

4.2.2 कतार की लम्बाई

चयनित खेत में सबसे लम्बी कतार की जाँच कर लम्बाई को कदमों में नापते है (आकृति–4.2.2)।



आकृति–4.2.2: सबसे लम्बी कतार की लम्बाई

4.2.3 कतारों की औसत संख्या

प्रयोगात्मक प्लॉट की चौड़ाई के अनुसार चयनित खेत में यादृच्छिक रुप से तीन स्थानों पर कतारों का वास्तविक अवलोकन (Physical observation) कर, कतारों की औसत संख्या की गणना की जाती है।

4.2.4 यादृच्छिक संख्या का चयन

4.2.4.1 यादृच्छिक कतार हेतु यादृच्छिक संख्या

चयनित खेत की अन्तिम कतार को प्रतिदर्श में शामिल करने के लिए चयनित खेत की कुल कतारों से प्रयोगात्मक प्लॉट की चौड़ाई में कतारों की औसत संख्या को घटाने के बाद शेषफल में **एक** जोड़ने के बाद प्राप्त संख्या के अनुसार यादृच्छिक संख्या तालिका के निर्दिष्ट स्तंभ **एक** का उपयोग कर यादृच्छिक संख्या का चयन किया जाता है।

मानाकि, चयनित खेत में कतारों की कुल संख्या 65 और प्रयोगात्मक प्लॉट में कतारों की औसत संख्या 6 है तो यादृच्छिक संख्या ज्ञात करने हेतु गणना निम्न प्रकार होगी।

उदाहरण: 4.2.4.1

चयनित खेत में कतारों की कुल संख्या	=	65
प्रयोगात्मक प्लॉट में कतारों की औसत संख्या	=	6
चयनित खेत की कुल कतारों से प्रयोगात्मक प्लॉट की औसत कतारों को घटाने पर संख्या	=	59
प्रतिदर्श में अन्तिम कतार को शामिल करने के लिए जोड़ी जाने वाली संख्या	=	1
जोड़ने पर कतारों की संख्या	=	60

उदाहरण 4.2.4.1 में कतारों की संख्या 60 के अनुसार दो अंकों की यादृच्छिक संख्या तालिका के निर्दिष्ट स्तंभ एक का उपयोग करते हुए 60 अथवा 60 से कम यादृच्छिक संख्या का चयन किया जाता है। माना निर्दिष्ट स्तंभ में सबसे पहले 60 अथवा 60 से कम यादृच्छिक संख्या 22 आती है। अतः 22वीं कतार को यादृच्छिक कतार माना जाता है।

4.2.4.2 लम्बाई हेतु यादृच्छिक संख्या

चयनित खेत की सबसे लम्बी कतार की लम्बाई (कदमों में) से प्रयोगात्मक प्लॉट की लम्बाई (5 मीटर =7 कदम और 10 मीटर = 13 कदम) को घटाने पर प्राप्त शेष संख्या के अनुसार यादृच्छिक संख्या तालिका के निर्दिष्ट स्तंभ एक का उपयोग कर यादृच्छिक संख्या का चयन करना चाहिए।

मानाकि, चयनित खेत में सबसे लम्बी कतार की लम्बाई 115 और प्रयोगात्मक प्लॉट की लम्बाई 7 कदम है। तो यादृच्छिक संख्या के चयन हेतु गणना निम्न प्रकार होगी।

उदाहरणः 4.2.4.2

चयनित खेत में सबसे लम्बी कतार की लम्बाई	=	115 कदम
प्रयोगात्मक प्लॉट की लम्बाई	=	7 कदम
चयनित खेत की सबसे लम्बी कतार की लम्बाई से प्रयोगात्मक प्लॉट की लम्बाई घटाने पर शेष लम्बाई	=	108 कदम

उदाहरण 4.2.4.2 के अनुसार चयनित खेत की सबसे लम्बी कतार की लम्बाई से प्रयोगात्मक प्लॉट की लम्बाई घटाने पर शेष लम्बाई 108 के अनुसार तीन अंकों की यादृच्छिक संख्या तालिका के निर्दिष्ट स्तंभ **एक** का उपयोग करते हुए 108 अथवा 108 से कम जो भी यादृच्छिक संख्या सबसे पहले आती है उसका चयन करना चाहिए।

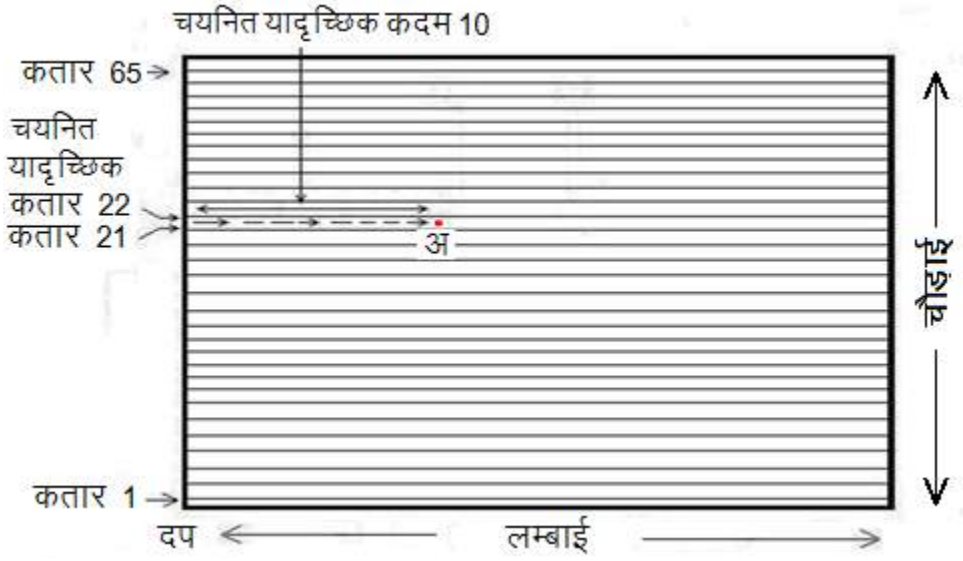
माना निर्दिष्ट स्तंभ **एक** का उपयोग करते हुए सबसे पहले 10 यादृच्छिक संख्या आती है। जो 108 से कम है। अतः 10 यादृच्छिक संख्या को यादृच्छिक कदम के रुप चयनित माना जायेगा।

प्रयोगात्मक प्लॉट का दक्षिण–पश्चिम कोना चयनित यादृच्छिक संख्या जोड़ा (22, 10) की सहायता से चिन्हित करते है।

4.2.5 चयनित खेत में प्रयोगात्मक प्लॉट तैयार करना

4.2.5.1 प्रयोगात्मक प्लॉट का दक्षिण–पश्चिम कोना

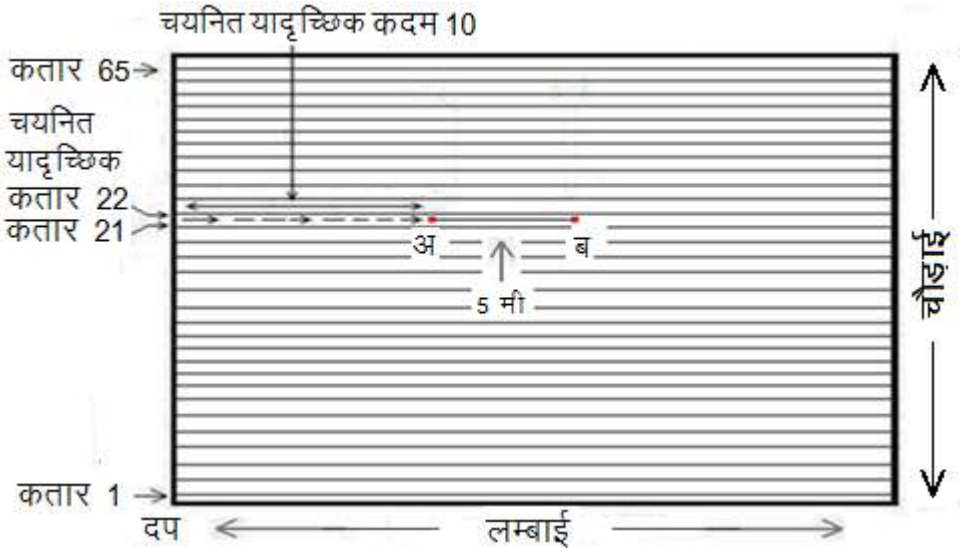
चयनित खेत के दक्षिण–पष्चिम कोने से कतारों को गिनते हुए चयनित यादृच्छिक कतार (22) पर रुकते है तथा यहाँ से चयनित यादृच्छिक कतार (22) और इससे पहली (21) कतार के बीच चलते हुए, 10 चयनित यादृच्छिक कदमों की दूरी पर जो बिन्दु आता है वह प्रयोगात्मक प्लॉट का दक्षिण–पश्चिम कोना कहलाता "**अ**" है। इसको प्रयोगात्मक प्लॉट का मुख्य–बिन्दु (Key-Point) एवं पहला कोना भी कहते है। इस मुख्य–बिन्दु "**अ**" पर चयनित यादृच्छिक कतार (22) और इससे पहली वाली कतार (21) के बीच पहली खूँटी गाड़ देते है (आकृति–4.2.5.1)।



आकृति–4.2.5.1: प्रयोगात्मक प्लॉट का दक्षिण–पश्चिम कोना

4.2.5.2 प्रयोगात्मक प्लॉट का दूसरा कोना

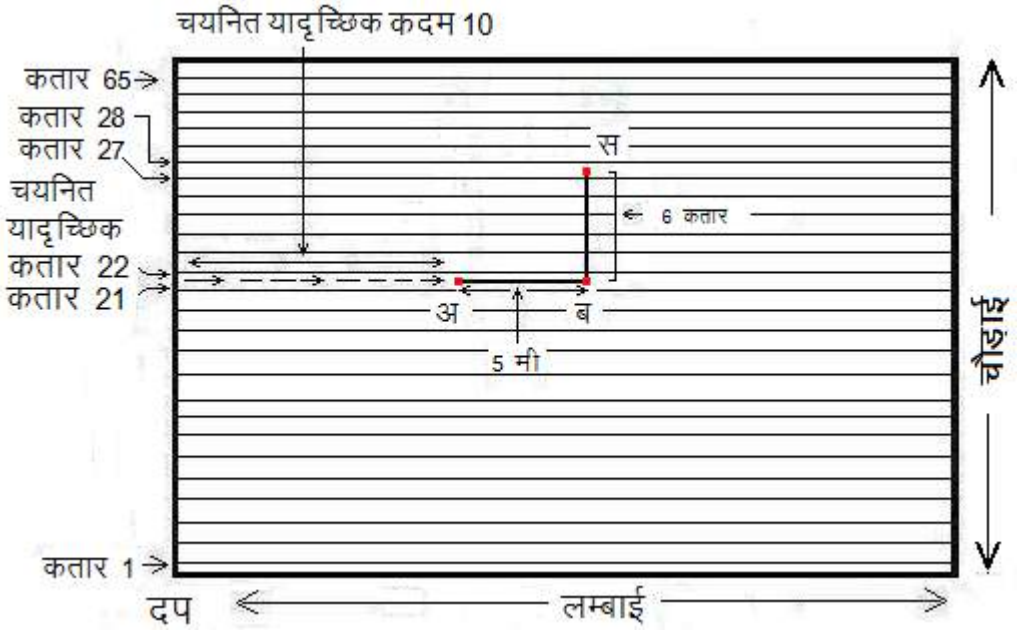
चयनित खेत की लम्बाई की दिशा में प्रयोगात्मक प्लॉट के मुख्य–बिन्दु "**अ**" से प्रयोगात्मक प्लॉट की लम्बाई (5 अथवा 10 मीटर) पर प्रयोगात्मक प्लॉट का दूसरा कोना "**ब**" कहलाता है। **ब**" बिन्दु पर दूसरी खूँटी गाड़ देते है (आकृति–4.2.5.2)।



आकृति–4.2.5.2: प्रयोगात्मक प्लॉट का दूसरा कोना

4.2.5.3 प्रयोगात्मक प्लॉट का तीसरा कोना

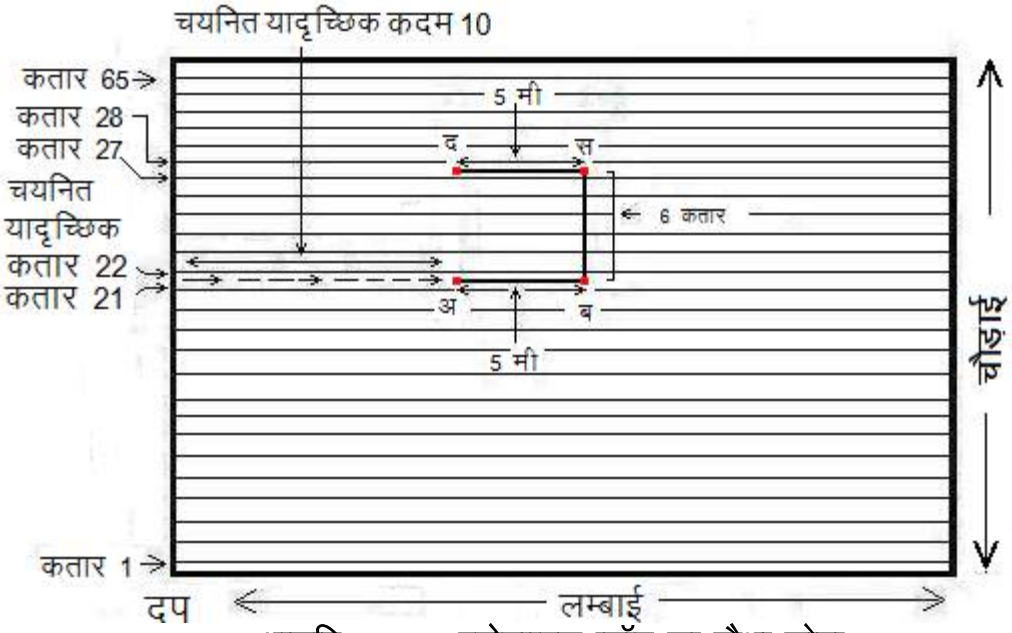
प्रयोगात्मक प्लॉट का तीसरा कोना चिन्हित करने हेतु "**ब**" बिन्दु से चयनित खेत की चौड़ाई की दिशा में प्रयोगात्मक प्लॉट की चौड़ाई में कतारों की औसत संख्या (6) गिनकर, प्रयोगात्मक प्लॉट की अन्तिम छठी कतार अर्थात चयनित खेत की सत्ताईसवीं कतार और इससे अगली अठाईसवीं कतार के बीच का बिन्दु प्रयोगात्मक प्लॉट का तीसरा कोना "**स**" कहलाता है। "**स**" बिन्दु पर तीसरी खूँटी गाड़ देते है (आकृति–4.2.5.3)। "**स**" बिन्दु पर खूँटी गाढ़ते समय ध्यान रहें कि "**स**" बिन्दु से "**ब**" बिन्दु को मिलाने वाली रेखा "**ब**" बिन्दु पर लम्ब बनाये।



आकृति–4.2.5.3: प्रयोगात्मक प्लॉट का तीसरा कोना

4.2.5.4 प्रयोगात्मक प्लॉट का चौथा कोना

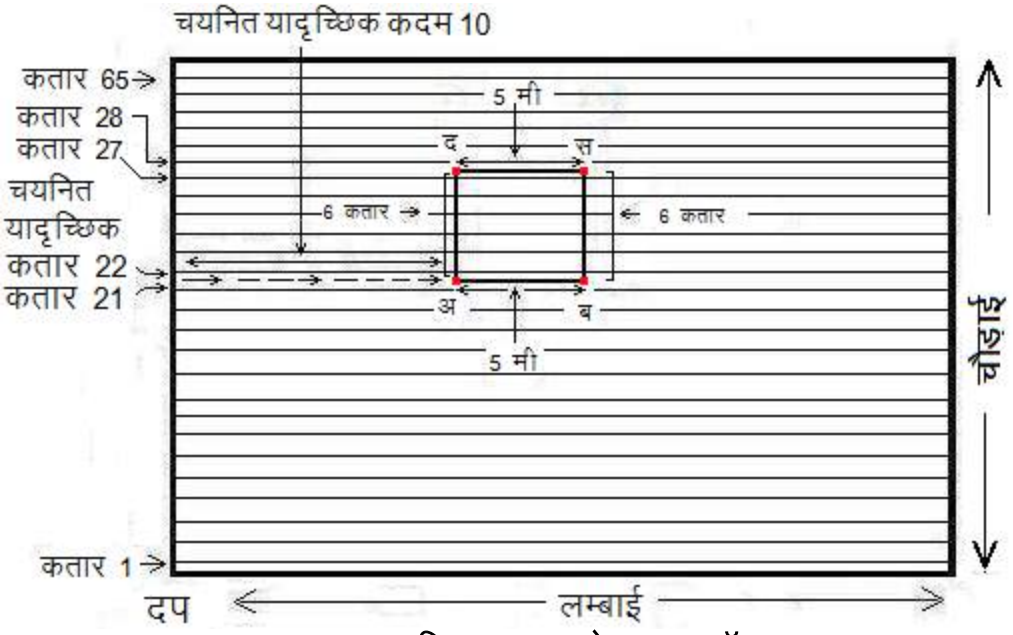
प्रयोगात्मक प्लॉट का चौथा कोना चिन्हित करने हेतु "**स**" बिन्दु से चयनित खेत की लम्बाई की दिशा में "**ब**" "**अ**" के समानांतर प्रयोगात्मक प्लॉट की लम्बाई (5 अथवा 10 मीटर) प्रयोगात्मक प्लॉट की अन्तिम छठी कतार अर्थात चयनित खेत की सत्ताईसवीं कतार और इससे अगली अठाईसवीं कतार कतार के बीच नापते हुए जिस बिन्दु पर पहुँचते है वह प्रयोगात्मक प्लॉट का चौथा कोना "द" कहलाता है। **"द"** बिन्दु पर चौथी खूँटी गाड़ देते है (आकृति–4.2.5.4)। **"द"** बिन्दु पर खूँटी गाढ़ते समय ध्यान रहें कि **"द"** बिन्दु से "**अ**" बिन्दु को मिलाने वाली रेखा "**अ**" बिन्दु पर लम्ब होनी चाहिए।



आकृति–4.2.5.4: प्रयोगात्मक प्लॉट का चौथा कोना

4.2.5.5 प्रयोगात्मक प्लॉट

प्रयोगात्मक प्लॉट के चारों कोनों ''अ ब स द'' पर खूँटों को अच्छी प्रकार भूमि में सीधा एवं मजबूत गाड़कर अ ब एवं स द के बीच की दूरी (5 मीटर) की ठीक से जाँच करनी चाहिए। इनके बीच की दूरी, प्रयोगात्मक प्लॉट की लम्बाई के बराबर होनी चाहिए। अ द एवं ब स के बीच कतारों की संख्या, प्रयोगात्मक प्लॉट की चौड़ाई में कतारों की औसत संख्या के बराबर होनी चाहिए (आकृति–4.2.5.5)।



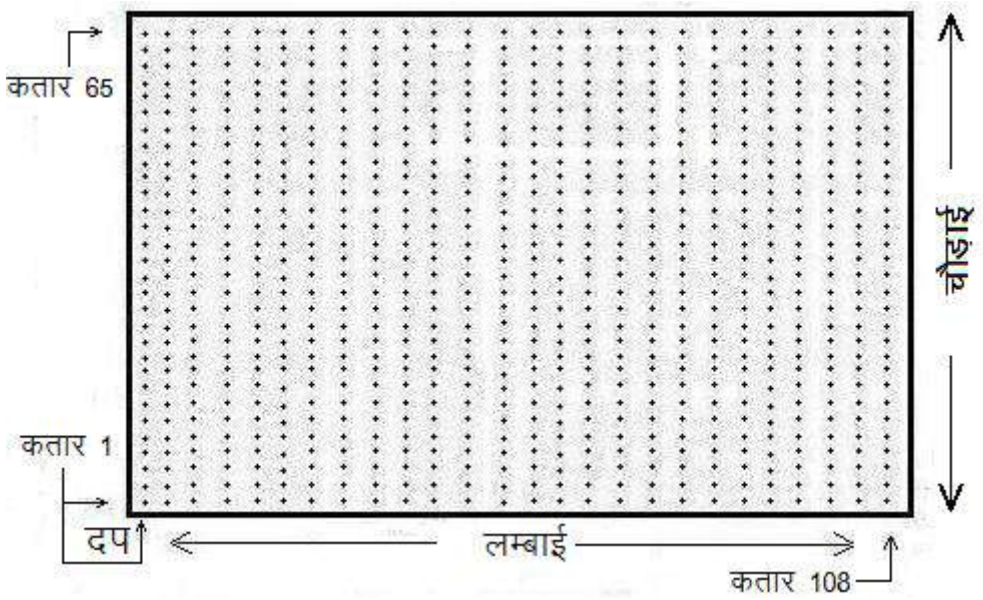
आकृति–4.2.5.5: प्रयोगात्मक प्लॉट

4.3 दोनों दिशाओं में स्पष्ट पंक्तियों में बोई गई फसलों की फसल कटाई प्रयोग विधि

तम्बाकू जैसी कुछ फसल स्पष्ट पंक्तियों में दोनों दिशाओं में की बोई जाती हैं। दोनों दिशाओं में बोई गई फसलों की फसल कटाई प्रयोग विधि बिना पंक्तियों एवं एक दिशा में पंक्तियों में बोई गई फसलों की फसल कटाई प्रयोग विधि से अलग हैं। खेत के चयन से लेकर चयनित खेत के दक्षिण–पश्चिम कोने के चिह्नांकन तक 3.2, 3.2.1, 3.2.2, 3.2.3, 3.2.4, 3.2.5, 4.1.1 में किये गये उल्लेख के समान ही है।

4.3.1 कतारों की गणना

चयनित खेत के दक्षिण–पश्चिम कोने से लम्बाई एवं चौड़ाई की दिशा में कतारों की गणना की जाती हैं (आकृति–4.3.1)। मानाकि चयनित खेत की लम्बाई एवं चौड़ाई की दिशा में क्रमशः 108 और 65 कतारें हैं।



आकृति–4.3.1: कतारों की गणना

4.3.2 लम्बाई एवं चौड़ाई में कतारों की औसत संख्या

प्रयोगात्मक प्लॉट की लम्बाई एवं चौड़ाई की दूरी में चयनित खेत की लम्बाई एवं चौड़ाई में यादृच्छिक रुप से तीन स्थानों पर कतारों के भौतिक अवलोकन कर कतारों की औसत संख्या की गणना की जाती है। मानाकि वर्गाकार आकृति के प्रयोगात्मक प्लॉट कीएक भुजा की लम्बाई 5 मीटर है और प्रयोगात्मक प्लॉट की लम्बाई एवं चौड़ाई में कतारों की औसत संख्या क्रमशः 6 एवं 8 है।

4.3.3 यादृच्छिक संख्याओं का चयन

प्रयोगात्मक प्लॉट का दक्षिण–पश्चिम कोने के चिह्नांकन हेतु चयनित खेत की लम्बाई एवं चौड़ाई की दिशा में यादृच्छिक कतार हेतु यादृच्छिक संख्या का चयन किया जाता है।

4.3.3.1 लम्बाई में यादृच्छिक कतार

चयनित खेत की लम्बाई की दिशा में कुल कतारों से प्रयोगात्मक प्लॉट की लम्बाई में औसत कतारों को घटाने पर शेषफल में चयनित खेत की लम्बाई की दिशा में अन्तिम कतार को प्रतिदर्श में शामिल करने के लिए **एक** जोड़ा जाता है। एक जोड़ने के बाद प्राप्त संख्या के अनुसार यादृच्छिक संख्या तालिका के निर्दिष्ट स्तंभ **एक** का उपयोग करके यादृच्छिक संख्या का चयन करना चाहिए। प्रयोगात्मक प्लॉट की लम्बाई की दिशा में यादृच्छिक कतार ज्ञात करने के लिए गणना निम्न प्रकार होगी।

उदाहरणः 4.3.3.1

चयनित खेत की लम्बाई की दिशा में कतारों की कुल संख्या	=	108
प्रयोगात्मक प्लॉट की लम्बाई की दिशा में कतारों की औसत संख्या	=	6
चयनित खेत की कुल कतारों से प्रयोगात्मक प्लॉट की औसत कतार घटाने पर शेषफल	=	102
प्रतिदर्श में अन्तिम कतार को शामिल करने के लिए जोड़ी जाने वाली संख्या	=	1
शेषफल में एक जोड़ने पर प्राप्त संख्या	=	103

उदाहरणः 4.3.3.1 में कतारों की कुल संख्या 103 के अनुसार तीन अंकों की यादृच्छिक संख्या तालिका के निर्दिष्ट स्तंभ **एक** का उपयोग करते हुए 103 अथवा 103 से कम जो भी यादृच्छिक संख्या सबसे पहले आती है उसका चयन करना चाहिए। माना सबसे पहले 48 यादृच्छिक संख्या आती है। जो 103 से कम है। अतः 48वीं कतार को यादृच्छिक कतार के रुप में चयनित माना जाता है।

4.3.3.2 चौड़ाई में यादृच्छिक कतार

चयनित खेत की चौड़ाई की दिशा में कुल कतारों की संख्या से प्रयोगात्मक प्लॉट की चौड़ाई में औसत कतारों को घटाने पर शेषफल में चयनित खेत की चौड़ाई की दिशा में अन्तिम कतार को प्रतिदर्श में शामिल करने के लिए **एक** जोड़ा जाता है। एक जोड़ने के बाद प्राप्त संख्या के अनुसार यादृच्छिक संख्या तालिका का निर्दिष्ट स्तंभ **एक** का उपयोग करके यादृच्छिक संख्या का चयन करना चाहिए। प्रयोगात्मक प्लॉट की चौड़ाई की दिशा में यादृच्छिक कतार ज्ञात करने के लिए गणना निम्न प्रकार होगी।

उदाहरणः 4.3.3.2

चयनित खेत की चौड़ाई की दिशा में कतारों की कुल संख्या	=	65
प्रयोगात्मक प्लॉट की चौड़ाई की दिशा में कतारों की औसत संख्या	=	8
चयनित खेत की कुल कतारों से प्रयोगात्मक प्लॉट की औसत कतार घटाने पर शेषफल	=	57
प्रतिदर्श में अन्तिम कतार को शामिल करने के लिए जोड़ी जाने वाली संख्या	=	1
शेषफल में एक जोड़ने पर प्राप्त संख्या	=	58

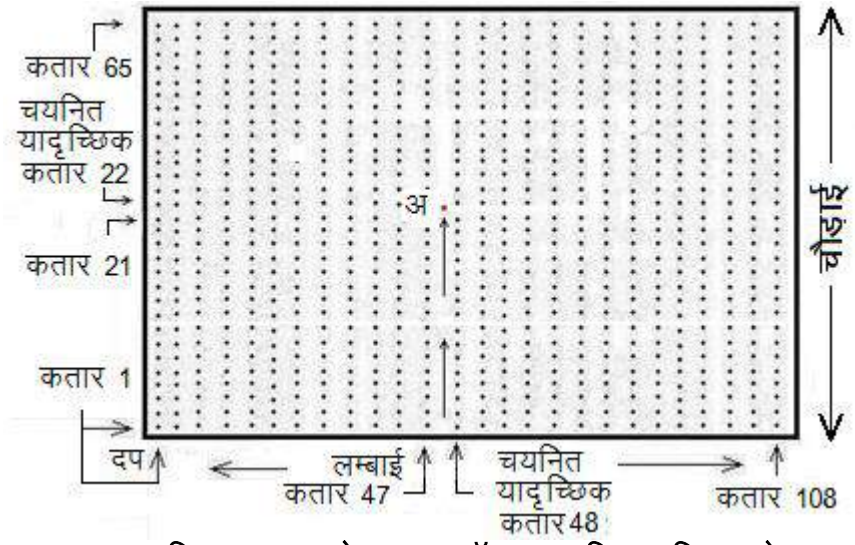
उदाहरणः 4.3.3.2 में कतारों की कुल संख्या 58 के अनुसार दो अंकों की यादृच्छिक संख्या तालिका के निर्दिष्ट स्तंभ एक का उपयोग करते हुए 58 अथवा 58 से कम यादृच्छिक संख्या का चयन करना चाहिए। माना निर्दिष्ट स्तंभ का उपयोग करते हुए सबसे पहले 22 यादृच्छिक संख्या आती है। जो 58 से कम है। अतः बाईसवीं कतार को यादृच्छिक कतार के रुप में

चयनित माना जाता है। प्रयोगात्मक प्लॉट का दक्षिण–पष्चिम कोना चयनित यादृच्छिक संख्या जोड़ा (48, 22) की सहायता से निर्धारित करते है।

4.3.4 प्रयोगात्मक प्लॉट तैयार करना

4.3.4.1 प्रयोगात्मक प्लॉट का दक्षिण–पश्चिम कोना

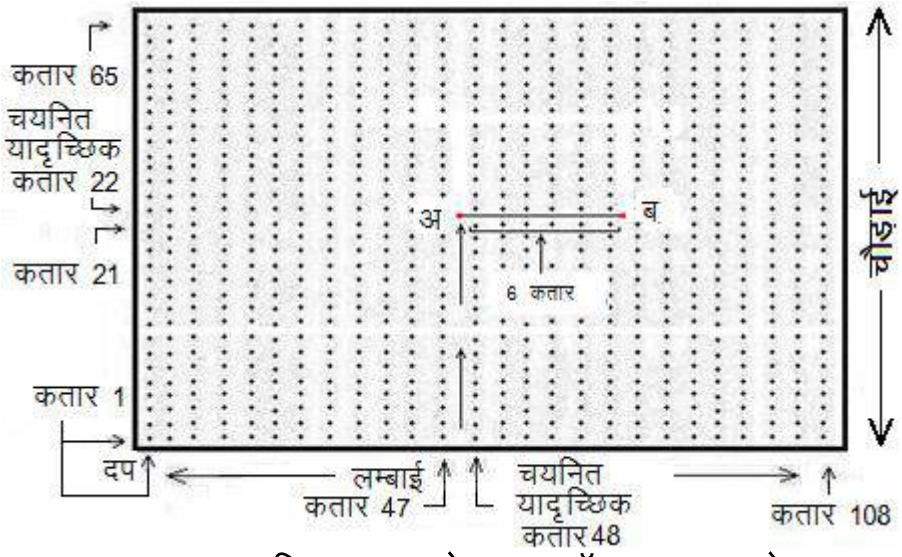
चयनित खेत के दक्षिण–पष्चिम कोने से लम्बाई की दिशा में कतारों को गिनते हुए लम्बाई के लिए चयनित यादृच्छिक कतार (48) पर रुकते है तथा यहाँ से चयनित यादृच्छिक कतार (48) और इससे पहली कतार (47) के बीच चयनित खेत की लम्बाई के लम्बवत्त एवं चयनित खेत की चौड़ाई के समानांतर चलते हुए, चौड़ाई के लिए चयनित यादृच्छिक कतार (22) पर रुकते है। चयनित यादृच्छिक कतार (22) एवं इससे पहली कतार (21) के बीच बिन्दु "**अ**" पहली खूँटी गाड़ते है। बिन्दु "**अ**" प्रयोगात्मक प्लॉट का दक्षिण–पश्चिम कोना अथवा मुख्य–बिन्दु (Key-Point) अथवा पहला कोना कहलाता है (आकृति–4.3.4.1)।



आकृति–4.3.4.1: प्रयोगात्मक प्लॉट का दक्षिण–पश्चिम कोना

4.3.4.2 प्रयोगात्मक प्लॉट का दूसरा कोना

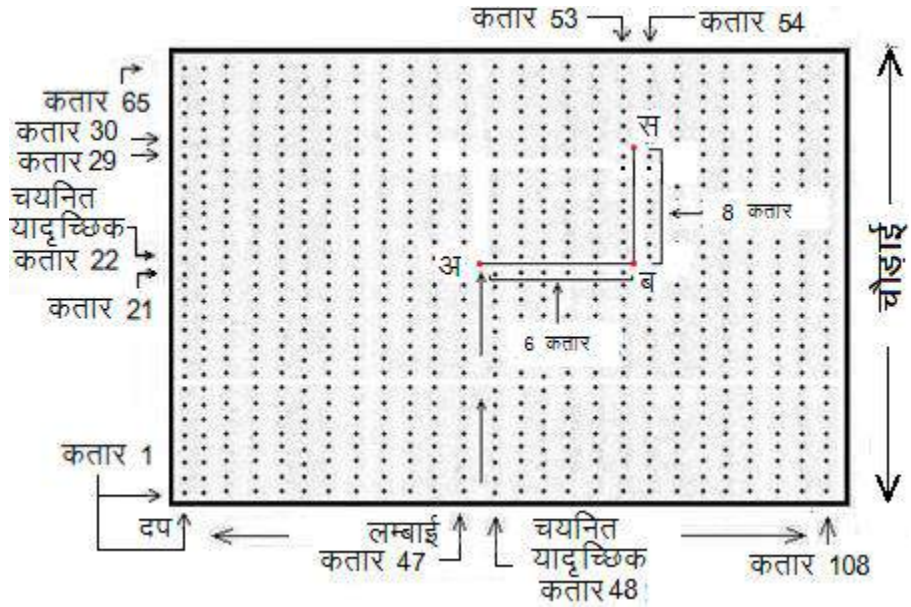
प्रयोगात्मक प्लॉट के मुख्य–बिन्दु "**अ**" से चयनित खेत की लम्बाई की दिशा में प्रयोगात्मक प्लॉट की लम्बाई में कतारों की औसत संख्या (6) की गिनती चयनित यादृच्छिक कतार (48) से की जाती है। चयनित खेत की अड़तालीसवीं एवं तिरपनवीं कतार प्रयोगात्मक प्लॉट की क्रमशः पहली और अन्तिम अर्थात छठ वीं कतार है। चयनित खेत की तिरपनवीं कतार और इससे अगली चौवनवीं कतार के बीच दूसरी खूँटी गाड़ते है। यह प्रयोगात्मक प्लॉट का दूसरा कोना "**ब**" कहलाता है (आकृति–4.3.4.2)।



आकृति–4.3.4.2: प्रयोगात्मक प्लॉट का दूसरा कोना

4.3.4.3 प्रयोगात्मक प्लॉट का तीसरा कोना

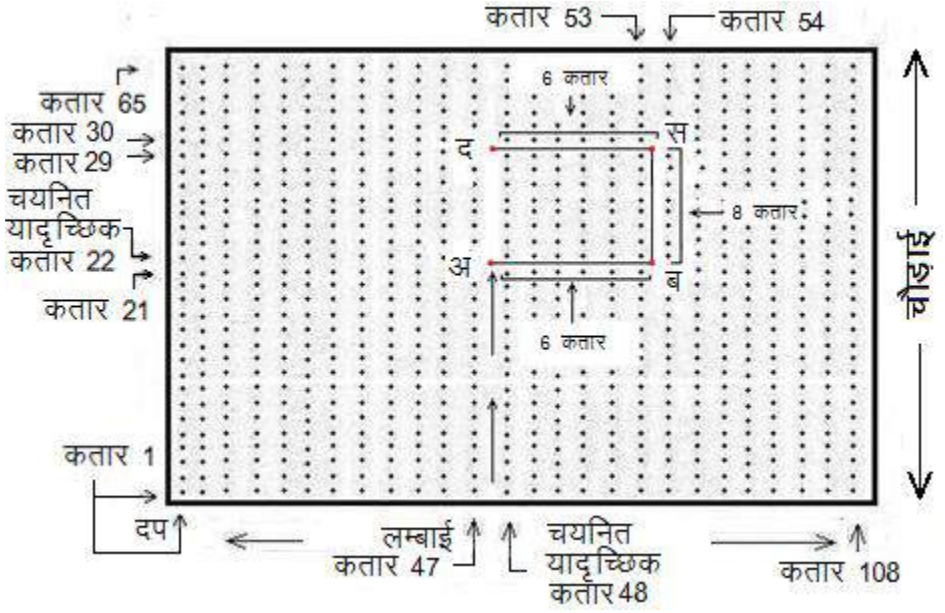
प्रयोगात्मक प्लॉट की चौड़ाई में कतारों की औसत संख्या आठ है। अतः प्रयोगात्मक प्लॉट का तीसरा कोना चिन्हित करने हेतु "**ब**" खूँटी से चयनित खेत की चौड़ाई की दिशा में बाईसवीं चयनित यादृच्छिक कतार प्रयोगात्मक प्लॉट की पहली और उनतीसवीं अर्थात आठवीं कतार है। चयनित खेत की उनतीसवीं और इससे अगली तीसवीं कतार के बीच तीसरी खूँटी गाड़ते है। यह प्रयोगात्मक प्लॉट का तीसरा कोना "**स**" कहलाता है (आकृति–4.3.4.3)। "**स**" बिन्दु पर खूँटी गाढ़ते समय ध्यान रहें कि "**स**" बिन्दु से "**ब**" बिन्दु को मिलाने वाली रेखा "**ब**" बिन्दु पर लम्ब बनाये।



आकृति–4.3.4.3: प्रयोगात्मक प्लॉट का तीसरा कोना

4.3.4.4 प्रयोगात्मक प्लॉट का चौथा कोना

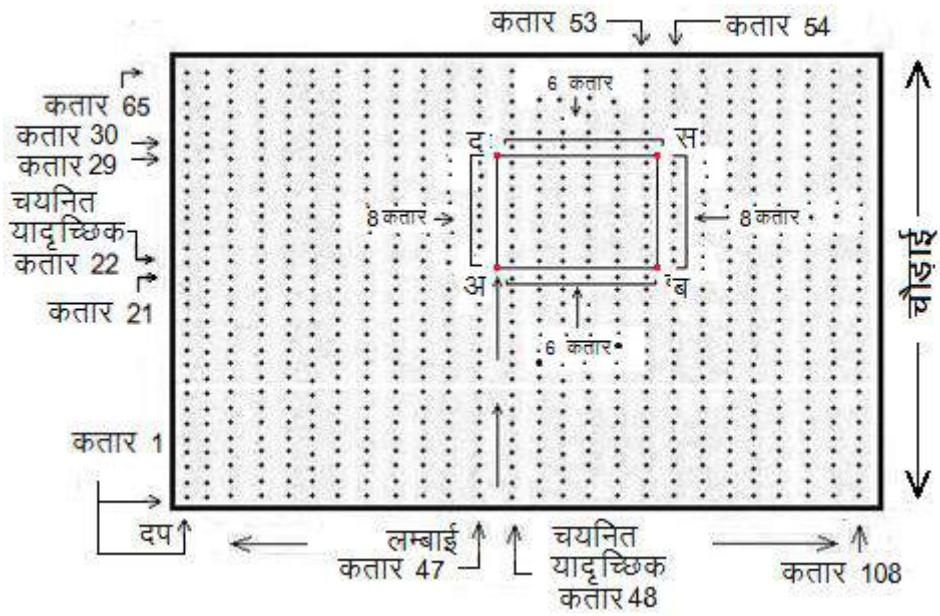
प्रयोगात्मक प्लॉट का चौथा कोना चिन्हित करने हेतु "**स**" बिन्दु से चयनित खेत की लम्बाई में "**ब**" "**अ**" के समानांतर उनतीसवीं एवं तीसवीं कतार के बीच चलते हुए प्रयोगात्मक प्लॉट की लम्बाई में कतारों की औसत संख्या (छह) गिनकर चौथी खूँटी गाड़ देते है। यह प्रयोगात्मक प्लॉट का चौथा कोना "द" कहलाता है। चयनित खेत की ये वही छह कतारें है जो प्रयोगात्मक प्लॉट की खूँटी "**अ**" और "**ब**" के बीच ("**अ**" से "**ब**" की दिशा में, चयनित खेत की अड़तालीसवीं और तिरपनवीं कतार क्रमशः प्रयोगात्मक प्लॉट की पहली और छठवीं अर्थात "**स**" खूँटी से **"द"** खूँटी की दिशा में, तिरपनवीं कतार पहली और अड़तालीसवीं कतार छठी) हैं। प्रयोगात्मक प्लॉट का चौथा कोना, पहले कोने "**अ**" की तरह अड़तालीसवीं और सैंतालीसवीं कतार के बीच होगा (आकृति–4.3.4.4)।



आकृति–4.3.4.4: प्रयोगात्मक प्लॉट का चौथा कोना

4.3.4.5 प्रयोगात्मक प्लॉट

प्रयोगात्मक प्लॉट के चारों कोनों ''अ ब स द'' पर खूंटों को अच्छी प्रकार भूमि में सीधा गाड़कर, अ ब एवं स द और अ द एवं ब स के बीच कतारों की औसत संख्या की ठीक से जाँच करनी चाहिए। अ ब एवं स द के बीच कतारों की संख्या, प्रयोगात्मक प्लॉट की लम्बाई में कतारों की औसत संख्या छह और अ द एवं ब स के बीच कतारों की संख्या, प्रयोगात्मक प्लॉट की चौड़ाई में कतारों की औसत संख्या आठ के बराबर होनी चाहिए (आकृति–4.3.4.5)। प्रयोगात्मक प्लॉट की लम्बाई एवं चौड़ाई को नापकर भी लिखना चाहिए।



आकृति–4.3.4.5: प्रयोगात्मक प्लॉट

5. प्रयोगात्मक प्लॉट की फसल कटाई

एक खूंटे के बाहरी किनारे से दूसरे खूंटे के बाहरी किनारे तक की दूरी, चारों कोनों पर भूमि में गड़ें खूटें की सीधाई एवं मजबूती की जाँच के बाद प्रयोगात्मक प्लॉट की सीमा के निर्धारण हेतू खींचने पर लम्बाई न बढ़ने वाली एक मजबूत रस्सी को चारों खूटों के बाहरी ओर से होते हुए बाँधकर धीरे–धीरे भूमि के समीप लाना चाहिए। जब रस्सी भूमि के समीप आये तो पौधों की जडों की स्थिति का निर्णय करना चाहिए। केवल उन पौधों की कटाई करनी चाहिए जिनकी जडें प्रयोगात्मक प्लॉट की सीमा पर आधे से ज्यादा अन्दर की ओर हैं। प्रयोगात्मक प्लॉट के पौधों की सावधानीपूर्वक कटाई करके, बंड़ल बनाकर, बंड़लों पर उचित बिल्ला लगाकर, खलिहान में सुखाने, गाहना और सफाई हेतु रखना चाहिए। ध्यान रहें कि कटाई करने, बंड़ल बनाने, बिल्ला लगाने और प्रयोगात्मक प्लॉट से खलिहान तक ले जाने में पौधों से बालियाँ एवं बालियों से दाने गिरने नहीं चाहिए।

5. प्रयोगात्मक प्लॉट की फसल कटाई

6. गाहना

खलिहान में पौधों को टाट–पट्टी पर सुखाने के बाद प्रचलित गाहाई विधि द्वारा गाहना चाहिए। ध्यान रहें कि खलिहान में पौधों को गाहते समय दाने इधर–उधर नहीं बिखरने चाहिए।

6. प्रयोगात्मक प्लॉट की फसल की गाहाई

7. ओसाई एवं सफाई

गाहने के बाद, उत्पाद को हवा अथवा ओसाई पंखा की मद्द से अलग करना चाहिए। ध्यान रहें, भूसा के साथ दाने न रहे। उत्पाद में किसी अन्य फसल के दाने, खरपतवार के बीज, कूड़ा–कर्कट, धूल, मिट्टी, कंकड़ इत्यादि भी नहीं होना चाहिए।

7. ओसाई एवं सफाई

8. उत्पाद का वजन

ओसाई के तुरन्त बाद उत्पाद का वजन करना चाहिए। पहली बार किया गया वजन उत्पाद का गिला वजन कहलाता है। वजन तौल की न्यूनतम इकाई तक करना चाहिए।

8. उत्पाद का वजन

9. शुष्कता प्रयोग एवं पुनः वजन

उपज दर के अन्तिम अनुमान के लिए उत्पाद पूर्ण रुप से सूखा होना चाहिए। जिन फसलों के उत्पाद में ओसाई एवं सफाई के समय नमी की मात्रा अधिक होती है उनके लिए शुष्कता प्रयोग करना अति आवश्यक है। शुष्कता प्रयोग प्रत्येक प्राथमिक कार्यकर्ता को करना आवश्यक नहीं है। जिला सांख्यिकीय स्तर पर जिला सांख्यिकीय पर्यवेक्षकों द्वारा पर्यवेक्षित फसल कटाई

प्रयोगों को शुष्कता प्रयोग के लिए चुना जाता है। विशिष्ठ फसलों के शुष्कता प्रयोग उनके निर्धारित फसल कटाई प्रयोगों की संख्या का 15 प्रतिशत अथवा प्रति फसल न्यूनतम 4 शुष्कता प्रयोग किये जाते है। जिला सांख्यिकीय पर्यवेक्षक द्वारा यादृच्छिक विधि से प्रयोगात्मक प्लॉट के उत्पाद से नमूना लिया जाता है। आमतौर पर नमूना की मात्रा एक किलोग्राम होती है। यदि कुल उत्पाद एक किलोग्राम से कम है तो संपूर्ण उत्पाद शुष्कता प्रयोग के लिए लिया जाता है। उत्पाद का नमूना कपड़े की थैली में लेकर एक निशचित अवधि के लिए सामान्य रुप से धूप में सूखने के बाद उत्पाद का दूसरी बार वजन करते है। दूसरी बार किया गया वजन उत्पाद का सूखा वजन कहलाता है। गन्ना के अन्तिम उत्पाद को गन्नो के वजन और कपास के अन्तिम उत्पाद को रुई के वजन के रुप में व्यक्त करते है। कपास का केवल एक बार वजन किया जाता है जो कपास को चुनकर उसी समय थोड़ी देर टाट–पट्टी पर धूप में सुखाने के बाद किया जाता है। कपास को रुई में परिवर्तित करने के लिए, ओटाई प्रतिशत ओटाई कारखानों से लिया जाता है। इसके अलावा गैंहू, जौ, मक्का, सरसों, बाजरा, ज्वार इत्यादि में पहली बार वजन करने पर उत्पाद में नमी की मात्रा अधिक होती है अतः इन्हे सुखाने के बाद दूसरी बार वजन किया जाता है।

संदर्भः

1. महालानोविश, पी. सी. (1945): "ए रिपोर्ट ऑन बिहार क्रॉप सर्वे" , 1943–44, सांख्या, 7, 29–118.
2. पानसे वी.जी. (1946 बी): "प्लॉट साईज इन ईल्ड सर्वेज ऑन कॉटन", कर्नट साइंस, 15, 218–19।
3. पानसे वी.जी. (1947): " प्लॉट साईज इन ईल्ड सर्वेज", नेचर, 15, 159, 820।
4. सूखात्मे पी.वी. (1946 ए): बायस इन दि यूज ऑफ स्मॉल साईज प्लॉटस इन सैम्पल सर्वेज फॉर ईल्ड", कर्नट साइंस, 15, 119–20 ।
5. सूखात्मे पी.वी. (1946 बी): "बायस इन दि यूज ऑफ स्मॅल साईज प्लॉटस इन सैम्पल सर्वेज फॉर ईल्ड", नेचर, 15, 7, 630।
6. सूखात्मे पी.वी. (1947 ए): "दि प्राब्लम ऑफ प्लॉट साईज इन लार्ज–स्केल ईल्ड सर्वेज", जर्नल ऑफ अमेरिकन स्टैटिस्टिकस एसोसिऐसन, 42, 297–310.
7. सूखात्मे पी.वी. एवं पान्से वी.जी. (1951): "क्रॉप सर्वेज इन इन्डिया–।।", स्टैटिस्टिकल जर्नल ऑफ इन्डियन सोसायटी एग्रीकल्चरल स्टैटिस्टिकस, वोल्यूम।।।:(2), 95–168।

अध्याय 12

आँकड़ों का विज़ुअलाइज़ेशन एवं अन्वेषणात्मक विश्लेषण

दीपक सिंह
वैज्ञानिक, प्रतिदर्श सर्वेक्षण प्रभाग
भा.कृ.अनु.प.-भारतीय कृषि साँख्यिकी अनुसंधान संस्थान, लाइब्रेरी एवेन्यू, नई दिल्ली-110012

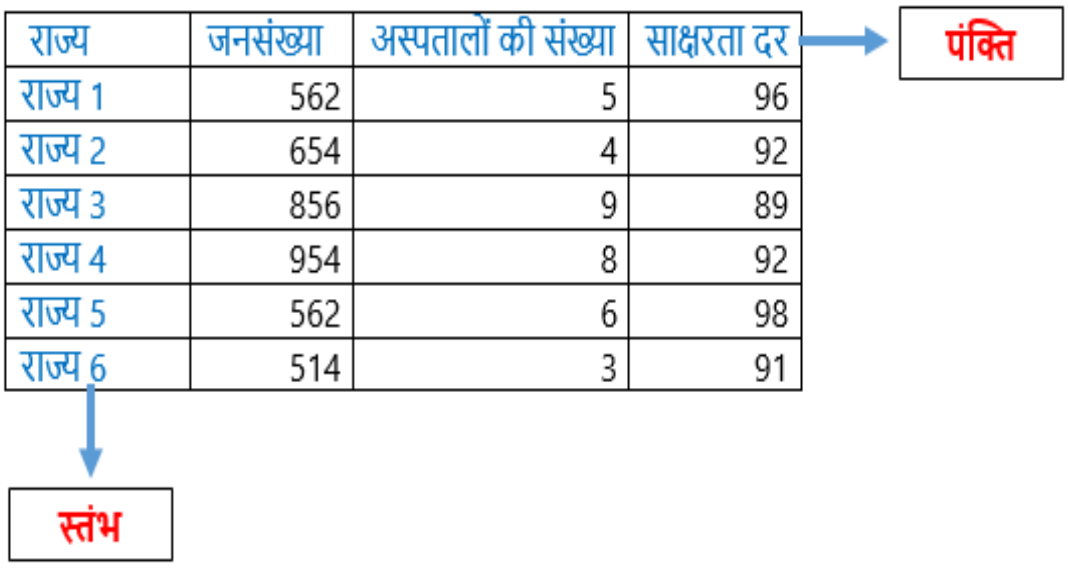
1. प्रस्तावना (Introduction):

आँकड़ों का विज़ुअलाइज़ेशन एवं अन्वेषणात्मक विश्लेषण, विश्लेषण के क्षेत्र में महत्वपूर्ण है। यह किसी भी डेटा सेट के अंदर छिपे जानकारी को सुलभ तरीके से प्रस्तुत करने में मदद करता है। आँकड़ों को दृश्यात्मक रूप में प्रस्तुत करने से उन्हें समझने में आसानी होती है। यह डेटा के पैटर्न, प्रवृत्ति और आउटलायर्स की पहचान करने में मदद करता है। सामान्यतः आँकड़े तीन प्रकार से प्रस्तुत किये जा सकते हैं :वर्णनात्मक प्रस्तुतीकरण, सारणीबद्ध प्रस्तुतीकरण, आरेखीय प्रस्तुतीकरण ।

1.1 वर्णनात्मक प्रस्तुतीकरण (Descriptive Presentation): पाठ-विषयक या वर्णनात्मक प्रस्तुतीकरण में पाठ के रूप में आँकड़ों का विवरण किया जाता है। यह विधि आँकड़ों के प्रस्तुतीकरण के लिए उपयोगी होती है जब आँकड़ों का परिणाम अधिक न हो।

1.2 सारणीबद्ध प्रस्तुतीकरण (Tabular Presentation): सारणीबद्ध प्रस्तुतीकरण डेटा को एक सारणी (पंक्ति और स्तंभों) के रूप में प्रस्तुत करने का एक तरीका है, जिसमें डेटा के विभिन्न पहलुओं को श्रेणियों में विभाजित किया जाता है। उदाहरण के लिए:

राज्य	जनसंख्या	अस्पतालों की संख्या	साक्षरता दर
राज्य 1	562	5	96
राज्य 2	654	4	92
राज्य 3	856	9	89
राज्य 4	954	8	92
राज्य 5	562	6	98
राज्य 6	514	3	91

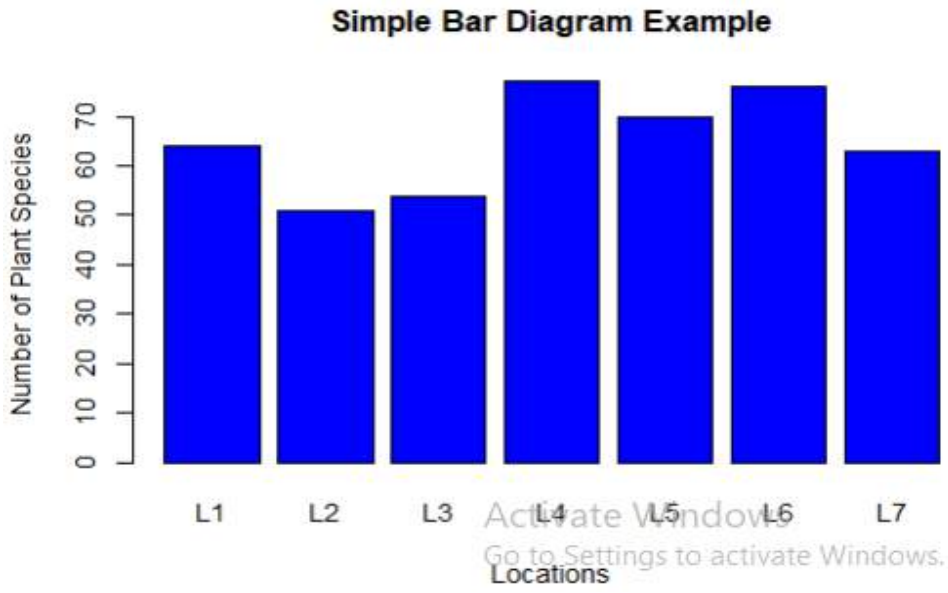


चित्र 1: आँकड़ों का सारणीबद्ध प्रस्तुतीकरण

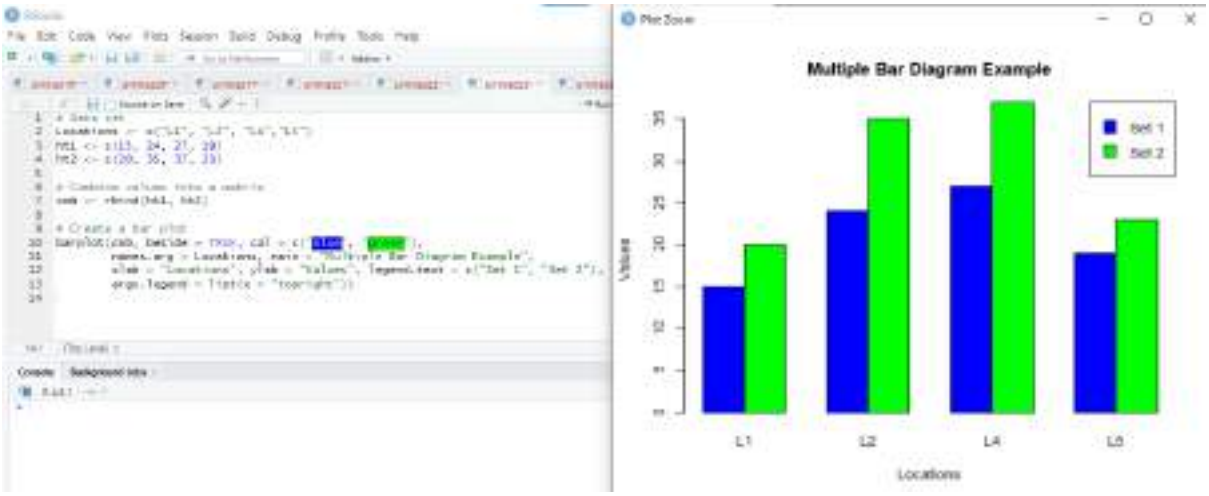
1.3 आरेखीय प्रस्तुतीकरण (Graphical Presentation): प्रस्तुतीकरण की यह विधि वर्णनात्मक एवं सारणीबद्ध प्रस्तुतीकरण की तुलना में आँकड़ों के आधार पर वस्तु-स्थिति को समझने में अधिक उपयोगी है। आंकड़ों के प्रस्तुतीकरण के लिये विभिन्न प्रकार के आरेखो का उपयोग किया जाता है जिनमें से कुछ इस प्रकार हैं :

दंड आरेख (Bar diagrams): दंड आरेख आँकड़ों को दृश्यात्मक रूप में प्रस्तुत करने का एक प्रमुख तरीका है जो डेटा के विभिन्न श्रेणियों या विकल्पों के बीच तुलना करने में मदद करता है। यह आरेख बारों की ऊंचाई को उपयोग करके डेटा को प्रस्तुत करता है, जिनमें प्रत्येक बार एक श्रेणी के मान को दर्शाता है। दंड आरेख के विभिन्न प्रकार के होते हैं जैसे कि सरल दंड आरेख, बहु दंड आरेख एवं घटक दंड आरेख इत्यादि ।

```
> library(readxl)
> data <- read_excel("C:/Users/Bharti/Desktop/Hindi workshop 2023/Hindi Workshop 2023/Lectures/Book1 Hindi w
orkshop.xlsx",
+     sheet = "Sheet4")
> View(data)
> attach(data)
> barplot(data$`Number of Plant Species`, names.arg = data$Location, col = "blue", main = "Simple Bar Diagra
m Example", xlab = "Locations", ylab = "Number of Plant Species")
>
```



चित्र 2: R में सरल दंड आरेख

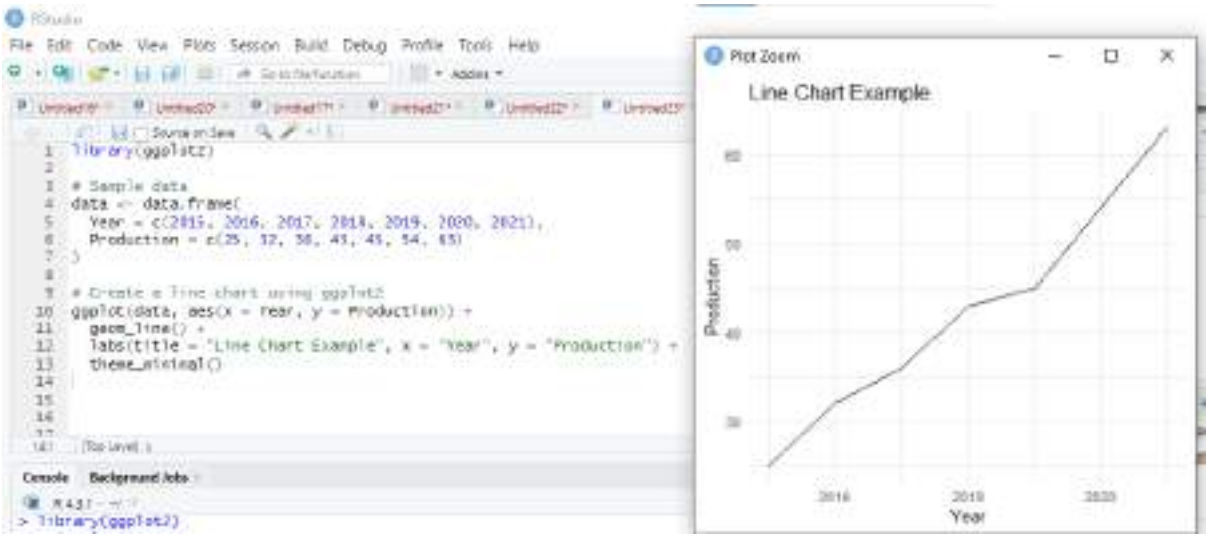


चित्र 3: R में बहु दंड आरेख

वृत्त आरेख (Pie Chart): वृत्त आरेख भी एक घटक आरेख है लेकिन घटक दंड आरेखों के स्थान पर इसे एक ऐसे वृत्त द्वारा प्रस्तुत किया जाता है, जिसके क्षेत्र को आनुपातिक रूप से उन घटकों में विभाजित किया जाता है, जिन्हें यह दर्शाता है।

चित्र 4: R में बहु वृत्त आरेख

रेखा चित्र (Line Diagram): रेखा चित्र का अधिकतर उपयोग डेटा के समय के साथ परिवर्तन को प्रदर्शित करने के लिए किया जाता है, जिससे ट्रेंड और पैटर्न की पहचान की जा सकती हैं। यह डेटा के स्थायिता और परिवर्तनों को दिखाने का माध्यम होता है।



चित्र 5: R में रेखा चित्र

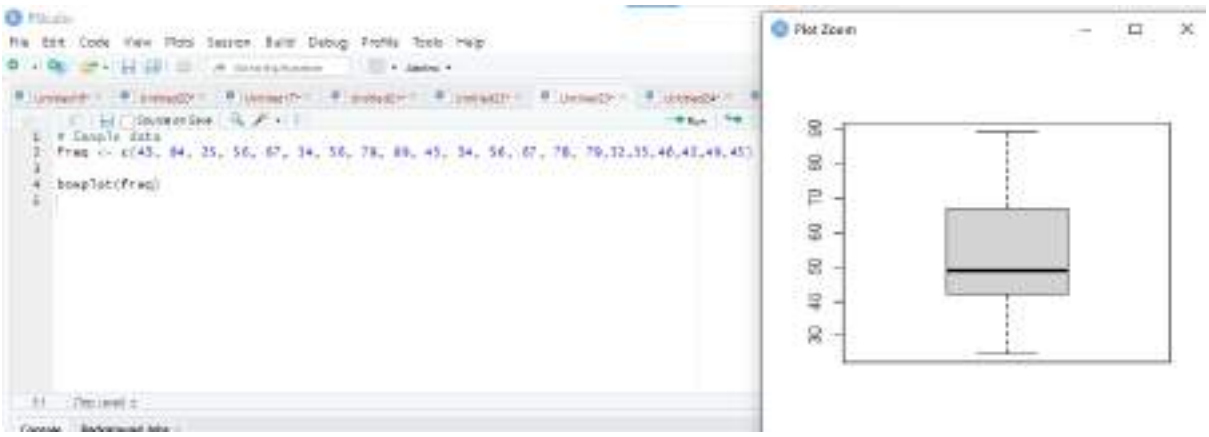
2. आयतचित्र (Histogram):

सांख्यिकी में आयतचित्र आँकड़े वितरण का ग्राफीय निरुपण है। आयतचित्र सारणीबद्ध आवृत्तियों का निरूपण है जिसे असतत अंतराल पर खड़े आयत द्वारा निरुपित किया जाता है। इसमें सम्बंधित अन्तराल का क्षेत्रफल प्रेक्षण की आवृत्ति के अनुक्रमानुपाति होता है। आयत की ऊँचाई भी सम्बंधित अन्तराल के आवृत्ति घनत्व अर्थात् आवृत्ति और अन्तराल चौड़ाई के अनुपात के समान होती है। आयतचित्र का कुल क्षेत्रफल आँकड़ों की कुल संख्या के समान होता है। एक आयतचित्र को सम्बंधित आवृत्ति के साथ प्रसामान्यकृत करके भी प्रदर्शित किया जा सकता है।

चित्र 6: R में आयतचित्र

3. **बॉक्स प्लॉट (Boxplot)**: बॉक्स प्लॉट (Boxplot) एक रेखा चित्र है जिसे डेटा के एक सेट की पांच-संख्या का सारांश प्रदर्शित करने के लिए उपयोग किया जाता है। पांच-संख्या के सारांश में न्यूनतम,

अधिकतम, माध्य, और पहला व तीसरा चतुर्थक होता है। यह डेटा के एक सेट का सांख्यिकी वितरण दिखाता है।



चित्र 7: R में Boxplot

संदर्भ (Reference):

R विकास कोर टीम (2010): सांख्यिकीय गणना के लिए एक भाषा एवं वातावरण, संख्यिकीय गणना के लिए **R** फाउण्डेशन, वियाना, आस्ट्रिया, URL: http://www.R-project.org

अध्याय 13

सहसंबंध एवं प्रतिगमन विश्लेषण तकनीकें

राहुल बनर्जी
वैज्ञानिक, प्रतिदर्श सर्वेक्षण प्रभाग
भा.कृ.अनु.प.-भारतीय कृषि साँख्यिकी अनुसंधान संस्थान, लाइब्रेरी एवेन्यू, नई दिल्ली-110012

1. प्रस्तावना (Introduction):

दो या दो से अधिक परिमाणों के बीच संबंध का अध्ययन करने के लिए सहसंबंध के माध्यम से, उनके बीच संबंध को ग्राफिकल रूप में दर्शाना महत्वपूर्ण है। संबंध की गुणवत्तात्मक जांच के लिए सबसे सरल तरीका डेटा को चित्रित करना है। बिखरती चित्र एक द्विपरिमाणीय डेटा के चित्रित प्रतिनिधित्व का सबसे सरल तरीका है। यह डेटा के वितरण की विचार को देता है जैसे कि स्पष्ट रूप से सकारात्मक या नकारात्मक रूप से रेखांश संबंध, गैर-रेखांश संबंध या कोई प्रत्यक्ष संबंध नहीं।

2. सहसंबंध विश्लेषण (Correlation Analysis):

वस्तुओं के प्रत्येक सेट पर संबंधित मापों (एक्स और वाई) की एक जोड़ी को देखते हुए, सहसंबंध गुणांक (आर) उस डिग्री का एक सूचकांक प्रदान करता है जिससे युग्मित माप एक रैखिक फैशन में सह-भिन्न होते हैं। सामान्य तौर पर r सकारात्मक होगा जब X के बड़े मान वाले आइटम में Y का मान भी बड़ा होगा जबकि X के छोटे मान वाले आइटम में Y का मान छोटा होगा। तदनुसार, r तब नकारात्मक होगा जब X के बड़े मान वाले आइटम में Y का मान बड़ा होगा Y के छोटे मान होते हैं जबकि X के छोटे मान वाले आइटम में Y के बड़े मान होते हैं। संख्यात्मक रूप से, r रैखिक संबंध की डिग्री के आधार पर -1 और +1 के बीच कोई भी मान मान सकता है। प्लस और माइनस एक पूर्ण सकारात्मक और नकारात्मक संबंधों को दर्शाता है जबकि शून्य इंगित करता है कि एक्स और वाई मान किसी भी रैखिक फैशन में सह-भिन्न नहीं होते हैं। इसे पियर्सन-उत्पाद-आघूर्ण सहसंबंध गुणांक भी कहा जाता है। सहसंबंध गुणांक के मानों की कोई इकाई नहीं होती। सहसंबंध विश्लेषण में हम किसी भी चर को प्रभावित नहीं करते हैं, बल्कि केवल उन्हें मापते हैं और चर के कुछ सेट, जैसे रक्तचाप और कोलेस्ट्रॉल स्तर के बीच संबंधों (सहसंबंध) की तलाश करते हैं। सहसंबंधी अनुसंधान के डेटा की हमारे पास मौजूद कुछ सिद्धांतों के आधार पर केवल कारणात्मक संदर्भ में "व्याख्या" की जा सकती है, लेकिन सहसंबंधी डेटा निर्णायक रूप से कार्य-कारण को साबित नहीं कर सकता है।

3. द्विचर एवं आंशिक सहसंबंध (Bivariate and Partial Correlation):

कोरलेशन संख्या दो मात्रात्मक परिमाणों के बीच एक रैखिक संबंध की मजबूती को मापती है। दो चरणीय परिमाणों के बीच रैखिक संबंध का सबसे आमतौर पर उपयोग किया जाने वाला माप "पियरसन-उत्पाद-क्षण कोरलेशन संख्या" या संक्षेप में नमूना कोरलेशन संख्या होता है और इसे r से दर्शाया जाता है। कोरलेशन संख्या के मानों को डेटा की इकाइयों में व्यक्त नहीं किया जाता है, लेकिन -1 से +1 के बीच होते हैं। हालांकि बिखरती चित्र आदान-प्रदान का एक चित्र प्रदान करती है, कोरलेशन का मान वैसा ही होता है यदि आप Y (लंबवत) और X (क्षैतिज) मापों को बदल दें। सहसंबंध गुणांक $r = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$ का अनुमान सूत्र द्वारा लगाया जाता है। नमूने के आकार n के लिए, ऊपर का अभिव्यक्ति $r = \sum_{i=1}^{n} \frac{(x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y}$ लिखा जा सकता है, जहाँ s_x और s_y दोनों परिमाणों के नमूने के मानक विचलन होते हैं।

4. रैंक सहसंबंध (Rank Correlation):

कुछ मामलों में, डेटा को मापने में संभावन नहीं होता है और केवल रैंकिंग की जाती है। ऐसे परिस्थितियों में, रैंक संबंध निर्धारित किया जाता है जो केवल पियरसन के उत्पाद-क्षण कोरलेशन संख्या होता है और यह व्यक्तियों के रैंक के बीच संबंध को व्यक्त करता है। इसे स्पियरमैन के रैंक संबंध संख्या भी कहा जाता है और यह -1 से +1 के बीच होता है। अगर d_i आईथ व्यक्ति के रैंकों के बीच का अंतर दर्शाता है और n व्यक्तियों की संख्या को दर्शाता है, तो स्पियरमैन के रैंक संबंध संख्या निम्नलिखित होता है:

$$r = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

अगर रैंकों में बंधन होता है, तो दिए गए रैंकों को उन व्यक्तियों के रैंकों का औसत रैंक होता है जिनके पास बंधन नहीं होता था। बंधनों के मामले में, रैंक संबंध संख्या निम्नलिखित रूप में होता है:

$r = 1 - \frac{6\left(\sum_{i=1}^{n} d_i^2 + \sum (m^3 - m)/12\right)}{n(n^2-1)}$; जहाँ m वे व्यक्तियाँ हैं जिनके पास एक ही रैंक है।

5. आंशिक सहसंबंध (Partial Correlation):

कभी-कभी दो चरणीय चरों Y और X_1 के बीच संबंध तीसरे चर X_2 के साथ के संबंध के कारण भी हो सकता है। Y और X_1 के बीच सच्चा संबंध केवल तब देखा जा सकता है जब X_2 के प्रभाव को हटा दिया गया है। हम नमूना आंशिक कोरलेशन संख्या के माध्यम से इसे पूरा करते हैं। इस प्रकार, आंशिक कोरलेशन दो चरणिकाओं के बीच रैखिक संबंध को एक या एक से अधिक चरों के प्रभाव को हटाने के बाद मापता है। आंशिक कोरलेशन उन चरों को प्रकट कर सकता है जो दो विशिष्ट चरों के बीच संबंध को बढ़ावा देते हैं या दबा देते हैं। उदाहरण के लिए, यदि प्रतिवर्ष, एक समुद्र तटीय केबिन में रसोई में एक वर्ष में प्रतिवार की गई चीनी की संख्या और घर के पास पास कारों की संख्या की जाती है, तो संबंध एक के पास हो सकता है। क्या कार चींटियों को लाते हैं? क्या यह अव्वल लगता है? एक तीसरा चर, तापमान, उपेक्षित है। जब मौसम गर्म होता है, तो चींटियाँ फलती हैं और समुंदर किनारे की तरफ लोग बहुत आते हैं; जब यह शीत होता है, तो न कारों की संख्या होती है और न ही चींटियों की। यदि तापमान का रैखिक प्रभाव नियंत्रित किया जाता है, तो चींटियों और कारों के बीच संबंध गायब हो जाता है। शेष बचा हुआ विभिन्न चर Y और X_1 को चर X_2 के प्रभाव को हटाए बिना दर्शाता है। चर X_2 के प्रभाव को हटाने के बाद चर Y और X_1 के बीच आंशिक कोरलेशन संख्या को प्रतीक r_{Y1} से दर्शाया जाता है। यदि हम चर Y और X_1 और चर Y और X_1 और X_2 के लिए सामान्य कोरलेशन संख्याओं को r_{Y1}, r_{Y2} और r_{12} के रूप में लिखें, तो X_2 को निश्चित रखकर Y और X_1 के लिए नमूना आंशिक कोरलेशन संख्या निम्नलिखित परिभाषा द्वारा दिया जाता है।

6. आंशिक सहसंबंध गुणांक (Partial Correlation Coefficient):

चर Y और X_1 के बीच रैखिक संबंध का माप, चर X_2 के साथ उनके संबंध को बनाने के बाद, नमूना आंशिक कोरलेशन संख्या $r_{y1.2}$ द्वारा आकलन किया जाता है।

जहाँ: $r_{Y1.2} = \frac{r_{Y1} - r_{Y2} r_{12}}{\sqrt{\left(1 - r_{Y2}^2\right)\left(1 - r_{12}^2\right)}}$

7. प्रतिगमन (Regression):

कई सांख्यिकीय अध्ययनों में, लक्ष्य एक चर के मूल्य को किसी दूसरे चर(ओं) के मूल्य के आधार पर पूर्वानुमानित करने के लिए एक माध्यम द्वारा व्यक्त किए गए एक संबंध स्थापित करना होता है। इस प्रकार की स्थितियों में, हमें रिग्रेशन विश्लेषण हमारी मदद कर सकता है। शब्द "रिग्रेशन" का उगम मूल रूप से सर फ्रांसिस गॉल्टन (1822-1911) द्वारा की गई मूल वंशानुक्रमण अध्ययनों से है, जिनमें उन्होंने पिता की ऊंचाई को बेटों की ऊंचाई से तुलना की। गॉल्टन ने दिखाया कि लम्बे पिता के बेटों

की ऊंचाई पीढ़ी-पीढ़ी से जनसंख्या की औसत ऊंचाई की ओर प्रतिसारण करती है। दूसरे शब्दों में, आमतौर पर लम्बे पिता के बेटे अपने पिता से छोटे होते हैं और आमतौर पर छोटे पिता के बेटे अपने पिता से ऊँचे होते हैं। आजकल, शब्द "रिग्रेशन" प्रमाणन समस्याओं के बहुतायत हो चुका है और इसका अर्थ आवश्यकता से निश्चित रूप से जनसंख्या की औसत की ओर प्रतिसारण नहीं करता है। इस खंड में, हम एक निर्भर चर के मूल्य की अनुमानित करने या पूर्वानुमानित करने की समस्या से निपटते हैं जिसमें एक सेट के स्वतंत्र चरों के मूल्य को दिया गया है। हम एकल स्वतंत्र चर के मामले से शुरू करते हैं।

8. सरल रेखीय प्रतिगमन (Simple Linear Regression):

जो प्रतिक्रिया चर (y) में परिवर्तन होता है, उसे स्वतंत्र चर (x) जिसे रिग्रेसर कहा जाता है द्वारा स्पष्ट किया जाता है। y की एकल रिग्रेशन या सामाजिक मॉडल के रूप में सीधी रेखा की समीकरण, एक यादृच्छिक त्रुटि (ε) के लिए एक शब्द जोड़ती है, क्योंकि बिना यह कि बिना यह त्रुटियों के प्वाइंट रेखा पर नहीं पड़ते हैं:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

स्लोप (β_1) एक रेखा पर ऊर्ध्वाधिकारी परिवर्तन और रेखा के साथीक परिवर्तन के बीच अनुपात होता है। एक परीक्षण $\beta_1=0$ वही है जैसा कि एक परीक्षण हो कि संबंध संख्या (r) शून्य है जैसा कि $r = \hat{\beta}_1 s_x / s_y$

शीर्षक ($\beta_0=0$) या निरंतरता, जैसा कि यह आमतौर पर कहा जाता है, वह स्थान है जहाँ रेखा ऊर्ध्वाधिकारी अक्ष को x = 0 पर काटती है।

मॉडल में त्रुटियों (ε) को प्रतिनिधित्व करने के लिए, प्रत्येक बिंदु से रेखा तक एक छोटी ऊर्ध्वाधिकारी रेखा खींचें। इन रेखा की लंबाई रेखा और प्लॉट बिंदुओं के बीच की लंबाई को "शेष" कहा जाता है और ये सच्ची त्रुटियों के लिए आकलन होते हैं।

उपरोक्त समीकरण में, Y आपकी निर्भर या परिणाम या पूर्वानुमानित चर है, जिसे आप पूर्वानुमान करने की कोशिश कर रहे हैं; x स्वतंत्र या पूर्वानुमानक चर है; और शीर्षक (β_0) और ढाल (β_1) प्रतिष्ठान होते हैं। यदि मॉडल चरों के बीच संबंध का अच्छा वर्णनकारक है, तो कूटनिष्ठ चरों के आकलन का उपयोग नए मामलों के लिए निर्भर चर के मूल्य की पूर्वानुमानित करने के लिए किया जा सकता है।

9. एकाधिक रेखीय प्रतिगमन (Multiple Linear Regression):

एक से अधिक स्वतंत्र चरों, उदाहरण के लिए, $X_1, X_2, \ldots, X_p$ जिन्हें Y में परिवर्तन के कारण कहा जा सकता है, के साथ, हम इस परिवर्तन को ध्यान में रखने के लिए y के x's पर मल्टिपल रिग्रेशन को फिट करते हैं। y के x's पर मल्टिपल रिग्रेशन को निम्नलिखित रूप में दिखाया जाता है:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \varepsilon$$

जहाँ, β_0 शीर्षक को प्रतिस्थान करता है और β_i को आंशिक रिग्रेशन संख्याएँ कहा जाता है। ε यादृच्छिक त्रुटि है। β_i की मदद से x_i में एकक परिवर्तन के प्रति y में औसत परिवर्तन को दिखाते हैं, अन्य x's को स्थिर रखकर।

R सॉफ्टवेयर का उपयोग करके सहसंबंध और प्रतिगमन पर व्यावहारिक अभ्यास

Practical Exercises on Correlation and Regression using R Software

प्रस्तावना (Introduction):

इस अध्याय में, हमने संबंध और रिग्रेशन विश्लेषण अध्ययन में प्रयुक्त तकनीकों को R सॉफ़्टवेयर का प्रयोग करके प्राक्टिकल प्रदर्शन के माध्यम से कवर किया है। इस संदर्भ में, हमने व्यावहारिक समस्याओं, डेटा और R कोडों की चर्चा की है साथ ही परिणाम और आउटपुट्स को उचित व्याख्यान और स्पष्टीकरण के साथ।

लाइब्रेरियाँ लोड करें (Load the Libraries): हम आपको R का उपयोग करके संबंध और रिग्रेशन विश्लेषण से संबंधित कोई भी अध्ययन करने के लिए निम्नलिखित लाइब्रेरियाँ लोड करने की सिफारिश करते हैं। हालांकि इसके लिए कई अन्य विकल्प भी उपलब्ध हैं।

```
library(ppcor)
library(car)
library(agricolae)
```

डेटा (Data): निम्नलिखित डेटा पायलट सैम्पल सर्वेक्षण के माध्यम से हाइब्रिड ज्वार फसल की पैदावार और जैविक विशेषताओं पर संकलित किया गया था। जैविक विशेषताएँ औसत पौधों की जनसंख्या (PP), औसत पौधों की ऊंचाई (PH), औसत हरे पत्तियों की संख्या (NGL) और पैदावार (किलो / प्लॉट) थी। (Source: https://drs.icar.gov.in/Analysis%20of%20data/correlation&regression.html)

S.No.	PP	PH	NGL	Yield		S.No.	PP	PH	NGL	Yield
1	142.00	0.525	8.2	2.470		**24**	55.55	0.265	5.0	0.430
2	143.00	0.640	9.5	4.760		**25**	88.44	0.980	5.0	4.080
3	107.00	0.660	9.3	3.310		**26**	99.55	0.645	9.6	2.830
4	78.00	0.660	7.5	1.970		**27**	63.99	0.635	5.6	2.570
5	100.00	0.460	5.9	1.340		**28**	101.77	0.290	8.2	7.420
6	86.50	0.345	6.4	1.140		**29**	138.66	0.720	9.9	2.620
7	103.50	0.860	6.4	1.500		**30**	90.22	0.630	8.4	2.000
8	155.99	0.330	7.5	2.030		**31**	76.92	1.250	7.3	1.990
9	80.88	0.285	8.4	2.540		**32**	126.22	0.580	6.9	1.360
10	109.77	0.590	10.6	4.900		**33**	80.36	0.605	6.8	0.680
11	61.77	0.265	8.3	2.910		**34**	150.23	1.190	8.8	5.360
12	79.11	0.660	11.6	2.760		**35**	56.50	0.355	9.7	2.120
13	155.99	0.420	8.1	0.590		**36**	136.00	0.590	10.2	4.160
14	61.81	0.340	9.4	0.840		**37**	144.50	0.610	9.8	3.120
15	74.50	0.630	8.4	3.870		**38**	157.33	0.605	8.8	2.070
16	97.00	0.705	7.2	4.470		**39**	91.99	0.380	7.7	1.170
17	93.14	0.680	6.4	3.310		**40**	121.50	0.550	7.7	3.620
18	37.43	0.665	8.4	1.570		**41**	64.50	0.320	5.7	0.670
19	36.44	0.275	7.4	0.530		**42**	116.00	0.455	6.8	3.050
20	51.00	0.280	7.4	1.150		**43**	77.50	0.720	11.8	1.700
21	104.00	0.280	9.8	1.080		**44**	70.43	0.625	10.0	1.550
22	49.00	0.490	4.8	1.830		**45**	133.77	0.535	9.3	3.280
23	54.66	0.385	5.5	0.760		**46**	89.99	0.490	9.8	2.690

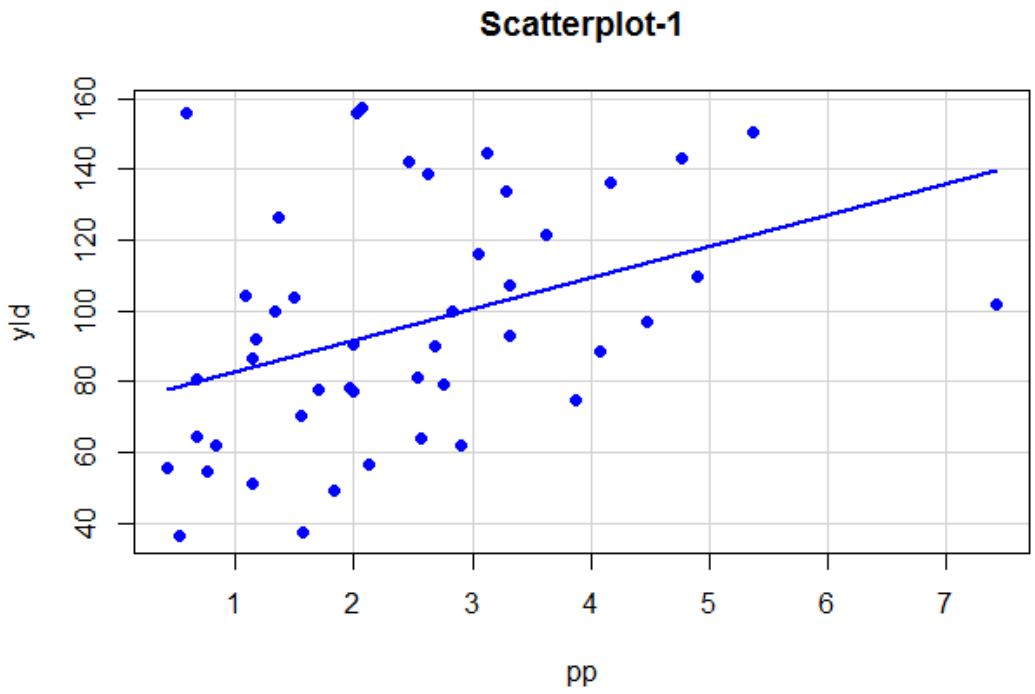
हमारे प्रदर्शन को सुविधाजनक बनाने के लिए, हम एक डेटासेट बना सकते हैं जिसमें केवल ये चर शामिल होते हैं:

```
d4correg <- read.csv(file.choose(), header=T)
d4correg
attach(d4correg)
```

प्रकीर्ण आरेख (Scatter Plot): कोरलेशन को गुणसूत्री रूप से पता करने का सबसे सरल तरीका डेटा को चित्रित करना है। निम्नलिखित कोड्स का उपयोग चर पैदावार को अन्य चरों के साथ चित्रित करने के लिए किया जा सकता है:

```
scatterplot(yld, pp, main="Scatterplot-1", xlab="pp", ylab="yld", pch=19, smooth=FALSE, boxplots=FALSE)
```

परिणामस्वरूप बिंदुग्राफ निम्नलिखित है:



जैसा कि ऊपर दिए गए चित्र में दिखाया गया है, पैदावार और पौधों की जनसंख्या के बीच मजबूत सकारात्मक संबंध प्रकट होता है, यानी चित्र में दिखाया गया है कि जैसे-जैसे पौधों की जनसंख्या बढ़ती है, पैदावार भी बढ़ता है। अन्य बिंदुग्राफ भी निम्नलिखित रूप में प्राप्त किए जा सकते हैं:

```
scatterplot(ph, yld, main="Scatterplot-2", xlab="ph", ylab="yld ", pch=19, smooth=FALSE, boxplots=FALSE)
scatterplot(ngl, yld, main="Scatterplot-3", xlab="ngl", ylab="yld ", pch=19, smooth=FALSE, boxplots=FALSE)
```

R का उपयोग करके संबंध विश्लेषण: संबंध विश्लेषण के अंतर्गत हम संबंधित आसान और आंशिक संबंध संख्याओं की गणना कैसे करें उसके साथ ही परीक्षण करेंगे।

सामान्य संबंध संख्या: सामान्य संबंध संख्या की गणना और परीक्षण निम्नलिखित कोड का उपयोग करके किया जा सकता है:

```
with(d4correg,correlation(yld,ngl,method="pearson", alternative = "two.sided"))
```

परिणाम है (Results):

```
> with(d4correg,correlation(yld,ngl,method="pearson", alternative = "two.sided"))
Pearson's product-moment correlation
data: yld and ngl
t = 1.925588 , df = 44 , p-value = 0.06063374
alternative hypothesis: true rho is not equal to 0
sample estimates:
cor
0.2787842
```

"yld" और "ngl" चरों के बीच सामान्य संबंध संख्या 0.28 है। इसके अलावा, आवेदनिक अध्ययन की आवश्यकता के अनुसार सामान्य संबंध संख्या (Pearson, Kendal और Spearman) की गणना और परीक्षण (कम, अधिक और दो-पक्षीय) की जा सकती है इस कोड का उपयोग करके।

आंशिक सहसंबंध गुणांक (Partial Correlation Coefficient): आंशिक सहसंबंध गुणांक की गणना और परीक्षण निम्नलिखित कोड का उपयोग करके किया जा सकता है:

```
partial <- pcor(d4correg)
partial
```

यह महत्वपूर्ण है कि पूरी जानकारी हमें आवश्यक नहीं हो सकती है। हम अपनी आवश्यकता के अनुसार आउटपुट को अनुकूलित कर सकते हैं। मान लीजिए कि हम चरों के आपसी आसानी से आपसी संबंध की गणना करना चाहते हैं और सीरियल नंबर नहीं, तो हम निम्नलिखित का उपयोग कर सकते हैं:

```
partial <- pcor(d4correg[,c(2:4,5)])
partial
The output is as follows:
> partial <- pcor(d4correg[,c(2:4,5)])
```

```
> partial
$estimate
          pp          ph         ngl       yld
pp  1.0000000  0.13153060  0.20278219 0.2825348
ph  0.1315306  1.00000000 -0.03094463 0.2686277
ngl 0.2027822 -0.03094463  1.00000000 0.1919946
yld 0.2825348  0.26862765  0.19199462 1.0000000

$p.value
           pp         ph       ngl       yld
pp  0.00000000 0.39473355 0.1867864 0.06313615
ph  0.39473355 0.00000000 0.8419484 0.07787612
ngl 0.18678642 0.84194836 0.0000000 0.21183369
yld 0.06313615 0.07787612 0.2118337 0.00000000

$statistic
          pp         ph        ngl      yld
pp  0.0000000  0.8598863  1.3420617 1.908805
ph  0.8598863  0.0000000 -0.2006402 1.807336
ngl 1.3420617 -0.2006402  0.0000000 1.267855
yld 1.9088054  1.8073364  1.2678546 0.000000

$n
[1] 46
$gp
[1] 2
$method
[1] "pearson"
```

और विशेष रूप से, हम पीसीओआर फ़ंक्शन से पूछ सकते हैं कि वे चरों के सभी संभावित जोड़ी वार्य परिसंघों के बीच आंशिक संबंधों को उत्पन्न करें, और फिर हमें उन आंशिक संबंधों को प्रदान

करें जो हमें चाहिए। यह एक और स्थिति का विचार करें जहाँ हम केवल यह आकलन करना चाहते हैं कि पैदावार का केवल आपसी संबंध अन्य चरों के साथ होता है, तो हम निम्नलिखित का उपयोग कर सकते हैं:

```
partial <- pcor(d4correg[,c(2:4,5)])$estimate[1:3,4]
partial
The output is given as
> partial <- pcor(d4correg[,c(2:4,5)])$estimate[1:3,4]
> partial
      pp       ph       ngl
0.2825348 0.2686277 0.1919946
```

प्रतिगमन विश्लेषण (Regression Analysis): रिग्रेशन विश्लेषण के अंतर्गत हम मॉडल फिटिंग से संबंधित विभिन्न समस्याओं पर चर्चा करेंगे।

मॉडल फिटिंग (Model Fitting): मान लीजिए हमें एक बहुवार्ती रिग्रेशन मॉडल को पैदावार के रूप में निर्भर चर के साथ फिट करना है, तो निम्नलिखित कोड का उपयोग इस उद्देश्य के लिए किया जा सकता है:

```
model1.yield <- lm(yld ~ pp + ph + ngl + sn, data=d4correg)
summary(model1.yield)
coef(model1.yield)
confint(model1.yield)

The output is given as:
> model1.yield <- lm(yld ~ pp + ph + ngl + sn, data=d4correg)
> summary(model1.yield)

Call:
lm(formula = yld ~ pp + ph + ngl + sn, data = d4correg)

Residuals:
    Min      1Q  Median      3Q     Max
-2.4147 -0.7710 -0.3766  0.8382  5.3645
```

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.766269   1.081573  -0.708   0.4827
pp           0.011911   0.006349   1.876   0.0678 .
ph           1.703157   0.933128   1.825   0.0753 .
ngl          0.158329   0.121658   1.301   0.2004
sn          -0.006520   0.015167  -0.430   0.6696
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.345 on 41 degrees of freedom
Multiple R-squared:  0.2425,         Adjusted R-squared:  0.1686
F-statistic: 3.282 on 4 and 41 DF,  p-value: 0.0201

> coef(model1.yield)
 (Intercept)          pp          ph         ngl          sn
-0.766268930  0.011910681  1.703157080  0.158329170 -0.006519609
> confint(model1.yield)
                  2.5 %     97.5 %
(Intercept) -2.9505509097 1.41801305
pp          -0.0009116661 0.02473303
ph          -0.1813341216 3.58764828
ngl         -0.0873644651 0.40402280
sn          -0.0371504564 0.02411124
```

अध्याय 14

लघु क्षेत्र आकलन विधि-एक अवलोकन

कौस्तव आदित्य
वरिष्ठ वैज्ञानिक, प्रतिदर्श सर्वेक्षण प्रभाग
भा.कृ.अनु.प.-भारतीय कृषि साँख्यिकी अनुसंधान संस्थान, लाइब्रेरी एवेन्यू, नई दिल्ली-110012

1. प्रस्तावना (Introduction):

सार्वजनिक व निजी क्षेत्रों में लघु क्षेत्र के विश्वसनीय आँकड़ों की बढ़ती हुई मांग के कारण प्रतिदर्श सर्वेक्षण में लघु क्षेत्र आकलन एक प्रमुख भूमिका निभाता है। प्रतिदर्श सर्वेक्षण, चाहे वे सरकारी संगठनों या गैर सरकारी संस्थाओं द्वारा संचालित हो, का उद्देश्य, न केवल संपूर्ण समष्टि बल्कि सभी उप–समष्टि या क्षेत्रों की किस्मों के लिए यथोचित सही प्रत्यक्ष आकलों को उत्पन्न करना है। ये प्रत्यक्ष आकलक क्षेत्र विशिष्ट प्रतिदर्श आँकड़ों पर आधारित है। हालांकि अनुसंधानकर्ताओं तथा नीति निर्माताओं सभी लघु क्षेत्र के आँकडों को प्राप्त करना चाहते हैं। पर्याप्त परिशुद्धता (precision) के प्रत्यक्ष आकलक के सर्मथन के लिए यदि क्षेत्र विशिष्ट प्रतिदर्श बड़ा नहीं है तो क्षेत्र (domain) को लघु के रुप में माना जाता है । इन छोटे क्षेत्रों को लघु क्षेत्र भी कहते हैं क्योंकि सर्वेक्षण में इन क्षेत्रों का प्रतिदर्श आकार छोटा होता है। इसलिए हमें इन लघु क्षेत्रों के आकलन करने के लिए एक विशेष विधि की आवश्यकता होती है। इस विधि को लघु क्षेत्र आकलन के नाम से जानते है। इस लेख में साहित्य में उपलब्ध लघु क्षेत्र आकलन की कुछ मौजूदा तकनीकों का उल्लेख किया गया है।

2. लघु क्षेत्र आकलन विधि (Small Area Estimation Technique):

लघु क्षेत्र समष्टि का एक उप–समूह है जिसके लिए प्रतिदर्श सर्वेक्षण से बहुत कम सूचना उपलब्ध होती है। इन उप–समूहों को भौगोलिक क्षेत्र (उदाहरणार्थ एक नगर पालिका, एक समष्टि प्रभाग, खण्ड़, तहसील, ग्राम पंचायत आदि) या जन–सांख्यिकीय समूह (जैसे एक विशाल भौगोलिक क्षेत्र में लोगों की आयु, लिंग, जाति का वर्ग विशिष्ट समूह) या दोनों का परस्पर वर्गीकरण से उल्लिखित करते है। इन लघु क्षेत्रों से सम्बन्धित आँकड़ों को प्रायः लघु क्षेत्र सांख्यिकी कहा जाता है। बढ़ती हुई मांग के मद्दे नजर, सर्वेक्षण संगठन मौजूदा प्रतिदर्श सर्वेक्षण के आँकड़ों से लघु क्षेत्र आकलन उत्पन्न करने में कठिनाईयों का सामना करते है। दुर्भाग्य से, इन लघु क्षेत्रों के लिए लघु विशिष्ट विश्वसनीय प्रत्यक्ष अनुमान प्रदान करने के लिए, लघु क्षेत्र में प्रतिदर्श आकार बहुत छोटा होता है और कभी–2 प्रतिदर्श उपलब्ध भी नहीं होता है। लघु क्षेत्र आकलन

में सही प्रत्यक्ष अनुमान के लिए प्रतिदर्श आकार में प्रर्याप्त वृद्धि करने की आवश्यकता है जो पहले से बजट की कमी में वृद्धि करेगा तथा आँकड़ों के विश्लेषण में अधिक समय लगेगा।

लघु क्षेत्र आकलन की दोहरी समस्या है, पहला मूल प्रश्न कि इन क्षेत्रों से लिए गये बहुत छोटे प्रतिदर्श के आधार पर लघु क्षेत्र के लिए इच्छित प्राचलों का विश्वसनीय अनुमान कैसे निकाला जाये। दूसरा सम्बन्धित प्रश्न है कि आकलन त्रुटि की गणना कैसे करे। दिये गये क्षेत्र में केवल एक छोटा प्रतिदर्श होने पर (और संभवतः प्रतिदर्श न होने पर) अन्य सम्बन्धित आँकड़ों के समूहों से सूचना लेकर आकलन की समस्या का केवल एक संभव समाधान है। लघु क्षेत्रों या क्षेत्रों के पर्याप्त परिशुद्धता के साथ अनुमान उत्पन्न करने के लिए, एक आकलन प्रक्रिया के माध्यम से सम्बन्धित क्षेत्रों या समय अवधि (अथवा दोनो) से मान लेकर (borrowing the strength) लघु क्षेत्र आकलन के तरीको पर नजर डाले सकते है और इस प्रकार कुल (प्रभावी) प्रतिदर्श आकार एवं परिशुद्धता में वृद्धि होती है। ये आकलन प्रक्रिया या तो अन्तर्निहित (implicit models) या स्पष्ट मॉडलों (explicit models) पर आधारित है जो अनुपूरक डाटा (सहायक सूचना) के उपयोग के माध्यम से सम्बन्धित क्षेत्रों या समय अवधि (अथवा दोनों) से जोड़ता है, जैसे हाल की जनगणना और वर्तमान प्रशासनिक अभिलेखों {राव (2003)}।

लघु क्षेत्र आकलन के लिए उपयोग में लायी गयी विधियों को सम्बन्धित आँकड़ो स्रोतों के उपयोग जोकि आपस में, भूतकाल आँकड़ों या दोनों के अनुसार विभाजित किया जा सकता है। निष्कर्ष के प्रकार के अनुसार विधियों को एक और प्रकार से विभाजित किया हैः अभिकल्पना–आधारित (design-based), मॉडल–निर्भर (model-dependent) (फ्रीक्यून्टिस और बायसन में उप–विभाजित) (with sub-division into the frequentist and Bayesian approaches) या दोनों का मिश्रण (combination of the two)। उपलब्ध सहायक सूचना की उपलब्धता के स्तर के आधार पर विधियों को क्षेत्र स्तर (area level) और एकक स्तर लघु क्षेत्र (unit level small area) मॉडल में भी विभाजित किया जा सकता है {राव (2003)}।

2.1 प्रत्यक्ष आकलक (Direct Estimator):

मान लीजिए प्रतिदर्श महत्व/मान (weights) $\{w_j; j \in s\}$ पर आधारित रैखिक आकलक (estimator) का उपयोग जनसंख्या स्तर की मात्राओं के निष्कर्ष निकालने में होता है। यहाँ s प्रतिदर्श के आकार n को दर्शाता है जिसका चयन N आकार की समष्टि $U = (1,\ldots\ldots, N)$ से प्रतिचयन अभिकल्पना $p(s)$ द्वारा किया गया है। आगे, यदि $\pi_j = \sum_{j \in s} p(s)$ प्रथम क्रम समावेश प्रायिकता (first order inclusion probabilities) है तब $W_j = \pi_J^{-1}$ तत्व j के अभिकल्पना महत्व/मान को दर्शाता है। सामान्य यादृच्छिक प्रतिचयन (simple random sampling) के अर्न्तगत, $\pi_j = nN^{-1}$ और $w_j = Nn^{-1}$ है i लघु क्षेत्र

$(i=1,\ldots m)$ को दर्शाता है। माना कि समष्टि m गैर–अतिव्यापी (non-overlapping) क्षेत्र या लघु क्षेत्र U_i, प्रत्येक N_i आकार की समष्टि के साथ जैसे $U=\bigcup_{i=1}^{m}U_i$ और $N=\sum_{i}^{m}N_i$, । माना S_i, n_i आकार के प्रतिदर्श का हिस्सा है जोकि लघु क्षेत्र i और $n=\sum_{i=1}^{m}n_i$ में आता है। हम इच्छित प्राचल Y के लिए लघु क्षेत्र i में j^{th} समष्टि ईकाई के मान को y_{ij} से दर्शाते है। लघु क्षेत्र i में Y का समष्टि माध्य, तब समान मान का उपयोग $\bar{y}_i=N_i^{-1}\sum_{j\in i}y_i$ करके आकलक का अनुमान हो सकता है

$$\hat{\bar{Y}}_i^{H\ddot{a}jek}=\left(\sum_{j\in s_i}w_j\right)^{-1}\left(\sum_{j\in s_i}w_jy_j\right) \qquad (2.1)$$

यदि लघु क्षेत्र i का समष्टि आकार N_i ज्ञात है तो

$$\hat{\bar{Y}}_i^{HT}=N_i^{-1}\left(\sum_{j\in s_i}w_jy_j\right) \qquad (2.2)$$

आकलक (2.1) और (2.2) को कभी–2 लघु क्षेत्र i माध्य $\bar{y}_i$ का प्रत्यक्ष आकलक के रूप में संर्दभित करते है। अधिक शुद्ध रूप में, आकलक (2.1) को प्रत्यक्ष आकलक का हेजक टाईप (*Häjek Type*) और आकलक (2.2) को प्रत्यक्ष आकलक का होरविटज–थोम्पसन टाईप [(Horvitz-Thompson (HT) Type] के नामों से संर्दभित करते है इन नामों को क्लासिकल प्रतिचयन में सीमित/ज्ञात समष्टि माध्य का आकलन करने के लिए वैकल्पिक विधि से संर्दभित करते हैं, कोकरान (1977) और सामदाल, स्वानसन और रेतमेन (1992)। प्रत्यक्ष आकलक का चाहे कोई भी रुप उपयोग में लाया जाये है। यह देखना आसान है कि इसका प्रसरण (variance) अधिक हो सकता है जब क्षेत्र प्रतिदर्श आकार n_i छोटा है उदाहरणातः सामान्य यादृच्छिक प्रतिचयन में, सहायक सूचना न होने पर, i लघु क्षेत्र $(\bar{Y}_i=N_i^{-1}\sum_{j\in U_i}y_j)$ के लिए Y के माध्य का प्रत्यक्ष अनुमानक $\hat{\bar{Y}}_i=\bar{y}_i$ है।

माना कि सर्वेक्षण चर Y के अलावा, p-सहायक चरों का मान भी ज्ञात है। हम क्षेत्र i में j इकाई के लिए सहायक चर X एक p x1 सदिश (vector) को x_{ij} के द्वारा व्यक्त करते है। तब ज्ञात सहायक सूचना के साथ, $\bar{Y}_i$ के लिए एक अति दक्ष अभिकल्पना–आधारित प्रत्यक्ष आकलक समाश्रयण आकलक को निम्न प्रकार परिभाषित कर सकते है।

$$\hat{\bar{y}}_i^{reg} = \bar{y}_i + \left(\bar{X}_i - \bar{x}_i\right)' \beta \qquad (2.3)$$

जहां β_1 क्षेत्र i में समाश्रयण गुणांक का सदिश है, $\bar{x}_i = n_i^{-1}\sum_{j\in s_i} x_j$ और $\bar{X}_i = N_i^{-1}\sum_{j=1}^{N_i} x_j$ क्षेत्र i में सहायक चर X के क्रमशः प्रतिदर्श माध्य और समष्टि माध्य है। (2.1) का विचरण निम्न प्रकार है।

$$Var_p(\hat{\bar{Y}}_i^{reg}) \approx n_i^{-1}(1-f_i)S_i^2(1-\rho_i^2) = (1-\rho_i^2)Var_p(\hat{\bar{Y}}_i) \qquad (2.4)$$

जहाँ ρ_i क्षेत्र i में सर्वेक्षण चर Y और सहायक चर X के मध्य बहु सहसम्बन्ध है। विचरण (2.3) का आकल $v(\hat{\bar{Y}}_i^{reg} \mid n_i) = (1-\hat{\rho}_i^2)(1-f_i)\,s_i^2/n_i$ है। सहायक चरों के उपयोग द्वारा, विचरण को कारक $\left(1-\rho_i^2\right)$ तक कम किया जा सकता है। इससे यह पता चलता है कि अच्छी सहायक सूचना के उपयोग से, सर्वेक्षण चर Y के साथ उच्च सहसम्बन्ध होने पर, लघु क्षेत्र आकलन में शुद्धता बढ़ जाती है।हालाँकि, समाश्रयण आकलक (2.4) के साथ समस्या यह है कि व्यवहारिक रुप में समाश्रयण गुणांक β_1 कभी ज्ञात होता है। β_1 के स्थान पर इसके साधारण न्यूनतम वर्ग (Ordinary Least Square) का इस्तेमाल करने पर, प्रत्येक क्षेत्र i में प्रतिदर्श आकार कम होने के कारण, अनुमान $\hat{\beta}_1$ प्रभावी नहीं होता।

2.2 अप्रत्यक्ष आकलन (Indirect Estimators):

समुचित रुप से परिशुद्ध अनुमान के लिए प्रत्येक लघु क्षेत्र में जब प्रतिदर्श आकार पर्याप्त बड़ा होता है तो प्रत्यक्ष अनुमानक सबसे वांछनीय है। आँकड़ों के स्रोतों के रूप में प्रायः प्रतिदर्श सर्वेक्षण बड़े या उच्च स्तर की सांख्यिकी उत्पन्न करने के लिए तैयार किये जाते है, लघु क्षेत्रों के लिए प्रतिदर्श आकार प्रायः कम होता है। परिणामस्वरूप, इन अनुमानकों का संबद्धित विचरण अस्वीकार्य अधिक होने की संभावना है। अतः लघु क्षेत्रों के अनुमान के लिए यह जरूरी है कि संबन्धित क्षेत्रों से गुण लेकर (Borrow strength) आकलन विधि प्रयोग में लायी जायें। इन अनुमानकों को प्रायः अप्रत्यक्ष अनुमानकों के रूप में उल्लिखित करते है चूंकि हम लघु क्षेत्रों या समय (काल) और सम्भवतः दोनों से सर्वेक्षण चरों (सहायक चर) के मानों का प्रयोग करते है। अन्तर्निहित(implicit) या स्पष्ट मॉडलों (explicit models) पर आधारित सांख्यिकी मॉडल जोकि संबद्धित लघु क्षेत्रों को सहायक जानकारी द्वारा जोड़ते है के उपयोग द्वारा लघु क्षेत्रों या काल (समय) (अथवा दोनों) से सूचना (आँकड़ें) लेते हैं। अंतर्निहित संबन्ध मॉडलों (implicit linking models) पर आधारित पारंपरिक अप्रत्यक्ष अनुमान विधि संश्लेषित/कृत्रिम और मिश्रित/संयुक्त अनुमान है जिनका वर्णन नीचे दिया गया है।

2.2.1 संश्लेषित / कृत्रिम अनुमानक (Synthetic Estimators):

लघु क्षेत्रों के लिए संश्लेषित / कृत्रिम अनुमान उत्पन्न करने में, समष्टि के बृहद् क्षेत्र के समुच्चय के लिए प्रत्यक्ष अनुमान की उपलब्धता कल्पित है। उचित मान या अनुपात को वांछित लघु क्षेत्र अनुमान प्राप्त करने के लिए बृहद् समष्टि क्षेत्र अनुमान में प्रयुक्त करते है। संश्लेषित / कृत्रिम अनुमान विधि का प्रयोग पहली बार राष्ट्रीय स्वास्थ्य सांख्यिकी केन्द्र संयुक्त राज्य द्वारा राष्ट्रीय स्वास्थ्य साक्षातकार सर्वेक्षण (1968) के आधार पर दीर्घ और अल्प कालिक शारीरिक विकलांगता के अनुमान के लिए किया था।

संश्लेषित / कृत्रिम अनुमानक में, हम बृहद् वर्ग के योग $g(g=1,.....G)$ के लिए विश्वसनीयता प्रत्यक्ष आकलन की उपलब्धता $\hat{T}_{y.g}=\sum_{i=1}^{m}\hat{T}_{y_{ig}}$ मान लेते है कि दिये गये सर्वेक्षण के लिए लघु क्षेत्र $i(i=1,.....m)$ को शामिल करता है, जहाँ $\hat{T}_{yig}$ $\mathrm{N_{ig}}$ आकार की समष्टि के साथ $(\mathrm{i,g})^{\mathrm{th}}$ कक्ष में Y का समष्टि योग का अनुमान ($T_{y_{ig}}=\sum_{j=1}^{N_{ig}}y_{jig}$) है। यहाँ y_{jig} कक्ष (i,g) में Y इच्छित चर के लिए इकाई j (j=1, ...$\mathrm{N_{ig}}$) का मान है। समष्टि $\hat{T}_{y.g}$, के लिए उपलब्ध अनुमान से समूह g के लिए समष्टि माध्य का अनुमान $\hat{\bar{Y}}_{.g}=\left(\sum_{i=1}^{m}\hat{T}_{y_{ig}}\right)/\left(\sum_{i=1}^{m}N_{ig}\right)=\hat{T}_{y.g}/N_{.g}$ के रुप में है। जनगणना या किसी अन्य स्रोत से उपलब्ध एक उपयुक्त सहायक सूचना $\mathrm{W_{ig}}$ एक श्रृखला या अनुपात की गणना में प्रयोग होते है जैसे $\Sigma_g\ w_{ig}=1$ । मान Wig को $i-th$ लघु क्षेत्र माध्य $\bar{Y}_i$ जैसे $\hat{\bar{Y}}_i^{syn}=\sum_{g=1}^{G}w_{ig}\hat{\bar{Y}}_{.g}$ के लिए संश्लेषित / कृत्रिम अनुमानक निकलाने के लिए वर्ग माध्य में प्रयुक्त करते है इस अनुमानक को अभिकल्पना–आधारित संश्लेषित / कृत्रिम अनुमानक के रुप में उल्लेखित करते है (गोन्जलेज और होजा (1978))।

अब मॉडल–आधारित संश्लेषित / कृत्रिम अनुमान का उल्लेख करते है। समाश्रयण मॉडल पर विचार करते है।

$$y_{ij}=x_{ij}'\beta+e_{ij} \qquad (2.5)$$

जहाँ y_{ij} लघु क्षेत्र में i (i=1,---,m) इकाई $j^{th}(j=1,...,n_i)$ के लिए इच्छित चर का मान है। और x_{ij} सहायक चर के P $^{\mathrm{X1}}$ सदिश है, β समाश्रयण गुणांक का P $^{\mathrm{X1}}$ सदिश है। आमतौर पर त्रुटि शब्द e_{ij} को माध्य (mean) शून्य और विचरण σ^2 के साथ सामान्य रुप से वितरित मानते है। मॉडल (2.5) के अर्न्तगत, लघु क्षेत्र के लिए दो अप्रत्यक्ष अनुमानक हैः–

लघु क्षेत्र i में Y के माध्य के लिए समाश्रयण संश्लेषित अनुमानक को निम्न प्रकार से परिभाषित करते है।

$$\hat{\bar{Y}}_i^{SynREG} = \bar{y}_i + (\bar{X}_i - \bar{x}_i)'\hat{\beta} \qquad (2.6)$$

जहाँ β सम्पूर्ण प्रतिदर्श का आकलन है क्योंकि गणना में सम्पूर्ण क्षेत्र के आँकड़ों का उपयोग हुआ है और इस प्रकार यह प्रत्यक्ष समाश्रयण अनुमानक (2.3) से अलग है। उस क्षेत्र के लिए, जहाँ प्रतिदर्श आँकडें न हो, $\bar{Y}_i$ के लिए मॉडल–आधारित संश्लेषित अनुमानक को निम्न प्रकार परिभाषित करते है।

$$\hat{\bar{Y}}_i^{MSyn} = \bar{X}_i'\hat{\beta} \qquad (2.7)$$

अनुमानक (2.7) बहुत प्रभावशाली होगा जब लघु क्षेत्र i समाश्रयण गुणांक के लिए मजबूत अलग–2 प्रभाव नहीं दिखाता है।

2.2.1 मिश्रित/संयुक्त अनुमानक (Composite Estimators):

ज्यों ही एक लघु क्षेत्र में प्रतिदर्श आकार बढ़ता है, एक अप्रत्यक्ष अनुमानक, संश्लेषित अनुमानक की अपेक्षा अधिक वांछनीय हो जाता है। यह सच है प्रतिदर्श लघु क्षेत्रों के अनुमान के लिए तैयार किया गया था या नहीं। यह एक दूसरे पर चुनने की अपेक्षा एक वांछनीय विकल्प रूप में प्रत्यक्ष और संश्लेषित अनुमानक के एक मानित जोड़ के उपयोग को प्ररित करता है। इस मानित अनुमानक (weighted estimator) को मिश्रित अनुमानक कहते है। ये अनुमानक बहुत रुचि के है क्योंकि वे प्रत्यक्ष और संश्लेषित अनुमानकों के मानित संयोजन के माध्यम से लाभ और हानियों के बीच संतुलन (trade off) की अनुमति देते है। सामान्यतः लघु क्षेत्र i में Y की सम्पूर्ण समष्टि के लिए मिश्रित अनुमानक निम्न रूप में परिभाषित है।

$$\hat{T}_{y_i}^{c} = \phi_i \hat{T}_{y_i}^{d} + (1-\phi_i)\hat{T}_{y_i}^{syn} \qquad (2.6)$$

जहाँ लघु क्षेत्र i के लिए Y की सम्पूर्ण समष्टि के लिए $\hat{T}_{y_i}^{d}$ प्रत्यक्ष अनुमानक और $\hat{T}_{y_i}^{syn}$ संश्लेषित अनुमानक है, और $\phi_i\ (0 \le \phi_i \le 1)$ एक उपयुक्त चुना हुआ मान लिया है। अनुमानक (2.6), दो घटक अनुमानकों का एक मानित जोड़ (weighted sum) की माध्य वर्ग त्रुटि (MSE) दोनों घटक अनुमानकों से अपेक्षाकृत अधिक छोटी हो सकती है जब एक समुचित भराकन/मानाकन (Weighting) की योजना का प्रयोग किया जाता है। तथापि लघु क्षेत्र आकलन में इस्टतम भाराकन/मानाकन उत्पन्न करना सामान्यतः चुनौतीपूर्ण कार्य है क्योंकि ये अनुमानक इस्टतम भराकन/मानाकन कमजोर अनुमानों के प्रति आश्चर्यजनक रुप से संवेदनशील होते है। साहित्य में भार/मान चयन के कई तरीकों को प्रस्तावित किया गया है राव

(2003)। परम्परागत अप्रत्यक्ष अनुमानकों जैसे संश्लेषित और मिश्रित अनुमानक का लाभ यह है कि इनको आसानी से प्रयोग किया जा सकता है ।

3. मिश्रित मॉडल (Mixed models):

लघु क्षेत्र आकलन में स्पष्ट मॉडल का उपयोग रैखिक मिश्रित मॉडल का एक विशेष मामला है और लघु क्षेत्र आकलन की जटिल समस्याओं को प्रतिपादित और संचालित करने में बहुत लचीला है (राव (2003))। समूह त्रुटि समाश्रयण मॉडल (nested error regression model) (बेटसे, हारटर और फुलर (1988)) (Battese, Harter and Fuller, 1988), यादृच्छिक समाश्रयण गुणांक मॉडल (डेमप्स्टर, रूबिन और सुतकावा (1981)) (Dempster, Rubin and Tsutakawa, 1981) और साधारण यादृच्छिक प्रभाव मॉडल (फे और हेरिओत (1979)) (Fay and Herriot, 1979) पर आधारित लघु क्षेत्र आकलन के लिए अनेक विधियाँ को मिश्रित मॉडल के विशेष मामलों के रुप में साहित्य में प्रस्तावित किया गया है। उपलब्ध सहायक सूचना और उपयोग के स्तर के आधार पर, लघु क्षेत्र आकलन के लिए दो प्रकार के यादृच्छिक प्रभाव मॉडल साहित्य में वर्णित है।

(1) क्षेत्र स्तर मिश्रित प्रभाव मॉडल (Area Level Mixed Effect Model) या क्षेत्र स्तर मॉडल (Area Level Model) जिसमें क्षेत्र–विशिष्ट सहायक सूचना का उपयोग होता है और
(2) इकाई स्तर मिश्रित प्रभाव मॉडल (Unit Level mixed effect model) या इकाई स्तर मॉडल (Unit Level model) जिसमें इकाई स्तर सहायक सूचना का उपयोग होता है।

ये रैखिक मिश्रित मॉडल के विशेष उदाहरण है, आमतौर पर इन्हें क्षेत्र स्तर और इकाई स्तर लघु क्षेत्र मॉडल के रूप में निर्दिष्ट है।

3.1 क्षेत्र स्तर मिश्रित प्रभाव मॉडल (Area Level Mixed Effect Models):

बहुतायत रूप में प्रयुक्त होने वाले लघु क्षेत्र मॉडल का साहित्य में व्यापक रूप में वर्णित है। इनका उपयोग उन परिस्थितियों में किया जाता है जब सहायक चरो के लिए अलग–2 माप उपलब्ध नहीं है अथवा जब सहायक सूचना केवल क्षेत्र स्तर पर उपलब्ध है। मॉडल (फे और हेरिओत (1979)) (Fay और Herriot (1979)) निम्न रूप में परिभाषित है।

$$\tilde{\theta}_i = \theta_i + e_i \; ; \;\; \theta_i = x_i'\beta + u_i \qquad (3.1)$$

जहाँ $\tilde{\theta}_i$ प्रत्यक्ष प्रतिदर्श अनुमानक (उदाहरण के लिए, प्रतिदर्श माध्य $\bar{y}$) दर्शाता है, अतः इस मामले में e_i प्रतिचयन त्रुटि का प्रतिनिधित्व करता है, माना, माध्य शून्य और ज्ञात विचरण $Var_D(e_i) = \sigma_{Di}^2$ और x_i

क्षेत्र स्तरीय सूचना दर्शाता है। इस मॉडल के अन्तर्गत श्रेष्ठतम रैखिक अनभिनत आकलक ((Best Linear Unbiased Estimator/ Predictor) (BLUP)) निम्न है।

$$\hat{\theta}_i = \gamma_i \tilde{\theta}_i + (1-\gamma_i) x_i' \hat{\beta}_{GLS} = x_i' \hat{\beta}_{GLS} + \gamma_i (\tilde{\theta}_i - x_i' \hat{\beta}_{GLS}) \qquad (3.2)$$

जहाँ $\gamma_i = \sigma_u^2 / (\sigma_{Di}^2 + \sigma_u^2)$ है। व्यवहारतः, विचरण σ_4^2 और σ_{Di}^2 प्रायः अज्ञात होते है और वे आनुभविक–श्रेष्ठतम रैखिक अनभिनत आकलक (Best Linear Unbiased Estimator/Predictor) (BLUP)) (अथवा EBLUPs) के समतुल्य प्रतिदर्श अनुमान का स्थान लेते है। प्रसाद और राव (1990) में वर्णित विधि का उपयोग कर माध्य वर्ग त्रुटि (Mean Squared Error (MSE)) का अनुमान करते है।

3.2 इकाई स्तर मिश्रित प्रभाव मॉडल (Unit Level mixed effect model) (या इकाई स्तर मॉडल) (Unit Level model):

समूह त्रुटि इकाई स्तर समाश्रयण मॉडल (nested error unit level regression model) आम रुप से उपयोग होने वाला एक सरल मॉडल है, बेटसे आदि (1988) (Battese *et at.* 1988) ने ईओवा राज्य के 12 देशों में सोयाबीन और मक्का के क्षेत्रफल के पूर्वानुमान के लिए इसका मूल रुप से उपयोग किया था। मान लिजिए, प्रतिदर्श में प्रत्येक इकाई के सहायक चरो के मान ज्ञात है और इन चरो के सही क्षेत्र माध्य भी ज्ञात है। i क्षेत्र में J ईकाई के लिए सहायक मानों को $\mathrm{x_{ij}}$ द्वारा दर्शाते हुए, मॉडल का निम्न रूप है।

$$y_{ij} = x_{ij}'\beta + u_i + e_{ij} \qquad (3.3)$$

जहाँ $\mathrm{y_{ij}}$, $\mathrm{i(i=1....m)}, \mathrm{x_{ij}}$ क्षेत्र में ली गयी प्रतिदर्श (sampled) ईकाई $j^{th}(j=1,....,n_i)$ के लिए इच्छित/रुचि के चर के मान को दर्शाता है, x_{ij} इकाई स्तर सहायक चर का $_P\mathrm{x1}$ (vector) सदिश है, β अज्ञात स्थायी (fixed) प्रभाव का $_P\mathrm{x1}$ (vector) सदिश है, $\mathrm{n_i}$ क्षेत्र i में प्रतिदर्श इकाईयों की संख्या है, u_i विशिष्ट क्षेत्र यादृच्छिक प्रभाव है जोकि, माध्य शून्य और विचरण σ_{ij}^2 के साथ क्षेत्र i से सम्बन्धित और $\mathrm{e_{ij}}$ माध्य शून्य और विचरण σ_{ij}^2 के साथ अलग–2 स्तर यादृच्छिक त्रुटि है। ये दोनों त्रुटि परस्पर स्वतन्त्र है। मॉडल (3.3) गैर–प्रतिदर्श इकाईयों को भी और सम्पूर्ण समष्टि को भी नियन्त्रित रखता है। हम यह मान ले कि प्रतिदर्श एक विनिर्दिष्ट प्रतिचयन अभिकल्पना के अनुसार सम्पूर्ण लघु क्षेत्र से स्वतंत्र रूप से लिया गया है इसलिए प्रतिदर्श अभिकल्पना लघु क्षेत्र के भीतर ध्यान न देने योग्य है। क्षेत्र i में Y का समष्टि माध्य $\bar{Y}_i = \bar{X}_i'\beta + u_i + \bar{e}_i$ है माना जहाँ $\bar{X}_i = N_i^{-1}\sum_{j=1}^{N_i} x_j$ ज्ञात है। प्रर्याप्त रुप में अधिक $\mathrm{N_i}$ के लिए, $\bar{e}_i = N_i^{-1}\sum_{j=1}^{N_i} e_j \approx 0$ और तब क्षेत्र i में Y का माध्य $\mu_i = \bar{X}_i'\beta + u_i$ के निकट है। ज्ञात

$\theta = (\sigma_u^2, \sigma_e^2)$ के लिए, लघु क्षेत्र i के लिए Y का माध्य के लिए निम्न हेनडरसन (1975) (Hinderson (1975)) श्रेष्ठतम रैखिक अनभिनत आकलक (BLUP), $\bar{Y}_i$ है।

$$\hat{\mu}_i = \bar{X}_i'\hat{\beta} + \gamma_i(\bar{y}_i - \bar{x}_i'\hat{\beta}) = \gamma_i\left\{\bar{y}_i + (\bar{X}_i - \bar{x}_i)'\hat{\beta}\right\} + (1-\gamma_i)\bar{X}_i'\hat{\beta} \qquad (3.4)$$

जहाँ β का BLUE $\hat{\beta} = \left(\sum_i X_i' V_i^{-1} X_i\right)^{-1}\left(\sum_i X_i' V_i^{-1} Y_i\right)$ और $\gamma_i = \sigma_u^2\left(\sigma_u^2 + n_i^{-1}\sigma_e^2\right)^{-1}$ है। $\gamma_i\ (0 \le \gamma_i \le 1)$ मान को संकुचन कारक (shrinkage factor) कहते है। जो अभिकल्पना–अनभिनत समाश्रयण अनुमानक (approximately design-unbiased regression estimator) और संश्लेषित /कृत्रिम अनुमानक (synthetic estimator) के मध्य समंजन कराता है तथा कुल विचरण $\left(\sigma_u^2 + n_i^{-1}\sigma_e^2\right)$ से सम्बन्धी मॉडल विचरण σ_4^2 को मापता है। σ_i^2 के छोटे मान के लिए, मान γ_i छोटा होगा तथा परिणाम स्वरूप संश्लेषित भाग ज्यादा मान प्राप्त करता है और इसके विपरीत। $n_i = 10$ के लिए, अर्थात बिना प्रतिदर्श का क्षेत्र, $\gamma_i \to 0$ और $\hat{\mu}_i = \bar{X}_i'\hat{\beta}$।

इसके अतिरिक्त β और $\hat{\mu}_i$ विचरण घटकों θ पर आश्रित रहते है जो आव्यूह सह–प्रसरण (co-variance matrix) V_i को परिभाषित करता है। व्यवहार रुप में विचरण घटक अज्ञात होते है और प्रतिदर्श आँकड़ों से आकलन की मानक पद्धतियों जैसे (ANOVA), अधिकतम सम्भावना (Maximum Likelyhood (ML)) अथवा प्रतिबंधित अधिकतम सम्भावना (REML) का उपयोग करके अनुमानित/प्राक्कलित करते है। हम एक अनुमान को निर्दिष्ट करने के लिए ''हेट (hat)'' का उपयोग करते है और तब, एक दो चरण अनुमानक जिसे लघु क्षेत्र i के लिए Y के माध्य के आनुभविक–श्रेष्ठतम रैखिक अनभिनत आकलक (EBLUP) के रूप में जानते है जो निम्न प्रकार है.

$$\hat{\mu}_i = \bar{X}_i'\hat{\beta} + \hat{\gamma}_i(\bar{y}_i - \bar{x}_i'\hat{\beta}) \qquad (3.5)$$

जहाँ $\hat{\gamma}_i$ और $\hat{\beta}$ क्रमशः γ_i और $\tilde{\beta}$ के अनुमान है जिनको $\hat{\theta}$ को θ के स्थान पर रखकर प्राप्त किया है। प्रसाद और राव (1990) में वर्णित विधि का उपयोग कर माध्य वर्ग त्रुटि (Mean Squared Error (MSE)) का अनुमान करते है।

4. लघु क्षेत्र अनुमान के लिए मान/भार की विधि (Weighting Methods for SAE):

मॉडल (3.3) में माना था कि प्रतिदर्श एक विनिर्दिष्ट प्रतिचयन अभिकल्पना के अनुसार सम्पूर्ण लघु क्षेत्र से स्वतंत्र रूप से लिया गया है अतः प्रतिदर्श अभिकल्पना लघु क्षेत्र के भीतर ध्यान न देने योग्य है अथवा

वैकल्पिक रुप में चयन अभिनति (selection bias) नहीं है। आकलन ऐसे मॉडलों पर आधारित है जिनमें इकाई स्तर सर्वेक्षण मान का उपयोग नहीं होता और समरुपी अनुमानक अभिकल्पना संगत नहीं है जब तक कि लघु क्षेत्र में प्रतिदर्श अभिकल्पना स्वभराकित (self weighting) है। इसके विपरित, अभिकल्पना–आधारित प्रत्यक्ष अनुमानक अभिकल्पना संगत है लेकिन संबन्धित क्षेत्रों से मान लेने में असफल है। हाल के वर्षों में, साहित्य में मॉडल आधारित लघु क्षेत्र आकलन की गणना में सर्वेक्षण मानों के उपयोग की कुछ विधियाँ प्रस्तावित है। कोट्ट (1989) ((Kott (1989)) द्वारा एक अभिकल्पना संगत अनुमानक प्रस्तावित है, (3.3) की तरह यादृच्छिक त्रुटियों के समान कल्पना के साथ सरल यादृच्छिक प्रभाव मॉडल के अर्न्तगत मॉडल अनभिनत भी है। उसने सिद्ध किया था कि यह अनुमानक विशेष परिस्थितियों में मॉडल की विफलता के संदर्भ में सुदृढ़ (robust) है और यादृच्छिक प्रभाव घटक को शामिल किये बगैर माध्य वर्ग त्रुटि का एक अनुमानक उत्पन्न किया। आनुभविक परिणाम दर्शाता है कि माध्य वर्ग त्रुटि अनुमान काफी अस्थिर है और यहाँ तक कि नकारत्मक मान लेते हैं। इसके परिणामस्वरूप, इस पद्धति का उपयोग प्रस्ताविक अभिकल्पना संगत लघु क्षेत्र अनुमानक और पारंपरिक अभिकल्पना–आधारित प्रत्यक्ष अनुमानक की तुलना के लिए नहीं किया जा सकता। प्रसाद एवं राव (1999) और यू एवं राव (2002) ने लघु क्षेत्र अनुमान के लिए एक मॉडल सहायता प्राप्त अनुमानक (model assisted estimator) प्रस्तावित किया जिसे कृत्रिम/आभासी आनुभविक श्रेष्ठतम रैखिक अनभिनत आकलक (*pseudo* empirical best linear unbiased predictor) (*pseudo*-EBLUP) के नाम से पुकारते है जो सर्वेक्षण मान पर निर्भर करता है और लघु क्षेत्र में प्रतिदर्श आकार में वृद्धि होने पर अभिकल्पना संगत बना रहता है।

चन्द्रा और चम्बर्स (2005) ((Chandra and Chambers (2005)) ने लघु क्षेत्र आकलन में अंशशोधित मान आधारित पद्धति प्रस्तुत की और लघु क्षेत्र के लिए मॉडल–आधारित प्रत्यक्ष ((modal-based direct)(MBD)) अनुमानक के रुप में परिभाषित किया। इस पद्धति में माध्य वर्ग त्रुटि (MSE) की गणना के लिए मान रैखिक लघु क्षेत्र अनुमानक के साथ–साथ सरल अभिव्यक्ति को परिभाषित करने के लिए रैखिक मिश्रित मॉडल के समष्टि स्तर रूपान्तरण के अंर्तगत व्युत्पन्न अंशशोधित प्रतिदर्श मानो का उपयोग होता है इसके विपरित अभिकल्पना–आधारित प्रत्यक्ष अनुमानक, ये अनुमानक मान को परिभाषित करने में प्रयुक्त रैखिक मिश्रित मॉडल के द्वारा अन्य क्षेत्रों से मान लेते है। इस पद्धति के बहुत से व्यवहारिक लाभ है, इस तथ्य से उत्पन्न होता है कि इच्छित लघु क्षेत्र से वास्तविक प्रतिदर्श आंकड़ों के मान रैखिक संयोजनों के रुप में अनुमानक की गणना करते हैं। शायद, अनुमान प्रक्रिया तथा माध्य वर्ग त्रुटि अनुमान दोनों इनकी सबसे महत्वपूर्ण सरलता हैं। इसके अतिरिक्त, मॉडल–आधारित प्रत्यक्ष ((modal-based direct) (MBD)) अनुमानक की व्याख्या आसान है और एक सर्वेक्षण संसाधन प्रणाली का निर्माण करता है। मॉडल के गलत विनिर्देशों के मामले में, मॉडल–आधारित प्रत्यक्ष विधि आनुभविक श्रेष्ठतम रैखिक अनभिनत

आकलक (empirical best linear unbiased predictor) (EBLUP) या आभासी आनुभविक श्रेष्ठतम रैखिक अनभिनत आकलक (*pseudo* empirical best linear unbiased predictor) (*pseudo*-EBLUP) की तुलना में अधिक मजबूत लघु क्षेत्र अनुमान उपलब्ध करता है।

संदर्भ (References):

- कोकरन, डब्ल्यू. जी. (1977) – सैम्पलिंग टेक्निकसः जॉन वीले एण्ड सन्सः न्यूयॉर्क ।
- बैटसे, जी. ई., हार्टर, आर. एम. एवं फूल्लर, डब्ल्यू. ए. (1988) – एन एरर कम्पोनेन्ट मॉडल फॉर प्रेडिक्सन ऑफ काउन्टि क्रॉप एरियास यूसिंग सर्वे एन्ड सैटेलाईट डाटाः जर्नल ऑफ दि अमेरिकन स्टैटिसटिक्स एसोसिएसनः **83**: 28–36।
- चन्द्रा, एच. एण्ड चेम्बरस, आर. (2005) – कम्पेयरिंग ईबीएलयूपी एण्ड सी– ईबीएलयूपी फॉर स्मॉल एरिया एस्टिमेशनः स्टैटिसटिक्स इन ट्रान्सिशनः **7**: 637–648।
- डेम्पस्टर, ए. पी.; रूबिन, डी.बी. एण्ड सूटेकवा, आर. के. (1981) – एस्टिमेशन इन कोवैरियेन्स कॅम्पोनेन्टस मॉडलसः जरनल ऑफ दि अमेरिकन स्टैटिसटिकल एसोसिएसनः **76**: 341–353।
- फे, आर. ई. एवं हेरिअट, आर. ए. (1979) – एस्टिमेसन ऑफ इनकम फ्रॉम स्मॉल प्लेसेस–एन एप्लिकेशन ऑफ जेम्स–स्टेन प्रोसिजर्स टू सेन्सस डाटाः जर्नल ऑफ दि अमेरिकन स्टैटिस्टिक्स एसोसिएसनः 74: 269–277।
- फे, आर.ई. एण्ड हेरिओट, आर. (1979) – एस्टिमेटस ऑफ इनकम फॉर स्मॉल प्लेसेसः एन अप्लिकेशन ऑफ जेमस–स्टेन प्रोसिजरस टू सेन्सस डाटाः जर्नल ऑफ दि अमेरिकन स्टैटिसटिक्स एसोसिएसनः **74**: 269–277।
- गेनजालस, एम. ई. एण्ड होज़ा, सी. (1978) – स्मॉल एरिया एस्टिमेसन विद एप्लिकेशनस टू अनएमप्लोयमेन्ट एण्ड हाउसिंग एस्टिमेटसः जरनल ऑफ दि अमेरिकन स्टैटिस्टिकल एसोसिएसनः **73**: 7–15 ।
- हेन्डरसन सी. आर. (1975) – बेस्ट लिनियर अनबायसड एस्टिमेशन एण्ड प्रिडिक्सन अन्डर ए सलेक्सन मॉडलः बायोमेट्रिकसः **31**: 423–447।
- कोट, पी. (1989) – रोबस्ट स्मॉल डोमेन एस्टिमेसन यूजिंग रेण्डम ईफेक्टस मॉडललिंगः सर्वे मेथोडोलोजीः **15**: 1–12 ।
- प्रसाद, एन. जी. एन. एण्ड राव जे. एन. के. (1990) – दि एस्टिमेसन ऑफ दि मीन स्कोयर एरर ऑफ स्मॉल एरिया एस्टिमेटरसः जरनल ऑफ दि अमेरिकन स्टैटिसटिकल एसोसिएसनः **85**: 163–171 ।
- राव, जे.एन.के. (2003) – स्मॉल एरिया एस्टिमेशनः न्यू यार्कः विले ।
- सारनडल, सी. ई. स्वेनसन, बी. एण्ड रेटमैन, जे. (1992)–मॉडल एस्सिटेड सर्वे सैम्पलिंगः स्प्रीनजर–वरलगः न्यू यॉर्क ।
- यू, वाई एण्ड राव, जे. एन. के. (2002) – ए स्यूडो–एम्पेरिकल बेस्ट लिनियर अनबायसड प्रिडिक्सन एपरोच टू स्मॉल एरिया एस्टिमसन यूजिंग सर्वे वेटसः कैनाडियन जरनल ऑफ स्टैटिसटिक्सः **30**: 431–439 ।

अध्याय 15

मशीन लर्निंग के अनुप्रयोग का परिचय

पंकज दास

वैज्ञानिक, प्रतिदर्श सर्वेक्षण प्रभाग

भा.कृ.अनु.प.-भारतीय कृषि साँख्यिकी अनुसंधान संस्थान, लाइब्रेरी एवेन्यू, नई दिल्ली-110012

1. प्रस्तावना (Introduction):

प्रौद्योगिकी की विकास के साथ-साथ मानवता ने भी नए और उन्नत दिशाओं की ओर कदम बढ़ाए हैं। मशीन लर्निंग ऐसा ही एक क्षेत्र है जो तकनीकी उन्नति के साथ-साथ विज्ञान की दुनिया में भी एक महत्वपूर्ण परिवर्तन लाने का काम कर रहा है। यह विज्ञान का ऐसा अद्वितीय क्षेत्र है जो कंप्यूटरों को सिखने और स्वयं को सुधारने की क्षमता प्रदान करता है। मशीन लर्निंग का मतलब होता है कि कंप्यूटर सिस्टम को स्वयं से सिखने की क्षमता प्रदान की जाती है, जिससे वे डेटा पर आधारित निर्णय ले सकें और नई सूचनाओं को समझ सकें। इसका मुख्य लक्ष्य होता है कि सिस्टम स्वयं को सुधार सके जिससे कि वह कार्यों को और भी सही और तेजी से कर सके।

2. मशीन लर्निंग क्या है? (What is Machine Learning?):

"Machine Learning" शब्द Machine और learning शब्दों के मेल से बना है। Machine का मतलब है एक कंप्यूटर या किसी भी इलेक्ट्रॉनिक डिवाइस जिसे हम इस्तेमाल करते हैं, और Learning का मतलब है सीखना। Machine Learning का अर्थ होता है कि एक मशीन को सिखाया जाता है कि कैसे किसी विशिष्ट कार्य को पूरा करना है, और इसके लिए उसके algorithm के जरिए बार-बार सिखाया जाता है और वह algorithm खुद सीखता रहता है। यह होता है Machine Learning.

चित्र: मशीन लर्निंग और सामान्य प्रोग्रामिंग के बीच अंतर

मशीन लर्निंग एक विशेष प्रकार की कृत्रिम बुद्धिमता है जिसमें कंप्यूटर प्रोग्राम्स स्वतंत्र रूप से सिख सकते हैं और उन्हें अनुभवों से सिखने की क्षमता होती है। इसका उपयोग डेटा प्रसंस्करण, डेटा विश्लेषण, और पैटर्न पहचान के लिए किया जाता है। मशीन लर्निंग के अंतर्गत, अलग-अलग प्रकार की एल्गोरिदम्स और मॉडल्स का उपयोग होता है, जिन्हें सिस्टम को डेटा से सिखने में मदद मिलती है। इन मॉडल्स को ट्रेन किया जाता है ताकि वे डेटा के आधार पर निर्णय ले सकें। जैसे-जैसे सिस्टम को अधिक डेटा मिलता है, वह अपने आप में बेहतर होता जाता है और नए और बेहतर निर्णय लेने की क्षमता प्राप्त करता है।

मशीन लर्निंग एक गतिशील क्षेत्र है जिसमें विशेष विशेषताएँ हैं जो इसे जटिल समस्याओं के समाधान के लिए एक शक्तिशाली और बदलावपूर्ण उपकरण बनाती है। यहां कुछ मशीन लर्निंग की मुख्य विशेषताएँ हैं:

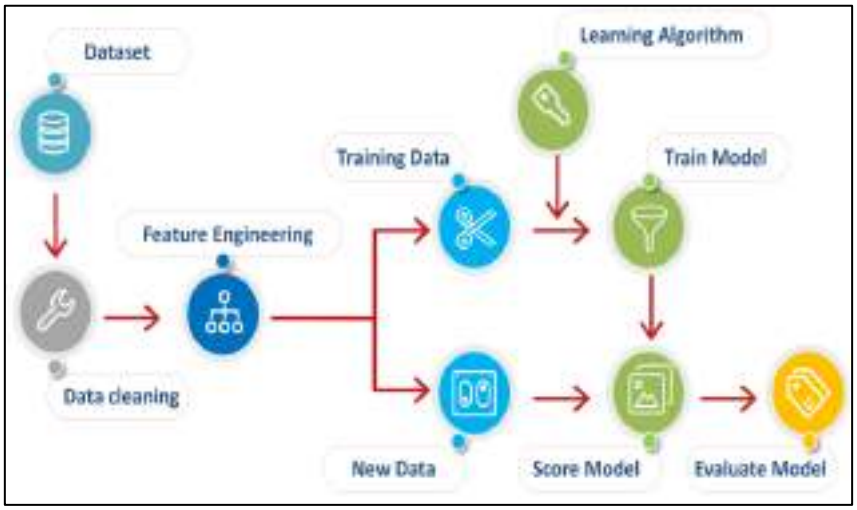
- **अनुकूलनशीलता (Adaptability):** मशीन लर्निंग की एक अद्वितीय विशेषता यह है कि यह समय के साथ अनुकूलित हो सकता है और सुधार सकता है। मशीन लर्निंग एल्गोरिथम अपने पैरामीटर्स को समायोजित कर सकते हैं और नई डेटा से सीख सकते हैं, जिससे वे जानकारी बढ़ते समय बेहतर पूर्वानुमान और निर्णय ले सकते हैं।
- **स्वचालन (Automation)** : मशीन लर्निंग कार्यों की स्वचालनीकरण की अनुमति देता है जो अन्यथा मैनुअल हस्तक्षेप की आवश्यकता होती। एल्गोरिथम डेटा से पैटर्न सीख सकते हैं और पूर्वनिर्धारण या निर्णय बिना स्पष्ट प्रोग्रामिंग के कर सकते हैं, जिससे कार्यक्षमता और उत्पादकता बढ़ती है।
- **पैटर्न पहचान (Pattern Recognition)** : मशीन लर्निंग एल्गोरिथम डेटा में पैटर्न की पहचान में उत्कृष्ट होते हैं, यहां तक कि जब ये पैटर्न मानवों के द्वारा नहीं पहचाए जा सकते हैं। यह सूक्ष्म संबंध और प्रवृत्तियों की पहचान की क्षमता मशीन लर्निंग की मौलिक मजबूती है।
- **सामान्यीकरण (Generalization)**: मशीन लर्निंग मॉडल्स प्रशिक्षण डेटा से नए, अदृश्य डेटा तक सामान्यीकरण करने के लिए डिज़ाइन किए गए हैं। वे प्रशिक्षण डेटा में मौजूद मूल पैटर्न और प्रवृत्तियों को पकड़ने का लक्ष्य रखते हैं ताकि वे पूर्व अदृश्य उदाहरणों पर सटीक पूर्वानुमान कर सकें।
- **पूर्वानुमान और पूर्वानुसंधान(Prediction and Forecasting)** : मशीन लर्निंग को अक्सर पूर्वानुमान और पूर्वानुसंधान कार्यों के लिए उपयोग किया जाता है। ऐतिहासिक डेटा से सीखकर, एल्गोरिथम भविष्य की घटनाओं के बारे में पूर्वानुमान बना सकते हैं, जैसे कि स्टॉक मूल्य, मौसम की स्थिति, या बीमारी के प्रकोप।

- **गैर-रेखांकन (Non-linearity):** कई मशीन लर्निंग एल्गोरिथम चरणों के बीच गैर-रेखांकन संबंध को पकड़ सकते हैं। इसका मतलब है कि वे जटिल इंटरैक्शन को पहचान और मॉडल कर सकते हैं जो पारंपरिक रैखिक तरीकों के माध्यम से स्पष्ट नहीं हो सकते।

3. मशीन लर्निंग कैसे काम करता है? (How Machine Learning works):

मशीन लर्निंग एक कृत्रिम बुद्धिमता की प्रक्रिया है जिसमें कंप्यूटर सिस्टम को स्वतंत्र रूप से सीखने और सुधारने की क्षमता दी जाती है। इस प्रक्रिया में, मॉडल नामक कंप्यूटर प्रोग्राम को डेटा से सिखने की जाती है ताकि वह नए और अजाने डेटा पर भी ठीक उत्तर दे सके। आइए देखें कि मशीन लर्निंग कैसे काम करता है:

- **डेटा संग्रहण (Data Storage)**: पहले, उस डेटासेट की आवश्यकता होती है जिस पर मॉडल को सिखाना है। यह डेटा संग्रहित किया जाता है जिसमें प्रशिक्षित और प्रशिक्षण डेटा शामिल होता है। प्रशिक्षित डेटा में सही उत्तर या लेबल भी शामिल होता है।
- **फीचर इंजिनीयरिंग (Feature Engineering):** डेटा को समझने के लिए, उसे विशेष विशेषताओं या फीचर्स में बदलना जाता है जिनसे मॉडल समझ सके कि कौनसे डेटा विशिष्ट पैटर्न को पहचानते हैं।
- **मॉडल चयन (Model Selection):** आवश्यकतानुसार, सही मॉडल का चयन किया जाता है जिसमें डेटा के साथ फीचर्स का संबंध बनाने के लिए अल्गोरिदम्स का उपयोग किया जाता है। कुछ मशीन लर्निंग मॉडल्स उपरक्रित डेटा पर पैटर्न पहचानते हैं और फिर नए डेटा पर उन पैटर्नों को लागू करते हैं।



चित्र: मशीन लर्निंग मॉडल कार्य प्रक्रिया

- **प्रशिक्षण या सिखाना (Training Data Set):** मॉडल को प्रशिक्षित करने के लिए प्रशिक्षण डेटा का उपयोग किया जाता है, जिसमें लेबल्ड डेटा के साथ सही उत्तर शामिल होता है। मॉडल डेटा से पैटर्नों और रिलेशनशिप्स को सीखता है ताकि वह नए डेटा पर सही नतीजे दे सके।
- **मॉडल की मान्यता प्राप्त करना (Model Validation):** मॉडल को प्रशिक्षित करने के बाद, उसे अनप्रशिक्षित डेटा पर टेस्ट किया जाता है ताकि उसकी प्रदर्शन क्षमता का माप किया जा सके। अगर मॉडल अपर्याप्त रूप से काम कर रहा है, तो उसे सुधारने की कोशिश की जाती है और फिर से प्रशिक्षण दिया जाता है।
- **नतीजों का उत्पन्न होना (Results):** प्रशिक्षित मॉडल को अनप्रशिक्षित डेटा पर टेस्ट करने के बाद, वह नए डेटा पर भी सही उत्तर प्रदान कर सकता है। जैसे-जैसे मॉडल प्रशिक्षित होता है, वह अपने द्वारा सीखे गए पैटर्न और रिलेशनशिप्स को अपने नजरिये से देखकर उत्तर प्रदान करता है।

मशीन लर्निंग के द्वारा कंप्यूटर सिस्टम स्वतंत्र रूप से सीखता है और अच्छे और निष्पक्ष नतीजे प्रदान करता है, जो विभिन्न क्षेत्रों में उपयोगी होते हैं।

4. मशीन लर्निंग के प्रकार (Types of Machine Learning):

मशीन लर्निंग विभिन्न प्रकारों में आता है, जो डेटा प्रोसेसिंग और पैटर्न पहचान के विभिन्न तरीकों का उपयोग करके काम करते हैं। यहाँ कुछ मुख्य मशीन लर्निंग के प्रकार हैं:

- **सुपरवाइज्ड लर्निंग (Supervised Learning):** इस प्रकार की लर्निंग में, एक मॉडल को प्रशिक्षित करने के लिए लेबल्ड डेटा का उपयोग किया जाता है। अर्थात्, हम मॉडल को डेटा दिखाते हैं जिसमें हर उदाहरण के साथ सही उत्तर या लेबल भी होता है। मॉडल को इस प्रकार से प्रशिक्षित किया जाता है कि वह नई डेटा पर सही उत्तर प्रदान कर सके। उदाहरण: डिजिट रिकॉग्निशन, बायोमेडिक सिग्नल प्रोसेसिंग।
- **अनुपामित लर्निंग (Unsupervised Learning):** इस प्रकार की लर्निंग में, मॉडल को लेबल नहीं दिया जाता है। यह मॉडल डेटा में पैटर्न और संबंधों को स्वतंत्र रूप से पहचानने का प्रयास करता है। क्लस्टरिंग और डिमेंशनलिटी रिडक्शन जैसे तकनीकों का उपयोग इस प्रकार की लर्निंग में होता है। उदाहरण: ग्रुपिंग, आवश्यकता की पहचान।
- **सेमी-सुपरवाइज्ड लर्निंग (Semi-Supervised Learning):** इस प्रकार की लर्निंग में, डेटा का केवल एक हिस्सा ही लेबल्ड होता है और बाकी डेटा अनलेबल्ड होता है। मॉडल को प्रशिक्षित करने के लिए लेबल्ड और अनलेबल्ड डेटा का उपयोग किया जाता है। इस प्रकार की लर्निंग में

मॉडल को स्वयं को सिखने की क्षमता मिलती है। उदाहरण: वाचन यातायात में गाड़ियों की पहचान।

- **रिइनफोर्समेंट लर्निंग (Reinforcement Learning):** इस प्रकार की लर्निंग में, मॉडल को किसी कार्रवाई के परिणामस्वरूप में प्राप्त इनसेंटिव्स (रिवॉर्ड्स) के आधार पर सिखाया जाता है। मॉडल का लक्ष्य सही क्रियाएँ करना होता है जो कि उसे मैक्सिमम रिवॉर्ड प्रदान करे। उदाहरण: गेम प्लेंग, रोबोटिक्स।

ये केवल कुछ मशीन लर्निंग के प्रकार हैं, बाकी भी विभिन्न प्रकार की तकनीकें हैं जो विभिन्न डेटा प्रोसेसिंग और पैटर्न पहचान के काम में आती हैं।

5. मशीन लर्निंग एल्गोरिथ्म के उदाहरण (Examples of Machine Learning Algorithm):

यहाँ कुछ मशीन लर्निंग एल्गोरिथ्म के उदाहरण हैं:

- **लीनियर रीग्रेशन (Linear regression):** यह एल्गोरिथ्म रीग्रेशन कार्यों के लिए उपयोग होता है, जहाँ लक्ष्य एक निरंतर मूल्य की पूर्वानुमान लगाना होता है। यह डेटा बिंदुओं को एक सीधी रेखा में फिट करता है जो सबसे अच्छी तरह से इनपुट फीचर्स और लक्षित चर के बीच संबंध को प्रतिनिधित करती है।
- **निर्णय पेड़ (Decision Tree):** निर्णय पेड़ वर्गीकरण और रीग्रेशन कार्यों के लिए प्रयुक्त होते हैं। वे डेटा को विभिन्न फीचर्स के आधार पर शाखाओं में विभाजित करके निर्णय लेते हैं। प्रत्येक शाखा एक निर्णय या परिणाम को प्रतिनिधित करती है, जो एक अंतिम पूर्वानुमान तक पहुँचता है।
- **रैंडम फ़ॉरेस्ट (Random forest):** रैंडम फ़ॉरेस्ट एक एन्सेम्बल लर्निंग एल्गोरिथ्म है जो अधिकांश निर्णय पेड़ों को मिलाकर अधिक सटीक पूर्वानुमान बनाने के लिए प्रयुक्त होता है। यह ओवरफ़िटिंग को कम करता है और व्यक्तिगत पेड़ों की पूर्वानुमानों की औसतनीकी को बेहतर बनाता है।
- **सपोर्ट वेक्टर मशीन (SVM):** SVM एक वर्गीकरण एल्गोरिथ्म है जो विभिन्न वर्गों को सर्वश्रेष्ठ तरीके से अलग करने वाली हाइपरप्लेन को खोजता है। यह विशेष तरह के सम्मिलन और गैर-सम्मिलन वर्गणित समस्याओं के लिए अच्छा काम करता है।
- **के-नियरेस्ट नेबर्स (KNN):** KNN एक सरल वर्गीकरण एल्गोरिथ्म है जो डेटा बिंदुओं को उनके k-नियरेस्ट नेबर्स के बहुमत वर्ग में वर्गीकृत करता है। यह इंस्टेंस-आधारित होता है और निष्पक्ष मॉडल प्रशिक्षण में शामिल नहीं होता है।

- **न्यूरल नेटवर्क्स (Neural network):** न्यूरल नेटवर्क्स मानव मस्तिष्क के संरचना और क्रियान्वयन से प्रेरित अल्गोरिथ्म परिवार हैं। इसमें एक संबंधित नोड (न्यूरॉन) की लेयरों की पट्टियों से मिली हुई होती है जो डेटा से जटिल पैटर्न सीखते हैं। डीप लर्निंग, न्यूरल नेटवर्क्स का एक उपयोग है, ने विभिन्न क्षेत्रों में महत्वपूर्ण उपलब्धियाँ प्राप्त की है।

6. कृषि में मशीन लर्निंग के अनुप्रयोग (Application of Machine Learning in Agriculture):

प्रौद्योगिकी की अद्भुत प्रगति ने मानवता की जीवनशैली को बदल दिया है, और इसका असर कृषि क्षेत्र में भी दिखाई दे रहा है। विज्ञान और प्रौद्योगिकी के साथ मिलकर, मशीन लर्निंग नामक तकनीकी उपकरण ने कृषि क्षेत्र में क्रांतिकारी बदलाव किया है। मशीन लर्निंग का अद्भुत उपयोग कृषि में समृद्धि की दिशा में कदम बढ़ा रहा है, जो कि इस लेख में हम देखेंगे।

समापन के रूप में, मशीन लर्निंग के अनुप्रयोग कृषि क्षेत्र में न केवल फसलों की उपज को बढ़ावा देने में मदद कर रहे हैं, बल्कि समृद्धि की दिशा में नए दरवाजे भी खोल रहे हैं। किसानों के जीवन को सुगम और आर्थिक रूप से सशक्त बनाने के लिए मशीन लर्निंग का यह उपयोग एक महत्वपूर्ण कदम है। कृषि में मशीन लर्निंग के ये उपयोग न केवल किसानों के लिए बल्कि खेती के सम्पूर्ण अनुभाग के लिए फायदेमंद हैं। यह तकनीक कृषि सेक्टर में नए और सुगम तरीकों को उपयोग करके समृद्धि और विकास की दिशा में कदम बढ़ा रही है।

मशीन लर्निंग के कुछ अनुप्रयोग क्षेत्र नीचे हैं-

- फसल उपज की भविष्यवाणी/ पूर्वानुमान: मशीन लर्निंग के उपयोग से किसान फसलों की उपज का पूर्वानुमान लगा सकते हैं। वे मौसम, मिट्टी की गुणवत्ता, पानी की आपूर्ति आदि के आधार पर फसल उपज को पूर्वानुमानित कर सकते हैं और इसके आधार पर उपयुक्त कदम उठा सकते हैं।
- रोग और कीट प्रबंधन: मशीन लर्निंग के द्वारा किसान फसलों में रोग और कीटों की पहचान कर सकते हैं। डेटा विश्लेषण के माध्यम से, संकेतों को पहचानकर उन्हें सही समय पर प्रबंधन करने की सलाह दी जा सकती है।
- स्वच्छ जल संसाधन: मशीन लर्निंग का उपयोग पानी की बचत और उपयोग में किया जा सकता है। सेंसर्स और डेटा विश्लेषण के माध्यम से, पौधों की आवश्यकता के आधार पर पानी का उपयोग किया जा सकता है, जिससे पानी की बर्बादी को कम किया जा सकता है।
- उत्पाद पुनर्निर्माण: मशीन लर्निंग के द्वारा उत्पादों के पुनर्निर्माण में मदद की जा सकती है। उत्पाद की गुणवत्ता को सुनिश्चित करने के लिए डेटा विश्लेषण और पैटर्न पहचान का उपयोग किया जा सकता है।
- जलवायु परिवर्तन की पहचान: मशीन लर्निंग के द्वारा जलवायु परिवर्तन की पहचान की जा सकती है। मौसम और जलवायु डेटा का विश्लेषण करके, किसान उन उपायों का अनुसरण कर सकते हैं जिनसे उनकी फसलों को जलवायु परिवर्तन की प्रतिक्रिया में सहायता मिल सके।
- ऑटोमेटेड खेती: मशीन लर्निंग और रोबोटिक्स के संयोजन से, ऑटोमेटेड खेती को संभावित किया जा सकता है। रोबोट्स और ड्रोन्स के सहायता से खेत में काम किया जा सकता है, जैसे कि बीज बोने, कीटनाशक छिड़कने, और सिंचाई करने का काम।
- बिजली व्यवस्था की पहचान: मशीन लर्निंग का उपयोग बिजली व्यवस्था की पहचान में भी किया जा सकता है। कृषि क्षेत्र में बिजली की आपूर्ति की गुणवत्ता की जांच करने के लिए डेटा विश्लेषण का उपयोग किया जा सकता है और इसके परिणामस्वरूप खेती में बेहतर बिजली उपयोग की जा सकती है।
- अनुशासन और संबंध प्रबंधन: मशीन लर्निंग के द्वारा किसानों के बेहतर अनुशासन और संबंध प्रबंधन की जा सकती है। डेटा विश्लेषण के माध्यम से, उन्हें किसान समूहों और बाजार की जानकारी मिल सकती है जो उनके उत्पादों की बिक्री में मदद कर सकती है।

- विपणन और उत्पादों की गुणवत्ता: मशीन लर्निंग के द्वारा उत्पादों की गुणवत्ता की निगरानी की जा सकती है। उत्पादों के आधार और उनकी गुणवत्ता के आधार पर, उनका विपणन किया जा सकता है और उन्हें बेहतर मूल्य पर बेचने में मदद मिल सकती है।

7. कृषि में मशीन लर्निंग अनुप्रयोग की चुनौतियाँ (Challenges of Machine Learning Application in Agriculture):

मशीन लर्निंग के अनुप्रयोग कृषि में अपनाने के बावजूद, इसके साथ ही कुछ चुनौतियाँ भी प्राकृतिक रूप से उत्पन्न होती हैं। कृषि क्षेत्र में मशीन लर्निंग के अनुप्रयोग करते समय निम्नलिखित चुनौतियों का सामना करना पड़ता है:

- डेटा की कमी: कृषि क्षेत्र में डेटा की कमी एक महत्वपूर्ण चुनौती हो सकती है। सही और पर्याप्त डेटा की आवश्यकता होती है ताकि मॉडल्स सही पूर्वानुमान कर सकें। कई बार किसानों के पास डिजिटल डेटा नहीं होता है और यह डेटा को संकलित और साफ़ रूप में प्राप्त करने की चुनौती पैदा कर सकता है।
- वैशिष्ट्यीकरण (Characterization): विभिन्न क्षेत्रों में कृषि उत्पादन की वैशिष्ट्यीकरण की आवश्यकता होती है, क्योंकि प्रत्येक क्षेत्र की आवश्यकताएँ और परिस्थितियाँ भिन्न होती हैं। मॉडल्स को विभिन्न जलवायु, मिट्टी प्रकृति, फसल प्रकार आदि के अनुसार तैयार करना चुनौतीपूर्ण हो सकता है।
- डेटा असंघटन: कृषि डेटा अक्सर असंघटित और अशुद्ध हो सकता है, जो अधिक संज्ञानात्मक और निष्कर्षक बनाने की क्षमता को प्रभावित कर सकता है। डेटा की यह असमानता मॉडल्स की प्रदर्शन क्षमता को भी प्रभावित कर सकती है।
- विशेषताओं की पहचान: कृषि डेटा में विशेषताएँ समय के साथ बदल सकती हैं, जैसे कि पौधों के विकास की अवधि आदि। मॉडल्स को इन बदलती विशेषताओं को पहचानने की क्षमता होनी चाहिए ताकि उनके पूर्वानुमान में सुधार किया जा सके।
- कृषि की विविधता: कृषि में विविध प्रकार की फसलें उगाई जाती हैं, और हर फसल की खासी विशेषताएँ होती हैं। फसलों की यह विविधता मॉडल विकसित करते समय एक सामान्य समस्या बन सकती है।

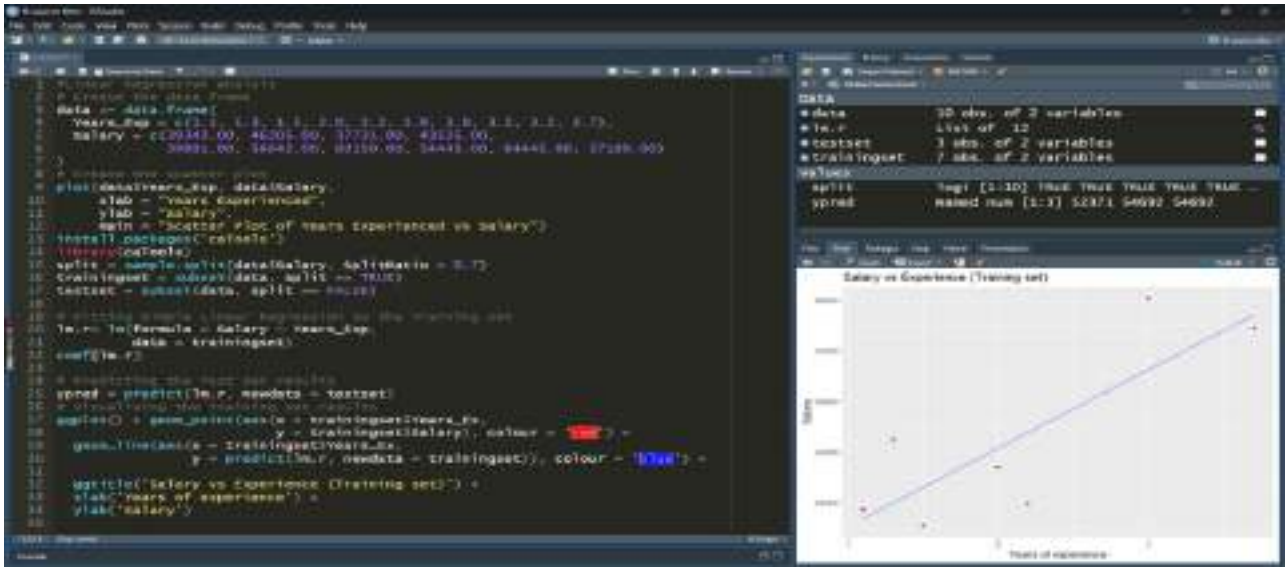
- अनुपस्थित डेटा: कई बार कृषि क्षेत्र में डेटा की अनुपस्थिति हो सकती है, जैसे कि फसलों की नियमित मॉनिटरिंग की असमर्थता या उपयुक्त डेटा संग्रहण की कमी। यह अप्रत्याशित परिणामों के कारण हो सकता है और मॉडल्स की आधारशिलता को कम कर सकता है।
- कृषि में मशीन लर्निंग के अनुप्रयोग को इन चुनौतियों का सामना करके उन्नति करने की कोशिश की जा रही है, ताकि किसानों को बेहतर उत्पादन और प्रबंधन के लिए अधिक समर्थन मिल सके।

8. **निष्कर्ष (Conclusion):**

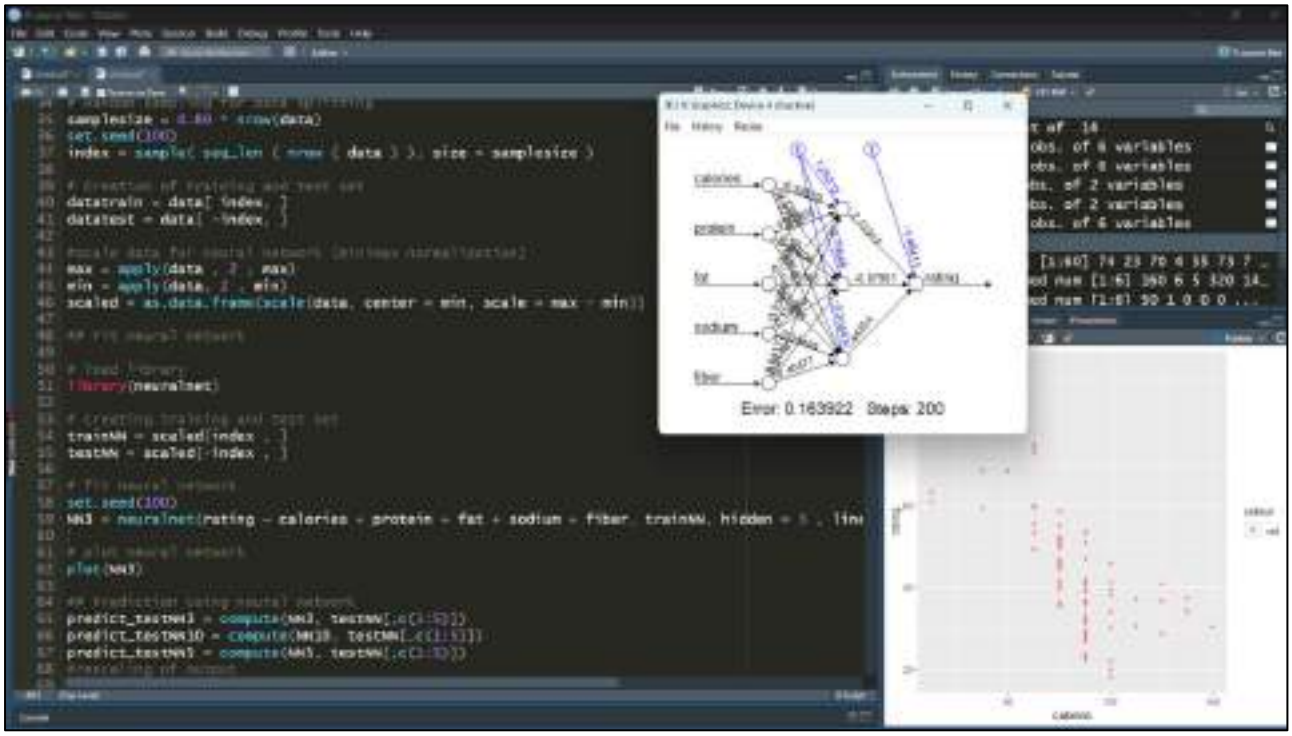
कृषि में मशीन लर्निंग का अनुप्रयोग वास्तव में एक योगदान है जो किसानों के लिए सूचना प्रौद्योगिकी के क्षेत्र में नया दौर लाने का प्रयास कर रहा है। यह तकनीकी उन्नति को बढ़ावा देने में महत्वपूर्ण भूमिका निभा रही है, जो उच्च उत्पादन, संसाधन सफलता, और कृषि उद्यमिता को बढ़ावा देने के लिए आवश्यक है। मशीन लर्निंग की विशेषताएँ और उनकी योग्यता कृषि क्षेत्र में कई समस्याओं का समाधान प्रदान करने में मदद कर रही हैं। विभिन्न फसलों के उत्पादन, फसलों के रोग और कीटों की पहचान, जलवायु पूर्वानुमान, और समाप्त उपयोगकर्ता की आवश्यकताओं की पूर्ति में मशीन लर्निंग का योगदान महत्वपूर्ण है। हालांकि इस क्रांतिकारी प्रौद्योगिकी का उपयोग कृषि में बढ़ते मात्र में हो रहा है, वह अपने साथ कुछ चुनौतियाँ भी लेकर आ रहा है।

डेटा की कमी, पैटर्न की पहचान में कठिनाइयाँ, तकनीकी संवाद की कठिनाइयाँ आदि कुछ चुनौतियाँ हैं जिन्हें हमें पार करना होगा। इसके बावजूद, मशीन लर्निंग कृषि में विकसिति की दिशा में एक बड़ी कदम है। यह न केवल किसानों के लिए बेहतर उत्पादन और प्रबंधन का माध्यम है, बल्कि यह खाद्य सुरक्षा को बढ़ावा देने और समृद्धि को समर्थन करने में भी महत्वपूर्ण योगदान प्रदान कर सकता है। कृषि में मशीन लर्निंग के अनुप्रयोग से उच्च गुणवत्ता वाले उत्पाद, सुरक्षित फसल प्रबंधन, और संभावित खतरों की पूर्वानुमानी में मदद मिल सकती है। आखिरकार, मशीन लर्निंग कृषि क्षेत्र को नई दिशाएँ देने का साधन है, जो किसानों की जीवनस्तर को उन्नत करने और खेती क्षेत्र को सामृद्धि प्राप्त करने में महत्वपूर्ण योगदान कर सकता है। इसके बदलते समय के साथ, मशीन लर्निंग का प्रयोग और अधिक व्यापक होगा और यह कृषि क्षेत्र में नए संभावनाओं की ओर प्रेरित करेगा।

R codes for linear regression



R codes for Artificial Neural Network



संदर्भ (Reference):

R विकास कोर टीम (2010): सांख्यिकीय गणना के लिए एक भाषा एवं वातावरण, संख्यिकीय गणना के लिए **R** फाउण्डेशन, वियाना, आस्ट्रिया, URL: http://www.R-project.org